Tampere University

Lauri Paulamäki

CHARACTERIZING THE ZEBRAFISH ORAL MICROBIOTA USING 16S SEQUENCING

Faculty of Medicine and Health Technology

Master's thesis

May, 2019

PRO GRADU -TUTKIELMA
Paikka: Tampereen yliopisto
Lääketieteen ja terveysteknologian tiedekunta
Tekijä: PAULAMÄKI, LAURI ILMARI
Otsikko: Seeprakalan suun mikrobiston karakterisointi käyttäen 16S sekvensointia
Sivumäärä: 58 sivua, 6 hakemistosivua
Ohjaajat: Dosentti Mataleena Parikka ja Leena-Maija Vanha-aho, FT
Tarkastajat: Apulaisprofessori Vesa Hytönen ja Dosentti Mataleena Parikka
Aika: Toukokuu 2019

TIIVISTELMÄ

Tässä pro gradu -tutkielmassani esittelen menetelmät, joita käytettiin ensimmäisissä yrityksissä määrittää taksonomisesti seeprakalan suun ja suoliston mikrobiston kolmesta seeprakalayksilöstä. Tavoitteena tutkimuksessa on auttaa selvittämään mahdollisuutta käyttää seeprakalaa malliorganismina suun mikrobiston ja sen tasapainon järkkymisen tutkimisessa ja hoitostrategioiden kehittämisessä. Tutkimuksessa käytettiin hyväksi mikrobistojen tutkimuksessa oletusarvoiseksi menetelmäksi vakiintunutta bakteerien ja arkkien ribosomaalisen pienen alayksikön RNA:ta koodaavan 16S geenin sekvensointia Illumina MiSeq sekvensaattorilla. Sekvensointikirjaston luomiseen kokeiltiin eri lähestymistapoja, joista lopulta päädyttiin käyttämään Illuminan valmistamaa kirjastonvalmistuspakkausta. Tulosten perusteella seeprakalan suun ja suolen mikrobistot vaikuttavat olevan samankaltaisia, mutta analyyseissa kuitenkin selkeästi omiksi kokonaisuuksiksi ryhmittyviä. Suun mikrobistossa runsaslukuisimpana esiintyivät firmikuutteihin kuuluvat peptostreptococcaceae heimon bakteerit ja toisiksi yleisimpinä proteobakteereihin lukeutuvat aeromonadaceae heimon edustajat. Suolen mikrobistossa kahden yleisimmän heimon järjestys on päinvastainen. Suun mikrobistot olivat tutkimuksen perusteella suolen mikrobistoja lajirikkaampia ja niiden määräsuhteet olivat myös tasaisemmin jakautuneet Shannonin indeksillä mitattuna. Tulosten perusteella seeprakalan käyttömahdollisuudet suun mikrobiston mallintamiseen vaikuttavat lupaavilta.

MASTER'S THESIS
Place: TAMPERE UNIVERSITY
Faculty of Medicine and Health Technology
Author: PAULAMÄKI, LAURI ILMARI
Title: Characterising the zebrafish oral microbiota using 16S sequencing
Pages: 58 pages, 6 index pages
Supervisors: Docent Mataleena Parikka and Dr. Leena-Maija Vanha-aho
Reviewers: Assistant professor Vesa Hytönen and Docent Mataleena Parikka
Date: May 2019

ABSTRACT

In this master's thesis I present the methods that was used in the first efforts to taxonomically characterise the zebrafish oral and gut microbiotas from three individual zebrafish. Purpose of the study is to assess the use of the zebrafish as a model organism for studying the oral microbiota and for testing treatment strategies for microbiome disruption in the oral environment. In the study, sequencing of bacterial and archaeal ribosomal small subunit 16S RNA coding gene with Illumina MiSeq sequencer was used, a method that has been established as a standard tool in microbiome studies. Different approaches for creating the sequencing library were tested, of which the Illumina library preparation kit was eventually used for creating the sequencing library. Based on the results, the microbiotas in zebrafish oral and gut environments are rather similar, yet distinguishable. The most abundant microbes in the oral environment were bacteria from the Firmicutes phylum, which belongs to the Peptostreptococcaceae family and the second most abundant family was Aeromonadaceae that is part of the Proteobacteria phylum. In the gut microbiota two most abundant families were the same as in the oral environment, but in the opposite order. In oral microbiota there were higher taxonomic diversity than in the gut microbiota and the relative abundances between taxons were more evenly distributed when measured with Shannon's index. Based on these results, the potential of the zebrafish for modelling oral microbiota seem promising.

ACKNOWLEDGEMENTS

I conducted this project as part of Infection Biology group in Faculty of Medicine and Health Technology of Tampere University. I am grateful for Mataleena Parikka for the opportunity to work with this wonderful group of people, and all the colleagues for discussing with me about the countless problems that I encountered and cheering me up when things felt overwhelming. Thanks for Leena-Maija for proofreading the literature review; I want to note that all the remaining miss-spellings and oddities are my mistakes. I would also like to thank Henna and Hanna for giving me advice for the experiments, Aleksandra for the help in the lab and the contagious positivity, Sonja for drawing the spectacular image of the biofilm and all the friends and family for helping me to have a break of the work when I needed it the most.

TABLE OF CONTENTS

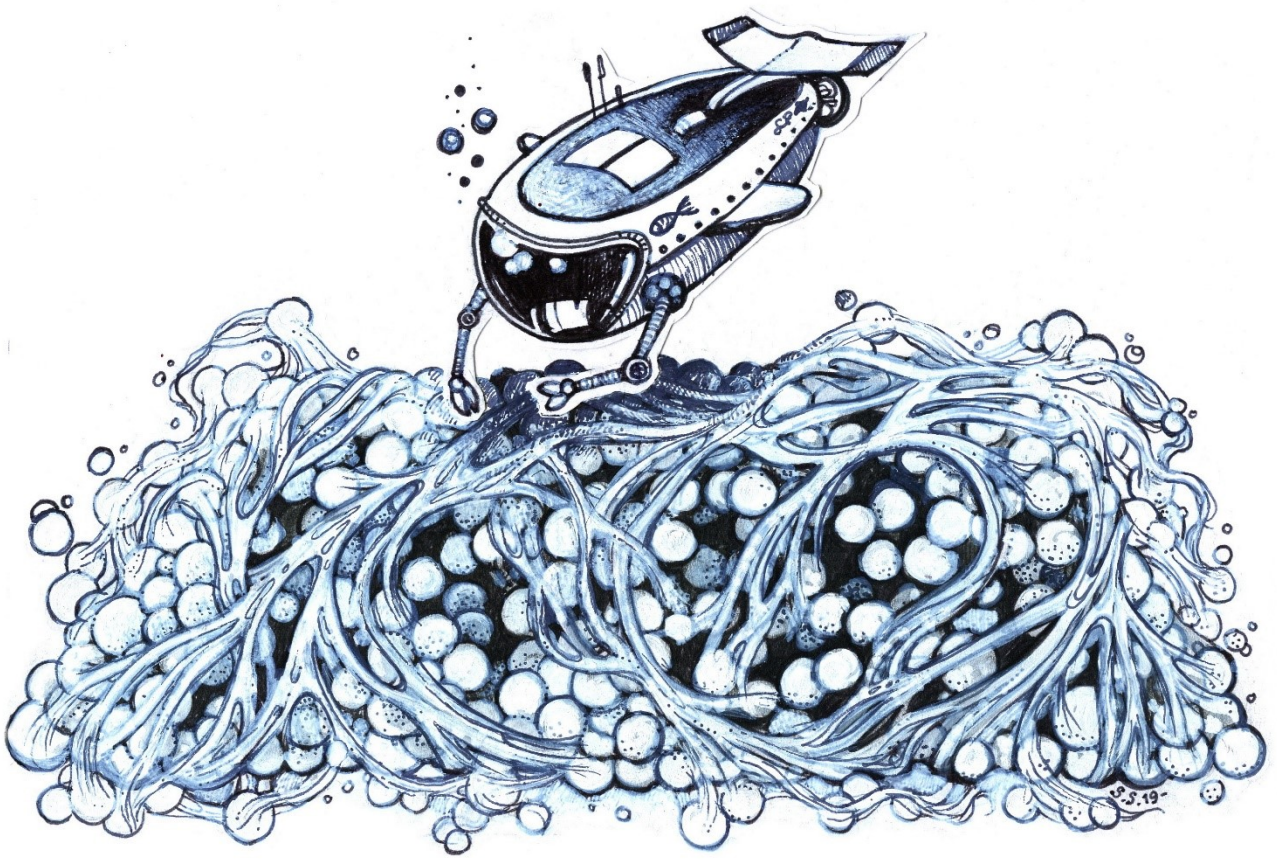

Figure 1. The artist rendering of the oral biofilms. *The bacteria (round) has excreted extracellular components (stringy fibres) for protection and attachment. The submarine is not in scale and exist solely for aesthetic purposes. Illustration by courtesy of artist Sonja Suokko, 2019.*

1 ABBREVIATIONS

CD	Crohn's disease
CNS	Central nervous system
CPR	Candidate phyla radiation
Dpf	Days post fertilization
EAE	Experimental autoimmune encephalomyelitis
FMT	Faecal microbiota transplantation
GALT	Gut-associated lymphoid tissue
IBD	Inflammatory bowel disease
NGS	Next generation sequencing
OMT	Oral microbial transplantation
PTU	1-phenyl-2-thiourea
SCFA	Short chain fatty acids
Treg	Regulatory T-cell
UC	Ulcerative colitis

2 INTRODUCTION

Microbial inhabitants residing on human epithelia have recently been understood having a role in various conditions affecting people worldwide, both in health and disease. For long time it has been known how certain bacteria associates with disease. Robert Koch formed the famous Koch's postulates about how microbial species derived from patient, grown in laboratory and injected into healthy individual causes similar disease, confirms that microbe to be causal factor for the disease. Newly acquired knowledge of microbiology has taught us to see deeper into the issue and those postulates that stayed long as unchanged rule of nature, have become if not fully outdated at least insufficient to explain all the microbial associated diseases known today. Last few decades have brought massive leaps for technology in microbiology and these new tools have given birth to entirely new branch inside the field of science. The key new tool from microbiology perspective has been the next generation sequencing that has become more efficient and cheaper to use during recent years. The taxonomic characterisations by sequencing conserved microbial genes have provided unprecedented amount of data of complex microbial colonies. It has assisted in gaining insight into microbial interactions and why dysbiosis; the state of microbiome where the microbial community and host cannot maintain the beneficial composition and metabolism of the microbiome, leads to the disease states that are visible in macro perspective. Studies have been conducted with healthy individuals, patients with diseases that are potentially associated with microbiome and in dedicated animal models. Each method contributes to the growing pool of information but has their limitations. Studies on healthy individuals tells how microbiota-host interactions are organised while homoeostasis is intact but are unable to tell which factors are necessary to maintain it. Studies on patients provide valuable data on disease states, yet the practical reasons usually limit the data to the state of microbiota where disease has already emerged, thus animal models are needed to be able to study the progression of microbial dysbiosis and its effect on disease phenotype in controlled and systematic manner. However, animal models are inevitably different from humans and that reflects also into composition and functions of anatomic microbiomes. Since neither of these means to gather information is not enough on its own, combination of them all is needed to uncover the secret life of microbes on human anatomic microbial environments. A lot is known about the gut microbiome and there are already studies with all the methods described above. As an example, microbiome disruption in gut is known to be associated with inflammatory bowel disease, which is confirmed with patients in clinical settings

and studied systematically with mouse and rat models. Despite of the anatomic differences between rodents and human, those studies have provided means to study the effect of genes and environmental factors to changing microbiome and microbiome's effect on disease progression. While gut is relatively well known, other anatomic microbiomes are less studied. Oral microbiome dysbiosis is known to relate to certain oral diseases, caries and periodontitis being the most commonly known examples. Both of those are multifactorial diseases that have strong association with dysbiosis and are among the most common diseases on the planet. While patient samples are abundant and easy to obtain, there is shortage of animal models that could be used to test treatment strategies for restoring a health enhancing microbiome composition in oral environment. This thesis presents efforts made for establishing zebrafish as an animal model for studying the oral microbial dysbiosis.

3 LITERATURE REVIEW

3.1 Microbiomes

The definition of microbiome in short is all the microscopic species that inhabit certain environment, the genes they have and express, interactions between species and the environment itself, and the environmental conditions that affect the inhabitant species. Microbiome includes all species that are contributing to the composition and function of the microbiome including yeasts, fungi, protozoans, helminths, viruses and not just bacteria that usually is the most abundant kingdom among the microbiome. As bacteria is the most well-known kingdom among microbial organisms, this thesis focuses especially on bacteria as part of the microbiome. One characteristic feature that is common for microbiomes in every environment is the dynamic nature of these micro ecosystems. Figuratively speaking, the seemingly static surface is revealed to be in constant motion once spectated closer. State of the microbiome is determined by the environmental factors and its ability to adapt into the changing conditions. Microbes are not isolated entities, only affected by the environment, but they are a part of a complex web of interactions that determines their fate in the larger microbial community. The environment and the resources it can provide for inhabitants dictates which features are required to fill available ecological niches. This thesis aims to further the knowledge about microbiomes as a part of human-microbiota supraorganism. Microbiomes of anatomic sites of vertebrates are having either direct or instrumental importance for achieving the goal and are therefore discussed in this text.

Human microbiomes can be divided roughly into four categories, i.e. skin, respiratory tract, gastrointestinal tract and urogenital microbiomes, which all are characteristically different in species composition and diversity (Huttenhower *et al.*, 2012). Each of these anatomic sites is affected by different environmental forces that are shaping the survival requirements for microbes. For example, the skin microbiota is further divided into three groups based on the properties of skin, i.e. sebaceous, moist and dry areas of skin, which all have distinct microbiomes (Byrd, Belkaid and Segre, 2018). A metagenomics study of different skin sites found that the metabolic profiles of microbiomes between these sites have major differences (Oh *et al.*, 2014). Despite differences, certain core metabolic functions exist in all skin microbiomes, the amino acid, nucleotide and lipid metabolism being the most widely shared pathways between different sites, yet significant variations in gene compositions of these pathways exist (Oh *et al.*, 2014). These differences arise from different environmental factors affecting each skin microbiomes. Moist skin, characterized by

an abundance of sweat glands, is acidic and contains antimicrobial molecules secreted in sweat. This favours microbes capable of tolerating high salt concentration and utilizing urea as nitrogen source, such as *Staphylococcus* spp. or *Corynebacterium* (Scharschmidt and Fischbach, 2013). Sebaceous skin regions favours lipophilic microbes (Grice *et al.*, 2009). *Propionibacterium acnes* is capable of cleaving arginine from skin proteins with proteases (Holland, Greenman and Cunliffe, 1979) and utilizing sebaceous gland excretions with triglyceride degrading lipases (Brüggemann *et al.*, 2004). In addition to acting as an energy source lipase degraded triglycerides forms free fatty acids, which have antimicrobial properties (Yoon *et al.*, 2018) and thus contributes to the host resistance. Microbial profiles also differ throughout life in response to changing environmental factors, such as increased sebum excretion in puberty (Oh *et al.*, 2012).

3.1.1 The oral microbiome

Similarly to skin microbiomes, which were discussed above, the microbial environment is not uniform throughout the oral cavity. On the contrary, vast differences in species composition and dominant species are evident along the salivary gradient and on different surfaces (Proctor *et al.*, 2018). A few genera of microbes make the most of the abundance of the species diversity and only a few species make up a significant proportion of the total mass of the oral microbiota (Huttenhower *et al.*, 2012; Hall *et al.*, 2017). As in other anatomic sites, each individual has a unique microbiota that is relatively stable over time (Flores *et al.*, 2014; Hall *et al.*, 2017). Some individuals have higher tendency for oral diseases due to inhabitant species of microbiome and their metabolic properties (Rosier, Marsh and Mira, 2018).

The oral microbiota is characteristically a diverse community that encounters a wide array of microbes throughout the day (David *et al.*, 2014). Most of these species are merely passing by (David *et al.*, 2014) lacking the advantageous traits needed for establishing themselves as a permanent occupant: adhesion molecules to attach themselves to mucosal pellicle (Hannig *et al.*, 2017) and to existing biofilm (**Figure 1.**) (Nobbs, Jenkinson and Jakubovics, 2011) and molecular mechanisms to process available nutrients in molecular networks (Marsh and Zaura, 2017).

Planktonic bacteria, unable to attach to oral biofilms are targeted by numerous antimicrobial substances (Bechinger and Gorr, 2017) and drained to the oesophagus by the constant flow of saliva. Saliva is a complex mix of peptides, glycoproteins, antibodies, ions, amino acids and other small molecules (Marsh *et al.*, 2016) that maintains oral homeostasis by buffering the pH (Li *et al.*, 2007) and preventing microbial growth in a number of ways including direct killing (Roversi *et al.*, 2014)

and scavenging important nutrients such as iron (Rosa *et al.*, 2017). These properties make saliva one of the most important defence systems in the oral environment, which ensures that the temporal existence of potentially harmful microbial species does not evolve into lasting inhabitation.

Where the exact species composition in each mouth is unique, there are species of certain key genera that are likely inhabitants of every microbiome. The most characteristic genera of microbes are *Streptococcus* spp., (Huttenhower *et al.*, 2012) which include members that are associated with beneficial properties for the host (Ho, Lamont and Xie, 2017) but also notorious pathogens, such as caries inducing *Streptococcus mutans* (Esberg *et al.*, 2019). Streptococcus species are important for cross species interactions, since they have a crucial role in formation of biofilms on oral surfaces (Nobbs, Jenkinson and Jakubovics, 2011). They express adhesion molecules on their cell membrane, which have a binding specificity for pellicle proteins (Brady *et al.*, 2010; Lizcano, Sanchez and Orihuela, 2012). Thus *Streptococcus* spp. act as early colonizers (Mashima and Nakazawa, 2015), which defines the further composition of the forming biofilms (Mishra *et al.*, 2010)

Streptococci are major phyla responsible for early colonization and adhesion to the acquired enamel pellicle. In the formation of oral multispecies biofilms, a group of microbial species play a key role by binding to the early colonizers. These bridge species share a common feature of expressing a set of adhesion molecules on their surface that can bind a vast selection of other bacteria (Nobbs, Jenkinson and Jakubovics, 2011). *Fusobacteria* are the most common example of these adapter species that can be found in most of the population. They attach themselves to the early biofilm with adhesins such as RadD that interacts with the streptococcal SpaP adhesin (Guo *et al.*, 2017) and expresses other adhesins, such as Fap2 (Coppenhagen-Glazer *et al.*, 2015) and FomA (Liu *et al.*, 2010) that readily bind to other microbes with a nearly universal binding specificity. *Fusobacteria* can be found in nearly all mouths and the members in this genus have both commensal and pathogenic properties (Brennan and Garrett, 2019)

Microbe-microbe interactions in oral environments are not limited to the previously described formation and adhesion to existing biofilms. One example of this kind of relationship is between *Actinomyces odontolyticus* XH001 and member of relatively recently found Candidate phyla radiation (CPR) species TM7x, where the latter is acting as an ectoparasite for XH001 (He *et al.*, 2015). TM7x lacks many essential genes, including genes required for amino acid synthesis and is thus dependent on the host for providing nutrients. Newly found interaction of these two species

indicates that there is still much to learn about oral microbial ecology. Many other CPR species are found in the oral environment as well and due to their small genome and physical size, it seems to be possible that they might be reliant on parasitic relationships with other microbes (Baker *et al.*, 2017).

3.2 The effect of the microbiome on health

3.2.1 Overview

Commensal microbes are important for the development and maturation of the immune system and therefore the disruption of these interactions might lead to severe consequences. In the gut, the commensal microbiome can induce the maturation of regulatory T-cells (Treg) with at least a few different mechanisms (Petersen and Round, 2014). The commensal gut bacterium *Bacteroides fragilis* is able to induce the conversion of gut peripheral $CD4^+$ T-cells into Tregs by producing polysaccharide A (PSA) which binds to TLR2 receptors on dendritic and T-cells. This leads to the production of IL-10 (Round and Mazmanian, 2010), a known anti-inflammatory cytokine. Certain *Clostridium* species generate the same outcome by inducing production of TGF-β cytokine, which guide the naïve T-cell to differentiate towards Tregs, with short chain fatty acids (SCFA) in the gut epithelia (Narushima *et al.*, 2014). Tregs are important for a mechanism called oral tolerance, which allows food antigens and gut commensal bacteria to pass through gut without inducing inflammation (Commins, 2015). Therefore, the disruption of these processes can lead to a decreased tolerance of self-antigens and the commensal microbiome (Petersen and Round, 2014). These two microbiome-host interactions serve as examples of the ways how microbiota can positively affect the health of the host.

3.2.2 Dysbiosis

The microbiota plays a pivotal role in host health and therefore, it is not surprising to learn that an abnormal microbiota can lead to variety of consequences from mild to severe. This chapter gives a few examples of how the so called microbial dysbiosis can cause certain diseases. The state of the microbiota where the homeostasis is disrupted is called dysbiosis. It is the result of a change in key environmental factors that drives and maintains the default state of the microbiota. Changes that precede the dysbiotic microbiome includes shifts in behavioural or dietary habits (Vallès *et al.*, 2018; Xu *et al.*, 2019), antibiotic usage (Stanisavljević *et al.*, 2019) or non-antibiotic medicine (Le Bastard *et al.*, 2017), and they lead to a decrease in the overall species diversity, loss of beneficial microbes

and potentially, an introduction of pathogens or their increased abundance. Dysbiosis is reported to be a factor in many inflammatory and autoimmune diseases including cancer (Gagnière *et al.*, 2016). Dysbiosis can also act in concert with disease predisposing genes in causing disease phenotype. In inflammatory bowel syndrome (IBD) mutation in the TLR5 gene is associated with IBD. In a study that was conducted with $TLR5^{-/-}$ knockout mice, the IBD disease phenotype was only seen, when the mice gut microbiota contained certain proteobacteria such as *Escherichia coli* (Carvalho *et al.*, 2012).

In a study conducted with experimental autoimmune encephalomyelitis (EAE) Dark Agouti rats, orally administered antibiotics disturbed the gut microbiota, which led to increased inflammation in the central nervous system (CNS) (Stanisavljević *et al.*, 2019). A shift in the microbiome was observed along with increased gut-associated lymphoid tissue (GALT) activation. The proportion of *Firmicutes, Actinobacteria*, *Clostridia* and *Bacilli* were decreased, while *Proteobacteria*, *Bacteroidetes* and gammaproteobacteria increased in abundance. This led to a decrease in total abundance of SCFA in the gut and to a following inflammation in the GALT that resulted in the failure of regulating autoimmunity in CNS. This study shows that dysbiosis can have far reaching severe consequences.

Studies on the role of the environment on the human microbiome has shown that lower environmental diversity near home during childhood negatively correlates with increased susceptibility for atopic dermatitis (Hanski *et al.*, 2012). The general diversity of gammaproteobacteria in the skin microbiota was significantly lower in atopic study subjects than in the control group. The control group had also significantly higher anti-inflammatory cytokine IL-10 levels than the atopic group. IL-10 levels correlated with abundance of gammaproteobacteria *Acinetobacter* on skin. The study suggests that environmental microbes have a role in host immunomodulation and in the overall health of the population, thus enforcing the hygiene hypothesis.

3.2.3 Oral dysbiosis

Most of the dysbiosis studies concentrate on the gut microbiome, and much more data have been accumulated on gut dysbiosis related diseases than on other anatomical site microbial environments. The oral cavity has its share of diseases, where dysbiosis serves as a causative factor. All types of oral cancers combined are among the deadliest cancers in the world (Ferlay *et al.*, 2015), and recent studies have shown a correlation between microbial dysbiosis and oral cancers, yet no

causality can be drawn from the results (Banerjee *et al.*, 2017). Microbial profile changes have been notable in the abundances of certain key genera, while species composition have remained constant between tumour patients and controls (Wang *et al.*, 2017). Actinobacteria and especially the genus *Actinomyces* were decreased in abundance, while the genus *Parvimonas* was increased in tumour patients (Wang *et al.*, 2017).

One well-known and extensively studied oral disease is caries, a disease in which the dental enamel is slowly demineralized by acidifying bacterial metabolism. Caries has long been known to be caused by bacteria, but only relatively recently has the multifactorial nature of the disease begun to unravel (Pitts *et al.*, 2017). The likelihood of caries formation is increased when following conditions are met: a frequent dietary sugar intake (Moynihan and Kelly, 2014), poor oral hygiene (Kumar, Tadakamadla and Johnson, 2016), and also the caries inducing composition of microbiome . A group of acid producing and acid tolerating bacteria, not limited to the notorious caries pathogen *Streptococcus mutans,* are co-operatively forming the disease promoting state (Simón-Soro and Mira, 2015). Many bacterial species have been shown by 16S gene sequencing to co-exist in caries lesions, yet mechanistic studies of the interactions are not abundant. However, in one *in vitro* study, *S. mutans* and *A. actinomycetemcomitans* were co-cultivated in artificial saliva and the latter species was shown to activate *S. mutans* quorum sensing, which led to the formation of a dual species biofilm and *S. mutans to* change its gene expression profile from oxidative stress related genes to anaerobic metabolism, chaperones and iron acquisition (Szafrański *et al.*, 2017). The virulence of both species was enhanced as the facultative anaerobe *S. mutans* could form anaerobic conditions inside the biofilm and *A. actinomycetemcomitans* was not able to form biofilm without the presence of *S. mutans* (Szafrański *et al.*, 2017).

Even though the oral environment contains a wide variety of fungi, there are relatively few studies concentrating on it. To date more than 75 genera have been found in healthy mouths (Ghannoum *et al.*, 2010). As the number of studies on mycobiome remains low, not much is known about all the inhabitant fungi in the mouth, but some of them are known opportunistic pathogens, which cause infections especially in immunocompromised patients. As an example, *Candida albicans* is commonly seen to cause oral candidiasis in AIDS patients (de Repentigny, Goupil and Jolicoeur, 2015). Apparently, the microbiome is regulating the abundance of oral *C. albicans*, *Aggregatibacter actinomycetemcomitans*, a known oral pathogen itself (Fine *et al.*, 2007), is inhibiting *C. albicans* growth via quorum sensing by binding to its hyphae (Bachtiar *et al.*, 2014).

3.2.4 Possibilities in medicine

As the research community has learned about the importance of host-microbiome interactions in health and disease, it has given rise to novel approaches to treat the affected conditions. To date the most successful example that has paved the way for the treatment of other conditions, is faecal microbiota transplantation (FMT), which has been used for the treatment of *Clostridium difficile* derived diarrhoea (Cammarota, Ianiro and Gasbarrini, 2014) and ulcerative colitis (UC) (Moayyedi *et al.*, 2015). UC is a gut disorder, which is caused by changed microbial environment in the gut that results in chronic inflammation (Ungaro *et al.*, 2017). The onset of UC includes genetic predisposition in certain leukocyte and barrier function associated genes and the administration of several drugs including non-steroidal anti-inflammatory drugs and oral contraceptives (Ungaro *et al.*, 2017). Fecal microbiota transplantation has shown significant increase in the remission rate compared to a placebo: 24% vs. 5% of patients respectively entered remission in this double-blinded clinical trial (Moayyedi *et al.*, 2015). After the treatment, FMT treated patients' gut microbiome was significantly more diverse and reminded that of the donor's. Depending on the donor, either the enrichment of Lachnospiraceae and *Ruminococcus* or *Escherichia* and *Streptococcus* was observed (Moayyedi *et al.*, 2015).

These early FMT treatment studies provide an interesting and novel approach to treat certain microbiome associated diseases. To date, most studies have concentrated on the gut microbiome. While some promising results have been seen, there remains a need to elucidate the mechanistic principles of microbe-microbe and microbe-host interactions to make these microbiome substitutions a viable and safe option for treatment. Similar strategies have been visioned for microbiomes colonizing other human anatomic locations, as basic principles behind dysbiosis associated diseases remain the same throughout microbiomes; a certain composition of the microbiota is a driving factor for disease and therefore substituting the dysbiotic microbiome with the microbiome of a healthy donor should alleviate symptoms of the disease.

Different strategies for the treatment of oral microbiome dysbiosis have been proposed, including treatment with probiotics or with oral microbial transplantation (OMT). The probiotic approaches have been attempts to affect the oral microbiota by introducing single or few characterised microbial species to dysbiotic subjects, but more evidence for their efficiency and safety are needed (Gruner, Paris and Schwendicke, 2016). The OMT strategy is very similar to FMT, yet the practical administration of the microbiome transplant has some key differences due to the characteristics of

the oral microbial environment (Nascimento, 2017). The proposed method for performing the transplantation can be divided into three steps: first, the collection of microbiota from healthy donor, second, the removing of the dysbiotic microbiota from the patient and third, the transplantation of the harvested microbiome to the patient (Pozhitkov *et al.*, 2015). In the first step the problem lies in recognising the composition of a healthy microbiome, although it has been the subject of many studies, which have provided the means for estimating the pathogenicity of the microbial composition (García-Jiménez and Wilkinson, 2019). Stabile microbiotas have intrinsic ability to resist changes in the species composition (Rosier, Marsh and Mira, 2018). Therefore, the attempts to reinstate a healthy microbiome requires disturbation of prevailing conditions. Multiple techniques for achieving that has been examined. In addition to rigorous oral hygiene measures, bactericidal mouth rinses have been used, chlorhexidine being one and sodium hypochlorite another tested substance. Chlorhexidine has been shown to be effective in killing oral bacteria in attempts to boost the efficiency of oral probiotic administration (Aminabadi *et al.*, 2011). However, due to its property of adhering to oral mucosal tissue, it remains bactericidal long after administration (Cousido *et al.*, 2010) and thus also affects the transplanted healthy microbiome. Sodium hypochlorite has been proven to be a better alternative to chlorhexidine as its bactericidal effects can be neutralised with a sodium ascorbate buffer (Pozhitkov *et al.*, 2015).

3.2.5 Animal models of dysbiosis

There are already some animal models for studying dysbiosis, the most common organism in use being mouse and rat, but also zebrafish (He *et al.*, 2013) and even *Drosophila melanogaster* (Lee and Lee, 2014) have been used. Different forms of IBD are the most studied of dysbiotic diseases and that reflects the abundance of animal models available. In mice, knock-out strains of certain IBD related genes have been made. The intracellular pathogen receptor $NOD2^{-/-}$ knock-out mice combined with an infection of protozoan *Giardia muris* and *Toxoplasma gondii*, and indomethacin results in a T-cell mediated disease state that resembles Crohn's disease (CD) in the mouse ileum (Craven *et al.*, 2012). In this model effect of different microbial compositions and treatment strategies for the disease are easily studied. In the same $NOD2^{-/-}$ knock-out study, another mouse model was presented for CD, in which the pro-inflammatory cytokine receptor CCR2 was knocked-out. These $CCR2^{-/-}$ mice show a significant decrease of the CD phenotype and dysbiosis when infected with *T. gondii* (Craven *et al.*, 2012)

Lee and Lee (2014) reviewed the use of fruit fly *D. melanogaster* for studying the role of gut microbial dysbiosis for IBD-like disease condition. The article also summarises the principles of performing gut microbial community studies by using a standardized *Drosophila* gut commensal microbiome in sterile conditions to control the effects of treatments or for example diets. The *Drosophila* has also been used for studying the effects of dysbiosis on Alzheimer's disease (Wu *et al.*, 2017). The study found that vacuolar degeneration in the human amyloid $A\beta_{42}$ expressing fly brains was significantly increased, when the gut microbiome was disturbed by colonization of non-pathogenic enterobacteria. Similar effects had been previously seen with antibiotic induced gut microbiome in the Alzheimer's disease murine model (Minter *et al.*, 2016), which strengthens the hypothesis of the role of gut dysbiosis in the disease progression and gives credibility for the use of *Drosophila* as a dysbiosis model.

Dysbiosis related diseases are a result of a complicated microbe-microbe and microbe-host relationships that pose a challenge for researchers as it is nearly impossible to take all the variables into account. One of the solutions for this problem, has been establishment of gnotobiotic, or in other words germ-free, animal models (Mooser, Gomez de Agüero and Ganal-Vonarburg, 2018). These models aim to minimise the randomness in datasets while still exploring the complex web of interactions. Zebrafish is one of the animals, in which a gnotobiotic research model has been established (Wang *et al.*, 2018) and it has been used to study microbial colonization, for example.

The above-mentioned examples illustrate that it is not only possible to model dysbiosis in taxonomically distant species, but that these models also provide valuable information as systematic clinical dysbiosis experiments with human patients are not possible.

3.3 Choice of methods

The methods used in this project have been widely used in the microbiome studies. This whole field of science is still relatively new and in practice emerged from the development of an easy and accessible sequencing technologies that could be used for the taxonomic classification of multiple species and samples in a single sequencing run. Thus, until recently most of the studies had been performed with a comparatively limited set of techniques that formed the golden standards of the field. Sequencing library is a pool of barcoded DNA fragments from different samples, which are

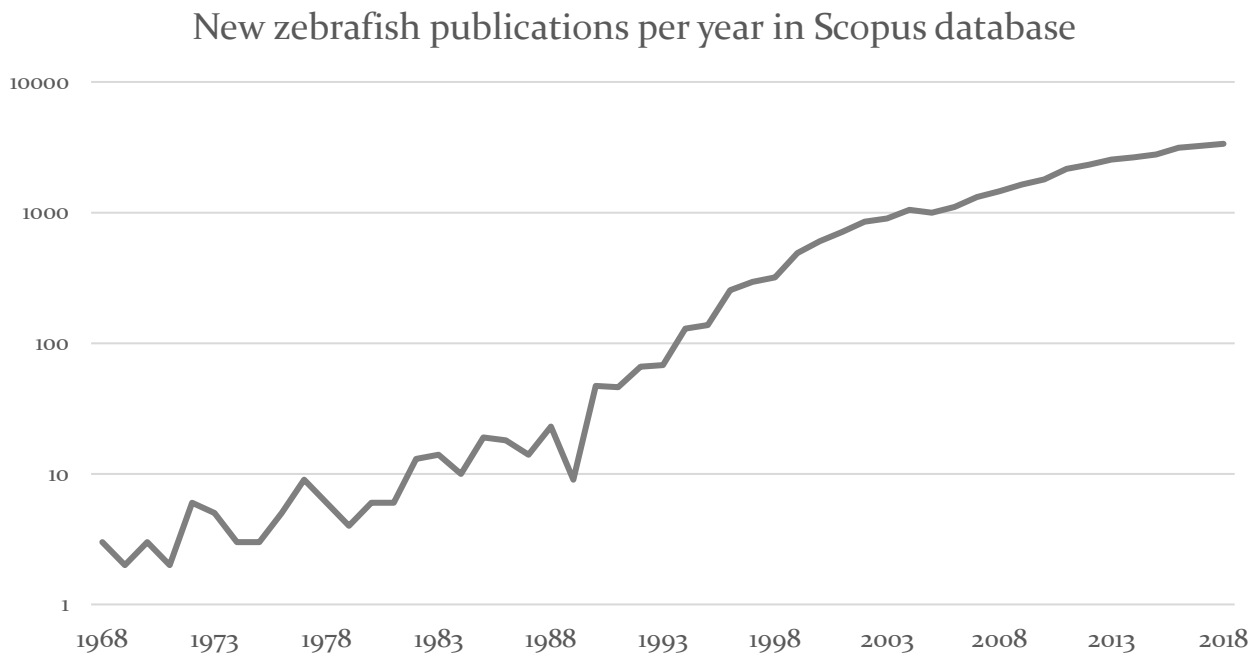

Figure 2. The number of publications in the Scopus database that has "zebrafish" in the title, abstract or in the keywords.

tagged with sequencing adapters. Library creation per se is the key to a successful sequencing run, as it determines the quality of reads and the sequencing run in general. The sequencing run itself requires little input from the researcher and is fully automated after the sample library is loaded into a machine and until the reads are generated. In this chapter, the methods and tools of choice are discussed considering previously published studies, especially the use of zebrafish as a model organism and the selection of the sequencing technology.

3.3.1 Zebrafish as a model organism

Zebrafish has been used as a model organism since 1960's and it has been gaining popularity during the past couple of decades, which can be seen in the increase of the number of publications in databases (**Figure 2.**) (Scopus, 2019). The zebrafish is a good model organism for several reasons. One of the advantages is the rapid reproduction that produces offspring in abundance (Lawrence, 2007). Females are spawning eggs continuously after reaching maturity (Lawrence, 2007) and a single breeding session with only few fish can produce hundreds of fertilized eggs. Eggs hatch a few days post fertilization (dpf) and the zebrafish reach maturity at around 3 months of age (Spence *et al.*, 2007). Zebrafish is also easily genetically manipulated (Garcia, Noyes and Tanguay, 2016) using technologies such as the CRISPR-Cas9 for genome editing or morpholinos for gene knock-outs.

Zebrafish larvae are not regulated by the legislature for research animals, thus a lot of research is carried out in the larval model. In Finland, the use of the zebrafish under the age of 7dpf does not

require permission, but experiments conducted with older zebrafish must be approved by an ethics committee. The use of zebrafish larvae is convenient also due to their innate lack of pigmented cells during the first days after fertilisation (Tobin, May and Wheeler, 2012). The transparency of the larvae makes it possible to follow the development of organs in real time, and the usage of fluorescent markers is not uncommon. It is possible to postpone the formation of pigment by exposing the larvae to 1-phenyl-2-thiourea (PTU) (Karlsson, von Hofsten and Olsson, 2001). In microbiome studies, the transparency of the larvae has been taken advantage of, by using fluorescently labelled microbes for studying the colonization of gut microbiota in real time (Wiles *et al.*, 2016).

One of the advantages over other non-mammalian model organisms is the immune system of the zebrafish, which includes both the innate and the adaptive arms of the immune defence (Tobin, May and Wheeler, 2012). Thus, the zebrafish immune system reminds the human immune system more than that of for example *D. melanogaster,* that solely relies on the innate immune defence against pathogens on innate immune cells (Parsons and Foley, 2016)

3.3.2 The zebrafish in microbiome studies

The zebrafish is not the obvious choice for conducting microbiome studies aiming to increase the understanding of human health. However, its gut microbiome has been first characterized using 16S sequencing in 2006 (Rawls *et al.*, 2006), and the zebrafish has since been used to study for example the effect of environmental anti-microbial agents for the gut microbiota (Gaulk*e et al*., 2016), and the bacterial colonization of gut (Wiles *et al.*, 2016). These studies have demonstrated that even if the core microbiota is drastically different at the taxonomical level from the human microbiota, the nature of the interactions between microbial species and between the microbes and the host, follows general patterns that are evident throughout the biosphere.

It could be argued that zebrafish grown in a clean laboratory environment cannot be compared to individuals living in their natural habitat, and that this decreases their suitability for microbial ecology studies. This assumption was demonstrated to be false, as the core gut microbiota is strikingly similar between populations from five different zebrafish research facility and a wild population (Rawls *et al.*, 2006). This finding suggests that it is possible to use laboratory-grown zebrafish in microbiome studies.

3.3.3 16S sequencing

The small ribosomal subunit 16S RNA coding gene has provided microbiologist with an invaluable tool to characterize the species diversities of different bacterial environments (Klindworth *et al.*, 2013). The 16S gene is well suited for the purpose, since it has both highly variable regions and conserved regions that are relatively similar across the bacterial kingdom (Van de Peer, Chapelle and De Wachter, 1996). By using primers targeting the conserved regions that are flanking the more variable regions, researchers can multiply and sequence the region in between (D'Amore *et al.*, 2016). 16S gene have enough diversity and length to justify taxonomic characterization on a genus level based on the sequence (Callahan, McMurdie and Holmes, 2017), and it has been used for the purpose since 1985 (Pace *et al.*, 1985). True potential of taxonomically characterizing microbes, was unleashed when the development of massive parallel sequencing techniques lowered the cost of sequencing per base to fit to a budget of a medium sized research group (Klindworth *et al.*, 2013). This led to the establishment of a new field of science that studies the microbial environment as a complex system.

The whole 16S gene is approximately 1500bp long, and therefore the most common sequencing technologies used for the purpose are not capable of sequencing the whole gene with single read. There are nine different variable regions in the gene; V1-V9 (D'Amore *et al.*, 2016) which allows selecting suitable sized fragment for each technology. The most common region to be sequenced with the Illumina MiSeq sequencer (Illumina Inc., San Diego, California, USA), is the V3-V4 region between bases 341 and 806 (Zheng *et al.*, 2015), which totals a fragment of 465bp. This fragment is frequently selected since the single primer pair is able to target a large proportion of the bacterial species listed in SILVA database of microbial rRNA genes (Yilmaz *et al.*, 2014).

3.3.4 Illumina MiSeq

Due to its availability and affordability, the Illumina Miseq has been a widely used sequencing technology for 16S sequencing since its introduction in 2011. The technology allows multiple reads to be sequenced simultaneously in a single flow cell without the need of cloning fragments into prokaryote cells (Mardis, 2013). The technology relies on so called sequencing by synthesis, which uses reversible dye terminators in the growing DNA strand to determine the incorporated nucleotide (Bentley *et al.*, 2008).

This chapter summarizes the core concepts of the Illumina MiSeq sequencing technology, as described by Bentley *et al.,* (2008). On the sequencing flow cell, library DNA fragments are annealed

to adapter sequences conjugated onto the surface of the flow cell. These fragments are then amplified into clusters using bridge PCR, which is a type of PCR that utilizes surface conjugated DNA adapter sequences for binding library molecules onto the flow cell and keeping them spatially organized while multiplying the sequence. When these fragments are amplified, they form distinct and separated clusters of around 1 µm in diameter containing only copies of a single DNA sequence**.** After cluster generation, the orientation of the reads is made uniform by enzymatically cleaving the reads oriented in opposite direction. The sequencing occurs by synthesizing the new strand by incorporating fluorophore coupled 3'-O-azidomethyl 2'-deoxynucleoside triphosphates into the growing sequence. Fluorophores are excited with lasers and the whole flow cell is then imaged. Tris(2-carboxyethyl) phosphine is used to remove the fluorophores from the incorporated nucleotides and to recover the 3' OH-group for the next round of synthesis. It is possible to increase the sequence length by using paired end sequencing, where the fragment is first sequenced in one direction, flipped to another orientation and sequenced again. This decreases the drop in quality as the sequencing results in a longer overall read length while the overlaps are merged in data analysis.

3.3.5 Data analysis; OTUs versus ASVs

The sequencing data from Illumina MiSeq paired end sequencing is stored in two paired Fastq files that contains sequences and respective quality information of each sequenced base (Schloss *et al.*, 2009). The sequencing parameters can be set to organize the data from each sample into own folders, which eases the division of the reads to correct samples during later steps. For the analysis, the sequence data is stripped of the adapter and index sequences and the order of the sequence in these two paired files correspond to the location in a flow cell, which allows the merging of reads in downstream analysis (Callahan *et al.*, 2016).

The first standard and still a widely used method for 16S data processing has been to compare the sequence data together to map the reads into clusters of similar sequences using algorithms such as Mothur (Schloss *et al.*, 2009) or UPARSE (Edgar, 2013). By selecting the threshold for maximum differences, these reads are clustered into groups of closely related sequences, so called Operational Taxonomic Units that should not be confused with species. These OTUs are then compared to databases such as SILVA (Yilmaz *et al.*, 2014) or Greengenes (DeSantis *et al.*, 2006) to make a taxonomic annotation for the data. (Schloss *et al.*, 2009; Edgar, 2013; Callahan *et al.*, 2016) The stricter the threshold, the better resolution the method gives, yet lowering the abundance of reads in a single unit, whereas a broader threshold lowers the resolution and increases the reliability

(Edgar, 2018). However, a recently published article speculates, whether the universally accepted 97% threshold is out-dated as the databases have improved due to the increase in the amount and the quality of data (Edgar, 2018) This approach relies on taxonomic characterizations for the depth of sequencing in a given data set as OTUs are not constructed in relation to previous data sets, thus annotations can be reliably done on a rather general level (Callahan, McMurdie and Holmes, 2017).

Another method for classification is to divide the sequencing reads into so called Amplicon Sequence Variants (ASV) (Callahan, McMurdie and Holmes, 2017)**.** This approach differs from the OTUs in that reads are not clustered into groups of similar sequences, but each variant is dealt as an individual entity (Rosen *et al.*, 2012)**.**The rationale behind this approach is much better resolution since as the knowledge of the species and sequence data accumulates, these ASVs can be compared to existing data in the databases. This makes it possible to compare data between different sequencing runs in a way the OTUs could not, as the OTUs are defined again in each data set.

4 AIMS OF THE STUDY

Aim of my master's thesis project was to assess the suitability of the use of zebrafish as a model organism for oral microbiota and its dysbiosis. There are ambitions to find versatile and cost-effective model for screening of treatment strategies for the microbial dysbiosis in oral environment, and for that purpose zebrafish is promising candidate. The main goal of this project was to find suitable methods for sample collection from the zebrafish oral and gut microbiotas, DNA extraction of small sample volume, an efficient library creation method, and optimize these methods into efficient protocol that can be used for future studies. Secondary goal was to characterise the oral and gut microbiotas.

5 MATERIALS AND METHODS

The samples were collected from healthy adult female zebrafish from wild type population. Zebrafish in this study were grown in Zebrafish Core Facility of Tampere University (Faculty of Medicine and Health Technology), according to the laws and regulations for ethical treatment of laboratory animals. To avoid pain, which the sampling methods could inflict, the fish were euthanized in overdose of anaesthetic tricaine methanesulfonate before any samples were collected. All the experiments in this study were conducted implementing 3R principles, by using as few individual animals as possible without reducing quality of results. Guidelines for euthanasia and minimizing the production of unnecessary stress for the laboratory animals were followed through the study.

5.1 Sample Collection

Samples from zebrafish skin were collected by scraping with sterile scalpel and transferring the mucous secretion into bead beating tube. For collection of oral sample, the fish was embedded into foamed plastic bed and the sample was collected using sterile interdental brush. Bristles of the brush was cut into bead beating tube. The gastrointestinal sample was collected by dissecting the fish with scalpel and transferring the whole gut into bead beating tube. All the samples were kept on dry ice until DNA extraction. Each sample was given code, which was describing which sample set it was from, the anatomic location of the sample i.e. gut or mouth, and fish identifier. For example, gut sample from fish number six of sixth sample set is named 6G-6.

5.2 DNA extraction methods

DNA extraction kits were tested to find method that preserves as much of the DNA sample as possible. Five extraction methods were used; phenol-chloroform extraction, QIAamp DNA Stool Mini kit (Qiagen, Hilden, Germany) and NucleoSpin DNA Stool (Macherey-Nagel, Düren, Germany) kit for gut samples and phenol-chloroform, PureLink Microbiome DNA Purification Kit (Thermo Fisher, Waltham, Massachusetts, USA), Quick-DNA Fungal/Bacterial Microprep Kit (Zymo Research Corporation, Irvine, California, USA), and QIAamp DNA Stool Mini kit for oral and skin samples. Every silica column-based DNA extraction kits were used following the protocols.

5.2.1 Phenol-chloroform extraction

Phenol-chloroform extractions were conducted according to the protocol by Zhu (2014). 1300µl of Tri-reagent (Molecular Research Center Inc., Cincinnati, Ohio, USA) was added to each bead beating

tube. To enhance the yield, bead beating of 3000 rpm for 1:30 minutes with PowerLyzer 24 homogenizer (Qiagen, Hilden, Germany) was performed twice. Samples were then sonicated in water bath for 9 minutes to also break the gram-negative bacteria. After these initial steps the phase separation was performed, and the DNA was eluted to 50 µl of sterile Milli-Q water.

5.2.2 Comparison of DNA extraction methods

The DNA concentration was measured by UV absorbance with NanoDrop 2000 (Thermo Fisher, Waltham, Massachusetts, USA) instrument. The instrument was blanked for each sample with respective elution buffer and measuring was done using 1.5 µl of sample. Each droplet was read three times to calculate the mean. Mean was calculated to minimize possible variability caused by instrument and low sample concentration.

5.3 Primer design

Even though Illumina has its own Nextera XT DNA Library Prep Kit (Illumina Inc., San Diego, California, USA) for creating a sequencing library for its MiSeq sequencer, there are numerous alternative ways reported in the literature (Fadrosh *et al.*, 2014; Zheng *et al.*, 2015). Each of the approaches has their own advantages and disadvantages, but common nominator is usage of PCR for attaching the sequencing adapters for the target sequence. Main reasons for the existence of multitude of approaches is arguably in the costs of the Illumina kits and the rigidity of the protocol. Alternative sequencing library protocols allow creating a library from thousands of samples for the same price as the Illumina Library Preparation kit for 24 samples costs.

Sequencing library primers consists of multiple functional parts (**Table 1.)** First part from 5' to 3' are the adapter sequences. Adapters are essential for the sequencing reaction, as they are used to generate the clusters by attaching the library molecules to the sequencing flow cell (Bentley *et al.*, 2008). There are two different adapter sequences to ensure that all the sequencing reactions within cluster happens to the same direction. Next part is 8 nt long index sequence, which is required for multiplexing of samples in one sequencing run. Every sample is amplified using unique pair of index

Table 1. Library PCR primer design. *First parts are the P7 and P5 sequences, that are used to attach libraries to the flow cell during sequencing, then comes 8 nt long barcode sequence, which are used to identify sequences from different samples in multiplex PCR, pad sequences are for sequencing primer attachment, spacer is 0-3 nt long sequence that reduces homogeneity while sequencing highly conserved 16S gene, linker is selected to be as dissimilar to sequence before targeted gene region and the last portion of the primers are complementary to the targeted gene region.*

P7 seq	**i7**	**pad**	**Spacer**	**linker**	**Rev primer**
CAAGCAGAAGACGGCATACGAGAT	TAAGGCGA	GCAGGTCGGTCGGTCGACCG	TGA	CA	GGACTACHVGGGTWTCTAAT
P5 seq	**i5**	**pad**	**Spacer**	**linker**	**Fwd primer**
AATGATACGGCGACCACCGAGATCTACAC	CTCTCTAT	CGTAGCCTGCCTCCGGTGAGG	GAC	TG	CCTACGGGAGGCAGCAG

sequences, which allows for the sequencing reads to be identified and allocated to correct samples during the data processing. The same index sequences as in the official Illumina index kit were used in this project, since those are carefully selected to function in multiplexing when varied number of samples are used.

Pad sequences are used to prime sequencing reaction by binding the sequencing primers after cluster formation. Pad sequence in P7 end of the library has dual function as it is also used for priming the i7 index. T_m of the library PCR primers are most easily modified by altering the pad sequence.

Spacers are 0–3 nt long sequences designed to desynchronise the sequencing reaction while sequencing highly homogenous fragments, such as 16S gene (Fadros*h et al*., 2014). Identifying nucleotides during sequencing is based on four fluorescence dyes, each excited using different wavelength. To record each nucleotide, the flow cell is then excited with all these different wavelengths and then imaged. In each sequencing cycle, these dyes should be represented in equal proportions to ensure that sequencer is able to resolve the signals. Using the spacer also decrease the need for PhiX library, which is homogeneity increasing sequencing library added to the final sequencing pool.

Linker sequence is a pair of nucleotides that are selected based on the target sequence. Optimal linker has minimum complementarity to the target region, which allows it to sequester sequencing overhang from the target gene, thus enhancing the PCR amplification of libraries.

Last part of the primer is approximately 20 nucleotides long and it is designed to be complementary to the target sequence.

5.3.1 Testing of primer pairs

16S region targeting primer sequences were designed with TestPrime function in SILVA database (Yilmaz *et al.*, 2014). The function compares primer sequences from user input to the 16S sequences in the database and returns information of the matched sequences. The search results of the function can be fine-tuned by adjusting few parameters, including number of allowed miss-matches per primer. Function returns list of organisms which have 16S gene sequence matching the primer pair within user defined threshold and the percentage value of the species coverage. f341 (Muyzer, de Waal and Uitterlinden, 1993) forward primer and 806R (Caporaso *et al.*, 2011) reverse primer, resulting in 465 bp fragment targeting V3-V4 regions, was selected based on the TestPrime function.

5.4 Library creation in single PCR run

Every PCR reaction in this project was done using AccuPrime Pfx SuperMix (Thermo Fisher, Waltham, Massachusetts, USA). Other components and their respective concentrations varied between experiments as aim of these experiments were to find optimal conditions for the PCR. Evaluation of PCR amplification was made using agarose gel electrophoresis.

5.4.1 Mycobacterium marinum culture for positive control

A 1ul loop of cultured *M. marinum* from 7H10 agar dish with 10% OADC enrichment and 0.5%v/v glycerol was diluted into 10 ml of 7H9 medium, supplemented with 10% ADC enrichment, 0.2% tween80 and 0.2% v/v glycerol in 25 ml cell culture flask. The culture was grown in 28 °C for 3 days at dark without shaking. Then the optical density with 600 nm wavelength (OD600) was measured and culture diluted 1:10 to fresh 7H9 medium. After 2 days incubation OD600 was measured again to evaluate the bacterial concentration. *M. marinum* from 1 ml of liquid culture was collected by centrifugation at 10,000 g for 3 min and the bacterial pellet was resuspended and diluted into 1:10000 of fresh 0.2 µm sterile filtered PBS solution.

5.4.2 Annealing temperature optimization

The annealing temperature was searched with temperature gradient PCR. The total volume of the reaction was 10 µl, of which 8.5 µl was AccuPrime Pfx SuperMix, 0.5 µl of forward and reverse primer, and 0.5 µl of the template DNA. The reaction conditions were as follows; initial denaturation in 95 °C for 2 minutes after which 30 cycles of denaturation in 95 °C for 20 seconds, annealing with annealing gradient from 54°C to 70°C for 15 seconds, and extension in 72 °C for 5 minutes. The final extension was in 72 °C for 10 minutes. Same protocol was performed for gut samples and *M. marinum* control sample to assess the suitability of the positive control.

		95°C	2 min
x 30	[	95°C	20 sec
		54 - 70°C	15 sec
		72°C	5 min]
		72°C	10 min
		4°C	→

5.5 Library creation in single PCR run with two annealing temperatures

The adjustment of annealing temperature mid-run can be performed to increase the yield with marginal loss of specificity. When the annealing temperature is lowered for few initial rounds, the specificity is slightly decreased but the amplification is more effective. Therefore, the ratio of template to background DNA is elevated prior applying the higher annealing temperature, which leads to better efficiency of the reaction without sacrificing the specificity.

5.5.1 *Annealing gradient experiment*

Table 2. Two- stage PCR annealing gradient panel. *Each column represents one 1. stage annealing temperature and each row 2. stage annealing temperature. The program used is presented on the right.*

2. Gradient \ 1. annealing radient				
	43,3°C 44,4°C	44,6°C 44,4°C	46,3°C 44,4°C	47,0°C 44,4°C
	43,3°C 46,4°C	44,6°C 46,4°C	46,3°C 46,4°C	47,0°C 46,4°C
	43,3°C 49,0°C	44,6°C 49,0°C	46,3°C 49,0°C	47,0°C 49,0°C
	43,3°C 50,0°C	44,6°C 50,0°C	46,3°C 50,0°C	47,0°C 50,0°C

	98°C	5 min	
x 6 [	98°C	30 sec	
	43,3-47°C	30 sec	
	72°C	2 min	]
x 35 [	98°C	30 sec	
	44,4-50°C	30 sec	
	72°C	2 min	]
	72°C	5 min	
	4°C	→	

Different combinations of annealing temperatures were tested in two-dimensional temperature gradient PCR (**Table 2.**). In the experiment the first annealing temperature was aligned along the rows of the PCR plate and second annealing temperature along the columns. Due to the limitations of the machine, the plate had to be turned 90° between the switch of the annealing temperature. For this test two different runs were performed, another with sample collected from the oral cavity of the zebrafish and another from the extracted gut sample of the same individual. The PCR instrument was programmed to run following program: initial denaturation in 98 °C for 5 minutes, then 6 cycles of denaturation in 98 °C for 30 seconds, annealing in temperature gradient from 43.3 °C to 47 °C for 30 seconds, extension in 72 °C for 2 minutes. Then the protocol was paused, and the plate turned 90° so that the temperature gradient was orthogonal compared to the first stage. Then protocol was immediately continued with 36 cycles of denaturation in 98 °C for 30 seconds, annealing gradient from 44.4 °C to 50 °C and extension in 72 °C for 2 minutes. Final extension is performed in 72 °C for 5 minutes.

5.5.2 *Additive PCR panel*

Multiplying the 16S library has proven to be difficult, possibly due to inhibitors in the remaining in the samples after extraction (Schrade*r et al*., 2012). To overcome the effects of these PCR inhibitors, certain additives can be used to increase the reaction efficiency. In this PCR experiment, different combinations of these additives were used. The total reaction volume in each sample was 10 µl of which 7.5µl AccuPrime Pfx SuperMix, 0.5µl of each primer, 0.5µl of template and 1 µl of additives or sterile Milli-Q water. The **table 3.** visualises the configuration of the additive panel. The additives used in this experiment were DMSO (Thermo Fisher, Waltham, Massachusetts, USA) and formamide

Table 3. Additive PCR panel. *First six rows represent gut samples with different concentrations of template DNA and additives. Last two rows are for oral DNA and additives, and negative controls.*

GI tract sample	200 ng/µl	100 ng/µl	50 ng/µl	25 ng/µl
	200 ng/µl, 5%DMSO	100 ng/µl, 5%DMSO	50 ng/µl, 5%DMSO	25 ng/µl, 5%DMSO
	200 ng/µl, 10% DMSO	100 ng/µl, 10% DMSO	50 ng/µl, 10% DMSO	25 ng/µl, 10% DMSO
	200 ng/µl, 1,25% formamide	100 ng/µl, 1,25% formamide	50 ng/µl, 1,25% formamide	25 ng/µl, 1,25% formamide
	200 ng/µl, 2,5% formamide	100 ng/µl, 2,5% formamide	50 ng/µl, 2,5% formamide	25 ng/µl, 2,5% formamide
	200 ng/µl, 5% formamide	100 ng/µl, 5% formamide	50 ng/µl, 5% formamide	25 ng/µl, 5% formamide
Oral sample	Oral DNA	Oral DNA, 1,25% formamide	Oral DNA, 2,5% formamide	Oral DNA, 5% formamide
	Oral DNA, 5% DMSO	Oral DNA, 10% DMSO	Water	No primers

(Thermo Fisher, Waltham, Massachusetts, USA). The run parameters were selected based on the previous PCR experiment. The initial denaturation was performed in 98 °C for 5 minutes, then 6 cycles of denaturation in 98 °C for 30 seconds, annealing in 46 °C for 30 seconds and extension in 72 °C for 2 minutes. Then the second stage was immediately performed by running 35 cycles of denaturing in 98 °C for 30 seconds, annealing in 50 °C for 30 seconds and extension in 72 °C for 2 minutes. The final extension was performed in 72 °C for 5 minutes.

5.6 Exonuclease treatment

The libraries contained bright band of primers in agarose gel after the PCR. To get rid of this, the samples were treated with exonucleases. 5 µl of sample, 1 µl FastAP Alkaline phosphatase (Thermo Fisher, Waltham, Massachusetts, USA) and 0.5 µl of ExoI 5'-exonuclease (Thermo Fisher, Waltham, Massachusetts, USA). Reaction mix was first incubated in 37 °C for 15 minutes using Biometra TPersonal (Biometra, Göttingen, Germany) thermocycler, to activate the enzymes and then in 85 °C for 15 minutes to halt the enzymatic activity. As the initial tests with the recommended protocol did not give desired outcome, incubation time optimization test was performed. In the test, four different 37 °C incubation times were tested 0.5 h, 1 h, 1.5 h and 2 h, while the termination phase was kept constant.

Table 4. Pre-amplification primers used for extracted DNA sample prior to Nextera XT Library Preparation protocol. *Overhang sequence is used for enzymatic attachment of sequencing adapters during Nextera XT protocol tagmentation step.*

	Tagmentation overhang	16S complementary sequence
F341_NeXT	**5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**	CCTACGGGAGGCAGCAG
R805_NeXT	**5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**	GGACTACHVGGGTWTCTAAT

5.7 Nextera XT Library Preparation Kit

The extracted zebrafish oral and gastrointestinal DNA of ten fish, were preamplified using AccuPrime Pfx SuperMix, F341_NeXT forward primer and 806R_NeXT reverse primer with Nextera tagmentation overhang sequences (**Table 4.**). For the reaction mix, 17µl of AccuPrime Pfx SuperMix, 1 µl of 4 µM forward primer, 1 µl of 4 µM reverse primer and 1 µl of DNA template were mixed to total volume of 20 µl. The initial denaturation was performed in 95°C for 5 minutes and then 25 cycles of the denaturation in 95°C for 30 seconds, the annealing in 47°C for 30 seconds and the extension in 72°C for 2 minutes. The final extension in 72°C was 5 minutes.

5.7.1 AMPure XP Bead purification

AMPure XP Beads (Beckman-Coulter, Brea, California, USA) can be used for the fragment size selection by altering the ratio between samples and the beads. The shorter the fragment, the more beads there must be compared to the sample. As the amplified DNA fragment was approximately 500 bp, the beads to sample ratio was 1.8, thus for 10µl of sample 18µl of AMPure XP Beads were used. The protocol was performed using PCR-strips and a magnetic stand for 1.5 ml microcentrifuge tubes, which was stripped from the microcentrifuge tube manifold. Fresh 70% ethanol was prepared, and all the steps were performed as in the protocol. Final elution to 30µl of Milli-Q water was used instead of the 40µl, suggested by the protocol.

5.7.2 Fragment analyzer

To assess the success of the library preparation protocol, the DNA fragment lengths of the samples, were characterized by using Fragment Analyzer (Agilent Technologies, Santa Clara, California, USA) instrument by Agilent Technologies. The method is based on dsDNA intercalating fluorescent dye, which marks the DNA during gel separation (Agilent, 2019). This technique gives relatively good estimation of mass per fragment size and thus is invaluable tool for defining the loading concentration for final library pool before sequencing

The fragment analyser was used for defining size distribution of the DNA fragments in the sequencing libraries, by using the DNA NGS Standard sensitivity kit. The reagents were thawing for at least 30 minutes before running the analysis. The separation gel was prepared by mixing 1µl of intercalating dye to 10ml of separation gel per row of samples. Capillary conditioning solution is prepared from the 5x stock solution by mixing 50 ml of the stock to 200 ml of Milli-Q water. Inlet buffer is diluted by mixing 50 ml of 5x stock solution to 200 ml of Milli-Q water. 1ml of this working solution is pipetted to each well of the first row of a MIDI storage plate, to replace the old inlet buffer in the machine. 100 µl of the rinse buffer is pipetted into the first row of wells of a semi-skirted 96-well PCR plate. For each sample set, one row of semi-skirted 96-well PCR plate is filled with 22 µl of NGS diluent marker solution. 2 µl of NGS Standard sensitivity Ladder is pipetted to position 12 and 2 µl of samples to positions 1-11, any position that does not have sample is filled with 24 µl of specific blank solution. Before starting the analysis, the software of the machine was used to set the run parameters and naming the samples. After the run, data was ready to be exported to various data formats.

5.8 Sequencing

5.8.1 Library pooling

For the pooling, loading concentrations needed to be normalized to 4 nM to ensure equal representation of different samples. The concentrations from the fragment analyser were used for calculating the loading volume for the final library pool. For total volume of 100 µl amount of each library was calculated with the following formula:

$$\text{Library volume} = \frac{\text{Final volume} \times \text{final concentration}}{\text{Number of pooled libraries} \times \text{concentration of pooled library}}$$

The rest of the volume was filled with low-TE buffer.

5.8.2 Denaturation of library and PhiX control

For the denaturation, 1 ml of 0.2N NaOH was prepared from the 1.0 N NaOH stock solution by diluting it to the Milli-Q water. 5 µl of the normalized library pool was mixed with 5 µl of 0.2 N NaOH and incubated in RT for 5 minutes, after which 990µl of chilled HT1 buffer from the kit was mixed to halt the denaturation and dilute the library to 20 pM. As 6 pM was the desired final concentration,

20 pM library pool was further diluted by adding 420 µl of chilled HT1 buffer to 180 µl of library pool.

PhiX control was next diluted to compensate for the homogeneity of sequencing 16S gene. 2 µl of 20 nM PhiX control and 3 µl of 20 mM Tris-Cl pH 8.5 + 0.1 % Tween 20 was mixed. Then the PhiX was denatured by adding 5 µl of 0.2 N NaOH and incubating 5 minutes in room temperature. 990 ml of chilled HT1 is added to stop denaturation and diluting the PhiX control to 20 pM. PhiX is further diluted to 12.5 pM by mixing 225 µl of chilled HT1 into 375 µl 20 pM PhiX.

5.8.3 Sequencing run

The sample sheet was created by using the Illumina Experiment Manager Software. With the software the index sequences were connected to the correct samples, serial numbers of the reagents and the sequencing cartridge was documented and the parameters for the run was selected. The 251 bp paired end sequencing was set to be the sequencing run length and the data format was set to FASTQ with adapter trimming applied. Once the sample sheet was created and carefully inspected, it was loaded into sequencing machine.

5.8.4 Preparation of sequencing run

The reagent cartridge was thawed in room temperature water bath for an hour before the loading of the denatured library pool. Once thawed, reagents were mixed by inversions of the cartridge. By tapping the cartridge to benchtop, possible air bubbles were removed from the wells. Cartridge was kept on ice until used. The foil seal of the load sample well was pierced with sterilized 1 ml pipette tip and 600 ml of pooled and denatured library was pipetted in. The flow cell was taken from the storage buffer with plastic forceps and thoroughly rinsed with Milli-Q water to remove excess salts. To avoid scratching the surface of the flow cell, rinse water was carefully dried with lint-free lens cleaning tissues. Storage flow cell in the sequencing machine was replaced with rinsed and dried flow cell. In the sequencing machine, wash bottle was replaced by PR2 buffer solution and the waste bottle was emptied. Next the sequencing cartridge was loaded into the machine.

5.9 Sequencing data analysis

Data-analysis steps was performed with dedicated R-packages, the filtering and merging of paired end reads were done with DADA2 package (Callahan *et al.*, 2016) and the visualization of the data with phyloseq package (McMurdie and Holmes, 2013). The data after sequencing was organized

into folders, which each contained reads for one sample stored in two FASTQ files, one for forward reads and another for reverse reads. These FASTQ files contained both the nucleic acid sequence data, already trimmed of the primer sequences, and the quality information.

5.9.1 DADA2

In the DADA2 pipeline, first step was to assess the quality of the reads with inbuilt quality assessment algorithm, which visualises the quality data by drawing a plot. Based on this information, the reads were then truncated to suitable length by cropping the poor-quality positions out of the data. In the same filtering step, the threshold for maximum number of erroneous bases for each read was set. The filtering parameters used for cropping was 250 and 235 bp for forward and reverse reads respectively, and the error threshold was set to maximum of two errors per read.

Next, the machine learning algorithm was used to learn about the error rates in the sample set. The error rate information was used for combining copies of unique sequences into one, while maintaining the quality data. This dereplication phase greatly reduces the required computing power in downstream analysis. DADA2 algorithm differs from other dereplication methods, by combining the quality data of each copy of a unique read into the dereplicated sequence. Preserving the quality data improves the reliability of the data analysis, as sequence variants can be compared and assessed whether the difference between two closely related variants are truly different and not artefact of the PCR or another phase of the library preparation.

After the dereplication, paired end reads were merged to form full length sequences. The read pairs, which did not align properly were discarded. The merged sequences are then arranged into ASV table, which was later used for searching database for matching sequences. First ASVs were filtered for primer dimers that may have occurred during the library creation. 16S genes have some variations in length, but when sequence length is significantly beyond normal range it was justified to consider such a sequence as an artefact. Before the taxonomic annotation of the ASVs, the sequences can be filtered based on the sequence length thus removing the artefacts from the sample set.

5.9.2 Phyloseq

Phyloseq package for R is flexible tool that uses ggplot2 graphics package for visualization of microbial data. It can be used for turning the ASV table into easily readable publication ready graphs and visualizations of the data. Reasonable way to illustrate 16S data is to assess the microbial

diversity in studied environment by plotting rarefaction curves and alpha diversities. These plots are easily visualized as the variables are already in the memory after previous analysis phases. Alpha diversity is important metrics for diversity assessment in the studied microbial environment, depending on equation used, it measures how rich the variety of species exist in the sample and how evenly they are spread. Shannon index takes both, number of species and their relative abundances into account when calculating the alpha diversity value. The Index is calculated with following formula:

$$H = -\sum_{i=1}^{s} p_i \, lnp_i$$

Where p_i is the number of observed members of certain ASV divided by number of total members of all ASVs in the sample. The p_i is then multiplied by its own natural logarithm, this calculation is repeated for all unique ASVs and results are summed together. The diversity index achieved by multiplying the sum by -1.

5.9.3 VEGAN

The assessment of sequencing depth was done using Vegan R-package (Dixon, 2003). The rarecurve() function was used to draw rarefaction curve. The function goes through reads one at a time and cumulatively tracks the encountered new unique reads. The resulting figure shows whether the sequencing depth was enough for revealing all the different ASVs in the sample, if the curve plateaus, there were enough sequencing reads to cover the whole diversity.

6 RESULTS

6.1 Extraction methods

6.1.1 Oral samples

The oral microbial DNA were extracted using four different methods. In **figure 3.** DNA yields determined from UV-absorbance readings measured with NanoDrop 2000 instrument, were compared in a boxplot. As the elution volume differed between extraction sets, the chart portrays the yield as total mass of DNA instead of concentration.

Highest yields from single samples were acquired with phenol-chloroform extraction (*n = 8, median = 669 ng*), although there were lot of deviation between samples. From silica column-based methods PureLink Microbiome DNA Purification Kit was used the most and resulted in relatively good and consistent yields (*n = 27*, *median = 985 ng*). Quick-DNA Fungal/Bacterial Microprep kit (*n = 8, median = 79 ng*) and QIAamp DNA Stool Mini Kit (*n = 4, median = 121.25 ng*) produced lowest yields of tested methods, but the sample sizes were also small.

6.1.2 Gut samples

Gut microbial DNA extraction methods were compared (**Figure 4.**) to find most effective method to purify DNA for downstream analysis. Concentrations were calculated from UV-absorbances measured with NanoDrop 2000 instrument. Three methods were used, phenol-chloroform

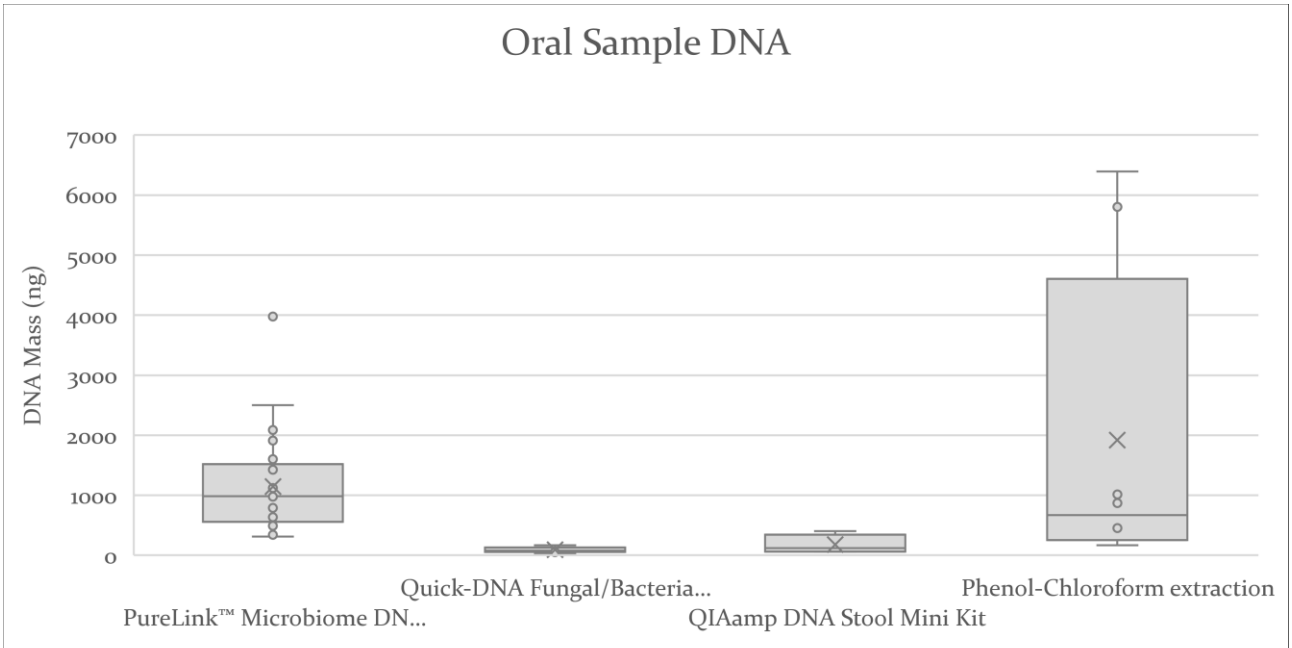

Figure 3. Oral sample DNA yields. *Zebrafish oral microbial DNA sample total DNA yields drawn as boxplot figure. Four different extraction methods were tested, and their effectiveness were evaluated by measuring the total DNA by calculating concentrations from UV-absorbance readings measured by NanoDrop 2000 instrument. PureLink Microbiome DNA Purification Kit (n = 27, median = 985 ng), Quick-DNA Fungal/Bacterial Microprep kit (n = 8, median = 79 ng), QIAamp DNA Stool Mini Kit (n = 4, median = 121.25 ng) and phenol-chloroform extraction (n = 8, median = 669 ng) were the methods used.*

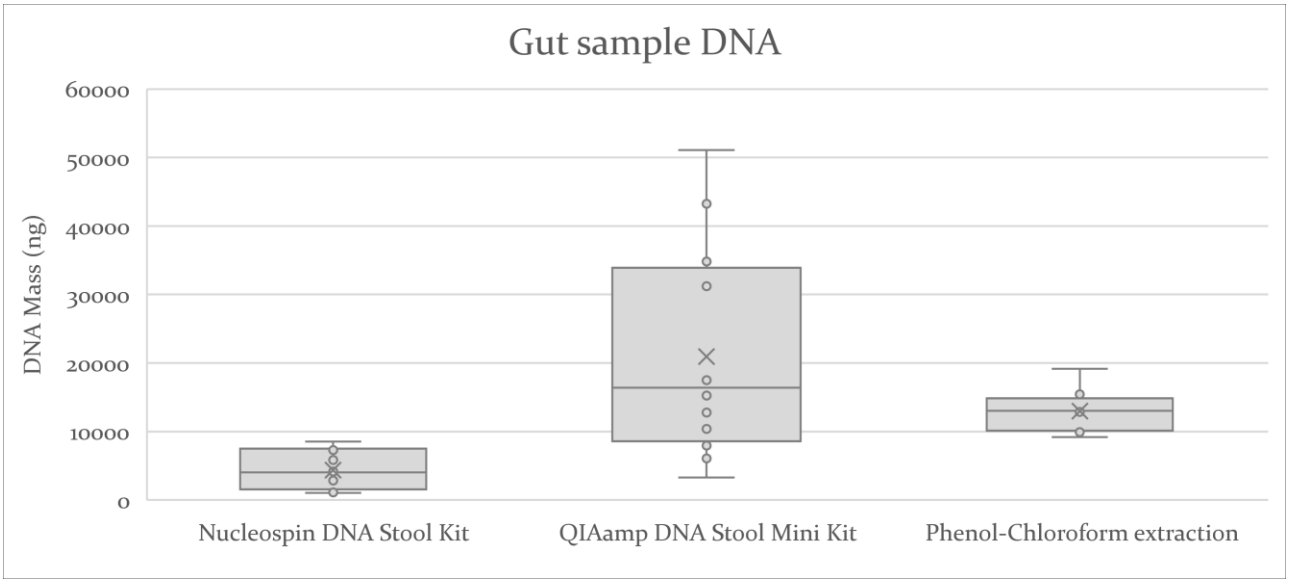

Figure 4. Gut sample DNA yields. *Three different methods for zebrafish gut microbiome DNA sample extraction were piloted throughout the study. Total DNA yields from each method are portrayed in this boxplot. Y-axis is the mass of the total DNA in nanograms and DNA from each method are shown in their respective figures. NucleoSpin DNA Stool Kit (n = 20, median =4037.5 ng), QIAamp DNA Stool Mini Kit (n = 12, median = 16375 ng) and phenol-chloroform extraction (n = 8, median =13000 ng were the methods used).*

extraction (*n = 8, median =13000 ng*), NucleoSpin DNA Stool Kit (*n = 20, median =4037.5 ng*) and QIAamp DNA Stool Mini Kit (*n = 12, median = 16375 ng*). QIAamp DNA Stool Mini Kit produced the best yields of tested methods, although the deviation was relatively high. The phenol-chloroform extraction yields were lower than the QIAamp kit, although being more consistent between samples. The NucleoSpin DNA Stool Kit produced relatively consistent results while yields were lower than with the other methods.

6.2 Library creation in single PCR run

16S sequencing library PCR conditions were optimized in annealing temperature gradient PCR, by using *M. marinum* DNA as a template (**Figure 5.**). Annealing temperatures ranged from 54 °C to 70 °C. 64.1 °C was the highest temperature where 16S band was visible in a gel electrophoresis and all temperatures lower than that had intense band. Results were not reproducible with sample DNA.

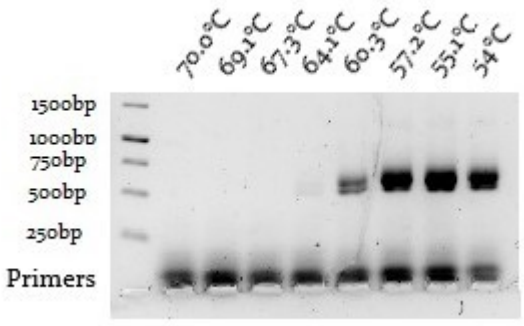

Figure 5. Gel electrophoresis of PCR amplified 16S gene of M. marinum. *PCR protocol was optimized for annealing temperature, which are shown on top of the image. Annealing temperature greatly affects the 16S DNA multiplication efficiency. Band intensity is elevated towards lower temperatures as is expected.*

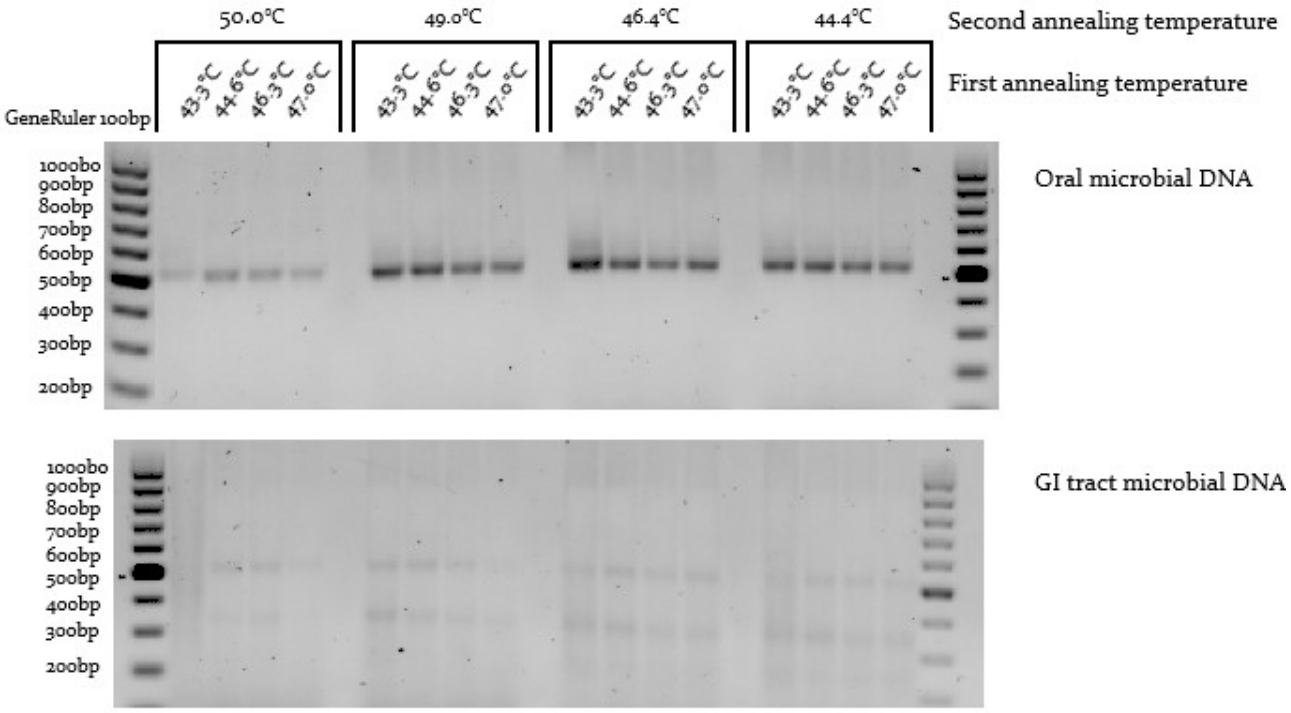

Figure 6. Gel electrophoresis of two phase annealing gradient PCR experiment. *Combinations of first and second stage annealing temperatures were tested in single PCR experiment. Each cluster of four bands represents the second annealing temperature in decreasing order and inside the clusters are the first annealing temperatures in increasing order. Upper image is from oral microbial samples and lower from the gut samples.*

6.3 Library creation in single PCR run with two annealing temperatures

6.3.1 Annealing gradient experiment

P PCR protocol with two different annealing temperatures was used to increase specificity and yield of PCR reaction. DNA samples from the zebrafish oral cavity and gut were used as templates in an experiment where aim was to find suitable combination of first and second annealing temperature (**Figure 6.**). In the oral samples the 16S gene was more effectively amplified than in the gut samples. Noticeable increase in the amplification can be seen when the second annealing temperature was decreased from 50 °C to 49°C. Also, when first stage annealing temperature was decreased, the band intensity was slightly increased. Similar trends could be seen in the gut samples, where the overall band intensity was significantly lower, and reaction appeared to be less specific, indicated by an unknown second band approximately at 350bp.

6.3.2 Applying protocol for sample set

Based on the previous experiment, the annealing temperatures were selected for the protocol that was used for five oral and skin DNA samples. The protocol produced visible bands in the gel electrophoresis in all but one individual oral sample (**Figure 7.**). Band intensities, however, varied among the samples that had successfully amplified the 16S gene. Skin samples had lower rate of success than oral samples. One of skin samples had similar bands to the ones present in gut samples

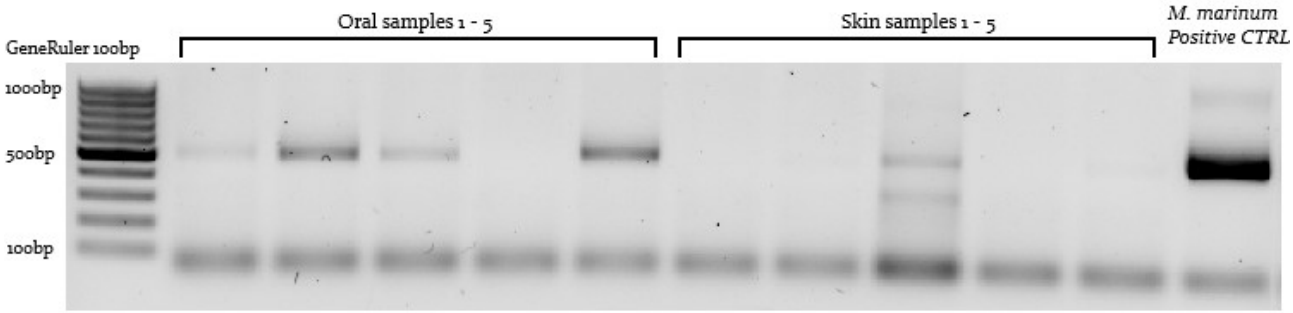

Figure 7. Gel electrophoresis of the zebrafish oral and skin microbial sample DNA. 16S gene amplified with PCR. *Using the conditions optimized during the previous experiment the resulting program was tested with larger sample set.*

in previous experiment including a visible smear of the contaminating background DNA. Another sample did show similar behaviour with barely visible bands. One sample represented more of the oral sample in that it had only one band around 500bp. It was faint enough to not be considered as a success.

6.3.3 PCR additive panel

For enhancing the effectivity of PCR protocol for the gut samples, different PCR additives were tested in variety of concentrations and combinations. However, no 16S DNA bands were visible in the resulting gel electrophoresis except two oral samples that had one clear band at around 500bp

6.4 Exonuclease treatment

For primer removal from the PCR amplified samples, exonuclease treatment with exonuclease I and FastAP alkaline phosphatase was tested with 0,5 h; 1 h; 1,5 h and 2 h treatment time. None of the tested times were enough to reduce the amount of primers in sample, when visually inspected after gel electrophoresis.

6.5 Illumina library kit

6.5.1 16S pre-amplification with PCR

Illumina protocol begun with pre-amplification PCR of 16S V3-V4 region. Sequence specific primers were attached to 20 nucleotides long overhang sequences, designed for sequencing adapter tagmentation in later phases of library preparation. Optimized PCR conditions from previous experiments were utilized when the program was designed. All the samples had amplified the 16S sequence, yet in the oral samples the DNA bands were faint and nearly non-distinguishable (**Figure 8.**). Gut samples had more intense bands in each sample and more non-specific background. All three negative control samples did not show any visible DNA band, whereas *M. marinum* positive control had intense band at 500 bp.

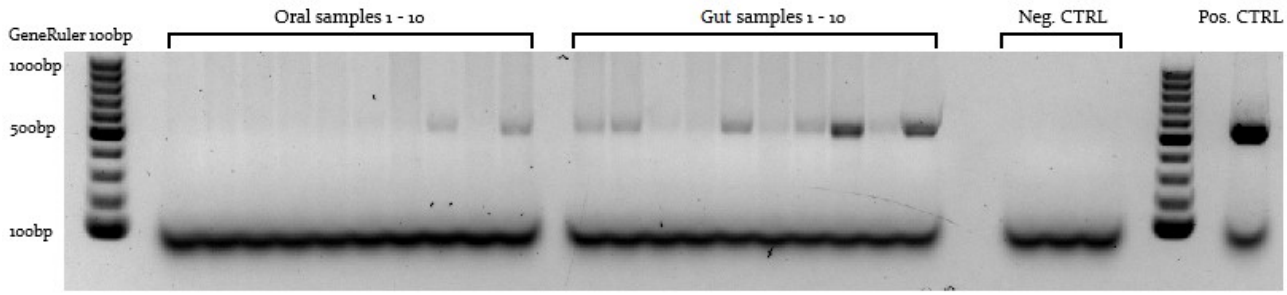

***Figure 8. Limited cycle pre-amplification PCR**. Zebrafish oral and gut microbial sample DNA was amplified with sequencing adapter conjugated 16S primers. All the samples have at the very least faint band of 16S DNA amplified in the gel*

6.5.2 Fragment analyser

The first round of fragment analysis was performed for all the samples that had been used through library preparation protocol. As an output, the instrument draws a graph of the emitted fluorescence, emission intensity on y-axis in relative fluorescence units and fragment length in base pairs on x-axis. The program also gives the mass of each peak calculated from the peak surface area. **Figure 9.** is the fragment analyser graph of gut sample 6G-8, after the gel extraction.

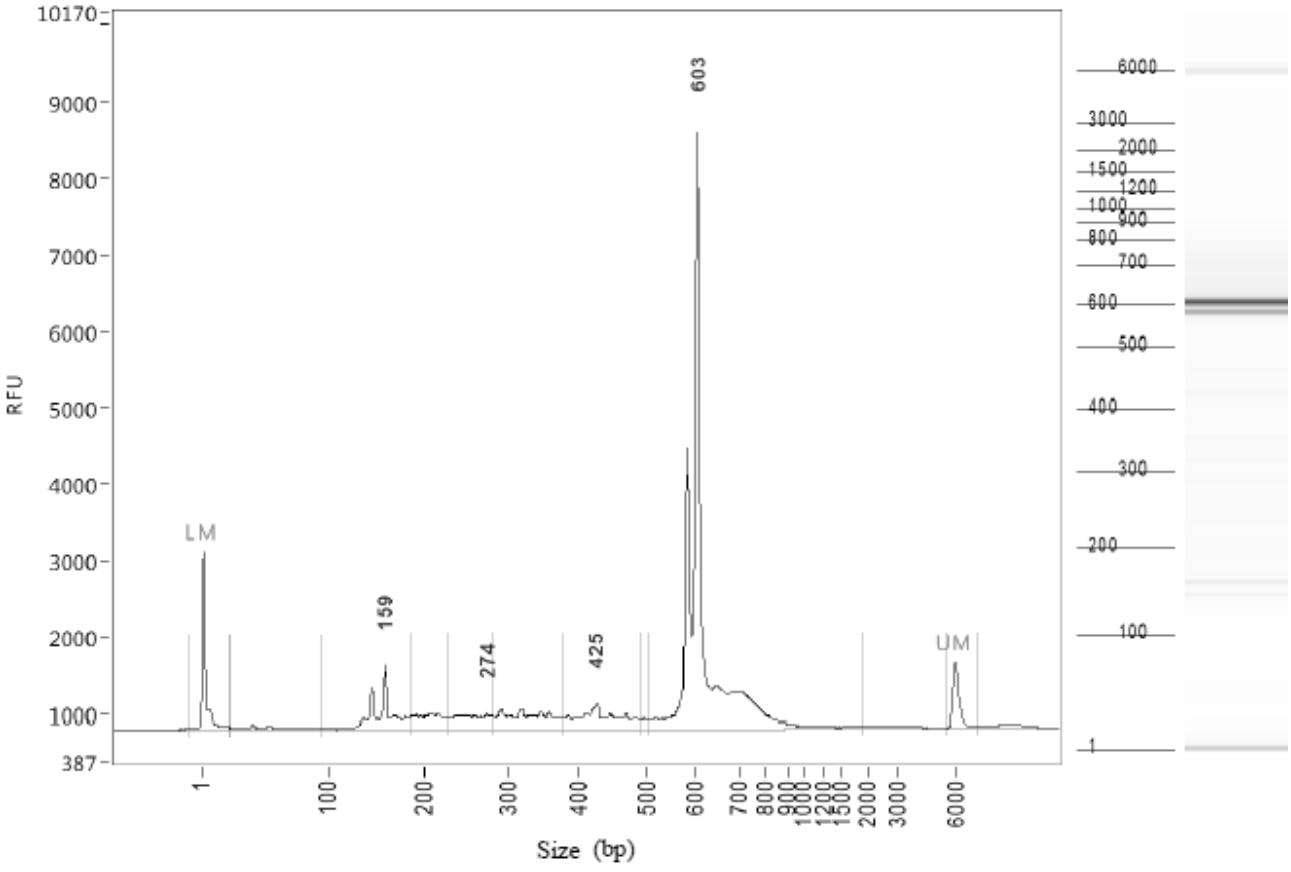

***Figure 9. Fragment analyzer fluorescence intensity curve.** Gut sample 6G-8 measured with fragment analyser. 1 bp marker is named as LM and 6000 bp marker as UM. The highest peak at 603 bp is the amplified 16S library. Some non-insignificant amount of contaminating DNA is present in samples, which led to gel extraction purification step*

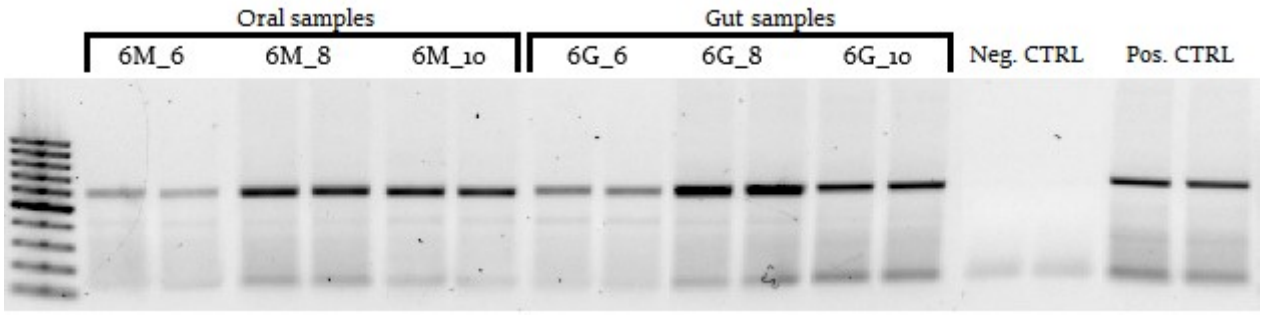

Figure 10. Gel extraction of 16S libraries. *As the sample purity of amplified 16S libraries was not satisfactory, gel extraction of the samples with highest yield was performed. Number of samples was lowered from 20 to 6 samples and the controls. Each sample was divided into two wells due to high volume.*

6.5.3 Gel extraction of library fragments

As the samples contained lot of noise in addition to the library fragment, the samples had to be purified running the samples to agarose gel and extracting the 16S bands with GeneJet Gel Purification Kit (**Figure 10.**). The number of samples was decreased from 20 to 6 as most of the samples had insufficient amount of amplified 16S library after the gel extraction.

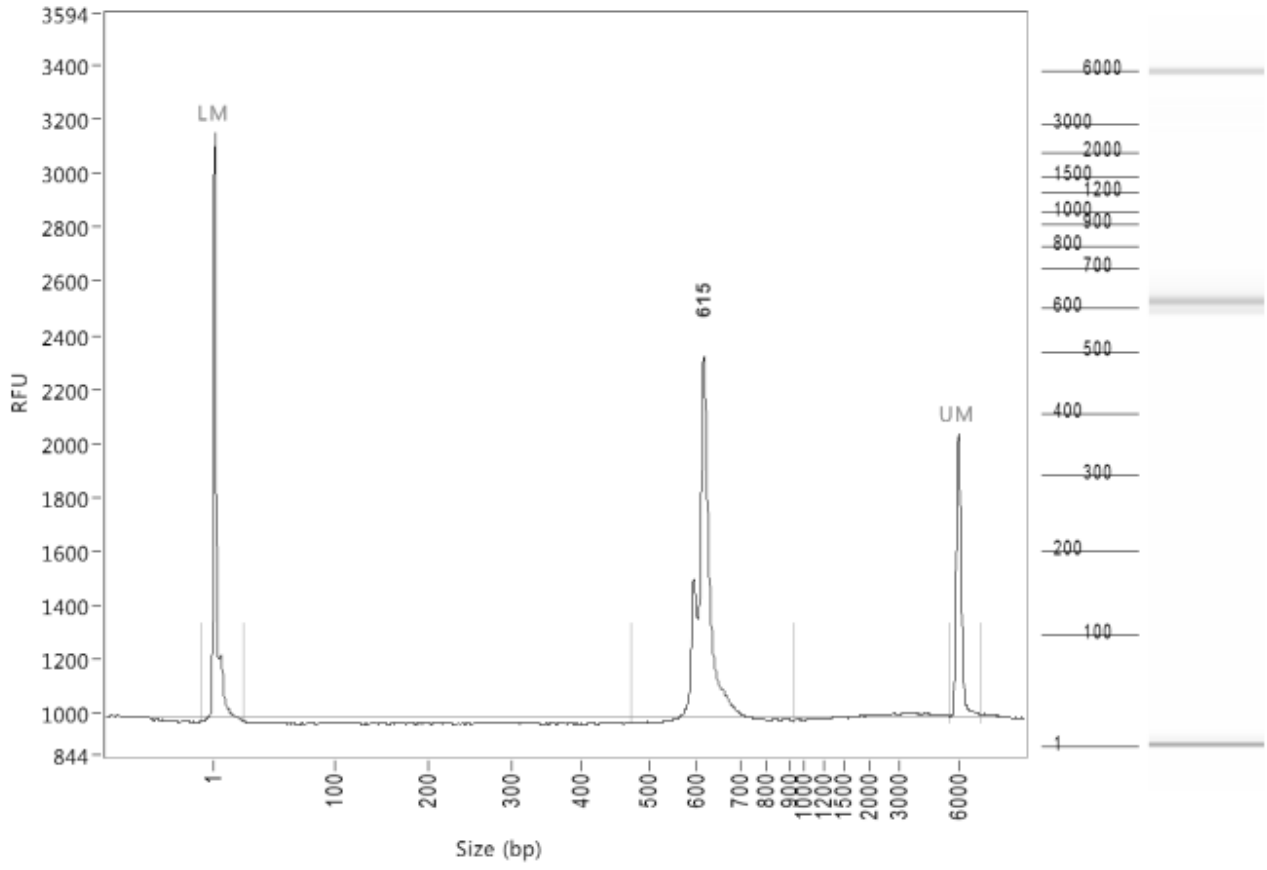

Figure 11. Fragment analysis of gel extracted 16S library. *After the gel extraction step most of the impurities are successfully removed from the sample. The sample shown here is the 6G-8.*

Table 5. Number of reads per sample after each filtering step. *Samples which are marked as "6G" refers to gut samples and "6M" refers to oral samples, numbers in the sample names are identifiers for individual fish.*

	Input	Filtered	DenoisedF	DenoisedR	Merged	Nonchim
6G-10	70812 (100)	58919 (83,2)	58757 (83,0)	58713 (82,9)	57894 (81,8)	43644 (61,6)
6G-6	98530 (100)	76418 (77,6)	75431 (76,6)	74501 (75,6)	72971 (74,1)	41538 (42,2)
6G-8	122847 (100)	102970 (83,8)	102670 (83,6)	102558 (83,5)	101794 (82,9)	75260 (61,3)
6M-10	154335 (100)	129715 (84,0)	128739 (83,4)	128551 (83,3)	127353 (82,5)	93127 (60,3)
6M-6	126814 (100)	101303 (79,9)	98867 (78,0)	97234 (76,7)	93686 (73,9)	49668 (39,2)
6M-8	154886 (100)	128808 (83,2)	128208 (82,8)	127682 (82,4)	126596 (81,7)	91472 (59,1)
CTRLneg	1223 (100)	734 (60,0)	678 (55,4)	624 (51,0)	557 (45,5)	524 (42,8)
CTRLpos	110130 (100)	87700 (79,6)	87634 (79,6)	87582 (79,5)	87223 (79,2)	65781 (59,7)

6.5.4 Fragment analysis of purified samples

When the library fragments were purified, the fragment analysis was run. The resulting graph of gut sample 6G-8 can be seen in **figure 11**. Most of the impurities were removed in gel extraction according to fragment analysis results. However substantial amount of library fragment was also lost in the purification step.

6.6 Sequencing data

6.6.1 DADA2

Sequences were filtered in multiple steps to remove the sequencing artefacts from the reads. The **table 5.** shows how much each filtering step reduced the number of reads. The mean number of reads per sample before the filtering were approximately 121,000 reads. The first filtering step had considerable effect on the number of reads, as approximately 20% of the reads were discarded. Denoising and merging of paired reads had only limited effect on for total number of reads. Chimera

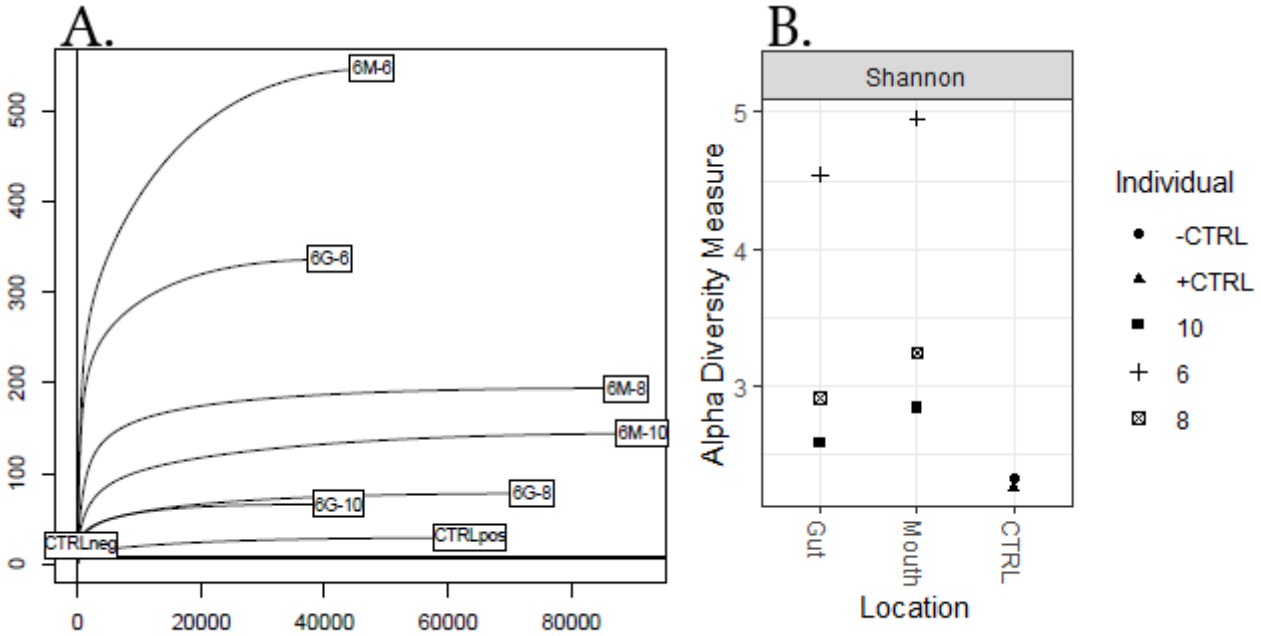

Figure 12. Rarefaction curve and measured alpha diversities in samples. ***A.*** *The sequencing depth is shown in rarefaction curve. Total reads in the x-axis and observed unique ASVs in y-axis* ***B.*** *Alpha diversity is presented here as function of Shannon diversity index.*

removal discarded another 20% of the reads. After the filtering steps approximately 65,000 reads per sample or 53.3% of the reads in total were remaining. For assessing the sequencing depth, the rarefaction curve was drawn (**Figure 12. A**). All the samples do plateau in the curve, and therefore, the sequencing depth was sufficient to reveal the microbial diversity in the samples.

6.6.2 Alpha diversity

The first taxonomic characterisation for the samples was determining the alpha diversity values as a function of the Shannon index (**Figure 12. B**). Based on the index metrics, individual's microbiomes were comparatively diverse in different anatomic locations, and microbiomes from the same individual clustered together in the graph. The Shannon diversity index also suggests that the gut microbiota was systematically slightly less diverse than the oral cavity microbiotas, although the sample size was too small to make definite conclusions. 6G-6 and 6M-6 appears to be distinctly more diverse when compared to the other samples.

6.6.3 Most abundant microbes

In the **figure 13.** the relative abundance of microbial orders is presented as percentage from total number of reads in sample that were annotated to the genus level and each order is coloured with variation of colour theme of the phyla it belongs to. Proteobacteria, Firmicutes and Actinobacteria were the most abundant phyla in the studied samples and most of the abundance was made up by Proteobacteria and Firmicutes. Considerable amount of Actinobacteria was only present in microbiotas of fish where Propionibacteria was the most common. Most of Proteobacteria belongs to order Aeromonadales and the rest of the phylum consists of rather tiny proportion of the reads. Oral cavity microbiota has more diversity within Proteobacteria than gut, except the 6G-6 and 6M-6 which had distinct diversity in both environments and had oral microbiota, in which other orders of Proteobacteria made up roughly 40% of total reads. In the gut microbiota Firmicutes made up 8%-27% of the reads, which made it the second largest phylum after Proteobacteria. In the oral environment Firmicutes was the most abundant except in the 6M-6. Most of the firmicutes belongs to the order Clostridiales, but Bacillales and Lactobacillales have also significant representation within many samples. Even rare taxonomic groups, which only have few reads can be important for the functions of the microbiota. As these groups are easily buried when abundances are shown in absolute scale, in **figure 14.** microbial phyla are portrayed in logarithmic scale to reveal rare taxa. Some phyla such as Chlamydiae is present in nearly all samples, yet neither had even 100 reads for it and some were only present in just one or few samples.

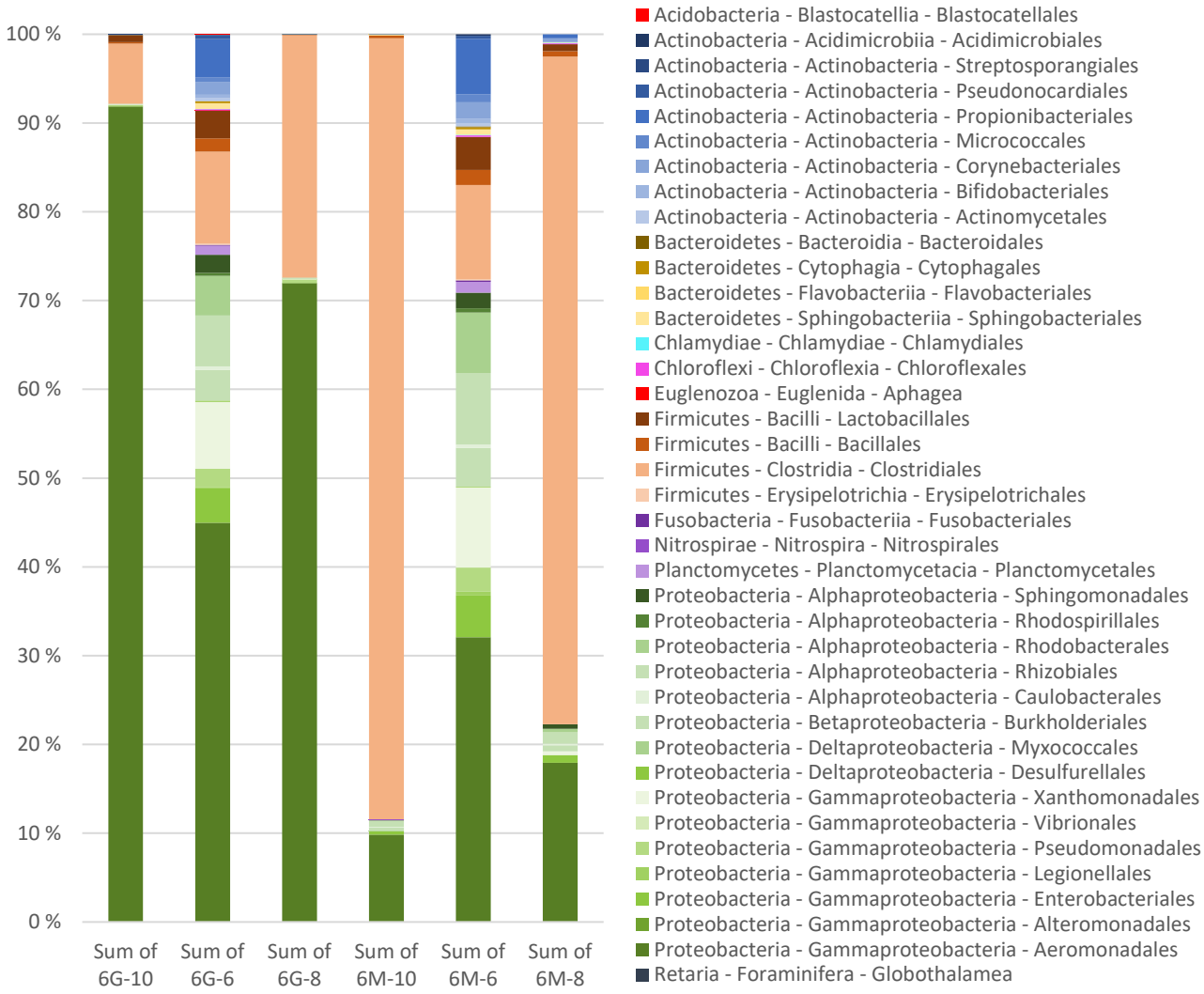

Figure 13. Relative abundances of reads annotated for each microbial order. *The chart combines all the reads that were annotated down to genus level and presents them as percentage of total reads in the sample. Taxonomic ranks are visualised so that each order belonging to same phylum are variations of the same colour theme*

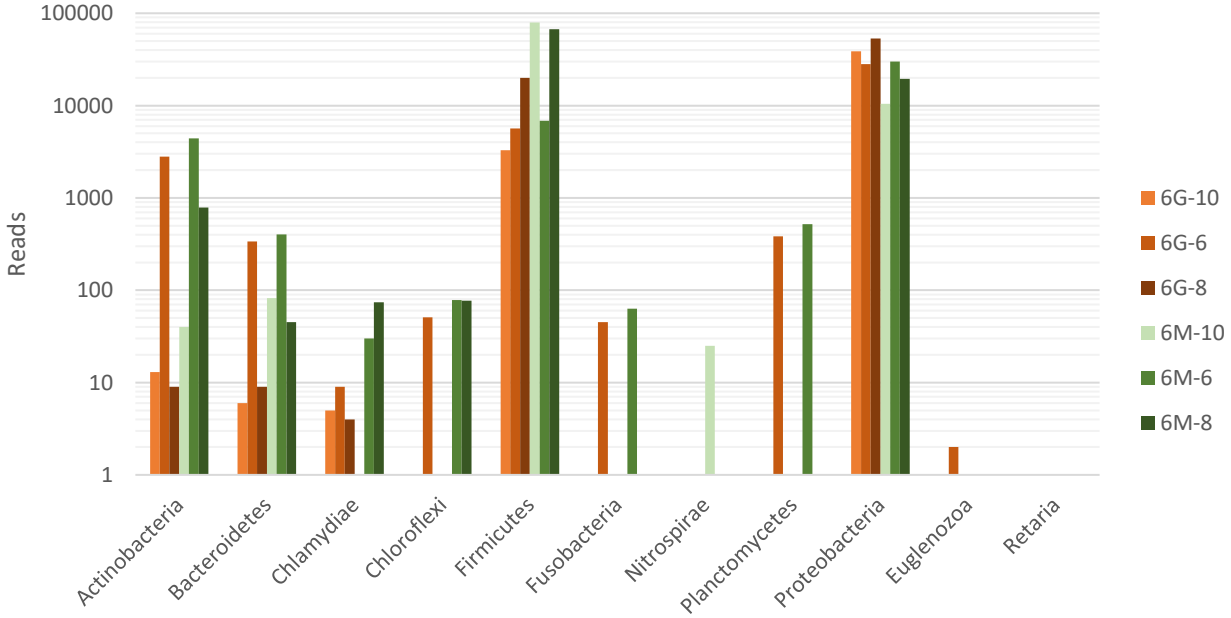

Figure 14. Microbial phyla in sequenced samples. *The y-axis shows number of reads per phyla in logarithmic scale.*

6.6.4 Genus level diversity

In the **table 6.** 71 of the most common genera are organised in decreasing order the genera with the highest total number of reads in the top. The two most common genera were *Paraclostridium* and *Aeromonas*, present in high numbers in every samples. Previous figures and tables have shown 6G-6 and 6M-6 to have distinct microbiota and this table confirms the observation. 6G-10 had in total 21 genera present in gut and 6M-10 had 25 genera in oral microbiota, where 6G-6 and 6M-6 had 76 and 77, and 6G-8 and 6M-8 had 21 and 56 respectively. Every fish had more genera in oral environment, but number of species in total was significantly higher in 6M-6, which also had the number of reads more evenly spread between genera. In the genus level, some obvious differences could be observed when taxonomic ranks were inspected in the genus level. For example, number of *Methylobacterium* and *Escherichia/Shigella* were nearly ten times higher in oral samples than in gut. Another example being *Burkholderia-Paraburkholderia* which had abundant reads in oral samples but were not present in gut samples except in the 6G-6. One can also observe that gut microbiota had no genera which are abundant in gut but non-existent or only present in lower numbers in oral environment. The gut microbiota also seems to be less homogenous outside of the two most common genera that makes most of the reads.

Table 6. Most abundant genera of sequenced microbiome samples. *Genera are organized to show highest total amount of reads in the top. Also, taxonomic ranks are shown up to the phylum level. Note that only reads annotated down to the genus level are shown in this table.*

Phylum	Class	Order	Family	Genus	6G-10	6G-6	6G-8	6M-10	6M-6	6M-8
Firmicutes	Clostridia	Clostridiales	Peptostreptococcac.	*Paraclostridium*	2614	2124	19836	79333	2111	64922
Proteobacteria	Gammaproteobact.	Aeromonadales	Aeromonadaceae	*Aeromonas*	38693	16786	52649	8880	13603	15678
Proteobacteria	Gammaproteobact.	Xanthomonadales	Xanthomonadaceae	*Stenotrophomon.*	0	2808	21	0	3833	343
Proteobacteria	Alphaproteobact.	Rhizobiales	Methylobacteriaceae	*Methylobacter.*	40	1424	90	445	2188	717
Proteobacteria	Gammaproteobact.	Enterobacteriales	Enterobacteriaceae	*Escherichia/Shig.*	39	1420	0	366	1958	716
Actinobacteria	Actinobacteria	Propionibacteriales	Nocardioidaceae	*Nocardioides*	0	1571	0	0	2572	300
Proteobacteria	Alphaproteobact.	Rhodobacterales	Rhodobacteraceae	*Paracoccus*	4	1406	0	0	2392	364
Proteobacteria	Gammaproteobact.	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	0	763	299	2	1135	48
Proteobacteria	Alphaproteobact.	Sphingomonadales	Sphingomonadaceae	*Sphingomonas*	0	754	0	110	759	427
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Burkholderia-Par.*	0	499	0	274	721	308
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Achromobacter*	0	621	0	0	950	109
Firmicutes	Bacilli	Bacillales	Bacillaceae	*Bacillus*	66	421	59	265	454	332
Firmicutes	Clostridia	Clostridiales	Peptostreptococcac.	*Peptostreptococ.*	0	480	9	0	926	143
Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	*Mycobacterium*	7	500	0	0	733	249
Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	*Vagococcus*	219	355	0	0	413	252
Proteobacteria	Alphaproteobact.	Rhizobiales	Rhizobiaceae	*Rhizobium*	0	299	0	159	666	113
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0	408	5	0	467	13
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Epulopiscium*	239	0	113	0	0	402
Firmicutes	Clostridia	Clostridiales	Peptostreptococcac.	*Asaccharospora*	0	271	0	0	350	69
Proteobacteria	Alphaproteobact.	Rhodobacterales	Rhodobacteraceae	*Defluviimonas*	0	184	16	0	386	0
Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	*Weissella*	0	181	0	6	280	117
Actinobacteria	Actinobacteria	Micrococcales	Microbacteriaceae	*Microbacterium*	0	122	7	0	278	147
Proteobacteria	Alphaproteobact.	Caulobacterales	Caulobacteraceae	*Caulobacter*	13	144	4	96	160	136
Planctomycete	Planctomycetacia	Planctomycetales	Planctomycetaceae	*Singulisphaera*	0	192	0	0	295	0
Firmicutes	Clostridia	Clostridiales	Family_XI	*Gallicola*	0	217	0	0	218	9
Planctomycete	Planctomycetacia	Planctomycetales	Planctomycetaceae	*Gemmata*	0	190	0	0	223	0
Firmicutes	Clostridia	Clostridiales	Clostridiaceae_1	*Clostridium sensu*	0	192	0	0	125	62
Proteobacteria	Alphaproteobact.	Rhizobiales	Methylobacteriaceae	*Meganema*	0	138	0	37	161	36
Proteobacteria	Betaproteobact.	Burkholderiales	Comamonadaceae	*Xylophilus*	16	115	7	41	112	64
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	*Asinibacterium*	0	149	0	47	117	27
Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0	116	0	4	214	0
Proteobacteria	Alphaproteobact.	Rhodospirillales	Rhodospirillales Inc.	*Reyranella*	0	130	0	0	200	0
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	5	106	0	0	152	64
Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	*Enterococcus*	124	0	0	0	0	197
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	*Carnobacterium*	0	94	0	0	178	46
Firmicutes	Clostridia	Clostridiales	Family_XI	*Tepidimicrobium*	0	120	0	0	179	10
Actinobacteria	Actinobacteria	Pseudonocardiales	Pseudonocardiaceae	*Pseudonocardia*	0	158	2	0	146	0
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0	138	0	0	153	0
Proteobacteria	Alphaproteobact.	Rhizobiales	Phyllobacteriaceae	*Mesorhizobium*	0	0	0	0	0	217
Proteobacteria	Alphaproteobact.	Rhizobiales	Xanthobacteraceae	*Labrys*	0	70	0	0	127	17
Firmicutes	Clostridia	Clostridiales	Family_XI	*Peptoniphilus*	7	130	0	0	44	29
Chloroflexi	Chloroflexia	Chloroflexales	Chloroflexaceae	*Candidatus Chlor.*	0	51	0	0	78	77
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	*Sediminibacter.*	6	69	0	26	100	5
Firmicutes	Clostridia	Clostridiales	Clostridiaceae_1	*Clostridium sensu*	2	53	4	25	76	39
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*	0	66	0	0	133	0
Proteobacteria	Alphaproteobact.	Rhizobiales	Rhizobiales Inc.	*Phreatobacter*	0	73	0	0	111	8
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	*Spirosoma*	0	69	0	0	105	0
Actinobacteria	Actinobacteria	Propionibacteriales	Propionibacteriaceae	*Propionibacter.*	0	49	0	5	81	35
Proteobacteria	Alphaproteobact.	Rhizobiales	Phyllobacteriaceae	*Aminobacter*	0	69	0	0	89	0
Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	*Rothia*	0	59	0	0	86	10
Proteobacteria	Gammaproteobact.	Legionellales	Legionellaceae	*Legionella*	0	67	0	0	87	0
Firmicutes	Clostridia	Clostridiales	Peptostreptococcac.	*Paeniclostridium*	0	58	0	0	89	0
Firmicutes	Clostridia	Clostridiales	Clostridiaceae_1	*Hathewaya*	0	60	0	0	85	0
Firmicutes	Bacilli	Bacillales	Bacillaceae	*Anoxybacillus*	0	70	0	0	73	0
Firmicutes	Clostridia	Clostridiales	Clostridiaceae_1	*Clostridium sensu*	0	61	0	0	77	0
Proteobacteria	Alphaproteobact.	Rhodobacterales	Rhodobacteraceae	*Albimonas*	0	42	0	0	83	0
Chlamydiae	Chlamydiae	Chlamydiales	Parachlamydiaceae	*Neochlamydia*	5	9	4	0	30	74
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Aquabacterium*	0	66	0	0	55	0
Proteobacteria	Gammaproteobact.	Legionellales	Coxiellaceae	*Coxiella*	0	0	0	0	120	0
Firmicutes	Clostridia	Clostridiales	Peptostreptococcac.	*Proteocatella*	0	53	0	0	65	0
Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	*Corynebacterium*	0	44	0	0	44	23
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0	45	0	0	63	0
Actinobacteria	Acidimicrobiia	Acidimicrobiales	Iamiaceae	*Iamia*	0	34	0	0	72	0
Proteobacteria	Alphaproteobact.	Rhizobiales	Bradyrhizobiaceae	*Bradyrhizobium*	0	32	0	0	73	0
Firmicutes	Bacilli	Bacillales	Planococcaceae	*Domibacillus*	0	0	0	0	0	101
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	*Ruminiclostrid.*	0	0	0	0	100	0
Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	*Erysipelothrix*	0	46	0	0	50	0
Proteobacteria	Betaproteobact.	Burkholderiales	Burkholderiaceae	*Cupriavidus*	0	0	0	0	0	93
Firmicutes	Clostridia	Clostridiales	Family_XI	*Anaerococcus*	0	47	0	0	40	0
Proteobacteria	Deltaproteobact.	Desulfurellales	Desulfurellaceae	*H16*	4	46	0	0	35	0
Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	*Leuconostoc*	0	20	0	0	57	0

7 DISCUSSION

The objective of the project was to characterize the oral microbiota of the healthy zebrafish and establish a protocol that can be used in further studies of microbe-microbe and microbe-host interactions in oral environment. In recent years, sequencing the ribosomal small subunit RNA coding gene variable regions, has been the prevalent method for microbial taxonomical classification (D'Amore *et al.*, 2016), due to its low cost and an abundance of data it provides. Selecting this technique was easy choice as there are already wide selection of published papers, which use this technique in variety of different study settings, spanning from environmental to clinical samples (Caporaso *et al.*, 2011; Byrd, Belkaid and Segre, 2018).

7.1 Extraction methods

As remarked earlier in this text, the oral samples from the zebrafish is challenging task. Volume in the mucous sample is diminishingly small and it also reflects in the DNA extraction step and all the other downstream workflows, while also affecting the UV-absorbance based measurements of concentration. As expected based on literature on the subject, phenol-chloroform extraction provided good yield in both sample types (Rosenbaum *et al.*, 2019), although in oral samples there were substantial variation between individual samples. In silica column-based methods there is always lower DNA extraction yield (Rosenbaum *et al.*, 2019), yet the purity and consistency between samples and swiftness of the protocols are considerable benefits that outweighs loss of yield compared to traditional phenol-chloroform extraction. Based on these experiments PureLink Microbiome DNA Purification Kit was selected for oral sample DNA extraction method for upcoming experiments. For gut samples QIAamp DNA Stool Mini Kit produced the best yield out of all methods but was unfortunately discontinued during the study. It was replaced by NucleoSpin DNA Stool Kit, which was the worst performing of the tested gut DNA extraction method measured in total DNA yield. However, it was still selected due to its fast and convenient protocol and yield that was enough for producing sequencing library. The PCR optimization in this thesis project was mainly for finding optimal conditions, but in hindsight it would have been interesting to compare DNA extracted with different methods for amplification efficiency. Now the only comparison was the total yield of DNA and not the quality of the DNA.

7.1.1 Optimizing the PCR protocol

PCR as a technique is rather robust, although some samples do require laborious optimization to function properly, which was evident when working with zebrafish microbiome samples. The difficulties for multiplication of nucleic acid sequences may arise from multiple different factors. The primers might form dimers, secondary structures, or the Tm's of the primer pair does not match each other. Sample can also have substances that inhibits the efficiency of the reaction (Schrader *et al.*, 2012). It is not clear which factors contributed to the difficulty of finding suitable reaction conditions in this study.

PCR is essential for successful sequencing run and there are numerous factors that can affect the outcome, potentially even obstructing the success of whole sequencing reaction. Potential source of error for misidentifying reads is cross-contamination of library primers during purification steps of manufacturing process or later by the final user, which can potentially have drastic effect for the interpretation of the results. User have little influence over the manufacturing process of primers, so extra care should be taken during performing PCR and other library creation steps to have as pure library as possible.

7.2 AccuPrime Pfx SuperMix

From the literature different alternatives for PCR reagents were compared and AccuPrime Pfx SuperMix was eventually selected due to its high-fidelity Thermococcus polymerase, which has very low error rate. The fidelity for sequencing is especially important, as few nucleotides differences in sequences may affect the proper allocation of closely related taxa. This master mix has advantage of the reaction conditions being pre-optimized. However, the same pre-optimized conditions have disadvantage in being relatively rigid for fine tuning the master mix for the reactions that are more challenging, such as the minute amount of microbial DNA present in zebrafish oral microbial samples. The benefit of high-fidelity outweighed rigidity in the assessment and all the PCR steps were performed using this master mix.

Homebrew sequencing library primers were selected as the initial approach for this project. Designing the primers itself provides flexibility in study design and good quality sequencing reads. However, the optimization of the PCR conditions proved to be more challenging than was initially thought. Problems in PCR might have been associated with length of the primers, which were around 80 nucleotides long. In series of PCR reactions, conditions were optimized for concentration of reaction components, temperatures for denaturation and annealing steps and experimented

with differing lengths of each step. None of the above approaches resulted into PCR protocol, which would successfully and consistently multiply sequencing library from the oral and gut samples that were mainly used for the pilot study.

Readily available cultured *Mycobacterium marinum* DNA was used as a positive control throughout the study. *Mycobacterium* DNA performed well as positive control, it was much more effectively amplified in PCR than sample DNA, thus indicating if adjustments to the protocol were bringing working protocol closer or not. The first successes of PCR came from protocol where two different annealing temperatures were used. The intention was to first use annealing temperature that was calculated based on the annealing temperature of the portion of the primers that are complementary to the 16S gene and then raise annealing temperature to match whole primer sequence.

Two-step process described above appeared initially promising when tested with *M. marinum* samples. However, test runs with oral and gut microbial samples did not yield consistent results with yield and specificity of the reaction. Some samples produced bright bands of the correct length once run on 1% agarose gel, but increasing the cycles up to 40, did not result in visible bands in others. Also, the protocol in those samples, that show good amplification, were not consistent between different runs. One of the factors leading to inconsistent results could be sample DNA deterioration. With each DNA extraction sets, multiple PCR runs were performed. Through earliest sample extraction sets, DNA was not divided into aliquots, thus those samples went through multiple freeze-thawing cycles, which have potentially accelerated the degradation. As two sample sets were aliquoted and stored in -80°C with not much better success in amplification efficiency, it remains speculation if degradation of template DNA affected the results in PCR runs. It is possible to speculate, that if the tested protocols would have worked, there would have been at least some visible evidence of the success even in the sample set most exposed to degradative conditions, since PCR rounds in the test runs were abundant.

7.2.1 Exonuclease treatment

Consistently, in addition to successful PCR amplification, there was leftover primers that were not consumed during PCR. As the sequencing efficiency with MiSeq is reliant on the mass of the sequencing tags, the primers would have had major undesired impact on the cluster formation. One approach for purifying unbound primers from the samples after PCR is treating them with exonucleases, which digest the single stranded DNA including 5'-overhangs on double stranded

fragments. As this protocol was used in our lab before, it was logical way of treating the amplified samples before sequencing. However, the standard protocol had poor efficiency, as it failed to remove excess primers. The incubation time was increased to make the reaction perform better. Neither gave this method good results. The enzymes used had been proved to work earlier, so this has to do with the reaction conditions or the sheer amount of substrate, which was present in abundance in these test samples. However, it was concluded that other measures are worth exploring for sample purification as continuing with exonuclease would have meant impractically lengthy treatment times or some optimization for reaction conditions before the working protocol would have been found.

7.3 Nextera XT Library Preparation Kit

Due to failures in attempts to utilize PCR for library creation, the enzymatic tagmentation approach with Nextera XT Library Preparation Kit was used. This method produced sequencing library that was enough to complete this project, while still not being optimal protocol for library creation.

7.3.1 AMPure XP Beads

AMPure XP beads are purification method for double stranded DNA, in which DNA is bound to metallic beads that can be then separated from other substances, by discarding the supernatant once beads are bound to the wall of microcentrifuge tube with help of magnetic stand. With the method, there is also possibility to do size selection of the purified DNA fragments by altering the beads to DNA ratio. In this project the method was used two times for sample purification. There are no data for purity concerning other substances than DNA, but the size selection which was sought after, was not as efficient as expected and thus gel extraction was needed to specifically select the correct sized library for sequencing.

7.3.2 Fragment analyser

The quantitation of the sequencing library is essential for the success of optimal clustering in the sequencer flow cell. Since cluster density is dependent on the molarity of library molecules, information of sheer w/v concentration is not enough for properly defining final library volume as small fragments can overfill the channels and fragments larger than anticipated library would leave clustering density lower than optimal. Therefore, libraries must be analysed with methods that can measure the exact size distribution in the sample. Fragment analyser was selected for the purpose due to its easy availability and for the high-quality data it provides. Also, the fragment analyser can run the analysis with low amount of sample, thus preserving valuable sample. In the project

fragment analyser was used twice as first round revealed lot of background DNA that had to be removed, as AMPure XP beads failed to remove it even when it was performed according to protocol. The fragment analysis proved to be powerful method for examining the fragment size, but it also has some limitations that prevent its usage in some occasions. If the library contains many different sizes of library fragment, it could be difficult to count the amount of library needed for sequencing run. In these instances, the library quantification is reasonable to be performed using quantitative PCR.

7.4 Sequencing

Where creating the sequencing library proved to be difficult, the sequencing itself was successful. The cluster density in the flow cell was close to perfect and three fourths of all reads passed the quality filtering. The chemistry was nearly six months past the expiry date when used, but it did not seem to significantly affect the quality, which proved that good results can also be obtained using expired sequencing chemistry.

7.4.1 Data analysis

The data analysis was performed using DADA2 R-package (Callahan *et al*., 2016), which filtered irrelevant reads in multiple steps. Quality filtering removed on average 21% and chimera removal another 22% of the reads, and in total all the filtering steps reduced the number of reads to just 53% of the original reads. Even after filtering there were on average approximately 65000 reads per sample. When the sequencing depth was assessed with rarefaction curve, the amount of reads per sample was proved to be enough to cover the diversity in each sample.

7.5 Zebrafish microbiome

For microbial ecology studies three individuals are undeniably small sample size, yet sufficient for providing valuable information in this pilot study where primary objective was to test the sequencing protocol. Despite its secondary nature, the sequencing itself revealed some interesting features of the zebrafish microbiota. Species diversity appears to be different between the microbial environments, gut samples had 288 unique ASVs whereas oral samples had 401. The taxon composition was also different, which can be seen in **table 5.** It is worth noting that this observed species diversity may have been distorted by process of creating the library and potential contaminants. Multiple rounds of PCR could have introduced mutations by polymerase, despite the high-fidelity polymerase used and amplification bias could have distorted the relative abundances. Another source of bias and potentially diversity increasing factor could be contaminating microbial

DNA from kits and reagents used for library creation. Salter *et al.* (2014) studied some commonly used extraction kits and reagents and found that especially with small amount of starting material the proportion of contaminating DNA can even became dominant.

According to previous microbial ecology studies conducted with zebrafish gut microbiota (Roeselers *et al.*, 2011), different zebrafish populations even between laboratory strains and fish recently caught from nature seems to share certain core microbiota. Fusobacteria was systematically present in all populations, whereas in this study it had only marginal importance covering only 0.1% of the reads. Proteobacteria phylum, which forms majority of reads in gut samples in this study and has considerable representation in oral samples, was also present with major percentage in Roeselers et al. study, where *Aeromonas* and *Pseudomonas* were making significant contribution to the microbial abundance. In this study *Pseudomonas* played smaller role.

The most interesting observation made based on this study was finding significant proportion of Firmicutes from both oral and gut environments. Majority of those reads were characterized as uncharacterized Peptostreptococcaceae species that is closely related to genus *Paraclostridium*. One previous study has detected the nearly identical ASV of 16S V3-V4 region from faeces of endangered Yangtze finless porpoise (Wan *et al.*, 2016). The genus has two known obligate anaerobe member species *Paraclostridium benzoelyticum* extracted from the marine sediment (Sasi Jyothsna *et al.*, 2016) and *P. bifermentas* which was previously classified as *Clostridium* but was proposed by Sasi Jyothsna *et al*., (2016) to be moved under this new *Paraclostridium* genus. *P. bifermentas* can potentially have pathogenic properties. The occurrence of *P. bifermentas* in mice gut correlates with the worsening UC symptoms and decreasing microbiome diversity in UC mouse model (Kutsuna *et al.*, 2018). The role of this novel *Paraclostridium* species for zebrafish microbiotas would be interesting to assess in upcoming studies along with the *Aeromonas* species, which was present in great numbers in all samples and which made majority of the bacterial abundance in gut. *Aeromonas* species was not characterized down to species level, but BLAST search of some of the ASVs found high similarity >99,5% to *A. veronii, yet* similarity to several other *Aeromonas* species were also over 99%. As the genus has some known pathogens, it might be worth studying further.

8 CONCLUSIONS

This project ended up with successfully sequenced the zebrafish oral and gut microbiota. PCR amplified sequencing library protocol and library creation kit with enzymatic sequencing adapter tagmentation was tested. PCR based method failed to give consistent results with library creation, where library creation kit produced library that was eventually used for sequencing. There were some non-optimally functioning steps during the library creation, which led to deviation from the kit protocol. The sequencing revealed higher diversity in oral samples compared to gut samples, also the dominant genera were different in both environments. ASVs annotated to *Paraclostridium,* the most common genus of oral microbiome might belong to uncharacterized species, which is not present in other studies of zebrafish microbiota.

9 REFERENCES

AGILENT (2019) *Assessment of Genomic DNA Quality with the Agilent 5200 Fragment Analyzer System*. Available at: https://www.agilent.com/cs/library/applications/application-genomic-dna-quality-fragment-analyzer-5994-0511en-agilent.pdf (Accessed: 22 May 2019).

AMINABADI, N. A., ERFANPARAST, L., EBRAHIMI, A. AND OSKOUEI, S. G. (2011) 'Effect of Chlorhexidine Pretreatment on the Stability of Salivary Lactobacilli Probiotic in Six- to Twelve-Year-Old Children: A Randomized Controlled Trial', *Caries Research*. Karger Publishers, 45(2), pp. 148–154. doi: 10.1159/000325741.

BACHTIAR, E. W., BACHTIAR, B. M., JAROSZ, L. M., AMIR, L. R., SUNARTO, H., GANIN, H., MEIJLER, M. M. AND KROM, B. P. (2014) 'AI-2 of Aggregatibacter actinomycetemcomitans inhibits Candida albicans biofilm formation.', *Frontiers in cellular and infection microbiology*. Frontiers Media SA, 4, p. 94. doi: 10.3389/fcimb.2014.00094.

BAKER, J. L., BOR, B., AGNELLO, M., SHI, W. AND HE, X. (2017) 'Ecology of the Oral Microbiome: Beyond Bacteria', *Trends in Microbiology*. Elsevier Ltd, 25(5), pp. 362–374. doi: 10.1016/j.tim.2016.12.012.

BANERJEE, S., TIAN, T., WEI, Z., PECK, K. N., SHIH, N., CHALIAN, A. A., O'MALLEY, B. W., WEINSTEIN, G. S., FELDMAN, M. D., *ET AL.* (2017) 'Microbial Signatures Associated with Oropharyngeal and Oral Squamous Cell Carcinomas', *Scientific Reports*, 7(1), pp. 1–20. doi: 10.1038/s41598-017-03466-6.

LE BASTARD, Q., AL-GHALITH, | G A, GR EGOIRE, | M, CHAPELET, | G, JAVAUDIN, | F, DAILLY, | E, BATARD, | E, KNIGHTS, | D AND MONTASSIER, | E (2017) 'Systematic review: human gut dysbiosis induced by non-antibiotic prescription medications'. doi: 10.1111/apt.14451.

BECHINGER, B. AND GORR, S. U. (2017) 'Antimicrobial Peptides: Mechanisms of Action and Resistance', *Journal of Dental Research*, 96(3), pp. 254–260. doi: 10.1177/0022034516679973.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., *et al.* (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature*, 456(7218), pp. 53–59. doi: 10.1038/nature07517.

Brady, L. J., Maddocks, S. E., Larson, M. R., Forsgren, N., Persson, K., Deivanayagam, C. C. and Jenkinson, H. F. (2010) 'The changing faces of Streptococcus antigen I/II polypeptide family adhesins', *Molecular microbiology*. NIH Public Access, 77(2), p. 276. doi: 10.1111/J.1365-2958.2010.07212.X.

Brennan, C. A. and Garrett, W. S. (2019) 'Fusobacterium nucleatum — symbiont, opportunist and oncobacterium', *Nature Reviews Microbiology*. Nature Publishing Group, 17(3), pp. 156–166. doi: 10.1038/s41579-018-0129-6.

Brüggemann, H., Henne, A., Hoster, F., Liesegang, H., Wiezer, A., Strittmatter, A., Hujer, S., Dürre, P. and Gottschalk, G. (2004) 'The complete genome sequence of Propionibacterium acnes, a commensal of human skin.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 305(5684), pp. 671–3. doi: 10.1126/science.1100330.

Byrd, A. L., Belkaid, Y. and Segre, J. A. (2018) 'The human skin microbiome', *Nature Publishing Group*, 16. doi: 10.1038/nrmicro.2017.157.

Callahan, B. J., McMurdie, P. J. and Holmes, S. P. (2017) 'Exact sequence variants should replace operational taxonomic units in marker-gene data analysis', *The ISME Journal*. Nature Publishing Group, 11(12), pp. 2639–2643. doi: 10.1038/ismej.2017.119.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. and Holmes, S. P. (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Cammarota, G., Ianiro, G. and Gasbarrini, A. (2014) 'Fecal microbiota transplantation for the treatment of clostridium difficile infection: A systematic review', *Journal of Clinical Gastroenterology*, 48(8), pp. 693–702. doi: 10.1097/MCG.0000000000000046.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N. and Knight, R. (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 108(Supplement 1), pp. 4516–4522. doi: 10.1073/PNAS.1000080107.

Carvalho, F. A., Koren, O., Goodrich, J. K., Johansson, M. E. V, Nalbantoglu, I., Aitken, J. D., Su, Y., Chassaing, B., Walters, W. A., *et al.* (2012) 'Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice.', *Cell host & microbe*. Elsevier, 12(2), pp. 139–52. doi: 10.1016/j.chom.2012.07.004.

Commins, S. P. (2015) 'Mechanisms of Oral Tolerance.', *Pediatric clinics of North America*. NIH Public Access, 62(6), pp. 1523–9. doi: 10.1016/j.pcl.2015.07.013.

Coppenhagen-Glazer, S., Sol, A., Abed, J., Naor, R., Zhang, X., Han, Y. W. and Bachrach, G. (2015) 'Fap2

of Fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth.', *Infection and immunity*. American Society for Microbiology (ASM), 83(3), pp. 1104–13. doi: 10.1128/IAI.02838-14.

COUSIDO, M. C., TOMÁS CARMONA, I., GARCÍA-CABALLERO, L., LIMERES, J., ÁLVAREZ, M. AND DIZ, P. (2010) 'In vivo substantivity of 0.12% and 0.2% chlorhexidine mouthrinses on salivary bacteria', *Clinical Oral Investigations*. Springer-Verlag, 14(4), pp. 397–402. doi: 10.1007/s00784-009-0320-2.

CRAVEN, M., EGAN, C. E., DOWD, S. E., MCDONOUGH, S. P., DOGAN, B., DENKERS, E. Y., BOWMAN, D., SCHERL, E. J. AND SIMPSON, K. W. (2012) 'Inflammation Drives Dysbiosis and Bacterial Invasion in Murine Models of Ileal Crohn's Disease', *PLoS ONE*. Edited by M. M. Heimesaat. Public Library of Science, 7(7), p. e41594. doi: 10.1371/journal.pone.0041594.

D'AMORE, R., IJAZ, U. Z., SCHIRMER, M., KENNY, J. G., GREGORY, R., DARBY, A. C., SHAKYA, M., PODAR, M., QUINCE, C., *ET AL.* (2016) 'A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling', *BMC Genomics*. BioMed Central, 17(1), p. 55. doi: 10.1186/s12864-015-2194-9.

DAVID, L. A., MATERNA, A. C., FRIEDMAN, J., CAMPOS-BAPTISTA, M. I., BLACKBURN, M. C., PERROTTA, A., ERDMAN, S. E. AND ALM, E. J. (2014) 'Host lifestyle affects human microbiota on daily timescales', *Genome Biology*. BioMed Central, 15(7), p. R89. doi: 10.1186/gb-2014-15-7-r89.

DESANTIS, T. Z., HUGENHOLTZ, P., LARSEN, N., ROJAS, M., BRODIE, E. L., KELLER, K., HUBER, T., DALEVI, D., HU, P., *ET AL.* (2006) 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 72(7), pp. 5069–72. doi: 10.1128/AEM.03006-05.

DIXON, P. (2003) 'VEGAN, a package of R functions for community ecology', *Journal of Vegetation Science*, 14(6), pp. 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x.

EDGAR, R. C. (2013) 'UPARSE: highly accurate OTU sequences from microbial amplicon reads', *Nature Methods*. Nature Publishing Group, 10(10), pp. 996–998. doi: 10.1038/nmeth.2604.

EDGAR, R. C. (2018) 'Updating the 97% identity threshold for 16S ribosomal RNA OTUs', *Bioinformatics*. Edited by A. Valencia. Narnia, 34(14), pp. 2371–2375. doi: 10.1093/bioinformatics/bty113.

ESBERG, A., HAWORTH, S., BRUNIUS, C., LIF HOLGERSON, P. AND JOHANSSON, I. (2019) 'Carbonic Anhydrase 6 Gene Variation influences Oral Microbiota Composition and Caries Risk in Swedish adolescents', *Scientific Reports*. Nature Publishing Group, 9(1), p. 452. doi: 10.1038/s41598-018-36832-z.

FADROSH, D. W., BING MA, P. G., SENGAMALAY, N., OTT, S., BROTMAN, R. M. AND RAVEL, J. (2014) 'An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform', *Microbiome*, 2(6), pp. 1–7. doi: 10.1186/2049-2618-2-6.

FERLAY, J., SOERJOMATARAM, I., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D. M., FORMAN, D. AND BRAY, F. (2015) 'Cancer incidence and mortality worldwide: Sources, methods and major

patterns in GLOBOCAN 2012', *International Journal of Cancer*. John Wiley & Sons, Ltd, 136(5), pp. E359–E386. doi: 10.1002/ijc.29210.

Fine, D. H., Markowitz, K., Furgang, D., Fairlie, K., Ferrandiz, J., Nasri, C., McKiernan, M. and Gunsolley, J. (2007) 'Aggregatibacter actinomycetemcomitans and its relationship to initiation of localized aggressive periodontitis: longitudinal cohort study of initially healthy adolescents.', *Journal of clinical microbiology*. American Society for Microbiology, 45(12), pp. 3859–69. doi: 10.1128/JCM.00653-07.

Flores, G. E., Caporaso, J. G., Henley, J. B., Rideout, J. R., Domogala, D., Chase, J., Leff, J. W., Vázquez-Baeza, Y., Gonzalez, A., *et al.* (2014) 'Temporal variability is a personalized feature of the human microbiome', *Genome Biology*. BioMed Central, 15(12), p. 531. doi: 10.1186/s13059-014-0531-y.

Gagnière, J., Raisch, J., Veziant, J., Barnich, N., Bonnet, R., Buc, E., Bringer, M. A., Pezet, D. and Bonnet, M. (2016) 'Gut microbiota imbalance and colorectal cancer', *World Journal of Gastroenterology*, 22(2), pp. 501–518. doi: 10.3748/wjg.v22.i2.501.

García-Jiménez, B. and Wilkinson, M. D. (2019) 'Robust and automatic definition of microbiome states', *PeerJ*. PeerJ Inc., 7, p. e6657. doi: 10.7717/peerj.6657.

Garcia, G. R., Noyes, P. D. and Tanguay, R. L. (2016) 'Advancements in zebrafish applications for 21st century toxicology', *Pharmacology & Therapeutics*. Pergamon, 161, pp. 11–21. doi: 10.1016/J.PHARMTHERA.2016.03.009.

Gaulke, C. A., Barton, C. L., Proffitt, S., Tanguay, R. L. and Sharpton, T. J. (2016) 'Triclosan exposure is associated with rapid restructuring of the microbiome in adult zebrafish', *PLoS ONE*, 11(5), pp. 1–20. doi: 10.1371/journal.pone.0154632.

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A. and Gillevet, P. M. (2010) 'Characterization of the oral fungal microbiome (mycobiome) in healthy individuals.', *PLoS pathogens*. Public Library of Science, 6(1), p. e1000713. doi: 10.1371/journal.ppat.1000713.

Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., NISC Comparative Sequencing Program, N. C. S., Bouffard, G. G., Blakesley, R. W., *et al.* (2009) 'Topographical and temporal diversity of the human skin microbiome.', *Science (New York, N.Y.)*. NIH Public Access, 324(5931), pp. 1190–2. doi: 10.1126/science.1171700.

Gruner, D., Paris, S. and Schwendicke, F. (2016) 'Probiotics for managing caries and periodontitis: Systematic review and meta-analysis', *Journal of Dentistry*. Elsevier, 48, pp. 16–25. doi: 10.1016/J.JDENT.2016.03.002.

Guo, L., Shokeen, B., He, X., Shi, W. and Lux, R. (2017) 'Streptococcus mutans SpaP binds to RadD of Fusobacterium nucleatum ssp. polymorphum.', *Molecular oral microbiology*. NIH Public Access, 32(5), pp. 355–364. doi: 10.1111/omi.12177.

Hall, M. W., Singh, N., Ng, K. F., Lam, D. K., Goldberg, M. B., Tenenbaum, H. C., Neufeld, J. D., Beiko, R. G. and Senadheera, D. B. (2017) 'Inter-personal diversity and temporal dynamics of dental, tongue,

and salivary microbiota in the healthy oral cavity', *npj Biofilms and Microbiomes*. Springer US, 3(1), pp. 0–1. doi: 10.1038/s41522-016-0011-0.

Hannig, C., Hannig, M., Kensche, A. and Carpenter, G. (2017) 'The mucosal pellicle – An underestimated factor in oral physiology', *Archives of Oral Biology*. Pergamon, pp. 144–152. doi: 10.1016/j.archoralbio.2017.04.001.

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., *et al.* (2012) 'Environmental biodiversity, human microbiota, and allergy are interrelated.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(21), pp. 8334–9. doi: 10.1073/pnas.1205624109.

He, Q., Wang, L., Wang, F., Wang, C., Tang, C., Li, Q., Li, J. and Zhao, Q. (2013) 'Microbial fingerprinting detects intestinal microbiota dysbiosis in Zebrafish models with chemically-induced enterocolitis', *BMC Microbiology*. BioMed Central, 13(1), p. 289. doi: 10.1186/1471-2180-13-289.

He, X., McLean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S.-Y., Dorrestein, P. C., Esquenazi, E., Hunter, R. C., *et al.* (2015) 'Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 112(1), pp. 244–249. doi: 10.1073/PNAS.1419038112.

Ho, M.-H., Lamont, R. J. and Xie, H. (2017) 'A novel peptidic inhibitor derived from Streptococcus cristatus ArcA attenuates virulence potential of Porphyromonas gingivalis', *Scientific Reports*, 7(16217). doi: 10.1038/s41598-017-16522-y.

Holland, K. T., Greenman, J. and Cunliffe, W. J. (1979) 'Growth of Cutaneous Propionibacteria on Synthetic Medium; Growth Yields and Exoenzyme Production', *Journal of Applied Bacteriology*. John Wiley & Sons, Ltd (10.1111), 47(3), pp. 383–394. doi: 10.1111/j.1365-2672.1979.tb01198.x.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*. Nature Publishing Group, 486(7402), pp. 207–214. doi: 10.1038/nature11234.

Karlsson, J., von Hofsten, J. and Olsson, P.-E. (2001) 'Generating Transparent Zebrafish: A Refined Method to Improve Detection of Gene Expression During Embryonic Development', *Marine Biotechnology*. Springer-Verlag, 3(6), pp. 0522–0527. doi: 10.1007/s1012601-0053-4.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. (2013) 'Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.', *Nucleic acids research*. Oxford University Press, 41(1), p. e1. doi: 10.1093/nar/gks808.

Kumar, S., Tadakamadla, J. and Johnson, N. W. (2016) 'Effect of Toothbrushing Frequency on Incidence and Increment of Dental Caries', *Journal of Dental Research*. SAGE PublicationsSage CA: Los Angeles, CA, 95(11), pp. 1230–1236. doi: 10.1177/0022034516655315.

Kutsuna, R., Tomida, J., Morita, Y. and Kawamura, Y. (2018) 'Paraclostridium bifermentans

exacerbates pathosis in a mouse model of ulcerative colitis', *PLOS ONE*. Edited by X. Han. Public Library of Science, 13(5), p. e0197668. doi: 10.1371/journal.pone.0197668.

Lawrence, C. (2007) 'The husbandry of zebrafish (Danio rerio): A review', *Aquaculture*. Elsevier, 269(1–4), pp. 1–20. doi: 10.1016/J.AQUACULTURE.2007.04.077.

Lee, K. A. and Lee, W. J. (2014) 'Drosophila as a model for intestinal dysbiosis and chronic inflammatory diseases', *Developmental and Comparative Immunology*. Elsevier Ltd, 42(1), pp. 102–110. doi: 10.1016/j.dci.2013.05.005.

Li, H., Thompson, I., Carter, P., Whiteley, A., Bailey, M., Leifert, C. and Killham, K. (2007) 'Salivary nitrate - An ecological factor in reducing oral acidity', *Oral Microbiology and Immunology*, 22(1), pp. 67–71. doi: 10.1111/j.1399-302X.2007.00313.x.

Liu, P.-F., Shi, W., Zhu, W., Smith, J. W., Hsieh, S.-L., Gallo, R. L. and Huang, C.-M. (2010) 'Vaccination targeting surface FomA of Fusobacterium nucleatum against bacterial co-aggregation: Implication for treatment of periodontal infection and halitosis', *Vaccine*. Elsevier, 28(19), pp. 3496–3505. doi: 10.1016/J.VACCINE.2010.02.047.

Lizcano, A., Sanchez, C. J. and Orihuela, C. J. (2012) 'A role for glycosylated Serine-rich repeatproteins in Gram-positive bacterial pathogenesis', *Molecular Oral Microbiology*. NIH Public Access, 27(4), p. 257. doi: 10.1111/J.2041-1014.2012.00653.X.

Mardis, E. R. (2013) 'Next-Generation Sequencing Platforms', *Annual Review of Analytical Chemistry*. Annual Reviews, 6(1), pp. 287–303. doi: 10.1146/annurev-anchem-062012-092628.

Marsh, P. D., Do, T., Beighton, D. and Devine, D. A. (2016) 'Influence of saliva on the oral microbiota', *Periodontology 2000*. Wiley/Blackwell (10.1111), pp. 80–92. doi: 10.1111/prd.12098.

Marsh, P. D. and Zaura, E. (2017) 'Dental biofilm: ecological interactions in health and disease', *Journal of Clinical Periodontology*. John Wiley & Sons, Ltd (10.1111), 44, pp. S12–S22. doi: 10.1111/jcpe.12679.

Mashima, I. and Nakazawa, F. (2015) 'The interaction between Streptococcus spp. and Veillonella tobetsuensis in the early stages of oral biofilm formation.', *Journal of bacteriology*. American Society for Microbiology (ASM), 197(3), pp. 2104–2111. doi: 10.1128/JB.02512-14.

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLOS ONE*. Public Library of Science, 8(4), p. e61217. doi: 10.1371/JOURNAL.PONE.0061217.

Minter, M. R., Zhang, C., Leone, V., Ringus, D. L., Zhang, X., Oyler-Castrillo, P., Musch, M. W., Liao, F., Ward, J. F., *et al.* (2016) 'Antibiotic-induced perturbations in gut microbial diversity influences neuro-inflammation and amyloidosis in a murine model of Alzheimer's disease', *Scientific Reports*. Nature Publishing Group, 6(1), p. 30028. doi: 10.1038/srep30028.

Mishra, A., Wu, C., Yang, J., Cisar, J. O., Das, A. and Ton-That, H. (2010) 'The Actinomyces oris type 2

fimbrial shaft FimA mediates co-aggregation with oral streptococci, adherence to red blood cells and biofilm development.', *Molecular microbiology*. NIH Public Access, 77(4), pp. 841–54. doi: 10.1111/j.1365-2958.2010.07252.x.

MOAYYEDI, P., SURETTE, M. G., KIM, P. T., LIBERTUCCI, J., WOLFE, M., ONISCHI, C., ARMSTRONG, D., MARSHALL, J. K., KASSAM, Z., *ET AL.* (2015) 'Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial', *Gastroenterology*. W.B. Saunders, 149(1), pp. 102-109.e6. doi: 10.1053/J.GASTRO.2015.04.001.

MOOSER, C., GOMEZ DE AGÜERO, M. AND GANAL-VONARBURG, S. C. (2018) 'Standardization in host–microbiota interaction studies: challenges, gnotobiology as a tool, and perspective', *Current Opinion in Microbiology*. Elsevier Current Trends, 44, pp. 50–60. doi: 10.1016/J.MIB.2018.07.007.

MOYNIHAN, P. J. AND KELLY, S. A. M. (2014) 'Effect on caries of restricting sugars intake: Systematic review to inform WHO guidelines', *Journal of Dental Research*. doi: 10.1177/0022034513508954.

MUYZER, G., DE WAAL, E. C. AND UITTERLINDEN, A. G. (1993) 'Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 59(3), pp. 695–700. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7683183 (Accessed: 30 April 2019).

NARUSHIMA, S., SUGIURA, Y., OSHIMA, K., ATARASHI, K., HATTORI, M., SUEMATSU, M. AND HONDA, K. (2014) 'Characterization of the 17 strains of regulatory T cell-inducing human-derived Clostridia.', *Gut microbes*. Taylor & Francis, 5(3), pp. 333–9. doi: 10.4161/gmic.28572.

NASCIMENTO, M. M. (2017) 'Oral microbiota transplant: a potential new therapy for oral diseases.', *Journal of the California Dental Association*. NIH Public Access, 45(10), pp. 565–568. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29497269 (Accessed: 29 April 2019).

NOBBS, A. H., JENKINSON, H. F. AND JAKUBOVICS, N. S. (2011) 'Stick to your gums: Mechanisms of oral microbial adherence', *Journal of Dental Research*, 90(11), pp. 1271–1278. doi: 10.1177/0022034511399096.

OH, J., BYRD, A. L., DEMING, C., CONLAN, S., KONG, H. H., SEGRE, J. A., BLAKESLEY, R., BOUFFARD, G., BROOKS, S., *ET AL.* (2014) 'Biogeography and individuality shape function in the human skin metagenome', *Nature*. Nature Publishing Group, 514(7520), pp. 59–64. doi: 10.1038/nature13786.

OH, J., CONLAN, S., POLLEY, E. C., SEGRE, J. A. AND KONG, H. H. (2012) 'Shifts in human skin and nares microbiota of healthy children and adults', *Genome Medicine*. BioMed Central, 4(10), p. 77. doi: 10.1186/gm378.

PACE, N. R., STAHL, D. A., LANE, D. J. AND OLSEN, G. J. (1985) 'Analyzing natural microbial populations by rRNA sequences.', *ASM News*, (51), pp. 4–12.

PARSONS, B. AND FOLEY, E. (2016) 'Cellular immune defenses of Drosophila melanogaster', *Developmental & Comparative Immunology*. Pergamon, 58, pp. 95–101. doi:

10.1016/J.DCI.2015.12.019.

VAN DE PEER, Y., CHAPELLE, S. AND DE WACHTER, R. (1996) 'A quantitative map of nucleotide substitution rates in bacterial rRNA.', *Nucleic acids research*. Oxford University Press, 24(17), pp. 3381–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8811093 (Accessed: 14 April 2019).

PETERSEN, C. AND ROUND, J. L. (2014) 'Defining dysbiosis and its influence on host immunity and disease.', *Cellular microbiology*. Wiley-Blackwell, 16(7), pp. 1024–33. doi: 10.1111/cmi.12308.

PITTS, N. B., ZERO, D. T., MARSH, P. D., EKSTRAND, K., WEINTRAUB, J. A., RAMOS-GOMEZ, F., TAGAMI, J., TWETMAN, S., TSAKOS, G., *ET AL.* (2017) 'Dental caries', *Nature Reviews Disease Primers*, 3(May). doi: 10.1038/nrdp.2017.30.

POZHITKOV, A. E., LEROUX, B. G., RANDOLPH, T. W., BEIKLER, T., FLEMMIG, T. F. AND NOBLE, P. A. (2015) 'Towards microbiome transplant as a therapy for periodontitis: an exploratory study of periodontitis microbial signature contrasted by oral health, caries and edentulism', *BMC Oral Health*. BioMed Central, 15(1), p. 125. doi: 10.1186/s12903-015-0109-4.

PROCTOR, D. M., FUKUYAMA, J. A., LOOMER, P. M., ARMITAGE, G. C., LEE, S. A., DAVIS, N. M., RYDER, M. I., HOLMES, S. P. AND RELMAN, D. A. (2018) 'A spatial gradient of bacterial diversity in the human oral cavity shaped by salivary flow', *Nature Communications*. Nature Publishing Group, 9(1), p. 681. doi: 10.1038/s41467-018-02900-1.

RAWLS, J. F., MAHOWALD, M. A., LEY, R. E. AND GORDON, J. I. (2006) 'Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection.', *Cell*. Elsevier, 127(2), pp. 423–33. doi: 10.1016/j.cell.2006.08.043.

DE REPENTIGNY, L., GOUPIL, M. AND JOLICOEUR, P. (2015) 'Oropharyngeal Candidiasis in HIV Infection: Analysis of Impaired Mucosal Immune Response to Candida albicans in Mice Expressing the HIV-1 Transgene.', *Pathogens (Basel, Switzerland)*. Multidisciplinary Digital Publishing Institute (MDPI), 4(2), pp. 406–21. doi: 10.3390/pathogens4020406.

ROESELERS, G., MITTGE, E. K., STEPHENS, W. Z., PARICHY, D. M., CAVANAUGH, C. M., GUILLEMIN, K. AND RAWLS, J. F. (2011) 'Evidence for a core gut microbiota in the zebrafish.', *The ISME journal*. Nature Publishing Group, 5(10), pp. 1595–608. doi: 10.1038/ismej.2011.38.

ROSA, L., CUTONE, A., LEPANTO, M. S., PAESANO, R. AND VALENTI, P. (2017) 'Lactoferrin: A natural glycoprotein involved in iron and inflammatory homeostasis', *International Journal of Molecular Sciences*, 18(9). doi: 10.3390/ijms18091985.

ROSEN, M. J., CALLAHAN, B. J., FISHER, D. S. AND HOLMES, S. P. (2012) 'Denoising PCR-amplified metagenome data', *BMC Bioinformatics*. BioMed Central, 13(1), p. 283. doi: 10.1186/1471-2105-13-283.

ROSENBAUM, J., USYK, M., CHEN, Z., ZOLNIK, C. P., JONES, H. E., WALDRON, L., DOWD, J. B., THORPE, L. E. AND BURK, R. D. (2019) 'Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota', *Scientific Reports*. Nature Publishing Group, 9(1), p. 1531. doi: 10.1038/s41598-018-

38049-6.

Rosier, B. T., Marsh, P. D. and Mira, A. (2018) 'Resilience of the Oral Microbiota in Health: Mechanisms That Prevent Dysbiosis', *Journal of Dental Research*, 97(4), pp. 371–380. doi: 10.1177/0022034517742139.

Round, J. L. and Mazmanian, S. K. (2010) 'Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(27), p. 12204. doi: 10.1073/PNAS.0909122107.

Roversi, D., Luca, V., Aureli, S., Park, Y., Mangoni, M. L. and Stella, L. (2014) 'How many antimicrobial peptide molecules kill a bacterium? The case of PMAP-23', *ACS Chemical Biology*, 9(9), pp. 2003–2007. doi: 10.1021/cb500426r.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., *et al.* (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology*. BioMed Central, 12(1), p. 87. doi: 10.1186/s12915-014-0087-z.

Sasi Jyothsna, T. S., Tushar, L., Sasikala, C. and Ramana, C. V (2016) 'Paraclostridium benzoelyticum gen. nov., sp. nov., isolated from marine sediment and reclassification of Clostridium bifermentans as Paraclostridium bifermentans comb. nov. Proposal of a new genus Paeniclostridium gen. nov. to accommodate Clostridium sordellii and Clostridium ghonii', *International Journal of Systematic and Evolutionary Microbiology*, 66, pp. 1268–1274. doi: 10.1099/ijsem.0.000874.

Scharschmidt, T. C. and Fischbach, M. A. (2013) 'What Lives On Our Skin: Ecology, Genomics and Therapeutic Opportunities Of the Skin Microbiome.', *Drug discovery today. Disease mechanisms*. NIH Public Access, 10(3–4). doi: 10.1016/j.ddmec.2012.12.003.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., *et al.* (2009) 'Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.', *Applied and environmental microbiology*. American Society for Microbiology, 75(23), pp. 7537–41. doi: 10.1128/AEM.01541-09.

Schrader, C., Schielke, A., Ellerbroek, L. and Johne, R. (2012) 'PCR inhibitors - occurrence, properties and removal', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd (10.1111), 113(5), pp. 1014–1026. doi: 10.1111/j.1365-2672.2012.05384.x.

Scopus (2019) *Scopus*, *Elsevier B.V.* Available at: https://www.scopus.com/results/results.uri?numberOfFields=0&src=s&clickedLink=&edit=&editSaveSearch=&origin=searchbasic&authorTab=&affiliationTab=&advancedTab=&scint=1&menu=search&tablin=&searchterm1=zebrafish&field1=TITLE_ABS_KEY&dateType=Publication_Dat (Accessed: 13 April 2019).

Simón-Soro, A. and Mira, A. (2015) 'Solving the etiology of dental caries.', *Trends in microbiology*. Elsevier, 23(2), pp. 76–82. doi: 10.1016/j.tim.2014.10.010.

Spence, R., Gerlach, G., Lawrence, C. and Smith, C. (2007) 'The behaviour and ecology of the zebrafish, Danio rerio', *Biological Reviews*. John Wiley & Sons, Ltd (10.1111), 83(1), pp. 13–34. doi: 10.1111/j.1469-185X.2007.00030.x.

Stanisavljević, S., Čepić, A., Bojić, S., Veljović, K., Mihajlović, S., Đedović, N., Jevtić, B., Momčilović, M., Lazarević, M., *et al.* (2019) 'Oral neonatal antibiotic treatment perturbs gut microbiota and aggravates central nervous system autoimmunity in Dark Agouti rats', *Scientific Reports*. Nature Publishing Group, 9(1), p. 918. doi: 10.1038/s41598-018-37505-7.

Szafrański, S. P., Deng, Z.-L., Tomasch, J., Jarek, M., Bhuju, S., Rohde, M., Sztajer, H. and Wagner-Döbler, I. (2017) 'Quorum sensing of Streptococcus mutans is activated by Aggregatibacter actinomycetemcomitans and by the periodontal microbiome', *BMC Genomics*. BioMed Central, 18(1), p. 238. doi: 10.1186/s12864-017-3618-5.

Tobin, D. M., May, R. C. and Wheeler, R. T. (2012) 'Zebrafish: A See-Through Host and a Fluorescent Toolbox to Probe Host–Pathogen Interaction', *PLoS Pathogens*. Edited by J. Heitman. Public Library of Science, 8(1), p. e1002349. doi: 10.1371/journal.ppat.1002349.

Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L. and Colombel, J.-F. (2017) 'Ulcerative colitis', *The Lancet*. Elsevier, 389(10080), pp. 1756–1770. doi: 10.1016/S0140-6736(16)32126-2.

Vallès, Y., Inman, C. K., Peters, B. A., Ali, R., Wareth, L. A., Abdulle, A., Alsafar, H., Anouti, F. Al, Dhaheri, A. Al, *et al.* (2018) 'Types of tobacco consumption and the oral microbiome in the United Arab Emirates Healthy Future (UAEHFS) Pilot Study', *Scientific Reports*. Nature Publishing Group, 8(1), p. 11327. doi: 10.1038/s41598-018-29730-x.

Wan, X., Ruan, R., McLaughlin, R. W., Hao, Y., Zheng, J. and Wang, D. (2016) 'Fecal Bacterial Composition of the Endangered Yangtze Finless Porpoises Living Under Captive and Semi-natural Conditions', *Current Microbiology*, 72(3), pp. 306–314. doi: 10.1007/s00284-015-0954-z.

Wang, A. R., Ran, C., Ringø, E. and Zhou, Z. G. (2018) 'Progress in fish gastrointestinal microbiota research', *Reviews in Aquaculture*. John Wiley & Sons, Ltd (10.1111), 10(3), pp. 626–640. doi: 10.1111/raq.12191.

Wang, H., Funchain, P., Bebek, G., Altemus, J., Zhang, H., Niazi, F., Peterson, C., Lee, W. T., Burkey, B. B., *et al.* (2017) 'Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas', *Genome Medicine*. BioMed Central, 9(1), p. 14. doi: 10.1186/s13073-017-0405-5.

Wiles, T. J., Jemielita, M., Baker, R. P., Schlomann, B. H., Logan, S. L., Ganz, J., Melancon, E., Eisen, J. S., Guillemin, K., *et al.* (2016) 'Host Gut Motility Promotes Competitive Exclusion within a Model Intestinal Microbiota', *PLOS Biology*. Edited by J. Gore. Public Library of Science, 14(7), p. e1002517. doi: 10.1371/journal.pbio.1002517.

Wu, S.-C., Cao, Z.-S., Chang, K.-M. and Juang, J.-L. (2017) 'Intestinal microbial dysbiosis aggravates the progression of Alzheimer's disease in Drosophila', *Nature Communications*. Nature Publishing Group, 8(1), p. 24. doi: 10.1038/s41467-017-00040-6.

Xu, Z., Wang, C., Dong, X., Hu, T., Wang, L., Zhao, W., Zhu, S., Li, G., Hu, Y., *et al.* (2019) 'Chronic alcohol exposure induced gut microbiota dysbiosis and its correlations with neuropsychic behaviors and brain BDNF/Gabra1 changes in mice', *BioFactors*. John Wiley & Sons, Ltd, 45(2), pp. 187–199. doi: 10.1002/biof.1469.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., *et al.* (2014) 'The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks', *Nucleic Acids Research*. Narnia, 42(D1), pp. D643–D648. doi: 10.1093/nar/gkt1209.

Yoon, B. K., Jackman, J. A., Valle-González, E. R. and Cho, N.-J. (2018) 'Antibacterial Free Fatty Acids and Monoglycerides: Biological Activities, Experimental Testing, and Therapeutic Applications.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 19(4). doi: 10.3390/ijms19041114.

Zheng, W., Tsompana, M., Ruscitto, A., Sharma, A., Genco, R., Sun, Y. and Buck, M. J. (2015) 'An accurate and efficient experimental approach for characterization of the complex oral microbiota', *Microbiome*. BioMed Central, 3(1), p. 48. doi: 10.1186/s40168-015-0110-9.

Zhu, S. (2014) *DNA extraction from TRIZOL _ organic _ phase*, *Stanford Medicine, Department of Pathology*. Available at: http://med.stanford.edu/labs/vanderijn-west/Protocols.html (Accessed: 30 April 2019).