The Caries and Caries-free Archaeome

Naile Dame-Teixeira[1], Jack Lynch[2], Xia Yu[2], Jessica Alves de Cena[1], Thuy Do[2]

[1] Department of Dentistry, School of Health Sciences, University of Brasilia, Brasilia, Brazil

[2] Division of Oral Biology, School of Dentistry, University of Leeds, Leeds, UK

Short Title: Archaea signals in dental caries

Abstract word count: 300

Total word count (Abstract to Conclusion): 3,200

Total number of tables/figures: 5

Number of references: 38

Keywords: *Archaea* domain, Oral Archaeome, Methanogens, Oral microbiome, Dental Caries, bioinformatic analysis.

Abstract

Aim: The difficulty of establishing a relationship between archaea and oral diseases such as dental caries stems from the challenges of detecting, identifying, and isolating these microorganisms. This study aimed to detect archaea in publicly available datasets comprising caries and caries-free saliva and dental plaque by using a tailored bioinformatic pipeline for shotgun sequencing analysis.

Methods: A systematic search was performed to identify studies using shotgun metagenomics or metatranscriptomics on samples obtained from individuals with dental caries. Two reviewers selected studies based on eligibility criteria. Sequencing and metadata from each study were retrieved from their SRA Bioproject. A count table was generated for each database by mapping reads against an archaea genome database, specifically tailored for this study, using stringent filtering parameters of above 97% similarity and 90% query coverage. Archaeal prevalence was determined using an arbitrary cut-off point (>500 reads). An effect size meta-analysis was performed to determine the overall prevalence. Phyloseq and DESeq2 packages were used to determine alpha and beta diversities, differential abundance in different taxonomic levels, and differential expression comparing caries and caries-free samples. Spearman correlation was performed with the bacteriome.

Results: The search yielded 154 titles, from which a collection of 7 datasets from 8 studies was obtained. N=63 samples out of 397 samples were positive for archaea using a post-filtering, comprising a putative prevalence of 20% (CI = 0-40%), and identifying *Euryarchaeota*, *Thermoplasmatota*, and *Nitrosphaeria*. Methanogens were present in both caries and caries-free groups (*Methanobrevibacter* spp., *Methanosarcina* and *Methanosphaera*), and positively correlated with *Strephotrophomonas*, *Streptococcus*, *Actinomyces*, *Abiotrophia*, *Gemella*, and *Corynebacterium*. Several methanogenesis genes, including Methyl-coenzyme M reductase, which catalyzes the final step of methane production in methanogens, were underexpressed in caries-active samples compared to caries-free samples.

Conclusion: Saliva and dental plaque emerged as sites of low-abundance archaea, with methanogenesis genes underexpressed in caries-active samples.

Introduction

Woese *et al*(1990) revolutionized our understanding of life on Earth by proposing a 3-domain system, classifying living organisms into Bacteria, Archaea and Eucarya(Woese et al. 1990). Prior to this, owing to their prokaryote classification, archaea were taxonomically grouped with bacteria. The discovery of important distinctions, such as the absence of peptidoglycan cell walls and the ribosomal RNA sequences, prompted the recognition of the need for this reclassification. There was an initial belief that all archaea were exclusively extremophiles. However, the later detection of archaea in several environments, including in humans(Koskinen et al. 2017), supported the hypothesis that archaea are ubiquitous and recyclers of organic and inorganic matter. We recently reviewed the history of the *Archaea* domain along with a summary of the biogeographic patterns of the human archaeome(Cena et al. 2024).

It is already known that members of the *Archaea* domain constitute a minor portion of the oral microbiome(Belmok et al. 2020), with extensive literature in archaeal association with periodontal diseases(Cena et al. 2022). Indeed, small proportions of DNA sequences from methanogenic archaea have been identified in ancient dental calculus; for example, a nearly complete genome of *Methanobrevibacter oralis* was discovered in a Neanderthal(Weyrich et al. 2017), indicating that the archaea have been part of the human/humanoid oral microbiome for a long time. Evidence suggests that their prevalence has diminished significantly due to changes in the diet and lifestyle of *Homo sapiens*(Gancz et al. 2023), although studies have demonstrated that archaea are present in at least 40% of subgingival biofilms associated with periodontitis in modern samples(Cena et al. 2022).

There is limited evidence of archaea presence and potential functions in dental caries, potentially representing fundamental gaps in understanding cross-domain interactions in either homeostasis or dysbiosis status(Butcher et al. 2022). In a pilot study, we have shown the presence of DNA sequences belonging to the *Archaea* domain in dentinal caries lesions, including low-abundant *Thaumarchaeota* and methanogenic archaea(Dame-Teixeira et al. 2019). Archaea likely have implications in countless biogeochemical processes, including the nitrogen and carbon cycles, as well as increasing the efficiency of bacterial metabolism by contributing to energy flux using cross-feeding strategies(Oren and Garrity 2021). While *Thaumarchaeota* (recently reclassified as *Nitrososphaerota*, with the majority comprised by ammonia-oxidizers) could be disturbing a natural buffer system, potentially influencing the supragingival biofilm balance, methanogenic archaea have an energy metabolism related to methane production(able to generate methane as

a byproduct of their energy metabolism processes). These archaea play a role in the carbon cycle, with some species capable of thriving under microoxic conditions(Seedorf et al. 2004).

The difficulty of establishing a relationship between archaea and oral diseases such as dental caries stems from their identification challenges due to low-abundance (underreported low-abundant microbiota less than 1% of relative abundance in next generation sequencing-NGS studies)(Cena et al. 2021), primers mismatches(Pausan et al. 2019), unsuitable DNA extraction methods, incomplete reference databases often missing archaeal sequences, and the lack of clinical interest on archaea as no archaeal pathogens are yet identified(Cena et al. 2024). Also, metagenomic bioinformatic pipelines aims at defining the core microbiome often overlooking rare microbial taxa(Cena et al. 2021). Therefore, this study aims to develop a rigorous bioinformatic pipeline to identify archaea in shotgun sequencing datasets related to dental caries. To investigate the presence of *Archaea* domain members in dental caries, we systematically selected datasets and reanalyzed their sequences using our tailored pipeline.

Methods

Included studies comprised those that employed either DNA shotgun metagenomics or cDNA metatranscriptomics in samples of individuals with dental caries, with or without a control group. Excluded studies were *in vitro*, without available datasets, or involving systemic diseases or conditions that alter the microbiota (e.g., Sjogren's, severe hyposalivation, and others). More details including the question design and eligibility criteria are in **Supplementary Table-1**. A comprehensive systematic search was performed on PubMed in April 2023(**Supplementary table-2**). Two reviewers independently selected studies. The availability of sequencing data and metadata from each study were retrieved from their SRA Bioproject. A flowchart detailing the process of identification, inclusion, and exclusion of studies is shown in **Figure 1A.**

Bioinformatics and metanalysis

Figure 1B illustrates the tailored pipeline used in our archaeal analysis. Further details can be found in the Methods section of the **Supplementary file-1**. A full non-redundant archaea database was specifically tailored for this study, containing 519 archaeal genomes(**Supplementary file-2)**. After downloading each dataset and its corresponding

metadata from NCBI, and processing the raw read data using fastqc and cutadapt (quality cutoff was set to trim 10 base pairs at both 5' and 3' from each read before adapter removal; reads shorter than 30 base pairs were discarded), a count table was generated for each dataset. Reads were mapped against our archaea genome-database using the DIAMOND v2.0.15.153 alignment tool. Stringent filtering parameters were applied: 97% sequence similarity and 90% query coverage(Buchfink et al. 2021).

Archaeal prevalence was determined using a post-filtering cut-off, considering samples with over 500 archaeal genus counts. The overall prevalence was meta-analysed using the restricted likelihood model for crude proportions with 95% confidence interval(Jamovi v.1.6). To identify potential associations between archaea and bacteria, a Spearman correlation analysis was conducted using outputs from analysis by MetaPhlAn 4. Bacterial genera were correlated with the top 20 most abundant archaeal genera(post-filtered), using Benjamini-Hochberg p-adjustment.

For the metatranscriptomic analysis, mRNA count tables were imported into R for visualization with Phyloseq(McMurdie and Holmes 2013) and DESeq2 R packages(Love et al. 2014). Kyoto Encyclopedia of Genes and Genomes(KEGG) was used to map the methanogenesis within the carbon metabolism pathway(map01200) using KEGG mapper to identify the methanogenesis differentially-expressed genes(https://www.genome.jp/kegg-bin/color_pathway_object).

Results

Following a thorough assessment and retrieval of sequencing data, 7 studies were included, providing a total of 211 caries-active and 168 caries-free DNA-seq samples, and 88 caries-active and 98 caries-free RNA-seq samples**(Table 1)**. For clarity in communication, "caries-active" refers to samples collected from donors with caries, regardless of the specific lesion activity status or index used(**Figure 1B)**.

The studies employed various methods for diagnosing caries, resulting in a highly heterogeneous disease group(**Table 1**). ICDAS system was the most commonly used for detailed visual classification of caries stages. ICCMS and definition disease group as early caries defined as having at least one tooth with early, non-cavitated carious lesion was also used(Al-Hebshi et al. 2019). Another study identified untreated caries affecting three or more surfaces, highlighting

severe cases(Belstrøm et al. 2017), and others used the DMFT index(Baker et al. 2021), d1mft for early caries stages in children(Blostein et al. 2022).

DNA-seq samples were used to calculate putative prevalence of archaea, independently of their taxonomy across datasets. As a result of filtering, n=63 out of 379 samples were archaea-positive. Using our post-filtering strategy, most studies had only 1-3 samples from each dataset positive for archaea(Carda-Diéguez et al. 2022; Espinoza et al. 2018; Pang et al. 2021) and one had no signs of archaea(Belstrøm et al. 2017). Meanwhile, two studies concentrated the highest prevalence of archaea-positive samples(Al-Hebshi et al. 2019; Blostein et al. 2022), interestingly both using the same DNA extraction kit (ZymoBiomics miniprep kit). It appears that both saliva and dental plaque can harbor archaea: out of 284 dental plaque samples, 52 were archaea-positive, while 11 of the 95 saliva samples were archaea-positive. A relative prevalence of 20%(CI=0.00-40.00%) was found, independently of their caries status(**Figure 2A**).

No differences in alpha-diversity between groups, caries and caries-free, were found(**Figure 2B-** filtered samples, **Supplementary Figure-1A-** overall samples). **Figure 2C** shows the beta diversity in of the total archaeal gene profiles, and for phyla and genus levels. For the pairwise comparison of caries and caries-free samples, a significant variability was observed(p=0.038). As a surrogate variable for age, 63 samples were divided according to type of dentition. Despite the disparity in sample sizes (59 deciduous vs. 4 permanent), the deciduous teeth group had significantly higher archaeal alpha-diversity, but not beta-diversity(**Supplementary Figure-1B)**.

Figure 3A presents the taxonomic classification at the phylum level for post-filtered samples. *Euryarchaeota* was the most common phylum, followed by *Thermoplasmatota*, *Nitrososphaerota*, and *Thermoproteota*. No significant taxonomy differences between caries and caries-free groups were observed at the phylum level. Among 150 non-redundant taxa at the genus level, *Methanobrevibacter* had the highest number of reads across all samples. It was succeeded by *Methanosarcina, Thermococcus, Methanothermobacter, Methanobacterium, Pyrococcus,* and *Methanococcus.*

Figure 3 shows the differential abundance of archaea in caries versus caries-free samples. While *Nitrososphaeraceae* was enriched in caries samples, the *Euryarchaeota* families *Halococcaceae* and *Thermococcaceae* were depleted(**Figure 3B**). At the genus level(**Figure 3C**), *Nitrososphaera* (*Thaumarchaeota* phylum) and *Ferroplasma* (*Euryarchaeota* from *Thermoplasmata* class) were enriched in caries.

Figure 3D shows the Spearman correlation of the archaeome and the core bacteriome. *Strephotrophomonas* was significantly correlated with the highest number of archaea taxa. Interestingly, *Streptococcus* has shown correlations with methanogens, while *Actinomyces* is associated specifically with *Methanobrevibacter*. Other bacteria like *Abiotrophia* and *Gemella* have positive correlations with archaea *Methanococcus*, *Pyrococcus*, and *Thermococcus*, with *Corynebacterium* being linked to the latter two. In contrast, *Saccharibacteria* is negatively correlated with *Halorussus* and *Methanosphaera*. *Porphyromonas* and *Tannerella* did not show significant correlations with any archaea.

Archaea functions in caries vs. caries-free

Genes frequently identified across all DNA-seq samples included ATP synthase subunit A, PFL family protein, ribulose-bisphosphate carboxylase, exonuclease ABC subunit UvrB, sodium-translocating pyrophosphatase, and ATP-dependent chaperone ClpB(**Supplementary Figure-2**). Several genes were overrepresented in caries samples(**Supplementary Table-3**): cation-transporting P-type ATPase, ATP synthase subunit-A, replication-associated recombination protein A, V-type ATP synthase subunit-A, pyruvate phosphate dikinase, NAD-dependent protein deacylase, cell division protein FtsZ, DNA topoisomerase subunit-B, anthranilate synthase component I, and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase.

RNA-seq analysis showed underexpression of various genes in caries-active samples(**Supplementary Table-3**). Genes with potentially significant roles in oral biofilms, such as adhesins, glycosyltransferases, growth factors, and agmatinase, were found to be more active in caries-free RNA-seq samples(**Table 2**). There was underexpression of several genes related to methanogenesis in caries-active samples, such as CoB--CoM heterodisulfide reductase iron-sulfur, tetrahydromethanopterin S-methyltransferase, coenzyme-B, and especially coenzyme-M (**Table 2**). **Supplementary Figure-3** shows the KEGG carbon metabolism pathway confirms that methanogenesis is underexpressed in caries-active samples, including genes from the beginning to the end of the pathway.

Discussion

Archaea are often overlooked in oral microbiome studies, largely due to methodological constraints including primer sensitivity. Our recent *in silico* analysis of primers revealed significant

gaps in coverage, highlighting the need for improved methodologies(Cena et al. 2024). Omics studies can overcome primers issues, expanding our understanding of the oral archaeome. We re-examined shotgun sequencing data from 379 samples using a pipeline tailored to avoid bioinformatics noise in mapping reads. The domain was considered as "detected" only if it met our threshold, resulting in 63 archaea-positive samples. Even if each taxon is present in low abundance, archaea could still be classified as positive if their signal surpasses the threshold of 500 reads in total. The relative prevalence found here aligns with findings from endodontic samples analyzed through various molecular methods, where archaea were similarly present in 20% of individuals(Cena et al. 2023). This suggests that the actual prevalence of archaea in most oral samples can be ~10-20%, although this can be influenced by the extraction method used. Therefore, it seems that our post-filtering could be applied in metagenomics to detect archaea ubiquity. Archaea appear more prevalent and diverse (higher alpha-diversity) in children with deciduous teeth compared to permanent teeth. However, these findings should be interpreted cautiously due to the limited sample size and potential confounding factors, such as lifestyle differences and variations in nucleic acid extraction methods. Further research with larger cohorts is warranted.

There has been no exploration of archaea commensal or protective contribution within the oral microbiome, despite extensive discussions on the role of archaea in oral pathogenesis. One hypothesis suggests that archaeal pathogenesis does not occur due to the absence of essential virulence factor genes or because their cell walls do not allow interdomain DNA changes with the host, however, the "not-yet-discovered hypothesis" proposes that the full spectrum of archaeal functions, including potential commensal or protective roles, remains to be uncovered(Duller and Moissl-Eichinger 2024). To better explore the oral archaeome, we advocate for the development of multifaceted approaches that encompass both abundance and prevalence metrics, alongside advancements in bioinformatic pipelines and primer design. For instance, we did not detect archaea in the same samples when using the MetaPhlAn 4 pipeline. This outcome was expected, especially for low-abundance organisms, as prior studies analyzing the human microbiome with different pipelines have consistently yielded variable results(Sun et al. 2021). By using our curated database and a tailored pipeline designed for archaea, we were able to detect them and identify several of their potentially relevant pathways within the oral microbiome. As next steps, we plan to compare the pipeline on environmental microbiomes known for high archaea abundance, using other conventional pipelines as controls. Also, we are planning to check if our pathway is capable of identifying low abundance archaea in species level using longer *de novo* assembled contigs,

as this will provide higher confidence levels in the genes aligned though with the loss of some diversity. Another strategy will be using metagenome assembled genomes(MAGs) tools to find novel oral archaea.

Euryarchaeota was the most abundant phylum, while *Thermoplasmatota* and *Nitrososphaerota* were less abundant. Some *Euryarchaeota* representatives are methanogenic and can act in periodontal and endodontic biofilms in syntropy with disease-associate bacteria, however, here they seem to be acting in favor to the host in caries-free individuals. On the other hand, *Nitrososphaeraceae*, a member *Thaumarchaeota* family, was enriched in caries samples. Our previous pilot study identified representatives of the *Thaumarchaeota* phylum in samples of coronal caries in low abundance(Dame-Teixeira et al. 2019). These archaea were characterized by their ability to nitrify ammonia, contributing to the disruption the natural buffering system that prevents demineralization processes, which can lead to the formation of caries lesions. Other archaeal groups were depleted in caries when compared to caries-fee samples: *Halococcaceae* includes halophilic archaea that thrive in high-salt environments, and *Thermococcaceae* are hyperthermophilic archaea, typically found in environments with high temperatures. It is rather surprising that archaea commonly associated with extreme environments may not tolerate low pH caries environment, given their low-porosity cell walls, which would seemingly offer protection against acidic conditions. Therefore, it is likely that another factor contributes to their depletion in caries environments, beyond their inability to withstand low pH conditions.

To observe cross-kingdom interactions, we made correlation between archaeome and core bacteriome. Archaea might have the ability to team up with bacteria typically found in mature supragingival dental plaque forming microbial networks, probably due to their metabolic capacity at the end of the food chain and the presence of microoxic niches within mature biofilms. Despite the aerobic nature of the mouth, early bacterial colonizers rapidly consume oxygen, creating conditions suitable for obligate anaerobes such as methanogens, particularly in dense dental plaque. Furthermore, correlations with *Strephotrophomonas* and several methanogens were found. In environmental microbiomes, *Strephotrophomonas* plays a role in nitrogen fixation and the oxidation of sulfur, highlighting its contribution to nutrient cycling(Ryan et al. 2009).

The underexpression of adhesins in caries-active samples suggests that archaea may also play a minor role in dental plaque formation in caries-free. This make sense considering that certain strains can form biofilms on various substrates, including mucous membranes, by producing relatively low amounts of extracellular polysaccharides composed of glucose, mannose, and

galactose(Bang and Schmitz 2018). These organisms possess cell surface structures, such as type IV pili and the archaellum, that allow them to move, adhere to surfaces, and interact with other microbial cells(van Wolferen et al. 2022).

The formation of biofilms by archaea is an emerging field of study. For instance, we found methanogens RNA reads showing its activity in laboratory-engineered biofilms(Dame-Teixeira et al. 2024). Other *in vitro* studies showed that *Methanobrevibacter* and *Methanosarcina* form monospecies biofilms with low amounts of extracellular polysaccharides(Bang et al. 2014). Also, it has been reported that *Methanothermobacter thermautotrophicus* develops monospecies biofilms on nickel grids, using Mth60 fimbriae for robust adhesion(Thoma et al. 2008). Other archaea has been shown to be involved in biofilms using archaella for adhesion or a sulfated extracellular polymeric substance, ensuring stability in harsh environments(van Wolferen et al. 2018). All of these genera were present and with gene expression here.

We showed for the first time that methanogenesis can be an active function in supragingival dental plaque and saliva in homeostasis. Methanogenesis is an unique metabolic process in which archaea use carbon dioxide(CO_2) and hydrogen(H_2), produced by anaerobic bacterial fermentation, as an electron donor to generate methane(CH_4) as a metabolic byproduct. This process involves several stages, including the formation of intermediates like formic acid, acetic acid, and methyl coenzyme M (Sogodogo et al. 2019). The proteins involved in methanogenesis are categorized into three essential phases: substrate activation and conversion, methyl-coenzyme M formation, and methane reduction. Key enzymes such as formate dehydrogenase and formylmethanofuran dehydrogenase activates substrates converting them into intermediates. Enzymes like tetrahydromethanopterin S-methyltransferase facilitate the formation of methyl-coenzyme M, while proteins such as CoB--CoM heterodisulfide reductase and methyl-coenzyme M reductase operon proteins drive the final steps in methane production(Ferry 2010).

Methanobrevibacter are not commonly prevalent in the oral microbiome of modern industrialized societies, but highly prevalent in ancient dental calculus suggesting its relation to lifestyle (oral hygiene, diet- refined carbohydrate, less protein, etc)(Gancz et al. 2023). The overexpression of carbohydrate metabolism functions in the caries archaeome reinforces the biological plausibility of our analysis and confirms the presence of archaea in supragingival samples, likely in very low abundance. Methane production can be increased in the presence of sucrose *in vitro*, especially under conditions of high pH(Yang et al. 2017). While there is limited data on the characteristics of the caries archaeome, methanogenic archaea may play a role in maintaining the environment

in caries by regulating pH levels, which could be essential for the survival of the main microorganisms involved in the disease.

Limitations

The datasets exhibit significant heterogeneity due to the differences in caries classification. We have previously discussed this issue of inconsistent definitions and descriptions of caries cases, even for the same condition, when using NGS data for meta-analysis(Dame-Teixeira 2021). The criteria for determining caries cases should be standardized for further NGS studies in caries. As this study is based on secondary data, we had no control on metadata quality, and more information on oral health of participants would have add value to the understanding of the findings.

Datasets containing RNA are not included among the post-filtered samples of DNA in the three studies where both nucleic acids were analyzed. While this indicates a significant risk of false negatives resulting from the arbitrary nature of our post-filtering strategy, which may seem contrary to our concept to find low abundance microorganisms, prioritizing the avoidance of false positives is more crucial than minimizing false negatives at this stage of the oral archaeome evidence. Sensitivity tests for post-filtering will be incorporated as a next step in our pipeline development.

Conclusion

Saliva and supragingival dental plaque emerged as sites of low archaeal abundance, and methanogenesis genes were underexpressed in caries-active samples. This analysis, tailored to minimize false-positive outcomes, uncovered compelling evidence of archaea in dental caries-associated samples. Although putative prevalence of archaea could be influenced by the extraction method used, it was similar to that observed in other oral sites in metanalysis, confirming that our pipeline seems suitable for integration with other methods to screen for archaea in shotgun analysis of oral samples.

Acknowledgements

ND-T and TD acknowledge the UK's Academy of Medical Sciences Newton International Fellowship (grant no. NIF\R5\242). ND-T and JAC acknowledge FAP-DF (grant no. 00193-00001824/2023-45). JAC acknowledges CAPES (Coordination for the Improvement of Higher Education Personnel) and DPG/DPI/UnB.

Author contribution

Naile Dame-Teixeira contributed to conception, design, data acquisition and interpretation, drafted and critically revised the manuscript; Jack Lynch contributed to data acquisition and interpretation, and critically revised the manuscript; Xia Yu contributed to data acquisition and interpretation, and critically revised the manuscript; Jessica Alves de Cena contributed to data acquisition and interpretation, drafted and critically revised the manuscript; Thuy Do contributed to conception, design, data acquisition and interpretation, and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

Conflict of Interest statement

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

References

Al-Hebshi NN, Baraniya D, Chen T, Hill J, Puri S, Tellez M, Hasan NA, Colwell RR, Ismail A. 2019. Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. Journal of oral microbiology. 11(1):1557986.

Baker JL, Morton JT, Dinis M, Alvarez R, Tran NC, Knight R, Edlund A. 2021. Deep metagenomics examines the oral microbiome during dental caries, revealing novel taxa and co-occurrences with host molecules. Genome Res. 31(1):64-74.

Bang C, Ehlers C, Orell A, Prasse D, Spinner M, Gorb SN, Albers S-V, Schmitz RA. 2014. Biofilm formation of mucosa-associated methanoarchaeal strains. Frontiers in microbiology. 5:353.

Bang C, Schmitz RA. 2018. Archaea: Forgotten players in the microbiome. Emerging Topics in Life Sciences. 2(4):459-468.

Belmok A, Cena JA, Kyaw CM, Damé-Teixeira N. 2020. The oral archaeome: A scoping review. J Dent Res.22034520910435.

Belstrøm D, Constancias F, Liu Y, Yang L, Drautz-Moses DI, Schuster SC, Kohli GS, Jakobsen TH, Holmstrup P, Givskov M. 2017. Metagenomic and metatranscriptomic analysis of

saliva reveals disease-associated microbiota in patients with periodontitis and dental caries. NPJ biofilms and microbiomes. 3(1):23.
Blostein F, Bhaumik D, Davis E, Salzman E, Shedden K, Duhaime M, Bakulski KM, McNeil DW, Marazita ML, Foxman B. 2022. Evaluating the ecological hypothesis: Early life salivary microbiome assembly predicts dental caries in a longitudinal case-control study. Microbiome. 10(1):240.
Blostein F, Zou T, Bhaumik D, Salzman E, Bakulski KM, Shaffer JR, Marazita ML, Foxman B. Bacterial community modifies host genetics effect on early childhood caries. medRxiv [Preprint]. 2023. 10:2023.01.11.23284235.
Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using diamond. Nat Methods. 18(4):366-368.
Butcher MC, Short B, Veena CLR, Bradshaw D, Pratten JR, McLean W, Shaban SMA, Ramage G, Delaney C. 2022. Meta-analysis of caries microbiome studies can improve upon disease prediction outcomes. APMIS. 130(12):763-777.
Carda-Diéguez M, Moazzez R, Mira A. 2022. Functional changes in the oral microbiome after use of fluoride and arginine containing dentifrices: A metagenomic and metatranscriptomic study. Microbiome. 10(1):159.
Cena JA, Alexandra de Vasques Castro J, Belmok A, Sales LP, Alves de Oliveira L, Stefani CM, Dame-Teixeira N. 2023. Unraveling the endodontic archaeome: A systematic review with meta-analysis. J Endod.
Cena JA, Belmok A, Kyaw CM, Dame-Teixeira N. 2024. The archaea domain: Exploring historical and contemporary perspectives with in silico primer coverage analysis for future research in dentistry. Arch Oral Biol. 161:105936.
Cena JA, Silvestre-Barbosa Y, Belmok A, Stefani CM, Kyaw CM, Damé-Teixeira N. 2022. Meta-analyses on the periodontal archaeome. Adv Exp Med Biol. 1373:69-93.
Cena JA, Zhang J, Deng D, Damé-Teixeira N, Do T. 2021. Low-abundant microorganisms: The human microbiome's dark matter, a scoping review. Frontiers in Cellular and Infection Microbiology. 11(478).
Dame-Teixeira N, Cena JA, Côrtes DA, Belmok A, Dos Anjos Borges LG, Marconatto L, Giongo A, Kyaw CM. 2019. Presence of archaea in dental caries biofilms. Arch Oral Biol. 110:104606.
Dame-Teixeira N, Almeida AK, Do T, Stefani CM. 2021. Meta-analysis using ngs data: The veillonella species in dental caries. Frontiers in Oral Health.
Dame-Teixeira N, El-Gendy R, Oliveira A, Holanda C, Romeiro L, Do T. 2024. Engineering a dysbiotic biofilm model for testing root caries interventions through microbial modulation. *Microbiome*.
Duller S, Moissl-Eichinger C. 2024. Archaea in the human microbiome and potential effects on human infectious disease. Emerging Infectious Diseases. 30(8):1505.
Espinoza JL, Harkins DM, Torralba M, Gomez A, Highlander SK, Jones MB, Leong P, Saffery R, Bockmann M, Kuelbs C. 2018. Supragingival plaque microbiome ecology and functional potential in the context of health and disease. MBio. 9(6):10.1128/mbio. 01631-01618.
Ferry JG. 2010. The chemical biology of methanogenesis. Planetary and Space Science. 58(14-15):1775-1783.
Gancz AS, Farrer AG, Nixon MP, Wright S, Arriola L, Adler C, Davenport ER, Gully N, Cooper A, Britton K et al. 2023. Ancient dental calculus reveals oral microbiome shifts associated with lifestyle and disease in great britain. Nat Microbiol. 8(12):2315-2325.
Koskinen K, Pausan MR, Perras AK, Beck M, Bang C, Mora M, Schilhabel A, Schmitz R, Moissl-Eichinger C. 2017. First insights into the diverse human archaeome: Specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin. mBio. 8(6).

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol. 15(12):550.
McMurdie PJ, Holmes S. 2013. Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data.
Oren A, Garrity GM. 2021. Valid publication of the names of forty-two phyla of prokaryotes. Int J Syst Evol Microbiol. 71(10).
Pang L, Wang Y, Ye Y, Zhou Y, Zhi Q, Lin H. 2021. Metagenomic analysis of dental plaque on pit and fissure sites with and without caries among adolescents. Frontiers in Cellular and Infection Microbiology. 11:740981.
Pausan MR, Csorba C, Singer G, Till H, Schöpf V, Santigli E, Klug B, Högenauer C, Blohs M, Moissl-Eichinger C. 2019. Exploring the archaeome: Detection of archaeal signatures in the human body. Front Microbiol. 10:2796.
Ryan RP, Monchy S, Cardinale M, Taghavi S, Crossman L, Avison MB, Berg G, Van Der Lelie D, Dow JM. 2009. The versatility and adaptation of bacteria from the genus stenotrophomonas. Nature reviews microbiology. 7(7):514-525.
Seedorf H, Dreisbach A, Hedderich R, Shima S, Thauer RK. 2004. F420h2 oxidase (fpra) from methanobrevibacter arboriphilus, a coenzyme f420-dependent enzyme involved in o2 detoxification. Arch Microbiol. 182(2-3):126-137.
Sogodogo E, Doumbo O, Aboudharam G, Kouriba B, Diawara O, Koita H, Togora S, Drancourt M. 2019. First characterization of methanogens in oral cavity in malian patients with oral cavity pathologies. BMC Oral Health. 19(1):232.
Sun, Z., Huang, S., Zhang, M. et al. 2021. Challenges in benchmarking metagenomic profilers. Nat Methods 18:618.
Thoma C, Frank M, Rachel R, Schmid S, Näther D, Wanner G, Wirth R. 2008. The mth60 fimbriae of methanothermobacter thermoautotrophicus are functional adhesins. Environmental microbiology. 10(10):2785-2795.
van Wolferen M, Orell A, Albers S-V. 2018. Archaeal biofilm formation. Nature Reviews Microbiology. 16(11):699-713.
van Wolferen M, Pulschen AA, Baum B, Gribaldo S, Albers S-V. 2022. The cell biology of archaea. Nature Microbiology.1-12.
Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, Morris AG, Alt KW, Caramelli D, Dresely V et al. 2017. Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus. Nature. 544(7650):357-361.
Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. Proc Natl Acad Sci U S A. 87(12):4576-4579.
Yang B, Wang Q, Zhao X, Wang C, Liu J, Yang H, Liu S, Yin F, Zhang W. 2017. Increased energy production from sucrose by controlled hydrogen-producing fermentation followed by methanogenic fermentation. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 39(6):576-582.

Figures Captions and Tables

Figure 1. Flowchart (adapted from PRISMA): 1A) Selection process of the studies and their databases. Two independent reviewers selected studies based on the assessment of titles and abstracts. In the second stage, same reviewers selected eligible studies based on full-text reading. Studies refer to the original research conducted, whereas study reports are the documents or publications where the details of those studies are presented. Sequencing data and metadata from each study were retrieved from their SRA Bioproject. Studies lacking data availability were excluded. In the initial search, 185 articles were identified. After removing duplicates, 184 articles were screened by title and abstract, and 33 were selected for full-text review; **1B) Tailored archaea pipeline.** A full non-redundant archaea database was specifically tailored for this study, extracted from the National Center for Biotechnology Information (NCBI) RefSeq. A total of 519 different genomes were included. After downloading each dataset and its corresponding metadata from NCBI, and processing the raw read data using fastqc and cutadapt, a count table was generated for each dataset. Reads were mapped against a curated archaea genome database using the DIAMOND v2.0.15.153 sequence alignment tool. Stringent filtering parameters were applied, requiring a minimum of 97% sequence similarity and 90% query coverage. Archaeal prevalence was determined using an arbitrary post-filtering cut-off point, considering the proportion of samples with over 500 archaeal genus counts in total. This filter was used to avoid false-positive data. Samples with more than 500 overall archaeal reads from four studies were used for further analysis. For alpha diversity, ANOVA and post hoc Tukey HSD test were used for normally distributed data, while Kruskal-Wallis and post hoc Dunn's test were used when data were not normally distributed. For beta diversity, permutational multivariate analysis of variance tests were used. The number of samples in total and according to presence of archaea for DNA is presented. For the metagenomic analysis, 10 non-caries related samples from one dataset under the BioProject PRJNA396840 were excluded (subgingival samples). For the metatranscriptomic analysis, three studies with available mRNA sequences were included and analyzed for differential gene expression.

Figure 2. Prevalence (2A) and diversity (2B and 2C) of the archaeome in caries and caries-free samples. A) Overall prevalence of the *Archaea* domain in caries and caries-free dental plaque and saliva samples in 7 metagenomic (DNAseq) studies. The box in the middle of each horizontal line (CI [confidence interval]) represents the point estimate of the effect for a single study, while the diamond represents the overall effect estimate of the meta-analysis. RE (random-effects) Model. **B)** Alpha and **C)** Beta-Diversity of the archaeome in caries and caries-free samples in general, and in genus and phyla levels. The compositional homogeneity of multivariate dispersions among archaea-positive samples was calculated for 999 permutations.

Figure 3. Taxonomy of the caries and caries-free archaeome. Averages at the phylum level **(3A)** (n=63 post-filtered samples using the >500 reads cut-off); Differential abundance of archaea in caries vs. caries free samples, calculated by DESeq at family **(3B)** and genus **(3C)** levels; **3D)** Spearman correlation of the archaeome and bacteriome.

Table 1. Characteristics of the selected studies, datasets and metadata.

NCBI SRA Bioproject Code	Authors	Year	Country	DNA/RNA extraction method	Sampling	N samples caries*	N sample caries free*	Age	Caries diagnosis
PRJNA766357	Pang et al.	2021	China	DNA Cetyltrimethylammonium bromide (CTAB), phenol/chloroform isoamyl alcohol	Dental plaque collected from all pit and fissure sites	20	20	12–13 years old	ICDAS-II
20180420	Al-Hebshi et al.	2019	USA	DNA ZymoBiomics miniprep kit	Whole-mouth dental plaque	20	10	6–10 years old	ICCMS
PRJNA712952	Carda-Diéguez et al.	2022	Spain	DNA, RNA Lysozyme, lysostaphin, and mutanolysin, bead beating, Ambion mirVana miRNA isolation kit	Dental plaque (pool of from 3 sites) collected longitudinally	52	47	≥ 16 years old	ICDAS
PRJNA383868	Espinoza et al.	2018	Australia	DNA, RNA Lysozyme, beat beading, phenol/chloroform isoamyl alcohol	Dental plaque collected before dental examination	58	29	5–11 years old	ICDAS II
PRJNA396840	Belstrøm et al.	2017	Denmark	DNA, RNA MasterPure Complete DNA and RNA Purification Kit	Stimulated saliva samples collected between 8:00am and 11:00am	10	10	37.6 years old (22–70)	Untreated caries ≥ 3 surfaces
PRJNA478018	Baker et al.	2021	USA	DNA QIAmp Microbiome (Qiagen) and DNA Clean & Concentrator (Zymo Research) kit	Stimulated saliva samples (by chewing on sterile parafilm) collected between 8:00am and 11:00am	22	23	≥3 years old	dmft
PRJNA752888 **	Blostein et al. Blostein et al.	2022 2023	USA USA	DNA Zymobiomics miniprep kit	Dental plaque from enamel lesions and saliva collected longitudinally at or after the 36-month visit)	29	29	2 months to 5 years old 2 months to 5 years old	d1mft Prevalence

*Number of samples downloaded from the SRA Bioproject and analyzed in the archaea-tailored pipeline (not refer to the samples described in the reports).

** Two reports from the same study by Blostein et al. were found: two articles for a single study, with the study being the primary unit of interest and analysis. Although data registered under PRJNA752888 included amplicon sequencing samples, only the shotgun sequencing samples were downloaded and analyzed.

Table 2. Methanogenesis genes underexpressed in caries-active samples compared to caries-free samples (n=186 samples), and their potential functions described for archaea.

Protein Name	Function Description	Methanogenesis phase	Log2 Fold Change	p-adjusted
Formate dehydrogenase subunit alpha	Catalyzes the oxidation of formate to carbon dioxide.	Activation and Conversion of Substrates	1.39	1.83E-05
Formylmethanofuran dehydrogenase subunit A	Catalyzes the reversible reduction of CO2 and methanofuran to N-formylmethanofuran.	Activation and Conversion of Substrates	1.15	8.51E-05
Formylmethanofuran dehydrogenase subunit B	Catalyzes the reversible reduction of CO2 and methanofuran to N-formylmethanofuran.	Activation and Conversion of Substrates	0.97	0.0008
Methanol--corrinoid protein MtaC	Involved in methanogenesis from methanol, transferring methyl groups from methanol to a specific corrinoid protein.	Activation and Conversion of Substrates	0.89	0.0011
Methanol--corrinoid protein co-methyltransferase MtaB	Methyltransferase involved in the methanol pathway, catalyzing the methylation of the MtaC-bound cob(I)amide.	Activation and Conversion of Substrates	0.85	0.0044
Coenzyme F420 hydrogenase subunit beta	Participates in the formatehydrogenylase system in methanogenic archaea.	Activation and Conversion of Substrates	1.34	9.08E-06
Coenzyme F420 hydrogenase/dehydrogenase, beta subunit C-terminal domain	Functions as a bidirectional enzyme in the formatehydrogenylase system.	Activation and Conversion of Substrates	1.31	9.31E-06
Tetrahydromethanopterin S-methyltransferase subunit C	Coenzyme M methyltransferase that drives a sodium ion pump.	Formation of Methyl-Coenzyme M	1.65	1.14E-07
Tetrahydromethanopterin S-methyltransferase subunit E	Coenzyme M methyltransferase that drives a sodium ion pump.	Formation of Methyl-Coenzyme M	1.54	5.83E-07
Tetrahydromethanopterin S-methyltransferase subunit H	Part of a complex catalyzing the formation of methyl-coenzyme M from coenzyme M and methyl-tetrahydromethanopterin.	Formation of Methyl-Coenzyme M	0.79	0.0062
F420-dependent methylenetetrahydromethanopterin dehydrogenase	Catalyzes the fourth reaction step of CO2 reduction to methane.	Formation of Methyl-Coenzyme M	0.67	0.0144
CoB--CoM heterodisulfide reductase iron-sulfur subunit A family protein	Catalyzes the reversible reduction of heterodisulfide of methanogenic thiol-coenzymes.	Reduction to Methane	1.98	6.58E-09
CoB--CoM heterodisulfide reductase subunit B	Catalyzes the reversible reduction of CoM-S-S-CoB to thiol-coenzymes.	Reduction to Methane	0.71	0.0114
Methyl-coenzyme M reductase operon protein D	Catalyzes the last step of methane reduction in methanogens.	Reduction to Methane	0.69	0.0111

Methyl-coenzyme M reductase I operon protein C	Catalyzes the last step of methane reduction in methanogens.	Reduction to Methane	0.57	0.0289
Coenzyme-B sulfoethylthiotransferase subunit gamma	Catalyzes the last step in methanogenesis.	Reduction to Methane	1.35	9.08E-06
Coenzyme-B sulfoethylthiotransferase subunit alpha	Catalyzes the final step in methane formation.	Reduction to Methane	0.75	0.0178
5,10-methenyltetrahydromethanopterin hydrogenase	Catalyzes the reversible reduction of methenyl-H4MPT+ to methylene-H4MPT.	Reduction to Methane	0.89	0.0012