Title: gbpC Gene Repertoire Variation among Mutans Streptococci.

Author(s): Kojima, Y; Okamoto-Shibayama, K; Sato, Y; Azuma, T


URL: http://hdl.handle.net/10130/2833
Original Article

**gbpC Gene Repertoire Variation among Mutans Streptococci**

Yuki Kojima*,**, Kazuko Okamoto-Shibayama*, Yutaka Sato* and Toshifumi Azuma*

*Department of Biochemistry and Oral Health Science Center, Tokyo Dental College, 1-2-2 Masago, Mihama-ku, Chiba 261-8502, Japan
**Dentistry and Oral Surgery, Asahi General Hospital, 1326, i, Asahi, Chiba 289-2511, Japan

Received 11 November, 2011/Accepted for publication 26 December, 2011

**Abstract**

The human dental caries pathogen *Streptococcus mutans* harbors one glucan-binding wall-anchored protein gene, *gbpC*, and another human pathogen, *Streptococcus sobrinus*, has 4 *gbpC* homologues. As no 4 *gbpC* homologues have so far been detected in other mutans streptococci, however, we cannot say whether those in *S. sobrinus* are paralogous or orthologous. Therefore, the purpose of this study was to identify and sequence the *gbpC/dbl* genes in *Streptococcus downei* and *Streptococcus criceti*. The findings revealed that the former harbored 5 *gbpC/dbl* homologues and the latter 4. In addition, another *dbl* homologue, the *dblC* gene, was found in some strains of *S. sobrinus*. We performed a phylogenetic analysis with the *gbpC/dbl* genes in these mutans streptococci. These results indicate that the *gbpC* and *dbl* genes have an orthologous relationship.

**Key words:** Mutans streptococci — *gbpC* — *dbl* — Glucan-binding wall-anchored proteins — Phylogenetic analysis

**Introduction**

Mutans streptococci (MS), major pathogens associated with dental caries not only in human, but also in monkey and rodent⁷, commonly conserve genes encoding major virulence factors involved in dental caries, including surface protein antigens (SpaP, PAc, antigen I/II) and glucosyltransferases.⁸ In addition to these two factors, *Streptococcus mutans* and *Streptococcus sobrinus* express other virulence-related factors in human, including the glucan-binding cell wall-anchored proteins encoded by the glucan-binding protein C (*gbpC*) and dextran-binding lectin (*dbl*) genes.⁹,¹² The *gbpC* gene was initially identified as the gene solely involved in dextran (α-1,6 glucan)-dependent aggregation of...
S. mutans. No homologous sequences were detected in the genome of this organism. In contrast, 4 gbpC gene homologues (gbpC1, gbpC2, dblA, and dblB) were detected in S. sobrinus, which can exhibit more active dextran-dependent aggregation than S. mutans. The monkey (Macaca fascicularis) and hamster microbial parasites Streptococcus downei and Streptococcus criceti are more closely related to S. sobrinus than S. mutans according to phylogenetic analysis with 16S ribosomal RNA sequences and base composition of deoxyribonucleic acid. Therefore, these two organisms are expected to harbor multiple gbpC gene homologues. The purpose of this study was to identify and sequence the gbpC/dbl genes in Streptococcus downei and Streptococcus criceti. Furthermore, we discuss the evolutionary development and significance of gbpC homologues in MS.

Materials and Methods

1. Bacterial strains

S. sobrinus strain OM55d (serotype d clinical isolate from our laboratories), S. downei strain Mfe28 (ATCC33748), and S. criceti HS6 (ATCC19642) were used in this study. The streptococci were maintained and cultured in Todd-Hewitt or Brain Heart Infusion broth/agar plates.

2. PCR amplification, nucleotide sequencing, and sequence analysis

Most of the regions corresponding to the gbpC1/gbpC2 and dblA/dblB genes in S. sobrinus strain OM55d and S. downei strain Mfe28 were amplified with primers previously used to identify these genes in S. sobrinus strains; some regions that could not be amplified with these primers were obtained by the PCR-based genome walking method (BD Biosciences Clontech, Palo Alto, CA) as described previously. The gbpC2 and dblA internal fragments of S. criceti HS6 were amplified with degenerate primers YF2 and YR2 as previously described and the flanking regions were amplified by the genome walking method as described above. The amplified fragments were purified and directly sequenced with the BigDye Terminator Cycle Sequencing FS Ready Reaction Kit and the ABI PRISM Genetic Analyzer 3130 Avant (Applied Biosystems, Foster City, CA). The sequence analyses were carried out with the DNASIS-Mac (Hitachi Software Engineering, Yokohama, Japan), GENETYX-MAC (Genetyx Corp., Tokyo, Japan), and ClustalW (DDBJ) programs. A phylogenetic tree was drawn with the TreeExplorer program (Ver. 2.12) based on the ClustalW results.

3. Nucleotide sequence accession numbers

The nucleotide sequence data reported in this communication will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases under accession numbers AB237533 (100-4 gbpC1/gbpC2), AB237534 (100-4 dblA/dblC), AB576862 (OM55d gbpC1/gbpC2), AB447999 (OM55d dblA/dblC), AB557647 (Mfe28 gbpC1/gbpC2), AB557646 (Mfe28 dblA/dblB), AB237528 (HS6 gbpC2), and AB557645 (HS6 dblA/dblB).

Results

1. Amplification and nucleotide sequences of gbpC/dbl gene homologues in S. sobrinus strain OM55d

The gbpC1, gbpC2, and dblA genes of S. sobrinus strain OM55d were amplified and sequenced with the primers used to identify these genes in strains 100-4 and 6715. In contrast, the dblB gene could not be amplified with these primers. Therefore, fragments downstream from the dblA gene were obtained by genome walking. A 4,026-bp open reading frame (ORF) sequence similar to those of the dblA or dblB genes but distinct from these two sequences was present 211 bp downstream from the dblA gene. This gene was 58.4% identical to the 6715 dblA gene over 3,621 bps and 58.5% identical to the 6715 dblB gene over 3,956 bps. Therefore, this gene was designated as the dblC gene. In addition, a highly
Variation among Mutans Streptococci

A conserved ORF sequence encoding a putative cation-transporting P-type ATPase, which was located 165 bp downstream from the termination codon of the \(\text{dblB}\) gene in strain 6715, started at 417 bp downstream from the termination codon of the \(\text{dblC}\) gene in OM55d. These data suggest that \(S.\ sobrinus\) strains harboring the \(\text{dblC}\) gene instead of the \(\text{dblB}\) gene may be present in natural \(S.\ sobrinus\) populations. To confirm this hypothesis, we performed PCR amplification with \(\text{dblB}\)- and \(\text{dblC}\)-specific primer sets in our \(S.\ sobrinus\) stock culture collection (OM-prefixed strains, except for OMZ65 in Fig. 1, were isolated from 3rd-year students of our dental college in 1978). The previously sequenced \(S.\ sobrinus\) \(\text{dblB}\) genes contained approximately 1.2-kb direct repeats near the 5′ region (5′DR). A \(\text{dblB}\)-specific primer set was designed to amplify these 5′DRs. The \(\text{dblB}\) gene was not amplified in 5 of 17 strains examined and the other 12 strains as well as previously reported strain 100-4 were \(\text{dblB}\)-positive (Fig. 1A). Furthermore, the \(\text{dblC}\) gene was amplified from the 5 \(\text{dblB}\)-negative strains with a PCR primer set designed for fragments corresponding to the extracellular domain of the OM55d

Fig. 1 Electrophoretic analyses of \(\text{dblB}\) and \(\text{dblC}\) PCR amplifications

(A) (D) \(\text{dblB}\) 5′ direct repeat regions amplified with primers, 5′-ACTAAATATGAAAGTTGCAAAGTCATT-3′ and 5′-GTGATAATGATTGGCTTGAGCT-3′. (B) extracellular domain-encoding regions of \(\text{dblC}\) gene amplified with primers, 5′-CAGTGGATGTCAGTCAATTCC-3′ and 5′-TCAAGTCTGCGGCAAGGTGTATT-3′. OM30d were negative controls. (C) extracellular domain-encoding regions of \(\text{dblB}\) gene amplified with primers, 5′-ACCTCGAGATGAAACAGGTAATTG-3′ and 5′-ATGAATTCGTCTGTGGCAGAGTTTTTTCAGATG-3′. \(\text{dblB}\) gene fragments were not amplified from \(\text{dblC}\)-positive strains in contrast to OM50d as positive control. (D) UA-, CH-, and SW-prefixed strains represent USA, China, and Sweden strains, respectively. Asterisks in (A)–(C) represent \(\text{dblA}\)/\(\text{dblC}\)-harboring strains.
DblC protein without the signal sequences (Fig. 1B). The results suggest that the 5 strains including OM55d contained the dblC gene instead of the dblB gene. We have previously obtained chromosomal DNA samples isolated from USA-, China-, and Sweden-S. sobrinus strains (3 strains each) by Dr. Caufield at New York University. All of these were dblB-positive (Fig. 1C) and dblC-negative strains.

Gene arrangements concerning the gbpC1, gbpC2, dblA, and dblB/dblC genes in S. sobrinus are summarized in Table 1.

### Table 1  gbpC/dbl gene arrangements among mutans streptococci

<table>
<thead>
<tr>
<th></th>
<th>gbpC genes</th>
<th></th>
<th>dbl genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. mutans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109cS</td>
<td></td>
<td>583 aa</td>
<td></td>
</tr>
<tr>
<td>S. sobrinus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6715</td>
<td>gbpC1</td>
<td>621 aa</td>
<td>1118 aa</td>
</tr>
<tr>
<td></td>
<td>gbpC2</td>
<td>632 aa</td>
<td>1425 aa</td>
</tr>
<tr>
<td>S. sobrinus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100-4</td>
<td>gbpC1</td>
<td>621 aa</td>
<td>1270 aa</td>
</tr>
<tr>
<td></td>
<td>gbpC2</td>
<td>638 aa</td>
<td>1425 aa</td>
</tr>
<tr>
<td>S. sobrinus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OM55d</td>
<td>gbpC1</td>
<td>621 aa</td>
<td>1264 aa</td>
</tr>
<tr>
<td></td>
<td>gbpC2</td>
<td>638 aa</td>
<td>1341 aa</td>
</tr>
<tr>
<td>S. downei</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mfe28</td>
<td>gbpC1</td>
<td>636 aa</td>
<td>1270 aa</td>
</tr>
<tr>
<td></td>
<td>gbpC3</td>
<td>602 aa</td>
<td>1569 aa</td>
</tr>
<tr>
<td></td>
<td>gbpC2</td>
<td>639 aa</td>
<td></td>
</tr>
<tr>
<td>S. criceti</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS6</td>
<td>gbpC1</td>
<td>615 aa</td>
<td>1093 aa</td>
</tr>
<tr>
<td></td>
<td>ORF1</td>
<td>204 aa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gbpC2</td>
<td>637 aa</td>
<td>1717 aa</td>
</tr>
</tbody>
</table>

Shaded lines represent nucleotide sequence regions similar to gbpC3 3’ sequence region of S. downei.

DbfC protein without the signal sequences (Fig. 1B). The results suggest that the 5 strains including OM55d contained the dbfC gene instead of the dbfB gene. We have previously obtained chromosomal DNA samples isolated from USA-, China-, and Sweden-S. sobrinus strains (3 strains each) by Dr. Caufield at New York University. All of these were dbfB-positive (Fig. 1C) and dbfC-negative strains. Gene arrangements concerning the gbpC1, gbpC2, dbfA, and dbfB/dbfC genes in S. sobrinus are summarized in Table 1.

### 2. Amplification and nucleotide sequences of gbpC/dbl gene homologues in S. downei strain Mfe28

Regions corresponding to the gbpC1, dbfA, and dbfB genes of S. downei strain Mfe28 were amplified and sequenced with the primers used to determine these gene regions of S. sobrinus, except for sequencing several minor regions in the gbpC1 gene, for which we designed new primers specific for S. downei nucleotide sequences. Nucleotide sequence identities of the gbpC1, dbfA, and dbfB genes between S. downei and S. sobrinus were respectively 91.2, 98.7, and 96.9%. Although the nucleotide sequence identity of the gbpC1 gene between these two species was slightly lower than those for the dbfA and dbfB genes, the identity of a ~1,500-bp upstream region of the gbpC1 gene was 97.1%, which corresponds to those for the dbfA and dbfB genes. A region downstream from the divergent 3’ portion of the gbpC1 gene was therefore obtained by PCR-based genome walking and nucleotide sequencing revealed two other ORFs similar to the gbpC2 gene. The second ORF was more similar (67% identity) to the gbpC2 gene of S. sobrinus than the first ORF (54.5% identity). Therefore, these two ORFs were respectively designated as the gbpC3 and gbpC2 genes. Gene arrangements of the gbpC/dbl gene homologues in S. downei strain Mfe28 are also depicted in Table 1.

### 3. Amplification and nucleotide sequences of gbpC/dbl gene homologues in S. criceti strain HS6

S. criceti exhibits significant dextran-depen-
dent aggregation similar to *S. sobrinus* and *S. downei*. Therefore, we presumed that *S. criceti* might have gbpC/dbl gene homologues similar to the latter organisms. Amplification with a primer set (YF2 and YR2) with which the *S. sobrinus* gbpC partial fragments could be amplified revealed approximately 0.3, 0.7, and 4-kb bands. The 0.3 and 0.7-kb fragments were excised from the gel, purified, and directly sequenced. The 0.3 and 0.7-kb sequences were highly similar to the corresponding sequences within the *S. sobri-

**Table 2**  Phylogenetic trees of gbpC/dbl genes (A) and 16S ribosomal RNA genes (B)

Phylogenetic analysis was performed using DDBJ ClustalW program with default setting, except for Bootstrap-On. Trees were drawn with TreeExplore program (http://evolgen. biol.metro-u.ac.jp/TE/). Bars represent genetic distances.

nus* gbpC2 and dblA genes, respectively. The entire gbpC2 and dblA-dblB gene sequences in *S. criceti* were determined following PCR-based genome walking (Kojima, Y., IADR abstract, 2009, http://iadr.confex.com/iadr/2009miami/webprogram/Paper115741.html). The gbpC1 gene was not detected 1.4 kb upstream or 1.7 kb downstream from the gbpC2 gene on the *S. criceti* chromosome. However, the *S. criceti* genome project (Project ID: 51461) has recently been in progress and nucleotide sequence data are now available.
with the BLAST homology search. When the *S. sobrinus* gbpC1 gene was searched against this database, a similar sequence was detected 1,790 bp upstream from the initiation codon of the gbpC2 gene. Within this intergenic 1,790 bp region, one large open leading frame (ORF1 depicted in Table 1) encoding a 204-amino acid protein was detected. However, this amino acid sequence was not similar to any part of the GbpC protein. The dblB gene was located 356 bp downstream from the dblA termination codon on the same strand and encoded the largest DblB protein among these three species. Gene arrangements of the gbpC/dbl gene homologues in *S. criceti* strain HS6 are also depicted in Table 1.

4. Phyllogenetic analysis of GbpC/Db protein homologues

Sequence similarities of GbpC/Db protein homologues between two MS species were higher than those between homologues within a species. For example, the similarities of DblA (71.4%) or DblB (80.2%) between *S. sobrinus* and *S. criceti* were higher than similarity between DblA and DblB within *S. sobrinus* (46.1%) or *S. criceti* (45.6%). Phylogenetic analysis was performed with 21 gbpC/dbl gene homologue sequences together with 16S ribosomal RNA genes of MS using the DDBJ ClustalW program. The unrooted phylogenetic tree constructed for these genes is shown in Table 2. The gbpC1, gbpC2, dblA, or dblB gene was classified in a different monophyletic cluster, irrespective of MS species. In addition, the branching patterns of the gbpC1 and dblB genes were very similar to that of the 16S rRNA genes. These branch shapes were similar to that of the glucosyltransferase gene gtfI from these species\(^1\). These results suggest that the divergence of gbpC/dbl genes occurred in a common ancestral species of these MS. A node with *1 indicates divergence of the gbpC1 and gbpC/dblC genes, and a node *2 indicates divergence of the dblA/dblB genes. In contrast, node *3 represents divergence of *S. criceti* and other MS species.

5. Multiple alignment of gbpC1-gbpC2 intergenic regions

The entire gbpC gene regions from *S. sobrinus*, *S. criceti*, and *S. downei* were analyzed by the ClustalW program at DDBJ. In addition to the highly conserved alignment of the gbpC1 and gbpC2 gene regions of these three organisms, a 3' region of the *S. downei* gbpC3 gene was aligned with several stretches of intergenic regions between the gbpC1 and gbpC2 genes from both *S. sobrinus* and *S. criceti*, as indicated in Table 3. The regions between the gbpC1 termination codons and the gbpC2 initiation codons in *S. sobrinus*, *S. criceti* and *S. downei* were respectively 686, 1,790, and 3,159 bp (in which the gbpC3 gene was located). These results suggest that a portion of the gbpC3 gene region harbored by an ancestral species of these three species may have been deleted from the chromosomes of *S. sobrinus* and *S. criceti* following species divergence, in contrast to *S. downei*.

Discussion

We previously identified the tandemly located gbpC1-gbpC2 and dblA-dblB genes in *S. sobrinus*\(^\text{10,11}\). Paralogous genes are often organized in a tandem localization following unequal crossover recombination with highly similar sequences between the two daughter DNAs during replication in a species. Orthologous genes are those in different species that are derived from a common ancestor (http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/Orthology.html). However, we could not determine whether these two homologous gene pairs were paralogous in this study, as no similar genes were detected in multiple species related to *S. sobrinus* using the BLAST program of the NCBI against the international DNA database (EMBL/GenBank/DDBJ). Therefore, we initially intended to identify gbpC/dbl genes in *S. criceti* and *S. downei* as related species. Phylogenetic analysis indicated that the gbpC1, gbpC2, dblA and dblB genes from *S. sobrinus*, *S. criceti* and *S. downei* are positioned in clusters. Therefore, these
gbpC/dbl genes from these three species appeared to be orthologous and derived from a common ancestor. In addition, it is of interest that *S. downei* possessed three *gbpC* genes distinct from those in two other species. Multiple alignment of *gbpC1-gbpC2* intergenic regions analyzed by the ClustalW program at DDBJ suggested that an ancestor species of these three species harbored the *gbpC3* gene rather than *S. downei* acquiring the *gbpC3* gene from another unknown species by horizontal transfer.

It is also of interest that *S. sobrinus* populations can be classified into two groups relative to *dbl* gene organization. One harbors the *dblA-dblB* tandem genes and the other has the *dblA-dblC* tandem genes. Downstream regions from the *dblB* and *dblC* gene regions are conserved between the two groups of *S. sobrinus* strains as described above. These results suggest that an ancestor species of these three species might also have possessed a *dblA-dblB-dblC* cluster similar to the *gbpC1-gbpC3-gbpC2* genes. The ratio of *dblA-dblC* harboring strains to *dblA-dblB* harboring strains was 5 to 12 in our stock culture collection and strain SL1.

---

**Table 3**  
Table 3. Multiple alignment of *gbpC3′* region with regions upstream from *gbpC2* genes in *S. criceti* and *S. sobrinus*

<table>
<thead>
<tr>
<th><em>gbpC3′ region</em></th>
<th><em>S. sobrinus</em></th>
<th><em>S. criceti</em></th>
<th><em>S. downei</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>gbpC1</td>
<td>5′-AGGACCTTTAATGACAGCAGATGAGA-3′</td>
<td>5′-TCTGCCAGATGAGA-3′</td>
<td>5′-TTGACAGATGAGA-3′</td>
</tr>
<tr>
<td>gbpC2</td>
<td>5′-TTTACAGTTCTCGGATATTGAGA-3′</td>
<td>5′-TCTGCCAGATGAGA-3′</td>
<td>5′-TTTACAGTTCTCGGATATTGAGA-3′</td>
</tr>
<tr>
<td>gbpC3</td>
<td>5′-AGAGAAATGACAGCAGATGAGA-3′</td>
<td>5′-TCTGCCAGATGAGA-3′</td>
<td>5′-TTTACAGTTCTCGGATATTGAGA-3′</td>
</tr>
</tbody>
</table>

Regions between *gbpC1* termination codons and *gbpC2* initiation codons in *S. sobrinus*, and *S. downei* were respectively 686, 1,790, and 3,159 bp (in which *gbpC3* gene was located), and are respectively numbered from nucleotides immediately following *gbpC1* termination codons. Aligned *gbpC1/gbpC2* intergenic regions indicated in this figure are: 215–479 in *S. sobrinus*, 1,221–1,516 in *S. criceti*, and 2,286–2,704 (*gbpC3′*; 2,286–2,539) in *S. downei*. Box represents termination codon of *S. downei* *gbpC.*
was the sole \( \text{dblA/dblC} \) harboring strain among the reference strains (OMZ65, 6715, B13) and each of the three China, USA, and Sweden strains. Therefore, geographically biased distribution of the two groups may be possible, although sample numbers may be too small.

Recently, new organisms classified as MS have been reported in other animals, including pig, bat, and wild boar\(^{14,15}\), although their pathogenicity in their hosts remains to be fully characterized. It will be of interest to determine which types of \( \text{gbpC/dbl variation} \) exist in these species.

**Acknowledgements**

We would like to thank P.W. Caufield at New York University for providing chromosomal DNA samples isolated from USA, China, and Sweden \( S.\text{sobrinus} \) strains and H.K. Kuramitsu (State University of New York at Buffalo, NY) for critical review of the manuscript. We would also like to thank Associate Professor Jeremy Williams, Tokyo Dental College, for his assistance with the English of the manuscript.

**References**


*Reprint requests to:*  
Dr. Yutaka Sato  
Department of Biochemistry and  
Oral Health Science Center,  
Tokyo Dental College,  
1-2-2 Masago, Mihama-ku,  
Chiba 261-8502, Japan  
E-mail: yusato@tdc.ac.jp