This is the published version of a paper published in *Scientific Reports*.

Citation for the original published paper (version of record):

Carbonic Anhydrase 6 Gene Variation influences Oral Microbiota Composition and Caries Risk in Swedish adolescents
*Scientific Reports*, 9: 452
https://doi.org/10.1038/s41598-018-36832-z

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-155838
Carbonic Anhydrase 6 Gene Variation influences Oral Microbiota Composition and Caries Risk in Swedish adolescents

A. Esberg1, S. Haworth2,3, C. Brunius4, P. Lif Holgerson5 & I. Johansson1

Carbonic anhydrase VI (CA6) catalyses the reversible hydration of carbon dioxide in saliva with possible pH regulation, taste perception, and tooth formation effects. This study assessed effects of variation in the CA6 gene on oral microbiota and specifically the acidophilic and caries-associated Streptococcus mutans in 17-year old Swedish adolescents (n = 154). Associations with caries status and secreted CA6 protein were also evaluated. Single Nucleotide Polymorphisms (27 SNPs in 5 haploblocks) and saliva and tooth biofilm microbiota from Illumina MiSeq 16S rDNA (V3-V4) sequencing and culturing were analysed. Haploblock 4 (rs10864376, rs3737665, rs12138897) CCC associated with low prevalence of S. mutans (OR (95% CI): 0.5 (0.3, 0.8)), and caries (OR 0.6 (0.3, 0.9)), whereas haploblock 4 TTG associated with high prevalence of S. mutans (OR: 2.7 (1.2, 5.9)) and caries (OR: 2.3 (1.2, 4.4)). The TTG-haploblock 4 (represented by rs12138897(G)) was characterized by S. mutans, Scardovia wiggsiae, Treponema sp. HOT268, Tannerella sp. HOT286, Veillonella gp.1 compared with the CCC-haploblock 4 (represented by rs12138897(C)). Secreted CA6 in saliva was weakly linked to CA6 gene variation. In conclusion, the results indicate that CA6 gene polymorphisms influence S. mutans colonization, tooth biofilm microbiota composition and risk of dental caries in Swedish adolescents.

Carbonic anhydrases (CAs) are zinc enzymes that catalyse the reversible hydration of carbon dioxide in the reaction CO₂ + H₂O ⇌ HCO₃⁻ + H⁺ and are important for pH homeostasis in body tissues and fluids and for removal of intracellular carbon dioxide1. In mammalian cells, CAs are involved in several biological processes, including HCO₃⁻ dependent metabolic processes, secretion of electrolytes, respiration, pH regulation, bone resorption, biomineralization, and odontogenesis2. Today, 16 different CA isozymes have been identified in humans1, of which 13 have been shown to be enzymatically active. Some isozymes are expressed in most tissues, while others are tissue- or organ-specific. Five isozymes are cytosolic (I, II, III, VII and XIII), five are membrane-bound (IV, IX, XII, XIV and XV), two are present in the mitochondria (VA and VB), and one (VI, CA6) is a secreted isozyme3,4.

CA6 is secreted into saliva with considerable individual variation in concentration and activity5, and incorporated in the protein pellicle on tooth enamel and in dental biofilms6–8. CA6 supports neutralization of lactic and other bacteria-produced acids through conversion of saliva HCO₃⁻ to water and carbon dioxide and pH maintenance through the subsequent phase buffering9. CA6 is therefore suggested to be an important enzyme in oral physiology and tooth tissues integrity, i.e., resistance to caries and dental erosion10. However, data are conflicting as both high and low concentrations and activities of CA6 have been associated with caries4,11,12. Single nucleotide polymorphisms (SNPs) and multisite haploblocks in the CA6 gene have been linked to CA6 concentration and activity in saliva and to caries status13–17, but findings are not consistent18. However, no study has confirmed the association between CA6 in saliva and S. mutans colonization in humans and no study has investigated CA6 polymorphism and levels of caries associated bacteria or overall oral microbiota.

1Department of Odontology/Section of Cariology, Umeå University, Umeå, Sweden. 2Medical Research Council Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom. 3Bristol Dental School, University of Bristol, Bristol, United Kingdom. 4Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. 5Department of Odontology/Section of Pedodontics, Umeå University, Umeå, Sweden. Correspondence and requests for materials should be addressed to A.E. (email: anders.esberg@umu.se)
### Table 1. Characteristics of the 17-year-old participants all together and by S. mutans and caries status. Caries information was missing for 1 participant who was excluded. Continuous measures are presented as medians (10, 90 percentiles) and group differences were tested with Mann Whitney U test. Differences between group numbers (presented as %) were tested with Chi2 or Fisher’s exact test. P-values were considered significant at FDR < 0.25 but significant differences at FDR < 0.06 are also indicated in bold. CFU for colony forming units and NGS for Next Generation Sequencing of DNA extracted from saliva (n = 152) or tooth biofilm (n = 139).

<table>
<thead>
<tr>
<th>Variable</th>
<th>All participants</th>
<th>S. mutans status by NGS</th>
<th>Caries status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 154)</td>
<td>S. mutans free (n = 38)</td>
<td>S. mutans present (n = 116)</td>
</tr>
<tr>
<td>Gender, % male</td>
<td>44.2</td>
<td>50.0</td>
<td>42.2</td>
</tr>
<tr>
<td>Smoker, %</td>
<td>2.6</td>
<td>0</td>
<td>3.5</td>
</tr>
<tr>
<td>Tooth brushing twice daily, %</td>
<td>83.8</td>
<td>89.5</td>
<td>81.9</td>
</tr>
<tr>
<td>Sweet snack frequency/day</td>
<td>1.1 (0.5, 2.7)</td>
<td>1.2 (0.6, 3.8)</td>
<td>0.9 (0.4, 3.6)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>21.6 (18.5, 26.0)</td>
<td>22.4 (18.2, 26.7)</td>
<td>21.4 (18.6, 25.9)</td>
</tr>
</tbody>
</table>

#### Dental status

- Number of tooth surfaces: 128 (108, 128) for Caries-free, 128 (128, 129) for Caries-affected, with p-value <0.001.
- DeFS: 2.0 (0.170) for Caries-free, 3.0 (0.0, 18.4) for Caries-affected, with p-value <0.001.
- S. mutans: 75.3% for Caries-free, 61.4% for Caries-affected, with p-value <0.001.

#### Saliva

- Protein concentration, mg/mL: 0.6 (0.3, 0.9) for Caries-free, 0.5 (0.3, 0.9) for Caries-affected, with p-value <0.001.
- Flow rate, ml/min: 1.5 (0.7, 2.5) for Caries-free, 1.4 (0.6, 2.5) for Caries-affected, with p-value 0.066.
- CA6 protein in saliva: 0.6 (0.3, 0.9) for Caries-free, 0.6 (0.3, 0.9) for Caries-affected, with p-value 0.070.

The primary aim of this study was to evaluate the effects of genetic variation in the CA6 gene region on oral microbiota and specifically the acidophilic and caries-associated S. mutans in Swedish adolescents. Secondary aims were to evaluate associations between variation in the CA6 gene region and secreted CA6 protein and caries status.

### Results

#### Participant characteristics

Using next-generation sequencing (NGS), 75.3% of participants had detectable S. mutans, with significantly higher proportions among the caries-affected than caries-free subjects (p < 0.001; Table 1). Overall, the median caries score (DeFS) was 2.0 tooth surfaces and 67% of participants were affected by caries (DeFS ≥ 1) and 33% were caries-free (DeFS = 0) (Table 1). There were no significant differences between those with detectable S. mutans versus not or those who were caries-affected versus not with respect to sex, smoking, sweet snacking, BMI, and number of tooth surfaces. The proportion who reported brushing the teeth twice a day did not differ between those with S. mutans versus not, but it was significantly lower among caries-affected than caries-free participants (p < 0.001; Table 1).

#### CA6 single nucleotide polymorphic sites and gene structure

After quality control measures, 27 out of 30 Single Nucleotide Polymorphisms (SNPs) were retained for analyses. Two SNP (rs2274333 and rs17032942) were excluded due to sample call rate of 0%. An additional SNP (rs6697763) was excluded from analysis for deviation from Hardy-Weinberg equilibrium (p = 0.033), leaving 27 SNPs for further analyses.

Evaluating potential multisite haploblocks within or around the CA6 gene resulted in five haploblocks (Fig. 1). Haploblock 1, comprised two SNPs (rs6688840 and rs2781087) and was located upstream of the first coding exon (+6387 to +6505bp), whereas Haploblock 5 comprised rs6577546 and rs6680186 and was located downstream of the stop codon (−3871 to −4553bp) of the CA6 gene. Haploblocks 2, 3, and 4 where within the CA6 gene; haploblock 2 with four SNPs (rs6577541, rs2274327, rs2274328, and rs17032907) covered the 2477 to 4512bp region; haploblock 3 with three SNPs (rs7545200, rs1832262, and rs6691526) covered the 12,141 to 15,160bp region; and haploblock 4 with three SNPs (rs10864376, rs3737665, and rs12138897) covered the 24,479 to 26,010bp region of the CA6 gene.

#### CA6 gene variation and S. mutans colonization by NGS

Five SNPs, i.e., rs10864376 (T), rs3737665 (T), rs12138897 (G), rs3765964 (A), and rs6680186 (A) were associated with increased odds of detectable S.
**S. mutans** in saliva or tooth biofilm by NGS (Table 2). Different genetic models (allelic-, dominant- and recessive model) yielded qualitatively similar conclusion, but with some variation in test statistics (Fig. 2).

Further, the CCC-haploblock 4 and the AG-haploblock 5 associated with reduced odds to have **S. mutans**, whereas the GCAC-haploblock 2, and TCG and TTG haploblock 4 associated with increased odds (Table 2). The odds ratio (95% CI) to have **S. mutans** in saliva or tooth biofilm if carrying the TCG allele versus the CCC haploblock 4 was 5.0 (1.5–17.2) (p = 0.011), and if carrying the TTG allele versus the CCC haploblock 4 it was 3.3 (1.5–7.6) (p = 0.004).

**CA6 gene variation and viable mutans streptococci in saliva.** Four SNPs, rs10864376 (T), rs12138897 (G), rs6680186 (A), and rs7513804 (T), were associated with having mutans streptococci by saliva culturing (Table 3), as was the TCG-haploblock 4, whereas, the CCC-haploblock 4 and AG-haploblock 5 were associated with being **S. mutans** free (Table 3). The odds ratio (95% CI) to have mutans streptococci by culturing if carrying the TCG allele versus the CCC haploblock 4 was 4.3 (1.5–12.2) (p = 0.006). Notably, the association pattern with CA6 gene variation was similar for having **S. mutans** by NGS and having viable **S. mutans** by culture.

**CA6 gene variation rs12138897 (G/C) and tooth biofilm microbiota.** The association of the CCC-haploblock 4 and the TCG-haploblock 4 with **S. mutans** status (as well as caries status as described in the next paragraph) led us to evaluate the association between the C and G allele of rs12138897 (C representing the CCC- and G- the TTG haploblock) and overall tooth biofilm microbiota by multivariate Partial Least Squares (PLS) analysis of NGS determined taxa. PLS modelling (crossvalidated prediction Q2 = 24%) displayed separation of participants carrying the G versus C allele based on their tooth biofilm microbiota (Fig. 3a). The separation was explained by strong associations (Variable Importance in Projection (VIP) > 1.5) with the G allele for eight species, i.e., higher detection rate of *Treponema* sp. HOT268, *Actinomyces* sp. HOT1896, *Tannerella* sp. HOT286, **S. mutans**, GN02[G-1] sp. HOT972, *Sardonia* wiggsiae, and species detected by Veillonella genus probe1 (Fig. 3b). The C allele associated with higher detection rate of *Bacteroidales* [G-2] sp. HOT152, *Prevotella* sp. HOT315, *Tannerella* forsythia, *Fretibacterium* fastidiosum, *Filifactor* alocis, and species detected by *Treponema* genus probe 2 (Fig. 3c). Five of the eight species associating with the G allele, i.e., *Treponema* sp. HOT268, *Tannerella* sp.
Table 2. CA6 SNP and haploblock associations with S. mutans detection in both saliva and tooth biofilm by NGS. Alleles in bold indicate significant positive associations and in italic negative associations at FDR of < 0.25 (bold underline for differences significant at FDR ≤ 0.06 too). * indicates SNPs also associated with caries (see Table 4).

CA6 gene variation and caries status. Following that CA6 gene variation was associated with colonization of caries associated bacteria, such as S. mutans and S. wiggiae, the association between the 27 SNPs and caries status (yes or no) was tested in an allelic association model. The rs12021597 (A), rs10864376 (T), rs3737665 (T), and rs17032912 (G) associated positively with having caries (Table 4). The strongest effects on having caries was observed for the rs10864376 (T) allele with an odds ratio (95% CI) of 2.1 (1.2, 3.7) (p = 0.031), and rs3737665 (T) (p = 0.031) but none was significant at FDR < 0.25 (Table S3).

CA6 gene variation and saliva CA6 protein levels. Saliva concentration of CA6 was measured by ELISA in the first 90 sampled participants (for further information see material and methods). The median (10%, 90% percentile) concentration of CA6 in saliva was 8.4 (0.3, 17.0) µg/mL and secreted CA6 11.0 (0.4, 25.4) µg/min (Table 1). High concentration tended to associate with rs10864376 (T) (p = 0.031) and rs3737665 (T) (p = 0.031), as well as the GCA-haploblock 3 (p = 0.030) and TTG-haploblock 4 (p = 0.051) but none was significant at FDR < 0.25 (Table S3). CA6 in saliva (concentration or secreted amounts) was not associated with having S. mutans or not, but secreted amounts (µg/min) was significantly higher in caries-free than caries-affected (p = 0.012) (Table 1). In accordance, twice as many caries-free, compared to caries-affected, participants were identified in the highest tertile based on the distribution of secreted CA6 amounts (p between groups = 0.007) with an OR to be caries-free of 3.5 (1.4, 9.1).
Haploblock 4 summary. A summary of characteristics for the various allele variants in haploblock 4 is presented in Table 5 together with comparisons between the allele variants groups as well as trends among the allele variants. There were no significant differences between the alleles with respect to sex, BMI, smoking, frequency of sweet snacking, or number of tooth surfaces. The proportions with *S. mutans* by NGS and culturing, respectively, differed significantly between the allele variant groups ($p < 0.001$) as did the trend from the TCG to CCC variant ($p < 0.001$). The association with caries was less strong, but the proportions caries-affected participants tended to differ between the allele variant groups ($p = 0.054$) with a significant trend ($p = 0.021$), whereas caries prevalence (DeFS) did not differ between the groups though the trend did ($p = 0.018$).

Discussion

This study tested the hypothesis that variation in the CA6 gene is relevant to dental health and found that CA6 gene polymorphisms and haploblocks of CA6 were associated with *S. mutans* colonization, overall microbiota composition and dental caries in Swedish adolescents. Secreted CA6 in saliva tended to be linked to CA6 gene variation. Thereby, it is the first study to demonstrate effects of CA6 gene polymorphisms on the oral bacterial communities in a population-based setting.

Collectively, the present findings support a role of CA6 polymorphisms in oral health, i.e., balancing less, against more, aciduric species, e.g., cariogenic *S. mutans* and *S. wiggsiae*, and de- and remineralization in the caries process. The results are in line with previous studies reporting a role of CA6 in caries development in children. Though several of the physiological effects of CA6 are relevant for both oral microbiota ecology and the caries process, the detailed mechanisms for the effects remain to be elucidated. The underlying mechanisms for oral health of CA6 and CA6 gene variations may, alone or in concert, include tooth formation, local pH regulation, bacteria adhesion and metabolism and taste induced food preferences.

A tooth formation mechanism appears plausible because CA6 is expressed in the bud, cup, and bell stage of early tooth development, and by pre- and early differentiated odontoblasts and dental papilla mesenchyme cells, but most avidly in late tooth development with formation and maturation of the enamel. Differences in enamel quality affects the initial tooth covering protein pellicle, and consequently the profile of attachment epitopes for bacteria and possibly oral microbiota composition. Differences in enamel quality also affects resistance of tooth tissue to destruction in an acidic environment.

Given the importance of pH homeostasis for the oral environment, local pH regulation by CA6 may be a contributory mechanism. Several of the most abundant bacteria in the mouth are glycolytic with avid acid formation, leading to pH fall and enrichment of aciduric, cariogenic species (dysbiosis). The CA6 protein is suggested to regulate pH in secreted fluids, and saliva buffer capacity was reported to differ by CA6 gene variation at rs2274327, and CA6 protein activity in some studies, but not others, studies. This aspect could not be tested in the present study as neither saliva buffer capacity nor pH were analyzed, but some gene variants were associated with...
presence of acidic and acidogenic species, such as *S. mutans* and *S. wiggisae*, which are promoted at low pH. Since reported intake of sweets did not differ by CA6 gene variation this may reflect a less well functioning pH regulation by some CA6 gene variants. Impaired pH regulation may also increase or decrease the risk of tooth demineralisation.

Other important factors in oral biofilm ecology include access to attachment sites on host- and partner bacteria structures, and exposure to oxygen and nutrients. CA6 incorporated in pellicle and tooth biofilm may influence bacterial ecology and metabolism through effects unrelated to its enzymatic activity. Such effects may influence bacterial structures, and exposure to oxygen and nutrients. CA6 incorporated in pellicle and tooth biofilm may influence bacterial ecology and metabolism through effects unrelated to its enzymatic activity. Such effects may be twofold; i.e., by proteolytic release of bioactive peptides or by providing a bacteria binding site. Here, the TTG bases of CA6 are linked with mixed panels of oral bacterial such as *Treponema sp*. HOT268, *Treponema sp*. HOT268, and *Treponema sp*. HOT268, but *S. mutans* and *S. wiggisae*, but also *Treponema sp*. HOT268, and *Tannerella sp*. HOT286 which thrive at a neutral to slightly basic pH. However, *Treponema sp*. HOT286 and *Tannerella sp*. HOT286 which thrive at a neutral to slightly basic pH. However, these suggested mechanisms are theoretical and need to be evaluated in future studies.

Finally, CA6 gene variation has been suggested to link to bitter taste and smell perception21–27. To date, reported associations with food selection are inconsistent and the present study did not find any effect of CA6 polymorphisms on sweet snack intake. This may reflect a “true” lack of association or the inherent error in measuring sugar intake in general and especially in the dental office. This aspect as well as the reported association between bitter taste receptor stimulation and innate immunity function28 should be addressed by a study design that is less prone to bias, such as using a biomarker or genetic marker reflecting sweet intake.

Compared to previous investigations, the present study included detailed genotyping of a relatively large number of SNPs within and near CA6, allowing fine-mapping of both single genetic variants and recombination blocks. The phenotypes were also detailed, including a range of salivary, microbiome, and host characteristics which may affect oral microbiota ecology and the caries process directly or indirectly. Additional strengths were that (i) all participants were of the same age (17 years old), (ii) participants were recruited in connection with their regular dental survey which reduces the risk for selection bias, (iii) participants permanent teeth have been exposed to the oral environment for years allowing ecology stabilization and potential development caries, and (iv) have minimal confounding by diseases or medications. Further, the study design, i.e., consecutive inclusion
In this study 154 17-year-old adolescents were recruited from three public dental health care clinics in the city of Umeå, Sweden. Adolescents who had other chronic medical conditions, used medication regularly, required antibiotic treatment during the latest six months, who did not consent, or were unable to communicate in Swedish or English were excluded. All participants had answered a questionnaire with information about their use of sweet snacks, which is prone to bias, such as the well-known under-reporting of diet intake.

There are some limitations in the study that should be acknowledged. The relative low number of participants, including that only a subset was available for assessing CA6 amounts, limits the possibility of comparisons in subgroups. Further, identification of microorganisms by NGS is associated with potential errors at various steps, such as the PCR amplification, sequencing per se, and bioinformatic accuracy. It should also be noted that in contrast to tooth biofilm samples, reflects bacteria from epithelial and tooth surfaces. These differences contributed to the variation in S. mutans detection frequency between the different identification strategies. To reduce the impact of the sampling source, being S. mutans positive was based on presence in either saliva or tooth biofilm.

Materials and Methods

Participants. In this study 154 17-year-old adolescents were recruited from three public dental health care clinics in the city of Umeå, Sweden. Adolescents who had other chronic medical conditions, used medication regularly, required antibiotic treatment during the latest six months, who did not consent, or were unable to communicate in Swedish or English were excluded. All participants had answered a questionnaire with information.

### Table 4: CA6 SNP and haplblock associations with being caries-affected or caries-free.

<table>
<thead>
<tr>
<th>SNP number</th>
<th>Name</th>
<th>Allele</th>
<th>Caries-affected vs. Caries-free OR (95%CI)</th>
<th>P excl.</th>
<th>Haploblock number</th>
<th>Allele</th>
<th>Frequency</th>
<th>Caries-affected vs. Caries-free OR (95%CI)</th>
<th>P excl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs6688840</td>
<td>T/C</td>
<td>1.1 (0.5, 2.3)</td>
<td>0.895</td>
<td>Block 1</td>
<td>CT</td>
<td>0.717</td>
<td>0.9 (0.5, 1.6)</td>
<td>0.748</td>
</tr>
<tr>
<td>2</td>
<td>rs2781087</td>
<td>G/T</td>
<td>1.1 (0.6, 1.9)</td>
<td>0.749</td>
<td></td>
<td>CG</td>
<td>0.167</td>
<td>1.1 (0.6, 2.2)</td>
<td>0.784</td>
</tr>
<tr>
<td>3</td>
<td>rs1475734</td>
<td>G/T</td>
<td>1.0 (0.5, 2.1)</td>
<td>0.955</td>
<td></td>
<td>TG</td>
<td>0.116</td>
<td>1.1 (0.5, 2.3)</td>
<td>0.894</td>
</tr>
<tr>
<td>4</td>
<td>rs6692694</td>
<td>C/T</td>
<td>1.1 (0.6, 2.1)</td>
<td>0.783</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>rs17032875</td>
<td>C/T</td>
<td>1.1 (0.6, 1.8)</td>
<td>0.744</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12067941</td>
<td>C/A</td>
<td>1.2 (0.5, 2.6)</td>
<td>0.663</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>rs12568902</td>
<td>A/T</td>
<td>1.0 (0.4, 2.5)</td>
<td>0.955</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs6577541</td>
<td>G/C</td>
<td>1.1 (0.6, 1.8)</td>
<td>0.821</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs2274327</td>
<td>C/T</td>
<td>1.2 (0.7, 2.0)</td>
<td>0.514</td>
<td>Block 2</td>
<td>CCAC</td>
<td>0.348</td>
<td>1.0 (0.6, 1.7)</td>
<td>0.948</td>
</tr>
<tr>
<td>10</td>
<td>rs2274328</td>
<td>C/A</td>
<td>1.0 (0.5, 1.7)</td>
<td>0.982</td>
<td></td>
<td>GCAT</td>
<td>0.220</td>
<td>0.9 (0.5, 1.7)</td>
<td>0.828</td>
</tr>
<tr>
<td>11</td>
<td>rs17032907</td>
<td>C/T</td>
<td>1.1 (0.6, 1.9)</td>
<td>0.815</td>
<td></td>
<td>GCAC</td>
<td>0.043</td>
<td>1.7 (0.5, 6.8)</td>
<td>0.479</td>
</tr>
<tr>
<td>12</td>
<td>rs11121278</td>
<td>G/C</td>
<td>1.4 (0.6, 3.5)</td>
<td>0.468</td>
<td></td>
<td>GCCC</td>
<td>0.036</td>
<td>5.4 (0.7, 43.5)</td>
<td>0.065</td>
</tr>
<tr>
<td>13</td>
<td>rs12021597</td>
<td>A/G</td>
<td>1.9 (1.1, 3.4)</td>
<td>0.023</td>
<td></td>
<td>GCCT</td>
<td>0.011</td>
<td>1.3 (0.1, 14.4)</td>
<td>0.854</td>
</tr>
<tr>
<td>14</td>
<td>rs17032912</td>
<td>G/C</td>
<td>1.5 (0.8, 3.0)</td>
<td>0.214</td>
<td></td>
<td>ACA</td>
<td>0.062</td>
<td>1.4 (0.5, 4.3)</td>
<td>0.508</td>
</tr>
<tr>
<td>15</td>
<td>rs7545200</td>
<td>G/A</td>
<td>1.0 (0.6, 1.8)</td>
<td>0.930</td>
<td>Block 3</td>
<td>GATA</td>
<td>0.401</td>
<td>0.6 (0.4, 1.1)</td>
<td>0.063</td>
</tr>
<tr>
<td>16</td>
<td>rs1832262</td>
<td>C/T</td>
<td>1.7 (1.0, 2.8)</td>
<td>0.057</td>
<td></td>
<td>GCA</td>
<td>0.296</td>
<td>1.9 (1, 3, 3)</td>
<td>0.037</td>
</tr>
<tr>
<td>17</td>
<td>rs6691526</td>
<td>A/G</td>
<td>1.2 (0.6, 2.1)</td>
<td>0.637</td>
<td></td>
<td>ACG</td>
<td>0.240</td>
<td>0.9 (0.5, 1.6)</td>
<td>0.636</td>
</tr>
<tr>
<td>18</td>
<td>rs3765965</td>
<td>T/C</td>
<td>1.2 (0.6, 2.3)</td>
<td>0.538</td>
<td></td>
<td>ACA</td>
<td>0.062</td>
<td>1.4 (0.5, 4.3)</td>
<td>0.508</td>
</tr>
<tr>
<td>19</td>
<td>rs2274334</td>
<td>G/T</td>
<td>1.1 (0.4, 3.6)</td>
<td>0.850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>rs10864376</td>
<td>T/C</td>
<td>2.1 (1.2, 3.7)</td>
<td>0.009</td>
<td>Block 4</td>
<td>CCC</td>
<td>0.461</td>
<td>0.6 (0.3, 0.9)</td>
<td>0.028</td>
</tr>
<tr>
<td>21</td>
<td>rs3737665</td>
<td>T/C</td>
<td>2.3 (1.2, 4.4)</td>
<td>0.014</td>
<td></td>
<td>CCG</td>
<td>0.217</td>
<td>0.9 (0.5, 1.7)</td>
<td>0.776</td>
</tr>
<tr>
<td>22</td>
<td>rs12138897</td>
<td>G/C</td>
<td>1.8 (1.1, 2.9)</td>
<td>0.028</td>
<td>Block 5</td>
<td>TTG</td>
<td>0.217</td>
<td>2.3 (1, 4, 4)</td>
<td>0.014</td>
</tr>
<tr>
<td>23</td>
<td>rs3765964</td>
<td>A/G</td>
<td>1.5 (0.9, 2.4)</td>
<td>0.160</td>
<td></td>
<td>TCG</td>
<td>0.105</td>
<td>1.3 (0.6, 3.1)</td>
<td>0.518</td>
</tr>
<tr>
<td>24</td>
<td>rs6577546</td>
<td>A/G</td>
<td>0.9 (0.5, 1.7)</td>
<td>0.851</td>
<td></td>
<td>AG</td>
<td>0.475</td>
<td>0.8 (0.5, 1.3)</td>
<td>0.314</td>
</tr>
<tr>
<td>25</td>
<td>rs6680186</td>
<td>A/G</td>
<td>1.4 (0.8, 2.2)</td>
<td>0.242</td>
<td></td>
<td>AA</td>
<td>0.301</td>
<td>1.3 (0.8, 2.4)</td>
<td>0.319</td>
</tr>
<tr>
<td>26</td>
<td>rs7533317</td>
<td>G/A</td>
<td>1.1 (0.6, 2.1)</td>
<td>0.697</td>
<td></td>
<td>GA</td>
<td>0.219</td>
<td>1.1 (0.6, 2.0)</td>
<td>0.754</td>
</tr>
<tr>
<td>27</td>
<td>rs7513804</td>
<td>T/C</td>
<td>1.2 (0.7, 1.9)</td>
<td>0.573</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CA6 was originally found in the salivary glands, but has now been found to be expressed by the lacrimal, tracheobronchial, nasal, and mammary glands (with especially high content in colostrum, and in epithelial linings of the oesophagus, stomach, and large intestine. This has led to a wider search of its functions. Thereby, CA6 has been linked to acinic cell carcinoma, colorectal cancer, and Sjögrens syndrome. Based on the findings from the current study it may be hypothesized that CA6 and CA6 gene variations contribute to the microbial environment in both the nasopharynx and gastro-intestinal tracts.

In conclusion, the results from the present study support that CA6 gene polymorphisms rs10864376 (T), rs3737665 (T), rs12138897 (G) and haplblock TTG of CA6 are associated with S. mutans colonization, overall microbiota composition and dental caries in Swedish adolescents. Further studies need to elucidate the mechanisms.
on general health status, medication use, oral hygiene, dietary habits, and tobacco use, and samples of whole simulated saliva and supragingival biofilm were collected and stored at −80 °C until use.

Ethical approval. The study was approved by the Regional Ethical Review Board in Umeå, Sweden (Dnr 2012-111-31 M, Dnr 2015-389-32 M and Dnr 2017-450-31) and abided by the Declaration of Helsinki and by the European General Data Protection Regulation (GDPR). Data collections and experiments were performed in accordance with relevant guidelines and regulations, including that all adolescents and their caregivers provided informed consent and data collection and handling followed the Helsinki declaration and the Swedish Law on personal data act (PuL).

CA6 genotyping. Single nucleotide polymorphisms in the region of the CA6 gene were selected and genotyped on DNA isolated from whole stimulated saliva. The criteria for SNP selection were a) minor allele frequency (MAF) ≥ 5% in the CEU population b) not in complete linkage disequilibrium (r² ≤ 0.8) and c) in the 10,000 bp up-stream to 10,000 bp down-stream region of the CA6 gene. Using the NIH snpinfo resource (https://snpinfo.niehs.nih.gov/snpinfo/snpinfo.html) 30 SNPs met these criteria. Genotyping was performed using a multiplexed primer extension chemistry of the iPLEX assay with detection of the incorporated allele by mass spectrometry using a MassARRAY analyzer (Agena Bioscience, Hamburg, Germany). Raw data from the mass reader

Figure 3. CA6 gene variation and tooth biofilm microbiota. (a) PLS scatter plot illustrating separation of participants carrying either the G or C of the rs12138897 SNP based on their tooth biofilm microbiota using the dominant SNP model (G/G + G/C vs. C/C); (b) PLS loading scatter plot of the X and Y weights (w* and c) illustrating the relation between the microbiota species and the G (red) and C (green) genotype of CA6. Red/green dots refer to taxa with a VIP-value > 1.5, and grey dots to taxa with a VIP value < 1.5. (c) Correlation coefficients (mean (95% CI; shown on the left x-axis) for taxa with a VIP value > 1.5 (shown on the right x-axis) in plot B. Red bars show association with G and green bars with C genotypes of CA6. The dots on the lines represent median prevalence for each taxon and the stars (*) statistically significant difference between the two genotypes using univariate Chi² test and p-values < 0.05.
Culturing, were stored at −80 °C until used.

Mutants refrained from oral hygiene on the morning of their visit to the dental clinic. All samples, except those for Sweden) as previously described38. Briefly, samples were collected and lysed in buffer with lysozyme and mutanolysin from bacteria mock communities with the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, Stockholm, Sweden) protocol 42.

rDNA amplicon sequencing was conducted at the Forsyth Research Institute with the HOMING protocol 43. Taxa targeted by the genus probes (gp) are described in Table S2. Based on which targeted recognition of 538 species (using 638 species-level probes) and 129 groups of closely-related species (using 129 genus-level probes) 43. Taxa were identified with the ProbeSeq customized BLAST program, which targeted recognition of 538 species (using 638 species-level probes) and 129 groups of closely-related species (using 129 genus-level probes) 43. Taxa targeted by the genus probes (gp) are described in Table S2. Based on sequencing results for the mock communities, taxa that were represented by at least 100 sequences with >98.5% similarity were included in further analyses.

Determination of CA6 concentration in saliva. For the first 90 participants’ saliva, samples were collected and stored at −80 °C until used for CA6 concentration measurements. CA6 protein concentration in saliva was measured by ELISA (CA6 ELISA Kit, OKCD01736, Nordic Biosite, Täby, Sweden) following the manufacturer’s instructions. Briefly, 100 µL saliva, diluted 1:5 was used for analysis, CA6 protein was captured by a target

<table>
<thead>
<tr>
<th>Proportion with detection by NGS, %</th>
<th>92.6</th>
<th>87.5</th>
<th>71.4</th>
<th>63.9</th>
<th>&lt;0.001</th>
<th>&lt;0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutans streptococci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFU/mL, log10 value</td>
<td>3.0 (0, 5.2)</td>
<td>2.8 (0, 4.6)</td>
<td>2.9 (0, 5.1)</td>
<td>0.7 (0, 4.4)</td>
<td>0.021</td>
<td>0.014</td>
</tr>
<tr>
<td>Proportion with detection by culture, %</td>
<td>81.5</td>
<td>64.3</td>
<td>69.6</td>
<td>50.4</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>Caries status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeFS</td>
<td>1.0 (0, 14.0)</td>
<td>2.0 (0, 13.9)</td>
<td>1.5 (0, 12.0)</td>
<td>1.0 (0, 10.0)</td>
<td>0.101</td>
<td>0.021</td>
</tr>
<tr>
<td>Proportion caries-affected, %</td>
<td>66.7</td>
<td>75.0</td>
<td>59.3</td>
<td>53.8</td>
<td>0.054</td>
<td>0.018</td>
</tr>
</tbody>
</table>

Table 5. Characteristics of the 17-year-old participants in Haploblock 4. Continuous measures are presented as medians (10, 90 percentiles) and group differences were tested with Mann Whitney U test. Differences between group numbers (presented as %) were tested with Chi² or Fisher's exact test. P-values were considered significant at FDR < 0.25 but significant differences at FDR ≤0.06 are also indicated in bold. CFU for colony forming units and NGS for Next Generation Sequencing of DNA extracted from saliva (n = 152) or tooth biofilm (n = 139). All analysis included participants who reported brushing their teeth twice daily.
specific antibody, then detected by a biotinylated secondary antibody, followed by an avidin-horseradish peroxidase (HRP) conjugate and tetramethylbenzidine (TMB) for colour development. The colour development reaction was terminated by sulfuric acid addition and the optical density (OD) measured at 450 nm using a MultiscanGO spectrophotometer (Thermo Scientific, Abinolab, Upplands-Väsby, Sweden). A standard curve using purified CA6 protein ranging from 0–1,000 pg/mL was run.

**Determination of total saliva protein concentration.** Total saliva protein concentration was determined using Pierce Coomassie (Bradford) protein assay kit (ThermoFisher). Briefly, 5 µL diluted (1:1) saliva, was mixed with 250 µL protein dye solution and incubated for 30 min at room temperature. The optical density of each sample at 595 nm wavelength was assessed using a MultiscanGO spectrophotometer (Thermo Scientific) and compared to the optical density of a bovine serum albumin standard curve from known antigen concentrations to determine the sample concentration.

**Data handling and statistical inference.** NGS sequencing of saliva was performed in 139 tooth biofilm samples and 154 saliva samples where 2 samples failed. Caries information was missing for one participant who was excluded. All analyses with caries as outcome were restricted to the 129 subjects who reported brushing their teeth twice a day.

SPSS software version 24.0 was used for descriptive statistics, including medians (10%, 90% percentiles), frequencies (n), proportions (%) and odds ratios (OR) with 95% confidence intervals (CI). Group comparisons of continuous variables were done using Mann-Whitney U test and for categorical variables Chi² or Fisher’s exact test. All tests were two-tailed. Correction for multiple comparison was done by the Benjamini Hochberg false discovery rate (FDR). P-values were considered significant at FDR < 0.25 and ≤ 0.06. The higher FDR was applied to avoid missing potential CA6 SNPs of interest as described earlier.44,45

Haplovieview software (version 4.2) was used to evaluate characteristics of SNPs (observed heterozygosity, predicted heterozygosity, Hardy-Weinberg equilibrium, minor allele frequency, pairwise linkage disequilibrium), as well as evaluation of potential haploblocks using the algorithm described by Gabriel et al.46. Haploblocks with frequencies < 1% were not included in the analyses. Association between genetic variation (SNPs and haploblocks) and phenotypes was evaluated using Haplovie and Chi2 tests. The tests were two-tailed tests and corrected for multiple comparison by FDR as described above. Phenotype penetrance was assessed in common allele, recessive and dominant models.47

Clustering of participants by presence (0/1) of bacterial taxa in saliva and tooth biofilm, respectively, was modelled using partial least square (PLS) regression. Saliva taxa did not yield a significant model and therefore PLS analyses employed tooth biofilm taxa. Variables with a Variable Importance in the Projection (VIP) > 1.5 were considered influential. SIMCA P+ version 14.1 (Umetrics AB) was used for these analyses.

**Data Availability** NGS sequence data can be found at [https://doi.org/10.6084/m9.figshare.5794989](https://doi.org/10.6084/m9.figshare.5794989). Genotyping CA6 datasets generated and analysed during the current study are available from the corresponding author on reasonable request and with appropriate ethical approvals.

**References**


Acknowledgements

The study was supported by the Swedish Patent Revenue Foundation, and TUA, the County Council of Västerbotten (Sweden). None of the funding bodies had any influence on the design, data collection, analysis, interpretation, or writing of the manuscript. Carina Öhman and Agnetha Rönnlund are acknowledged for skilful laboratory work and Linda Eriksson for clinical examinations.

Author Contributions

Esberg A. and Johansson I. contributed to conception, design and data analysis, drafted and critically revised the manuscript. Haworth S., Lif Holgersson P. and Brunius C. contributed to data analysis and critically revised the manuscript; All authors gave final approval and agree to be accountable for all aspects of the work.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-36832-z.

Competing Interests: The authors declare no competing interests.