Research Article

Streptococcus mutans gene expression and functional profile in root caries: an RNA-seq study

Heitor Sales de Barros Santos [a], Thuy Do[b], Clarissa Cavalcanti Fatturi Parolo[a], Joice de Faria Poloni[c,d], Marisa Maltz[a], Rodrigo Alex Arthur[a*], Nailê Damé-Teixeira[e]

[a]Department of Preventive and Community Dentistry, School of Dentistry, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[b]School of Dentistry, University of Leeds, Leeds, UK

[c]Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[d]Laboratory of Bioinformatics in Bioenergy, EMBRAPA Agroenergy, Distrito Federal Brasília, Brazil

[e]Department of Dentistry, School of Health Sciences, University of Brasília, Brasília, Brazil

Short Title: *S. mutans* in root caries: an RNA-seq study

Corresponding Author:

Rodrigo Alex Arthur

Department of Preventive and Community Dentistry

School of Dentistry, Federal University of Rio Grande do Sul

Rua Ramiro Barcelos 2492,

Porto Alegre, RS, 90035-003, Brazil

Tel:+55 (51)3308-5193

E-mail: rodrigoarthur.ufrgs@gmail.com

Number of Tables: 03 (as online supplementary material).

Number of Figures: 05 (in the manuscript); 03 (as online supplementary material).

Word count: 5,110

Keywords: Biofilms, Microbiology, RNA-seq, Root caries, *Streptococcus mutans*

Abstract

The literature is still scarce on studies describing *S. mutans* global gene expression under clinical conditions such as those found on complex biofilms from sound root surfaces (SRS) and carious root surfaces (RC). This study aimed to investigate the *S. mutans* gene expression and functional profile within the metatranscriptome of biofilms from SRS and from RC in an attempt to identify enriched functional signatures potentially associated with healthy to disease transitioning process. Total RNA was extracted, and prepared libraries (SRS=10 and RC=9) were paired-end sequenced using the Illumina HiSeq2500. Read count assigned to each gene of the *S. mutans* UA159 strain were obtained. Differentially expressed genes (DEG) between SRS and RC were identified using the DESeq2 R package and weighted gene co-expression network analysis (WGCNA) was performed to explore and identify functional modules related to SRS and RC. We found seventeen DEG between SRS and RC samples, with three overexpressed in RC and related to membrane protein, alanyl-tRNA synthetase and GTP-binding protein with the remaining ones overexpressed in SRS samples and related to hypothetical protein, transposon integrase, histidine kinase, putative transporter, bacteriocin immunity protein, response regulator, 6-phospho-beta-galactosidase, purine metabolism and to transcriptional regulator. Key-functional modules were identified for SRS and RC conditions based on WCGNA, being 139 hub genes found on SRS key-module and 17 genes on RC key-module. Functional analysis of *S. mutans* within the metatranscriptome of biofilms from sound root and from carious root revealed a similar pattern of gene expression, and only a few genes have been differentially expressed between biofilms from sound root surfaces and from root carious lesions. However, *S. mutans* presented a greater functional abundance in the lesion samples. Some functional patterns related to sugar (starch, sucrose, fructose, mannose and lactose) and heterofermentative metabolisms, to cell-wall biosynthesis and to acid tolerance stress seem to be enriched on carious root surfaces conferring ecological advantages to *S. mutans*. Altogether, the present data suggest that a functional signature may be associated with carious root lesions.

Introduction

The oral cavity is colonized by a wide variety of microbial species which establish a symbiotic relationship with the host being compatible with healthy conditions [Belda-Ferre et al., 2012]. This beneficial host-microbiota interplay is disrupted by several factors, including frequent dietary sugars intake which leads to the selection of highly acidogenic and highly acid tolerant microorganisms in the detriment of acid sensitive ones [Marsh, 2003]. Ultimately, this microbial dysbiosis leaves dental surfaces with a significantly increased risk of developing dental caries [Takahashi and Nyvad, 2011]. *Streptococcus mutans* has the ability to elicit an efficient acid tolerance response, which gives it a competitive advantage and makes it a prevalent organism in cariogenic biofilms [Lemos et al., 2019]. Reports have found evidence of its higher abundance in root caries lesions compared to sound root surface biofilms [Bowden et al., 1990; Preza et al., 2008].

There is an understanding that both microbial interactions and physiological functions of poly-microbial communities may exert a decisive role on the onset of biofilm-associated diseases, such as dental caries [Nyvad et al., 2013]. This is especially important for complex microbiota such as that found in root caries lesions. Dental biofilm/carious dentine collected from root carious lesion surfaces present high abundances of *Actinomyces* spp. *Atopobium* spp., *Enterococcus faecalis,* lactobacilli, *Olsenella profusa*, *Oribacterium* spp., *Propionibacterium acidifaciens*, *Pseudoramibacter alactolyticus Selenomonas* spp., *Prevotella multisaccharivorax*, *Streptococcus mutans* among others [Preza et al., 2008; Chen et al., 2015]. By analyzing the transcriptome of some micro-organisms in the metatranscriptome of root caries, a potential cariogenicity attributed to *Candida albicans*, symbiotic relationships between *Lactobacillus* spp. and *Scardovia* spp. and/or an antagonism between the latter and *S. mutans* have been suggested [Dame-Teixeira et al., 2020; Ev et al., 2020]. *Actinomyces* spp. was found as *core* microbiome in the root biofilm irrespective to the presence or absence of root caries [Dame-Teixeira et al., 2016]. However, increasing evidence suggest that root carious lesion microbial composition is site-specific being dependent on whether the assessed surface is supragingival or whether it extends beyond the gingival margin [Takenaka et al., 2021]. This way, while *Actinomyces* spp., *Prevotella* spp. and *Streptococcus* spp. are more abundant in supragingival lesions, some periodontitis-associated pathogens are commonly found on lesions extending beyond the gingival margin [Takenaka et al., 2021]. Considering these tissue-dependent differences, the identification of potential microbial signatures for root caries lesions has not been fully addressed. Moreover, these findings implied an underestimated complexity of microbial interactions in root caries. Moreover, an increased *S. mutans* gene expression related to collagenolytic proteases was found within a poly-microbial community associated with root caries [Dame-Teixeira et al., 2018]. To date, despite the

evidence there is not a clear indication of microbial or functional biomarkers associated with root caries lesions. Moreover, a comprehensive global *S. mutans* transcriptomic analysis carried out under clinically relevant condition is still lacking. Taken *S. mutans* is highly cariogenic, understanding its global gene expression within complex biofilms collected from sound root surfaces and root carious lesions might help identify functional patterns potentially associated with healthy to disease transitioning process.

In this context, metatranscriptome analysis allows the investigation of a large number of genes and associated functional patterns present in the biofilm samples. This highly sensitive technique uses next-generation sequencing platforms that can be aligned to bacterial genomes, enabling transcriptomic analysis of individual species [Do et al., 2015]. More recently, gene co-expression network analysis has been suggested as a useful method to depict genome-wide function [van Dam et al., 2018]. This analysis may reveal genes with their co-expressed partners that could be functionally correlated with healthy or disease states [Wang et al., 2020 and others]. Interestingly, this method has not been used so far to address the relationship between dental biofilm functional activity and dental caries.

Therefore, this study aimed to investigate the *S. mutans* gene expression and functional profile within the metatranscriptome of biofilms from sound root and from carious root surfaces to identify enriched functional signatures, by both differential gene expression and by gene co-expression network analysis.

Materials and Methods

Samples origin and metatranscriptome data

Data from this study were originated from the project "Metatranscriptome of root caries". Complete data is available in the National Center for Biotechnological Information (NCBI), under accession numbers SRS779973 and SRS796739. Patient selection, collection and sample preparation were performed as described elsewhere [Dame-Teixeira et al., 2016; 2018; 2020; Ev et al., 2020], and depicted in the Supplementary Material. Briefly, biofilm samples were collected from participants who had at least one exposed root surface without dental caries lesions (group with sound root surface – SRS) or who had at least one primary active cavitated root lesion that needed restorative treatment (group of root caries – RC), resulting in 10 libraries of SRS and 9 of RC. Libraries were sequenced with Illumina HiSeq2500. The genes belonging to the *S. mutans* UA159 genome and the read count assigned to each gene were then manually extracted from the 162 genomes identified on the studied samples [Ajdić et al., 2002]. Genes that were not expressed in at least in three different samples to each

condition were excluded from downstream analysis [Rau et al., 2013]. Differentially expressed genes (DEGs) between SRS and RC were obtained by DESeq2 R package considering as cut-off adjusted p-value < 0.05 and |log2 fold change| ≥ 1 (Love et al. 2014). To explore and identify functional modules related to SRS and RC conditions, Weighted gene co-expression network analysis (WGCNA) was performed [Zhang and Horvath 2005; Langfelder and Horvath 2012]. Pearson's correlation between trait (SRS and RC) and module eigengene was performed to identify modules significantly associated with the clinical condition. The node relevance was defined by considering gene significance ≥ 0.6 and module membership ≥ 0.6 as a cut-off. The R package clusterProfiler with the function *enrichKEGG* was used to identify the most relevant biological processes related to each of the clinical condition.

Results

Overall and Differential Gene Expression

All samples presented countable mRNA reads attributed to *S. mutans* in both SRS and in RC surfaces (Fig. 1). Two RC libraries originated from Leeds Dental School (U.K.) (RC_D and RC_E), while all the others originated from School of Dentistry of Federal University of Rio Grande do Sul (Brazil). The total *S. mutans* read count (normalized using the DESeq2 R package) per sample ranged between 274.96 – 120,459.78 for SRS and between 117,217.27 – 7,701,534.27 for RC. Median read count for SRS (1,123.89) was statistically lower than for RC (1,441,964.47) according to Mann-Whitney Rank Sum Test ($p<0.001$) (Fig. 1). Samples were grouped into two different clusters according to the general gene expression variability found in each clinical trait (SRS or RC) (Fig. 2). A sample-to-sample distance analysis for gene expression also showed an analogous trend for RC and SRS clustering, which indicates that there is high similarity between samples from the same condition (Supplementary Fig. 1).

To assess the differential gene expression, a cut-off point of presence in at least 3 samples of each trait was adopted, resulting in the analysis of 304 genes for which overall expression is shown in Supplementary Figure 2. a. Seventeen genes were differentially expressed between SRS and RC samples (|log2 fold change|>1 and $p<0.05$) (Supplementary Fig. 2. b and Supplementary Table 1). Most of the DEGs are related to proteins located as integral component of membrane. DNA recombination, maintenance of CRISPR repeat elements, translation, signal transduction system, regulation of transcription and carbohydrate/IMP metabolic processes were some of the biological processes associated with DEGs. Some of DEGs are related to molecular functions of DNA, t-RNA, GTP and ATP binding, as well with integrase, kinase, t-RNA ligase, endoribonuclease, ATPase, GTPase and galactosidase activities. Three genes were overexpressed in RC and their putative functions are related to membrane protein, alanyl t-RNA aminoacylation and translation. The other fourteen genes were

overexpressed in SRS. Putative functions of ten out 14 genes are related to DNA recombination and binding, histidine kinase, bacteriocin immunity, carbohydrate and purine metabolisms and kinase activity. The remaining four overexpressed genes are related to hypothetical protein (Supplementary Table 1).

WCGNA and *Functional analysis within key-modules*

After processing the data, all samples were included in the co-expression analysis, except for sample RC_7 which was excluded as an outlier (Supplementary Fig. 3). Highly correlated genes were clustered by the hierarchical clustering tree into 9 modules which were represented with different colors (Fig. 3. a). Correlation coefficients were calculated between the eigengenes modules and the clinical conditions (SRS or RC). One module (blue color) correlated with the SRS condition and another one (green color) correlated with the RC condition ($r \geq 0.6$ and $p < 0.05$) (Fig. 3. b-d) being further referred as "SRS key-module" and "RC key-module", respectively. One-hundred third-nine genes in the SRS key-module and 17 genes in the RC key-module presented GS ≥ 0.6 and MM ≥ 0.6 (Fig. 4. a-b). Several functions are putatively attributed to these hub genes. Whereas most of SRS key-module hub genes encode for uncharacterized protein, some of them are related to transposon/transposase proteins, transcriptional regulators or endonucleases activities. Moreover, DNA integration/recombination/transposition and histidine biosynthetic process are some of the biological functions attributed to those genes. In respect to the RC key-module, some of the hub genes are related to the processes of biofilm formation, intracellular transmembrane transport, carbohydrate/organic acids metabolic processes, leucine biosynthesis along with cell cycle/cell division, cell wall organization and peptidoglycan biosynthetic process. Both SRS and RC hub-genes are also related to DNA replication, translation and regulation of transcription (Supplementary Tables 2 and 3).

Overall, histidine metabolism, metabolic pathways, biosynthesis of secondary metabolites, microbial metabolism in diverse environments, biosynthesis of amino acids, ABC transporter, *quorum sensing*, ribosome and protein export are likely to be found on the SRS key-module (Fig. 5. a) and glycolysis/gluconeogenesis, fructose and mannose metabolism, lysine biosynthesis, starch and sucrose metabolism, peptidoglycan biosynthesis, one carbon pool by folate, aminoacyl tRNA biosynthesis, metabolic pathways, biosynthesis of secondary metabolites, microbial metabolism in diverse environments, carbon metabolism, biosynthesis of amino acids, ABC transporters, two-component system, *quorum sensing*, ribosome and RNA degradation are likely to be found on the RC key-module (Fig. 5. b). Considering the node relevance within each key-module (GS ≥ 0.6 and MM ≥ 0.6; adjusted

p<0.05), histidine metabolism, ABC transporter, *quorum sensing*, microbial metabolism in diverse environments, biosynthesis of amino acids, biosynthesis of secondary metabolites and metabolic pathways were likely to be found on the SRS key-module (Fig. 5. c), and *quorum sensing* and ABC transporter were likely to be found on the RC key-module (Fig. 5. d).

Both WCGNA and the functional analyses described above mean that the genes within the same key-module behave similarly in terms of expression (either up- or downregulated). In this case, and compared with DEGs, the data show that a robust and intricate group of genes is actively involved in *S. mutans* metabolic activity on both SRS and on RC samples. It is likely specific groups of genes are activated/deactivated promoting a concerted response (represented by the above-mentioned functional pathways) that either contribute to the health (in the case of SRS samples) or to the disease status (in the case of RC samples).

Discussion

The literature is still scarce on studies describing the physiology of *S. mutans* under clinical conditions such as those found on naturally formed complex dental biofilms. Available evidence from *in vitro* studies have shown an altered *S. mutans* gene expression profile in response to distinct environmental conditions [Gong et al., 2009; Baker et al., 2015]. By investigating the transcriptome of *S. mutans* within the metatranscriptome landscape of biofilms from SRS and from RC samples we found reads attributed to *S. mutans* suggesting it is metabolically active in both conditions (Fig. 1), although with a significantly higher relative abundance in RC.

The higher number of reads found in RC surfaces indicates greater *S. mutans* functional abundance in comparison with SRS (Fig. 1). The heat map of the co-expressed genes showed that most of the samples presented a similar pattern within each correlated key-module (Figure 3C and 3D) which could be representative of each clinical condition. Interestingly, the SRS key-module presented a higher number of hub genes compared with the RC key-module (Fig. 4. a-b and Supplementary Tables 2 and 3). Moreover, most of the identified DEGs were found in SRS condition (Supplementary Table 1). This way, it is likely the development of a cariogenic biofilm, besides involving a microbiological shift towards the increase on the abundance of more acidogenic and acid-tolerant micro-organisms (Nyvad & Takahashi, 2020), may also modify the expression of a specific and minor group of *S. mutans* genes eliciting specific functional patterns. This hypothesis may be confirmed by observing that most of the identified functional patterns were shared by both SRS and RC biofilms, and were related to global metabolism (such as metabolic pathways, biosynthesis of secondary metabolites, microbial

metabolism in diverse environments, biosynthesis of amino acids and ribosome), ABC transporter and *quorum sensing* (Fig. 5). It is important to point out that samples within each distinct sample group (SRS or RC) presented a certain level of similarity on the overall gene expression being grouped into different clusters (Fig. 2). Overall, these findings may indicate that *S. mutans* may behave similarly in both SRS and RC conditions but some functional activities found in RC could potentially be associated with the development of a cariogenic biofilm.

Some specific functional patterns were likely to be enriched in RC samples. This is the case of carbohydrate metabolism associated pathways, such starch, sucrose, fructose and mannose metabolism. Sugar uptake by oral streptococci is mediated by phosphoenolpyruvate:sugar phosphotransferase system (PEP:PTS) [Vadeboncoeur and Pelletier, 1997; Ajdić and Pham, 2007]. Overall, PTS for glucose, fructose, maltose and sucrose is constitutively expressed on *S. mutans* (Ajdić and Pham, 2007), but, higher expression of inducible fructose/mannose transporter operon which encodes for EII subunit of PTS is expected to be found whenever fructose or mannose is available [Ajdić and Pham, 2007]. It is plausible to consider then that the higher $EII^{Fru/Man}$ transporter gene expression might be a consequence of sucrose breakdown by fructosyltransferase (encoded by *ftf* genes) [Hamada and Slade, 1980] which lead to further fructose uptake via PTS. Conversely, the fructose PTS activity is also up-regulated in the presence of mannose [Abranches et al., 2003]. Optimal mannose PTS transport is also associated to an optimal expression of *S. mutans* glucosyltransferase (*gtf*BC) and *ftf* genes [Abranches et al., 2003], being those compatible to a high cariogenic condition represented in RC samples. This interplay between mannose/fructose metabolism is also observed on the high correlation among co-expressed genes *man*A (coding for mannose-6-phosphate isomerase involved in fructose and mannose metabolism), *leu*B (coding for 3-isopropylmalate dehydrogenase) and *ack*A (coding for acetate kinase) (Fig. 4. c-c1), being *leu*B and *ack*A interconnected with pyruvate metabolism. Interestingly, genes *pik*F (coding for pyruvate kinase) and *act* (related to the pyruvate-formate lyase pathway) were within the hub genes of RC key-module (Supplementary Table 3). These genes, along with *ack*A, are related to pathways that might be activated under anaerobic condition [Abbe et al., 1982; Yamada et al., 1985]. This means that localized anaerobic pockets found on RC biofilms seems to induce *S. mutans* heterofermentative metabolism through the conversion of fructose/mannose into acetate, formate and ethanol. Yet, considering the ability of *S. mutans* to bind to extracellular polysaccharide via cell-surface non-glucosyltransferase glucan-binding proteins (Gbp), especially GbpC [Banas and Vickerman, 2003; Lynch et al., 2007], it is important to highlight that *gbp*C gene was found as one with the highest gene significance among the hub genes of the RC key-module (Supplementary Table 3). This suggests that *gbp*C might then play an important role on the colonization of RC associated biofilms by *S. mutans* and this needs to be further studied.

S. mutans cell-wall biosynthesis and signaling system known as two component signaling transduction system (TCSTS) pathways are also enriched in RC samples. The RC key-module hub gene *mtl*G encodes for endolytic murein transglycosylase which is involved on peptidoglycan biosynthetic process (Supplementary Table 3). The incorporation of specific amino-acids into the interpeptide bridge of the peptidoglycan is performed by tRNA-dependent aminoacyl transferases being L-alanyl tRNA ligase an intermediate of this process [de Pascale et al., 2008]. In agreement with this, we observed that the gene *ala*S (coding for alanine-tRNA ligase) was overexpressed on RC samples (Supplementary Table 1). Besides, the enrichment on lysine biosynthesis pathway found on RC samples is also important for cell-wall biosynthesis since lysine, in addition to be involved in the synthesis of macromolecules, such as protein, DNA and RNA [Schleifer and Kandler, 1972; Mattingly et al., 1976], is an essential amino-acid that is precursor of peptidoglycan. LiaSR, one of the TCSTS pathways present in *S. mutans*, is encoded by the operon *lia*FSR-*ppi*B-*pnp*B that is also showed to be involved on peptidoglycan synthesis [Suntharalingam et al., 2009]. Moreover, there is some evidence showing that the total amount of GTP-binding proteins increases and a substantial portion of them attach to *S. mutans* cell membrane under acidic pH [Baev et al., 1999]. Recent investigation has also indicated that genes coding for GTP-binding may play a role on microbial acid-tolerance [Zhang et al., 2018]. These findings agree with the overexpression of gene *SMU_546* (coding for a putative GTP-binding protein) on RC samples (Supplementary Table 1). A high correlation on the co-expression of *gtl*X (coding glutamyl-tRNA synthetase), *SMU_75* (coding D-alanyl-D-alanine carboxypeptidase in peptidoglycan biosynthesis) and *htp*X (coding heat shock protein) was also observed on the RC key-module (Fig. 4. c-c2). The gene *fts*Y (a hub gene of the RC key-module that encodes for membrane-associated receptor protein FtsY), a conserved part of the signal recognition particle (SRP) translocation pathway [Hasona et al., 2005], seems to participate on acid stress tolerance [Hasona et al., 2005]. It is important to highlight that in *Streptococcus pneumoniae*, a Gram-positive bacteria, the above mentioned tRNA-dependent aminoacyl transferases is also connected to enhanced penicilin resistance [Aggarwal et al., 2021]. Furthermore, the gene *pbp*2b plays an important role in *S. pneumoniae* peptidoglycan formation [Berg et al., 2013]. Similarly, *S. mutans pbp*2b was found as hub gene in RC key-module (Supplementary Table 3). The gene *sec*A (a hub gene of the RC key-module; Supplementary Table 3) encodes for an ATPase transmembrane carrier component (SecA) that belongs to the microbial general secretion pathway (GSP) which is involved in the secretion of polypetides [van Wely et al., 2001]. SecA has been found to aid the secretion of glucosyltransferases and other *S. mutans* virulence factors [Huang et al., 2018]. This carrier was also found as component of *S.mutans* extracellular membrane vesicle (EMV) associated proteins [Morales-Aparicio et al., 2020] which also carries e-DNA and proteins, both playing a pivotal role on biofilm formation [Liao et al., 2014; Wen et al., 2021].

Altogether, these data indicate that an interplay between cell-wall biosynthesis, heat-shock [Lemos et al., 2019] and GTP-binding proteins, as well as SRP pathway (in respect to *fts*Y gene) may be related to *S. mutans* acid tolerance response on RC samples. Moreover, *S. mutans* metabolism is also directed to aid on biofilm formation under the cariogenic conditions found on RC samples. Yet, NrdD (which is encoded by *ndr*D, a hub gene found in the RC key-module; Supplementary Table 3) is a class III ribonucleotide reductases linked to the conversion of ribonucleotides to the deoxyribonucleotides [Jordan & Reichard, 1998]. It seems that this enzyme is important to anaerobic growth and virulence in *Staphylococcus aureus* and in *Streptococcus sanguinis* [Paik et al., 2005; Kirdis et al., 2007]. But, besides its common role allowing DNA replication and repair, not much is unknown about its activity in *S. mutans* virulence, which needs to be further investigated.

Histidine kinase is the transmembrane sensor that belongs to TCSTS which is linked to gene expression regulation. Both genes *SMU_1145c* and *SMU_1146c*, which were overexpressed on SRS samples (Supplementary Table 1), being *SMU_1145c* also found as a hub gene of SRS key-module (Supplementary Table 2), encode for a histidine kinase (HK3) and its response regulator (RR3), respectively, forming *S. mutans* TCSTS-3 [Levesque et al., 2007]. The gene *SMU_1261c* (that encodes for histidine biosynthetic process) was also part of the same group of hub genes (Supplementary Table 2). This specific TCSTS-3 shares some identity with the ScnRK system present in *Streptococcus pyogenes* which is responsible for bacteriocin production [McLaughlin et al. 1999]. *S. mutans* produces a type of bacteriocin known as mutacin that could suppress the growth of other competing bacteria found in dental biofilm [Zhang et al., 2009; Lemos et al., 2019] being this an important mechanism responsible for microbial co-existence in the biofilm [Kawada-Matsuo et al., 2013]. However, bacteriocin-producing organisms are resistant to their own bacteriocins through the action of bacteriocin immunity proteins. These immunity proteins are encoded not only by *SMU_1148* and *SMU_1149* genes overexpressed on SRS samples (Supplementary Table 1), but also by *SMU_1150* ,a hub gene of SRS key-module (Supplementary Tables 1 and 2). Therefore, we consider that, under a homeostatic condition represented by SRS samples and in the absence of selective pressure induced by frequent environmental acidification which could favor the growth of acid-tolerant microorganisms, the ability to produce bacteriocin might exert a certain ecological defense for *S. mutans*.

Transposons are self-replicated genetic elements responsible for genome plasticity contributing to microorganism´s diversification and adaptation [Juhas et al., 2009].The motility of those elements depends on the integration among DNA/protein complexes and the activity of many enzymes. Several hub genes present on SRS key-module seems to be related to those mobile elements, such as transposon proteins, integrases, transposases and excisionase (Supplementary Tables 1 and

2). On the other hand, CRISPR-*cas* system (whose *cas* gene was overexpressed on SRS samples; Supplementary Table 1) provides resistance against phages and non-self-mobile genetic elements due to an RNA-interference-like mechanism [Sorek et al., 2008]. Defense against exogenous DNA is also played by *hsd* type I restriction–modification system (RMS) [Koonin et al., 2017]. RMS is formed by two DNA methyltransferase subunits (M; involved in DNA methylation), two restriction endonuclease subunits (R; encoded by *hsd*R and involved in DNA cleavage) and one specifity subunit (S; encoded by *hsd*S and involved in targeting a specific DNA sequence)[Murray, 2000]. Both *hsd*M and *SMU_897* (hub genes of SRS key-module) encode for subunits M and R of type I RMS (Supplementary Table 2). Yet, *S. mutans* locus *SMU.816-SMU.817* (that exhibits some similarities with *Escherichia coli rel*BE genes) is embedded in the *hsd-prr* locus [Lemos et al., 2005], being *rel*BE considered a putative toxin-anti-toxin (TA) system in *S. mutans* [Lemos et al., 2005] that acts as stress response elements under adverse conditions. In addition to all these above mentioned functions, transposon motility, *tnr*5 expression, as well as *rel*BE transcription are directly dependent on environmental stress conditions [Chattoraj et al., 2010; Lemos et al., 2005; Foster, 2009]. CRISPR-*cas* also acts in response to oxidative stress [Serbanescu et al., 2015]. Altogether, these findings may suggest that the combined effect of mobilome, CRISPR-*cas* and RMS defense systems may yield *S. mutans* on colonizing non-cariogenic biofilms. However, the exact role played by these elements on the physiological and on the fitness aspects of *S. mutans* found on biofilms associated with sound root surfaces is still unclear and it needs to be further investigated.

Moreover, an overexpression of *lac*G gene also suggests that lactose metabolism with lactic acid as the final metabolic product [Calmes and Brown, 1979] seems to be enriched in SRS samples (Supplementary Table 1). Lactose catabolism is activated by low levels of glucose [Zeng et al., 2018]. Nucleotide synthesis is also attributed to *pur*N gene (overexpressed in SRS samples; Supplementary Table 1) that is associated with purine metabolism. As previously discussed [Baker et al., 2015], this pathway might be associated with synthesis of adenine which is further used to generate ATP. We hypothesize that both lactose catabolism and nucleotide synthesis are important *S. mutans* pathways in biofilms associated with sound root surfaces but this needs to be confirmed by further studies.

Considering the role played by *S. mutans* in the health-to-disease transitioning process and in the establishment of a highly cariogenic environment, and taking into account the above mentioned evidences and the differences found between SRS and RC conditions, genes related to carbohydrate metabolism (*act*), to cell-adhesion (*gpb*C), to cell-wall biosynthesis (*mtl*G and *ala*S), acid tolerance stress (*fts*Y and SMU_546), to sugar transport as well as those DEG presenting the highest Log2foldchange (such as SMU_1399; *tnr*5, *SMU_959* and *SMU*_2133) might be further studied to

address whether they represent only an ecological advantage for *S. mutans* or they are consequence of *S. mutans* virulence. Transcriptional regulator genes found as hub genes in both SRS and RC-key modules (Supplementary Tables 2 and 3) might also be considered as promising targets for antimicrobial compounds. Future mechanistic studies should assess how and whether the modulation of those genes and/or their associated functional patterns affects dental caries onset.

Another concern that may be raised is related to what extent the distinct geographic regions interfere with general biofilm metabolic profile taking into account some differences on microbial composition among distinct regions. It is important to note, two RC libraries were originated from England (RC_D and RC_E), while all others were originated from Brazil. However, in terms of microbial function, no difference between geographic regions was found since libraries were similar in terms of *S. mutans* gene expression (Fig. 1-3) and a high similarity was observed among all tested samples from the same condition, suggesting a close relationship among the recruited patients (Supplementary Fig. 1). This finding agrees with others showing minor genetic differentiation among *S. mutans* isolates obtained from different geographical locations [Do et al., 2010; Cornejo et al., 2013]. Another point of concern is that most of the co-expressed genes found on the SRS key-module encoded for unknown proteins being the functional analysis associated with this module underestimated. We recognize this as a limitation of this study that does not enable us to have a clear and complete characterization of metabolic patterns associated with health tooth surfaces. Nevertheless, the comparative analysis between the RC and SRS key-modules revealed that a different group of genes co-expressed in RC samples led to the enrichment of specific functional patterns potentially associated with dental caries. In order to achieve a minimal concentration of total RNA (30ng/RNA) samples of RC group were pooled. We acknowledge the small number of samples (10 for SRS and 9 for RC) might have limited the significance of the present study and that a larger cohort of participants would have strengthened the data. Anyhow, despite the sample size limitation, the present data were able to show that *S. mutans* of both SRS and RC conditions are functionally different indicating future direction on the search of potential biomarker candidates for the health-to-caries transitioning process.

Conclusion

Functional analysis of *S. mutans* within the metatranscriptome of biofilms from sound root and from carious root revealed a similar pattern of gene expression, and only a few genes have been differentially expressed between biofilms from sound root surfaces and from root carious lesions. However, *S. mutans* presented a greater functional abundance in the lesion samples. Some functional patterns related to sugar (starch, sucrose, fructose, mannose and lactose) and heterofermentative metabolisms, to cell-wall biosynthesis and to acid tolerance stress seem to be enriched on carious root

surfaces conferring ecological advantages to *S. mutans*. Altogether, the present data suggest that a functional signature may be associated with carious root lesions.

Statements

Acknowledgement

We thank the Brazilian National Counsel of Technological and Scientific Development (CNPq) and the Coordination for the Improvement of Higher Level Education Personnel (CAPES) which conceived a scholarship to NDT and to HSBS.

Statement of Ethics

This study was conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

Study approval statement: This study protocol was reviewed and approved by the Ethics Committee of the Federal University of Rio Grande do Sul (protocol 8.427.168) and by the National Research Ethics Service Committee Yorkshire & the Humber - Leeds West (protocol 8 2012002DD

Consent to participate statement: All participants consented to donate samples after reading and signing an informed consent.

Conflict of Interest Statement

The authors have no conflicts of interest to declare

Funding Sources

The Dunhill Medical Trust (R245/0212) and Leeds Teaching Hospitals Charitable Foundation Trust (R&D/PP/12011) for providing financial support to this research. The funder had no role in any of the following: study design; collection, analysis, interpretation of data and writing of the report.

Author Contributions

Thuy Do, Clarissa Cavalcanti Faturi Parolo, Marisa Maltz and Nailê Damé-Teixeira contributed to conception and design. Thuy Do and Nailê Damé-Teixeira contributed to data acquisition. Joice de Faria Poloni performed the statistical analysis. Heitor Sales de Barros Santos, Joice de Faria Poloni and Rodrigo Alex Arthur contributed to data analysis and data interpretation. Heitor Sales de Barros Santos and Rodrigo Alex Arthur. All authors contributed to manuscript revision, read, and approved the submitted version.

Data Availability Statement

The data that support the findings of this study are openly available in the National Center for Biotechnological Information (NCBI), under access numbers SRS779973 and SRS796739.

References

Abbe K, Takahashi S, Yamada T: Involvement of oxygen-sensitive pyruvate formate-lyase in mixed-acid fermentation by *Streptococcus mutans* under strictly anaerobic conditions. J Bacteriol 1982;152:175–182.

Abranches J, Chen YYM, Burne RA: Characterization of *Streptococcus mutans* strains deficient in EIIAB Man of the sugar phosphotransferase system. Appl Environ Microbiol 2003;69:4760–4769.

Aggarwal SD, Lloyd AJ, Yerneni SS, Narciso AR, Shepherd J, Roper DI, et al.: A molecular link between cell wall biosynthesis, translation fidelity, and stringent response in *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A 2021;118. DOI: 10.1073/pnas.2018089118

Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, et al.: Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci U S A 2002;99:14434–14439.

Ajdić D, Pham VTT: Global transcriptional analysis of *Streptococcus mutans* sugar transporters using microarrays. J Bacteriol 2007;189:5049–5059.

Baev D, England R, Kuramitsu HK: Stress-induced membrane association of the *Streptococcus mutans* GTP- binding protein, an essential G protein, and investigation of its physiological role by utilizing an antisense RNA strategy. Infect Immun 1999;67:4510–4516.

Baker JL, Abranches J, Faustoferri RC, Hubbard CJ, Lemos JA, Courtney MA, et al.: Transcriptional profile of glucose-shocked and acid-adapted strains of *Streptococcus mutans*. Mol Oral Microbiol 2015;30:496–517.

Banas JA, Vickerman MM: Glucan-binding proteins of the oral streptococci. Crit Rev Oral Biol Med 2003;14:89–99.

Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simón-Soro A, Pignatelli M, et al.: The oral metagenome in health and disease. ISME J 2012;6:46–56.

Berg KH, Stamsås GA, Straume D, Håvarstein LS: Effects of low PBP2b levels on cell morphology and peptidoglycan composition in *Streptococcus pneumoniae* R6. J Bacteriol 2013;195:4342–4354.

Bowden GH, Ekstrand J, McNaughton B, Challacombe SJ: Association of selected bacteria with the lesions of root surface caries. Oral Microbiol Immunol 1990;5:346–351.

Calmes R, Brown AT: Regulation of lactose catabolism in *Streptococcus mutans*: purification and regulatory properties of phospho-beta-galactosidase. Infect Immun 1979;23:68–79.

Chattoraj P, Banerjee A, Biswas S, Biswas I. ClpP of *Streptococcus mutans* differentially regulates expression of genomic islands, mutacin production and antibiotic tolerance. J Bacteriol 2010; 192: 1312-1323.

Chen L, Qin B, Du M, Zhong H, Xu Q, Li Y, et al.: Extensive description and comparison of human supra-gingival microbiome in root caries and health [Internet]. . PLoS One 2015;10. Available from: http://dx.doi.org/10.1371/journal.pone.0117064

Cornejo OE, Lefébure T, Pavinski Bitar PD, Lang P, Richards VP, Eilertson K, et al.: Evolutionary and population genomics of the cavity causing bacteria *Streptococcus mutans* [Internet]. . Mol Biol Evol 2013;30:881–893.

Dame-Teixeira N, Parolo CCF, Maltz M, Tugnait A, Devine D, Do T: *Actinomyces* spp. gene expression in root caries lesions. J Oral Microbiol 2016;8:32383

Dame-Teixeira N, Fatturi Parolo CC, Maltz M, Rup AG, Devine DA, Do T: Gene expression of bacterial collagenolytic proteases in root caries. J Oral Microbiol 2018;10. DOI: 10.1080/20002297.2018.1424475

Damé-Teixeira N, Parolo CCF, Maltz M, Devine DA, Do T. Gene expression profile of *Scardovia* spp in the metatranscriptome of root caries. Braz Oral Res 2020;34:e42.

De Pascale G, Lloyd AJ, Schouten JA, Gilbey AM, Roper DI, Dowson CG, et al.: Kinetic characterization of lipid II-Ala:Alanyl-tRNA ligase (MurN) from *Streptococcus pneumoniae* using semisynthetic aminoacyl-lipid II substrates. J Biol Chem 2008;283:34571–34579.

Do T, Sheehy EC, Mulli T, Hughes F, Beighton D: Transcriptomic analysis of three *Veillonella* spp. present in carious dentine and in the saliva of caries-free individuals. Front Cell Infect Microbiol 2015;5:25.

Do T, Gilbert SC, Clark D, Ali F, Fatturi Parolo CC, Maltz M, et al.: Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST [Internet]. . PLoS One 2010;5. Available from: http://dx.doi.org/10.1371/journal.pone.0009073

Ev LD, Damé-Teixeira N, Do T, Maltz M, Parolo CCF: The role of *Candida albicans* in root caries biofilms: An RNA-seq analysis. J Appl Oral Sci 2020;28:1–10.

Foster PL: Stress-Induced Mutagenesis in Bacteria. Crit Rev Biochem Mol Biol 2007;42:373–397.

Gong Y, Tian XL, Sutherland T, Sisson G, Mai J, Ling J, et al.: Global transcriptional analysis of acid-inducible genes in *Streptococcus mutans*: Multiple two-component systems involved in acid adaptation. Microbiology 2009;155:3322–3332.

Hamada S, Slade HD: Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol Rev 1980;44:331–384.

Hasona A, Crowley PJ, Levesque CM, Mair RW, Cvitkovitch DG, Bleiweis AS, et al.: Streptococcal viability and diminished stress tolerance in mutants lacking the signal recognition particle pathway or YidC2. Proc Natl Acad Sci U S A 2005;102:17466–17471.

Huang M, Meng L, Fan M, Hu P, Bian Z: Effect of biofilm formation on virulence factor secretion via

the general secretory pathway in *Streptococcus mutans* [Internet]. . Arch Oral Biol 2008;53:1179–1185.

Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW: Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev 2009;33:376–393.

Kawada-Matsuo M, Oogai Y, Zendo T, Nagao J, Shibata Y, Yamashita Y, et al.: Involvement of the novel two-component nsrrs and lcrrs systems in distinct resistance pathways against nisin a and nukacin isk-1 in *Streptococcus mutans*. Appl Environ Microbiol 2013;79:4751–4755.

Kirdis E, Jonsson IM, Kubica M, Potempa J, Josefsson E, Masalha M, et al.: Ribonucleotide reductase class III, an essential enzyme for the anaerobic growth of *Staphylococcus aureus*, is a virulence determinant in septic arthritis. Microb Pathog 2007;43:179–188.

Koonin E V., Makarova KS, Wolf YI: Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu Rev Microbiol 2017;71:233–261.

Langfelder P, Horvath S: Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw 2012; 46(11):i11.

Lemos JA, Palmer SR, Zeng L, Wen ZT, Kajfasz JK, Freires IA, et al.: The Biology of *Streptococcus mutans*. Microbiol Spectr. 2019 Jan;7(1): 10.1128/microbiolspec.GPP3-0051-2018.

Lemos JA, Brown TA, Abranches J, Burne RA: Characteristics of *Streptococcus mutans* strains lacking the MazEF and RelBE toxin-antitoxin modules. FEMS Microbiol Lett 2005;253:251–257.

Lévesque CM, Mair RW, Perry J., Lau PCY, Li Y-H, Cvitkovitch DG: Systemic inactivation and phenotypic characterization of two-component systems in expression of *Streptococcus mutans* virulence properties. Lett Appl Microbiol 2007;45:398–404.

Liao S, Klein MI, Heim KP, Fan Y, Bitoun JP, Ahn SJ, et al.: *Streptococcus mutans* extracellular DNA is upregulated during growth in biofilms, actively released via membrane vesicles, and influenced by components of the protein secretion machinery. J Bacteriol 2014;196:2355–2366.

Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

Lynch DJ, Fountain TL, Mazurkiewicz JE, Banas JA: Glucan-binding proteins are essential for shaping *Streptococcus mutans* biofilm architecture. FEMS Microbiol Lett 2007;268:158–165.

McLaughlin RE, Ferretti JJ, Hynes WL: Nucleotide sequence of the streptococcin A-FF22 lantibiotic regulon: Model for production of the lantibiotic SA-FF22 by strains of *Streptococcus pyogenes*. FEMS Microbiol Lett 1999;175:171–177.

Marsh PD: Are dental diseases examples of ecological catastrophes?. Microbiology 2003;149:279–294.

Mattingly SJ, Dipersio JR, Higgins ML, Shockman GD: Unbalanced growth and macromolecular synthesis in *Streptococcus mutans* FA 1. Infect Immun 1976;13:941–948.

Morales-Aparicio JC, Lara Vasquez P, Mishra S, Barrán-Berdón AL, Kamat M, Basso KB, et al.: The Impacts of Sortase A and the 4'-Phosphopantetheinyl Transferase Homolog Sfp on *Streptococcus mutans* Extracellular Membrane Vesicle Biogenesis. Front Microbiol 2020;11:1–19.

Murray NE: Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). Microbiol Mol Biol Rev 2000;64:412–434.

Nyvad B, Crielaard W, Mira A, Takahashi N, Beighton D: Dental caries from a molecular microbiological perspective. Caries Res 2013;47:89–102.

Nyvad B, Takahashi N: Integrated hypothesis of dental caries and periodontal diseases. J Oral Microbiol 2020;12:1710953.

Paik S, Senty L, Das S, Noe JC, Munro CL, Kitten T: Identification of virulence determinants for endocarditis in *Streptococcus sanguinis* by signature-tagged mutagenesis. Infect Immun 2005;73:6064–6074.

Preza D, Olsen I, Aas JA, Willumsen T, Grinde B, Paster BJ: Bacterial profiles of root caries in elderly patients. J Clin Microbiol 2008;46:2015–2021.

Rau A, Gallopin M, Celeux G, Jaffrézic F: Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics 2013;29:2146–2152.

Schleifer KH, Kandler O: Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev 1972;36:407–477.

Serbanescu MA, Cordova M, Krastel K, Flick R, Beloglazova N, Latos A, et al.: Role of the *Streptococcus mutans* CRISPR-Cas systems in immunity and cell physiology. J Bacteriol 2015;197:749–761.

Sorek R, Kunin V, Hugenholtz P: CRISPR - A widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol 2008;6:181–186.

Suntharalingam P, Senadheera MD, Mair RW, Levesque CM, Cvitkovitch DG: The LiaFSR system regulates the cell envelope stress response in *Streptococcus mutans*. J Bacteriol 2009;191:2973–2984.

Takahashi N, Nyvad B: The Role of Bacteria in the Caries Process: Ecological Perspectives. J Dent Res 2011;90:294–303.

Takenaka S, Edanami N, Komatsu Y, Nagata R, Naksagoon T, Sotozono M, et al.: Periodontal pathogens inhabit root caries lesions extending beyond the gingival margin: A next-generation sequencing analysis. Microorganisms 2021;9. DOI: 10.3390/microorganisms9112349

Vadeboncoeur C, Pelletier M: The phosphoenolpyruvate:sugar phosphotransferase system of oral streptococci and its role in the control of sugar metabolism. FEMS Microbiol Rev 1997;19:187–207.

van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP: Gene co-expression analysis for

functional classification and gene-disease predictions. Brief Bioinform 2018;19:575–592.

Van Wely KHM, Swaving J, Freudl R, Driessen AJM: Translocation of proteins across the cell envelope of Gram-positive bacteria. FEMS Microbiol Rev 2001;25:437–454.

Wang Q, Luo Q, Yang Z, Zhao Y-H, Li J, Wang J, et al.: Weighted gene co-expression network analysis identified six hub genes associated with rupture of intracranial aneurysms. PLoS One 2020;15:e0229308.

Wen ZT, Jorgensen AN, Huang X, Ellepola K, Chapman L, Wu H, et al.: Multiple factors are involved in regulation of extracellular membrane vesicle biogenesis in *Streptococcus mutans*. Mol Oral Microbiol 2021;36:12–24.

Yamada T, Takahashi-Abbe S, Abbe K: Effects of oxygen on pyruvate formate-lyase in situ and sugar metabolism of *Streptococcus mutans* and *Streptococcus sanguis*. Infect Immun 1985;47:129–134.

Zeng L, Chen L, Burne RA: Preferred Hexoses Influence Long-Term Memory in and Induction of Lactose Catabolism by *Streptococcus mutans*. Appl Environ Microbiol 2018; 84(14):e00864-18

Zhang B, Horvath S: A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4:Article17.

Zhang M, Zheng Y, Li Y, Jiang H, Huang Y, Du M: Acid-resistant genes of oral plaque microbiome from the functional metagenomics [Internet]. . J Oral Microbiol 2018;10. Available from: http://dx.doi.org/10.1080/20002297.2018.1424455

Zhang K, Ou M, Wang W, Ling J: Effects of quorum sensing on cell viability in *Streptococcus mutans* biofilm formation. Biochem Biophys Res Commun 2009;379:933–938.

Figure Legends

Fig. 1. *S. mutans* m-RNA read count in SRS and RC. Log10 plotted graph. The median read count for SRS (1,123.89) was statistically lower than for RC (1,441,964.47) according to Mann-Whitney Rank Sum Test ($p<0.001$).

Fig. 2. Principal Component Analysis of gene expression variability between SRS and RC samples with the corresponding 95% confidence ellipse of the SRS and RC conditions.

Fig. 3. Consensus network analysis and modules of gene co-expression. **(a)** Hierarchical clustering tree and co-expression analysis. Different colors represent different modules. **(b)** Module eigengenes-trait (SRS or RC) relationship. Correlation coefficient along with *p*-value in parenthesis underneath; **(c)** heatmap of SRS key-module across SRS and RC; **(d)** heatmap of RC key-module across SRS and RC. Sample RC_7 was excluded from the analysis as an outlier.

Fig. 4. Module Membership (MM) vs gene significance (GS) and network construction. Scatterplots of the correlations between MM and GS values for RC key-module **(a)** and SRS key-module **(b)**. Dotted-lines retangules on A abd B graphs represent hub genes in each module. **(c)** Network visualization of genes in RC key-module. **(d)** Network visualization of genes in SRS key-module. Diamond nodes denote hub genes. Dotted-line circles on Figure Crepresent highly correlated genes. **(c1)** *man*A, *leu*B and *ack*A; **(c2)** SMU_75, *htp*X and *glt*X.

Fig. 5. KEGG pathway analysis considering all co-expressed genes **(a-b)** and only the hub genes **(c-d)** in SRS key-module and in RC key-modules. The abscissa represents the rich factor; the ordinate represents the enriched pathway term