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.

In humans, dental caries is defined as the demineralisation of teeth by the bacterial production of organic acid from fermentable substrates such as monosaccharides, oligosaccharides, and processed polysaccharides¹³. The dental caries pandemic has been previously discussed from an anthropological and archaeological viewpoints in terms of the consumption of fermentable carbohydrates⁴–⁷. Indeed, the emergence of dental caries positively correlates with the increased consumption of carbohydrates through the main source, grain⁸–⁹, following the advent of prehistoric farming. At that time, the degree of caries was comparatively mild, since oral bacteria produced only low amounts of organic acid. However, the widespread use of refined sugar during the industrial revolution due to improvements in food processing techniques, stabilisation of plantation-based sugar cane supplies, and the abolition of the Corn Law¹⁰ led to a sudden increase in the prevalence of caries.

Thus far, only anthropological factors, like those described above, have been considered in the development of dental caries in humans, while factors affecting the disease-causing bacteria themselves have been largely ignored. This may be based on the firmly held belief that caries-causing bacteria preceded the development of caries. However, presently, there are no studies describing the origin of caries-causing bacteria. Today, *Streptococcus mutans* is considered the principal causative organisms of dental caries. These bacteria have the ability not only to metabolise large amounts of acid from various sugars but also to firmly adhere to the tooth surface using adhesive glucans produced from refined sugar by glucosyltransferases (GTFs)¹¹. This latter property of caries-causing bacteria is the principal characteristic distinguishing prehistoric caries from modern caries.

Because of their role in the formation of dental biofilms, streptococcal GTF enzymes are thought to be one of the primary virulence factors of *Str. mutans* responsible for the development of dental caries¹¹–¹⁵. Thus, phylogenetic analysis of the origin of GTFs may provide novel insights into the acquisition of cariogenicity by *Str. mutans*. GTFs (EC: 2.4.1.5), which are encoded by *gtf* genes, belong to the glycosyl hydrolase family 70 and catalyse the transfer of D-glucopyranosyl units from sucrose to acceptor molecules. To predict the ancestry of streptococcal GTFs, phylogenetic trees for the putative amino acid sequences of glycosyl hydrolase family 70 proteins were constructed using the neighbour-joining (NJ), minimum evolution (ME), and maximum parsimony (MP) methods. The surrounding sequences of the genes encoding the glycosyl hydrolase family 70 proteins were also analyzed. Our findings suggest that the genus *Streptococcus* acquired the *gtf* genes via
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 dextranrases from *Leuconostoc* and GTFs from *Lactobacillus* and *Lactococcus*, suggesting a common ancestry for the genes encoding these enzymes.

---

**Figure 1 | Phylogenetic analysis of glycosyl hydrolase family 70 enzymes.** A phylogenetic tree of the amino acid sequences of glycosyl hydrolase family 70 enzymes was constructed using the maximum parsimony (MP) method. Genes bracketed with bold lines are adjacent to each other in the bacterial genome. The value on each branch is the estimated confidence limit (expressed as a percentage) for the position of the branches, as determined by bootstrap analysis. Only values exceeding 50% are shown.
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 gtf180 from Lactobacillus reuteri and gtfKg15 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). Trees based on both the gene and deduced amino acid sequences indicated that the streptococcal GTF family could be classified into 3 clusters: the water-soluble glucan-synthesising group (WSG), the water-insoluble glucan-synthesizing group (WIG), and the intermediate group (INT). 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.

Phylogenetic analysis of the functional regions of 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). Trees based on both the gene and deduced amino acid sequences indicated that the streptococcal GTF family could be classified into 3 clusters: the water-soluble glucan-synthesising group (WSG), the water-insoluble glucan-synthesizing group (WIG), and the intermediate group (INT). 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 GTF groups, 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>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>WIG</td>
<td>0.555</td>
<td>0.366</td>
<td>0.015</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>INT</td>
<td>0.571</td>
<td>0.393</td>
<td>0.381</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>gtfB</td>
<td>0.539</td>
<td>0.017</td>
<td>0.021</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>gtfC</td>
<td>0.560</td>
<td>0.144</td>
<td>0.022</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>gtfD</td>
<td>0.537</td>
<td>0.397</td>
<td>0.338</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
</tbody>
</table>

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

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 Lactobacillus, Lactobacillaceae, and Streptococcus. More specifically, our data suggests that Lcb. reuteri GTFB or Lcb. reuteri GTFML41 would be the practical ancestor of the streptococcal GTFs. Our previous finding that only 51% of 41 Str. oralis clinical isolates possess the gtf gene supports the idea that some of the Streptococcus spp. are in the process of acquiring 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.

It may seem likely that glycosyl hydrolase family 70 enzymes were produced by a combination of sucrase and glucan-binding proteins, such as Gbp from Str. mutans. However, in silico analyses using the DNA and protein databases, revealed that streptococcal GTF catalytic regions are homologous to α-amylase, but not to sucrase. These analyses further showed that the glucan-binding regions are categorized as CW_binding, including 1,4-β-N-acetylmuraminidase and N-acetylmuramoyl-l-alanine amidase. These findings suggest that the common ancestor of the glycosyl hydrolase family 70 enzymes would be an enzyme group whose general function was sugar hydrolysis and sugar chain binding. In addition, glucan-binding proteins, such as Gbp, were found in Str. mutans only in the present database. For this reason, it has been suggested that Gbp, one of the virulence factors allowing adherence to the tooth surface, was first produced due to the lack of the first half of glucosyltransferase in Str. mutans.

Homologues can be acquired through gene duplication (paralogs) or horizontal gene transfer (orthologs), and evidence exists to support both methods of acquisition in the glycosyl hydrolase family 70 genes. Our study located a transposase encoding gene upstream or downstream of the gtf gene in both Lcb. reuteri and Lcb. sakei (Fig. 2A). A sequence highly homologous to streptococcal transposase genes was also found downstream of Str. oralis gtfR (Fig. 2A and B). It has been reported that intra- and interspecies genomic variation in Streptococcus spp. is associated with transposase and that the bacterial population is a flexible gene pool. These findings suggest that the streptococcal gtf gene was acquired through a horizontal gene transfer via transposons, while periodic extinguishing of transposable elements may explain their absence around the gtf gene in some species of bacteria. On the other hand, glycosyl hydrolase family 70 genes are believed to easily duplicate, since many bacteria possess multiple glycosyl hydrolase family 70 genes within the same genome. Tandem-arranged genes such as gtfB and gtfC (Fig. 2) are considered to have replicated by tandem duplication, a phenomenon observed in Str. mutans, Streptococcus salivarius, and Streptococcus criceti and in the lactic acid bacterium, Lcb. reuteri (Fig. 1). Based on these findings, Str. mutans may have acquired the 3 gtf genes as orthologs from other bacteria and/or as paralogs by duplication.

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 and 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 (WSG, WIG, 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 GTGB 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...
Figure 4 | Hypothetical model of GTF enzyme acquisition by oral Streptococci. The genus Streptococcus acquired the gtf gene via horizontal gene transfer when it encountered lactic acid bacteria, such as Lactobacillus and Leuconostoc, in the human oral cavity. This event coincided with the introduction of fermented foods into the human diet, which was brought about by the development of agricultural practices and food preservation methods.

glucans would only be acquired if the ancestral character was strongly retained and highly specialised.

As shown in Figure 3A and B and Table 1, both gtfB and gtfC belong to the WIG cluster and adjoin each other in the phylogenetic tree, with only a short distance of 0.144 separating them. gtfC is probably replicated after gtfB, as the phylogenetic distances between the 2 genes and the root are 0.560 and 0.539, respectively. In contrast, gtfB and gtfD are divided into separate clusters, and the phylogenetic distance between them is relatively high (0.397). Thus, it is likely that these 2 genes were acquired separately; however, the similar phylogenetic distances between these genes and the root (gtfB, 0.539; gtfD, 0.537) suggest that they were acquired at similar time points.

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. pneumoni-tiae and pyogenic group streptococci do not possess the gtf gene. At least, the gtf gene does not exist in the genome of those streptococci on the current database. This suggested that those streptococci that did not encounter the lactic acid bacteria carrying the gtf gene could not acquire the gene. We propose that Streptococcus obtained the first gtf gene via gene transfer as an initial step towards cariogenicity. The character acquisition was then completed following the consumption of refined sugar by humans, which acted as a secondary selection pressure prompting the acquisition of multiple gtf genes (Fig. 4). We also propose that this acquisition allowed Str. mutans to develop a niche on the tooth surface, as opposed to other oral regions, such as the saliva, tongue, pharynx, or larynx.

In this study, we suggest that the reasons underlying the emergence of dental caries were not only anthropological but also bacteriological, i.e., genetic events leading to the evolution of cariogenic Str. mutans. We therefore recommend that a molecular biological approach, such as PCR-based detection of the gtf genes, be adopted for future anthropological and archaeological studies. This would enable a more precise distinction between caries and tooth wear as well as analysis of the historical expansion of Str. mutans.

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,26,30, 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 Planck 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/FAQ). 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 corresponding parameter of the NJ algorithm was set at ‘complete deletion’, and the ‘nucleotide: p-distance’ model and ‘bootstrap method’ were used. The mean interpopulation 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.jicv.org/) or by referring to previous reports.19,24 The upstream and downstream sequences of the gfr gene from Str. oralis UO5 genome (GenBank: FR720062.1) 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. 2365968 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/.