Title
Evolution of Cariogenic Character in Streptococcus mutans: Horizontal Transmission of Glycosyl Hydrolase Family 70 Genes

Author(s)
Hoshino, Tomonori; Fujiwara, Taku; Kawabata, Shigetada

Citation
Scientific Reports, 2, 00518; 2012

Issue Date
2012-07-18

URL
http://hdl.handle.net/10069/30085

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/

NAOSITE: Nagasaki University’s Academic Output SITE
http://naosite.lb.nagasaki-u.ac.jp
Evolution of Cariogenic Character in *Streptococcus mutans*: Horizontal Transmission of Glycosyl Hydrolase Family 70 Genes

Tomonori Hoshino¹, Taku Fujiwara¹ & Shigetada Kawabata²

¹Department of Pediatric Dentistry, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8588, Japan, ²Department of Oral and Molecular Microbiology, Osaka University Graduate School of Dentistry, 1–8 Yamadaoka, Suita, Osaka 565-0871, Japan.

Acquisition of the ability to produce polysaccharides from sucrose, i.e. the *gtf* gene encoding glucosyltransferase (GTF), is the key evolutionary event enabling dental biofilm formation by streptococci. To clarify the ancestry of streptococcal GTFs, time of its occurrence, and order of specific events, we investigated the distribution of GTFs among bacteria by phylogenetic analysis of the glycosyl hydrolase family 70 enzymes. We found that streptococcal GTFs were derived from other lactic acid bacteria such as *Lactobacillus* and *Leuconostoc*, and propose the following evolutionary model: horizontal gene transfer via transposons occurred when streptococci encountered lactic acid bacteria contained in fermented food. Intra-genomic gene duplication occurred by a secondary selection pressure such as consumption of refined sugar. Our findings concerning this evolution in *Streptococcus mutans* provide an important background for studies of the relationship between the historical spread of dental caries and anthropological factors.
horizontal gene transfer. Through the acquisition of GTFs, Str. mutans became capable of forming cariogenic dental biofilms. Thus, our data support the idea that the pandemic of dental caries is likely to have been caused by not only anthropological factors but also the evolution of Str. mutans.

Results
Phylogenetic analyses of glycosyl hydrolase family 70 enzymes. To determine the origin of genes contributing to the cariogenic potential of Str. mutans, 3 phylogenetic trees of glycosyl hydrolase 70 family enzymes (Supplementary Table S1) were constructed using the NJ, the ME, and the MP methods. The tree generated using the MP method is shown in Figure 1, and the other 2 are presented in supplemental information. The 3 resultant trees showed complete congruence. Streptococcal GTF enzymes exhibit homology with dextransucrases from Leuconostoc and GTFs from Lactobacillus and Lactococcus, suggesting a common ancestry for the genes encoding these enzymes.
Mapping of the genomic location of glycosyl hydrolase family 70 genes. The upstream and downstream regions of gtf genes from 4 species of bacteria are shown in Figure 2A. Transposase sequences were observed in the upstream and/or downstream regions of gtf from Lactobacillus reuteri and gtfKg from Lactobacillus sakei. The downstream of gtfR from Streptococcus oralis also exhibited sequence homology to streptococcal transposase genes (Fig. 2B). The identity between this region and the transposase genes from Str. pneumoniae, Str. suis, Str. gordonii, and Str. mitis, indicated 80, 80, 78, and 76%, respectively. However, sequences homologous to transposase genes were not detected downstream or upstream of gtfB, gtfC, and gtfD in Str. mutans.
Phylogenetic analysis of the catalytic domain in streptococcal GTFs. To infer age associated with ecological events, the mean interpopulational evolutionary diversity in the sequences encoding the catalytic domain of glycosyl hydrolase family 70 genes among Streptococcus, Leuconostoc and Lactobacillus was estimated as 0.043% (S.E.; 0.003). Further, to identify the most evolved enzyme among the streptococcal GTFs, we reconstructed the phylogenetic tree with only the streptococcal GTF catalytic domains and their coding sequences using the NJ method. The catalytic domain of Lcb. reuteri GTFB and its coding sequences were used as a root (Fig. 3A and B). The phylogenetic distances based on the catalytic domain DNA sequences are shown in Table 1. The distances between the root and each streptococcal gtf gene group were found to be almost identical. In addition, the distance between the root and gtfB was found to be the same as the distance between the root and gtfD. In contrast, the distance between the root and gtfC was slightly larger in comparison.

Further analysis of the streptococcal GTF catalytic domains using a BLAST search showed that these regions are homologous to...
Table 1 | Mean phylogenetic distances (lower triangle) and standard error (upper triangle) of the catalytic regions among Lcb. reuteri GTFB (LR GTFB), streptococcal GTFs, and 3 gtf genes from *Str. mutans*

<table>
<thead>
<tr>
<th></th>
<th>root</th>
<th>WSG</th>
<th>WIG</th>
<th>INT</th>
<th>gtfB</th>
<th>gtfC</th>
<th>gtfD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR GTFB</td>
<td>0.020</td>
<td>0.021</td>
<td>0.019</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>WSG</td>
<td>0.558</td>
<td>0.015</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIG</td>
<td>0.555</td>
<td>0.366</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>0.571</td>
<td>0.393</td>
<td>0.381</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gtfB</td>
<td>0.539</td>
<td></td>
<td>0.017</td>
<td>0.021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gtfC</td>
<td>0.560</td>
<td>0.144</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gtfD</td>
<td>0.537</td>
<td>0.397</td>
<td>0.338</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: WSG, water-soluble glucon-synthesising group; WIG, water-insoluble glucon synthesizing group; INT, intermediate group (INT).

α-amylase, but not to sucrase. In the blastp analysis of the catalytic domain of *Str. mutans* GTFB, the E-value for the α-amylase catalytic domain was 8.25e-03. This finding suggests that the catalytic regions of streptococcal GTFs are not phylogenetically related to sucrase, despite having the same function.

### Phylogenetic analysis of the glucan-binding domain in streptococcal GTFs

Since the glucan-binding domains in streptococcal GTFs are closely associated with cariogenicity, phylogenetic analyses of their coding genes and putative amino acid sequences of the first unit of the glucan-binding domains constructed of 6 direct repeating units were also performed. The phylogenetic trees were constructed using the NJ method, and the glucan-binding domains of *L. reuteri* GTFB were used as a root (Fig. 3C and D). The trees based on both the coding genes and the putative amino acid sequences could not be classified into 3 clusters as shown in Figure 3A and B. Although the periodicity of clustering was not observed in the tree based on the genes, the tree based on the putative amino acid sequence indicated that GTFs from the same or closely related species could be divided into the same cluster.

In the SYSTERS Protein Family Database, the glucan-binding regions were categorized as Pfam CW_binding_1 (PF01473), which was a cell wall binding protein family that included glucan-binding protein (Gbp), choline-binding protein, 1,4-β-N-acetylmuramidase, and N-acetylmuramoyl-1-alanine amidase (http://systers.molgen.mpg.de/cgi-bin/nph-fetchcluster.pl?PFAM=CW_binding_1).

### Discussion

Phylogenetic analyses of glycosyl hydrolase family 70 enzymes were performed. Glucosyltransferase-S enzymes from *Lactococcus lactis*, encoded by pspA or pspB, have the simplest structures and were consequently defined as the roots of the dendrograms. However, these enzymes do not actually belong to the glycosyl hydrolase family 70 enzymes, since they do not possess the catalytic and glucan-binding domains characteristic of this enzyme family shown in the alignment analysis. The constructed trees indicated that the streptococcal GTFs were derived from other lactic acid bacteria following their spread through the genera in the order of *Lactococcus, Lactobacillus, Leuconostoc*, and *Streptococcus*. More specifically, our data suggests that *Lcb. reuteri* GTFB or *Lcb. reuteri* GTFML4 would be the practical ancestor of the streptococcal GTFs. Our previous finding that only 51% of 41 *Str. oralis* clinical isolates possessed the gtf gene supports the idea that some of the *Streptococcus* spp. are in the process of acquired the gtf gene. The average rate of sequence divergence at synonymous sites as determined by a comparison of homologous protein-coding regions, is 0.90% per million year. Given that the interpopulational divergence in the sequences encoding the catalytic domain of GTFs is 0.043%, their expansion would have occurred approximately 48,000 years ago. Thus, *Streptococcus* acquired the gene only recently in bacterial evolutionary time.

The acquisition of a gene is caused as the result of adaptation to habitat. During this process, it is thought that some parts of the acquired gene are modified to adapt to the species-specific circumstances that the other parts are conserved to keep the primary function. For this reason, it is thought that the diversity of the conserved region reflects the systematic evolution of the acquired gene without species-specific modification. The catalytic regions of GTFs are highly homologous among streptococcal species and are highly conserved to keep sucrase activity. Thus, to observe the systematic evolution of streptococcal gtf genes, the phylogenetic analyses of catalytic region were carried out (Fig. 3A and B). Our results show that each phylogenetic distance between the root and the 3 streptococcal gtf gene groups (WIG, WSG, and INT) was almost identical, suggesting that these groups differentiated within the same period. This finding is in contrast with the commonly held view by dental researchers that GTFB and GTFc are the most advanced *S. mutans* GTF enzymes, since they primarily synthesize the water-insoluble α-1,3-linked glucans directly associated with cariogenicity.

Here, we have shown that enzymes in the upper part of the phylogenetic tree synthesize glucans with various linkage types such as α-1,3; α-1,6; α-1,2; and α-1,4, while those in the lower part of the tree synthesize only water-soluble α-1,6-linked glucans (Fig. 1 and Supplementary Table S1). Previous studies have shown that streptococcal GTF produces soluble glucans as a result of the displacement of amino acids in the catalytic domain and/or a decrease in the number of repeating units in the glucan-binding domain. Thus, it was suggested that streptococcal GTF evolved to synthesize water-soluble glucans and that the ability to synthesize water-insoluble...
also propose that this acquisition allowed pressure prompting the acquisition of multiple refined sugar by humans, which acted as a secondary selection gtfs not to acquire the gene. We propose that an important role in the varying environmental adaptations of Streptococci.

Character acquisition is an evolutionary response that follows exposure to different selection pressures such as starvation conditions and thermal stress. The resultant functional divergence plays an important role in the varying environmental adaptations of organisms. For oral streptococci, it is conceivable that the availability of fermented foodstuffs and encounters with other lactic acid bacteria that possessed glycosyl hydrolase family 70 enzymes were major environmental selection pressures contributing to their acquisition of gtf genes. Non-oral streptococci, such as Str. pneumoniae and pyogenic group streptococci do not possess the gtf gene. At least, the gtf gene does not exist in the genome of those streptococci. The corresponding parameter of the NJ algorithm was set at ‘complete deletion’, and the ‘nucleotide: p-distance’ model and ‘bootstrap method’ were used. The mean interpopulational evolutionary diversity of the sequences encoding the catalytic domain of GTFs among Streptococcus, Leuconostoc, and Lactobacillus was also calculated with MEGA 5. The ‘complete deletion’, ‘bootstrap method’, and ‘p-distance’ model were used in this analysis.

Mapping of the genomic location of glycosyl hydrolase family 70 genes. The locations of the genes encoding glycosyl hydrolase family 70 proteins from Str. mutans, Lcb. reuteri, and Lcb. sakei were obtained using the Annotation Search Tool in the Comprehensive Microbial Resource of J. Craig Venter Institute (http://www.jcvib.org/) or by referring to previous reports. The upstream and downstream sequences of the gtfB gene from Str. oralis UO5 genome (GenBanK: FR720602.1) were analysed using BLAST (NCBI; http://www.ncbi.nlm.nih.gov/BLAST/).

Methods
Glycosyl hydrolase family 70 sequences. The DNA and amino acid sequences of glycosyl hydrolase family 70 proteins used in this study were obtained from GenBank at NCBI (http://www.ncbi.nlm.nih.gov/) with cross-reference to Pfam (http://pfam.sanger.ac.uk/). Sequencing analyses of unknown streptococcal gtf genes was performed as described previously, and the sequences were deposited in the DNA Data Bank of Japan (http://sakura.ddbj.nig.ac.jp/sakura_en.html). We analysed 20 GTFs from Streptococcus; 2 GTFs, 9 dextran sucrases, and 1 alternan sucrase from Leuconostoc; 10 glucan sucrases from Lactobacillus; and 2 GTFs from Lactococcus. The NCBI accession numbers of these GTFs are provided in Figure 1 and Supplementary Table S1.

Homology and protein family search. Motifs from GTFs were analyzed by homology and protein family searches using BLAST in NCBI and SYSTERS Protein Family Database in Max Plank Institute for Molecular Genetics, Computational Molecular Biology (http://systers.molgen.mpg.de/), respectively.

Phylogenetic analysis. Sequence alignment was performed using ClustalX software version 1.83 (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/#G). Multiple alignment files saved by ClustalX in the Clustal format (*.aln) were converted to the MEGA format (*.meg) using the MEGA version 5 software (http://www.megasoftware.net/). Phylogenetic analysis was performed by the NJ, ME, and MP methods using MEGA version 5 software. Phylogenetic distances were calculated by the NJ method using the same software. The upstream and downstream sequences of the gtfB gene from Str. oralis UO5 genome (GenBanK: FR720602.1) were analysed using BLAST (NCBI; http://www.ncbi.nlm.nih.gov/BLAST/).


Acknowledgments
The authors would like to thank Ikuri Konishi for providing technical assistance. This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) from the Japan Society for the Promotion of Science (no. 23659968 and 24659912) and the JSPS International program for Young Researcher Oversea Visits.

Author contributions
T.H. conceived the study. S.K. and T.F. supervised this project. T.H. analyzed the genetic data and constructed phylogenetic trees. T.H. and T.F. wrote the paper. All authors discussed the results and edited the manuscript.

Additional information
Accession code: The datum of Str. sanguinis gtfP has been deposited in the GenBank database under accession code AB252650.
Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.
License: This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/