How do the changes in the environment alter the oral mycobiome of adolescents

Submitted to the University College London in partial fulfillment of the requirement for the Degree of Clinical Doctorate in Paediatric Dentistry

Eastman Dental Institute
University College London
47- 49 Huntley Street
London
WC1E

Submitted by:
Tengku Maryam Fatimah Tengku Ab Malek
BDS (Adelaide)

Table of Contents

Declaration

I hereby declare that the work that is being presented in this thesis titled "How do the changes in the environment alter the oral mycobiome of adolescents", in partial fulfillment of the requirement for the Degree of Clinical Doctorate in Paediatric Dentistry has been carried out by myself and is entirely my own work, unless stated in references and acknowledgements. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

.

Tengku Maryam Fatimah Tengku Ab Malek
24/7/22

Abstract

INTRODUCTION:

Oral microbiota residing in oral cavity play important roles in health and disease and its components are affected by various factors, including host and external environments. Research on oral microbiota often focused on bacteria, with fungi trailing behind, leading it to be understudied and poorly characterized in comparison. Minimal studies on oral mycobiome led to an interest in how and when oral mycobiome is established and stabilised in early life and how the environment would affect the composition and diversity of oral mycobiome.

AIMS & OBJECTIVES:

The aim of this study is to analyse the oral fungal microbiota (mycobiome) of adolescents, its diversity and stability over time, and how the environment might influence its composition. A systematic review was also conducted to explore how environmental and host factors are related to diversity of human mycobiome in children, specifically in oral cavity and gut.

DESIGN:

Collection of saliva samples from 17 subjects were done at nine different time points over the span of ten months. Subjects were students at an independent boarding school, with age ranging from 11 to 16 years. They were grouped as boarders and non-boarders. 135 saliva samples were collected and processed for analysis of ITS sequences.

RESULTS:

The boarders group showed a higher alpha diversity of oral mycobiome in comparison to those in the non-boarders group but could be attributed to more subjects in the boarders group. In terms of beta diversity, there was no statistically significant dissimilarity between both groups. The low number of participants in total (n=17) indicated data was more indicative rather than definitive. Ascomycota was the most abundant phylum in both groups. No *Candida* species was found in the taxonomic analysis of both groups, which suggested potential issues during sample collection, processing, and analysis. Systematic review carried out identified caries, age, obesity, diet, Crohn's disease, antibiotics and probiotics as potential host and environmental factors related to oral/gut mycobiome diversity in children. All included studies were also deemed to be of fair quality with high risk of bias.

CONCLUSION:

This study served as a pilot investigation into the role of environmental factors on the oral mycobiome of adolescents and highlighted the need for further research with considerations of collection of saliva samples over a longer period, with a larger and more evenly distributed subjects. The systematic review revealed other potential host and environmental factors worth exploring to determine their association and potential effects on oral/gut mycobiome.

Impact Statement

As a newborn, the oral cavity is sterile but with contact with the outside world, colonisation and establishment of microbiota will start to take place. The oral microbiota then continues to reform throughout life, depending on host and environmental factors. There is still limited knowledge and understanding of the oral fungal microbiota, specifically how it is established early in life, how it evolves in composition and diversity with time, and its role in health and disease.

To our knowledge, this is the first study in which the oral mycobiome of adolescents with different living environments (boarders and non-boarders) were investigated over a period of ten months. The participants involved in this study gave us the opportunity to investigate whether a shared living environment had any influence on the composition and diversity of oral fungal microbiota.

This current study has its limitations, such as that the length of the study was reduced to less than one year from the original plan of at least two years, the small sample size, which was heavily skewed towards one group, lack of proper planning of optimal sampling times and difficulty in sequencing and analysis of data, to name a few. The lack of *Candida* species, which is usually ubiquitous was a surprise find as well, which indicated some issues during the different steps in the study. However, the study served its purpose as a pilot investigation into this area of interest and where specific areas of investigations and analysis of data can be improved to achieve substantial results. The data generated from the results of the study are also indicative of the possible results that can be achieved in future studies.

This study also implicated the need for and importance of a more streamlined approach, or perhaps a standardised flowchart of mycobiome analysis, from DNA

extraction to sequencing, ITS and taxonomic analysis as there are currently various approaches employed by different studies which makes it difficult to ascertain whether results generated can be compared and validated against previous studies with similar points of interest. A more organised bioinformatic analysis could also limit the potential for bias at different stages of the process.

The systematic review carried out identified the different host and environmental factors previously studied and their associations with oral or gut mycobiome, particularly in children which helped to determine which exposure factors are worth pursuing in future studies. It also highlighted the presence and diversity of specific oral/gut mycobiome in the presence of specific host or environmental factors in adolescents which would be useful as points of reference in future studies.

Overall, this study presented some indicative data which showed that a shared living environment can lead to an increase in richness and evenness of oral mycobiome and no difference in terms of beta diversity, compared to those who stayed at home. It can be considered a pilot investigation into this area of interest and helped to inform us of the limitations and how its study can be improved. It is hoped that with these types of studies, we will have a greater understanding of oral mycobiome and specifically, its role in health and disease.

Acknowledgements

The completion of this project would not be possible without the help and guidance of my supervisors, Professor Paul Ashley, Professor David Spratt, and Professor Andrew Smith. My sincerest gratitude and thanks for your patience and wisdom. I would also like to thank Dr Mehmet Davrandi and Dr Silke Rath for their generosity in sharing their expertise, particularly in answering my queries about the data analysis aspect of this project which proved to be most challenging. I would also like to thank my senior fellow postgrad Violeta Kakiora, who was involved in this project prior to me, completed the ethics approval and carried out the early aspects of the study. Her explanation of the fine details of the study was crucial to my involvement in this project.

Words cannot adequately express how thankful I am to both my parents for their prayers, unwavering support, and confidence in me over the past three years in completing this challenging yet rewarding course. They are my pillars of strength, and I am very proud to call them my parents.

Lastly, I am thankful for the friendship and support from my friends and fellow postgrads who have now felt like family. This has been a long and arduous journey, but I am grateful for the experience and glad to have made it through to the end.

List of tables

List of figures

List of abbreviations

ANOSIM	Analysis of similarity
ASV	Amplicon sequence variant
CA	Caries-free with dentine lesions
CAE	Caries-free with enamel lesions
CDHS	Children Dental Health Survey
CF	Caries-free
DFT	Decayed Filled Teeth
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
HIV	Human Immunodeficiency Virus
IBD	Inflammatory bowel disease
ICDAS	International Caries Detection and Assessment System
ITS	Internal transcribed spacer
NGS	Next-generation sequencing
NMDS	Non-metric dimensional scaling
NOS	Newcastle-Ottawa Scale
OTU	Operational taxonomic unit
PCOA	Principal coordinate analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
PICO	Population Intervention Comparison Outcomes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PUFA	Pulp Ulcer Fistula Abscess
QIIME	Quantitative Insights Into Microbial Ecology
RNA	Ribonucleic acid

Introduction

After the gut, the oral cavity has the second largest collection of microbiota in the human body with high diversity and complexity. Oral microbiota can be classified as a collection of microorganisms that can be found in the oral cavity and this includes bacteria, viruses, protozoa, archaea and fungi (Deo and Deshmukh, 2019). In the understanding of the oral microbiota, research is often targeted towards the study of bacteria. The abundance of bacteria in the oral cavity, in addition to their ease of cultivation and detection are some of the reasons for this trend.

In contrast, fungi remain to be a component of the oral microbiota that is lagging behind in terms of its investigation. This is attributed to their low number, with less than 0.1% of the oral microbiota and difficulty for their genetic material to be isolated. Furthermore, in some species of fungi, they are unable to be cultivated by the current methods (Baker et al., 2017). *Candida* species specifically has dominated most of the research done on oral fungi due to the fact that they are the most abundant in the oral cavity and responsible for several oral infections (Sharma et al., 2018). Due to this, very little is known about the role of fungi, specifically in the establishment and maintenance of a healthy oral cavity.

The oral cavity is the entry point for fungi to reach both the gastrointestinal and respiratory tracts which makes its study crucial to allow the understanding of what constitutes fungal microbiota communities in health and disease. The recognition of these patterns allow for proper diagnostic methods and technologies to be developed to identify and treat specific fungal microbiota responsible for disease (Zarco et al., 2011).

A large number fungal species (>600) are reported to infect humans and correlations between fungi and a wide range of diseases, ranging from asthma to skin conditions have been noted. They are environmental opportunists that can occupy and colonise different organs in individuals who come into contact with them. This can be in an external environment, where fungi are ever present and indoors, where fungi can be found in the air, on food and on many surfaces. In recent years, fungi has emerged as a prevalent class of human pathogens and thought to be facilitated by the increasing prevalence of immunocompromised individuals (Dupuy et al., 2014). To fully understand their direct and indirect roles on human health, a full characterisation of their presence, whether commensal or pathogenic, in heathy and diseased population, is vital.

The stability of the oral microbiota in relation to age and environment, relationship between bacteria and fungi microbiota in the oral cavity in health and disease, and whether presence of specific oral fungi are transient or part of an established population are some of the key areas worth pursuing. An increase in knowledge with regards to oral fungal microbiota in healthy individuals and those affected by host factors or environmental factors can have positive implications, not only in the future of dental practice, but also therapeutic approaches to treating pathogenic oral fungi.

1 Review of literature

1.1 Oral cavity

The oral cavity is an initial access but major gateway to the body for microbiota colonisation of the oral and gut. Because of this, it is an accessible site for assessment of the microbiota community and a potentially useful biomarker for prediction and diagnosis of oral and systemic diseases as well as monitoring its progress (Xiao et al., 2020). The multiple essential functions of the oral cavity, such as eating, communicating and defence against infection, affect the oral microbiota ecosystem. Oral hygiene practices, such as brushing with toothpaste, rinsing with mouthwash and flossing would also affect the ecological balance of the oral microbiota (Zarco et al., 2011).

On average, oral cavity typically has a temperature of 37°C which is relatively stable without much changes. This provides oral microbiota with a stable environment to thrive in. Saliva also has a steady pH of 6.5 to 7 which is favourable to a majority of oral microbiota. Microbiota are kept hydrated and saliva functions as a transport medium of nutrients to the oral microbes (Deo and Deshmukh, 2019).

There are two types of surfaces in the oral cavity where oral microbiota can colonise, which are the oral mucosal soft tissue and hard teeth surfaces. The different structures in the oral cavity such as teeth, tongue, cheeks, gingival sulcus, hard and soft palates and tonsils all provide a niche environment for the oral microbiota to colonise and flourish. Some studies have found that different oral cavity sites in the same host can have different bacteria diversity as each location would have specific conditions and nutrients that are optimal to its colonising oral microbe (Zarco et al., 2011).

1.1.1 Biofilm

Biofilm or plaque is inhabited by a group of microorganisms that adhere together to hard and soft tissues surfaces in the oral cavity. Bacterial cells within the biofilm attach to each other and extracellular polymeric substances surround them. It can begin to form on teeth over the span of a few hours in the absence of any mechanical removal by tooth brushing. If left as is, it can result in gingivitis within 10 to 21 days. The formation of biofilm is said to be an essential characteristics of oral diseases. *Streptococci* is found to be the most recognised initial biofilm coloniser which provide metabolites and signals that attract subsequent bacterial species to colonise. The subsequent microbes interact with the initial colonising oral bacteria as well as each other as the biofilm mature. The late colonisers are usually more pathogenic and are the ones that cause inflammation and lead to development of oral disease such periodontal disease (Xu & Gunsolley, 2014).

1.1.2 Saliva

Saliva is a type of body fluid and produced and secreted by three pairs of major salivary glands which are parotid, submandibular and sublingual glands as well as multitudes of minor salivary glands. After secretion of primary saliva, it is modified with serum exudates and other constituents which originated from mucosal cells, immune cells, and oral microbiota. Constituents of blood which are present in the oral cavity through gingival crevicular fluid, oral mucosa, and intraoral bleeding, are also mixed in. This forms what is called a mixed or whole saliva which comprised of complex mixture of molecules with elevated diversity. It is thought to heavily influence oral microbiota colonisation and clearance. Saliva also plays a vital role in the physical, chemical, and immunological defences in the oral cavity (Fábián et al., 2012).

Saliva forms a thin (about 0.5 to 1 µm) protective liquid layer, the acquired pellicle, on tooth surfaces. The acquired pellicles are made up of several salivary proteins which have calcium hydroxide binding properties and play a role in microbial adhesion and colonisation processes. This has a disadvantage on the tooth surfaces that it is attached to as it may lead to dental caries and periodontal disease, especially amongst those with poor oral hygiene. However, it may have an advantage as a defender and protector of tooth surfaces from being attacked by transient pathogenic oral microbiota entering the oral cavity. Some defence proteins which are correlated to the immunity within the oral cavity include antimicrobial peptides, lysozyme, α-amylase, salivary immunoglobulins, mucin, peroxidases, statherin and cystatin. Saliva also serves an essential role in the healing of mucosal lesions, ulcers and wounds in the oral cavity (Fábián et al., 2012).

Saliva sample is considered an easy, non-invasive method of oral sample collection, especially for vulnerable population such as babies and young children. It currently presents as an optimal diagnostic medium with great potential to be use as an effective diagnostic tool, specifically in predicting future oral and systemic diseases (Xiao et al., 2020).

1.2 Oral microbiota

The oral cavity harbours one of the most diverse collection of microbiota in the body. Besides bacteria, it is also made up of fungi, viruses, protozoa, and archaea (Wade, 2013). It forms an ecological community of symbiotic, commensal and pathogenic microbiota which exists in biofilm form and maintains a relationship with dynamic equilibrium within the oral cavity of its host. However, any ecological shift of this balance could result in dysbiosis, causing oral pathogens to manifest and cause oral diseases such as dental caries, periodontal disease, endodontic infections such as

periapical periodontitis, and pericoronitis. There is also an association between oral microbiota and systemic diseases such as diabetes, rheumatoid arthritis, cardiovascular disease, irritable bowel disease, cancer and premature births. (Xin et al., 2015).

1.2.1 Bacteria

Bacteria that could be found in the oral cavity are highly complex, with six billion bacteria from about 1000 species found. It has the potential to contribute to the health and psychological status of the oral cavity. Because if its continued interaction with external environment, the population of oral bacteria is a dynamic one and can change rapidly. Bacterial biofilm which are established can be found in specific locations in the oral cavity, such as tongue surfaces, buccal mucosa, gingival crevices, and on prosthesis such as dentures and orthodontic appliances. These population of bacteria are relatively stable but their population dynamics can still affected by external factors such as diet, age and underlying systemic conditions. Preference for bacterial colonisation is also dictated by environmental factors and host factors. Majority of these are commensal bacteria which means that the host and microbes have a symbiotic relationship and both share biological benefits. It plays an important role in maintaining oral and overall systemic health. However, some can have the potential to be pathogenic and are capable of causing diseases within and beyond the perimeters of the oral cavity (Parahitiyawa et al, 2010).

1.2.2 Initial colonisation and development of oral microbiota

Prenatal

The oral microbiota is a complex system not only because the oral cavity has multiple niches which are significantly different from one another but also because the oral cavity changes with time (Krom et al, 2014). Prior to the rupturing of the amniotic sac, a fetus has always been considered essentially sterile. However, recent studies has

shown that oral microbiota colonisation of intrauterine environment, specifically the amniotic fluid was found in almost 70% of pregnant women. The most prevalent microbe found was *Fusobacterium nucleatum*, a species linked to periodontal diseases. This supports the notion that periodontal disease is a risk factor for low birth weight and prematurely born babies (Sampaio & Monteiro-Silva, 2014). Another study of 12 mother-neonate pairs found that the oral microbiota of the infants were similar to the microbiota of the placenta, which suggests that infant microbiota might have a prenatal origin (Sedghi et al., 2021).

Birth and early life

The initial colonisation of oral microbiota would firstly depend on the method of delivery, whether it is vaginal or Caesarean section (C-section) (Cho & Blaser, 2012). The initial oral microbial population of the infants born through vaginal delivery will be dominated by bacterial taxa such as *Prevotella, Bacteroides* and *Saccharibacteria* (TM7) while those born through C-section were colonised by *Propionibacterium Staphylococcus, Slackia* and *Veillonella* (Gomez & Nelson, 2017). Another study found an association between mode of delivery and specific oral microbiota diversity in three to six month old babies, specifically with abundances of *Streptococcus, Fusobacterium* and *Slackia.* Higher bacterial taxa were found in infants delivery vaginally (79 species) compared to infants delivered through Caesarean section (54 species). However, only those delivered by Caesarean section had presence of *Slakia exigua* in their oral cavity (Lif Holgerson et al., 2011). More research is required to confirm the prediction of whether these would affect their oral health later on in life (Gomez & Nelson, 2017). Following birth, colonisation of the oral cavity rapidly occurs within 8 to 16 hours through vertical transmission (exposure to maternal skin), diet by oral fixation and horizontal transmission (human interactions additional to mother) by microbiota community (Sedghi et al., 2021). The oral cavity of new-borns may also be colonised by fungi, specifically *Candida* in the first 24 hours of their life and in their

first year of life, *Candida* colonisation varies between 40-82%. However, as children age, this was found to be reduced to values between 3-36% (Sampaio & Monteiro-Silva, 2014).

With eruption of the first tooth, the environment in the oral cavity will change due to presence of a new surface for adhesion, which also causes changes in the oral microbiota. This was initially predicted to cause the emergence of *Streptococcus* species such as *S. mutans* because of their preference for adhesion on teeth surfaces. However, studies have also found presence of this species in children who are edentulous which suggests that this species can also adhere to soft tissue surfaces which can act as a reservoir for these types of pathogenic oral microbiota. These findings highlight the importance of maintaining good oral practices in infants, even prior to tooth eruption (Sampaio & Monteiro-Silva, 2014). The oral microbiota changes in the oral cavity continues to evolve and adapt with eruption of more teeth, from primary dentition, to mixed dentition and finally permanent dentition.

1.2.3 Fungi

In addition to bacteria, fungi is also present in the oral cavity, estimated to comprise of about 100 species (Ghannoum et al., 2010). The study of oral fungi microbiota, in terms of their role in health and disease, has always lagged behind oral bacteria and most of the researches on oral mycobiome has predominantly focused specifically *Candida* species. Most of the other oral fungi and their role have remained largely uncharacterised. An understanding of their potential effects human health, either directly or indirectly would be beneficial but require an in depth investigation and profiling of mycobiome in both healthy populations and those with specific health issues. A search on oral fungal microbiota studies which were focused on healthy adults and children respectively, was conducted in January 2020 and a summary of the studies found is included in Table 1.

Table 1: Summary of oral fungal microbiota studies in healthy adults and children

Study	Study type	Participants	Exposure	Comparison	Method of sample collection	Outcome/s measured	Outcomes
Ghannoum et al. (2010)	Cross-sectional	Adults, n = 20, aged 21-60 years	Healthy individuals	No comparison	Oral rinse samples	Presence and types of basal oral mycobiome	74 culturable, 11 non-culturable fungi genera present. Candida species were the most frequently present (isolated from 75% of participants), followed by Cladosporium (65%), Aureobasidium, Saccharomycetales (50% for both), Aspergillus (35%), Fusarium (30%), and Cryptococcus (20%). 101 total species identified.
Dupuy et al. (2014)	Cross-sectional	Adults, n = 6, aged > 18 years	Healthy individual	No comparison	Saliva samples	Presence and types of basal oral mycobiome	Genus *Malassezia* was present in all 6 samples, with high abundance (13-96%). 5 genera were found in high frequency which were not present in study by Ghannoum et al. (2010) – *Malassezia, Irpex (0.02-4.07%), Cytospora/Valsa (0.005-.92%), Lenzites/Trametes (0.02-1.18%)* and *Sporobolomyces/Sporidiobolus (0.01-2.87%)*.
Monteiro-da-Silva et al.(2014)	Cohort	Adults, n = 40, aged 24.0 ± 2.8 years	Healthy individuals	No comparison	Oral rinse	Characteristics and stability of oral fungi over a period of time (28 weeks and 30 weeks after baseline)	Most common oral fungi isolated were Candida (67.5%), Rhodotorula (75%), Penicillium (85%), Aspergillus (75%), Cladosporium (72.5%), Trichoderma (10%), Scedosporium (7.5%), Alternaria (5%), and Rhizopus (2.5%). Oral fungi community showed high interindividual variability but consistent intraindividual stability over time.

Fechney et al. (2019)	Cross-sectional	Children, n = 17, aged 7-10 years	Caries, n = 6	No caries, n = 11	Supragingival dental plaque samples	Presence and abundance of oral mycobiome	17 fungi species were more abundant in children with no caries ($p<0.001$). Alpha and beta diversity analyses showed no differences in richness and evenness of oral mycobiome between both groups. *C. albicans* was present in all samples regardless of whether there was caries present.
O'Connell et al. (2020)	Cross-sectional	Children, n = 33, aged 2-7 years old	Caries active with enamel lesions (CAE) n = 28, caries active with dentine lesions (CA), n = 46	Caries free (CF) n = 8	Supragingival dental plaque samples (n = 82)	Presence and abundance of oral mycobiome	*C. albicans, C. dubliniensis, Nigrospora oryzae* and an unclassified *Microdichium* sp. were associated with caries. *Candida albicans* most abundant in all plaque groups. C. *dubliniensi*s is second most abundant mycobiome species and oral mycobiome community richness (alpha diversity) reduced as caries progressed.

1.2.3.1 Oral fungal microbiota in healthy adults

In terms of previous research done specifically to investigate the presence of fungi in healthy adults, three studies were found that addressed this (Ghannoum et al. (2010), Dupuy et al. (2014), Monteiro-da-Silva et al.(2014)). Ghannoum et al. (2010) found 74 culturable and 11 non-culturable fungi genera as a result of their study of the oral cavity of 20 healthy individuals. Among the genera that were culturable, 61 were represented by one species each while 13 genera comprised between 2 to 6 different species. The total number of species identified were 101. In each individual, the number of species in the oral cavity of ranged between 9 and 23. *Candida* species were the genera most frequently obtained and was isolated from 75% of the total participants. This was followed by *Cladosporium* (65%), *Aureobasidium (50%)*, *Saccharomycetales* (50%), *Aspergillus* (35%), *Fusarium* (30%), and *Cryptococcus* (20%). One downside of this study was the use of oral rinse samples which could possibly underestimate the oral fungi diversity specifically from subgingival sulcus or in periodontal pockets.

In a study done by Dupuy et al. (2014), they investigated saliva samples of 6 healthy individuals and after use of different isolation techniques and sequence based taxonomy assignments, found five genera in high frequency that were not present in the study done by Ghannoum et al. (2010). The most notable was the presence of *Malassezia* which was highly abundant and prevalent in all 6 participants and ranged from 13% to 96%. Apart from the five genera, this study confirmed the presence of nearly every oral fungal species mentioned in the study by Ghannoum et al. (2010).

Another study done by Monteiro-da-Silva et al. (2014) investigated the presence of oral fungal microbiota in 40 healthy individuals. This study was unique in comparison to the previous two studies due to the fact that a follow-up of 10 participants at two

different time periods, 28 weeks after baseline and 30 weeks after baseline was carried out as well. This was done to study the variability and stability of oral fungi microbiota, within and amongst participants over time. In the initial samples of 40 participants, the most common isolated fungi were *Candida* (67.5%), *Rhodotorula* (75%), *Penicillium* (85%), *Aspergillus* (75%), *Cladosporium* (72.5%), *Trichoderma* (10%), *Scedosporium* (7.5%), *Alternaria* (5%), and *Rhizopus* (2.5%). They found that the oral fungal microbiota were stable intra-individually over time but were varied between different individuals. All three studies previously mentioned used samples from adults, ages ranging from 21 to 60 years old.

1.2.3.2 Oral fungal microbiota in healthy children

In terms of studies investigating the presence of oral fungi microbiota in healthy children, only two were found that addressed this, both involved in characterising the oral fungal microbiota associated with healthy children and those with dental caries. A study by Fechney et al. (2019) identified the presence of oral fungi in plaque samples taken from 17 children who were aged between seven and ten years old and examined whether there were any differences between children with and without dental caries. This was the first study that used the Next Generation Sequencing (NGS) of the primary oral fungi DNA barcode, the internal transcribed spacer 2 (ITS2) region, to study fungi as part of children's oral microbiota and its contribution to dental caries. The study found over 40 fungal species in the oral microbiota of children with six core fungi that were present in children with and without caries. *Candida albicans* and an unclassified *Saccharomycetes* species were ubiquitous in all the samples, followed by *Naganishia Diffluens* (94%), *Rhodotorula mucilaginosa* (88%), *Malassezia globosa* (59%) and *Cladosporium cladosporioides* (53%). The abundance of *Candida albicans* was comparable between individuals with and

without caries which suggests that this species may not be involved in the progression of caries.

This study also discussed similarities in some of the types of core members of oral fungal microbiota between the children examined in this study and previous studies by Ghannoum et al. (2010) and Dupuy et al. (2014) that involved oral fungal microbiota of adults and all three studies identified *Candida*, *Saccharomyces, Malassezia*, and *Cladosporium* species as “core members” of oral fungal microbiota. The identification of *Rhodotorula mucilaginosa* in this study as a core member of the oral fungal microbiota was similar to a finding in a study by Monteiro-da-Silva et al. (2014) where this species was amongst the ones most frequently identified in the examination of the oral fungal microbiota in 40 adults.

An interesting find from the study was that while as a whole, the diversity of oral fungi was similar in children with and without dental caries, presence of caries affected the abundance of specific oral fungal microbiota. Children with healthy dentitions had a significantly higher abundance of 17 fungi species compared to 3 enriched species in children with caries. With this information, it might be worthwhile to investigate the relationship between the changes in oral fungal profile from a healthy dentition to a carious state, to the changes in oral bacteria composition and diversity in the development of caries (Fechney et al., 2019).

The second study found which was carried out by O'Connell et al. (2020) investigated the presence of oral fungal microbiota by collecting supragingival plaque samples from 33 children aged between 2 and 7 years with different caries status. The caries status were divided into three types, caries-free (CF), caries-active with enamel lesions (CAE), or caries-active with dentin lesions (CA). Plaque samples were collected from caries-free surfaces (PF), and from enamel (PE) and dentin (PD)

lesions. Taxonomic profiles which represented the different categorisations (CF-PF, CAE-PF, CAE-PE, CA-PF, CA-PE and CA-PD) were utilised to characterise the mycobiome and its changes through disease progression.

139 oral fungi species in total were found with *Candida* Albicans being the most abundant in all plaque groups, followed by *Candida dubliniensis, Debaryomyces* sp and *Clasdosporium exasperatum*. Severely progressed plaque communities (CA-PD) were found to be significantly different from healthy plaque communities (CF-PF). 32 taxa in total were differentially abundant across the plaque categories. C. albicans, C. *dubliniensis, Nigrospora oryzae* and an unclassified *Microdichium* sp. were associated with caries while 12 other taxa were associated with health.

Candida albicans was found to be an unreliable indicator of early childhood caries due to its presence in high frequencies in both caries free and caries-active participants. Other the other hand, *Candida dubliniensis* is found to be a potential species indicator for caries. It was found in low abundance/absence in caries free sample and a steady increase in abundance as caries progressed. Due to the morphological and biochemical similarities of *Candida dubliniensis* to *Candida albicans*, differentiating between these two species proved to be difficult. This brought up a possibility that in earlier studies, *Candida dubliniensis* could have been misidentified as *Candida albicans* and caused its potential importance in dental caries to be overlooked. Nevertheless, further studies using different analysis method are necessary to fully understand *Candida dubliniensis* and its role in dental caries progression. Another interesting find was the presence four health-associated fungal taxa which have the potential to antagonise the cariogenic *S. mutans* by xylitol production which hypothesised a possible fungal mechanism that could contribute to maintenance of dental health. Further insight into this complex relationship between

oral fungi and bacteria members of the plaque biofilm involved with caries can be beneficial (O'Connell et al., 2020).

1.2.3.3 Oral fungi collection and analysis

In terms of sample collection of oral fungi, studies that were previously done have used several different techniques. Oral rinse, saliva or supragingival and subgingival plaque samples are some of the methods used. These differences could account for the varying results found especially in terms of the DNA quantity collected and diversity of oral fungi.

In the analysis of oral fungi samples, the process has traditionally been culture-dependent techniques. However, with technological advances and knowledge that the oral microbiota is dominated by non-culturable species, more culture independent methods are now being done in studies involving oral fungal microbiota. One of the most frequently employed techniques involves high-throughput, massively parallel amplicon-based sequencing and subsequent taxonomic assignment based on publicly available reference databases. The characterisation of oral fungi encompasses sample collection, DNA extraction, PCR amplification, sequencing, data processing and statistical analysis and can be influenced at any of the steps mentioned (Rosenbaum et al., 2019).

In terms of DNA extraction methods, one study was found by Vesty et al. (2017) that evaluated four commonly used microbial DNA extraction methods (MoBio PowerSoil® DNA Isolation Kit, QIAamp® DNA Mini Kit, Zymo Bacterial/Fungal DNA Mini PrepTM and phenol: chloroform-based DNA isolation) and their impact on oral bacterial and fungi communities. The performance of the other three DNA extraction kit were found to be inferior when compared to the phenol: chloroform-based DNA

isolation method in the analysis of the oral fungal microbiota. This method was the only one of the four assessed DNA extraction methods to yield sufficient fungal sequences for analysis from the saliva replicates. This finding was identical to the results of another study that tested and evaluated eight different DNA extraction methods on bacterial and fungal microbiota (Rosenbaum et al., 2019). Vesty et al. (2017) also found that DNA extraction method had an impact on the diversity of the oral fungi which suggested that not all DNA extraction protocols are suitable for the study of oral fungal microbiota. Several reasons were mentioned by the authors that accounted for this. Firstly, due to the fact that fungi and yeast have cell walls which are harder to lyse compared to bacterial cell walls, some of the kits mentioned in the study may not be suitable for DNA extraction of fungi (Vesty et al., 2017). This was also reiterated in a study by Dupuy et al. (2014) who concluded that cell lysis methodology was likely to have a significant impact on identifying certain oral fungi species, specifically Malassezia, in saliva. Another reason that was pointed out was that because of the uncertainty of oral fungi load in healthy individuals, the sensitivity of the DNA extractions kits in the study may not be suitable (Vesty et al., 2017).

In terms of PCR amplification and sequencing, different versions were found in the literatures. In one of the earlier researches on oral microbiota that included some oral fungi profiling, a PCR-based approach using the 18S rDNA primers was carried out in the investigation of subgingival plaque in 14 HIV-infected patients. For cloning of 18S rDNA, universally conserved fungal primers, forward primer B2F (5′-ACT TTC GAT GGT AGG ATA G-3) and reverse primer B4R (5′-TGA TCG TCT TCG ATC CCC TA-3) were used to amplify around 690-base 18S rDNA from clinical samples. These primers were only able to amplify the presence of *Candida* spp. and eight divergent fungal genera including *Hansenula* spp., S. *cerevisiae*, *Cryptococcus neoformans*, *Trichosporon beigelii, Malassezia furfur, P. carinii, Aspergillus* spp., and *Penicillium* spp. A total of 306 clones with an 18S rDNA insert of the correct size of approximately

700 bases were analysed and only *C. albicans* and *S. cerevisiae* were found (Aas et al., 2007). This method was found to only provide a snapshot of the true oral fungal microbiota. Furthermore, identical host and fungal sequences at 18S rRNA gene primer binding sites make this gene an unsuitable target, as sequences obtained may be predominantly human rather than fungal-derived (Vesty et al., 2017).

A superior method mentioned in the literature that was thought to provide a more comprehensive profile of the oral fungal microbiota was the targeting of the oral fungi in the oral cavity using Universal Internal Transcribed Spacer (ITS) primers. In one of the earliest studies of oral fungi in healthy individuals by Ghannoum et al. (2010), the ITS1 region from DNA sample extracts was amplified in triplicate using primers with high specificity for Ascomycete fungi (fluorescently-labelled forward primer ITS1F (CTTGGTCATTTAGAGGAAGTAA) and unlabelled reverse primer ITS2 (GCTGCGTTCTTCATCGATGC). ITS primers have broad fungal specificity and with the use of sequence analysis of the ITS region of ribosomal genes, the study was able to characterise the “basal” oral fungal microbiota in 20 healthy individuals and identify 101 total number of species (Ghannoum et al., 2010). ITS regions also have greater sequence variability which differentiates it from the host DNA and also allows for greater taxonomic resolution. However, there is still an inadequate amount of information and understanding of the oral fungal microbiota even with the advantages of ITS sequencing. This is in most part due to a minimal amount studies being done to describe this community and even less with emphasis on how DNA extraction might affect the representation of the oral fungal microbiota (Vesty et al., 2017).

After amplification of 18S rDNA, ITS1 or ITS2 regions, sequencing of the data is either done by Sanger sequencing or Next Generation Sequencing (NGS). Next Generation Sequencing (NGS) have been used in most of the recent studies on oral fungal microbiota due to its advantage in allowing large scale sequencing to be completed

in days and sometimes hours. The main NGS technologies are 454 pyrosequencing, Applied Biosystems, Illumina, Pacific Biosciences and Oxford Nanopore (Deo and Deshmukh, 2019). Most of the studies found that investigated oral fungal microbiota have used Illumina, specifically Illumina's MiSeq platform with the exception of a study by Ghannoum et al. (2010) which used 454 pyrosequencing.

1.2.3.4 Reasons for difference in number of oral fungi found between different studies and why low amount of studies have been done on oral fungi

Several reasons accounted for the differences in number of oral fungi species found in different studies. Firstly, as mentioned previously, would be the sampling method chosen. Oral rinse enables collection of organisms from the oral mucosal environment which may have a completely different oral microbiota environment compared to those within the subgingival biofilm plaque. Another reason is the sensitivity of the detection probe. 18S rDNA probes were found to only be able to detect *Candida* and eight other fungi genera in comparison to the ITS1/ITS2 probes which could detect a wider variety of fungi. The type of sequencing method chosen could account for this as well (Ghannoum et al., 2010).

Even with the technological advances specifically in the metagenomics and microbiomics, there are several challenges in the characterisation of profiles for fungi which makes studying them particularly challenging. The binary naming and phylogenetic classifications of fungi have been one of the challenges that have plagued biologists for a while. A single organism can sometimes have multiple names based on their varied stages or the ecosystem they originated from and this is further exacerbated by genomic approaches. Failure to correctly identify fungi according to their proper species will result in a complex list of fungi community members and present a very misleading view of the actual abundance of some fungi species in

studies. The universal issue of the recognition of process-induced sequencing errors should not be excluded as well (Dupuy et al., 2014). There is also a need for well-curated databases that is comparable to the ones used for bacterial 16S rRNA gene-based studies which is currently lacking for fungi (Vesty et al., 2017)

1.2.3.5 Gaps in current literature and potential areas for research

In researching current studies available on oral fungal microbiota, a few limitations were found which would necessitate more research on this particular topic. Firstly, is the limited amount of longitudinal research done on oral fungi that compared its presence and diversity in the oral cavity over a significant period of time. There was only one study found by Monteiro-da-Silva et al. (2014) which investigated oral fungal microbiota over 30 weeks. A comparison of the presence of the types of oral fungi present over an extensive period of time would be useful in identifying whether the presence of the certain oral fungi is just a result of transient colonisation or if the oral fungi found are permanent residents of the oral microbiota.

A continuing challenge is also the identification of oral fungi present in healthy individuals. Although a few studies have addressed this, the number of samples used were low and only involved adults. Consideration of younger children as participants and increasing the number of the sample size could yield better and more reliable results. Current studies that have been done in investigating oral fungi in healthy children have always been in done in comparison to children with caries. So far, no research on oral fungal microbiota has been done specifically just targeting healthy children. A further look into this could result in more understanding regarding the role of oral fungi in healthy young individuals. This can then be expanded to compare the oral fungal microbiota of healthy individuals to those with systemic diseases to investigate any links between them. Apart from this, other variables such as

geography, different age groups, diet etc. can be investigated as potential factors affecting the oral fungal microbiota.

A wider age range in the sample of participants, from young children to adults and investigating them over a period of time can also be useful in determining when and how the oral fungal microbiota change over time and whether age plays a part in the diversity of the oral fungal microbiota.

Considering the fact that a high number of studies have been carried out with regards to the oral bacteria, not much is known about the link between oral bacteria and oral fungi in a healthy individual. With the findings found from the studies done previously that compared oral fungal microbiota between healthy children and children with caries, it could be worthwhile to study how the change relates to the change within the oral bacterial community with the development of caries.

2 Systematic Review

A systematic review of how environmental and host factors are related to diversity of human mycobiome in children, specifically in oral cavity and gut

2.1 Abstract

Objective: The primary objective of this systematic review was to determine, in oral cavity and gut of children below 18 years of age, any associations of environmental and host factors with presence and diversity of mycobiome.

Methods: MEDLINE Ovid, EMBASE Ovid and ISI Web of Science databases were searched in March 2021 and 12 articles which met the inclusion and exclusion criteria were independently reviewed. 5 cohort, 4 cross-sectional and 3 case-control studies were included and risk of bias was assessed using the Newcastle-Ottawa scale (NOS).

Results: In the 12 studies selected, the host factors identified included caries, gender, age, obesity, preterm birth and inflammatory bowel disease, specifically Crohn's' disease, while environmental factors found in studies were dietary milk intake, antibiotics and probiotics. 3 studies were specific to only the study of *Candida*. High risk of bias were found in all studies. Some studies noted differences in mycobiome composition and diversity between exposed and unexposed participants to factors such as caries, age, diet, and obesity, particularly dominance of Ascomycota and *Candida spp*. However, other studies found minimal correlation between exposure and presence and diversity of mycobiome of subjects. Meta-analysis was not possible due to heterogeny of exposures in the included studies.

Conclusion: Even with high risk of bias of included studies**,** this systematic review elucidates the different host and environmental factors which could potentially impact

oral and gut mycobiome of children and highlights the potential for further research in this understudied but promising area.

2.2 Introduction

The gut, followed by the oral cavity both hold the two largest collections of microbiota in the whole body with high diversity and complexity. (Deo and Deshmukh, 2019). The study of fungal microbiota has lagged behind that of bacteria and reasons for this include their low number (comprising of less than 0.1% in in oral microbiota), complex genetic composition which complicates isolation of their genetic material and for some species, technique sensitivity and difficulty of cultivation by current methods available (Baker et al., 2017). The low number of fungi in comparison to bacteria in culture samples also led to them being overlooked in microbiota analysis.

In researches on the oral fungal microbiota, *Candida* species havemostly been the dominant focus. This is due to the fact that they are the most abundant in the oral cavity and responsible for several common oral infections (Sharma et al., 2018). Due to this, very little is known about the role of fungi, specifically in the establishment and maintenance of a healthy oral cavity as well as their role in presence and progression of disease.

This review aimed to explore the current literature available that addressed any associations between host or environmental factors on oral or gut mycobiome of children under the age of 18. The inclusion of gut mycobiome in addition to oral mycobiome was done after a previously conducted literature review, specific to oral mycobiome in children and adults, revealed minimal available resources. Additionally, the oral cavity is an initial access but major gateway to the body for microbiota colonisation of the oral and gut. Based on its function as one of the organs in the

alimentary canal, the gut mycobiome is therefore thought to be an important area of interest in this review.

2.3 Methods

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Liberati et al., 2009).

2.3.1 Criteria for considering studies for this review

2.3.1.1 Types of studies

All human studies addressing human mycobiome in children up to 18 years of age in two specific body sites, oral cavity and gut were accepted. The type of studies included were cohort, cross-sectional, case-control, prospective and retrospective comparative cohort studies. Exposures in studies must be host factors, environmental factors or interaction between both. Animal and in vitro studies, case reports and case series, abstracts, editorials, expert opinion, letters, guidelines, protocols, seminars, reports, books or book chapters and review studies were excluded from this review.

2.3.1.2 Types of participants

Only studies which involved children and adolescents from 0 up to 18 years old that measured human mycobiome after participants were exposed to different host and environmental factors were included. Studies that included cohorts more than 18 years in age and adults were excluded.

2.3.1.3 Types of exposures

Test group

Environmental and host factors which could include lifestyle, diet, gender, age, site, underlying condition, preterm or low birth weight, underweight, obesity, use of

antibiotics, inflammatory bowel disease (IBD), immune system and response among others.

Control group

No exposure/intervention, healthy participants.

2.3.1.4 Types of outcome measures

Primary outcome

- Identification and profiling of human mycobiome.

This was measured by presence of any mycobiome, diversity of mycobiome, abundance, evenness and changes (increase or decrease) in oral cavity or gut. Both culture dependent and culture independent methods to detect presence of mycobiome will be accepted.

Secondary outcome

- Comparison of human mycobiome in test group(s) to control group
- Comparison of human mycobiome in different sites in the body

2.3.1.5 Search methods for identification of studies

Electronic searches of MEDLINE Ovid (Ovid MEDLINE®, Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations) (1946 to 22 March 2021), Embase Ovid (1980 to 22 March 2021) and ISI Web of Science (1900 to 22 March 2021) were conducted. Subject strategies were modelled on the search strategy designed for MEDLINE Ovid (Appendix 1) but revised based on each database where appropriate. No restrictions were placed on the language or date of publication when searching the electronic databases.

2.3.2 Data collection and analysis

2.3.2.1 Selection of studies

Following the electronic search, the software EndNote was used to organise the list of references retrieved during searches and to identify and remove duplicates. The titles and abstracts were then for inclusion in the review. If the titles and abstracts did not provide adequate information to judge their eligibility for inclusion in the review, full texts of the studies were acquired and assessed. Following full text assessment, the studies which fulfilled the eligibility criteria were included. Any disagreements were resolved by discussion. A manual search was also conducted by cross-referencing relevant articles. The flowchart of included studies were summarised using the PRISMA flow diagram (Moher et al., 2009).

2.3.2.2 Data extraction and management

All studies that met the inclusion criteria were included, regardless of the study quality. A specifically designed data extraction form adapted from the Cochrane Handbook of Systematic Reviews of Interventions (Higgins & Green, 2011) was used to extract information that was relevant to the objectives, exposures and outcome measures. The journal of publication or authors' name on the papers was not masked prior to paper screening or data extraction. Descriptive data were collected where available in addition to that already outlined in the data extraction form.

The following data were collected:

- First author's last name
- Country where study was carried out
- Year study started (if not available, year it was published)
- Aim of study
- Design of study and how participants are allocated into groups

- Allocation into groups by outcome or exposure?
- Ethical approval needed?
- Population description
- Methods of recruitment
- Total participants at start and end of study
- Type and duration of exposure
- How outcome was assessed
- Method used to measure outcome
- Assessment of human mycobiome in each group, such as presence, levels, and diversity
- Source of funding
- Possible conflicts of interest

A meta-analysis was not planned for this review as it was anticipated that based on the aims of the systematic reviews, the studies included would be too heterogenous in terms of the participants, intervention/exposure, comparisons and outcomes (PICO) framework, study designs and site of interest, to be merged into a useful summary.

2.3.2.3 Assessment of risk of bias in included studies

The risk of bias in the included studies were assessed using the Newcastle-Ottawa Scale (NOS) (Wells et al, 2013). NOS is a validated tool for assessing the quality of non-randomized studies. Following the NOS, each study was judged on three specific sections which are the selection of groups studied, comparability of the groups and outcomes of interest based on a star scoring system". The NOS tool for the quality assessment of the studies is included in Appendix 2. An adapted version of the Newcastle-Ottawa Quality Assessment Scale for cohort studies was used for cross-

sectional studies. For cohort and case-control studies, a maximum of 1 star can be awarded for each numbered item within the selection and outcome categories, whereas for cross-sectional studies, one numbered item in the outcome category can be awarded a maximum of 2 stars. For all three types of studies, a maximum of 2 stars can be awarded in the comparability category.

A maximum of nine stars in total can be awarded for cohort and case-control studies whereas 8 stars would be the highest score possible for cross-sectional studies. The scale for risk of bias are noted as such:

0-3 = very high risk of bias, poor quality study

4-6 = high risk of bias, fair quality study

≥7 = low risk of bias, good/high quality study

2.4 Results

2.4.1 Characteristics of included studies

In total, an initial data set of 189 were retrieved from all relevant databases. After removal of duplicates (n = 49), the numbers decreased to 140, which were then reviewed. After screening of title and abstracts, 37 records were identified and full text papers were retrieved and screened for relevancy to this review. A total of 12 studies were included in this review for qualitative analysis. A flowchart detailing this process is represented in Figure 1. Studies removed were those which involved populations above the age of 18, non-human subjects, editorials, case reports and case series and body sites other than oral and gut.

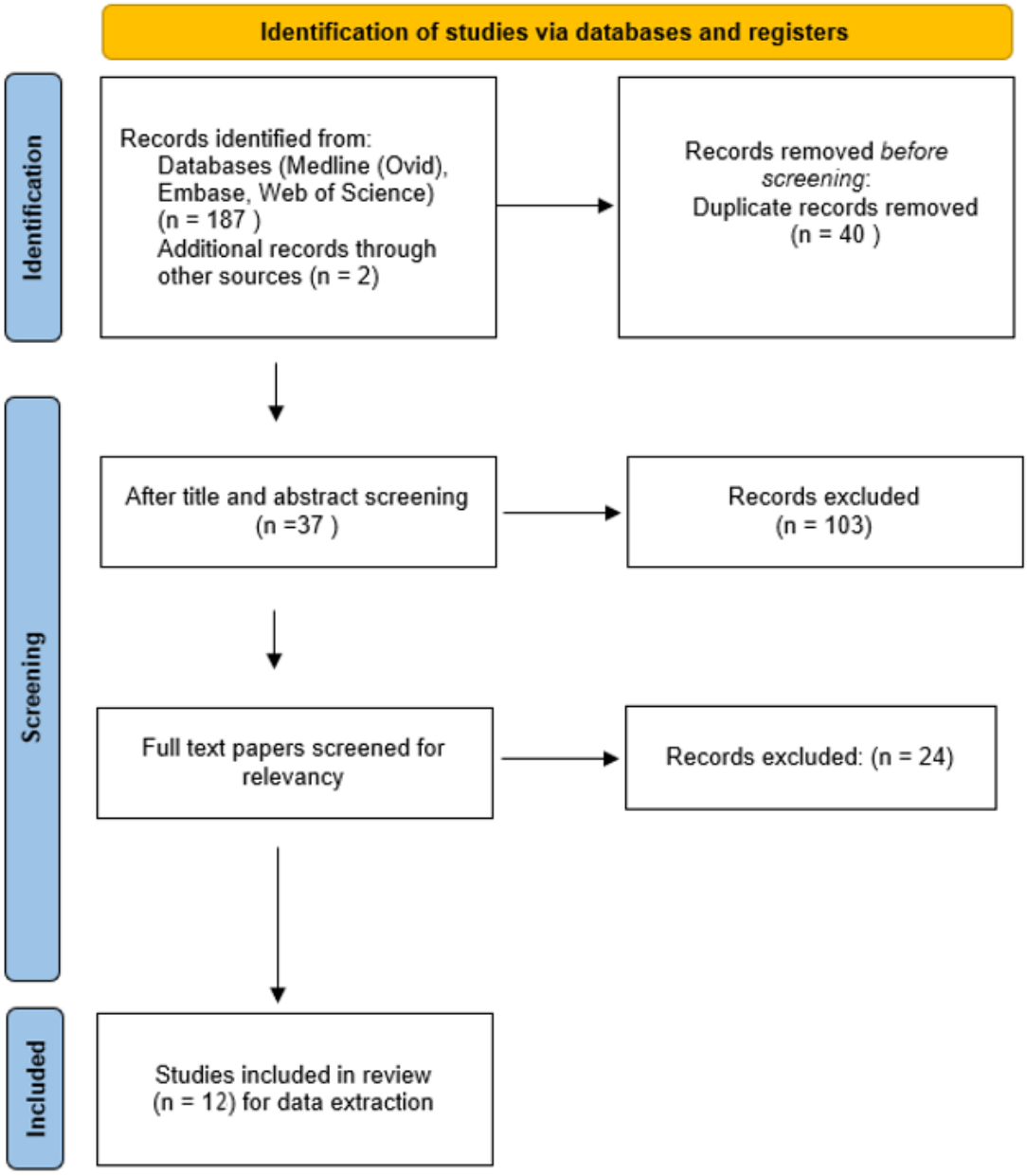


Figure 1: PRISMA flowchart of selection process for included studies (Moher et al., 2009)

The 12 studies included were distributed across 10 countries; 2 in the U.S.A, 2 in Italy, 1 in United Kingdom, 1 in Canada, 1 in India, 1 in Saudi Arabia, 1 in Australia, 1 in Turkiye, 1 in Pakistan, and 1 in Norway. Dates of publication ranged from 2005 to 2020. In terms of research design, 5 were cohort studies, 3 were case-control studies and 4 were cross-sectional studies. 8 studies involved studying the oral mycobiome by collecting oral cavity samples while 4 studied investigated the gut mycobiome by collection of stool and/or gut mucosa samples. The age range of the population included in the studies ranged from premature infants born at 25 weeks to adolescents 18 years of age. The number of participants in the included studies ranged from 11 to 298 subjects. 3 studies were only concerned with the presence of oral *Candida sp*. 5 studies were interested in presence of both bacteria and fungi in participants. No study was found which investigated both oral and gut mycobiome. A summary of the 12 included studies is included in Table 2.

Table 2: Summary of included studies

Study	Study type	Participants	Exposure	Comparison	Method of sample collection	Outcome/s measured	Outcomes
Beena et al. (2017)	Cross-sectional	n = 100, < 6 years	Early childhood caries (ECC) (DMF ≥1), n = 50	Caries-free, n = 50	Oral swabs	Presence of oral *Candida*	*Candida* carriage of ECC children (84%) is significantly higher than the non-ECC group (24%). No significant difference of *Candida albicans* (p>0.05) between both groups.
Bender et al. (2019)	Prospective cohort	n = 25, all infants recruited at birth	Antibiotics given, n = 7	No antibiotics, n = 18	Stool samples (collected 1 week apart)	Presence and diversity of gut mycobiome	No identifiable fungi found (only 3 samples had identifiable fungi which were too few for ITS sequencing)
Borgo et al. (2017)	Case-control	n = 61, mean age 10.03 ± 0.68	Obesity, n = 28	Normal-weight children, n = 33	Stool samples	Presence and abundance of gut mycobiome	Obese children had lower abundance in *Candida* spp and *Saccharomyces* spp (p=0.047 and p=0.034) in comparison to normal weight children.
De Jesus et al. (2020)	Cross-sectional	n = 80, < 72 months of age	Early childhood caries (ECC), gender (n = 40, 15 males, 25 females)	Caries-free children, opposite gender (n = 40, 19 males, 21 females)	Supragingival dental plaque samples	Presence and diversity of oral mycobiome	Children with ECC had higher level of *Candida dubliniensis* compared with those caries-free (*P* < 0.05). No significant difference was observed with *Candida albicans* (P > 0.05). *Candida* was the most abundant fungal genus in the ECC group. No sex-based differences were observed with alpha and beta diversity analyses.

El Mouzan et al. (2017)	Case-control	n = 35	Crohn's disease (CD), n = 15, median age 15 (7.3–17.8)	Non-IBD controls, n = 20, median age 16.3 (3.9–18.6)	Gut tissues and stool samples	Presence and diversity of gut mycobiome	Fungal taxa significantly more abundant in CD stools included *Cortinariaceae* family (p = 0.02), *Hymenochaete* (p = 0.03), and *Gymnopilus* genera (p = 0.02). There was no significant difference in fungal diversity between CD and controls.
Fechney et al. (2019)	Cross-sectional	n = 17, aged 7-10 years	Caries, n = 6	No caries, n = 11	Supragingival dental plaque samples	Presence and abundance of oral mycobiome	17 fungi species were more abundant in children with no caries (p<0.001). Alpha and beta diversity analyses showed no differences in richness and evenness of oral mycobiome between both groups. *C. albicans* was present in all samples regardless of whether there was caries present.
Kadir et al. (2005)	Prospective cohort	n = 64, aged 0-2 years	Different types of milk (Breast milk, bottled milk, and other fluids), n = 38	Only breast milk, n = 26	Oral swabs (samples collected within 8-week period)	Presence of oral *Candida*	Prevalence of oral *Candida* in children feed different types of milk (18.5%) was significantly higher (p<0.01) than those who only had breast milk (0%).
Oba et al. (2020)	Prospective cohort	n = 12, new-borns	Diet (exclusive breast feeding until 4 months, n = 6),	Non-exclusive breast feeding until 4 months, n =6, no comparison for age	Oral swabs (over 6 months)	Presence and abundance of oral mycobiome	*Ascomycota* phylum was present 48 hours after birth and higher at 2 months (p = 0.0475) compared to 4 and 6 months of age. *Candida* spp. was low at birth (only 17% of infants), 25% of infants at 2 and 4 months, and 75% of infants at 6 months.

			Age (48 hours after birth), n = 12				No statistical significance in presence of oral mycobiome was found between the two different dietary groups.
Raja et al. (2010)	Case-control	n = 100, 6-12 years	Caries (DMF≥1), n = 50	No caries, n = 50	Oral swabs and rinses	Presence of oral *Candida*	Children with caries had higher frequency of *Candida* ($p<0.01$) with odds ratio of 67.37. *C. albicans* accounted for 65% of positive samples followed by *C. tropicalis* (25%), *C. parapsilosis* (8.33%) and *C.rugosa* (1.66%).
James et al. (2020)	Prospective cohort	n = 11 infants, aged less than 2 years, median age 30 + 4 weeks	Preterm babies (born 25 – 36 weeks)	Not applicable	Stool samples (collected over 12-month period; at 6, 12, and 18 months)	Presence and abundance of gut mycobiome	At phylum level, Ascomycota was the most dominant phylum at all 3 timepoints. *Candida* accounted for 28.8% of all oral mycobiome reads, was present in every sample and ranged from 0.02% to 97.1% in abundance. Diversity of preterm infant gut mycobiome was low at birth and varied between infants but alpha diversity increased with age (similar to findings in studies on full term infants).
O'Connell et al. (2020)	Cross-sectional	n = 33, aged 2-7 years old	Caries active with enamel lesions (CAE) n = 28, caries active with dentine lesions (CA), n = 46	Caries free (CF) n = 8	Supragingival dental plaque samples (n = 82)	Presence and abundance of oral mycobiome	*C. albicans, C. dubliniensis, Nigrospora oryzae* and an unclassified *Microdichium* sp. were associated with caries. *Candida albicans* most abundant in all plaque groups. C. *dubliniensis* is second most abundant mycobiome species and oral mycobiome community richness (alpha diversity) reduced as caries progressed.

Schei et al. (2017)	Prospective cohort	n = 298, infants 36 weeks in gestation to 2 years of age	Probiotic drank by mother during and after pregnancy, n = 149	No probiotics, n = 149	Stool samples	Presence and diversity of gut mycobiome	Alpha diversity of gut mycobiome of infant increased from its lowest at 10 days after birth (1st sampling point). Ascomycota spp. Was 86.4% of the fungi in all the age groups. The odds of DNA gut mycobiome of infants being detected is higher if mother had a detectable gut mycobiome DNA (OR = 1.54, p = 0.04). However, no significant association was found between probiotics and infant gut mycobiome.

2.4.2 Risk of bias within included studies

Results of quality assessment for the included studies according NOS are presented in Table 3 for cohort and case-control studies and Table 4 for cross-sectional studies. Most of the studies included scored at least one star in each of the three categories, selection, comparability and outcome. The number of stars awarded ranged from four to six stars which meant that all 12 studies included were considered to be of fair quality with high risk of bias. Some of the weaknesses found within some of the included studies were small sample sizes, inadequate follow-up (less than 6 months), minimal description of the cohort included and lack of control of possible confounding factors.

Author	Year	Selection	Comparability	Outcome
Bender et al.	2019	★★	★	★★
Borgo et al.	2017	★★★	★	★★
El Mouzan et al.	2017	★★★	★	★★
Kadir et al.	2005	★ ★	★ ★	★
Oba et al.	2020	★ ★	★	★
Raja et al.	2010	★★★	★	★★
James et al.	2020	★ ★ ★	★	★ ★
Schei et al.	2017	★ ★	★	★ ★ ★

Table 3: Quality assessment of included cohort and case-control studies according to Newcastle-Ottawa Scale with a maximum possible score of 9 stars (NOS) (Wells et al, 2013).

Author	Year	Selection	Comparability	Outcome
Beena et al.	2017	★★★	★	★★
De Jesus et al.	2020	★★★	★	★★
Fechney et al.	2019	★	★	★★★
O'Connell et al.	2020	★	★	★★★

Table 4: Quality assessment of included cross-sectional studies according to Newcastle-Ottawa Scale (NOS) with 8 stars in total as the highest likely score (Wells et al, 2013).

2.4.3 Body site for collection of samples

In 7 out of the 12 studies included, mycobiome of the participants were collected by oral cavity samples. The types of sampling method were oral swabs (3 studies), supragingival dental plaque samples (3 studies) and one study collected both dental plaque and oral rinse. Gut mycobiome were analysed in 5 studies with collection of stool samples in 4 studies while one study took samples of both stool and gut mucosa samples of participants.

2.4.4 Host factors and environmental factors

In the 12 studies selected for this review, host factors were investigated in 9 studies. Four studies, by Beena et al. (2017), Fechney et al. (2019), Raja et al. (2010) and O'Connell et al. (2020) were concerned with effects of caries on oral mycobiome. One study involved studying the effects of inflammatory bowel disease (IBD), specifically newly diagnosed Crohn's disease, on gut mycobiome (El Mouzan et al., 2017). One study looked at associations of caries as well as gender on diversity of oral mycobiome (de Jesus et al., 2020). James et al. (2020) investigated the causal effects of preterm birth on mycobiome of gut. The impact of age on oral fungi was looked at in a study by Oba et al. (2017). This study also included an environmental factor, specifically dietary milk. Another study by Kadir et al. (2005) was also concerned with the effects of dietary milk on mycobiome. Borgo et al. (2017) also considered the effects of obesity on gut. The effects of the use of probiotics on infant gut mycobiome were investigated by Schei et al. (2017).

2.4.3.1 Caries

In Beena et al. (2017), comparison of *Candida* between two groups of children between 6 years of age, 50 with early childhood caries (ECC) and 50 caries free children showed presence of *Candida* in 84% in ECC children compared to 24% non-caries children, which was statistically significant. Fechney et al. (2019) found Ascomycota and Basidiomycota to be the most dominant phyla in all participants in

their study of 17 children (7–10-year-olds) with and without caries. C. albicans was present in all samples, no matter whether caries was present or absent. Assessment of “core” fungi in participants with and without caries was done and 6 OTUs were present at minimum 50% of samples with both groups. Out of 6 OTUs, all plaque samples had presence of *C. albican* and an unclassified *Saccharomycetes* species. The next pervasive “core” members were *N. _diffluens, R. mucilaginosa, M. globose* and *Cladosporium cladosporoides*. Alpha and beta diversity measurements showed no difference in richness, evenness and dissimilarity of the oral mycobiome between carious and non-carious dentitions. However, the oral mycobiome profile differed between the two groups in terms of abundance. 20 OTUS were differently abundant between them with healthy groups having more abundance of majority of the species (17 out of 20). A third study, done by Raja et al. (2010) which compared oral *Candida* between 100 children (age 6-12 years old) with and without caries found that carious dentition has a higher frequency of *Candida* ($p<0.01$) and an odds ratio of 67.37. C. albicans were present in 65% of positive samples. High number of *Candida Dubliniensis* were found in children (<72 months in age) with early childhood caries compared to non-carious children but no difference in *Candida albicans* ($p>0.05$) was noted in a study by De Jesus et al. (2020). O’Connell et al. (2020) studied 33 children aged 2-7 years old and with different caries severity and found association of severity of caries with *C. albicans, C. dubliniensis, Nigrospora oryzae* and an unclassified *Microdichium* sp., while 12 other taxa were associated with health.

2.4.3.2 Inflammatory bowel disease (IBD)

A study was done by El Mouzan et al. (2017) which investigated the gut mycobiome of 15 children with Crohn’s disease (median age 15) compared to 15 healthy children (median age 16.3) and found that although some difference in abundances of Basidiomycota and Ascomycota levels were noted between the two groups, these differences were not statistically significant ($p>0.05$). However, stool samples of

children with Crohn's' disease had a higher abundance of fungal taxa, which included Cortinariaceae family (p = 0.02), Hymenochaete (p = 0.03), and Gymnopilus genera (p = 0.02). However, there were no difference in gut mycobiome diversity between the two groups.

2.4.3.3 Preterm birth

One study by James et al. (2020) which studied the gut mycobiome of 11 infants under the age of 2 years, who were preterm babies who were born from 25 weeks to 36 weeks. Presence of *Candida* were found in 28.8% of all reads, ranging from 0.02% to 97.1% in terms of abundance. Ascomycota phyla dominated at all three collection points over the course of 12 months. Four *Candida* species namely *C. albicans, C. metapsilosis, C. parapsilosis* and *C. tropicalis* species were identified in the preterm infant group but overall, preterm infant gut fungal microbiota had low diversity with each infant having a specific fungal profile. However, the alpha diversity was found to increase with age, like those of full-term infants. These findings suggest that low presence of gut fungal microbiota at birth and increase in its presence and diversity are more likely to be affected by age, rather than prematurity of infants.

2.4.3.4 Gender

De Jesus et al. (2020), in addition to studying the effects of caries on oral fungal microbiota, also compared association between different genders of children under 72 months in age with oral fungal microbiota. Alpha and beta diversity analyses found no gender-based differences but *Candida* was noted to be the most abundant fungal genus in all samples.

2.4.3.5 Age

Oba et al. (2020) monitored the oral mycobiome of 12 new-borns over the course of 6 months and noted *Ascomycota* phylum was present within 24-48 hours after birth and increased with age but was highest (p = 0.0475) at 2 months. *Candida* spp were detected in low amounts at birth but at 6 months, was present in 75% of infants.

2.4.3.6 Antibiotics

The effects of antibiotics on gut mycobiome were investigated by one study which recruited 25 infants where 7 were on antibiotics during at least one of the sample points of collection (2 sample collections, 1 week apart). Limited amounts of gut mycobiome were found in the samples whereby only 3 samples out of 51 had species which were identifiable. This was associated with a lack of bacterial diversity but also reiterated findings from other studies which found low gut mycobiome present at birth (James et al., 2020).

2.4.3.7 Probiotics

A prospective cohort study by Schei et al. (2017) monitored 298 babies and their mothers (half of whom were given probiotics) from 36 weeks to 2 years of age to study the effect of probiotics on gut mycobiome. Consumption of probiotics increased the abundance of gut mycobiome in mothers and the DNA gut mycobiome of infants had a higher chance of being detected if mother had a detectable gut mycobiome. However, no significant correlation was found between probiotics taken by mother and infant gut mycobiome.

2.4.3.7 Diet

There were 2 studies which investigated the effects of diet, specifically milk on mycobiome. A study by Kadir et al. (2005) compared the effect of dietary milk types on oral mycobiome, specifically *Candida*. 64 children under the age of 2 were recruited and separated into 2 groups, those who only had breast milk (n= 26) and those who drank both breast milk and other milk and fluids (n=38). *Candida* was prevalent in 18.5% of children who drank different type of milk while 0% of *Candida* was found in children who only had breast milk. This statistically significant finding ($p<0.01$) showed a likely association of oral mycobiome with dietary intake. However, Oba et al. (2020) monitored the oral mycobiome of 12 new-borns over the course of 6 months and found no significant correlation between diet and oral fungi

communities and noted that presence and abundance of oral mycobiome were more likely to be affected by age.

2.4.3.8 Obesity

Borgo et al. (2017) compared the abundance of gut mycobiome between 28 obese children and 33 normal weight children and found that children with obesity was associated with a lower richness of *Candida* spp ($p = 0.047$) and *Saccharomyces* spp ($p = 0.034$). However, no other gut mycobiome was found to be affected by this exposure.

2.5 Discussion

This review included 12 studies which were all judged to be of fair quality with high risk of bias, although the reporting of studies included was mostly poor. Most of the studies presented results which were lacking in specificity, with minimal statistical data and were difficult to interpret. This complicated the analysis of the relevance and significance of the findings presented. The heterogeneity of the aims of this systematic review, along with variable types of exposures, differing outcome measures and findings as well, precluded meta-analysis of any homogenous groups of exposures and related outcomes.

2.5.1 Host factors

There were some discrepancies noted among the 5 studies which had investigated the association of caries to oral mycobiome. While presence of *Candida* in children with caries was found to be statistically significant in the related studies, there were some discrepancies about the association of *Candida albicans* with caries. O'Connell et al. (2020) and Raja et al. (2010) found *Candida albicans* to be related to severity of caries, while the rest of the included studies found presence of *Candida albicans* in both carious and non-carious children. O'Connell et al (2020) and de Jesus et al. (2020) also found *C. dubliniensis* to be associated with caries. Differences in method of collection of samples, whereby some were collected by oral swabs while others

were done by oral rinses could affect the results of the studies. Apart from that, the low number of subjects in some of the studies included could potentially skew the analysis of the results.

The different host factors found present in the studies selected for this review were obesity, Crohn's disease, preterm birth, and gender. Only one study was found for each of the factors mentioned, with inconclusive correlations between the host factors and presence and diversity of the mycobiome. Again, this was likely due to the low sample size, poor control of confounding factors as well as poor use of statistical methods.

A study by Oba et al. (2020) found that age is a host factor that could be related to presence and abundance of oral mycobiome, with Ascomycota and *Candida* spp increasing with age. However, its low sample size and other confounding factor such as diet, warrants more investigation into its true effects on presence of oral mycobiome and its diversity.

2.5.2 Environmental factors

There were 3 environmental factors which were identified from the included studies, which were antibiotics, probiotics, and diet. Only one study each were found for antibiotics and prebiotics, respectively and neither one showed any correlation with gut mycobiome presence and diversity. Two studies looked at the effects of dietary milk on oral mycobiome and while Kadir et al. (2005) found that level of *Candida* in children who drank different types of milk to be significantly higher than those who drank only breast milk, Oba et al. (2020) found no difference in oral mycobiome between the two dietary groups. There is insufficient evidence to determine which findings were more accurate and significant.

2.5.3 Strengths and limitations of included studies

The strength of this systematic review is that, to our knowledge, it is the first to consider the possible host and environmental factors associated with oral or gut mycobiome, specifically in children. Studies on mycobiome, as mentioned before, are scarce and have been mostly specific to a narrow age group or combination of children and adults with mostly, very small sample sizes. Furthermore, mycobiome are usually added as an afterthought in most studies, with much of the focus being on bacterial microbiota. This makes it difficult to ascertain the true effects of different host and environmental factors on its presence and diversity.

The small number of samples in the included studies, short period of follow-ups, retrospective non-randomised designs, lack of standardised criteria for analysis of mycobiome and poor control of confounding factors were major limitations of this systematic review. The diverse host and environmental exposure factors found but with limited number of studies makes it difficult to regard these studies as being sufficient to draw definite conclusions.

2.5.4 Implication for future studies

From the included studies, presence of *Ascomycota* phylum and *Candida* were noted in most of the studies, with varying significance in terms of their associations with the specific host or environmental factors being investigated. This knowledge helps to better plan for future studies, to include their presence and diversity as one of the areas of interest. Studies included in this review can also be used as reference points in the planning of future studies so that the designs, collection methods and reporting of those studies can be improved. Some of the results from the included studies, particularly those which reported findings of oral/gut mycobiome in healthy participants can be compared with results from future studies on similar healthy cohorts from the same population, of the same age, gender, and health status.

The environmental and host factors from the included studies are only some of potential factors which could affect the oral and gut mycobiome of children. Other unincluded potential factors can also be explored for their potential effect, perhaps with careful consideration of the study design. A more robust study with a stricter protocol, larger sample size which followed participants over a longer period time with more sampling time points could provide a more accurate representation of the correlations between exposure factors and oral/gut mycobiome.

2.6 Conclusion

This review indicated that current researches which considered the effects of host and/or environmental factors on oral or gut mycobiome are still lacking and the ones identified in this study have high risk of bias, low sample size, inadequate description of protocol and results of study, minimal sampling points and inadequate length of follow-up. However, they were also helpful in the understanding of the different oral and gut mycobiome taxa, abundance and diversity in children when exposed to different host or environmental factors as well as the composition and diversity of oral and gut mycobiome in healthy children, specifically Ascomycota and *Candida.* The host and environmental factors identified in this systematic review are useful as guides for the study design of future investigations into oral or gut mycobiome.

3 SysMIC Course, QIIME2 & R Studio

3.1 SysMIC Course

SysMIC is an online course that offers training in the areas of mathematics, statistics, and computing skills specifically for the bioscience research community. There are a total of 3 modules available for enrolment, distinguished by increasing difficulty as well as specific requirements of the participants. I enrolled myself in Module 1 – Introduction in Quantitative Skills for Bioscience, which I felt would help in improving my understanding of different methods available for analysis of the types of data involved in this study. This module was conducted online over the course of 6 months with a total of 12 sessions. Each session was conducted every 2 weeks. The modules were made up of a series of lectures, quizzes, and assignments. Three different programs for data analysis were introduced in this Module which are MATLAB, R package and Python.

MATLAB is a type of computing program that is used for development of algorithm, analysis, visualisation, and exploration of data as well as creation of scripts and customize functions, among others. It also has built-in functions for common mathematical, engineering, and scientific calculations which can be accessed by executing specific coding commands. R package on the other hand, is an extension of the R statistical programming which bundles together a multitude of data, code, tests, and documentations including those from other users. Packages from other users can be shared in R and be used to solve similar data analysis. Lastly, Python is an alternative programming language to the former two that is specifically useful for data analysis. It has an emphasis on code readability and uses general purpose language which makes it a suitable for a variety of different programs. All three of these programs uses distinct and specific coding methods which are unique to each program and are mostly not interchangeable with one another. A knowledge and

understanding of one program would not necessarily transfer to the understanding the others.

Module 1 was made of three topics: Networks, Maths and Modelling, and Statistics. The first topic, Networks was taught in session 1 to 4 and involved introduction to the MATLAB program and how biological systems are described and analysed as networks. Maths and Modelling were taught next in session 5 to 8, which included how biological systems and mechanisms can be modelled and simulated using MATLAB and Python programs. Sessions 9-12 focused on Statistics, specifically the introduction of the statistical package R and how to deal with statistical analysis. The 12th session, which is the last one encompassed a longer mini project which involved integration and application of knowledge and skills learnt in the prior sessions in solving a specific modelling problem.

Although I did initially find the program course useful in terms of providing an introduction to the different data analysis programs available, the data analysis involved in this study required a more detailed and comprehensive understanding of oral fungal microbiota analysis which were not fully covered in this module. The module also focused heavily on the MATLAB program which I did not find to be relatable to this study or as user friendly compared to other programs available. As a result, other resources available online such as Github and QIIME2 forums were accessed to provide a better understanding of possible analysis specific to these types of studies and ultimately, QIIME2 was the program that I chose to carry out the data processing and analysis.

Although training in R program was included as part of this Module, I felt that an introduction to the program at the beginning of the module would have been more useful for me considering that most of my statistical analysis revolved around script

writing in R studio. A late introduction to the R program meant that it was a steep learning curve to understand the program and apply the concepts taught to the data I have on hand. In the end, I found it easier to seek help from a PhD student on how to manipulate data on R studio to obtain the statistical results I required for this study.

Overall, even with its limitations, the opportunity given to be enrolled in this course was much appreciated as it gave me a starting point in my journey in understanding the scale, depth, and complexities in managing the data involved in bioscience, specifically in the study of microbiology.

3.2 QIIME2

As mentioned previously, QIIME2 was the program that I chose to carry out the processing and analysis of oral fungal microbiota in this study. QIIME2, which stands for Quantitative Insights Into Microbial Ecology is a next-generation, open source, microbiome bioinformatics software package which is useful for comparison and analysis of microbial communities. It is a free program, community developed, and extensible. It is able to produce statistical results and publishable interactive visuals within the program and can also process different types of microbiome data. This includes raw sequencing data, such as those generated from Illumina platform and other types, including amplicons of other markers such as 18S rRNA, internal transcribed spacers (ITS), and cytochrome oxidase I (COI), shotgun metagenomics, and untargeted metabolomics (Bolyen et al, 2019).

QIIME2 analysis can be carried out in a multitude of ways, depending on the raw data at hand, goal of experiment and analysis, as well as how the data was collected. Within the QIIME2 website (https://qiime2.org/) are different interfaces that can be selected to suit the needs of the microbiome data analysis. The q2cli,

the command line interface was the one I chose as it allowed me to execute specific QIIME analyses from different codings that I have researched, generated and adapted from tutorials on the QIIME2 website and through Github to suit the data I have. All the tutorials that were available on the QIIME2 website were specific to bacteria and hence, it was a bit of a challenge to adapt them to be used for analysis of oral fungal microbiota. The biggest issue faced while using the QIIME2 program was calculating alpha and beta diversity. Even though it was possible for these statistical analyses to be done in the QIIME2, I did not have any success in generating a phylogenetic tree using the program, which was an essential part to proceed with further statistical calculations. Hence, a decision was made to continue the alpha and beta diversity analyses using the R studio, which had an option to perform the necessary calculations required without the need for a phylogenetic tree. QIIME2 was instead used for taxonomic analysis and visualisations which showed different taxonomic levels and their abundance, according to each sample taken and sampling time point.

Different QIIME2 plugins are also available which utilises other software packages and can be downloaded to assist in data analysis. However, this was not necessary in the analysis involved in this study and was not utilised. An advantage of using the QIIME2 program was that the visuals can be shared and viewed even by those who did not install the program. This is because the website (https://view.qiime2.org/) has interface which allowed ".qza" and ".qzv" files to be viewed directly on the browser without the need to upload them onto a server (Bolyen et al, 2019). This proved useful when I needed to share a few of my initial results with my supervisors.

Even with the challenges mentioned, QIIME2 proved to be a useful tool in the bioinformatic analysis of oral fungal microbiota with its ease of use compared to the

MATLAB program. However, more tutorials should be generated in the future specific to fungal microbiota to help those like myself who are involved in studies involving fungi.

3.3 R Studio

The R studio is an open source, free, integrated development environment for R, which is a type of programming language specifically for statistical computing and graphics. There are packages that can be installed within R studio which are essentially extensions to the R programming language. They contain code, data and documentation in a standardised format and are useful in carrying out specific data calculations and generate visualisations (R Core Team, 2020).

Due to issues faced when performing statistical analysis using QIIME2, I chose to use R studio to carry out alpha and beta diversity measurements of the oral fungal microbiota. I was successful in calculating the alpha and beta diversity by following the R studio workflow of a PhD student who had carried out a statistical analysis of microbiota and by several discussions with another PhD student on how to generate the best analysis using the data I have.

Admittedly, there were steep learning curves involved in the understanding of the different software used in this study, each one with its own complexities. In the end, I was able to appreciate the pros and cons of each own and make best use of them to fit the overall goal of my data analysis.

4 Aims and Objectives

4.1 Aims

The aims of this project were to investigate the oral mycobiome composition in adolescence, their diversity, evenness, richness over a specific length of time and whether their living environment played any significant part in any changes noted. Saliva was chosen for its ease of collection. The data in this study was concurrently used in a thesis of a DDent Paediatric Dentistry postgraduate student (Kakiora, 2021) which focused on the environmental effects on oral microbiota, specifically bacteria in adolescents.

4.2 Objectives

A collection of saliva samples was taken from students attending Haileybury, an independent boarding school in Hertfordshire at different time points over a period of 12 months which allowed for analysis of how the oral mycobiome changed over time. This timeline was reduced from the initial plan of 24 months due to the disruption from the Covid-19 pandemic.

Students were divided into two categories: boarders and non-boarders. Boarders were students who stayed in the boarding school during school term and returned to their respective homes during term break and non-boarders were those who stayed at home during school term and commute to school. Division of students into these two categories were done to:

1. Compare the effects of their living environment on their oral mycobiome and whether students who experienced a change in their living environment would

have an altered oral mycobiome in comparison to students who did not experience a change in their living conditions.

2. Determine any possible effects of a shared school environment amongst boarding students on their oral mycobiome.

5 Methodology

5.1 Sponsorship

This research project was funded by a research grant which was awarded by the Biomedical Research Centre (BRC), University College London Hospital.

5.2 Ethics statement

This human study was ethically approved by the Research Ethics Committee (REC) of University College London (UCL) (Project identification: 7567/001). The ethical approval for this study was received on 18th of July 2018 was valid up to 31st of August 2020 (Appendix 3). A request for extension of the ethical approval was done by submission of the Annual Continuing Review Approval Form. This request was approved by the REC and the ethical approval was extended to 31st of August 2021 (Appendix 4).

Two versions of the participant information sheet which included aims and objectives of the study, as well as other related information were made. One version was for young adults, which were given to the participants (Appendix 5). Another version was tailored specifically for parents/guardians (Appendix 6). All parents or guardians of the students who have volunteered to be involved in this study gave their approval and informed consent by completion of consent form in Appendix 7.

5.3 Participants

The study was conducted at Haileybury, an independent boarding school in Hertfordshire in the south of England. It is a co-educational school for pupils aged eleven to eighteen years which also offers boarding for both boys and girls. All the participants in the study were students enrolled in this school. Inclusion criteria for

participants to be accepted into the study were those who recently enrolled into the school. Initially (in September 2019), 17 students were recruited in the study, ranging from year 7, 9 or 11. Each one of them were assigned to a specific ID code, in the form of H* which ranged from H1 to H18 which maintained the same throughout the study. At the start of the subsequent year of study in September 2019, a total of 9 additional students were recruited. Unfortunately, these new cohort of participants had to be excluded from the study due to limited samples taken caused by the unexpected interruption of the study from the Covid-19 pandemic.

An oral health questionnaire (Appendix 8) was created by a postgraduate student (Violeta Kakiora) and was given to each participant by her to complete. The questionnaire encompassed information about participant's demographic, boarding status, their general and oral health. They were divided into two groups based on their boarding status. Boarders were the test group which were students who stayed in Haileybury boarding school during study term and experienced changes in household. The second group were the control group, non-boarders. They were students whose households remain unchanged and were the ones who stayed at home and travelled to and from school daily.

5.4 Clinical screening

A dental clinical examination was carried out alone by a postgraduate student (Violeta Kakiora) on all the participants involved in the study which comprised of a basic oral examination that included assessment of the oral mucosa, detection of presence of any caries, restorations as well as an assessment of amount of plaque present in the oral cavity.

A clinical screening protocol was also developed by Violeta Kakiora and adhered to for every participant (Appendix 9). A clinical data collection form was prepared for the purpose of recording findings from clinical examination of participants (Appendix 10). All the information and data were then shared with me to be included as part of my study.

Examination of the oral mucosa was done systematically in the same order for each participant, beginning from the upper right, upper left, lower left and lastly, the lower right with adequate retraction of the lips and cheeks. Assessment of the oral mucosa which included the soft tissues, lips and cheeks was recorded following the Pulp, Ulceration, Fistula and Abscess (PUFA) index. Pulp (P) was denoted by presence of an open pulp exposure was noted in the permanent dentition while Ulceration (U), Fistula (F) and Abscess (A) were indicated as present if any were found in the oral cavity.

The PUFA index scores and representation of codes were indicated as such:

0 – no lesions

1 – single lesion

2 – two or more lesions

The Decay, Filled Teeth (DFT) index were used to record presence of caries and restorations following the modified ICDAS criteria. Apart from this, any unrestorable decay were also noted if present.

Caries score following the ICDAS criteria as recorded as followed:

0 – No caries evident

A – Incipient caries

3 – Localised enamel breakdown with no dentine visible

4 – Underlying dark shadow from dentine

5 – Distinct cavity, dentine visible

6 – Extensive distinct cavity, dentine visible

The Silness-Löe plaque index was used to measure plaque, whereby presence of soft debris and mineralised deposits were recorded 6 permanent teeth, the upper right first molar, upper right lateral incisor, upper left first molar, lower left first molar , lower left lateral incisor and lower right first molar. If the permanent tooth has not yet erupted, a primary tooth can be substituted for it. However, if the tooth was missing, no substitution was necessary. This was recorded for all four surfaces the teeth which were buccal, lingual, mesial and distal and each surface was allocated a score of 0-3.

The criteria for the allocated scores are as such:

0 – No plaque

1 – A film of plaque adhering to the free gingival margin and adjacent area of tooth. The plaque may be seen in situ only after application of disclosing solution or by using the probe on the tooth surface

2 - Moderate accumulation of soft deposits within the gingival pocket, or the tooth and gingival margin which can be seen with the naked eye

3 - Abundance of soft matter within the gingival pocket and/ or on the tooth and gingival margin

The plaque index for each tooth were calculated by dividing the total scores from all four surfaces of each tooth by four. Additionally, if any participants were wearing any orthodontic appliances and dentures, this was recorded as well.

5.5 Sampling of specimens

Sampling of saliva was planned to be carried out at 3 different time points in each of the 3 terms of the school year from 2018 – 2019, the beginning, middle and end, making up a total of nine collection time points. Each sampling cycle was identified as S^, beginning from S1 to S10. The postgraduate student (Violeta Kakiora) collected the first 8 samples by herself but I was present and assisted in the collection of the last set of samples.

Terms (2018-2019)	Start	End	Start of Half term	End of Half term
Autumn	03/09/2018	14/12/2018	22/10/2018	02/11/2018
Spring	07/01/2019	12/04/2019	16/02/2019	22/02/2019
Summer	23/04/2019	28/06/2019	25/05/2019	02/06/2019
Autumn	04/09/2019	13/12/2019	21/10/2019	3/11/2019

Table 5: Haileybury School term dates for year 2018-2019

For collection of saliva sample, each participant was given a 14ml sterile tube which was pre-filled with 2ml of saliva buffer, the composition of which is denoted in Table 6. A printed sticker labelling the participant identification code and sampling cycle (H*S^) was pasted onto each tube. Participants were asked to deposit approximately 3ml of saliva sample into the tubes each time. If participants were absent or unavailable during a sample collection time point, a school teacher would attempt to collect a sample at a time as close as possible to the original date of collection. If not possible, then no samples would be collected for this cycle for those participants.

Reagent	**Final volume**	**Final concentration**
1 M Tris pH 8.0	50 ml	50 mM
0.5 M EDTA pH 8.0	100 ml	50 mM

Sucrose	17.2 gr	50 mM
3 M NaCl	33.33 ml	100 mM
10% SDS	100 ml	1%
dH2O	Top up to 1 L	

Table 6: Saliva buffer mixtures (Marlina, 2021)

These samples were then transported on ice and kept in the laboratory under a low temperature of -20 ^{o}C in preparation for the second phase of processing of samples, the extraction of DNA, RNA, and sequencing.

5.6 DNA extraction

The extraction and isolation of DNA was conducted using the PurElute™ Bacterial Genomic Kit (EdgeBio, USA), a bacterial genomic DNA purification kit and was carried out completely by a postgraduate student (Violeta Kakiora). The steps and process involved were then shared with me when I became involved in this current study.

Figure 2: Components of PurElute™ Bacterial Genomic Kit by EdgeBio
(Image taken from https://www.medicalexpo.com/prod/edge-biosystems/product-128193-943152.html)

The five components that make up the kit are 25ml of Spheroplast Buffer, 6ml of Lysis 1, 6ml of Lysis 2, 6ml of Advamax™ 2 Beads, and 6ml of Extraction Buffer. Following the manufacturer's instructions, the Spheroplast Buffer was stored at -20ºC while the other components of the Bacterial Genomic DNA Purification Kit were stored at 4ºC. According to the manual, under these conditions, the kit would be stable up to one year.

Other equipment and materials used for the processing of this kit, as per the instruction of the manufacturer were microcentrifuge which is capable of reaching ≥10,000 x g, 2ml tube, isopropanol and 70% ethanol.

Following the recommended protocol, 2ml of culture with an OD_{600} value between 2.0 and 3.0 was centrifuged for 5 minutes at 4ºC to obtain pellet, after which the supernatant was discarded. The centrifuging process was then repeated one more time. 400 µl of Spheroblast Buffer was added and vortexed at the highest speed until the pellet was resuspended. The pellet was then incubated for 10 minutes at 37ºC. 100 µl of Lysis and 100 µl of Lysis 2 were added to the pellet, mixed and incubated for 5 minutes at 65ºC. After that, 100 µl of Extraction Buffer were added, the mixture was vortexed for 10 seconds at low speed before being centrifuged for 3 minutes at 18,000 g at 4ºC. 100 µl of Advamax 2 Beads was added and the tube was gently inverted 10 times to mix the solutions. This was followed by another 3 minutes of centrifuging cycle. Supernatant was then transferred to a clean and well-labelled 2ml tube and equal volume of isopropanol (800 µl) was mixed in, Centrifugation was done yet again at 18,000 g for 2 minutes. Following this, supernatant was carefully decanted to remove it without disturbing the DNA pellet. DNA was washed by adding 750 µl of 70% ethanol, inverting the tube 2-3 times and centrifugation at 18,000 g for 2 minutes. The DNA samples were left to air dry for about 40 minutes, until no ethanol was left prior to being resuspended in 100 µl of dH_2O. A NanoDrop ND-1000

Spectrophotometer (Thermo Scientific) was used to measure the concentration of the total DNA of each samples. The extracted DNA were kept in the sampling tubes and stored at -80ºC in preparation for DNA sequencing.

5.7 DNA sequencing and Polymerase Chain Reaction (PCR)

In the next step of data processing, the isolated DNA samples were handed to UCL Genomics for sequencing. Each sample was suspended in a 1.5ml Eppendorf tube and labelled clearly with the details of the sample. A spreadsheet was also made which outlined each sample's eluted volume, DNA concentration (ng/ul), 280/260 and 230/260 ratios which were measured using the NanoDrop Spectophotometer (see Appendix 11). Qubit dsDNA HS (High Sensitivity) Assay Kit and Qubit dsDNA BR (Broad-Range) Assay Kit by ThermoFisher Scientific Inc (2015) were then used for quantitation of DNA samples.

The Swift Amplicon 16S+ITS Panel by Swift Biosciences was the chosen kit for the next generation sequencing (NGS) analysis of both bacterial and fungal microbiota to produce a data library of quality. This kit does so by using a single multiplexed primer pool that targets both the V1-V9 variable 16S rRNA genes as well as fungal ITS 1 and ITS 2 spacer regions which would result in a more balanced sequence read depth for each DNA region. This is superior when compared to a conventional 16S rRNA gene assays that primarily targets only variable regions V3-V4 or single-plex primers which are region-specific, need phased reverse primers and produce libraries with low complexity, reduced sensitivity and poor quality sequencing.

Specification	Feature	16S + ITS Panel
Input	Input DNA required	10pg-50ng
Workflow	Required time Multiplexing on MiSeq v2 Standard	2 hours Up to 96
Design	Amplicon No. Average amplicon size Genes covered	5 16S rRNA +2 Fungal 475 bp Bacterial 16S + Fungal ITS

Table 7: Key characteristics of Swift Amplicon 16S + ITS Panel

A 2-hour time frame is required for a single-tube workflow set out by Swift Biosciences (2019) from DNA sample to the creation of next generation sequencing (NGS) library. In this study, this was initiated by a multiplex polymerase chain reaction (PCR) which was completed within 70 minutes. A pre-programmed thermal cycler was set at 98ºC prior to loading of samples and later increased to 105ºC when lid is confirmed to be turned on. Over the course of 70 minutes, the pre-programmed thermal cycler underwent a variety of temperature changes following the course of 18 cycles at variable time length as described in Table 8.

	Pre-program thermal cycler		
Multiplex PCR thermal cycler program	Lid heating on		
		98ºC	30 sec
	4 cycles	98ºC	10 sec
		63ºC	5 min
		65ºC	1 min
	14 cycles	98ºC	10 sec
		64ºC	1 min
		65ºC	1 min
		4ºC	Hold

Table 8: Multiplex PCR thermal cycler program and cycles

Following this, post-multiplex PCR clean-up was carried out by making sure the samples are at room temperature. Then, 30 µl of magnetic beads were added to each 30 µl sample and mixed by process of vortex. They were then placed in a microcentrifuge where pulse-spinning was done for collection of contents. The samples were then incubated for 5 minutes at room temperature and the sample tubes were placed on magnetic rack until a pellet was formed in each tubes and solution has cleared which took a total of 5 minutes. Supernatants were removed and discarded from the samples using a clean pipette tip. 180 µl freshly prepared 80% ethanol solution were added to the pellet while it is still on the magnetic rack and repeated for all the tubes. Incubation of 30 seconds was performed and ethanol solution was then carefully removed. A second wash with ethanol solution was done. The samples were placed in the microcentrifuge again, pulse-spun for 30 seconds, placed back on the magnetic rack and any remaining ethanol solution was removed from the tube. Each bead pellet was then resuspended in 17.4 µl post-PCR TE Buffer in preparation for indexing PCR (Swift Biosciences, 2019).

Indexing PCR was performed by programming the thermal cycler with the Indexing PCR program as shown on Table 9. The thermal cycler temperature was allowed to reach 37ºC prior to loading of samples. The program was run for a total of 7 cycles over the span of 20 minutes following the manufacturer's instruction. Post-indexing PCR clean-up was performed following the same methods as post-multiplex PCR clean-up. 20 µl of post-PCR TE buffer were added to the cleaned-up samples, incubated at room temperature for 2 minutes, placed back on the magnet and the clean 20 µl library eluate were transferred to a fresh tube. These freshly prepared libraries were then stored at ~20ºC (Swift Biosciences, 2019).

		Thermal cycler program	
Indexing PCR		37ºC	20 mins
		98ºC	30 secs
	7 cycles	98ºC	10 sec
		60ºC	30 secs
		66ºC	1 min
		4ºC	Hold

Table 9: Indexing PCR program

SparQ Universal Library Quant Kit was used for quantification of these library molecules and this was done using real-time quantitative PCR (qPCR) to quantify how many of the library molecules possess appropriate adapter tag on their ends (Beverly Q, 2020). This was done over the span of 45 minutes and after which the libraries were ready for sequencing on Illumina MiSeq platform with a V2 500 cycle run (2x250bp reads). The resulting raw sequencing data were uploaded onto BaseSpace, a cloud platform that is integrated directly with Illumina MiSeq systems.

The data were uploaded as FASTQ files. Further analysis of the raw sequencing data were split into two, bacterial and fungi. The analysis of the bacterial data was carried out by another postgraduate student (Violeta Kakiora) as part of her thesis and the fungal analysis was done by myself and will be described further in the analysis section.

5.8 Bioinformatics ITS analysis

In consideration of the next phase of the analysis, there are a few bioinformatic tools which are suitable for analysis of ITS gene sequences. The two commonly used tools are Mothur (Schloss et al, 2009) and Quantitative Insights Into Microbial Ecology (QIIME2) (Bolyen et al, 2019). Both uses similar bioinformatic pipelines and are able to produce comparable results (Lopez-Garcia et al, 2018). Analysis was initially attempted using both tools but QIIME2 was ultimately chosen for its ease of use. QIIME2 is a next-generation microbiome bioinformatics platform with a microbiome analysis package that focuses on data and analysis transparency. It also enables analysis of raw DNA sequences which can finish with figures and statistical results (Bolyen et al, 2019). QIIME2 version 2021.4 was used for this analysis, following the workflow in Figure 3.

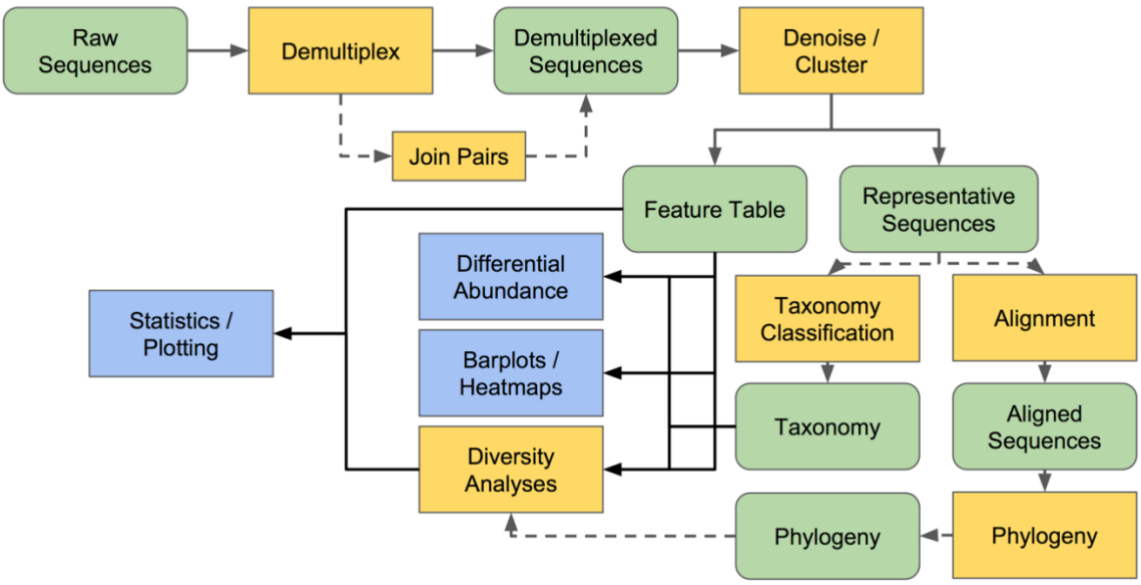


Figure 3: Conceptual overview of workflows using QIIME2 2021.4 (Bolyen et al., 2019)

The 250bp paired-end sequences (FASTQ files) were first downloaded from BaseSpace. The FASTQ file format keeps biological sequences and their quality scores in a plain and simple text format and in recent times has been more commonly used as the standard file format for storing sequence data with high-throughput (Cock et al, 2010). The FASTQ files uploaded into BaseSpace were formatted as decompressed, paired-end demultiplexed fastq files. This meant that there were two ".fastq" files for each sample in the study, where one contained the forward read and the other contained the reverse read for that sample. The sample identifier formed part of the file name and the rest of the file name encompassed further details which are separated by underscore. An example of a single sample with forward and reverse read file names would be "H1S1_S1_L001_R1_001.fastq" and "H1S1_S1_L001_R2_001.fastq" respectively. The underscore-separated fields are:

1. The sample identifier
2. The barcode sequence/identifier
3. The lane number
4. The reading direction which is either R1 or R2
5. The set number

The ITS analysis in QIIME2 was conducted following an open workflow pipeline from Swift Snap ITS1 prep kit (Biosciences Inc, 2019) which is specific for processing amplicon sequencing data and was accessed on GitHub. This pipeline correlates with the workflow shown on Figure 3 which is available on the QIIME2 website. Prior to running the pipeline, three output directories were made:

1. Qobj which is a folder for all QIIME2 data objects
2. Fastq, a folder for intermediate fastq files and
3. Export, a folder to export QIIME2 data objects.

A total of 270 fastq files (135 x2) were imported into QIIME2. A metadata file was created which contains information about the samples which included the sample ID as the heading in the first column, the month and subject as headings in the second and third columns. This was saved in a tab-separated (tsv) text file format. This was then imported into QIIME2. For the next step, the samples were checked for presence of primers and if they were present, they would be trimmed as necessary. The fastq files contained the SNAP primers which were trimmed off. A manifest file was also created which was saved in tsv format and used for importing data into QIIME2 upon completion of primer trimming of fastq files. It contained the sample ID in the first column and the absolute filepath for both the forward and reverse reads. The manifest file is also compatible with metadata and can be used as a metadata file as file but in this analysis, the files were separated. If any samples contained zero sequences, they would be skipped and not imported into QIIME2. The samples all contained sequences and thus, were included. The trimmed forward read (R1) and reverse read (R2) files for every sample were combined into a single file to be processed and read as single read file. The fastq manifest format was set to SingleEndFastqManifestPhred33V2 for import of data into QIIME2. This meant that the read directions must all be either forward or reverse and assumed that the PHRED offset of the positional quality scores is set at 33 for all the fastq files. The files were then demultiplexed with the help of the manifest file to generate a summary of how many sequences were obtained per sample. This was saved in "swift_seqs_qual_qzv" format which can be visualised on the Qiime2view website (Appendix 12). A summary of the distribution of the sequence qualities of the data can also be visualised. The mean of demultiplexed sequence count of the samples is 298861.

The next step would be to perform quality control of the data and the options available are DADA2, Deblur and basic-quality score-based filtering. For our analysis, DADA2 was chosen. DADA2 is a pipeline that functions to detect and correct Illumina amplicon sequence data (Bolyen et al, 2019). The data was denoised by DADA2 whereby noisy sequences were filtered, errors in marginal sequences were corrected and chimeras were removed. The output after use of DADA2 was saved as "swift_seqs_dada2_stats.qzv" which formulated a table that included input, filtered sequence count, percentage of input that passed the filter, denoised data, non-chimeric data and percentage of input (in Appendix 13).

Following this step, a classifier was used for prediction of the taxonomic assignment of each of the denoised amplicon sequence variants. In this analysis, a few classifiers were considered, and after testing of several classifiers, the fungal sequences were classified to the UNITE dynamic ITS database (unite_ver8_dynamic_02.02.2019_dev_ phylum_2020_11.qza) as it gave the highest results in terms of classified phyla (more than two-third of total phyla). This resulted in taxonomic classification as well as a taxa bar plot, saved as "swift_seqs_dada2_barplot.qzv" which was converted into a feature table to be exported for further analysis in R studio, a software for statistical computing and graphics (R Core Team, 2020).

5.9 Taxonomic analysis

The taxonomic classification and analysis were visualised by uploading the taxa bar plot (swift_seqs_dada2_barplot.qzv) on the Qiime2view website. Different visualisations of data can be produced on the website by simply altering the taxonomic level (1-7), taxonomic abundance, index, months and samples tabs. The visualisations also showed the relative frequency of specific taxonomy in each sample

taken. A visualisation of taxonomic level 2 which represent Phylum level is shown in Figure 4.

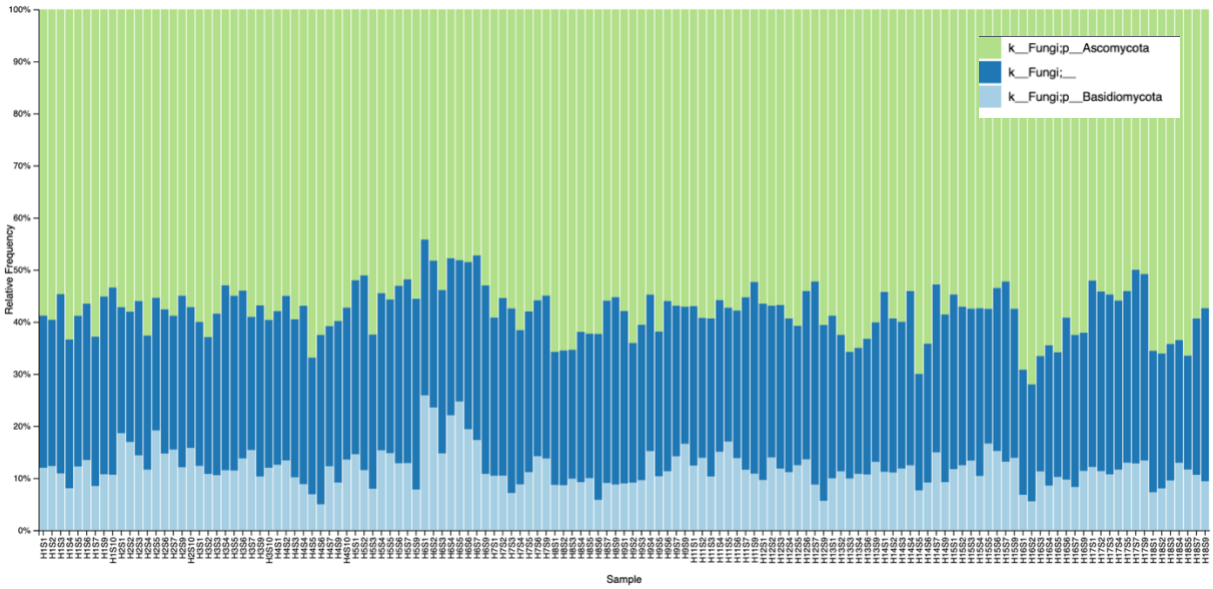


Figure 4: Taxonomic level 2 (phylum level)

Visualisations generated on QIIME2view of species relative abundance for each participant at every sampling points and all participants at S1 and S9 are included in Appendix 14.

5.10 Statistical analysis

Further statistical analysis and visualisation of data were conducted on Microsoft Excel and R studio to assess the relationship between the samples in the study, namely the alpha diversity and beta diversity. The “phyloseq” package was used in the R software environment.

5.10.1 Alpha diversity

Alpha diversity refers to the measurement of species richness and evenness within each sample. Species richness refers to the number of different species present in a sample but neglects the relative distributions or abundance of the species.

Measurement of species evenness considers the relative abundance of different species that make up the richness.

5.10.1.1 Chao1, Shannon and Simpson Indices

Chao1 which is an appropriate measurement for abundance data makes assumption that the number of organisms that are identified for a taxa has a Poisson distribution and corrects for variance. It is appropriate for analysing mycobiome data as it is useful for data sets skewed toward low-abundance species which is typical for microbes. It only estimates the total richness.

The Shannon index assumes that all species are randomly sampled and represented in a sample and summarises the diversity in a population. An increase in Shannon index reflects an increase in both richness and evenness of the mycobiome community.

The Simpson Index takes into account the number of taxa and abundance in its measurement of diversity. It gives more weightage to the more dominant species which means that the diversity of the sample will not be affected by a few rare species with few representatives. It accounts for proportion of species in a sample.

5.10.1.2 Wilcoxon rank sum test

The Wilcoxon rank sum test is a non-parametric statistical test which can be used to investigate whether two independent cohorts consist of samples from populations that have the same alpha diversity distribution. The null hypothesis would be that the probability that a randomly selected value from one cohort is lower than a randomly selected value from the other cohort is equal to the probability of being greater.

P-values of less than 0.05 would suggest that the null hypothesis can be rejected and confirms that the samples from the two cohorts are selected from populations with varying distributions of alpha diversity.

5.10.2 Beta diversity

Beta diversity on the other hand, is a comparison of samples to each other and how different they are. It measures the distance or dissimilarity between each sample pair using a distance matrix, where the input matrix is relative abundance to reflect the fungal composition of the samples.

5.10.2.1 Binary Jaccard Index

The Binary Jaccard Index measures how dissimilar two sets of data are but does not include any information about its abundance. Values are set from 0 to 1. A value of 0 informs us that both samples share the same exact species whereas a value of 1 shows that both samples have no species in common.

5.10.2.2 Bray-Curtis Index

Bray-Curtis Index is a non-phylogenetic metric that takes abundance into account and is used to measure dissimilarity between two different sites/samples. It is calculated with the following formula:

$BC_{ij} = 1- 2C_{ij} / S_i + S_j$

Where i & j are the two sites, S_i is the total number of species counted on site i, S_j is the total number of species counted on site j, and C_{ij} is the sum of only the lesser counts for each species found in both sites. The Bray-Curtis dissimilarity is set at values of 0 to 1 with 0 being the two samples sharing all the same species and a value of 1 is when the two separate samples do not share any similar species.

5.10.2.3 Permutational Multivariate Analysis of Variance (PERMANOVA)

Permutational Multivariate Analysis of Variance (PERMANOVA) is a non-parametric multivariate statistical test which is based on a prior calculation of the distance between any two groups included to the Principal Coordinate Analysis (PCoA) of beta diversity.

6 Results and Analysis

6.1 Samples

Saliva samples included in the study were collected at eight separate time-points, beginning from September 2018 up to June 2019, as noted in Figure 5. The time points were labelled S1 to S9 with S8 being excluded as no saliva samples were collected in May 2019. Figure 5 also marked the number of days participants were in school during school terms and the dates of their school holidays. For every school holiday period, boarders would return to their home.

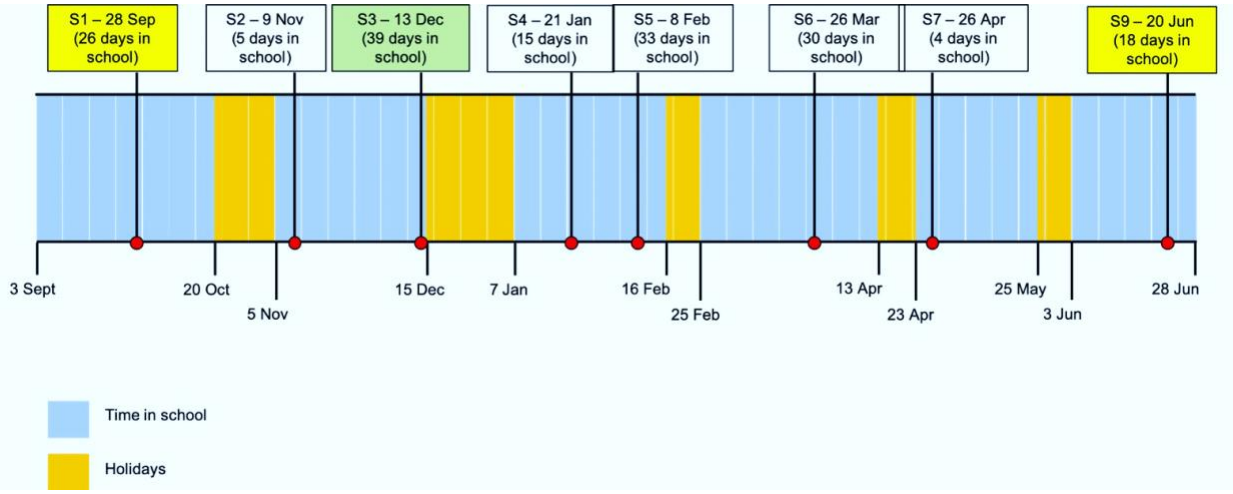


Figure 5: Timeline of sample collection included in the study (September 2018 – June 2019)

For the purpose of this study, 17 participants were recruited which comprised of 13 boarders and 4 non-boarders, of whom 13 were males and 4 were females. They had an average age of 12.23 years. They were each assigned to a specific identification code which were consistent throughout the study, in the form of H* (H1 to H18). No saliva sample was collected from H10 so this participant was excluded from the study.

135 saliva samples in total were collected from the 17 participants between September 2018 to June 2019 and sent for sequencing. 7 saliva samples were unable

to be collected. The saliva samples collected from the participants are depicted in Table 10. A cross meant that saliva sample was collected at that time period. A shaded cell meant that no saliva sample was collected. A blank box meant that saliva samples were taken but not included for sequencing. A total of 10 saliva samples were excluded from sequencing. This was due to the limitations of the Swift Amplicon 16S+ITS Panel, which could only run up to 96 samples. Only two kits were available and this had to be shared with two other projects which were running concurrently in the laboratory at this time.

Participant	Group	Gender	School year	S1	S2	S3	S4	S5	S6	S7	S9	S10
H1	NB	F	7	X	X	X	X	X	X	X	X	X
H2	B	F	7	X	X	X	X	X	X	X	X	X
H3	B	M	7	X	X	X	X	X	X	X	X	X
H4	NB	M	7	X	X	X	X	X	X	X	X	X
H5	NB	M	7	X	X	X	X	X	X	X	X	X
H6	B	M	7	X	X	X	X	X	X	X	X	
H7	B	F	12	X	X	X	X	X	X		X	
H8	B	M	7	X	X	X	X	X	X	X	X	
H9	B	M	7	X	X	X	X	X	X	X	X	
H11	B	M	9	X	X	X	X	X	X	X	X	
H12	B	M	9	X	X	X	X	X	X	X	X	
H13	B	F	9	X	X	X	X		X		X	
H14	B	M	9	X	X	X	X	X	X	X	X	
H15	B	M	9	X	X	X	X	X	X	X	X	
H16	NB	M	7	X	X	X	X	X	X	X	X	
H17	B	M	9	X	X	X	X	X		X	X	
H18	B	M	9	X	X	X	X	X		X	X	

Table 10: Saliva samples collected from participants

6.2 Oral health questionnaire

The oral health questionnaire were given and completed by the participants at the start of the study, in September 2018. Boarders were made up of 13 students (H2, H3, H6, H7, H8, H9, H11, H12, H13, H14, H15, H17, H18), 3 were females and 10 were males. The 4 non-boarders (H1, H4, H5, H16) were made up of 1 female and 3 males. The average age of the boarders was 12.6 years whereas non-boarders were all 11 years of age.

GROUP	No of children, (%)	Gender	Age in years (mean)
Boarders	13 (76%)	Female = 3 Male = 10	12.6
Non- Boarders	4 (24%)	Female = 1 Male = 3	11

Table 11: Distribution of boarders and non-boarders

In terms of country of origin, a large majority of participants were from United Kingdom and made up 70.5% of the total participants. Two participants (H6 & H12) came from Croatia and were also siblings, one from India (H7), another from Germany (H13) and one originated from Malaysia (H15). Participant H8 and H17 were also siblings.

Country of origin	No of children, (%)
United Kingdom	12 (70.5%)
Croatia	2 (11.76%)
India	1 (5.9%)
Germany	1 (5.9%)
Malaysia	1 (5.9%)

Table 12: Country of origin of participants

Fifteen participants were owners of pets where seven owned cats, five had dogs, two with hamsters and one owned a fish. All except two participants (H6, H14) played sports and have mouthguards. All participants brushed their teeth twice daily and six of them (H5, H7, H9, H13, H14, H16) use a mouthwash. Four participants are wearing fixed orthodontic appliances (H6, H12, H13, H18) and none have a false tooth or wearing a denture to replace a missing tooth. Probiotic drinks are taken by participant H5 and H16. In terms of medical history, only one participant (H5) reported having asthma and three participants (H3, H5, H6) have eczema.

Participant	Own pet (Y/N)	Own mouthguard (Y/N)	Brush teeth twice daily (Y/N)	Use mouthwash (Y/N)	Fixed orthodontic appliance (Y/N)
H1	N	Y	Y	N	N
H2	Y (dog)	Y	Y	N	N
H3	Y (dog)	Y	Y	N	N
H4	Y (hamster)	Y	Y	N	N
H5	Y (cat, fish)	Y	Y	Y	N
H6	N	N	Y	N	Y
H7	Y (dog)	Y	Y	Y	N
H8	Y (cat)	Y	Y	N	N
H9	Y (cat)	Y	Y	Y	N
H11	Y (cat)	Y	Y	N	N
H12	N	Y	Y	N	Y
H13	Y (cat)	Y	Y	Y	Y
H14	Y (hamster)	N	Y	Y	N
H15	Y (dog)	Y	Y	N	N
H16	Y (cat)	Y	Y	Y	N
H17	Y (cat)	Y	Y	N	N
H18	Y (dog)	Y	Y	N	Y

Table 13: Oral habits of participants and whether they own pets

Participant	Probiotics (Y/N)	Asthma (Y/N)	Eczema (Y/N)
H1	N	N	N
H2	N	N	N
H3	N	N	Y
H4	N	N	N
H5	Y	Y	Y
H6	N	N	Y
H7	N	N	N
H8	N	N	N
H9	N	N	N
H11	N	N	N
H12	N	N	N
H13	N	N	N
H14	N	N	N
H15	N	N	N
H16	Y	N	N
H17	N	N	N
H18	N	N	N

Table 14: Participants' medical history and whether they take probiotics

6.3 Clinical examination

The International Caries Detection and Assessment System (ICDAS) which is a clinical scoring system was used for measurement of caries experience in participants. Three participants (H1, H8, H18) had initial caries in one of their permanent first molar tooth (two uppers and one lower), denoted by score A. One participant (H2) had two permanent first molars which had an ICDAS score of 3 which represented localised enamel breakdown with no dentine exposure. H2 was also the only participant with an existing restoration on one permanent first molar (upper right). No teeth were found to have an underlying dark shadow from dentine, distinct, or extensive cavity with visible dentine detected which is represented by ICDAS scores 4, 5 or 6, respectively.

The average plaque score of the participants was 0.125. The highest plaque score was 1.4, from participant H2 followed by 1.0 from participant H9. Three participants (H5, H6, H7) had a plaque score of 0.6, followed by 0.3 in participant H15 and 0.1 in participants H8, H12 and H18. The lowest plaque score was 0 which was reported in seven participants (H1, H3, H11, H13, H14, H16, H17).

The Pulp, Ulceration, Fistula & Abscess (PUFA) Index examination only noted one ulcer lesion, present in participant H15.

6.4 Taxonomic analysis

6.4.1 Phylum level

At phylum level, samples primarily consisted of fungi from the Ascomycota and Basidiomycota phyla with Ascomycota being the most abundant phylum at 57%, followed by Basidiomycota at 12%. They both accounted for 69% of the total phyla observed. Missing taxonomic information in the database accounted for the remaining fungal OTUs being unclassified and were labelled as k_Fungi.

6.4.2 Class to Genus level

The fungi taxa found in samples which are able to be classified from class level up to genus level are listed in Table 15.

Taxonomic rank	Classified fungi taxa found in sample
Class	Agaricomycetes, Xylonomycetes, Tremellomycetes, Pezizomycetes, Eurotiomycetes
Order	Symbiotaphrinales, Agaricales, Cantharellales, Tremellales, Pezizales, Boletales, Eurotiales

Family	*Symbiotaphrinaceae, Strophariaceae, Tulasnellaceae, Bulleribasidiaceae, Pyronemataceae, Helvellaceae, Trichocomaceae*
Genus	*Symbiotaphrina, Galerina, Lamprospora, Helvella, Talaromyces*
Species	*Galerina_triscopa, Symbiotaphrina_buchner*

Table 15: Taxonomic rank of classified fungi taxa

However, as the taxonomic rank increase in specificity, the classified fungi taxa found in total samples decrease exponentially in their relative abundances. At Class level, the five classified fungi taxa identified only made up 13% of the total, with Agaricomycetes being the most abundant class (10.8%), followed by Xylonomycetes (2.1%). Tremellomycetes, Pezizomycetes and Eurotiomycetes made up the rest of the remaining abundance total. Three unclassified fungi Class made up 87% of the total. At genus level, even though five generas were able to be identified, they only made up 2.7% of the total relative abundance while the rest were eleven unclassified generas. At species level, the two classified species only accounted for a measly 0.7% of the total abundance and fifteen unclassified species made up the rest.

6.4.3 Abundance level between boarders and non-boarders at S1 and S9

Phyla abundance between boarders and non-boarder were compared at two different sampling points, S1 and S9 and shown in Figure 6. These two sampling points were specifically chosen to compare the initial mycobiome composition (S1) and any changes over the period of ten months due to the different environments the two groups (boarders and non-boarders) were exposed to. In both groups, both Ascomycota and Basidiomycota phyla showed a slight decrease when comparing

between sampling points S1 and S9, while the unclassified fungi showed a slight increase.

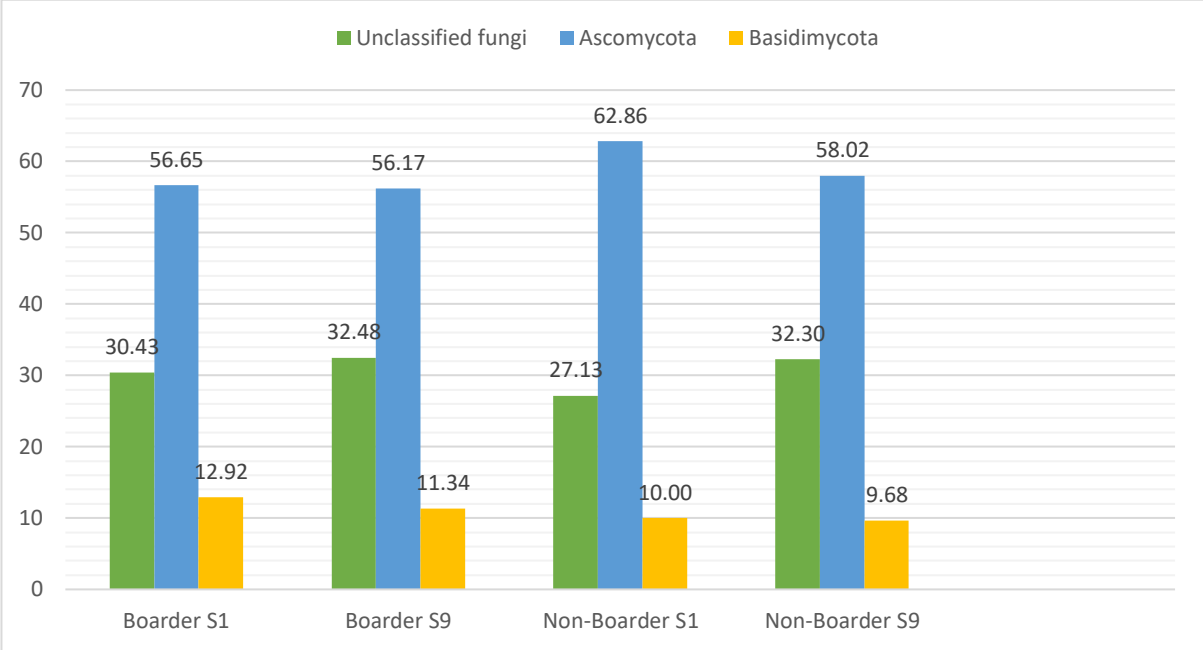


Figure 6: Mean abundance at phyla level for boarders and non-boarders at sampling points S1 and S9

6.4.4 Comparison of two pairs of siblings at S1 and S9

Amongst the 17 participants in this study were two pairs of siblings, H6 and H12 as well as H8 and H17. The oral fungi composition and abundance were compared among the respective siblings to determine whether a shared environment over a period of ten months would have a similar effect on the development and stabilisation of their fungi level. However, no significant correlations or trends were found in both pair of siblings.

6.5 Alpha diversity

6.5.1 Alpha diversity between boarders and non-boarders

Species richness within the boarders (labelled “Yes”) and non-boarders (labelled as “No”) were measured using four different alpha diversity measurements, namely observed OTUs, Chao1, Shannon and Simpson indices using the “Phyloseq” package in R studio, which are shown as boxplots with associated p-values in the

Figure 7. For each measurements, pairwise comparisons were carried out using Wilcoxon rank sum test with continuity correction data and holm method was used for p value adjustment. Outliers are represented by black points.

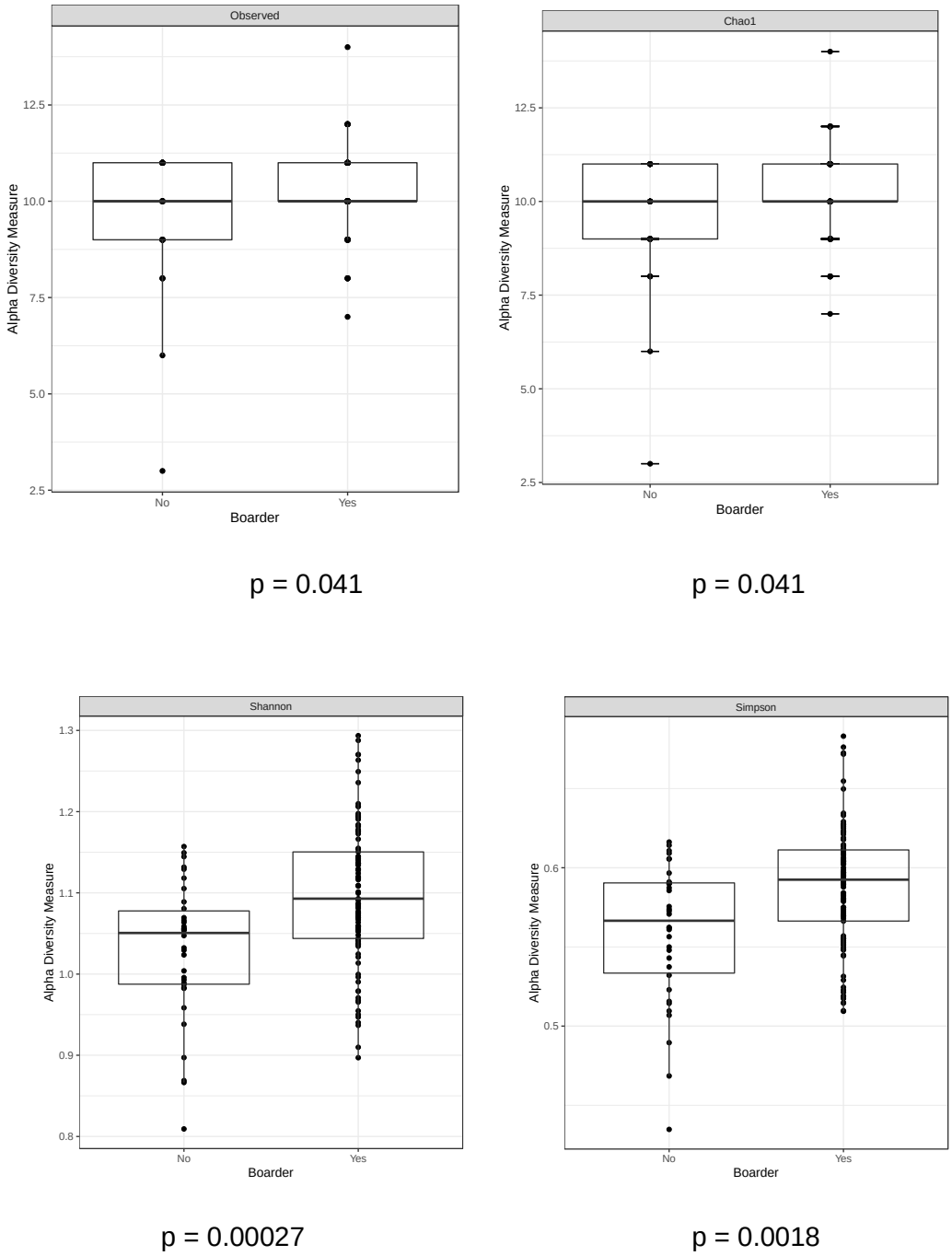


Figure 7: Alpha diversity measurements (observed OTUs, Chao1, Shannon and Simpson Indices) of boarders and non-boarders

The different alpha diversity measurements account for their own unique assumptions. Chao1 index estimates the number of species but are sensitive to rare OTUs while Shannon Index estimates the effective number of species and assumes that all species are represented in a sample and are randomly sampled. On the other hand, the Simpson Index accounts for the proportion of species in samples, accounts for number of taxa and abundance and gives more weight to more common species. All four metrics showed statistically significant results with p-values of <0.05. The OTU and Chao1 plots are exactly alike, meaning that the richness in each sample and total samples of the boarders group are higher than those in the non-boarders group. Both the plots of the Shannon and Simpson Indices showed that the boarder group has a higher species richness (number of taxonomic group) and evenness (distribution of abundances of the taxonomic groups) compared to non-boarders group.

6.5.2 Alpha diversity between at S1 and S9

Comparison of the alpha diversity between the initial saliva collection point (S1) and the last collection point which was ten months later (S9), revealed no significant change in oral mycobiome diversity between these two points of interest (p>0.01).

6.6 Beta diversity

Several methods were used to compare the mycobiome composition in the two separate groups.

6.6.1 Principal Coordinates Analysis (PCoA)

Bray Curtis dissimilarity matrix was used as input for Principal Coordinates Analysis (PCoA) which is also known as metric multidimensional scaling (mMDS). PCoA performs an Eigen analysis on the data, summarises and attempts to represent inter-dimensional dissimilarity which is visualised in Figure 8. Interpretation of a PCoA plot

is fairly straightforward whereby objects which are ordinated closer to one another have smaller dissimilarity values than those ordinated further away.

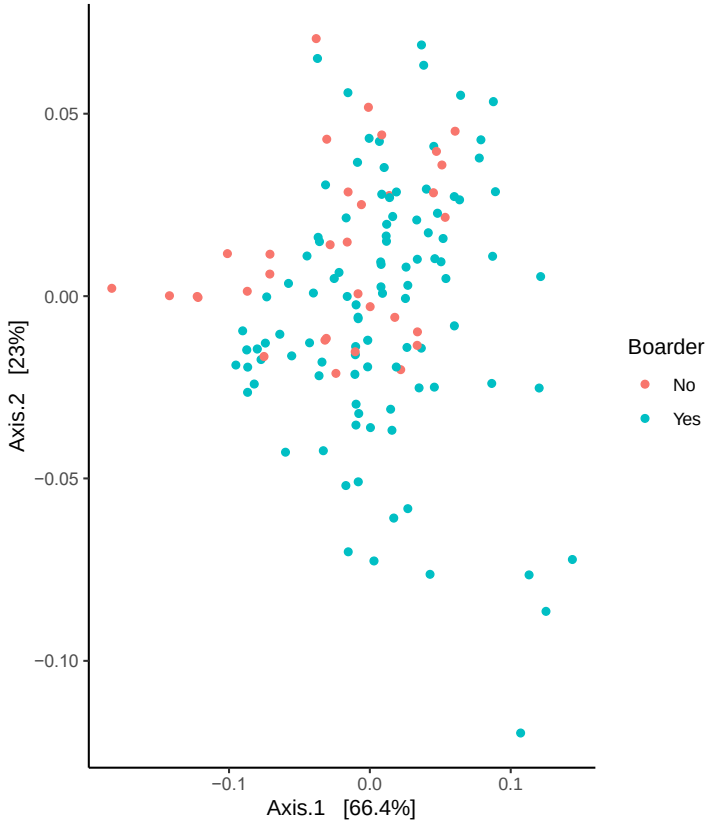


Figure 8: Principal coordinate analysis (PCoA) ordination of a Bray-Curtis dissimilarity matrix

From the plots recorded in Figure 8, although some of the plots representing the boarders seem to be concentrated in the middle of the graph, no solid trends were found in both boarder and non-boarder groups that would suggest a strong dissimilarity values of one group in comparison to the other.

6.6.2 Non-parametric multidimensional scaling (NMDS)

NMDS is an approach that uses indirect gradient analysis which produces an ordination that is based on a distance or dissimilarity matrix. It attempts to represent pairwise dissimilarity between objects in a low-dimensional space as closely as possible. It also uses any dissimilarity or distance measures as input. In the

measurement of NMDS, Jaccard Index was used as input and outcome is represented in Figure 9.

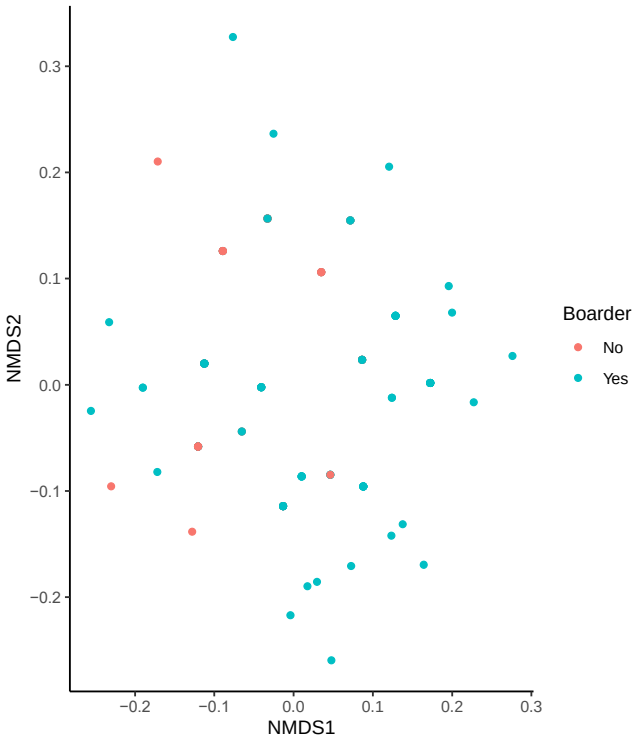


Figure 9: Non-parametric multidimensional scaling (NMDS) between boarders and non-boarders

Similar to the PCoA plot in Figure 8, the NMDS plots in Figure 9 did not portray a strong enough result to indicate that the participants in the boarders group share similar mycobiome composition.

6.6.3 Permutational Multivariate Analysis of Variance (PERMANOVA)

The null hypothesis for Permutational Multivariate Analysis of Variance (PERMANOVA) is that there are no differences in the presence/absence or relative magnitude of a set of variables among objects from different groups or treatments. The p-value for boarders was 0.02 (<0.05) which meant that their environment had a

significant effect on their fungi composition. Pairwise PERMANOVA calculations for boarders and non-boarders both gave the same p and p-adj values of 0.035.

6.6.4 Analysis of Similarity (ANOSIM) & Pairwise ANOSIM

Analysis of Similarity (ANOSIM) test poses the null hypothesis that there is no difference between the means of two or more groups of (ranked) dissimilarities. ANOSIM test is similar to that of ANOVA-like hypothesis test but is used to evaluate a dissimilarity matrix rather than raw data. The ANOSIM statistic compares the mean of ranked dissimilarities between groups to the mean of ranked dissimilarities within groups. An R value close to 1.0 suggests dissimilarity between groups. In contrast, an R value close to 0 suggests an even distribution of high and low ranks within and between groups. Values of R below 0 suggest that dissimilarities are greater within groups than between groups. Comparisons between boarder and non-boarder groups produced an R value of 0.06141 with a p-value (significance level) of 0.039. The pairwise ANOSIM of boarders and non-boarders were found to be the same to the results of pairwise PERMANOVA, with both having p values of 0.035.

7 Discussion

The demographic and characteristic of the 17 participants included in this study were 11 to 16 years of age with a mean age of 12.23, 13 males and 4 females, one with asthma and three with eczema. Based on the caries assessment of all participants using the International Caries Detection and Assessment System (ICDAS) scoring system, three participants had initial caries (Score 1) in one of their permanent first molar tooth. One participant had an ICDAS score of 2 on two permanent first molars which is described as a distinct visual change in enamel but no cavitation. Only one participant had one restoration in one of their permanent first molar. The Pulp, Ulceration, Fistula & Abscess (PUFA) Index examination was clear for all participants with the exception of one ulcer lesion in one participant. Using the mean age of the cohort, comparisons were made to the Children's Dental Health Survey (CDHS) 2013 of children in the 12 year old age bracket. In the CDH survey, the mean caries rate of 12 year old which was defined as an obvious decay experience was 34% and the 19% had caries into dentine of permanent teeth. This was well above the mean caries rate for the cohort in the study which was calculated to be 23.5%. No participants had any caries which progressed into dentine. The mean DMFT for the 12 year olds in the CDH survey was 0.8 and the mean number of teeth affected was 2.5. In comparison, the participants in this study had a much lower average DMFT score of 0.29 and mean number of teeth affected was 0.3. However, due to a low sample size, it is difficult to ascertain this as a statistically significant results.

Another important aspect to consider when comparing the findings in this study to the results in the CDH survey is the effects of socioeconomic status on caries status and severity. The CDH survey had an oversample of students from schools with higher rates of free school meal eligibility so that they could compare the children from lower

income families. with the inclusive criteria being children who were eligible for free meals in 2013 with other children who are of the same age. In 2013, a free school meal was one of the statutory benefits that were offered to families who are on other qualifying benefits, such as income support. This comparison was carried out to evaluate any similarities and differences in terms of oral health, attitudes, perceptions and experiences related to the oral health of each respective groups. It was found that socioeconomic status did play a part in this but was also compounded by certain demographic characteristics such as ethnicity and country of birth. As this study was conducted in Haileybury school, a private boarding school, the cohort of this study are more likely to come from a higher socioeconomic background which would explain the low incidence of caries and reiterated the findings in the CDH survey. The CDH survey also noted that socioeconomic status also affected oral hygiene status and this was confirmed by a low mean plaque score of 0.13 of the cohort in the study.

7.1 Findings of study

7.1.1 Phylum level

At phylum level, 69% of the total phyla observed could be identified and this consisted of Ascomycota and Basidiomycota phyla. Ascomycota was the most abundant phylum at 57% and Basidiomycota accounted for 12% of the total phyla observed. This is consistent with the phyla observed in healthy children (Fechney et al., 2019). Ascomycota was also the highest phylum found in all subjects in studies related to oral mycobiome by James et al. (2020) and Schei et al. (2017).

7.1.2 Class to genus level

As the taxonomic classification became more specific, from Class level to genus level, the classification of the classified taxonomic fungal taxa found in samples significantly differed from previous studies. The number of classified taxa also dropped to 13% at Class level and at species level, only 2 out of 17 taxa were

identified (*Galerina_triscopa, Symbiotaphrina_buchner*) making up 0.7% of the total abundance. These taxa were not noted in previous studies on oral mycobiome and their presence could be attributed to them being inhaled into the oral cavity or from food ingested. Another likelihood is possible contamination of samples by spore producing mycobiome.

7.1.3 Lack of *Candida* genus

A surprising find from the taxonomic analysis was the lack of presence of *Candida* genus in any of the samples which does not correlate with the results of previous studies on oral mycobiome. Although it can be argued that a low sample size could account for this, a previous study analysing the oral mycobiome in children with and without caries with the similar sample size (although the mean age is lower than this study), *Candida* was found in over 50% of samples in both children and without caries (Fechney et al., 2019). Hence, it is possible that *Candida* spp are present in the samples but are not classified due to the type of classifier used.

7.1.4 Alpha diversity

The different measurements of alpha diversity within the boarders group and the non-boarders group used, specifically observed OTUs, Chao1, Shannon and Simpson indices all revealed that the boarders group has a higher species richness and evenness, which are the number of taxonomic groups and distribution of abundances of the taxonomic groups, respectively. All four measurements were statistically significant, with p-values of <0.05.

These findings could be explained by boarders experiencing a more frequent change in their living environment compared to non-boarders. The different environments experienced by them would be a daily change between school setting and living environment with other boarding students during school term to a home environment during school holidays. However, these results could potentially be due to a higher

number of participants in the boarders group (n=13) in comparison to the non-boarders (n=4) which would skew the results towards the boarders.

7.1.5 Beta diversity

The beta diversity of the boarders and non-boarders were measured using several methods. The results of the Principal Coordinates Analysis (PCoA) with the use of Bray Curtis Index showed no definitive trends on the plots on the graph which meant that the boarders group did not show a lower dissimilarity compared to the non-boarders group, which is a result contrary to what is expected. Another method, non-parametric multidimensional scaling (NMDS) also presented similar findings to PCoA.

However, these findings contradicted the finding from the Permutational Multivariate Analysis of Variance (PERMANOVA) calculation which showed a p-value of 0.02 (<0.05) for boarders group which meant that their environment had a significant effect on their fungi composition. Pairwise PERMANOVA calculations for boarders and non-boarders both gave the same p and p-adj values of 0.035 which meant that there is no difference in the composition of the two groups of samples which reflected the results of PCoA and NMDS analysis.

Analysis of Similarity (ANOSIM) test comparing boarder and non-boarder groups was carried out and produced an R value of 0.06141 with a p-value of 0.039. Although the p-value is significant which suggests that the null hypothesis can be rejected, the low R value (close to 0) is more meaningful to this calculation and confirmed the null hypothesis that there were no differences in between the dissimilarity means of the boarders and non-boarders and within their own groups and an even distribution of high and low ranks within and between the two groups.

Overall, it can be deduced that there is low beta diversity when comparing between the two groups, meaning that they both show high level of similarity of oral mycobiome diversity.

7.2 Limitations of study

Looking back at how the study was conducted, there were several shortcomings found which could be improved in future studies. First would be the sample size of the cohorts included in the study. The sample size of seventeen participants was quite small and also skewed heavily towards to the boarders group. This meant that any results or data analysis collected could potentially be affected by this and makes it challenging to interpret, even those that are supposedly statistically significant. For example, in the calculation of alpha diversity, boarders were found to have a higher species richness and evenness in comparison to non-boarders which could be explained by the effects of their changing living environment. However, it is also plausible that this is could be due to the higher number of boarder subjects. Based on these reasons, the results achieved from the analysis are found to be more indicative rather than definitive and would require further similar research in the future to prove their validity. The discrepancy in the distribution of boarders and non-boarders also makes it difficult to take into account considerations of the effects of other variables, such as age and gender. A higher number of participants with a more equal distributions between the two sample groups could provide a more confident analysis and interpretation of the results.

When analysing the characteristics of the samples, the number of male participants (n = 13) were at a much higher number compared to female participants (n = 4). As mentioned previously, age could be a potential variable that could affect the results of the study and hence, in future, a more even distribution of gender in both groups should be considered. Apart from that, a lot more information obtained from the oral

health questionnaire such as other dental condition or anomalies, dental appliance such as orthodontic brackets can be utilised to observe their significance on the findings. The small sample size limited the use of these variables in the analysis of the results in this study.

The lack of *Candida* genus in any samples as noted the taxonomic analysis was surprising and can be considered as a limitation of this study as this is a genus that is thought to be quite ubiquitous in adolescents. It is likely that its lack of presence can be attributed to issues during any of the steps in the methodology of this study.

Issues with increase number of unidentified taxa, absence of *Candida* genus and low similarity of oral mycobiome taxa compared to previous studies was discussed with a PhD student with knowledge in microbiology analysis who noted several issues that could account for these findings. Firstly is the discrepancy of the ITS Qiime2 workflow which was recommended and provided by Swift Biosciences (2019). The steps in this workflow for downstream analysis of ITS was carried out within the QIIME2 platform. Potential issues were noted in the workflow for importing fastq files containing Illumina paired-ends (PE) which sometimes cannot be merged into single reads for further analysis. Some solutions include just using the first reads for taxonomy annotation but can waste potential relevant information in the second reads. Another solution would be to classify the paired reads separately and combine the taxonomic annotations later. However, the tracking of these reads are difficult especially with the high amount of fastq files included in this study, not to mention the likelihood of this step in severely slowing down the downstream analysis. These suggested methods, although helpful, were beyond the scope of my current capabilities for these types of data analysis at this stage and was put on hold for now but noted down for consideration in future study designs.

The unexpected pause of sample collection from participants was another limitation of this study. As mentioned previously, the original plan of study was to collect and analyse samples over the span of two years. However, due to the pandemic, this was cut short to one year with only a maximum of nine samples collected from participants for further analysis. A period of one year could potentially be too short of a time to observe significant findings on the change in oral mycobiome over time. With regards to the collection point within the one year period, the gaps between each collection were also quite close to one another which meant that the effects of environment may not be accurately represented in the results. In retrospect, perhaps more precise calculations of the timing of sample collection could be determined beforehand to ensure that an adequate amount of time is given to accurately represent the effects of a shared environment.

As highlighted in a study by Vesty et al. (2017), DNA extraction methods had an impact on the diversity of oral fungi in samples and certain protocols for DNA extraction may not be suitable for the study of oral fungal microbiota. This could explain the high number of unclassified fungi and why *Candida*, a relatively common fungi, was not identified in the samples. The DNA extraction kit chosen, the PurElute™ Bacterial Genomic Kit (EdgeBio, USA) as the name suggested, is specific and perhaps optimised for bacterial DNA extraction and hence, could affect the extraction of DNA from fungal cultures. This was the only available DNA extraction kit available at the start of the study and decision was made at the time to proceed with this kit due to the initial aim of focusing on oral microbiota, specifically bacteria, while the decision to expand the study to include oral fungal microbiota was decided later. Nevertheless, a rich dataset was able to be obtained for both oral bacteria and fungi. No studies were found that included the use of this DNA extraction kit, specifically for DNA extraction from fungal samples, so it difficult to ascertain the efficacy of this kit and its limitations.

Due to the low number of studies to date on oral fungal microbiota in comparison to oral bacterial microbiota, the amount of information available in the process of bioinformatics ITS analysis was relatively limited. This prolonged the analysis process which admittedly was difficult, with a lot of trial and errors and a lot of adaptations of the steps outlined specifically for 16S rRNA analysis to fit the analysis of fungal ITS sequences.

7.3 COVID-19 Impact Statement

The declaration from World Health Organisation (WHO) of COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020, characterisation of it as a pandemic on 11 March 20202 and associated restrictions has caused some significant impact on this study. The study was initially planned to be completed by collection of samples from participants over the span of two years. At the beginning of the study in September 2018, 17 students were recruited, consisting of 13 boarders and 4 non-boarders. There were 13 males and 4 females with an average age of 12.23 years. On September 2019, a new cohort with a total 9 students were added to the study, of which 7 were boarders and 2 were non-boarders. There were 5 males and 4 females with an average age of 13.8 years. Due to the pandemic, no samples were able to be collected after December 2019. Due to the limited number of samples taken, the second cohort was excluded from the study and total number of participants were down to 17 from 26. This also meant that the original plan of follow up of participants over two years was reduced to less than one year which affected not only the aims of the study but potentially the quality and significance of the results as well.

The handling of samples was also impacted by the Covid-19 announcement. Due to Covid-19, University College London (UCL) changed the classification of saliva as a

Category 3 biosafety hazard (Cat 3). This meant that any saliva samples would be required to be handled in a biosafety Level 3 facility (Cat 3 facility). The classification of samples and their containment are described in Table 16.

Sample Type	Containment Level of Respiratory Samples
Confirmed or presumptive COVID-19 positive samples	Level 3
Samples from people of unknown status	Level 3
Convalescent or recovered COVID-19 patient samples (without negative test results)	Level 2 (If >42 days after onset of symptoms AND if asymptomatic)
Confirmed COVID-19 negative samples	Level 2
Inactivated samples	Level 2

Table 16: Sample types and containment levels

Biosafety level 3 (BSL-3) encompassed all clinical, teaching, research, diagnostic and production facilities whereby work is carried out with agents which have the potential to contaminate the environment and cause serious or lethal diseases by way of inhalation to the persons. These restrictions meant that saliva samples were unable to be collected from participants or processed unless they reported a negative PCR result for COVID-19 particularly because we did not have access to any Category 3 facilities to store and process these types of samples. Therefore, there were 39 samples which had already been collected but not processed which unfortunately, had to be excluded from the study.

Another setback to this study from the COVID-19 pandemic was that Haileybury School had to close and hence, boarders had to return home to their families. This

interrupted our planned sampling times as we did not access to these students to collect their saliva samples. Furthermore, their unexpected change in environment for the unforeseeable future at the time could have an unintended effect on the aims and results of our study.

Based on the limitations of this study which was caused by the unexpected COVID-19 pandemic, the study length had to be shortened and analysis was carried out based on the samples that we had on hand. Any results achieved were analysed with the acknowledgement of the low sample size and reduced timeline of sample collection. A systematic review was also added on alongside this study to investigate other host and environmental factors which are associated with oral or gut mycobiome, specifically in children. It was thought that this systematic review would be helpful in moving our future studies on mycobiome in children in the right direction.

8 Future work

Even with its limitations, this study characterised the potential for further research into the effects of environment on the oral mycobiome of adolescents, which is an area that is currently still understudied but has a lot of potential to reveal some insights in the development and change in oral mycobiome in children and young adults and its effects on their health and disease. An understanding of the diversity of oral mycobiome in healthy children could be beneficial in the detection of specific fungal species associated with certain diseases, such as caries, diabetes, irritable bowel syndrome (IBS) to name a few, as well as the effects of certain environments such as diet or specific lifestyles on the presence and change in oral mycobiome. It may also have practical applications in the diagnosis, characterisation and monitoring of progression of different diseases. The fact that oral mycobiome still considered an understudied area in microbiology makes it an exciting endeavour with a lot of potential for meaningful findings. Further research should be considered to further clarify any importance in the role of oral mycobiome in adolescents.

With the results of this study, which is considered as a pilot investigation in to this area of interest, future plans for this project would be to continue to investigate the effects of the environment on the oral mycobiome of adolescents by replicating the study with a larger cohort of students with diverse age groups, and incorporating a more similar distribution of boarder and non-boarder groups as well as male and female participants. A longer study period of at least two years is planned, but with limiting sampling points to three times a year. This will be set to the beginning, middle and end of the year to allow participants to stay in a specific type or types of living condition long enough to allow time for the environment to potentially affect their oral mycobiome. These are warranted to improve the power of the dataset and potentially produce more reliable and significant results and limit the effects of undesired

variables. A state school with a mix of boarders and non-boarders, linked to Haileybury School is considered to assess whether similar results could be reproduced, in addition to answering our initial question of whether the environment had any effects on oral mycobiome. The results from this school could also be compared to that of Haileybury to assess for any similarities or differences and what environmental effects could potentially account for that.

After several discussions with a PhD student during the challenging data analysis stage, it was suggested that it would be wise to include the help of a PhD student or someone with more experience in the study of bioinformatics in the early stages of the study to ascertain whether the steps followed in the sample collection, DNA extraction and sequencing, and finally analysis, were done in the correct manner and following certain standards to avoid any errors that could possibly affect the results of the study. Apart from that, a more well-known kit for the next generation sequencing (NGS) analysis instead of the current one (Swift Amplicon 16S+ITS Panel) could be considered to avoid the problems encountered in this study during the analysis process of the samples. The use of the same kits as those used in other studies of oral mycobiome could help lower any potential confounding factors which could affect the results of the study.

9 Conclusion

The systematic review conducted showed the different host and environmental factors that could potentially be associated with oral or gut mycobiome of adolescents. Specific host and environmental factors, such as caries, age, diet and obesity were found to have a significant effect on the levels of oral/gut mycobiome, specifically *Candida.* However, there were other studies which contradicted these findings and found no differences in oral/gut mycobiome composition, alpha and beta diversity. This further proves our initial knowledge about the lack of substantive research in this area and the need for further investigations. The outcome of this systematic review is useful in providing information about presence of specific oral and gut mycobiome in health and specific diseases. It also highlighted other potential host and environmental factors which have yet to be explored in substantial studies.

In the current study, a report of the oral mycobiome of adolescents who stayed in boarding schools compared to adolescents who stayed at home over a period of 10 months was presented. The strength in the study lies in it being the first study, to our knowledge, that has looked at the effects of living environment on the oral mycobiome of adolescents.

Two phyla were found in both boarder and non-boarder groups, the most abundant being Ascomycota, followed by Basidiomycota, which represented 69% of total phyla observed and correlated with findings of previous studies on both healthy and diseased participants. The presence of Agaricomycetes, Xylonomycetes, Tremellomycetes, Pezizomycetes, and Eurotiomycetes fungi at Class level differed greatly from other studies on oral mycobiome. These were also present in very low amounts, making up only 13% of the total. Only 2 of 17 taxa (*Galerina_triscopa, Symbiotaphrina_buchner)* were able to be classified at species level, which only

accounted for 0.7% of total abundance. Although these results were not what was expected, they represent the need for more careful consideration of the extraction process and ITS analysis of oral mycobiome samples. There were multiple steps involved in the bioinformatics ITS analysis and any possible error in the steps could have affected the final results. Application and training of different classifiers to achieve a higher percentage of assigned fungi or trying out other databases other than UNITE database could be considered as well to assess detection of specific oral mycobiome that were commonly found in other oral mycobiome studies, such as *Candida*.

This study revealed that a shared living environment, which was represented by those in the boarders group, showed higher alpha diversity although this data is thought to be more indicative rather than definitive due to the low number of participants. Boarders group had higher species richness and evenness in comparison to those who lived at home during school term. Calculation of the beta diversity however, mostly showed low dissimilarity of diversity between the two groups. No similarities were found in the composition and diversity of oral mycobiome of the siblings in this study.

This study served as a pilot investigation into the potential effects of the living environment on the oral mycobiome of adolescents and helps us to tailor our future study into this area. A comparison can be made from these results to future investigations with similar cohort groups to ascertain whether our results were accurate or just anomalous findings. Furthermore, the difficulties faced during the taxonomic and statistical analysis of these data help us to better prepare ourselves, specifically in terms of the study design process of our future study to avoid such issues.

References

AAS, J. A., BARBUTO, S. M., ALPAGOT, T., OLSEN, I., DEWHIRST, F. E. & PASTER, B. J. 2007. Subgingival plaque microbiota in HIV positive patients. *J Clin Periodontol,* 34**,** 189-95.

BAKER, J. L., BOR, B., AGNELLO, M., SHI, W. & HE, X. 2017. Ecology of the Oral Microbiome: Beyond Bacteria. *Trends Microbiol,* 25**,** 362-374.

BEVERLY LLC, Q. 2020. SparQ Universal Libnrary Quant Kit manual. Accessed 23rd Novermber 2021. https://www.quantabio.com/wp-content/uploads/2021/12/IFU-130.1_REV_01_95210_sparQ_Universal_Library_Quant_Kit_1020.pdf

BEENA, M. S., PEEDIKAYIL, F. C., GUFRANAFMED, M. B., CHANDRU, T. P., SONI, K. & DHANESH, N. 2017. Comparison of Candida species isolated from children with and without early childhood caries: A descriptive cross-sectional study. Journal of the Indian Society of Pedodontics & Preventive Dentistry, 35, 296-300.

BENDER, J. M., LI, F., PURSWANI, H., CAPRETZ, T., CERINI, C., ZABIH, S., HUNG, L., FRANCIS, N., CHIN, S., PANNARAJ, P. S. & ALDROVANDI, G. 2019. Early exposure to antibiotics in the neonatal intensive care unit alters the taxonomic and functional infant gut microbiome. Journal of Maternal-Fetal & Neonatal Medicine, 1-9.

BIOSCIENCES INC, S. 2019. q2_ITS Pipeline using Qiime 2 for processing amplicon sequencing data from Swift SNAP ITS1 prep kit. Accessed on 15th February 2021. https://github.com/swiftbiosciences/q2_ITS.git

BIOSCIENCES INC, S. 2019. Swift Amplicon 16S+ITS Panel data sheet. Accessed 23rd November 2021. http://www.biosystems.com.ar/archivos/folletos/429/16s.pdf

BOLYEN E, RIDEOUT JR, DILLON MR, BOKULICH NA, ABNET CC, AL-GALITH GA, ALEXANDER H, ALM EJ, ARUMUGAM M, ASNICAR F, BAI Y, BISANZ JE, BITTTINGER K, BRENJROD A, BRISLAWN CJ, BROWN CT, CALLAHAN BJ, CARABALLO-RODRIGUEZ AM, CHASE J, COPE EK, DA SILVA R, DIENER C, DORRESTEIN PC, DOUGLAS GM, DURALL DM, DUVALLET C, EDWARDSON CF, ERNST M, ESTAKI M, FOUQUIER J, GAUGLITZ JM, GIBBONS SM, GIBSON DL, GONZALEZ A, GORLICK K, GUO J, HILLMANN B, HOLMES S, HOLSTE H, HUTTONHOWER C, HUTTLEY GA, JANSSEN S, JARMUSCH AK, JIANG L, KAEHLER BD, KANG KB, KEEFE CR, KEIM P, KELLEY ST, KNIGHTS D, KOESTER I, KOSLIOCEK T, KREPS J, LANGILLE MGI, LEE J, LEY R, LIU YX, LOFTFIELD E, LOZUPONE C, MAHER M, MAROTZ C, MARTIN BD, MCDONALD D, MCIVER LJ, MELNIK AV, METCALF JL, MORGAN SC, MORTON JT, NAIMEY AT, NAVAS-MOLINA JA, NOTHIAS LF, ORCHANIAN SB, PEARSON T, PEOPLES SL, PETRAS D, PREUSS ML, PRUESSE E, RASMUSSEN LB, RIVERS A, ROBESON MS, ROSENTHAL P, SEGATA N, SHAFFER M, SHIFFER A, SINHA R, SONG SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, and Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9

BORGO, F., VERDUCI, E., RIVA, A., LASSANDRO, C., RIVA, E., MORACE, G. & BORGHI, E. 2017. Relative Abundance in Bacterial and Fungal Gut Microbes in Obese Children: A Case Control Study. Childhood Obesity, 13, 78-84.

CHILDREN'S HEALTH DENTAL SURVEY. 2015. Child Dental Health Survey 2013, England, Wales and Northern Ireland. National statistics, Survey. https://digital.nhs.uk/data-and-information/publications/statistical/children-s-dental-health-survey/child-dental-health-survey-2013-england-wales-and-northern-ireland#summary

COCK, P.J.A., FIELDS, C.J., GOTO, N., HEUER, M.L., RICE, P.M. 2009. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucl. Acids Res. (2010) 38 (6): 1767-1771. first published online December 16, 2009. doi:10.1093/nar/gkp1137 http://nar.oxfordjournals.org/content/38/6/1767$. Nat Rev Genet. 13(4):260-70. doi: 10.1038/nrg3182. PMID: 22411464; PMCID: PMC3418802.

DE JESUS, V. C., SHIKDER, R., ORYNIAK, D., MANN, K., ALAMRI, A., MITTERMULLER, B., DUAN, K., HU, P., SCHROTH, R. J. & CHELIKANI, P. 2020. Sex-Based Diverse Plaque Microbiota in Children with Severe Caries. Journal of Dental Research, 99, 703-712.

DEO, P. N. & DESHMUKH, R. 2019. Oral microbiome: Unveiling the fundamentals. *J Oral Maxillofac Pathol,* 23**,** 122-128.

DUPUY, A. K., DAVID, M. S., LI, L., HEIDER, T. N., PETERSON, J. D., MONTANO, E. A., DONGARI-BAGTZOGLOU, A., DIAZ, P. I. & STRAUSBAUGH, L. D. 2014. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of Malassezia as a prominent commensal. *PLoS One,* 9**,** e90899.

EL MOUZAN, M., WANG, F., AL MOFARREH, M., MENON, R., AL BARRAG, A., KOROLEV, K. S., AL SARKHY, A., AL ASMI, M., HAMED, Y., SAEED, A., DOWD, S. E., ASSIRI, A. & WINTER, H. 2017. Fungal Microbiota Profile in Newly Diagnosed Treatment-naive Children with Crohn's Disease. Journal of Crohn's & colitis, 11, 586-592.

FÁBIÁN TK, HERMANN P, BECK A, FEJÉRDY P, FÁBIÁN G. 2013. Salivary defense proteins: their network and role in innate and acquired oral immunity. Int J Mol Sci. 13(4):4295-320. doi: 10.3390/ijms13044295. Epub 2012 Apr 2. PMID: 22605979; PMCID: PMC3344215.

FECHNEY, J. M., BROWNE, G. V., PRABHU, N., IRINYI, L., MEYER, W., HUGHES, T., BOCKMANN, M., TOWNSEND, G., SALEHI, H. & ADLER, C. J. 2019. Preliminary study of the oral mycobiome of children with and without dental caries. *J Oral Microbiol,* 11**,** 1536182.

GHANNOUM, M. A., JUREVIC, R. J., MUKHERJEE, P. K., CUI, F., SIKAROODI, M., NAQVI, A. & GILLEVET, P. M. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog,* 6**,** e1000713.

GOMEZ A, NELSON KE. 2017. The Oral Microbiome of Children: Development, Disease, and Implications Beyond Oral Health. Microb Ecol. 73(2):492-503. doi: 10.1007/s00248-016-0854-1. Epub 2016 Sep 14. PMID: 27628595; PMCID: PMC5274568.

HIGGINS, J & GREEN, S. 2009. Cochrane Handbook for Systematic Reviews of Interventions. 10.1002/9780470712184.

JAMES, S.M., PHILLIPS, S., TELATIN, A., BAKER, D., ANSORGE, R., CLARKE, P., L, J. H. & S, R. C. 2020. Preterm Infants Harbour a Rapidly Changing Mycobiota That Includes Candida Pathobionts. Journal of Fungi, 6, 09.

KADIR, T., UYGUN, B. & AKYUZ, S. 2005. Prevalence of Candida species in Turkish children: relationship between dietary intake and carriage. Archives of Oral Biology, 50, 33-7

KROM BP, KIDWAIi S, TEN CATE JM. 2014. Candida and Other Fungal Species: Forgotten Players of Healthy Oral Microbiota. Journal of Dental Research. 93(5):445-451. doi:10.1177/0022034514521814

LAMONT RJ, KOO H, HAJISHENGALL G. 2018. The Oral Microbiota: Dynamic Communities and Host interactions. Nat Rev Microbiol. 16(12):745-759. doi: 10.1038/s41579-018-0089-x. PMID: 30301974; PMCID: PMC6278837.

LIBERATI, A., ALTMAN, D.G., TETZLAFF, J., MULROW, C., GØTZCHE, P.C., IOANNIDIS, J.P., CLARKE, M., DEVEREAUX, P.J., KLEIJNEN, J. and MOHER, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology, 62(10), pp.e1-e34.

LIF HOLGERSON P, HARVENIK L, HERNELL O, TANNER AC, JOHANSSON I. Mode of Birth Delivery Affects Oral Microbiota in Infants. J Dent Res. 2011 Oct;90(10):1183-8. doi: 10.1177/0022034511418973. Epub 2011 Aug 9. PMID: 21828355; PMCID: PMC3173012.

LOPEZ-GARCIA, A., PINEGA-QUIROGA, C., ATXAERANDIO, R., PEREZ, A., HERNANDEZ, I., GARCIA-RODRIGUEZ, A., GONZALES-RECIO, O. 2018. Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Front. Microbiol.* 9:3010. doi: 10.3389/fmicb.2018.03010

Marlina, E. 2021. The clinical and biological effects of the use of probiotic VSL#3 in patients with oral lichen planus. Doctoral thesis (Ph.D), UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/10128767

MOHER D, LIBERATI A, TETZLAFF J, ALTMAN D G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement BMJ 2009;339 :b2535 doi:10.1136/bmj.b2535

MONTEIRO-DA-SILVA, F., ARAUJO, R. & SAMPAIO-MAIA, B. 2014. Interindividual variability and intraindividual stability of oral fungal microbiota over time. *Med Mycol,* 52**,** 498-505.

MUKHOPADHYA, I., HANSEN, R., MEHARG, C., THOMSON, J. M., RUSSELL, R. K., BERRY, S. H., EL-OMAR, E. M. & HOLD, G. L. 2015. The fungal microbiota of de-novo paediatric inflammatory bowel disease. Microbes & Infection, 17, 304-10.

OBA, P. M., HOLSCHER, H. D., MATHAI, R. A., KIM, J. & SWANSON, K. S. 2020. Diet Influences the Oral Microbiota of Infants during the First Six Months of Life. Nutrients, 12, 05.

O'CONNELL, L. M., SANTOS, R., SPRINGER, G., BURNE, R. A., NASCIMENTO, M. M. & RICHARDS, V. P. 2020. Site-specific profiling of the dental mycobiome reveals strong taxonomic shifts during progression of early childhood caries. *Appl Environ Microbiol***,** AEM.02825-19.

PARAHITIYAWA, N., SCULLY, C., LEUNG, W., YAM, W., JIN, L. and SAMARANAYAKE, L. 2010. Exploring the oral bacterial flora: current status and future directions. Oral Diseases, 16: 136-145. https://doi.org/10.1111/j.1601-0825.2009.01607.x

RAJA, M., HANNAN, A. & ALI, K. 2010. Association of oral candidal carriage with dental caries in children. Caries Research, 44, 272-6.

R CORE TEAM. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

ROSENBAUM, J., USYK, M., CHEN, Z., ZOLNIK, C. P., JONES, H. E., WALDRON, L., DOWD, J. B., THORPE, L. E. & BURK, R. D. 2019. Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota. *Sci Rep,* 9**,** 1531.

SAMPAIO-MAIA B, MONTEIRO-SILVA F. 2104. Acquisition and maturation of oral microbiome throughout childhood: An update. Dent Res J (Isfahan). 11(3):291-301. PMID: 25097637; PMCID: PMC4119360.

SCHEI, K., AVERSHINA, E., OIEN, T., RUDI, K., FOLLESTAD, T., SALAMATI, S. & ODEGARD, R. A. 2017. Early gut mycobiota and mother-offspring transfer. Microbiome, 5, 107.

SCHLOSS PD, WESTSCOTT SL, RYABIN T, HALL JR, HARTMANN M, HOLLISTER EB, LESNIEWSKI RA, OAKLEY BB, PARKS DH, ROBINSON CJ, SAHL JW, STRES B, THALLINGER GG, VAN HORN DJ, WEBER CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 75: 7537-41. DOI: 10.1128/AEM.01541-09

SCIENTIFIC INC TF. 2015. Qubit ds DNA BR Assay Kits. Catalog No: Q32850, Q32853. Accessed 23rd November 2021. https://www.thermofisher.com/document-connect/documentconnect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_BR_Assay_UG.pdf

SCIENTIFIC INC TF. 2015. Qubit ds DNA HS Assay Kits. Catalog No: Q32851, Q32854. Accessed 23rd November 2021. https://www.thermofisher.com/document-connect/documentconnect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_HS_Assay_UG.pdf

SEDGHI, L, DIMASSA, V, HARRINGTON, A, LYNCH, SV, KAPILA, YL. 2021. The Oral Microbiome: Role of Key Organisms and Complex Networks in Oral

Health and Disease. Periodontol 2000. 2021; 87: 107– 131. https://doi.org/10.1111/prd.12393

SHARMA, N., BHATIA, S., SODHI, A. S. & BATRA, N. 2018. Oral microbiome and health. *AIMS Microbiol,* 4**,** 42-66.

SCHEI, K., AVERSHINA, E., OIEN, T., RUDI, K., FOLLESTAD, T., SALAMATI, S. & ODEGARD, R. A. 2017. Early gut mycobiota and mother-offspring transfer. Microbiome, 5, 107.

SYSMIC. 2019. SysMIC Module 1 https://sysmic.ac.uk/moodle/course/view.php?id=30

VESTY, A., BISWAS, K., TAYLOR, M. W., GEAR, K. & DOUGLAS, R. G. 2017. Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. *PLoS One,* 12**,** e0169877.

WADE, WG. 2013. The oral microbiome in health and disease. Pharmacological Research. Volume 69, Issue 1, 137-143, ISSN 1043-6618, https://doi.org/10.1016/j.phrs.2012.11.006.

WELLS, G, SHEA, B, O'CONNELL, D, PETERSON, JE., WELCH, V., LOSOS, M., TUGWELLI, P. (2000). The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of Non-Randomized Studies in Meta-Analysis.

XIAO, J., FISCELLA, K.A. & GILL, S.R. 2020. Oral microbiome: Possible Harbinger for Children's Health. *Int J Oral Sci* **12,** 12. https://doi.org/10.1038/s41368-020-0082-x

XIN X, JUNZHI H, XUEDONG Z. 2015. Oral microbiota: a promising predictor of human oral and systemic diseases. Hua Xi Kou Qiang Yi Xue Za Zhi. 33(6):555-60. Chinese. doi: 10.7518/hxkq.2015.06.001. PMID: 27051943; PMCID: PMC7030375.

XU P, GUNSOLLEY J. 2014. Application of Metagenomics in Understanding Oral Health and Disease. Virulence. Apr 1;5(3):424-32. doi: 10.4161/viru.28532. Epub 2014 Mar 18. PMID: 24642489; PMCID: PMC3979870.

ZARCO, M., VESS, T. and GINSBURG, G. 2012. The Oral Microbiome in Health and Disease and The Potential Impact On Personalized Dental Medicine. Oral Diseases, 18: 109-120. https://doi.org/10.1111/j.1601-0825.2011.01851.x

Appendices

Appendix 1: MEDLINE OVID search strategy

MEDLINE Ovid Search strategy

#	Searches
1.	"fung* microbio*2.mp
2.	mycobiome.mp or Mycobiome/
3.	oral.mp
4.	oral.tw
5.	Intestinal mucosa/ or Gastrointestinal Tract/ or gut.mp
6.	Gut.tw
7.	Gastrointestinal Tract.tw
8.	Intestinal mucosa.tw
9.	child.mp. or Child/
10.	Infant/ or infant.mp
11.	toddler.mp
12.	baby.mp
13.	adolescent/ adolescent.mp
14.	("paediatric*" or "paedatric").mp.[mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
15.	lifestyle.tw
16.	diet.tw
17.	sex.tw
18.	age.tw
19.	weight.tw
20.	inflammat*.tw
21.	antibiotics.tw
22.	"immune system".tw

23. 1 or 2
24. 3 or 4 or 5 or 6 or 7 or 8
25. 9 or 10 or 11 or 12 or 13 or 14
26. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
27. 23 and 25
28. 24 and 27
29. 26 and 28
30. ((fungal or fungi) adj4 (oral or gut or intestinal mucosa or gastrointestinal tract)).mp [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
31. 1 or 2 or 30
32. 25 and 31
33. 24 and 32
34. 26 and 33

Appendix 2: Newcastle Ottawa Scale

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✵
   b) yes, eg record linkage or based on self-reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✵
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✵
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (beginning/endpoint) ✵
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for other host or environmental factors ✵
   b) study controls for any additional factor ✵

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✵
   b) structured interview where blind to case/control status ✵
   c) interview not blinded to case/control status
   d) written self-report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✵
   b) no

3) Non-Response rate
   a) same rate for both groups ✵
   b) non respondents described
   c) rate different and no designation

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

1) Representativeness of the exposed cohort
   a) truly representative of the average population in the community ✵
   b) somewhat representative of the average population in the community ✵
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort

2) Selection of the non-exposed cohort
   a) drawn from the same community as the exposed cohort ✵
   b) drawn from a different source
   c) no description of the derivation of the non-exposed cohort

3) Ascertainment of exposure
   a) secure record (eg surgical records) ✵
   b) structured interview ✵
   c) written self-report
   d) no description

4) Demonstration that outcome of interest was not present at start of study
   a) yes ✵
   b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for ✵
   b) study controls for any additional factor ✵ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome
   a) independent blind assessment ✵
   b) record linkage ✵
   c) self-report
   d) no description

2) Was follow-up long enough for outcomes to occur
   a) yes (>6 months) ✵
   b) no (<6 months)

3) Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ✵
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > (<15 % follow up, or description provided of those lost) ✵
   c) follow up rate < 85% and no description of those lost
   d) no statement

Appendix 3: Ethics Approval

17th July 2018

Dr Andrew Smith Microbial Diseases Eastman Dental Institute UCL

Dear Dr Smith

Notification of Ethics Approval with Provisos
Project ID/Title: 7567/001: Do changes in environment alter the oral microbiome?

Further to the review of your application at the recent meeting of the UCL REC, I am pleased to confirm in my capacity as Joint Chair of the UCL Research Ethics Committee (REC) that your study has been ethically approved by the UCL REC until **31st August 2020.**

Ethical approval is subject to the following conditions:

Notification of Amendments to the Research

You must seek Chair's approval for proposed amendments (to include extensions to the duration of the project) to the research for which this approval has been given. Each research project is reviewed separately and if there are significant changes to the research protocol you should seek confirmation of continued ethical approval by completing an 'Amendment Approval Request Form' http://ethics.grad.ucl.ac.uk/responsibilities.php

Adverse Event Reporting – Serious and Non-Serious

It is your responsibility to report to the Committee any unanticipated problems or adverse events involving risks to participants or others. The Ethics Committee should be notified of all serious adverse events via the Ethics Committee Administrator (ethics@ucl.ac.uk) immediately the incident occurs. Where the adverse incident is unexpected and serious, the Joint Chairs will decide whether the study should be terminated pending the opinion of an independent expert. For non-serious adverse events the Joint Chairs of the Ethics Committee should again be notified via the Ethics Committee Administrator within ten days of the incident occurring and provide a full written report that should include any amendments to the participant information sheet and study protocol. The Joint Chairs will confirm that the incident is non-serious and report to the Committee at the next meeting. The final view of the Committee will be communicated to you.

Final Report

At the end of the data collection element of your research we ask that you submit a very brief report (1-2 paragraphs will suffice) which includes in particular issues relating to the ethical implications of the research i.e. issues obtaining consent, participants withdrawing from the research, confidentiality, protection of participants from physical and mental harm etc.

Office of the Vice Provost Research, 2 Taviton Street University College London
Tel:+44(0)2076798717
Email: ethics@ucl.ac.uk

http://ethics.grad.ucl.ac.uk/

In addition, please:

- ensure that you follow all relevant guidance as laid out in UCL's Code of Conduct for Research: http://www.ucl.ac.uk/srs/governance-and-committees/resgov/code-of-conduct-research

- note that you are required to adhere to all research data/records management and storage procedures agreed as part of your application. This will be expected even after completion of the study.

With best wishes for the research. Yours sincerely

Professor Michael Heinrich
Joint Chair, UCL Research Ethics Committee

Cc: Professor Dave Spratt

Appendix 4: Ethics Approval Extension

Ethics ID Number: 7567/001

Dear Violeta,

The REC has approved your attached extension request and the ethics approval of this study has therefore been extended to **31/08/2021**. Please take this email as confirmation of that approval.

IMPORTANT: For projects collecting personal data only

You should inform the Data Protection Team – data-protection@ucl.ac.uk of your proposed amendments to include a request to extend ethics approval for an additional period.

Best wishes,
Lola

Lola Alaska
Research Evaluation Administrator

Appendix 5: Participation Information Sheet For Young Adults

Participant Information Sheet For Young Adults

UCL Research Ethics Committee Approval ID Number: 7567/001

Version 1 25/05/2018

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Do changes in environment alter the oral microbiome?

Secondary Title: Do changes in environment that you grow up in change the bacterial community in your mouth?

Department: Microbial Diseases, Eastman Dental Institute
Name and Contact Details of the Researcher(s):

Name and Contact Details of the Principal Researchers:
Dr Andrew Smith (Andrew.m.smith@ucl.ac.uk)
Prof David Spratt (d.spratt@ucl.ac.uk)

Invitation

We'd like you to help us with a research project. You need to understand why we are doing this and what taking part will mean. So please take time to read the following information carefully and discuss it with your parent(s)/guardian and teachers if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the project's purpose?

Background

The oral microbiome is the name given to all the microscopic creatures or bugs (bacteria, fungi and viruses) that live in our mouths. Everybody has an oral microbiome and each one is unique in the same way as a fingerprint. A number of these bugs can be very beneficial to our health and scientists and dentists are very interested in understanding how the different types of bugs in our mouths change as we age.

From birth to now, your mouth has gone through a lot of changes. We know that these physical changes (e.g. teeth) will influence and change the types of bugs that live in our mouths. We now think that other factors might affect the types of bacteria we have, and where you live could be one of them.

Aims

The aim of this study is to see if your surroundings (ie where you live) can affect the type of bacteria you have in your mouth

Why is this research important?

Currently we do not understand enough about how the oral bacteria develop and how stable they are in children and young adults. But we do know that the type of bacteria we have can influence everything from caries through to major health problems like diabetes. We have designed this study to help investigate this.

What do we want to do? / What will happen to me if I take part?

We want to recruit new students entering the school at Year's 7, 9 and 11. We will ask students to complete a general dental health questionnaire in the first appointment and we will also collect nine saliva samples (each one is less than a teaspoon of saliva). The saliva samples will be collected at the beginning, middle and end of each term. This will allow us to work out what bugs are living in your mouth and monitor any changes that may occur over the school year. We will also conduct two dental examinations (one at the beginning of term and a second at the end of term three) in order to determine your overall oral health during the study. The students will be put into one of two groups depending if they are non-boarders (Group 1, a stable environment) or boarders (Group 2, a changed environment). We will then determine the identity of the oral bacteria at each sampling point (for each student) and see if it changes with time. We will then compare the changes observed between Group 1 and Group 2. This will help us understand if the student's environment influences the bacterial community in the mouth.

Why have I been chosen?

You have been asked to take part because you are a new student attending secondary school and are enrolled in either year 7, 9 or 11. In order to take part in the study you will have to get permission from your parent or legal guardian.

Do I have to take part?

The study is entirely voluntary, and it is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep. You and a parent or primary carer will need to read and sign a consent form. You can withdraw at any time without giving a reason and without any penalty or impact on your education. If you decide to withdraw you will be asked what you wish to happen to the information you have provided up that point.

Are there any risks to me?

There are no risks that we can see. If the dental examination finds a problem which might need looking at in more detail, we will contact your parents and dentist to follow this up.

What are the possible benefits of taking part?

There won't be any obvious benefits for you. But the information we get from this study might help children in the future.

What if something goes wrong?

If you have any complaints relating to the study or any of the members of the research team, you can contact in confidentiality the Principal Investigators Dr Andrew Smith or Professor David Spratt (contact details at the end of this document). However, you can also take your complaint to the Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk), if you are not satisfied with how you have been treated or if the issue is with principle investigators.

Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential.

Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

What will happen to the results of the research project?

The results from this study will be presented at clinical, dental and scientific meetings in the UK and abroad. We will also publish the findings in international peer reviewed journals. We will also return to your school the following year and present the findings to the students and teachers. We will explain the findings and participate in a question and answer session. Any data we present will be anonymised.

Data Protection Privacy Notice

Notice:

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal

What are the possible benefits of taking part?

There won't be any obvious benefits for you. But the information we get from this study might help children in the future.

What if something goes wrong?

If you have any complaints relating to the study or any of the members of the research team, you can contact in confidentiality the Principal Investigators Dr Andrew Smith or Professor David Spratt (contact details at the end of this document). However, you can also take your complaint to the Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk), if you are not satisfied with how you have been treated or if the issue is with principle investigators.

Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential.

Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

What will happen to the results of the research project?

The results from this study will be presented at clinical, dental and scientific meetings in the UK and abroad. We will also publish the findings in international peer reviewed journals. We will also return to your school the following year and present the findings to the students and teachers. We will explain the findings and participate in a question and answer session. Any data we present will be anonymised.

Data Protection Privacy Notice

Notice:

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal

data, and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer is Lee Shailer and he can also be contacted at data-protection@ucl.ac.uk.

Your personal data will be processed for the purposes outlined in this notice. The legal basis that would be used to process your personal data will be the provision of you and your parents/legal guardians consent. You can provide your consent for the use of your personal data in this project by completing the consent form that has been provided to you.

Your personal data will be processed so long as it is required for the research project. We will remove all identifiable information from the personal data you provide and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at data-protection@ucl.ac.uk. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

Who is organising and funding the research?

This research project was designed by the principal investigators and is being funded by a research grant awarded by Biomedical Research Centre (**BRC**) University College Hospital.

Contact for further information

Principle Investigators

Dr Andrew Smith
andrew.m.smith@ucl.ac.uk
020 76796008

Professor David Spratt
d.spratt@ucl.ac.uk
020 34561107

Thank you for reading this information sheet and for considering taking part in this research study. If you take part in the study you will be able to keep a copy of the information sheet and signed consent form for your records.

Appendix 6: Parent/ Guardian Information Sheet

Parent/Guardian Information Sheet

UCL Research Ethics Committee Approval ID Number: 7567/001

Version 1 25/05/2018

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Do changes in environment alter the oral microbiome?

Secondary Title: Do changes in environment that you grow up in change the bacterial community in your mouth?

Department: Microbial Diseases, Eastman Dental Institute
Name and Contact Details of the Researcher(s):

Name and Contact Details of the Principal Researchers:
Dr Andrew Smith (Andrew.m.smith@ucl.ac.uk)
Prof David Spratt (d.spratt@ucl.ac.uk)

Invitation

Your Child is being invited to take part in a scientific research project. Before you decided if you would be willing for your child to enrol in the study, it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with your child. Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you would be happy for your child to take part in the study.

If you are happy for your child to participate in the study, could you please sign the consent form and return it to the school. We will then invite your child to a recruitment meeting where we will go over the study answer any questions they may have and then get them to fill in the consent form that you have signed.

What is the project's purpose?

Background

We are very interested in how the community of microbes (bacteria, fungi and viruses) in our mouths changes as we age. When we are born our mouths lack any oral microbes. In the early period of our growth it becomes colonised by bacteria, which are typically acquired from our parents or primary care giver. As we grow up, our oral bacterial community develops to become a complex and diverse ecosystem with hundreds of species present.

We have recently acquired new evidence showing that where we live and our immediate family has a significant effect on our oral microbes. What is also surprising is that our own genes (DNA) play almost no role. Previously your DNA was thought to play a crucial role in forming and defining oral microbial communities. These new findings therefore provide the possibility of altering an individual's oral microbes through changing the environment they live in. It is now clear that both oral disease and indeed other diseases, such as heart disease and diabetes, are influenced by the types of the bacteria that live in the mouth. If we were able to "re-programme" the microbial community within the mouth we could be able to confer substantial health benefits. However, in adults it is very difficult to alter the oral microbial communities permanently. What we don't know is when, in our development, the oral microbial communities become fixed.

Aims

The aim of this study is to determine if the oral bacterial community is stable in adolescents and how the environment may influence the composition of this bacterial community.

Why is this research important?

Despite being almost completely preventable by good oral hygiene, a third of children in England suffer from tooth decay (caries). It is the most common reason why 5-9 year olds are admitted to hospital and undergo general aesthetic. The NHS spends £30 million per year on hospital based tooth extractions for children. Caries is caused by acid producing bacteria that destroy the tooth's enamel. We believe that while the oral bacterial community is still developing, we can intervene to create a community that does not cause caries, this would be a major leap forward and save all of the pain and suffering associated with tooth decay and treatment. Currently we do not understand enough about how the oral bacteria develop and how stable they

are in children and young adults. We have designed this study to help answer these questions.

Proposed study

We believe that there must be a period in human development in which the oral microbial community can be altered permanently and point where it becomes fixed. To accurately find this period we need to follow subjects over time. There are only a few points in human development where this type of study can easily be performed one of which is when people of similar age come together and reside in a new environment. In the first year of secondary school students have a dramatic change in lifestyle in some cases. Students who enrol in boarding school will have a major change in environment and this will enable us to determine how much of an effect environment has on the oral microbial community and how stable it becomes over time.

We will recruit new students entering the school at Year's 7, 9 and 11. We will require the students to complete a general oral health questionnaire during their first appointment and we will also collect nine saliva samples (each one is less than a teaspoon of saliva). The saliva samples will be collected at the beginning, middle and end of each term. The students will be put into one of two groups depending if they are non-boarders (Group 1, a stable environment) or boarders (Group 2, a changed environment). We will then determine the identity of the oral bacteria at each sampling point (for each student) and see if it changes with time. We will then compare the changes observed between Group 1 and Group 2. This will help us understand if the student's environment influences the bacterial community in the mouth.

Why have you chosen my child?

We have identified your child as a potential participant in our study because they are a new student attending secondary school and are enrolled in either year 7, 9 or 11. In order for your child to take part in the study they will need to get permission from their parent or legal guardian, agree to answer the questionnaire and provide nine saliva samples during the coming school year. We will also conduct two oral examinations (one at the beginning of term and a second at the end of term three) in order to determine your child's overall oral health during the study.

Does my child have to take part?

The study is entirely voluntary and it is up to you and your child to decide whether or not to take part. If your child takes part, you will be given this information sheet to keep. You and your child will need to read and sign a consent form. They can withdraw at any time without giving a reason and without prejudice on their education. If they decide to withdraw, they will be asked what they wish to happen to the data they have provided up that point.

What will happen to your child if they take part?

The study will last for a year and your child will be required to attend nine appointments (beginning, middle and end of each term). All the appointments will be conducted at school and within term time. The appointments will last approximately 5-10 minutes and they will be required to provide a saliva sample (less than a teaspoon in volume and they will just need to spit into a sterile tube that we provide at the meeting) and answer a lifestyle questionnaire. On the first and last appointments they will also be asked to undergo an oral examination, which should last approximately 5 minutes. In the first meeting they will be assigned into group 1 (non-boarders) or group 2 (boarders) depending on their residential status at the school.

What are the possible disadvantages and risks of taking part?

The saliva samples that are collected for the study will only be used to isolate bacterial DNA and carries no identified risks to the participants. The oral examination may identify a hitherto unknown condition, which may require additional investigation. In these circumstances the fully qualified paediatric dentist who conducted the examination will send a letter to you as their legal guardian explaining the findings and advising on the best course of action.

What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will lead to a greater understanding of how the oral bacterial community develop and how much of this is influenced by a person's environment. If the oral bacteria are found to be dynamic (not fixed) in adolescents, then it could be possible to artificially manipulate the bacterial composition in order to provide

beneficial effects on an individual. An indirect benefit will be the greater knowledge of the mouth and the importance of oral microbiome that the participant will gain.

What if something goes wrong?

If you or your child has any complaints relating to the study or any of the members of the research team, you can contact in confidentiality the Principal Investigators Dr Andrew Smith or Professor David Spratt (contact details at the end of this document). However, you can also take your complaint to the Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk), if you are not satisfied with how you have been treated or if the issue is with principle investigators.

Will my child taking part in this project be kept confidential?

All the information that we collect about your child during the course of the research will be kept strictly confidential. Upon enrolment your child will be assigned a unique project code which will be linked to their data. This pseudonymised data means that they cannot be identified in any ensuing reports or publications.

Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

What will happen to the results of the research project?

The results from this study will be presented at clinical, dental and scientific meetings in the UK and abroad. We will also publish the findings in international peer reviewed journals. We will also return to your school the following year and present the findings to the students and teachers. We will explain the findings and participate in a question and answer session. All data will be presented in a pseudonymised fashion where no individual can be identified.

Data Protection Privacy Notice

Notice:

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer is Lee Shailer and he can also be contacted at data-protection@ucl.ac.uk.

Your child's personal data will be processed for the purposes outlined in this notice. The legal basis that would be used to process your child's personal data will be the provision of your child's and your consent. You and your child can provide consent for the use of your child's personal data in this project by completing the consent form that has been provided.

Your child's personal data will be processed so long as it is required for the research project. We will pseudonymise the personal data your child provides and will endeavour to minimise the processing of personal data wherever possible.

If you or your child are concerned about how personal data is being processed, please contact UCL in the first instance at data-protection@ucl.ac.uk. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

Who is organising and funding the research?

This research project was designed by the principal investigators and is being funded by a research grant awarded by Biomedical Research Centre (**BRC**) University College Hospital.

Contact for further information

Principle Investigators

Dr Andrew Smith
andrew.m.smith@ucl.ac.uk
020 76796008

Professor David Spratt
d.spratt@ucl.ac.uk
020 34561107

Thank you for reading this information sheet and for considering taking part in this research study. If you take part in the study you will be able to keep a copy of the information sheet and signed consent form for your records.

Appendix 7: Consent Form for Adolescents in Research Studies

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Do changes in environment alter the oral microbiome?

Department: Microbial Diseases, Eastman Dental Institute

Name and Contact Details of the Researcher(s):
Principle researchers
Professor David Spratt
d.spratt@ucl.ac.uk
Dr Andrew Smith
Andrew.m.smith@ucl.ac.uk

Name and Contact Details of the UCL Data Protection Officer: Sara Green S.green@ucl.ac.uk

This study has been approved by the UCL Research Ethics Committee:
Project ID number: 7567/001

Thank you for considering taking part in this research. The person organising the research must explain the project to you and your parent(s)/guardian before you agree to take part. If you or your parent(s)/guardian have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You and your parent(s)/guardian will be given a copy of this Consent Form to keep and refer to at any time.

		Tick Box
	I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction and would like to take part in the study.	
	I consent to the processing of my personal information associated with my lifestyle and oral health for the purposes explained to me. I understand that such information will be handled in accordance with all applicable data protection legislation.	
	I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified.	
	I understand that my information may be subject to review by responsible individuals from the University for monitoring and audit purposes.	
	I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason, without my legal rights being affected. I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.	
	I understand the potential risks of participating and the support that will be available to me should I become distressed during the course of the research.	
	I understand that no promise or guarantee of benefits have been made to encourage you to participate.	
	I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researcher(s) undertaking this study.	
	I understand that I will not benefit financially from this study or from any possible outcome it may result in in the future.	
	I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No	
	I hereby confirm that I understand the inclusion criteria as detailed in the Information Sheet and explained to me by the researcher.	
	I hereby confirm that: (a) I understand the exclusion criteria as detailed in the Information Sheet and explained to me by the researcher; and (b) I do not fall under the exclusion criteria.	

	I have informed the researcher of any other research in which I am currently involved or have been involved in during the past 12 months.	
	I am aware of who I should contact if I wish to lodge a complaint.	
	I agree to complete the questionnaire associated with the study on oral health (see Information sheet)	
	I agree to provide saliva samples at specified time points throughout the study (see information sheet)	
	I would be happy for all of my data and information generated during the study to be pseudonymised and retained for 20 years in accordance with current University College London records Management policy. I would be happy for the data I provide to be archived at the Department of Microbial Diseases, University College London. I understand that other authenticated researchers will have access to my pseudonymised data.	

I confirm that I understand that by ticking/initialling each box below I am consenting to this element of the study. I understand that it will be assumed that unticked/initialled boxes means that I DO NOT consent to that part of the study. I understand that by not giving consent for any one element that I may be deemed ineligible for the study.

If you would like your contact details to be retained so that you can be contacted in the future by UCL researchers who would like to invite you to participate in follow up studies to this project, or in future studies of a similar nature, please tick the appropriate box below.

	Yes, I would be happy to be contacted in this way	
	No, I would not like to be contacted	

_____	_____	_____
Name of participant	Date	Signature
_____	_____	_____
Name of Parent/Guardian	Date	Signature
_____	_____	_____
Researcher	Date	Signature

Appendix 8: Oral Health Questionnaire

Oral health questionnaire

Do changes in environment alter the oral microbiome?

Principal Investigators: Dr Andrew Smith, Prof Dave Spratt, Dr Paul Ashley

Version 1 (22/05/2018)

UCL Research Ethics Committee Approval ID Number: 7567/001

In order to understand better how the bugs in your mouth can interact with your body, we need to know a little bit more about you. Please try and answer the questions below as best as you can.

The information you give us is confidential.

1. About you

1.1 Unique Identification code

……..…..

1.2 Age (years)

……

1.3 Gender (please circle)

Male Female

1.4 What year are you in?

…….

2. About your home

2.1 Are you a boarder? (please circle)

Yes No

2.2 Which country is home in?

…….....

2.3 How many brothers/sisters do you have?

……...

2.4 Do you have any pets at home? (please circle)

Yes No

If the answer was yes to 2.4, what pets do you have?

……

3. About your health and fitness

3.1 Have you had antibiotics in the last 3 months?

Yes No

3.2 Do you have asthma? (please circle)

Yes No

3.3 Do you have eczema? (please circle)

Yes No

3.4 Do you drink Yakult or Actimel more than once a week? (please circle)

Yes No

3.5 Are you on any of the school sports teams? (please circle)

Yes No

If the answer was yes to 3.5, which teams?

……

3.6 How many times a month do you swim in a swimming pool

……

3.7 Do you ever wear a sports mouthguard? (please circle)

Yes No

4. About your mouth

4.1 How many times a day do you clean your teeth?

…….

4.2 How many times a day do you use mouth rinse?

…….

4.3 Have you got any fillings? (please circle)

Yes No

4.4 Have you had any teeth extracted? (please circle)

Yes No

4.5 Have you ever had dental braces? (please circle)

Yes No

If the answer was yes to 4.5, do you have retainers? (please circle)

Yes No

4.6 Do you have dental braces now? (please circle)

Yes No

4.7 Do you wear a denture with a false tooth? (please circle)

Yes No

Appendix 9: Screening Protocol

Do changes in environment alter the oral microbiome?
Screening protocol V2
Paul Ashley 6/11/18

Environment and Cross-infection control

The participant should be seated in a comfortable chair, which has good head support, and which allows the examiner to access and examine the head and mouth.

A suitable external bright light source should be used such as a head torch or lamp.

The instruments should be laid out on a clean tissue out of sight of the participant (if possible) and allowing easy access.
The light source should be set at the highest power setting and dark eye protection glasses placed on the subject.
Each examiner should carry sufficient sets of sterile **disposable** instruments to ensure that there are sterile instruments for every examination. Following the examination these should be disposed. Examiners should wear a clean pair of **latex-free gloves** for the examination of each participant along with a mask and eye protection. Appropriate cross-infection procedures must be followed throughout.

An equipment list is in Appendix 1

Data collection

Where possible questionnaires for additional data should be completed by participants as they wait for the oral examination.
When scoring clinical outcomes, if in doubt score 'low'.
Examples of data collection forms are in appendix 2.

Examiner training

The training should be structured to provide a clear understanding of the nature and aims of the study or screening exercise together will the

assessment/examination procedures and completion of appropriate documentation. Preferably, examiners should be trained and compared to a gold-standard examiner. Furthermore, it will be important to assess their consistency in measurements (within-individual repeatability).

Caries and restorations (DFT)

Caries and restorations should be scored using modified ICDAS criteria (https://www.iccms-web.com/). Caries will be recorded at tooth level with the most severe score on any surface being recorded. Caries will be regarded as more severe than a restoration if both are recorded on one tooth.

MODIFIED ICDAS

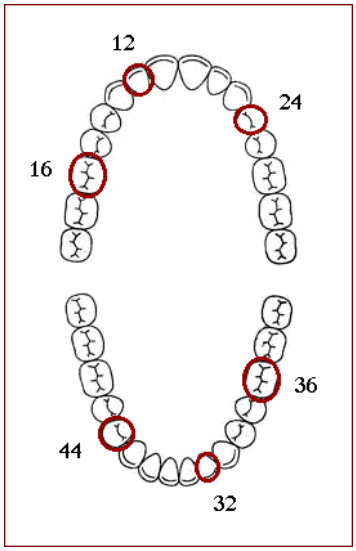
0 No evidence of caries
A Initial caries
3 Localised enamel breakdown due to caries with no visible dentine
4 Underlying dark shadow from dentine
5 Distinct cavity with visible dentine
6 Extensive distinct cavity with visible dentine

In addition, teeth with unrestorable decay should be noted. This can be defined as:
“Untreated teeth with extensive dentinal decay have obvious loss of tooth structure, with a cavity both deep and wide so that dentine is clearly visible on the walls and at the base. Such a cavity would involve at least half of a tooth surface, and teeth coded in this way are so broken down that it is inconceivable that there is not pulp involvement and so restoration of the tooth would be very involved or impossible.” (Adult Dental Health Survey, 2009)

Plaque

Plaque will be measured using the Silness-Löe Index (https://www.mah.se/CAPP/Methods-and-Indices/Oral-Hygiene-Indices/Silness-Loe-Index/)

The measurement of the state of oral hygiene by Silness-Löe plaque index is based on recording both soft debris and mineralized deposits on the following teeth (Primary teeth can be substituted for permanent teeth. Missing teeth are not substituted).



Each of the four surfaces of the teeth (buccal, lingual, mesial and distal) is given a score from 0-3. The scores from the four areas of the tooth are added and divided by four in order to give the plaque index for the tooth with the following scores and criteria:

The Plaque Index System

Scores	Criteria
0	No plaque
1	A film of plaque adhering to the free gingival margin and adjacent area of the tooth. The plaque may be seen in situ only after application of disclosing solution or by using the probe on the tooth surface.

2	Moderate accumulation of soft deposit s within the gingival pocket, or the tooth and gingival margin which can be seen with the naked eye.
3	Abundance of soft matter within the gingival pocket and/or on the tooth and gingival margin.

Swellings/abscesses/infections

Pulp, Ulceration, Fistula and Abscess (PUFA Index)

Description of conditions to be recorded in PUFA

P = open pulp in permanent dentition

U = obvious ulceration

F = fistula in permanent dentition

A= abscess in permanent dentition

Codes and criteria: PUFA

0 = No lesions evident

1 = A single lesion present

2 = 2 or more lesions present

The mouth should be examined in the following order (upper right, upper left, lower left, lower right), ensuring that the lips or cheeks are gently retracted to allow the soft tissues to be examined.

A single code (0, 1 or 2) will be called for each of the four conditions examined.

Other

In addition the presence or absence of any orthodontic appliances or any other dentures will be recorded.

Appendix 10: Clinical Data Collection Form

Clinical Data Collection form

Patient ID……

Clinician ……..

Date ……

PUFA index (Pulp, Ulceration, Fistula, Abscess)

	Code **0=no lesions, 1=single lesion, 2=2 or more lesions**
Pulp	
Ulcer	
Fistula	
Abscess	

Plaque examination (0,1,2,3)

UR6	UR2	UL4
LR4	LL2	LL6

DFT (ICDAS criteria for caries, code restoration as R)

Tooth	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Tooth	**48**	**47**	**46**	**45**	**44**	**43**	**42**	**41**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**

Tooth	55	54	53	52	51	61	62	63	64	65
Tooth	85	84	83	82	81	71	72	73	74	75

Appendix 11: DNA concentration measured with NanoDrop Spectophotometer

CODING	260/280	260/230	ng/uL
H1S1	0.94	1.13	35.7
H2S1	2.19	1.45	29.4
H3S1	2.12	1.07	14.8
H4S1	1.81	1.35	138.2
H5S1	1.52	2.01	4.8
H6S1	1.84	1.49	125.0
H7S1	1.65	0.55	18.8
H8S1	1.71	0.86	7.7
H9S1	2.00	1.20	6.2
H11S1	2.14	1.83	4.8
H12S1	1.64	3.51	4.1
H13S1	1.58	0.85	217.9
H14S1	1.97	2.32	35.3
H15S1	1.38	2.33	2.3
H16S1	1.74	0.86	9.2
H17S1	2.01	2.38	25.8
H18S1	2.04	1.56	10.6
H1S2	0.63	1.22	3.5
H2S2	1.75	2.24	6.6
H3S2	2.38	1.78	7.5
H4S2	1.96	2.38	190.1
H5S2	1.80	1.09	16.4
H6S2	1.43	1.66	24.2

H7S2	1.89	1.85	45.3
H8S2	1.53	3.53	6.8
H9S2	1.60	1.04	99.7
H11S2	2.41	1.63	8.2
H12S2	-1.99	9.09	1.0
H13S2	2.18	2.30	6.7
H14S2	2.00	2.39	46.6
H15S2	1.50	0.87	182.6
H16S2	4.75	1.15	2.4
H17S2	1.64	1.04	366.4
H18S2	0.74	1.25	11.8
H1S3	-5.32	2.29	2.2
H2S3	2.08	1.85	14.5
H3S3	2.35	1.06	8.3
H4S3	1.87	2.29	29.3
H5S3	-0.7	-0.73	0.5
H6S3	6.13	3.23	1.8
H7S3	1.62	2.15	15.5
H8S3	1.73	1.38	425.8
H9S3	1.69	1.81	8.8
H11S3	6.66	-1.1	0.8
H12S3	4.49	2.85	1.8
H13S3	-0.91	5.87	0.5
H14S3	1.91	2.21	47.0
H15S3	-4.25	2.72	2.2
H16S3	2.99	1.43	6.7

H17S3	1.93	1.58	19.9
H18S3	4.36	1.76	6.1
H1S4	0.60	1.05	-1.1
H2S4	2.34	1.72	14.1
H3S4	0.29	0.38	-0.7
H4S4	-0.17	-0.33	0.2
H5S4	0.44	0.91	-0.5
H6S4	2.28	0.75	5.1
H7S4	3.36	1.06	2.8
H8S4	3.06	1.69	7.6
H9S4	1.67	1.33	205.5
H11S4	1.23	1.91	6.1
H12S4	1.64	1.40	4.1
H13S4	0.04	0.35	0.0
H14S4	2.02	2.58	370.2
H15S4	2.24	2.39	2.8
H16S4	1.94	2.25	6.6
H17S4	1.70	1.75	8.0
H18S4	3.51	2.21	10.2
H1S5	0.75	0.73	3.1
H2S5	1.85	2.52	16.7
H3S5	1.61	1.27	52.8
H4S5	1.81	1.53	214.9
H5S5	0.56	1.03	2.5
H6S5	0.45	1.64	1.4
H7S5	1.30	2.13	14.9

H8S5	1.36	1.79	14.0
H9S5	1.67	2.39	33.1
H11S5	1.11	1.40	10.3
H12S5	0.32	0.59	0.6
H14S5	1.71	2.40	37.4
H15S5	1.52	1.84	21.6
H16S5	0.67	0.94	2.6
H17S5	1.33	1.51	8.8
H18S5	0.54	1.08	3.7
H1S6	0.76	0.47	-2.1
H2S6	-0.65	-0.27	1.1
H3S6	1.87	1.56	31.2
H4S6	1.85	2.22	236.2
H5S6	1.80	1.40	16.2
H6S6	1.92	2.58	67.7
H7S6	-9.36	0.34	-1.2
H8S6	1.99	1.80	16.4
H9S6	2.02	1.92	82.4
H11S6	1.78	-2.79	4.7
H12S6	3.72	8.39	14.6
H13S6	5.48	0.81	3.6
H14S6	2.07	2.22	120.0
H15S6	1.85	1.15	44.5
H16S6	1.71	1.57	12.0
H1S7	4.09	0.72	2.2
H2S7	1.83	2.01	35.4

H3S7	1.27	0.95	4.3
H4S7	1.79	1.73	1076.8
H5S7	1.87	1.54	30.3
H6S7	1.58	1.01	13.8
H8S7	1.71	1.56	14.2
H9S7	2.04	2.03	99.0
H11S7	2.98	1.13	6.9
H12S7	1.93	0.99	4.8
H14S7	1.80	1.33	21.5
H15S7	1.60	1.06	11.8
H16S7	1.46	1.03	8.8
H17S7	2.08	2.23	47.1
H18S7	2.04	1.55	12.6
H1S9	-0.44	0.12	0.1
H2S9	2.38	1.36	7.1
H3S9	1.81	14.06	9.6
H4S9	2.51	-1.69	2.8
H5S9	-0.05	0.04	-0.1
H6S9	9.43	1.40	3.8
H7S9	2.15	1.22	16.5
H8S9	6.46	18.35	7.6
H9S9	-4.52	0.43	-0.6
H11S9	2.55	1.67	6.9
H12S9	1.86	1.65	11.6
H13S9	2.48	1.04	3.2
H14S9	2.00	1.12	3.5

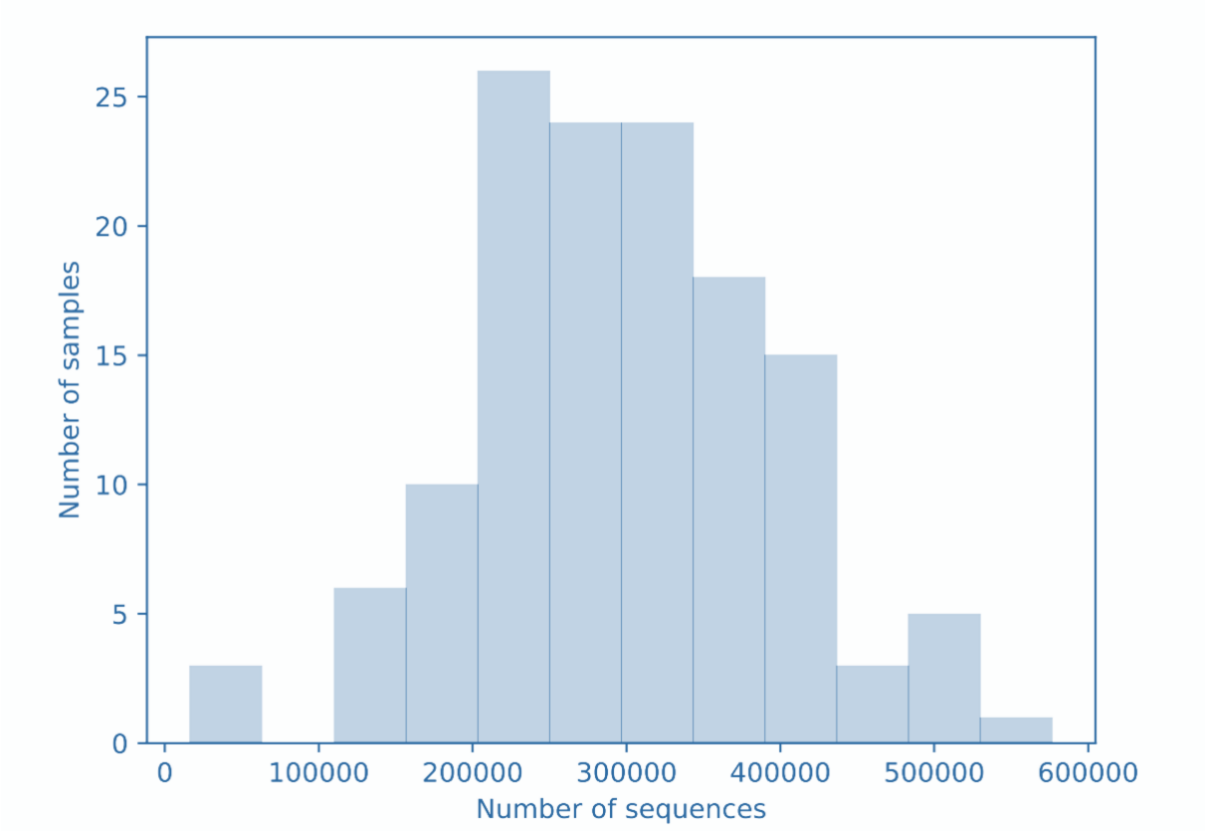
H15S9	1.72	0.64	4.8
H16S9	0.99	-0.11	0.4
H17S9	2.06	2.03	15.1
H18S9	1.54	1.62	10.9
H1S10	-7.53	0.81	1.9
H2S10	2.34	1.66	8.4
H3S10	-2.22	0.63	1.3
H4S10	1.95	1.19	5.5

Appendix 12: Summary of sequences after demultiplexing

Demultiplexed sequence counts summary

	forward reads
Minimum	16416
Median	294036
Mean	298861
Maximum	576796
Total	40346282

Forward Reads Frequency Histogram



Per-sample sequence counts

Total samples: 135 (forward)

sample ID	Forward sequence count
H2S1	576796
H2S7	521838
H14S3	516686
H14S1	513650
H6S6	511442
H8S3	495062
H18S7	482008
H14S9	481258
H7S1	462140
H6S1	429340
H9S5	419526
H7S2	415774
H8S2	409540
H2S4	408168
H11S1	407868
H9S3	407576
H14S5	407490
H6S9	403172
H4S1	403166
H11S9	402062

H7S9	397612
H15S1	392080
H4S2	390792
H18S2	390626
H18S5	385858
H3S2	384984
H13S1	382838
H8S1	381864
H15S4	373424
H3S1	371832
H12S9	370902
H16S4	366938
H5S2	362266
H16S1	362158
H1S5	356806
H2S3	354768
H11S2	349482
H15S9	348264
H16S5	348244
H16S2	348116
H4S7	344508
H9S9	343688
H17S1	341484

H13S3	341470
H17S4	340792
H6S3	340426
H11S6	338934
H9S2	338616
H13S9	337790
H3S3	336764
H15S2	334864
H2S10	334680
H5S7	334438
H15S3	331502
H4S9	330054
H8S5	328310
H9S4	325824
H12S3	325310
H12S1	325154
H13S2	318752
H17S9	311240
H11S4	308986
H7S5	303646
H3S5	303338
H14S4	300108
H4S4	297110

H8S9	296382
H7S3	294036
H14S7	290574
H2S5	289700
H1S7	286596
H5S6	285516
H15S6	282982
H7S4	282056
H3S9	280330
H6S5	277978
H1S9	275436
H1S6	274850
H14S2	272182
H2S2	269322
H5S5	267486
H6S4	263090
H4S3	262602
H18S1	262002
H1S10	261642
H17S3	260106
H1S4	259116
H5S4	255396
H1S3	253964

H16S3	251118
H3S10	249628
H11S3	246074
H1S2	244774
H2S9	243296
H12S2	241932
H17S5	238062
H18S9	237592
H9S6	235340
H6S2	234912
H3S6	234218
H11S5	234196
H2S6	232156
H9S1	229512
H5S3	228144
H9S7	227884
H13S6	225958
H17S7	225764
H5S9	225648
H8S4	224976
H18S4	224526
H6S7	223626
H11S7	221018

H3S4	218982
H17S2	215782
H12S7	206272
H15S7	204920
H12S6	199244
H8S7	196356
H3S7	191528
H16S7	191250
H14S6	189788
H12S4	189366
H18S3	188094
H12S5	183898
H16S9	179164
H16S6	178186
H4S10	153882
H13S4	151250
H15S5	141952
H1S1	130314
H4S5	117352
H8S6	114170
H4S6	58462
H7S6	33752
H5S1	16416

Appendix 13: DADA2 denoising statistics

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H11S1	407868	301386	73.89	294173	254326	62.35
H11S2	349482	270709	77.46	264683	223319	63.9
H11S3	246074	175353	71.26	170444	149342	60.69
H11S4	308986	228170	73.84	222473	186151	60.25
H11S5	234196	181499	77.5	177498	156936	67.01
H11S6	338934	267598	78.95	261171	217642	64.21
H11S7	221018	161803	73.21	159054	144081	65.19
H11S9	402062	299985	74.61	292729	242912	60.42
H12S1	325154	241053	74.14	234585	207226	63.73
H12S2	241932	182378	75.38	177591	162725	67.26
H12S3	325310	243180	74.75	237879	209621	64.44
H12S4	189366	138931	73.37	134518	119186	62.94
H12S5	183898	132699	72.16	128391	119688	65.08

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H12S6	199244	150794	75.68	147681	138468	69.5
H12S7	206272	145290	70.44	142974	130107	63.08
H12S9	370902	268592	72.42	264095	226405	61.04
H13S1	382838	284142	74.22	277895	240536	62.83
H13S2	318752	241679	75.82	236780	213023	66.83
H13S3	341470	270937	79.34	265680	233253	68.31
H13S4	151250	107309	70.95	105019	99342	65.68
H13S6	225958	167851	74.28	164087	142569	63.1
H13S9	337790	263995	78.15	256447	211542	62.63
H14S1	513650	390352	76	379830	321992	62.69
H14S2	272182	210662	77.4	204592	174581	64.14
H14S3	516686	409746	79.3	397955	301471	58.35
H14S4	300108	225631	75.18	219855	188953	62.96

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H14S5	407490	316757	77.73	307452	228014	55.96
H14S6	189788	137150	72.26	133154	111754	58.88
H14S7	290574	230256	79.24	221233	172721	59.44
H14S9	481258	363897	75.61	352965	264525	54.97
H15S1	392080	296476	75.62	289764	248329	63.34
H15S2	334864	251661	75.15	245712	207324	61.91
H15S3	331502	252661	76.22	246879	218410	65.88
H15S4	373424	274865	73.61	268023	226799	60.73
H15S5	141952	97313	68.55	94607	87451	61.61
H15S6	282982	210285	74.31	203249	161414	57.04
H15S7	204920	148830	72.63	146366	134002	65.39
H15S9	348264	264044	75.82	257946	227955	65.45
H16S1	362158	271761	75.04	266695	248299	68.56

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H16S2	348116	264718	76.04	260906	230139	66.11
H16S3	251118	178872	71.23	175195	153761	61.23
H16S4	366938	270406	73.69	266563	238980	65.13
H16S5	348244	259046	74.39	255353	229793	65.99
H16S6	178186	132664	74.45	130993	118998	66.78
H16S7	191250	136198	71.21	134774	123958	64.81
H16S9	179164	133479	74.5	131312	119242	66.55
H17S1	341484	254536	74.54	248078	220710	64.63
H17S2	215782	154494	71.6	149889	136344	63.19
H17S3	260106	186657	71.76	180798	157539	60.57
H17S4	340792	260648	76.48	253729	219172	64.31
H17S5	238062	177910	74.73	172098	151045	63.45
H17S7	225764	160732	71.19	154448	134387	59.53

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H17S9	311240	228015	73.26	221274	180121	57.87
H18S1	262002	178556	68.15	174231	151653	57.88
H18S2	390626	293307	75.09	287349	241800	61.9
H18S3	188094	133354	70.9	129880	118234	62.86
H18S4	224526	150438	67	147822	138530	61.7
H18S5	385858	283975	73.6	277483	235208	60.96
H18S7	482008	356062	73.87	346616	262012	54.36
H18S9	237592	174667	73.52	169331	136260	57.35
H1S1	130314	79841	61.27	76169	65200	50.03
H1S10	261642	191454	73.17	187354	162001	61.92
H1S2	244774	168782	68.95	164510	143006	58.42
H1S3	253964	183118	72.1	178310	153431	60.41
H1S4	259116	179926	69.44	176864	161553	62.35

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H1S5	356806	252723	70.83	248796	220111	61.69
H1S6	274850	198078	72.07	193437	171403	62.36
H1S7	286596	213005	74.32	208205	175466	61.22
H1S9	275436	203089	73.73	198934	176276	64
H2S1	576796	452893	78.52	442258	360341	62.47
H2S10	334680	252103	75.33	246020	204837	61.2
H2S2	269322	202309	75.12	196722	176303	65.46
H2S3	354768	267951	75.53	262000	228838	64.5
H2S4	408168	303784	74.43	296051	252249	61.8
H2S5	289700	223225	77.05	218344	196811	67.94
H2S6	232156	178397	76.84	174930	161641	69.63
H2S7	521838	408315	78.25	397768	314858	60.34
H2S9	243296	175232	72.02	172189	158097	64.98

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H3S1	371832	290132	78.03	283562	237949	63.99
H3S10	249628	188037	75.33	184091	162446	65.08
H3S2	384984	274397	71.27	267823	237637	61.73
H3S3	336764	245544	72.91	240157	210157	62.4
H3S4	218982	161344	73.68	157103	131791	60.18
H3S5	303338	214224	70.62	209706	187308	61.75
H3S6	234218	164493	70.23	159573	139844	59.71
H3S7	191528	144016	75.19	139294	119837	62.57
H3S9	280330	199803	71.27	196033	176049	62.8
H4S1	403166	303242	75.22	297083	262335	65.07
H4S10	153882	111395	72.39	109604	102724	66.76
H4S2	390792	276314	70.71	269132	222257	56.87
H4S3	262602	197058	75.04	190887	169709	64.63

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H4S4	297110	210200	70.75	203599	174730	58.81
H4S5	117352	55789	47.54	53444	47689	40.64
H4S6	58462	17044	29.15	16229	15599	26.68
H4S7	344508	262407	76.17	257785	225609	65.49
H4S9	330054	249213	75.51	244648	216987	65.74
H5S1	16416	1038	6.32	490	473	2.88
H5S2	362266	263733	72.8	256945	217067	59.92
H5S3	228144	161272	70.69	156809	137566	60.3
H5S4	255396	182547	71.48	177931	162878	63.77

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H5S5	267486	211457	79.05	205115	175553	65.63

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H5S6	285516	217581	76.21	212785	187144	65.55
H5S7	334438	252815	75.59	247755	217600	65.06
H5S9	225648	160456	71.11	158094	146390	64.88
H6S1	429340	341844	79.62	334010	288396	67.17
H6S2	234912	156292	66.53	152371	138513	58.96
H6S3	340426	253291	74.4	247353	219445	64.46
H6S4	263090	194453	73.91	188210	163407	62.11
H6S5	277978	211968	76.25	205795	184477	66.36
H6S6	511442	399410	78.09	385169	278444	54.44
H6S7	223626	166475	74.44	162454	146211	65.38
H6S9	403172	292278	72.49	281117	224287	55.63
H7S1	462140	349318	75.59	339926	284599	61.58
H7S2	415774	292221	70.28	284089	245257	58.99
H7S3	294036	220023	74.83	213800	189227	64.36
H7S4	282056	206553	73.23	201929	184794	65.52
H7S5	303646	240109	79.08	235172	211071	69.51

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H7S6	33752	22188	65.74	20916	20365	60.34
H7S9	397612	308954	77.7	292795	197458	49.66
H8S1	381864	289905	75.92	281412	227067	59.46
H8S2	409540	322474	78.74	313557	255149	62.3
H8S3	495062	386069	77.98	378654	305893	61.79
H8S4	224976	165551	73.59	161023	136410	60.63
H8S5	328310	247157	75.28	242977	214661	65.38
H8S6	114170	61178	53.59	58736	51156	44.81
H8S7	196356	141963	72.3	138704	118978	60.59
H8S9	296382	219672	74.12	214479	185699	62.66
H9S1	229512	163446	71.21	159652	145436	63.37
H9S2	338616	257917	76.17	252192	223413	65.98
H9S3	407576	320425	78.62	313439	266852	65.47
H9S4	325824	231323	71	224617	187247	57.47
H9S5	419526	320267	76.34	313016	263986	62.92
H9S6	235340	174814	74.28	171472	161201	68.5

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
H9S7	227884	172968	75.9	166676	138231	60.66
H9S9	343688	266604	77.57	258728	211289	61.48

Appendix 14: Species relative abundance of each participants at every sampling points and all participants at S1 and S9 sampling point