



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/83999/>

Version: Accepted Version

---

**Article:**

Do, T, Sheehy, EC, Mulli, T et al. (2015) Transcriptomic analysis of three *Veillonella* spp. present in carious dentine and in the saliva of caries-free individuals. *Frontiers in Cellular and Infection Microbiology*, 5. 25.

<https://doi.org/10.3389/fcimb.2015.00025>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## **Transcriptomic analysis of three *Veillonella* spp. present in carious dentine and in the saliva of caries-free individuals.**

Thuy Do, Evelyn C Sheehy, Tonnie Mulli, Francis Hughes and David Beighton

<b>Journal Name:</b>	Frontiers in Cellular and Infection Microbiology
<b>ISSN:</b>	2235-2988
<b>Article type:</b>	Original Research Article
<b>First received on:</b>	27 Jun 2014
<b>Revised on:</b>	04 Mar 2015
<b>Frontiers website link:</b>	<a href="http://www.frontiersin.org">www.frontiersin.org</a>

## 1 ABSTRACT

2  
3 *Veillonella* spp. are predominant bacteria found in all oral biofilms. In this study, a  
4 metatranscriptomic approach was used to investigate the gene expression levels of three oral  
5 *Veillonella* spp. (*V. parvula*, *V. dispar* and *V. atypica*) in whole stimulated saliva from caries-free  
6 volunteers and in carious lesions (n=11 for each group). In the lesions the greatest proportion of  
7 reads were assigned to *V. parvula* and genes with the highest level of expression in carious  
8 samples were those coding for membrane transport systems. All three *Veillonella* spp.  
9 increased expression of genes involved in the catabolism of lactate and succinate, notably the  
10 alpha- and beta-subunits of L(+)-tartrate dehydratase (EC 4.2.1.32). There was also  
11 significantly increased expression of histidine biosynthesis pathway in *V. parvula*, suggesting  
12 higher intra-cellular levels of histidine that could provide intra-cellular buffering capacity and,  
13 therefore, assist survival in the acidic environment. Various other systems such as potassium  
14 uptake systems were also up regulated that may aid in the survival and proliferation of *V.*  
15 *parvula* in carious lesions.

## 16 17 INTRODUCTION

18  
19 *Veillonella* are obligate anaerobic Gram-negative small cocci isolated from the oral cavity and  
20 intestinal tract of humans and animals that gain energy from the utilization of short-chain  
21 organic acids, particularly lactate and succinate (Delwiche et al., 1985). The human *Veillonella*  
22 are *Veillonella parvula*, *V. atypica*, *V. dispar*, *V. montpellierensis*, *V. denticariosi* and *V. rogosae*  
23 (Mays et al., 1982; Rogosa 1984; Jumas-Bilak et al., 2004; Byun et al., 2007; Arif et al., 2008).  
24 The predominant *Veillonella* species on the tongue were *V. rogosae*, *V. atypica* and *V. dispar*  
25 (Beighton et al., 2008; Mashima et al., 2011). *V. parvula* has often been detected as the  
26 predominant *Veillonella* species isolated from active occlusal carious lesions (Arif et al.,  
27 2008; Beighton et al., 2008). Based on these studies, each *Veillonella* species seems to occupy  
28 different intra-oral habitats with limited degree of overlap between species. With the pH of  
29 carious lesions reported to be below 5 (Hojo et al., 1994), the bacteria's ability to colonise and  
30 proliferate in such an environment necessitates them to exhibit a phenotype characterised by  
31 acid resistance. The objective of this study was to determine and compare the transcriptome of  
32 three of the predominant human oral *Veillonella* (*V. parvula*, *V. dispar* and *V. atypica*) present in  
33 caries lesions and in the saliva of caries-free individuals.

34 Many bacterial genome sequence data are now publicly available, making it possible to exploit  
35 the opportunities offered by next generation sequencing (NGS) approaches to determine the *in*  
36 *vivo* expression of specific bacterial genes of individual species present in mixed-population  
37 biofilms. The short reads obtained from NGS can be aligned to bacterial genomes, enabling  
38 transcriptomic analysis of species without the need for species-specific protocols, as is  
39 necessary with the micro-array approach. The functional potential of the oral microbiome has

40 been investigated using metagenomic approaches in which genomic DNA is extracted,  
41 sequenced and the resulting sequences annotated by comparison to extant complete and  
42 partial genome sequences (Belda-Ferre et al., 2012;Luo et al., 2012). To investigate gene  
43 expression, the metatranscriptome of an individual species within a natural biofilm may be  
44 determined using RNA sequencing (RNA-seq). The application of RNA-seq to the study of  
45 bacterial transcriptomes has been reviewed by (Pinto et al., 2011) and (McLean, 2014). Several  
46 studies have also recently described the use of RNA-Seq as a tool to investigate the oral  
47 microbiome in health and disease (Duran-Pinedo et al., 2014;Jorth et al., 2014) as well as  
48 interrogate specific metabolic pathways in oral bacterial species *in vitro* (Zeng et al., 2013).  
49 In this study, we adopted a metatranscriptomic approach to investigate the level of genes  
50 expressed by the three *Veillonella* in both active carious lesions and saliva of caries-free  
51 subjects, in order to observe metabolic activities occurring in their natural environment, which  
52 may give an insight into their intra-oral distribution.

53

## 54 **MATERIALS & METHODS**

55

### 56 **Samples collection and RNA isolation**

57 Ethical approval was obtained for the collection of carious lesions (n=11) and saliva (n=11)  
58 samples. All subjects (n=22) gave informed consent prior to collection of the clinical material.  
59 Extracted teeth with large occlusal soft, active carious lesions were obtained from patients  
60 attending dental clinics at Guy's Hospital dental surgery. The teeth were immediately placed in  
61 5 ml RNAprotect® Bacteria Reagent (Qiagen) and transferred to the laboratory. The superficial  
62 biofilm was carefully removed and discarded. The infected soft dentine was collected using  
63 sterile excavators, and placed in 1 ml RNAprotect reagent, disaggregated, centrifuged (4°C at  
64 10,000xg) and the pellets stored at -80°C. Whole mouth wax-stimulated saliva samples,  
65 collected for 5 mins, were obtained from caries-free volunteers who refrained from eating for at  
66 least 2 hours prior to sampling. Immediately after collection, RNAprotect reagent was added to  
67 the saliva (1:1 v/v), the samples were centrifuged and the pellets stored at -80°C until further  
68 processing. Total RNA was extracted using the UltraClean® Microbial RNA isolation kit (MOBIO  
69 Laboratories, Inc.), including a DNase treatment step using the RNase-Free DNase Set  
70 (Qiagen) prior RNA elution.

71

### 72 **cDNA synthesis and library preparation for high-throughput sequencing**

73 A minimum of 100 ng of total RNA was extracted from each clinical sample. The total RNA was  
74 processed using reagents provided in the Illumina® TruSeq™ RNA Sample Preparation Kit.  
75 Briefly, the RNA extracts were further purified, and fragmented. First and second strands cDNA  
76 were synthesised with Superscript II Reverse Transcriptase (Invitrogen). End repair was  
77 performed on the nucleic acid fragments, 3' ends were adenylated and adapter indexes ligated.

78 The processed cDNA were amplified and further purified, prior library validation with the Agilent  
79 DNA 1000 Bioanalyzer (Agilent Technologies) and dsDNA BR Qubit assays (Invitrogen). The  
80 resulting libraries were processed for cluster generation using the TruSeq paired end cluster kit  
81 v.2, Illumina Inc., and an equimolar amount of each library was run in a separate flowcell lane.  
82 Paired end sequencing was then carried out using a Genome Analyzer Ix Illumina platform to  
83 produce 76bp reads.

84

#### 85 **Data handling and gene expression analyses**

86 A FASTQ file was obtained for each of the 22 cDNA libraries. Initial checks of the sequencing  
87 read base qualities were done via the local server provided by the Genomics facilities at Guy's  
88 Hospital Biomedical Research Centre, the data were then imported into the CLC Genomics  
89 Workbench software (CLC Bio, Qiagen). Within the CLC environment, adapter sequences were  
90 removed and for each sample file a short read mapping was performed simultaneously against  
91 144 annotated oral bacterial genomes which were previously imported from various databases  
92 (the DNA Data Bank of Japan, NCBI, the Broad Institute and HMD databases)  
93 (Supplementary file 5). The read mapping was carried out using the RNA-Seq analysis package  
94 default settings (mismatch cost: 2, insertion cost: 3, deletion cost: 3, length fraction: 0.8, and  
95 similarity fraction: 0.8; with the maximum number of hits for a read set to 1) within the CLC  
96 software, which employs the CLC Assembly Cell (CLC3) read mapper  
97 (<http://www.clcbio.com/products/clc-assembly-cell/>).

98 In this study, we are concerned with reads that mapped to 3 Veillonella strains: *V. parvula*  
99 DSM2008, *V. dispar* ATCC 17748 and *V. atypica* ACS 0049 V Sch6 only (Table 1). In order to  
100 facilitate comparison between these Veillonella strains, a RAST annotation (Rapid Annotation  
101 using Subsystem Technology) (Aziz et al., 2008) was carried out on their genomes and used  
102 with the other oral strains in the read mapping. All of the 22 sequence data files were processed  
103 for read mapping against the 144 oral genomes. Results were exported as excel files containing  
104 raw read counts determined for each of the genes from the 144 oral strains (total of 351,456  
105 genes) (Supplementary file 1). In order to compare expression levels between the 22 biological  
106 samples, the raw count data were gathered into a single excel spreadsheet for normalization  
107 (Supplementary file 1). The read counts were scaled by determining the effective library size of  
108 each sample, using the estimateSizeFactors and counts accessor functions within the  
109 Bioconductor R package DESeq (Anders and Huber, 2010), which provided an output table  
110 displaying normalized expression values for each gene and for each of the 22 samples. Data  
111 corresponding to the 3 Veillonella strains were manually extracted from the spreadsheet and  
112 used separately for further analysis to infer on their gene expression levels in caries lesions and  
113 caries-free saliva samples. Median values were calculated for both caries and saliva sample  
114 groups (n=11 each) (Supplementary file 2), which we called relative median expression (RME)  
115 values. The RME values of identical genes found in the 3 Veillonella strains were summed and  
116 ranked from highest to lowest values, to observe the most highly expressed Veillonella

117 transcripts in caries and caries-free saliva samples (Supplementary file 2). The gene identities  
118 were obtained from the RAST annotations and was supplemented by BLAST searching within  
119 Uniprot (<http://www.uniprot.org/>), InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) and  
120 PATRIC (<http://patricbrc.org/>) when necessary.

121 The raw read count data from all 144 oral strains were also used to carry out differential gene  
122 expression analysis between both sample groups using the statistical software R package  
123 DESeq2 (Love et al., 2014) based on the negative binomial model. Differential expression  
124 analysis results for the 3 *Veillonella* strains were manually extracted from the total R result  
125 outputs into excel spreadsheets, and the largest negative and positive Log<sub>2</sub> Fold Change  
126 values, with adjusted p values (padj) < 10<sup>-3</sup> were considered as significant.

127 The supplementary file 1 contains the raw count input information used for the DESeq and  
128 DESeq2 analyses.

129

### 130 **Sequence data accession numbers**

131 RNA-Seq sequencing data are available from the National Center for Biotechnology  
132 Information (NCBI) Sequence Read Archive; biosamples accession numbers for this study are  
133 SRS741215 and SRS752041.

134

135

## 136 **RESULTS & DISCUSSION**

137

### 138 **Analysis of read count and ecological considerations**

139 Here we have determined gene expression levels by mapping reads to bacterial species which  
140 form part of the oral microbial populations. The total number of mapped reads ranged between  
141 25,593,022 and 88,238,546 for the caries-free saliva samples and between 20,088,245 and  
142 32,910,299 for the caries samples (Supplementary file 1). In the carious lesions, 16.62±11.17  
143 per cent, 2.18±1.13 per cent and 0.91±0.43 percent of the mapped reads were assigned to *V.*  
144 *parvula*, *V. dispar* and *V. atypica*, respectively, compared with 4.76±7.21, 7.08±5.07 and  
145 4.09±3.47 in the saliva samples (all p<0.05) (Table 1). The pattern of the distribution of reads  
146 mirrored the reported distribution of these three species based on cultivable bacterial studies  
147 (Arif et al., 2008;Beighton et al., 2008). Belda-Ferre et al. (2012) also reported *V. parvula* to be  
148 the most predominant species in biofilm infecting dentine, with 166 contigs (>500 bp) assigned  
149 to *V. parvula* from their metagenomic data.

150 The major environmental factors affecting the *Veillonella* strains in the carious lesions and in  
151 saliva are suspected to be the low pH and availability of organic acids (lactate and succinate)  
152 required for the generation of ATP. The acidic environment within carious lesions is unlikely to  
153 be homogenous despite lactic acid being the major organic acid present (Palmer et al., 2006),  
154 resulting in areas that might be more alkaline (i.e. pH>6). Nevertheless, it should be expected  
155 that the concentration of organic acids in saliva is less than that of carious lesions, since  
156 subjects had refrained from eating for 2h prior sample collection, hence organic acids and

157 dietary components should have cleared from the mouth. Moreover, we should emphasise that  
158 the microbiota present in wax-stimulated saliva is likely to derive from the intra-oral mucosal  
159 surfaces and from the supra-gingival plaque, providing an average composition of intra-oral  
160 surfaces, but mostly of the tongue surface (Simon-Soro et al., 2013). These ecological aspects  
161 have been taken into account and explain the differences in metabolic activities occurring within  
162 the Veillonella species in both sample groups.

163

### 164 **Gene expression analysis**

165 The combined expression level of the 3 Veillonella species was determined for each condition.  
166 The relative mean expression (RME) values for each identical gene product were added and  
167 the top 30 most highly expressed gene products in saliva were ranked. The corresponding  
168 values for the caries samples are also displayed together in Figure 1.

169 Overall, the 3 Veillonella species present in the caries and saliva samples display a similar  
170 profile of transcripts. *V. parvula* expressed more genes in the caries samples, whereas *V.*  
171 *dispar* expressed more genes in the saliva samples (Table 1, Supplementary file 2).

172 The most abundant transcripts were related to the production of cell surface proteins (RME =  
173 13175), outer membrane synthesis (S-layer proteins, RME=6592), translation elongation factors  
174 (G and Tu, RME=11860 and 7731 respectively), transport systems (RME=5481), ribosomal  
175 subunit proteins (protein biosynthesis, RME=3028), and carbohydrate metabolism (particularly  
176 the glyoxylate and dicarboxylate metabolism, EC 4.2.1.32, EC 6.4.1.2, EC 1.1.1.37; RME=  
177 2147, 2126 and 1461 respectively). These results are consistent with those described by  
178 (Peterson et al., 2014) in plaque biofilm. Similarly, (Benítez-Páez et al., 2014) found evidence of  
179 overrepresentation of translation functions, together with high expressions of elongation factors  
180 Tu and G, emphasising their importance and involvement in oral biofilm formation especially in  
181 early biofilms.

182 We also report high levels of transcripts encoding membrane transport proteins (cadmium-  
183 exporting ATPase, RME= 3861; autotransporter adhesin, RME= 5481; ABC transporters, RME=  
184 1998), as well as transcripts involved in oxidative stress protection (rubrerythrin, RME=1577;  
185 and alkyl hydroperoxide reductase protein C, EC 1.6.4.-, RME=1700), in both caries and saliva  
186 groups (Supplementary file 2). The overall similarity in transcription profiles in both sample  
187 groups suggest that the selected Veillonella species are actively expressing genes that are  
188 involved in cellular maintenance and survival within diverse environments.

189

### 190 **Differential expression analysis**

191 Differential gene expression between the caries and saliva groups was investigated using the R  
192 package DESeq2 (Love et al., 2014).

193 Sample to sample distances were calculated within the DESeq2 package. The principal  
194 components analysis and the heatmap of Euclidian distance between samples were based on  
195 the metatranscriptomic data mapped to 144 oral strains, and show caries and saliva samples to

196 form distinct clusters. The PCA plot displays larger differences between saliva samples than  
197 between caries samples (Figure 2), suggesting that metabolic functions in the caries lesions are  
198 more conserved than in the caries-free samples. Likewise, the heatmap indicates the overall  
199 similarity between samples of the same group, with the exception of saliva sample number 9  
200 (H9 in Figure 3), which seems to cluster with the caries samples, indicating that it shares similar  
201 functions found in caries.

202 Jorth et al. (2014) found more similar functional features in microbiota associated with disease  
203 compared to health-associated microbiota, even though great variations in the oral microbial  
204 composition were observed between and within patients. Other papers have described inter-  
205 patients variations in terms of bacterial profiles, and these seem to reduce in diversity when  
206 changing from healthy to a disease status (Munson et al., 2004;Preza et al., 2008). Our data  
207 suggest that in the caries lesions, metabolic functions in the 3 *Veillonella* species are more  
208 similar, than in caries-free saliva samples.

209 In order to identify the main functional differences between the caries lesions and saliva  
210 samples, output data from the DESeq2 analysis were sorted according to the log<sub>2</sub> fold change  
211 values (Supplementary file 3). Since the transcriptomic data (n=22) were analysed with the  
212 caries-free vs caries condition (used as the default DESeq2 condition setting), negative log<sub>2</sub>  
213 fold change values, with corresponding Benjamini-Hochberg (BH) adjusted p values (padj) <  
214 10<sup>-3</sup> considered as significant (Benjamini and Hochberg, 1995), indicate genes with the  
215 strongest down-regulation in saliva (or strongest up-regulation in caries). Conversely, the  
216 largest log<sub>2</sub> fold change values, with corresponding significant BH padj < 10<sup>-3</sup>, indicate genes  
217 which are the most differentially expressed in saliva. Only the top 15 genes in both conditions  
218 are displayed in Table 2, and ranked according to the log<sub>2</sub>FoldChange values. A heatmap was  
219 also constructed within the DESeq2 package, and displays the top 30 differentially expressed  
220 genes across all 22 samples for the 3 *Veillonella* species (Supplementary file 4).

221 Genes that were differentially expressed in caries lesions (padj<10<sup>-3</sup>) were those expressed by  
222 *V. parvula*, and were mainly involved in pyruvate metabolism, transferases and membrane  
223 transport systems (including the biosynthesis of efflux pump components, ABC transporter and  
224 sulfur carrier proteins) (Table 2), inferring a role of these functions in disease. Similar findings  
225 were reported by Benitez-Paez et al. (2014) who found that ABC transporters were significantly  
226 up-regulated in mature biofilms, with cell motility function associated with bacterial chemotaxis,  
227 whereas Duran-Pinedo et al. (2014) reported significant levels of ABC transporters in  
228 periodontitis samples that seem associated with high levels of expression of virulence factors.  
229 Other specific pathways associated with disease have also been reported. In the case of  
230 periodontitis, a significant enrichment in butyrate production was detected (Jorth et al., 2014),  
231 iron acquisition and membrane synthesis have also been described as important metabolic  
232 activities defining disease (Duran-Pinedo et al., 2014).

233

234 However, in our data all 3 *Veillonella* species (especially *V. parvula*) expressed genes involved  
235 in glyoxylate and dicarboxylate metabolism, and alanine aspartate and glutamate metabolism,  
236 in particular genes encoding the alpha- and beta-subunits of L(+)-tartrate dehydratase (EC  
237 4.2.1.32). These are involved in the production of ATP through catabolism of lactate and  
238 succinate. Overall, the data suggest that all species responded to growth in the carious lesions  
239 by increasing the expression of many genes associated with the utilization of lactate and  
240 succinate with the consequent generation of ATP via the sodium ion-translocating  
241 methylmalonyl-CoA decarboxylase (Buckel, 2001). We also found significant up-regulation of  
242 genes encoding aspartate aminotransferases (Vpar\_1105, Vpar\_0075, HMPREF9321\_0571,  
243 HMPREF9321\_1684) in both caries and saliva samples (Supplementary file 3); these enzymes  
244 catalyse the reaction L-aspartate + 2-oxoglutarate into oxaloacetate + L-glutamate and may be  
245 an alternative method of entering intermediates into the lactate metabolic pathway, for  
246 producing ATP.

247 Genes involved in histidine metabolism were also up-regulated in caries by *V. parvula*, but not  
248 in the other 2 species (Supplementary file 3). Of particular importance is the up-regulation of  
249 ATP phosphoribosyltransferase (EC 2.4.2.17) which has a central role in histidine biosynthesis.  
250 Similar up-regulation was observed in *Corynebacterium glutamicum* and *Salmonella*  
251 *typhimurium*, as well as in *Lactobacillus casei*, in response to acid adaptation (20 min at pH 4.5)  
252 (Foster, 1995; Brockmann-Gretza and Kalinowski, 2006; Broadbent et al., 2010). It was  
253 suggested that the up-regulation of the histidine operon resulted in increased intra-cellular  
254 levels of His which may contribute to intracellular buffering capacity as the pK<sub>a</sub> value of the  
255 imidazole groups of histidine and histidine-containing peptides is near 6.0 and these have been  
256 shown to contribute to intracellular buffering in vertebrate cells (Abe, 2000).

257 Additionally, a potassium uptake system in *V. parvula* (Vpar\_1334 and Vpar\_1335; KtrA and  
258 KtrB) was also significantly up-regulated in caries, but not by the other 2 species. K<sup>+</sup> uptake in  
259 prokaryotes is essential for maintenance of cytoplasmic pH (Csonka and Epstein, 1996; Stumpe  
260 et al., 1996), this system may also assist in the survival of *V. parvula* in the acidic environment  
261 of the carious dentine. In *V. dispar* and *V. dispar*, these genes were significantly up-regulated in  
262 saliva, which may explain their lower ability to control their intracellular pH in the caries lesions.  
263 Clearly, *V. parvula* exhibits several distinct systems for intracellular pH control which do not  
264 appear to function as well in either *V. atypica* or *V. dispar*, and this may explain the ability of *V.*  
265 *parvula* to be better fitted to growth and proliferation in the acidic environment of carious lesions  
266 compared to the other two species.

267 Most of the differentially expressed genes in the saliva samples are those expressed by *V.*  
268 *atypica* and *V. dispar*, and encode for the oligopeptide, sulfonate transporter systems, and  
269 cysteine and methionine metabolism (EC 2.1.1.10). Others include genes involved in purine  
270 metabolism (EC 3.6.1.11), ferrichrome and other transport systems, molybdenum cofactor  
271 biosynthesis, as well as oxidoreductases which involve the use of NAD<sup>+</sup> or NADP<sup>+</sup> as acceptor

272 in the chemical reaction leading to the formation of siroheme from uroporphyrinogen III (EC  
273 1.3.1.76) (Table 2).

274 General stress response genes have also been identified in the carious lesions and saliva  
275 samples in all 3 *Veillonella* species (Supplementary file 3). Several genes encoding heat shock  
276 and chaperonin proteins were found up-regulated in caries (Vpar\_1034, Vpar\_1035,  
277 Vpar\_0881), and others up-regulated in saliva (VEIDISOL\_01212, HMPREF9321\_0106,  
278 VEIDISOL\_01142). Stress proteins such as chaperonin heat shock protein 33  
279 (HMPREF9321\_0536) and putative peroxide-responsive repressor PerR (HMPREF9321\_0995)  
280 were up-regulated by *V. atypica* in the saliva samples (Supplementary file 3). In *V. parvula* and  
281 *V. dispar*, several genes associated with extracellular S-layer formation, were significantly up-  
282 regulated, which is a well characterized stress-associated response (Xiao et al., 2012).

283

## 284 **CONCLUSION**

285

286 Recent reports of metagenomic and metatranscriptomic analyses of oral samples are providing  
287 extensive information on the microbial populations and functions characterising health and  
288 disease. These studies have also confirmed previous culturable observations regarding the  
289 intra-oral distribution of particular species but also found novel taxa which have not previously  
290 been identified amongst cultured bacteria and phyla for which only limited cultivated isolates are  
291 extant.

292 Here we have applied a metatranscriptomic approach to study 3 predominant oral *Veillonella*  
293 spp. in their natural habitat and have shown that their gene expression profiles are overall  
294 similar in both caries lesions and saliva (caries-free) samples. However, through differential  
295 expression analysis, *V. parvula* seems to exhibit a distinct method of intra-cellular pH control  
296 not evident in the other two species investigated, which might explain the preponderance of *V.*  
297 *parvula* in carious lesions and the reduced ability of *V. atypica* and *V. dispar* to proliferate in this  
298 acid environment. Other important functions related to membrane transport systems are  
299 reported to be over-expressed in the caries lesions inferring a role in disease.

300 We have shown here that RNA-Seq is a powerful technique that can be used to observe the  
301 transcriptome of selected species or strain, provided their genome sequence data are available.  
302 The obvious drawbacks from such technique relate to the limited number of reference genomes  
303 available for reads mapping, and also to the fact that non-core genome sequences are not  
304 captured using the current methodology. Further analyses including larger samples and  
305 samples from similar biofilms such as plaque instead of saliva would be beneficial to add to our  
306 understanding of the oral microbial functions during initiation and development of disease.

307

308

309 **ACKNOWLEDGEMENTS.**

310 This research was supported by the Dental Institute, King's College London and the  
311 Department of Health via the National Institute for Health Research (NIHR) comprehensive  
312 Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in  
313 partnership with King's College London and King's College Hospital NHS Foundation Trust. The  
314 funders had no role in study design, data collection and analysis, decision to publish, or  
315 preparation of the manuscript. The authors declare no potential conflicts of interest with respect  
316 to the authorship and/or publication of this article.

317

## 318 REFERENCES

319

320

- 321 Abe, H. (2000). Role of histidine-related compounds as intracellular proton buffering constituents in  
322 vertebrate muscle. *Biochemistry (Mosc)* 65, 757-765.
- 323 Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol*  
324 11, R106.
- 325 Arif, N., Do, T., Byun, R., Sheehy, E., Clark, D., and Gilbert, S.C. (2008). *Veillonella rogosae* sp. nov., an  
326 anaerobic, Gram-negative coccus isolated from dental plaque. *Int J Syst Evol Microbiol* 58, 581-  
327 584.
- 328 Aziz, R.K., Bartels, D., Best, A.A., Dejongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass,  
329 E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., Mcneil, L.K.,  
330 Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein,  
331 V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems  
332 technology. *BMC Genomics* 9, 75.
- 333 Beighton, D., Clark, D., Hanakuka, B., Gilbert, S., and Do, T. (2008). The predominant cultivable  
334 *Veillonella* spp. of the tongue of healthy adults identified using *rpoB* sequencing. *Oral*  
335 *Microbiol Immunol* 23, 344-347.
- 336 Belda-Ferre, P., Alcaraz, L.D., Cabrera-Rubio, R., Romero, H., Simón-Soro, A., Pignatelli, M., and Mira, A.  
337 (2012). The oral metagenome in health and disease. *ISME J* 6, 46-56.
- 338 Benítez-Páez, A., Belda-Ferre, P., Simón-Soro, A., and Mira, A. (2014). Microbiota diversity and gene  
339 expression dynamics in human oral biofilms. *BMC Genomics* 15, 311.
- 340 Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful  
341 Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57, 289-300.
- 342 Broadbent, J.R., Larsen, R.L., Deibel, V., and Steele, J.L. (2010). Physiological and transcriptional  
343 response of *Lactobacillus casei* ATCC 334 to acid stress. *J Bacteriol* 192, 2445-2458.
- 344 Brockmann-Gretza, O., and Kalinowski, J. (2006). Global gene expression during stringent response in  
345 *Corynebacterium glutamicum* in presence and absence of the *rel* gene encoding (p)ppGpp  
346 synthase. *BMC Genomics* 7, 230.
- 347 Buckel, W. (2001). Sodium ion-translocating decarboxylases. *Biochim Biophys Acta* 1505, 15-27.
- 348 Byun, R., Carlier, J.P., Jacques, N.A., Marchandin, H., and Hunter, N. (2007). *Veillonella denticariosi* sp.  
349 nov., isolated from human carious dentine. *Int J Syst Evol Microbiol* 57, 2844-2848.
- 350 Csonka, L.N., and Epstein, W. (1996). "Osmoregulation.," in *Escherichia coli and Salmonella: cellular and*  
351 *molecular biology* eds. F. Neidhardt, R. Curtis Iii, J. Ingraham, E. Lin, K. Low & E.A. Magasanik.  
352 (Washington DC: American Society for Microbiology Press), 1210-1223.
- 353 Delwiche, E.A., Pestka, J.J., and Tortorello, M.L. (1985). The veillonellae: gram-negative cocci with a  
354 unique physiology. *Annu Rev Microbiol* 39, 175-193.
- 355 Duran-Pinedo, A.E., Chen, T., Teles, R., Starr, J.R., Wang, X., Krishnan, K., and Frias-Lopez, J. (2014).  
356 Community-wide transcriptome of the oral microbiome in subjects with and without  
357 periodontitis. *ISME J* 8, 1659-1672.
- 358 Foster, J.W. (1995). Low pH adaptation and the acid tolerance response of *Salmonella typhimurium*. *Crit*  
359 *Rev Microbiol* 21, 215-237.

360 Hojo, S., Komatsu, M., Okuda, R., Takahashi, N., and Yamada, T. (1994). Acid profiles and pH of carious  
361 dentin in active and arrested lesions. *J Dent Res* 73, 1853-1857.

362 Jorth, P., Turner, K.H., Gumus, P., Nizam, N., Buduneli, N., and Whiteley, M. (2014). Metatranscriptomics  
363 of the human oral microbiome during health and disease. *MBio* 5, e01012-01014.

364 Jumas-Bilak, E., Carlier, J.P., Jean-Pierre, H., Teyssier, C., Gay, B., Campos, J., and Marchandin, H. (2004).  
365 *Veillonella montpellierensis* sp. nov., a novel, anaerobic, Gram-negative coccus isolated from  
366 human clinical samples. *Int J Syst Evol Microbiol* 54, 1311-1316.

367 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for  
368 RNA-seq data with DESeq2. *Genome Biol* 15, 550.

369 Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. (2012). Direct comparisons of  
370 Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample.  
371 *PLoS One* 7, e30087.

372 Mashima, I., Kamaguchi, A., and Nakazawa, F. (2011). The distribution and frequency of oral veillonella  
373 spp. in the tongue biofilm of healthy young adults. *Curr Microbiol* 63, 403-407.

374 Mays, T.D., Holdeman, L.V., Moore, W.E.C., Rogosa, M., and Johnson, J.L. (1982). Taxonomy of the  
375 genus *Veillonella*. *Int J Syst Bacteriol* 32.

376 Mclean, J.S. (2014). Advancements toward a systems level understanding of the human oral  
377 microbiome. *Front Cell Infect Microbiol* 4, 98.

378 Munson, M.A., Banerjee, A., Watson, T.F., and Wade, W.G. (2004). Molecular analysis of the microflora  
379 associated with dental caries. *J Clin Microbiol* 42, 3023-3029.

380 Palmer, R.J.J., Diaz, P.I., and Kolenbrander, P.E. (2006). Rapid succession within the *Veillonella*  
381 population of a developing human oral biofilm in situ. *J Bacteriol.* 188, 4117-4124.

382 Peterson, S.N., Meissner, T., Su, A.I., Snesrud, E., Ong, A.C., Schork, N.J., and Bretz, W.A. (2014).  
383 Functional expression of dental plaque microbiota. *Front Cell Infect Microbiol* 4, 108.

384 Pinto, A.C., Melo-Barbosa, H.P., Miyoshi, A., Silva, A., and Azevedo, V. (2011). Application of RNA-seq to  
385 reveal the transcript profile in bacteria. *Genet Mol Res* 10, 1707-1718.

386 Preza, D., Olsen, I., Aas, J.A., Willumsen, T., Grinde, B., and Paster, B.J. (2008). Bacterial profiles of root  
387 caries in elderly patients. *J Clin Microbiol* 46, 2015-2021.

388 Rogosa, M. (1984). "Anaerobic Gram-negative cocci," in *Bergey's Manual of Systematic Bacteriology*,  
389 ed. N. Baltimore. Williams & Wilkins), 680-685.

390 Simon-Soro, A., Tomas, I., Cabrera-Rubio, R., Catalan, M.D., Nyvad, B., and Mira, A. (2013). Microbial  
391 geography of the oral cavity. *J Dent Res* 92, 616-621.

392 Stumpe, S., Schlosser, A., Schleyer, M., and Bakker, E. (1996). "K<sup>+</sup> circulation across the prokaryotic cell  
393 membrane: K<sup>+</sup>-uptake systems. Transport processes in eukaryotic and prokaryotic organisms,"  
394 in *Handbook of biological physics*, eds. K. Konings, H. Kaback & J. Lolkema. (Amsterdam, The  
395 Netherlands.: Elsevier), 473-499.

396 Xiao, J., Klein, M.I., Falsetta, M.L., Lu, B., Delahunty, C.M., Yates Iii, J.R., Heydorn, A., and Koo, H. (2012).  
397 The Exopolysaccharide Matrix Modulates the Interaction between 3D Architecture and  
398 Virulence of a Mixed-Species Oral Biofilm. *PLoS Pathog* 8, e1002623.

399 Zeng, L., Choi, S.C., Danko, C.G., Siepel, A., Stanhope, M.J., and Burne, R.A. (2013). Gene regulation by  
400 CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. *PLoS One* 8,  
401 e60465.

402  
403  
404  
405  
406  
407  
408  
409

410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448

**Figure 1:** Relative median expression (RME) levels in 3 Veillonella strains (*V. parvula* DSM2008, *V. dispar* ATCC 17748 and *V. atypica* ACS 0049 V Sch6). RME were calculated from the median values of normalized read counts in the caries (n=11) and saliva (n=11) samples. The 30 highest RME values were sorted in ascending order for the genes in saliva samples and are displayed with the RME values of corresponding genes in caries samples.

**Figure 2:** Principal components analysis plot displaying sample-to-sample distances for caries and saliva samples. The PCA plot is based on the differential expression analysis of 144 oral bacterial strains, carried out using the R package DESeq2 (Love et al., 2014).

**Figure 3:** Heatmap of Euclidian distances between samples (n=22). The heatmap was constructed using the R package DESeq2 (Love et al., 2014), and is based on the differential expression analysis of 144 oral bacterial strains.

**Table 1:** Characteristics of the 3 Veillonella strains selected in this study, with their relative proportions in caries and saliva metatranscriptomes.

**Table 2:** Up-regulated genes in the caries samples (top of table) and in the saliva samples (bottom of table). Genes expressed with the strongest down-regulation in the saliva samples (or up-regulation in the caries lesions) and genes with the strongest up-regulation in saliva samples were determined using the R package DESeq2 (Love et al., 2014). The list of genes is ranked according to the Log2FoldChange values from the negative lowest values (strongest down-regulation in saliva) to the positive highest values (strongest up-regulation). The baseMean corresponds to the average of the normalized count values (divided by size factors), the log2FoldChange corresponds to the effect size estimate indicating the change in gene expression between both sample groups; lfcSE corresponds to the standard error of the

449 log2FoldChange estimate, and padj corresponds to the Benjamini & Hochberg adjusted p-  
450 values.

451

452 **Supplementary file 1:** Read count data for the 144 oral strains (including *V. parvula* DSM2008,  
453 *V. dispar* ATCC 17748 and *V. atypica* ACS 0049 V Sch6), used as input file for analyses using  
454 DESeq and DESeq2 in R.

455

456 **Supplementary file 2:** Relative median expression (RME) values based on the median and  
457 75th percentile (Q3) values of normalized read counts in the caries (n=11) and saliva samples  
458 (n=11), for the 3 Veillonella species.

459

460 **Supplementary file 3:** Data output from the differential expression analysis for the 3 Veillonella  
461 species, obtained from DESeq2 analysis.

462

463 **Supplementary file 4:** Heatmap constructed using the R package DESeq2, displaying the  
464 highest most variable genes across all samples (n=22), the analysis was based on the 3  
465 Veillonella strains only.

466

467 **Supplementary file 5:** List of reference strains used in the CLC Genomics Workbench (CLC  
468 Bio, Qiagen) for RNA-Seq analysis.

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488  
489  
490  
491  
492  
493  
494  
495

496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530

Table 1

<b>Feature code</b>	<b>Veillonella species</b>	<b>strain</b>	<b>Number of CDS (from RAST annotation)</b>	<b>Relative proportion (%) of transcripts in caries (n=11) ±SD</b>	<b>Relative proportion (%) of transcripts in saliva (n=11) ±SD</b>
<b>HMPREF</b>	<i>V. atypica</i>	ACS-049-V-Sch6	1840	0.91±0.43	4.09±3.47
<b>VEIDISOL</b>	<i>V. dispar</i>	ATCC 17748	1954	2.18±1.13	7.08±5.07
<b>Vpar</b>	<i>V. parvula</i>	DSM 2008	1904	16.62±11.17	4.76±7.21

531

532 Table 2

Feature ID	Gene product	baseMean	log2Fold Change	lfcSE	padj
Vpar_1291	L(+)-tartrate dehydratase beta subunit (EC 4.2.1.32)	357.26	-8.90	0.72	2E-31
Vpar_1292	L(+)-tartrate dehydratase alpha subunit (EC 4.2.1.32)	646.86	-7.83	0.74	2E-22
VEIDISOL_00680	L(+)-tartrate dehydratase beta subunit (EC 4.2.1.32)	84.70	-6.60	0.81	1E-13
VEIDISOL_00681	Possible membrane transport protein	5.72	-6.28	1.09	2E-07
Vpar_0720	hypothetical protein	2.99	-6.17	1.32	4E-05
Vpar_1308	Ornithine carbamoyltransferase (EC 2.1.3.3)	41.52	-5.84	1.03	3E-07
Vpar_0455	Sulfur carrier protein ThiS	6.55	-5.73	1.18	2E-05
Vpar_1307	N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)	36.32	-5.51	1.02	1E-06
Vpar_1306	Acetylglutamate kinase (EC 2.7.2.8)	28.84	-5.28	1.11	3E-05
Vpar_0164	FIG01197475: hypothetical protein	20.37	-5.28	1.04	7E-06
Vpar_1004	Alpha-aspartyl dipeptidase Peptidase E (EC 3.4.13.21)	18.93	-5.26	1.12	3E-05
Vpar_0330	FIG01197189: hypothetical protein	159.84	-5.00	0.80	2E-08
VEIDISOL_00679	L(+)-tartrate dehydratase alpha subunit (EC 4.2.1.32)	63.25	-4.96	0.94	3E-06
Vpar_1022	Putative ATP:guanido phosphotransferase (EC 2.7.3.-)	114.18	-4.74	0.92	5E-06
Vpar_1367	RND efflux system, outer membrane lipoprotein, NodT family	40.60	-4.63	0.86	1E-06
HMPREF9321_0616	Siroheme synthase / Precorrin-2 oxidase (EC 1.3.1.76)	6.02	6.78	1.66	3.8E-04
VEIDISOL_01296	Cold shock protein CspC	5.35	6.80	1.63	2.8E-04
HMPREF9321_1331	Ferrichrome transport ATP-binding protein FhuC (TC 3.A.1.14.3)	14.49	6.86	1.21	3.4E-07
HMPREF9321_0811	FIG002958: hypothetical protein	5.70	6.90	1.62	2.0E-04
HMPREF9321_0294	Small-conductance mechanosensitive channel	5.82	6.94	1.62	1.7E-04
HMPREF9321_1565	Exopolyphosphatase (EC 3.6.1.11)	5.67	6.94	1.61	1.6E-04
HMPREF9321_0879	Homocysteine S-methyltransferase (EC 2.1.1.10)	18.40	7.04	1.21	1.5E-07
HMPREF9321_1453	Molybdenum cofactor biosynthesis protein MoaB	6.44	7.10	1.60	9.3E-05
VEIDISOL_01157	Sodium-dependent transporter	7.78	7.28	1.59	5.8E-05
HMPREF9321_0702	FIG01197118: hypothetical protein	8.58	7.33	1.60	5.4E-05
HMPREF9321_0668	NAD(P)HX epimerase / NAD(P)HX dehydratase	8.15	7.34	1.59	4.4E-05
HMPREF9321_0246	Mobile element protein	11.60	7.74	1.56	1.1E-05
HMPREF9321_1665	binding-protein-dependent transport systems inner membrane component	13.47	7.88	1.56	7.5E-06
VEIDISOL_00207	Alkanesulfonates/ Sulfonate ABC transporter, ATP-binding protein	14.99	7.96	1.57	6.2E-06
HMPREF9321_1134	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (TC 3.A.1.5.1)	17.60	8.24	1.53	1.6E-06

533

534

535

Figure 1.JPEG

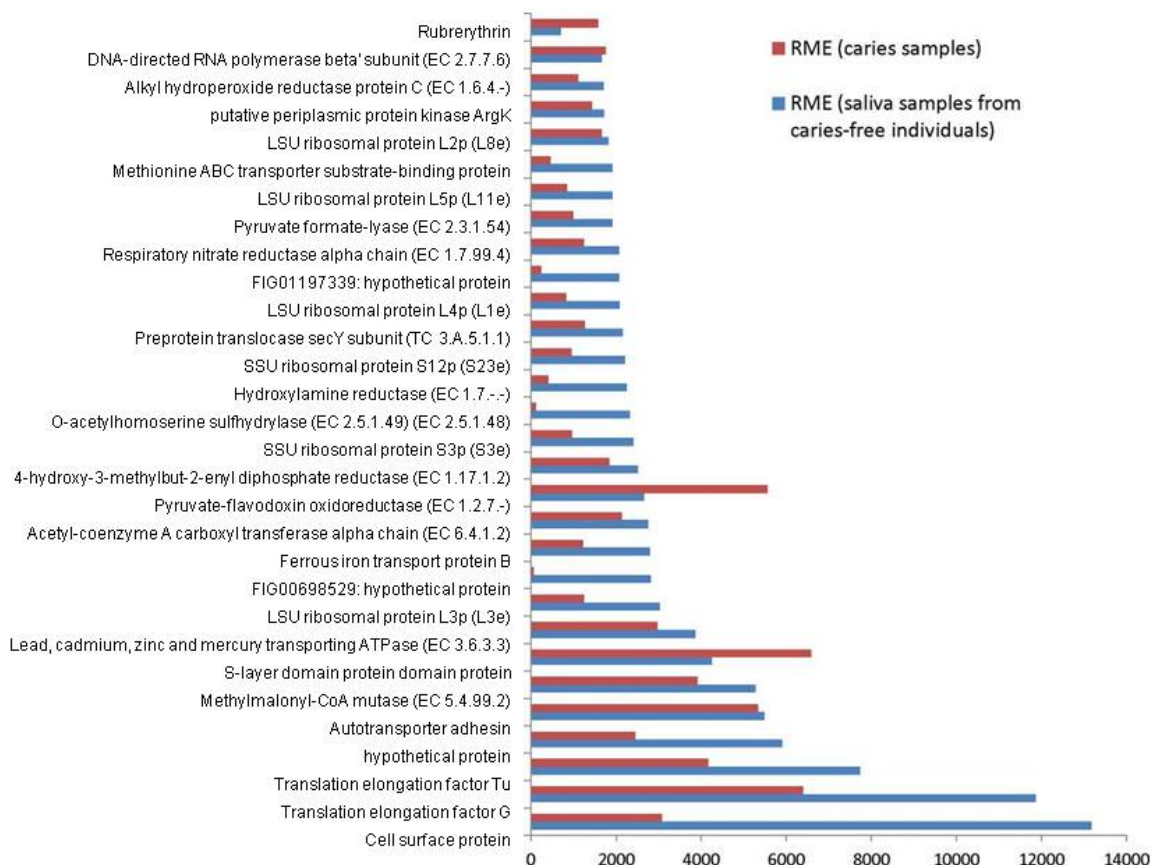


Figure 2.JPEG

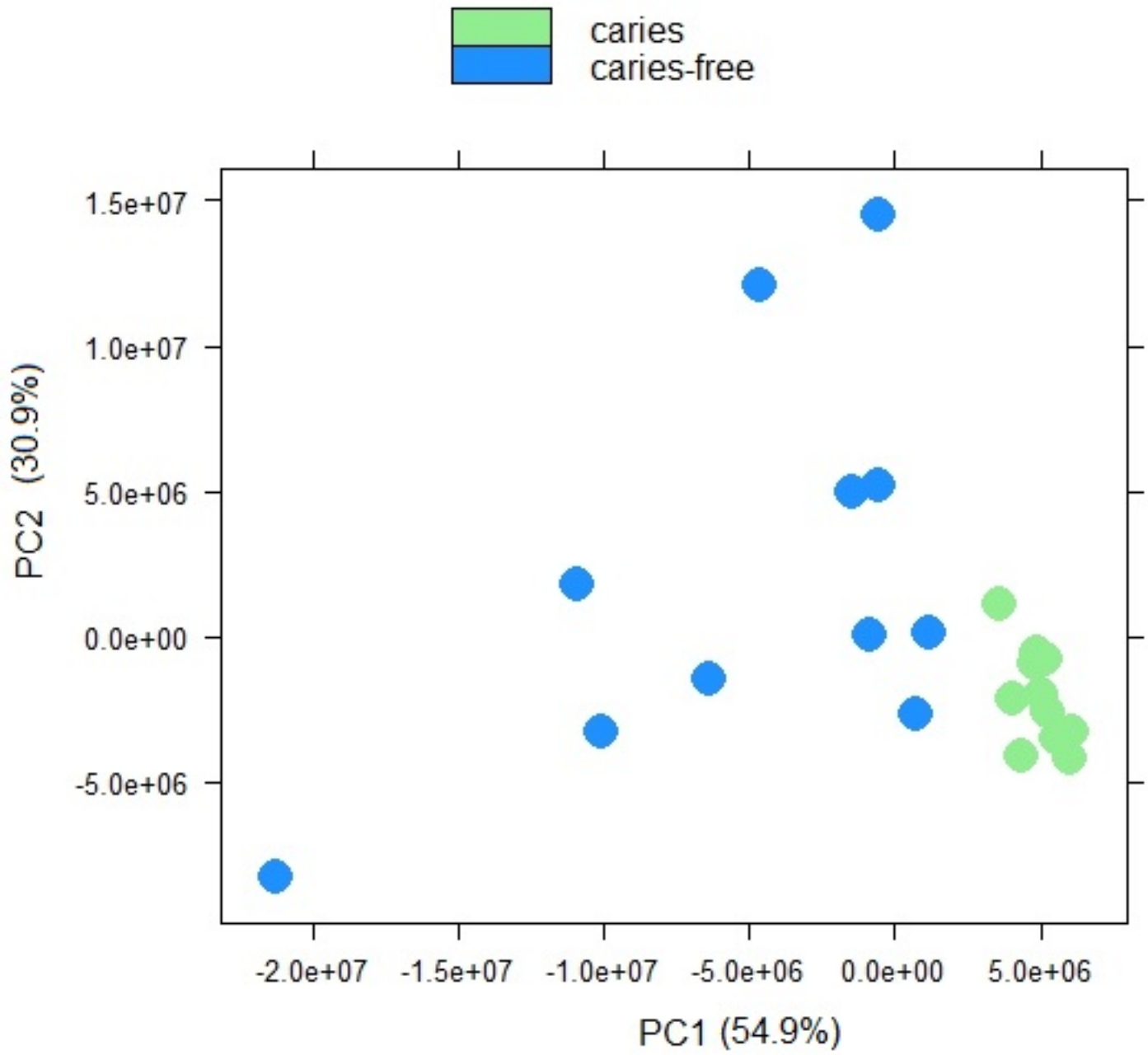


Figure 3.JPEG

