Prediction of Future Caries in Toddlers via Salivary Microbiome: A 1-Year Longitudinal Study.

Mr. Ratcha Raksakmanut

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Oral Biology
FACULTY OF DENTISTRY
Chulalongkorn University
Academic Year 2022

การทำนายการเกิดฟันผุในเด็กเล็กด้วยไมโครไบโอมในน้ำลาย: การศึกษาระยะยาวเป็นเวลา 1 ปี

นายรัชชา รักศักดิ์มนุษย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีววิทยาช่องปาก ไม่สังกัดภาควิชา/เทียบเท่า
คณะทันตแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

Thesis Title	Prediction of Future Caries in Toddlers via Salivary Microbiome: A 1-Year Longitudinal Study.
By	Mr. Ratcha Raksakmanut
Field of Study	Oral Biology
Thesis Advisor	Associate Professor PANIDA THANYASRISUNG, D.D.S., Ph.D.
Thesis Co Advisor	Associate Professor ORANART MATANGKASOMBUT, D.D.S., Ph.D. Associate Professor WARANUCH PITIPHAT, D.D.S.,M.P.H.M., M.S., Sc.D.

Accepted by the FACULTY OF DENTISTRY, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

..... Dean of the FACULTY OF DENTISTRY
(Professor PORNCHAI JANSISYANONT, D.D.S., M.S., Ph.D.)

DISSERTATION COMMITTEE

..... Chairman
(Associate Professor SORAYA CHATURONGAKUL, B.S., Ph.D.)

..... Thesis Advisor
(Associate Professor PANIDA THANYASRISUNG, D.D.S., Ph.D.)

..... Thesis Co-Advisor
(Associate Professor ORANART MATANGKASOMBUT, D.D.S., Ph.D.)

..... Thesis Co-Advisor
(Associate Professor WARANUCH PITIPHAT, D.D.S.,M.P.H.M., M.S., Sc.D.)

รัชชา รักศักดิ์มนุษย์ : การทำนายการเกิดฟันผุในเด็กเล็กด้วยไมโครไบโอมในน้ำลาย: การศึกษาระยะยาวเป็นเวลา 1 ปี. (Prediction of Future Caries in Toddlers via Salivary Microbiome: A 1-Year Longitudinal Study.) อ.ที่ปรึกษาหลัก : รศ. ทญ. ดร.พนิดา ธัญญศรีสังข์, อ.ที่ปรึกษาร่วม : รศ. ทญ. ดร. อรนาฎ มาตังคสมบัติ,รศ. ทญ. ดร.วรานุช ปิติพัฒน์

โรคฟันผุในเด็กจัดเป็นโรคเรื้อรังในช่องปากที่พบได้มากที่สุดและส่งผลเสียต่อคุณภาพชีวิตของเด็กรวมถึงการรบกวนพัฒนาการในด้านต่างๆ การป้องกันฟันผุตั้งแต่ระยะเริ่มต้นเป็นแนวทางสำคัญในการลดความชุกของโรคฟันผุในเด็กปฐมวัย ซึ่งมีความเป็นไปได้ในทางปฏิบัติเมื่อใช้วิธีการป้องกันฟันผุแบบกำหนดกลุ่มเป้าหมาย อย่างไรก็ดียังไม่มีตัวทำนายฟันผุที่น่าเชื่อถือเพียงพอสำหรับการใช้งานในเด็กกลุ่มอายุดังกล่าว การศึกษานี้จึงมุ่งที่จะพัฒนาการทำนายความเสี่ยงของการเกิดฟันผุโดยใช้ไมโครไบโอมในน้ำลายของเด็กอายุ 1 ขวบที่ปราศจากฟันผุ ตัวอย่างน้ำลาย 30 ตัวอย่างถูกคัดเลือกมาจากการศึกษากลุ่มประชากรแบบไปข้างหน้า โดยทั้งหมดได้มาจากเด็กอายุ 1 ขวบที่ปราศจากฟันผุที่มีปริมาณตัวอย่างที่เพียงพอต่อการวิเคราะห์ แบ่งเป็น 3 กลุ่ม กลุ่มละ 10 ตัวอย่าง ตามสถานะฟันผุเมื่อเด็กมีอายุ 2 ขวบที่แตกต่างกัน ได้แก่ เด็กที่พบรอยโรคฟันผุในระยะเริ่มต้น เด็กที่พบโพรงฟันผุ และเด็กที่ไม่พบรอยโรคฟันผุ นำมาวิเคราะห์ด้วยวิธีการจัดลำดับยีน 16S rRNA พบว่าไมโครไบโอมในน้ำลายของเด็กที่ไม่พบรอยโรคฟันผุในอนาคตมีความแตกต่างอย่างมีนัยสำคัญเมื่อเปรียบเทียบกับเด็กที่พบโพรงฟันผุในอนาคต (unweighted UniFrac, ANOSIM, Benjamini-Hochberg corrected P = 0.042) โดยพบว่าปริมาณเชื้อสัมพัทธ์ของ *Prevotella nanceiensis Leptotrichia sp. HMT 215 Prevotella melaninogenica* และ *Campylobacter concisus* ในเด็กที่ไม่พบรอยโรคฟันผุในอนาคตมีค่าสูงกว่าเด็กที่พบโพรงฟันผุในอนาคตอย่างมีนัยสำคัญ (Wilcoxon rank-sum test, P = 0.024 0.040 0.049 และ 0.049 ตามลำดับ) และเชื้อทั้ง 4 ชนิดดังกล่าวยังถูกระบุว่าเป็นตัวบ่งชี้ทางชีวภาพสำหรับเด็กที่ไม่พบรอยโรคฟันผุในอนาคต (LEfSe, LDA score = 3.69 3.74 3.52 และ 3.46 ตามลำดับ) แบบจำลองการทำนายฟันผุที่สร้างขึ้นโดยแมชชีนเลิร์นนิงโดยใช้ปริมาณเชื้อสัมพัทธ์ของเชื้อทั้ง 4 ชนิดดังกล่าวสามารถทำนายการเกิดโพรงฟันผุในอนาคตของเด็กอายุ 1 ขวบได้ที่ค่าความถูกต้องร้อยละ 80 ความไวร้อยละ 80 และความจำเพาะร้อยละ 80 (AUC, 0.8; 95% CI, 44.4-97.5) ผลการศึกษานี้แสดงให้เห็นว่าไมโครไบโอมในน้ำลายของเด็กอายุ 1 ขวบที่ปราศจากฟันผุสามารถใช้ทำนายความเสี่ยงในการเกิดฟันผุในอนาคตได้ ซึ่งสามารถพัฒนาต่อไปเพื่อใช้เป็นตัวทำนายฟันผุสำหรับการป้องกันฟันผุแบบกำหนดกลุ่มเป้าหมายได้ในอนาคต

สาขาวิชา	ชีววิทยาช่องปาก	ลายมือชื่อนิสิต .....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

6176055932 : MAJOR ORAL BIOLOGY

KEYWORD: dental caries, biomarkers, microbiota, saliva, machine learning, infant

Ratcha Raksakmanut : Prediction of Future Caries in Toddlers via Salivary Microbiome: A 1-Year Longitudinal Study.. Advisor: Assoc. Prof. PANIDA THANYASRISUNG, D.D.S., Ph.D. Co-advisor: Assoc. Prof. ORANART MATANGKASOMBUT, D.D.S., Ph.D.,Assoc. Prof. WARANUCH PITIPHAT, D.D.S.,M.P.H.M., M.S., Sc.D.

Dental caries in children is the most common chronic oral disease that could disturb their quality of life including their development. Early prevention is a key approach to reducing the prevalence of early childhood caries. However, a reliable caries predictor, as an essential tool for targeted prevention that is important to this approach, is still lacking for infants before caries onset. Therefore, we aimed to develop the caries risk prediction model based on the salivary microbiome of caries-free 1-year-old children to predict caries onset at 1-year follow-up. Using a nested case-control design within a prospective cohort study, 30 saliva samples, collected at a baseline time point, were selected based on sufficient sample quantity and caries status when children were 2-year-old: 10 children who developed non-cavitated caries lesions, 10 who developed cavitated caries lesions, and 10 who remained caries-free then 16S rRNA gene sequencing was performed. The salivary microbiota of the children who remained caries-free was significantly different when compared with those who developed cavitated caries lesions (unweighted UniFrac, ANOSIM, Benjamini-Hochberg corrected, P = 0.042). The relative abundance of *Prevotella nanceiensis, Leptotrichia sp. HMT 215, Prevotella melaninogenica*, and *Campylobacter concisus* were significantly higher in the children who remained caries-free compared with those who developed cavitated caries lesions (Wilcoxon rank-sum test, P = 0.024, 0.040, 0.049, and 0.049, respectively) and were identified as biomarkers for the children who remained caries-free (LEfSe, LDA score = 3.69, 3.74, 3.52, and 3.46, respectively). Caries prediction model generated by machine learning based on these 4 biomarkers differentiated the 1-year-old children between those who did and did not develop cavitated caries lesion at 2-year-old with an accuracy of 80%, sensitivity of 80%, and specificity of 80% (AUC, 0.8; 95% CI, 44.4-97.5). These findings suggest that the salivary microbiome of caries-free 1-year-old children could predict future caries onset in infants that could further develop into a promising caries predictor for targeted prevention.

Field of Study:	Oral Biology	Student's Signature .....
Academic Year:	2022	Advisor's Signature .....
		Co-advisor's Signature .....
		Co-advisor's Signature .....

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my primary supervisor, Assoc. Prof. Dr.Panida Thanyasrisung, and my co-supervisor, Assoc. Prof. Dr.Oranart Matangkasombut, for their invaluable advice, support, and patience during my Ph.D. study. Their immense knowledge and experience have encouraged and guided me throughout this project. I am also grateful to my co-supervisor, Assoc. Prof. Dr.Waranuch Pitiphat, who supported me and provided the samples from her previous project made this project possible.

I could not have undertaken this journey without my defense committee, Assoc. Prof. Dr.Soraya Chaturongakul, Prof. Dr.Thanaphum Osathanon, Assoc. Prof. Dr.Waleerat Sukarawan, and Asst. Prof. Dr.Nattanan Govitvattana, who generously provided knowledge and expertise.

I would like to extend my sincere thanks to Dr.Sirima Sritangsirikul for her effort in sample collection and data acquisition which was an important part of this project.

Thanks should also go to the staff in the Center of Excellence on Oral Microbiology and Immunology for their generous support in my laboratory work.

Additionally, this endeavor would not have been possible without the support from Rangsit University, which financed my Ph.D. study.

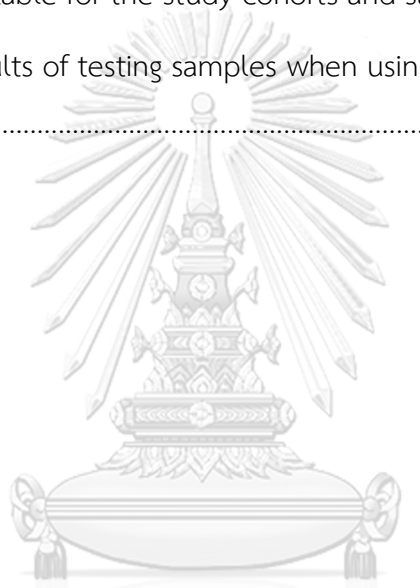
Lastly, I would be remiss in not mentioning my wife, who always support and believe in me and kept my spirit and motivation high during this process.

Ratcha Raksakmanut

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES



Chapter I

1. Introduction

1.1. Background and rationale

Dental caries is one of the most prevalent diseases in humans, affecting 97% of the population at least once in a lifetime (Berg, 2006) with an estimated 2.8 billion people suffering from this disease (James et al., 2018). In children, dental caries is the most common chronic disease globally. The data collected from worldwide studies reported that the average prevalence of dental caries in 3-year-olds was 43% (Tinanoff et al., 2019). In Thailand, the prevalence was higher than that average, according to the national survey in 2017, the prevalence of dental caries in Thai 3-year-olds was 52.9% (Keeddee et al., 2018). This disease is a hidden threat that usually has been overlooked. Although it is not life-threatening, it can cause many adverse effects on children both physically and psychologically that can have profoundly negative impacts on their growth and development, involving nutritional problems, and leading to oral health-related quality of life issues (AAPD, 2017; Zaror et al., 2022). Also, it is a reason that disturbs their learning development since it is associated with poor school performance and attendance (Rebelo et al., 2019). Moreover, this is a serious problem for their families, societies, and the health care system in terms of economic burden (Righolt et al., 2018). Therefore, preventing dental caries since childhood is likely to reduce the overall prevalence of caries and also help enhance the well-being of children and adults whom they become in the future.

In the past centuries, caries treatment had been focused on fixing the problems caused by dental caries, by removing caries lesions and restoring the functions of teeth, rather than controlling the process of disease. This might be a reason the caries prevalence is still high, in contrast to the current knowledge and advancements in dentistry. Recently, there have been increasing attempts to improve the effectiveness and efficiency of dental caries treatment. Nowadays, dentistry, including caries treatment, has entered an era of personalized care with a

customized treatment plan for each individual or group, based on their risks. Caries risk assessment tools have been developed, such as CAMBRA (Caries Management by Risk Assessment) (Featherstone et al., 2007) and CCI™ (CariesCare International) developed by ICDAS Foundation (Martignon et al., 2019). According to these systems, caries risk assessment is a crucial element that dictates the direction of disease management. The currently available caries risk assessments have a good performance in identifying the risk factors of each patient by considering the range of factors; including past caries experience, socio-economy, socio-demography, oral hygiene care, dietary habits, oral bacteria, fluoride usage, and saliva to predict the likelihood of caries progression and onset. Unfortunately, as disease predictors, most of them had limited accuracy, limited predictive value, or insufficiently supported evidence (Berkowitz et al., 2011; Tellez et al., 2013; Mejàre et al., 2014; Amin et al., 2015; Christian et al., 2020). Only the past caries experience was considered the most powerful predictor in all age groups (Q. Zhang & van Palenstein Helderman, 2006; Hänsel Petersson et al., 2013; Mejàre et al., 2014; Senneby et al., 2015; Du et al., 2017; Hu et al., 2018). However, this predictor is not practical to be used in the case of 1-year-old children with erupting baby teeth who mostly never experience caries (Tinanoff et al., 2019). The importance of this age group is emphasized in the "Early Childhood Caries: IAPD Bangkok Declaration" which recommended that providing preventive intervention within the first year of life is a key approach to reducing caries prevalence (Pitts et al., 2019).

The prediction of dental caries is complicated because dental caries is a multifactorial disease resulting from microbiological, genomic, behavioral, and social factors that are involved in the series of events that happen and last for years until the lesion is developed. Many caries risk assessment systems have been proposed but there is not sufficient scientific data that could confirm the effectiveness of these methods as a caries predictor (Cagetti et al., 2018) thus studies in this area are still in urgent need (Fontana et al., 2020). According to the ecological plaque hypothesis, dental caries is the result of an imbalance of the oral microbes within the biofilm due to ecological stress, resulting in an overgrowth of caries-related microbes (Marsh, 1994). When we consider the currently used caries risk assessments, most of them

focus on factors related to changes in the ecology of the oral microbes, such as sugar consumption, oral health care, dental plaque accumulation, the oral health of the caregiver, and dental appliances. It could be said that changes in the oral microbial population are the end result of the majority of currently known caries risk factors. Therefore, the analysis of the patient's oral microbes could be used as a simplified and comprehensive caries prediction.

Since the 1970s, many researchers had tried to identify caries-related microorganisms using culture-dependent techniques and microscopy with the hope that dental caries might be cured with antibiotics like other infectious diseases. Thereafter, *Streptococcus mutans* was recognized as a human odontopathogen since 1986 (Loesche, 1986) and have been focused on as a key pathogen (Emilson & Krasse, 1985; Balakrishnan et al., 2000; Marsh, 2003). Although *Streptococcus mutans* was often observed at a high level in caries lesions at the early stage, it could be found in some caries-free subjects and was not associated with caries progression (Gross et al., 2010). The use of *Streptococcus mutans* level alone was not recommended for caries prediction (Hong & Hu, 2010). The culture-independent techniques, such as metagenomics which could explore genetic materials recovered directly from biological samples revealed that enormous microbial diversity had been missed when using the culture-dependent methods alone (Hugenholtz et al., 1998). Accumulated evidence from the studies using genomic technologies, including next-generation sequencing (NGS) and bioinformatics, showed that a lot of other bacteria, such as *Veillonella spp., Scardovia wiggsiae, Slackia exigua, Firmicutes, Granulicatella elegans, Bifidobacterium, Corynebacterium matruchotii, Streptococcus cristatus, Streptococcus gordonii, Neisseria flavescens,* and *Neisseria mucosa* could be found in caries lesions and might relate to cariogenesis (Becker et al., 2002; Corby et al., 2007; Kreth et al., 2008; Ventura et al., 2009; Kanasi et al., 2010; Crielaard et al., 2011; Tanner et al., 2011; Gross et al., 2012; S. Jiang et al., 2016; Xiao et al., 2016; Agnello et al., 2017; Innes & Robertson, 2018; Xiao et al., 2018). These technologies allow us not just to identify novel disease-related microorganisms but give us a better understanding of the contribution of the oral microbiome to health. The relationship between microbiome and host is dynamic and is influenced by various

aspects of individual lifestyles, such as diet, smoking, and stress, which could create both healthy or dysbiotic ecology depending on the alteration of the microbiome and its properties (Kilian et al., 2016). This concept supports the ecological plaque hypothesis which stated that dental caries is developed as a result of an oral microbial imbalance (Hojo et al., 2009; Høiby et al., 2011; Marsh, 2012).

The microbial community structures between the caries-active and caries-free cohorts differ significantly (Gross et al., 2010; W. Jiang et al., 2013; W. Jiang et al., 2014). Even within a caries-active individual, the oral microbiome could be different at each stage of the caries process due to environmental acidification which could affect both the genotypic and phenotypic changes that occur in the oral microbiome (Takahashi & Nyvad, 2011). The caries process consists of several reversible stages that could be intervened to revert the process from disease to health. If we could identify the biomarkers in the oral microbiome that could predict the future caries status prior to the cavitation, preventing or remineralizing initial caries lesions without the surgical approaches could be possible, and the use of oral microbiome analysis for caries prediction would be an interesting model to enhance caries management. A similar approach has been used in many microbiome studies worldwide to classify and predict various host states in other diseases using human microbiome data (Knights et al., 2011; Human Microbiome Project, 2012).

In this study, we aimed to examine the differences between the salivary microbiome of caries-free 1-year-old children who remained caries-free vs those who developed caries at 1-year follow-up. Our goal is to identify potential microbial biomarkers and develop the caries prediction model for future caries in 1-year-old children since they are a key target group for targeted prevention to reduce caries prevalence worldwide.

1.2. Research Questions

1. Are there any differences in the salivary microbiome of caries-free 1-year-old children between those who remained caries-free and who progressed to caries-active states?

2. Are there any microbial biomarkers in the salivary microbiome of caries-free 1-year-old children that could reflect the future caries status in the next 1 year?
3. Can salivary microbiome be used as a predictor for future caries in caries-free 1-year-old children?

1.3. Objectives

1. To determine the differences in the salivary microbiome of caries-free 1-year-old children between those who remained caries-free and who progressed to caries-active states at 1-year follow-up.
2. To identify the microbial biomarkers in the salivary microbiome of caries-free 1-year-old children that could be used as a predictor for future caries.
3. To analyze the salivary microbiome as a predictor for future caries in caries-free 1-year-old children.

1.4. Hypotheses

1. The salivary microbiome of the caries-free 1-year-old children between those who remained caries-free and who progressed to caries-active states are different.
2. There are microbial biomarkers in the salivary microbiome of caries-free 1-year-old children that can reflect their future caries status.
3. The salivary microbiome can be used as a predictor for future caries in caries-free 1-year-old children.

1.5. Keywords

Dental Caries, Saliva, Microbiota, Next Generation Sequencing, 16S rRNA, Machine Learning, Prediction model, Infant

1.6. Conceptual framework

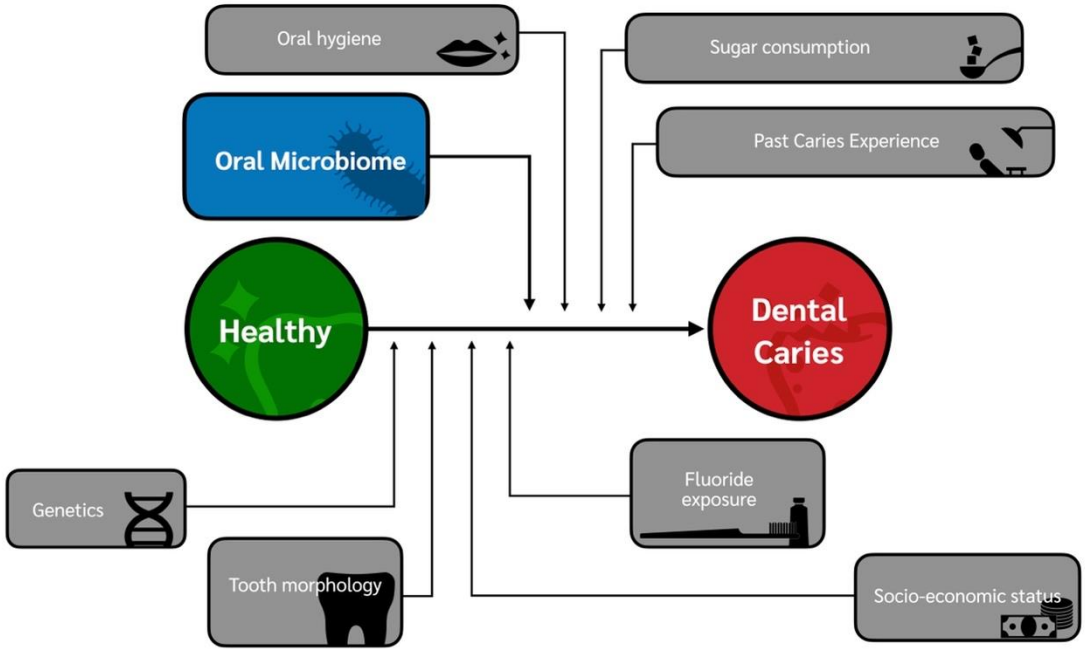


Figure 1: Conceptual framework (Caries predictors).

1.7. Benefits of study

This study evaluated the performance of the caries prediction models generated using the salivary microbiome of caries-free 1-year-old children. This information could be beneficial as a part of the collective evidence that may lead to the future applications of salivary microbiome-based caries prediction for 1-year-old children, an important target group for caries control strategy but lacks a reliable caries predictor, as an additional tool that combines with the existing caries management systems to promote caries management in children.

Chapter II

2. Literature reviews

2.1. Dental caries and current management

2.1.1. Current situation of Dental Caries

From the earliest evidence of dental caries found in human skulls from the Paleolithic era (40,000 to 25,000 years ago) (Lufkin, 1938) to the understanding of microorganisms' involvement in cariogenesis (Brock, 1961) and the discovery of the dynamic relationship within the dental biofilm (Marsh, 1994), this has been a long history of humanity and dental caries. However, even with accumulated knowledge on the pathogenesis of this disease, it is still the most prevalent oral disease in the world (James et al., 2018). Human is susceptible to caries as soon as the first tooth has erupted, which generally happens at the age of 6 months. The prevalence of dental caries greatly increased from 17% in 1-year-olds to 36% in 2-year-olds. Moreover, the caries prevalence continually increases to 43%, 55%, and 63% when children grow up to the age of 3, 4, and 5 years old, respectively (Tinanoff et al., 2019). Dental caries in children is often going untreated (Kassebaum et al., 2015). In Thailand, most caries cases in 3 and 5-year-old children were left untreated (Keeddee et al., 2018). This data illustrates the attitude toward dental caries, especially among children, that they and their parents are unaware of the detrimental health effects that are associated with impairing quality of life, both physically and mentally (Åkesson et al., 2016) and disturbing development of their children with reduced weight and delayed growth were reported (Fung et al., 2013). When a caries lesion progresses to the cavitation state, that lesion would no longer be prevented or remineralized by the medical approach, instead, surgical intervention would be necessary to stop the disease symptoms. At that point, the damage is irreversible and the correction is more complicated and expensive depending on the severity of the disease. Therefore, cost-effective caries preventive interventions starting at early age should improve the health-related quality of life and reduce the economic burden (Kastenbom et al., 2019).

2.1.2. Dental Caries: Slow-progressed multifactorial disease

Dental caries is a biofilm-mediated disease modulated by diet (Takahashi & Nyvad, 2011, 2016; Ferreira Zandoná et al., 2019). Dysbiosis in the oral biofilm, over time, can lead to demineralization of adjacent tooth surfaces. The repeated demineralization eventually progresses to the destruction of the intact surfaces, forming cavitation on enamel. Theoretically, any caries lesion with the proper conditions could be remineralized and healed since it is an initially reversible, chronic disease that process with a known multi-factorial etiology (Fejerskov & Nyvad, 2003). But once cavitation occurs, removal of biofilm is practically impossible then the surgical intervention will be necessary to stop caries progression. In the past, the treatment of dental caries had been focused on the removal and replacement of the damaged tooth structure rather than the correction of the root cause of the disease. In fact, the restoration of teeth can be found in ancient writings of many historical regions with the first description of the restoration of teeth being credited to Pierre Fauchard in 1728 (Ismail et al., 2001). This might be because dental caries is considered a complex and multifactorial disease, similar to diseases like cancer or diabetes with no single causation pathway (Fejerskov, 2004). Thus, it is easier to fix the obvious damage of caries lesions rather than control the complex causes of the disease that might not be successful. However, effective management strategies against dental caries that would give a long-lasting result are needed and should be based on the understanding of its complex etiology, and its multi-level influencing factors including biological, behavioral, and socioeconomic factors.

Dental caries is a chronic disease with slow progression in most cases. The caries lesion, which is the localized destruction of the tooth structure, is the sign of the disease (Fejerskov et al., 2015). The development of lesions is dynamic with the alternation of lesion progression and regression (Dirks, 1966; Nyvad et al., 2003). Lesion progression might be reverted at any stage of development when the proper environmental conditions are provided, such as biofilm control and an increase in topical fluoride exposure (Nyvad & Fejerskov, 1997). According to the extended caries

ecological hypothesis, the caries process consists of 3 reversible stages with the demineralization/remineralization balance of the process (Takahashi & Nyvad, 2008). If we can identify the current stage of the caries process, especially before the cavitation, it will be possible to design the appropriate caries management plan that is tailor-made for each patient.

2.1.3. Early childhood caries

Early childhood caries (ECC), formerly known as nursing caries and baby bottle tooth decay (Cleaton-Jones, 2002), is tooth decay in preschool children which is common and mostly untreated. It is defined as the presence of one or more caries lesions (non-cavitated or cavitated), missing (as a result of caries), or filling in any primary tooth of children younger than 6 years old (AAPD, 2017). In more severe cases, severe early childhood caries (S-ECC) is defined as any sign of smooth-surface caries in a child younger than 3 years old, and from ages 3-5, one or more cavitated, missing (due to caries), or filled smooth surfaces in primary maxillary anterior teeth or a dmfs (decayed, missing, or filled surfaces) score of greater than or equal to four (age 3), greater than or equal to five (age 4), or greater than or equal to six (age 5) (Drury et al., 1999). Nowadays, ECC is still a common chronic disease of childhood that is a worldwide health challenge (Zou et al., 2022), especially in low- and middle-income countries where there is a rapidly growing number of cases (Phantumvanit et al., 2018).

As well as dental caries in general, ECC is not entirely associated with poor feeding behavior but rather reflects its multifactorial etiology. The primary teeth are difficult to clean for many reasons, such as their anatomy that is constricted in the cervical portion and the cooperation of children to allow parents to clean their teeth regularly. It also has a lower level of calcium content and mineralization with a thinner thickness than the permanent teeth (De Menezes Oliveira et al., 2010) which could be responsible for its susceptibility to dental caries. The caries microbiome plays an important role as a primary factor in cariogenesis. Cariogenic bacteria produce weak acids after their metabolism of fermentable carbohydrates within the

biofilm attached to the tooth surfaces. As a result, there is a decrease in local pH values, which causes the demineralization of dental hard tissues. Thus, the study of ECC etiology is focused on oral microbial ecological imbalance, caries-related microbiome, and their relationships with host genetics, which might provide the theoretical basis that can be developed for more effective prevention and treatment of ECC. For this reason, the current approach to finding the predictors and biomarkers of ECC has been focused on the oral microbiome.

In order to reduce the prevalence and impact of ECC worldwide, the IAPD Bangkok Declaration recommends the following actions. First, the awareness of ECC should be raised among all stakeholders including parents/caregivers, dentists, physicians, nurses, etc. Second, sugar intake in foods and drinks should be limited and free sugars should be avoided for children younger than 2 years old. Third, tooth brushing with at least 1000 ppm fluoridated toothpaste should be performed twice a day in all children. Lastly, preventive guidance should be provided within the first year of life by a health professional or community health worker. Ideally, a referral to a dentist for comprehensive dental care should be provided (Pitts et al., 2019). However, to achieve this recommendation using only limited resources, a targeted prevention strategy is required to select only high caries-risk children to receive appropriate caries prevention. Therefore, the current approach to finding the predictors and biomarkers of ECC has been focused on the oral microbiome that should be developed for the better prevention and treatment of ECC (Zou et al., 2022).

2.1.4. Management of Early Childhood Caries

In the past decades, attempts to overcome this disease have been developed. Many methods were proposed upon the understanding that dental caries is not developed by any specific pathogens but initiated by several microorganisms in a complex biofilm that changes dynamically due to its environmental factors such as the habitat for the microbes, nutrition, and saliva. Moreover, factors like educational level, income, occupation, and socioeconomic status are also associated

with dental caries (Costa et al., 2012; Engelmann et al., 2016). Thus, the direction of caries management has been changed to investigate which of many factors is causing the disease in each situation and correct them instead of finding the ultimate universal treatment for everybody.

The integration of Caries-risk assessment (CRA) into caries management is an important milestone in dental health care. This assessment consists of identifying and analyzing various factors that are related to dental caries based on the most updated evidence and developing a personalized caries care plan for individuals. Several CRA models related to ECC were proposed worldwide since the starting of the 21st century and most of them are still being developed since then, including the caries-risk assessment tool (CAT) by the American Academy of Pediatric Dentistry (AAPD) (AAPD, 2002, 2021), Caries Management By Risk Assessment (CAMBRA) by California Dental Association (Featherstone et al., 2003; Featherstone et al., 2007; Featherstone et al., 2019), American Dental Association (ADA) caries-risk assessment (ADA, 2011), and Cariogram (Bratthall & Hansel Petersson, 2005). Another guideline that has been proposed to be used on a global scale is CariesCare International (CCI™) by the ICDAS foundation.

CariesCare International is the latest guideline developed from ICDAS (International Caries Detection and Assessment System) (Martignon et al., 2019). Starting from 2002, they tried to develop a simple, logical, evidence-based system for caries detection and classification that could be used in dental education, clinical practice, dental research, and dental public health (Pitts & Stamm, 2004; Pitts, 2009). In 2013, ICDAS was developed by adding the guideline for management, thus the ICCMS™ Guide for Practitioners and Educators has been proposed (Pitts et al., 2014). The fundamental concept that influenced all decisions in the ICCMS™ is to “Preserve tooth structure and restore only when indicated”. A comprehensive assessment and personalized caries care plan will be made specifically for each individual. In 2019, CariesCare International (CCI™) was developed as the simpler and shorter version of the full ICCMS guide (Martignon et al., 2019) with the hope that this practice-friendly consensus guideline will be widely adopted into routine dental practice all over the world.

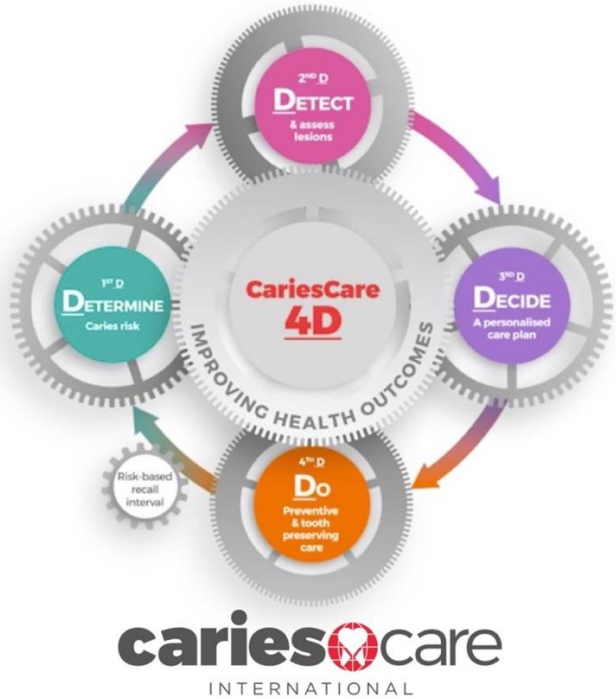


Figure 2: CariesCare 4D cycle (Martignon et al., 2019).

According to this concept, caries risk assessment is the first essential element (1st **D**: **D**etermine Caries risk, Figure 2) that will aid clinical decision-making and dictate the development of a personalized caries management plan. Moreover, the level of risk of caries could help each patient to understand and illustrate their caries prediction which might increase their motivation to engage with better health care and improve their behavior to enhance their oral health. The currently used caries risk assessments were well-performed to identify the factors responsible for the disease in a particular patient (Fontana & Zero, 2006) and could be the foundation for caries management in all age groups (Featherstone et al., 2021). However, as a caries predictor, there is not enough evidence to prove that the caries predictive ability of the existing caries risk assessment systems is valid enough to be implemented in a clinical situation (Tellez et al., 2013). Moreover, the categorization is mostly subjective based on the experience of evaluators, which is prone to human bias and error. Therefore, a reliable method that lacks bias with accurate caries prediction is needed, especially for dental caries in children, to improve the

efficiency of caries prevention under limited resource constraints. This development might fulfill the dream that everyone could stay cavity-free during their entire lives, as a goal of the Alliance for a Cavity-Free Future (ACFF) for every child born in 2026 and thereafter *(https://www.acffglobal.org)*.

2.2. Oral microbiome and dental caries

2.2.1. Oral microbes: a primary factor in cariogenesis

To predict any disease, we need to find the markers that are closely related to the disease condition. For dental caries, a caries lesion is occurred by the acid from bacteria in the biofilm that dissolves the mineralized tissues of the tooth (Segura et al., 2014). Caries lesions vary in severity, starting from clinically sound enamel surface, white spot lesion, and cavitated dentin lesion (Takahashi & Nyvad, 2011). At each stage, the microfloral members within the dental biofilm are different due to the change in the microenvironment. The initial colonizers of freshly cleaned tooth surfaces are a highly selective group of microbes, mainly *Streptococcus sanguinis*, *Streptococcus oralis*, and *Streptococcus mitis* (Nyvad & Kilian, 1987) as well as *Actinomyces* (J. Li et al., 2004; Dige et al., 2009). In contrast, mutans streptococci which comprise only 2% or less of the initial streptococcal population (Nyvad & Kilian, 1990), when a lesion is developed, the proportion of mutans streptococci in biofilm covering white spot enamel lesions was found to be higher when compared to the clinically healthy area (van Houte et al., 1991). However, the majority of bacteria found in white spot lesions are still non-mutans streptococci (Sansone et al., 1993; van Houte et al., 1996). In cavitated dentin lesions, mutans streptococci significantly increased to about 30% of the total flora (Loesche et al., 1984; Milnes & Bowden, 1985; Boue et al., 1987), suggesting that mutans streptococci are associated with extensive stages of caries. All of these studies proved that the microflora in the oral biofilm is changed during the caries lesions development. However, recent molecular identification methods have revealed that the microbial structure in the human oral cavity is much more diverse with over 700 prokaryote species (Paster et al., 2006; Chen et al., 2010), 50-60% of which are not cultivable (Aas et al., 2005; Aas

et al., 2008; Dewhirst et al., 2010). Compare with the traditional culture methods, next-generation sequencing of the 16S rRNA gene could discover about 3-times more unique bacterial species (S. Gupta et al., 2019). Therefore, using the evidence in the past that relied on culture-dependent techniques alone seems not enough to understand the relationship between microbes and the caries process. Instead, studying the oral microbiome as a whole microbial community using culture-independent techniques can uncover the complexities of the microbial community with new insights into the role of microbial variation during the caries process.

2.2.2. Human oral microbiome and dental caries

Humans are not autonomous organisms, instead, we are biological units that include abundant microbial symbionts and their genomes (Bordenstein & Theis, 2015). In fact, the human body accompanies roughly as many microbial cells as human cells (Sender et al., 2016). The community of our microbial residents is referred to as our "microbiome" to convey the ecology of commensal, symbiotic, and dysbiotic microbes that live with us and have been all but neglected as determinants of health and disease (Lederberg & McCray, 2001). The oral microbiome is the second most diverse microbial community in humans, behind the gut microbiome (Kilian et al., 2016), and plays an important role in maintaining oral homeostasis, protecting the oral cavity, and preventing disease development (Gao et al., 2018). The relationship between the oral microbiome and dental caries could be explained using historical evidence combined with oral microbiome studies. In humans, dental caries was assumed to be associated with the introduction of agriculture, resulting in the consumption of farmed plants and animals (Braidwood et al., 1961), with greater carbohydrate content compared with a natural diet (M. P. Richards et al., 2003). It is very rare for dental caries to be found among hunter-gatherers who broadly lived in the earlier period of time (Smith, 1984). Recent genetic analyses, both from the ancient fossils and currently available samples suggested that the transition of diet from hunter-gatherer to agricultural societies altered the composition of the oral microbiome and could be associated with the

development of dental caries (Contreras et al., 2010; Nasidze et al., 2011; Cornejo et al., 2013; Clemente et al., 2015). Moreover, during the Industrial Revolution, an increase in the consumption of processed flour and sugar did change the oral ecology in humans with the expansion of cariogenic bacteria which promoted dental caries to become a major endemic disease until this day (Adler et al., 2013).

2.2.3. Next-generation sequencing in caries research

Next-generation sequencing (NGS) of the 16S ribosomal RNA (rRNA) gene is a widely used technique for the investigation of the microbiome in the past decades, especially in the classification and identification of bacteria from biological samples. 16S rRNA is a component of the small subunit of ribosome in prokaryotes which is necessary for the synthesis of all prokaryotic proteins. The gene that encodes this RNA is one of the most conservative genes for all prokaryotes. However, the internal structure of this gene consists of both conserved and variable regions which interlace with each other (Figure 3). The conserved regions are shared by almost all prokaryotes and the variable regions are different among prokaryotes with different degrees due to their evolutionary relationships. According to this character, the identification could be performed by using the universal primers, designed to match the conserved regions, to pick up a specific region of 16S rRNA gene from all prokaryotes in samples, then amplify, sequence, and identify the generated sequences based on the similarity of variable regions compare to the reference 16S rRNA gene sequences available in public databases. This technique allows us to determine the relative abundance of all bacteria in the sample regardless of their cultivable ability. Due to the rapid increase in the availability of sequencing facilities and a decrease in sequencing cost over time, the use of 16S rRNA gene sequencing is not limited only to research fields, but may soon be implemented in clinical practices as well (Woo et al., 2008).

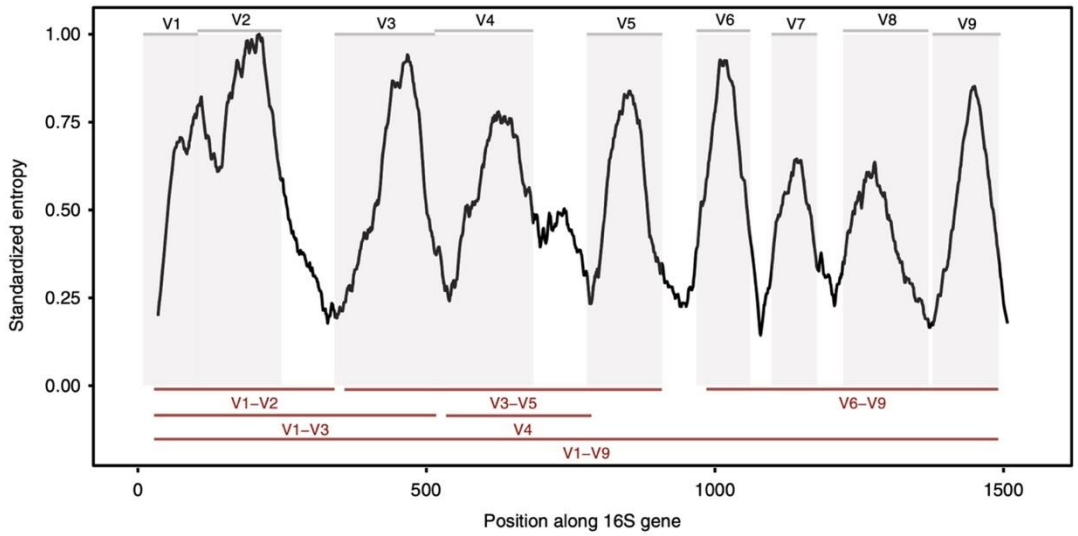


Figure 3: Shannon entropy across the 16S rRNA gene, based on the Greengenes database. V1-V9 refers to the variable regions of these genes (Johnson et al., 2019).

The NGS studies supported the findings from previous culture-based studies but also revealed a much more diverse bacteria associated with dental caries (Becker et al., 2002; Munson et al., 2004; Aas et al., 2008). Among the increasing numbers of studies about the oral microbiome in children with caries, even though members of mutans streptococci, in particular *Streptococcus mutans*, are still the key microbes found in many NGS studies (Kanasi et al., 2010; Soncini et al., 2010; Lif Holgerson et al., 2015; V. P. Richards et al., 2017; Wang et al., 2017; Qiao et al., 2018; Y. Zheng et al., 2018). However, there are a large number of acidogenic bacteria presented in the oral biofilm that could be caries-associated, such as members of the genera *Bifidobacterium, Propionibacterium,* and *Scardovia* (Munson et al., 2004; Downes et al., 2011; Tanner et al., 2011; Kaur et al., 2013). Moreover, *Veillonella* spp. has been found to have a role in caries-affected children (Kanasi et al., 2010; Xu et al., 2014; M. Zhang et al., 2015; Agnello et al., 2017; Wang et al., 2017; Y. Zheng et al., 2018) with the evidence suggesting that *Veillonella* may serve as a predictor for future caries (Gross et al., 2012). Furthermore, *Prevotella* spp. and *Lactobacillus* spp. were shown to have higher abundance in the caries-affected group compared with the caries-free group (Wang et al., 2017). In contrast, the abundance of specific taxa is reduced in the advanced stage of dental caries, for example, *Streptococcus mitis*

group, *Neisseria*, and *Streptococcus sanguinis* (Gross et al., 2012). At each stage of the caries process, gradual changes in the microbiota throughout the caries process had been found in previous NGS studies (W. Jiang et al., 2014; Kianoush et al., 2014; Rocas et al., 2016). The decrease in the diversity of bacteria was found as caries progressed from health to caries. The healthy subjects had a wider range of significantly enriched bacteria. In comparison, caries-affected subjects had a smaller number of enriched bacteria with different dominant genera at each stage, *Actinomyces* dominating white spot lesions, *Streptococcus* dominating enamel cavitated caries lesions, and *Lactobacilli* dominating dentin caries lesions (W. Jiang et al., 2014; Kianoush et al., 2014; Rocas et al., 2016). Nevertheless, many studies confirmed that dental caries was more complicated than being caused by a particular group of bacterial species (Yang et al., 2012; Kianoush et al., 2014; Rocas et al., 2016). The comparison of the oral microbiome between healthy and caries-active subjects showed that there were no "caries-specific" bacteria found completely absent in healthy subjects. Instead, the shifts in the abundance of bacteria were different between healthy and caries-active subjects and could be associated with dental caries (Yang et al., 2012).

Microbiomes across the body change rapidly in the first 3 years of life (Yatsunenko et al., 2012). In the oral cavity, the study about the maturation of the oral microbiome in children found a significant increase in species richness and taxa diversity from the age of 3 months to 3 years (Lif Holgerson et al., 2015). Moreover, specifically in 1-year-old children, their microbial diversity both within and between samples was significantly different when compared with other age groups (Dashper et al., 2019). In fact, the significant differences could be observed among the oral microbiome of 12, 18, and 24-month-old toddlers (F. Li et al., 2018), and fully eruption of all primary anterior teeth is a critical stage in the maturation of oral microbiota (Xu et al., 2022). Thus, data from children of a particular age could not apply to use with children of different ages. Most of the studies on children with caries focused on children aged below 7 years old or those with primary dentition with a few studies that collected samples from children younger than 1-year old (Ramli & Azmi, 2020). Therefore, oral microbiome studies in an early year of life are

needed to develop effective approaches in young children who are the target group for caries control strategy (Pitts et al., 2019).

Among the oral microbiome studies in children with caries, the variety of regions of the study was still limited, with most of the studies being conducted in the United States and China (Ramli & Azmi, 2020). However, there was considerable diversity in terms of racial differences because some studies focused on minorities in their country (Soncini et al., 2010; Han et al., 2016; Agnello et al., 2017), or mixed races in the same country were involved in their studies (Kanasi et al., 2010; Goldberg et al., 2015). Oral microbiome studies among healthy adults indicate that the diversity of the oral microbiome varies by geographical and racial variations (V. K. Gupta et al., 2017) indicating that the oral microbiome could be geographically dependent. Thus, information on the oral microbiome in children from various geographical regions is needed to expand the understanding of the relationship between the oral microbiome and dental caries in that particular population.

2.2.4. Salivary microbiome and dental caries

Traditionally, microbiological analysis of patients with dental caries has been focused on the supragingival plaque, which is closely related to cariogenesis. However, the acquisition of this type of sample could be difficult because the procedure requires both dental healthcare personnel and some specific instruments. Moreover, in the case of good oral hygiene individuals, dental plaque might not be enough for collection. Furthermore, the plaque microbiome is sensitive to the severity of caries lesions rather than the caries status of the host (V. P. Richards et al., 2017), so the plaque collecting procedure must be standardized to prevent the variation that could occur during the sample collection. Alternatively, a saliva sample is non-invasive and simple to collect and store which is easier for volunteers to cooperate (Bhattarai et al., 2018) even without the need of healthcare personnel as it is already available commercially as a self-collecting kit. Moreover, it is safe to handle, cost-effective, and contains high-quality DNA (Gura, 2008; C. Z. Zhang et al., 2016). Saliva is sterile when secreted from the glands (Schrøder et al., 2017).

However, after the secretion into the oral cavity, salivary microbiota consists of bacteria shed from the oral surface (Segata et al., 2012) that is considerably individualized (Hall et al., 2017) and temporarily stable (Cameron et al., 2015) regardless of the type of saliva that has been collected, unstimulated or stimulated (Jo et al., 2019). From the comparison between saliva and supragingival plaque, although saliva showed different microbial structures from supragingival plaque in terms of diversities, compositions, and functional characters, the salivary microbiota showed positive associations with the supragingival microbiota and might be possible to use for monitoring supragingival microbiota (Shi et al., 2018). Moreover, the analysis of saliva could provide insights into caries-causing microbes (Bhaumik et al., 2021).

Recent NGS-based studies reported caries-associated characteristics of salivary microbiota in children (Xiao et al., 2018; Hurley et al., 2019), adolescents (Eriksson, Lif Holgerson, Esberg, et al., 2017; Eriksson, Lif Holgerson, & Johansson, 2017), adults (Zhou et al., 2016), and elderly patients (Q. Jiang et al., 2018). The longitudinal study found that salivary microbiota has the potential to predict recurrent caries in 3-year-old children using the salivary levels of *Fusobacterium, Prevotella, Leptotrichia,* and *Capnocytophaga* species (Zhu et al., 2018). Moreover, when combined with the salivary levels of host defense peptides, the salivary microbiota could be used to predict the caries progression in 4-year-old children (Simon-Soro et al., 2018).

The ultimate goal of the caries prediction is to be able to make a precise prediction before the occurrence of a cavitated caries lesion. Thus, we can give proper prevention to the ones who need it right before the point of no return without wasting the limited resources, and the cost of prevention is much cheaper than the restoration or replacement. Moreover, it is easier to motivate the patients to improve themselves when we can show scientific data to illustrate their risk of disease, like blood cholesterol or blood pressure level. Finally, this information could be used to monitor the caries-risk status in each further recall interval.

2.3. Caries prediction in children

2.3.1. Current caries predictor

The etiology of dental caries in children is multifactorial, complicated, and can be viewed from various perspectives. Several comprehensive models had been proposed to explain the multilevel influences on children's oral health, related health disparities, and even connecting multiple factors to create unifying conceptual models (Fisher-Owens et al., 2007; Seow, 2012; Lee & Divaris, 2014). These models are exceptional representations of multilayered determinants of early childhood caries (ECC) at the population level, such as family education and socioeconomic disparity. Same as diet and specifically sugar intake that recently re-emerged as a major influence on caries incidence at the population level (Meyer & Lee, 2015; Sheiham & James, 2015). However, these population-derived determinants are both theoretically and practically different from the causes of individual cases (Rose, 1985). For example, dental caries can occur in both high and low socioeconomic children. To precisely predict the caries onset at the individual level, we need to find another model that does not use just the population-level parameters into consideration.

2.3.2. Cariogenic bacteria in currently used caries risk assessment

The well-known published caries risk assessment tools, including Cariogram (Bratthall & Hansel Petersson, 2005), CAMBRA (Featherstone et al., 2019), American Dental Association caries risk assessment (ADA, 2011), and American Academy of Pediatric Dentistry Caries risk assessment (AAPD, 2021), rely on the biological and environmental factors that mostly evaluated by the healthcare provider which gave the different result when predicting the future risk of dental caries (Featherstone et al., 2021). Besides identifying the risk factors derived from the details gathered during the risk assessment, the prediction of the risk level for future caries occurrence is also an important element for successful caries management. From the various predictors that have been chosen for caries prediction, the previous caries experience was the most powerful predictor, especially in pre-school children

(Mejàre et al., 2014). However, this predictor is not practical to be used in the case of toddlers since most of them are innocent of caries experience.

The microbial factors associated with the activity within biofilms are promising candidates for caries prediction because they could reflect conditions in the oral cavity deriving from other cariogenic factors, such as diet, oral hygiene, and the characteristics of saliva. However, the bacterial level shows a poor accuracy as a univariate model (Mejàre et al., 2014) and could only moderately improve the predictive ability when combine with other factors (Demers et al., 1990; Krasse, 1990; Hong & Hu, 2010). The reason is probably those microbial factors depended on a particular species count such as *Streptococcus mutans* and *Lactobacilli* spp. which could only reflect the current caries status (Hong & Hu, 2010; Sounah & Madfa, 2020) but might not be able to predict the caries onset in the future. The NGS studies in recent decades make us to capable of expanding the understanding of the microbial contributions to the etiology of dental caries beyond the knowledge from culture-dependent studies. Moreover, the differences in oral microbiota between healthy and caries-active children had been explained (Luo et al., 2012; S. Jiang et al., 2016; Hajishengallis et al., 2017). The discovery of various microbes associated with caries has been reported as mentioned suggests that the oral microbiome has the potential to be a robust predictor of dental caries in toddlers (Grassl et al., 2016; Hemadi et al., 2017; Kato et al., 2017).

2.3.3. Oral microbiome and caries prediction in children

There were studies that proposed caries predictive models based on oral microbiome data (Table 1). Teng and colleagues used both plaque and saliva samples to develop a model that could predict future ECC onsets with 81% of accuracy (Teng et al., 2015). This study collected the samples from 4-year-old children and then tracked them longitudinally for 2 years. Another study, which used the samples from 3-year-old children with a 1-year follow-up, showed that they could construct a caries-onset prediction model with an accuracy of 93.1% using supragingival microbiome profiles (Xu et al., 2018). Both studies, with closely related

race and age of the population, 4 to 6-year-old and 3 to 4-year-old Chinese children, respectively, showed that the genera *Streptococcus* and *Prevotella* were found to be most discriminatory. The most recent study, conducted in the United States in children aged 1 to 3 years, showed that salivary microbiota profile could be used for the prediction of ECC onset (Grier et al., 2021). This study showed different taxa of the important discriminants which *Rothia mucilaginosa, Streptococcus* sp., and *Veillonella parvula* were represented as biomarkers of risk for ECC onset. These differences might arise from several factors. First, the geographical difference could affect the oral microbiota both in terms of genetic and environmental, since these two regions have vastly different cultures and practices, especially about the infant feeding (Schulze et al., 2009). Second, the age of subjects was crucially different in terms of oral habitat for microflora, erupting vs full deciduous dentition, which reported that microbial richness and diversity were different (Lif Holgerson et al., 2015; Dashper et al., 2019). For example, *R. mucilaginosa* is a species that colonize and adhere to mucosal epithelial surfaces, so it could play a more important role in children with partial dentition than the complete-erupted dentition. Third, the types of samples that these studies used were all different, the combination of saliva and dental plaque (Teng et al., 2015) vs supragingival plaque (Xu et al., 2018) vs saliva (Grier et al., 2021). Lastly, they used different strategies to generate the prediction models. Although all of them used supervised machine learning based on oral microbiota composition but using different approaches. Grier and colleagues (2021) sought the signature of caries risk in the caries-free samples, rather than simpler implying to the oral microbiota of caries-free samples that are similar to caries-active samples are at higher risk of caries onset in the future, as were used in the previous studies (Teng et al., 2015; Xu et al., 2018). Moreover, they used the overall taxa to generate the prediction models rather than the selective biomarkers.

Table 1: Previous ECC prediction studies using machine learning based on oral microbiome.

Author (Year)	Type of model	Training group	Variables	Age (y)	Race	Algorithm	Accuracy (%)	Biomarkers
Grier et al. (2021)	Prediction	Caries-free	Salivary microbiome	1-5	American	RF, GB	73.2-85.5%, 71.4-83.6%	*R. mucilaginosa*, Streptococcus* sp.*, *V. parvula*,*

								etc.
Xu et al. (2018)	Prediction	Mixed caries status	Plaque microbiome	3-4	Chinese	RF	93.1%	*Streptococcus* sp.*, *Prevotella* spp.*, *Solobacterium* sp.*, *Kingella* sp.†, *Capnocytophaga* sp.†, *Neisseria* sp.†, *Fusobacterium* sp.†, etc.
Teng et al. (2015)	Prediction	Mixed caries status	Plaque and Salivary microbiome	4-6	Chinese	RF	81%	*Streptococcus* spp.*, *Prevotella* spp.*, etc.

Abbreviation: RF (Random Forest), GB (Gradient Booster)
**Enriched in caries-active children, †Enriched in caries-free children*

2.3.4. Machine learning in oral microbiome research

Microbiome data could be used for various analyzes such as to establish an association between microbiome and diseases, predict disease incidence, and classify various disease states. On this matter, machine learning (ML) can be used for in-depth analysis by generating models that can predict the outputs of interest. ML trains and evaluates the generated models to identify, classify, and predict patterns from provided data. Unsupervised ML analyzes and clusters unlabeled datasets by discovering the hidden pattern in data without intervention from humans, while supervised ML used the labeled datasets to train the model and learn the pattern of input features to classify or predict the outcome. ML techniques can be used for various purposes when applied to microbiome studies. First, for the classification of microbial taxa and taxonomic assignment. Second, for the prediction of the host phenotype by associating microbial relative abundance to disease incidence or severities, for example, disease prediction. Lastly, to understand the disease mechanisms, for example, biomarker-finding. The supervised learning methods used in this study were Naïve Bayes classifiers (NB), Linear discriminant analysis (LDA), and Random forests (RF). NB is integrated with QIIME2 (Bolyen et al., 2019) used for the taxonomic classification of 16S rRNA gene sequences. LDA is a generalization of Fisher's linear discriminant, a method used in statistics, to find a linear combination of variables by focusing on maximizing the separability among known categories. The LDA effect size (LEfSe) method, proposed by The Huttenhower Lab *(http://huttenhower.sph.harvard.edu/lefse)*, was designed for biomarker discovery in

metagenomic data, including 16S rRNA gene datasets (Segata et al., 2011). RFs are the ensemble method that combines multiple classifiers to achieve better performance compare with a single classifier. In this case, RFs are made by combining many decision trees. The final output of RFs is the majority voting of the individual decision trees. RF classifiers are the widely used ML algorithm in oral microbiome studies including caries prediction studies in children (Teng et al., 2015; Xu et al., 2018; Grier et al., 2021).

2.3.5. 1 year old: a critical age group for caries prevention

Caries prevalence in children could be controlled by giving preventive intervention within the first year of life (Pitts et al., 2019). With the limited resources, targeted prevention would be necessary for the distribution of adequate caries prevention to the high caries-risk children. Research on caries prediction in toddlers is largely missing and needed, especially in Thailand, to develop a caries prediction model that is reliable enough for clinical caries management in this region. The accurate caries risk prediction at the individual level, “precision dentistry”, is desirable and achievable but must be based on accumulated high-quality evidence (Divaris, 2016). In this study, we used the stored samples previously collected from Thai children in the project “Impact of prolonged and on demand breastfeeding on early childhood caries in Khon Kaen” (Sritangsirikul et al., 2021). This is a nested case-control design within a prospective cohort study focused on identifying the microbial biomarkers as the potential caries predictors and developing prediction models using a machine learning approach to predict future caries onset in caries-free 1-year-old children.

Chapter III

3. Materials and Methods

3.1. Research Design

A nested case-control within a prospective cohort study

3.2. Biosafety Consideration

This project was approved by The Institutional Biosafety Committee of the Faculty of Dentistry, Chulalongkorn University on May 27, 2021 (Approval No. DENT CU-IBC 016/2021).

3.3. Ethical Consideration

This project was approved by The Human Research Ethics Committee of the Faculty of Dentistry, Chulalongkorn University on May 7, 2021 (Study Code: HREC-DCU 2021-032). The cohort protocol from the previous study (Sritangsirikul et al., 2021) was approved by Khon Kaen University Ethics Committee for Human Research (HE592266) and Human Research and Ethics Committees of University of Washington (HSD52258).

3.4. Sample Size

With the limitation of sufficient sample quantity and caries status when children were 2 years old, a total of 40 saliva samples were used in this study. 30 samples, 10 samples from caries-free 1-year-old children in each of the 3 groups, were used for differences analyses, biomarker discovery, and machine learning training. The other 10 samples from 1-year-old children with different future caries development were used as a testing group for validating the performance of prediction models. The details of how to use samples for each analysis will be explained in the results section. According to a previous study on the salivary microbiome in children, the species accumulation curves suggested that the data

reached a saturated point at 10 samples/set (Zhu et al., 2018). This data indicated that the sample size of 10 samples is sufficient from the sequencing aspect.

3.5. Subjects and Grouping Criteria

The saliva samples in this study were the stored samples from the previous study (Sritangsirikul et al., 2021). The participants were recruited from the Thai Primary Health Centers and Mother and Child Health Center in Muang District, Khon Kaen province, Thailand. All participants were unrelated individuals, systemically healthy, and aged around 1 year old at the baseline time point (12.75m±1.10). The selected saliva samples used in this study were collected at the baseline time point in 2017. The information from oral examination and interview was collected at the baseline time point and repeated at the 6- and 12-month-follow-up visits. These time points were described as time point 1 (T1, baseline), time point 2 (T2, 6 months), and time point 3 (T3, 12 months) (Figure 4).

A total of 568 participants voluntarily agreed to the survey and sample collection and passed all of the following criteria;

1. The child was 1 year old when the survey was started, attending Thai Primary Health Centers and Mother and Child Health Center Muang District, Khon Kaen.
2. The child had a routine well-baby visit at the selected Thai Primary Health Care Centers and Mother and Child Health Center Muang District, Khon Kaen.
3. The child was a permanent resident in Khon Kaen without plans to relocate.
4. His/her caregiver can understand the Thai language.
5. The child had not taken any antibiotics for the last 30 days before the sample collection.

Of 331 one-year-old caries-free children, 30 samples from those children were selected for this study (age 12.57m 0.97), based on the sufficiency of the sample (volume at least 0.5 ml), the caries status at the baseline time point, and their longitudinal change of caries status at 1-year follow-up. From the oral examination, all of the selected subjects were not found any caries lesion at the baseline (T1) but had different caries statuses at the 1-year follow-up time point (T3),

including 10 children who remained caries-free, 10 children developed at least 1 non-cavitated caries lesion without any cavitated caries lesion, and 10 children who developed at least 1 cavitated caries lesion. These subjects were divided into three groups according to the longitudinal change of their caries statuses; caries-free to caries-free (**F2F**: caries-**F**ree to caries-**F**ree), caries-free to non-cavitated caries lesion (**F2W**: caries-**F**ree to **W**hite spot), and caries-free to cavitated caries lesion (**F2D**: caries-**F**ree to **D**ecay), respectively (Figure 4). In addition, another 10 saliva samples from 1-year-old cavitated caries lesion-free children were selected to validate the performance of caries prediction models (testing group).

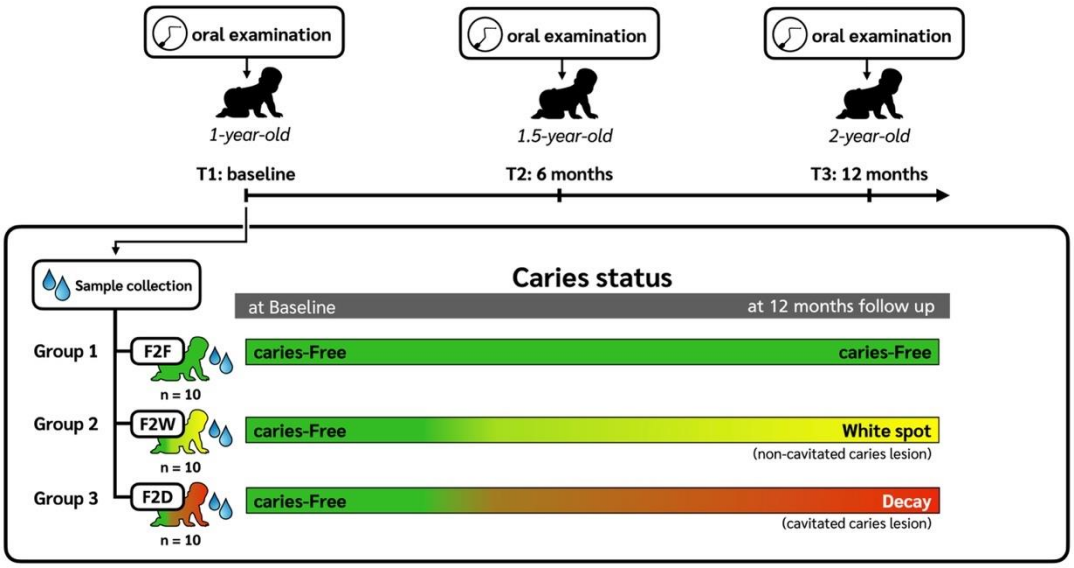


Figure 4: Schematic of sample collection, data collection, and grouping of samples. F2F, caries-Free-to-caries-Free; F2W, caries-Free-to-White spot (non-cavitated caries lesion); F2D, caries-Free-to-Decay (cavitated caries lesion), based on caries status at 12-month-follow-up.

3.6. Saliva Sampling

The saliva sampling protocol was described in the previous study (Sritangsirikul et al., 2021). In brief, the saliva collection was performed approximately at the same time of the day to avoid changes due to the circadian rhythm (Dawes, 1972). Approximately 1-3 ml of saliva was obtained from each subject by dropper and saturating gauze in saliva that pooled on the floor of mouth. The whole process took around 3-5 minutes to be completed. The pH of saliva was immediately

measured using LAQUA twin pH 22 Pocket meter (HORIBA Instruments, Singapore) The sensor was calibrated before each measurement using standard buffer solutions of pH 4.0 and 7.0. Measurements were repeated in duplicates for each sample. The remaining sample was then transferred into the glycerol solution in a sterile plastic tube and kept in the transport container with frozen packs at 2-6℃ until they were delivered to the laboratory at the Department of Oral Biology Laboratory, Faculty of Dentistry, Khon Kaen University to be stored at -80℃.

Before the experiment of this study was started, all of the available samples were transferred to the laboratory at the Department of Microbiology, Faculty of Dentistry, Chulalongkorn University to be stored at -80℃ for sample categorization prior to the DNA extraction.

3.7. Data Collection and Analysis

The data from each subject was collected in the previous study (Sritangsirikul et al., 2021), including caries experiences, age and sex of children, number of erupted teeth, human milk feeding status, and pH of saliva. At baseline (T1), 6-month-follow-up (T2), and 12-month-follow-up (T3) visits, each subject was examined his/her teeth by one calibrated dentist. Caries status was assessed using the WHO diagnostic criteria (WHO, 1997) for visible cavitated and non-cavitated caries lesions. The number of cavitated caries lesion on deciduous teeth (d) was used with the number of affected teeth (dt) and surfaces (ds) as the numerator. Moreover, the d_1 subgroup for d was used to record caries lesions that extend to non-cavitated caries lesions (d_1t and d_1s), d_1 is a detectable enamel lesion with a sound surface (Table 2). All tooth surfaces were carefully wiped with dry gauze to remove dental plaque before the examination. Under artificial light, the examination was conducted using a visual examination with a mouth mirror, while the subject was in the supine position. After each examination, the caregiver received a dental report card indicating further dental treatment if needed. In case of urgent or painful conditions, they had been referred to the Pediatric Department, Faculty of Dentistry, Khon Kaen University, or their family dentist, as appropriate. They also received the oral hygiene home-care

recommendation and the oral hygiene pack consisted of a toothbrush, toothpaste, and handkerchief at each visit. The amount and position of caries lesions at each time point were recorded separately.

Table 2: The number of caries lesions of subjects at each time point.

Sample #	Baseline (T1, 1-y-old)				6m-follow-up (T2, 1.5-y-old)				12m-follow-up (T3, 2-y-old)			
	dt	ds	d_1t	d_1s	dt	ds	d_1t	d_1s	dt	ds	d_1t	d_1s
F2F01	0	0	0	0	0	0	0	0	0	0	0	0
F2F02	0	0	0	0	0	0	0	0	0	0	0	0
F2F03	0	0	0	0	0	0	0	0	0	0	0	0
F2F04	0	0	0	0	0	0	0	0	0	0	0	0
F2F05	0	0	0	0	0	0	0	0	0	0	0	0
F2F06	0	0	0	0	0	0	0	0	0	0	0	0
F2F07	0	0	0	0	0	0	0	0	0	0	0	0
F2F08	0	0	0	0	0	0	0	0	0	0	0	0
F2F09	0	0	0	0	0	0	0	0	0	0	0	0
F2F10	0	0	0	0	0	0	0	0	0	0	0	0
F2W01	0	0	0	0	0	0	6	6	0	0	4	4
F2W02	0	0	0	0	0	0	2	2	0	0	8	8
F2W03	0	0	0	0	0	0	6	6	0	0	6	6
F2W04	0	0	0	0	0	0	0	0	0	0	6	6
F2W05	0	0	0	0	0	0	8	8	0	0	10	10
F2W06	0	0	0	0	0	0	4	4	0	0	4	4
F2W07	0	0	0	0	0	0	0	0	0	0	6	6
F2W08	0	0	0	0	0	0	0	0	0	0	6	6
F2W09	0	0	0	0	0	0	0	0	0	0	8	8
F2W10	0	0	0	0	0	0	4	4	0	0	3	3
F2D01	0	0	0	0	1	1	4	4	2	2	6	6
F2D02	0	0	0	0	6	6	6	6	12	13	12	13
F2D03	0	0	0	0	2	2	5	5	2	2	4	4
F2D04	0	0	0	0	0	0	0	0	1	1	11	11
F2D05	0	0	0	0	0	0	3	3	2	2	6	6

Sample #	Baseline (T1, 1-y-old)				6m-follow-up (T2, 1.5-y-old)				12m-follow-up (T3, 2-y-old)			
	dt	ds	d_1t	d_1s	dt	ds	d_1t	d_1s	dt	ds	d_1t	d_1s
F2D06	0	0	0	0	0	0	4	4	12	12	12	12
F2D07	0	0	0	0	0	0	6	6	4	12	10	18
F2D08	0	0	0	0	0	0	4	4	3	4	5	6
F2D09	0	0	0	0	0	0	0	0	4	4	4	4
F2D10	0	0	0	0	0	0	0	0	2	2	6	6

Yellow: found non-cavitated caries lesion without cavitated caries lesion

Red: found cavitated caries lesion

d is the number of cavitated caries lesions by teeth (dt) or surfaces (ds) in the deciduous teeth.

d_1 is the number of caries lesions both cavitated and non-cavitated by teeth (d_1t) or surfaces (d_1s).

The caregivers were interviewed by the structured questionnaire, modified from the questionnaire developed by the WHO and 44 consortium members from 18 countries and funded by the National Institute of Dental and Craniofacial Research (NIDCR) in 2006 and the Indicators for assessing infant and young child feeding practices: conclusions of a consensus meeting held 6-8 November 2007 in Washington DC., USA. (Sritangsirikul et al., 2021).

All data were described using percentages for categorical data and mean and standard deviation for continuous data. The comparisons of characteristics of subjects and samples between study groups were performed using Kruskal-Wallis test, One-Way Analysis of Variance (ANOVA) test, and Pearson's chi-squared test.

3.8. Extraction of Genomic DNA

The workflow of experiments from DNA extraction to bioinformatic analyses is illustrated in Figure 5. Total genomic DNA extraction was performed using DNeasy PowerSoil kit (Qiagen, Hilden, Germany) according to the manufacturer s instructions. Concentration and purity testing of the DNA were performed using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA samples were quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). The extracted DNA samples were stored at -80°C until further use.

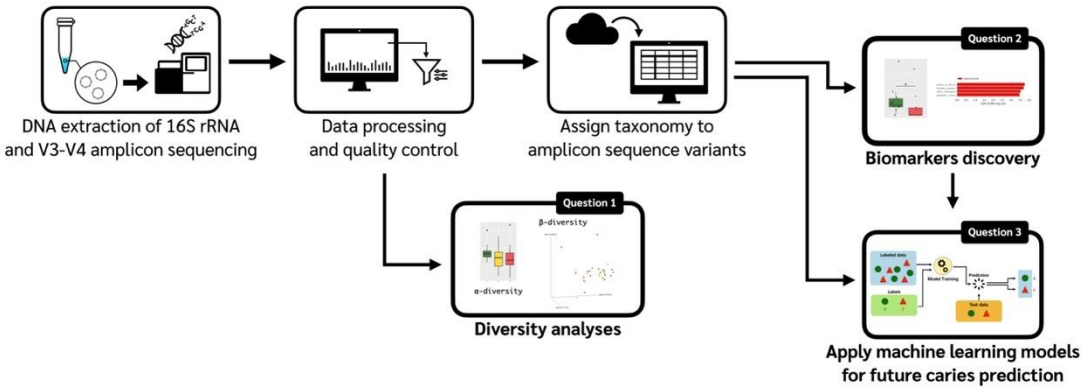


Figure 5: Workflow of the experiments and analysis pipelilne.

All of the extracted DNA samples were stabilized with active chemical protection using GenTegra-DNA (GenTegra LLC, Pleasanton, CA, USA) and shipped to Vishuo Biomedical laboratory (Singapore) for the further sequencing process.

3.9. Amplicon Generation, Library Preparation, and Sequencing

Next-generation sequencing library preparations and Illumina MiSeq sequencing were conducted at Vishuo Biomedical laboratory (Singapore). A total of 30-50 ng of DNA was used to generate 16S rRNA amplicons using a MetaVx™ Library Preparation kit (GENEWIZ, Inc., South Plainfield, NJ, USA). V3 and V4 hypervariable regions of prokaryotic 16S rRNA gene were selected for generating amplicons and following taxonomy analysis. The laboratory designed a panel of proprietary primers aimed at relatively conserved regions bordering the V3 and V4 hypervariable regions of bacteria and archaea 16S rRNA gene. The V3 and V4 regions were amplified using forward primers containing the sequence CCTACGGRRBGCASCAGKVRVGAAT" and reverse primers containing the sequence GGACTACNVGGGTWTCTAATCC" (Teng et al., 2018). The first-round PCR products were used as templates for the second-round amplicon enrichment PCR. At the same time, indexed adapters were added to the ends of the 16S rRNA gene amplicons to generate indexed libraries ready for downstream NGS sequencing on Illumina Miseq.

DNA libraries were validated by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit 2.0 Fluorometer

(Invitrogen, Carlsbad, CA, USA). DNA libraries were multiplexed and loaded on an Illumina MiSeq instrument according to the manufacturer s instructions (Illumina, San Diego, CA, USA). Sequencing was performed using a 2x300/2x250 paired-end (PE) configuration; image analysis and base calling were conducted by the MiSeq Control Software (MCS) embedded in the MiSeq instrument.

3.10. Bioinformatics and Statistical Analyses

3.10.1. Data processing

Raw data from the Illumina MiSeq was first converted into FASTQ format paired-end sequence files using the bcl2fastq program, version 1.8.4, provided by Illumina. QIIME 1.9.1 was used for demultiplexing, by extracting the barcodes and sorting sequenced reads into separate files for each sample, then imported into QIIME2 V2021.8 (Bolyen et al., 2019), which was used to perform all subsequent processing.

Denoising was performed using DADA2 (The Divisive Amplicon Denoising Algorithm) (Callahan et al., 2016): forward and reverse reads were truncated to 240 bps, inferred exact amplicon sequence variants (ASVs) from high-throughput amplicon sequencing data to the sequence variants and their sample-wise abundances after removing substitution and chimera errors. Phylogenetic trees were constructed for each cohort using MAFFT (Katoh & Standley, 2013) for sequence alignment and FastTree (Price et al., 2010) for tree construction.

3.10.2. Alpha Diversity and Beta Diversity Analyses

All diversity metrics were calculated on the rarefied samples using QIIME2. Alpha diversity (**α**-diversity) is the mean species diversity within the sample. The meaning of species diversity includes richness which simply quantifies how many different species contains within the sample, evenness which refers to how close an abundance of each species within a sample is, and diversity which takes both count of species and the abundance of each species into consideration. All of these 3 types of **α**-diversity were calculated using the Chao1, Pielou's evenness, and

Shannon indices, respectively. Comparisons of α-diversity between groups were performed with the Kruskal-Wallis test and pairwise comparison using their *P* values with Benjamini-Hochberg correction. Beta diversity (β-diversity) is the differentiation among samples. β-diversity was calculated using the weighted and unweighted UniFrac (the unique fraction) metrics and then was visualized using the principal coordinate analysis (PCoA). Dissimilarity analysis of β-diversity was calculated using Analysis of similarities (ANOSIM) (Clarke, 1993).

3.10.3. Taxonomic Assignment

Taxonomic classification was performed with a custom Naïve Bayesian classifier trained on the February 2021 release of the expanded Human Oral Microbiome Database (eHOMD, *https://www.homd.org*) (WHO, 2008; Escapa et al., 2018) which has taxonomic categories predicted from phylum to the species level that used for the further analyses. In addition, the OSU CORE oral microbiome database (Griffen et al., 2011), the SILVA 138 database (Quast et al., 2013), and the Greengenes database (McDonald et al., 2012) were used for taxonomic assignment for the comparison of different databases usage.

3.10.4. Biomarker Discovery

Multiple approaches were used for biomarker discovery. First, the comparison of differentially abundant microbial species between F2F and F2D groups was performed with the Wilcoxon rank-sum test using R software (version 4.1.0) with package stats 4.1.0. Next, the mean differences in the relative abundance of species between groups were evaluated using the linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al., 2011), with an alpha value of 0.05 for the Wilcoxon test and a threshold of 2.0 for logarithmic LDA scores using the Galaxy web application (*http://huttenhower.sph.harvard.edu/galaxy*).

3.10.5. Caries Prediction Model Development

3.10.5.1. All-taxa model

Caries prediction models were constructed using Random Forest (RF) classifiers in R software with package randomForest 4.6-14 based on species-level taxa. “All-taxa model” was first constructed based on the relative abundance of the all species-level taxa of F2F and F2D groups.

3.10.5.2. Important-features model

Next, the sets of biomarkers were selected using 2 approaches to further develop caries prediction models. First, “Important-features model”, the important taxa were determined by their importance to the accuracy of the All-taxa model, using two variable important measures, mean decrease in accuracy (MDA) and mean decrease Gini (MDG). The important features were selected using 2 methods, Boruta (package Boruta 7.0.0) and recursive feature elimination (RFE) algorithms (package caret 6.0-88).

3.10.5.3. Differential-abundance model

Second, “Differential-abundance model”, the microbial biomarkers, were identified as the differentially abundant species between F2F and F2D groups using the Wilcoxon rank-sum test and as the biomarkers of each group using LEfSe, as described above. The modified prediction models were constructed based on the relative abundance of both sets of selected species.

3.10.5.4. Single-species models

Furthermore, the simplified version of prediction models, “Single-species model”, was constructed based on the relative abundance of every single species of the important taxa and microbial biomarkers, selected as described above.

3.10.6. Model Validation

The performance of prediction models was validated using the salivary microbiome at the species level of the samples from unrelated 1-year-old cavitated-caries-lesion-free children in the testing group (Table 3).

Table 3: The number of cavitated caries lesion and d1s index in testing group.

Sample#	Baseline (T1, 1-y-old)		6m-follow-up (T2, 1.5-y-old)		12m-follow-up (T3, 2-y-old)	
	ds	d_1s	ds	d_1s	ds	d_1s
T01	0	4	0	12	2	8
T02	0	8	0	6	3	4
T03	0	4	0	12	1	7
T04	0	2	0	0	0	6
T05	0	2	0	4	0	4
T06	0	4	0	2	0	12
T07	0	4	4	5	6	6
T08	0	4	0	6	0	6
T09	0	4	0	9	0	2
T10	0	4	4	16	39	41

Red: found cavitated caries lesion

ds is the number of surfaces of cavitated caries lesion.

d_1s is the sum of the number of surfaces of cavitated and non-cavitated caries lesion.

Moreover, the cross-study validation was performed using the selected salivary microbiome data, retrieved from BioProject ID PRJNA622300 (Grier et al., 2021). The selection criteria were the age of children under 2 years old. 17 samples (age 20.47 months ± 2.61) were selected and used as the validating group (Table 4).

Table 4: The selected samples for validation were retrieved from the publicly available dataset through NCBI accession number PRJNA622300 (Grier et al., 2021).

Sample#	sample-id	Status	Host Age (months)
V01	SRR11458012	Pre-caries	22.67
V02	SRR11458024	Pre-caries	23.27
V03	SRR11458035	Pre-caries	19.17
V04	SRR11458047	Pre-caries	17.83
V05	SRR11458055	Healthy Caries Free	22.63
V06	SRR11458062	Healthy Caries Free	23.60
V07	SRR11458082	Healthy Caries Free	22.70
V08	SRR11458088	Pre-caries	21.27
V09	SRR11458091	Healthy Caries Free	20.07
V10	SRR11458098	Healthy Caries Free	16.03
V11	SRR11458102	Healthy Caries Free	20.70
V12	SRR11458105	Pre-caries	21.30
V13	SRR11458109	Pre-caries	17.17
V14	SRR11458139	Pre-caries	21.60
V15	SRR11516738	Healthy Caries Free	16.13
V16	SRR11516749	Healthy Caries Free	23.73
V17	SRR11516753	Healthy Caries Free	18.17

Healthy Caries free: Children who remained caries free for the 2-y study period

Pre-caries: Children with ECC onset within 2-y study period

Chapter IV

4. Results

4.1. Overview of Participants and Samples

The selected children who remained caries-free (F2F) versus whom developed non-cavitated caries lesions (F2W) and cavitated caries lesions (F2D) within 12 months were similar in age, sex, number of teeth, the status of human milk consumption, and pH of saliva but significantly different in terms of caries experience at 12-month-follow-up (Table 5).

Table 5: Characteristic table for the study cohorts and samples in each group at the baseline time point and the number of caries lesions at the 12-month-follow-up, presented as mean ± SD or %.

Variable	F2F (n = 10)	F2W (n = 10)	F2D (n = 10)	*P value*
Age, months	12.7 ± 1.3	12.6 ± 1.1	12.4 ± 0.5	0.99*
Sex				0.87†
male	50%	60%	50%	
female	50%	40%	50%	
No. of teeth at 1-y old	5.7 ± 2.2	5.4 ± 2.2	4.0 ± 1.9	0.19*
No. of teeth at 2-y old	17.4 ± 2.3	16.6 ± 1.3	15.8 ± 1.3	0.28*
Weaned (human milk)	90%	90%	60%	0.15†
pH of saliva	7.2 ± 0.3	7.2 ± 0.6	7.4 ± 0.6	0.57‡
ds§ at 12m-follow-up	0[a]	0[a]	5.4 ± 4.9[b]	<0.001*
d_1s¶ at 12m-follow-up	0[a]	6.1 ± 2.1[b]	8.6 ± 4.7[b]	<0.001*

**Kruskal-Wallis test, †Chi-squared test, ‡ANOVA test*

§ds is the number of cavitated caries lesion tooth surfaces in the deciduous teeth.

¶d_1s is the number of non-cavitated and cavitated caries lesion tooth surfaces in the deciduous teeth.

Lowercase letters indicate a significant difference among groups (P<0.05).

The salivary microbiota composition was determined by 16S rRNA gene sequencing. Illumina MiSeq sequencing produced an average of >57,000 reads per sample after quality control and amplicon sequence variant identification among all sequenced samples. The minimum frequency among all samples was 18,207 with an average of 28,489.7 per sample. SILVA (Quast et al., 2013), Greengenes (DeSantis et al., 2006), OSU CORE (Griffen et al., 2011), and eHOMD (Escapa et al., 2018) databases were used to assign the taxonomy of the dataset. The 15-most-abundance genera and the 25-most-abundance species based on each database are showen in figure 6 and 7, respectively. At the species level, the SILVA and Greengenes databases could classify the dominant species-level taxa as unclassified species with only 7 and 9 out of 25 most abundant species that could be classified, respectively (Figure 7A and B). The eHOMD and OSU CORE databases were able to classify the species-level taxa of 19 and 21 out of the 25 most abundant species, respectively (Figure 7C and D). Therefore, the eHOMD database was chosen for further analysis since it is the most up-to-date oral microbiome-specific database that was publicly available at that moment, the last update of eHOMD was made in 2021 *(https://ehomd.org/download#refseq)*, while the OSU CORE database was updated in 2017 *(http://microbiome.osu.edu)*. The reads from this dataset represent 945 unique features, 160 distinct species-level taxonomic assignments, 74 genera, 46 families, 31 orders, 20 classes, and 11 phyla, using the eHOMD database. The top 15 most abundant genera accounted for 94.70% of the overall composition across all samples, with *Streptococcus* being the most abundant (29.15% overall abundance), followed by *Alloprevotella* (12.52%), *Veillonella* (9.99%), *Haemophilus* (8.84%), *Leptotrichia* (6.18%), and *Neisseria* (6.09%) (Figure 6C), which all of the genera were not significantly different among these 3 groups ($P>0.05$, Kruskal-Wallis). The top 25 most abundant species accounted for 87.91% of the overall composition across all samples, of which 6 out of 25 were unclassified species (Figure 7C).

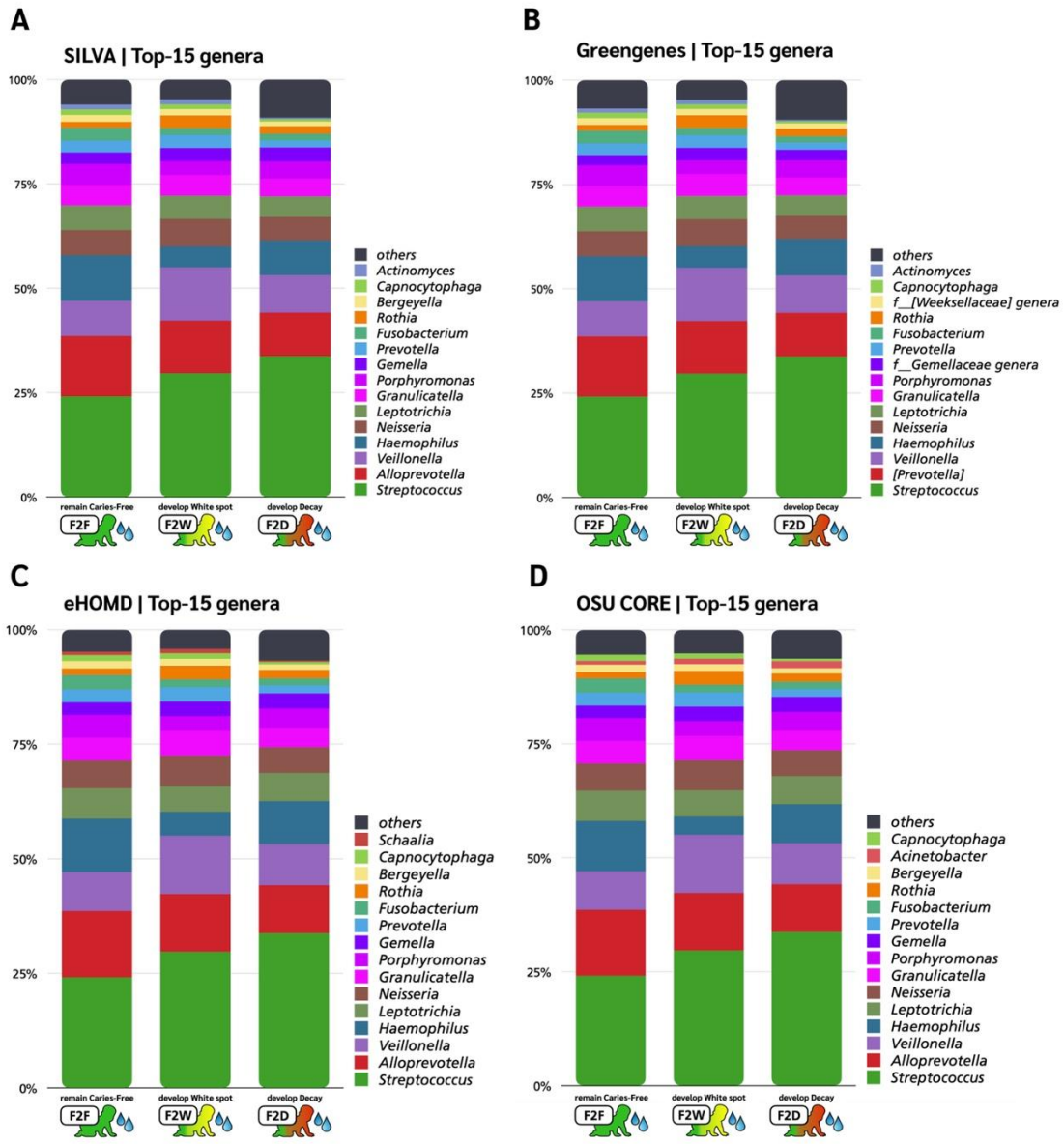


Figure 6: Relative abundance of bacterial genera in the salivary microbiome of caries-free 1-year-old children by caries status at 12-month-follow up using (A) SILVA, (B) Greengenes, (C) eHOMD, and (D) OSU CORE databases.

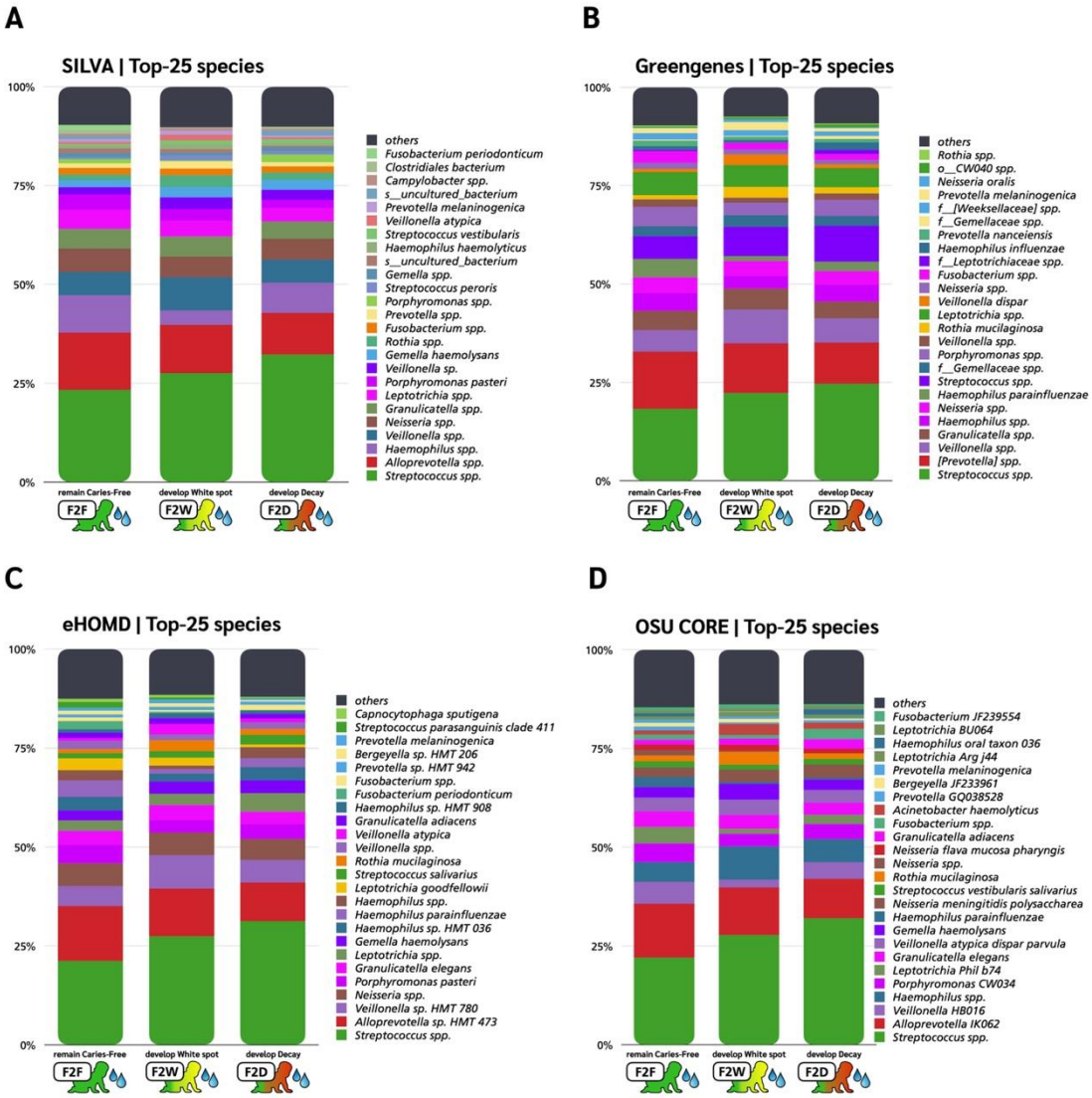


Figure 7: Relative abundance of bacterial species in the salivary microbiome of caries-free 1-year-old children by caries status at 12-month-follow up using (A) SILVA, (B) Greengenes, (C) eHOMD, and (D) OSU CORE databases.

4.2. Microbiota Composition prior to the Caries Development

The diversity analyses were performed both within-sample (**α**-diversity) and between-samples (**β**-diversity). For **α**-diversity, the children who remained caries-free (F2F) showed the most diverse salivary microbiome but were not significant in terms of diversity and richness (Shannon and Chao1 s alpha diversity, *P*=0.11 and 0.30, respectively, Kruskal-Wallis; Figure 8A-B). Only the relative evenness of species richness of the children who remained caries-free was significantly higher than those

who develop non-cavitated caries lesions (F2W), evaluated using Pielou's evenness (*P*=0.024, Kruskal-Wallis with Benjamini-Hochberg correction; Figure 8C).

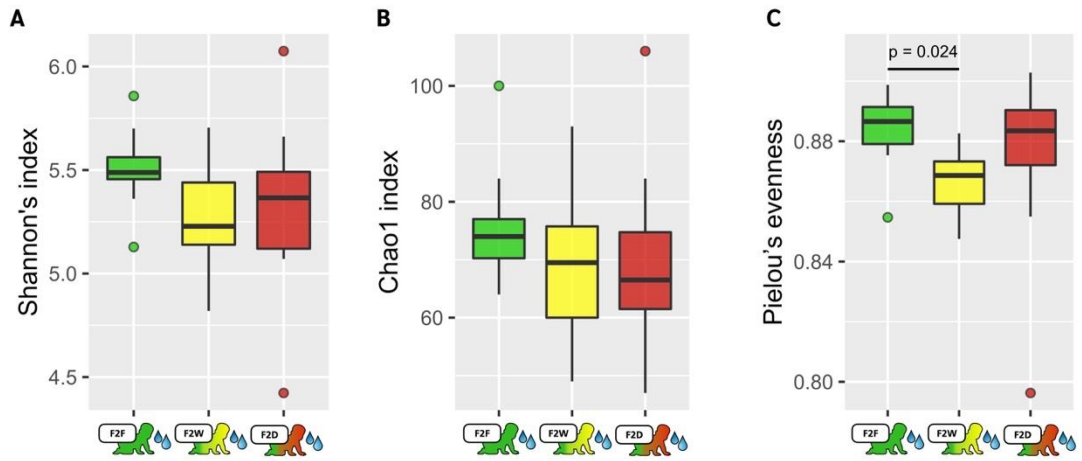


Figure 8: Boxplots of **α***-diversity, measured by (A) Shannon diversity index, (B) Chao 1 index, and (C) Pielou's evenness index. Green boxplots represent F2F group, yellow boxplots represent F2W group, and red boxplots represent F2D group. Comparison between F2F and F2W groups was significantly different in terms of evenness (P=0.024, Kruskal-Wallis with Benjamini and Hochberg correction).*

For **β**-diversity, principal coordinates analysis (PCoA) was conducted based on weighted and unweighted UniFrac distance. The result showed that, based on unweighted UniFrac distance, the salivary microbiota of the children who remained caries-free (F2F) was significantly different from those who developed cavitated caries lesions (F2D) (*P*=0.042, ANOSIM with Benjamini-Hochberg correction; Figure 9A). However, the differences were not significant when using weighted UniFrac distance (*P*=0.583, ANOSIM; Figure 9B). Therefore, the comparison between F2F and F2D groups was used for further analyses.

4.3. Microbial Biomarkers for Future Caries Prediction

To identify the potential microbial biomarkers for predicting future caries. First, the univariate analyses were used to compare the relative abundance of each individual species between the F2F and F2D groups using the Wilcoxon rank-sum test. *Prevotella nanceiensis, Leptotrichia sp. HMT 215, Prevotella melaninogenica*,

and *Campylobacter concisus* were found significantly higher in the F2F group (*P*=0.024, 0.040, 0.049, and 0.049, respectively, Wilcoxon rank-sum test; Figure 10).

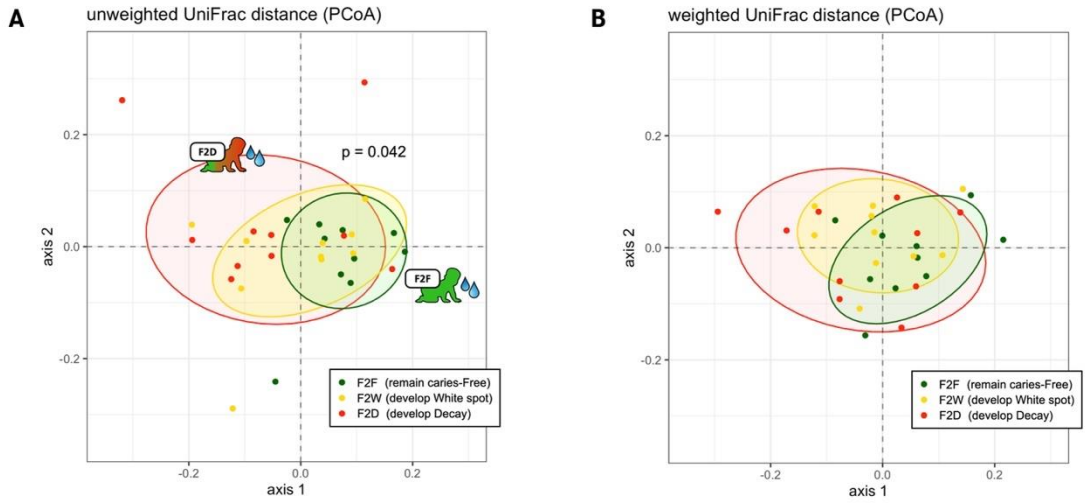


Figure 9: Principal coordinate analysis (PCoA) plot of β-diversity based on (A) the unweighted UniFrac and (B) the weighted UniFrac distance matrices, with samples colored by groups, green for F2F, yellow for F2W, and red for F2D. Based on the unweighted UniFrac distance, there was a significant difference between F2F and F2D groups (P=0.042, ANOSIM with Benjamini and Hochberg correction).

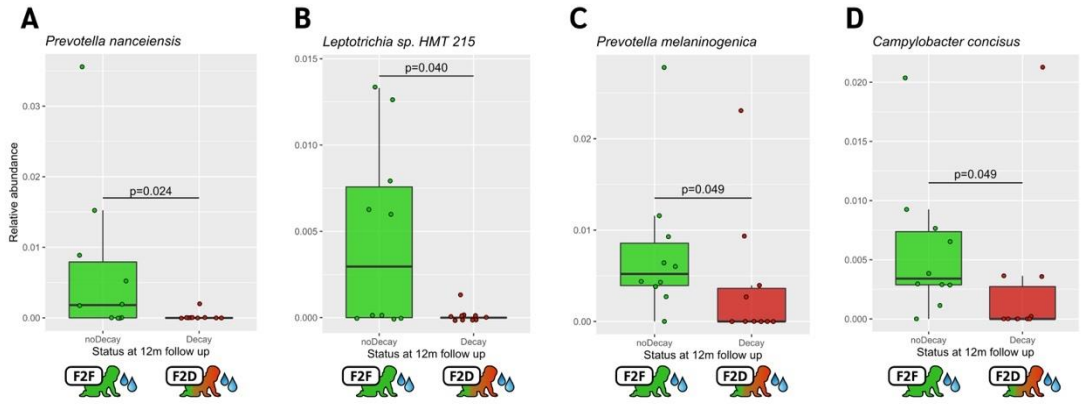


Figure 10: Boxplots of differential relative abundant microbes between F2F (green) and F2D (red) groups. There were 4 species found to be significantly different which were (A) Prevotella nanceiensis, (B) Leptocrichia sp. HMT 215, (C) Prevotella melaninogenica, (D) Campylobacter concisus.

Next, the specifically designed method for biomarker discovery in 16S rRNA gene sequencing data, Linear discriminant analysis (LDA) Effect size (LEfSe) (Segata et al., 2011), was performed to identify the biomarker species of the F2F and F2D groups. LDA focuses on maximizing the separation between the classes, by increasing the variability between the classes while decreasing it within the classes, to evaluate their differences. LEfSe identifies the important features that could explain those differences. The results were consistent with the previous analysis which revealed that *Leptotrichia sp. HMT 215, Prevotella nanceiensis, Prevotella melaninogenica,* and *Campylobacter concisus*, same as the previous result, are the biomarkers of the F2F group (LDA Score=3.74, 3.69, 3.52, and 3.46, respectively; Figure 11). It could imply that if these 4 species are found in relatively high abundance, there would be a lower risk of caries developing a year later suggesting their potential to be used as the caries predictor in the saliva of caries-free 1-year-old children.

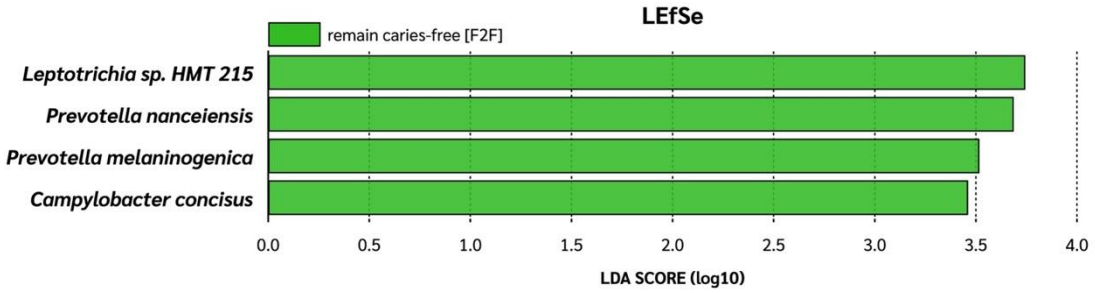


Figure 11: Histogram of the LDA scored for differentially abundant features between F2F and F2D groups. The threshold on the logarithmic LDA score for discriminative features was set to 2.0.

4.4. Caries Prediction Models Based on Salivary Microbiome

Based on our finding that the salivary microbiome structural differences could be observed prior to the onset of dental caries when compare between the children who remained caries-free (F2F) and those who developed cavitated caries lesions within 12 months (F2D). The supervised machine learning models (Random Forest; RF) for future caries prediction were developed using the combination of salivary microbiome data from these 2 groups, as the training group. Then validated their predictive performances using another set of salivary microbiome data from

unrelated 1-year-old cavitated-caries-lesion-free children, as the testing group. "All-taxa model" was first generated based on the relative abundance of overall microbiota composition including all species-level taxa as joint predictors. This model could give an accuracy of 70%, sensitivity of 60%, and specificity of 80% (AUC, 0.7; 95% confidence interval (CI), 34.8-93.3; Figure 12A).

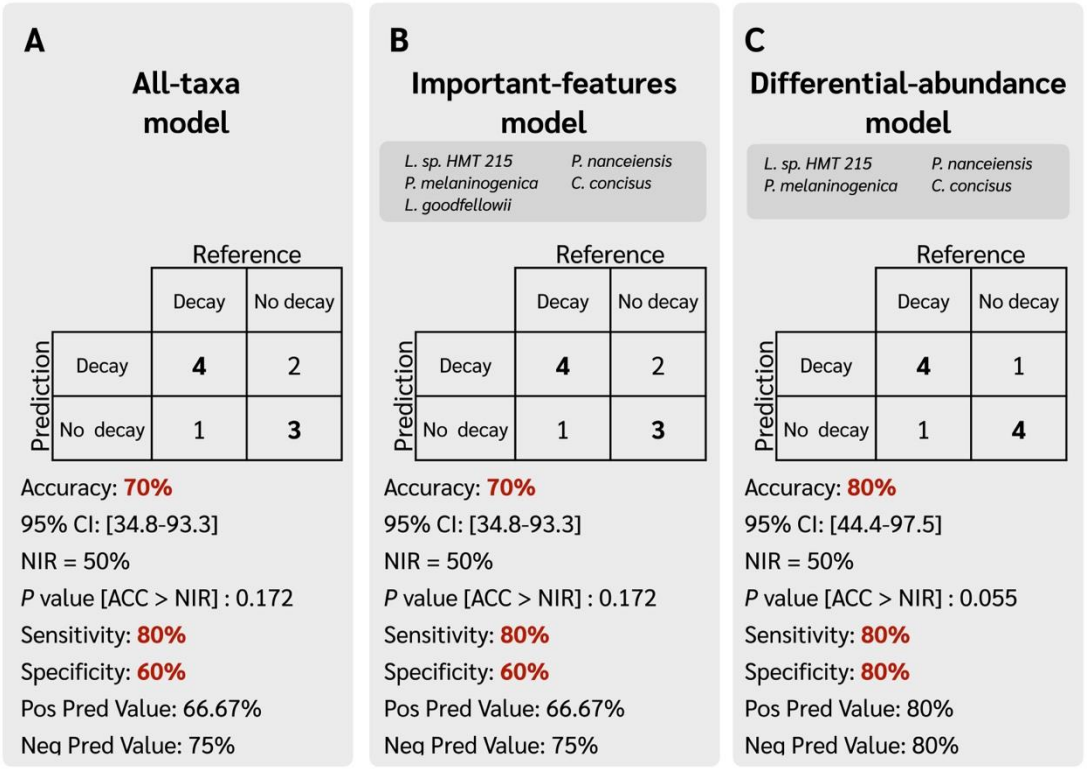


Figure 12: Confusion matrix evaluation of RF models based on multi-species relative abundance, including (A) All-taxa model, (B) Important-features model, and (C) Differential-abundance model. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and P value [Acc >No Information Rate] of models are shown, as calculated by caret package for R.

To improve the generalization of the model, the overfitting was corrected by developing RF models based on only the meaningful features to reduce the noise that could be picked up during the model training process, using 2 approaches. First, "Important-features model", this RF model was generated based on the relative abundance of the species that were important to the performance of the All-taxa model. *Prevotella melaninogenica, Campylobacter concisus, Prevotella nanceiensis,*

Leptotrichia sp. HMT 215, and *Leptotrichia goodfellowii* were selected as the important features due to their importance to the accuracy of the All-taxa model using Boruta (Kursa & Rudnicki, 2010) and recursive feature elimination (RFE) algorithms (Figure 13). This model exhibited an equal predictive performance when compared to the All-taxa model, giving an accuracy of 70%, sensitivity of 60%, and specificity of 80% (AUC, 0.7; 95% CI, 34.8-93.3; Figure 12B). Next, "Differential-abundance model", this RF model was generated based on the relative abundance of the identified biomarker species from previous analyses including *Prevotella nanceiensis, Leptotrichia sp. HMT 215, Prevotella melaninogenica,* and *Campylobacter concisus.* This approach could slightly improve the predictive performance of the model, giving an accuracy of 80%, sensitivity of 80%, and specificity of 80% (AUC, 0.8; 95% CI, 44.4-97.5; Figure 12C).

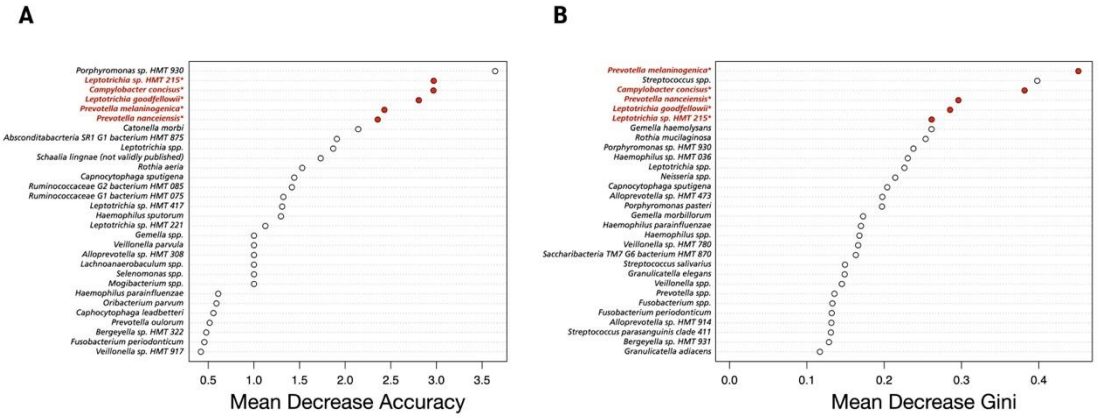


Figure 13: Caries-predictive taxa were determined by applying Random Forests analysis using the overall species dataset against the future caries status (All-taxa model). Bacterial taxa that are most discriminatory were ranked in descending order of their importance to the accuracy of the model. Their importance was determined based on (A) the mean decrease in accuracy (MDA) of microbiota prediction when the relative abundance of each species was randomly permuted and (B) mean decrease in Gini coefficient which is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest. The species showed in red with asterisk were the confirmed important features by Boruta algorithm and Recursive Feature Elimination (RFE).

Next, the simplified version of caries prediction models was further generated based on each species of potential biomarkers and important features, "Single-

species models". Only the model that used the relative abundance of *Campylobacter concisus* exhibited a predictive performance that was comparable to the multi-species models, giving an accuracy of 80%, sensitivity of 80%, and specificity of 80% (AUC, 0.8; 95% CI, 44.4-97.5; Figure 14A). For the other Single-species models, their performances were reduced drastically (Figure 14B-E).

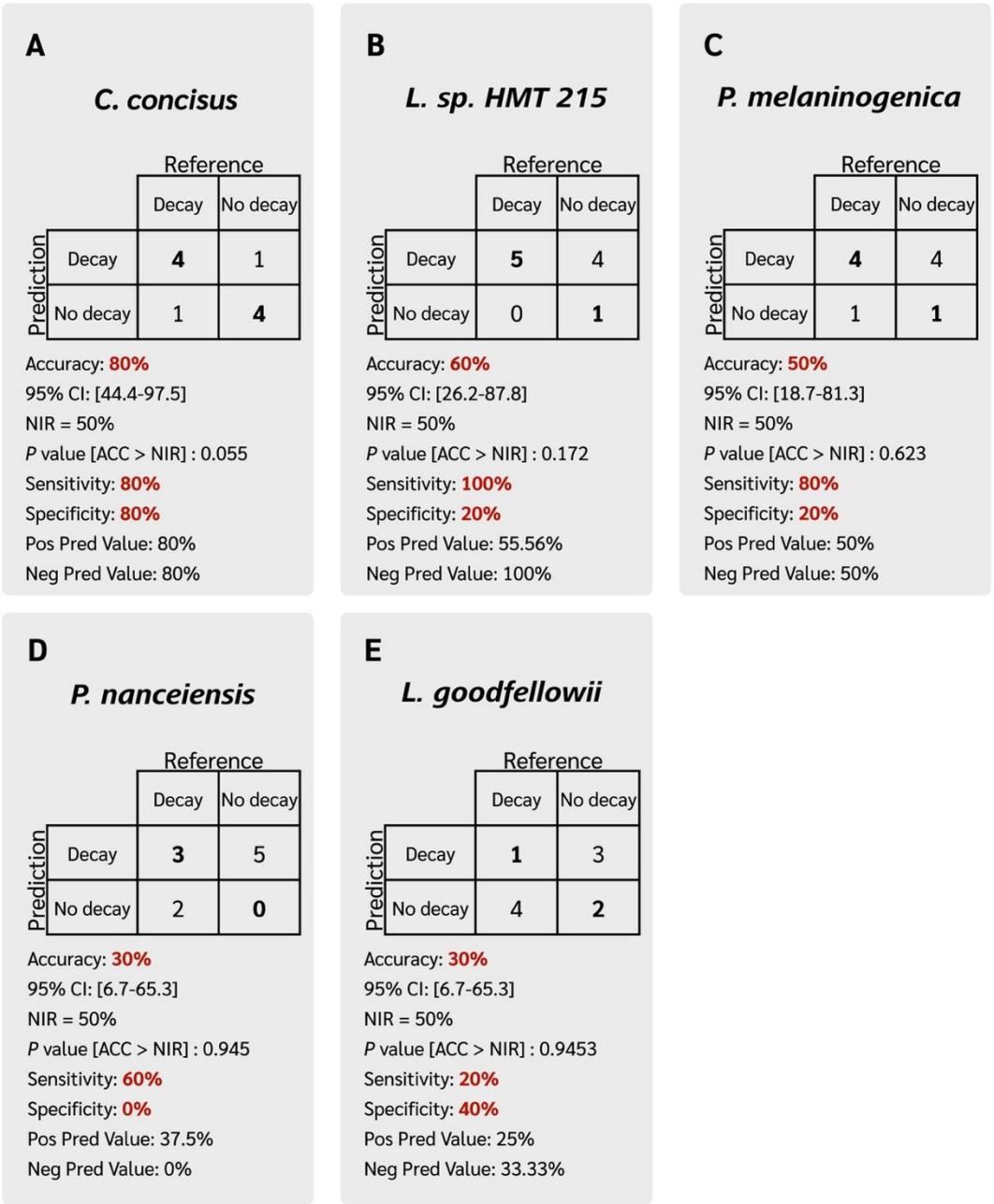
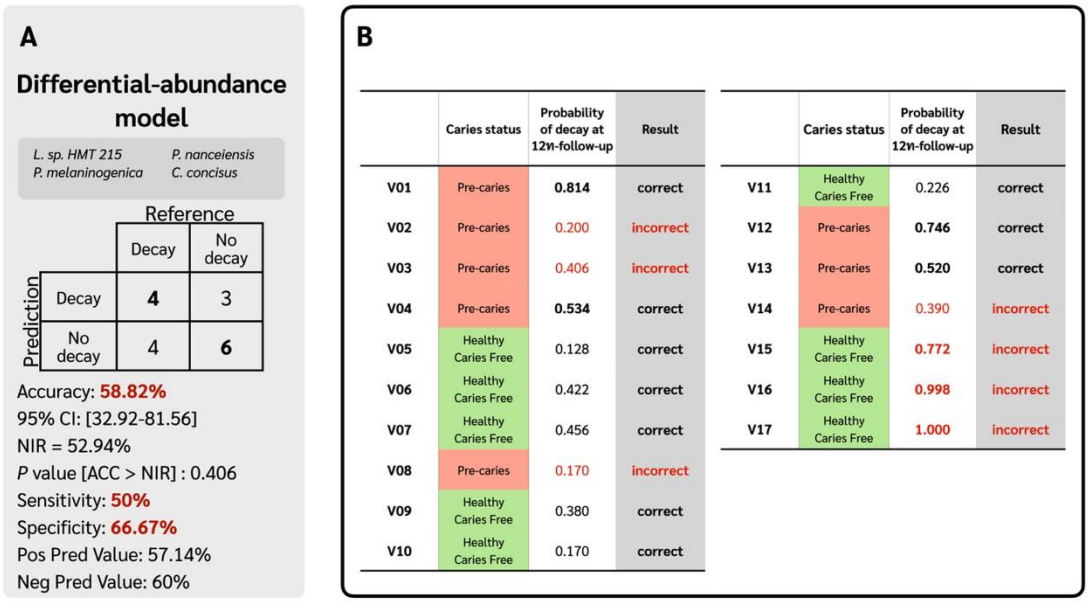


Figure 14: Confusion matrix evaluation of RF models based on single-species relative abundance, including (A) C. concisus model, (B) L. sp. HMT215 model, (C) P. melaninogenica model, (D) P. nanceiensis model, and (E) L. goodfellowii model. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and P value [Acc >No Information Rate] of models are shown, as calculated by caret package for R.

Lastly, our best-performed model, Differential-abundance model, was further challenged by validating its performance using a cross-study approach. The selected

salivary microbiome data, retrieved from the publicly available dataset of the previous study conducted in the United States (Grier et al., 2021), was used as the validating group. The predictive performance of this model was reduced with an accuracy of 58.82%, sensitivity of 50%, and specificity of 66.67% (AUC, 0.58; 95% confidence interval (CI), 32.92-81.56; Figure 15A).



	Caries status	Probability of decay at 12m-follow-up	Result
V01	Pre-caries	0.814	correct
V02	Pre-caries	0.200	incorrect
V03	Pre-caries	0.406	incorrect
V04	Pre-caries	0.534	correct
V05	Healthy Caries Free	0.128	correct
V06	Healthy Caries Free	0.422	correct
V07	Healthy Caries Free	0.456	correct
V08	Pre-caries	0.170	incorrect
V09	Healthy Caries Free	0.380	correct
V10	Healthy Caries Free	0.170	correct

	Caries status	Probability of decay at 12m-follow-up	Result
V11	Healthy Caries Free	0.226	correct
V12	Pre-caries	0.746	correct
V13	Pre-caries	0.520	correct
V14	Pre-caries	0.390	incorrect
V15	Healthy Caries Free	0.772	incorrect
V16	Healthy Caries Free	0.998	incorrect
V17	Healthy Caries Free	1.000	incorrect

Figure 15: (A) Confusion matrix evaluations of Differential- model in cross-study validation by predicting the selected samples retrieved from BioProject ID PRJNA622300 (Grier et al., 2021). Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and P value [Acc >No Information Rate] of models are shown, as calculated by caret package for R. (B) Prediction results of validating samples based on probability.

Chapter V

5. Discussion

5.1. Summary of Findings

Dental caries is a preventable disease but yet the most common oral disease, especially in children. Providing the proper preventive interventions as early as within the first year of life is a key strategy for reducing the caries prevalence in children (Pitts et al., 2019). With the lack of a reliable caries predictor for children at this age, targeted prevention could not be achieved. This will lead to the overall failure of caries prevention due to the constrained resource both in terms of healthcare personnel and allocated budget. The currently used Caries risk assessments have some limitations in data collection, most of the information is obtained from interviews with parents or caregivers, which is prone to error in the acquisition. For example, the information is often categorical data due to the limitations in the informant's self-assessment, such as the oral hygiene practice has often used the frequency of tooth brushing, which may not always reflect the quality of oral hygiene practice. Another example is the information about caries-associated diet, which is difficult to obtain accurate and complete information by short questioning. All of this may lead to incomplete information that will affect the caries risk prediction. Moreover, some risk factors require clinical examination, such as past caries experience, plaque score, and salivary flow rate, which some populations may not be able to access especially in populations of low socioeconomic status. In Thailand, the accessibility rate to dental care is very low with over 90% of the population that did not receive any dental service (Jaichuen, 2018). Using data collection that is proactive and does not rely on dental services should help to expand the survey to reach more populations, especially vulnerable groups.

The oral microbiome is a promising candidate for caries prediction because it could reflect conditions in the oral cavity deriving from various currently known caries risk factors by analyzing the biological sample. Previous studies demonstrated that the oral microbiome could be developed into the caries predictive models with

desirable performances (Teng et al., 2015; Xu et al., 2018; Grier et al., 2021). However, all of them were conducted in children older than 1 year, thus that information might not be applicable for 1-year-old children due to significant differences in the oral microbial compositions between children of different ages (F. Li et al., 2018; Dashper et al., 2019). In this study, we aimed to construct the dental caries prediction model for 1-year-old children using bacterial biomarkers that could reflect the ongoing change in the oral microbiome during the caries process prior to the occurrence of caries lesions.

The oral microbiome of the 1-year-old caries-free children who remained caries-free (F2F) was the most diverse microbial community within the sample when compared with those who developing caries, not significantly different except for the relative evenness of species richness between the children who remained caries-free (F2F) and who developed non-cavitated caries lesions (F2W) (Figure 8). Our finding is consistent with the previous evidence showing that the α-diversity is decreasing during the caries progression (Gross et al., 2010) and the children who suffer from dental caries had much less diverse oral microbiome than caries-free children (Kanasi et al., 2010). Our results showed that the difference in α-diversity could be detected as early as the children are still caries-free. When compare between groups, the microbial community structure exhibited significant differences between the children who remained caries-free (F2F) and who developed cavitated caries lesions (F2D) based on the unweighted UniFrac distance metric (Figure 9A). This difference is consistent with the previous study that found differences in the oral microbiome between the children who developed the recurrent ECC within 12 months and those who did not (Zhu et al., 2018). A recent study indicated that no significant differences were found between the children with different caries statuses (Caries Free, Pre-caries, and Caries Active) (Grier et al., 2021). However, that result was based on the weighted UniFrac distance metric which corresponded with our analysis when using the same index (Figure 9B). Unfortunately, they did not perform the unweighted UniFrac β-diversity analysis, thus we cannot compare the results from every aspect.

UniFrac distance is a phylogenetic-based β-diversity using the percentage of observed branch length that unique to either compared samples, in other words, the

unique fraction of the phylogenic tree. The weighted UniFrac distance uses abundance information and weights the branch length with abundance differences, which makes it most sensitive to detect the changes in abundant lineages. On the contrary, the unweighted UniFrac distance is suitable for detecting the changes in rare lineages. From the biomarker discovery analyses, 4 bacterial species were discovered as the microbial biomarkers for future caries prediction and all of these species were rare taxa with a relative abundance lower than 1% (Figure 7C). Based on univariate tests and LEfSe, we identified 4 species including *L. sp. HMT 215, C. concisus, P. nanceiensis,* and *P. melaninogenica* as biomarkers for caries-free 1-year-old children who remained caries-free for the next 12 months. In addition, these 4 species were selected as important features of the All-taxa prediction model using the Boruta algorithm, which is considered the most powerful method among the variable selection methods of random forest (Degenhardt et al., 2017), as well as RFE, a popular feature selection method that can discover a minimal set of variables with a good prediction (Diaz-Uriarte & Alvarez de Andres, 2006) (Figure 13). These findings indicated that these 4 species could be effective biomarkers for the early detection of dental caries up to 12 months before it appears.

The etiological factors are usually used interchangeably with the predictors because they share some common methodologies, such as multivariable modeling. However, the aim of usage and interpretation of results are very different (van Diepen et al., 2017), which should be clearly understood before being applied. Etiology aims to reveal the cariogenic effect of risk factors for caries that would help caries management by reducing each of the specific etiological factors. In contrast, prediction aims to precisely predict the future caries risk using caries predictors that may not necessarily be the causative factor of dental caries. For example, the previous caries experience is a strong predictor of caries (Mejàre et al., 2014) but it is not the causal pathway of the disease. The microbial biomarkers found in this study were diminished in the children who developed cavitated caries lesions. Although they are not the causal factors of dental caries, they could be used as potential biomarkers for caries prediction.

To assess the predictive performance of the oral microbiome of caries-free 1-year-old children as the caries predictor, the supervised machine learning models were developed using various combinations of multi-species and single-species biomarkers. When compared to the All-taxa model, the Important-feature model based on 5 important species (*L. sp. HMT 215, P. nanceiensis, P. melaninogenica, C. concisus,* and *L. goodfellowii*) could maintain the performance by using much fewer biomarkers. Moreover, using 4 species that were identified as the potential biomarkers (*L. sp. HMT 215, P. nanceiensis, P. melaninogenica,* and *C. concisus*), as in the Differential-abundance model, could improve the prediction accuracy to 80% (Figure 12C). Our findings, however, revealed that for single-species models, apart from *C. concisus*, their performances were reduced drastically (Figure 14B-E). These findings imply that the accuracy of the prediction model could have deteriorated when a model learns too much detail, including noise in the training process. Moreover, the smaller set of variables is preferable for implementation in practical situations. Also, using just one biomarker could reduce the accuracy of a prediction model due to the sparsity nature of microbiome data that some species could be missed in some samples. Therefore, an effective model should include an appropriate biomarker combination. The only single-species model based on *C. concisus* abundance could perform a comparable accuracy to the multi-species models (Figure 14A), suggesting that this species might be a member of the core microbiome that play an important role in inhibiting cariogenesis.

5.2. Microbial Biomarkers

Campylobacter concisus, which was found relatively higher in the remained-caries-free group (F2F), was one of the important features in the All-taxa model (Figure 14). The presence of these bacteria has been associated with gingival inflammation and the onset of periodontal disease (Macuch & Tanner, 2000), as well as inflammatory bowel disease (Li Zhang et al., 2009). However, the high prevalence of *C. concisus* could be detected in the saliva of healthy persons (L. Zhang et al., 2010) and it is a member of the core salivary microbiome in children from newborn

to 4-year-old (Dashper et al., 2019), suggesting that these bacteria could be part of the normal human oral microbiota. Moreover, *C. concisus* ATCC 51562 strain was identified as the biomarker for caries-free children in middle childhood (Al-Hebshi et al., 2019), and some *Campylobacter* species were significantly reduced in severe caries in teenagers (Gross et al., 2010). Because of their ability of biofilm formation (Lavrencic et al., 2012), combined with their survivability in a low pH environment (Kaakoush et al., 2016), they might compete with other acidic biofilm producers, such as *Streptococcus mutans*, that could reduce their virulence during the caries process. Based on this assumption, *C. concisus* could have inhibiting effects on cariogenesis that could lower the risk of having caries.

The overabundance of *Prevotella* spp. has been associated with caries-active individuals (Yang et al., 2012), including *P. melaninogenica* in middle childhood (Al-Hebshi et al., 2019). Using the relative abundance of seven *Prevotella* spp. (*P. pallens, P. denticola, P. verovalis, P. salivae, P. histicola, P. DO039, and P. maculosa),* which were found higher relative abundance in caries-active samples, was able to construct the caries prediction model for the new onset of ECC in preschool children with an accuracy of 74% (Teng et al., 2015). These studies showed the opposite results from what we found in our study. However, on the contrary, a recent study in Thai 3-year-old children reported that *P. melaninogenica* was more prevalent and enriched in caries-free children when compared with ECC children (Wu et al., 2022). Moreover, the relative abundance of the *Prevotella* genus was found significantly higher in the salivary microbiome of the preschool children who did not develop the recurrent ECC within 12 months than of those who did, which could be used as the predictor for the recurrence of ECC (Zhu et al., 2018). Therefore, further studies of these bacterial species, especially at the species level, both in terms of their existence and functions are needed to validate our findings and to understand their roles during the caries process.

The mutans streptococci (MS) have been associated with dental caries (Tanzer et al., 2001). However, it is understandable that the changes in their relative abundance were not recognized as the biomarkers in our study since we were investigating the saliva of caries-free children while the high MS level is positively

associated with increasing caries lesions in children (Thibodeau & O'Sullivan, 1999; Edelstein et al., 2016). Moreover, MS is rarely detected in 1-year-old children and is usually found in plaque earlier than in saliva (Seki et al., 2003). Using the culture-based analysis, there were no significant differences in the MS level were found when compared between groups (*P*=0.092, Kruskal-Wallis; Figure 16). Nevertheless, the most abundant species that we found in all groups were unclassified *Streptococcus* species (Figure 7). Using a more refined sequencing technique that provides a better taxonomic resolution might be able to reveal the hidden biomarkers in this data.

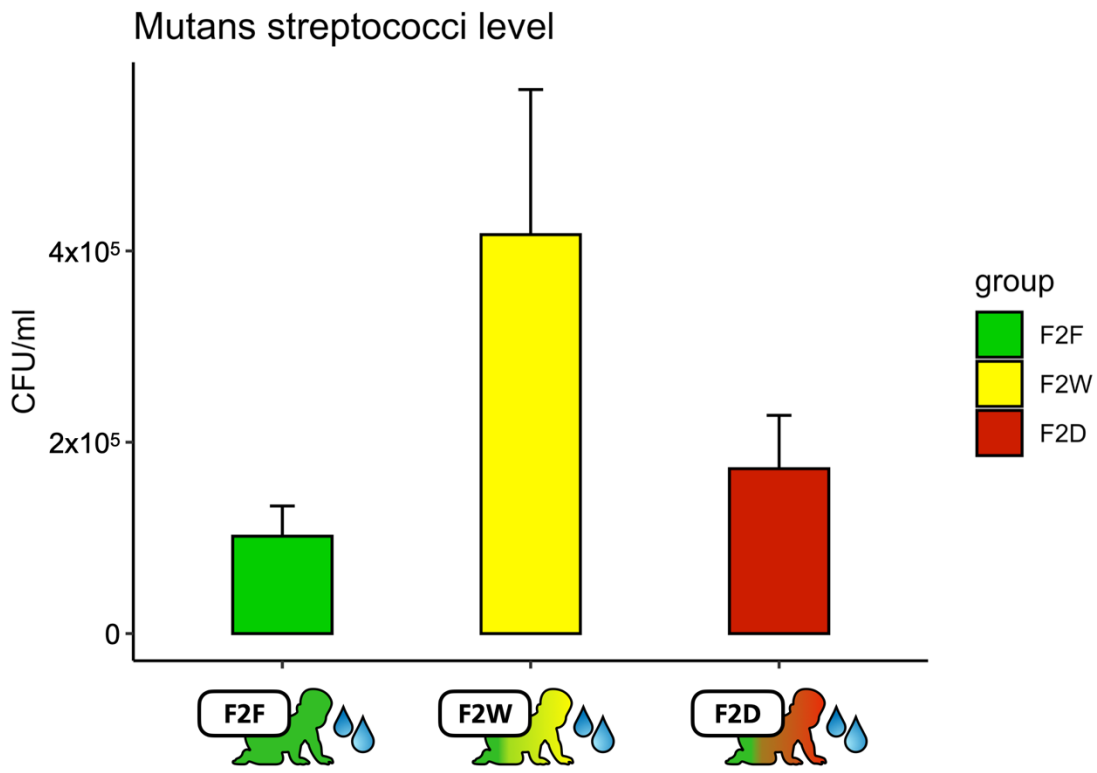


Figure 16: The mean (SD) of the mutans streptococci count (CFU/ml) in the saliva at the baseline time point from caries-free 1-y-old children in each group including the children who remained caries-free (F2F, green), those who developed non-cavitated caries lesions (F2W, yellow), and those who developed cavitated caries lesions (F2D, red). (P=0.092, Kruskal-Wallis)

Another essential factor that could affect biomarker discovery is the reference database used for taxonomy assignment. Taxonomic identification based on 16S

rRNA gene sequencing data requires taxonomic information from the public databases for the annotation. Greengenes database *(https://greengenes.secondgenome.com)* (DeSantis et al., 2006) is the most popular and widely used database that is the default database in the QIIME pipeline (Caporaso et al., 2010). However, it is considered obsolete as it has not been updated since 2013. The SILVA database *(https://www.arb-silva.de)* (Quast et al., 2013) provides comprehensive and updated datasets of aligned 16S rRNA sequences for bacteria, archaea, and eukarya, which were updated in 2020. Although these 2 databases were widely used in microbiome studies, the oral microbiome-specific database might be more appropriate to be used for oral microbiome analysis especially when species-level resolution is required. The Ohio State University (OSU) CORE database *(http://microbiome.osu.edu)* is a phylogenetically-curated database of 16S rRNA gene sequences that represent the core oral microbiome (Griffen et al., 2011). This database provides a comprehensive and minimally redundant representation of the oral bacteria with the classification at the level of genus and species, the latest version was released in 2017. The expanded Human Oral Microbiome Database (eHOMD) *(https://www.homd.org)* is the most up-to-date public database that provides comprehensive curated information on bacteria in the human oral cavity (Escapa et al., 2018). This database consists of 774 oral bacteria species, 58% are officially named with a provisional naming scheme for the unnamed taxa, based on the 16S rRNA sequence phylogeny with the latest update in 2021. We used these 4 databases to assign the taxonomy of our data at the genus and species level. At the genus level, all databases gave a comparable result both in terms of the type and abundance of each assigned genus (Figure 6). However, at the species level, SILVA and Greengenes databases could classify most of the species-level taxa as unclassified species with only 7 and 9 out of 25 species that could be classified from the top 25 most abundant species using Greengenes and SILVA databases (Figure 7B and A, respectively). In contrast, eHOMD and CORE databases were able to classify the species-level taxa of 19 and 21 out of 25 of the top 25 most abundant species (Figure 7C and D, respectively), suggesting that the oral microbiome-specific database has a good performance in identifying the species-level taxa from the oral

microbiome data. Therefore, we chose the eHOMD database for our analysis since it is the most up-to-date oral microbiome-specific database that was publicly available at that moment.

5.3. Caries Prediction Models

To validate the performance of each generated prediction model, the unrelated 1-year-old children who developed cavitated caries lesions differently at the age of 2 years old, with or without cavitated caries lesions, were used as the testing group (Table 3). The Differential-abundance model was highly predictive in the sample with an early change in the caries development when looking at the predictive probability. For example, in T07 and T10, these children had been early diagnosed with cavitated caries lesion at the 6-month-follow-up (Table 3) which could be correctly predicted with a probability of over 90% (Table 6).

Table 6: Prediction results of testing samples based on probability when using the top-performed prediction models.

		All-taxa model		Important-features model		Differential-abundance model		*C. concisus* model	
Sample #	Status at 12m-follow-up	Probability of decay at 12m-follow-up	Result	Probability of decay at 12m-follow-up	Result	Probability of decay at 12m-follow-up	Result	Probability of decay at 12m-follow-up	Result
T01	Decay	0.644	correct	0.636	correct	0.898	correct	0.992	correct
T02	Decay	0.566	correct	0.542	correct	0.740	correct	0.992	correct
T03	Decay	0.450	incorrect	0.200	incorrect	0.084	incorrect	0.006	incorrect
T04	No Decay	0.550	incorrect	0.518	incorrect	0.392	correct	0.022	correct
T05	No Decay	0.408	correct	0.426	correct	0.412	correct	0.876	incorrect
T06	No Decay	0.676	incorrect	0.660	incorrect	0.658	incorrect	0.272	correct
T07	Decay	0.530	correct	0.674	correct	0.994	correct	0.992	correct
T08	No Decay	0.464	correct	0.404	correct	0.164	correct	0.052	correct
T09	No Decay	0.420	correct	0.260	correct	0.404	correct	0.012	correct

T10	Decay	0.506	correct	0.674	correct	0.994	correct	0.992	correct

On the other hand, the testing samples that most of the models gave a wrong prediction could be related to the detail in the severity of caries during the observation. For example, the testing sample that all of the top-performed models gave the wrong prediction was T03, predicted as “no decay at 2-year-old” (Table 6). This child was diagnosed with only 1 cavitated caries lesion at the 12-month-follow-up, while the examination at 6-month-follow-up showed no cavity (Table 3). Another example is T06 which the $d_1$mfs index was 12, increased by 3 times compared to at the baseline time point (Table 3). All of the multi-species models gave the prediction as “have decay at 2-year-old” (Table 6), which was incorrect since there was no cavitated caries lesion was detected at the 12-month-follow-up. All of these show that the performance of prediction models was impaired when predicting samples in the borderline scenarios with obscure caries development. A way to improve a model to better predict the complex problem is to increase the training data. We recognize that the small sample size used in the current analysis might be a limitation. However, it could be improved in further studies when we can analyze the entire cohort of over a hundred caries-free 1-year-old children from the stored samples.

A practical prediction model should perform reasonably well across different datasets with comparable but not identical populations. A cross-study performance testing can be performed using sequencing data from previous studies deposited in public databases such as GenBank (www.ncbi.nlm.nih.gov/Genbank), EMBL (www.ebi.ac.uk/ena/browser/), and DDBJ (www.ddbj.nig.ac.jp/). Unfortunately, available 16S rRNA gene sequencing data of oral microbiome is limited, especially in infants. Moreover, the provided metadata of each dataset is mostly insufficient to be applied for cross-study comparison and the different experimental protocols, such as hyper-variable region selection, might affect the results of the analysis as well (W. Zheng et al., 2015). However, with these limitations, the salivary microbiome dataset from 17 samples (age 20.47 months ± 2.61), retrieved from BioProject ID

PRJNA622300 (Grier et al., 2021), was selected and used for model validation (Table 4). Our best-performed prediction model, the Differential-abundance model, was challenged by validating its performance using a cross-study approach. The predictive performance of this model was reduced with an accuracy of 58.82%, sensitivity of 50%, and specificity of 66.67% (AUC, 0.58; 95% confidence interval (CI), 32.92-81.56; Figure 15A). There are several factors that could affect the performance of our model. First, the inclusion criteria for the pre-caries sample in the validating group included non-cavitated caries lesions as diagnostic criteria for ECC, according to the American Academy of Pediatric Dentistry Guidelines (AAPD, 2011). Thus, we cannot differentiate the pre-caries samples with cavitated caries lesions to match our supervised training process which labeled only cavitated caries lesions as developed decay. As a result, our prediction model could predict future caries with only 50% of accuracy when analyzing the pre-caries samples (Figure 15B). Second, the ages of children were different when compared between the training and validating groups, 12.57 months ± 0.97 and 20.47 months ± 2.61, respectively (Table 5 and 4). The children of these two ages had different salivary microbiomes both in terms of within- and between-sample diversity (Dashper et al., 2019). Moreover, the races of children were different, with our study conducted on Thai children while the previous study was on American children. The diversity of the oral microbiome varies by geography and race/ethnicity (V. K. Gupta et al., 2017), which may be due to cultural differences such as diet and lifestyle. Moreover, genetic factors could influence the oral microbial community as well since their differences could be found in populations of different races who have shared similar environmental factors over several generations (Mason et al., 2014). Although this difference was found mostly in adults, it is likely to be found in young children as well because their oral microbiome was similar to those found in their mothers due to the vertical transmission that could share both commensal and disease-related bacteria (Jo et al., 2021). All of these reasons could explain the reduced performance in this validation. However, we found consistent results with our findings that the relative abundance of *C. concisus* and *P. nanceiensis* were higher in the children who remained caries-free compared with those who developed ECC within 2 years in

validating group, but not statistically different (*P*=0.743 and 0.160, respectively, Wilcoxon rank-sum test; Figure. 17A, B).

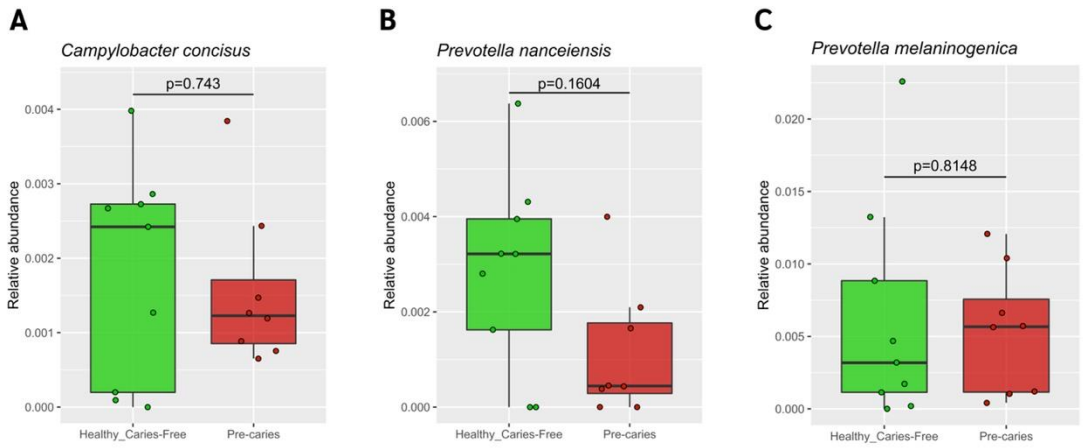


Figure 17: Relative abundance of the Differential-abundance species including (A) Campylobacter concisus, (B) Prevotella nanceiensis, and (C) Prevotella melaninogenica in the selected samples retrieved from BioProject ID PRJNA622300 (Grier et al., 2021), between Healthy caries-free subgroup (green) and Pre-caries subgroup (red). The comparison between these subgroups was performed using the Wilcoxon rank-sum test. Leptotrichia sp. HMT 215 was absent in this dataset.

5.4. Comparison to the Previous Evidence

There were only 3 longitudinal studies had proposed future caries prediction models based on oral microbiota using a supervised machine learning approach (Teng et al., 2015; Xu et al., 2018; Grier et al., 2021). Apart from the different ages of the children, several differences could distinguish our study from the previous reports. First, the prediction models reported by Teng and colleagues (2015), and Xu and colleagues (2018) used the differences between the caries-free and caries-active oral microbiota to define the high risk of having caries, by classifying the caries-free children who had the oral microbial composition that similar to caries-active state as a high risk of ECC. This interpretation should be validated since each stage of the caries process consists of different microbial members (Takahashi & Nyvad, 2008), the oral microbiota of pre-caries and caries-active children could be different as well. Our models exclusively used the oral microbiota of caries-free children who developed

caries differently for the training process, the same approach as Grier and colleagues (2021), which is a more straightforward approach for future caries prediction of caries-free children before the caries onset. Second, the prediction models based on the overall microbial taxa reported by Grier and colleagues (2021) show a good predictive power with an accuracy of 83.6-85.5%, a sensitivity of 77.1-85.7%, and a specificity of 85.0-95.0%. However, this performance could be accomplished only when the models learned from the caries-free samples 6 months prior to caries diagnosis. When the training group was changed to be the samples with a longer prediction period, 12 months, the accuracy was slightly decreased with a drastic reduction of specificity (an accuracy of 71.4-73.2%, a sensitivity of 80.6-86.1%, and a specificity of 45.0-60.0%). Our approach, using the selected biomarkers for model training, could achieve an accuracy of 80% with a good balance between sensitivity and specificity when predicting the caries-free children 12 months prior to the caries onset. This approach could be a robust alternative that could the predictive performance by solving some limitations of the training process of machine learning. Moreover, it is possible for the cross-study analysis and more practical when it comes to clinical practice. Furthermore, our study developed caries prediction models based on the Southeast Asian population for the first time. Since the oral microbiome varies by geography and race/ethnicity (V. K. Gupta et al., 2017), our study should expand this information to a region that has never been explored. Two previous studies on the Chinese population (Teng et al., 2015; Xu et al., 2018) reported that the genera *Streptococcus* and *Prevotella* were found to be the most discriminatory taxa in their models. While the most recent study conducted on the Americans (mixed both African and Caucasian) reported that *Streptococcus* sp., *Rothia mucilaginosa*, and *Veilonella parvula* were the important features in all of their prediction models (Grier et al., 2021). Our study could identify the biomarkers that are different from these previous reports, showing that the caries risk prediction could be geographically dependent. Therefore, the development of models should be based on specifically each population to ensure that they are suitable for use in that particular population.

5.5. Future Work

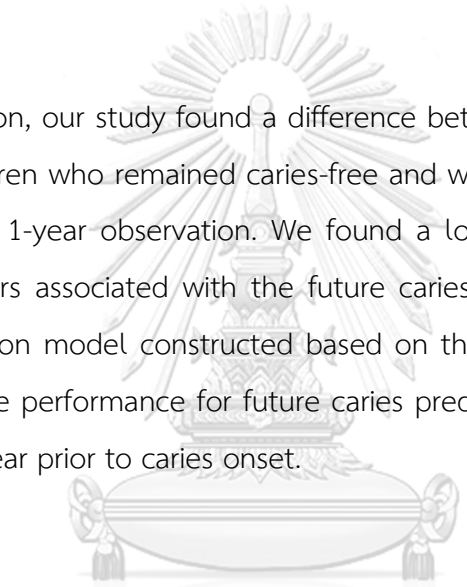
Some limitations should be considered for our study. Apart from the relatively small sample size, the sequencing technology used in this study was based on amplicon sequencing, in which reliable bacterial classification is mostly possible down to only at the genus level (Yarza et al., 2014; Winand et al., 2019), especially when using amplicons of selected parts of this gene (Caudill & Brayton, 2022) due to the false-positive results on classification caused by a high similarity of 16S rRNA gene sequences. Using the method that provides a more extensive read could be the alternative technique to accomplish a more refined biomarker discovery, such as 16S full-length-based synthetic long-read sequencing (sFL16S) that showed a better resolution in the analyses of α-diversity, relative abundance frequency, and identification accuracy (Jeong et al., 2021) and shallow shotgun metagenomic sequencing (SSMS), using few as 500K sequences per sample, that could recover more-accurate species-level taxonomic at possibly the same per-sample cost as 16S sequencing (Hillmann et al., 2018). The DNA extraction is also a factor involved in the result as well, specifically *Streptococcus* species that are tenacious to be lysed (Cho et al., 2021) which could be another reason that the genus *Streptococcus* was not classified well at the species level in our results.

Finally, using the saliva from the caries-free 1-year-old children, we found promising evidence of the microbiota-based caries prediction for future caries specifically for this targeted-age group that currently lacks a reliable caries predictor. However, this study is classified as level 3 of evidence with a narrow validation (McGinn et al., 2015) that we applied our models to validate a slightly different population. The next step will be a broad validation that the model should be applied in multiple clinical settings with varying prevalence and outcomes of dental caries, level 2 of evidence. Lastly, level 1 of evidence is needed to be accomplished before the clinical application by impact evaluation. The caries prediction model needs to be tested in randomized controlled trials to ensure that it improves the outcomes of prevention, is cost-effective, practical, and improves clinical decision-making (Moons et al., 2009). In caries management, caries prediction is not expected

to replace the currently used caries risk assessments, but rather be integrated to enhance caries management for better efficiency. Reliable and practical caries prediction can help in a proactive broad survey to find the children who have a high risk of having caries, then initiate an intervention to provide a thorough assessment to determine the caries risk factors and to provide appropriate caries prevention for each individual. With this approach, effective targeted prevention with limited resources could be possible and that would lead to a decrease in the overall caries prevalence in the future.

6. Conclusion

In conclusion, our study found a difference between the salivary microbiome of 1-year-old children who remained caries-free and who developed cavitated caries lesions during the 1-year observation. We found a low relative abundance of four bacterial biomarkers associated with the future caries onset in 1-year-old children. The caries prediction model constructed based on these four biomarkers yielded a desirable predictive performance for future caries prediction in 1-year-old cavity-free children up to 1 year prior to caries onset.



REFERENCES

VITA

NAME	Ratcha Raksakmanut
DATE OF BIRTH	7 July 1981
PLACE OF BIRTH	Thailand
INSTITUTIONS ATTENDED	- Department of Operative Dentistry, Chulalongkorn University, Thailand. - Faculty of Dentistry, Chulalongkorn University, Thailand.
HOME ADDRESS	11/174 Ratchaphruek road Bang Phlap Pak Kret Nonthaburi 11120
PUBLICATION	Raksakmanut R, Nithipratheep A, Weeriyanan O, Pittayaporn P, Komonpale W, Techakunakorn T, Luanprueg N, Promserane J. Effect of Bleaching on Color Change of Stained Resin Composites. Key Engineering Materials. 2017;751: 636-642.
AWARD RECEIVED	2022 IADR Pediatric Oral Health Research Group (POHRG) Young Investigator Award

1. AAPD. Policy on use of a caries-risk assessment tool (CAT) for infants, children, and adolescents, Reference Manual 31, (2002): 29-33.
2. AAPD. American Academy of Pediatric Dentistry reference manual 2011-2012, Pediatric Dentistry 33,6 Reference Manual (2011): 1-349.
3. AAPD. Policy on early childhood caries (ECC): Classifications, consequences, and preventive strategies, Pediatric Dentistry 38,6 (2017): 52-54.
4. AAPD. Caries-risk assessment and management for infants, children, and adolescents, American Academy of Pediatric Dentistry III, (2021): 252-257.
5. Aas, J. A., Griffen, A. L., Dardis, S. R., Lee, A. M., Olsen, I., Dewhirst, F. E., et al. Bacteria of dental caries in primary and permanent teeth in children and young adults, Journal of Clinical Microbiology 46,4 (2008): 1407-1417.
6. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity, Journal of Clinical Microbiology 43,11 (2005): 5721-5732.
7. ADA. (2011). Caries risk assessment form (Age 0-6). *American Dental Association.* Retrieved from https://www.ada.org/-/media/project/ada-organization/ada/ada-org/files/resources/public-programs/give-kids-a-smile/gkas_caries_risk_assessment_forms.pdf?rev=4b2191e7b6a4485aa0dbe6455cc00512&hash=431971C5061B2EF093A33C43E3BC59B5
8. Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions, Nature Genetics 45,4 (2013): 450-455, 455e451.
9. Agnello, M., Marques, J., Cen, L., Mittermuller, B., Huang, A., Chaichanasakul Tran, N., et al. Microbiome Associated with Severe Caries in Canadian First Nations Children, Journal of Dental Research 96,12 (2017): 1378-1385.
10. Åkesson, M.-L., Wärnberg Gerdin, E., Söderström, U., Lindahl, B., & Johansson, I. Health-related quality of life and prospective caries development, BMC Oral Health 16,1 (2016): 15.

11. Al-Hebshi, N. N., Baraniya, D., Chen, T., Hill, J., Puri, S., Tellez, M., et al. Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children, Journal of Oral Microbiology 11,1 (2019): 1557986.
12. Amin, M., Nouri, R., ElSalhy, M., Shah, P., & Azarpazhooh, A. Caries recurrence after treatment under general anaesthesia for early childhood caries: a retrospective cohort study, European Archives of Paediatric Dentistry 16,4 (2015): 325-331.
13. Balakrishnan, M., Simmonds, R. S., & Tagg, J. R. Dental caries is a preventable infectious disease, Australian Dental Journal 45,4 (2000): 235-245.
14. Becker, M. R., Paster, B. J., Leys, E. J., Moeschberger, M. L., Kenyon, S. G., Galvin, J. L., et al. Molecular analysis of bacterial species associated with childhood caries, Journal of Clinical Microbiology 40,3 (2002): 1001-1009.
15. Berg, J. H. The marketplace for new caries management products: dental caries detection and caries management by risk assessment, BMC Oral Health 6 Suppl 1,S1 (2006): S6.
16. Berkowitz, R. J., Amante, A., Kopycka-Kedzierawski, D. T., Billings, R. J., & Feng, C. Dental caries recurrence following clinical treatment for severe early childhood caries, Pediatric Dentistry 33,7 (2011): 510-514.
17. Bhattarai, K. R., Kim, H. R., & Chae, H. J. Compliance with Saliva Collection Protocol in Healthy Volunteers: Strategies for Managing Risk and Errors, International Journal of Medical Sciences 15,8 (2018): 823-831.
18. Bhaumik, D., Manikandan, D., & Foxman, B. Cariogenic and oral health taxa in the oral cavity among children and adults: A scoping review, Archives of Oral Biology 129, (2021): 105204.
19. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science using QIIME 2, Nature Biotechnology 37,8 (2019): 852-857.
20. Bordenstein, S. R., & Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes, PLoS Biology 13,8 (2015): e1002226.

21. Boue, D., Armau, E., & Tiraby, G. A bacteriological study of rampant caries in children, Journal of Dental Research 66,1 (1987): 23-28.
22. Braidwood, R. J., Howe, B., & Reed, C. A. The Iranian Prehistoric Project: New Problems Arise as More is Learned of the First Attempts at Food Production and Settled Village Life, Science (New York, N.Y.) 133,3469 (1961): 2008-2010.
23. Bratthall, D., & Hansel Petersson, G. Cariogram--a multifactorial risk assessment model for a multifactorial disease, Community Dentistry and Oral Epidemiology 33,4 (2005): 256-264.
24. Brock, T. D. Milestones in Microbiology. Prentice-Hall, 1961.
25. Cagetti, M. G., Bonta, G., Cocco, F., Lingstrom, P., Strohmenger, L., & Campus, G. Are standardized caries risk assessment models effective in assessing actual caries status and future caries increment? A systematic review, BMC Oral Health 18,1 (2018): 123.
26. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data, Nature Methods 13,7 (2016): 581-583.
27. Cameron, S. J., Huws, S. A., Hegarty, M. J., Smith, D. P., & Mur, L. A. The human salivary microbiome exhibits temporal stability in bacterial diversity, FEMS Microbiology Ecology 91,9 (2015): fiv091.
28. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. QIIME allows analysis of high-throughput community sequencing data, Nature Methods 7,5 (2010): 335-336.
29. Caudill, M. T., & Brayton, K. A. The Use and Limitations of the 16S rRNA Sequence for Species Classification of Anaplasma Samples, Microorganisms 10,3 (2022): 605.
30. Chen, T., Yu, W. H., Izard, J., Baranova, O. V., Lakshmanan, A., & Dewhirst, F. E. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information, Database: The Journal of Biological Databases and Curation 2010,0 (2010): baq013.
31. Cho, H. S., Choi, M., Lee, Y., Jeon, H., Ahn, B., Soundrarajan, N., et al. High-Quality Nucleic Acid Isolation from Hard-to-Lyse Bacterial Strains Using PMAP-

36, a Broad-Spectrum Antimicrobial Peptide, International Journal of Molecular Sciences 22,8 (2021): 4149.

32. Christian, B., Calache, H., Adams, G., Hall, M., Dashper, S., Gibbs, L., et al. A methodological study to assess the measurement properties (reliability and validity) of a caries risk assessment tool for young children, Journal of Dentistry 95, (2020): 103324.
33. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure, Austral Ecology 18,1 (1993): 117-143.
34. Cleaton-Jones, P. No consensus on definition of early childhood caries, Evidence-Based Dentistry 3,3 (2002): 75.
35. Clemente, J. C., Pehrsson, E. C., Blaser, M. J., Sandhu, K., Gao, Z., Wang, B., et al. The microbiome of uncontacted Amerindians, Science Advances 1,3 (2015): e1500183.
36. Contreras, M., Costello, E. K., Hidalgo, G., Magris, M., Knight, R., & Dominguez-Bello, M. G. The bacterial microbiota in the oral mucosa of rural Amerindians, Microbiology (Reading) 156,Pt 11 (2010): 3282-3287.
37. Corby, P. M., Bretz, W. A., Hart, T. C., Schork, N. J., Wessel, J., Lyons-Weiler, J., et al. Heritability of oral microbial species in caries-active and caries-free twins, Twin Research and Human Genetics 10,6 (2007): 821-828.
38. Cornejo, O. E., Lefebure, T., Bitar, P. D., Lang, P., Richards, V. P., Eilertson, K., et al. Evolutionary and population genomics of the cavity causing bacteria *Streptococcus mutans*, Molecular Biology and Evolution 30,4 (2013): 881-893.
39. Costa, S. M., Martins, C. C., Bonfim Mde, L., Zina, L. G., Paiva, S. M., Pordeus, I. A., et al. A systematic review of socioeconomic indicators and dental caries in adults, International Journal of Environmental Research and Public Health 9,10 (2012): 3540-3574.
40. Crielaard, W., Zaura, E., Schuller, A. A., Huse, S. M., Montijn, R. C., & Keijser, B. J. Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health, BMC Medical Genomics 4, (2011): 22.

41. Dashper, S. G., Mitchell, H. L., Le Cao, K. A., Carpenter, L., Gussy, M. G., Calache, H., et al. Temporal development of the oral microbiome and prediction of early childhood caries, Scientific Reports 9,1 (2019): 19732.
42. Dawes, C. Circadian rhythms in human salivary flow rate and composition, Journal of Physiology 220,3 (1972): 529-545.
43. De Menezes Oliveira, M. A., Torres, C. P., Gomes-Silva, J. M., Chinelatti, M. A., De Menezes, F. C., Palma-Dibb, R. G., et al. Microstructure and mineral composition of dental enamel of permanent and deciduous teeth, Microscopy Research and Technique 73,5 (2010): 572-577.
44. Degenhardt, F., Seifert, S., & Szymczak, S. Evaluation of Variable Selection Methods for Random Forests and Omics Data Sets, Briefings in Bioinformatics 20,2 (2017): 492-503.
45. Demers, M., Brodeur, J. M., Simard, P. L., Mouton, C., Veilleux, G., & Frechette, S. Caries predictors suitable for mass-screenings in children: a literature review, Community Dental Health 7,1 (1990): 11-21.
46. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, Applied and Environmental Microbiology 72,7 (2006): 5069-5072.
47. Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., et al. The human oral microbiome, Journal of Bacteriology 192,19 (2010): 5002-5017.
48. Diaz-Uriarte, R., & Alvarez de Andres, S. Gene selection and classification of microarray data using random forest, BMC Bioinformatics 7,1 (2006): 3.
49. Dige, I., Raarup, M. K., Nyengaard, J. R., Kilian, M., & Nyvad, B. *Actinomyces naeslundii* in initial dental biofilm formation, Microbiology (Reading) 155,Pt 7 (2009): 2116-2126.
50. Dirks, O. B. Posteruptive changes in dental enamel, Journal of Dental Research 45,3 (1966): 503-511.
51. Divaris, K. Predicting Dental Caries Outcomes in Children: A "Risky" Concept, Journal of Dental Research 95,3 (2016): 248-254.

52. Downes, J., Mantzourani, M., Beighton, D., Hooper, S., Wilson, M. J., Nicholson, A., et al. *Scardovia wiggsiae* sp. nov., isolated from the human oral cavity and clinical material, and emended descriptions of the genus *Scardovia* and *Scardovia inopinata*, International Journal of Systematic and Evolutionary Microbiology 61,Pt 1 (2011): 25-29.
53. Drury, T. F., Horowitz, A. M., Ismail, A. I., Maertens, M. P., Rozier, R. G., & Selwitz, R. H. Diagnosing and reporting early childhood caries for research purposes. A report of a workshop sponsored by the National Institute of Dental and Craniofacial Research, the Health Resources and Services Administration, and the Health Care Financing Administration, Journal of Public Health Dentistry 59,3 (1999): 192-197.
54. Du, Q., Yu, M., Li, Y., Du, H., Gao, W., Mei, H., et al. Permanent caries experience is associated with primary caries experience: a 7-year longitudinal study in China, Community Dentistry and Oral Epidemiology 45,1 (2017): 43-48.
55. Edelstein, B. L., Ureles, S. D., & Smaldone, A. Very High Salivary Streptococcus Mutans Predicts Caries Progression in Young Children, Pediatric Dentistry 38,4 (2016): 325-330.
56. Emilson, C. G., & Krasse, B. Support for and implications of the specific plaque hypothesis, Scandinavian Journal of Dental Research 93,2 (1985): 96-104.
57. Engelmann, J. L., Tomazoni, F., Oliveira, M. D., & Ardenghi, T. M. Association between Dental Caries and Socioeconomic Factors in Schoolchildren--A Multilevel Analysis, Brazilian Dental Journal 27,1 (2016): 72-78.
58. Eriksson, L., Lif Holgerson, P., Esberg, A., & Johansson, I. Microbial Complexes and Caries in 17-Year-Olds with and without *Streptococcus mutans*, Journal of Dental Research 97,3 (2017): 275-282.
59. Eriksson, L., Lif Holgerson, P., & Johansson, I. Saliva and tooth biofilm bacterial microbiota in adolescents in a low caries community, Scientific Reports 7,1 (2017): 5861.

60. Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., & Lemon, K. P. 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 (2018): e00187-00118.
61. Featherstone, J. D., Adair, S. M., Anderson, M. H., Berkowitz, R. J., Bird, W. F., Crall, J. J., et al. Caries management by risk assessment: consensus statement, April 2002, Journal - California Dental Association 31,3 (2003): 257-269.
62. Featherstone, J. D., Crystal, Y., Chaffee, B., Zhan, L., & Ramos-Gomez, F. An updated CAMBRA caries risk assessment tool for ages 0 to 5 years, Journal - California Dental Association 47,1 (2019): 37-47.
63. Featherstone, J. D., Crystal, Y. O., Alston, P., Chaffee, B. W., Domejean, S., Rechmann, P., et al. A Comparison of Four Caries Risk Assessment Methods, Frontiers in Oral Health 2, (2021): 656558.
64. Featherstone, J. D., Domejean-Orliaguet, S., Jenson, L., Wolff, M., & Young, D. A. Caries risk assessment in practice for age 6 through adult, Journal of the California Dental Association PMID - 18044378 35,10 (2007): 703-707, 710-703.
65. Fejerskov, O. Changing paradigms in concepts on dental caries: consequences for oral health care, Caries Research 38,3 (2004): 182-191.
66. Fejerskov, O., & Nyvad, B. Is dental caries an infectious disease? : diagnostic and treatment consequences for the practitioner, Nordic dentistry yearbook (2003): 141-151.
67. Fejerskov, O., Nyvad, B., & Kidd, E. Dental Caries. John Wiley & Sons, 2015.
68. Ferreira Zandoná, A. G., Ritter, A. V., & Eidson, R. S. (2019). Dental Caries. In *Sturdevant's Art and Science of Operative Dentistry* (pp. 40-94): Elsevier.
69. Fisher-Owens, S. A., Gansky, S. A., Platt, L. J., Weintraub, J. A., Soobader, M. J., Bramlett, M. D., et al. Influences on children's oral health: a conceptual model, Pediatrics 120,3 (2007): e510-520.
70. Fontana, M., Carrasco-Labra, A., Spallek, H., Eckert, G., & Katz, B. Improving Caries Risk Prediction Modeling: A Call for Action, Journal of Dental Research 99,11 (2020): 1215-1220.

71. Fontana, M., & Zero, D. T. Assessing patients' caries risk, Journal of the American Dental Association 137,9 (2006): 1231-1239.
72. Fung, M. H. T., Wong, M. C., Lo, E. C. M., & Chu, C. H. Early Childhood Caries: A Literature Review, Journal of Oral Hygiene & Health 1, (2013): 1-7.
73. Gao, L., Xu, T., Huang, G., Jiang, S., Gu, Y., & Chen, F. Oral microbiomes: more and more importance in oral cavity and whole body, Protein & Cell 9,5 (2018): 488-500.
74. Goldberg, B. E., Mongodin, E. F., Jones, C. E., Chung, M., Fraser, C. M., Tate, A., et al. The Oral Bacterial Communities of Children with Well-Controlled HIV Infection and without HIV Infection, PloS One 10,7 (2015): e0131615.
75. Grassl, N., Kulak, N. A., Pichler, G., Geyer, P. E., Jung, J., Schubert, S., et al. Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome, Genome Medicine 8,1 (2016): 44.
76. Grier, A., Myers, J. A., O'Connor, T. G., Quivey, R. G., Gill, S. R., & Kopycka-Kedzierawski, D. T. Oral Microbiota Composition Predicts Early Childhood Caries Onset, Journal of Dental Research 100,6 (2021): 599-607.
77. Griffen, A. L., Beall, C. J., Firestone, N. D., Gross, E. L., Difranco, J. M., Hardman, J. H., et al. CORE: a phylogenetically-curated 16S rDNA database of the core oral microbiome, PLoS ONE 6,4 (2011): e19051.
78. Gross, E. L., Beall, C. J., Kutsch, S. R., Firestone, N. D., Leys, E. J., & Griffen, A. L. Beyond *Streptococcus mutans*: dental caries onset linked to multiple species by 16S rRNA community analysis, PloS One 7,10 (2012): e47722.
79. Gross, E. L., Leys, E. J., Gasparovich, S. R., Firestone, N. D., Schwartzbaum, J. A., Janies, D. A., et al. Bacterial 16S sequence analysis of severe caries in young permanent teeth, Journal of Clinical Microbiology 48,11 (2010): 4121-4128.
80. Gupta, S., Mortensen, M. S., Schjorring, S., Trivedi, U., Vestergaard, G., Stokholm, J., et al. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing, Commun Biol 2,1 (2019): 291.

81. Gupta, V. K., Paul, S., & Dutta, C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity, Frontiers in Microbiology 8, (2017): 1162.
82. Gura, T. Just spit it out, Nature Medicine 14,7 (2008): 706-709.
83. Hajishengallis, E., Parsaei, Y., Klein, M. I., & Koo, H. Advances in the microbial etiology and pathogenesis of early childhood caries, Molecular Oral Microbiology 32,1 (2017): 24-34.
84. Hall, M. W., Singh, N., Ng, K. F., Lam, D. K., Goldberg, M. B., Tenenbaum, H. C., et al. Inter-personal diversity and temporal dynamics of dental, tongue, and salivary microbiota in the healthy oral cavity, NPJ Biofilms Microbiomes 3, (2017): 2.
85. Han, C. S., Martin, M. A., Dichosa, A. E. K., Daughton, A. R., Frietze, S., Kaplan, H., et al. Salivary microbiomes of indigenous Tsimane mothers and infants are distinct despite frequent premastication, PeerJ 4,Suppl 1 (2016): e2660.
86. Hänsel Petersson, G., Ericson, E., Isberg, P. E., & Twetman, S. Caries risk assessment in young adults: A 3-year validation of clinical guidelines used in Public Dental Service, Acta Odontologica Scandinavica 71,6 (2013): 1645-1650.
87. Hemadi, A. S., Huang, R., Zhou, Y., & Zou, J. Salivary proteins and microbiota as biomarkers for early childhood caries risk assessment, International Journal of Oral Science 9,11 (2017): e1.
88. Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., et al. Evaluating the Information Content of Shallow Shotgun Metagenomics, mSystems 3,6 (2018): e00069-00018.
89. Høiby, N., Ciofu, O., Johansen, H. K., Song, Z. J., Moser, C., Jensen, P., et al. The clinical impact of bacterial biofilms, International Journal of Oral Science 3,2 (2011): 55-65.
90. Hojo, K., Nagaoka, S., Ohshima, T., & Maeda, N. Bacterial interactions in dental biofilm development, Journal of Dental Research 88,11 (2009): 982-990.

91. Hong, X., & Hu, D. Y. Salivary *Streptococcus mutans* level: value in caries prediction for 11-12-year-old children, Community Dental Health 27,4 (2010): 248-252.
92. Hu, J., Jiang, W., Lin, X., Zhu, H., Zhou, N., Chen, Y., et al. Dental Caries Status and Caries Risk Factors in Students Ages 12-14 Years in Zhejiang, China, Medical Science Monitor 24, (2018): 3670-3678.
93. Hugenholtz, P., Goebel, B. M., & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity, Journal of Bacteriology 180,18 (1998): 4765-4774.
94. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome, Nature 486,7402 (2012): 207-214.
95. Hurley, E., Barrett, M. P. J., Kinirons, M., Whelton, H., Ryan, C. A., Stanton, C., et al. Comparison of the salivary and dentinal microbiome of children with severe-early childhood caries to the salivary microbiome of caries-free children, BMC Oral Health 19,1 (2019): 13.
96. Innes, N. P. T., & Robertson, M. D. Recent advances in the management of childhood dental caries, Archives of Disease in Childhood 103,4 (2018): 311-315.
97. Ismail, A. I., Hasson, H., & Sohn, W. Dental caries in the second millennium, Journal of Dental Education 65,10 (2001): 953-959.
98. Jaichuen, W. Accessibility to Dental Care According to Type of Service in Thai Population, Journal of Health Systems Research 12,4 (2018): 636-644.
99. James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., et al. 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, The Lancet 392,10159 (2018): 1789-1858.
100. Jeong, J., Yun, K., Mun, S., Chung, W.-H., Choi, S.-Y., Nam, Y.-d., et al. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology, Scientific Reports 11,1 (2021): 1727.

101. Jiang, Q., Liu, J., Chen, L., Gan, N., & Yang, D. The Oral Microbiome in the Elderly With Dental Caries and Health, Frontiers in Cellular and Infection Microbiology 8, (2018): 442.

102. Jiang, S., Gao, X., Jin, L., & Lo, E. C. Salivary Microbiome Diversity in Caries-Free and Caries-Affected Children, International Journal of Molecular Sciences 17,12 (2016):

103. Jiang, W., Ling, Z., Lin, X., Chen, Y., Zhang, J., Yu, J., et al. Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood, Microbial Ecology 67,4 (2014): 962-969.

104. Jiang, W., Zhang, J., & Chen, H. Pyrosequencing analysis of oral microbiota in children with severe early childhood dental caries, Current Microbiology 67,5 (2013): 537-542.

105. Jo, R., Nishimoto, Y., Umezawa, K., Yama, K., Aita, Y., Ichiba, Y., et al. Comparison of oral microbiome profiles in stimulated and unstimulated saliva, tongue, and mouth-rinsed water, Scientific Reports 9,1 (2019): 16124.

106. Jo, R., Yama, K., Aita, Y., Tsutsumi, K., Ishihara, C., Maruyama, M., et al. Comparison of oral microbiome profiles in 18-month-old infants and their parents, Scientific Reports 11,1 (2021): 861.

107. Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis, Nature Communications 10,1 (2019): 5029.

108. Kaakoush, N. O., Thomas, D. S., Ruzayqat, M. M., Lynch, D., Leach, S. T., Lemberg, D. A., et al. *Campylobacter concisus* utilizes blood but not short chain fatty acids despite showing associations with Firmicutes taxa, Microbiology 162,8 (2016): 1388-1397.

109. Kanasi, E., Dewhirst, F. E., Chalmers, N. I., Kent, R., Jr., Moore, A., Hughes, C. V., et al. Clonal analysis of the microbiota of severe early childhood caries, Caries Research 44,5 (2010): 485-497.

110. Kassebaum, N. J., Bernabe, E., Dahiya, M., Bhandari, B., Murray, C. J., & Marcenes, W. Global burden of untreated caries: a systematic review and metaregression, Journal of Dental Research 94,5 (2015): 650-658.

111. Kastenbom, L., Falsen, A., Larsson, P., Sunnegårdh-Grönberg, K., & Davidson, T. Costs and health-related quality of life in relation to caries, BMC Oral Health 19,1 (2019): 187.

112. Kato, I., Vasquez, A., Moyerbrailean, G., Land, S., Djuric, Z., Sun, J., et al. Nutritional Correlates of Human Oral Microbiome, Journal of the American College of Nutrition 36,2 (2017): 88-98.

113. Katoh, K., & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Molecular Biology and Evolution 30,4 (2013): 772-780.

114. Kaur, R., Gilbert, S. C., Sheehy, E. C., & Beighton, D. Salivary levels of Bifidobacteria in caries-free and caries-active children, International Journal of Paediatric Dentistry 23,1 (2013): 32-38.

115. Keeddee, C., Mongkolchaiarunya, S., Pochanukul, N., Jintakanon, P., & Sukhumalind, P. (2018). *The 8th National Oral Health Survey in Thailand*. Retrieved from Thailand: http://www.dent.chula.ac.th/upload/news/791/file_1_5834.pdf

116. Kianoush, N., Adler, C. J., Nguyen, K. A., Browne, G. V., Simonian, M., & Hunter, N. Bacterial profile of dentine caries and the impact of pH on bacterial population diversity, PloS One 9,3 (2014): e92940.

117. Kilian, M., Chapple, I. L., Hannig, M., Marsh, P. D., Meuric, V., Pedersen, A. M., et al. The oral microbiome - an update for oral healthcare professionals, British Dental Journal 221,10 (2016): 657-666.

118. Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C., & Knight, R. Human-associated microbial signatures: examining their predictive value, Cell Host & Microbe 10,4 (2011): 292-296.

119. Krasse, B. Microbiological and salivary risk factors, Risk assessment in dentistry. Proceedings of a conference (1990): 51-61.

120. Kreth, J., Zhang, Y., & Herzberg, M. C. Streptococcal antagonism in oral biofilms: *Streptococcus sanguinis* and *Streptococcus gordonii* interference with *Streptococcus mutans*, Journal of Bacteriology 190,13 (2008): 4632-4640.

121. Kursa, M. B., & Rudnicki, W. R. Feature Selection with the Boruta Package, Journal of Statistical Software 36,11 (2010): 1-13.

122. Lavrencic, P., Kaakoush, N. O., Huinao, K. D., Kain, N., & Mitchell, H. M. Investigation of motility and biofilm formation by intestinal *Campylobacter concisus* strains, Gut Pathogens 4,1 (2012): 22.

123. Lederberg, J., & McCray, A. T. 'Ome Swee 'Omics-- A Genealogical Treasury of Words, The Scientist 7,15 (2001): 8.

124. Lee, J. Y., & Divaris, K. The ethical imperative of addressing oral health disparities: a unifying framework, Journal of Dental Research 93,3 (2014): 224-230.

125. Li, F., Tao, D., Feng, X., Wong, M. C. M., & Lu, H. Establishment and Development of Oral Microflora in 12-24 Month-Old Toddlers Monitored by High-Throughput Sequencing, Frontiers in Cellular and Infection Microbiology 8, (2018): 422.

126. Li, J., Helmerhorst, E. J., Leone, C. W., Troxler, R. F., Yaskell, T., Haffajee, A. D., et al. Identification of early microbial colonizers in human dental biofilm, Journal of Applied Microbiology 97,6 (2004): 1311-1318.

127. Lif Holgerson, P., Ohman, C., Ronnlund, A., & Johansson, I. Maturation of Oral Microbiota in Children with or without Dental Caries, PloS One 10,5 (2015): e0128534.

128. Loesche, W. J. Role of *Streptococcus mutans* in human dental decay, Microbiological Reviews 50,4 (1986): 353-380.

129. Loesche, W. J., Eklund, S., Earnest, R., & Burt, B. Longitudinal investigation of bacteriology of human fissure decay: epidemiological studies in molars shortly after eruption, Infection and Immunity 46,3 (1984): 765-772.

130. Lufkin, A. W. A history of dentistry illustrated with 90 engravings. Lea & Febiger, 1938.

131. Luo, A. H., Yang, D. Q., Xin, B. C., Paster, B. J., & Qin, J. Microbial profiles in saliva from children with and without caries in mixed dentition, Oral Diseases 18,6 (2012): 595-601.

132. Macuch, P. J., & Tanner, A. C. *Campylobacter* species in health, gingivitis, and periodontitis, Journal of Dental Research 79,2 (2000): 785-792.

133. Marsh, P. D. Microbial ecology of dental plaque and its significance in health and disease, Advances in Dental Research 8,2 (1994): 263-271.

134. Marsh, P. D. Are dental diseases examples of ecological catastrophes?, Microbiology (Reading) 149,Pt 2 (2003): 279-294.

135. Marsh, P. D. Contemporary perspective on plaque control, British Dental Journal 212,12 (2012): 601-606.

136. Martignon, S., Pitts, N. B., Goffin, G., Mazevet, M., Douglas, G. V. A., Newton, J. T., et al. CariesCare practice guide: consensus on evidence into practice, British Dental Journal 227,5 (2019): 353-362.

137. Mason, M. R., Nagaraja, H. N., Camerlengo, T., Joshi, V., & Kumar, P. S. Correction: Deep Sequencing Identifies Ethnicity-Specific Bacterial Signatures in the Oral Microbiome, PloS One 9,6 (2014): e99933.

138. McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea, The ISME Journal 6,3 (2012): 610-618.

139. McGinn, T., Wyer, P., McCullagh, L., Wisnivesky, J., Devereaux, P. J., Stiell, I., et al. (2015). Clinical Prediction Rules. In G. Guyatt, D. Rennie, M. O. Meade, & D. J. Cook (Eds.), *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed*. New York, NY: McGraw-Hill Education.

140. Mejàre, I., Axelsson, S., Dahlén, G., Espelid, I., Norlund, A., Tranæus, S., et al. Caries risk assessment. A systematic review, Acta Odontologica Scandinavica 72,2 (2014): 81-91.

141. Meyer, B. D., & Lee, J. Y. The Confluence of Sugar, Dental Caries, and Health Policy, Journal of Dental Research 94,10 (2015): 1338-1340.

142. Milnes, A. R., & Bowden, G. H. The microflora associated with developing lesions of nursing caries, Caries Research 19,4 (1985): 289-297.

143. Moons, K. G., Altman, D. G., Vergouwe, Y., & Royston, P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice, BMJ 338, (2009): b606.

144. Munson, M. A., Banerjee, A., Watson, T. F., & Wade, W. G. Molecular analysis of the microflora associated with dental caries, Journal of Clinical Microbiology 42,7 (2004): 3023-3029.

145. Nasidze, I., Li, J., Schroeder, R., Creasey, J. L., Li, M., & Stoneking, M. High diversity of the saliva microbiome in Batwa Pygmies, PloS One 6,8 (2011): e23352.

146. Nyvad, B., & Fejerskov, O. Assessing the stage of caries lesion activity on the basis of clinical and microbiological examination, Community Dentistry and Oral Epidemiology 25,1 (1997): 69-75.

147. Nyvad, B., & Kilian, M. Microbiology of the early colonization of human enamel and root surfaces in vivo, Scandinavian Journal of Dental Research 95,5 (1987): 369-380.

148. Nyvad, B., & Kilian, M. Comparison of the initial streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals, Caries Research 24,4 (1990): 267-272.

149. Nyvad, B., Machiulskiene, V., & Baelum, V. Construct and predictive validity of clinical caries diagnostic criteria assessing lesion activity, Journal of Dental Research 82,2 (2003): 117-122.

150. Paster, B. J., Olsen, I., Aas, J. A., & Dewhirst, F. E. The breadth of bacterial diversity in the human periodontal pocket and other oral sites, Periodontology 2000 42,1 (2006): 80-87.

151. Phantumvanit, P., Makino, Y., Ogawa, H., Rugg-Gunn, A., Moynihan, P., Petersen, P. E., et al. WHO Global Consultation on Public Health Intervention against Early Childhood Caries, Community Dentistry and Oral Epidemiology 46,3 (2018): 280-287.

152. Pitts, N. B. Detection, Assessment, Diagnosis and Monitoring of Caries. Karger Medical and Scientific Publishers, 2009.

153. Pitts, N. B., Baez, R. J., Diaz-Guillory, C., Donly, K. J., Alberto Feldens, C., McGrath, C., et al. Early Childhood Caries: IAPD Bangkok Declaration, International Journal of Paediatric Dentistry 86,2 (2019): 384-386.

154. Pitts, N. B., Martignon, S., Douglas, G. V. A., Deery, C., Ellwood, R. G., Manton, D. J., et al. ICCMS™ Guide for Practitioners and Educators. in: 2014.

155. Pitts, N. B., & Stamm, J. W. International Consensus Workshop on Caries Clinical Trials (ICW-CCT) - Final consensus statements: Agreeing where the evidence leads. in C125-128. Journal of dental research, 2004.

156. Price, M. N., Dehal, P. S., & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments, PloS One 5,3 (2010): e9490.

157. Qiao, Y., Wu, M., Feng, Y., Zhou, Z., Chen, L., & Chen, F. Alterations of oral microbiota distinguish children with autism spectrum disorders from healthy controls, Scientific Reports 8,1 (2018): 1597.

158. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Research 41,Database issue (2013): D590-596.

159. Ramli, M. Z., & Azmi, N. S. (2020). Oral microbiome in children with caries: A systematic review. In *Materials Science Forum* (Vol. 981 MSF, pp. 271-279): Materials Science Forum.

160. Rebelo, M. A. B., Vieira, J. M. R., Pereira, J. V., Quadros, L. N., & Vettore, M. V. Does oral health influence school performance and school attendance? A systematic review and meta - analysis, International Journal of Paediatric Dentistry 29,2 (2019): 138-148.

161. Richards, M. P., Schulting, R. J., & Hedges, R. E. M. Archaeology: sharp shift in diet at onset of Neolithic, Nature 425,6956 (2003): 366.

162. Richards, V. P., Alvarez, A. J., Luce, A. R., Bedenbaugh, M., Mitchell, M. L., Burne, R. A., et al. Microbiomes of Site-Specific Dental Plaques from

Children with Different Caries Status, Infection and Immunity 85,8 (2017): e00106-00117.

163. Righolt, A. J., Jevdjevic, M., Marcenes, W., & Listl, S. Global-, Regional-, and Country-Level Economic Impacts of Dental Diseases in 2015, Journal of Dental Research 97,5 (2018): 501-507.

164. Rocas, I. N., Alves, F. R., Rachid, C. T., Lima, K. C., Assuncao, I. V., Gomes, P. N., et al. Microbiome of Deep Dentinal Caries Lesions in Teeth with Symptomatic Irreversible Pulpitis, PloS One 11,5 (2016): e0154653.

165. Rose, G. Sick individuals and sick populations, International Journal of Epidemiology 14,1 (1985): 32-38.

166. Sansone, C., Houte, J. v., Joshipura, K., Kent, R., & Margolis, H. C. The association of mutans streptococci and non-mutans streptococci capable of acidogenesis at a low pH with dental caries on enamel and root surfaces, Journal of Dental Research 72,2 (1993): 508-516.

167. Schrøder, S. A., Bardow, A., Eickhardt-Dalbøge, S., Johansen, H. K., & Homøe, P. Is parotid saliva sterile on entry to the oral cavity?, Acta Oto-Laryngologica 137,7 (2017): 762-764.

168. Schulze, P. A., Zhao, B., & Young, C. E. Beliefs About Infant Feeding in China and the United States: Implications for Breastfeeding Promotion, Ecology of Food and Nutrition 48,5 (2009): 345-368.

169. Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples, Genome Biology 13,6 (2012): R42.

170. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. Metagenomic biomarker discovery and explanation, Genome Biology 12,6 (2011): R60.

171. Segura, A., Boulter, S., Clark, M., Gereige, R., Krol, D. M., Mouradian, W., et al. Maintaining and Improving the Oral Health of Young Children, Pediatrics 134,6 (2014): 1224-1229.

172. Seki, M., Karakama, F., Terajima, T., Ichikawa, Y., Ozaki, T., Yoshida, S., et al. Evaluation of mutans streptococci in plaque and saliva: correlation with caries development in preschool children, Journal of Dentistry 31,4 (2003): 283-290.

173. Sender, R., Fuchs, S., & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans, Cell 164,3 (2016): 337-340.

174. Senneby, A., Mejare, I., Sahlin, N. E., Svensater, G., & Rohlin, M. Diagnostic accuracy of different caries risk assessment methods. A systematic review, Journal of Dentistry 43,12 (2015): 1385-1393.

175. Seow, W. K. Environmental, maternal, and child factors which contribute to early childhood caries: A unifying conceptual model, International Journal of Paediatric Dentistry 22,3 (2012): 157-168.

176. Sheiham, A., & James, W. P. Diet and Dental Caries: The Pivotal Role of Free Sugars Reemphasized, Journal of Dental Research 94,10 (2015): 1341-1347.

177. Shi, W., Tian, J., Xu, H., Zhou, Q., & Qin, M. Distinctions and associations between the microbiota of saliva and supragingival plaque of permanent and deciduous teeth, PloS One 13,7 (2018): e0200337.

178. Simon-Soro, A., Sherriff, A., Sadique, S., Ramage, G., Macpherson, L., Mira, A., et al. Combined analysis of the salivary microbiome and host defence peptides predicts dental disease, Scientific Reports 8,1 (2018): 1484.

179. Smith, B. H. Patterns of Molar Wear in Hunter-gatherers and Agriculturalists. American journal of physical anthropology, 1984.

180. Soncini, J. A., Kanasi, E., Lu, S. C., Nunn, M. E., Henshaw, M. M., & Tanner, A. C. Oral microbiota of children in a school-based dental clinic, Anaerobe 16,3 (2010): 278-282.

181. Sounah, S. A., & Madfa, A. A. Correlation between dental caries experience and the level of *Streptococcus mutans* and lactobacilli in saliva and carious teeth in a Yemeni adult population, BMC Research Notes 13,1 (2020): 112.

182. Sritangsirikul, S., Kitsahawong, K., Matangkasombut, O., Seminario, A. L., DeRouen, T. A., & Pitiphat, W. Breastfeeding Duration and the Risk of Dental Caries: A Prospective Cohort Study, International Journal of Clinical Preventive Dentistry 17,3 (2021): 189-189.

183. Takahashi, N., & Nyvad, B. Caries ecology revisited: microbial dynamics and the caries process, Caries Research 42,6 (2008): 409-418.

184. Takahashi, N., & Nyvad, B. The role of bacteria in the caries process: ecological perspectives, Journal of Dental Research 90,3 (2011): 294-303.

185. Takahashi, N., & Nyvad, B. Ecological Hypothesis of Dentin and Root Caries, Caries Research 50,4 (2016): 422-431.

186. Tanner, A. C., Kent, R. L., Jr., Holgerson, P. L., Hughes, C. V., Loo, C. Y., Kanasi, E., et al. Microbiota of severe early childhood caries before and after therapy, Journal of Dental Research 90,11 (2011): 1298-1305.

187. Tanzer, J. M., Livingston, J., & Thompson, A. M. The microbiology of primary dental caries in humans, Journal of Dental Education 65,10 (2001): 1028-1037.

188. Tellez, M., Gomez, J., Pretty, I., Ellwood, R., & Ismail, A. I. Evidence on existing caries risk assessment systems: are they predictive of future caries?, Community Dentistry and Oral Epidemiology 41,1 (2013): 67-78.

189. Teng, F., Darveekaran Nair, S. S., Zhu, P., Li, S., Huang, S., Li, X., et al. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling, Scientific Reports 8,1 (2018): 16321.

190. Teng, F., Yang, F., Huang, S., Bo, C., Xu, Z. Z., Amir, A., et al. Prediction of Early Childhood Caries via Spatial-Temporal Variations of Oral Microbiota, Cell Host & Microbe 18,3 (2015): 296-306.

191. Thibodeau, E. A., & O'Sullivan, D. M. Salivary mutans streptococci and caries development in the primary and mixed dentitions of children, Community Dentistry and Oral Epidemiology 27,6 (1999): 406-412.

192. Tinanoff, N., Baez, R. J., Diaz Guillory, C., Donly, K. J., Feldens, C. A., McGrath, C., et al. Early childhood caries epidemiology, aetiology, risk

assessment, societal burden, management, education, and policy: Global perspective, International Journal of Paediatric Dentistry 29,3 (2019): 238-248.

193. van Diepen, M., Ramspek, C. L., Jager, K. J., Zoccali, C., & Dekker, F. W. Prediction versus aetiology: common pitfalls and how to avoid them, Nephrology, Dialysis, Transplantation 32,suppl_2 (2017): ii1-ii5.

194. van Houte, J., Lopman, J., & Kent, R. The final pH of bacteria comprising the predominant flora on sound and carious human root and enamel surfaces, Journal of Dental Research 75,4 (1996): 1008-1014.

195. van Houte, J., Sansone, C., Joshipura, K., & Kent, R. In vitro acidogenic potential and mutans streptococci of human smooth-surface plaque associated with initial caries lesions and sound enamel, Journal of Dental Research 70,12 (1991): 1497-1502.

196. Ventura, M., Turroni, F., Zomer, A., Foroni, E., Giubellini, V., Bottacini, F., et al. The *Bifidobacterium dentium* Bd1 genome sequence reflects its genetic adaptation to the human oral cavity, Plos Genetics 5,12 (2009): e1000785.

197. Wang, Y., Zhang, J., Chen, X., Jiang, W., Wang, S., Xu, L., et al. Profiling of Oral Microbiota in Early Childhood Caries Using Single-Molecule Real-Time Sequencing, Frontiers in Microbiology 8, (2017): 2244.

198. WHO. (1997). Oral health surveys : basic methods. In (4th ed.). Geneva: World Health Organization.

199. WHO. (2008). *Indicators for assessing infant and young child feeding practices: part 1 definition*. Retrieved from

200. Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Braekel, J. V., et al. Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies, International Journal of Molecular Sciences 21,1 (2019): 298.

201. Woo, P. C. Y., Lau, S. K. P., Teng, J. L. L., Tse, H., & Yuen, K. Y. Then and now: use of 16S rDNA gene sequencing for bacterial identification and

discovery of novel bacteria in clinical microbiology laboratories, Clinical Microbiology and Infection 14,10 (2008): 908-934.

202. Wu, T. T., Xiao, J., Manning, S., Saraithong, P., Pattanaporn, K., Paster, B. J., et al. Multimodal Data Integration Reveals Mode of Delivery and Snack Consumption Outrank Salivary Microbiome in Association With Caries Outcome in Thai Children, Frontiers in Cellular and Infection Microbiology 12, (2022): 881899.

203. Xiao, J., Grier, A., Faustoferri, R. C., Alzoubi, S., Gill, A. L., Feng, C., et al. Association between Oral Candida and Bacteriome in Children with Severe ECC, Journal of Dental Research 33,7 (2018): 22034518790941.

204. Xiao, J., Moon, Y., Li, L., Rustchenko, E., Wakabayashi, H., Zhao, X., et al. *Candida albicans* Carriage in Children with Severe Early Childhood Caries (S-ECC) and Maternal Relatedness, PloS One 11,10 (2016): e0164242.

205. Xu, H., Hao, W., Zhou, Q., Wang, W., Xia, Z., Liu, C., et al. 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 (2014): e89269.

206. Xu, H., Tian, B., Shi, W., Tian, J., Wang, W., & Qin, M. Maturation of the oral microbiota during primary teeth eruption: a longitudinal, preliminary study, Journal of Oral Microbiology 14,1 (2022): 2051352.

207. Xu, H., Tian, J., Hao, W., Zhang, Q., Zhou, Q., Shi, W., et al. Oral Microbiome Shifts From Caries-Free to Caries-Affected Status in 3-Year-Old Chinese Children: A Longitudinal Study, Frontiers in Microbiology 9, (2018): 2009.

208. Yang, F., Zeng, X., Ning, K., Liu, K. L., Lo, C. C., Wang, W., et al. Saliva microbiomes distinguish caries-active from healthy human populations, The ISME Journal 6,1 (2012): 1-10.

209. Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences, Nature Reviews Microbiology 12,9 (2014): 635-645.

210. Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. Human gut microbiome viewed across age and geography, Nature 486,7402 (2012): 222-227.

211. Zaror, C., Matamala-Santander, A., Ferrer, M., Rivera-Mendoza, F., Espinoza-Espinoza, G., & Martinez-Zapata, M. J. Impact of early childhood caries on oral health-related quality of life: A systematic review and meta-analysis, International Journal of Dental Hygiene 20,1 (2022): 120-135.

212. Zhang, C. Z., Cheng, X. Q., Li, J. Y., Zhang, P., Yi, P., Xu, X., et al. Saliva in the diagnosis of diseases, International Journal of Oral Science 8,3 (2016): 133-137.

213. Zhang, L., Budiman, V., Day, A. S., Mitchell, H., Lemberg, D. A., Riordan, S. M., et al. Isolation and detection of *Campylobacter concisus* from saliva of healthy individuals and patients with inflammatory bowel disease, Journal of Clinical Microbiology 48,8 (2010): 2965-2967.

214. Zhang, L., Man, S. M., Day, A. S., Leach, S. T., Lemberg, D. A., Dutt, S., et al. Detection and Isolation of *Campylobacter* Species Other than *C. jejuni* from Children with Crohn's Disease, Journal of Clinical Microbiology 47,2 (2009): 453-455.

215. Zhang, M., Chen, Y., Xie, L., Li, Y., Jiang, H., & Du, M. Pyrosequencing of Plaque Microflora In Twin Children with Discordant Caries Phenotypes, PloS One 10,11 (2015): e0141310.

216. Zhang, Q., & van Palenstein Helderman, W. H. Caries experience variables as indicators in caries risk assessment in 6-7-year-old Chinese children, Journal of Dentistry 34,9 (2006): 676-681.

217. Zheng, W., Tsompana, M., Ruscitto, A., Sharma, A., Genco, R., Sun, Y., et al. An accurate and efficient experimental approach for characterization of the complex oral microbiota, Microbiome 3, (2015): 48.

218. Zheng, Y., Zhang, M., Li, J., Li, Y., Teng, F., Jiang, H., et al. 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, (2018): 361.

219. Zhou, J., Jiang, N., Wang, S., Hu, X., Jiao, K., He, X., et al. Exploration of Human Salivary Microbiomes--Insights into the Novel Characteristics of Microbial Community Structure in Caries and Caries-Free Subjects, PloS One 11,1 (2016): e0147039.

220. Zhu, C., Yuan, C., Ao, S., Shi, X., Chen, F., Sun, X., et al. The Predictive Potentiality of Salivary Microbiome for the Recurrence of Early Childhood Caries, Frontiers in Cellular and Infection Microbiology 8, (2018): 423.

221. Zou, J., Du, Q., Ge, L., Wang, J., Wang, X., Li, Y., et al. Expert consensus on early childhood caries management, International Journal of Oral Science 14,1 (2022): 35.

APPENDIX A

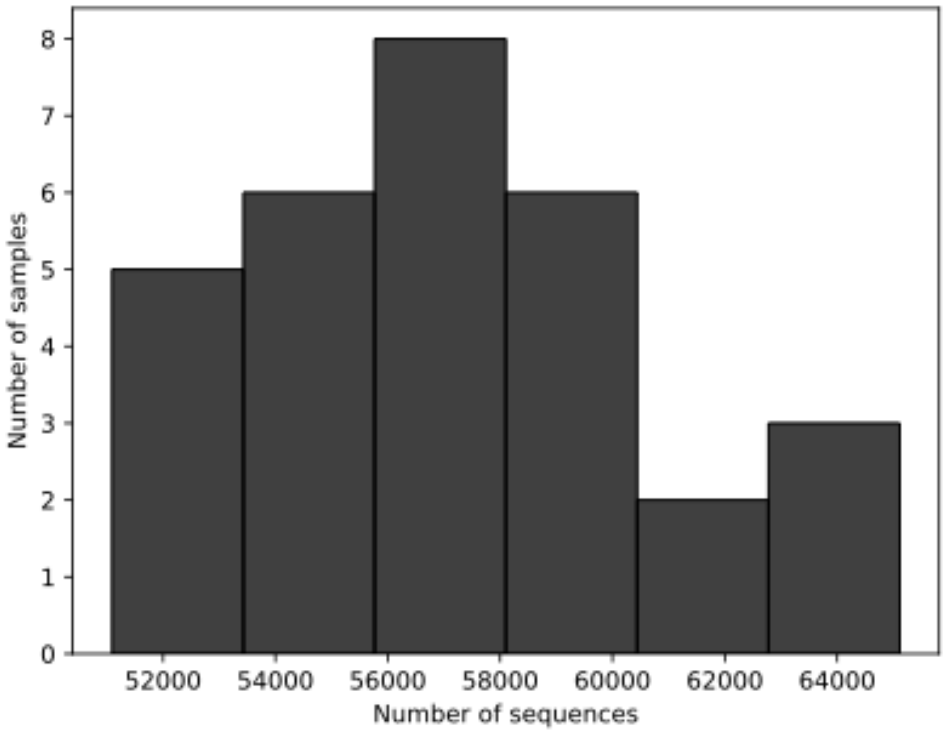
Results from QIIME2

1. Sequencing details

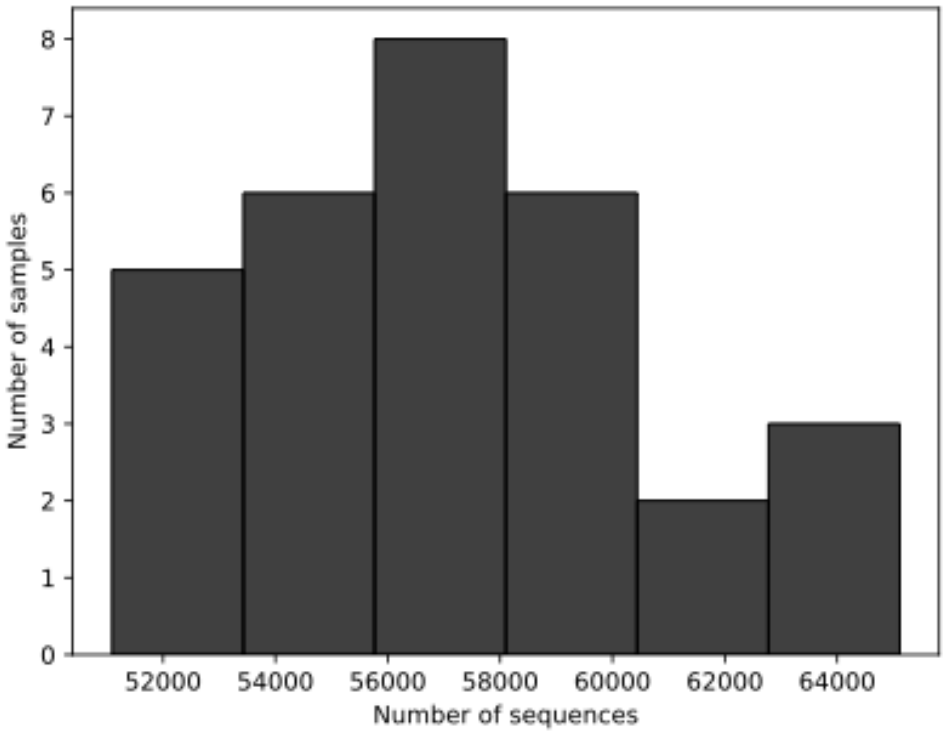
Demultiplexed sequence counts summary

	forward reads	reverse reads
Minimum	51093	51093
Median	56693.0	56693.0
Mean	57261.9	57261.9
Maximum	65114	65114
Total	1717857	1717857

Forward Reads Frequency Histogram



Reverse Reads Frequency Histogram



Per-sample sequence counts

Total Samples: 30 (forward) 30 (reverse)

sample ID	forward sequence count	reverse sequence count
F2D04	65114	65114
F2F04	64590	64590
F2W08	63802	63802
F2F06	62774	62774
F2D08	61587	61587
F2W09	60095	60095
F2F01	59893	59893

sample ID	forward sequence count	reverse sequence count
F2F10	59869	59869
F2D03	59691	59691
F2D07	59070	59070
F2W02	58556	58556
F2D10	57895	57895
F2F09	57696	57696
F2F07	57610	57610
F2D01	56701	56701
F2W06	56685	56685
F2D09	56352	56352
F2D05	56042	56042
F2D06	56015	56015
F2F02	55346	55346
F2W07	55233	55233
F2F05	54773	54773
F2D02	54252	54252
F2W04	53809	53809
F2F03	53547	53547
F2W05	52885	52885
F2W03	52883	52883
F2W10	52343	52343

sample ID	forward sequence count	reverse sequence count
F2F08	51656	51656
F2W01	51093	51093

Demultiplexed sequence length summary

Forward Reads

Total Sequences Sampled	10000.0
2%	250 nts
9%	250 nts
25%	250 nts
50% (Median)	250 nts
75%	250 nts
91%	250 nts
98%	250 nts

Reverse Reads

Total Sequences Sampled	10000.0
2%	250 nts
9%	250 nts
25%	250 nts
50% (Median)	250 nts
75%	250 nts

91%	250 nts
98%	250 nts

Summary of DADA2 run statistics

sample-id	input	filtered	percentage of input passed filter	denoised	non-chimeric	percentage of input non-chimeric
#q2:types	numeric	numeric	numeric	numeric	numeric	numeric
F2D01	56701	50696	89.41	49335	30543	53.87
F2D02	54252	48485	89.37	47650	28519	52.57
F2D03	59691	52936	88.68	52179	28503	47.75
F2D04	65114	59302	91.07	58441	28013	43.02
F2D05	56042	49684	88.65	48892	28308	50.51
F2D06	56015	50338	89.87	49505	29037	51.84
F2D07	59070	53188	90.04	52174	32297	54.68
F2D08	61587	54513	88.51	53656	28912	46.94
F2D09	56352	49778	88.33	48873	23402	41.53
F2D10	57895	52215	90.19	51522	31619	54.61
F2F01	59893	52927	88.37	52199	29680	49.56
F2F02	55346	49258	89	48415	24428	44.14
F2F03	53547	47963	89.57	47253	25408	47.45
F2F04	64590	58554	90.65	57271	37828	58.57
F2F05	54773	49233	89.89	48300	28602	52.22
F2F06	62774	56587	90.14	55783	32402	51.62
F2F07	57610	51912	90.11	51011	29765	51.67
F2F08	51656	46449	89.92	45767	28871	55.89

sample-id	input	filtered	percentage of input passed filter	denoised	non-chimeric	percentage of input non-chimeric
#q2:types	numeric	numeric	numeric	numeric	numeric	numeric
F2F09	57696	51781	89.75	51018	29924	51.86
F2F10	59869	53821	89.9	53161	27262	45.54
F2W01	51093	45685	89.42	45256	18207	35.64
F2W02	58556	51781	88.43	51136	29483	50.35
F2W03	52883	47115	89.09	46473	24621	46.56
F2W04	53809	48378	89.91	47760	24755	46.01
F2W05	52885	47583	89.97	46719	30959	58.54
F2W06	56685	51215	90.35	50397	28590	50.44
F2W07	55233	49845	90.24	49186	24056	43.55
F2W08	63802	58183	91.19	57657	27055	42.4
F2W09	60095	54398	90.52	53501	31893	53.07
F2W10	52343	47032	89.85	46336	31749	60.66

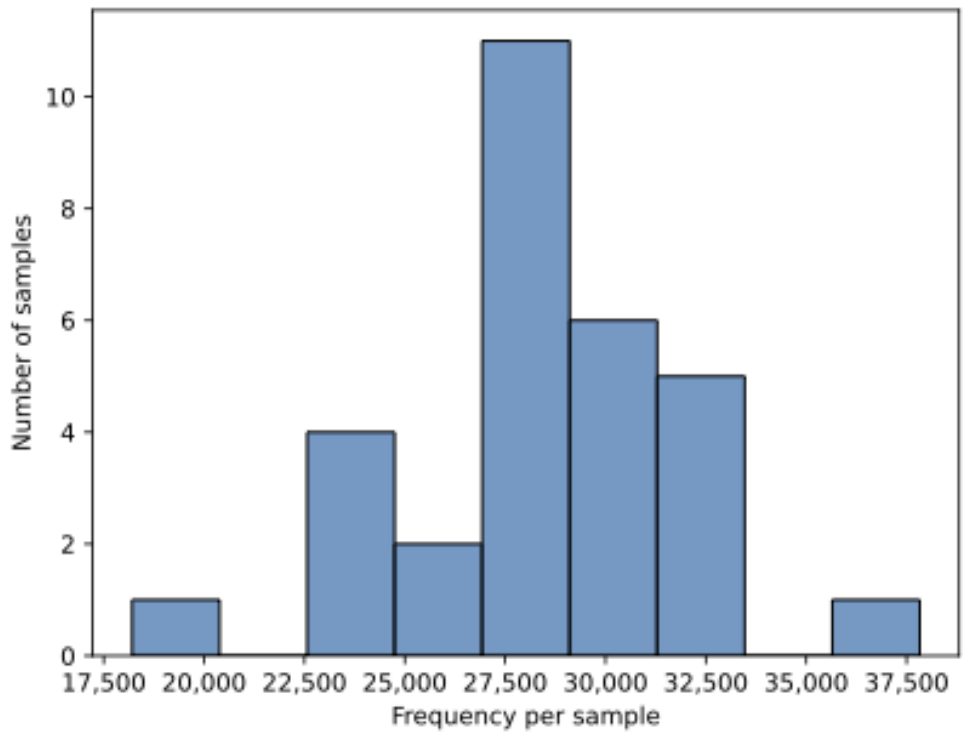
Table summary

Metric	Sample
Number of samples	30
Number of features	945
Total frequency	854,691

Frequency per sample

	Frequency
Minimum frequency	18,207.0
1st quartile	27,106.75
Median frequency	28,736.5
3rd quartile	30,388.25
Maximum frequency	37,828.0
Mean frequency	28,489.7

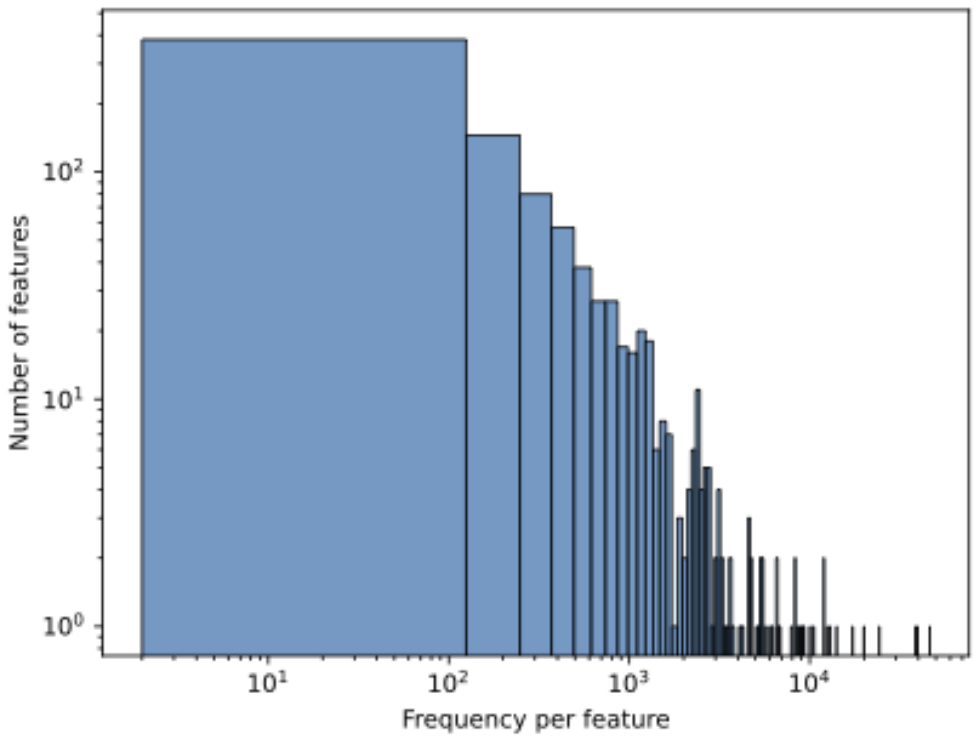
Frequency per sample detail



Frequency per feature

	Frequency
Minimum frequency	2.0
1st quartile	48.0
Median frequency	187.0
3rd quartile	652.0
Maximum frequency	46,195.0
Mean frequency	904.43

Frequency per feature detail



Feature count in each sample

Sample ID	Feature Count
F2F04	37828
F2F06	32402
F2D07	32297
F2W09	31893
F2W10	31749
F2D10	31619
F2W05	30959
F2D01	30543
F2F09	29924
F2F07	29765
F2F01	29680
F2W02	29483
F2D06	29037
F2D08	28912
F2F08	28871
F2F05	28602
F2W06	28590
F2D02	28519
F2D03	28503
F2D05	28308

Sample ID	Feature Count
F2D04	28013
F2F10	27262
F2W08	27055
F2F03	25408
F2W04	24755
F2W03	24621
F2F02	24428
F2W07	24056
F2D09	23402
F2W01	18207

Feature detail

	Frequency	# of Samples Observed In
4cf73cff34ec1e1fad4b3485b060f2a8	46,195	29
fc38a354a7217151ea797502380e976c	39,493	29
5411b490ac1ea4b24f99d790bfe98250	38,460	29
32d2b339d4f15316e8d9196976d393f8	24,295	17
9ce625cc75c536cab814de241d7a04de	19,932	10
b08a9b66256c867dceb6eaf682f2a64b	17,204	15
20443bad42654f6016224a61ad6f08bf	14,076	23
0f7c4fda4f4db224b4bf526229ae2d56	13,017	30
bd03ed65a8296835ebf9bc0bd49c1d45	12,503	12

	Frequency	# of Samples Observed In
100250149a161bdd6b0f91e2af354157	12,030	12
1a26690ed6103f33fafea102e270ef54	11,984	29
8d0b3f3c723b9f9e8776efddd5425559	10,537	10
7d81ce8699a74343a55d39f989e894a2	10,002	7
85b311bb5dde236817b6021bb9b4f1fe	9,315	6
703c33259c384682a0e092b26985e25b	9,171	14
23a00e8312408d4229316b83174e8da4	8,900	30
649613f7161ac9af412238df5ea5f5ad	8,775	15
422a9557d95fdb77f808bf36b0562442	8,492	20
c532ec32b0c0e01f556aab710eeac530	8,242	30
d43c3957d6ceac23cb16db39163b6dc1	8,239	9
d849304d9ebff36f1cc4be30596e3479	7,887	12
881a45e8228f920312d8788e0d1eb158	6,872	13
e5286d2bcbd90faaee13161f10421f28	6,728	19
b18de10b2d1c64dc6dea08b706f8bba0	6,605	19
8407da518e5703caff1fe314b1deecc2	6,549	30
0d2da84018b5ec61b1106ed2eec21324	6,228	19
d92c76b469e9cfbfa214b5e721107c62	6,053	12
6bf054aa5cb55226856496f403d57dd6	5,702	14
df6ee4a08f85b4cf4674284898f2a4c8	5,468	8
15954ffc3f5d9c5437272c19a74cb692	5,442	5
e9d5e493edbb20a1a18c25fae392d82e	5,372	5
557588347979532cd37abddaf70a2462	5,345	3

	Frequency	# of Samples Observed In
e304fd4c064b7021229b587a19691fc2	5,191	6
5723a380772a1c3fa0719f3b9d5367cb	5,124	16
3781316a3cd7289328ed973eab77ff51	4,697	18
115da0da7eeced4c6b168832ec616978	4,673	5
9f3122b9db83ac0a7d2f9974dec24d54	4,611	23
7190f13305c8837d8b9c9c193467d969	4,576	5
4fc5591b266c4274bac5911d57c47e36	4,549	22
84de3c2855e938b6a02d90fe08a29add	4,273	8
62f9c0f8be078d3dd34460e48d412c81	4,095	11
eb53e450d944dbaa69e6c4577d25831a	3,754	9
c5ff80489f369205419a19042a1902b1	3,672	4
c4773ebdb4173dca07bf3aaba0c84538	3,654	4
eceb098726d3066efc315ae749af7ce5	3,560	6
7d710be491d9d1dc8c0d46cd8837841c	3,435	3
d988a16828ec2a47b8507a796c64a2ce	3,315	13
abfe99a3b7196fcd6dbfca5857c31480	3,203	18
c97bfb2d73133acc09035bb2cf84a9c8	3,155	4
bdc423d8cd5c2236eec1a29d559ee498	3,114	6
53abab2d024f795dc7aea08a32a70beb	3,104	6
b538f0d4606cc874d5b6d1ad03a3aee6	3,084	3
dbd950c92afd0106567ee05d517302c7	3,046	2
81048cbc5dce5355af45d92f7eb90e90	3,026	16
4f60449b91f1840ef5323d912f1f1437	2,943	10

	Frequency	# of Samples Observed In
6aaa7a94e04d1b8830cdfb011bc4bf98	2,793	5
a576e2094a11183b6e7098ef2df66ee7	2,782	3
a140f920875450d955faa975cb697c0d	2,762	3
5216aeed2be23c6eddc5224a948782a5	2,742	3
e2d04a3cdf4f8effbdedc98b2f364b8f	2,713	8
f3fe6a265563583664050ac8a74a71b4	2,674	6
5a5f8e668f8d37039f4cf471d32de183	2,666	4
d648417d8ae27ef63cf87df9f5d1defa	2,632	3
44678eb70622de779c9730e7d9ffaba6	2,616	6
9d44c7350f634a4bbf2282d58e4f28ed	2,582	5
8164d42c05d73a67ef418b6c6ced6faa	2,564	4
07a7fda1379642dc5908ac3c7dd5957d	2,564	3
346b43eb18e06717b725cb75f21b41ec	2,537	6
2d72922c386a4518cbcd6bc94d0c5a12	2,464	5
5443ae28761fcc5725a25c1a09159dbb	2,459	12
3b3c82890e6dd54895d9088342eefc61	2,454	5
e36185fe279017c52513e2ef0a5e86d7	2,430	3
f5f709b2f303e66c552f4f2dd338a8a8	2,429	2
60aed765105aed7678299d66e045b726	2,415	2
c55a4fe1fe619140bea1a877686fff06	2,401	4
22c7e4325e8fd8dc7c1cb4b90ae7110d	2,381	2
3d8c1563a3e0a023f47688c2434b04a0	2,369	4
e85a8227df4dd759e42a760fad3c3e67	2,359	10

	Frequency	# of Samples Observed In
dda4a1cd8836e76789197267b5e705db	2,352	8
251c76de24e7714ae1faf89bfaea1e56	2,348	2
c53ce795226bd77a21d8c4b428a068ff	2,334	6
21ba99e1a58b905504d927d86875bba8	2,287	2
a952d4b4042c0d4e162cf57ece10016f	2,245	3
fe2f82a5344823387dbb4512823ba2e7	2,233	6
81764ffd3e58bbbb33fb968dfe46757c	2,226	7
33d139ea24c2a75f3bb4764dbc86526f	2,222	9
7eb5e02a48bc966103cae776ca64add9	2,180	2
04d5bffd94aa3c4572444d888802e619	2,151	9
df1971a0a8c06f3c5949e1ec928d089c	2,113	18
7c2a97c884f223f2d155787677d665e2	2,091	6
ca0f40d28db6cfaa8a86158b048abe0e	2,012	3
48f69e4344582efc2ba801ff3c1b3c7b	1,990	5
fb6aab3ac80742d780a51336e8a56114	1,901	5
644d34797a6d8912004f53a2a9b91934	1,868	9
a77d978333d690a86ced7a28ce57e529	1,847	4
d9310683843e0aa8bbc50b5f948e2bbe	1,763	3
3cbb483273842a92e4a3c3209c681ac2	1,700	2
d7754a67fd0a2a2b51dbf5fc19167af2	1,699	3
ebcae4fa2868c0d27b01061a91ad0b03	1,699	5
61fea6763759370d20b155df325cc7bc	1,662	2
2f4f4b4fcdfdc073d6d963ff2bf1b0af	1,636	3

	Frequency	# of Samples Observed In
13ceeba63d37722b6260b61722836d09	1,634	4
e56ed8c9b3b6bb463cdf16a2f30300ae	1,626	3
bf7ef3dc6a00561b6b92c1ff2f06ba26	1,553	3
93c91a2c0a61accc0e4e5693a7126cbc	1,550	1
ae56c826e826412e8fa8217751eb595d	1,543	3
f07e8c000c23e462b5bc592bccd85478	1,542	1
1c5dffb8938c304a570448bd04fd5cfa	1,522	1
37f9ad242ed72f927f4f1c577de73763	1,510	2
31acaee353c19e2da8b89724c3b73311	1,501	3
ed6231ed0f5f5c60f9a72aff03b8a686	1,484	1
3c1e9f7979bdee1d61204e9ae0e72519	1,429	4
6f715a7f32509807b94ace0158fee16a	1,400	2
2f27df775d541c0d877ee6898e0789f7	1,394	4
1127b1c5113e540ed28109c43714230d	1,384	2
f8b455b2c00715e8f9ed52b8914e6ad1	1,360	3
d7f3d864229eefe566d2d2270ed2913c	1,356	1
fb29b6851a27e3e0d654acf294b484ea	1,352	2
71ff8243573ae3adb272e891ced4dfdd	1,338	1
10d2662f22735b091059db7aff7d5b33	1,337	4
3ffbb98f77cac33f0f6c9df5fc5da11a	1,331	4
1456993147cd3d7a5ce63413bfc78f80	1,327	3
2c6fb94ab738cb7add923408c215ee42	1,326	1
fbe901934ab023bcda410ca2f25c3d5e	1,322	5

	Frequency	# of Samples Observed In
ff2cbad91d34a2b5b051df2b36083d1a	1,310	1
32a0d02ff594884708a07da5fc618818	1,302	1
d5515fb61179db25cc1a30950ac449ac	1,298	1
a6752365acb0ed90846e3eea97c537c5	1,294	1
29e423b29c029a48d160a24ee7f0cebe	1,289	1
db560ef2e2cf7be33c684de43aa48942	1,280	1
215fee3cfe20774769d72e5db016793d	1,276	2
655d24e16cb567adef4d50a8544c85f7	1,250	1
86357b11bd59ae9912c1d1e416cd8efc	1,245	1
23db4cb9f04c11a3c20da0c9dcdc507f	1,238	1
840f7cdd9a87e3f232590665a50059c7	1,231	4
4c62004242a77b2d76e72dee358eea32	1,222	4
13a2a23793e6b4583a53e2d8edb57c23	1,220	1
f43e05e185a5b7e116d171a624f4bfcb	1,203	1
e8c1a0e110b4db118ab7864c9959fa25	1,200	1
f66b05d3ecfb2be5b1e02d1379cac29e	1,199	1
6cd282be94801fc420ce43856c542346	1,191	3
03f0eec0d2fef1b0a59a40c6d2a5302e	1,189	1
38cd1a3661f5c15c4e76224928e21422	1,184	8
0b0938a63a2b9f169f17fbc89d330680	1,177	1
f99ae09068271675a18b6f04056c688b	1,174	1
6597213a9b6fef8da224f17a414665ba	1,166	4
699fa6b5b4a491d6b5c41ecd2fbdb6b0	1,156	1

	Frequency	# of Samples Observed In
677ccd5d6ef498a23babb02630b60f32	1,149	1
15f875e5464fdfe87144c0237b6d6acc	1,136	1
65ab6d5cfe155aa4949ef0bee78cbb46	1,131	2
bbb7909a4d2e725a230083e6f53c2e56	1,130	1
ea7e16b613a2f829cc301a4d8018763d	1,128	4
6bee4001f9da30080ef22c9e25ad6b6b	1,128	2
a7ba1ab79345294600e8e9f5e5347ffd	1,121	1
fa2381a453d641f38dc2fd49c3922fa6	1,117	14
f8eda33e8c63ae2274b8ccfc97e8875c	1,101	1
8b6ab29ff64b50c66644b56ed1a79a3e	1,100	1
e04f0110e83930eedddb1610c768242b	1,084	5
7de3603c80efd9c4f7ca0115b548339b	1,076	5
d765a44fea61a3b47be2ea9b5defde05	1,066	2
0357026e65a0255497f5d860ab762a35	1,058	4
64e406c068eb2e2f0d0865ba396df1b5	1,054	2
fec2153fa4d8b2be91f54c0a9f958bcd	1,038	2
489037f060a5d5a1c423ff850413c421	1,031	1
d552e6d7f4543d8eb8ad5c63d5409da5	1,026	1
6aa4e721b5d46d636039999891ff6f46	1,025	1
f72247585585d13420756f565cdaed1e	1,024	2
95f6585457205325f9292a0c93edea28	1,013	3
f76a4c64d1364e656ce8c1c4e703df37	1,008	1
1b3db460ff7801a9a41203f8f97685ed	994	1

	Frequency	# of Samples Observed In
b2bda9b755becbcfb6a847342aa923d7	989	1
a0713ff079dc11fc51cae1ccd58cb269	961	1
d9212ccf7295cf558da87b5f0d51ab77	952	3
0acd35896a14175bf8fd990d4362015e	941	5
183ead9d96861a09c5e2b6a5bd147d77	931	2
bbdcb524a2eb55e34f16fe2fe4447da5	923	1
a52da5b84a1481d92e05b32f7d16815a	921	2
e44b29e86c70d1c19ba4aad11710d9f9	911	1
5655f5645e5f6a07d0d62af974ccc3a0	910	1
9753c8d44cd57f31f8ed635417dc2b6c	907	1
17811ace7662a019fcef97ddab63c191	907	1
8e35dccf7d058315b2b4c612700c3a4c	899	1
89d7fdfffbd327219f3bce35f4fda14	895	2
b45f83146c3ceb0fe07f6b89c82dbf40	894	2
a52ed69d81ca5ab208f7a2076e3947df	894	1
758900d5be0d0ff03fdc533544b43cba	883	1
9b4ad931ff0ddd855046b0bacc788beb	880	2
7d8e6d8c302469f864cb2e255245b4dd	865	1
6190ae8a2c89a8bebba999a8de081997	855	2
1eaef510cc370187f7675a0e852cc9bf	854	3
0775de959985d62457a75343798ef7d3	851	1
2df41a9e219d216368b20343a398d601	851	2
b353012c3d041762d13e9b0b2b25c7f4	844	2

	Frequency	# of Samples Observed In
7f332539645bbd62bc6a0e2d61bc8c5e	843	1
f155a2a9cd6479146fef894da26db7d9	836	2
e20a4ab18ce6e3232e8ed4f8ae589ddf	828	4
033d9e287270833dbd272e919c3f13c1	825	1
51254e3411b27b03a03f779e67e5987d	816	1
b0fb78d824f10bdf4966e91c18590751	814	3
2ab4ce388f076e620667cc34e3707866	801	3
98e135ad2255ac91333072ea01ce65aa	799	1
6e99ffb781824b3901afe093e6e13b81	794	2
f1ee1e5c8fd29eeeb25fcb06d4ccd0cf	788	3
dd5fcc6566119ebee8a97938707d3e22	786	1
26faa6c4b2f60e8718b1c106366aac7a	782	1
bbc602b197ec62daf3093fc94c18921d	779	1
ee4b34c374c04d35057bac984b703f28	773	3
d3818dae776eac1ad69822525d0b5c77	771	10
6efdab2e34ee877d47fed621d7bc40c1	767	2
9421f8eba60d0c2202564422d70f3324	763	4
843c29cc8a795efb81ba959c1e5fe55b	754	3
bd21c36b70e1e9af4b1ba3fba0a84c90	754	1
d0fdfdd70419780017380cb894919e8a	748	1
00e510e968dcf023b70b158d10e25fd5	741	3
60c722450f8e3f4be80327f977f84af7	740	5
ef705168b866db69a2b1091cfba71e4a	736	3

	Frequency	# of Samples Observed In
6b118ca4ad65714569cd236032b9cac9	732	1
645a86e3db60b1b6eaba5bbf8a943aac	726	2
355d9e189c0d5459fb4487ad2a7261bc	720	2
1b924bf32a152e3b188273bd69e9283e	716	1
45dab101dc7ef071233605bdef2bf10f	711	3
4fb5b5c690b17e77f3bd257d5a0e25a7	709	2
965a0c01f78e36af3793b313df426f3e	708	7
99e7205b7f9f518773b14a96bbb716dd	707	1
6baa82b870f439cf572c386cff2040c0	697	3
bc26b843d94135c6d69e0ae01452baf6	697	1
27f2e2fd013379ec02ef33e590f0d2d0	682	1
2b739af2453c13f09881051d5f463591	682	1
44257e017d61aab5b621f0c028f91b21	678	1
adad18f2e1484f65a6f6ac8300dec3b5	672	2
274781a9b7a59d7ed4dc3489dae072c9	671	1
cf52f56d488117ac5e03c8ab5d4b0cdd	669	4
7c984317e7b98330265dbf37333bcace	664	1
794993bec9d7f9c171f14a729047ab8c	664	1
7412d9ec9a1e7b3f07a8b8d5e168acb4	662	1
09184d6dc484f07ae43b339db28179cb	658	4
c17868131c91a329a98b03393ca1c7e3	652	1
111403548eb91f6107d3f4dcf640ba8f	639	2
25421b96fd9ce2b756b68f7046ba1855	638	3

	Frequency	# of Samples Observed In
4636325fdc85efa47f287963a6df3cb3	630	1
73e7bf14e4ccde6f73f5dc561f1ba65d	624	1
85fddad762228ec866310951354e2b36	618	1
6d54daa041d47743168f7c28e308deed	616	3
ecb7abab734008e804aa7ce2fdfa7af2	615	1
d11a67a8957fb5b5e1d3976dd5349287	612	2
a2a7bbfe38abafd2f896dbe87a79334a	608	1
75a8739894c23a30d3d8acd6ff9450b2	607	1
ab4e981b308dc28a3c09686c44d35654	605	1
b7da704bfed3859ef3efa8bd188214a1	595	1
e443bc23d2e01daec7b8f708f2da4589	595	1
c4d36369a8146cb33074d7c0b89f192c	593	4
a0ec2c7825935897b8acbbec43c592e9	592	1
b6d718ca8b2e7dd959dd63549a33010d	588	3
cc3e1a1e0f459f2d610b4a336dce5d6b	587	1
5a2daabe2b2fec36455c0e67f1b97db5	586	2
dc118a34eb35363f28fd6622f488b14f	571	1
9ab429aa28d743ea920abd0a66d9fc74	570	3
058256901ac8365887c3e0c3c19cf469	564	1
b03d8ef52f7c2e6819eb02fdebb2e6fc	560	1
6aa49f4de24a79f80029fbc6237e3ba8	559	1
81be2fc44b238a8a64b8869f591e189e	552	1
c492c8b6316b0959109e7e2010083bc9	547	1

	Frequency	# of Samples Observed In
2272e25f719cea2b2101ef4114dd2286	546	1
30327184862e0449147204eb0b2acb90	546	2
13ba3e889ad58fe17ef0e12fb40a9757	544	1
616294db19f2fc4785a02648609592f5	542	3
653db4a1723b106eea69f340372be513	541	1
fc4d7ee9f5f4ed5da3a4a990d3a97a8a	538	2
c78a911eb4dbd6eea90be0086379e51a	535	1
b49bd6a1d281c77b8a5544147c302451	534	1
ba972647db1d99f21d21320fffa5466f	532	1
bcfc6c1782ffce9634029e2d47caabca	531	1
158a4612a4282a8746be0ed6b1ce5878	523	1
70dd4460693acae5dbc1399026b80586	522	2
249230558a60f2ec85cd876f61d7cf1f	519	1
cd54fc4000ec3c91b625d985944ecfc0	515	1
299bc5e935415d95a026d208987598f2	507	1
8d3cb75e8ca2d1dbbf98daf31be3e8a4	505	1
7c27be8390cf1def665727f81801febc	504	1
d54e5d4e9f9d9f4b400653861f9411d9	501	1
7801263af66a4b0c02e22d0e46bfa66c	492	1
544fc9778e9209d51f053b2cab6581c5	488	1
7d85cf02ae1f15bb8b4983986908df8f	488	1
683253b425187e988cb628480de7fae0	487	1
31197929d5a35310692efad17b4c3d3c	487	3

	Frequency	# of Samples Observed In
bd90542801791bd5827b9d2cad04b8f8	486	1
56e44373d36e5302e71a7b28e95d9a97	484	5
8715b89d2fbbd5ed2c378afe03496e22	482	1
8389224a0c3e8c424fb0c67195454cd8	479	1
cda3eefdb688d34c1a883c6daf43a3e5	478	1
d23d95e8edc0da3416efdd0fe0b60420	476	1
79109664ffccf55685f79194e8058b78	475	3
1eb0b862ba25f3749bfd9122e69e8d6d	470	2
6da6f665e32a8b0eace590b0936a964a	465	1
1e3320a94522ac0e15616c1c1eb47246	464	1
e18b2ba4965b17f16f79f0be205d9592	461	1
2de29a9bb8ab5471644b7cc167ee516b	461	1
912ecc50d78084373ad5d22d9ccc3cde	456	1
8f14b08e04d058809483582a07cfb13c	454	1
b28b3d82bfd4bc6329298200f89dbe16	454	1
45f84f9d834e248c33bc7609176318ea	451	1
6b43dd0f7ce0f15d60408a51f18aff80	446	1
f9b3b9031044af09ea40007c8b3296f3	445	1
23512897fe7ad9a9bb7c7acc97f2885c	445	1
dd93194efe67eec62bb5917c6987e3aa	444	1
4032b7083318c948a18c7ee60f7697f2	444	1
1b22e4bf6fa4b5fe23cf545f4f0aa773	442	1
d8526e3c6d8c5c8ee19ed4eb32eae9ea	440	3

	Frequency	# of Samples Observed In
f169416487ddece9d4e6643703ed7732	440	1
80985dddcd9731ec913ad6bcc87d6165	439	1
277226f5c06f3759d48511f375d5b8ed	436	1
d1ba999da67447c4cc93deee5b9e5b8f	436	1
c49b198f3cce09edf559c904a080767d	434	2
de8302683e1ce5f9325a00ea2e5bbbd0	434	2
5c7cb5e967272b7c20b3a3d3fbcce775	431	1
b4641aebf94899e104afc485eade8711	430	8
45bd89862406ea1da249de8377478372	429	1
0a79b1f053bd5ca2c53597b67676acff	428	1
66fe05eb8ae6dcf4c60806296bc9f1d9	424	1
84a2fc5004b72eef339be7b31c6cebf8	424	1
c5a91bae99c959ce5b5c23bd924578dd	422	1
8df3a2a90c77b864f21dcd7f07de0586	422	1
ddada3383b1b3b2d146c39080a6b6c4e	417	1
e93b1f6cc584c9664a9c1fda71d2ae45	412	2
4491efff629439b1051fd19259e779f7	409	1
0150c8b9569e2998e742dc80bbd8b9ea	397	1
8fa23b717b6d6909d0be7ca547e85ad6	394	1
de9b6b4b499795e7b8134985c17890e9	393	1
de5698e14c1ed2c01e392a6f08429b1c	385	1
e47202322d19467c2bfa78cf71fa2b07	383	3
aa3859ae9f06fefbba12e4a7420bb7aa	383	1

	Frequency	# of Samples Observed In
72d76ff4c0ae4027a35e93e4527f1806	380	2
7ff407f96400124f04c0e0eff70d7bb3	379	1
d39dabd4d38b27608f9a62b2eaf80e2d	377	2
aa8b2144ae0c63f134eac8122fea8720	375	1
4f9b95a143f71b08d4f25cfba2657f32	375	1
16c579388266debd23ba595a0d62b3b3	374	1
fda2b750e2203bbc012a545c280700ee	370	1
747b24a59088d2ac7c7fda24ebd43366	369	3
f29bde4ec11692c6b85dadf17d54cb07	364	1
11463f956b5943e769000c40ddc51c62	364	1
b6177d0e61c273bd2eb4c29bc738351a	363	1
1febeb3eb05f1a2e959ebc8cb4738437	361	2
8b8400e7d99b73cbdabd04f03bbef79e	360	1
98400c9c6b9ef7bd01c975f69b77ba99	358	1
f2fe7962bebded971fe1cc3cff0c1ad0	350	2
719a20fc31bbe704be36d06b89abe616	350	1
2758cae224d634e279238f14a395c125	346	1
ae86ad0a07cb6273e28366800c08362a	342	2
6d75310405c122ee67d5504824f4fa26	340	4
fd43f335c4e59f7d63d24360125d6df7	340	1
932997c11f6f0e47e2b7c3879bcaac73	340	1
b5cec4c0cab735a5bc9edb4bc4e9b0a3	339	1
2bdbf7e548a29525b7f622e44d49a9ff	338	1

	Frequency	# of Samples Observed In
40e8adf94124091eca4d66396a0b0fc1	333	1
191cd805d9877b0af7694f6aa3a659d4	332	1
da465f1f4cdb2f706a24883db48fee34	330	1
2973e95c2874780053362eda44e87f26	329	1
31833996528d5fedd3f2e8250cd88562	325	1
c0db3bddf4c7b0afb60dac2c466de505	325	1
237e171ebf574ddcbaa7b1980a3f4c54	323	1
bc9ec63394ae8fc7d14f6e7a5eb4dc97	323	2
7817ad1c2215d3d900a6e4703d494893	321	1
b659f07f7411a8d142c4590dbd4f3f6f	320	1
a7608034cf4a2cde02b1c2af8760f813	320	1
0e9b447237a98ad7573efb4dc43d0e37	319	1
16b55063b8a71be530be8e52c88d1d66	318	1
da8c2335f36250c8cffd892c66654896	318	1
2c6a3d81a116ae4ce6b187675071196b	318	1
2b9e5bfcc099ed5101b559486dc0ab3d	313	1
4970943f013dc3274fecd6e186adf0ff	310	1
516ec942895665d811353af5d052a5bf	309	1
bc641131cbbc5ec57671a4d55775055f	309	1
accb3a89a64d13bfa4b9bc62572b9a45	307	1
a8bcf4cdf5b35ac0a28199d2cefea203	306	1
6f9bd2e5b50ef3042656a60936188d77	302	1
5105f826e316bf50d3b1d2181b254cd6	299	1

	Frequency	# of Samples Observed In
2ed0b9764e0717ab3d8e799b1182450d	298	1
10790e5126b4bd2db6e9d1b34e9fd2f3	295	5
5d603a63feb7cae6a2f4fe7c043108c1	293	1
5ab0738efeeb581acad9bd6d9c801822	292	1
8ec8bfcb66095f975ba8c9bf3c485f0a	288	1
0aeadd2522eea3c6f677fe1fbc7d98df	286	1
05d383a0209b10f54806b3b8e11bfb7c	285	1
fd0ed24def3e9782f8945791932d023a	284	1
d093161fe8da7871806658216f9192c0	284	1
ca816aedcce3ad50804f4affdaeba637	283	1
60be133672fe16ebfccec837a97a0564	281	1
f0d8e41db1291e11c0f25824e2e34842	280	1
55ca69e34c379bfb49da82138fe799dc	279	1
84c644018cae5b5fc83910faee7c18b6	277	1
1e04fd2bb0c35b64705f3a98c58c1c09	277	1
2ad3892d93cc46fd5404c7cc6f37f549	277	1
621dc8ea3922809e3d5bb726d2102076	275	1
0f967c3cdf2aa6bd7ae8e15ec9051459	272	2
b3da695a65b7b1ba9d457da4e565cf6a	271	1
64d4510d116168167aca968d79e7d407	269	2
6086360609bbf267564e8068768813b7	269	1
9dbe9bbbf5ef4a1776df2989bdaeda89	269	1
ac2782870a1419b6ba7115c1385d99a6	268	1

	Frequency	# of Samples Observed In
ce5b548a8a6e872f096cd9c9c56c3586	266	2
38286ae48eca2ee15d5e753ce09ff804	266	1
b9c3f60f7846829d726789dbf4aae8f2	264	1
e1297d42591f9f6dd60d7b8e324e6a2b	263	1
4968b42bc17d62cc76074514a5bdaf9f	262	1
3728ba756fb18824392759b1560ad149	261	1
ecf8c46c1af810a41031356c0eb2d068	260	1
e9eb3cd36ec834e3f7ef4b8f9983d876	259	1
c43d8ca507770d99d2474f37abdfa307	259	1
dabf6955d97e8007c5672efe0c6c9df7	258	1
88dcb5f3342206b0d262da57c4cfefe3	257	1
21f6dedac2a137835ae9ccf42797a4bc	257	1
e3cef310df8f472e2ca6c8e8f82a4e34	255	1
dc401640b279ebaf57c61f590554cf04	255	1
7c65d6385227f20efd9eedc1310bd57c	251	1
d66b6424dc3a4e4a25c80eb44df1d0f5	251	1
3ac2d4b89c5758b468d14a97652e0702	248	1
2a2ed58a0b9e8f6be9e2eeba4b015a28	246	1
da0fdc25b33b08b84619c3afb10aabbe	245	5
974aeae7064a2a97ca1813af50314d08	245	1
d1ea8b834d8c94110785dc8b091ac57d	243	1
0e4313ddca6ab4c1721747c46450c4e7	242	1
e6e83bc745b5a7f22910cb3a433e4320	241	1

	Frequency	# of Samples Observed In
ddfe66912fe0e9f05ee8172e7ccf9d05	239	1
0ca304494871c8422d60d2ccc73ed14f	239	1
a1da6d20889a67feb434169329d72f10	238	1
78e33f1503d385757946592f9fa348e6	237	1
76e6545b8a6cb7f99a3815e717385302	234	1
48beec2c94dd1c6d7d3872f3e50cfaa6	232	2
9eff0a4fa1ed1e4f911c6a05feaff3fa	231	2
382a18b40f74a049e99c5f463cc98ffa	230	1
beac40c63e4fa9c9ff0b543cb2451a49	229	1
81b800de0622923306c7992860d37a1d	228	1
33be0eeeb744c7fa3537be81a58ae6e6	227	2
eb980ea4ae622e8df3790185aa1213b5	227	1
decc2ea371605e9239598b687580e3a0	227	1
aab787522a0def952d283155e5209ed8	227	1
592f0f093007de2672553482d7b5a9d4	225	1
c39c1bb7c0bbb0fa7bad31e0a89af87e	224	1
1fa271b4f0164797f331577117e62e09	223	1
db010d3e05c06787665b321f168ffc55	223	1
108be5192e0ca38a1694544efc3775fe	222	1
2f4595c5a4518a264a8e78ad388c5e9b	222	1
74bad92b8dd3c70f2f85e82a8cd34d43	221	1
f53dcb10fe00c6510e75dc6a9652f859	220	1
97ff21939b956d0b58dca0d54dc8c5ac	216	1

	Frequency	# of Samples Observed In
e418f5d8282f01e36695b571291de8a1	215	1
d292f4ef0f9c4d13880db3d0c4973e34	213	1
558bf707e258b69ef1a004a67390699c	213	1
77a82685c190925fa5f76877dc4588ed	212	1
f3d9ac50c5a82d0368104ebcad607e3b	210	4
f8e2bc2a15b7943fe0d8dd1dce771a8b	210	1
1fdf2ea71bea2b52a4489f53b6bbf37c	209	2
acff88777a210543cd0eab85ce9646af	206	1
9e3ad279b06b66a8d8cefb9caab205af	206	1
841ac7c62fdfd8936ec25c8cddfb6d0f	205	1
a1ae82d937c38480d885dc9b024a5f25	201	1
9e958b42b96d1b843dfecfda68d79f81	200	1
b0ec342bc905b3997987c47e797c4d60	200	1
0ae7b576a66192741bfade44706234c4	200	1
238d25d4e8397a7a1af1210e15a5609e	198	1
a2dedf642ef6dc68fed94fb568dd9a8f	197	3
47bb94fc2bf455d1ce96832d8e756945	197	1
e591fc3dd991d516b9facac91bed22e0	197	1
c8904746ab24082a975a5dc8b0c1ec7a	195	1
bcafed60059e76273dce9984a2fa0615	194	1
e76db4d0ceead9bd07bf8a903f866161	193	1
f8aba27bdbd6c3ba9a8917aa30cd76ba	193	2
d0ff1043c5970bbd4cf48560bb84b7d5	192	1

	Frequency	# of Samples Observed In
89b019c7ecf7fe95733c66b03da762cb	192	1
9cab3715f64993d159e6ee3a62598fc8	191	1
2ae0ffc1e19eed8bd000aaa7c34ac75f	190	1
6552dde3e0c73a783233a3d5fc84f9dd	187	1
678f763f6b4c8ca27dda357e8f57db06	185	1
32a42d4d29ab3611302550300c0f8ae7	185	1
f75b2d49fba0563704f78fb127c2a554	184	1
3ee7620c980edddeab6ccf9422640c0c	183	5
ffcffe55f2719bf20b7a8ab1040bd922	181	1
dd85a9362458e1af9e667707510750a4	181	1
1e8d1139569050b5911ca78622d06830	180	1
1ce8ed8ae6615218b902ef7e7e61f6a3	180	1
dad218d3ed8d5f56a72a2dbbd6130f33	180	1
8b344e6cdc20b919dcf445ec3690e5e7	179	1
4267b39e24b4e11c799973a4afe7c03b	178	1
4aaedba49b8cd7d050e0b0c098035cfe	178	1
83ba39e8927ac01f3f9c43bd7c3a589f	178	1
3a18041f30d1dc39542a3ec1ec019ecb	178	7
5d2addaaff770de4ba4bbe43f323a8a7	176	3
2b7728df4e23cfe8dac1fe25567bf62d	176	1
b1436232f3e091f215147d283541f332	175	1
d922057ea6d8a14b79a32ebfca50ce2a	174	3
c476b00e26d0b15ba5b96cb424f96af0	173	1

	Frequency	# of Samples Observed In
6b0bf33726fbfbd78dcf70ce05cfc93c	173	1
b5a12d12394e4bdfc4250b150b78d57e	172	1
0136edd19b6116e1b1f51b2c9f502f0b	171	1
5f3ef951f64050ce814a72b5848b18b0	169	1
c90b2ba95b88136f5f4c407b6690d47b	169	1
5b2fc65fef51b415e71a43ef94bc1e33	169	1
71ef4c47ffe35ae2624705ec169fa0eb	168	1
b84a11390f172a8958e63c7ac3275171	167	1
4a5bc1d41e40250fbe74b69d35758c53	167	2
3d37c203448a855c42440f7985578ba2	165	1
24cdeefbc6b38c10bc51adc2734029b4	164	1
ac6fd7c1fccad58cc6d4496fabbecbb2	164	1
9fa6b787e43e41f1b23f93de12435476	164	1
e89703dca71ef978b852a76f5ad33e5a	163	1
8ad554bd1df04f85ff7d8f0e347d9dcd	162	2
7b3931b3e4141a3c3011d26229dc2973	162	2
969f247af14d5258ba691787e8a9841c	161	1
ed6487fa5e444e0e801f7defc3fda3f4	161	1
54c1a0d548ea4d141290701d1397ec96	161	1
cb6470394cf73d69cb4326636398cabc	160	2
f7139ddbae2d74b1863111a6a670a539	160	1
eb666c0091a87e79f79701a6d45d45f7	158	1
1e80056e9ebf91a0eea4ab58661c2863	158	3

	Frequency	# of Samples Observed In
55b5da798124cf921d3f7e78ccbf83b8	158	1
3833dca3b8e23088ca141c42a9e304b5	157	1
5a078fbe863eb17d511191de3548a8e0	156	1
1742d12de08a73e35cbfa9ab09364249	154	1
9e3750234d9760450fcbdcd3a98449f3	154	3
bf949c3ce2e58063778fec435a13f2f5	154	1
fba7cceaf882b8d74dbf543e2249a924	153	1
d6a18a6eeb321e4e884721c858ac6dab	153	1
f772c6cab604d209140824dbe4075aac	152	1
a49691ef0139cfd8ab8f655286cdda49	151	1
69b4cb916eac3a98e4fd9314f8b4c6ac	151	1
9709c95e737b1f1ab80e94949fb14b40	151	1
747fac4e413b740150fdd7dd89bca983	151	1
4e5e4762643a3ada4f18377af41c8767	145	1
e0259e94b8943b6646842b9ab1ff1695	145	1
bfd0c791e84940737b24df4e1b2daab2	145	1
f920e41e815e5a1e31fde01bdd873ca6	144	1
346a822df956d834aeb5561ce6528eec	144	1
164ce7a1560ba9aab32873fc7f558380	143	1
b31285a97747c759d2248c3be68a57fe	143	1
27ae7b72987af334135a0bb16ba68757	142	1
6ada132dd037bcb4e09f074f2748180a	141	1
7fbc39f634636ae08461cefcba70d08d	141	1

	Frequency	# of Samples Observed In
a57efee0b70943a24f80f12bc856adcb	140	1
5886b3f2277135adad46300dba86a6d3	140	1
95b78451b40831688bccf91745e2b6dd	140	1
cad5d3b06c064be2fdb2b687c045f62d	139	1
d9331e83abe8d1b1763cb8007c91bd0c	139	1
7c6524a576b867bc2b05e903bbf95edf	137	1
8db38fb7c5c617c308fec20ad5b3675e	137	1
f77d0b3f4681ba6d83c3daeaeb0a0d3d	137	1
e6d9701fac813123b92c38c252532b67	136	1
e9af5c565342971f32fb7d940b99142a	133	1
40898f612c0398550af7d723fb490662	132	2
6a1c977142858d1148bc41bec207d462	131	1
1f266e9f3639f651278a44515077e372	131	1
10bb3f13d98930f51c21e58747fa5026	131	1
accd96aca10031d84ee13fea206d31e3	130	2
db45b3580bfa8f29abd95547d673508e	130	4
3664cd8742f96a93a83c58c988395dc5	130	1
0ea1d75cc3b133594fc22bbe83b47a62	129	1
190e48aa7eef5757db172a3844d554bf	128	1
33fee3b186495367bf294bdc9329d47b	128	1
ecf99bed296dbdcace50651d8d31d5f1	127	1
6ed934d7624673bcf8a12626cf0412d0	127	6
2fa163361a19c86a1310db5450302e32	126	1

	Frequency	# of Samples Observed In
9889101906c8d5674eb0f1568be4cd44	126	5
27858bf0d08f0c2fa18d4a3d58c63b0e	122	1
2fd30e29f009b5ca3272fb2dd09287fa	122	1
9e407f5c61e03ea8dad0162e08c9b4e4	122	1
47948ab2032e9155c87048e159251fc2	122	1
d967125bb6a28843d2257fcfd0eef876	121	4
35459fb96913eeb7f117c9da1e3dc87c	121	1
c4510c7cc6f68ed0023a78b4dc7e8422	121	1
13791f4f7650485718bae707e883e261	120	1
b143ca2b409db08388cf3bf0ec0105e9	120	1
f7862ea4d057098bac02b0a58cb45caf	119	1
40bb838beb1405e01f5f79eb4c6d5d96	117	1
a0e579591262e7e74111807f38e43ef1	116	2
2fb1ecc78075fe7bdfc90f21c2b27e54	116	1
008c50b71483761f65b8301411c46c04	116	1
30e01a525fa4630b1222e37fe4b74b13	115	1
45d2358b70b407700f9ea2122a21f2fc	114	2
11f867b6c41eb061b6c817dd39b22a42	113	1
d47a8ca3cecf46c87a0e889f79cf9382	111	1
a66e2b5dafb3d72ea7c7025d2533afe2	111	1
267f3130ff411f3d2b9584a0c2a7138b	111	1
fb495c5d973e733bcad11f86f53bffa5	109	1
bc34e2c4ce8d0692925cb39a68576637	109	1

	Frequency	# of Samples Observed In
3fba03e2f606f60d1fbf8e66197828a4	109	1
2f33ef0f1e322ded6bbd05b78cc026c1	108	1
304a6475097536c08abe20ce071cb0a5	106	4
e4877d8d6fa9e60c4fe4b857a4604465	106	1
7de9230551c73d4e597d25e52152743c	105	1
d74c5f18bd231431fbff25f4623546c9	105	1
000760fd0df204d8af62efefdac99a29	104	1
df8aa33ac159c19dd8e0ef074005c0e0	104	1
a2c53066b5015311021bcf8ca589e5fd	103	1
aafa38e0f95aff1b8f65614dbc847e21	103	1
db289130865951d0fc8df1b06548478c	103	1
e5b84a81de95fa83e6c7daab08084cd5	102	1
bc114e94819ff44207566732f7bc9c61	101	1
a32b38f2c848ae383ca52d3c441a754c	101	1
99d9967d7a09d67c427323bde61b28e2	101	1
55ef87814a69a1e0a9defff416aef9d4	100	1
7b74ae0861fab3990e20df721c879a30	100	1
349c047c567abb99f28d008091e97321	99	1
a0988b91063e8507a3c1f29e3a8425a5	99	2
2a40c63047f9221af6d6d0a1091b629e	99	1
a2555a1261c9209c54a4645b6a21a359	98	3
d553e851e782e5028c5106f2b9d86408	98	1
71e3fa22192bdddd68a1ed51e40a3468	97	1

	Frequency	# of Samples Observed In
da2de2adbdfa157bd62776433d6c50a9	97	1
561587a8db1beeb2928fda92f6be63ce	95	1
a63d7bdec51e1b9318ef67408f08beaf	95	1
f248f6d53dbb93b82efb0e2fa762ad41	95	1
7eda57bce9679d6bb7c950d528715feb	94	1
a5e9e52ed592028707083a32e92db10f	94	1
b2e12d4020f790ffb6e362c99d7b820c	94	1
7b8f79e290e7eaa96b8f4db4c0610aad	93	1
f08dd769fe370f83ec37d93f92362982	93	1
7dce551de1d79168678e76af0af6fd2f	93	1
d4a90a494496bb7dd4fcfc540d6a0af4	93	1
2369cbc196871fdeab74c3fb56661b67	93	1
1383c12970b73e9aad78883dd966e4f0	93	1
59444dddae7d2e9b7ee21ef709744c2c	93	1
241b187a6b57c435f3323eaff9e69a18	92	1
f94cb99ded9bc1f974f76e2c971d2712	92	1
652fafb39f24ccf5d5d720df3813cf60	92	1
d6591defe7248b8a6791466a6a246c88	92	4
47d21323164abdb57ae5da03cd7e2782	92	1
1b28191f1b07710cff8c4641946a1d94	91	1
832c53c8ad9ec723f59a5374f47ecd61	91	1
39796e0d1217f3938bac2b9bd272e5cf	91	1
55031d8c22abf6f8769287b50a948418	89	1

	Frequency	# of Samples Observed In
036497d57758d4a8b437b985aab86265	88	1
867de7c9328cc775428f311ed66993e4	87	1
3a8ccb79922c8c46190eeced0154ef01	86	1
75faa77c92262607a397181918b867da	86	1
7a9ac30748131e321ddca43193b68704	85	1
da51318dd4ab8ce85e4d65632ef3c854	84	1
78d5db145ea8d27dac6a729f68749ccc	84	1
e0e242b7800226249ad148d0878f9736	84	1
d1e9b9fe7f3a1ca87dc7a31d73605dae	83	1
5f92fac893e5412ba120d58197e69cbc	83	1
1c2c3d1e9d84efb94f685d60583b8344	83	1
e7283ecc878d827e3c2550ee11efecc4	83	1
38f5039b94347fd4b272a28d0ee628b3	83	1
1a31a96286eceb156435b88a9a9de702	82	1
adf7e01cab9032bd0c54bb5624b33c71	80	1
d6c116e865a05bd41a3dc7c8a3f309f5	80	1
4dc5ac768a688d2e304e137024d3004d	80	1
1f0ddf269388ad8ca07fb70755f9ce4b	79	1
4130a3874a94a2fd6412e9e4a9e53409	78	1
7eae189444cda928bd5a0ef6ba8f75a2	78	1
36845defdc8c74a1edd0c124f279512a	77	1
4701ab4d14efcb855ef3a8789104f465	77	1
309280d205275981908deec0177091b5	76	1

	Frequency	# of Samples Observed In
13eb5360d9ffacd41dbf8fe61a1f1108	76	2
6dd385e2c2f0bdc6880a9c227533eea2	75	1
5d1c95ad28a69aa3eaf67fafcf06d84f	75	1
4a7ca189fe0d0f5afee33c55ce26053b	75	1
df3241e0e1926e65b6275114aa948cd4	74	1
9d52b5cafe2e235a0b492ee70f3728f0	74	1
7cecaa1f10917429193451bf95acbe52	74	2
540a9dc9669b1a77766799c6f859b9d9	73	1
e694b2189d2134adceebd3d1c097ef1c	72	1
42f9f5c76555dbc80ec30d14cc1bd293	71	1
6624aa58d845eb9fa7de1097b1ad25be	71	1
1ece76d118de4bf7f10d537c7f4f913d	70	1
a6a8515bb0dca8100fe8abf727c012fb	69	1
f57699d89e2cda2a118058c657a0ebb4	68	1
07ebd1ee5cbaa7716106df0c7b46da44	68	1
0bd40067bd02ae5da73b7f6ccd68b807	66	1
4dd9abd5fe396101b4c29f2611bc846f	66	1
bee8087742cb2cf27bdcfabc4fa40535	65	1
9c21401dfefca70da18b5f92803e0d7f	65	1
450d93d1ae2bb1287b03b8433c0e5fea	65	1
b85f32e9788b57ad07e3cbc658f03e5a	65	1
f608f0da0277ae3dc1d643c0d3c4d30b	65	1
cb002dcf78d0b2fc3ce098105d04a345	65	1

	Frequency	# of Samples Observed In
3225da7190308bf09edb25d8c6c63427	64	1
dcb2d79a3a0c0157c0169eca6d6193e8	64	1
ef0b10fafdeb152f699fd1ac87a01024	64	1
6140570d52371d4c1605d5aa2c7e9278	63	2
8ac93c2d771e4250f7f54cce868312d5	63	1
d35adaea845961635ef9da979f670800	62	1
4059a21cc96afd6997ebfa04c6e1bcd5	61	2
e84fc83610a9d4b8a1582166f3996747	61	3
5149c23501b97a7904c547d508df2314	61	1
66a9732bd49e4130f1b67d74a9890167	60	1
8ec0e5c3653b0e0d0fba84e657f771b2	60	1
72cb297bd4fe6c60b693a7639e3523fb	60	1
9dbe45a9e1471967a842a22ea3d225e9	58	1
c4b82668dbd417c78318120fd1a38050	58	1
e4c8c68a9c69cd0ccccf715b010cb214	57	1
7b97c7ab2a4ea9f2b3f12e4c1004da44	57	2
24a9b9c0504bd56339974d2696229fc3	57	1
c825e2b4e52e41704fe47480bc42a67e	56	1
99a4b96e8d14ada9789c6e46f960c18e	56	1
a26931635040a0ff4086f8b6bc72560d	56	1
41321ce62bd0cf90954235544dfbe583	56	1
6fdbbedd9d7b83a3069c08e844b46e80	56	4
f8e8420ce06c8c0637467b695a71ba49	54	1

	Frequency	# of Samples Observed In
1933a75feb05764413367060624f384a	54	1
368f0faf6c9c592fd774fa097dd12590	53	1
f8a80efc00f962a5000faf6d15260206	52	3
2e1e28a5eef4b3868ec18f11bf58aadc	52	2
6b5f35bb6734b666ac02075c406eb62b	52	3
d2372b90b22377b41786cfae1ec54b72	52	1
9ac6d338661e86dec5cfb3b434237790	51	1
e720f9b84810d590da73c5f398bc7b00	50	1
6b59b25157d5f0e93a6505173e1664bc	49	1
b70b46e823236c0f052a5982d2e11597	48	1
c82068a53dd45730b64884cf5728c74e	48	1
c9db2c0cb2a78dd477630130ea6faca7	47	1
3ba0a61327ab7603e6457506ea47c76a	46	1
64018917b341c76abf5a26eb50965433	46	1
8d66750badfe99dc6efe704434b4d015	46	1
64cc6e10f88fe9a17c7032974badc31c	45	1
abeebf2cb1f8a6d9b776be8770bdd067	45	1
c9221537e35169901ab09403072d8d48	45	1
7f1cc0cec693188073dec4e2594d5e82	44	1
fe694ca30247fbf379bfd88f9adfe686	44	1
77b393df3fcae00fd9aaffa0af36cfa2	43	1
c36d31d7526f2510c5524a30b655bd05	43	1
11d92d3f9ef489cc00b8d5900147d4e7	42	1

	Frequency	# of Samples Observed In
38137b00d0869d9a4584e39b3a385764	42	1
8be3eba94a0aab17ec8a9205f177b344	42	1
9646cbd39cf13ff74aa2498f74d33269	42	1
23a7599e6f9741899b150d92614ef01b	41	1
550d7a5fd7cfef4660ffb56dd42217a3	41	1
b941f7d41581fcd500076ff463cad580	40	1
95c0423a0dcd37a1c8e16d1c82530f58	40	1
f133a85952341aebdce67e67e3c9d778	40	1
2faf5efe675bf50fc2305756b4713871	40	1
f10e74f27c0d859a47f3fe1f8d528f27	40	1
e0378f74f896e563f87e80d490c5ff49	40	1
3e663cf2fa3aecb5301836b1fc46f03b	39	1
acbb339a63ba2954be9b099b0faeffbe	39	1
0b77f8b406074583d271464bbff2a72a	39	1
27161108f541611c024bc1efa6b23726	39	1
a2067d9ff08e884e803745516e377f46	38	1
9c1c3ff8f8c02a322206b9f3bb89d32e	38	1
0be63e7cca84b251fe573e591f06caa4	38	1
af9d5b71ad29ad99a78bb96aae5f3bef	38	2
4e2a22c909a7fc8b50112efc8195ca25	38	2
ee85b01154af6a8a26ff9eb9fe8d19fb	38	1
4f18ac7e0ad9a3a376a14a1e39891670	37	1
5ebd4275c44cbb3fcb80fca68a3095bf	37	1

	Frequency	# of Samples Observed In
d33da4aff45460ece3ee14546fe95276	37	1
95c3ac5e47b846f16d05c87beb10934a	37	1
77b8f83c891d9d147048588487c51b3f	37	1
462e3114fe9e135bee202bd19152d50b	36	1
d6068797252a5b7d994d8c090d4985a0	36	4
1a1cacc8890eb72a61ca075c23db1fbd	36	1
53263eaa91b8c66aa308b796419d78ea	36	1
b05e9f1ac744b64d6e35c3ee65e379e7	35	1
5192e1f5ec47e817dc3c3c9ea1f7e97d	35	1
3f376cd7c6af18b1b45f23a912eb7efb	35	1
30a1f44816a74e4a188d322fc3deeb86	35	1
d93c99dd814396b089ec21076b91ffab	35	1
6619853c10993fe80d513455bec5dc02	35	1
c35981f8b56a75fc44de3a5180b36ace	34	1
daa114d72149b0d760614edfba75e05f	34	1
cd08fa7c19c774477b2412d1fa2ebf1c	34	2
e3210eefe2301699d14de70a76c8a614	34	1
59e8378fafc9b8637835c6ea5778c808	34	1
0b4843367d2550cd3cefda8ca0378ef8	34	1
548ff7f9fc41bc51a6f7e6cdd14397ff	33	1
a550996afb7652eb4595d5192f19d788	33	2
06ddbbc9242f08a43db5a088b9d88e27	33	1
a8851447f9e53cabec81e4e01eb8eb76	33	1

	Frequency	# of Samples Observed In
65651271976a1fe04849fe649c2fcaba	32	1
2f020018878638b9b0a088170f1999e8	32	2
26328e87c7906b844284a19fc7c608e1	32	1
e1123a4a0827959776421957f5bdbaa5	32	1
ecab0d288c06f32ddce4c8c808cc37b1	32	1
7d2f0951586f7192b3c6be0a7b31fe1e	32	1
14520ad2a846f78708d193e4d621026e	31	1
098ed39b13cfca9703a19b27038575da	31	1
63911fc79fac2d1133168e744b8d2fee	31	1
b025b64a27db72d0a231619eac056ecd	31	1
9f8f19c6f999752b06fc985f59019d48	31	1
b5764c68d157ff5b09ca6ed2137fbfc2	31	1
a478e128e6641961c98ade127d0bb267	31	1
61a487495d720402cb70fa9d527ec75a	30	1
adf54af13ddf38b69d98677092579b72	30	1
2d4c12d3c44f749e4aebbb35ceb5b670	30	1
d5c7ad0d2ab67e5ae373324298baaa79	30	1
6b308a79fadaefb6d2fe7c3720de305f	30	1
a736eb7a13acce17431149250c952e7a	29	1
89a412837abe692bb7e7528f07b6a5ac	29	1
b34b95ed17a20f8db6988dac6ec4ac26	29	1
babcec1ae7485a0c31f0428469b1d963	28	1
1718a3465a56102da49b564450841cc7	28	1

	Frequency	# of Samples Observed In
30d428fda673ad8bcabc5a4184d1f287	28	1
0e643f06cbaa06999656faf244e36272	27	2
3c03774aa8cd76dfd791af651d7ec41a	27	1
bd6d9a19652697b491c6d9f47c50e1a1	27	1
71af8fa501f19b1f06c0687704413dfe	27	1
af6983ba85555f3e47eb9ab8db824cd7	27	1
54191b4777e24de27d34a737b4c7842f	27	1
61fb0f64dde933104a4df347a9faf9b4	26	1
5fac07edfb1ed313f53e9eaa55f550af	26	1
eeed50b8965d8fb00cc1075ecd908f16	26	1
7832316426a02a0b19648deb088cc18a	25	1
b7dd8d79a097f9e04a9f2917d23e1507	25	1
04da17e85384dbdd2ecdc675195401f4	25	1
48384847ebef96cbf26a0be13212a9aa	25	2
83fc1dbaea3ca289e34ab992c19466d2	25	1
182c3994239333a1896904302ef69c6a	25	1
be45ee678f894eb9dc264206f829b934	25	1
c74965dda48c4e300e5d01e3e3638893	25	1
44e9d50158a8042aa015dbc2db19aded	25	1
f4e4609a25565469e96bb18ef8eaaf6d	25	1
d421dba7a16ffbd061e46d440786dc1d	25	1
bce4b9ff0cb1a29bec45ccc23ece74ff	24	1
64f8588a5beaa3f544892ba991066e26	24	1

	Frequency	# of Samples Observed In
e93ab8e6d49c2454b3698ed48ed5f76c	24	1
96a97cea91fd3da909a6cb2273c5af3b	24	1
2a4c0d66e241830699d99f7ed8783f39	24	1
c78d2ebbe048ceb807b559e60639fd9e	24	1
51d4c464ec9cb72dfba0adca24948a59	24	1
e7ee430e93800004f21293a6becb99c7	23	1
9beb14cbd8fc89da7aceda26095cfd68	23	1
700b99ca548341cac9edda0cdbce2e6b	23	1
a7a7dfc94fa3b677e6354e367a0f046e	23	1
e3b6ddf335a69df05de1720f2906707e	22	1
bfce5724f071fb1571fc76f57161f25b	22	1
30cb6b0e6d0ed0a982b5b88b9fe6b7ff	22	1
59b424ab669b3af910fd1bac7bec31eb	22	1
6daae29652f538adfc5cf36316a9cdf3	22	1
812122a21304f9f57dc225056d472c11	22	1
43effe984f9d140b134f0697ad894f15	22	2
ab1a2a53d53956b3c17988eb713fab36	22	1
987c05ae3d48eb97e7977a166da2d816	21	2
ca7a84dbf3b99cb90b15bcf4fc965cbd	21	1
698969486f5018dbfbf79f0425b50e04	21	1
fd01239ab2c5e655728c4d0cfe59f4c3	21	1
217c68b6365a52e87e2c33156ef40aeb	21	1
6403f3b0ab729a3321d58162eea91ecb	21	1

	Frequency	# of Samples Observed In
19a15460f24e454ce7ba173bfa852f66	21	1
f15237fcca0b92b2f812553f4adddcfc	21	1
3c4489ecb0ffcc250b3f091626bcf0df	20	1
6ca9686a9aeaa800e704673112d01684	20	1
70a8c6b8f02502cd28e404f5fa0fc106	20	1
da1c0514e8af3ae21018c95fd0db49d4	20	1
a114134c503b90b894f96b36819c0e30	20	1
e5ff2fcf167a2e81b2257aa09dd7f5a9	20	1
05106fa1067a64ecb2218c89a6da313d	19	1
022afc77021bc0654a163d423ffdacf2	19	1
aee85f6bb07988d3b5f7e2c93d252835	18	1
6db31ae1f4add429aabc84420bdeec84	18	1
c1fa01d749832ed1b6737e11191d9ecb	18	1
316a22ccca6d73751628d33a45735aee	18	1
3cc54305ba9c9e11597593761e7be6ff	18	1
81dd9d3cabb1ae61e19abbd25e874dce	18	1
c3ab0a201aa6669a4add9cbffd565584	18	1
bd1749818ad9d523909e75b06e74ffb8	17	1
70575a8cb61ffb99eb1246653faca194	17	1
d824eef4449ec68a9cc28b809b353877	17	1
18481445ece46196fa6cd62b782f3ed8	17	1
57289f53648fee82e6a843a6a7b157e7	17	1
713993e57c9fe537def4e472811638ab	17	1

	Frequency	# of Samples Observed In
e946c36266ceb50e0193c9c7f08b5b51	17	1
76b250fe7f2082147de2b48d90dd079b	16	1
ccaa42dde153a8157033a6f31088e889	16	1
9efa507dacd5b14a70f023bad0f87c57	16	2
67870fb42912dcd87b232664a1f173d2	16	2
52cfada5f10419e4f6a3695dfb61bdf1	16	1
7a0e3d360ac317f1c95b22bb196a7b24	15	1
d280fd356fdf76ab1df44814b0e9bdf8	15	1
4ddd8007f515fa0383bd9849074bbea9	14	1
3975b5d434076b273706b0ccb0b581e8	14	1
447a2be0578af6180d43706381d50c5d	13	1
417062f8f06c8fa367115cf3c3a38438	13	1
cfb3eabb0cd4166fbf4708afb6233f09	13	1
e1921cb7e4080e3b518699c8a861a128	12	1
2bdb4924fdd70d688557c80ea7fe6b29	12	1
c63f01a33331ca04016aa4e597de1546	12	1
c9c83e97fe9e72d43e36bab1a106e119	12	1
a705501a0abe4b868aaa826ea86c48b8	12	1
dbe0d4f2c5ac4d10320b9f8d0e38ea5b	12	1
752bbb762d73090b12a2416a825dd332	11	1
c8e73d7f93a10554c557a1b31159043b	11	1
077db4f17f6e777407f21fd32e12168b	11	1
bdb8d9da33cdc819f2f0172d36328d82	11	1

	Frequency	# of Samples Observed In
443e1d6cdd34d009482d26547c6cbfee	11	1
1d23abb636c036d453713406630d474b	11	1
aa8e87e9eafc67991279936fc2dc4cf1	10	1
2032956e8e4b02f45105cd3f398a74d4	10	1
18bf1231b8256ca54ef2cac512ddc63e	10	1
7ab22d1f6a60b32aadc63dba790aef30	10	1
73992e0423f0f8bf2bd6955682e7b500	10	1
af9bec14ef4a9443b1a24d36e3012308	10	1
de0269f5b94562cb84d16ec6aa1ca941	10	1
7bd40d2d447ea0e287f379d54b97d81d	10	1
dcf5fc016dbb665c322dd526b73935d9	10	1
d4f7d1295b56600a6a750874eadbc9d9	9	1
88eb24c01d1ebfb092e8f23d0a95f4cf	9	1
7aa4ee0cc11aaf77beed6d156523b99b	9	1
9ac453bd29dd646b33a1b0c8e1195db2	9	1
466681ddefe252851d7c62c1526fae6c	9	1
edbf50cb3a56598c9c681981b7813ffd	9	1
5ca83d5a217afc8ecbb0b19b223f0b99	9	1
d0e9a55189967f0674caec81458812d1	9	1
b4212576a04f219a8611792e2580e283	9	1
e2a2f02e47dd7d6282e87918f2b96152	8	1
a658afd6439e256a7984f82023ad9116	8	1
06ea9e331fcbe9bea1c902097af83235	8	1

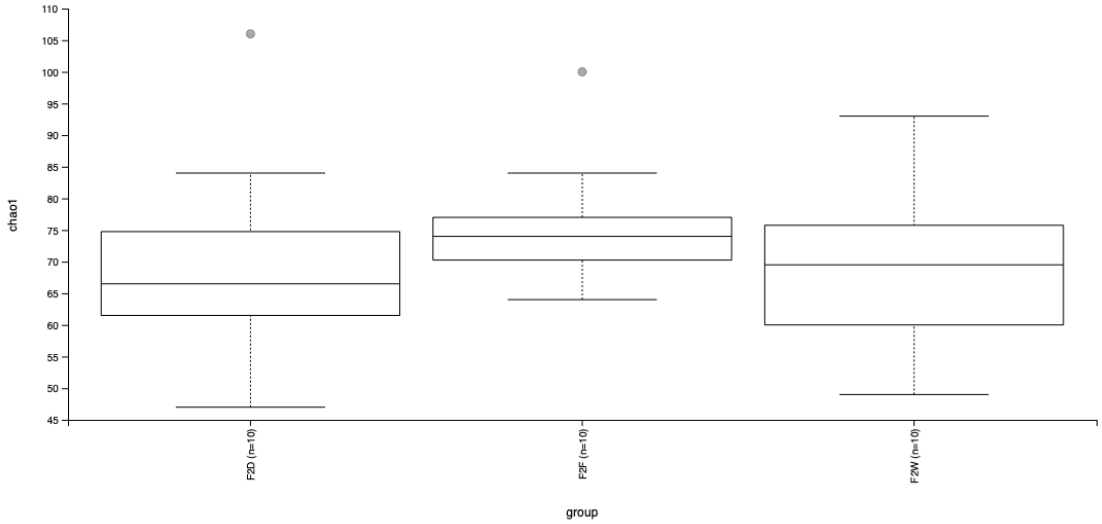
	Frequency	# of Samples Observed In
e56275709ef8d009370ade472970cb73	7	1
0e646ce871146a86cb6c1102d10cc1ee	7	1
60f4177fb9bc9b70f14621a08fd10e6f	7	1
16583df801aa25cd7874b16eebea0a8f	7	1
2f4679a1b75b6d91a308b5c1a404521e	7	1
aa06b573f0a6b4fbc36ce94a7268d4f0	6	1
477cde4762ee0adf2eff820ebd98c3e8	6	1
608556ecff0e8838f76a41d72b8ac9d1	6	1
58d877a60ea863a45a20f4f49174df3a	6	1
208ded2edcc52ad570f37df372f9d2d3	6	1
86989d0d580fc80c8712c110fa60b39b	6	1
68e6c50907f4582c51d34edc8461acfc	6	1
39b688211bf3a44c46f06cbcbfb9ef5a	6	1
c3fbc69e12d855426f9c6c7c8ec684dd	6	1
cb9e35554c906453c167be8301255af3	6	1
0257462ea3f63df940407090e5cf86c6	6	1
8101c01c85ade81233d4196036bbf5ba	5	1
aa134ce70df1c9fa3f9b7e7be8a4eb4c	5	1
63fedf988bf99809a44ac77c09dc04b0	5	1
b487b550c323044942e04979bb5ff91c	5	1
bc5ad9e6c93b66388a925794145a46c2	4	1
647239b3e7c30dd8cfdd9d44974d42e1	4	1
4f24db4c29b44008c54158c17015a732	4	1

	Frequency	# of Samples Observed In
7d7f7a98dc58d8485ba2607226168614	4	1
1740c35f8a62dee2e94ddb090c80e1d5	4	1
1959d6fcafd2a4326797a4b2360fe819	4	1
7c40ac18c0cda71804963d069531227e	3	1
133dc3a639631b8305842da7bb1db221	3	1
130b5c5d84dd19d822d5b16e09fe694f	3	1
04ec7fa83aedf3ef2d3bb8a3e0655d17	3	1
3b459dfb825eb508c8b43ad41447604b	3	1
336c0e09827717afff030223af32c8b9	3	1
c7be18139e645dbf8235991f7bae85c8	3	1
d2164aa481aba405008d30648c2bd324	3	1
381399920a70388dc32db7756e8af804	3	1
308b57dc29911c7a62bca5a77ac05b0d	2	1
a42c07abe9fef9e738c2002f5dfd595f	2	1
789b4e59738f78b5327f1efe4c3b3f34	2	1
f37661c842c974a8881c8d96ac8e2619	2	1

2. α-diversity

2.1. Chao1 Index

Boxplots



Kruskal-Wallis (all groups)

	Result
H	2.40901182244033
p-value	0.2998401105546

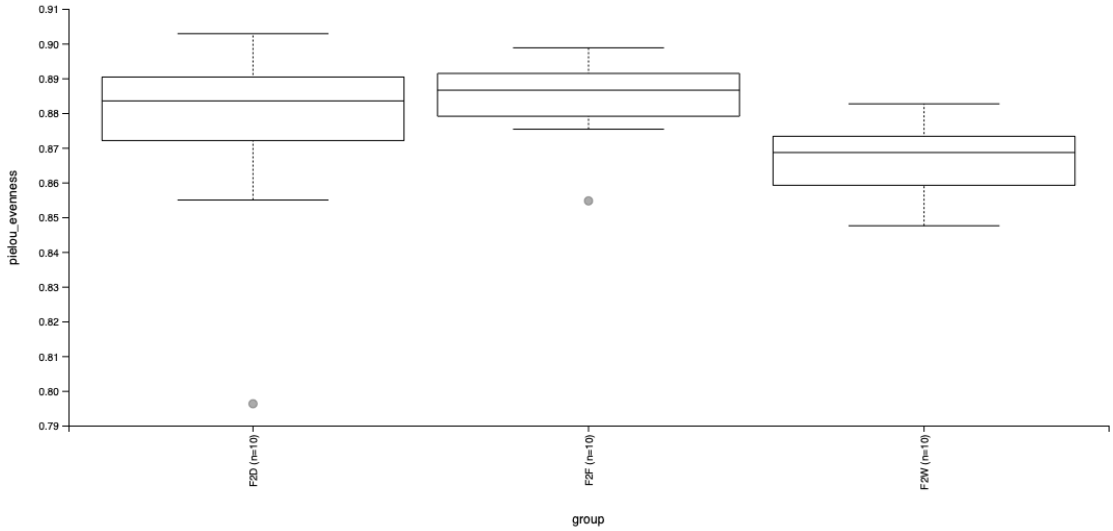
Kruskal-Wallis (pairwise)

		H	p-value	q-value*
Group 1	Group 2			
F2D (n=10)	F2F (n=10)	1.961614	0.161341	0.317150
	F2W (n=10)	0.035795	0.849939	0.849939
F2F (n=10)	F2W (n=10)	1.561585	0.211433	0.317150

**q-value: p-value with a Benjamini & Hochberg correction*

2.2. Pielou's evenness index

Boxplots



Kruskal-Wallis (all groups)

	Result
H	7.09161290322579
p-value	0.0288453507475363

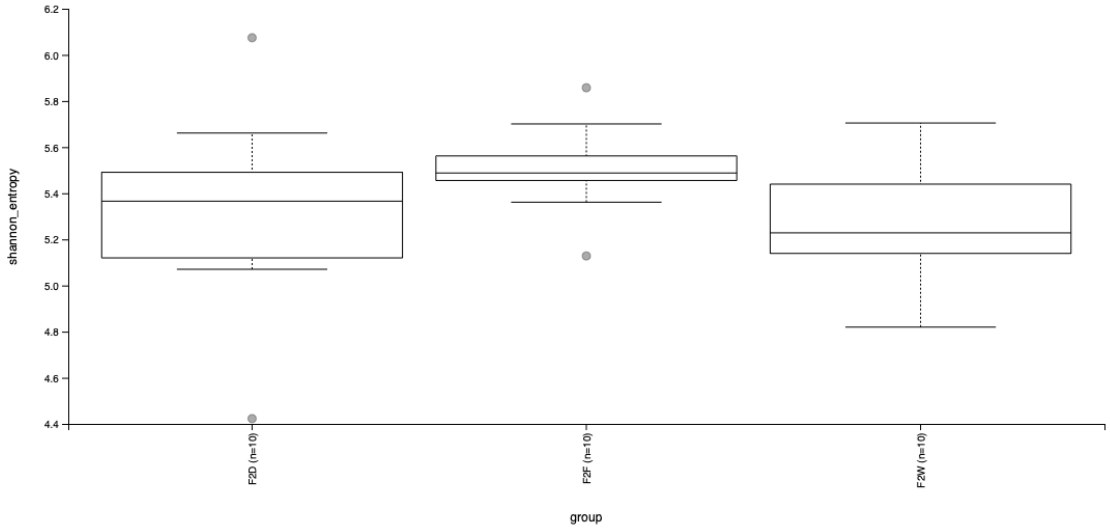
Kruskal-Wallis (pairwise)

		H	p-value	q-value*
Group 1	Group 2			
F2D (n=10)	F2F (n=10)	0.462857	0.496292	0.496292
	F2W (n=10)	3.022857	0.082099	0.123148
F2F (n=10)	F2W (n=10)	7.000000	0.008151	0.024453

**q-value: p-value with a Benjamini & Hochberg correction*

2.3. Shanon index

Boxplots



Kruskal-Wallis (all groups)

	Result
H	4.35354838709677
p-value	0.113406769517536

Kruskal-Wallis (pairwise)

		H	p-value	q-value*
Group 1	Group 2			
F2D (n=10)	F2F (n=10)	2.285714	0.130570	0.195855
	F2W (n=10)	0.280000	0.596701	0.596701
F2F (n=10)	F2W (n=10)	3.862857	0.049366	0.148099

**q-value: p-value with a Benjamini & Hochberg correction*

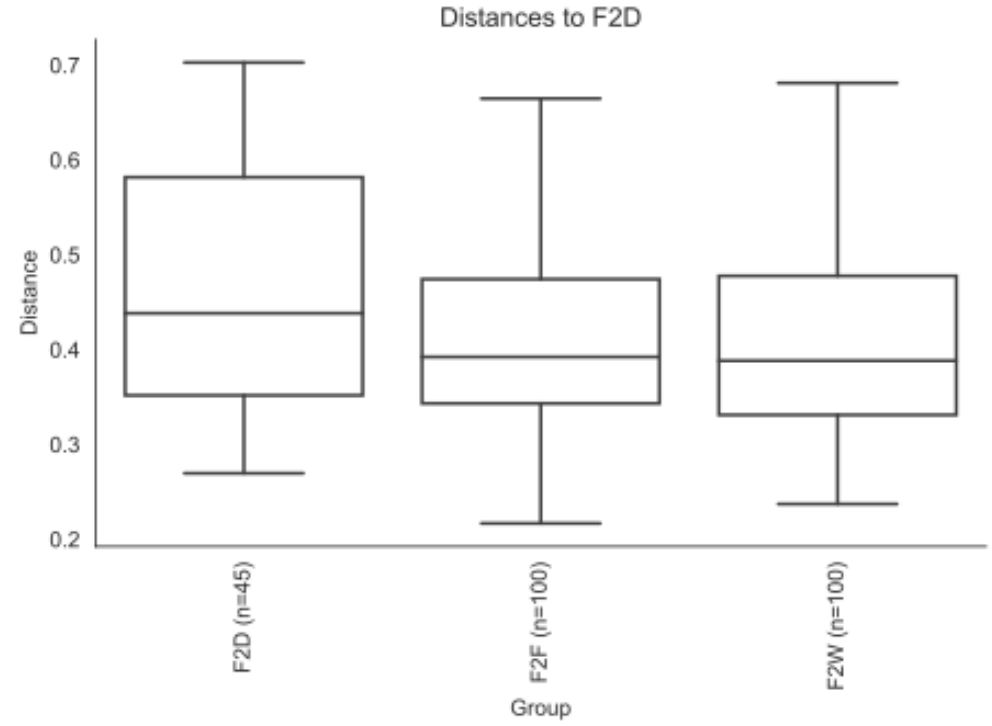
3. β-diversity

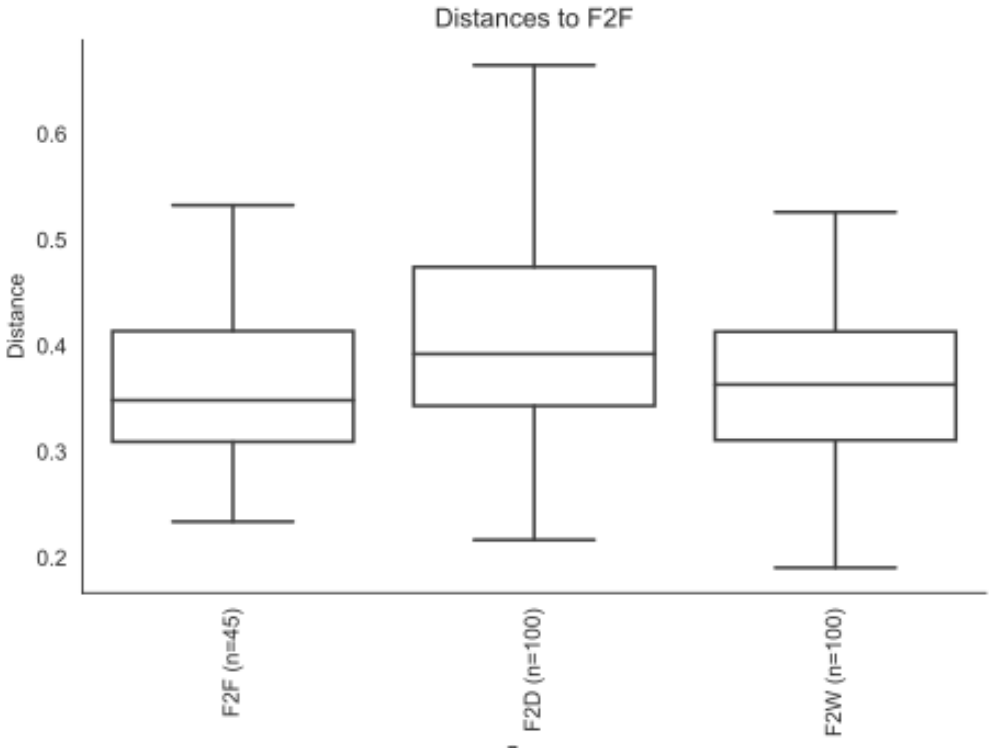
3.1. unweighted UniFrac distance

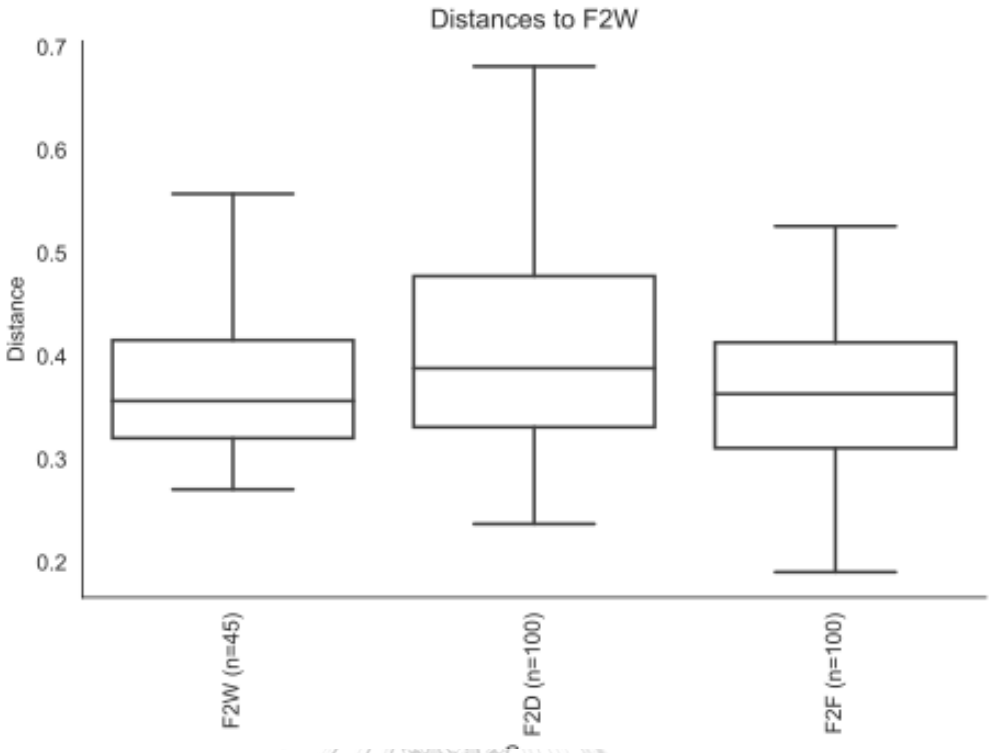
Overview

	ANOSIM results
method name	ANOSIM
test statistic name	R
sample size	30
number of groups	3
test statistic	0.065481
p-value	0.048
number of permutations	999

Group significance plots


Distances to F2D
Distance
0.7
0.6
0.5
0.4
0.3
0.2
F2D (n=45)
F2F (n=100)
F2W (n=100)
Group


Distances to F2F
Distance
0.6
0.5
0.4
0.3
0.2
F2F (n=45)
F2D (n=100)
F2W (n=100)



Pairwise anosim results

		Sample size	Permutations	R	p-value	q-value*
Group 1	Group 2					
F2D	F2F	20	999	0.124000	0.014	0.042
	F2W	20	999	0.012000	0.344	0.344
F2F	F2W	20	999	0.055778	0.152	0.228

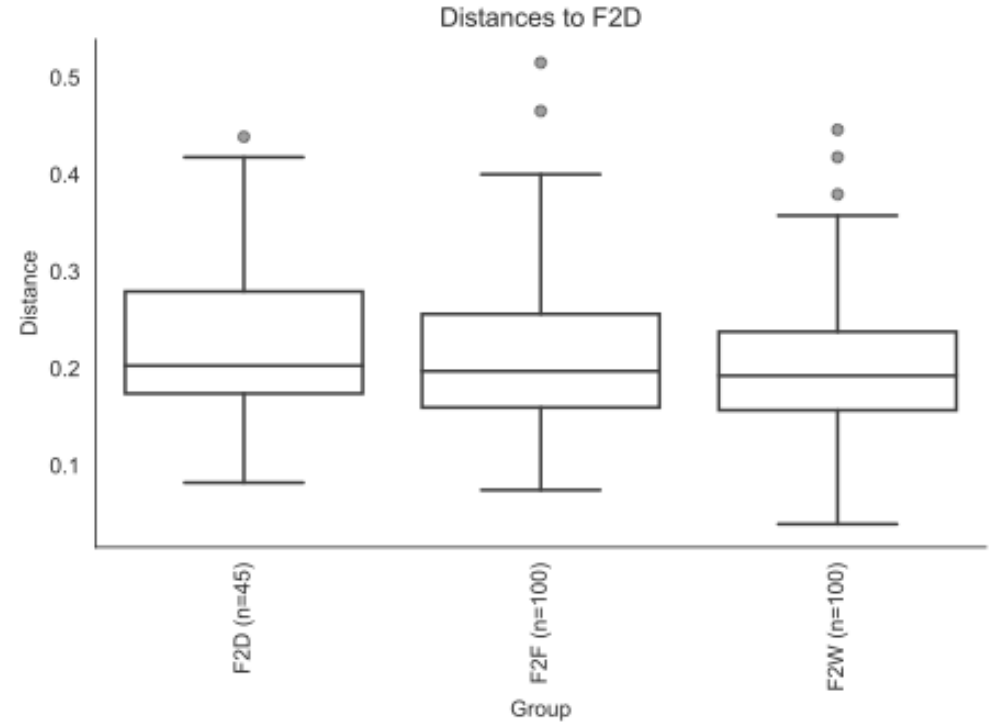
**q-value: p-value with a Benjamini & Hochberg correction*

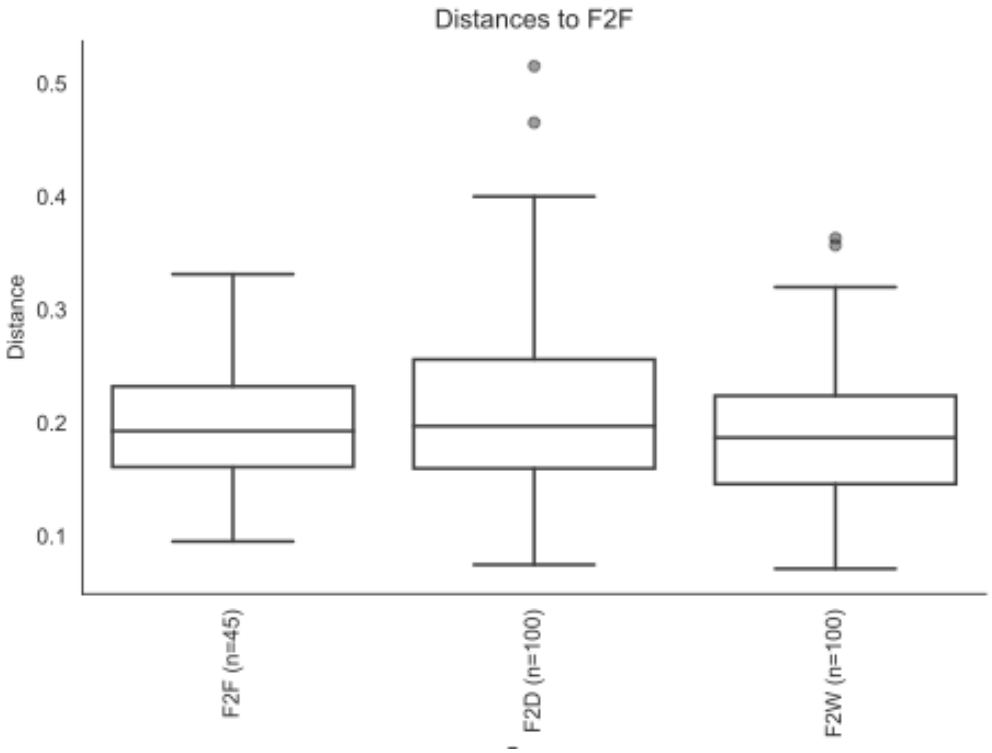
3.2. weighted UniFrac distance

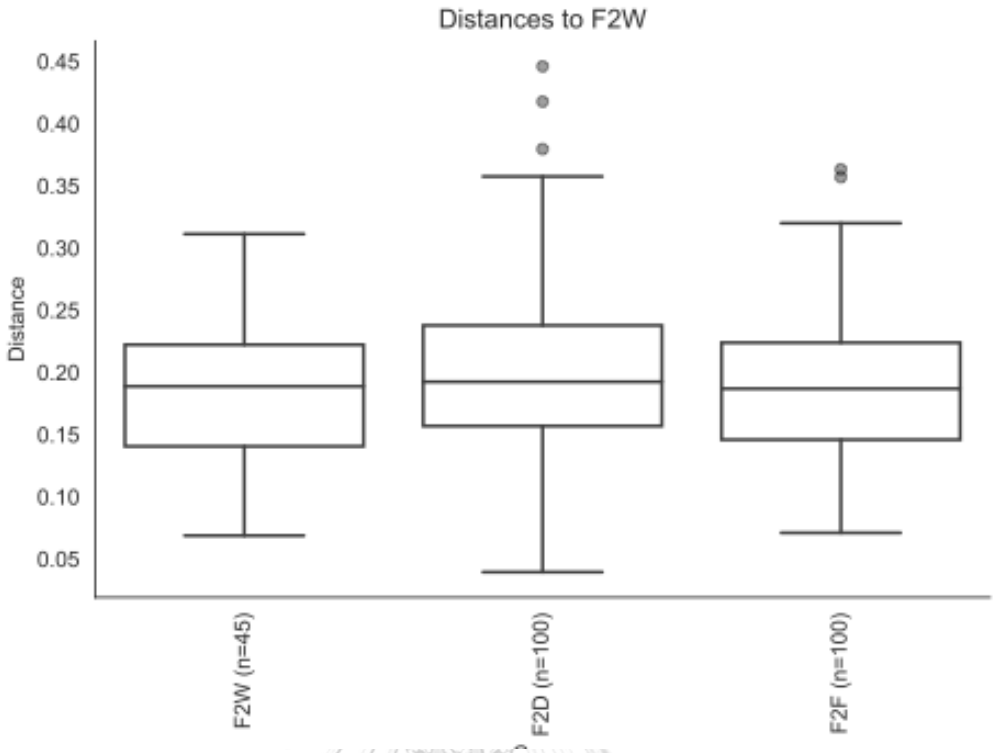
Overview

	ANOSIM results
method name	ANOSIM
test statistic name	R
sample size	30
number of groups	3
test statistic	-0.017778
p-value	0.583
number of permutations	999

Group significance plots







Pairwise anosim results

		Sample size	Permutations	R	p-value	q-value*
Group 1	Group 2					
F2D	F2F	20	999	-0.018889	0.553	0.553
	F2W	20	999	-0.023111	0.552	0.553
F2F	F2W	20	999	-0.019111	0.526	0.553

**q-value: p-value with a Benjamini & Hochberg correction*

APPENDIX B

Results from LEfSe

LDA Effect Size (LEfSe)

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Alloprevotella__	1.74440591437			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Bacillales_f__Gemellaceae_g__Gemella_s__haemolysans	4.51972723813			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Schaalia_s__lingnae__Not_Validly_Published_	3.34547542045			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Oribacterium_s__parvum	3.29346085957			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pseudomonadales_f__Moraxellaceae_g__Moraxella_s__osloensis	2.02868703366			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Alloprevotella_s__sp__HMT_914	3.81458312814			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1_____	1.74611032159			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__shahii	1.93747818195			-
k__Bacteria_p__Proteobacteria_c__Alphaproteobacteria_o__Sphingomonadales_f__Sphingomonadaceae_____	2.25399631539			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Capnocytophaga_s__gingivalis	1.59426955388			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Actinomyces_s__graevenitzii	3.26079258853			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Porphyromonadaceae_g__Porphyromonas_s__sp__HMT_930	3.56290695294			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Fusobacteriaceae_g__Fusobacterium_s__nucleatum_subsp__vincentii	0.0			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus_s__paraphrohaemolyticus	3.43640316448			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Peptostreptococcaceae__XI__g__Peptostreptococcus_s__stomatis	2.05565507522			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Kingella__	2.93338959915			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__melaninogenica	**3.88239880062**	**remained CariesFree**	**3.51754532769**	**0.0446277077963**
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__sp__HMT_917	3.62061102086			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Enterobacterales_f__Enterobacteriaceae_____	3.15837692586			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Corynebacteriales_f__Corynebacteriaceae_g__Corynebacterium_s__matruchotii	3.66672925264			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Bergeyella_s__sp__HMT_931	3.84159958755			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Kingella_s__sp__HMT_012	2.62528169546			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Aggregatibacter_s__sp__HMT_513	3.77872437642			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__histicola	3.17303461034			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Schaalia_s__sp__HMT_180	2.8646663887			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella___	4.29677733373			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV_____	3.38327725974			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Fusobacteriaceae_g__Fusobacterium_s__sp__HMT_248	3.23096026709			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Sneathia_s__amnii__Not_Validly_Published_	2.65673555342			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__sp__HMT_472	2.5775226571			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia___	4.65778710396			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Burkholderiales_f__Burkholderiaceae_g__Lautropia_s__mirabilis	3.77911517226			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pseudomonadales_f__Moraxellaceae_g__Acinetobacter_s__baumannii	3.9435588512			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Ruminococcaceae_g__Ruminococcaceae__G_2__s__bacterium_HMT_085	2.38597444698			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Actinomyces_s__sp__HMT_170	2.52569359905			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseria_s__flavescens	3.06411320819			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Eikenella_s__corrodens	0.0			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Propionibacteriales_f__Propionibacteriaceae_g__Arachnia_s__propionica	2.92869165204			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Fusobacteriaceae_g__Fusobacterium___	4.09750904571			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Peptostreptococcaceae__XI__g__Mogibacterium___	2.22453608613			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Enterobacterales_____	2.12210871882			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_____	3.1290643272			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Micrococcaceae_g__Rothia_s__mucilaginosa	4.20457785946			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseria___	4.76315668126			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Aggregatibacter_s__paraphrophilus	0.0			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pseudomonadales_f__Moraxellaceae_g__Moraxella___	3.04219823374			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Bacillales_f__Staphylococcaceae_g__Staphylococcus___	2.35502289459			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Fusobacteriaceae_g__Fusobacterium_s__periodonticum	4.2975002849			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Aggregatibacter_s__aphrophilus	2.00465735811			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella___	3.69450428801			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Cardiobacteriales_f__Cardiobacteriaceae_g__Cardiobacterium_s__hominis	2.55269364971			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__sp__HMT_942	3.92170441671			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Aerococcaceae_g__Abiotrophia_s__defectiva	3.18254126072			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Porphyromonadaceae_g__Porphyromonas_s__pasteri	4.65670352075			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Capnocytophaga___	3.26938791948			-
k__Bacteria_p__Actinobacteria_c__Coriobacteriia_o__Coriobacteriales_f__Coriobacteriaceae_g__Atopobium_s__parvulum	2.93280092862			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Aggregatibacter___	3.2785163014			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Bacteroidaceae_g__Bacteroides_s__zoogleoformans	1.10786827971			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_225	3.17973028339			-
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1__g__Saccharibacteria__TM7___G_1____	3.25558373532			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__oulorum	1.79452418804			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_221	3.51105132611			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_392	3.25068720753			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Capnocytophaga_s__leadbetteri	3.56653692902			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__shahii	2.30300021192			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Capnocytophaga_s__sputigena	3.87854658175			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus_s__sp__HMT_036	4.55175375519			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Megasphaera_s__micronuciformis	3.06481522009			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Selenomonadales_f__Selenomonadaceae_g__Selenomonas_s__noxia	2.7232166153			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Kingella_s__oralis	3.32720246387			-
k__Bacteria_p__Proteobacteria_c__Epsilonproteobacteria_o__Campylobacterales_f__Campylobacteraceae_g__Campylobacter___	2.99621788736			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__salivae	2.79140247705			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1__g__Saccharibacteria__TM7___G_6__s__bacterium_HMT_870	3.79739709729			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus_s__parainfluenzae	4.61215578308			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Actinomyces___	2.27096316351			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Carnobacteriaceae_g__Granulicatella_s__adiacens	4.13325203433			-
k__Bacteria_p__Spirochaetes_c__Spirochaetia_o__Spirochaetales_f__Spirochaetaceae_g__Treponema_s__lecithinolyticum	0.0			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1__g__Saccharibacteria__TM7___G_1__s__bacterium_HMT_347	2.96139057715			-
k__Bacteria_p__Absconditabacteria__SR1__c__Absconditabacteria__SR1___C_1__o__Absconditabacteria__SR1___O_1__f__Absconditabacteria__SR1___F_1__g__Absconditabacteria__SR1___G_1__s__bacterium_HMT_875	3.13300603336			-
k__Bacteria_p__Absconditabacteria__SR1__c__Absconditabacteria__SR1___C_1__o__Absconditabacteria__SR1___O_1__f__Absconditabacteria__SR1___F_1__g__Absconditabacteria__SR1___G_1__s__bacterium_HMT_874	2.9990854239			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseria_s__oralis	3.38963038323			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Gracilibacteria__GN02__c__Gracilibacteria__GN02___C_1__o__Gracilibacteria__GN02___O_1__f__Gracilibacteria__GN02___F_1__g__Gracilibacteria__GN02___G_1__s__bacterium_HMT_872	2.48712412194			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_212	0.0			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_215	**3.66321481977**	**remained CariesFree**	**3.74447320547**	**0.0359991932769**
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus___	4.43049402431			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Peptoniphilaceae_g__Parvimonas_s__micra	0.0			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_218	0.0			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Porphyromonadaceae_g__Tannerella_s__sp__HMT_286	1.67745912474			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Micrococcaceae_g__Rothia_s__aeria	3.53938912718			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__atypica	3.9570151157			-
k__Bacteria_p__Proteobacteria_c__Epsilonproteobacteria_o__Campylobacterales_f__Campylobacteraceae_g__Campylobacter_s__gracilis	1.42218661964			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__goodfellowii	4.50419988003			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Stomatobaculum_s__longum	0.0			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__hongkongensis	2.26898906687			-
k__Bacteria_p__Cyanobacteria_c__Oscillatoriophycideae_o__Oscillatoriophycideae_f__Oscillatoriales_g__Arthrospira_s__platensis	2.40244192362			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Lachnoanaerobaculum___	3.16048301469			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Streptococcaceae_g__Streptococcus_s__salivarius	4.39284756872			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Bergeyella_s__sp__HMT_322	3.40817248706			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Actinomyces_s__oris	3.2960882175			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Streptococcaceae_g__Streptococcus___	5.50863617092			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Bergeyella_s__sp__HMT_206	3.96219702504			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Selenomonadales_f__Selenomonadaceae_g__Selenomonas_s__artemidis	1.89529954955			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Peptococcaceae_g__Peptococcus_s__sp__HMT_167	0.0			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__rogosae	3.19185871876			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__oris	0.0			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Bacillales_f__Gemellaceae_g__Gemella___	3.32817151782			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__sp__HMT_780	4.75205873579			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Lachnospiraceae__G_2__s__bacterium_HMT_088	1.02198696152			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Actinomyces_s__naeslundii	0.0			-
k__Bacteria_p__Proteobacteria_____	0.0			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__dispar	2.12975679573			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Porphyromonadaceae_g__Tannerella___	0.992209562554			-
k__Bacteria_p__Bacteroidetes_c__Flavobacteriia_o__Flavobacteriales_f__Flavobacteriaceae_g__Capnocytophaga_s__sp__HMT_332	2.56816675132			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Fusobacteriaceae_g__Fusobacterium_s__sp__HMT_203	3.00645000167			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseriaceae__G_1__s__bacterium_HMT_174	1.67213971015			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Lachnoanaerobaculum_s__orale	2.37735564318			-
k__Bacteria_p__Proteobacteria_c__Alphaproteobacteria_o__Rhizobiales_f__Brucellaceae_g__Ochrobactrum_s__anthropi	2.16222594203			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Enterobacterales_f__Enterobacteriaceae_g__Enterobacter___	3.43352075028			-
k__Bacteria_p__Gracilibacteria__GN02__c__Gracilibacteria__GN02___C_2__o__Gracilibacteria__GN02___O_2__f__Gracilibacteria__GN02___F_2__g__Gracilibacteria__GN02___G_2__s__bacterium_HMT_873	2.71098796024			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Sneathia___	0.0			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus_s__sp__HMT_908	3.89418177591			-
k__Bacteria_____	1.21417208			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Veillonellales_f__Veillonellaceae_g__Veillonella_s__parvula	3.17971538854			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Enterobacterales_f__Enterobacteriaceae_g__Escherichia_s__coli	2.23789186912			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseria_s__lactamica	0.0			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Actinomycetales_f__Actinomycetaceae_g__Schaalia___	3.60835327436			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Ruminococcaceae_____	1.14566358902			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Bacillales_f__Gemellaceae_g__Gemella_s__morbillorum	2.79612122506			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Bacillales_f__Staphylococcaceae_g__Staphylococcus_s__aureus	0.0			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Ruminococcaceae_g__Ruminococcaceae__G_1__s__bacterium_HMT_075	3.91264911949			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Lactobacillaceae_g__Lactobacillus_s__salivarius	2.4979632571			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Lactobacillaceae_g__Lactobacillus_s__gasseri	2.3840199048			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Haemophilus_s__sputorum	3.81894848595			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__nanceiensis	**3.8360676946**	**remained CariesFree**	**3.68751960309**	**0.0211010572811**
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Burkholderiales_f__Comamonadaceae_g__Ottowia_s__sp__HMT_894	2.48357125639			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1__g__Saccharibacteria__TM7___G_3__s__bacterium_HMT_351	2.41817829483			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Prevotella_s__sp__HMT_306	0.0			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Lachnospiraceae__G_3__s__bacterium_HMT_100	1.89529954955			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Streptococcaceae_g__Streptococcus_s__sanguinis	2.98328800328			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_o__Pasteurellales_f__Pasteurellaceae_g__Aggregatibacter_s__actinomycetemcomitans	0.0			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Catonella_s__morbi	2.994418954			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Alloprevotella_s__sp__HMT_308	2.92124815637			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Corynebacteriales_f__Corynebacteriaceae_g__Corynebacterium_s__durum	2.67262503681			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Streptococcaceae_g__Streptococcus_s__parasanguinis_clade_411	4.17592995699			-
k__Bacteria_p__Proteobacteria_c__Betaproteobacteria_o__Neisseriales_f__Neisseriaceae_g__Neisseria_s__elongata	2.95947152477			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Porphyromonadaceae_g__Porphyromonas___	3.38650076318			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Saccharibacteria__TM7__c__Saccharibacteria__TM7__C_1__o__Saccharibacteria__TM7__O_1__f__Saccharibacteria__TM7___F_1__g__Saccharibacteria__TM7___G_1__s__bacterium_HMT_352	1.21417208			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Prevotellaceae_g__Alloprevotella_s__sp__HMT_473	5.13487830683			-
k__Bacteria_p__Bacteroidetes_c__Bacteroidia_o__Bacteroidales_f__Bacteroidales__F_2__g__Bacteroidales__G_2__s__bacterium_HMT_274	1.69117956689			-
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Selenomonadales_f__Selenomonadaceae_g__Selenomonas___	2.88516952814			-
k__Bacteria_p__Firmicutes_c__Clostridia_o__Clostridiales_f__Lachnospiraceae__XIV__g__Catonella_s__sp__HMT_451	0.0			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Propionibacteriales_f__Propionibacteriaceae_g__Cutibacterium_s__acnes	2.13589700331			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_498	2.26327633484			-
k__Bacteria_p__Fusobacteria_c__Fusobacteriia_o__Fusobacteriales_f__Leptotrichiaceae_g__Leptotrichia_s__sp__HMT_417	3.0487167801			-
k__Bacteria_p__Proteobacteria_c__Gammaproteobacteria_____	0.0			-
k__Bacteria_p__Actinobacteria_c__Actinobacteria_o__Micrococcales_f__Brevibacteriaceae_g__Brevibacterium_s__paucivorans	2.19894874906			-
k__Bacteria_p__Firmicutes_c__Bacilli_o__Lactobacillales_f__Carnobacteriaceae_g__Granulicatella_s__elegans	4.56297609903			-

	The highest mean among all the classes (log10)	The class with the highest mean	LDA score (log10)	p-value*
k__Bacteria_p__Firmicutes_c__Negativicutes_o__Selenomonadales_f__Selenomonadaceae_g__Mitsuokella_s__sp__HMT_521	1.72128673787			-
k__Bacteria_p__Proteobacteria_c__Epsilonproteobacteria_o__Campylobacterales_f__Campylobacteraceae_g__Campylobacter_s__concisus	**3.7600567867**	**remained CariesFree**	**3.46201126053**	**0.0446277077963**
k__Bacteria_p__Proteobacteria_c__Epsilonproteobacteria_o__Campylobacterales_f__Campylobacteraceae_g__Campylobacter_s__rectus	3.10726506523			-

The class with the highest mean, LDA score (log10), and p-value are shown if the feature is discriminative.

**p-value: Wilcoxon test*