Integrative Exploration of Taste Genetics, Oral Microbiome, and Socioeconomic Factors in Early Childhood Caries

By

Mohd Wasif Khan

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Biochemistry and Medical Genetics
Rady Faculty of Health Sciences
University of Manitoba
Winnipeg

Copyright © 2025 by Mohd Wasif Khan

DISCLAIMER

This modified sandwich thesis includes multi-authored manuscripts, each containing a substantial amount of original work by the author, as evidenced by their first authorship. Collaborating colleagues contributed to varying aspects of the work. The individual contributions are detailed in the respective published chapters. Since the manuscripts were published separately, each manuscript had to be independent, and therefore may contain some repeated information, mostly limited to the introduction and methodology sections.

THESIS ABSTRACT

Early childhood caries (ECC) refers to dental cavities or decay in children less than 72 months of age. Nearly half of the children are affected by ECC globally, which can reach up to 98% in some disadvantaged communities. ECC is multifactorial in nature, with various etiological factors, including the microbiome, socioeconomic factors (SEF), and host genetics. Among host genetic factors, taste receptors are crucial as they modulate host taste preference and may also play a role in interactions with microbial species through innate immune responses. To address whether host genetics can influence oral microbes, this study investigated the association of taste receptors with ECC and the dental plaque microbiome. Furthermore, we explored the role of various SEF in ECC and their impact on the oral microbiome. We also assessed interkingdom associations through species co-occurrence.

For this analysis, data were collected from children's parents or caregivers for socioeconomic and behavioral information, along with dental plaque and buccal swab samples for microbiome and 55 taste-related gene sequencing, respectively. Sequencing was performed using next-generation sequencing, with data analyzed using QIIME2 for microbiome and GATK pipelines for variants in taste-related genes. Various analytical, statistical, and machine learning tools in R were employed for the analysis.

Our results indicate that SEF, specifically the Socioeconomic Factor Index (SEFI) and rural-urban status, are highly associated with ECC status. We found a strong correlation between the fungal species *Candida dubliniensis* and the bacteria *Neisseria bacilliformis* in ECC. Through integrative machine learning-based prediction of ECC, the bacterial species *Streptococcus mutans* and the fungi *Candida dubliniensis*, a bitter taste receptor variant *TAS2R60*-rs35195910, along with SEFI and rural-urban status, emerged as top markers. We also evaluated the generalizability of machine learning based ECC prediction using previously published ECC microbiome studies.

In conclusion, we identified microbial markers that are highly associated with ECC and the potential interactions among them. Although taste genetic markers did not influence ECC-associated microbial species, they were associated with other species in the dental plaque microbiome. Lastly, due to the complex role of multiple factors, we recommend integrating multiomics data analysis in future ECC-related studies. These findings help us understand the etiological landscape of ECC and identify susceptible individuals, enabling personalized therapeutic approaches and preventative strategies.

ACKNOWLEDGEMENTS

I want to thank my parents for loving me unconditionally, providing all the support I ever needed, and encouraging me to never stop learning. I also want to thank my siblings for motivating me by considering me their ideal, even though they are far better than me in every way.

To my PhD supervisors and committee members, thank you for your patience, for giving me the space to grow, and for standing by me when I needed it the most. Your presence, words, and support have meant far more than you may realize.

Sincere thanks to the study participants for your time, trust, and willingness to contribute. Your involvement was essential to this research. I gratefully acknowledge the University of Manitoba as the host institution for this work, and thank the university and office staff for their kindness and for making many things easier along the way. My deep appreciation also goes to the funding agencies and taxpayers, whose support made this work possible. Science thrives because of your generosity and continued commitment to advancing knowledge.

To all my friends, for giving me more than I could ever give in return in this journey. Thank you for accepting me despite my limitations and for being a constant source of fun and encouragement. You've all been truly awesome, and I'm proud to have you in my life. Each of you, with your unique qualities and your annoying habits, has taught me life lessons I couldn't have understood any other way. Thank you from the bottom of my heart, though I can never thank you enough.

As the saying goes, it takes a village to raise a child, but I believe it takes even more to raise a PhD scholar. This journey has been filled with countless moments shaped by the people around me. No contribution has been too small, and I'm grateful for every kind gesture and every bit of support along the way.

Lastly, and very much the least, I want to thank myself, for believing in me even when I wasn't sure I could. Those moments were tough and required a strong mindset. Thank you for being resilient during those times and for being my own support. Let this also be a reminder to my future self: no challenge is too big, as long as I am with me.

DEDICATION

***To my parents**, for instilling in me the value of education and for their unwavering belief in me.*

***To all my teachers,** whose guidance has played a vital role in shaping the raw version of me into my better self.*

***To Dr. Vikas Divyakirti,** a continuous source of inspiration, whose teachings have greatly influenced my way of thinking.*

***To A R Rahman** (the Mozart of Madras), for showing what striving for the best can truly achieve, and for his transcendent melodies that brought the touch of art to the long hours of doing science.*

TABLE OF CONTENTS

DISCLAIMER II

THESIS ABSTRACT III

ACKNOWLEDGEMENTS IV

DEDICATION V

LIST OF TABLES VIII

LIST OF FIGURES IX

LIST OF ABBREVIATIONS XI

CHAPTER 1: INTRODUCTION 1

1.1 EARLY CHILDHOOD CARIES 1

1.2 SOCIOECONOMIC, ENVIRONMENTAL, AND BEHAVIORAL FACTORS IN ECC 6

1.3 ORAL MICROBIOME AND ECC 12

1.4 TASTE RECEPTORS AND ECC 20

1.5 HOST GENETICS AND MICROBIOME ASSOCIATION 24

1.6 KNOWLEDGE GAPS AND THESIS CONTRIBUTION 26

CHAPTER 2: THESIS OBJECTIVES AND RATIONALE 28

2.1 AIM 1 28

2.2 AIM 2 28

2.3 AIM 3 29

CHAPTER 3: STUDY POPULATION AND DATA SOURCES 31

3.1 STUDY POPULATION 31

3.2 PARTICIPANTS' INFORMATION 31

3.3 SAMPLE COLLECTION 32

3.4 MICROBIOME DATA 32

3.5 TASTE GENETICS DATA 33

3.6 DATA FROM EXTERNAL STUDIES/SOURCES 33

CHAPTER 4: A CROSS-COHORT ANALYSIS OF DENTAL PLAQUE MICROBIOME IN EARLY CHILDHOOD CARIES 35

4.1 SUMMARY 37

4.2 INTRODUCTION 38

4.3 RESULTS 40

4.4 DISCUSSION 51

4.5 METHODS 55

4.6 SUPPLEMENTARY DATA 59

BRIDGE TO CHAPTER 5 66

CHAPTER 5: ROLE OF SOCIOECONOMIC FACTORS AND INTERKINGDOM CROSSTALK IN THE DENTAL PLAQUE MICROBIOME IN EARLY CHILDHOOD CARIES 67

5.1 Summary 69
5.2 Introduction 70
5.3 Results 71
5.4 Discussion 84
5.5 Methods 90
5.6 Supplementary data 96
Bridge to chapter 6 101

CHAPTER 6: INTEGRATIVE ANALYSIS OF TASTE GENETICS AND THE DENTAL PLAQUE MICROBIOME IN EARLY CHILDHOOD CARIES 102

6.1 Summary 104
6.2 Introduction 105
6.3 Results 107
6.4 Discussion 121
6.5 Methods 127
6.6 Supplementary data 134

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS 150

7.1 Conclusions 150
7.2 Limitations 152
7.3 Future Directions 153

REFERENCES 156

List of Tables

Main Tables

Table 4.1: Information about the datasets used..... 41

Table 5.1: Descriptive statistics for the host variables analyzed. 78

Table 6.1: Descriptive statistics for the host variables used in the study. 108

Table 6.2: Significant loci identified in 55 taste-related genes to be associated with ECC status. 110

Supplementary Tables

Table S4.1: Additional details of datasets and Qiime2 processing options. 59

Table S4.2: Key resources table with additional description. 61

Table S5.1: Explanation of host variables analyzed in the study. 96

Table S6.1: List of the genes considered as candidate genes for targeted sequencing. 135

List of Figures

Main Figures

Figure 4.1: Schematic overview of data processing and analysis pipeline. 40
Figure 4.2: UpSet plot for the number of shared OTUs across different datasets. 42
Figure 4.3: Relative abundance of top taxa between caries and ECC samples. 43
Figure 4.4: Comparison of species-level alpha diversity by Shannon index. 44
Figure 4.5: Analysis of similarities (ANOSIM) test using Bray diversity metrics on species-level OTUs to illustrate the extent of similarities between five studies. 45
Figure 4.6: Principal component analysis (PCA) plots for species-level datasets. 46
Figure 4.7: Meta-analysis for differentially abundant taxa. 47
Figure 4.8: Model performance by AUROC for 5 different machine learning models on pooled dataset from 5 studies. 49
Figure 4.9: Heatmap for the performance of random forest classifier by AUROC on species-level data. 50
Figure 4.10: Heatmap for feature relative weights along with their rank obtained from the random forest model. 51
Figure 5.1: Microbiome composition and alpha and beta diversity in caries-free and early childhood caries. 73
Figure 5.2: Differential network analysis for CF and ECC samples on combined bacterial and fungal datasets 75
Figure 5.3: Machine learning performance for the classification of ECC and CF status. 77
Figure 5.4: Association between host variables and early childhood caries. 80
Figure 5.5: Association of the host variables with diversity metrics. 82
Figure 5.6: Associations between microbial features and various socioeconomic factors. 84
Figure 6.1: Taste genetic variants and genes association with ECC status. 111
Figure 6.2: Association between dental plaque microbial taxa and ECC outcome. 114
Figure 6.3: Spearman correlation of the ECC associated variants with alpha and beta diversity. 115
Figure 6.4: Association between ECC-associated variants and microbial taxa. 116
Figure 6.5: Mediation analysis for genetic variants and ECC association with microbial species as mediator 118
Figure 6.6: Machine learning based classification for ECC status using microbiome, genetic variants and covariate data 121

Supplementary Figures

Figure S4.1: Differentially abundance taxa on raw abundance without batch correction values at species-level using DESeq2. 62
Figure S4.2: Meta-analysis for differentially abundance taxa without batch correction values. 62
Figure S4.3: Random Forest performance in terms of AUROC with combined genus and species-level OTUs for Pooled dataset with batch-corrected values. 63
Figure S4.4: AUROC values with species-level data with CLR normalization without batch correction. 63
Figure S4.5: AUROC values with batch corrected values using genus-level data. 64
Figure S4.6: AUROC values with imputed data with CLR normalization with species-level data. 65
Figure S5.1: Exploratory data analysis for the distribution of host variables used as socioeconomic and behavioral factors after imputation. 98
Figure S5.2: Association between microbial features and various socioeconomic factors. 99
Figure S5.3: LinDA association between microbial species and socioeconomic factors. 100
Figure S6.1: Batch processing, batch effects, covariates, and principal components from genetic data. 137
Figure S6.2: Covariates used in this study. 138
Figure S6.3: Genomic inflation rate for *p*-values for all variants using different covariates. 139
Figure S6.4: Model evaluation for different covariates using the top five ECC-associated taste genetic variants. 140
Figure S6.5: Significant genetic variants across different association models for ECC outcome. 143
Figure S6.6: Genetic variants and gene association with ECC status using a logistic model. 143
Figure S6.7: Linkage disequilibrium (LD) analysis of the variants identified in the model with age and sex as covariates. 145
Figure S6.8: Allele frequencies of ECC-associated significant taste-related genetic variants. 146
Figure S6.9: Association between ECC-associated variants and microbial taxa in ECC and CF stratified samples. 147
Figure S6.10: Association between 31 top variants (5% of all quality-filter passing variants) based on *p*-values for association with ECC and all microbial taxa. 148
Figure S6.11: Machine learning based classification for ECC status using microbiome data on external datasets. 149

List of abbreviations

Abbreviation	Full Form
AAPD	American Academy of Pediatric Dentistry
ACME	Average Causal Mediation Effect
ADE	Average Total Effect
AIC	Akaike Information Criterion
ANOSIM	Analysis of Similarities
AUROC	Area Under the Receiver Operating Characteristic Curve
BH	Benjamini-Hochberg
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CF	Caries-Free
CLR	Centered Log Ratio
CRA	Caries Risk Assessment
CV	Cross-Validation
DDE	Developmental Defects of Enamel
DZ	Dizygotic
ECOHIS	Early Childhood Oral Health Impact Scale
FDR	False Discovery Rate
GPCR	G-Protein-Coupled Receptor
GWAS	Genome-Wide Association Studies
HOMD	Human Oral Microbiome Database
HWE	Hardy-Weinberg Equilibrium
ITS	Internal Transcribed Spacer
KNN	K-Nearest Neighbor
LM	Linear Mixed Effect Models
LODO	Leave-One-Dataset-Out
MAF	Minor Allele Frequency
MCHP	Manitoba Centre for Health Policy
MICE	Multiple Imputation by Chained Equations
ML	Machine Learning
MRS	Microbial Risk Score

MZ	Monozygotic
NGS	Next-Generation Sequencing
OHRQoL	Oral Health-Related Quality of Life
OR	Odds Ratio
OTU	Operational Taxonomic Units
PCR	Polymerase Chain Reaction
PCs	Principal Components
QSM	Quorum-Sensing Molecules
S-ECC	Severe Early Childhood Caries
SDF	Silver Diamine Fluoride
SEF	Socioeconomic Factors
SEFI	Socioeconomic Factor Index
SES	Socioeconomic Status
SRA	Sequence Read Archive
SVM	Support Vector Machines
TWAS	Transcriptome-Wide Association Studies
VCF	Variant Calling Format

Chapter 1: Introduction

1.1 Early childhood caries

1.1.1 Definition and Epidemiology

Early childhood caries (ECC) refers to dental decay or cavities occurring in the primary dentition of children younger than 72 months. According to the American Academy of Pediatric Dentistry (AAPD), ECC is characterized by the presence of at least one decayed tooth surface (whether non-cavitated or cavitated), a missing tooth due to caries, or a filled tooth surface in any primary tooth in children under six years of age (AAPD, 2024). ECC is now a widely accepted term for caries in the primary dentition, replacing earlier terminology such as nursing caries, baby bottle syndrome, and baby bottle tooth decay, terms currently discouraged by pediatric dentists and researchers (Tinanoff, 2015). ECC affects approximately half of the children globally (El Tantawi *et al.*, 2018), with prevalence rates exceeding 90% in certain socially disadvantaged and minority group populations (Amin *et al.*, 2014; Pierce *et al.*, 2019). The Global Burden of Disease Study reported in 2017 that untreated caries in deciduous teeth affects approximately 532 million children worldwide (GBD Disease and Injury Incidence and Prevalence Collaborators, 2018), highlighting its significance as one of the most common chronic diseases.

ECC typically starts in the primary incisors and spreads to other areas of the dentition following the pattern of tooth eruption. ECC often progresses to a severe form, termed severe early childhood caries (S-ECC), which is classified according to the child's age, total number of affected teeth, and location of carious lesions (AAPD, 2024). Due to its severe nature, S-ECC often necessitates treatment under general anesthesia, particularly for children requiring extensive surgical interventions who may be uncooperative or for whom local anesthesia is ineffective, as determined by urgency and surgery risk assessments (Oubenyahya and Bouhabba, 2019). Nonetheless, such surgical treatments are complicated and do not address the underlying etiological factors, resulting in many children experiencing recurrent or new carious lesions (Schroth and Smith, 2007; Edmonds, Williams and Carrico, 2019). If left

untreated, ECC can lead to severe pain, abscess formation, and irreversible damage to the dental hard tissues and pulp.

The consequences of ECC extend beyond oral discomfort and impaired nutritional intake, including difficulties in sleeping, learning, and increased irritability (Wong *et al.*, 2017; Grant *et al.*, 2019). Furthermore, ECC heightens host susceptibility to infections in other body regions, including the brain (Hibberd and Nguyen, 2012). These complications, compounded by the high prevalence of ECC, create a substantial economic burden on healthcare systems and affected families (Schroth *et al.*, 2016).

1.1.2 Phenotypic characterization and pathophysiology

Dental enamel, the hardest tissue in the human body, has a highly mineralized structure. During amelogenesis, the process of enamel formation, extracellular matrix proteins secreted by ameloblasts act as scaffolds that facilitate the deposition of calcium and phosphate ions (Yu and Klein, 2020). Mature enamel primarily comprises calcium hydroxyapatite crystals containing calcium (Ca^{2+}) ions. ECC initially presents as white-yellow subsurface lesions, commonly termed white spots, which signify subsurface enamel demineralization (Khoroushi and Kachuie, 2017; Lopes *et al.*, 2024). Due to the relatively thin enamel layer in primary dentition, these lesions can rapidly progress into cavitations characterized by yellow or brown discoloration accompanied by enamel breakdown. With further advancement, ECC can cause irreversible damage to the dental hard tissues and dental pulp, leading to extensive tooth destruction, structural compromise, and eventual tooth loss. S-ECC manifests more aggressively in the presence of enamel hypoplasia, a condition characterized by inadequate enamel formation (Caufield, Li and Bromage, 2012). Typically, the maxillary primary incisors are the first and most severely affected teeth due to their early eruption and prolonged exposure to cariogenic substrates.

The primary causative factor of ECC is microbial activity. Under normal conditions, the dental plaque microbiome maintains homeostasis through balanced synergistic and antagonistic microbial interactions; however, disruptions can alter the microbial composition. According to the ecological plaque hypothesis, dental plaque functions as a dynamic microbial ecosystem that responds to environmental changes, such as pH fluctuations and nutrient

availability (Marsh, 1994). A shift in the oral microbiome structure predisposes specific sites to disease (Marsh, 1994; Marsh and Zaura, 2017). The continuous presence and fermentation of sugars in the oral cavity create acidic conditions, promoting the predominance of aciduric (acid-tolerant) species through the selection of acidogenic (acid-producing) bacteria and the subsequent elimination of acid-sensitive microorganisms. This process is further intensified by the dominance of acid-tolerant microbes within the plaque microbial community, promoting the sustained progression of carious lesions, which is also described as the extended ecological plaque hypothesis (Becker *et al.*, 2002; Corby *et al.*, 2005; Fakhruddin, Ngo and Samaranayake, 2019).

Microorganisms present within dental biofilms ferment sugars available on tooth surfaces, producing acids (Takahashi, 2005; Sheiham and James, 2015). Prolonged exposure to these acidic conditions leads to enamel demineralization, characterized by the loss of calcium and phosphate ions. Under normal physiological conditions, this mineral loss is counterbalanced by remineralization through calcium replenishment from the salivary flow, resulting in a dynamic balance between demineralization and remineralization (Pitts *et al.*, 2017). However, under consistently high acidic conditions, salivary calcium levels become insufficient to maintain this equilibrium or buffer against mineral loss, marking the initiation of carious lesions. Immature or underdeveloped enamel in children further exacerbates susceptibility, contributing significantly to the high prevalence of ECC (Selwitz, Ismail and Pitts, 2007).

1.1.3 Current preventing strategies

Regular tooth brushing reduces dental biofilm formation, and the addition of fluoride helps in enamel remineralization, thus increasing resistance to caries formation. Brushing should begin as soon as the first tooth erupts, and due to the delicate structure of children's teeth, a soft toothbrush is often recommended to prevent enamel damage. Moreover, restricting sugary food intake, especially fruit juices, in accordance with the recommendations of the AAPD is advised. Fluoride remains the principal preventive measure, which, in addition to toothpaste, can easily be incorporated into the community water supply (Wright *et al.*, 2014; Walsh *et al.*, 2019). Communities with fluoridated water demonstrate significant reductions in caries and ECC prevalence, with up to a 35% decrease in ECC cases (McDonagh *et al.*, 2000;

Iheozor-Ejiofor *et al.*, 2024). Additional fluoride sources include oral supplements, topical gels, and varnishes (Tubert-Jeannin *et al.*, 2011). As a treatment strategy, Silver Diamine Fluoride (SDF) is routinely applied in clinical settings because of its high silver content in addition to fluoride (Hafiz *et al.*, 2022). While fluoride in SDF facilitates enamel remineralization, silver functions as an antibacterial agent that inhibits microbial growth (Rosenblatt, Stamford and Niederman, 2009). SDF is applied topically and is particularly beneficial for children who are unable to undergo restorative dental treatments. However, the primary disadvantage of SDF is its cosmetic effect, as it stains carious lesions black. The AAPD recommends that the effectiveness of SDF treatment be re-evaluated 2–4 weeks after the initial application to assess lesion arrest, and reapplication should be considered as necessary (Crystal *et al.*, 2018).

Another recommendation by the AAPD to prevent ECC incidence or severity is the incorporation of caries risk assessment (CRA) during a child's first dental examination by healthcare providers, facilitating personalized preventive plans (Ramos-Gomez *et al.*, 2012; Schroth *et al.*, 2021). The CRA tool evaluates biological, social, and behavioral risk factors, including indicators such as the presence of decayed/missing/filled surfaces (dmfs), carious lesions or enamel defects, and visible dental plaque. Behavioral factors contributing to caries development include frequent consumption of sugary snacks or drinks between meals, the practice of putting a child to bed with a sugar-containing bottle, and active caries in the primary caregiver or the mother. Social factors collected through the CRA include special healthcare needs, recent immigration status, and low socioeconomic status. Utilizing CRA assists healthcare providers in identifying caries risk levels and developing appropriate preventive and management strategies in collaboration with the parents.

Fisher-Owens et al. (2007) proposed a comprehensive conceptual framework addressing the multifaceted determinants influencing children's oral health. The framework emphasizes that children's oral health outcomes are determined not only by individual practices, behaviors, and biological factors, such as genetics and microbial composition, but also by broader influences involving family, community, and healthcare systems. Thus, a holistic approach with coordinated efforts is essential (Fisher-Owens *et al.*, 2007). Such an integrated model is

crucial for oral health management, as social determinants and disparities significantly influence ECC prevalence (Lee and Divaris, 2014).

Understanding the risk and susceptibility of ECC in children requires the identification of specific factors contributing to oral disease development. A proper assessment of these risk factors provides valuable insights into children's vulnerability to ECC.

1.1.4 Risk factors

ECC is a multifactorial disease resulting from a complex interplay between sociocultural, environmental, microbial, and host genetic factors. Socioeconomic status (SES) significantly influences the risk of developing ECC (Vasireddy *et al.*, 2021). Children from families with lower SES consistently exhibit higher susceptibility, primarily due to factors such as limited access to dental care, poor nutritional habits, and fewer opportunities for oral hygiene education. Parental or caregiver educational levels also significantly contribute to the risk of ECC (Folayan *et al.*, 2024). While family income critically affects access to regular dental check-ups and preventive care, cultural practices related to diet, feeding behaviors, and oral hygiene habits are often influenced by geographic location or place of residence (El Tantawi *et al.*, 2024). Additionally, factors such as previous caries experience, fluoride exposure, dental defects, feeding practices, dentist-patient relationships, and patterns of dental visits also substantially contribute to caries risk and can be assessed using caries risk assessment tools (Schroth *et al.*, 2021).

The most extensively studied microorganisms associated with ECC include *Streptococcus mutans* and *Lactobacillus* species due to their acidogenic properties (Richards *et al.*, 2017). Furthermore, recent studies have identified specific fungal species that aggravate ECC, highlighting potential interkingdom interactions with bacteria in biofilm formation and acid production (Menon *et al.*, 2022). These microbes ferment carbohydrates into lactic acid, resulting in enamel demineralization and subsequent cavity formation. Behavioral factors, such as inadequate oral hygiene practices and frequent consumption of sugary foods, exacerbate the microbial burden, thus elevating the risk of ECC.

Host genetic factors contribute to ECC susceptibility through various mechanisms. The most investigated genetic contributors include genes associated with tooth formation, enamel composition, salivary flow rate, and immune responses to cariogenic bacteria (Ozturk, Famili and Vieira, 2010; Vieira, Modesto and Marazita, 2014; Lips *et al.*, 2017). Genetic variations in tooth formation genes may affect enamel strength and predispose teeth to demineralization. Additionally, variations in salivary composition and immune response genes can modulate the oral microbiome and influence the risk of caries. Variations in taste-related genes also significantly influence dietary preferences and potentially mediate interactions between the host and microbial peptides, suggesting host-microbe interplay (Behrens and Meyerhof, 2011; Medapati, Singh, *et al.*, 2021).

Understanding the intricate interactions among the microbial, socioeconomic, and genetic risk factors of ECC is crucial for developing targeted preventive and intervention strategies. The following sections comprehensively explore these factors, emphasizing their roles in ECC pathogenesis.

1.2 Socioeconomic, environmental, and behavioral factors in ECC

Socioeconomic, environmental, and behavioral factors, including dietary habits, tooth brushing practices, bedtime snacking, and access to early dental care, significantly contribute to ECC development. A substantial proportion of ECC cases are associated with frequent sugar consumption, especially from between-meal juices and packaged sweetened foods. The risk increases when bedtime snacking habits coincide with poor oral hygiene, such as not brushing before sleep, as reduced salivary flow during sleep exacerbates the condition. While prolonged breastfeeding is generally protective against ECC, nighttime breastfeeding extending beyond 12 months has been associated with an increased caries risk. Caregivers' health literacy and beliefs also play an essential role in determining children's dental care practices. Additionally, caregivers' stress responses, which influence their adoption of preventive oral health measures, are critical in shaping childhood caries experiences. Caregiver anxiety and depression, which are frequently outcomes of chronic stress, are similarly linked to ECC prevalence. Furthermore, individuals from socially disadvantaged communities often have lower self-efficacy, leading to the underutilization of available

preventive oral health services, which in turn results in a higher ECC occurrence among their children.

1.2.1 Sex

Many diseases exhibit sex-based differences in prevalence, severity, and underlying mechanisms, prompting researchers to investigate similar patterns in ECC. Efforts have been made to understand how biological, behavioral, and social factors contribute to sex-related differences in the risk and progression of ECC (Lukacs, 2008). Evidence regarding sex-based differences in ECC prevalence has been mixed; some studies have reported a higher prevalence among males than females, though without statistical significance (C. Acuña *et al.*, 2019; Papadaki *et al.*, 2021). In a meta-analysis examining sex-based ECC prevalence among Canadian children, five out of six studies found no significant differences, whereas the remaining study reported that male children were 38% more susceptible to caries (Schroth and Cheba, 2007). Despite inconsistent prevalence data, significant sex-related differences in microbial composition have been observed, with certain species being more frequently found in males (de Jesus *et al.*, 2020). These sex-based differences have often been attributed to behavioral variations between male and female children regarding hygiene practices and eating habits, rather than biological distinctions (Shaffer *et al.*, 2015; Amato *et al.*, 2023). However, some researchers have suggested that these differences may also be influenced by hormonal changes associated with female sex hormones (Lukacs, 2011).

1.2.2 Age

Age significantly influences the prevalence and severity of ECC. ECC may initiate soon after the eruption of the first tooth and continue to progress if not managed effectively during its early stages (Rosenblatt and Zarzar, 2002). The prevalence of ECC is age-dependent (Anderson *et al.*, 2021), while relatively low among children aged 2–3 years, ECC prevalence notably increases in the 4–6-year-old age group (El Tantawi *et al.*, 2018; Uribe, Innes and Maldupa, 2021). Older children have a greater number of erupted teeth, resulting in a longer exposure to cariogenic factors and consequently higher susceptibility to caries (Fernandes *et al.*, 2023). Additionally, research indicates that the frequency of sugary food intake increases with age, contributing to a higher caries prevalence (Li *et al.*, 2023). Moreover, independent

oral hygiene practices adopted by children can significantly impact their oral health, positively or negatively. Late detection of ECC also contributes to ambiguity in categorizing caries severity, complicating the classification of severe and non-severe ECC (Paglia, 2018). Studies have consistently demonstrated that early intervention through specialized programs, such as the "age-one" dental visit, is critical for ECC prevention (Finlayson *et al.*, 2018). Such early interventions and timely detection allow healthcare providers to deliver proper guidance to parents and caregivers, substantially reducing the risk of ECC development.

1.2.3 Place of residence

Differences in ECC prevalence between rural and urban settings have been attributed to multiple factors, including the higher consumption of junk food and sugars in urban areas and limited access to dental care, poor oral hygiene, and lack of parental education in rural areas (Mafuvadze, Mahachi and Mafuvadze, 2013; Narang *et al.*, 2016; Anil and Anand, 2017; Yadav *et al.*, 2022). These disparities are predominantly influenced by SES and access to nutritional resources. A study conducted in Britain reported similar findings, emphasizing the higher prevalence of caries in urban areas associated with feeding practices influenced by social and environmental factors (Rohnbogner and Lewis, 2018). Conversely, a study from Poland found a higher prevalence of ECC in rural areas, which was attributed to poor oral hygiene, nocturnal bottle feeding, and lower parental education levels (Olczak-Kowalczyk, Gozdowski and Kaczmarek, 2020). In Australia, increased ECC prevalence in rural communities has been linked to limited access to fluoridated water, inadequate dental care services, and higher sugar consumption (Ha *et al.*, 2020). Similarly, research from Colombia indicated that caries status was significantly worse in rural children, highlighting toothbrushing frequency, dietary habits, and socioeconomic factors as major contributors (Cerón-Bastidas, Suarez and Guauque-Olarte, 2018). Another factor influencing these disparities is the variation in access to healthcare services (Peña-Alegre *et al.*, 2023). A notable finding from China reported a high ECC prevalence in rural areas coupled with lower dental care indices, indicating healthcare disparities (Zhang *et al.*, 2016). Research from the United States has presented conflicting outcomes, demonstrating higher ECC prevalence in some rural regions, while a greater prevalence occurred in certain metropolitan areas (Kandel, Richards and Binkley, 2012).

These studies collectively suggest marked differences between residential settings, primarily driven by variations in dietary habits, oral hygiene practices, dental care accessibility, and socioeconomic disparities. Therefore, ECC prevention and management strategies must consider these differences in prevalence and distribution and be customized to address the unique requirements of diverse population groups.

1.2.4 The Early Childhood Oral Health Impact Scale (ECOHIS)

The Oral Health-Related Quality of Life (OHRQoL) in young children is assessed using a validated tool known as the Early Childhood Oral Health Impact Scale (ECOHIS) (Pahel, Rozier and Slade, 2007). The ECOHIS comprises a comprehensive set of questions divided into two sections: the Child Impact Section and the Family Impact Section. These sections address factors such as pain, discomfort, distress, emotional well-being, and financial burdens. The score ranges from 0 to 52, with a score of 0 indicating no impact on quality of life and 52 indicating the highest impact. Since its development in 2007, the ECOHIS has demonstrated consistent reliability, leading to its translation into multiple languages, including Swedish, Brazilian Portuguese, Chinese, Italian, Thai, and German. The widespread acceptance of the ECOHIS across diverse populations underscores its reliability as a measure for evaluating and comparing oral health impacts internationally (Lee *et al.*, 2009; Sabel *et al.*, 2024). ECC negatively influences OHRQoL, as indicated by higher ECOHIS scores in children with caries than in those without (Sabel *et al.*, 2024). ECC severity shows a strong correlation with elevated ECOHIS scores, reflecting significant impacts on the daily lives of children and their caregivers (Lee *et al.*, 2020; Lara *et al.*, 2022). Furthermore, ECOHIS scores are strongly associated with poor nutritional and low SES, factors known to increase ECC prevalence (Shrutha *et al.*, 2025).

1.2.5 Host nutrition

Numerous studies have explored the relationship between child nutrition and ECC outcomes. Malnutrition, specifically underweight conditions, has been associated with ECC (Pacey, Nancarrow and Egeland, 2010; So *et al.*, 2017; Feldens *et al.*, 2018). However, some studies have identified a positive association between body mass index (BMI) and ECC (Davidson *et al.*, 2016). This association may result from high sugar consumption, which is consistently

linked to ECC development, as indicated by a large number of studies (Moynihan and Petersen, 2004; Bencze *et al.*, 2021; Chou *et al.*, 2024; Sandy, Helmyati and Amalia, 2024). Cross-sectional surveys have revealed a strong correlation between sugar intake and increased caries prevalence among Australian children, although this association was significantly mitigated by the presence of fluoride in drinking water (Moynihan and Petersen, 2004; Armfield *et al.*, 2013; Sheiham and James, 2015).

Breastfeeding practices also influence ECC outcomes. A meta-analysis indicated that prolonged breastfeeding beyond 12 months is associated with a higher ECC risk, and bedtime breastfeeding further increases this risk (Shrestha *et al.*, 2023). Additionally, vitamin D deficiency may elevate the risk of ECC by influencing tooth development and promoting enamel defects (Williams *et al.*, 2021). Vitamin D's role in immune responses and antimicrobial peptide synthesis may also impact oral microbiome composition, further affecting ECC susceptibility (Seminario and Velan, 2016).

Nutrition education within maternal-child health programs and related health policies can also affect ECC prevalence, as demonstrated by studies emphasizing their role in ECC prevention (Sokal-Gutierrez *et al.*, 2016; Amalia *et al.*, 2022). The Healthy Eating Index (HEI) derived from the Third National Health and Nutrition Examination Survey (NHANES III) found that children with low HEI scores were more susceptible to severe ECC (S-ECC) than those with high HEI scores, representing healthier dietary practices (Nunn *et al.*, 2009).

However, studies on nutrition have not reached a consensus, largely due to inconsistent methodologies for measuring nutritional intake (Lui *et al.*, 2023). Consequently, recent efforts have advocated for standardized nutritional assessment tools, such as NutriSTEP, which can be generalized across geographically diverse populations (Randall Simpson *et al.*, 2008). Similar to the ECOHIS, NutriSTEP evaluates factors related to food and nutrient intake, growth, and physical activity among preschool children. It is a validated tool for the early identification of nutrition-related issues, with scores ranging from 0 to 68. Scores between 0-20 indicate low risk, 21-25 moderate risk, and scores of 26 or higher suggest a high risk of poor nutrition. Thus, NutriSTEP serves as an effective tool for examining the role of nutrition in ECC.

Many researchers have highlighted the importance of dietary patterns and eating behaviors in the occurrence of ECC. Hu et al. emphasized that adherence to the WHO-recommended dietary patterns from 6 to 12 months of age within a multi-ethnic Asian cohort protects against ECC compared to diets dominated by breastmilk, packaged foods, and seafood (Hu *et al.*, 2019).

1.2.6 Socioeconomic conditions

SES is consistently recognized as a major determinant of ECC. As previously mentioned, SES affects various aspects of life, influencing ECC outcomes, including nutrition, oral hygiene practices, access to dental care, and parental education and guidance. In the US, lower SES and food insecurity have been associated with higher rates of dental caries (Chi *et al.*, 2014), leading researchers to recommend social interventions that specifically focus on child nutrition. A consistent observation is that socioeconomic inequalities, low-income households, and individual poverty are strongly linked to ECC (Folayan *et al.*, 2020; Chou *et al.*, 2024; Dettori *et al.*, 2024; El Tantawi *et al.*, 2024; Maklennan *et al.*, 2024). Consequently, incorporating socioeconomic factors into ECC-related studies has been recommended (El Tantawi *et al.*, 2024).

Geographical locations and cultural differences further add to ECC-related challenges, along with national income disparities and income inequality (Maklennan *et al.*, 2024). Additionally, factors such as parental education, health knowledge and attitudes, parental distress, and sense of coherence, a measure of one's ability to handle life's challenges and stressors, were identified as significant correlates of ECC in a meta-analysis (Rai and Tiwari, 2018). Parents or caregivers managing children’s chronic conditions also frequently report higher dental anxiety and nicotine dependency, which are exacerbated by low SES (Knoblauch *et al.*, 2019). ECC-related mouth and jaw pain leads to difficulties in eating, associated reductions in body mass index, and poorer nutrition compared to healthy children. Consequently, ECC negatively impacts children's quality of life, resulting in missed school days and compromised academic performance. Furthermore, ECC causes parents or caregivers to miss work, creating a significant socioeconomic burden on society due to its high prevalence.

1.2.7 Other factors

In addition to socioeconomic and behavioral determinants, several other factors have been consistently evaluated in the context of ECC, including fluoride exposure, tooth brushing habits, frequency of dental visits, and the ongoing relationship between dentists and patients (Schroth *et al.*, 2021). Furthermore, factors such as the mode of childbirth, vaginal delivery or cesarean section, can influence the oral microbiome, potentially impacting ECC risk. However, current evidence regarding the association between mode of delivery and ECC is inconsistent; some studies have reported a weak association (Boustedt *et al.*, 2021), while others have found no significant relationship (Xiang *et al.*, 2024). Additionally, parental oral health literacy, public health policies, and accessibility of preventive dental services have been implicated in ECC prevalence (Hooley *et al.*, 2012). Beyond these behavioral and environmental factors, the oral microbiome remains a critical contributor to ECC development, as imbalances characterized by the overgrowth of cariogenic bacteria significantly promote its progression. This aspect is elaborated in the subsequent sections.

1.3 Oral microbiome and ECC

1.3.1 Microbiome definition

Microbial community is a general term that refers to a collection of microorganisms cohabiting within a single environment. These typically comprise multiple species capable of interacting with each other and their surroundings (Konopka, 2009). To standardize terminology in the field, the term microbiota has been proposed as the "assembly of microorganisms belonging to different kingdoms," including bacteria, archaea, fungi, and viruses, inhabiting a specific environment (Lederberg and McCray, 2001; Berg *et al.*, 2020). In contrast, the microbiome refers not only to the microorganisms but also includes their theatre of activity, encompassing structural elements, metabolites/signaling molecules, and surrounding environmental conditions (Berg *et al.*, 2020). However, the term "metagenome," has a narrower scope and specifically denotes the complete set of genetic material from the microbiota members within a given environment (Marchesi and Ravel, 2015; Berg *et al.*, 2020). Despite these distinctions, the terms microbiome and microbiota are frequently used interchangeably in the literature, particularly when collectively studying all microorganisms

within an environment using advanced sequencing techniques (Marchesi and Ravel, 2015; Berg *et al.*, 2020).

1.3.2 Strategies for investigating the microbiome

Traditional methods of studying microbiomes involve the *in vitro* cultivation of dental plaque microorganisms using selective media, followed by species identification through phenotypic features or first-generation genotypic methods, typically polymerase chain reaction (PCR) amplification and Sanger sequencing. These techniques are still employed when the research focus is limited to selected microbial species. However, with advances in next-generation sequencing (NGS), traditional approaches are no longer primarily used for comprehensive microbiome analyses. NGS provides a robust understanding of the intricate microbial diversity across various habitats (Caporaso *et al.*, 2011).

Currently, two primary NGS approaches, amplicon sequencing and metagenomic sequencing, are used in microbiome studies. Amplicon sequencing involves the amplification and sequencing of specific DNA regions common to certain microbial groups, followed by classification and quantification using reference databases. Typical targets include the highly conserved 16S rRNA gene for bacteria and archaea, and the internal transcribed spacer (ITS) region for fungi, both of which form essential components of the transcriptional machinery. Notably, the primers for bacteria and archaea differ, preventing their simultaneous amplification in the same sequencing run. Furthermore, differences in sequencing methods and selected regions for bacterial 16S and fungal ITS amplification introduce biases that often complicate comparisons across studies. Conversely, metagenomic sequencing employs a target-free approach, sequencing the entire DNA content within a sample, and thus requires comprehensive genome databases for analysis (Quince *et al.*, 2017). However, the high cost of metagenomic sequencing compared to amplicon sequencing makes it prohibitively expensive for large study designs.

Owing to its targeted nature, amplicon sequencing is cost-effective, simpler to analyze, and highly sensitive for selected microbial groups, whereas metagenomic sequencing facilitates the detection of novel organisms, enabling further exploration of their origin and phylogenetic relationships. Nevertheless, the simplicity and sensitivity of amplicon sequencing make it a

viable and widely used method. In the future, third-generation sequencing may become more prevalent because of its long-read capability, which simplifies taxonomic classification. Additionally, its low cost and minimal logistical requirements enhance its accessibility. Nonetheless, the convenience and sensitivity of amplicon sequencing ensure its continued relevance in microbiome studies. To address existing methodological inconsistencies, standardized protocols may be necessary for future research (Knight *et al.*, 2018).

1.3.3 Oral microbiome composition and dynamics

The human body has approximately the same number of microorganisms as host cells across various sites, including the gut, skin, oral cavity, and nasal passages (Sender, Fuchs and Milo, 2016). The microbiomes of these sites differ significantly. Distinct microbial communities inhabit the oral cavity, including supragingival and subgingival plaque, saliva, and buccal mucosa. While there are minor compositional differences among these communities, the eHOMD database reports approximately 800 species in the oral cavity, including both cultivable and non-cultivable species (Escapa *et al.*, 2018). At any given time, nearly half of these species may be present in an individual (Palmer, 2014). The oral microbiome is highly dynamic in nature and undergoes significant modifications from infancy to early childhood before stabilizing at approximately two years of age (Dzidic *et al.*, 2018).

Dental plaque is a structured biofilm on tooth surfaces that provides a habitat for microorganisms, some of which are cariogenic. A shift toward a more acidogenic and aciduric bacterial community within the plaque increases the risk of ECC. Additionally, factors such as diet, oral hygiene, and fluoride exposure influence plaque microbiome composition. Saliva plays multiple roles in the oral cavity, including buffering calcium and phosphate and providing antimicrobial components (e.g., lysozyme and immunoglobulins) (Humphrey and Williamson, 2001). It also serves as a reservoir for bacteria shed from dental plaque and reflects the microbial shifts occurring on tooth surfaces. Due to the significant differences in microbiota among the enamel, dentine, and subgingival sites, a single universal model cannot be applied. Nevertheless, the plaque microbiome directly reflects caries onset and progression, as it is the primary site for sugar metabolism and caries initiation.

Dental plaque contains the highest microbial concentration after the colon (Sender, Fuchs and Milo, 2016). While the mode of delivery influences the initial composition of the oral microbiome, it normalizes with age. Other factors affecting microbiome composition at later stages include breastfeeding habits and antibiotic treatments during infancy (Sender, Fuchs and Milo, 2016).

Under healthy conditions, the oral microbiome maintains a symbiotic or eubiotic relationship with the host. However, dysbiosis, an imbalance in this relationship, can lead to various dental diseases (Fakhruddin, Ngo and Samaranayake, 2019). Some microorganisms contribute to enamel remineralization through alkali production via the arginine deiminase pathway, whereas acidogenic bacteria disrupt this balance and aggravate enamel demineralization (Burne and Marquis, 2000). Therefore, caries is an endogenous infection resulting from ecological imbalance, characterized by an overabundance of cariogenic species leading to disease (Marsh, 2006; Seneviratne, Zhang and Samaranayake, 2011).

In children up to seven years of age, *Streptococcus*, *Lactobacillus*, and *Veillonella* are early colonizers, appearing between 3 to 6 months, while *Gemella*, *Rothia*, *Haemophilus*, and *Granulicatella* increase in abundance over time and are considered constant colonizers (Dzidic *et al.*, 2018). Late colonizers, appearing after one year, include *Actinomyces*, *Neisseria*, *Abiotrophia*, and *Porphyromonas* species. In breastfed neonates, *Candida* and *Streptococcus* species are less abundant in the oropharyngeal cavity (Pattaroni *et al.*, 2022), potentially due to higher levels of IgA and human milk oligosaccharides, which may protect against biofilm formation (Gonia *et al.*, 2015; Bertoldo *et al.*, 2017; Pattaroni *et al.*, 2022). In healthy adults, *Streptococcus*, *Veillonella*, *Actinomyces*, *Corynebacterium*, *Fusobacterium*, and *Rothia* constitute approximately 50% of the dental plaque species (Keijser *et al.*, 2008).

Compared to bacterial microbiome studies, research on the oral fungal microbiome (mycobiome) is limited. Recent studies have sought to characterize the oral mycobiome and identified diverse fungal compositions across oral sites. One of the first studies utilizing amplicon sequencing of the ITS region identified 85 fungal genera, with *Candida*, *Cladosporium*, *Aureobasidium*, *Saccharomycetales*, *Aspergillus*, *Fusarium*, and *Cryptococcus* being the most prevalent (Ghannoum *et al.*, 2010). Other studies have

identified multiple *Candida* species, *Saccharomyces cerevisiae*, and *Cladosporium velox* as predominant in the oral cavities of infants (Ward *et al.*, 2018). Notably, *Candida dubliniensis* has gained importance due to its association with oral candidiasis, oral lichen planus, periodontal disease, and dental caries (Diaz *et al.*, 2019; Molkenthin *et al.*, 2022). It has been suggested that *C. dubliniensis* may have been underreported in previous studies due to misidentification, given its high similarity to *Candida albicans* (Al-Ahmad *et al.*, 2016).

Most ECC studies have identified mutans streptococci as the primary cariogenic microorganisms because of their high acidogenic and aciduric properties. However, these studies were largely limited to bacterial species only. Recent investigations incorporating fungal species have identified *C. albicans* and *C. dubliniensis* to be associated with ECC. The acid produced by these species modulates the microenvironment of dental biofilms, attracting other acid-tolerant species, including *Scardovia wiggsiae*, bifidobacteria, and lactobacilli.

1.3.4 Bacteria in ECC

Most studies suggest that *S. mutans* is highly prevalent in ECC, which is both acidogenic and aciduric, and is considered the primary species responsible for caries (Ardiyantini, Pramesti and Gartika, 2023). The production of water-insoluble glucans, extracellular polysaccharides, and surface antigen proteins allows *S. mutans* to persist in saliva and dental biofilms (Al Shukairy *et al.*, 2006). The role of *Streptococcus* species in ECC is well documented (Aas *et al.*, 2005, 2008; Klein *et al.*, 2015). The presence of *S. mutans* in both surface caries and deeper dentinal lesions suggests its involvement in all stages of caries progression (Lin, Wang and Zhang, 2024). Children susceptible to *S. mutans* infections also exhibit a higher risk of developing caries in adulthood. However, not all carious lesions contain *S. mutans*, and some individuals harboring *S. mutans* remain caries-free, raising questions about its role as a primary colonizer and suggesting a more complex interplay of microbial species in caries development (Fakhruddin, Ngo and Samaranayake, 2019).

Another closely related species found in ECC is *Streptococcus sobrinus*, which contributes to the severity of ECC and is often co-detected with *S. mutans* (Palmer *et al.*, 2010; Colombo *et al.*, 2017). Other *Streptococcus* species implicated in caries include *S. sanguis*, *S. gordonii*, *S.*

mitis, and *S. oralis*. Similarly, *Lactobacillus* species, known for their high acidogenicity and acid tolerance, are frequently found at elevated levels in children with ECC (Al Shukairy *et al.*, 2006). Notably, higher levels of *Lactobacilli* have also been identified in the mothers of children with ECC, suggesting vertical transmission.

Other bacterial species associated with ECC include *Veillonella* species, such as *V. dispar*, *V. atypica*, and *V. parvula* (Teng *et al.*, 2015). Although *Veillonella* species are not acidogenic and thus not considered primary caries pathogens, they may play a role in biofilm formation (Kanasi *et al.*, 2010). *Scardovia wiggsiae* has been detected independently from *S. mutans* in both early white spot lesions and advanced caries (Kressirer *et al.*, 2017; Jasim Al-Adool, Muwafaq Hamdoon and Faiz Hamodat, 2024). Its high acidogenicity and acid tolerance make *S. wiggsiae* a potential risk factor for caries. Additionally, *Actinomyces* and *Fusobacterium* species have been linked to ECC through their roles in biofilm formation and caries initiation (Becker *et al.*, 2002; Tu *et al.*, 2022).

Prevotella and *Alloprevotella* species, which are commonly associated with periodontitis, have been frequently detected in ECC lesions (Jiang, Zhang and Chen, 2013; Ardiyantini, Pramesti and Gartika, 2023). These bacteria appear to interact synergistically with *Veillonella* and *Fusobacterium*, forming an anaerobic network that promotes the progression of carious lesions. At the phylum level, ECC is associated with a decrease in *Actinobacteria* and an increase in *Firmicutes* (Gross *et al.*, 2012; Jiang *et al.*, 2014; Agnello *et al.*, 2017).

Several bacterial taxa associated with oral health have also been identified. For example, *Corynebacterium matruchotii*, a gram-positive bacillus, utilizes lactate in biofilms, reducing acidity and thereby promoting a healthy oral microbiome (Li *et al.*, 2022; Almeida *et al.*, 2023). Other species with protective effects include *S. sanguinis* and *S. gordonii*, which thrive under aerobic conditions and produce hydrogen peroxide, inhibiting biofilm formation and bacteriocin production by *S. mutans* (Ogawa *et al.*, 2011; Pham and Zeng, 2024).

Species from the *Neisseria* genus are associated with both health and dental caries. While *Neisseria bacilliformis* has been linked to caries, species such as *N. oralis*, *N. mucosa*, and *N. pharyngis* are associated with oral health. This highlights the importance of these species in maintaining homeostasis of the plaque microbiome.

1.3.5 Fungi in ECC

The role of fungi in ECC has been a topic of ongoing research, gaining attention in recent years due to their crucial involvement in the disease (Beena *et al.*, 2017; Fechney *et al.*, 2019; Cui *et al.*, 2021). Evidence suggests significant alterations in fungal composition and diversity in ECC, with reduced diversity indicating the enrichment of keystone species (Fechney *et al.*, 2019). *Candida* species, in particular, have been found to be more prevalent in ECC than in caries-free (CF) children, with their abundance increasing with caries severity (Beena *et al.*, 2017; Cui *et al.*, 2021). *Candida* species were detected in 40–60% of patients with caries (Macfarlane, 1990).

Normally, *Candida* species exist as commensals in the oral cavity of healthy individuals. However, these infections are considered opportunistic and predominantly affect immunocompromised individuals. Initially, *C. albicans* was identified as a key player in the development of ECC (Marchant *et al.*, 2001; Rozkiewicz *et al.*, 2006; Raja, Hannan and Ali, 2010). Later, other species, such as *C. dubliniensis*, were also found to be associated with ECC (Kneist *et al.*, 2015; Al-Ahmad *et al.*, 2016). Due to their close genetic similarity and shared phenotypic traits, *C. dubliniensis* was initially misidentified as *C. albicans* (Sullivan and Coleman, 1998; Al Mosaid *et al.*, 2001; McManus *et al.*, 2008; Al-Ahmad *et al.*, 2016).

A meta-analysis revealed that the presence of *C. albicans* increases the likelihood of developing caries by 5–7 times (Rozkiewicz *et al.*, 2006; Raja, Hannan and Ali, 2010; Xiao *et al.*, 2018). *C. albicans* exhibits high acidogenic potential, fermenting glucose and maltose to produce acids (Misra, Tahmassebi and Brosnan, 2007). Recent studies have identified the roles of both *C. albicans* and *C. dubliniensis* in caries. Due to their frequent co-occurrence in ECC cases, they have been proposed as "keystone pathogens" in ECC (Cvanova *et al.*, 2022). The same study reported that *C. dubliniensis* was nearly twice as prevalent in severe ECC cases (Cvanova *et al.*, 2022). The ability of *Candida* species to form robust biofilms, produce acids, and secrete hydrolytic enzymes underscores their active role in the progression of dental caries. However, whether fungal species act as true pathogens, secondary agents, or mere passenger species in ECC remains unclear (Pereira *et al.*, 2017).

1.3.6 Interkingdom interactions between bacteria and fungi

Interkingdom interactions between fungi and bacteria in ECC pathogenesis are a significant area of research, playing a crucial role in disease progression (Pereira *et al.*, 2017; Baraniya *et al.*, 2020; Tu *et al.*, 2022). These interactions, which can be symbiotic or competitive, involve both intra- and interkingdom species of bacteria and fungi. Their increased presence in dental biofilms is associated with increased cariogenicity and resistance to treatment.

Among these interactions, the most extensively studied are those between *S. mutans* and *C. albicans*, as these species frequently coexist in dental plaque biofilms, particularly in ECC (Baraniya *et al.*, 2020; Liu *et al.*, 2023; Siddaiah and Sinha, 2023; Weng *et al.*, 2024). These interactions enhance acid production and biofilm formation through synergistic mechanisms. Additionally, *Candida* species have been reported to be associated with other *Streptococcus* species in the oral cavity, such as *S. oralis*, S. sanguinis, and *S. gordonii*, as well as with root caries-associated pathogens such as *Actinomyces viscosus* (Diaz *et al.*, 2012; Deng *et al.*, 2019; Dige and Nyvad, 2019).

Beyond *Candida*, other fungal species have been identified in cross-kingdom associations. Positive correlations have been observed between health-associated taxa, such as *Leptospira* and *Streptococcus*, whereas negative correlations have been found between *Naganishia*, *Leptospira*, and *Prevotella*, the latter being primarily caries-associated (Tu *et al.*, 2022). *Malassezia globosa*, which is associated with healthy teeth, was negatively correlated with cariogenic bacteria, particularly *S. mutans* (Baraniya *et al.*, 2020).

Moreover, *Corynebacterium durum* modulates interkingdom interactions between *S. mutans* and *C. albicans*. Interestingly, *C. durum* has been found to disrupt the interactions between *S. mutans* and *C. albicans*, suggesting its potential as a beneficial commensal species with therapeutic implications for polymicrobial diseases (Treerat *et al.*, 2024). Understanding the mechanisms behind these interactions presents opportunities for developing novel therapeutic strategies. For instance, targeting GtfB-mannan interactions using mannan-degrading enzymes can disrupt biofilm assembly and reduce acidity without microbicidal effects, offering a potential approach for the management of dental caries (Kim *et al.*, 2021).

1.4 Taste receptors and ECC

1.4.1 Heritability of caries

The role of taste genetics in dental caries or ECC involves a complex interplay of genetic, environmental, and behavioral factors. Genetics has long been suggested as a significant contributor to ECC (Fisher-Owens *et al.*, 2007). Additionally, advancements in tools for studying host genetics have led to renewed interest in identifying genetic markers associated with ECC susceptibility (Wang *et al.*, 2010; Abbasoğlu *et al.*, 2015; Holla *et al.*, 2015; Tim Wright, 2019).

To investigate the genetic components of dental caries in children, several studies have compared their prevalence in monozygotic (MZ) and dizygotic (DZ) twins. These studies have consistently reported a higher prevalence of caries in MZ twins than in DZ twins, suggesting a genetic basis for dental caries (Lovelina, Shastri and Kumar, 2012; Anu *et al.*, 2018; Dos Anjos *et al.*, 2023). However, heritability estimates vary, ranging from 24% to 63%, with a concordance rate of 0.64 in MZ twins and 0.44 in first-degree relatives (Kuppan *et al.*, 2017; Haworth *et al.*, 2020; Shrestha *et al.*, 2024). Furthermore, heritability differs across various tooth surface clusters, ranging from 41.9% to 54.3% (Haworth *et al.*, 2020).

1.4.2 Genetics factors in ECC and Caries

Developmental defects of enamel (DDE) are major risk factors for ECC, where enamel layering is deficient in quantity, a condition known as enamel hypoplasia. These defects result in hypomineralized sites on the tooth surface with reduced mineral content, leading to structurally compromised or rough surfaces that facilitate microbial colonization. Some cases of DDE can be attributed to genetic factors (Smith *et al.*, 2017; Sharma *et al.*, 2023), while others result from congenital medical conditions, such as premature birth, infancy infections, or exposure to certain drugs, particularly cytotoxic agents (Arrow, 2009; Krasuska-Sławińska *et al.*, 2016; Alshehhi *et al.*, 2020).

Genetic polymorphisms in *AMBN*, *AMELX*, *TUFT1*, and *KLK4* affect amelogenesis and increase the susceptibility to ECC (Smith *et al.*, 2017; Sharma *et al.*, 2023). Dentin, produced by odontoblasts, is less mineralized than enamel and shares protein composition similarities

with bones (Narayanan *et al.*, 2001; Yu and Klein, 2020). Genetic mutations affecting bone formation (e.g., *GNAS*, *TAS1R3*, and *FFAR1*) may also disrupt dentin mineralization (Zhang *et al.*, 2017; Luo *et al.*, 2019; Nguyen *et al.*, 2021).

Other factors contributing to ECC include genes related to tooth structure and saliva production. For example, *PRH1* and *PRH2* polymorphisms in proline-rich protein receptors categorize children into different susceptibility groups. Highly susceptible children develop caries due to autoimmune conditions, whereas those in low-susceptibility groups acquire it primarily due to poor oral hygiene and nutritional practices (Strömberg *et al.*, 2017). Additionally, a direct correlation has been identified between ferritin levels and dental caries, with elevated ferritin levels observed in children with ECC (Kotian and Gurunathan, 2021). Several immune response genes have also been investigated for their role in ECC (Aruna *et al.*, 2023). Notably, genetic variants of *ALOX15*, *TRAV4*, and *TNF* have been associated with increased susceptibility to ECC (Aruna *et al.*, 2023).

Genome-wide association studies (GWAS) have been instrumental in identifying novel loci associated with ECC. Numerous GWAS have explored the role of host genetics in dental caries by identifying specific variants linked to disease susceptibility. In 2018, a consortium-based study investigating both primary and permanent dentition identified intronic variants, *ALLC*-rs1594318 for primary teeth and *NEDD9*-rs7738851 for permanent teeth (Haworth *et al.*, 2018). However, consortium-wide heritability estimates were significantly lower than those from within-study estimates for both primary and permanent dentitions. A GWAS conducted in a Japanese population identified two additional intronic variants, rs10974056 in *GLIS3* and rs4801882 in *SIGLEC5*, which are associated with dental caries (Nogawa *et al.*, 2024).

A recent meta-analysis of five cohorts targeting ECC provided suggestive evidence for genes previously associated with caries in the primary dentition. This cohort was also analyzed using transcriptome-wide association studies (TWAS), identifying genes, *CDH17* (cadherin 17) and *SMIM10L1* (small integral membrane protein 10-like 1), which were predicted to be associated with ECC (Orlova *et al.*, 2022). Another GWAS conducted in a large multiethnic population identified several genes associated with caries in both primary and permanent

dentition, including multiple bitter taste receptor genes at suggestive significance thresholds. However, most identified loci were intronic or located in non-coding regions (Alotaibi *et al.*, 2021). A multi-ancestry GWAS identified several genomic loci associated with ECC, explaining 24% of the phenotypic variance among unrelated individuals. Given the profound influence of environmental factors on dental caries, some recent GWAS have also focused on genotype-by-environment interactions (Zou *et al.*, 2023; Shrestha *et al.*, 2024). These studies have highlighted how ECC-associated genes interact with environmental conditions, including sex, socioeconomic status, and oral hygiene behaviors.

The role of taste genetics has been a focal point in several studies due to the multiple functions of taste receptors, including taste perception, innate immunity, and bacterial recognition. A recent meta-analysis of GWAS identified multiple taste receptors associated with ECC (Orlova *et al.*, 2022). Given the strong genetic influence on ECC, it is essential to explore how specific genetic factors, such as taste receptors, contribute to caries susceptibility. The following sections discuss the broader significance of taste receptors, their molecular mechanisms, and their direct role in ECC development.

1.4.3 Molecular Mechanisms and Functions of Taste Perception

In humans, taste perception involves the detection of chemical stimuli through taste receptors on the cell surfaces. There are five primary taste modalities: sweet, salty, sour, bitter, and umami tastes. The molecular mechanisms of these taste receptors are diverse, with three T1Rs mediating sweet and umami tastes, whereas 25 T2Rs are responsible for bitter taste perception. These proteins are encoded by the TAS1R and TAS2R gene families and designated as T1Rs and T2Rs. Both T1Rs and T2Rs belong to the G-protein-coupled receptor (GPCR) family. Upon ligand binding, these receptors activate specific intracellular pathways that are essential for signal transduction. For example, the PLCβ2/IP3 signaling pathway responds to sweet, bitter, and umami compounds by releasing calcium from internal stores and transmitting signals to the brain (Zhang *et al.*, 2003).

In contrast, sour and salty tastes are mediated by ion channels. Sour taste is detected by proton-sensitive channels, most notably Otopetrin 1 (Otop1), whereas salty taste is primarily mediated by the epithelial sodium channel (ENaC) (Zhang *et al.*, 2019; Taruno and Gordon,

2023). Additionally, fatty acid and carbonic anhydrase receptors have been proposed as candidates for additional taste modalities, with research suggesting that they may represent distinct tastes such as "fat taste" and "carbonation," respectively (Chandrashekar *et al.*, 2009; Besnard, Passilly-Degrace and Khan, 2016). The fatty acid translocase *CD36* and *GPR120* have been identified as potential receptors for fat taste (Cartoni *et al.*, 2010).

Polymorphisms in taste receptors can influence taste sensitivity, thereby affecting dietary preferences, which may contribute to various health conditions, such as obesity, depression, and certain cancers (Shaji and Saraswathy, 2023).

1.4.4 Importance of taste receptors

Taste receptors play a crucial role in human survival, nutrition, and overall health by influencing dietary choices and, consequently, overall nutrition. Dysfunction in taste perception can contribute to malnutrition, obesity, and other metabolic disorders. Among the five basic taste modalities, bitter taste receptors (T2Rs) detect a wide range of potentially harmful substances—peptides, alkaloids, terpenoids, and amines, and thus serve as an important first line of defense against ingesting toxins.

In addition to their role in taste perception within the oral cavity, taste receptors are also expressed in other tissues, such as the gastrointestinal tract, reproductive organs, and respiratory tract. This widespread distribution suggests that their functions extend beyond traditional taste perception, playing roles in metabolic regulation and host defense against pathogens (Behrens and Lang, 2022; Li *et al.*, 2024). Previous studies have linked these receptors to various biological functions (Clark, Liggett and Munger, 2012). For instance, T2Rs contribute to innate immunity in epithelial and ciliated cells that interact with the environment. These receptors are activated by microbe-derived quorum-sensing molecules (QSM), triggering the production of bactericidal molecules (R. J. Lee *et al.*, 2012). Thus, T2Rs play a recognized role in bacterial recognition and chemosensing, contributing to oral immunity (Medapati, Bhagirath, *et al.*, 2021; Medapati, Singh, *et al.*, 2021). The broad distribution of T2Rs in extraoral tissues and the availability of well-established natural and synthetic agonists make these receptors attractive targets for future drug development (Clark, Liggett and Munger, 2012).

1.4.5 Taste receptors and caries

The most direct influence of taste genetics on ECC may be through dietary preferences mediated by taste-related genetic polymorphisms (Wendell *et al.*, 2010; Öter *et al.*, 2011; Jurczak *et al.*, 2020). This notion suggests that children with heightened sensitivity to sweet taste may prefer sugary foods, increasing the risk of caries in the presence of inadequate oral hygiene (Ashi *et al.*, 2019). Furthermore, genetic polymorphisms in all three TAS1Rs have been associated with the risk of caries (Haznedaroğlu *et al.*, 2015; Arid *et al.*, 2021). However, the same study also reported that nutritional status was not associated with genetic polymorphisms and dental caries.

A meta-analysis confirmed that polymorphisms in *TAS1R2*, which influences sweet taste acceptance, are associated with dental caries (Motahari, Molaei and Adhami, 2024). Additionally, a meta-analysis of multiple GWAS and TWAS identified the bitter taste receptor genes, *TAS2R43* and *TAS2R14*, as being strongly associated with ECC (Orlova *et al.*, 2022). Notably, *TAS2R14* showed one of the strongest associations and was also involved in innate immunity and antimicrobial defense through the recognition of competent stimulation peptides released by cariogenic bacteria.

Moreover, *TAAR6*, an olfactory receptor specific for trace amines, has been identified in association with ECC, alongside the taste receptor activity gene set (Orlova *et al.*, 2022; Shrestha *et al.*, 2024).

1.5 Host genetics and microbiome association

1.5.1 Host genetics and oral microbiome

While environmental and behavioral factors primarily shape the oral microbiome, host genetics also contribute to its composition (Freire *et al.*, 2020). Several gene loci, including *IL32*, *GALK2*, and *CELF4*, have been linked to the presence of *S. mutans* and ECC severity (Meng *et al.*, 2019). A 2005 study by Slaton et al. demonstrated a significant association between tuftelin (*TUFT1*) and *S. mutans*, explaining 27% of the variation in ECC severity through their interactions. Additionally, a higher similarity in the oral microbiome was observed in MZ twins than in DZ twins, suggesting a genetic component in the microbiome

composition (Zheng *et al.*, 2018). The same study also indicated an interaction between host genetics and dietary habits, as sugar consumption influenced the microbiome composition.

Conversely, Gomez et al. (2017) reported that while certain microbial species exhibit higher heritability, those associated with ECC do not follow this trend (Gomez *et al.*, 2017). The oral microbiome composition in MZ and DZ twins remained similar in the caries cohorts, suggesting that host genetics has a limited influence on cariogenic microbiota. Similar conclusions were drawn by Blostein et al., who found that the presence of cariogenic bacterial community state types (CSTs) was significantly higher in ECC, regardless of genetic susceptibility (Blostein *et al.*, 2023).

The role of genetics in ECC has long been recognized. Moreover, individuals with hypoplastic amelogenesis imperfecta, a hereditary disorder characterized by enamel malformations, exhibited lower caries experience than those without this condition. This was attributed to the enrichment of *Bacillus* spp. and lower levels of Streptococci, likely due to a more alkaline oral pH (Kammoun *et al.*, 2019).

1.5.2 Taste genetics and oral microbiome

Genetic polymorphisms in taste and taste-related genes involved in taste signaling pathways influence the composition of the oral microbiome (de Jesus *et al.*, 2022). Variants in *PLCB2* and *TAS2R4* have been found to be protective, whereas variants in *TAS2R38* are associated with cariogenic microbes. Another study demonstrated that genetic variants influencing sensitivity to sweet, salty, and bitter tastes are correlated with differences in habitual food consumption and oral microbiota composition (Cattaneo, Riso, *et al.*, 2019). Individuals with low sensitivity to sweet taste preferred sweet foods, whereas those with high sensitivity to bitter taste consumed more carbohydrates and high-energy foods.

Variations in sugar preference genes (*TAS1R1* and *GNAT3*) are linked to differences in oral microbiota clusters, with higher sucrose intake associated with acidogenic bacterial species (Esberg *et al.*, 2020). Individuals on protein-rich diets exhibited an abundance of *Clostridia*, whereas those on vegetable-rich diets were enriched with *Prevotella*. Recent studies have also highlighted the interactions between ECC-associated microbial species and taste

receptors (Medapati, Singh, *et al.*, 2021). These interactions contribute to host defense by regulating the growth of specific microbial species. For instance, the bitter taste receptor *TAS2R14* of the T2R family plays a role in innate immunity and antimicrobial defense, influencing oral microbial composition (Orlova *et al.*, 2022).

1.6 Knowledge gaps and thesis contribution

Despite significant advances in the understanding of ECC, there are still important areas that require exploration. One primary challenge in microbiome research is the inconsistency in outcomes regarding species identified as associated with ECC or healthy tooth conditions. A key limitation is determining whether these variations represent actual differences in disease etiology or arise from technical discrepancies in sample collection, DNA extraction, and sequencing methodologies. These differences introduce batch effects, complicating cross-study comparisons and hindering the establishment of robust microbial signatures. To minimize these discrepancies, a unified methodology from the start to the end is required. Although variations arising from sample collection and preprocessing are irreversible, *in silico* methods and bioinformatics tools applied to sequenced data can be standardized, potentially mitigating inconsistencies and improving the robustness of the findings. Such an approach could elucidate microbial differences across cohorts, facilitating the identification of shared and unique microbial factors influencing ECC development in diverse populations. Additionally, although individual studies have identified potential microbial contributors to ECC, the lack of consensus on a core set of microbial species introduces uncertainty in defining microbial biomarkers for diagnostic and therapeutic applications.

Another critical gap is the limited exploration of species interactions, particularly interkingdom interactions. Existing studies have primarily focused on a small number of species dominating the ECC microbiome, overlooking the broader multispecies ecological dynamics. Furthermore, research has predominantly concentrated on microbial changes associated solely with ECC or caries-free conditions, neglecting microbial shifts driven by sociodemographic and behavioral factors. As discussed earlier, many sociodemographic and behavioral factors influence microbial composition; therefore, a deeper understanding of

these variables and their associated microbial shifts is essential for distinguishing changes specifically related to ECC.

Moreover, although taste genetics has been implicated in ECC susceptibility, most studies investigating these associations rely on microarray studies constrained by fixed panels of genetic variants, limiting the identification of novel variants linked to ECC. To address this gap, deep sequencing methods can be useful in capturing previously uncharacterized genetic variants. Despite the growing interest in the interplay between host genetics and microbiome composition, studies integrating genetic variants and microbiome data from the same samples remain scarce. This lack of integrated investigations within the same cohort hinders a comprehensive understanding of how host genetics shape microbiome composition and its collective role in ECC etiology.

Lastly, studies employing multiomics data with incorporating microbiome, host genetics, and environmental factors, to investigate ECC remain underexplored, limiting insights into their relative contributions to ECC. This knowledge gap restricts the development of approaches useful for therapeutic interventions and policymaking. Addressing these gaps is crucial for advancing a holistic understanding of ECC pathogenesis and developing targeted preventive and therapeutic strategies.

Due to the knowledge gaps in the existing literature and the limitations of prior work, this thesis sets out to address several of these questions. The specific objectives, along with the hypotheses and rationale designed to address these research gaps, are outlined in the following chapter.

Chapter 2: Thesis Objectives and Rationale

2.1 Aim 1

To integrate data from multiple microbiome studies to enhance the understanding of microbial patterns associated with ECC and improve the robustness of ECC prediction.

Rationale: Numerous studies have sought to identify caries-causing microbes through metagenomic analysis; however, most are limited by their small sample sizes. Effective disease classification using metagenomics requires a larger sample size. This can be achieved by consolidating the data from existing studies. Data consolidation introduces heterogeneity, which can be minimized using batch correction methods. Machine learning methods can identify patterns in complex, high-dimensional data. Therefore, by applying suitable machine learning techniques to analyze cross-cohort data, ECC classification accuracy and the generalizability of predictive models across cohorts can be improved, facilitating the identification of robust biomarker species for ECC.

Hypothesis: The application of machine learning techniques to cross-cohort oral microbiome profiles can improve ECC classification and identification of common microbial biomarkers across studies.

Objectives:

1.1: Design a unified pipeline to analyze raw microbiome data from multiple previously published studies.

1.2: Compare differentially abundant species across datasets to obtain a consensus.

1.3: Apply machine learning methods to evaluate ECC prediction by combining multiple datasets.

2.2 Aim 2

To characterize microbial profiles and interkingdom interactions associated with ECC and assess their associations with sociodemographic and behavioral determinants.

Rationale: This approach helps elucidate potential species interactions, particularly interkingdom interactions between bacteria and fungi in ECC. As bacterial and fungal data

originate from the same samples, measuring the correlations in their relative abundances aids in identifying co-occurrence patterns. Additionally, investigating how various sociodemographic and behavioral factors modulate microbiome composition and their overall impact on ECC would enhance our understanding of ECC pathogenesis.

Hypothesis: Interkingdom interactions within the oral microbiome contribute to ECC development and progression. These interactions can be studied by comparing shifts in the relative abundances between the CF and ECC samples, and sociodemographic and behavioral factors influence ECC outcomes and oral microbiome composition.

Objectives:

2.1: Analyze the microbiome in a large cohort to identify microbial differences and modulations in interkingdom interactions between ECC and CF conditions.

2.2: Investigate associations of various sociodemographic and behavioral factors with ECC and oral microbiome composition.

2.3: Use machine learning models to evaluate the predictive performance for ECC outcomes using combined bacterial and fungal data and identify species important for ECC.

2.3 Aim 3

To evaluate the relationships and interactions among taste genetics, oral microbiome, and environmental factors influencing susceptibility to ECC.

Rationale: Variations in TAS1R2 and TAS2Rs genes have been linked to increased caries risk and overabundance of various bacterial genera (de Jesus et al., 2022; Kulkarni et al., 2013). Significant progress has been made in understanding genetic and microbiome associations over the past decade, driven by technological advancements. However, the interactions between host genetics and the ECC microbiome remain understudied. We focused on taste genetics because of its potential influence on taste preferences and biological interactions with microbial molecules. The outcomes of this study could facilitate the prediction of individual susceptibility to caries, thereby improving personalized disease management and care.

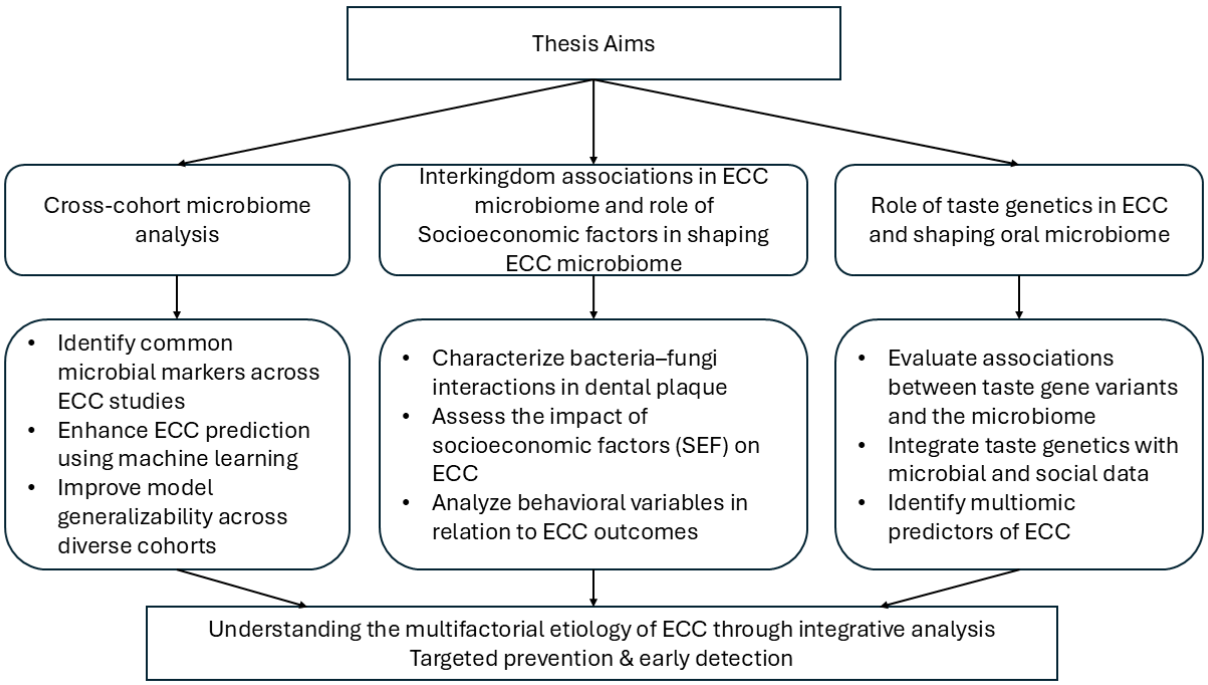
Hypothesis: Variations in taste-associated genes may influence the abundance of microorganisms in the oral cavity.

Objectives:

3.1: Identify associations between taste genes and genetic variants with ECC.

3.2: Investigate the interactions between taste genetic variants and oral microbiome composition through linear association and mediation analyses.

3.3: Combine microbiome, taste genetics, and environmental variables to assess their predictive performance for ECC and determine their relative importance in ECC development.



Thesis overview

Chapter 3: Study Population and Data Sources

3.1 Study population

The samples for the study were collected in a cross-sectional manner, including both children with ECC and CF, between December 2017 and July 2022. Samples from children with ECC were considered cases, whereas those from children with CF served as controls. Over this period, 554 children were recruited; however, data from only 538 children were used for analysis due to an age cut-off of 72 months (6 years). Other exclusion criteria included recent antibiotic use and children who did not satisfy the ECC case definition according to the AAPD (AAPD, 2017). According to the AAPD guidelines, participants with one or more decayed, missing (due to caries), or filled tooth surfaces in any primary tooth were considered to have ECC. Most children included in the study had severe ECC, and their samples were collected before surgical appointments at the Misericordia Health Centre (MHC), the Children's Hospital Research Institute of Manitoba (CHRIM), and various community dental clinics in Winnipeg, Manitoba, Canada. Additionally, CF children were recruited from community clinics in Winnipeg. CF status was determined by clinical examination performed by an experienced dentist (R.J.S.) using artificial light, without radiographs, with a dmft (decayed, missing, and filled primary teeth) index equal to zero, indicating no decayed, missing, or filled primary tooth surfaces, and no incipient (non-cavitated) lesions. Written informed consent for participation in the study was obtained from the parents or legal caregivers. The study protocol was approved by the University of Manitoba's Health Research Ethics Board (Ethics number HS24058 [H2020:310]) and the Misericordia Health Centre.

3.2 Participants' information

An interviewed questionnaire, proctored by members of the study team, was completed by parents or caregivers, gathering information including the child's age, sex, demographic characteristics, place of residence (urban or rural), nutritional habits, oral hygiene practices, mode of birth, and other variables as summarized in **Table 5.1**. Socioeconomic information was represented by three different variables: material deprivation score, SEFI, and social deprivation score, which are all measures of socioeconomic inequities derived from Manitoba Centre for Health Policy (MCHP) data based on participants' postal codes (Metge *et al.*, 2009).

The SEFI is calculated from Canadian Census data, combining four areas: unemployment rate, average household income, proportion of single-parent households, and proportion of the population without high school graduation. When interpreting the SEFI, a lower score indicates a more favorable socioeconomic status, while a score greater than 0 reflects less favorable socioeconomic conditions. Information about gender was not collected, as the participants were toddlers and preschoolers. Additionally, information on ancestry and race was not collected. Limited ethnicity data were collected but not included in the analyses due to the lack of ethics approval for their use.

3.3 Sample collection

Two samples, supragingival plaque (dental plaque) and buccal mucosa, were collected from each participant by a dentist or research staff member. To avoid contamination by food or blood, samples were collected at least 30 minutes after eating or drinking and prior to surgery for children undergoing surgical procedures. Dental plaque samples were collected using a sterile interdental brush from all tooth surfaces, irrespective of cavitated or non-cavitated status. Oral swab samples were collected using sterile polyester-tipped applicators (Fisher Scientific) from the buccal mucosa and anterior floor of the mouth, under the tongue. Both samples were immediately placed in 1 mL RNAprotect Reagent solution (Qiagen, Cat. #76506, Hilden, Germany) and stored at -80°C until DNA extraction.

3.4 Microbiome data

Dental plaque samples were used for microbiome sequencing in this study. Total DNA was extracted from these samples using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) or DNeasy PowerSoil Kit (Qiagen), following the manufacturer's protocol. When using the QIAamp DNA Mini Kit, an additional lysozyme treatment step was performed before DNA extraction, using 20 µg/ml lysozyme in a buffer containing 20 mM Tris HCl (pH 8), 1.2% Triton X-100, and 2 mM EDTA.

The extracted DNA samples were loaded onto 96-well plates and shipped on dry ice to the McGill University–Génome Québec Innovation Center for library preparation and sequencing. Sequencing was performed using either the MiSeq or NovaSeq6000 platforms (Illumina Inc., San Diego, CA, USA) with paired-end sequencing (250 bp × 2). Amplicon sequencing targeted

the V4 hypervariable region of bacterial 16S rRNA and fungal ITS1 (internal transcribed spacer-1). Primers used for the 16S rRNA V4 region were 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). Primers targeting the ITS1 region were ITS1-30 (5'-GTCCCTGCCCTTTGTACACA-3') and ITS1-217 (5'-TTTCGCTGCGTTCTTCATCG-3'). Raw sequencing data were processed using QIIME2 to generate an amplicon sequence variant (ASV) table containing the microbial abundance data. Further details on sequencing data processing are provided in Chapter 5 (Methods Details).

3.5 Taste genetics data

Buccal mucosa samples were used to sequence 55 candidate genes selected based on their roles in taste signal transduction, including receptors for bitter, sweet, umami, salt, sour, carbonation, and fatty acids, as well as some associated signaling proteins (**Table S6.1**). DNA was isolated using the QIAamp Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Candidate genes were amplified using primers designed specifically for the Fluidigm Access Array technology platform (Fluidigm, South San Francisco, CA, USA) to facilitate Illumina sequencing. These primers were designed to fully cover the exonic regions and include some flanking intronic regions of the selected genes, generating amplicons approximately 250–350 base pairs in length, compatible with 250 × 2 paired-end sequencing. Further details regarding the preprocessing steps to identify genetic variants in taste-related genes are provided in the Methods section in Chapter 6.

3.6 Data from external studies/sources

The data used for the cross-cohort analysis in Chapter 4 consisted partly of samples from our own study, owing to the limited number available at the time of analysis, and partly from datasets obtained from previously published studies. The 80 samples included in our study were previously analyzed and published by V.C.J. (de Jesus *et al.*, 2020). External datasets were derived from studies by Agnello et al. (Agnello *et al.*, 2017), Gomez et al. (Gomez *et al.*, 2017), Teng et al. (Teng *et al.*, 2015), and Kalpana et al. (Kalpana *et al.*, 2020). Details regarding the number of participants, data sources, and additional information are summarized in **Table 4.1**. The inclusion criteria for selecting external studies for comparison with our dataset were as follows: studies designed as case-control investigations comparing CF and ECC; studies

utilizing 16S rRNA sequencing, irrespective of hypervariable regions, with accessible raw data; and studies analyzing dental plaque samples. Data were downloaded from the NCBI repositories using Fasterq-dump (NCBI SRA-toolkit). Some raw sequence data were obtained directly from the authors upon request, as indicated **Table 4.1**. The preprocessing steps for these datasets are described in the Methods section of Chapter 4.

Chapter 4: A Cross-Cohort Analysis of Dental Plaque Microbiome in Early Childhood Caries

Mohd Wasif Khan, Daryl Fung Lerh Xing, Robert J. Schroth, Prashen Chelikani, Pingzhao Hu

iScience, Volume 27, Issue 8, 110447

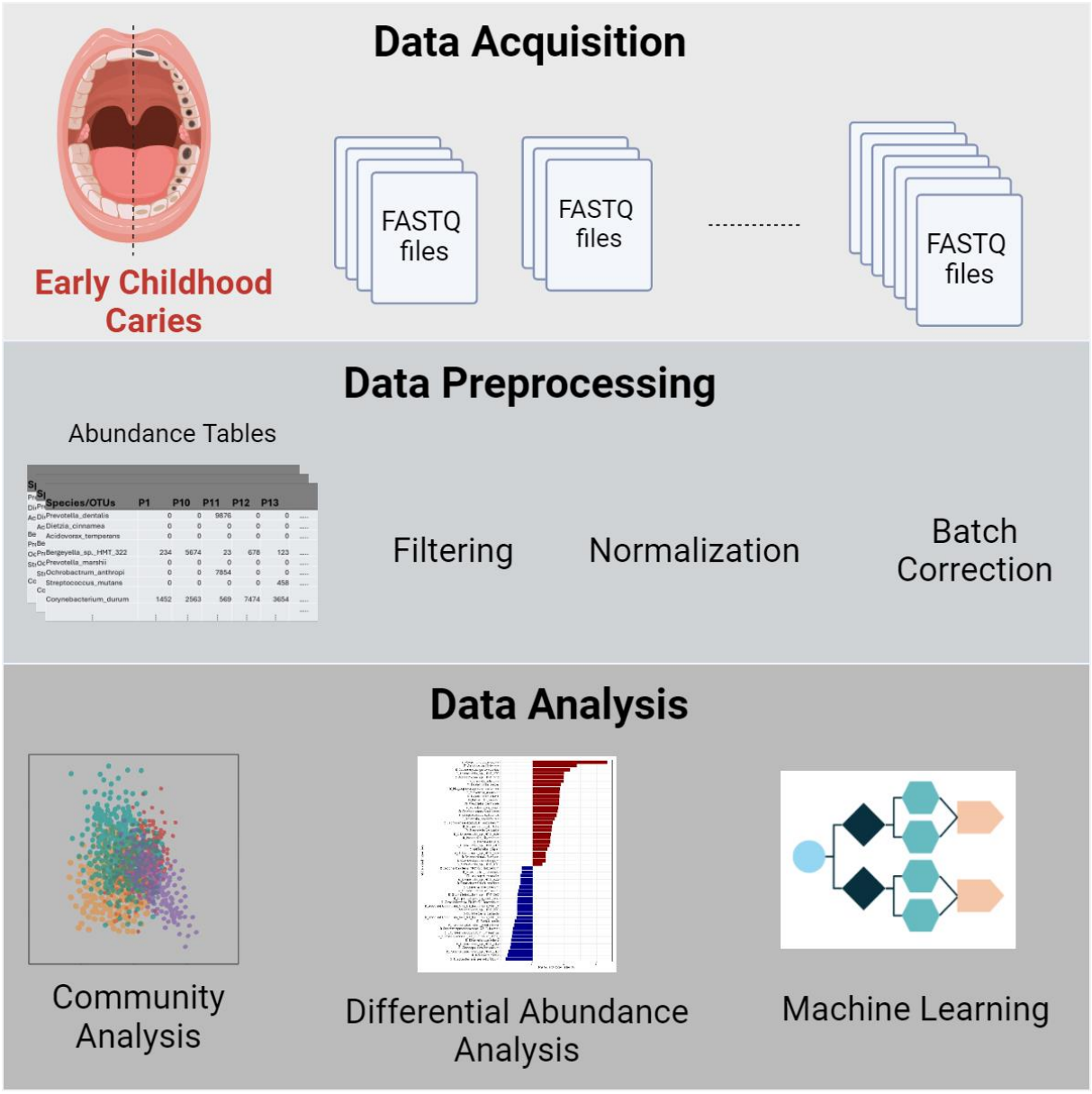
DOI: https://doi.org/10.1016/j.isci.2024.110447

In accordance with Elsevier's author rights policy, the author is permitted to include this article in their thesis without obtaining written permission from the publisher.

AUTHORS CONTRIBUTIONS

Conceptualization: P.H., M.W.K.; Methodology: M.W.K., P.H., D.F.L.X.; Investigation: M.W.K., P.H., D.F.L.X., P.C., R.J.S.; Writing original draft preparation: M.W.K.; Writing review and editing: M.W.K., P.H., P.C., R.J.S., D.F.L.X.; Visualization: M.W.K.; Supervision: P.H., P.C., R.J.S. All authors have read and agreed to the published version of the manuscript.

Graphical Abstract



4.1 Summary

Early childhood caries (ECC) is a multifactorial disease with microbiome playing a significant role in caries progression. Understanding changes at the microbiome level in ECC is required to develop diagnostic and preventive strategies. In our study, we combined data from small independent cohorts to compare microbiome composition using a unified pipeline and applied batch correction to avoid the pitfalls of batch effects. Our meta-analysis identified common biomarker species consistently reported across different studies, including *Megasphaera micronuciformis*, *Prevotella salivae*, and *Selenomonas sputigena*. We identified that the random forest was the best machine learning method for the classification of ECC versus caries-free samples and compared the performance of this method using a leave-one-dataset-out approach. Our random forest model was found to be generalizable when used in combination with other studies. While our results highlight the potential microbial species involved in ECC and disease classification, we also mentioned the limitations that can serve as a guide for future researchers to design and use appropriate tools for such analyses.

4.2 Introduction

The human body comprises roughly equal numbers of microorganisms as of host cells (Rizkallah *et al.*, 2012; Sender, Fuchs and Milo, 2016). The microorganisms and their genomes found in the oral cavity are collectively referred to as the oral microbiome (Dewhirst *et al.*, 2010; Berg *et al.*, 2020). Approximately 800 species have been identified in the human oral microbiome database (HOMD), including both cultivable and not-yet-cultivated species (Escapa *et al.*, 2018). Dysbiosis of the microbiome causes several oral diseases, including dental caries. Caries is the term used for tooth decay, and it is the most prevalent infectious disease in the oral cavity. Caries in children less than 72 months involving the primary dentition is known as early childhood caries (ECC) and affects about half of children worldwide (Casamassimo *et al.*, 2009; Teng *et al.*, 2015).

Dental caries is multifactorial in origin and one of main drivers for caries development is the biochemical transformations caused by acidogenic microbes on the dental surface (Marsh, 2006; Selwitz, Ismail and Pitts, 2007). There is a hypothesis of relatively lower bacterial diversity in carious lesions due to constant acidic exposure. This change can be observed in the dominance of aciduric microbes. *S. mutans* is found to be the main caries-causing bacteria, but previous studies have identified several other bacteria involved in caries, such as *Actinomyces* and *Lactobacillus* (Li *et al.*, 2007; Tanner, Kent, *et al.*, 2011; Teng *et al.*, 2015). Several acidogenic species have been identified in the microbiome of children with ECC, including *Scardovia wiggsiae*, *Lactobacillus salivarius*, *S. mutans*, and *Parascardovia denticolens* (Richards *et al.*, 2017).

Since the arrival of next-generation sequencing (NGS), amplicon sequencing using the 16S rRNA gene has been the predominant method for identifying and quantifying bacterial communities in complex biological samples (Lozupone *et al.*, 2013). NGS provides a platform to sequence the culturable or non-culturable strains using short sequence reads to study the oral microbial community. The most common sites to study the oral microbiome are saliva and supragingival plaque, although there are site-based differences in microbiomes especially in the context of ECC (de Jesus, Khan, *et al.*, 2021). Furthermore, there is no standard approach yet regarding which region of 16S rRNA has to be used, and this introduces

biases among studies and also makes it difficult to resolve it beyond the genus level (Baker *et al.*, 2021).

Microbiome biomarkers have been extensively studied in gut microbiomes for colorectal cancer, diabetes, obesity, and inflammatory bowel disease. Given the high dimensionality of microbiome profiles in any study, robust models are required to identify predictive features and ensure the reproducibility of the analysis. One of the goals of supervised classification methods in microbiome analysis is to identify the predictive features for a given condition and produce a predictive model. In addition to the high dimensionality and high sparsity of microbiome data, one of the key challenges is the limited number of common species among samples. Based on the popularity of machine learning methods, a recent review suggested four common methods for microbiome-based classification analysis: random forest, support vector machines (SVM), logistic regression, and k-nearest neighbor (KNN) (Marcos-Zambrano *et al.*, 2021).

Meta-analysis can help address discrepancies that arise due to technical and/or biological inconsistencies between studies (Bisanz *et al.*, 2019). It also enables the identification of the universality of the biomarker for specific disease diagnosis or prognosis. On the other hand, it is also important to understand the factors underestimating the effects of these analyses. Different procedures for sample handling, sequencing methods, and data preprocessing methods can lead to discrepancies in the datasets. Hence, it is important to address such issues and minimize these factors to the extent possible. Several studies have applied meta-analysis to various diseases, such as Parkinson's disease (Romano *et al.*, 2020), cancer (Thomas *et al.*, 2019; Limeta *et al.*, 2020), and urolithiasis (Kachroo *et al.*, 2021). To our knowledge, no previous study has combined microbiome data from different studies in order to undertake a meta-analysis of the microbiome of children with ECC. Such an analysis would enable us to understand the complexity of microbial profiles and site-specific complexity in ECC.

In this study, we analyzed five published studies on ECC, which made the raw data publicly available based on 16S rRNA amplicon sequencing. We processed the raw data using the same approach wherever possible. This combined analysis helps to consolidate the diverse

outcomes observed across individual studies. The meta-analysis pipeline we report in this study provides robust results for ECC classification and identification of taxonomic biomarkers. **Figure 4.1** illustrates an overview of the overall study design. Our results suggested that *Megasphaera micronuciformis*, *Prevotella salivae*, *S. mutans*, and *S. sputigena* are common microbial markers across studies, and that the random forest model performs better for ECC vs CF classification, helping identify the associated microbial features.

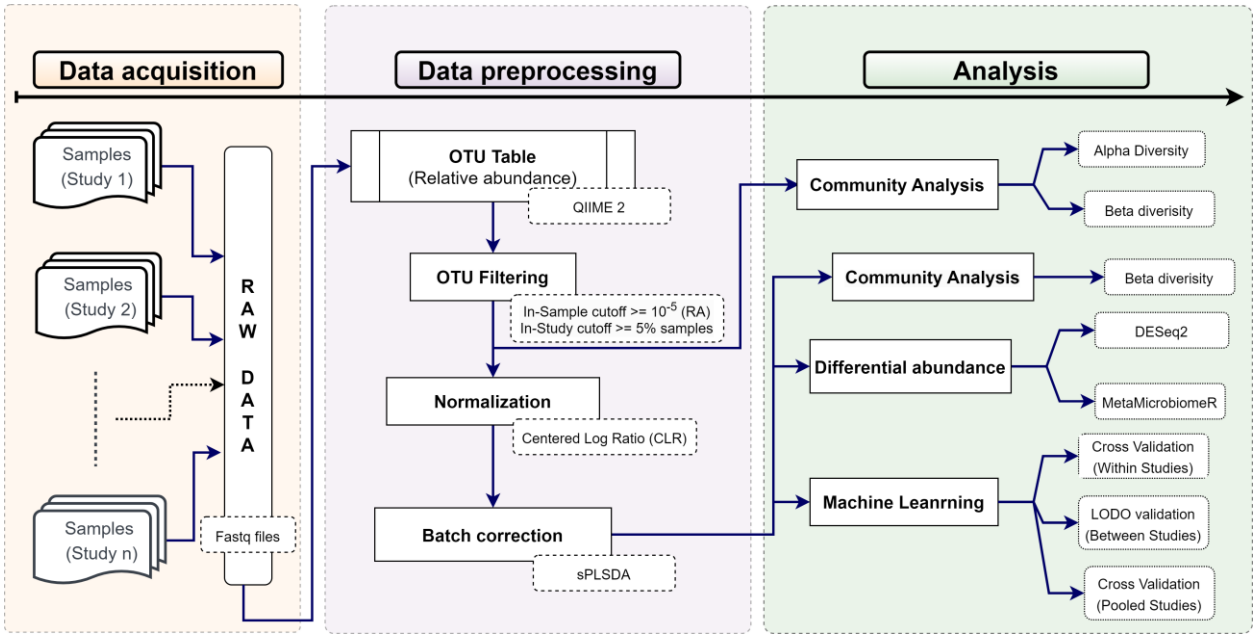


Figure 4.1: Schematic overview of data processing and analysis pipeline.

4.3 Results

4.3.1 Datasets

Based on our selection criteria, we identified five studies that were included in this analysis (Teng *et al.*, 2015; Agnello *et al.*, 2017; Gomez *et al.*, 2017; de Jesus *et al.*, 2020; Kalpana *et al.*, 2020). Data collection from these datasets resulted in a total of 223 samples:117 caries-free (CF) and 106 ECC. The data for all the datasets except Agnello_2017 were downloaded from sequence read archive (SRA) in form of FASTQ format and the data for Agnello_2017 was

obtained from the authors upon request (Agnello *et al.*, 2017). The description of the datasets about the sample size, case-control number, additional metadata given about the samples, and accession number for the raw data included in the analysis is given in **Table 4.1**.

In the operational taxonomic units (OTU) tables of five processed datasets, the genus-level taxa ranged from 59 to 101 with 50 common genera (**Figure 4.2A**). While at species level, the total number of OTUs ranged from 133 to 342 with an intersection of 96 species common across all the samples (**Figure 4.2B**).

Table 4.1: Information about the datasets used.

Study name	Samples		16S rRNA region	SRA accession	Metadata information available	Location of the sample collection
Agnello_2017	20 CF	30 ECC	V3-V4	Obtained from the authors	None	Winnipeg, Canada.
Gomez_2017	20 CF	12 ECC	V4	PRJNA383868	Age, Sex	Adelaide/ Melbourne/ Sydney, Australia
Kalpana_2020	10 CF	11 ECC	V3-V4	PRJNA454811	None	Tiruchengode, India
Teng_2015	27 CF	13 ECC	V1-V3	SRP040945 and SRP040947	Age, dmfs	Guangzhou, China
DeJesus_2020	40 CF	40 ECC	V4	PRJNA555320	Age, Sex	Winnipeg, Canada

CF: caries-free; ECC: early childhood caries;
dmfs: decayed, missing, filled surfaces score

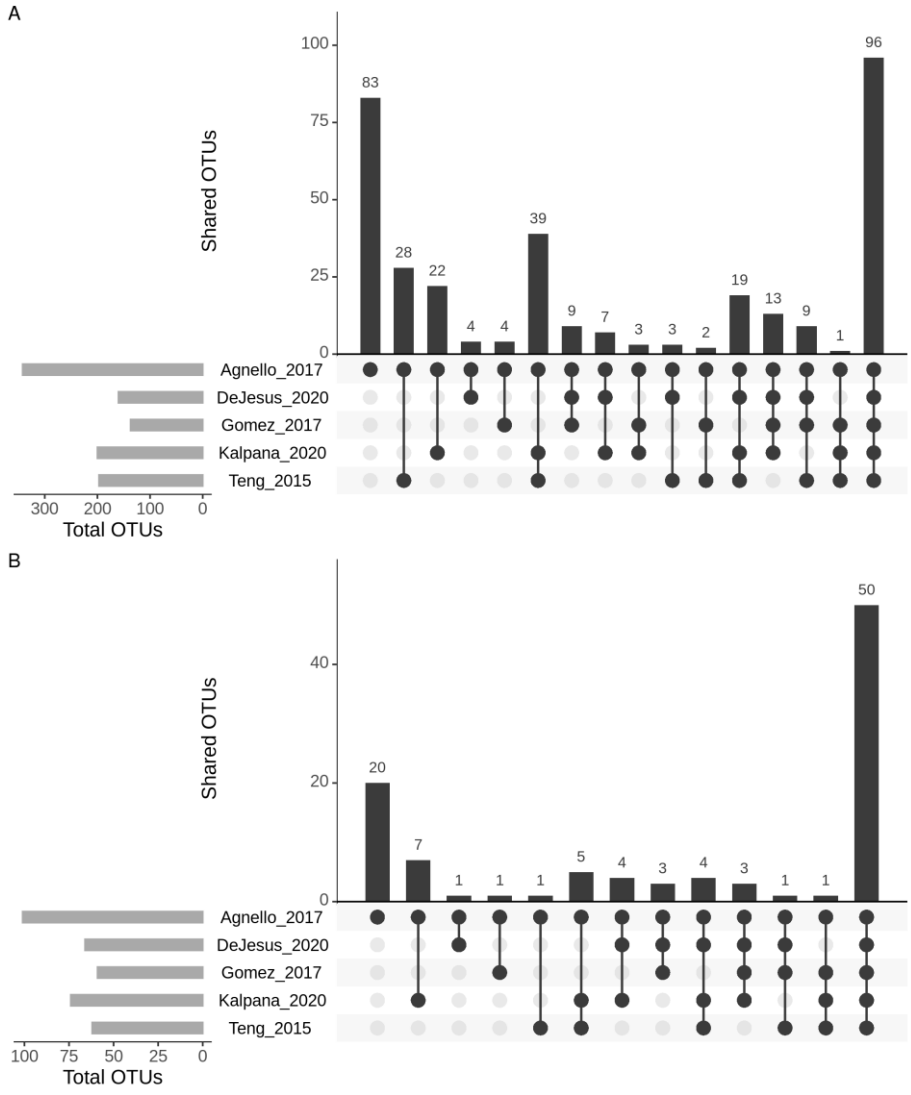


Figure 4.2: UpSet plot for the number of shared OTUs across different datasets.
(A) Species level (B) Genus level. The smaller subplot represents the total number of OTUs at the respective taxonomic level. The main bar plot shows the number of unique and shared OTUs between the studies. The connecting dots below indicate the intersection of the studies indicated by each bar.

4.3.2 Diversity analysis

The most common genera in both CF and ECC samples were *Streptococcus*, *Neisseria*, and *Veillonella* (**Figure 4.3**). Among all species, *Veillonella* species were abundant. The meta-analysis for Shannon diversity at the species level revealed that the diversity for disease status did not significantly change between CF and ECC samples, and the heterogeneity for diversity was found to be 48 percent (**Figure 4.4**). However, a significant change among the CF samples between the studies can be observed. While this is not the case among ECC

samples, a significant diversity change is only observed between Teng_2015 and Kalpana_2020 datasets. For the species-level beta diversity analysis, the analysis of similarities (ANOSIM) statistic R-value was 0.64 and 0.49 for CF and ECC groups, respectively (**Figure 4.5**). A reduction in R values was observed after batch correction (CF = 0.075 and ECC = 0.091), which is suggestive of a decrease in dissimilarity or increased homogeneity in beta diversity (**Figure 4.5** and **Figure 4.6**).

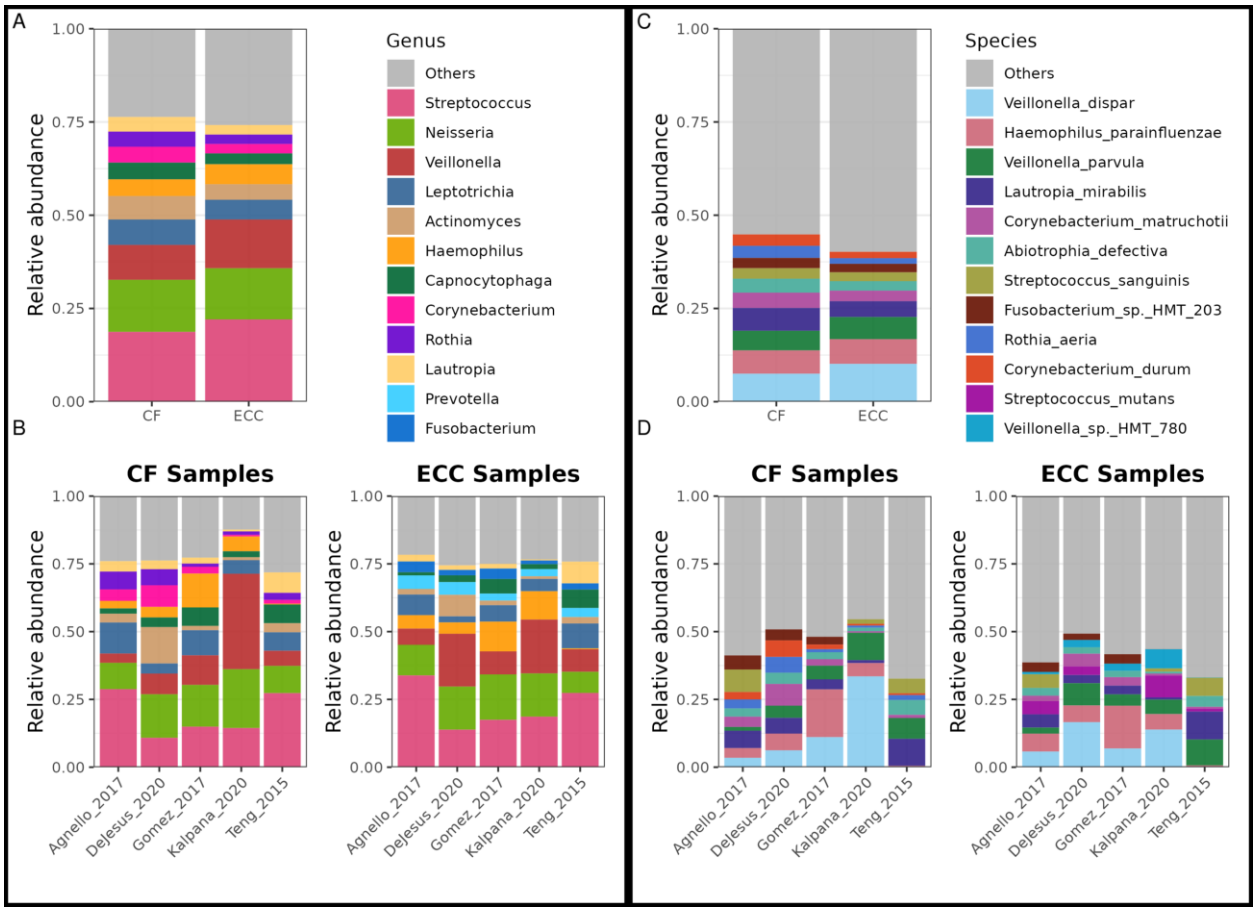


Figure 4.3: Relative abundance of top taxa between caries and ECC samples.
(A) Combined datasets at the genus level. (B) Individual datasets at the genus level (C) Combined datasets at the species level. (D) Individual datasets at the species level.

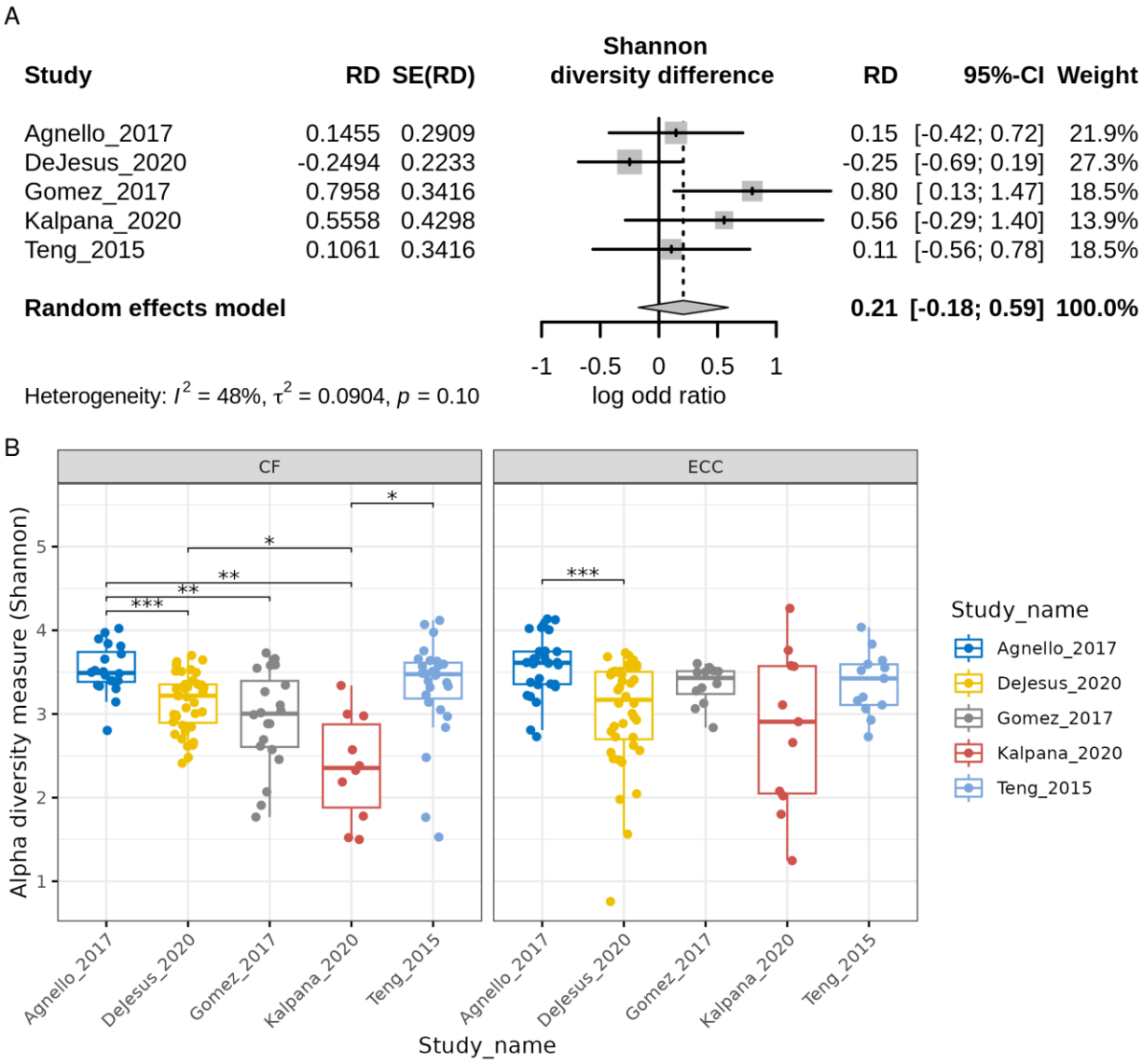


Figure 4.4: Comparison of species-level alpha diversity by Shannon index.
(A) The forest plot compares the log odd ratios (with a 95% confidence interval) of the Shannon diversity difference between CF and ECC samples for each dataset. The TE and seTE columns provide the total estimate and standard error of the total estimate values, respectively. (B) Comparison of alpha diversity for CF and ECC samples between datasets using a t-test with significance levels of * = p<0.05, **= p<0.01, and *** = p<0.005.

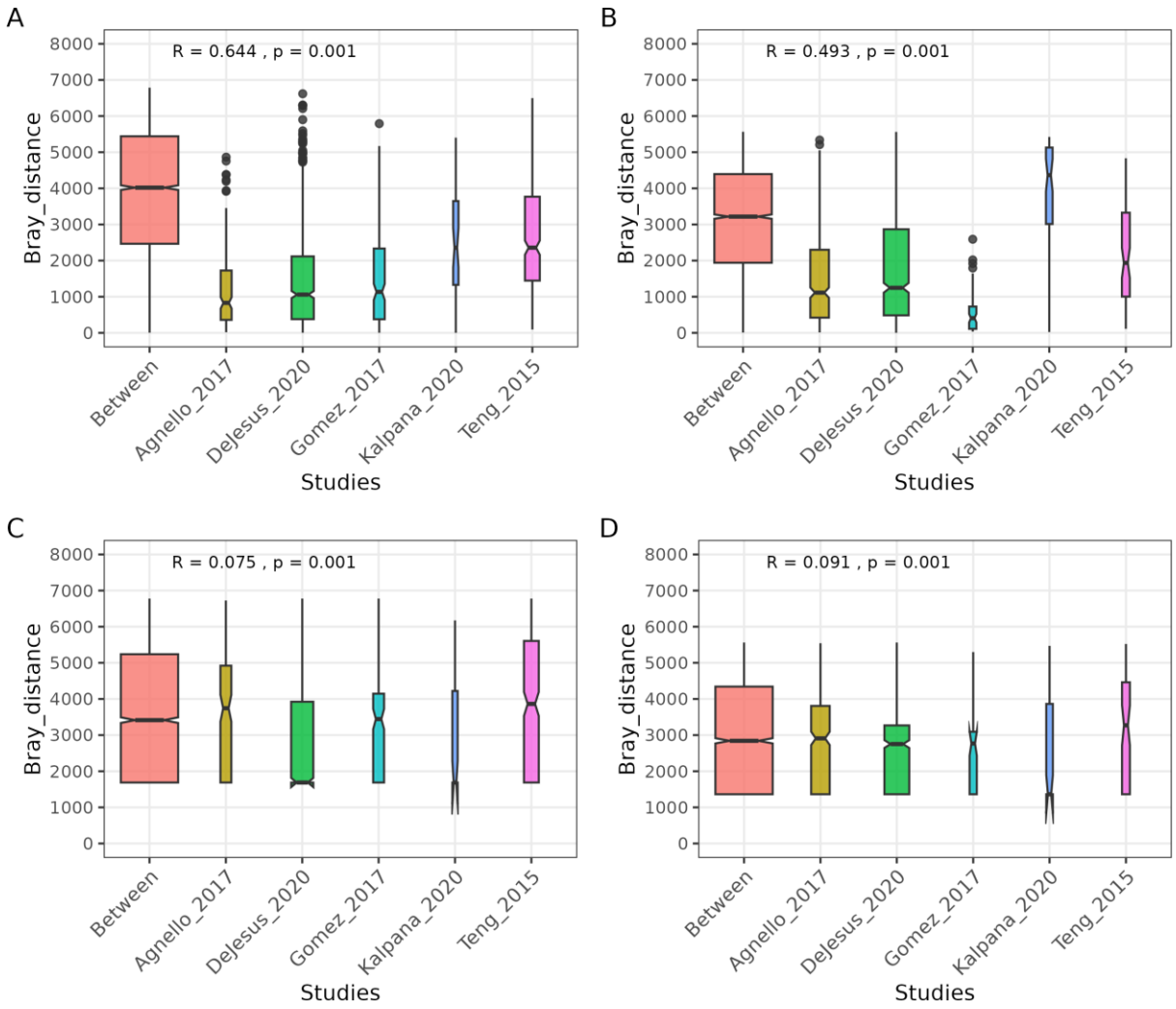


Figure 4.5: Analysis of similarities (ANOSIM) test using Bray diversity metrics on species-level OTUs to illustrate the extent of similarities between five studies.

(A) Caries-free group before batch correction. (B) ECC group before batch correction, (C) CF group after batch correction, (D) ECC group after batch correction. An R value close to "1.0" suggests dissimilarity between groups while an R value close to "0" suggests an even distribution of high and low ranks within and between groups". The *p*-values represent the significance of the R value. The first column in each group of boxplots represents the between-group dissimilarity, and the other five columns represent the beta diversity within each study. The figure shows that there was a significant drop in between-study dissimilarity after the sPLSDA batch correction method. The size of the box is proportional to the sample size in each dataset.

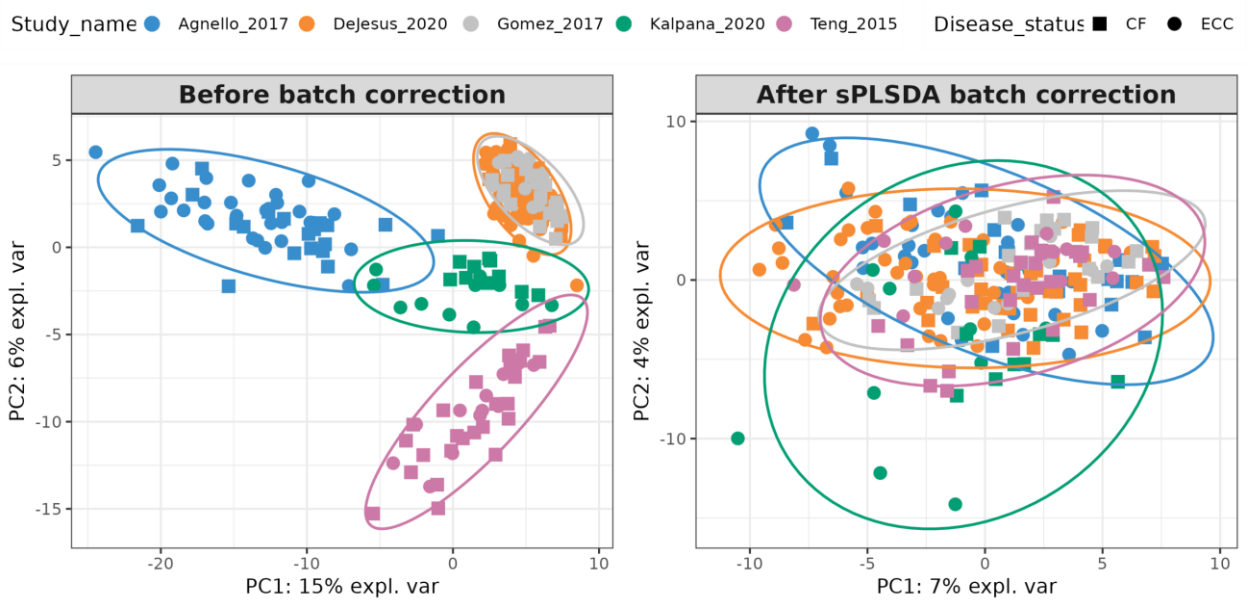


Figure 4.6: Principal component analysis (PCA) plots for species-level datasets. Using relative abundance data before any batch correction (left) and after sPLSDA batch correction on centered log ratio (CLR) transformed values (right). The PC1 and PC2 axes also include the explained variance of these PC components. The batch correction can be noticed by the reduction in explained variance on the PC1 and PC2 axes.

4.3.3 Differential abundance analysis

For differential abundance analysis, Datasets DeJesus_2020, Agnello_2017, and Kalpana_2020 resulted in a higher number of differentially abundant taxa than the remaining two datasets at both species and genus levels. In our analysis, we found that DESeq2 identified more species as differentially abundant than the LefSe method. The species identified by DESeq2 methods in all the datasets are illustrated in supplementary **Figure S4.1**. The meta-analysis of differentially abundant species revealed a higher abundance of *Prevotella salivae*, *Selenomonas sputigena, and Prevotella oulorum* in the ECC group, while at the genus level, *Alloprevotella* and *Megasphaera* were found to be associated with ECC (**Figure 4.7**). A comparison of differential abundance taxa for relative abundance and batch-corrected data suggests that the significantly altered OTUs are more consistent with batch-corrected values, and the number of such OTUs is higher for adjusted *p*-values. The batch-corrected data provided nine significant species in comparison with two species with relative abundance data (**Figure 4.7** and supplementary **Figure S4.2**).

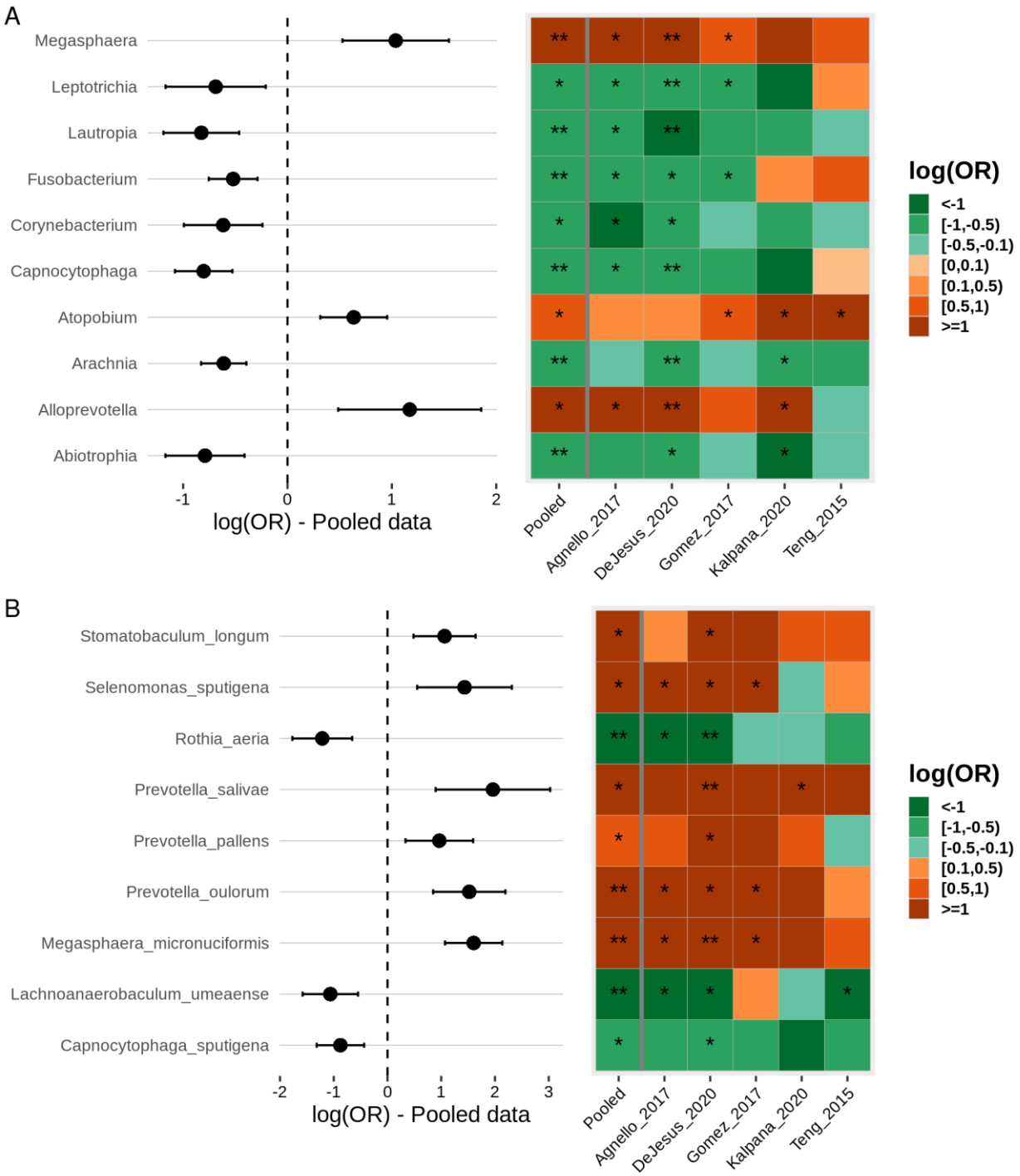


Figure 4.7: Meta-analysis for differentially abundant taxa.
(**A**) Genus level (**B**) Species level The heatmap represents the log odd ratios of differentially abundant with significant adjusted *p*-value (*p*-adjusted < 0.05) in pooled dataset along with the odd ratio estimates in each dataset. The forest plot signifies a 95% confidence interval for the log odd ratio values for each taxon in the pooled dataset.

4.3.4 Classification of CF and ECC using machine learning

For the classification of CF and ECC samples using machine learning methods, we tested some commonly used machine learning methods in metagenomic studies, which are logistic regression with Lasso penalty, random forest, XGBoost, SVM, and decision trees. The performance for cross-validation and test set validation was the best with random forest and XGBoost, while lasso results were very close to these two methods (**Figure 4.8B**). The average classification performance in terms of area under the receiver operating characteristic curve (AUROC) value is 0.85 at the species level and 0.83 at the genus level using random forest. When the OTUs for both genus and species levels were combined, the AUROC value was 0.84 (supplementary **Figure S4.3**). The performance of random forest was best with 20-40 OTUs selected from random forest models based on feature weights, and it did not change significantly with a higher number of OTUs (**Figure 4.9).** Here, we used the term within-study-cross-validation (CV) for the performance within the same study, and leave-one-dataset-out (LODO) analysis when applying a model trained on one study and tested on another study. The random forest method outperformed the other machine learning methods used in this study for LODO analysis as well. We also explored the features importance obtained from the models for individual studies and pooled datasets (**Figure 4.10**). Comparatively, genus-level OTUs were more common among studies than species-level OTUs. At the species level, Agnello_2017 and DeJesus_2020 shared many OTUs used in model development.

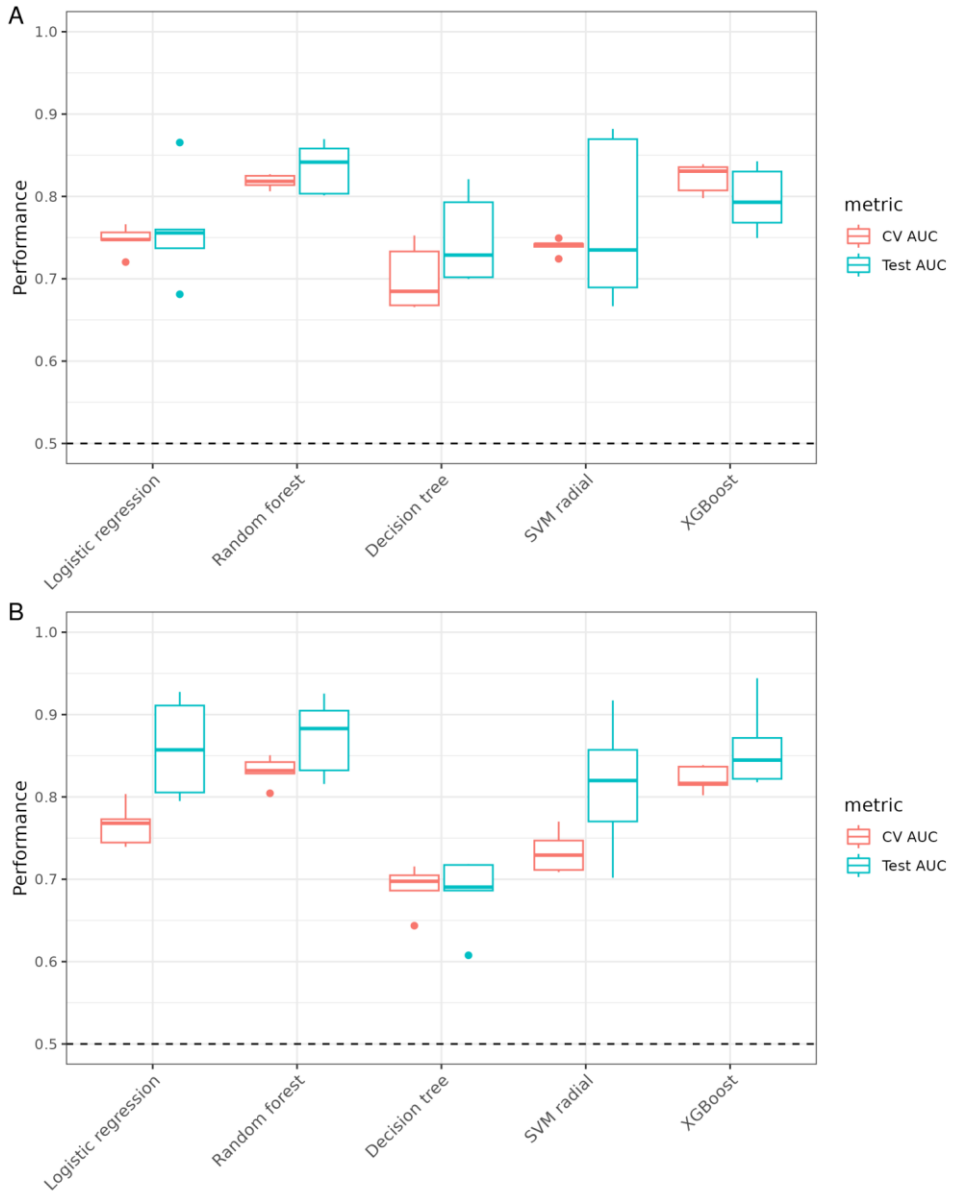


Figure 4.8: Model performance by AUROC for 5 different machine learning models on pooled dataset from 5 studies.

(A) Genus level (B) Species level. The x-axis represents the models used for the comparison of the performance value in terms of AUROC. The methods are glmnet-logistic regression with lasso penalty, rf- random forest, rpart2-decision tree, svm Radial- support vector machine with radial kernel, and xgbTree- extreme gradient boosting. CV-AUROC and test-AUROC boxes denote the AUROC values from cross-validation and test data, respectively, from pooled datasets. These results indicate that the average performance of the random forest method outperforms the other methods compared here.

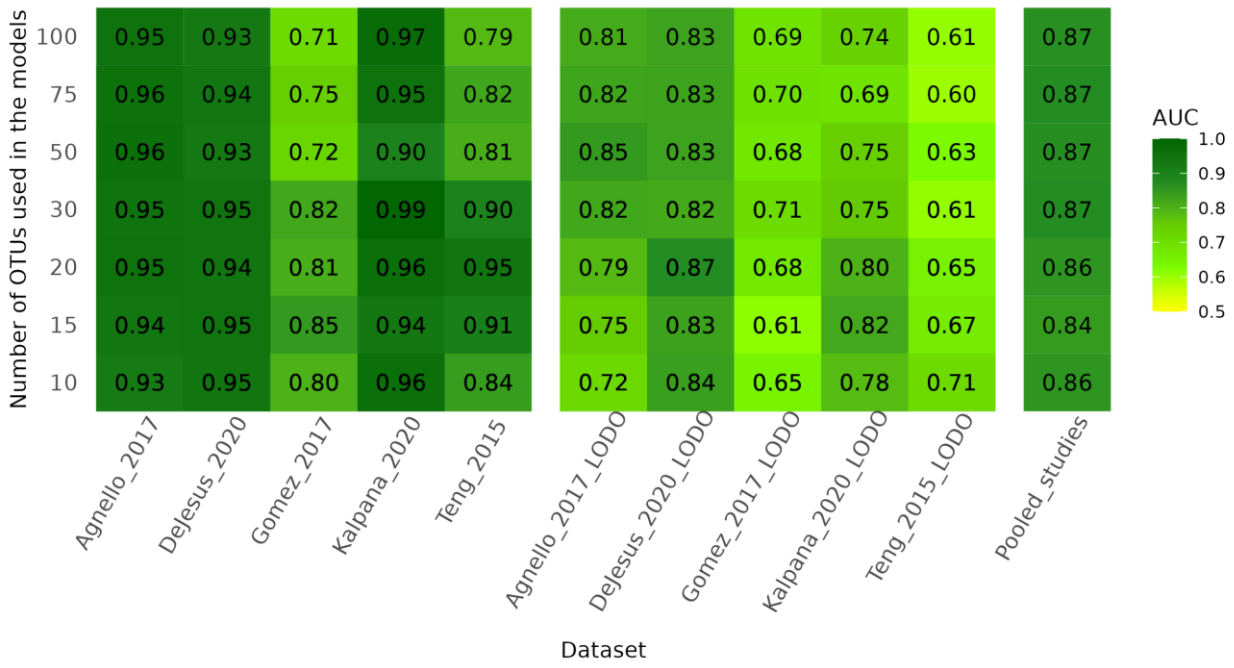


Figure 4.9: Heatmap for the performance of random forest classifier by AUROC on species-level data.

The left panel shows the cross-validation results for each dataset. The middle panel shows the model performance for LODO analysis. In the LODO analysis, all datasets except one were used for training, and the left-out dataset was then used for testing to assess the generalizability of the model. The rightmost column shows the cross-validation performance of the pooled dataset. The y-axis represents the number of top OTUs used for model assessment.

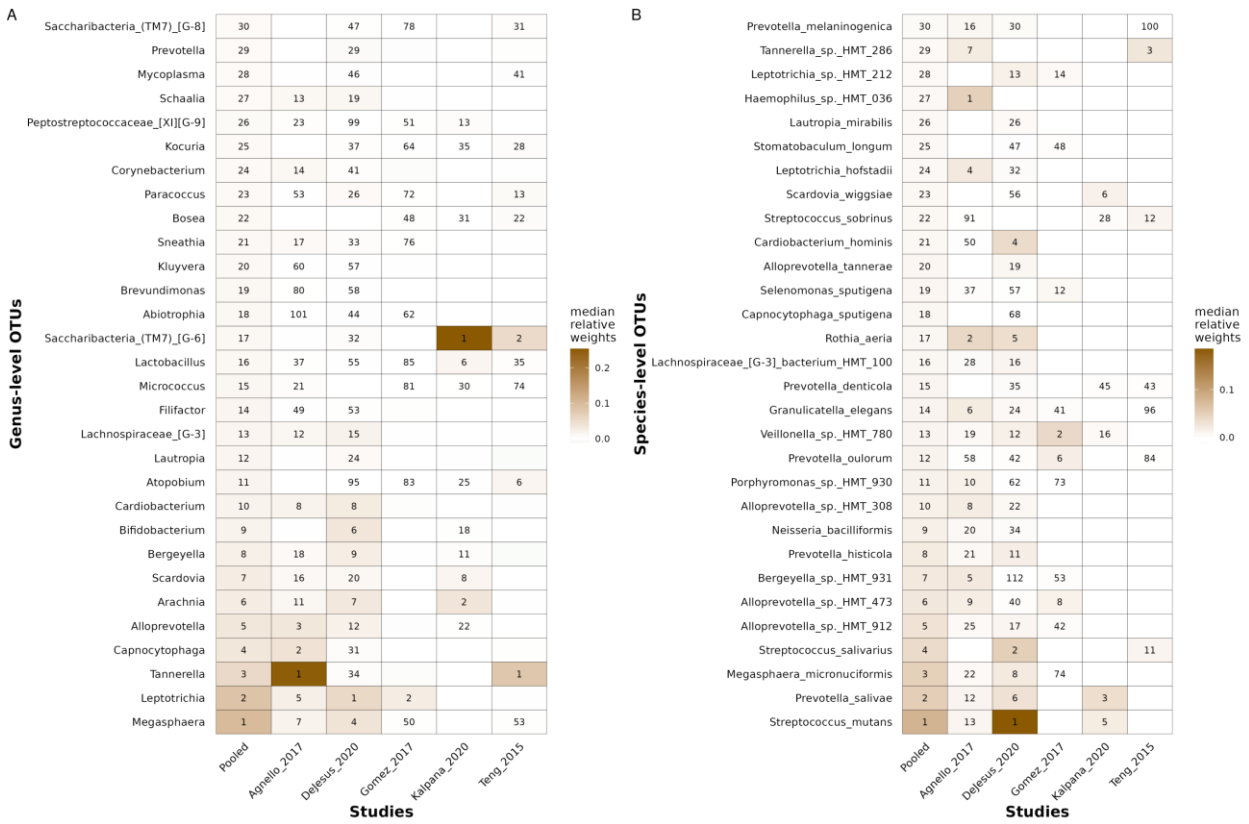


Figure 4.10: Heatmap for feature relative weights along with their rank obtained from the random forest model.
(A) Genus-level OTUs and (B) Species-level OTUs. The top 30 features in the pooled dataset were used to compare the feature weights from the individual datasets. The random forest model was used for each dataset with a 5-fold cross validation and five repeats. The weights were obtained by the median of the feature weights from each model.

4.4 Discussion

Small and heterogeneous cohorts provide distinct opportunities to study disease-specific microbiomes (Thomas *et al.*, 2019). We extended this idea to study the ECC-associated microbiome by combining them to achieve a better representation of ECC cases and controls. In our analysis, we combined previously published dental plaque microbiome datasets for ECC children and analyzed them using a common data processing pipeline to understand the patterns across cohorts and test the generalizability of machine learning classifiers across datasets.

A general practice in microbiome data analysis is to transform read counts to relative abundance or to rarify the data to a certain number of reads. However, significant variation

can be observed among the studies in terms of OTUs, even after such transformations. These differences can arise from discrepancies in sample handling, DNA extraction methods, and sequencing techniques. To minimize these batch effects, we first processed the raw reads using a common pipeline whenever possible. Second, we used the sparse Partial Least Squares Discriminant Analysis (sPLSDA) batch correction method, which has shown better efficiency in reducing batch effects in microbiome data than other batch correction methods (Wang and Cao, 2020). sPLSDA methods captures more variance due to the treatment than the commonly used methods such as ComBat (Wang and Lê Cao, 2023). For compositional data, centered log ratio (CLR) method addresses the variability in library sizes and diminishes the skewness present in the data (Susin *et al.*, 2020). Moreover, CLR transformation approximates a Gaussian distribution on the count data which provides a better performance using sPLSDA method (Wang and Lê Cao, 2023).

We identified very few significant species in the differential abundance analysis in some of the datasets. One reason could be the power of the statistical test – studies with more samples resulted in a higher number of significant taxa, for example, the DeJesus_2020 dataset (**Table 4.1**). Another reason for that could be the limited number of total reads in certain samples. For optimal taxonomic coverage, it is recommended to have approximately 5000 reads per sample. Although most of the samples in our study had more than the specified number of reads, sometimes samples with a smaller number of reads, as low as 3000, were also included due to the small sample size in some of the datasets.

Commonly used machine learning methods in microbiome studies include logistic regression models, SVM, random forest, and gradient boosting methods, such as XGBoost. In our analysis, the random forest method outperformed the other tested methods in both cross-validation and the train-test strategy when applied to the pooled dataset (**Figure 4.8)**. The observation of lower CV-AUROC values in these results can be attributed to the diversity within each fold which will lead to the poorer average performance of CV-AUROC. However, the final model may have leveraged from the combined coefficients of all folds which captures more nuances and generalizes better on the test dataset. Although comparable with relative abundance, the batch-corrected data give more consistent results with random

forest, and the performance linearly improves with the number of features. However, it did not change significantly after 20-30 features in terms of OTUs (**Figure 4.9**). Previous studies have also shown that random forest provides better classification than other commonly used machine learning methods in meta-analysis of microbiome data (Thomas *et al.*, 2019; Zhou and Gallins, 2019).

Our meta-analysis and machine learning analysis revealed some important taxa identified in this cross-cohort analysis. *Streptococcus* species (for example, *S. mutans*, *S. salivarius*, *S. oralis* and S. sp. HMT-074) were dominant in the top species from the machine learning models. Interestingly, *Streptococcus* species were not identified in the meta-analysis and genus-level machine learning analysis. The role of *Streptococcus* species has been highlighted in previous studies; for example, *S. mutans* and *S. oralis* were discussed as high-risk factors in ECC (Teng *et al.*, 2015; Meriç *et al.*, 2020). *Streptococcus* species ferment sugars to produce lactic acid, which reduces the pH and causes demineralization of enamel (Zafar *et al.*, 2020). A novel host-microbe interaction between the cariogenic *S. mutans* and taste receptors on the oral gingiva leading to innate immune responses was reported (Medapati, Singh, *et al.*, 2021). It was suggested that *S. mutans* might be modulating the immune system to inhibit other gram-positive bacteria and involved in mediating autophagy flux in gingival epithelial cells (Medapati, Bhagirath, *et al.*, 2021; Singh *et al.*, 2024). The next most significant genera were *Prevotella* and *Alloprevotella*. The species from these genera were significant in both meta-analysis and machine-learning analysis at the species level. *Prevotella* and *Alloprevotella* are genera enriched in increased dental caries (Downes *et al.*, 2013; Uchida-Fukuhara *et al.*, 2020). *Alloprevotella* species were found to be mild saccharolytic and acid-producing as a result of fermentation (Downes *et al.*, 2013). Another genus, *Tannerella*, was also found to be important from machine learning-based feature analysis at both the species and genus levels.

To illustrate that the classifier performance with batch correction data is largely due to the batch correction method and not the underlying CLR normalization, we also compared the AUROC values for CLR-transformed data and batch-corrected values (**Figure 4.9** and supplementary **Figure S4.4**). From these results, it is evident that although CLR improves

cross-validation within studies, the LODO performance of the CLR method alone is poorer than the machine learning performance on sPLSDA batch-corrected data.

Our results demonstrate that LODO analysis provides generalizability for the classification of CF and ECC using dental plaque microbiome data. Results for LODO analysis are very promising, reaching as high as 0.88 in terms of AUROC values for the classification of ECC and CF status. Another conclusion is that species-level analysis provides better results than genus-level analysis (**Figure 4.9**and supplementary **Figure S4.5**). We also compared the features identified in each study for within-study-CV at the genus level for plaque samples and observed that there were very few common features among studies, and no single feature was common across all five datasets (**Figure 4.10**).

These results also suggest that the classification performance and differentially abundant species could be more robust with larger datasets, as observed in the DeJesus_2020 and Agnello_2017 datasets. Additionally, microbiome-based diagnostic biomarkers can help in the early prediction of severe ECC, which can help clinicians and parents to adopt timely preventive measures. It can help improve the quality of life at both the individual and societal levels. Future prospective studies are needed to establish a causal link between these biomarkers and ECC.

Machine learning classifiers work better when the dataset is the union of all studies for OTUs instead of only common OTUs (data not shown). This is potentially due to the loss of a substantial number of features when considering the intersection of the datasets for OTUs, which reduced to 96 from 391. However, union of datasets induced 0 inflation for the OTUs that were not present in the original OTU table. To circumvent this problem, several machine learning methods have relied on imputation techniques. In microbiome data analysis, the only method available specifically for microbiome data is mbImpute. We included the results of random forest performance imputing the missing values using the mbImpute method (supplementary **Figure S4.6**). However, we did not observe any improvement in the classifier's performance. Further exploration using imputation methods is required in microbiome studies.

In summary, we conducted a meta-analysis of dental plaque ECC-associated microbiomes from five previously published studies. We assessed the importance of normalization and batch correction methods when conducting such analyses. Our meta-analysis included the differential abundance and machine learning-based identification of features in combined studies. Based on high value AUROC value in LODO analysis, we conclude that the ECC microbiome shows a good extent of common features across different studies worldwide, which can be studied with appropriate data merging techniques and modern machine learning tools.

Limitations of the study

Some possible limitations of our analysis include the relatively small sample size even with meta-analysis, which is a common constraint in omics research, as well as the considerable variation in the number of reads between samples. In addition, each study used a particular variable region or combination of variable regions of 16S rRNA amplicon sequencing to identify the bacterial community (Abellan-Schneyder *et al.*, 2021). The differences in the datasets might be due to the different regions selected during sequencing, as shown in **Table** 4.1 (Abellan-Schneyder *et al.*, 2021). Another potential variance in the caries microbiome may stem from different areas of the teeth. These differences have been observed between occlusal and proximal caries, as well as between enamel and dentin caries in ECC (Richards *et al.*, 2017; Fakhruddin *et al.*, 2022). However, the samples used in our analysis were not collected from specific tooth surfaces, which may limit the applicability of our findings to the differences arising from distinct sites.

4.5 Methods

4.5.1 Experimental model and study participant details

This meta-analysis is based on previously published studies. Participant details can be found in the articles related to their respective studies (**Table 4.1**).Inclusion and exclusion criteria used in this study are provided in the sections below and in supplementary **Table S4.1**.

4.5.2 Inclusion and exclusion of studies

Studies were selected from publications that conducted microbiome comparisons between caries-free (CF) individuals and those with ECC using a case-control design. Among these studies, only those employing 16S rRNA amplicon sequencing and providing the raw sequencing data were selected. The raw data in such studies can be obtained from one of two techniques:16S rRNA amplicon sequencing and shotgun metagenomic sequencing. Generally, amplicon sequencing using 16S rRNA sequences can be obtained using one or a combination of more than one hypervariable region of the 16S rRNA gene or with full-length 16S rRNA sequencing. For our analysis, studies with 16S rRNA amplicon sequencing data, irrespective of the hypervariable region, were selected. Studies that utilized shotgun metagenomic sequencing methods were excluded from our analysis to minimize the known technical effects. The two most common oral sites for amplicon sequencing are the supragingival plaque (also referred to as dental plaque) and saliva. To avoid site-based differences, we focused on studies that collected their samples from dental plaque.

4.5.3 Collection of data and raw data processing

The data were downloaded from NCBI repositories using Fasterq-dump (NCBI SRA-toolkit). Some of the raw sequences were obtained directly from the authors upon request **Table 4.1**.

Raw FASTQ sequences were analyzed using the QIIME2 pipeline to obtain the operational taxonomic units (OTUs) table for each study. For the datasets in each study, several combinations of left and right trims were optimized using the DADA2 plugin in QIIME2 to obtain the maximum number of non-chimeric reads. The performance of DADA2 trimming was manually assessed for forward and reverse read trimming. The combination that retained the maximum number of non-chimeric reads in the samples, particularly samples with a low number of total reads, was used for final trimming. To classify the amplicon sequencing variants (ASVs) obtained from DADA2 into OTUs, HOMD 15.22 was used as a reference database. The sequences from the reference database were extracted using the primer sequences cited in each study's original publication (Supplementary **Table S4.1).** The final OTU tables were obtained at both the genus and species levels. All the tools used to process the datasets can be found in the **Table S4.2**.

4.5.4 Preprocessing

Samples with fewer than 3000 amplicon sequencing variants (ASVs) after the DADA2 step were excluded from all downstream analyses. The raw read counts in all samples were scaled with total sum scaling relative abundance. Further filtering was applied for the OTUs present in less than 5 percent of the samples or OTUs with a maximum value of 10^{-5} relative abundance within a study. Since the samples in our analyses came from different studies, we performed batch correction between studies for our analysis. For batch correction, the relative abundance values were first normalized using centered log-ratio (CLR) normalization and batch correction was performed using the sPLSDA method from the mixOmics package in R (Wang and Cao, 2020). To test the role of imputation on the classification performance, missing taxa values in each dataset were estimated using the mbImpute method (Jiang, Li and Li, 2021). For the imputation step, the phylogenetic distance between OTUs was calculated from 16S sequences using with 'Jukes-Cantor' distance in DECIPHER package (Wright, 2016).

4.5.5 Diversity Analysis

We compared the alpha and beta diversities of species-level OTUs using relative abundance. To explore the change in diversity for caries status, alpha diversity with the Shannon index was estimated using the R-package Phyloseq. A meta-analysis was conducted on Shannon diversity changes for individual and pooled studies, and alpha diversity heterogeneity was assessed for these datasets. Shannon diversity differences within samples of CF and ECC children across the datasets were also examined. The significance of the difference between the groups was calculated by t-test, and *p*-values were adjusted by the "holm" method in the R-package rstatix, with adjusted $p<0.05$ considered significant. For beta diversity comparison, owing to the different group sizes, we used analysis of similarities (ANOSIM) with the Bray distance method and compared the beta diversity before and after batch correction.

4.5.6 Differential abundance Analysis

For differential abundance analysis, two very commonly used methods, LefSe and DESeq2, were applied with multiple hypotheses testing corrections using the false discovery rate (FDR) with adjusted-$p<0.05$, using the R package microbiomeMarker (Cao *et al.*, 2022). UpSet plots were used to illustrate the shared and unique differentially abundant species between

studies. Furthermore, for the meta-analysis of differentially abundant taxa, we used metamicrobiomeR in R (Ho *et al.*, 2019). For the meta-analysis of differentially abundant species, we used two approaches: GAMLSS-BEZI (Generalized Additive Models for Location Scale and Shape with beta zero-inflated family) on relative abundance and LM (linear mixed effect models) approach on batch-corrected data from the R package metamicrobiomeR.

4.5.7 Machine learning modeling

For the classification of CF and ECC samples, logistic regression with the lasso penalty, random forest, decision trees, extreme gradient boosting (XGBoost), and SVM methods were used. For machine learning, 5-fold cross-validation (CV) with five repeats was used. The leave-one-dataset-out (LODO) strategy was used for machine learning-based cross-cohort analysis. In the LODO analysis, all datasets, except one, were used for training with CV, and the left-out dataset was then used for testing to assess the generalizability of the model. The model performance was assessed by area under the receiver operating characteristic curve (AUROC). We extracted the OTU importance for each model in the form of feature weights obtained from the model. The model performance was also compared to the machine learning classifier performance for relative abundance data, batch-corrected data without imputation, and with imputation.

4.5.8 Machine learning hyperparameter tuning

The machine learning models were used from the mikroml pipeline, which used the caret package from R (Topçuoğlu *et al.*, 2021). For the training of random forest for LODO analysis and holdout dataset testing, the SIAMCAT package was used (Wirbel *et al.*, 2021). The logistic regression model was used with a lambda value of 10^{-4} to 10 and alpha=1 for the lasso penalty. The number of trees in random forest was tried between 100-1000 with mtry = c(round(sqrt.mdim / 2), round(sqrt.mdim), round(sqrt.mdim * 2)), where mtry and sqrt.mdim are the square root of number of random variables in each tree and the total number of features, respectively. In XGBoost, the learning rate (eta), the maximum depth of a tree (max_depth), and the fraction of samples used to train each boosting round (subsample) were evaluated for which a specific range of values. For the SVM models, the cost parameter (C) and the Gaussian kernel's standard deviation (sigma) were tuned.

4.6 Supplementary data

Supplementary Table and Figures

Table S4.1: Additional details of datasets and Qiime2 processing options.
Differences in QIIME2 processing options used for dereplication of raw sequences and primer sequences to extract reference sequences for classifier generation. Primer sequences were taken from the respective papers. Dereplication options were optimized to maximize the number of non-chimeric reads

Study	Dereplication option and Primer sequences used in Qiime2	Additional information about the datasets
Agnello_2017	vsearch cluster-features-de-novo • perc-identity 0.99 f-primer CCTACGGGNGGCWGCAG r-primer GACTACHVGGGTATCTAATCC	The authors provided the FASTA files and not the FASTQ files. Since the dada2 option cannot be used with FASTA file, we used the vsearch option in Qiime2 with 99 percent similarity for the binning of similar reads.
DeJesus_2020	dada2 denoise-paired • p-trunc-len-f 200 • p-trunc-len-r 150 • p-trim-left-f 15 • p-trim-left-r 15 f-primer GTGCCAGCMGCCGCGGTAA r-primer GGACTACHVGGGTWTCTAAT	
Gomez_2017	dada2 denoise-paired • p-trunc-len-f 200 • p-trunc-len-r 150 • p-trim-left-f 15 • p-trim-left-r 15 f-primer GTGCCAGCMGCCGCGGTAA r-primer GGACTACHVGGGTWTCTAAT	This dataset was composed of participants of 5 to 11 years old. Since the age restriction to be considered as ECC is less than 6 years old children only, participants with age less than 6 years were selected from this database which were 20 Caries free and 12 with caries.

Kalpana_2020	dada2 denoise-paired • p-trunc-len-f 250 • p-trunc-len-r 230 • p-trim-left-f 5 • p-trim-left-r 5 f-primer CCTACGGGNBGCASCAG r-primer GACTACNVGGGTATCTAATCC	The NCBI repository had raw FASTQ reads for 10 caries-free and 11 ECC participants only. These numbers were also confirmed with the authors, and they mentioned that only good-quality samples were deposited in the NCBI SRA repository.
Teng_2015	dada2 denoise-pyro • trim-left 20 • trunc-len 490 • max-ee 5 f-primer GAGTTTGATCCTGGCTCAG r-primer TACCGCGGCTGCTGGCAC	For this longitudinal study, the authors divided the children into three groups: H2H- Children who remained healthy for the entire duration, H2C- Children who were caries-free at the time of recruitment but developed caries later, C2C- Children who had caries at the time of recruitment and continued to have caries for the duration of this study. To mimic the samples for case-control type data, we selected all samples from H2H at different data points with confident health (27 samples) and from caries only the samples with dmfs value greater than or equal to 5 were selected which can be considered as severe ECC (SECC). A further sample filtering was done for the samples that retained less than 3000 non-chimeric reads after the dada2 step in Qiime2 processing.

Table S4.2: Key resources table with additional description.

RESOURCE	IDENTIFIER	DESCRIPTION
fasterq-dump v 2.9.6	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software	Converts SRA data format to FASTQ format.
QIIME2 v 2021.2	https://docs.qiime2.org/2021.2/	Microbial analysis from raw FASTQ sequences to OTU/ASV table.
HOMD v 15.22	https://v2.homd.org/	Human Oral Microbiome Database for curated information on bacteria in the human mouth.
mixOmics v 6.20.0	https://bioconductor.org/packages/mixOmics/	Multivariate analysis of omics biological studies.
mbImpute v 0.1.0	https://github.com/ruochenj/mbImpute	Missing data imputation for microbiome analysis.
DECIPHER v 2.24.0	https://bioconductor.org/packages/DECIPHER/	A toolset for deciphering and managing biological sequences.
Phyloseq v 1.40.0	https://joey711-github-io.uml.idm.oclc.org/phyloseq/	A tool to import, store, analyze, and graphically display microbiome data.
metamicrobiomeR v 1.2	https://github.com/nhanhocu/metamicrobiomeR	R package for analyses and meta-analyses of other microbiome studies.
Vegan v 2.6-4	https://github.com/vegandevs/vegan	This package provides tools for descriptive community ecology.
microbiomeMarker v 1.4.0	https://bioconductor.org/packages/microbiomeMarker/	R package for microbiome marker identification and visualization.
UpSetR v 1.4.0	https://github.com/hms-dbmi/UpSetR/	Visualization of complex intersections of datasets.
Mikropml v 1.4.0	https://github.com/SchlossLab/mikropml	Machine learning pipeline for microbiome data.
SIAMCAT v 2.0.1	https://github.com/zellerlab/siamcat	Statistical inference of associations between microbial communities and host phenotypes.

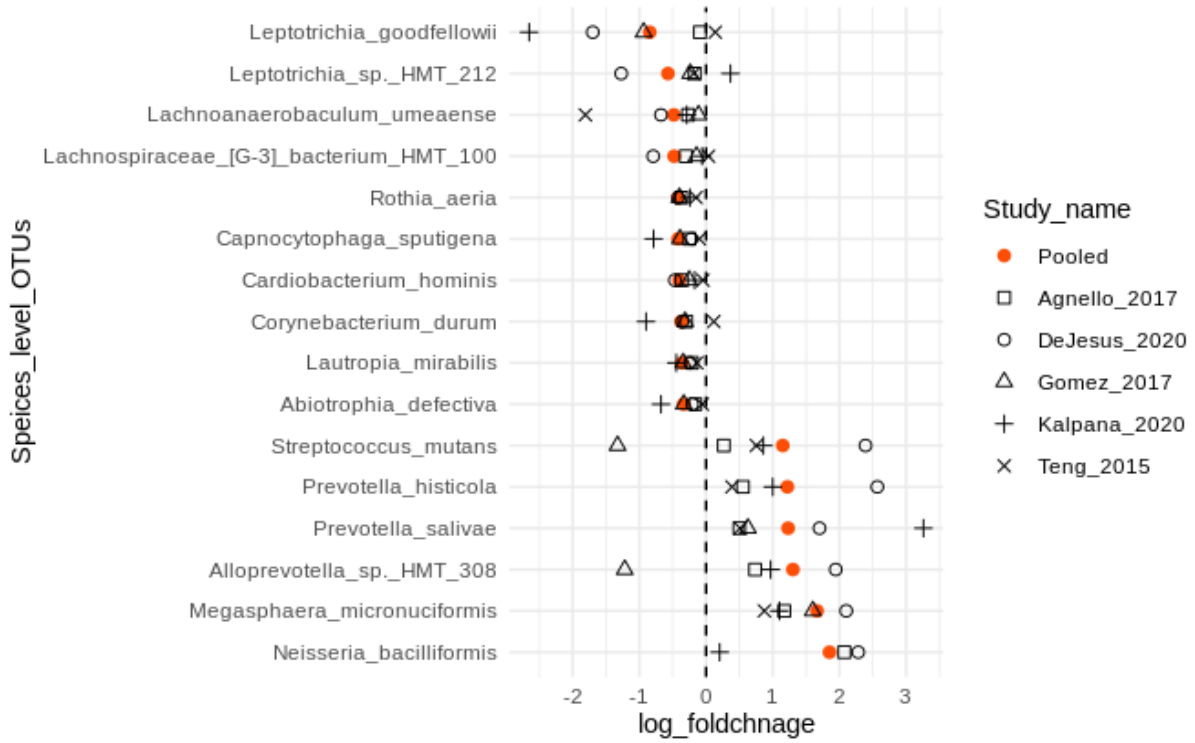


Figure S4.1: Differentially abundance taxa on raw abundance without batch correction values at species-level using DESeq2.
Only the significant taxa in Pooled studies were selected for plotting.

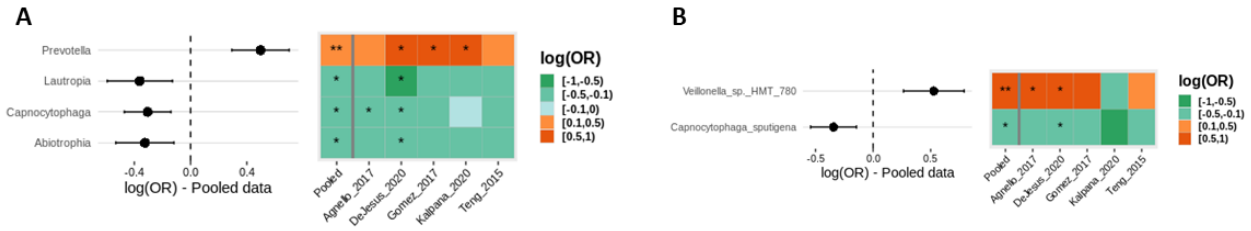


Figure S4.2: Meta-analysis for differentially abundance taxa without batch correction values.
(A) Genus-level. (B) Species-level. The heatmap represents the log odd ratios of differentially abundant with significant *p*-adjusted value (*p*-adjusted < 0.05) in pooled dataset along with the odd ratio estimates in each dataset. The forest plot signifies 95% confidence interval for the log odd ratio values for each taxon in pooled dataset.

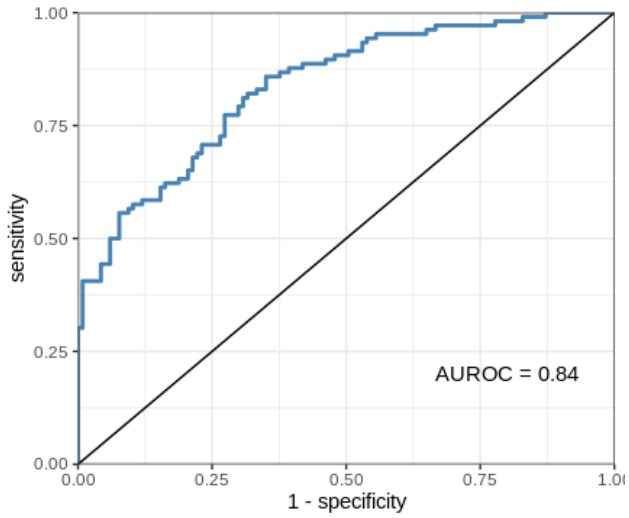


Figure S4.3: Random Forest performance in terms of AUROC with combined genus and species-level OTUs for Pooled dataset with batch-corrected values.

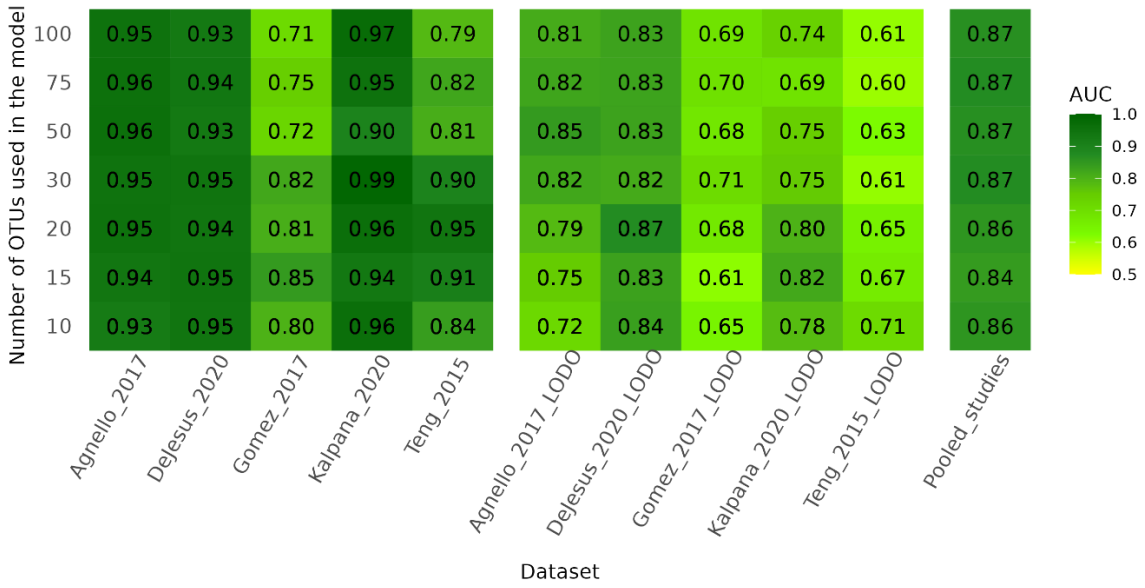


Figure S4.4: AUROC values with species-level data with CLR normalization without batch correction.

The left panel represents the cross-validation results for each dataset. The middle panel is for model performance for LODO analysis. In LODO analysis, all datasets except one were used for the training and the left-out dataset was then used for the testing to assess the generalizability of the model. The rightmost column is for the cross-validation performance of the pooled dataset. The y-axis represents the number of top OTUs used for model assessment.

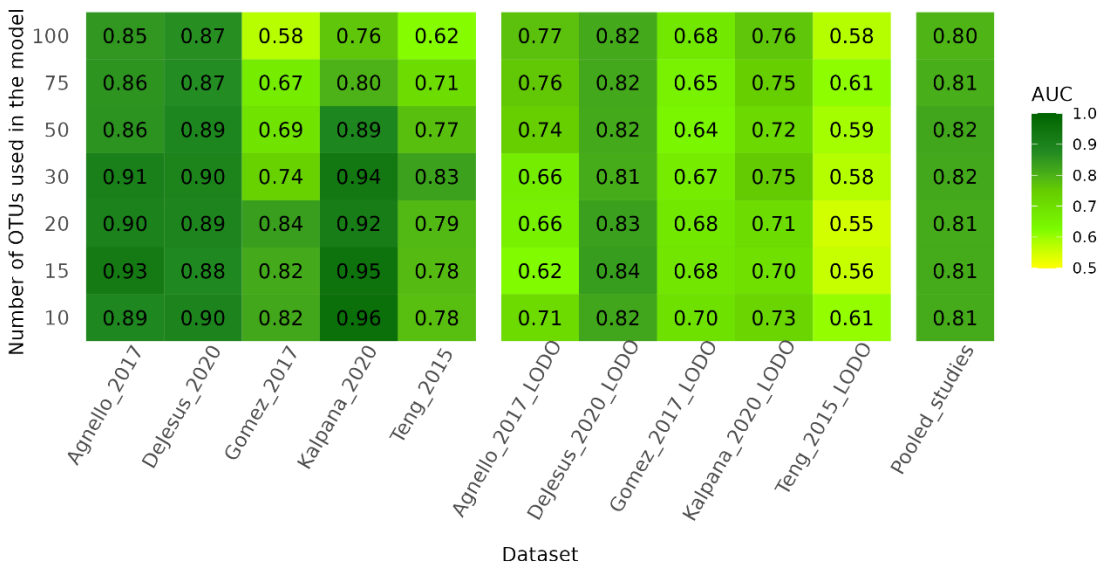


Figure S4.5: AUROC values with batch corrected values using genus-level data.
The left panel represents the cross-validation results for each dataset. The middle panel is for model performance for LODO analysis. In LODO analysis, all datasets except one were used for the training and the left-out dataset was then used for the testing to assess the generalizability of the model. The rightmost column is for the cross-validation performance of the pooled dataset. The y-axis represents the number of top OTUs used for model assessment.

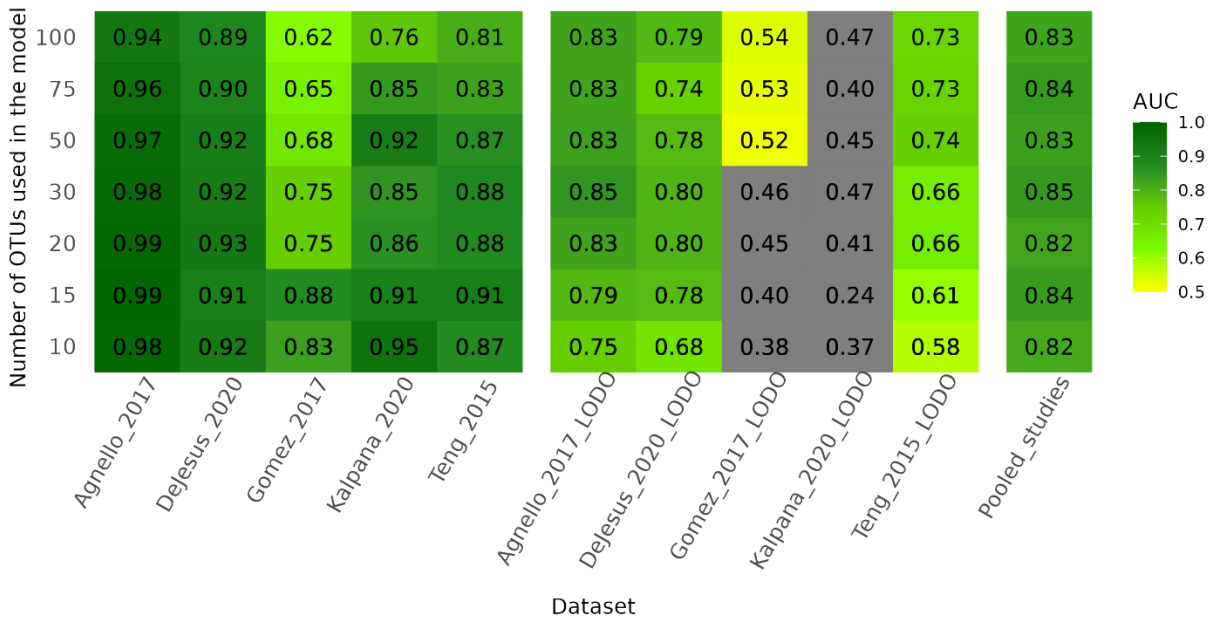


Figure S4.6: AUROC values with imputed data with CLR normalization with species-level data.

The left panel represents the cross-validation results for each dataset. The middle panel is for model performance for LODO analysis. In LODO analysis, all datasets except one were used for the training and the left-out dataset was then used for the testing to assess the generalizability of the model. The rightmost column is for the cross-validation performance of the pooled dataset. The y-axis represents the number of top OTUs used for model assessment.

Bridge to chapter 5

The previous chapter conducted a cross-cohort comparison using dental plaque datasets with available raw data and included 80 samples from our pilot study. Whereas, the analyses in the following chapter are based entirely on samples collected from and processed in our lab, comprising 534 participants. These analytical approaches explored in the previous chapter are utilized in the next chapter. Using a larger dataset of 534 participants, including both ECC and CF children, we investigate the interkingdom associations and the relative importance of bacterial and fungal markers for ECC prediction. As we found that the generalizability of ECC prediction is moderate to good on external datasets, the results from the previous chapter also inform how generalizable the findings from the next chapter may be to other populations.

Chapter 5: Role of Socioeconomic Factors and Interkingdom Crosstalk in the Dental Plaque Microbiome in Early Childhood Caries

Mohd Wasif Khan, Vivianne Cruz de Jesus, Betty-Anne Mittermuller, Shaan Sareen, Victor Lee, Robert J. Schroth, Pingzhao Hu, Prashen Chelikani

Cell Reports, Volume 43, Issue 8, 114635

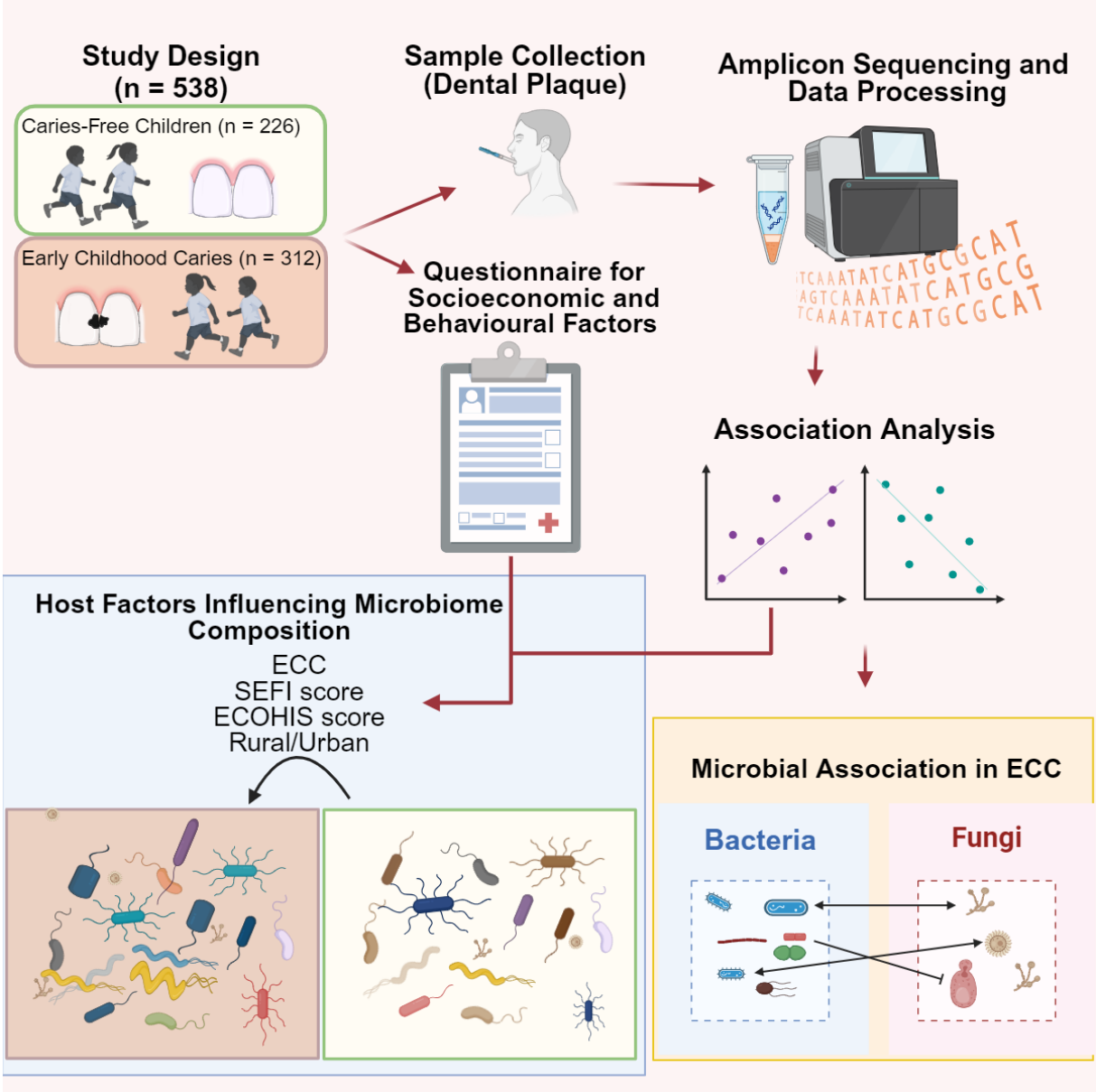
DOI: https://doi.org/10.1016/j.celrep.2024.114635

In accordance with Elsevier's author rights policy, the author is permitted to include this article in their thesis without obtaining written permission from the publisher.

AUTHOR CONTRIBUTIONS

Conceptualization: M.W.K., R.J.S., P.H., P.C.; Methodology: M.W.K., V.C.J., P.H., P.C.; Investigation: M.W.K., V.C.J., B.A.M., V.L., R.J.S.; Data curation: B.A.M., V.L., S.S.; Writing original draft preparation: M.W.K., V.C.J., P.H., P.C.; Writing review and editing: M.W.K., V.C.J., P.H., P.C., R.J.S.; Supervision: R.J.S., P.H., P.C.; Funding acquisition: P.H., R.J.S., P.C. All authors have read and agreed to the published version of the manuscript.

Graphical Abstract



5.1 Summary

Early childhood caries (ECC) is influenced by microbial and host factors, including social, behavioral, and oral health. In this cross-sectional study, we analyze interkingdom dynamics in the dental plaque microbiome and its association with host variables. We use 16S rRNA and ITS1 amplicon sequencing on samples collected from preschool children and analyze questionnaire data to examine the social determinants of oral health. The results indicate a significant enrichment (adjusted-p<0.05) of *Streptococcus mutans* and *Candida dubliniensis* in ECC samples, in contrast to *Neisseria oralis* in caries-free children. Our interkingdom correlation analysis reveals that *Candida dubliniensis* is strongly correlated with both *Neisseria bacilliformis* and *Prevotella veroralis* in ECC. Additionally, ECC shows significant associations with host variables, including oral health status, age, place of residence, and mode of childbirth. This study provides empirical evidence associating the oral microbiome with socioeconomic and behavioral factors in relation to ECC, offering insights for developing targeted prevention strategies.

5.2 Introduction

Early childhood caries (ECC) is a condition involving tooth decay in the primary dentition in children under 72 months of age. The prevalence of ECC is a major public health concern worldwide, affecting nearly half of all children globally (El Tantawi *et al.*, 2018; Uribe, Innes and Maldupa, 2021). In Canada, the prevalence of ECC in some disadvantaged communities can be as high as 98.5% (Pierce *et al.*, 2019). This high prevalence imposes significant socioeconomic and psychological burdens on families and society (Carvalho *et al.*, 2018; Bencze *et al.*, 2021). The link between ECC and future caries experiences in permanent dentition highlights its importance as a critical factor in long-term oral health (Jordan *et al.*, 2016; Tsai *et al.*, 2023). Furthermore, ECC can increase the risk of other dental diseases, malocclusion, negative oral health related quality of life (OHRQoL), and nutritional status (Davidson *et al.*, 2016; Lee *et al.*, 2022; Schroth *et al.*, 2022; Khan *et al.*, 2023).

The oral cavity is known to harbor more than 700 species of bacteria and 100 species of fungi (Peters *et al.*, 2017; Escapa *et al.*, 2018; Diaz and Dongari-Bagtzoglou, 2021). Acidogenic and aciduric microorganisms are crucial in the pathogenesis of dental caries and form an integral part of dental plaque biofilms identified in caries (Alam *et al.*, 2000; Struzycka, 2014; Chen *et al.*, 2020). Dysbiosis in microbial composition, characterized by an overabundance of acidogenic microbes, increases the risk of caries (Grier *et al.*, 2020; Chen *et al.*, 2021). Constant exposure to acidic conditions leads to the demineralization of tooth enamel (Marsh and Nyvad, 2008; Takahashi and Nyvad, 2011). This effect can be partially reversed by using fluoride, which alters the dental plaque microenvironment and promotes remineralization of the enamel (Zhang *et al.*, 2022). Host variables, encompassing various demographic, socioeconomic, and behavioral factors, also influence the development of caries (Arora, Schwarz and Blinkhorn, 2011; Pierce *et al.*, 2019; Imes *et al.*, 2021). Therefore, researchers are increasingly interested in investigating the comprehensive role of these risk factors in ECC, such as diet, fluoride exposure, limited access to care, and poor oral hygiene (Schroth, Harrison and Moffatt, 2009; Ganesh *et al.*, 2020; Adler *et al.*, 2021; Handsley-Davis *et al.*, 2021; Kahharova *et al.*, 2023).

Bacteria and fungi, present on all human sites in contact with the external environment, exhibit a diverse range of interactions. These interactions, both symbiotic and competitive, play a crucial role in maintaining the microenvironment (Barbosa *et al.*, 2016; He *et al.*, 2017). A similar trend is observed in ECC-associated bacteria and fungi, where *Streptococcus mutans* and opportunistic *Candida albicans* can exhibit both synergistic and antagonistic behaviors in ECC (Lu *et al.*, 2023). These interactions can affect the host through pathogenesis induced by ecological shifts (Balakrishnan *et al.*, 2021). However, previous studies have generally limited their focus to interkingdom interactions at the genus level or specific microbial species (Krzyściak *et al.*, 2017; Kim *et al.*, 2021; Tu *et al.*, 2022). Therefore, further investigations are necessary to enhance our understanding of interkingdom microbial signaling and their interactions with host risk factors in caries development. This study highlights that interkingdom associations exist between cariogenic species *Neisseria baciliformis* and *Candida dubliniensis*, suggesting a potential role of microbial interactions in caries development. It also shows that both bacterial and fungal species are important for caries prediction using machine learning. Furthermore, the socioeconomic index is associated with both caries status and specific caries-associated microbial markers.

5.3 Results

5.3.1 Microbiome community analyses

A total of 230 bacterial and 95 fungal species were consisted in the filtered amplicon sequence variant (ASV) tables obtained from QIIME2 analysis. Based on the number of bacterial species, our sample size of 538 would achieve a power of 0.93 to detect at least a two-fold increase in the top 10 species associated with ECC at a significant level 0.05. The most abundant bacterial species identified in bacterial data were *Haemophilus parainfluenzae, Corynebacterium matruchotii, and Lautropia mirabilis* (**Figure 5.1A**). The alpha diversity metric, Chao1, revealed significant differences between CF and ECC status suggesting a higher number of total species in ECC samples, however, these differences were not observed in the Shannon and Simpson diversity indices. Chao1 diversity estimates the total number of species in a community based on the number of rare species, that is, species that are present in only a few individuals (**Figure 5.1B**). Upon comparing the beta diversity

differences among the microbial communities, a significant difference was found between CF and ECC (**Figure 5.1C**). For fungi, the most common species were *Candida dubliniensis*, *C. albicans,* and *Blumeria graminis* with the former two exhibiting clear differences in their abundance between CF and ECC (**Figure 5.1A**). The fungal data was found to be significantly different for Chao1, Shannon, and diversity metrics, although the diversity was lower than the corresponding bacterial diversities (**Figure 5.1B**). Similar to bacterial beta diversity, the diversity in fungi was also significantly different between CF and ECC. However, the distinction in fungal diversity was more clearly defined compared to that in the bacteriome. (**Figure 5.1C**).

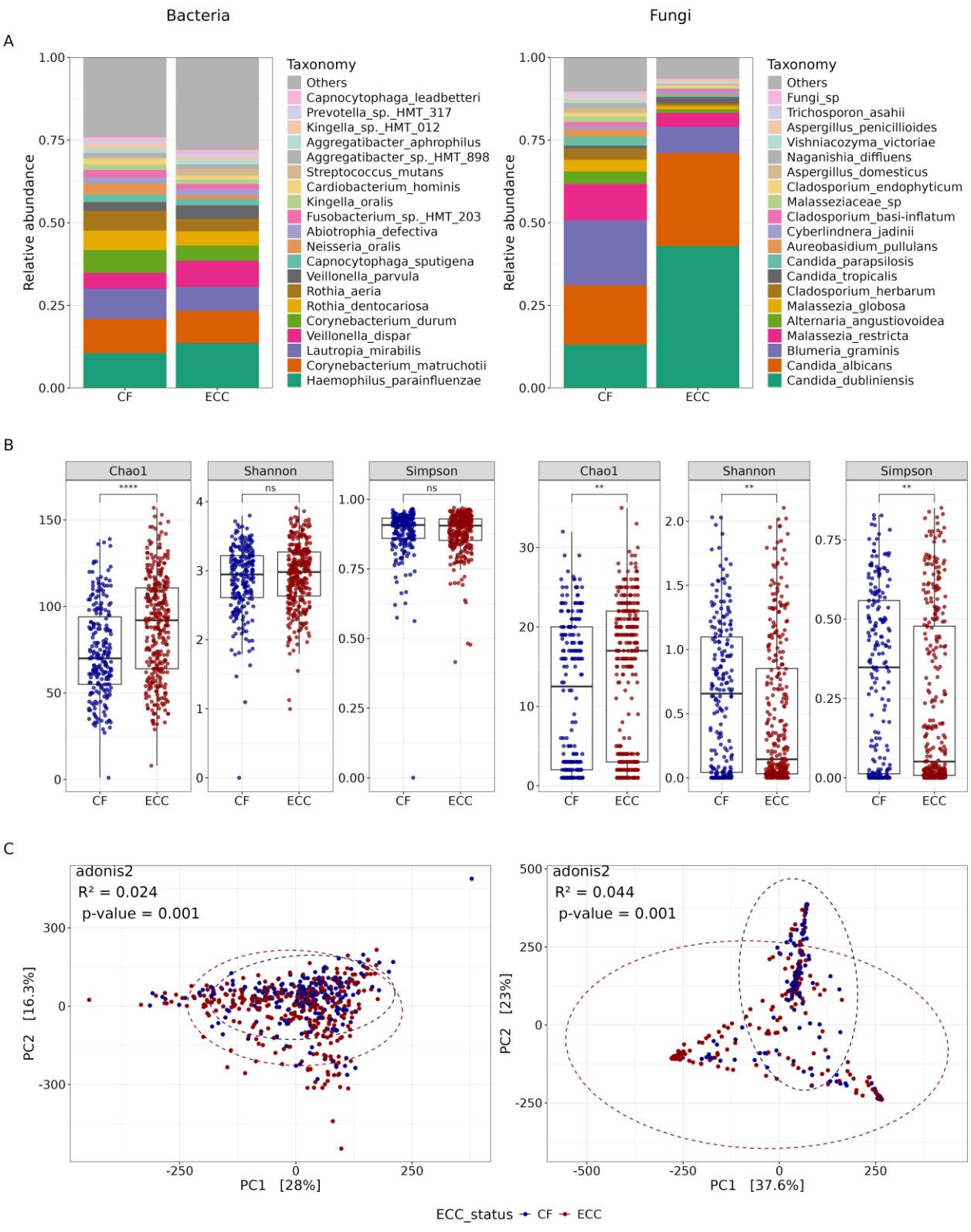


Figure 5.1: Microbiome composition and alpha and beta diversity in caries-free and early childhood caries.

(Left panel for bacteria and right panel represents fungi)

(A) Relative abundance of most abundant species. (B) Boxplot for alpha diversity comparison for Chao1, Shannon and Simpson. Each dot represents one sample and boxplot representing median and 1.5 interquartile range. (C) Principal component analysis for beta diversity comparison using adonis2 function for bacteria (left) and fungi (right).

5.3.2 Interkingdom network analysis

To study the correlation among species abundance at both intra and interkingdom levels, we conducted network analyses for ECC and CF samples separately. To visualize ECC-relevant species, we selected only those species found to be associated with ECC status using the MaAsLin2 model. When comparing the networks of the CF samples with those of the ECC samples, more positive connections were observed in the ECC samples than in the CF samples. *Prevotella* species exhibited numerous novel and strong positive associations in ECC samples compared to CF samples. In the differential network plot, *C. dubliniensis* demonstrated positive associations with *Neisseria bacilliformis* and negative associations with *Streptococcus salivarius* and *Prevotella veroralis* (**Figure 5.2**). Another interkingdom negative correlation in ECC was observed between ECC-associated *C. albicans* and CF-associated *Corynebacterium durum*, *Cladosporium herbarum*, and *Lautropia mirabilis*. Species exhibiting positive correlations in ECC group included *Selenomonas sputigena*, *P. veroralis*, *Selenomonas flueggei*, *Prevotella salivae*, and *Streptococcus salivarius*. Additional connections involved *S. sputigena*, *P. veroralis*, and *N. bacilliformis*. A third path connecting ECC-associated species and originating from *S. sputigena* encompassed *Leptotrichia shahii*, and *Prevotella olurum*. Another distinct connection, not related to these species, was between *S. mutans* and *Scardovia wiggsiae*. Interestingly, *C. albicans* was positively associated with S. *mutans* in CF samples, this association was absent in ECC samples while associations between *C. dubliniensis* and *S. mutans* were negative in CF which shifted to no interactions in ECC. The differential network revealed that numerous associations between the CF and ECC samples changed from a state of no interaction to either positive or negative. However, it is noteworthy that we did not identify any instances of drastic shifts in which associations changed from a negative to positive state, or vice versa. Moreover, ECC-associated species appeared more dynamic, exhibiting a higher number of both positive and negative associations than those associated with CF.

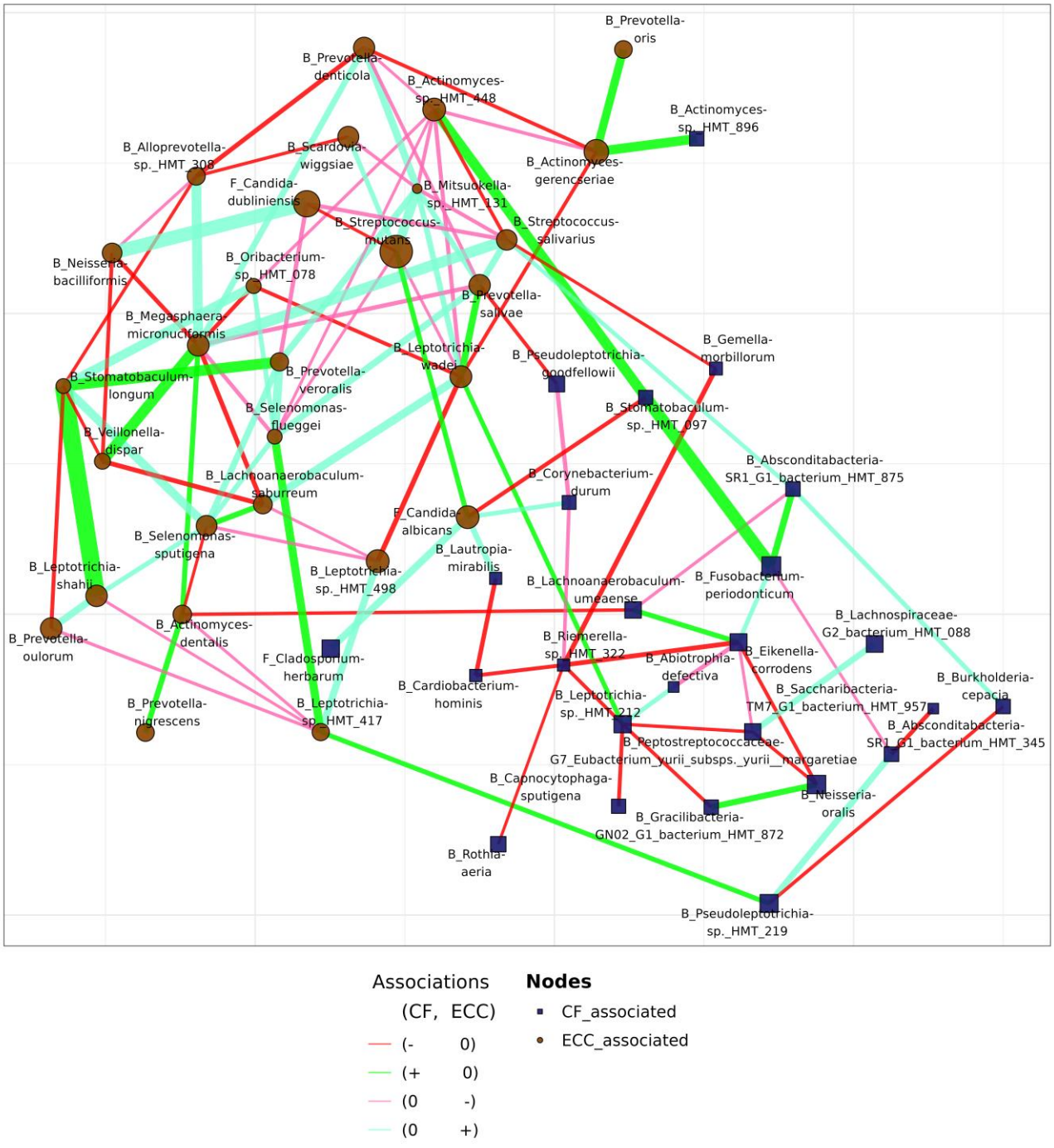


Figure 5.2: Differential network analysis for CF and ECC samples on combined bacterial and fungal datasets

The network plot visualizes the associations between species identified as differentially abundant, highlighting their relevance to CF and ECC conditions. Each node represents one species, and the size of the node is proportional to its coefficient value, indicating the strength of association with either CF or ECC conditions. Edges between nodes signify the interactions from the differential matrix, with each edge type categorized by the nature of interaction within CF or ECC samples: '0' denotes no correlation, '-' indicates a negative correlation, and '+' signifies a positive correlation between two

species. To limit the edge connections, only the interactions involving species identified as differentially abundant through MaAsLin2 analysis are included.

5.3.3 Classification of caries status using microbiome data

We evaluated the performance of the classification model separately for bacterial and fungal data, and then on the combined dataset. Among the five machine learning classifiers selected for the classification of CF and ECC, the random forest algorithm emerged as the most effective for bacteria and the combined data of bacteria and fungi, achieving Area Under the Receiver Operating Characteristic curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC) values of 0.92 and 0.93, respectively (**Figure 5.3A**). The performance of random forest and LightGBM was comparable for fungi, with AUROC and AUPRC values of 0.85 and 0.89, respectively. Other classifiers, such as Lasso, Ridge, and SVM, displayed a performance similar to that of random forest for bacteria; however, they did not perform as well as random forest or LightGBM for fungi. Except for LightGBM, the overall performance of these classifiers on bacterial data was slightly better than that on fungal data, while the performance on combined data did not improve the results further. The AUPRC values remained consistent for both bacteria and combined data; however, they were notably higher for fungi than their respective AUROC values. Further investigation into the random forest model, focusing on combined data for feature importance based on the "permutations" measure, revealed that the species identified as most significant were *S. mutans* and *C. dubliniensis*, with *S. mutans* being significantly more important than any other species (**Figure 5.3B**). The other top species that appeared in the random forest model were *N. bacilliformis*, *Prevotella denticola*, *C. albicans*, and *S. wiggsiae*.

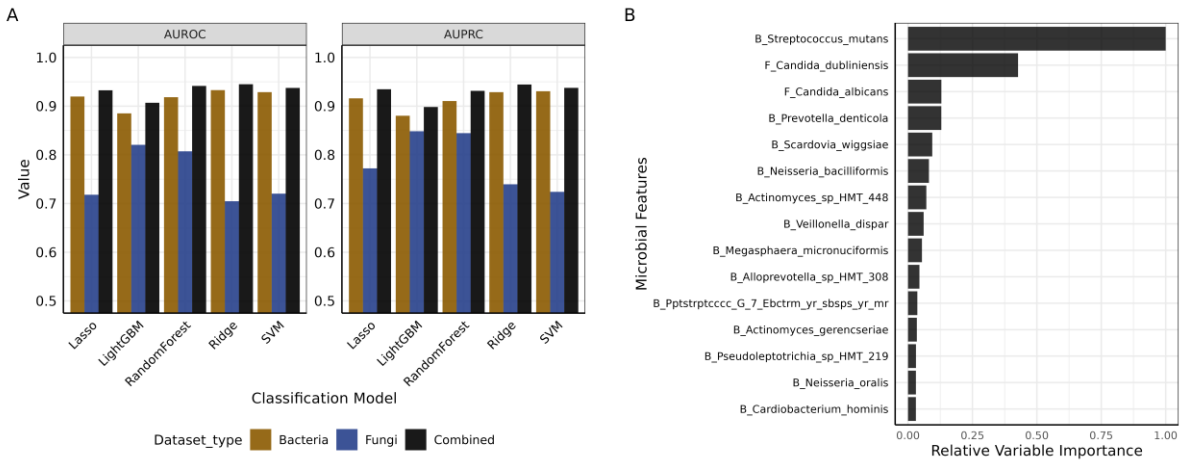


Figure 5.3: Machine learning performance for the classification of ECC and CF status. (A) Barplot for Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPRC) values of five machine learning classifiers for the classification of ECC and CF samples, evaluated for the combined dataset and for bacterial and fungal datasets separately. (B) Barplot for the variable importance derived from Random Forest model on combined bacterial and fungal datasets.

5.3.4 Association of host variables with caries status

A total of 16 host variables from 538 participants were included in the analysis. A descriptive statistic of these variables is presented in **Table 5.1**. The ECC status, age, and sex information were not missing from any of the samples, on the other hand, NutriSTEP and ECOHIS scores were not available in 79 and 80 entries, respectively. The distributions of these variables after using MICE imputation for the missing values are shown in **Figure S5.1.** Spearman correlation analysis revealed a 0.95 coefficient value between the material deprivation score and the SEFI score (**Figure 5.4A**); therefore, out of these two, only the SEFI score was used in all subsequent analyses. ECC status, ECOHIS score, and child dental health were also correlated with a correlation coefficient of > 0.65. A negative correlation was observed between the ECC status and urban residence. Our multivariable logistic regression analysis identified a strong association between ECC and other host variables such as child dental health, bedtime snacking, and residency status (**Figure 5.4B**). Of these variables, child dental health, fluoride toothpaste, vaginal birth, bedtime snacking, age, and ECOHIS score were

positively associated with ECC, whereas social deprivation score exhibited a negative trend, though not statistically significant.

Table 5.1: Descriptive statistics for the host variables analyzed.

Variable	N	CF, N = 226[1]	ECC, N = 312[1]	*p*-value[2]
Sex	538			>0.9
Female (0)		114 (50%)	157 (50%)	
Male (1)		112 (50%)	155 (50%)	
Age	538	43 (28, 55)	49 (36, 58)	<0.001
Vaginal_birth	530			0.002
No (0)		72 (32%)	62 (20%)	
Yes (1)		152 (68%)	244 (80%)	
Breastfeed_child	530			<0.001
No (0)		40 (18%)	113 (37%)	
Yes (1)		184 (82%)	193 (63%)	
Bottlefeed_child	531			0.012
No (0)		64 (29%)	59 (19%)	
Yes (1)		160 (71%)	248 (81%)	
Bedtime_snack	536			<0.001
No (0)		143 (64%)	127 (41%)	
Yes (1)		81 (36%)	185 (59%)	
Child_dental_health	534			<0.001
Very good (1)		118 (53%)	12 (3.8%)	
Good (2)		91 (41%)	64 (21%)	
Fair (3)		12 (5.4%)	113 (36%)	
Poor (4)		1 (0.5%)	108 (35%)	
Very poor (5)		0 (0%)	15 (4.8%)	
Times_day_brushed	536			0.003
Never (1)		2 (0.9%)	2 (0.6%)	
Sometimes (2)		5 (2.2%)	29 (9.3%)	
Once daily (3)		85 (38%)	130 (42%)	
Twice daily (4)		129 (58%)	147 (47%)	
>2 time daily (5)		3 (1.3%)	4 (1.3%)	
Fluoride_toothpaste	492			<0.001
No (0)		63 (30%)	38 (14%)	
Yes (1)		148 (70%)	243 (86%)	
Urban_status	534			<0.001

Variable	N	CF, N = 226[1]	ECC, N = 312[1]	*p*-value[2]
Rural (0) Urban (1)		8 (3.6%) 216 (96%)	156 (50%) 154 (50%)	
Material_depr_score (Material deprivation score)	498	-0.10 (-0.69, 0.44)	0.75 (0.06, 1.72)	<0.001
SEFI_score (Socioeconomic Factor Index score)	498	-0.06 (-0.67, 0.71)	0.87 (0.04, 1.89)	<0.001
Social_depr_score (Social deprivation score)	498	0.17 (-0.36, 1.01)	-0.01 (-0.57, 0.63)	0.003
ECOHIS_total_score (Early Childhood Oral Health Impact Scale)	454	0 (0, 1)	6 (2, 12)	<0.001
NutriSTEP_total_score Nutrition Screening Tool for Toddlers and Preschoolers)	455	20 (16, 24)	23 (19, 27)	<0.001

N, Number of non-missing entries for each variable
CF, Caries-free
ECC, Early childhood caries
[1]n (%); Median (IQR)
[2]Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test
Explanation of variables and their coding is given in **Table S5.1**

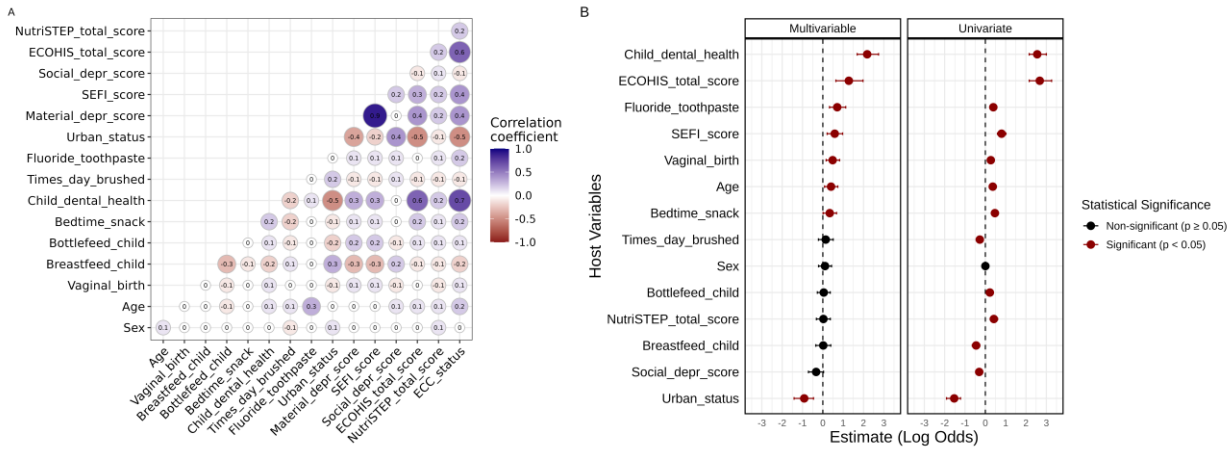


Figure 5.4: Association between host variables and early childhood caries.
(A) The correlation heatmap displays the relationships between various socioeconomic and health-related factors. Darker blue circles indicate a stronger positive correlation, while darker red circles represent a stronger negative correlation between the variables. (B) The forest plot illustrates the log odds ratios for the association of each factor with ECC status. Factors significantly associated with disease status are highlighted in red, with the horizontal lines representing the 95% confidence intervals. Features with positive log odds values indicate the likelihood of ECC, while negative values signify protective factors against ECC. The left panel shows the estimates for a multivariable model and the right panel shows the result for models with individual variables.

5.3.5 Microbial diversity and host variables

To assess the impact of each host variable on the dental plaque microbiome, we examined the Spearman correlation between each host variable and Shannon diversity, a measure of alpha diversity that explains diversity within a sample. For bacteria, three variables, SEFI score, frequency of tooth brushing, and breastfeeding status, had a significant impact on Shannon diversity (**Figure 5.5A**). For fungi, the variables significantly associated with Shannon diversity index were SEFI score, ECC status, and ECOHIS score (**Figure 5.5A**). The observed relationship between Shannon diversity for bacteria and fungi due to the host variables demonstrated mostly contrasting trends; variables positively correlated with bacterial diversity were inversely correlated with fungal diversity, and the reverse was also true.
In the beta diversity analysis, PERMANOVA, where we tested a multivariable model using Bray-Curtis distance as the response variable, identified most of the variables as significant,

with the exceptions of sex and use of Fluoride toothpaste for both bacterial and fungal samples, vaginal birth in bacterial samples, and age, NutriSTEP score, and bedtime snacking for fungal samples (**Figure 5.5B**). However, the proportion of variance (R^2) explained by the host variables was higher in the fungal microbiome than in the bacterial microbiome.

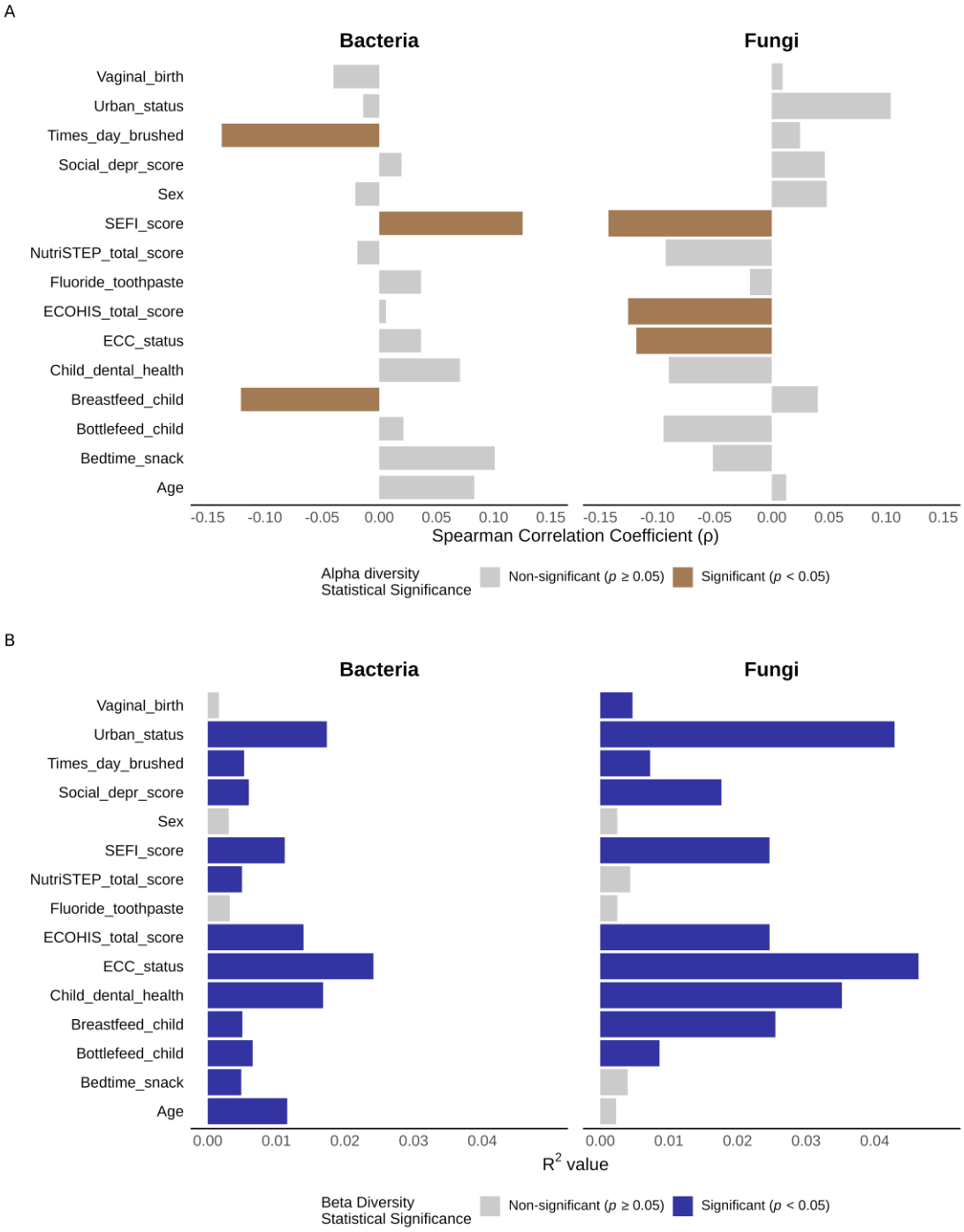


Figure 5.5: Association of the host variables with diversity metrics.
(A) Host variables and their Spearman correlation coefficient (ρ) values for alpha diversity, assessed using Shannon's metric in bacteria and fungi. (B) Beta diversity analyses of host variables via PERMANOVA, with R^2 values indicating variance explained in bacterial and fungal communities. The colored bar in each plot represents the statistical significance at $p<0.05$.

5.3.6 Association between microbiome and host variables

The associations of different species-level taxa with host variables are shown in **Figure 5.6**. The majority of microbial species associated with at least one host variable were linked to ECC status, exhibiting a similar pattern to species influenced by child dental health. After controlling for age, sex, and urban status, the mode of childbirth, NutriSTEP score, and fluoride toothpaste use showed no association with any species. While other variables related to eating habits such as breastfed child, bottle-fed child, and bedtime snacking habit were associated with not more than three species. *S. mutans*, which also showed the most prominent association with ECC status, was also found to be strongly associated with child dental health, SEFI score, and ECOHIS score. *Prevotella* and *Alloprevotella* (formerly classified as Prevotella) species were among the most prevalent in ECC-associated species, for instance, *Prevotella* species *P. salivae, P. oulorum, P. denticola*, *P. nigrescens, P. veroralis*, and *Alloprevotella* sp. HMT-308. *Alloprevotella* sp. HMT-912, *P. maculosa,* and *P. oris* were unique in that they were positively associated only with socioeconomic variables, SEFI score and social deprivation scores. For fungi, *C. dubliniensis* and *C. albicans* were found to be significantly associated with ECC status, with *C. dubliniensis* being also associated with SEFI score. Another fungal species, *Cladosporium herbarum*, was found to be significantly associated with CF status but not with any other host variables. Many species strongly associated with the SEFI score were also strongly associated with ECC status, even after adjusting for place of residence, suggesting that ECC is highly linked to socioeconomic conditions. This further emphasizes that socioeconomic factors are associated with both caries-causing and caries-protective species.

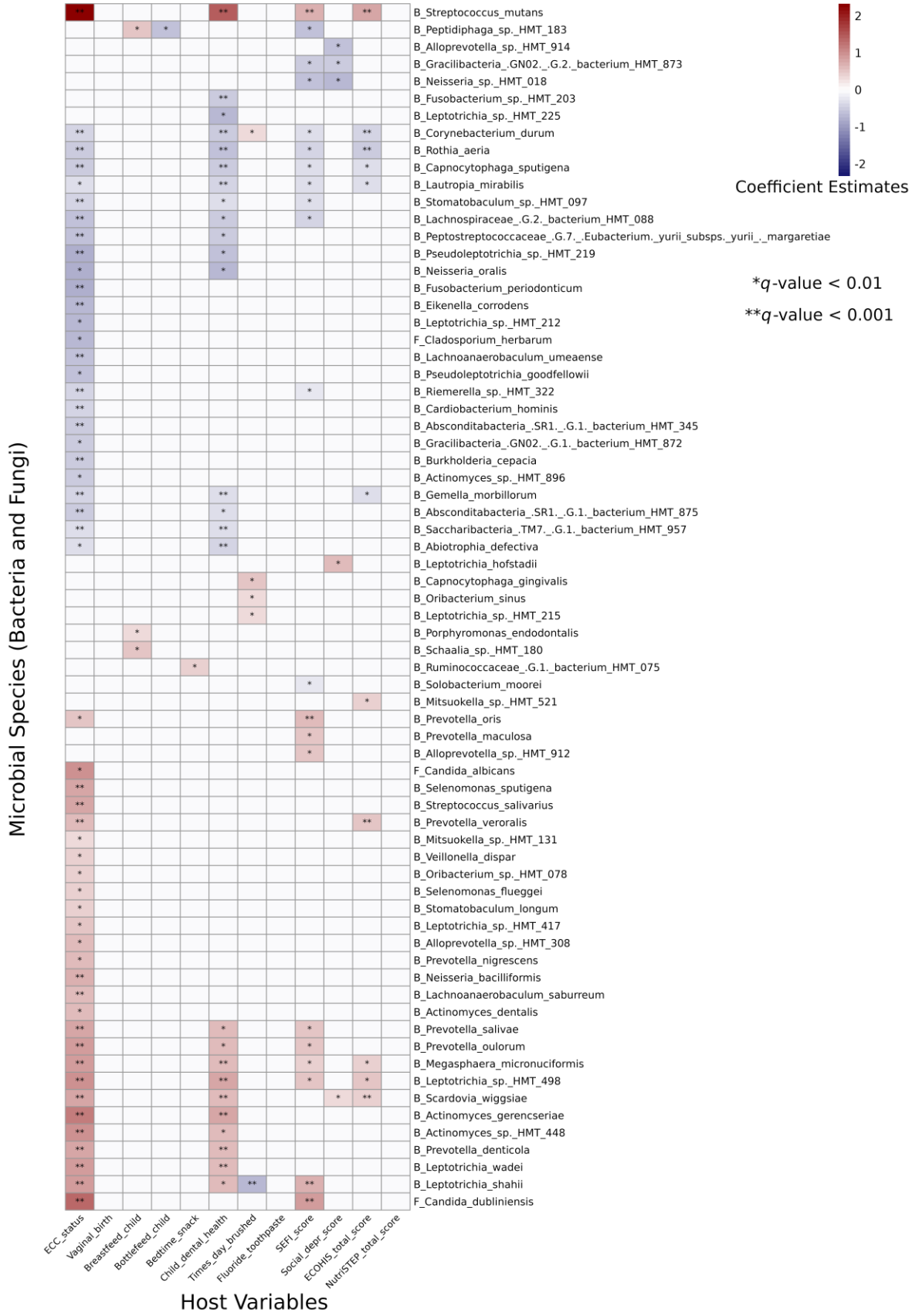

Microbial Species (Bacteria and Fungi)
B_Streptococcus_mutans
B_Peptidiphaga_sp._HMT_183
B_Alloprevotella_sp._HMT_914
B_Gracilibacteria_.GN02._.G.2._bacterium_HMT_873
B_Neisseria_sp._HMT_018
B_Fusobacterium_sp._HMT_203
B_Leptotrichia_sp._HMT_225
B_Corynebacterium_durum
B_Rothia_aeria
B_Capnocytophaga_sputigena
B_Lautropia_mirabilis
B_Stomatobaculum_sp._HMT_097
B_Lachnospiraceae_.G.2._bacterium_HMT_088
B_Peptostreptococcaceae_.G.7._.Eubacterium._yurii_subsps._yurii_._margaretiae
B_Pseudoleptotrichia_sp._HMT_219
B_Neisseria_oralis
B_Fusobacterium_periodonticum
B_Eikenella_corrodens
B_Leptotrichia_sp._HMT_212
F_Cladosporium_herbarum
B_Lachnoanaerobaculum_umeaense
B_Pseudoleptotrichia_goodfellowii
B_Riemerella_sp._HMT_322
B_Cardiobacterium_hominis
B_Absconditabacteria_.SR1._.G.1._bacterium_HMT_345
B_Gracilibacteria_.GN02._.G.1._bacterium_HMT_872
B_Burkholderia_cepacia
B_Actinomyces_sp._HMT_896
B_Gemella_morbillorum
B_Absconditabacteria_.SR1._.G.1._bacterium_HMT_875
B_Saccharibacteria_.TM7._.G.1._bacterium_HMT_957
B_Abiotrophia_defectiva
B_Leptotrichia_hofstadii
B_Capnocytophaga_gingivalis
B_Oribacterium_sinus
B_Leptotrichia_sp._HMT_215
B_Porphyromonas_endodontalis
B_Schaalia_sp._HMT_180
B_Ruminococcaceae_.G.1._bacterium_HMT_075
B_Solobacterium_moorei
B_Mitsuokella_sp._HMT_521
B_Prevotella_oris
B_Prevotella_maculosa
B_Alloprevotella_sp._HMT_912
F_Candida_albicans
B_Selenomonas_sputigena
B_Streptococcus_salivarius
B_Prevotella_veroralis
B_Mitsuokella_sp._HMT_131
B_Veillonella_dispar
B_Oribacterium_sp._HMT_078
B_Selenomonas_flueggei
B_Stomatobaculum_longum
B_Leptotrichia_sp._HMT_417
B_Alloprevotella_sp._HMT_308
B_Prevotella_nigrescens
B_Neisseria_bacilliformis
B_Lachnoanaerobaculum_saburreum
B_Actinomyces_dentalis
B_Prevotella_salivae
B_Prevotella_oulorum
B_Megasphaera_micronuciformis
B_Leptotrichia_sp._HMT_498
B_Scardovia_wiggsiae
B_Actinomyces_gerencseriae
B_Actinomyces_sp._HMT_448
B_Prevotella_denticola
B_Leptotrichia_wadei
B_Leptotrichia_shahii
F_Candida_dubliniensis
ECC_status
Vaginal_birth
Breastfeed_child
Bottlefeed_child
Bedtime_snack
Child_dental_health
Times_day_brushed
Fluoride_toothpaste
SEFI_score
Social_depr_score
ECOHIS_total_score
NutriSTEP_total_score
Host Variables
2
1
0
-1
-2
Coefficient Estimates
*q-value < 0.01
**q-value < 0.001

Figure 5.6: Associations between microbial features and various socioeconomic factors. The coefficient estimate values are after adjusting for age, sex and rural/urban status using MaAsLin2 method for differential abundance analysis. Features are categorized as B_ for bacterial and F_ for fungal. Significance levels are denoted as: * *q*-value < 0.01 and ** *q*-value < 0.001, where *q*-value = Benjamini-Hochberg adjusted *p*-value.

5.4 Discussion

This study represents a large-scale investigation into how socioeconomic and behavioral factors are associated with the diversity and composition of the dental plaque microbiome in the context of caries and CF conditions in the Canadian preschool population. Previous research on the factors shaping the oral microbiome has primarily focused on ECC status, whereas other host variables included in this study remain largely underexplored. Our research found that socioeconomic status and place of residence (rural or urban), along with certain behavioral factors such as bedtime snacking habits, are associated with ECC, as well as with variations in plaque microbial diversity (**Figure 5.4B** and **Figure 5.5**). Additionally, we identified specific taxonomic differences attributable to these factors (**Figure 5.6**). Furthermore, we identified potential shifts in the interkingdom interactions between bacteria and fungi in the presence and absence of ECC (**Figure 5.2**).

The differences in oral microbiome composition were very prominent in fungi, as the two most common species, *C. dubliniensis* and *C. albicans*, occupied close to 75% abundance in ECC, whereas their abundances were very low in children who were CF. In contrast, the taxonomic changes in the bacterial data were more subtle. Fungal data also showed significantly higher Shannon and Simpson diversities in samples from CF children compared to ECC; however, these differences were not significant for bacterial composition. This indicates that while the overall number of species and their distribution within groups may be similar, as suggested by alpha diversity, the specific species present or their relative abundances across samples vary between the groups. For beta diversity, both fungal and bacterial data showed a similar pattern, where differences between the CF and ECC groups were significant but had a low effect on overall differences. The results for alpha diversity are similar to those of our previous publication, where a significant change was observed in Chao1 diversity but not in Shannon diversity in bacteria (de Jesus, Khan, *et al.*, 2021). A similar trend has been previously

observed in the salivary bacteriomes of CF children and those with caries (Grier *et al.*, 2020). Although not significant, the mean Shannon diversity value in our bacteriome was slightly increased in ECC, which is similar to an earlier report on the dental plaque microbiome (Richards *et al.*, 2017). We also found Shannon diversity to be the most commonly used metric for comparing alpha diversity (Richards *et al.*, 2017; Jiang *et al.*, 2018; Grier *et al.*, 2020; de Jesus, Khan, *et al.*, 2021; Alyousef *et al.*, 2023). Owing to its standardized, relative, and comprehensive analysis, which considers both richness and evenness, we suggest including it in future publications for proper diversity comparison.

Our present study suggests that a species-level analysis is crucial for understanding the etiology of ECC. This importance extends beyond species-specific biochemical transformations, as species belonging to the same genus can exhibit contrasting effects leading to disease pathology. For instance, *N. bacilliformis* was found to be associated with ECC, whereas *Neisseria oralis* was associated with CF status. Similar distinctions were observed with *L. shahii* (ECC) compared to *Leptotrichia* sp. HMT-212 and *Leptotrichia* sp. HMT-225 (CF) and among some *Prevotella* species (**Figure 5.6)**. This conclusion is supported by findings from other studies (Xu *et al.*, 2014; Jiang *et al.*, 2016). This pattern also applies to the genus *Streptococcus*. While *S. mutans* and *Streptococcus salivarius* were significantly associated with ECC, the genus *Streptococcus* as a whole was not (**Figure S5.2B**). These results are consistent with those of other studies (Jiang *et al.*, 2016; Yang *et al.*, 2023).

Several studies have investigated the interkingdom associations, particularly between bacteria and fungi, in the oral microbiome (Sztajer *et al.*, 2014; Balakrishnan *et al.*, 2021). We found an interaction between *C. dubliniensis* and *N. bacilliformis*, which was exclusively observed in the ECC cohort. Notably, these two species were not found to be positively associated with any other species in the ECC group. Such interkingdom associations within ECC microbiomes derived from dental plaque samples have rarely been reported (de Jesus *et al.*, 2020). The significance of adhesins in the *Neisseria* genus was highlighted by Hung and Christodoulides (2013) (Hung and Christodoulides, 2013). The adhesins may help facilitate the co-existence of *Neisseria* and *Candida* species. Given that such interactions have not yet been documented in other studies and that research on *Candida* species other than *C.*

albicans remains limited (Montelongo-Jauregui and Lopez-Ribot, 2018), the role of these interactions in ECC pathogenesis are yet to be experimentally confirmed. Furthermore, most studies on interactions in the oral microbiome are restricted to *C. albicans* and *Streptococci*, specifically, the interactions between *C. albicans* and *S. mutans* and their co-localization in ECC (Montelongo-Jauregui and Lopez-Ribot, 2018; Du *et al.*, 2022). In contrast, our findings suggest that the interactions between *C. albicans* and *S. mutans* are more prominent in the CF group than in the ECC group (**Figure 5.2**). A potential explanation for this observation could be that both synergistic and antagonistic behaviors are observed between *C. albicans* and *S. mutans* (Krüger *et al.*, 2019). These species act synergistically in biofilm formation, while antagonistic interactions arise due to the inhibition of filamentation in *C. albicans* by *S. mutans*. It is important to note that the graphical representations in this study illustrate the co-abundances of individual microbial species and may not be interpreted as direct evidence of biological interactions or causation (**Figure 5.2**). Moreover, the abundances of both *C. albicans* and *S. mutans* are higher in ECC samples when compared to CF samples, highlighting their positive associations with ECC (**Figure 5.6**). Differences between our findings and others could also be due to the sampling technique used in this study. While we focused on the effects of the disease and socioeconomic factors on the overall dental plaque microbiome, studies looking at the tooth-level microbiome may provide a localized picture of the interactions between bacteria and fungi (Mann *et al.*, 2023; O'Connell *et al.*, 2023).

Our intrakingdom associations support recent findings where species correlations within the same genus of *Leptotrichia* and *Prevotella* were high (Cho *et al.*, 2023). Furthermore, Cho et al. showed *S. mutans* interactions with *P. salivae* and *Leptotrichia* wadei, which in our results was not direct but appears to be mediated by another species *Mitsuokella* sp. HMT-131. While their *in vivo* study showed that *S. sputigena* interacts with *S. mutans*, a clear correlation was not observed in their metagenomic analysis. Similarly, in our analysis, both *S. sputigena* and *S. mutans* were found to be significantly associated with ECC status, but a correlation between them was not observed. On the other hand, *S. mutans* was found to be significantly correlated with *S. wiggsiae*, supporting a previous study in adolescents' caries (Eriksson *et al.*, 2018).

Machine learning-based classifiers are instrumental in distinguishing between the two conditions based on microbiome data. Previous research on caries has used machine learning to identify caries-related biomarkers (de Jesus *et al.*, 2020; Grier *et al.*, 2020; de Jesus, Khan, *et al.*, 2021; Butcher *et al.*, 2022). Moreover, these models help to determine the contribution of each species to this classification, which indirectly suggests their empirical importance. We identified random forest as the best among the machine learning algorithms selected in our study. Other studies have also found random forest to provide good predictive potential. In addition to its prediction capacity, random forest also provides the importance of the variables used in the model, which aids in identifying the species ranking in outcome prediction. Given the variations in performance shown by other classifiers, hyperparameter spaces, and model interpretability, we recommend using the random forest method for such analyses. In our machine learning analysis for feature importance, in addition to *Streptococcus*, *Candida*, and *Neisseria* species, *S. wiggsiae* emerged as one of the top feature species for classifying ECC and CF samples. The prominence of ECC-associated species in the model can potentially be attributed to the significant shifts observed in these species between CF and ECC conditions. Even though bacterial samples demonstrated superior classification performance for CF and ECC compared to fungal samples, we suggest including both bacterial and fungal species in future studies (**Figure 5.3A**). This recommendation is based on the observation that fungal species are among the top-ranking features based on their variable importance (**Figure 5.3B**). The AUROC metric is better suited for perfectly balanced classes. As our samples were not perfectly balanced for disease outcome, we also included the AUPRC metric, which is less affected by class imbalances. The performance remained consistent across both metrics, potentially owing to the large sample size, which is critical for machine learning, especially when the feature count is nearly 300, as in our case. Our large sample size, supporting the application of machine learning methods, is evident from the high and similar AUROC and AUPRC values. These limitations and considerations were previously discussed (Grier *et al.*, 2020).

Our study found no significant association between ECC status and sex, consistent with findings from a review of Canadian studies, in which only one of five studies reported sex as a significant variable in ECC (Schroth and Cheba, 2007; Pierce *et al.*, 2019). This finding

suggests an equal prevalence of caries among male and female children. We also observed a pattern of differentially abundant species linked to child dental health, similar to that observed in ECC, attributable to a high correlation between these variables. We also identified that a higher SEFI score, which indicates less favorable socioeconomic conditions, is associated with an increased risk of ECC. Since socioeconomic variables were derived from the participants' postal codes, they were assessed for correlations. The material deprivation score exhibited a high correlation with the SEFI score (Spearman correlation=0.94); consequently, only the SEFI score was selected for further analysis. Due to this high correlation, we anticipate a strong convergence between these two variables. In contrast, the third social variable, the social deprivation score, was not significantly correlated with the SEFI score (Spearman correlation=0.18), therefore no convergence was observed in the multivariable model. Additionally, ECC's association with bedtime snacking habits shows that ECC is influenced by an intricate interplay of biological, behavioral, and socioeconomic factors (Anil and Anand, 2017; Hussein *et al.*, 2017). For differential species analysis, we used age, sex, and place of residence as confounding factors. The influence of age and sex on ECC microbial differences has been shown previously (Xu *et al.*, 2018; de Jesus *et al.*, 2020). The variability in differentially abundant species with and without confounding factors is compared in **Figure 5.6** and **Figure S5.2A**. The association between ECC and bedtime snacking habits underscores the importance of dental hygiene, especially before bedtime, in preventing caries. Furthermore, the association between fluoride toothpaste and ECC could be attributed to the more prevalent use of fluoride toothpaste by children with existing poor dental health. Future studies could also benefit from including variables such as dental flossing, the length of breastfeeding, and bottle-feeding on the ECC outcome.

A high prevalence of *Prevotella* spp. associated with ECC was observed in our analysis. *Prevotella* species serve as biomarkers for ECC detection and their role as potential periodontal pathogens has been widely recognized (Teng *et al.*, 2015; He *et al.*, 2018; Yang *et al.*, 2023). In contrast, none of the *Lactobacillus* species were identified as ECC-associated. *Lactobacillus* species have been previously found to be associated with dental caries (Chen *et al.*, 2020). From our results, *N. oralis* can be considered a potential biomarker for CF status,

while the role of *N. bacilliformis* in ECC has been further emphasized by recent studies (Cherkasov *et al.*, 2019; Lee *et al.*, 2021; Fakhruddin *et al.*, 2022).

Our findings on fungi are consistent with both our previous study and other research where *C. albicans* and *C. dubliniensis* were the most prevalent species (Al-Ahmad *et al.*, 2016; de Jesus *et al.*, 2020). These two species also appeared as biomarkers in both differential abundance analysis and random forest-based classification. The differential abundance results obtained from our analysis were also compared with those derived using an alternative method, LinDA (Zhou *et al.*, 2022). This comparison revealed a high degree of consistency between the outcomes of both the models (**Figure 5.6** and **Figure S5.3**). This consistency strengthens the validity of our results and contributes to a broader understanding of the microbial interactions in ECC.

In summary, our study investigated the impact of dental plaque microbiome, socioeconomic, and behavioral factors on the presence of ECC. We reported several novel interactions between the bacteriome and mycobiome in the ECC and CF groups. This includes interkingdom interactions between *C. dubliniensis* and *N. bacilliformis*, and *C. albicans* and *C. durum*. Our study identified key species associated with ECC, including *S. mutans*, *C. dubliniensis*, *N. bacilliformis*, among others. This analysis was facilitated using microbiome association analysis methods and machine learning models, made possible by inclusion of mycobiome sequencing and a large sample size. This study provides empirical evidence linking socioeconomic and behavioral factors, such as SEFI score and bedtime snacking habits, to ECC and offers new insights into the contributors to ECC. Our research will aid oral healthcare providers in the development and implementation of targeted intervention strategies for ECC based on the specific host variables included in the study.

Limitations of the study

Our analysis was based on 16S rRNA and ITS sequencing rather than metagenomic data. ITS sequencing is widely accepted for studying fungal composition owing to its reasonable discriminatory power and well-defined reference databases (Schoch *et al.*, 2012; Conti *et al.*, 2023). However, metagenomic sequencing offers additional advantages, as it enhances strain-level resolution, which is somewhat challenging with 16S rRNA and ITS sequencing. We

also recognize the value of longitudinal studies and matching participants by sex, age, and socioeconomic status to capture temporal differences, while reducing confounding effects. The lack of species-level information for fungi in some samples, attributed to a low number of reads, may have resulted from insufficient fungal DNA in the original samples, owing to low fungal biomass. Additionally, the UNITE database, commonly used for fungal taxonomic assignments, does not offer as high taxonomic resolution as bacterial databases such as HOMD and Silva (Quast *et al.*, 2013; Nilsson *et al.*, 2016). Furthermore, enhancing the understanding of ECC could be achieved through longitudinal birth cohort studies. Such an analysis would yield more specific differences and capture dynamic changes over time, while effectively controlling for numerous confounding variables. While many ECC-associated species and interactions identified in our analysis corroborated with previous studies, these findings may not be generalized across different populations due to the variations identified by place of residence and socioeconomic conditions. Additionally, the biases introduced by class imbalance for variables selected in the study, the non-calibrated dentists who collected the initial samples from the participants, sample processing and sequencing at different time points are also inevitable in microbiome studies, however, some biases were minimized using centered log ratio transformation. Furthermore, due to data privacy, the host's socioeconomic and behavioral data are not included in the NCBI data repository, except for host age and sex. Some studies have highlighted the association between host taste genetics and the incidence of caries, including ECC (de Jesus *et al.*, 2022; Orlova *et al.*, 2022). Therefore, the inclusion of genetic data, especially in the field of taste genetics, may help to identify the role of nutrition and food intake in the susceptibility to ECC. Additionally, multimodal machine learning tools can identify ECC contributors from the microbiome, host behaviors, socioeconomic status, and genetic components.

5.5 Methods

5.5.1 Experimental model and study participant details

This study enrolled 553 participants, who were recruited between December 2017 and July 2022 at the Misericordia Health Centre (MHC), Children’s Hospital Research Institute of Manitoba (CHRIM), and various community dental clinics in Winnipeg, Manitoba, Canada.

Among the participants, 538 children were under the age of 72 months. Children aged > 72 months or missing age information (n = 15) were excluded from the study. The remaining participants included 226 caries-free (CF) children and 312 children with ECC, implementing a cross-sectional study design. Participants with one or more decayed, missing (due to caries), or filled tooth surfaces in any primary tooth were considered as having ECC, based on the American Academy of Pediatric Dentistry definition criteria (AAPD, 2017). The dental examination to determine the CF and ECC status was performed by experienced dentists who adhered to these criteria.

A supragingival plaque sample was collected from all tooth surfaces, irrespective of cavitated or non-cavitated teeth, by brushing the teeth with a sterile interdental brush according to our previously published methods (Agnello *et al.*, 2017; de Jesus *et al.*, 2020). Children on any antibiotic and who did not meet the case definition for ECC were not selected for this study (American Academy of Pediatric Dentistry, 2003). At MHC, the samples were collected before the scheduled surgery, and for other locations, the samples were obtained during standard oral examinations by a dentist or research staff. The participants' information was collected as metadata through a questionnaire completed by parents or caregivers who also provided written informed consent. Participant recruitment, sample collection, and information collection for this study were approved by the University of Manitoba's Health Research Ethics Board (HS22388 – H2018:472). Information about age, sex and socioeconomic status of the human participants is available in **Table 5.1**. Socioeconomic status was considered in 3 different variables - Material deprivation score, Socioeconomic Factor Index score, and Social deprivation score, which are all measures of inequities. Information about gender was not collected as participants were toddlers and preschoolers. Information about ancestry and race was not collected. Limited information about ethnicity was collected but was not included in **Table 5.1** due to lack of ethics approval for including them. This study adhered to the checklist provided by the STROBE guidelines for conducting cross-sectional studies. The methodology for dental plaque sample collection and DNA extraction was consistent with the procedures described in our previous studies (de Jesus, Khan, *et al.*, 2021; de Jesus *et al.*, 2022). To estimate the statistical power of our microbiome study, an online tool "micropower"

was used for supragingival plaque at a significance level of 0.05, while maintaining other parameters at their default settings (Kelly *et al.*, 2015).

5.5.2 Amplicon sequencing

Library preparation and sequencing were performed using either MiSeq or NovaSeq6000 instruments (Illumina Inc., San Diego, CA, USA), employing paired-end (250 × 2 bp in length) sequencing technology at the Genome Quebec Innovation Center (Montreal, Canada), using amplicon sequencing targeting the V4 hypervariable region of bacterial 16S rRNA and fungal ITS1 (internal transcribed spacer 1) spacer DNA. Primers for the 16S rRNA V4 region were 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). Primers targeting the ITS1 region were ITS1-30 (5'-GTCCCTGCCCTTTGTACACA-3') and ITS1-217 (5'-TTTCGCTGCGTTCTTCATCG-3') (Usyk *et al.*, 2017; de Jesus *et al.*, 2020).

5.5.3 Sequencing data processing and taxonomic assignment

The sequencing reads received from the sequencing center were provided as demultiplexed, barcode-removed, and paired-end FASTQ files. These reads were processed using QIIME2 (version 2022.11) to create separate ASV tables for 16S and ITS1 reads (Bolyen *et al.*, 2019). For the 16S sequencing data, quality trimming, filtering, removal of chimeric reads, and merging were performed using DADA2 within the QIIME2. The read trimming length for each sequencing batch was optimized using different combinations of forward and reverse reads, and the best combination was selected for final trimming. For ITS data, the Q2-ITSxpress QIIME2 plugin was used to trim the conserved regions around ITS1 before applying DADA2 for merging and chimera removal (Bengtsson-Palme *et al.*, 2013; Rivers *et al.*, 2018). The Human Oral Microbiome Database (HOMD, version 15.23) and UNITE database, which uses dynamic clustering thresholds (version 9; QIIME release), were employed for taxonomic assignment of bacterial and fungal ASVs, respectively. Separate ASV tables consisting of the abundance at each taxonomic level were generated for bacteria and fungi. The ASV tables were filtered to retain features that exhibited a minimum prevalence of 5% across all the samples.

5.5.4 Microbiome community and network analysis

All community analyses were performed using the phyloseq object in R package “phyloseq” (version 1.40.0). For alpha diversity, the Chao1, Shannon, and Simpson diversity indices were used. Beta diversity was calculated using the adonis2 function in the “vegan” package (version 2.6-4). For the plotting of the most abundant species “microeco” (version 1.4.0) options were considered. All figures were generated in the “ggplot2” (version 3.4.3) library in R.

To identify the differentially abundant species between CF and ECC, association analysis with a linear model was implemented using MaAsLin2 (version 1.14.1) in R (Mallick *et al.*, 2021). MaAsLin2 is a statistical model designed to identify the associations between microbial taxa and clinical metadata. It employs general linear models and incorporates methods for data normalization and transformations to identify multivariable associations, while controlling for the false discovery rate. In our application of MaAsLin2, the "LM" method was utilized on center log-ratio (CLR) transformed abundance data as the response variable and ECC status as the independent variable. In this analysis age, sex, and place of residence were considered as confounding variables along with the Benjamini-Hochberg (BH) method to control for false discovery rates. To further minimize the number of false positives, associations based on a *q*-value (adjusted *p*-value) threshold of less than 0.01 were considered significant.

For the correlation network analysis, the species identified by MaAsLin2 were used to generate a network plot illustrating the associations between species using the “NetCoMi” package (version 1.1.0) in R. First, network associations for the CF and ECC groups were obtained separately, employing the association measure "SPRING". Subsequently, a differential network was generated from these two groups using “discordant” as the differential network method.

5.5.5 Classification of CF and ECC samples

To discriminate between CF and ECC samples, machine learning classifiers were constructed in R using the "tidymodels" package (version 1.1.1). To perform the classification between CF and ECC samples, we applied commonly used machine learning methods (Wirbel *et al.*, 2020; Topçuoğlu *et al.*, 2021). We tested five machine learning classifiers: LASSO logistic

regression, ridge logistic regression, Support Vector Machine (SVM), Random Forest, and LightGBM. A stratified 5-fold cross-validation approach with 10 repeats was employed. The features used in these classifiers were the CLR normalized abundance data for 16S and ITS sequences.

For the lasso and ridge logistic regression, the 'glmnet' package was utilized for L1 and L2 regularization, respectively, with fine-tuning of the loss-function hyperparameter. In the SVM model, the cost and degree hyperparameters were fine-tuned, with the other parameters set to their default values. For the random forest model, the number of trees, the minimum node size, and the number of variables at each split were tuned. In the LightGBM model, the tree depth and learning rate were fine-tuned along with the hyperparameters used in the random forest model. To ensure a comprehensive analysis, a grid size of 50 was used for tuning the hyperparameters in all selected classifiers. The performance of each method was compared using the AUROC and AUPRC metrics.

5.5.6 Analysis of socioeconomic and behavioral variables

Demographic, lifestyle, socioeconomic, and behavioral variables were collected using a comprehensive questionnaire obtained from parents or caregivers at the time of participant recruitment. These variables will be referred to as 'host variables' and include age, sex, feeding, and oral hygiene habits, among others (**Table 5.1** and **Table S5.1)**. Residential information using the postal codes of the participants was utilized for rural-urban designation as well as to determine the Socio-Economic Factor Index (SEFI) score, social deprivation score, and material deprivation score (Fransoo *et al.*, 2013). These data were obtained from the Manitoba Centre for Health Policy (MCHP) (Metge *et al.*, 2009). The Early Childhood Oral Health Impact Scale (ECOHIS) and NutriSTEP scores were determined based on responses to a standard set of questions (Pahel, Rozier and Slade, 2007; Randall Simpson *et al.*, 2008).

To assess the number of missing entries prior to imputation, the R packages “mice” (version 3.16.0) was used. Missing information in the data was imputed using the default methods in the Multiple Imputation by Chained Equations (MICE) approach for binary, categorical, and numerical data. The data was then checked for correlations among the variables using Spearman’s correlation analysis. Variables were tested for associations with ECC status using

both multivariable and univariate analytical approaches. In both models, ECC status was assigned as the outcome variable. The multivariable model incorporated all other host variables as independent variables within a single model. While the univariate approach analyzed each variable independently in separate models.

5.5.7 Microbiome and socioeconomic variables

The data was analyzed to determine the association between host variables and various microbiome phenotypes. For the alpha diversity index, a non-parametric Spearman correlation test was used. For beta diversity dissimilarities, PERMANOVA was applied using adonis2 on the Bray-Curtis distance with 999 permutations. The *p*-value for each variable was adjusted using the BH method. All statistical analyses were performed using R version 4.3.2 (R Core Team, 2023). To identify the associations between individual taxa and host variables, MaAsLin2 was used, as described above.

5.6 Supplementary data

Table S5.1: Explanation of host variables analyzed in the study.

1	ECC Status	Binary	Indicates the dental caries status of the child.	Caries free (0), ECC (1)
2	Sex	Binary	The biological sex of the child.	Female (0), Male (1)
3	Vaginal Birth	Binary	Mode of childbirth.	Cesarian (0), Vaginal (1)
4	Breastfed Child	Binary	Whether the child was breastfed.	No (0), Yes (1)
5	Bottlefed Child	Binary	Whether the child was bottle-fed.	No (0), Yes (1)
6	Bedtime Snack	Binary	Consumption of snacks by the child at bedtime.	No (0), Yes (1)
7	Urban-Rural	Binary	The residential setting of the child's home.	Rural (0), Urban (1)
8	Times Day Brushed	Ordinal	Number of times the child's teeth are brushed per day.	Never (1), Sometimes (2), 1x (3), 2x (4), >2x (5)
9	Child Dental Health	Ordinal	Parental perception of the child's dental health.	Very good (1), Good (2), Fair (3), Poor (4), Very poor (5)
10	Age	Continuous	The age of the child.	Measured in months
11	ECOHIS Total Score	Continuous	Score from the Early Childhood Oral Health Impact Scale measures the oral health-related quality of life (OHRQoL) of preschool children and their families.	Higher ECOHIS score signifies worsening oral health related quality of life.
12	NutriSTEP Total Score	Continuous	Nutrition Screening Tool for Toddlers and Preschoolers. A set of 17 questions about a child's typical food choices, eating behaviors, as well as physical	A higher score indicates a greater nutritional risk.

			activity, and growth to assess the nutrition risk for preschoolers	
13	SEFI score	Continuous	Socioeconomic Factor Index score. Derived from Census data that reflects non-medical social determinants of health and is used as a proxy measure of socioeconomic status	Lower scores indicate more favorable conditions
14	Material deprivation score	Continuous	Calculated from Canadian Census data which reflects the deprivation of wealth, goods, and conveniences	Lower scores indicate less deprivation (better status)
15	Social deprivation score	Continuous	Calculated from Canadian Census data which reflect the deprivation of relationships among individuals in the family, the workplace, and the community	Lower scores indicate less deprivation (better status)
16	Fluoride toothpaste	Binary	Use of fluoride toothpaste	No (0), Yes (1)

Footnote: Binary variables are encoded as 0 or 1 based on the alphabetical order of their categories

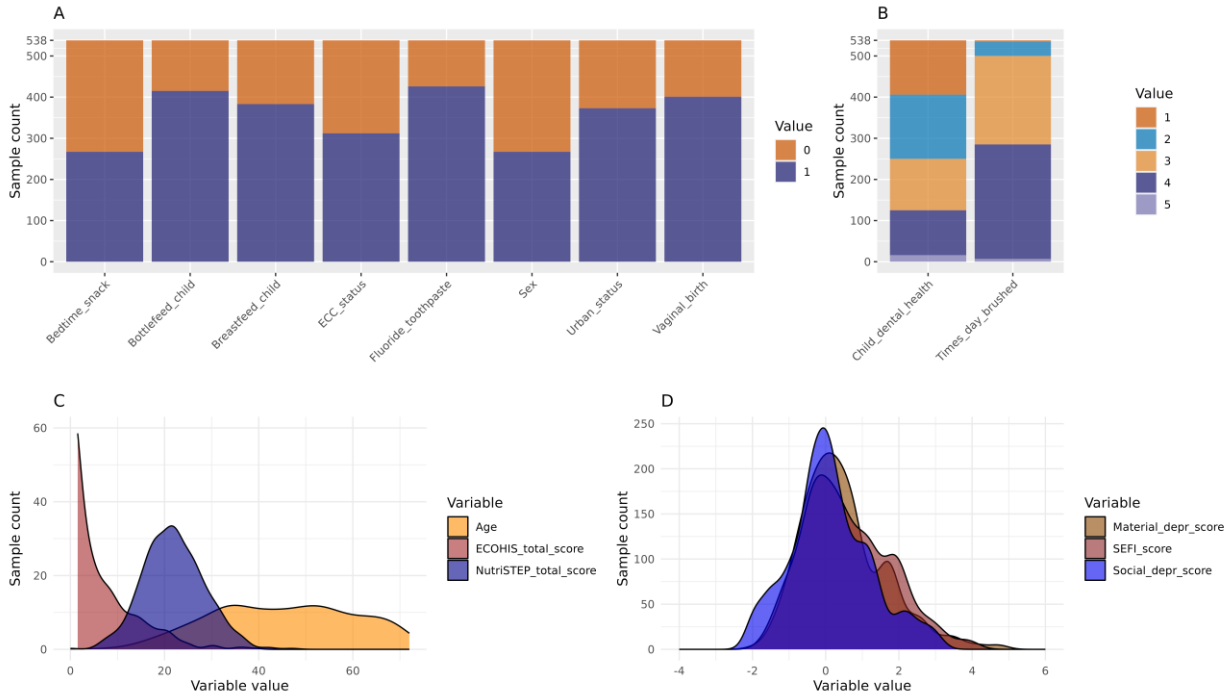


Figure S5.1: Exploratory data analysis for the distribution of host variables used as socioeconomic and behavioral factors after imputation.
(A) Binary variables. (B) Ordinal Variables. (C) Continuous variables for age, ECOHIS and NutriSTEP score. (D) Continuous variables for socioeconomic factors. The detail of these variables is provided in Table S5.1.



Figure S5.2: Association between microbial features and various socioeconomic factors. (A) Species level. (B) Genus level. The coefficient estimates values from MaAsLin2 differential abundance analysis method without adjusting for any confounder. Features are categorized as B_ for bacterial and F_ for fungal. Significance levels are denoted as: * *q*-value < 0.01 and ** *q*-value < 0.001, where *q*-value = Benjamini-Hochberg adjusted *p*-value.

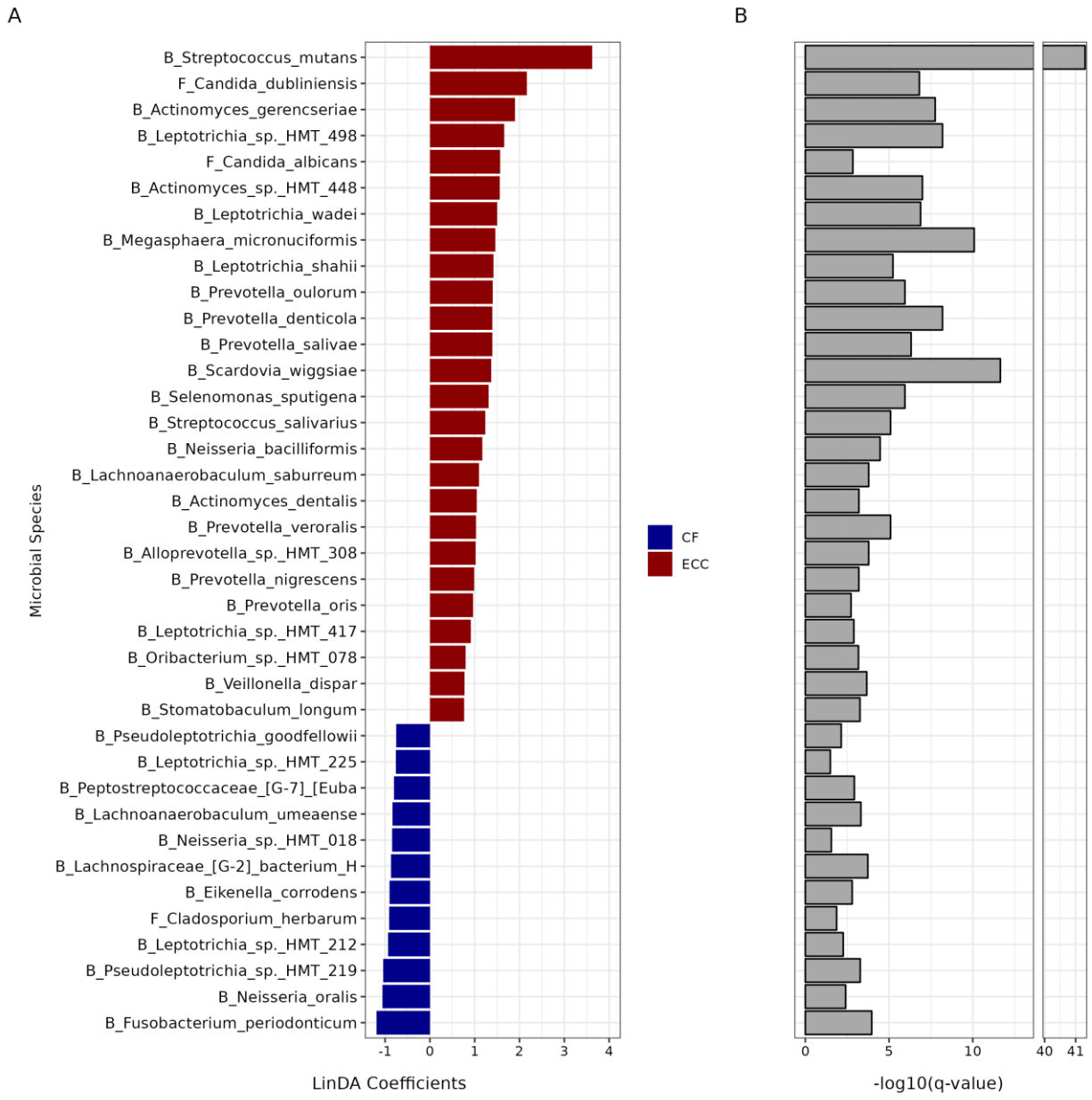


Figure S5.3: LinDA association between microbial species and socioeconomic factors. (A) The coefficient estimates values from LinDA differential abundance analysis method after adjusting for age, sex and rural/urban status. Features are categorized as B_ for bacterial and F_ for fungal. (B) Adjusted -log *p*-values (*q*-values) calculated using the Benjamini-Hochberg correction method for LinDA association.

Bridge to chapter 6

In the previous chapter, we identified the socioeconomic factors associated with ECC, along with the ECC-associated microbiome. In the next chapter, we investigate the role of taste genetic factors in ECC and the ECC microbiome. The findings from the previous chapter regarding ECC-associated sociodemographic factors guide us including them as covariates while exploring these associations. Furthermore, the machine learning methods used previously will be expanded to incorporate taste genetic factors and sociodemographic variables to investigate the relative contribution of these factors in a unified model.

Chapter 6: Integrative Analysis of Taste Genetics and the Dental Plaque Microbiome in Early Childhood Caries

Mohd Wasif Khan, Vivianne Cruz de Jesus, Betty-Anne Mittermuller, Robert J. Schroth, Pingzhao Hu, Prashen Chelikani

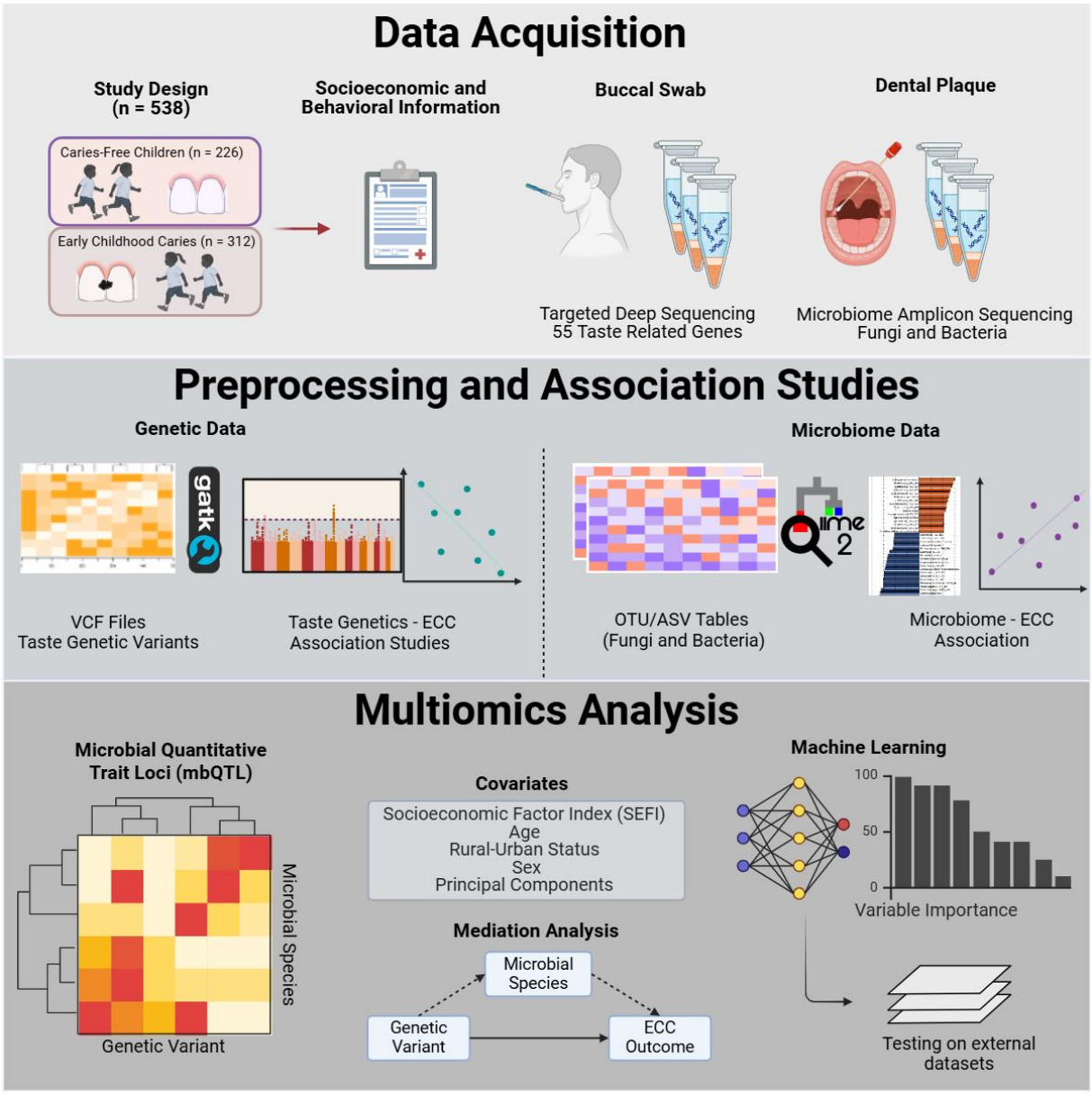
bioRxiv 2025.03.10.642456; (in revision at *Cell Reports*)

DOI: https://doi.org/10.1101/2025.03.10.642456

AUTHOR CONTRIBUTIONS

Conceptualization: M.W.K., R.J.S., P.H., P.C.; Methodology: M.W.K., V.C.J., P.H., P.C.; Investigation: M.W.K., V.C.J., B.A.M., R.J.S.; Data curation: B.A.M.; Writing—original draft preparation: M.W.K., V.C.J., P.H., P.C.; Writing—review and editing: M.W.K., V.C.J., P.H., P.C., R.J.S.; Supervision: R.J.S., P.H., P.C.; Funding acquisition: P.H., R.J.S., P.C. All authors have read and agreed to the published version of the manuscript.

Graphical Abstract



6.1 Summary

Early childhood caries (ECC) is a multifactorial disease mainly caused by the oral microbiome; however, it is also influenced by host genetics and environmental factors. The combinatorial analysis of these multiple factors influencing ECC susceptibility requires further research. This study investigated the interplay between genetic variants in taste-related genes and the microbiome in ECC, targeting taste genes because of their role in taste preference and potential interactions with oral fungi and bacteria. Using a case-control design involving 538 children, we obtained dental plaque microbiome profiles and genetic variants across 55 candidate genes through next-generation sequencing. Our association analysis for taste genetics and ECC outcome used the socioeconomic factor index (SEFI) and rural-urban status as confounders. We observed a few taste gene variants associated with ECC and microbial diversity. However, no specific association was observed between the variants and cariogenic species. Furthermore, our analysis indicated that *Streptococcus mutans* is a partial mediator between these genetic variants and ECC outcomes. Machine learning models integrating microbiome, genetics, and covariates achieved robust ECC vs. caries-free classification (AUROC = 0.96), with *Streptococcus mutans*, rural-urban status, a bitter taste receptor variant, *Candida dubliniensis*, and SEFI as the top ECC-associated factors. Our findings highlight the association between host genetics and the oral microbiome, underscoring the need for multiomics approaches in ECC risk assessment.

6.2 Introduction

Early childhood caries (ECC) refers to dental caries that affect children under 72 months of age (AAPD, 2024). It is a condition of the primary dentition. The high prevalence of ECC poses a significant socioeconomic burden on both families and society (Pierce *et al.*, 2019; Tinanoff *et al.*, 2019). Although microorganisms in the oral cavity, particularly in dental plaque, play a major role in ECC, numerous studies have highlighted additional factors, including dietary habits, oral hygiene practices, and host-related factors such as salivary composition, immune response, and tooth susceptibility (Opal *et al.*, 2015; Baker *et al.*, 2021). The etiology of ECC is complex; however, many of these contributing factors are closely linked to host genetics.

ECC is associated with dysbiosis of the oral microbiome, and diverse microbial species influence the progression of the disease. While *Streptococcus mutans* is considered a key cariogenic species, other species of genera, such as *Veillonella, Prevotella, Scardovia,* and *Candida,* consistently appear to be highly abundant in the oral microbiome (Tanner, Mathney, *et al.*, 2011; Wang *et al.*, 2017; Lee *et al.*, 2025). Furthermore, the role of socioeconomic and behavioral factors in shaping the oral microbiome has gained a lot of attention recently (Khan, de Jesus, *et al.*, 2024; Lee *et al.*, 2025).

ECC is a complex health issue with a strong role of socioeconomic, behavioral and environmental factors. Numerous studies have identified a strong association between low socioeconomic status (SES) and ethnicity-deprived neighborhoods (Willems *et al.*, 2005; Seow, 2012). Furthermore, behavioral factors such as poor feeding habits, infrequent tooth brushing, and lack of fluoride exposure significantly contribute to ECC development (Khan *et al.*, 2023; Khan, de Jesus, *et al.*, 2024).

Host genetics plays a significant role in caries, with nearly half of the prevalence attributed to genetic factors (Haworth *et al.*, 2020). Genetic predisposition has been linked to increased susceptibility to ECC, with specific genetic markers associated with this condition (Stanley *et al.*, 2014; Morrison *et al.*, 2016; Shungin *et al.*, 2019; Olatosi *et al.*, 2021). Twin studies examining host genetic influence have revealed higher similarities in the oral microbiome of monozygotic twins than in dizygotic twins, with more than 50% heritability of microbiome phenotypes (Demmitt *et al.*, 2017). Additional studies have demonstrated a connection

between host genetics and the oral microbiome (Blekhman *et al.*, 2015; Gomez *et al.*, 2017; Kolde *et al.*, 2018; Hashim *et al.*, 2021; Liu *et al.*, 2021; de Jesus *et al.*, 2022). Recent metagenome-genome association studies have identified several genomic loci associated with the oral microbiome (Liu *et al.*, 2021), suggesting an indirect role of host genetics in the etiology of ECC mediated by microbiome interactions.

Recent studies have suggested that taste perception, mediated by T1R2 and T1R3 receptors for sweet and umami tastes and T2R receptors for bitter taste, may influence the oral microbiome and potentially affect caries susceptibility (Chandrashekar *et al.*, 2000; Chen *et al.*, 2011; Holla *et al.*, 2015; Cattaneo, Gargari, *et al.*, 2019; Chisini *et al.*, 2021; de Jesus *et al.*, 2022; Shrestha *et al.*, 2024). Notably, single nucleotide variations (SNVs) in *TAS1R2* and *TAS2R38* have been linked to an increased risk of caries and the prevalence of specific bacterial genera (Kulkarni *et al.*, 2013; Cattaneo, Gargari, *et al.*, 2019; de Jesus, Singh, *et al.*, 2021).

Despite these findings, the influence of host taste genetics on ECC and the associated microbiome remains largely underexplored. To address this, we aimed to utilize next-generation sequencing (NGS) technology, which can capture novel genetic variants across diverse populations. Additionally, we sought to identify the empirical changes in microbial profiles driven by these genetic variations. Research exploring the relationship between taste genetics and the oral microbiome is limited.

This study explored the relationship between the dental plaque microbiome, taste genetics, and ECC, addressing the following key questions. First, we investigated whether variants in taste and taste-related genes (referred to as "taste genes") are associated with ECC and caries-free (CF) outcomes. Second, we examined whether these genetic variants influence the microbial composition of dental plaque, thereby mediating ECC outcomes through host genetics. Further, we compared the relative contributions of microbial and genetic factors using machine learning models. Third, we analyzed the role of the socioeconomic factor index (SEFI) and rural-urban status as confounders. Our results suggest that certain taste gene variants are associated with ECC, and a variant in the bitter taste receptor *TAS2R60* influences the abundance of multiple dental plaque microbial species. Understanding these

associations can contribute to deeper insights into disease etiology, children's susceptibility to ECC, and strategies for risk assessment.

6.3 Results

6.3.1 Social demographical variables included in the study

Initially, 16 variables were selected for this study, as outlined in our previous publication (Khan, de Jesus, *et al.*, 2024). However, based on their relevance and association with ECC, only age, sex, place of residence (rural-urban status), SEFI score, and ECC status were retained. The distributions of the selected variables are presented in (**Table 6.1**). Age, rural-urban status, and SEFI score showed significant associations with ECC. These variables, along with the first five principal components (PCs), were included as covariates in all subsequent statistical models. The first five PCs accounted for 73% of the total genetic variation in the dataset (**Figure S6.1**). A conceptual framework representing the relationships among covariates, independent variables, and outcomes of the study is presented in **Figure S6.2A**. Notably, the first PC effectively captured batch information with a coefficient of 0.81 (**Figure S6.2B**). Therefore, PC1 was used as a proxy for batch effects, and the batch variable was not included as a separate covariate.

Table 6.1: Descriptive statistics for the host variables used in the study.

Variable	N	CF, n = 226[1]	ECC, n = 312[1]	*p*-value[2]
Sex	538			>0.9
Female (0)		114 (50%)	157 (50%)	
Male (1)		112 (50%)	155 (50%)	
Age (months)	538	43 (28, 55)	49 (36, 58)	<0.001
Urban_status	534			<0.001
Rural (0)		8 (3.6%)	156 (50%)	
Urban (1)		216 (96%)	154 (50%)	
SEFI_score (Socioeconomic Factor Index score)	498	-0.06 (-0.67, 0.71)	0.87 (0.04, 1.89)	<0.001

n, Number of non-missing entries for each variable
CF, Caries-free
ECC, Early childhood caries
1n (%); Median (interquartile range)
2Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test

6.3.2 Gene and genetic variant association testing with ECC

The sequencing data achieved a minimum depth of 30× for 95.4% of the reads across the 55 selected genes, covering an average of 99.4% of their exonic regions. Mapping these reads to the reference sequence for the selected genes using the standard GATK pipeline (McKenna *et al.*, 2010), combined with initial quality control using VCFtools, identified 6,843 variants across 538 individuals. Subsequent filtering in PLINK removed 969 variants for --geno, 229 variants for --hwe, and 4,966 variants using --maf filters, retaining 679 variants (Purcell *et al.*, 2007), including single nucleotide variants (SNVs) and insertions/deletions (InDels), which were further tested for their association with ECC status.

The association of variants with ECC status was tested for different combinations of covariates (**Figure S6.3**, **Figure S6.4**, and **Figure S6.5**). These results show that while rural-urban status and SEFI score improve the model performance by decreasing the genomic inflation and Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, the inclusion of the batch effect does not. Based on this optimization, the final model included the following covariates: age, sex, urban-rural status, SEFI score, and first five PCs. Using a *q*-value (Benjamini-Hochberg adjusted *p*-values) threshold of 0.05, three variants were significantly associated with ECC in the dominant model (**Figure 6.1A** and **Figure 6.1B**,

Table 6.2): rs111819661 in *SCNN1D*, rs35195910 in *TAS2R60*, and rs2305645 in *PLCB2*. These variants also displayed significant deviations from the expected *p*-values in the quantile-quantile (QQ) plot (**Figure 6.1A**). In the additive model, only one variant, rs111819661 in *SCNN1D*, remained significant, and no association was identified in the recessive model (**Table 6.2** and **Figure S6.6**). The genomic inflation factors for both the dominant (λ = 1.05) and additive (λ = 1.03) models were close to 1, indicating no inflation, thereby minimizing the risk of false-positive results. The identified variants were located in *TAS2R60* (in-frame deletion) and *SCNN1D* (one intronic and one missense variant), as shown in the volcano plot (**Figure 6.1B**). The intron variant in the *PLCB2* gene is located only 8 nucleotides away from the nearest exon, suggesting that it may play a role in splice site binding or be involved in gene regulatory mechanisms. Furthermore, variants in *SCNN1D* and PLCB2 exhibited mild linkage disequilibrium ($0.2 < r^2 < 0.5$) with other variants in their respective genes **(Figure S6.7**).

Sequence kernel association test (SKAT) analysis (Wu *et al.*, 2011; S. Lee *et al.*, 2012), which assessed gene-level associations by considering the cumulative effects of both common (minor allele frequency [MAF] >= 0.01) and rare (MAF < 0.01) variants, identified *ADGRG1* and *TAS2R60* as strongly associated with ECC outcomes at $q < 0.05$ (**Figure 6.1C**). The other significant genes were *RAC1*, *PLCB2*, *SCNN1D*, and *TAS2R19*. Although no significant genes were identified for rare variants after *p*-value adjustment, *ADGRG6* and *SCNN1B* showed suggestive associations before adjustment (**Figure 6.1D**).

Table 6.2: Significant loci identified in 55 taste-related genes to be associated with ECC status.

Location	Gene	SNP_rsID	A1	A2	Frq_ECC	Frq_CF	OR (95% CI)	*p*	*p*_BH	Variant_Type
Dominant model										
chr1: 1287578	SCNN1D	rs111819661	T	C	0.175	0.020	6.165 (2.52–15.12)	6.91 e-05	4.79 e-02	missense variant
chr7: 143443837	TAS2R60	rs35195910	G	GTCT	0.397	0.179	2.749 (1.61–4.68)	1.83 e-04	4.87 e-02	conservative inframe deletion
chr15: 40303364	PLCB2	rs2305645	T	C	0.151	0.295	0.371 (0.22–0.63)	2.11 e-04	4.87 e-02	intron variant
Additive model										
chr1: 1287578	SCNN1D	rs111819661	T	C	0.175	0.020	6.108 (2.53–14.76)	6.11 e-05	4.24 e-02	missense variant

Location, Chromosomal location of the variant in the GRCh38 genome assembly.
SNP_rsID, Reference SNP ID number, if available.
A1, Allele corresponding to measured effect on the outcome.
A2, Reference allele, not corresponding to measured effect on the outcome.
Frq_ECC/Frq_CF, Allele frequencies in ECC and caries-free (CF) groups.
OR, Odds ratio, indicating the effect size.CI, confidence interval for odds ratio
p, *p*-value for the association.
*p*_BH, Benjamini-Hochberg-corrected *p*-value for multiple testing.
Variant_Type, Functional annotation of the variant

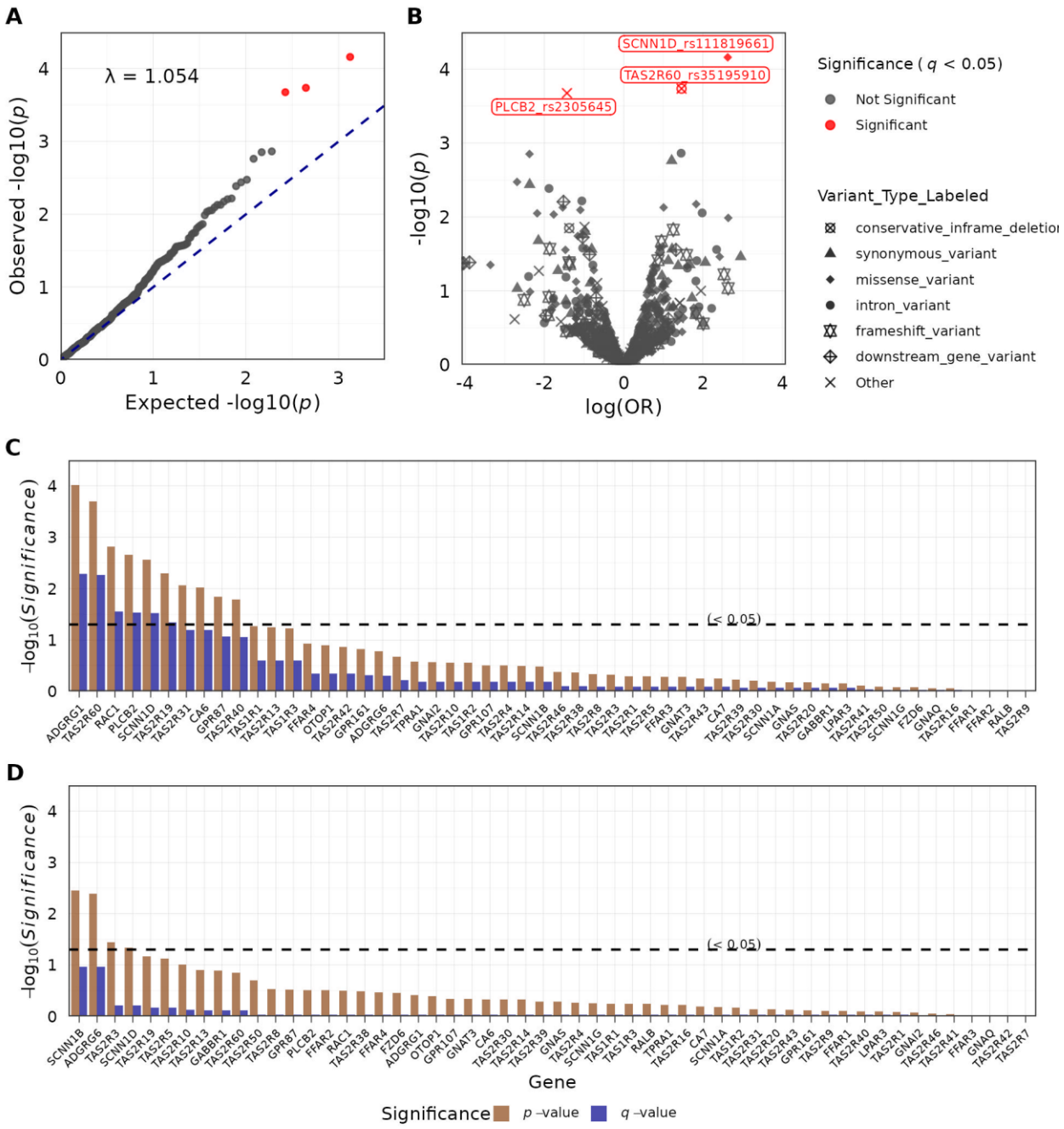


Figure 6.1: Taste genetic variants and genes association with ECC status.
(A) Dominant model. The Q-Q plot illustrates the distribution of observed -log10(*p*-value) values against the expected -log10(*p*-value) values under the null hypothesis of no association between the variables. Each point represents a single nucleotide polymorphism (SNP). The blue dashed line indicates the null expectation, where the observed equals the expected.
(B) Volcano plot for the dominant model. The x-axis represents the log2 odds ratio (OR), and the y-axis shows the -log10(*p*-value). Each point represents a variant, with the shape indicating the variant type. Red points denote significant associations for $q < 0.05$ (BH-adjusted *p*-value < 0.05). The Q-Q and volcano plots are from logistic regression models in PLINK under a dominant model for 55 taste-associated candidate genes, with age, sex, rural-urban status, SEFI score, and PCs 1–5 as covariates.

(C) Gene-wise association for rare (minimum allele frequency, MAF < 0.01) and common (MAF > 0.01) variants with equal weights in the SKAT method in R. The bar plot shows *p*-values and *q*-values for the association between variants within each gene after removing singletons. The analysis was performed in the SKAT R-package using the SKAT-O option for optimal association, adjusting for age, sex, rural-urban status, SEFI score, and the top five principal components. The horizontal line marks the significance cutoff of 0.05.
(D) Same as (C), but only for rare variants (MAF < 0.01).

6.3.3 ECC and microbiome

Microbiome associations for both bacteria and fungi with ECC were performed using Microbiome Multivariable Associations with Linear Models version 2(MaAsLin2) with a *q*-value cutoff of 0.01 (Mallick *et al.*, 2021). After adjusting for age, sex, rural-urban status, and SEFI score, 39 bacterial species and two fungal species differed significantly between the CF and ECC groups (**Figure 6.2**). At the genus level, ten bacterial genera and one fungal genus were identified as being significantly associated with ECC outcomes (**Figure 6.2**).

The most prominent bacterial associations included *S. mutans* and two *Actinomyces* species, *Actinomyces gerencseriae* and an unclassified species, *Actinomyces* sp. HMT-448, which were strongly linked to the ECC condition. In contrast, *Fusobacterium periodonticum* and *Neisseria oralis* were significantly enriched in the CF samples. Among fungi, the genus *Candida* was the most abundant in ECC, although no specific fungal species were associated with CF conditions. Additionally, the bacterial genera *Scardovia* and *Fusobacterium* showed significantly increased abundance in ECC and CF groups, respectively.

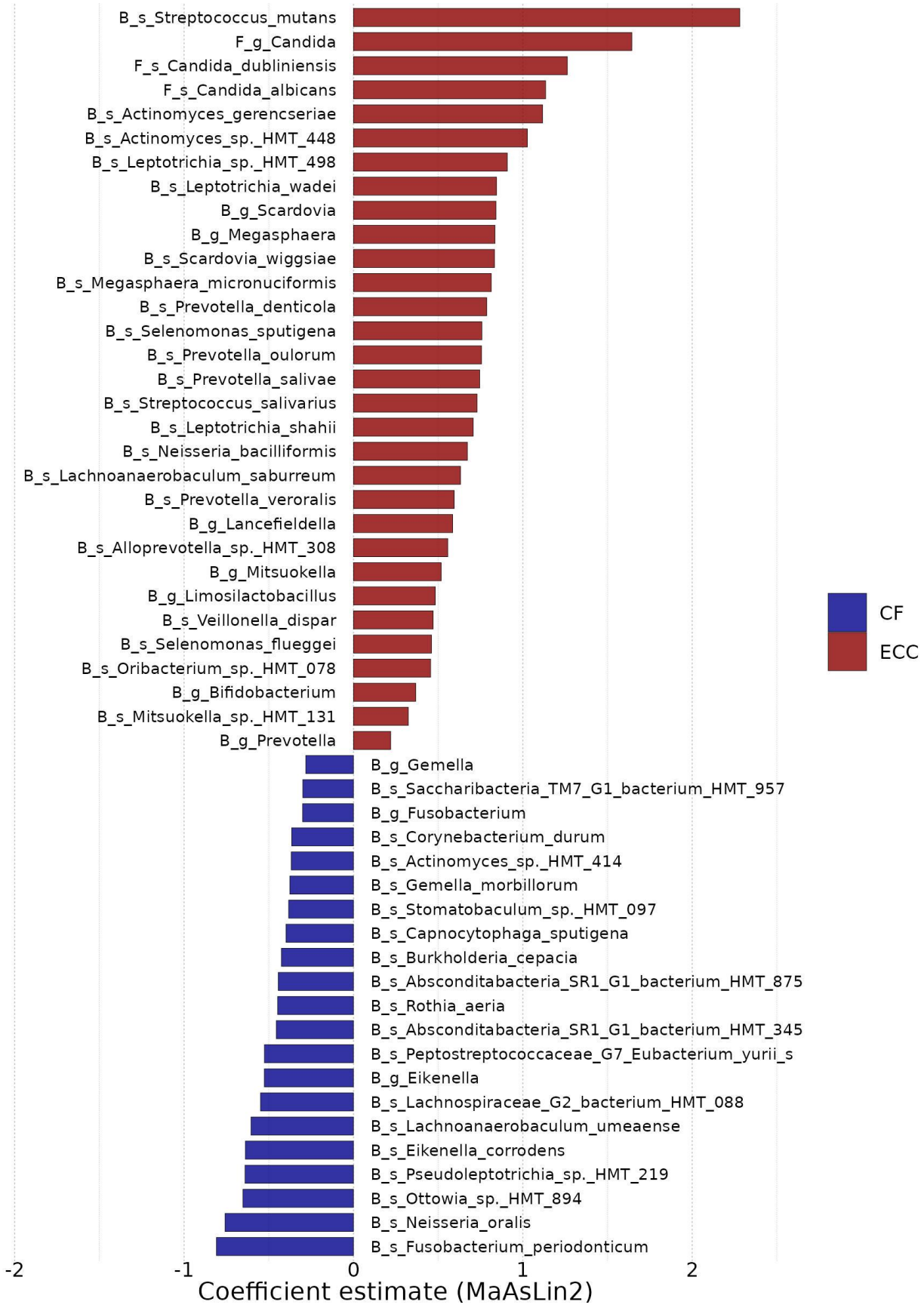

B_s_Streptococcus_mutans
F_g_Candida
F_s_Candida_dubliniensis
F_s_Candida_albicans
B_s_Actinomyces_gerencseriae
B_s_Actinomyces_sp._HMT_448
B_s_Leptotrichia_sp._HMT_498
B_s_Leptotrichia_wadei
B_g_Scardovia
B_g_Megasphaera
B_s_Scardovia_wiggsiae
B_s_Megasphaera_micronuciformis
B_s_Prevotella_denticola
B_s_Selenomonas_sputigena
B_s_Prevotella_oulorum
B_s_Prevotella_salivae
B_s_Streptococcus_salivarius
B_s_Leptotrichia_shahii
B_s_Neisseria_bacilliformis
B_s_Lachnoanaerobaculum_saburreum
B_s_Prevotella_veroralis
B_g_Lancefieldella
B_s_Alloprevotella_sp._HMT_308
B_g_Mitsuokella
B_g_Limosilactobacillus
B_s_Veillonella_dispar
B_s_Selenomonas_flueggei
B_s_Oribacterium_sp._HMT_078
B_g_Bifidobacterium
B_s_Mitsuokella_sp._HMT_131
B_g_Prevotella
B_g_Gemella
B_s_Saccharibacteria_TM7_G1_bacterium_HMT_957
B_g_Fusobacterium
B_s_Corynebacterium_durum
B_s_Actinomyces_sp._HMT_414
B_s_Gemella_morbillorum
B_s_Stomatobaculum_sp._HMT_097
B_s_Capnocytophaga_sputigena
B_s_Burkholderia_cepacia
B_s_Absconditabacteria_SR1_G1_bacterium_HMT_875
B_s_Rothia_aeria
B_s_Absconditabacteria_SR1_G1_bacterium_HMT_345
B_s_Peptostreptococcaceae_G7_Eubacterium_yurii_s
B_g_Eikenella
B_s_Lachnospiraceae_G2_bacterium_HMT_088
B_s_Lachnoanaerobaculum_umeaense
B_s_Eikenella_corrodens
B_s_Pseudoleptotrichia_sp._HMT_219
B_s_Ottowia_sp._HMT_894
B_s_Neisseria_oralis
B_s_Fusobacterium_periodonticum
CF
ECC
-2
-1
0
1
2
Coefficient estimate (MaAsLin2)

Figure 6.2: Association between dental plaque microbial taxa and ECC outcome.
The coefficient estimates for significantly different taxa at *q*-values < 0.01 between the ECC and CF groups using the MaAsLin2 differential abundance analysis method after adjusting for age, sex, rural-urban status, and SEFI score. Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level and 'g' indicates genus level.

6.3.4 Microbiome quantitative trait loci

Regarding the association between ECC-associated variants and dental plaque microbiome diversity, the three identified variants did not exhibit significant effects on bacterial alpha diversity (**Figure 6.3A**). However, variants in *SCNN1D* and *TAS2R60* were negatively correlated with the alpha diversity of fungal species, indicating an overabundance of certain fungal species in the presence of these variants. Additionally, these two variants were significantly associated with beta diversity in both bacterial and fungal datasets, whereas the PLCB2 variant was only associated with beta diversity in fungal samples (**Figure 6.3B**).

Microbiome quantitative trait loci (mbQTL) analysis was performed using linear associations with a dominant model for significant variants. At the genus level, no bacterial or fungal genera were found to be associated with any of the ECC-related variants (**Figure 6.4A**). The *TAS2R60*-rs35195910 variant was significantly associated with the bacterial species *Streptococcus intermedius*, *Parascardovia denticollens*, *Aggregatibacter* sp. HMT-898, and the fungal species *Protrudomyces lateralis* (**Figure 6.4B**). Meanwhile, the variant in *SCNN1D* was associated with one bacterial species, *Capnocytophaga* sp. HMT-878. However, none of these species were identified as being differentially abundant in ECC and CF.

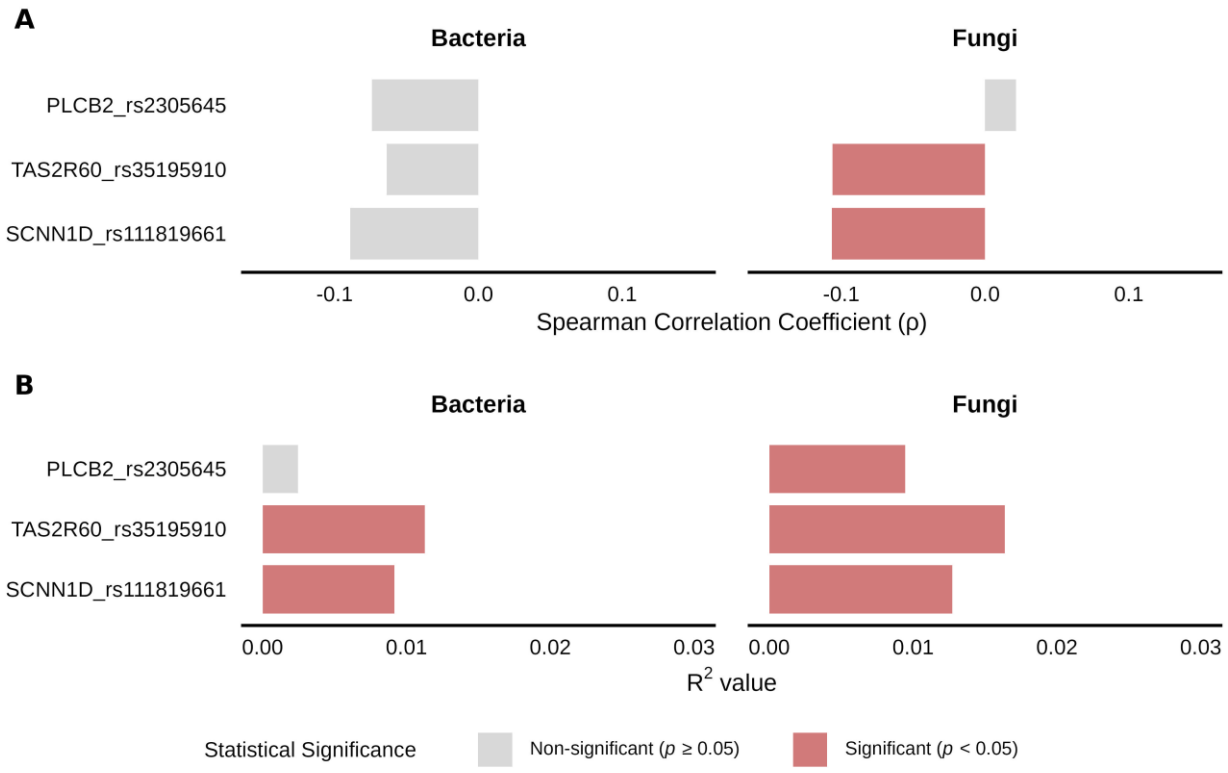


Figure 6.3: Spearman correlation of the ECC associated variants with alpha and beta diversity.

(A) Significant genetic variants and their Spearman correlation coefficient (ρ) values for alpha diversity assessed using Shannon's metric for bacteria and fungi.

(B) Beta diversity analyses with significant genetic variants via PERMANOVA. The R^2 values indicate the variance explained by the bacterial and fungal communities.

The colored bar in each plot represents the statistical significance of the Spearman correlation at $p<0.05$.

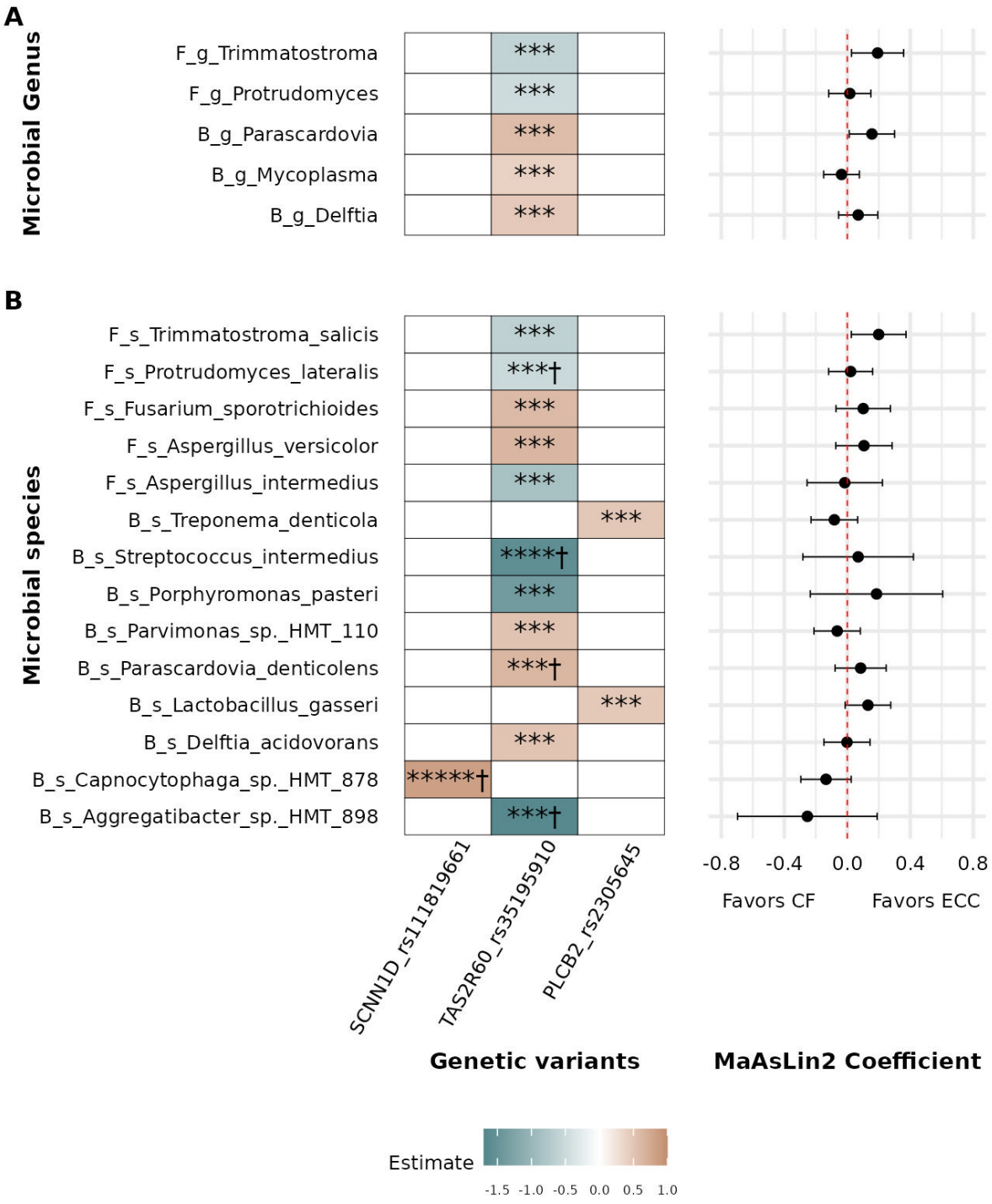


Figure 6.4: Association between ECC-associated variants and microbial taxa.
The heatmap illustrates the association between ECC-associated genetic variants and microbial taxa when adjusted for covariates such as age, sex, urban-rural status, SEFI score, and the top five principal components from the genetic data. The forest plot shows the coefficient of association from the MaAsLin2 method between the selected microbial species in the heatmap and ECC outcome.

Positive coefficients favor ECC, whereas negative coefficients favor CF status. The analysis was performed for significant variants identified in the dominant genetic model and all microbial species with CLR normalization using a linear model. The heatmap value signifies the estimates values, with red showing a positive association between variants with alternate alleles and green representing a negative association. The *p*-values are (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ ***** $p<0.00001$), and "†" indicates a significant association at BH adjusted $p<0.05$. Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level and 'g' indicates genus level.
(A) Genus level (B) Species level.

6.3.5 Mediation analysis

The linear decomposition model (LDM) was used to test the potential mediation of multiple microbial species in the association between ECC outcomes and specific genetic variants (Yue and Hu, 2022). Three ECC-associated variants were examined to identify the mediating species from the differentially abundant species (**Figure 6.5A**). These results suggest that the association between the *TAS2R60* variant and ECC status may be mediated by four bacterial species: *S. mutans*, *Peptostreptococcaceae* subsp. *yurii margaretiae*, *Fusobacterium periodonticum*, and *Eikenella corrodens*. Additionally, the ECC outcome associated with the *SCNN1D* variant was mediated by *S. mutans*. Further analysis of the single variant with single-taxon mediation revealed only two significant mediators after adjusting for *p*-values: *S. mutans* and *Fusobacterium periodonticum*, both of which mediated the association between the *TAS2R60* variant and ECC (**Figure 6.5B**). However, the average mediation effect (ACME) of these two species was lower than the average total effect (ADE) of the variants on ECC outcomes, suggesting a partial mediation (**Figure 6.5C**).

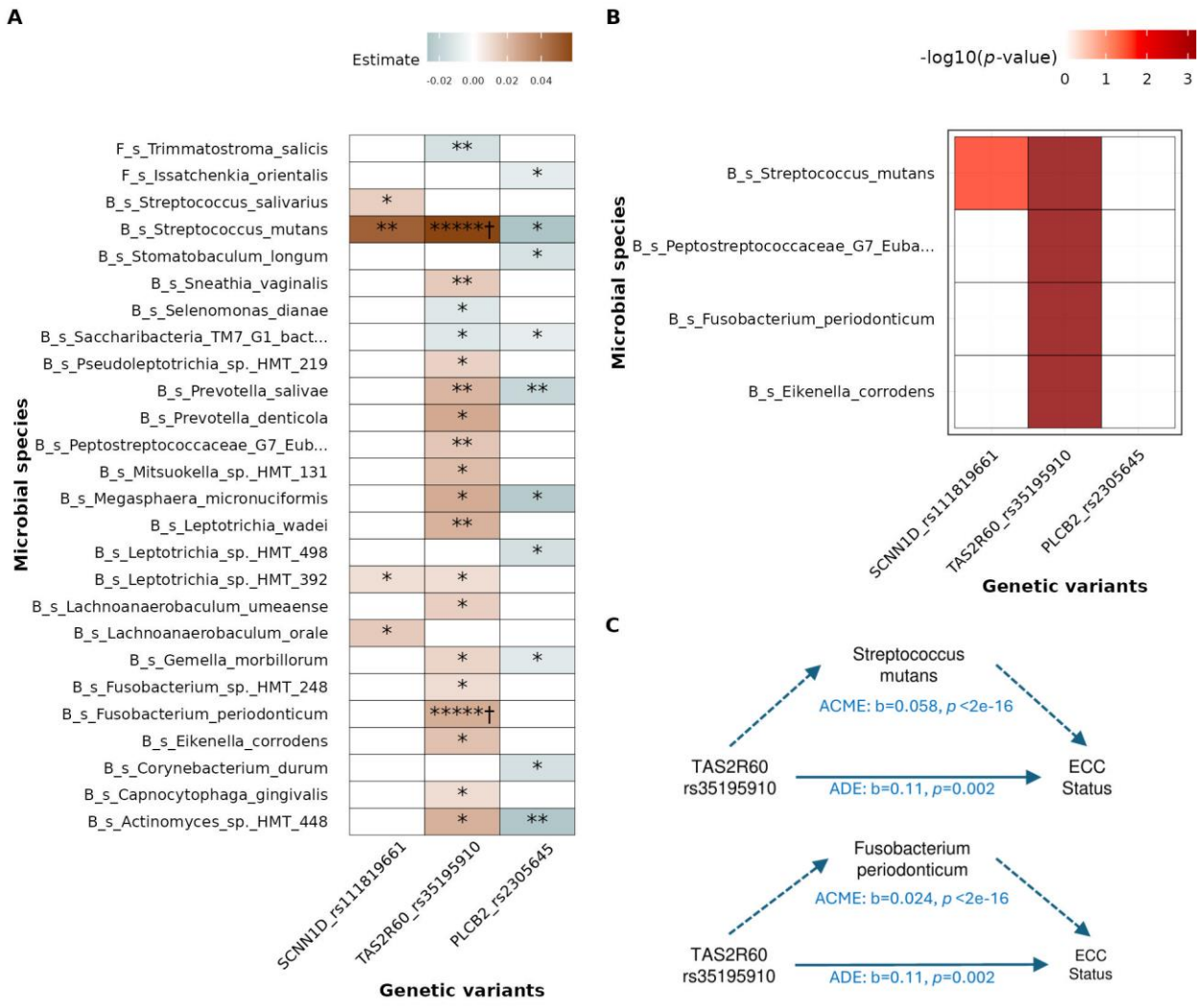


Figure 6.5: Mediation analysis for genetic variants and ECC association with microbial species as mediator

(A) Mediation analysis in the R package "mediation" for significant variants with each microbial species. The heatmap value signifies the estimate values, with red showing positive mediation between variants with alternate alleles and green representing a negative association. The *p*-values are (* $p<0.05$, ** $p<0.01$, ***** $p<0.00001$), and "†" indicates a significant association at BH adjusted $p<0.05$.

(B) Mediation was performed using LDM with one SNP and multiple mediator approach. This heatmap illustrates the significant mediation of microbial species between genetic variants and ECC outcomes. The x-axis shows the genetic variants, and the y-axis lists the microbial species. The color scale represents the negative log10 of *p*-values.

(C) Average mediation effect (ACME) and average direct effect (ADE) for significant mediators at BH adjusted $p<0.05$ in (A).

Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level and 'g' indicates genus level.

6.3.6 ECC prediction models

To assess the predictability of ECC outcomes, four machine learning methods were trained and tested on each dataset (microbiome, genetic variants, and covariates) individually and in combination. Among the methods, the random forest model demonstrated comparable or superior performance compared to the other models (**Figure 6.6A**). Using random forest, the prediction of ECC outcomes with microbiome data achieved the best classification performance, with the AUROC and AUPRC values both reaching nearly 0.94. For the genetic data, random forest provided AUROC and AUPRC values of 0.85 and 0.89, respectively. The covariate-based model, which included age, sex, rural-urban status, and SEFI score, yielded an AUROC of 0.8.

When features from the microbiome, genetic variants, and covariates for the prediction of ECC status were combined, the model performance for the prediction of ECC status for the prediction of ECC status improved to 0.96 for both AUROC and AUPRC. Feature importance analysis in the random forest model identified the top predictive features from all three data sets. The most influential factor was the bacterial species *S. mutans*, followed by urban-rural status and the variant *TAS2R60*-rs35195910 (**Figure 6.6B**). The fungal species *Candida dubliniensis* ranked among the top five variables in the combined dataset and was the second most important feature in the microbiome data after *S. mutans*. Among the other covariates, age ranked in the top variables, whereas sex was not identified as a significant feature.

The microbiome random forest model was further tested using external microbiome datasets with sPLSDA batch correction (**Figure 6.6C**) (Teng *et al.*, 2015; Agnello *et al.*, 2017; Gomez *et al.*, 2017; Kalpana *et al.*, 2020). The AUROC values for the test performance were 0.78, 0.82, and 0.71 for the Agnello-2017, Kalpana-2020, and Teng-2015 datasets, respectively. However, the Gomez-2017 dataset exhibited near-random performance with an AUROC of 0.47. For AUPRC, the performances of the Gomez-2017 and Teng-2015 datasets were lower than their respective AUROC values. Using sPLSDA with batch-corrected data, the performance slightly improved for the Kalpana-2020 dataset compared to the centered log-ratio (CLR)-normalized values (**Figure 6.6C and S6.11A**). However, a decrease in AUPRC performance was observed for the Teng-2015 dataset, which decreased from 0.54 to 0.48.

The predictive performance of the microbial risk score (MRS) was evaluated using batch-corrected datasets. On the holdout data from this study, the MRS demonstrated a strong performance, achieving AUROC and AUPRC values close to 0.9 (**Figure 6.6D**). The performance remained high on the Kalpana-2020 dataset, with an AUROC exceeding 0.8, and was moderate on the Agnello-2017 dataset, with an AUROC of approximately 0.75. In contrast, the performance was lower on the remaining datasets, with AUROC values below 0.68 and 0.58 for Teng-2015 and Gomez-2017, respectively. MRS was also tested on non-batch-corrected datasets (**Figure S6.11B**). While the AUROC values remained largely unchanged compared to the non-batch-corrected data, the AUPRC values showed slight improvements (**Figure 6.6D**). This was particularly evident in the two datasets with lower AUROC values: Teng-2015 and Gomez-2017 datasets.

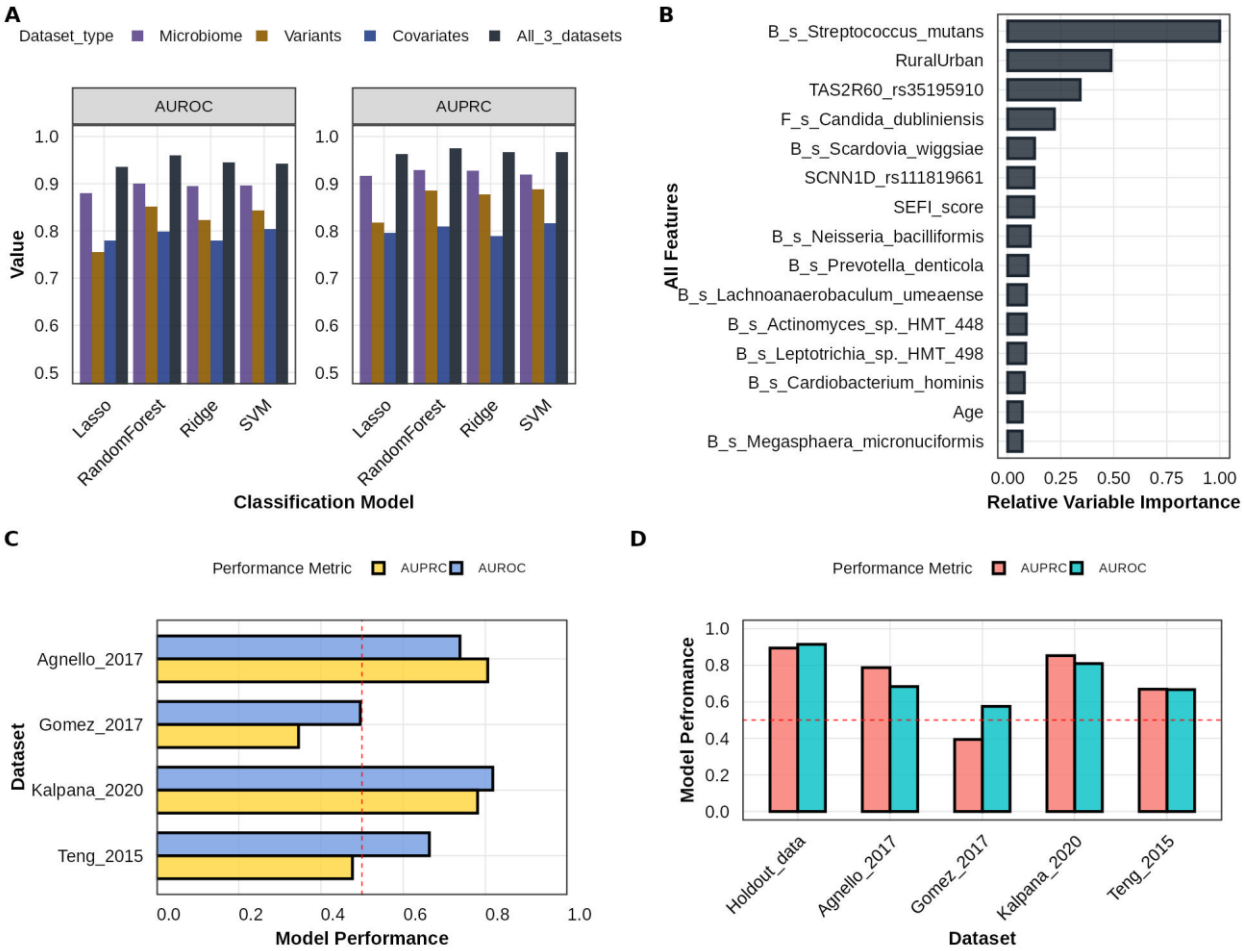


Figure 6.6: Machine learning based classification for ECC status using microbiome, genetic variants and covariate data
(A) Comparison of machine learning models based on individual and combined datasets for microbiome (fungi and bacteria), genetic variants, and covariate data using AUROC and AUPRC metrics. The machine learning models selected for analysis were lasso, random forest, ridge, and SVM.
(B) The random forest model with the combined dataset was chosen to plot the variable importance bar plot. The bars indicate the relative importance of each feature when all features are used to classify ECC_status.
(C) Random forest model performance on sPLSDA batch-corrected external datasets. Four external datasets were chosen to assess the random forest model based on only microbiome data, and the model performance was assessed using AUROC and AUPRC values.
(D) ECC prediction model performance on sPLSDA batch-corrected external datasets based on the MRS score on CLR-normalized datasets.
Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level and 'g' indicates genus level.

6.4 Discussion

This study examined the relationship between taste genetics and ECC, as well as taste genetics and the oral microbiome in ECC in the context of sociodemographic factors using an integrative analysis. We focused on taste-associated genes because of their impact on taste preferences, which influence dietary habits and subsequently shape the oral microbiome. Previous research has suggested potential interactions between receptors encoded by these genes and genetic variants that could modify these interactions (Medapati, Singh, *et al.*, 2021). This study employed robust statistical analyses, incorporating age, sex, SEFI score, and rural-urban status as covariates, to refine the identification of potential genetic variants and their association with the dental plaque microbiome. Although sex was not found to be associated with ECC status in Chapter 5, it was included as a covariate based on previous findings on a subset of this data suggesting some sex-based differences (de Jesus *et al.*, 2020). Our findings identified genetic variants in the bitter taste receptor gene *TAS2R60* (rs35195910), sodium ion transport gene *SCNN1D* (rs111819661), and phospholipase gene *PLCB2* (rs2305645) that were associated with ECC. These results align with findings from large population studies that highlight the role of host taste genetics in ECC (Orlova *et al.*, 2022; Shrestha *et al.*, 2024). Furthermore, these variants appear to be linked to changes in the microbial species within the dental plaque microbiome of the participants.

Focusing on specific genes allows us to sequence them using deep sequencing methods, which is more logistically feasible than genome-wide analysis. Deep sequencing facilitates the identification of novel variants, particularly frameshifts caused by insertions and deletions. Some novel variants, such as those in *TAS2R19* and *TAS2R42*, appeared significant in the base model adjusted for age and sex but lost significance when adjusted for additional covariates. Notably, the single nucleotide variation in *SCNN1D*-rs111819661 demonstrated both additive and dominant roles (**Table 6.2**). This dual mode of association has also been reported in genetic studies on rheumatoid arthritis and cognitive function in cancer patients treated with opioids (Jiang *et al.*, 2012; Kurita *et al.*, 2016). However, larger sample sizes are required to empirically determine whether this model reflects an additive or a dominant effect.

Additionally, we analyzed the cumulative association of both rare and common variants and found that variants in *ADGRG1* and *TAS2R60* were associated with ECC. The association with ECC may be attributed to the highly ECC-associated variant rs35195910 in *TAS2R60*, whereas the significance of *ADGRG1* highlights the potential polygenic nature of this relationship. However, none of the genes showed a significant association with ECC when only rare variant analyses were performed (**Figure 6.1D**). This finding suggests that rare variants, even in combination, may not influence ECC outcomes and that the association with *ADGRG1* is likely driven by multiple common alleles acting in combination.

The role of host genetics in ECC has yielded varying results in previous studies. While some studies suggest a potential role for host genetic composition, others have found minimal or no influence, emphasizing the dominant roles of environmental factors and time in shaping the microbial composition (Gomez *et al.*, 2017; Freire *et al.*, 2020; Mukherjee *et al.*, 2021). A recent meta-analysis of genome-wide association studies (GWAS) and transcriptome-wide association studies (TWAS) for ECC did not identify associations with variants in taste-associated genes. However, their TWAS analysis highlighted the association between the expression of several bitter taste receptors and ECC (Orlova *et al.*, 2022). Additionally, the pathway related to taste receptor activity was found to be associated with ECC, and the

GWAS association of the variant rs58016156 in *KCNU1*, which is involved in sweet taste signaling, further underscores the role of taste genetics (Shrestha *et al.*, 2024).

According to the Genome Aggregation Database (gnomAD) (Karczewski *et al.*, 2020), the frequencies of the significant variant *PLCB2*-rs2305645 in our samples were within the range of the average allele frequencies observed in the general population. In contrast, the variant *SCNN1D*-rs111819661 and *TAS2R60*-rs35195910 demonstrated a higher prevalence in ECC samples, exceeding the population average (**Figure S6.8**). While the highest frequency in any population for *SCNN1D*-rs111819661 is 0.022, it was 0.175 in the ECC samples from this study. Similarly, for *TAS2R60*-rs35195910, these frequencies were 0.098 and 0.397, respectively. This suggests that the population under consideration may be at a higher risk of taste genetics–associated ECC outcomes.

We did not identify any ECC-associated variants that showed a significant relationship with ECC-causing microbial species in the linear association analysis, including ECC status as one of the covariates. Although some dental plaque species were significantly associated with ECC-related variants, none of the taxa at the genus level exhibited significant associations. This suggests that interactions between receptors and microbes may occur at the species level, rather than at the genus level. Both mbQTL and mediation analyses indicated that *TAS2R60* influences a greater number of species than the other two variants (**Figure 6.4A**, **Figure 6.5B**). *TAS2R60* is an orphan taste receptor with no known agonists or antagonists. The high prevalence of this variant and its association with ECC and ECC-related microbiota highlight the importance of future research on the role of this receptor in host physiology and microbial interactions (Medapati, Singh, *et al.*, 2021; Xi, Zheng and Tizzano, 2022; Kouakou and Lee, 2023). In the stratified analysis of the ECC and CF samples, only one association, between *TAS2R60*-rs35195910 and *Streptococcus intermedius*, was common in both groups (**Figure S6.9**). Furthermore, the mediation observed between *S. mutans* and the bitter taste receptors, T2Rs, may be through competence-stimulating peptides secreted by *S. mutans* (Medapati, Singh, *et al.*, 2021). As previously identified, competence stimulating peptide-1 interaction with the bitter taste receptor induces an immunogenic response, such

as the secretion of CXCL-8/IL-8, TNF-α, and IL-6, and the inhibition of other oral microbes (Medapati, Bhagirath, *et al.*, 2021; Medapati, Singh, *et al.*, 2021).

The top ECC-associated species belong to genera such as *Streptococcus*, *Actinomyces*, and *Candida*, which have been previously linked to biofilm formation (Rickard *et al.*, 2003; Karkowska-Kuleta *et al.*, 2022). A recent study demonstrated associations between *Staphylococcus epidermidis* and the host genetic variation rs3828054 in the *TUFT1* gene, as well as between Candida species and the enamel gene *ENAM* variant rs3796704, suggesting a connection between host genes and microorganisms (Vieira and Modesto, 2022). In our study, despite the strong association between *Candida* species and ECC, no taste gene variants were found to be associated with *Candida* species. However, the significant negative correlation between both *TAS2R60*-rs35195910 and *SCNN1D*-rs111819661 with fungal diversity suggests the enrichment of certain species, potentially *C. dubliniensis* and *C. albicans*. Furthermore, although not statistically significant, we observed a negative association between the CF-associated species *Fusobacterium periodonticum* and a novel variant in *TAS1R1* at position 6574730 (**Figure S6.10**). This finding suggests that the presence of this variant reduces the abundance of ECC-protective species in the dental plaque microbiome.

Our results align with those of Gomez et al., where some microorganisms were associated with host genetics, but cariogenic species were not (Gomez *et al.*, 2017). Additionally, our previous research demonstrated that many of these taxa are influenced by socioeconomic and environmental factors, such as the SEFI score, rural-urban status, and bedtime snacking habits (Khan, de Jesus, *et al.*, 2024). Nonetheless, certain ECC-associated species may partially mediate the relationship between host genetics and disease outcomes.

Machine learning models enable the integration of different data types, such as microbiome, genetic, and covariate data, into a single model. Additionally, models such as random forest can capture the nonlinear patterns and variable interactions in the data. By investigating these models, we can assess the relative importance of each variable while controlling for others. The ranking of *S. mutans* as the top feature underscores its critical role in ECC etiology (**Figure 6.6B**). The appearance of covariates, specifically the SEFI score among the top-ranked

variables, highlights their association and the need to adjust for these factors when studying ECC. Although previous studies have highlighted the role of SES, our study compares these factors with microbial and taste genetic factors, providing the relative importance of these variables through an integrative model.

Our random forest model demonstrated that genetic variants alone can predict ECC outcomes, with an AUROC of 0.85 (**Figure 6.6A**). Although this is lower than the predictive performance achieved using microbiome data, the appearance of genetic variants, such as *TAS2R60* and *SCNN1D*, among the top features suggests that combining genetic and microbiome data can improve microbiome-based risk prediction. This finding also indicates that certain genetic markers may play a more significant role than some ECC-associated microbial species in shaping the etiology of ECC, further supporting the influence of taste genetics on caries progression. This is further evident from a Polish study that showed an AUC of 0.97 for ECC prediction using candidate genes (Zaorska *et al.*, 2021). Our results add to the potential candidate genetic markers that may further improve ECC prediction.

Moreover, MRS, which is based on a linear combination of ECC-associated microbial markers, achieved comparable performance on both the internal and external datasets. These results validate the robustness of microbial markers in ECC risk assessment. A combined ECC risk score, incorporating both microbial and genetic markers, could potentially improve predictive accuracy in external datasets where both genetic and microbiome data are available. Furthermore, as suggested by Blostein et al., MRS can be used as one of the covariate to identify the polygenic burden of ECC-associated variants (Blostein *et al.*, 2023). However, given the observed differences in the microbiome and genetic associations with ECC, the generalizability of these methods must be evaluated across diverse populations.

Future multiomics approaches, such as combining metatranscriptomics and metabolomics for both host and microbial communities, could provide deeper insights by capturing differences in gene expression and highlighting the effects of genetic variants on host gene expression and microbial abundance (Divaris *et al.*, 2019). These interactions may also extend to adult caries, as some genetic markers are consistent between ECC and caries experience in adulthood (Shrestha *et al.*, 2024). Furthermore, integrating gene-environment

studies, such as those on host-specific dietary influences, may further elucidate the environmental factors contributing to ECC. Collecting and including severity scores for ECC and incorporating standardized subtyping based on calibrated dentists' scoring systems is also suggested (Gormley *et al.*, 2023; Simancas-Pallares *et al.*, 2023). Moreover, recent advancements in deep learning methods for receptor-ligand interactions could facilitate the development of algorithms to study microbial ligands and their interactions with taste receptors. These interactions can be validated through laboratory testing, offering potential insights into the additional mechanisms underlying ECC.

While the role of the microbiome has been extensively studied in previous research, our study extends the knowledge of the genetic basis of ECC by focusing on taste genes along with socioeconomic variables. Genetic factors may influence ECC through host protein interactions with specific microbial species, potentially leading to microbial dysbiosis. Although ECC outcomes are primarily determined by the oral microbiome, certain genetic variants rank highly as biomarkers in our multiomics models. Our findings provide a rationale for integrated studies using both microbiome and host genetics, while considering socioeconomic conditions as covariates to study ECC etiology. These results can help identify susceptible individuals, enabling personalized therapeutic approaches and informing dental public health policymakers and oral healthcare providers to implement targeted early interventions at both societal and familial levels.

Limitations of the study

This study may have some potential limitations. As the study design is cross-sectional and not longitudinal, it will be difficult to determine whether the alteration of the microbiome is the cause or consequence of the disease. The external validation of our machine learning models showed moderate performance, suggesting that microbial profiles may not generalize well across different geographic populations owing to variability in microbial composition. Moreover, the limited association between the identified variants and ECC-associated microbes does not eliminate the possibility of interactions or the influence of these variants on such associations. Since our analyses relied on linear associations between variants and microbial abundance, which may not fully reflect biological complexity, further research using

host transcriptomics and microbial metatranscriptomics is required (Chetty and Blekhman, 2024).

6.5 Methods

6.5.1 Experimental model and study participant details

This study included 538 participants under the age of 72 months, recruited between December 2017 and July 2022 from the Children's Hospital Research Institute of Manitoba (CHRIM), Misericordia Health Center (MHC), and various community clinics across Winnipeg, Manitoba. The eligibility criteria required participants to be less than 72 months of age and not currently on antibiotics. In this cross-sectional study, samples from children with ECC were considered cases, whereas those from children with CF were considered controls. ECC was diagnosed by experienced dentists following the American Academy of Pediatric Dentistry (AAPD) definition, which identifies ECC as the presence of one or more decayed, missing (due to caries), or filled tooth surfaces in any primary tooth (AAPD, 2024). A questionnaire was completed by parents or caregivers gathering additional information, including the child's age, sex, place of residence (urban or rural), feeding habits, and mode of birth. The socioeconomic factor index (SEFI), an area-based socioeconomic measure, was derived from the Manitoba Centre for Health Policy (MCHP) data based on participants' postal codes (Metge *et al.*, 2009). SEFI is derived from Canadian Census data and is a combined score of four areas: unemployment rate, average household income, proportion of single-parent households, and proportion of the population without a high school graduation. When interpreting the SEFI, a lower score indicates a more favorable SES, while a score greater than 0 reflects less ideal socioeconomic conditions.

6.5.2 Sample collection, DNA extraction, and sequencing

Samples were collected from two sites for each participant: supragingival plaque (dental plaque) and buccal swabs to sequence the microbiome and host genetic factors, respectively. Dental plaque samples were obtained using sterile interdental brushes and collected from all tooth surfaces, regardless of cavity presence (Agnello *et al.*, 2017; de Jesus *et al.*, 2020; Khan, de Jesus, *et al.*, 2024). Buccal cavity samples were collected by swabbing

the buccal mucosa and the anterior floor of the mouth beneath the tongue (de Jesus, Khan, *et al.*, 2021).

Amplicon sequencing of bacteria and fungi was performed using MiSeq or NovoSeq6000 platforms (Illumina Inc., San Diego, CA, USA). For bacterial 16S rRNA (V4 region), the primer pair 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) was used. Fungal ITS1 regions were targeted using the primers ITS1-30 (5′-GTCCCTGCCCTTTGTACACA-3′) and ITS1-217 (5′-TTTCGCTGCGTTCTTCATCG-3′). Raw sequences were processed using QIIME2 to generate an amplicon sequence variant (ASV) table containing microbial abundance data.

For host genetic analysis, 55 candidate genes were selected based on their roles in taste signal transduction, including receptors for bitter, sweet, umami, salt, sour, carbonation, and fatty acids, and their associated signaling proteins (**Table S6.1**). DNA was isolated using the QIAamp Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The candidate genes were amplified using primers designed for the Fluidigm platform to facilitate Illumina sequencing, ensuring that each amplified region covered approximately 250–350 base pairs for compatibility with 250 × 2 paired-end sequencing. All sequencing was conducted at the Genome Quebec Innovation Center (Montreal, Canada), and the resulting sequences were provided as demultiplexed paired-end FASTQ files with the barcodes removed. Additional details regarding these methods are given the our previous publications (de Jesus, Khan, *et al.*, 2021; de Jesus *et al.*, 2022).

6.5.3 Raw sequence processing

Microbiome

The raw FASTQ sequences from amplicon sequencing were processed separately for 16S rRNA and ITS1 in QIIME2 to generate ASV tables for bacteria and fungi (Bolyen *et al.*, 2019). For bacterial analysis, the DADA2 plugin in QIIME2 was used, whereas fungal analysis included an additional preprocessing step using Q2-ITSxpress in QIIME2 to remove conserved regions flanking ITS1 before applying the DADA2. Taxonomic assignment was conducted using the Human Oral Microbiome Database (HOMD, version 15.23) for bacteria and UNITE

(version 9.0) for fungi. Taxonomies with less than 5% prevalence across the samples were excluded from the final ASV tables.

Differentially abundant species were identified using the MaAsLin2 method, which performs multivariable associations while accounting for confounding variables (Mallick *et al.*, 2021). This analysis used ECC status as the outcome variable, with age, sex, rural-urban status, and SEFI score as the confounders. MaAsLin2 applies CLR normalization internally before performing association tests and provides coefficients along with *q*-values (Gloor *et al.*, 2017), which are *p*-values adjusted using the Benjamini-Hochberg (BH) method. In this study, MaAsLin2 was applied separately to the bacterial and fungal ASV tables with a significance threshold of $q < 0.01$. A detailed description of microbiome analysis using amplicon sequencing and preprocessing steps to obtain taxonomic abundances at the species and genus levels has been provided in our recent publication (Khan, de Jesus, *et al.*, 2024).

Candidate genes

The raw sequences of the candidate genes were mapped to the GRCh38 human reference genome using the BWA mapping tool (Li and Durbin, 2009). The workflow adhered to the GATK best practice guidelines, with the exception of PCR duplicate removal, as this step is not recommended for targeted region amplification (McKenna *et al.*, 2010; Ebbert *et al.*, 2016; Fu *et al.*, 2018). A schematic of the GATK pipeline steps, including the integration of data from multiple batches during the analysis, is provided in the Supplementary data (**Figure S6.1A**). A combined VCF file containing genotyping information for the genes included in the study was generated for all 538 samples. This file was cleaned and filtered using VCFtools (v 0.1.16). Quality control steps included filtering for a Phred quality score greater than 25, individuals with less than 50 % missing genotype, and a sequencing depth greater than 5. SNPs were annotated using BCFtools with the human dbSNP138 database, and variant annotation was performed using SnpEff based on the human genome version, GRCh38. For missing variants, all missing values were considered as not available. For analyses in which missing genotypes were not permissible, the genotype of the reference genome was used. Genetic variants were named using the Gene_rsID approach when the rsID was available. For variants lacking rsID, the naming convention Gene_p followed by the

chromosomal position was used. The absence of rsID also indicates novel variants not assigned to the dbSNP138 database.

6.5.4 Genetic association testing

The association between the variants identified by the GATK pipeline and ECC was tested using PLINK (v1.90) (Purcell *et al.*, 2007; McKenna *et al.*, 2010). Quality control and variant selection criteria included a genotype missingness threshold of less than 1% (--geno), a MAF of at least 1% (--maf), and the removal of SNPs that significantly deviated from the Hardy-Weinberg equilibrium (HWE) with a *p*-value threshold of 0.00001 (--hwe). A logistic regression approach in PLINK was used to assess the association between genetic variants and ECC status.

The analysis was performed using additive, dominant, and recessive genetic models, with genotypes coded following standard practices used in PLINK. To identify the appropriate covariates, a base model was first constructed in PLINK, including only age and sex as covariates. Additional covariates, such as rural-urban status, SEFI score, the first five PCs, derived from the quality-filtered genetic variants in this study using PLINK, and batches, were subsequently added to compare their AIC and BIC values using a linear model. The model was tested using the top five variants identified in the base model and the genomic inflation factor, which quantifies the deviation of the observed *p*-values from those expected under the null hypothesis. The AIC and BIC criteria for model performance for covariate combinations suggested that while rural-urban status and SEFI score improved the model, batch information did not (**Figure S6.4**). The final model included age, sex, urban-rural status, SEFI score, and top five PCs. Logistic regression in PLINK was performed with a *q*-value (Benjamini-Hochberg adjusted *p*-value) below 0.05 considered significant.

Additionally, gene-level association analyses of ECC outcomes were conducted using the SKAT method. Equal weights were applied to all variants, including those with MAF greater than or less than 0.01. However, for gene-level analyses focusing only on rare variants, an MAF threshold of less than 0.01 was used. The SKAT analysis employed the "SKAT-O" option, which optimally combines the standard SKAT test and burden test, adapting to variant causality and effect directionality (Wu *et al.*, 2011; S. Lee *et al.*, 2012).

6.5.5 Microbiome quantitative trait loci

We first examined the changes in microbial diversity at both the alpha and beta levels associated with the three variants identified as ECC-associated. Microbial alpha diversity was measured using the Shannon index in Phyloseq, and correlations between Shannon diversity and variants were assessed using Spearman's correlation. For beta diversity, the distance matrix was calculated using the Bray method in Phyloseq, and the variance in the distance matrix explained by genetic variants was analyzed using the "adonis2" function in Vegan.

To perform the microbiome quantitative trait loci (mbQTL) analysis, linear regression analysis was performed using the "glm" function. In this analysis, variants were treated as independent variables, and microbial abundance served as the outcome variable, with ECC status included as an additional covariate along with the other covariates used in the genetic association testing. The inclusion of ECC status as a covariate accounted for its potential confounding effect on microbial abundance, thereby reducing bias in the associations. The *p*-values were adjusted using the BH method. Only significant variants identified under the dominant genetic model were selected for analysis. To further explore these associations, a stratified association analysis was conducted separately within the ECC and CF samples using the same methodology.

6.5.6 Mediation analysis

To determine whether the effect of genetic variants on ECC outcomes was mediated through the microbiome, we used LDM mediation in R, which applies a linear decomposition model to test the mediation effects (Yue and Hu, 2022). This approach evaluates the exposure-taxon and taxon-outcome associations, where the exposure is an individual genetic variant and the outcome is ECC status. This analysis assumes that each variant independently influences several microbial species. Therefore, while the microbiome data included all significant taxa identified using MaAsLin2, only one variant was analyzed at a time.

LDM mediation enables mediation testing at the community level, incorporating all taxa provided in the analysis. The results identified potential taxa mediating the relationship between a genetic variant and ECC outcomes based on a defined significance cutoff. All significant taxa identified using the MaAsLin2 method were tested for their mediating effects.

Furthermore, we conducted mediation analysis using the R package mediation to evaluate the mediation effect between a single genetic variant and a single microbial species. This approach provided the ACME as the mediation coefficient and the ADE to quantify the direct relationship.

6.5.7 Machine learning classifiers for ECC

To evaluate the relative importance of genotyping variants identified in our study alongside microbiome features, we incorporated genotyping data for quality filtered variants in a dominant fashion with microbiome data and constructed a classifier for CF and ECC classification in R using the tidymodels package (version 1.1.1). The random forest, Lasso, Ridge, and SVM models were tested owing to their frequent application to similar data types. The model was tuned for the minimum node size, number of trees, and number of variables at each split. A grid size of 50 was used for the hyperparameter tuning.

The model performance was assessed using the AUROC and AUPRC metrics. The AUROC value quantifies the models' ability to correctly predict the outcome classes, whereas the AUPRC suggests a trade-off between the precision and recall of the prediction. The variable importance in the final random forest model was evaluated based on the mean decrease in accuracy upon shuffling each feature using the VIP package (version 0.4.1) in R to measure the impact on performance. Additionally, the random forest model trained on bacterial species from our data was tested on external bacterial datasets from our previous study (Khan, Fung, *et al.*, 2024), and the performance was assessed using the same criteria as above. The prediction was conducted with and without batch correction to assess performance under different preprocessing conditions.

6.5.8 Microbiome risk score

In addition to the machine learning methods described in the previous section, we utilized MRS markers based on bacterial species that can be tested on external datasets. Similar to the polygenic risk score (Choi, Mak and O'Reilly, 2020), the MRS score calculates an individual's predisposition to ECC based on their microbiome profile by aggregating the effects of multiple microbial species identified as ECC-associated by MaAsLin2, with a *q*-value cutoff of < 0.01 (Choi, Mak and O'Reilly, 2020). Each microbial marker was weighted by

its MaAsLin2 coefficient and combined in a linear equation with the microbial profile. The MRS was calculated for 80% of the samples used in our study, and the resulting score represented the cumulative risk. The equation was then used as a predictor for the remaining 20% of the holdout test dataset. The performance of the MRS derived from our dataset was evaluated using four external datasets. Predictions using MRS were assessed using the AUROC and AUPRC scores.

6.6 Supplementary data

6.6.1 Inclusion criteria for genes

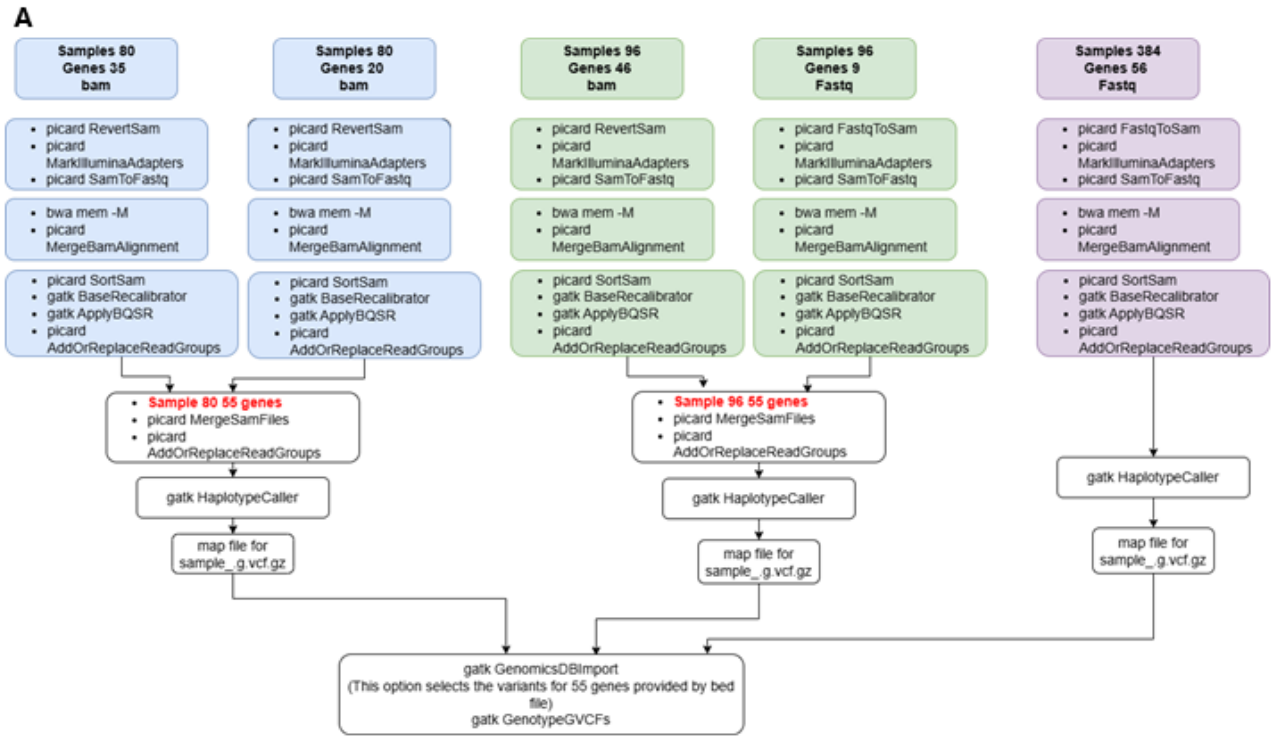
The inclusion criteria for selecting the candidate genes were based on their roles in taste perception. This included 25 genes for bitter taste receptors, collectively known as the TAS2R family, which are part of the GPCR superfamily and activated by bitter compounds. The other group of genes included the TAS1R family, which detect sweet and umami tastes through GPCR signaling. Additionally, several GPCRs involved in various taste signaling pathways were also included. The study also included genes encoding free fatty acid receptors, ion channels, and transporters for sodium balance, carbonic anhydrase genes involved in taste perception and pH regulation, and various other signal transduction proteins. A detailed list of these genes is provided in Supplementary **Table S6.1**. In our study, we utilized deep sequencing of gene exomes to identify novel variants, specifically frameshifts caused by insertions and deletions. However, no novel variants were found to be significantly associated with ECC in our analyses.

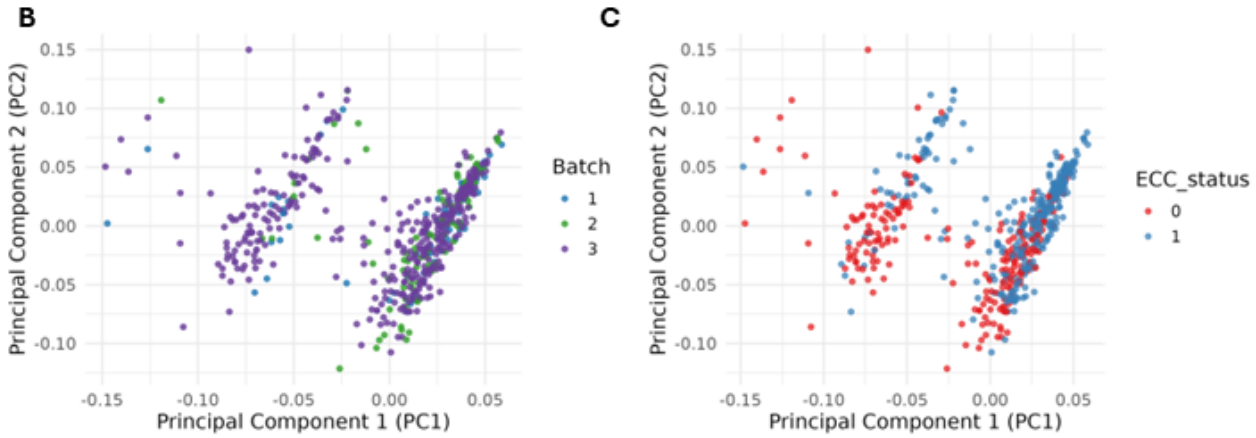
6.6.2 Covariates included in the analyses

In addition to age and sex, the SEFI score and rural-urban information were included as confounders because of their association with ECC outcomes (Khan, de Jesus, *et al.*, 2024). To address potential biases related to multi-ancestry in our samples due to diverse ethnic groups and immigration patterns over time, we also incorporated the top five principal components (PCs) (**Figure S6.2A**). Although sequencing was conducted at multiple time points, batch information was not included as a covariate because most batch effects were captured by PC1, which showed a Spearman correlation of 0.81. Consequently, PC1 in our covariates effectively served as a proxy for batch effects (**Figure S6.2B**). Furthermore, including batch effects as a covariate did not improve the model. A comparison of genomic inflation factors indicated that adding batch effects after incorporating PCs worsened the model by increasing genomic inflation (**Figure S6.3**). This likely reflects the bias introduced by the strong correlation between PC1 and the batch effects. This observation was further supported by the AIC and BIC values from covariate comparisons for individual models of the top variants (**Figure S6.4**).

Table S6.1: List of the genes considered as candidate genes for targeted sequencing.

Gene Symbol	Function/Role	Type
TAS2R1 TAS2R3 TAS2R4 TAS2R5 TAS2R7 TAS2R8 TAS2R9 TAS2R10 TAS2R13 TAS2R14 TAS2R16 TAS2R19 TAS2R20 TAS2R30 TAS2R31 TAS2R38 TAS2R39 TAS2R40 TAS2R41 TAS2R42 TAS2R43 TAS2R45 TAS2R46 TAS2R50 TAS2R60	Bitter taste receptors	Taste G-protein-coupled receptors (GPCRs)
TAS1R1 TAS1R2 TAS1R3	Sweet and umami taste receptors	Taste GPCRs
SCNN1A SCNN1B SCNN1D SCNN1G	Sodium ion transport	Epithelial Sodium Channels (ENaC)
GPR161 GPR107 GPR87 ADGRG1 ADGRG6	Signal transduction and cellular communication	GPCRs
FFAR1 FFAR2 FFAR3 FFAR4	Free fatty acid receptors	GPCRs - Fatty Acid Receptors
GNAI2 GNAT3 GNAS GNAQ	Intracellular signaling via G proteins	G-proteins
PLCB2	Secondary messenger generation (phosphoinositide pathway)	Phospholipase C
CA6 CA7	pH regulation through bicarbonate production	Carbonic Anhydrases
RALB RAC1	Small GTPase-mediated signaling	Small GTPases
OTOP1	Proton transport and pH sensing in taste	Ion Channel
FZD6	Wnt signaling	Frizzled class receptor
GABBR1	Synaptic inhibition via GABA signaling	GABA Receptor


A
Samples 80 Genes 35 bam
Samples 80 Genes 20 bam
Samples 96 Genes 46 bam
Samples 96 Genes 9 Fastq
Samples 384 Genes 56 Fastq
picard RevertSam
picard MarkIlluminaAdapters
picard SamToFastq
picard FastqToSam
bwa mem -M
picard MergeBamAlignment
picard SortSam
gatk BaseRecalibrator
gatk ApplyBQSR
picard AddOrReplaceReadGroups
Sample 80 55 genes
picard MergeSamFiles
Sample 96 55 genes
gatk HaplotypeCaller
map file for sample_.g.vcf.gz
gatk GenomicsDBImport
(This option selects the variants for 55 genes provided by bed file)
gatk GenotypeGVCFs


B
C
Principal Component 2 (PC2)
Principal Component 1 (PC1)
Batch
1
2
3
ECC_status
0
1

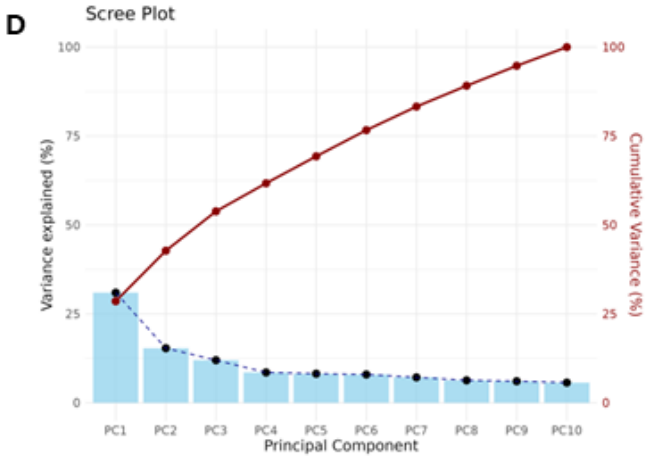

D
Scree Plot
Variance explained (%)
Cumulative Variance (%)
Principal Component
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10

Figure S6.1: Batch processing, batch effects, covariates, and principal components from genetic data.

(A) Flowchart of the steps used to combine data from different sequencing runs. Initially, 560 samples were run, but only 538 were retained after removing duplicated samples and age selection criteria of less than 72 months.

(B) Batch effects in the quality-filtered genetic variants after selection cutoff, as explained in the Methods section.

(C) PCA plot colored by CF (0) and ECC (1) status.

(D) Scree plot of principal components (PCs) showing the variance explained by individual components (left y-axis) and line plot of cumulative variance (right y-axis).

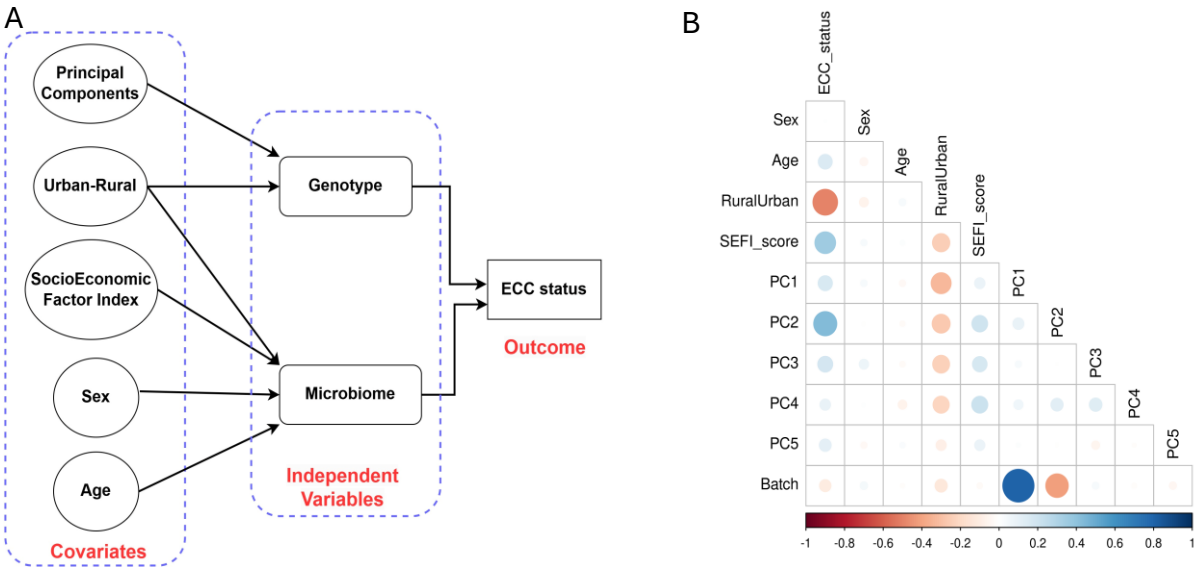


Figure S6.2: Covariates used in this study.
(A) Directed acyclic graph illustrating the hypothesized relationships between covariates, independent variables, and outcome in our study
(B) Spearman correlation between the covariates and the top five principal components (PCs) from the genetic data used in the study.

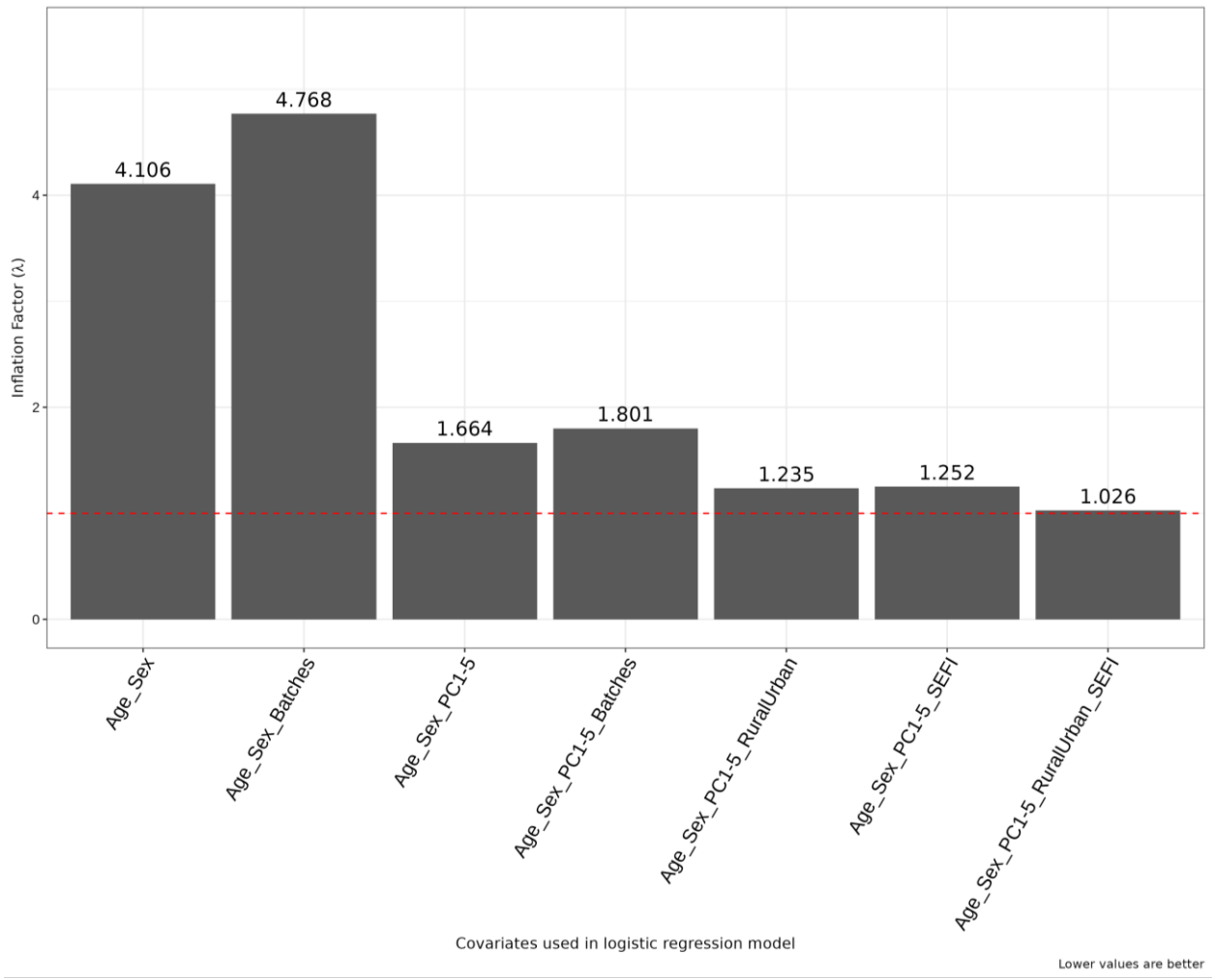


Figure S6.3: Genomic inflation rate for *p*-values for all variants using different covariates. The λ is a measure of the deviation of *p*-values from the null distribution, with λ = 1 indicating no inflation. Each bar represents the combinations of covariates, which include age, sex, batches, rural-urban status, SEFI score, and top 5 principal components (PC1-5).

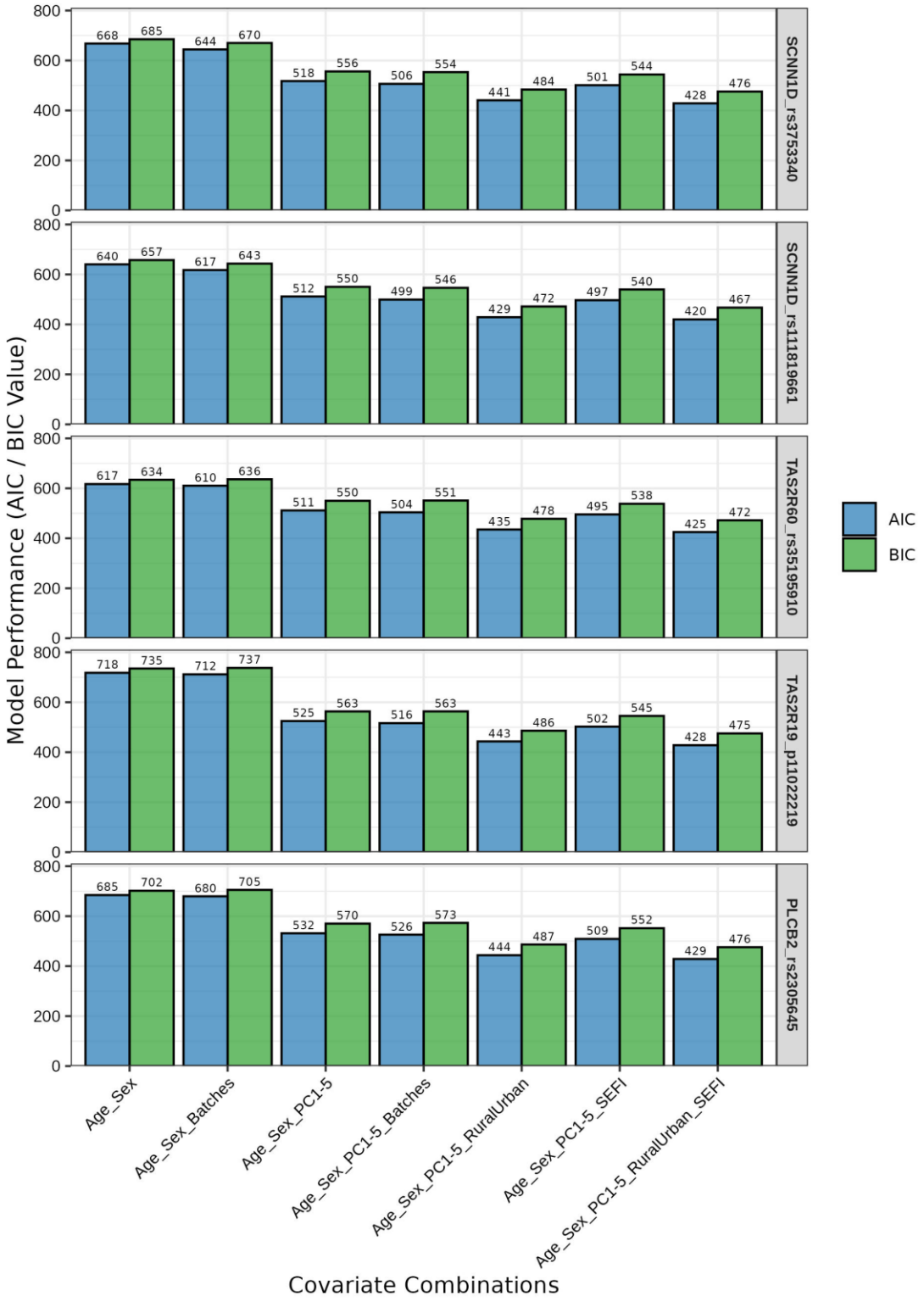


Figure S6.4: Model evaluation for different covariates using the top five ECC-associated taste genetic variants.

The top five variants identified using the base additive model for the association between variants and ECC status using age and sex as covariates. Various models were evaluated for combinations of

covariates, including batches, rural-urban status, SEFI score, and the top five principal components (PC1-5). The *glm* model was used with ECC status as the outcome, with each variant as an independent variable, and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were compared. (The lower values are better).

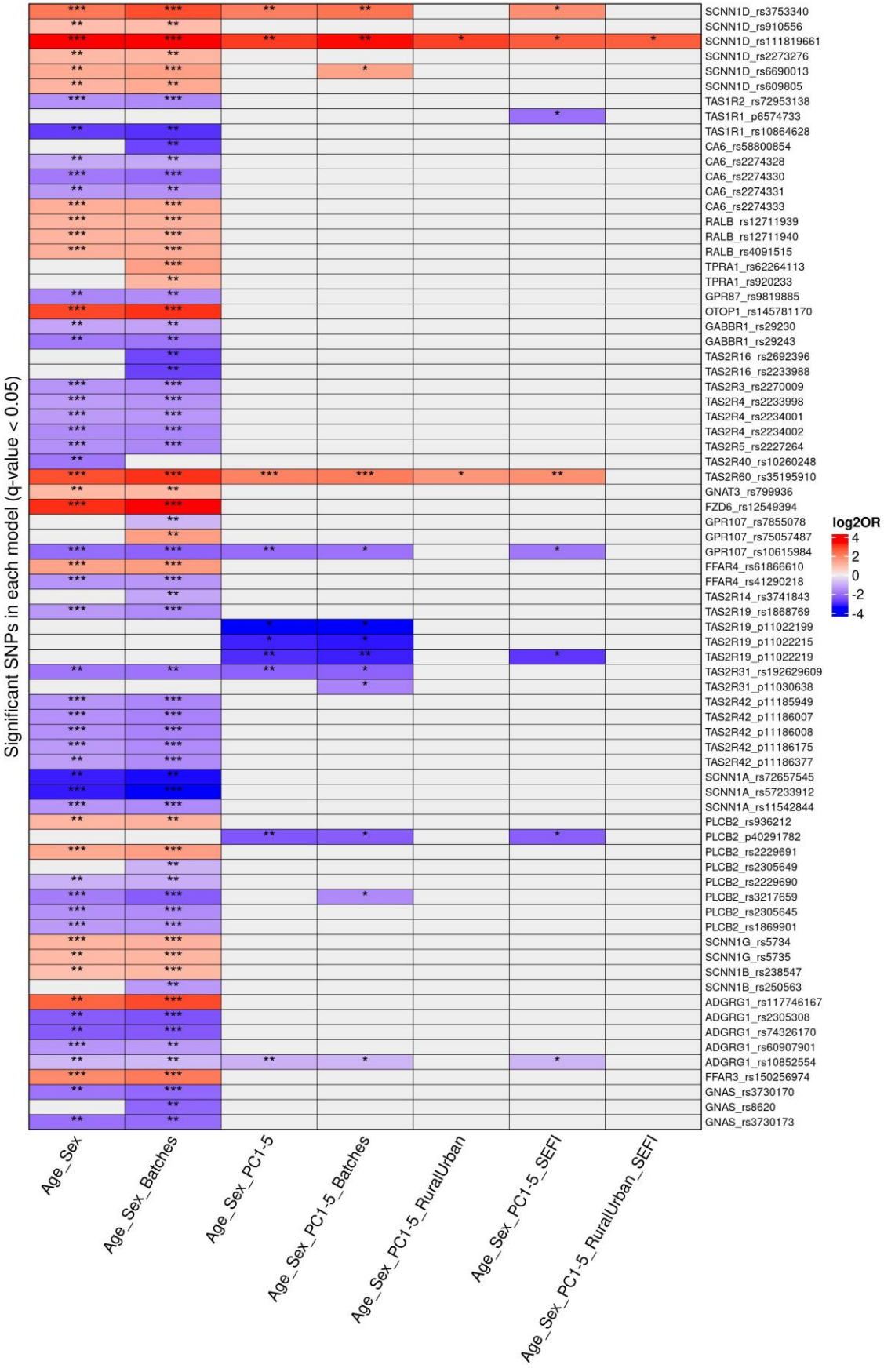

Significant SNPs in each model (q-value < 0.05)
SCNN1D_rs3753340
SCNN1D_rs910556
SCNN1D_rs111819661
SCNN1D_rs2273276
SCNN1D_rs6690013
SCNN1D_rs609805
TAS1R2_rs72953138
TAS1R1_p6574733
TAS1R1_rs10864628
CA6_rs58800854
CA6_rs2274328
CA6_rs2274330
CA6_rs2274331
CA6_rs2274333
RALB_rs12711939
RALB_rs12711940
RALB_rs4091515
TPRA1_rs62264113
TPRA1_rs920233
GPR87_rs9819885
OTOP1_rs145781170
GABBR1_rs29230
GABBR1_rs29243
TAS2R16_rs2692396
TAS2R16_rs2233988
TAS2R3_rs2270009
TAS2R4_rs2233998
TAS2R4_rs2234001
TAS2R4_rs2234002
TAS2R5_rs2227264
TAS2R40_rs10260248
TAS2R60_rs35195910
GNAT3_rs799936
FZD6_rs12549394
GPR107_rs7855078
GPR107_rs75057487
GPR107_rs10615984
FFAR4_rs61866610
FFAR4_rs41290218
TAS2R14_rs3741843
TAS2R19_rs1868769
TAS2R19_p11022199
TAS2R19_p11022215
TAS2R19_p11022219
TAS2R31_rs192629609
TAS2R31_p11030638
TAS2R42_p11185949
TAS2R42_p11186007
TAS2R42_p11186008
TAS2R42_p11186175
TAS2R42_p11186377
SCNN1A_rs72657545
SCNN1A_rs57233912
SCNN1A_rs11542844
PLCB2_rs936212
PLCB2_p40291782
PLCB2_rs2229691
PLCB2_rs2305649
PLCB2_rs2229690
PLCB2_rs3217659
PLCB2_rs2305645
PLCB2_rs1869901
SCNN1G_rs5734
SCNN1G_rs5735
SCNN1B_rs238547
SCNN1B_rs250563
ADGRG1_rs117746167
ADGRG1_rs2305308
ADGRG1_rs74326170
ADGRG1_rs60907901
ADGRG1_rs10852554
FFAR3_rs150256974
GNAS_rs3730170
GNAS_rs8620
GNAS_rs3730173
log2OR
4
2
0
-2
-4
Age_Sex
Age_Sex_Batches
Age_Sex_PC1-5
Age_Sex_PC1-5_Batches
Age_Sex_PC1-5_RuralUrban
Age_Sex_PC1-5_SEFI
Age_Sex_PC1-5_RuralUrban_SEFI
Models with Covariates

Figure S6.5: Significant genetic variants across different association models for ECC outcome.

Each row represents a specific variant identified by its chromosomal location and associated gene. The columns depict different statistical models based on covariate inclusion, which is given on the x-axis and includes a combination of age, sex, batch effects, rural-urban status, SEFI score, and the top five principal components from genetic variants in the 55 taste-associated candidate genes. The color represents the log2odds ratio, while asterisks denote levels of statistical significance (* p<0.05, ** p<10^{-3}, *** p<10^{-5}) with white indicating no significant association.

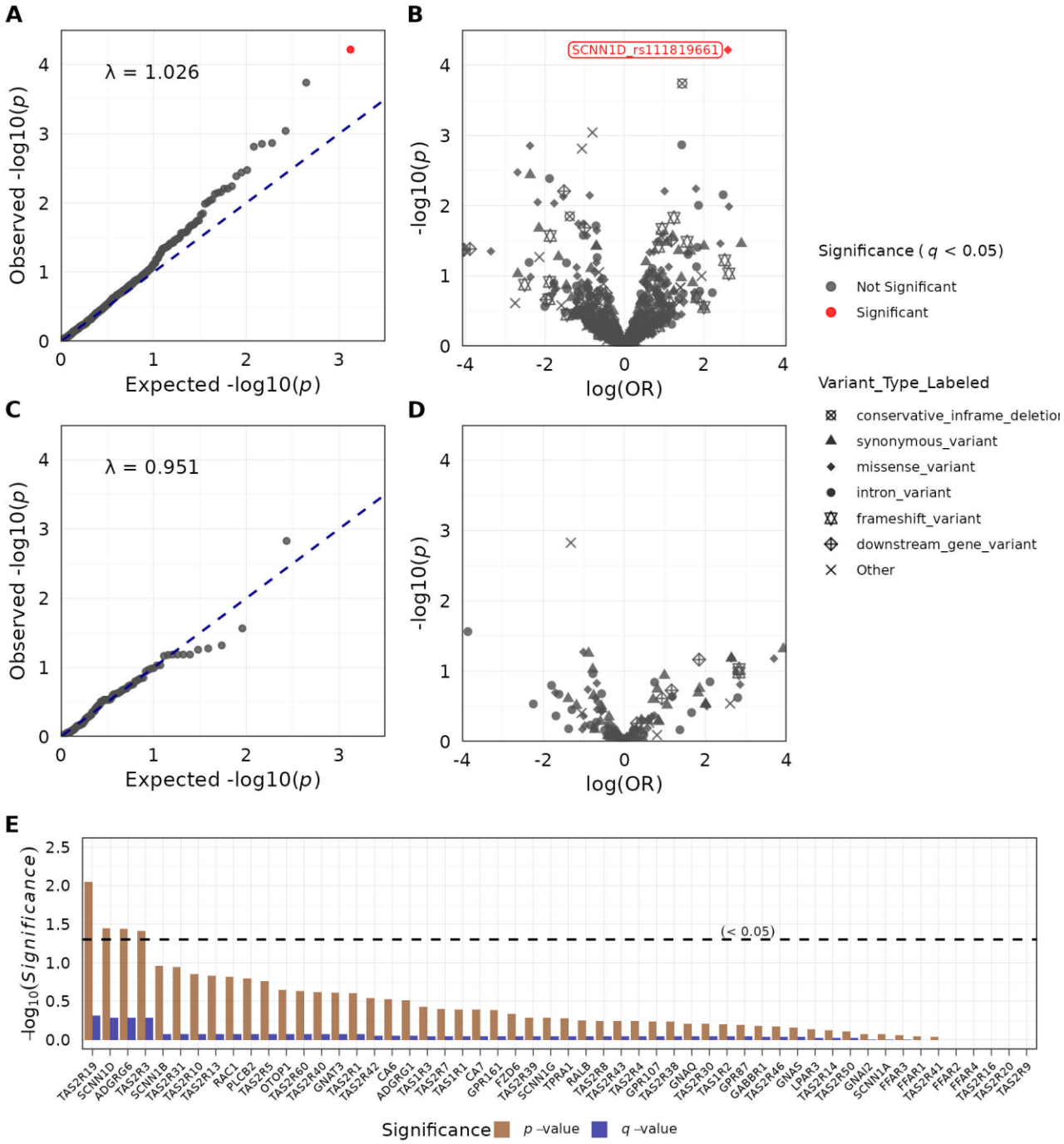


Figure S6.6: Genetic variants and gene association with ECC status using a logistic model.

(A) Additive model. The Q-Q plot illustrates the distribution of observed -log10(*p*-value) values against the expected -log10(*p*-value) values under the null hypothesis of no association between the variables. Each point represents a single nucleotide polymorphism (SNP) that was tested in the analysis. The red dashed line represents the null expectation, where the observed values equal the expected values.
(B) Volcano plot for additive model. The x-axis represents the log2 odds ratio (OR), and the y-axis shows the -log10(*p*-value). Each point represents a variant, with the shape indicating the variant type. Red points denote significant associations for q <0.05, where q-values are BH adjusted p-value <0.05.
(C) Similar to A, but for the recessive genetic model.
(D) Similar to B, but for the recessive genetic model.
The Q-Q and volcano plots are from logistic regression models in PLINK for 55 taste-associated candidate genes, with age, sex, rural-urban status, SEFI score, and PCs 1–5 as covariates.
(E) Gene-wise association for rare, minimum allele frequency (MAF) < 0.01) and common variants (MAF > 0.01) with default weights in the SKAT method in R. The bar plot represents the p-values and q-values for the association between variants within each gene after removing singletons. The analysis was performed in the SKAT R-package using the SKAT-O option for optimal association with age, sex, rural-urban status, SEFI score, and top 5 principal components as covariates. The horizontal line indicates the significance value cutoff of <0.05.

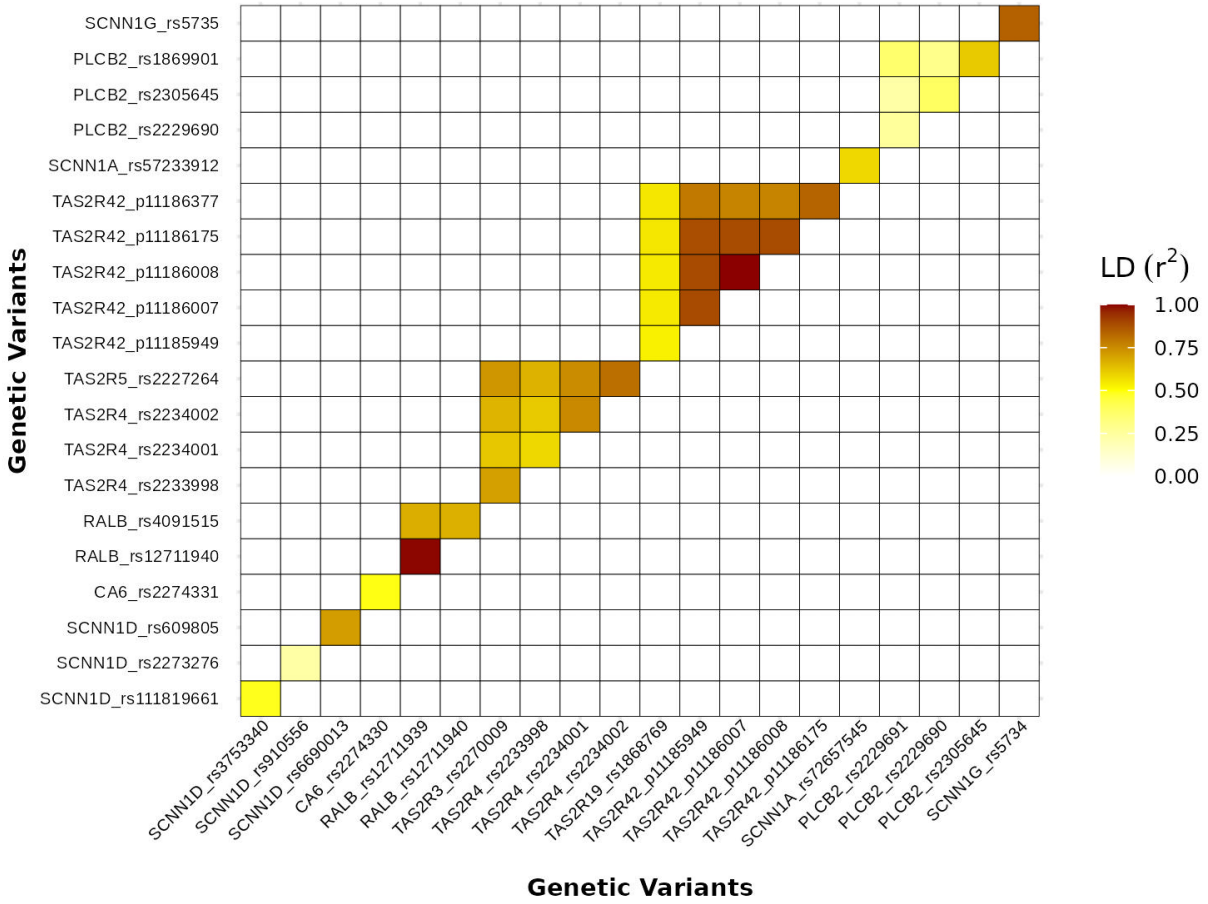


Figure S6.7: Linkage disequilibrium (LD) analysis of the variants identified in the model with age and sex as covariates.
Only the variants in *TAS2R3*, *TAS2R4*, and *TAS2R5* exhibited a moderate level of intergenic LD, whereas no significant intergenic LD was observed among the other variants.

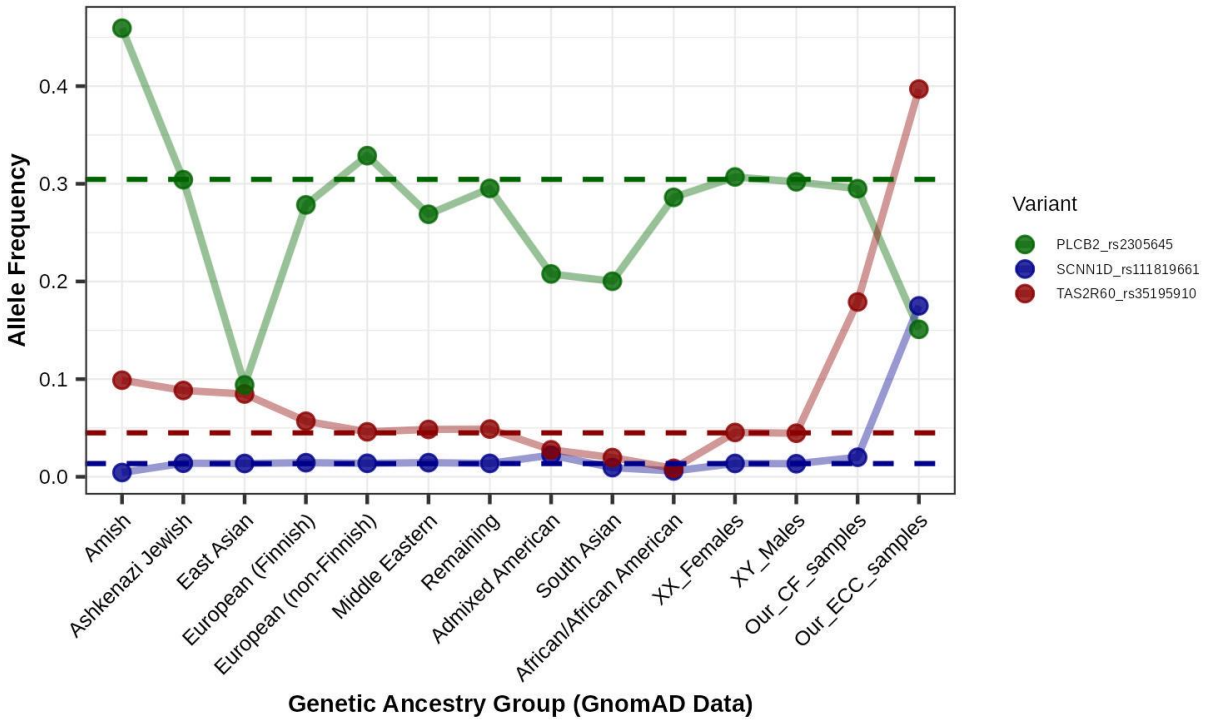


Figure S6.8: Allele frequencies of ECC-associated significant taste-related genetic variants.

The figure presents the allele frequencies of *SCNN1D*-rs111819661, *TAS2R60*-rs35195910, and *PLCB2*-rs2305645 across different genetic ancestry groups from the gnomAD database v4.1. Horizontal lines indicate the average allele frequency of the population for each variant.

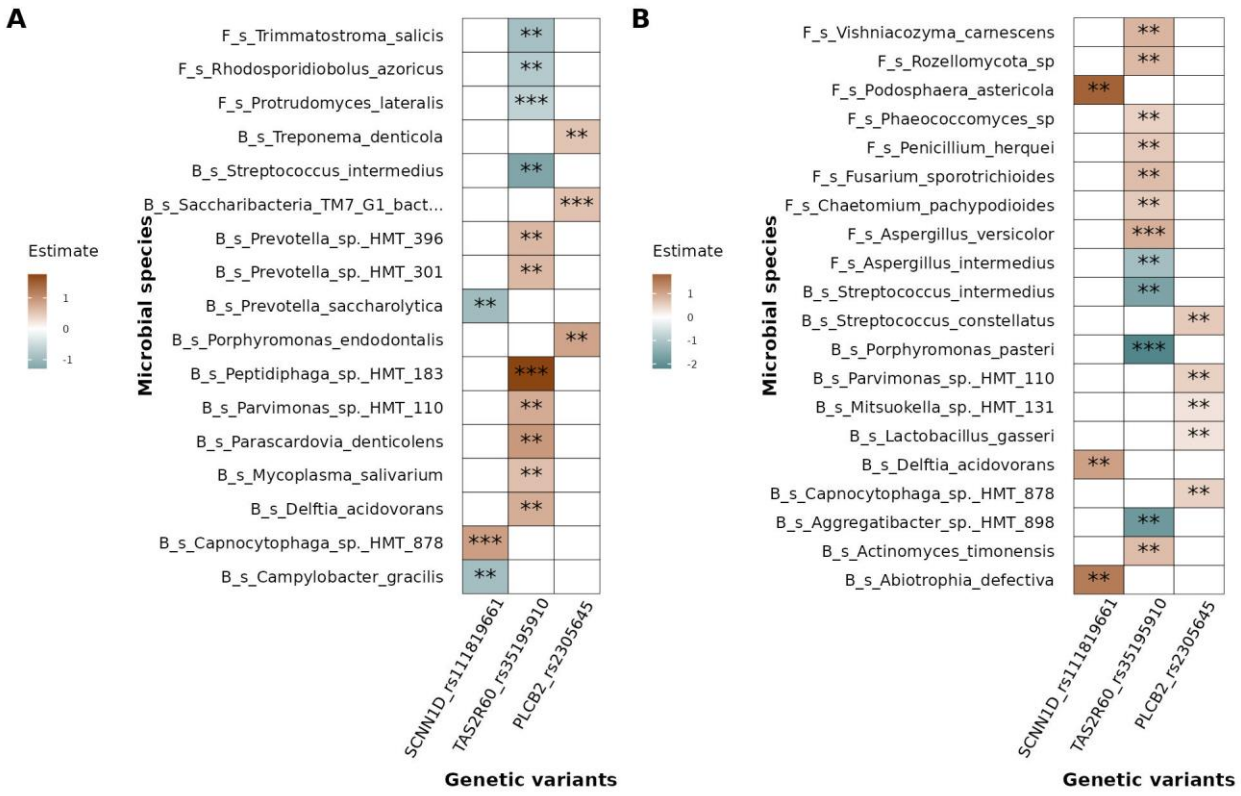


Figure S6.9: Association between ECC-associated variants and microbial taxa in ECC and CF stratified samples.

The heatmap illustrates the association between ECC-associated genetic variants and microbial taxa when adjusted for covariates such as age, sex, urban-rural status, SEFI score, and the top five principal components from the genetic data. The forest plot shows the coefficient of association from the MaAsLin2 method between the selected microbial species in the heatmap and ECC outcome. Positive coefficients favor ECC, whereas negative coefficients favor CF status. The analysis was performed for significant variants identified in the dominant genetic model and all microbial species with CLR normalization using a linear model. The heatmap value signifies the estimates values, with red showing a positive association between variants with alternate alleles and green representing a negative association. The *p*-values are (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level, and 'g' indicates genus level. (A) Genus level (B) Species level.

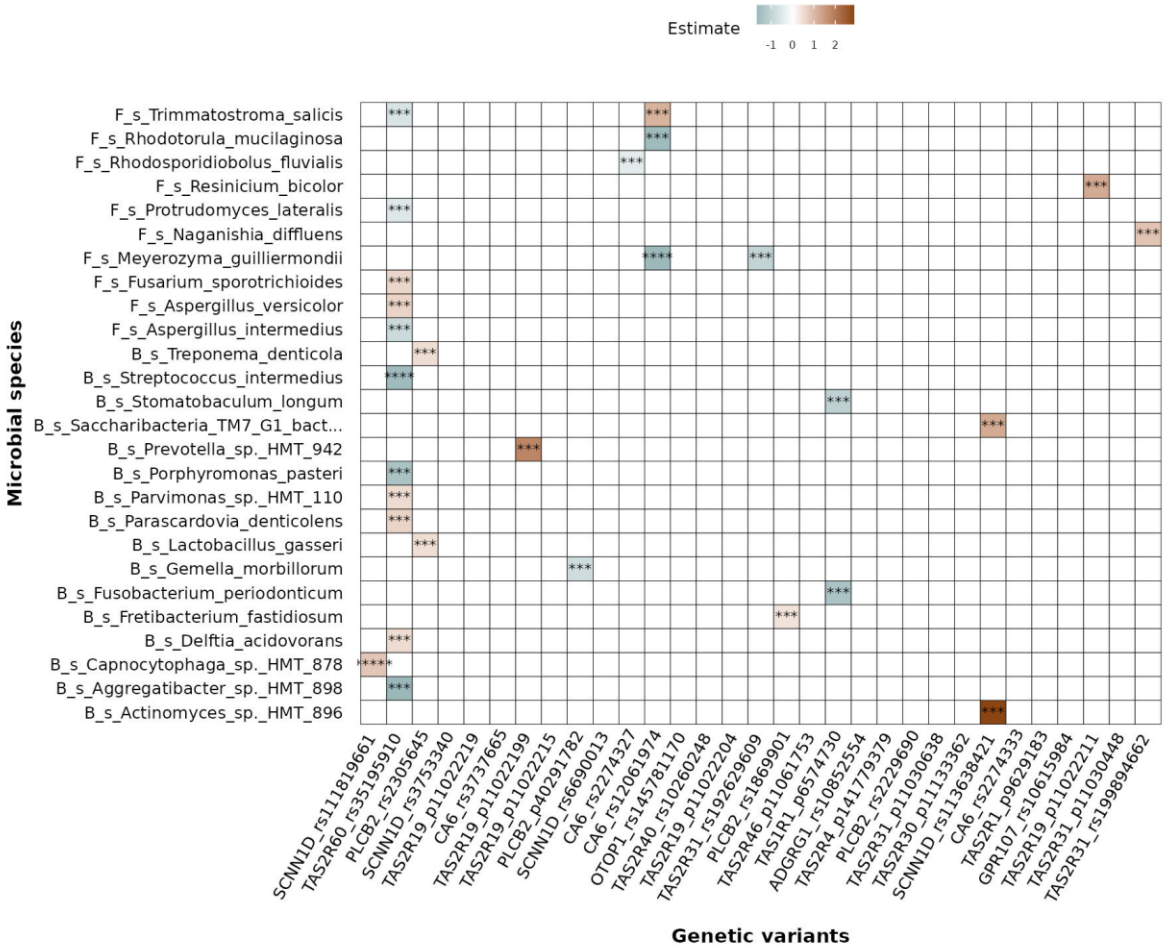


Figure S6.10: Association between 31 top variants (5% of all quality-filter passing variants) based on *p*-values for association with ECC and all microbial taxa.

With covariates age, sex, rural-urban status, SEFI score, top 5 principal components, and ECC status The analysis was performed for significant variants identified in the dominant genetic model and all microbial species with CLR normalization using a linear model. The heatmap value signifies the estimates values, with red showing a positive association between the variant with the alternate allele and blue representing a negative association. The *p*-values are (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$). Microbial features are labeled as 'B' for bacterial and 'F' for fungal taxa. The prefix 's' signifies species level, and 'g' indicates genus level. Genetic variants were named using the Gene_rsID (Reference SNP ID) approach when rsID based on dbSNP138 was available. For variants lacking rsID, the naming convention Gene_p followed by the chromosomal position was applied, indicating novel variants not assigned in the dbSNP138 database.

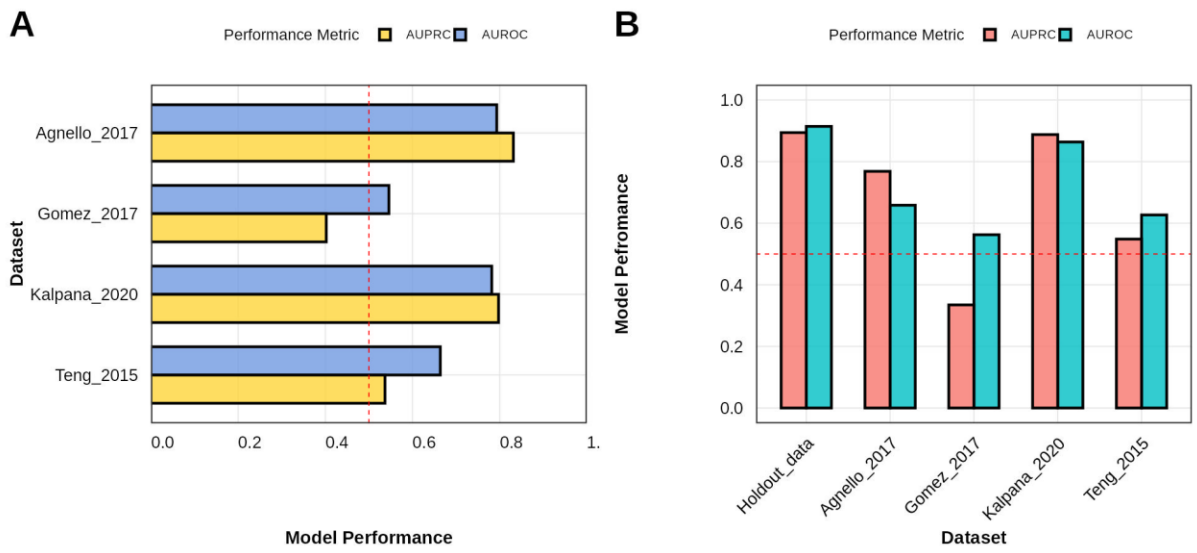


Figure S6.11: Machine learning based classification for ECC status using microbiome data on external datasets.

(A) Random forest model performance on external datasets (without batch correction). Four external datasets were chosen to assess the random forest model based on only microbiome data, and the model performance was assessed using AUROC and AUPRC values.

(B) ECC prediction model performance on external datasets (without batch correction) based on the MRS score on CLR-normalized datasets.

Chapter 7: Conclusions and Future Directions

7.1 Conclusions

Despite the inherent differences between studies, our research identified common microbial species associated with ECC across multiple studies using cross-cohort microbiome analysis. While alpha diversity did not substantially differ across datasets, beta diversity showed significant differences, indicating that competition for resources leads to distinct microbial compositions, even when overall species richness remains similar. Microbial species commonly enriched in ECC across multiple studies include *Megasphaera micronuciformis*, *Prevotella salivae*, *Selenomonas sputigena*, and *Prevotella oulorum*, with the genera *Alloprevotella* and *Megasphaera* being prominent. Conversely, the CF-enriched species included *Rothia aeria* and *Lachnoanaerobaculum umeaense*, with consistently enriched genera such as *Leptotrichia*, *Capnocytophaga*, and *Arachnia*. These taxa were robustly identified using both differential abundance methods and machine learning models.

Additionally, *Streptococcus mutans*, *Streptococcus salivarius*, *Neisseria bacilliformis*, and *Alloprevotella* species sp. HMT-912 and sp. HMT-473 were identified as the top ECC-associated species in machine learning analyses. Our findings suggest that random forest modeling provides superior and more consistent performance in terms of AUROC values when analyzing microbiome data, supporting its preferential use in future investigations. However, the generalizability of random forest models may be limited by the inherent differences in sequencing data.

We also observed improvements in species-level comparisons using traditional differential methods after batch correction. Nonetheless, random forest marginally improved classification even without batch correction, likely because of its capacity to capture nonlinear patterns.

In our second study, we investigated the association between various socioeconomic and behavioral factors and ECC. We found strong associations between ECC prevalence and socioeconomic factors, such as the SEFI score and place of residence (rural-urban status), with higher ECC occurrence in rural and socioeconomically deprived communities.

Additionally, specific behavioral habits, particularly bedtime snacking, were significantly associated with ECC.

Our comparative machine learning analysis evaluating the relative roles of bacteria and fungi in ECC identified *S. mutans* as the most influential species, followed by *Candida dubliniensis* and *Candida albicans*. Interkingdom association analysis revealed a high co-occurrence between *Neisseria bacilliformis* and *C. dubliniensis*, emphasizing the crucial role of fungi in dental caries pathogenesis and highlighting the need for further fungal-focused ECC research. Machine learning models, including LASSO, Ridge, and SVM, performed comparably to random forest when analyzing combined bacterial-fungal data; however, their performance decreased substantially when analyzing fungal data alone. This observation suggests the predominant role of bacteria relative to fungi in ECC.

While many previous studies have focused exclusively on bacterial microbiomes, our approach analyzed both bacteria and fungi from the same samples, facilitating direct comparisons of fungal components alongside bacterial communities, which most of the earlier research has been based on.

In our third study, we examined the association between taste gene variants and ECC status, along with their influence on the composition of the dental plaque microbiome. We selected 55 candidate genes involved in taste signaling and performed deep sequencing to identify genetic variants. Our association analyses identified three variants—rs111819661 in *SCNN1D*, rs35195910 in *TAS2R60*, and rs2305645 in *PLCB2*— that were significantly associated with ECC.

Gene-level association testing highlighted *ADGRG1* and *TAS2R60* as strongly associated with ECC, suggesting a combined effect of rare and common genetic variants on ECC outcomes. Although no significant associations were observed between these variants and ECC-associated microbial species, certain microbial species were significantly associated with the presence or absence of variants, indicating their potential role in shaping oral microbiome profiles. These findings imply that variations in taste-associated genes can influence microbial composition. Random forest modeling identified *S. mutans*, *TAS2R60*-rs35195910, and *C. dubliniensis* as the top markers influencing ECC. Upon validation with external

datasets from Chapter 4, these microbiome models performed well on two datasets but exhibited lower generalizability on the remaining two, underscoring the necessity of population-specific validation.

In conclusion, our study identified key microbial contributors to ECC, highlighted important interkingdom microbial interactions, demonstrated the significant impact of socioeconomic factors, and evaluated the relative importance of taste genetic variants alongside microbial species. Unlike previous studies limited to bacterial microbiome analysis, our research integrated fungal data, emphasizing the critical role of fungi and highlighting the importance of including fungi in ECC research designs. Furthermore, by incorporating socioeconomic factors, the dental plaque microbiome, and taste genetic variants, our study elucidated their relative contributions to ECC. Our results underscore the need for multiomics studies employing machine learning-based microbial and genetic interaction models to further advance our understanding of these complex relationships. These findings can guide the development of targeted prevention strategies, such as personalized risk assessment, early detection, and intervention in susceptible populations. They can help policymakers and oral health professionals design evidence-based, population-specific preventive programs and allocate resources more efficiently.

7.2 Limitations

In our cross-cohort analysis, we were limited by an unequal number of samples across studies and imbalanced CF and ECC status. Despite using batch correction, the underlying differences due to sample processing, different 16S rRNA regions, and varying sequencing depth make it difficult to determine whether the observed microbiome differences among cohorts are due to actual microbial variation or technical artifacts. Furthermore, while batch correction removes batch effects, it may also reduce some of the biological effects of interest.

Although socioeconomic status was significantly associated in our analyses, it was inferred from participants' postal codes and therefore lacks individual specificity. Similarly, while we participants' current fluoride-toothpaste use was recorded, this does not indicate whether fluoride use commenced before or after the ECC onset.

Our interkingdom interactions were based on species co-occurrence and their relative abundance. This approach does not capture the functional or dynamic aspects of species interactions, which may not be directly proportional to their abundances. Furthermore, the associations identified in our cross-sectional study design do not suggest causation, which requires longitudinal study designs to establish temporal relationships and causal links.

The non-homogeneous population in this study may have diluted the effects of certain variants on ECC. However, we accounted for rural and urban residency as covariates in the association studies, capturing some population-level differences. Additionally, only the taste-related genes were selected for association tests, so not all genes implicated in tooth formation, enamel composition, salivary flow rate, and immune responses to microorganisms were evaluated. Although our sample size is considered high for microbiome dysbiosis studies, genome-wide association studies (GWAS) typically require thousands of samples for robust results, especially when adjusting for multiple covariates, as in our analysis. Moreover, integrating multi-omics data increases the number of features relative to our limited sample size, which may limit the generalizability of some findings.

The external validation of our machine learning models showed moderate performance, suggesting that microbial profiles may not generalize well across different geographic populations due to variability in microbial composition. Moreover, the limited association between our identified variants and ECC-associated microbes does not eliminate the possibility of interactions or the influence of these variants on such associations. Since our analyses relied on linear associations between variants and microbial abundance, which may not fully capture biological complexity, further research using host transcriptomics and microbial metatranscriptomics is required.

7.3 Future Directions

The findings from this thesis open avenues for future exploration, highlighting several opportunities to build upon and extend the research presented here. First and foremost, for improved comparability across studies, researchers should adopt a unified strategy guided by a consortium that adheres to best practices in studying microbiome genetics and other omics as they become increasingly accessible. For example, collecting and including specific

metadata such as age, sex, ECOHIS score, NutriSTEP scores, SES, and standardized CRA, as well as targeting specific 16S rRNA variable regions or utilizing metagenomic sequencing for strain-level resolution. We also suggest identifying ECC status based on disease severity at the time of sample collection to facilitate comparisons of microbial differences and using dmft scores assessed by calibrated dentists following the International Caries Detection and Assessment System (ICDAS) guidelines (Gormley *et al.*, 2023; Simancas-Pallares *et al.*, 2023). These consistencies can limit inter-study variability and improve real biological signals. Expanding sampling approaches to include site-specific microbiome analyses from distinct tooth surfaces could further illuminate localized microbial patterns and their contributions to ECC. Given the significant influence of socioeconomic factors on microbiome composition, we recommend using socioeconomically homogenous populations in ECC-related association studies.

Future research should prioritize longitudinal and prospective study designs, which would allow for the capture of dynamic temporal changes and help determine whether alterations in the microbiome are causes or consequences of ECC. Moreover, given the prominent role of fungi in our findings, it is crucial to expand research on fungal contributions to ECC. Currently, tools to infer functional profiles from 16S rRNA data exist, including pathways and metabolic functions; however, to the best of our knowledge, standardized tools for fungal communities are lacking. We anticipate that with the development of such databases, the data from our studies could also be utilized to explore and compare pathways of both bacterial and fungal origin.

In addition to microbial factors, the role of host genetics warrants further exploration. The inclusion of genetic data, especially the genes involved with tooth formation and immunogenic responses in addition to taste receptor genes, may help identify mechanisms through which dietary preferences and immune signaling pathways contribute to ECC. Larger sample sizes will be necessary to confidently distinguish between additive and dominant models for these associations.

Furthermore, for future studies, we recommend incorporating ethnicity information, as the validity of the observed associations can be strengthened through ancestry-specific or

stratified analyses. Integrating gene-environment studies, such as those examining host-specific dietary influences, SES, and geographical differences may enhance our understanding of the environmental contributions to ECC. Expanding analyses to include gene-microbiome interaction studies can reveal complex interdependencies that influence disease susceptibility, underscoring the need for integrative approaches to understand disease etiology. This will help in identifying population-specific effects and improving the generalizability of findings. The ECC-associated genetic markers identified in our analysis can be incorporated into an ECC-specific genetic marker panel and tested in independent cohorts to evaluate their utility across different populations. Moreover, multimodal machine learning frameworks capable of incorporating microbiome, genetic, behavioral, and sociodemographic data can aid in identifying high-impact contributors to disease and support the development of robust prediction models. Combining microbial and genetic markers into a single ECC risk score may further improve predictive accuracy.

Additionally, multiomics approaches, such as combining metatranscriptomics and metabolomics for both host and microbial communities, could provide deeper insights by capturing differences in gene expression and elucidating the effects of genetic variants on host gene expression and microbial abundance through quantitative trait loci approaches. These interactions may also extend to adult caries, as some genetic markers are shared between ECC and caries experience in adulthood.

Moreover, recent advances in deep learning methods for receptor-ligand interactions may enable the development of algorithms to study microbial ligands and their interactions with taste receptors. These findings can be validated experimentally, offering insights into the additional mechanisms underlying ECC. Finally, future studies should explore interspecies associations and microbial–taste receptor interactions using experimental designs that incorporate molecular and functional validation. Investigating these mechanisms will provide a deeper understanding of ECC pathogenesis and may uncover potential targets for the development of preventive strategies.

References

AAPD (2017) “Policy on Early Childhood Caries (ECC): Classifications, Consequences, and Preventive Strategies,” *Pediatric dentistry*, 39(6), pp. 59–61. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29179321.

AAPD (2024) “Policy on Early Childhood Caries ECC: Consequences and Preventive Strategies,” *The Reference Manual of Pediatric Dentistry: American Academy of Pediatric Dentistry*, pp. 89–92. Available at: https://www.aapd.org/globalassets/media/policies_guidelines/p_eccconsequences.pdf.

Aas, J.A. *et al.* (2005) “Defining the normal bacterial flora of the oral cavity,” *Journal of clinical microbiology*, 43(11), pp. 5721–5732. Available at: https://doi.org/10.1128/JCM.43.11.5721-5732.2005.

Aas, J.A. *et al.* (2008) “Bacteria of dental caries in primary and permanent teeth in children and young adults,” *Journal of clinical microbiology*, 46(4), pp. 1407–1417. Available at: https://doi.org/10.1128/jcm.01410-07.

Abbasoğlu, Z. *et al.* (2015) “Early childhood caries is associated with genetic variants in enamel formation and immune response genes,” *Caries research*, 49(1), pp. 70–77. Available at: https://doi.org/10.1159/000362825.

Abellan-Schneyder, I. *et al.* (2021) “Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing,” *mSphere*, 6(1). Available at: https://doi.org/10.1128/mSphere.01202-20.

Adler, C.J. *et al.* (2021) “How does the early life environment influence the oral microbiome and determine oral health outcomes in childhood?,” *BioEssays: news and reviews in molecular, cellular and developmental biology*, 43(9), p. e2000314. Available at: https://doi.org/10.1002/bies.202000314.

Agnello, M. *et al.* (2017) “Microbiome Associated with Severe Caries in Canadian First Nations Children,” *Journal of dental research*, 96(12), pp. 1378–1385. Available at: https://doi.org/10.1177/0022034517718819.

Al Mosaid, A. *et al.* (2001) “Differentiation of Candida dubliniensis from Candida albicans on staib agar and caffeic acid-ferric citrate agar,” *Journal of clinical microbiology*, 39(1), pp. 323–327. Available at: https://doi.org/10.1128/JCM.39.1.323-327.2001.

Al Shukairy, H. *et al.* (2006) “A comparative study of Streptococcus mutans and lactobacilli in mothers and children with severe early childhood caries (SECC) versus a caries free group of children and their corresponding mothers,” *The journal of clinical pediatric dentistry*, 31(2), pp. 80–85. Available at: https://doi.org/10.17796/jcpd.31.2.w855524520216761.

Al-Ahmad, A. *et al.* (2016) “Prevalence of Candida albicans and Candida dubliniensis in caries-free and caries-active children in relation to the oral microbiota-a clinical study,”

Clinical oral investigations, 20(8), pp. 1963–1971. Available at: https://doi.org/10.1007/s00784-015-1696-9.

Alam, S. *et al.* (2000) "Genotypic heterogeneity of Streptococcus oralis and distinct aciduric subpopulations in human dental plaque," *Applied and environmental microbiology*, 66(8), pp. 3330–3336. Available at: https://doi.org/10.1128/AEM.66.8.3330-3336.2000.

Almeida, E. *et al.* (2023) "Bacterial multispecies interaction mechanisms dictate biogeographic arrangement between the oral commensals Corynebacterium matruchotii and Streptococcus mitis," *mSystems*, 8(5), p. e0011523. Available at: https://doi.org/10.1128/msystems.00115-23.

Alotaibi, R.N. *et al.* (2021) "Genome-Wide Association Study (GWAS) of dental caries in diverse populations," *BMC oral health*, 21(1), p. 377. Available at: https://doi.org/10.1186/s12903-021-01670-5.

Alshehhi, A. *et al.* (2020) "Enamel defects and caries prevalence in preterm children aged 5-10 years in Dubai," *The Libyan journal of medicine*, 15(1), p. 1705633. Available at: https://doi.org/10.1080/19932820.2019.1705633.

Alyousef, Y.M. *et al.* (2023) "Oral microbiota analyses of paediatric Saudi population reveals signatures of dental caries," *BMC oral health*, 23(1), p. 935. Available at: https://doi.org/10.1186/s12903-023-03448-3.

Amalia, R. *et al.* (2022) "Regulations on nutrition in Indonesia and its relation to early childhood caries," *Frontiers in public health*, 10, p. 984668. Available at: https://doi.org/10.3389/fpubh.2022.984668.

Amato, J.N. *et al.* (2023) "Relation between caries experience and the consumption of sweetened drinks and processed food in children: A population-based study," *International journal of dental hygiene*, 21(3), pp. 561–568. Available at: https://doi.org/10.1111/idh.12662.

American Academy of Pediatric Dentistry (2003) "Definition of Early Childhood Caries (ECC)." Council on Clinical Affairs. Available at: https://www.aapd.org/assets/1/7/d_ecc.pdf.

Amin, M.S. *et al.* (2014) "Improving the oral health of young children of newcomer families: a forum for community members, researchers, and policy-makers," *Journal (Canadian Dental Association)*, 80, p. e64. Available at: https://pubmed.ncbi.nlm.nih.gov/25437943/ (Accessed: March 18, 2025).

Anderson, M. *et al.* (2021) "Development of dental caries and risk factors between 1 and 7 years of age in areas of high risk for dental caries in Stockholm, Sweden," *European archives of paediatric dentistry: official journal of the European Academy of Paediatric Dentistry*, 22(5), pp. 947–957. Available at: https://doi.org/10.1007/s40368-021-00642-1.

Anil, S. and Anand, P.S. (2017) “Early Childhood Caries: Prevalence, Risk Factors, and Prevention,” *Frontiers in pediatrics*, 5, p. 157. Available at: https://doi.org/10.3389/fped.2017.00157.

Anu, V. *et al.* (2018) “Dental caries experience, dental anomalies, and morphometric analysis of canine among monozygotic and dizygotic twins,” *Contemporary clinical dentistry*, 9(Suppl 2), pp. S314–S317. Available at: https://doi.org/10.4103/ccd.ccd_345_18.

Ardiyantini, S.D., Pramesti, H.T. and Gartika, M. (2023) “Bacterial profile in children with early childhood caries,” *Journal of Syiah Kuala Dentistry Society*, 7(2), pp. 138–149. Available at: https://doi.org/10.24815/jds.v7i2.30239.

Arid, J. *et al.* (2021) “Association of taste receptor gene polymorphisms with dental caries.” SciELO journals. Available at: https://doi.org/10.6084/M9.FIGSHARE.14287642.V1.

Armfield, J.M. *et al.* (2013) “Water fluoridation and the association of sugar-sweetened beverage consumption and dental caries in Australian children,” *American journal of public health*, 103(3), pp. 494–500. Available at: https://doi.org/10.2105/AJPH.2012.300889.

Arora, A., Schwarz, E. and Blinkhorn, A.S. (2011) “Risk factors for early childhood caries in disadvantaged populations,” *Journal of investigative and clinical dentistry*, 2(4), pp. 223–228. Available at: https://doi.org/10.1111/j.2041-1626.2011.00070.x.

Arrow, P. (2009) “Risk factors in the occurrence of enamel defects of the first permanent molars among schoolchildren in Western Australia,” *Community dentistry and oral epidemiology*, 37(5), pp. 405–415. Available at: https://doi.org/10.1111/j.1600-0528.2009.00480.x.

Aruna, P. *et al.* (2023) “Association between polymorphisms of immune response genes and early childhood caries - systematic review, gene-based, gene cluster, and meta-analysis,” *Journal, genetic engineering & biotechnology*, 21(1), p. 124. Available at: https://doi.org/10.1186/s43141-023-00566-x.

Ashi, H. *et al.* (2019) “Childhood obesity in relation to sweet taste perception and dental caries - a cross-sectional multicenter study,” *Food & nutrition research*, 63(0). Available at: https://doi.org/10.29219/fnr.v63.1682.

Baker, J.L. *et al.* (2021) “Deep metagenomics examines the oral microbiome during dental caries, revealing novel taxa and co-occurrences with host molecules,” *Genome research*, 31(1), pp. 64–74. Available at: https://doi.org/10.1101/gr.265645.120.

Balakrishnan, K. *et al.* (2021) “Importance of interkingdom interactions among oral microbiome towards caries development – A review,” *Journal of immunological sciences*, 5(2), pp. 27–35. Available at: https://doi.org/10.29245/2578-3009/2021/2.1211.

Baraniya, D. *et al.* (2020) “Supragingival mycobiome and inter-kingdom interactions in dental caries,” *Journal of oral microbiology*, 12(1), p. 1729305. Available at: https://doi.org/10.1080/20002297.2020.1729305.

Barbosa, J.O. *et al.* (2016) “Streptococcus mutans Can Modulate Biofilm Formation and Attenuate the Virulence of Candida albicans,” *PloS one*, 11(3), p. e0150457. Available at: https://doi.org/10.1371/journal.pone.0150457.

Becker, M.R. *et al.* (2002) “Molecular analysis of bacterial species associated with childhood caries,” *Journal of clinical microbiology*, 40(3), pp. 1001–1009. Available at: https://doi.org/10.1128/JCM.40.3.1001-1009.2002.

Beena, M.S. *et al.* (2017) “Comparison of Candida species isolated from children with and without early childhood caries: A descriptive cross-sectional study,” *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, 35(4), p. 296. Available at: https://doi.org/10.4103/jisppd.jisppd_160_17.

Behrens, M. and Lang, T. (2022) “Extra-oral taste receptors-function, disease, and perspectives,” *Frontiers in nutrition*, 9, p. 881177. Available at: https://doi.org/10.3389/fnut.2022.881177.

Behrens, M. and Meyerhof, W. (2011) “Gustatory and extragustatory functions of mammalian taste receptors,” *Physiology & behavior*, 105(1), pp. 4–13. Available at: https://doi.org/10.1016/j.physbeh.2011.02.010.

Bencze, Z. *et al.* (2021) “The Burden of Early Childhood Caries in Children under 5 Years Old in the European Union and Associated Risk Factors: An Ecological Study,” *Nutrients*, 13(2). Available at: https://doi.org/10.3390/nu13020455.

Bengtsson-Palme, J. *et al.* (2013) “Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data,” *Methods in Ecology and Evolution*, 4(10), pp. 914–919. Available at: https://doi.org/10.1111/2041-210X.12073.

Berg, G. *et al.* (2020) “Microbiome definition re-visited: old concepts and new challenges,” *Microbiome*, 8(1), p. 103. Available at: https://doi.org/10.1186/s40168-020-00875-0.

Bertoldo, B.B. *et al.* (2017) “Comparisons of IgA response in saliva and colostrum against oral streptococci species,” *Brazilian oral research*, 31(0), p. e39. Available at: https://doi.org/10.1590/1807-3107BOR-2017.vol31.0039.

Besnard, P., Passilly-Degrace, P. and Khan, N.A. (2016) “Taste of fat: A sixth taste modality?,” *Physiological reviews*, 96(1), pp. 151–176. Available at: https://doi.org/10.1152/physrev.00002.2015.

Bisanz, J.E. *et al.* (2019) “Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet,” *Cell host & microbe*, 26(2), pp. 265-272.e4. Available at: https://doi.org/10.1016/j.chom.2019.06.013.

Blekhman, R. *et al.* (2015) “Host genetic variation impacts microbiome composition across human body sites,” *Genome biology*, 16, p. 191. Available at: https://doi.org/10.1186/s13059-015-0759-1.

Blostein, F. *et al.* (2023) “Bacterial Community Modifies Host Genetics Effect on Early Childhood Caries,” *Journal of dental research*, 102(10), pp. 1098–1105. Available at: https://doi.org/10.1177/00220345231175356.

Bolyen, E. *et al.* (2019) “Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2,” *Nature biotechnology*, 37(8), pp. 852–857. Available at: https://doi.org/10.1038/s41587-019-0209-9.

Boustedt, K. *et al.* (2021) “Is the mode of childbirth delivery linked to the prevalence of early childhood caries? A systematic review and meta-analysis,” *European archives of paediatric dentistry: official journal of the European Academy of Paediatric Dentistry*, 22(5), pp. 765–772. Available at: https://doi.org/10.1007/s40368-021-00621-6.

Burne, R.A. and Marquis, R.E. (2000) “Alkali production by oral bacteria and protection against dental caries,” *FEMS microbiology letters*, 193(1), pp. 1–6. Available at: https://doi.org/10.1111/j.1574-6968.2000.tb09393.x.

Butcher, M.C. *et al.* (2022) “Meta-analysis of caries microbiome studies can improve upon disease prediction outcomes,” *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica* [Preprint]. Available at: https://doi.org/10.1111/apm.13272.

C. Acuña, J.E. *et al.* (2019) “Prevalence of early childhood caries in children aged 1 to 5 years in the city of Quito, Ecuador,” *The open dentistry journal*, 13(1), pp. 242–248. Available at: https://doi.org/10.2174/1874210601913010242.

Cao, Y. *et al.* (2022) “microbiomeMarker: an R/Bioconductor package for microbiome marker identification and visualization,” *Bioinformatics* , 38(16), pp. 4027–4029. Available at: https://doi.org/10.1093/bioinformatics/btac438.

Caporaso, J.G. *et al.* (2011) “Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample,” *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 1(supplement_1), pp. 4516–4522. Available at: https://doi.org/10.1073/pnas.1000080107.

Cartoni, C. *et al.* (2010) “Taste preference for fatty acids is mediated by GPR40 and GPR120,” *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 30(25), pp. 8376–8382. Available at: https://doi.org/10.1523/JNEUROSCI.0496-10.2010.

Carvalho, T.S. *et al.* (2018) "Early childhood caries and psychological perceptions on child's oral health increase the feeling of guilt in parents: an epidemiological survey," *International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children*, 28(1), pp. 23–32. Available at: https://doi.org/10.1111/ipd.12306.

Casamassimo, P.S. *et al.* (2009) "Beyond the dmft: the human and economic cost of early childhood caries," *Journal of the American Dental Association* , 140(6), pp. 650–657. Available at: https://doi.org/10.14219/jada.archive.2009.0250.

Cattaneo, C., Riso, P., *et al.* (2019) "Exploring associations between interindividual differences in taste perception, oral Microbiota composition, and reported food intake," *Nutrients*, 11(5), p. 1167. Available at: https://doi.org/10.3390/nu11051167.

Cattaneo, C., Gargari, G., *et al.* (2019) "New insights into the relationship between taste perception and oral microbiota composition," *Scientific reports*, 9(1), p. 3549. Available at: https://doi.org/10.1038/s41598-019-40374-3.

Caufield, P.W., Li, Y. and Bromage, T.G. (2012) "Hypoplasia-associated severe early childhood caries--a proposed definition," *Journal of dental research*, 91(6), pp. 544–550. Available at: https://doi.org/10.1177/0022034512444929.

Cerón-Bastidas, X.A., Suarez, A. and Guauque-Olarte, S. (2018) "Differences in caries status and risk factors among privileged and unprivileged children in Colombia," *Acta stomatologica Croatica*, 52(4), pp. 330–339. Available at: https://doi.org/10.15644/asc52/4/7.

Chandrashekar, J. *et al.* (2000) "T2Rs function as bitter taste receptors," *Cell*, 100(6), pp. 703–711. Available at: https://doi.org/10.1016/s0092-8674(00)80706-0.

Chandrashekar, J. *et al.* (2009) "The taste of carbonation," *Science (New York, N.Y.)*, 326(5951), pp. 443–445. Available at: https://doi.org/10.1126/science.1174601.

Chen, J. *et al.* (2021) "Core Microbiota Promotes the Development of Dental Caries," *NATO Advanced Science Institutes series E: Applied sciences*, 11(8), p. 3638. Available at: https://doi.org/10.3390/app11083638.

Chen, X. *et al.* (2011) "A gustotopic map of taste qualities in the mammalian brain," *Science*, 333(6047), pp. 1262–1266. Available at: https://doi.org/10.1126/science.1204076.

Chen, X. *et al.* (2020) "Isolation and Identification of Potentially Pathogenic Microorganisms Associated with Dental Caries in Human Teeth Biofilms," *Microorganisms*, 8(10). Available at: https://doi.org/10.3390/microorganisms8101596.

Cherkasov, S.V. *et al.* (2019) "Oral microbiomes in children with asthma and dental caries," *Oral diseases*, 25(3), pp. 898–910. Available at: https://doi.org/10.1111/odi.13020.

Chetty, A. and Blekhman, R. (2024) “Multi-omic approaches for host-microbiome data integration,” *Gut microbes*, 16(1), p. 2297860. Available at: https://doi.org/10.1080/19490976.2023.2297860.

Chi, D.L. *et al.* (2014) “Socioeconomic status, food security, and dental caries in US children: mediation analyses of data from the National Health and Nutrition Examination Survey, 2007-2008,” *American journal of public health*, 104(5), pp. 860–864. Available at: https://doi.org/10.2105/AJPH.2013.301699.

Chisini, L.A. *et al.* (2021) “Single nucleotide polymorphisms of taste genes and caries: a systematic review and meta-analysis,” *Acta odontologica Scandinavica*, 79(2), pp. 147–155. Available at: https://doi.org/10.1080/00016357.2020.1832253.

Cho, H. *et al.* (2023) “Selenomonas sputigena acts as a pathobiont mediating spatial structure and biofilm virulence in early childhood caries,” *Nature communications*, 14(1), p. 2919. Available at: https://doi.org/10.1038/s41467-023-38346-3.

Choi, S.W., Mak, T.S.-H. and O’Reilly, P.F. (2020) “Tutorial: a guide to performing polygenic risk score analyses,” *Nature protocols*, 15(9), pp. 2759–2772. Available at: https://doi.org/10.1038/s41596-020-0353-1.

Chou, Y.C. *et al.* (2024) “Risk of severe early childhood caries over time in low-income preschoolers,” *JDR clinical and translational research*, p. 23800844241279264. Available at: https://doi.org/10.1177/23800844241279266.

Clark, A.A., Liggett, S.B. and Munger, S.D. (2012) “Extraoral bitter taste receptors as mediators of off-target drug effects,” *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 26(12), pp. 4827–4831. Available at: https://doi.org/10.1096/fj.12-215087.

Colombo, N.H. *et al.* (2017) “Quantitative assessment of salivary oral bacteria according to the severity of dental caries in childhood,” *Archives of Oral Biology*, 83, pp. 282–288. Available at: https://doi.org/10.1016/j.archoralbio.2017.08.006.

Conti, A. *et al.* (2023) “MinION Sequencing of Yeast Mock Communities To Assess the Effect of Databases and ITS-LSU Markers on the Reliability of Metabarcoding Analysis,” *Microbiology spectrum*, 11(1), p. e0105222. Available at: https://doi.org/10.1128/spectrum.01052-22.

Corby, P.M. *et al.* (2005) “Microbial risk indicators of early childhood caries,” *Journal of clinical microbiology*, 43(11), pp. 5753–5759. Available at: https://doi.org/10.1128/jcm.43.11.5753-5759.2005.

Crystal, Y.O. *et al.* (2018) “Use of silver diamine fluoride for dental caries management in children and adolescents, including those with special health care needs,” *Journal of the California Dental Association*, 46(1), pp. 45–55. Available at: https://doi.org/10.1080/19424396.2018.12221981.

Cui, Y. *et al.* (2021) “Oral mycobiome differences in various spatial niches with and without severe early childhood caries,” *Frontiers in pediatrics*, 9, p. 748656. Available at: https://doi.org/10.3389/fped.2021.748656.

Cvanova, M. *et al.* (2022) “Candida species and selected behavioral factors co-associated with severe early childhood caries: Case-control study,” *Frontiers in cellular and infection microbiology*, 12, p. 943480. Available at: https://doi.org/10.3389/fcimb.2022.943480.

Davidson, K. *et al.* (2016) “Higher body mass index associated with severe early childhood caries,” *BMC pediatrics*, 16, p. 137. Available at: https://doi.org/10.1186/s12887-016-0679-6.

Demmitt, B.A. *et al.* (2017) “Genetic influences on the human oral microbiome,” *BMC genomics*, 18(1), p. 659. Available at: https://doi.org/10.1186/s12864-017-4008-8.

Deng, L. *et al.* (2019) “Cross-kingdom interaction of Candida albicans and Actinomyces viscosus elevated cariogenic virulence,” *Archives of oral biology*, 100, pp. 106–112. Available at: https://doi.org/10.1016/j.archoralbio.2019.02.008.

Dettori, M. *et al.* (2024) “Impact of socioeconomic inequalities on dental caries status in Sardinian children,” *Children (Basel, Switzerland)*, 11(1), p. 96. Available at: https://doi.org/10.3390/children11010096.

Dewhirst, F.E. *et al.* (2010) “The human oral microbiome,” *Journal of bacteriology*, 192(19), pp. 5002–5017. Available at: https://doi.org/10.1128/JB.00542-10.

Diaz, P.I. *et al.* (2012) “Synergistic interaction between Candida albicans and commensal oral streptococci in a novel in vitro mucosal model,” *Infection and immunity*, 80(2), pp. 620–632. Available at: https://doi.org/10.1128/IAI.05896-11.

Diaz, P.I. *et al.* (2019) “Integrated analysis of clinical and microbiome risk factors associated with the development of oral candidiasis during cancer chemotherapy,” *Journal of fungi (Basel, Switzerland)*, 5(2), p. 49. Available at: https://doi.org/10.3390/jof5020049.

Diaz, P.I. and Dongari-Bagtzoglou, A. (2021) “Critically Appraising the Significance of the Oral Mycobiome,” *Journal of dental research*, 100(2), pp. 133–140. Available at: https://doi.org/10.1177/0022034520956975.

Dige, I. and Nyvad, B. (2019) “Candida species in intact in vivo biofilm from carious lesions,” *Archives of oral biology*, 101, pp. 142–146. Available at: https://doi.org/10.1016/j.archoralbio.2019.03.017.

Divaris, K. *et al.* (2019) “The Supragingival Biofilm in Early Childhood Caries: Clinical and Laboratory Protocols and Bioinformatics Pipelines Supporting Metagenomics, Metatranscriptomics, and Metabolomics Studies of the Oral Microbiome,” *Methods in molecular biology (Clifton, N.J.)*, 1922, pp. 525–548. Available at: https://doi.org/10.1007/978-1-4939-9012-2_40.

Dos Anjos, A.M.C. *et al.* (2023) "Is there an association between dental caries and genetics? Systematic review and meta-analysis of studies with twins," *Journal of dentistry*, 135(104586), p. 104586. Available at: https://doi.org/10.1016/j.jdent.2023.104586.

Downes, J. *et al.* (2013) "Description of Alloprevotella rava gen. nov., sp. nov., isolated from the human oral cavity, and reclassification of Prevotella tannerae Moore et al. 1994 as Alloprevotella tannerae gen. nov., comb. nov," *International journal of systematic and evolutionary microbiology*, 63(Pt 4), pp. 1214–1218. Available at: https://doi.org/10.1099/ijs.0.041376-0.

Du, Q. *et al.* (2022) "Cross-kingdom interaction between Candida albicans and oral bacteria," *Frontiers in microbiology*, 13, p. 911623. Available at: https://doi.org/10.3389/fmicb.2022.911623.

Dzidic, M. *et al.* (2018) "Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay," *The ISME journal*, 12(9), pp. 2292–2306. Available at: https://doi.org/10.1038/s41396-018-0204-z.

Ebbert, M.T.W. *et al.* (2016) "Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches," *BMC bioinformatics*, 17 Suppl 7(S7), p. 239. Available at: https://doi.org/10.1186/s12859-016-1097-3.

Edmonds, B., Williams, T. and Carrico, C. (2019) "The prevalence and factors associated with sibling-recurrent dental treatment under general anesthesia at an academic institution," *Pediatric dentistry*, 41(1), pp. 40–46. Available at: https://www.ingentaconnect.com/content/aapd/pd/2019/00000041/00000001/art00006 (Accessed: March 13, 2025).

El Tantawi, M. *et al.* (2018) "Prevalence and Data Availability of Early Childhood Caries in 193 United Nations Countries, 2007-2017," *American journal of public health*, 108(8), pp. 1066–1072. Available at: https://doi.org/10.2105/AJPH.2018.304466.

El Tantawi, M. *et al.* (2024) "A scoping review of early childhood caries, poverty and the first sustainable development goal," *BMC oral health*, 24(1), p. 1029. Available at: https://doi.org/10.1186/s12903-024-04790-w.

Eriksson, L. *et al.* (2018) "Microbial Complexes and Caries in 17-Year-Olds with and without Streptococcus mutans," *Journal of dental research*, 97(3), pp. 275–282. Available at: https://doi.org/10.1177/0022034517731758.

Esberg, A. *et al.* (2020) "Oral Microbiota profile associates with sugar intake and taste preference genes," *Nutrients*, 12(3), p. 681. Available at: https://doi.org/10.3390/nu12030681.

Escapa, I.F. *et al.* (2018) "New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract," *mSystems*, 3(6). Available at: https://doi.org/10.1128/mSystems.00187-18.

Fakhruddin, K.S. *et al.* (2022) “Diversity of site-specific microbes of occlusal and proximal lesions in severe- early childhood caries (S-ECC),” *Journal of oral microbiology*, 14(1), p. 2037832. Available at: https://doi.org/10.1080/20002297.2022.2037832.

Fakhruddin, K.S., Ngo, H.C. and Samaranayake, L.P. (2019) “Cariogenic microbiome and microbiota of the early primary dentition: A contemporary overview,” *Oral diseases*, 25(4), pp. 982–995. Available at: https://doi.org/10.1111/odi.12932.

Fechney, J.M. *et al.* (2019) “Preliminary study of the oral mycobiome of children with and without dental caries,” *Journal of oral microbiology*, 11(1), p. 1536182. Available at: https://doi.org/10.1080/20002297.2018.1536182.

Feldens, C.A. *et al.* (2018) “Feeding frequency in infancy and dental caries in childhood: a prospective cohort study,” *International dental journal*, 68(2), pp. 113–121. Available at: https://doi.org/10.1111/idj.12333.

Fernandes, J. de K.B. *et al.* (2023) “Small for gestational age and early childhood caries: the BRISA cohort study,” *Scientific reports*, 13(1), p. 14343. Available at: https://doi.org/10.1038/s41598-023-41411-y.

Finlayson, T.L. *et al.* (2018) “Dental service utilization among children in the child welfare system,” *Maternal and child health journal*, 22(5), pp. 753–761. Available at: https://doi.org/10.1007/s10995-018-2444-y.

Fisher-Owens, S.A. *et al.* (2007) “Influences on children’s oral health: a conceptual model,” *Pediatrics*, 120(3), pp. e510-20. Available at: https://doi.org/10.1542/peds.2006-3084.

Folayan, M.O. *et al.* (2020) “Association between early childhood caries and poverty in low and middle income countries,” *BMC oral health*, 20(1), p. 8. Available at: https://doi.org/10.1186/s12903-019-0997-9.

Folayan, M.O. *et al.* (2024) “Association between early childhood caries and parental education and the link to the sustainable development goal 4: a scoping review,” *BMC oral health*, 24(1), p. 517. Available at: https://doi.org/10.1186/s12903-024-04291-w.

Fransoo, R. *et al.* (2013) *The 2013 RHA Indicators Atlas*. Winnipeg, MB: Manitoba Centre for Health Policy.

Freire, M. *et al.* (2020) “Longitudinal study of oral microbiome variation in twins,” *Scientific reports*, 10(1), p. 7954. Available at: https://doi.org/10.1038/s41598-020-64747-1.

Fu, Y. *et al.* (2018) “Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers,” *BMC genomics*, 19(1). Available at: https://doi.org/10.1186/s12864-018-4933-1.

Ganesh, A. *et al.* (2020) “Risk Factors for Early Childhood Caries in Toddlers: An Institution-based Study,” *Cureus*, 12(4), p. e7516. Available at: https://doi.org/10.7759/cureus.7516.

GBD Disease and Injury Incidence and Prevalence Collaborators (2018) “Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017,” *Lancet*, 392(10159), pp. 1789–1858. Available at: https://doi.org/10.1016/S0140-6736(18)32279-7.

Ghannoum, M.A. *et al.* (2010) “Characterization of the oral fungal microbiome (mycobiome) in healthy individuals,” *PLoS pathogens*, 6(1), p. e1000713. Available at: https://doi.org/10.1371/journal.ppat.1000713.

Gloor, G.B. *et al.* (2017) “Microbiome Datasets Are Compositional: And This Is Not Optional,” *Frontiers in microbiology*, 8, p. 2224. Available at: https://doi.org/10.3389/fmicb.2017.02224.

Gomez, A. *et al.* (2017) “Host Genetic Control of the Oral Microbiome in Health and Disease,” *Cell host & microbe*, 22(3), pp. 269-278.e3. Available at: https://doi.org/10.1016/j.chom.2017.08.013.

Gonia, S. *et al.* (2015) “Human milk oligosaccharides inhibit candida albicans invasion of human premature intestinal epithelial cells,” *The journal of nutrition*, 145(9), pp. 1992–1998. Available at: https://doi.org/10.3945/jn.115.214940.

Gormley, A. *et al.* (2023) “Subtypes of early childhood caries predict future caries experience,” *Community dentistry and oral epidemiology*, 51(5), pp. 966–975. Available at: https://doi.org/10.1111/cdoe.12795.

Grant, C.G. *et al.* (2019) “Oral health-related quality of life of Canadian preschoolers with severe caries after dental rehabilitation under general anesthesia,” *Pediatric dentistry*, 41(3), pp. 221–228. Available at: https://www.ingentaconnect.com/content/aapd/pd/2019/00000041/00000003/art00010 (Accessed: March 13, 2025).

Grier, A. *et al.* (2020) “Oral Microbiota Composition Predicts Early Childhood Caries Onset,” *Journal of dental research*, p. 22034520979926. Available at: https://doi.org/10.1177/0022034520979926.

Gross, E.L. *et al.* (2012) “Beyond Streptococcus mutans: dental caries onset linked to multiple species by 16S rRNA community analysis,” *PloS one*, 7(10), p. e47722. Available at: https://doi.org/10.1371/journal.pone.0047722.

Ha, D.H. *et al.* (2020) “The impact of different determinants on the dental caries experience of children living in Australia rural and urban areas,” *Community dentistry and oral epidemiology*, 49(4), pp. 337–345. Available at: https://doi.org/10.1111/cdoe.12606.

Hafiz, Z. *et al.* (2022) “Effectiveness of silver diamine fluoride in arresting caries in primary and early mixed dentition: A systematic review,” *Children (Basel, Switzerland)*, 9(9), p. 1289. Available at: https://doi.org/10.3390/children9091289.

Handsley-Davis, M. *et al.* (2021) “Biocultural Drivers of Salivary Microbiota in Australian Aboriginal and Torres Strait Islander Children,” *Frontiers in oral health*, 2, p. 641328. Available at: https://doi.org/10.3389/froh.2021.641328.

Hashim, A. *et al.* (2021) “Loss of neutrophil homing to the periodontal tissues modulates the composition and disease potential of the oral Microbiota,” *Infection and immunity*, 89(12), p. e0030921. Available at: https://doi.org/10.1128/IAI.00309-21.

Haworth, S. *et al.* (2018) “Consortium-based genome-wide meta-analysis for childhood dental caries traits,” *Human molecular genetics*, 27(17), pp. 3113–3127. Available at: https://doi.org/10.1093/hmg/ddy237.

Haworth, S. *et al.* (2020) “Heritability of Caries Scores, Trajectories, and Disease Subtypes,” *Journal of dental research*, 99(3), pp. 264–270. Available at: https://doi.org/10.1177/0022034519897910.

Haznedaroğlu, E. *et al.* (2015) “Association of sweet taste receptor gene polymorphisms with dental caries experience in school children,” *Caries research*, 49(3), pp. 275–281. Available at: https://doi.org/10.1159/000381426.

He, J. *et al.* (2017) “RNA-Seq Reveals Enhanced Sugar Metabolism in Streptococcus mutans Co-cultured with Candida albicans within Mixed-Species Biofilms,” *Frontiers in microbiology*, 8, p. 1036. Available at: https://doi.org/10.3389/fmicb.2017.01036.

He, J. *et al.* (2018) “Taxonomic and Functional Analyses of the Supragingival Microbiome from Caries-Affected and Caries-Free Hosts,” *Microbial ecology*, 75(2), pp. 543–554. Available at: https://doi.org/10.1007/s00248-017-1056-1.

Hibberd, C.E. and Nguyen, T.D. (2012) “Brain abscess secondary to a dental infection in an 11-year-old child: case report,” *Journal (Canadian Dental Association)*, 78, p. c49. Available at: https://pubmed.ncbi.nlm.nih.gov/22558953/ (Accessed: March 13, 2025).

Ho, N.T. *et al.* (2019) “metamicrobiomeR: an R package for analysis of microbiome relative abundance data using zero-inflated beta GAMLSS and meta-analysis across studies using random effects models,” *BMC bioinformatics*, 20(1), p. 188. Available at: https://doi.org/10.1186/s12859-019-2744-2.

Holla, L.I. *et al.* (2015) “GLUT2 and TAS1R2 polymorphisms and susceptibility to dental caries,” *Caries research*, 49(4), pp. 417–424. Available at: https://doi.org/10.1159/000430958.

Hooley, M. *et al.* (2012) “Parental influence and the development of dental caries in children aged 0-6 years: a systematic review of the literature,” *Journal of dentistry*, 40(11), pp. 873–885. Available at: https://doi.org/10.1016/j.jdent.2012.07.013.

Hu, S. *et al.* (2019) “Infant dietary patterns and early childhood caries in a multi-ethnic Asian cohort,” *Scientific reports*, 9(1), p. 852. Available at: https://doi.org/10.1038/s41598-018-37183-5.

Humphrey, S.P. and Williamson, R.T. (2001) “A review of saliva: normal composition, flow, and function,” *The journal of prosthetic dentistry*, 85(2), pp. 162–169. Available at: https://doi.org/10.1067/mpr.2001.113778.

Hung, M.-C. and Christodoulides, M. (2013) “The biology of Neisseria adhesins,” *Biology*, 2(3), pp. 1054–1109. Available at: https://doi.org/10.3390/biology2031054.

Hussein, Y.A. *et al.* (2017) “Differences between children with early childhood caries and caries free children in oral health habits: A case control study,” *Alexandria dental journal: ADJ*, 42(1), pp. 28–32. Available at: https://doi.org/10.21608/adjalexu.2017.57853.

Iheozor-Ejiofor, Z. *et al.* (2024) “Water fluoridation for the prevention of dental caries,” *Cochrane database of systematic reviews*, 10(11), p. CD010856. Available at: https://doi.org/10.1002/14651858.CD010856.pub3.

Imes, E.P. *et al.* (2021) “Guardian Reports of Children’s Sub-optimal Oral Health Are Associated With Clinically Determined Early Childhood Caries, Unrestored Caries Lesions, and History of Toothaches,” *Frontiers in public health*, 9, p. 751733. Available at: https://doi.org/10.3389/fpubh.2021.751733.

Jasim Al-Adool, R., Muwafaq Hamdoon, S. and Faiz Hamodat, H. (2024) “Scardovia wiggsiae as recent player in caries process: A review,” *Mustansiria Dental Journal*, 19(1), pp. 117–123. Available at: https://doi.org/10.32828/mdj.v19i1.998.

de Jesus, V.C. *et al.* (2020) “Sex-Based Diverse Plaque Microbiota in Children with Severe Caries,” *Journal of dental research*, 99(6), pp. 703–712. Available at: https://doi.org/10.1177/0022034520908595.

de Jesus, V.C., Singh, M., *et al.* (2021) “Association of bitter taste receptor T2R38 polymorphisms, oral Microbiota, and rheumatoid arthritis,” *Current issues in molecular biology*, 43(3), pp. 1460–1472. Available at: https://doi.org/10.3390/cimb43030103.

de Jesus, V.C., Khan, M.W., *et al.* (2021) “Characterization of Supragingival Plaque and Oral Swab Microbiomes in Children With Severe Early Childhood Caries,” *Frontiers in microbiology*, 12, p. 1517. Available at: https://doi.org/10.3389/fmicb.2021.683685.

de Jesus, V.C. *et al.* (2022) “Genetic variants in taste genes play a role in oral microbial composition and severe early childhood caries,” *iScience*, 25(12), p. 105489. Available at: https://doi.org/10.1016/j.isci.2022.105489.

Jiang, Q. *et al.* (2018) “The Oral Microbiome in the Elderly With Dental Caries and Health,” *Frontiers in cellular and infection microbiology*, 8, p. 442. Available at: https://doi.org/10.3389/fcimb.2018.00442.

Jiang, R., Li, W.V. and Li, J.J. (2021) “mbImpute: an accurate and robust imputation method for microbiome data,” *Genome biology*, 22(1), p. 192. Available at: https://doi.org/10.1186/s13059-021-02400-4.

Jiang, S. *et al.* (2016) “Salivary Microbiome Diversity in Caries-Free and Caries-Affected Children,” *International journal of molecular sciences*, 17(12). Available at: https://doi.org/10.3390/ijms17121978.

Jiang, W. *et al.* (2014) “Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood,” *Microbial ecology*, 67(4), pp. 962–969. Available at: https://doi.org/10.1007/s00248-014-0372-y.

Jiang, W., Zhang, J. and Chen, H. (2013) “Pyrosequencing analysis of oral microbiota in children with severe early childhood dental caries,” *Current microbiology*, 67(5), pp. 537–542. Available at: https://doi.org/10.1007/s00284-013-0393-7.

Jiang, Y. *et al.* (2012) “Meta-analysis of 125 rheumatoid arthritis-related single nucleotide polymorphisms studied in the past two decades,” *PloS one*, 7(12), p. e51571. Available at: https://doi.org/10.1371/journal.pone.0051571.

Jordan, A.R. *et al.* (2016) “Early Childhood Caries and Caries Experience in Permanent Dentition: A 15-year Cohort Study,” *Swiss dental journal*, 126(2), pp. 114–119. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26915485.

Jurczak, A. *et al.* (2020) “Differences in sweet taste perception and its association with the streptococcus mutans cariogenic profile in preschool children with caries,” *Nutrients*, 12(9), p. 2592. Available at: https://doi.org/10.3390/nu12092592.

Kachroo, N. *et al.* (2021) “Meta-analysis of clinical microbiome studies in urolithiasis reveal age, stone composition, and study location as the predominant factors in urolithiasis-associated microbiome composition,” *mBio*, 12(4), p. e0200721. Available at: https://doi.org/10.1128/mBio.02007-21.

Kahharova, D. *et al.* (2023) “Microbial Indicators of Dental Health, Dysbiosis, and Early Childhood Caries,” *Journal of dental research*, 102(7), pp. 759–766. Available at: https://doi.org/10.1177/00220345231160756.

Kalpana, B. *et al.* (2020) “Bacterial diversity and functional analysis of severe early childhood caries and recurrence in India,” *Scientific reports*, 10(1), p. 21248. Available at: https://doi.org/10.1038/s41598-020-78057-z.

Kammoun, R. *et al.* (2019) “Dental caries and hypoplastic amelogenesis imperfecta: Clinical, structural, biochemical and molecular approaches,” *Microbial pathogenesis*, 135(103615), p. 103615. Available at: https://doi.org/10.1016/j.micpath.2019.103615.

Kanasi, E. *et al.* (2010) “Clonal analysis of the microbiota of severe early childhood caries,” *Caries research*, 44(5), pp. 485–497. Available at: https://doi.org/10.1159/000320158.

Kandel, E.A., Richards, J.M. and Binkley, C.J. (2012) “Childhood caries in the state of Kentucky, USA: a cross-sectional study,” *BMC oral health*, 12(1), p. 38. Available at: https://doi.org/10.1186/1472-6831-12-38.

Karczewski, K.J. *et al.* (2020) “The mutational constraint spectrum quantified from variation in 141,456 humans,” *Nature*, 581(7809), pp. 434–443. Available at: https://doi.org/10.1038/s41586-020-2308-7.

Karkowska-Kuleta, J. *et al.* (2022) “Fungi-A component of the oral microbiome involved in periodontal diseases,” *Advances in experimental medicine and biology*, 1373, pp. 113–138. Available at: https://doi.org/10.1007/978-3-030-96881-6_6.

Keijser, B.J.F. *et al.* (2008) “Pyrosequencing analysis of the oral microflora of healthy adults,” *Journal of dental research*, 87(11), pp. 1016–1020. Available at: https://doi.org/10.1177/154405910808701104.

Kelly, B.J. *et al.* (2015) “Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA,” *Bioinformatics* , 31(15), pp. 2461–2468. Available at: https://doi.org/10.1093/bioinformatics/btv183.

Khan, M.W., Fung, D.L.X., *et al.* (2024) “A cross-cohort analysis of dental plaque microbiome in early childhood caries,” *iScience*, 27(8), p. 110447. Available at: https://doi.org/10.1016/j.isci.2024.110447.

Khan, M.W., de Jesus, V.C., *et al.* (2024) “Role of socioeconomic factors and interkingdom crosstalk in the dental plaque microbiome in early childhood caries,” *Cell reports*, 43(8), p. 114635. Available at: https://doi.org/10.1016/j.celrep.2024.114635.

Khan, S.Y. *et al.* (2023) “A systematic review of caries risk in children <6 years of age,” *International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children* [Preprint]. Available at: https://doi.org/10.1111/ipd.13140.

Khoroushi, M. and Kachuie, M. (2017) “Prevention and treatment of White Spot lesions in orthodontic patients,” *Contemporary clinical dentistry*, 8(1), pp. 11–19. Available at: https://doi.org/10.4103/ccd.ccd_216_17.

Kim, H.E. *et al.* (2021) “Intervening in Symbiotic Cross-Kingdom Biofilm Interactions: a Binding Mechanism-Based Nonmicrobicidal Approach,” *mBio*, 12(3). Available at: https://doi.org/10.1128/mBio.00651-21.

Klein, M.I. *et al.* (2015) “Streptococcus mutans-derived extracellular matrix in cariogenic oral biofilms,” *Frontiers in cellular and infection microbiology*, 5, p. 10. Available at: https://doi.org/10.3389/fcimb.2015.00010.

Kneist, S. *et al.* (2015) “First-time isolation of Candida dubliniensis from plaque and carious dentine of primary teeth,” *European archives of paediatric dentistry: official journal of the European Academy of Paediatric Dentistry*, 16(4), pp. 365–370. Available at: https://doi.org/10.1007/s40368-015-0180-1.

Knight, R. *et al.* (2018) “Best practices for analysing microbiomes,” *Nature reviews. Microbiology*, 16(7), pp. 410–422. Available at: https://doi.org/10.1038/s41579-018-0029-9.

Knoblauch, U. *et al.* (2019) “The association between socioeconomic status, psychopathological symptom burden in mothers, and early childhood caries of their children,” *PloS one*, 14(10), p. e0224509. Available at: https://doi.org/10.1371/journal.pone.0224509.

Kolde, R. *et al.* (2018) “Host genetic variation and its microbiome interactions within the Human Microbiome Project,” *Genome medicine*, 10(1), p. 6. Available at: https://doi.org/10.1186/s13073-018-0515-8.

Konopka, A. (2009) “What is microbial community ecology?,” *The ISME journal*, 3(11), pp. 1223–1230. Available at: https://doi.org/10.1038/ismej.2009.88.

Kotian, N. and Gurunathan, D. (2021) “Estimation of salivary ferritin in children with childhood caries – a cross sectional study,” *Brazilian dental science*, 24(2). Available at: https://doi.org/10.14295/bds.2021.v24i2.2143.

Kouakou, Y.I. and Lee, R.J. (2023) “Interkingdom detection of bacterial quorum-sensing molecules by mammalian taste receptors,” *Microorganisms*, 11(5), p. 1295. Available at: https://doi.org/10.3390/microorganisms11051295.

Krasuska-Sławińska, E. *et al.* (2016) “Antineoplastic chemotherapy and congenital tooth abnormalities in children and adolescents,” *Contemporary oncology (Poznan, Poland)*, 20(5), pp. 394–401. Available at: https://doi.org/10.5114/wo.2016.64602.

Kressirer, C.A. *et al.* (2017) “Scardovia wiggsiae and its potential role as a caries pathogen,” *Journal of oral biosciences*, 59(3), pp. 135–141. Available at: https://doi.org/10.1016/j.job.2017.05.002.

Krüger, W. *et al.* (2019) “Fungal-Bacterial Interactions in Health and Disease,” *Pathogens*, 8(2). Available at: https://doi.org/10.3390/pathogens8020070.

Krzyściak, W. *et al.* (2017) “Effect of a Lactobacillus Salivarius Probiotic on a Double-Species Streptococcus Mutans and Candida Albicans Caries Biofilm,” *Nutrients*, 9(11). Available at: https://doi.org/10.3390/nu9111242.

Kulkarni, G.V. *et al.* (2013) “Association of GLUT2 and TAS1R2 genotypes with risk for dental caries,” *Caries research*, 47(3), pp. 219–225. Available at: https://doi.org/10.1159/000345652.

Kuppan, A. *et al.* (2017) “Prevalence and heritability of early childhood caries among monozygotic and dizygotic twins,” *Twin research and human genetics: the official journal of the International Society for Twin Studies*, 20(1), pp. 43–52. Available at: https://doi.org/10.1017/thg.2016.96.

Kurita, G.P. *et al.* (2016) “Genetic variation and cognitive dysfunction in opioid-treated patients with cancer,” *Brain and behavior*, 6(7), p. e00471. Available at: https://doi.org/10.1002/brb3.471.

Lara, J.S. *et al.* (2022) “Impact of early childhood caries severity on oral health-related quality of life among preschool children in Mexico: A cross-sectional study,” *International journal of paediatric dentistry*, 32(3), pp. 334–343. Available at: https://doi.org/10.1111/ipd.12889.

Lederberg, J. and McCray, A.T. (2001) “Ome SweetOmics--A genealogical treasury of words,” *The scientist*, 15(7), pp. 8–8. Available at: https://go.gale.com/ps/i.do?id=GALE%7CA73535513&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=08903670&p=.

Lee, E. *et al.* (2021) “Microbiome of Saliva and Plaque in Children According to Age and Dental Caries Experience,” *Diagnostics (Basel, Switzerland)*, 11(8). Available at: https://doi.org/10.3390/diagnostics11081324.

Lee, G.H.M. *et al.* (2009) “Translation and validation of a Chinese language version of the Early Childhood Oral Health Impact Scale (ECOHIS),” *International journal of paediatric dentistry*, 19(6), pp. 399–405. Available at: https://doi.org/10.1111/j.1365-263X.2009.01000.x.

Lee, J. *et al.* (2022) “Oral Health Status and Oral Health-Related Quality of Life of First Nations and Metis Children,” *JDR clinical and translational research*, 7(4), pp. 435–445. Available at: https://doi.org/10.1177/23800844211037992.

Lee, J.Y. and Divaris, K. (2014) “The ethical imperative of addressing oral health disparities: a unifying framework: A unifying framework,” *Journal of dental research*, 93(3), pp. 224–230. Available at: https://doi.org/10.1177/0022034513511821.

Lee, R.J. *et al.* (2012) “T2R38 taste receptor polymorphisms underlie susceptibility to upper respiratory infection,” *The journal of clinical investigation*, 122(11), pp. 4145–4159. Available at: https://doi.org/10.1172/JCI64240.

Lee, S. *et al.* (2012) “Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies,” *The American Journal of Human Genetics*, 91(2), pp. 224–237. Available at: https://doi.org/10.1016/j.ajhg.2012.06.007.

Lee, V.H.K. *et al.* (2020) “Association between early childhood oral health impact scale (ECOHIS) scores and pediatric dental surgery wait times,” *BMC oral health*, 20(1), p. 285. Available at: https://doi.org/10.1186/s12903-020-01263-8.

Lee, V.H.K. *et al.* (2025) “Analysis of Nutritional Status and Oral Veillonella With Caries Status in Preschool Children in Manitoba,” *Pediatric Dentistry*, 47(1), pp. 41–49. Available at: https://mspace.lib.umanitoba.ca/handle/1993/36525.

Li, C. *et al.* (2024) “Taste and its receptors in human physiology: A comprehensive look,” *Food frontiers*, 5(4), pp. 1512–1533. Available at: https://doi.org/10.1002/fft2.407.

Li, H. and Durbin, R. (2009) “Fast and accurate short read alignment with Burrows-Wheeler transform,” *Bioinformatics (Oxford, England)*, 25(14), pp. 1754–1760. Available at: https://doi.org/10.1093/bioinformatics/btp324.

Li, Q. *et al.* (2022) “Corynebacterium matruchotii: A confirmed calcifying bacterium with a potentially important role in the supragingival plaque,” *Frontiers in microbiology*, 13, p. 940643. Available at: https://doi.org/10.3389/fmicb.2022.940643.

Li, R.Y. *et al.* (2023) “Analysis of risk factors associated with early childhood caries,” *Risk management and healthcare policy*, 16, pp. 2369–2378. Available at: https://doi.org/10.2147/RMHP.S426712.

Li, Y. *et al.* (2007) “Genetic profiling of the oral microbiota associated with severe early-childhood caries,” *Journal of clinical microbiology*, 45(1), pp. 81–87. Available at: https://doi.org/10.1128/JCM.01622-06.

Limeta, A. *et al.* (2020) “Meta-analysis of the gut microbiota in predicting response to cancer immunotherapy in metastatic melanoma,” *JCI insight*, 5(23). Available at: https://doi.org/10.1172/jci.insight.140940.

Lin, B., Wang, J. and Zhang, Y. (2024) “Bacterial dynamics in the progression of caries to apical periodontitis in primary teeth of children with severe early childhood caries,” *Frontiers in microbiology*, 15, p. 1418261. Available at: https://doi.org/10.3389/fmicb.2024.1418261.

Lips, A. *et al.* (2017) “Salivary protein polymorphisms and risk of dental caries: a systematic review,” *Brazilian oral research*, 31(0), p. e41. Available at: https://doi.org/10.1590/1807-3107BOR-2017.vol31.0041.

Liu, X. *et al.* (2021) “Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome,” *Cell discovery*, 7(1), p. 117. Available at: https://doi.org/10.1038/s41421-021-00356-0.

Liu, Y. *et al.* (2023) “Addition of cariogenic pathogens to complex oral microflora drives significant changes in biofilm compositions and functionalities,” *Microbiome*, 11(1), p. 123. Available at: https://doi.org/10.1186/s40168-023-01561-7.

Lopes, P.C. *et al.* (2024) “White spot lesions: diagnosis and treatment - a systematic review,” *BMC oral health*, 24(1), p. 58. Available at: https://doi.org/10.1186/s12903-023-03720-6.

Lovelina, F.D., Shastri, S.M. and Kumar, P.D.M. (2012) “Assessment of the oral health status of monozygotic and dizygotic twins - A comparative study,” *Oral Health and Preventive Dentistry*, 10(2), pp. 135–139. Available at: https://doi.org/10.3290/j.ohpd.a28000.

Lozupone, C.A. *et al.* (2013) “Meta-analyses of studies of the human microbiota,” *Genome research*, 23(10), pp. 1704–1714. Available at: https://doi.org/10.1101/gr.151803.112.

Lu, Y. *et al.* (2023) “Roles of Streptococcus mutans-Candida albicans interaction in early childhood caries: a literature review,” *Frontiers in cellular and infection microbiology*, 13, p. 1151532. Available at: https://doi.org/10.3389/fcimb.2023.1151532.

Lui, D.-T. *et al.* (2023) “Association of early childhood caries and nutritional status: a scoping review,” *The journal of clinical pediatric dentistry*, 47(3), pp. 11–25. Available at: https://doi.org/10.22514/jocpd.2023.021.

Lukacs, J.R. (2008) “Fertility and agriculture accentuate sex differences in dental caries rates,” *Current anthropology*, 49(5), pp. 901–914. Available at: https://doi.org/10.1086/592111.

Lukacs, J.R. (2011) “Sex differences in dental caries experience: clinical evidence, complex etiology,” *Clinical oral investigations*, 15(5), pp. 649–656. Available at: https://doi.org/10.1007/s00784-010-0445-3.

Luo, J. *et al.* (2019) “The role of GPCRs in bone diseases and dysfunctions,” *Bone research*, 7(1), p. 19. Available at: https://doi.org/10.1038/s41413-019-0059-6.

Macfarlane, T.W. (1990) “Ecology and epidemiology of Candida,” in L.P. Samaranayake and T.W. Macfarlane (eds.) *Oral candidosis*. London, UK: Butterworth and Co., pp. 21–46. Available at: https://www.cabidigitallibrary.org/doi/full/10.5555/19911207985 (Accessed: March 16, 2025).

Mafuvadze, B.T., Mahachi, L. and Mafuvadze, B. (2013) “Dental caries and oral health practice among 12 year old school children from low socio-economic status background in Zimbabwe,” *The Pan African medical journal*, 14(164), p. 164. Available at: https://doi.org/10.11604/pamj.2013.14.164.2399.

Maklennan, A. *et al.* (2024) “A systematic review and meta-analysis on early-childhood-caries global data,” *BMC oral health*, 24(1), p. 835. Available at: https://doi.org/10.1186/s12903-024-04605-y.

Mallick, H. *et al.* (2021) “Multivariable association discovery in population-scale meta-omics studies,” *PLoS computational biology*, 17(11), p. e1009442. Available at: https://doi.org/10.1371/journal.pcbi.1009442.

Mann, A.E. *et al.* (2023) “Impact of HIV on the Oral Microbiome of Children Living in Sub-Saharan Africa, Determined by Using an rpoC Gene Fragment Metataxonomic Approach,” *Microbiology spectrum*, 11(4), p. e0087123. Available at: https://doi.org/10.1128/spectrum.00871-23.

Marchant, S. *et al.* (2001) “The predominant microflora of nursing caries lesions,” *Caries research*, 35(6), pp. 397–406. Available at: https://doi.org/10.1159/000047482.

Marchesi, J.R. and Ravel, J. (2015) “The vocabulary of microbiome research: a proposal,” *Microbiome*, 3(1), p. 31. Available at: https://doi.org/10.1186/s40168-015-0094-5.

Marcos-Zambrano, L.J. *et al.* (2021) “Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment,” *Frontiers in microbiology*, 12, p. 634511. Available at: https://doi.org/10.3389/fmicb.2021.634511.

Marsh, P.D. (1994) “Microbial ecology of dental plaque and its significance in health and disease,” *Advances in dental research*, 8(2), pp. 263–271. Available at: https://doi.org/10.1177/08959374940080022001.

Marsh, P.D. (2006) “Dental plaque as a biofilm and a microbial community - implications for health and disease,” *BMC oral health*, 6 Suppl 1(Suppl 1), p. S14. Available at: https://doi.org/10.1186/1472-6831-6-S1-S14.

Marsh, P.D. and Nyvad, B. (2008) “The oral microflora and biofilms on teeth,” in O. Fejerskov and E.A.M. Kidd (eds.) *Dental Caries*. 2nd ed. Blackwell Publishing, pp. 163–187.

Marsh, P.D. and Zaura, E. (2017) “Dental biofilm: ecological interactions in health and disease,” *Journal of clinical periodontology*, 44 Suppl 18, pp. S12–S22. Available at: https://doi.org/10.1111/jcpe.12679.

McDonagh, M.S. *et al.* (2000) “Systematic review of water fluoridation,” *BMJ (Clinical research ed.)*, 321(7265), pp. 855–859. Available at: https://doi.org/10.1136/bmj.321.7265.855.

McKenna, A. *et al.* (2010) “The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data,” *Genome research*, 20(9), pp. 1297–1303. Available at: https://doi.org/10.1101/gr.107524.110.

McManus, B.A. *et al.* (2008) “Multilocus sequence typing reveals that the population structure of Candida dubliniensis is significantly less divergent than that of Candida albicans,” *Journal of clinical microbiology*, 46(2), pp. 652–664. Available at: https://doi.org/10.1128/JCM.01574-07.

Medapati, M.R., Singh, N., *et al.* (2021) “Bitter taste receptor T2R14 detects quorum sensing molecules from cariogenic Streptococcus mutans and mediates innate immune responses in gingival epithelial cells,” *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 35(3), p. e21375. Available at: https://doi.org/10.1096/fj.202000208R.

Medapati, M.R., Bhagirath, A.Y., *et al.* (2021) “Bitter taste receptor T2R14 modulates gram-positive bacterial internalization and survival in gingival epithelial cells,” *International journal of molecular sciences*, 22(18). Available at: https://doi.org/10.3390/ijms22189920.

Meng, Y. *et al.* (2019) “Human genes influence the interaction between Streptococcus mutans and host caries susceptibility: a genome-wide association study in children with primary

dentition," *International journal of oral science*, 11(2), p. 19. Available at: https://doi.org/10.1038/s41368-019-0051-4.

Menon, L.U. *et al.* (2022) "Candida albicans and Early Childhood Caries," *Frontiers in dental medicine*, 3, p. 849274. Available at: https://doi.org/10.3389/fdmed.2022.849274.

Meriç, E. *et al.* (2020) "Evaluation of oral streptococci in saliva of children with severe Early Childhood Caries and caries-free," *European journal of paediatric dentistry: official journal of European Academy of Paediatric Dentistry*, 21(1), pp. 13–17. Available at: https://doi.org/10.23804/ejpd.2020.21.01.03.

Metge, C. *et al.* (2009) *Composite Measures/Indices of Health and Health System Performance*. Winnipeg, MB: Manitoba Centre for Health Policy.

Misra, S., Tahmassebi, J.F. and Brosnan, M. (2007) "Early childhood caries--a review," *Dental update*, 34(9), pp. 556–8, 561–2, 564. Available at: https://doi.org/10.12968/denu.2007.34.9.556.

Molkenthin, F. *et al.* (2022) "Factors influencing the presence of Candida dubliniensis and other non-albicans species in patients with oral lichen planus: a retrospective observational study," *Clinical oral investigations*, 26(1), pp. 333–342. Available at: https://doi.org/10.1007/s00784-021-04004-5.

Montelongo-Jauregui, D. and Lopez-Ribot, J.L. (2018) "Candida Interactions with the Oral Bacterial Microbiota," *Journal of fungi (Basel, Switzerland)*, 4(4). Available at: https://doi.org/10.3390/jof4040122.

Morrison, J. *et al.* (2016) "Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL)," *Human molecular genetics*, 25(4), pp. 807–816. Available at: https://doi.org/10.1093/hmg/ddv506.

Motahari, P., Molaei, Z. and Adhami, Z.E. (2024) "Association of TAS1R2 (rs35874116 or rs9701796) gene polymorphism with dental caries: A systematic review and meta-analysis," *The open dentistry journal*, 18(1). Available at: https://doi.org/10.2174/0118742106275130240109051833.

Moynihan, P. and Petersen, P.E. (2004) "Diet, nutrition and the prevention of dental diseases," *Public health nutrition*, 7(1A), pp. 201–226. Available at: https://doi.org/10.1079/phn2003589.

Mukherjee, C. *et al.* (2021) "Acquisition of oral microbiota is driven by environment, not host genetics," *Microbiome*, 9(1), p. 54. Available at: https://doi.org/10.1186/s40168-020-00986-8.

Narang, S. *et al.* (2016) "Oral health status and treatment needs among 6, 9 and 12 years old rural and urban school children in India: An epidemiological survey," *Journal of advanced oral research*, 7(2), pp. 27–31. Available at: https://doi.org/10.1177/2229411220160205.

Narayanan, K. *et al.* (2001) "Differentiation of embryonic mesenchymal cells to odontoblast-like cells by overexpression of dentin matrix protein 1," *Proceedings of the National Academy of Sciences of the United States of America*, 98(8), pp. 4516–4521. Available at: https://doi.org/10.1073/pnas.081075198.

Nguyen, H.T.T. *et al.* (2021) "Dentinogenesis imperfecta and caries in osteogenesis imperfecta among Vietnamese children," *Dentistry journal*, 9(5), p. 49. Available at: https://doi.org/10.3390/dj9050049.

Nilsson, R.H. *et al.* (2016) "Top 50 most wanted fungi," *MycoKeys*, (12), pp. 29–40. Available at: https://doi.org/10.3897/mycokeys.12.7553.

Nogawa, S. *et al.* (2024) "Genome-wide association meta-analysis identifies two novel loci associated with dental caries," *BMC oral health*, 24(1), p. 1003. Available at: https://doi.org/10.1186/s12903-024-04799-1.

Nunn, M.E. *et al.* (2009) "Healthy eating index is a predictor of early childhood caries," *Journal of dental research*, 88(4), pp. 361–366. Available at: https://doi.org/10.1177/0022034509334043.

O'Connell, L.M. *et al.* (2023) "Supragingival mycobiome of HIV-exposed-but-uninfected children reflects a stronger correlation with caries-free-associated taxa compared to HIV-infected or uninfected children," *Microbiology spectrum*, 11(6), p. e0149123. Available at: https://doi.org/10.1128/spectrum.01491-23.

Ogawa, A. *et al.* (2011) "Inhibition of Streptococcus mutans biofilm formation by Streptococcus salivarius FruA," *Applied and environmental microbiology*, 77(5), pp. 1572–1580. Available at: https://doi.org/10.1128/AEM.02066-10.

Olatosi, O.O. *et al.* (2021) "Replication of GWAS significant loci in a sub-Saharan African Cohort with early childhood caries: a pilot study," *BMC oral health*, 21(1), p. 274. Available at: https://doi.org/10.1186/s12903-021-01623-y.

Olczak-Kowalczyk, D., Gozdowski, D. and Kaczmarek, U. (2020) "Factors associated with early childhood caries in Polish three-year-old children," *Oral health & preventive dentistry*, 18(4), pp. 833–842. Available at: https://doi.org/10.3290/j.ohpd.a45088.

Opal, S. *et al.* (2015) "Genetic factors affecting dental caries risk," *Australian dental journal*, 60(1), pp. 2–11. Available at: https://doi.org/10.1111/adj.12262.

Orlova, E. *et al.* (2022) "Association of Early Childhood Caries with Bitter Taste Receptors: A Meta-Analysis of Genome-Wide Association Studies and Transcriptome-Wide Association Study," *Genes*, 14(1). Available at: https://doi.org/10.3390/genes14010059.

Öter, B. *et al.* (2011) "The relation between 6-n-propylthiouracil sensitivity and caries activity in schoolchildren," *Caries research*, 45(6), pp. 556–560. Available at: https://doi.org/10.1159/000332432.

Oubenyahya, H. and Bouhabba, N. (2019) “General anesthesia in the management of early childhood caries: an overview,” *Journal of dental anesthesia and pain medicine*, 19(6), pp. 313–322. Available at: https://doi.org/10.17245/jdapm.2019.19.6.313.

Ozturk, A., Famili, P. and Vieira, A.R. (2010) “The antimicrobial peptide DEFB1 is associated with caries,” *Journal of dental research*, 89(6), pp. 631–636. Available at: https://doi.org/10.1177/0022034510364491.

Pacey, A., Nancarrow, T. and Egeland, G.M. (2010) “Prevalence and risk factors for parental-reported oral health of Inuit preschoolers: Nunavut Inuit Child Health Survey, 2007-2008,” *Rural and remote health*, 10(2), p. 1368. Available at: https://doi.org/10.22605/rrh1368.

Paglia, L. (2018) “Child abuse: Awareness is the first step to action,” *European journal of paediatric dentistry: official journal of European Academy of Paediatric Dentistry*, 19(2), p. 89. Available at: https://doi.org/10.23804/ejpd.2018.19.02.00.

Pahel, B.T., Rozier, R.G. and Slade, G.D. (2007) “Parental perceptions of children’s oral health: the Early Childhood Oral Health Impact Scale (ECOHIS),” *Health and quality of life outcomes*, 5, p. 6. Available at: https://doi.org/10.1186/1477-7525-5-6.

Palmer, C.A. *et al.* (2010) “Diet and caries-associated bacteria in severe early childhood caries,” *Journal of dental research*, 89(11), pp. 1224–1229. Available at: https://doi.org/10.1177/0022034510376543.

Palmer, R.J., Jr (2014) “Composition and development of oral bacterial communities,” *Periodontology 2000*, 64(1), pp. 20–39. Available at: https://doi.org/10.1111/j.1600-0757.2012.00453.x.

Papadaki, S. *et al.* (2021) “Gender differences in caries and periodontal status in UK children,” *bioRxiv*. medRxiv. Available at: https://doi.org/10.1101/2021.03.24.21253842.

Pattaroni, C. *et al.* (2022) “Early life inter-kingdom interactions shape the immunological environment of the airways,” *Microbiome*, 10(1), p. 34. Available at: https://doi.org/10.1186/s40168-021-01201-y.

Peña-Alegre, M. *et al.* (2023) “Impact of oral conditions on quality of life in Peruvian preschoolers in rural and urban areas,” *Odovtos - International Journal of Dental Sciences*, 25(3), pp. 249–260. Available at: https://doi.org/10.15517/ijds.2023.55835.

Pereira, D. *et al.* (2017) “Is the oral fungal pathogen Candida albicans a cariogen?,” *Oral diseases*, 24(4), pp. 518–526. Available at: https://doi.org/10.1111/odi.12691.

Peters, B.A. *et al.* (2017) “The oral fungal mycobiome: characteristics and relation to periodontitis in a pilot study,” *BMC microbiology*, 17(1), p. 157. Available at: https://doi.org/10.1186/s12866-017-1064-9.

Pham, D. and Zeng, L. (2024) “Genetic Characterization of the Glucose-PTS in Streptococcus sanguinis, examining competitive fitness,” *Journal of undergraduate research (Gainesville, Fla.)*, 26. Available at: https://doi.org/10.32473/ufjur.26.135370.

Pierce, A. *et al.* (2019) “The Burden of Early Childhood Caries in Canadian Children and Associated Risk Factors,” *Frontiers in public health*, 7, p. 328. Available at: https://doi.org/10.3389/fpubh.2019.00328.

Pitts, N.B. *et al.* (2017) “Dental caries,” *Nature Reviews Disease Primers*, 3(1), p. 17030. Available at: https://doi.org/10.1038/nrdp.2017.30.

Purcell, S. *et al.* (2007) “PLINK: a tool set for whole-genome association and population-based linkage analyses,” *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at: https://doi.org/10.1086/519795.

Quast, C. *et al.* (2013) “The SILVA ribosomal RNA gene database project: improved data processing and web-based tools,” *Nucleic acids research*, 41(Database issue), pp. D590-6. Available at: https://doi.org/10.1093/nar/gks1219.

Quince, C. *et al.* (2017) “Shotgun metagenomics, from sampling to analysis,” *Nature biotechnology*, 35(9), pp. 833–844. Available at: https://doi.org/10.1038/nbt.3935.

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rai, N.K. and Tiwari, T. (2018) “Parental factors influencing the development of early childhood caries in developing nations: A systematic review,” *Frontiers in public health*, 6, p. 64. Available at: https://doi.org/10.3389/fpubh.2018.00064.

Raja, M., Hannan, A. and Ali, K. (2010) “Association of oral candidal carriage with dental caries in children,” *Caries research*, 44(3), pp. 272–276. Available at: https://doi.org/10.1159/000314675.

Ramos-Gomez, F.J. *et al.* (2012) “Minimal intervention dentistry: part 3. Paediatric dental care--prevention and management protocols using caries risk assessment for infants and young children,” *British dental journal*, 213(10), pp. 501–508. Available at: https://doi.org/10.1038/sj.bdj.2012.1040.

Randall Simpson, J.A. *et al.* (2008) “Nutrition Screening Tool for Every Preschooler (NutriSTEP): validation and test-retest reliability of a parent-administered questionnaire assessing nutrition risk of preschoolers,” *European journal of clinical nutrition*, 62(6), pp. 770–780. Available at: https://doi.org/10.1038/sj.ejcn.1602780.

Richards, V.P. *et al.* (2017) “Microbiomes of Site-Specific Dental Plaques from Children with Different Caries Status,” *Infection and immunity*, 85(8). Available at: https://doi.org/10.1128/IAI.00106-17.

Rickard, A.H. *et al.* (2003) “Bacterial coaggregation: an integral process in the development of multi-species biofilms,” *Trends in microbiology*, 11(2), pp. 94–100. Available at: https://doi.org/10.1016/s0966-842x(02)00034-3.

Rivers, A.R. *et al.* (2018) “ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis,” *F1000Research*, 7, p. 1418. Available at: https://doi.org/10.12688/f1000research.15704.1.

Rizkallah, M.R. *et al.* (2012) “The PharmacoMicrobiomics Portal: A Database for Drug-Microbiome Interactions,” *Current pharmacogenomics and personalized medicine*, 10(3), pp. 195–203. Available at: https://doi.org/10.2174/187569212802510030.

Rohnbogner, A. and Lewis, M. (2018) “Dental caries as a measure of diet, health, and difference in non-adults from urban and rural Roman Britain,” *Dental Anthropology Journal*, 29(1), pp. 16–31. Available at: https://doi.org/10.26575/daj.v29i1.32.

Romano, S. *et al.* (2020) “Meta-analysis of the gut microbiome of Parkinson’s disease patients suggests alterations linked to intestinal inflammation,” *bioRxiv*. medRxiv. Available at: https://doi.org/10.1101/2020.08.10.20171397.

Rosenblatt, A., Stamford, T.C.M. and Niederman, R. (2009) “Silver diamine fluoride: a caries ‘silver-fluoride bullet,’” *Journal of dental research*, 88(2), pp. 116–125. Available at: https://doi.org/10.1177/0022034508329406.

Rosenblatt, A. and Zarzar, P. (2002) “The prevalence of early childhood caries in 12- to 36-month-old children in Recife, Brazil,” *ASDC journal of dentistry for children*, 69(3), pp. 319–24, 236. Available at: https://pubmed.ncbi.nlm.nih.gov/12613320/ (Accessed: March 14, 2025).

Rozkiewicz, D. *et al.* (2006) “Oral Candida albicans carriage in healthy preschool and school children,” *Advances in medical sciences*, 51 Suppl 1, pp. 187–190. Available at: https://europepmc.org/article/med/17458089 (Accessed: March 6, 2025).

Sabel, N. *et al.* (2024) “Dental caries and oral health-related quality of life in Preschoolers - introducing the Swedish version of the early childhood oral health impact scale (ECOHIS),” *Acta odontologica Scandinavica*, 83, pp. 47–53. Available at: https://doi.org/10.1080/00016357.2023.2287235.

Sandy, L.P.A., Helmyati, S. and Amalia, R. (2024) “Nutritional factors associated with early childhood caries: A systematic review and meta-analysis,” *The Saudi dental journal*, 36(3), pp. 413–419. Available at: https://doi.org/10.1016/j.sdentj.2023.12.001.

Schoch, C.L. *et al.* (2012) “Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*,” *Proceedings of the National Academy of Sciences of the United States of America*, 109(16), pp. 6241–6246. Available at: https://doi.org/10.1073/pnas.1117018109.

Schroth, R.J. *et al.* (2016) “Treating early childhood caries under general anesthesia: A national review of Canadian data,” *Journal (Canadian Dental Association)*, 82, p. g20. Available at: https://pubmed.ncbi.nlm.nih.gov/27548666/ (Accessed: March 13, 2025).

Schroth, R.J. *et al.* (2021) “A systematic review to inform the development of a Canadian caries risk assessment tool for use by primary healthcare providers,” *International journal of paediatric dentistry*, 31(6), pp. 767–791. Available at: https://doi.org/10.1111/ipd.12776.

Schroth, R.J. *et al.* (2022) “Improvement in Iron and Iron-Related Nutritional Status Following Pediatric Dental Surgery To Treat Severe Early Childhood Caries,” *Pediatric dentistry*, 44(1), pp. 58–66. Available at: https://www.ncbi.nlm.nih.gov/pubmed/35232538.

Schroth, R.J. and Cheba, V. (2007) “Determining the prevalence and risk factors for early childhood caries in a community dental health clinic,” *Pediatric dentistry*, 29(5), pp. 387–396. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18027773.

Schroth, R.J., Harrison, R.L. and Moffatt, M.E.K. (2009) “Oral health of indigenous children and the influence of early childhood caries on childhood health and well-being,” *Pediatric clinics of North America*, 56(6), pp. 1481–1499. Available at: https://doi.org/10.1016/j.pcl.2009.09.010.

Schroth, R.J. and Smith, W.F. (2007) “A review of repeat general anesthesia for pediatric dental surgery in Alberta, Canada,” *Pediatric dentistry*, 29(6), pp. 480–487. Available at: https://www.ingentaconnect.com/content/aapd/pd/2007/00000029/00000006/art00007 (Accessed: March 13, 2025).

Selwitz, R.H., Ismail, A.I. and Pitts, N.B. (2007) “Dental caries,” *Lancet*, 369(9555), pp. 51–59. Available at: https://doi.org/10.1016/S0140-6736(07)60031-2.

Seminario, A. and Velan, E. (2016) “Vitamin D and Dental Caries in Primary Dentition,” *Journal of dentistry for children*, 83(3), pp. 114–119. Available at: https://www.ingentaconnect.com/content/aapd/jodc/2016/00000083/00000003/art00002.

Sender, R., Fuchs, S. and Milo, R. (2016) “Revised Estimates for the Number of Human and Bacteria Cells in the Body,” *PLoS biology*, 14(8), p. e1002533. Available at: https://doi.org/10.1371/journal.pbio.1002533.

Seneviratne, C.J., Zhang, C.F. and Samaranayake, L.P. (2011) “Dental plaque biofilm in oral health and disease,” *The Chinese journal of dental research: the official journal of the Scientific Section of the Chinese Stomatological Association (CSA)*, 14(2), pp. 87–94. Available at: https://pubmed.ncbi.nlm.nih.gov/22319749/ (Accessed: March 16, 2025).

Seow, W.K. (2012) “Environmental, maternal, and child factors which contribute to early childhood caries: a unifying conceptual model: A unifying conceptual model for early childhood caries,” *International journal of paediatric dentistry*, 22(3), pp. 157–168. Available at: https://doi.org/10.1111/j.1365-263X.2011.01186.x.

Shaffer, J.R. *et al.* (2015) “Caries experience differs between females and males across age groups in northern Appalachia,” *International journal of dentistry*, 2015(1), p. 938213. Available at: https://doi.org/10.1155/2015/938213.

Shaji, C.S. and Saraswathy, R. (2023) “Taste receptors influencing effective modalities in human health - A cutting edge update on TAS1R and TAS2R receptor polymorphisms in taste perception and disease risk,” *Nutrition and health*, p. 2601060231186865. Available at: https://doi.org/10.1177/02601060231186865.

Sharma, A. *et al.* (2023) “Single nucleotide polymorphisms of enamel formation genes and early childhood caries - systematic review, gene-based, gene cluster and meta-analysis,” *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, 41(1), pp. 3–15. Available at: https://doi.org/10.4103/jisppd.jisppd_78_23.

Sheiham, A. and James, W.P.T. (2015) “Diet and dental caries: The pivotal role of free sugars reemphasized: The pivotal role of free sugars reemphasized,” *Journal of dental research*, 94(10), pp. 1341–1347. Available at: https://doi.org/10.1177/0022034515590377.

Shrestha, P. *et al.* (2024) “Multiancestry genome-wide association study of early childhood caries,” *Journal of dental research*, p. 220345241291528. Available at: https://doi.org/10.1177/00220345241291528.

Shrestha, S.K. *et al.* (2023) “Association of breastfeeding and early childhood caries: A systematic review and meta-analysis,” *Nutrients*, 16(9), p. 1355. Available at: https://doi.org/10.3390/nu16091355.

Shrutha, S.P. *et al.* (2025) “Correlation of socioeconomic, nutritional, and hematological status in early childhood caries among pre-schoolers-a case-control study,” *International journal of contemporary pediatrics*, 12(2), pp. 307–311. Available at: https://doi.org/10.18203/2349-3291.ijcp20250101.

Shungin, D. *et al.* (2019) “Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data,” *Nature communications*, 10(1), p. 2773. Available at: https://doi.org/10.1038/s41467-019-10630-1.

Siddaiah, D.S.B. and Sinha, D.S. (2023) “Candida albicans - A partner in crime in early childhood caries,” *International Journal of Research Publication and Reviews*, 4(4). Available at: https://doi.org/10.55248/gengpi.234.4.36868.

Simancas-Pallares, M. *et al.* (2023) “Evidence for clinical subtypes of early childhood caries,” *Research Square*, p. Preprint. Available at: https://doi.org/10.21203/rs.3.rs-2629612/v1.

Singh, N. *et al.* (2024) “Bitter Taste Receptor T2R14 and Autophagy Flux in Gingival Epithelial Cells,” *Cells* , 13(6). Available at: https://doi.org/10.3390/cells13060531.

Smith, C.E.L. *et al.* (2017) “Amelogenesis imperfecta; Genes, proteins, and pathways,” *Frontiers in physiology*, 8, p. 435. Available at: https://doi.org/10.3389/fphys.2017.00435.

So, M. *et al.* (2017) “Early childhood dental caries, mouth pain, and malnutrition in the Ecuadorian Amazon region,” *International journal of environmental research and public health*, 14(5), p. 550. Available at: https://doi.org/10.3390/ijerph14050550.

Sokal-Gutierrez, K. *et al.* (2016) “Early childhood caries and malnutrition: baseline and two-year follow-up results of a community-based prevention intervention in Rural Ecuador,” *BMC nutrition*, 2(1), pp. 1–11. Available at: https://doi.org/10.1186/s40795-016-0110-6.

Stanley, B.O.C. *et al.* (2014) “Genetic association of MPPED2 and ACTN2 with dental caries,” *Journal of dental research*, 93(7), pp. 626–632. Available at: https://doi.org/10.1177/0022034514534688.

Strömberg, N. *et al.* (2017) “Genetic- and lifestyle-dependent dental caries defined by the acidic proline-rich protein genes PRH1 and PRH2,” *EBioMedicine*, 26, pp. 38–46. Available at: https://doi.org/10.1016/j.ebiom.2017.11.019.

Struzycka, I. (2014) “The oral microbiome in dental caries,” *Polish journal of microbiology / Polskie Towarzystwo Mikrobiologow = The Polish Society of Microbiologists*, 63(2), pp. 127–135. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25115106.

Sullivan, D. and Coleman, D. (1998) “Candida dubliniensis: characteristics and identification,” *Journal of clinical microbiology*, 36(2), pp. 329–334. Available at: https://doi.org/10.1128/JCM.36.2.329-334.1998.

Susin, A. *et al.* (2020) “Variable selection in microbiome compositional data analysis,” *NAR genomics and bioinformatics*, 2(2), p. lqaa029. Available at: https://doi.org/10.1093/nargab/lqaa029.

Sztajer, H. *et al.* (2014) “Cross-feeding and interkingdom communication in dual-species biofilms of Streptococcus mutans and Candida albicans,” *The ISME journal*, 8(11), pp. 2256–2271. Available at: https://doi.org/10.1038/ismej.2014.73.

Takahashi, N. (2005) “Microbial ecosystem in the oral cavity: Metabolic diversity in an ecological niche and its relationship with oral diseases,” *International congress series. Excerpta Medica*, 1284, pp. 103–112. Available at: https://doi.org/10.1016/j.ics.2005.06.071.

Takahashi, N. and Nyvad, B. (2011) “The role of bacteria in the caries process: ecological perspectives,” *Journal of dental research*, 90(3), pp. 294–303. Available at: https://doi.org/10.1177/0022034510379602.

Tanner, A.C.R., Mathney, J.M.J., *et al.* (2011) “Cultivable anaerobic microbiota of severe early childhood caries,” *Journal of clinical microbiology*, 49(4), pp. 1464–1474. Available at: https://doi.org/10.1128/JCM.02427-10.

Tanner, A.C.R., Kent, R.L., Jr, *et al.* (2011) “Microbiota of severe early childhood caries before and after therapy,” *Journal of dental research*, 90(11), pp. 1298–1305. Available at: https://doi.org/10.1177/0022034511421201.

Taruno, A. and Gordon, M.D. (2023) “Molecular and cellular mechanisms of salt taste,” *Annual review of physiology*, 85(1), pp. 25–45. Available at: https://doi.org/10.1146/annurev-physiol-031522-075853.

Teng, F. *et al.* (2015) “Prediction of early childhood caries via spatial-temporal variations of oral microbiota,” *Cell host & microbe*, 18(3), pp. 296–306. Available at: http://www.cell.com/article/S1931312815003339/abstract.

Thomas, A.M. *et al.* (2019) “Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation,” *Nature medicine*, 25(4), pp. 667–678. Available at: https://doi.org/10.1038/s41591-019-0405-7.

Tim Wright, J. (2019) “The Role of Genetics in Caries Risk and Resistance,” *Dimensions of Dental Hygiene*, 21 February. Available at: https://dimensionsofdentalhygiene.com/article/genetics-caries-risk/ (Accessed: January 14, 2025).

Tinanoff, N. (2015) “Introduction to the conference: Innovations in the prevention and management of early childhood caries,” *Pediatric dentistry*, 37(3), pp. 198–199. Available at: https://www.ingentaconnect.com/content/aapd/pd/2015/00000037/00000003/art00001.

Tinanoff, N. *et al.* (2019) “Early childhood caries epidemiology, aetiology, risk assessment, societal burden, management, education, and policy: Global perspective,” *International journal of paediatric dentistry*, 29(3), pp. 238–248. Available at: https://doi.org/10.1111/ipd.12484.

Topçuoğlu, B.D. *et al.* (2021) “mikropml: User-Friendly R Package for Supervised Machine Learning Pipelines,” *Journal of open source software*, 6(61). Available at: https://doi.org/10.21105/joss.03073.

Treerat, P. *et al.* (2024) “Corynebacterial membrane vesicles disrupt cariogenic interkingdom assemblages,” *Applied and environmental microbiology*, 90(11), p. e0088524. Available at: https://doi.org/10.1128/aem.00885-24.

Tsai, C. *et al.* (2023) “Early childhood caries sequelae and relapse rates in an Australian public dental hospital,” *International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children*, 33(1), pp. 1–11. Available at: https://doi.org/10.1111/ipd.12969.

Tu, Y. *et al.* (2022) “The Crosstalk Between Saliva Bacteria and Fungi in Early Childhood Caries,” *Frontiers in cellular and infection microbiology*, 12, p. 845738. Available at: https://doi.org/10.3389/fcimb.2022.845738.

Tubert-Jeannin, S. *et al.* (2011) “Fluoride supplements (tablets, drops, lozenges or chewing gums) for preventing dental caries in children,” *Cochrane database of systematic reviews*, 2019(12), p. CD007592. Available at: https://doi.org/10.1002/14651858.CD007592.pub2.

Uchida-Fukuhara, Y. *et al.* (2020) “Caries Increment and Salivary Microbiome during University Life: A Prospective Cohort Study,” *International journal of environmental research and public health*, 17(10). Available at: https://doi.org/10.3390/ijerph17103713.

Uribe, S.E., Innes, N. and Maldupa, I. (2021) “The global prevalence of early childhood caries: A systematic review with meta-analysis using the WHO diagnostic criteria,” *International journal of paediatric dentistry*, 31(6), pp. 817–830. Available at: https://doi.org/10.1111/ipd.12783.

Usyk, M. *et al.* (2017) “Novel ITS1 Fungal Primers for Characterization of the Mycobiome,” *mSphere*, 2(6). Available at: https://doi.org/10.1128/mSphere.00488-17.

Vasireddy, D. *et al.* (2021) “Socioeconomic factors associated with the risk and prevalence of dental caries and dental treatment trends in children: A cross-sectional analysis of National Survey of Children’s Health (NSCH) data, 2016-2019,” *Cureus*, 13(11), p. e19184. Available at: https://doi.org/10.7759/cureus.19184.

Vieira, A.R. and Modesto, A. (2022) “Oral microbiome dental caries associated genotypes analysis of 6- to 19-year-old individuals shows novel associations,” *Frontiers in dental medicine*, 3, p. 875953. Available at: https://doi.org/10.3389/fdmed.2022.875953.

Vieira, A.R., Modesto, A. and Marazita, M.L. (2014) “Caries: review of human genetics research,” *Caries research*, 48(5), pp. 491–506. Available at: https://doi.org/10.1159/000358333.

Walsh, T. *et al.* (2019) “Fluoride toothpastes of different concentrations for preventing dental caries,” *Cochrane database of systematic reviews*, 3(3), p. CD007868. Available at: https://doi.org/10.1002/14651858.CD007868.pub3.

Wang, X. *et al.* (2010) “Genes and their effects on dental caries may differ between primary and permanent dentitions,” *Caries research*, 44(3), pp. 277–284. Available at: https://doi.org/10.1159/000314676.

Wang, Y. and Cao, K.-A.L. (2020) “A multivariate method to correct for batch effects in microbiome data,” *bioRxiv*. Available at: https://doi.org/10.1101/2020.10.27.358283.

Wang, Y. and Lê Cao, K.-A. (2023) “PLSDA-batch: a multivariate framework to correct for batch effects in microbiome data,” *Briefings in bioinformatics*, 24(2). Available at: https://doi.org/10.1093/bib/bbac622.

Wang, Yuan *et al.* (2017) “Profiling of Oral Microbiota in Early Childhood Caries Using Single-Molecule Real-Time Sequencing,” *Frontiers in microbiology*, 8, p. 2244. Available at: https://doi.org/10.3389/fmicb.2017.02244.

Ward, T.L. *et al.* (2018) “Development of the human mycobiome over the first month of life and across body sites,” *mSystems*, 3(3). Available at: https://doi.org/10.1128/mSystems.00140-17.

Wendell, S. *et al.* (2010) “Taste genes associated with dental caries,” *Journal of dental research*, 89(11), pp. 1198–1202. Available at: https://doi.org/10.1177/0022034510381502.

Weng, L. *et al.* (2024) “Inter-kingdom interactions and environmental influences on the oral microbiome in severe early childhood caries,” *bioRxiv*. Available at: https://doi.org/10.1101/2024.09.26.615216.

Willems, S. *et al.* (2005) “The independent impact of household- and neighborhood-based social determinants on early childhood caries: A cross-sectional study of inner-city children,” *Family & community health*, 28(2), pp. 168–175. Available at: https://doi.org/10.1097/00003727-200504000-00008.

Williams, T.L. *et al.* (2021) “Association between vitamin D and dental caries in a sample of Canadian and American preschool-aged children,” *Nutrients*, 13(12), p. 4465. Available at: https://doi.org/10.3390/nu13124465.

Wirbel, J. *et al.* (2020) “SIAMCAT: user-friendly and versatile machine learning workflows for statistically rigorous microbiome analyses.” Available at: https://doi.org/10.1101/2020.02.06.931808.

Wirbel, J. *et al.* (2021) “Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox,” *Genome biology*, 22(1), p. 93. Available at: https://doi.org/10.1186/s13059-021-02306-1.

Wong, S. *et al.* (2017) “Quality of life changes in children following emergency dental extractions under general anaesthesia,” *International journal of paediatric dentistry*, 27(2), pp. 80–86. Available at: https://doi.org/10.1111/ipd.12241.

Wright, E. (2016) “Using DECIPHER v2.0 to analyze big biological sequence data in R,” *The R journal*, 8(1), p. 352. Available at: https://doi.org/10.32614/rj-2016-025.

Wright, J.T. *et al.* (2014) “Fluoride toothpaste efficacy and safety in children younger than 6 years: a systematic review,” *Journal of the American Dental Association (1939)*, 145(2), pp. 182–189. Available at: https://doi.org/10.14219/jada.2013.37.

Wu, M.C. *et al.* (2011) “Rare-variant association testing for sequencing data with the sequence kernel association test,” *The American Journal of Human Genetics*, 89(1), pp. 82–93. Available at: https://doi.org/10.1016/j.ajhg.2011.05.029.

Xi, R., Zheng, X. and Tizzano, M. (2022) “Role of taste receptors in innate immunity and oral health,” *Journal of dental research*, 101(7), pp. 759–768. Available at: https://doi.org/10.1177/00220345221077989.

Xiang, L. *et al.* (2024) “The association between mode of childbirth delivery and early childhood caries: A comprehensive systematic review and meta-analysis,” *PloS one*, 19(9), p. e0310405. Available at: https://doi.org/10.1371/journal.pone.0310405.

Xiao, J. *et al.* (2018) “Candida albicans and early childhood caries: A systematic review and meta-analysis,” *Caries research*, 52(1–2), pp. 102–112. Available at: https://doi.org/10.1159/000481833.

Xu, H. *et al.* (2014) “Plaque bacterial microbiome diversity in children younger than 30 months with or without caries prior to eruption of second primary molars,” *PloS one*, 9(2), p. e89269. Available at: https://doi.org/10.1371/journal.pone.0089269.

Xu, L. *et al.* (2018) “Dynamic Alterations in Salivary Microbiota Related to Dental Caries and Age in Preschool Children With Deciduous Dentition: A 2-Year Follow-Up Study,” *Frontiers in physiology*, 9, p. 342. Available at: https://doi.org/10.3389/fphys.2018.00342.

Yadav, S.P. *et al.* (2022) “Association of Early Childhood Caries with feeding, Dietary Habits, and Oral Hygiene Practices among rural and Urban School Children of Jaipur,” *International journal of clinical pediatric dentistry*, 15(3), pp. 273–279. Available at: https://doi.org/10.5005/jp-journals-10005-2396.

Yang, Z. *et al.* (2023) “Oral microbial communities in 5-year-old children with versus without dental caries,” *BMC oral health*, 23(1), p. 400. Available at: https://doi.org/10.1186/s12903-023-03055-2.

Yu, T. and Klein, O.D. (2020) “Molecular and cellular mechanisms of tooth development, homeostasis and repair,” *Development (Cambridge, England)*, 147(2), p. dev184754. Available at: https://doi.org/10.1242/dev.184754.

Yue, Y. and Hu, Y.-J. (2022) “A new approach to testing mediation of the microbiome at both the community and individual taxon levels,” *Bioinformatics (Oxford, England)*, 38(12), pp. 3173–3180. Available at: https://doi.org/10.1093/bioinformatics/btac310.

Zafar, N. *et al.* (2020) “Role of Streptococci as etiological agents of dental caries,” *Novel Research in Microbiology Journal*, 4(3), pp. 766–778. Available at: https://doi.org/10.21608/nrmj.2020.95320.

Zaorska, K. *et al.* (2021) “Prediction of early childhood caries based on single nucleotide polymorphisms using neural networks,” *Genes*, 12(4), p. 462. Available at: https://doi.org/10.3390/genes12040462.

Zhang, J. *et al.* (2019) “Sour sensing from the tongue to the brain,” *Cell*, 179(2), pp. 392-402.e15. Available at: https://doi.org/10.1016/j.cell.2019.08.031.

Zhang, L. *et al.* (2017) “Overexpression of GαS in Murine osteoblasts in vivo leads to increased bone mass and decreased bone quality,” *Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research*, 32(11), pp. 2171–2181. Available at: https://doi.org/10.1002/jbmr.3223.

Zhang, Q. *et al.* (2022) “Application of fluoride disturbs plaque microecology and promotes remineralization of enamel initial caries,” *Journal of oral microbiology*, 14(1), p. 2105022. Available at: https://doi.org/10.1080/20002297.2022.2105022.

Zhang, X. *et al.* (2016) “Prevalence and care index of early childhood caries in mainland China: evidence from epidemiological surveys during 1987-2013,” *Scientific reports*, 6(1), p. 18897. Available at: https://doi.org/10.1038/srep18897.

Zhang, Y. *et al.* (2003) “Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways,” *Cell*. Elsevier, pp. 293–301. Available at: https://doi.org/10.1016/s0092-8674(03)00071-0.

Zheng, Y. *et al.* (2018) “Comparative analysis of the microbial profiles in supragingival plaque samples obtained from twins with discordant caries phenotypes and their mothers,” *Frontiers in cellular and infection microbiology*, 8, p. 361. Available at: https://doi.org/10.3389/fcimb.2018.00361.

Zhou, H. *et al.* (2022) “LinDA: linear models for differential abundance analysis of microbiome compositional data,” *Genome biology*, 23(1), p. 95. Available at: https://doi.org/10.1186/s13059-022-02655-5.

Zhou, Y.-H. and Gallins, P. (2019) “A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction,” *Frontiers in genetics*, 10, p. 579. Available at: https://doi.org/10.3389/fgene.2019.00579.

Zou, T. *et al.* (2023) “Genome-wide analysis of dental caries variability reveals genotype-by-environment interactions,” *Genes*, 14(3), p. 736. Available at: https://doi.org/10.3390/genes14030736.