QATAR UNIVERSITY

COLLEGE OF HEALTH SCIENCES

PROFILING THE SALIVARY MICROBIOME IN THE QATARI POPULATION

BY

SARA FAHAD MOHAMMED AL-AHMAD

A Thesis Submitted to

the Faculty of the College of Health Sciences

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Biomedical Sciences

June 2019

© 2019 Sara Al-Ahmad. All Rights Reserved.

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Sara Al-Ahamd defended on 17/04/2019.

Souhaila Al Khodor, Marwan Abu-Madi
Thesis/Dissertation Supervisor

Maha Abdulla Al-Asmakh
Committee Member

Muhammed Umar Sohail
Committee Member

Asma Al-Thani
Committee Member

Add Member

Approved:

Asma Al-Thani, Dean, College of Health Science

ABSTRACT

AL-AHMAD, SARA, F.M., Masters of Science: June : 2019, Biomedical Sciences

Title: PROFILING the SALIVARY MICROBIOME in the QATARI POPULATION

Supervisor ofThesis: Souhaila, Al Khodor.

Humans are living ecosystems composed of human cells and microbes. The microbiome is the collection of microbes and their genes. Recent breakthrough in the high throughput sequencing technologies made it possible for us to understand the composition of the human microbiome. Launched by the National Institutes of Health in USA, the human microbiome project indicated that our bodies harbor a wide array of microbes, specific to each body site with inter and intra-personal variabilities. Numerous studies have indicated that, the microbiome composition plays an important role in health and disease, thus highlighting the significance of microbiome research in human health.

Saliva is a biofluid secreted from salivary glands composed of water, electrolytes, mucus, DNA, RNA, proteins, enzymes and microbes. Several studies assessed the role of the salivary microbiome in many conditions ranging from local diseases of the oral cavity such as dental carries and gingivitis to neurodevelopmental disease such as autism, indicating the potential of applying the knowledge generated from the salivary microbiome projects towards a better understanding of various pathological conditions.

In this study, we aim to profile the salivary microbiome of the Qatari population and identify the oral microbial communities in individuals with diabetes or obesity.

100 saliva samples collected from Qatari participants, selected randomly, were retrieved from Qatar Biobank repository. Samples were collected by spitting in a tube. After microbial DNA extraction, 16S rRNA gene was sequenced using Illumina Miseq. Microbial profiles were then correlated with the individual phenotypic and clinical data to identify the microbial signatures associated with health and disease conditions, with special focus on diabetes and obesity due to the increasing prevalence rate of both conditions in Qatar.

DEDICATION

I would like to dedicate this work to those who supported me during my study. I am grateful for having them in my life.

ACKNOWLEDGMENTS

I would like to especially thank all members of the Microbiome and Host-Microbe Interaction team in the Research Department at Sidra Medicine.

Special thanks to Dr. Selvasankar Murugesan for his immense help in the analysis part of the project.

TABLE OF CONTENTS

DEDICATION v

ACKNOWLEDGMENTS vi

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: REVIEW OF THE LITERATURE 3

1. The human microbiome project 3

2. Development of the microbiome: Inception from the early days of life 3

3. The salivary microbiome 4

4. The gut microbiome 6

5. Factors that affect the salivary microbiome composition 6

5.1. Oral hygiene 8

5.2. Smoking 9

6. Microbiome in disease: Microbial dysbiosis 9

7. Tools used to assess the microbiome composition: 16S rRNA gene sequencing 11

8. Aims of the project 12

CHAPTER 3: METHODOLOGY 13

1. Description of the study participants 13

2. Inclusion/ Exclusion criteria 13

3. Sample collection 14

4. DNA extraction from saliva using QIAsymphony 14

5. DNA Quality control 16

6. Amplification of 16S rRNA gene fragment by polymerase chain reaction 16

7. PCR thermal conditions and duration 18

8. Gel electrophoresis 18

9. 16S library magnetic bead purification 19

10. AmpPure magnetic bead purification 19

11. Assessment of the PCR product quality using Agilent bioanalyzer and Qubit fluorometer 20

11.1. Agilent High sensitivity kit 20

11.2. Qubit Fluorometer 21

12. Sequencing 21

13. Sequencing Procedure 23

14. Data Analysis 24

CHAPTER 4: RESULTS 25

1. **Description of the study participants** 25

2. Microbial DNA Quality and Quantity 26

3. **PCR product visualization using agarose gel electrophoresis** 26

4. **Taxa summary** 27

5. **Relative abundance of the salivary microbiome at the Genus level** 30

6. **Comparative relative abundance between various categories** 32

7. **Interindividual and intraindividual variability: Alpha and beta-diversity** 34

8. **Correlation between clinical variables and microbial data: HALLA** 37

9. **Significant differences among clinical categories using LDA (Linear discriminant analysis) Effect Size (LEfSe)** 41

10. **Predictive functional profiling of microbial communities using PiCrust** 42

CHAPTER 5: DISCUSSION 45

REFRENCES 49

LIST OF TABLES

Table 1 17

Table 2 17

Table 3 25

LIST OF FIGURES

Figure 1:Summerized method workflow 13

Figure 2: Schematic representation of the QIAsymphony SP protocol used for DNA extraction.......... 15

Figure 3: 16S rRNA gene: the red boxes represent the variable regions, whereas the blue areas in between the boxes represents the conserved region. 16

Figure 4: AMPure XP magnetic beads purification of the PCR product.......... 19

Figure 5: Sequencing by synthesis, Source:(Mardis, 2013) 23

Figure 6: Gel electrophoresis for PCR products 26

Figure 7: Relative abundance of bacterial phyla among salivary samples 27

Figure 8: Comparison of the microbial taxa at the phylum level in normal weight, overweight and obese.......... 29

Figure 9: Taxa summary of the phylum level for Diabetic Versus Normal 30

Figure 10: Taxa summary of the genus level for BMI categories with most abundant genera 31

Figure 11: Taxa summary of the genus level for Diabetic versus Normal with most abundant genera 32

Figure 12: Hierarchical clustering heatmap of top 25 genera between Normal, Overweight and Obese group.......... 33

Figure 13: Hierarchical clustering heatmap of top 25 genera between diabetic and normal group. 34

Figure 14: Alpha diversity of diabetic and normal (non-diabetic) subjects.......... 35

Figure 15: Chao1 analysis for alpha diversity of BMI categories 36

Figure 16: Principle of Coordinate Analysis (PCoA) by applying weighted UniFrac of

BMI category; red triangle: normal, blue: obese, circle: overweight 37

Figure 17: Principle of Coordinate Analysis (PCoA) by applying weighted UniFrac of diabetic and normal; blue: normal, red: diabetic 37

Figure 18: Associations between clinical parameters and microbial composition (genus level) for all participants 38

Figure 19: Associations between clinical parameters and microbial composition (genus level) for the Diabetic group 39

Figure 20: Associations between clinical parameters and microbial composition (genus level) for the normal weight group 40

Figure 21: Associations between clinical parameters and microbial composition (genus level) for the obese group 40

Figure 22: Associations between clinical parameters and microbial composition (genus Level) for Overweight group 41

Figure 23: LefSe analysis of BMI groups and their microbial composition................42

Figure 24: PiCrust analysis of Normal and diabetic in relation to N-glycan biosynthesis 43

Figure 25: PiCrust analysis of normal weight and obese in relation to N-glycan biosynthesis (significant difference was observed, p value <0.05) 44

CHAPTER 1: INTRODUCTION

The term microbiome refers to the collection of genomes annotated for all the microbes that form a habitat in a specific body site (Ursell, Metcalf, Parfrey, & Knight, 2012). The other term that has been used extensively in the literature is Microbiota, which refers to the collection of the microbes themselves (Ursell et al., 2012). The advent in the molecular techniques especially sequencing methods, has made the study of the microbiome possible, since not all microbes are cultivable (Consortium, 2012a).

In an effort to introduce a profile of the human microbiota, the Human Microbiome Project (HMP) was initiated in 2008 with an aim to identify and characterize the core microbiota in humans using samples collected from different body sites (Peterson et al., 2009). The project achieved major milestones introducing 3000 reference sets of genome sequences that represent microbial isolates from human samples (Peterson et al., 2009). This database is now used as a reference for researchers aiming to assess the microbiome composition in various study designs (Consortium, 2012a). The microbiome composition can be compared in health and disease (Althani et al., 2016), in longitudinal studies, microbial communities can be compared between seasons, between different subjects, or within the same subject but at different time periods (Cameron, Huws, Hegarty, Smith, & Mur, 2015).

Qatar Biobank (QBB) is an organization that aims to promote public health in Qatar by recruiting participants, collecting and storing a series of biological samples along with a comprehensive assessment of the individual phenotypic, dietary and clinical data among others. Saliva samples used in this study were collected from participants recruited in QBB (discussed in more details in the chapter 3: Materials and Methods) for the purpose of investigating the salivary microbiome.

In this study, we aim to profile the salivary microbiome of the Qatari population and identify the oral microbial communities in individuals with diabetes or obesity. A total of 100 saliva samples randomly selected from the QBB participants, were obtained for microbial profiling. The microbiome profiles of the 100 saliva samples were correlated with the clinical data and phenotypic data, in order to identify microbial signatures associated with various diseases or pathological conditions. In this project, we focused on diabetes and obesity due to their increasing prevalence rate in Qatar.

HYPOTHESIS:

In this study, we hypothesize that saliva samples have a microbial composition that vary between individuals, and that specific microbial signatures can be associated with various pathological conditions and diseases.

To test this hypothesis, we aim to:

1. Study the microbiome composition of 100 saliva samples collected from Qatari participants in QBB.
2. Correlate the microbial composition with the clinical data in order to identify specific signatures associated with various pathological conditions.
3. Identify salivary microbial communities associated with diabetes and obesity.

CHAPTER 2: REVIEW OF THE LITERATURE

1. The human microbiome project

The human microbiome has gained a considerable attention especially after the initiation of the HMP (Turnbaugh et al., 2007). In order to characterize the normal microbiota of different body habitats in healthy adults, an extensive sampling was planned across two time points. (Consortium, 2012b). The HMP recruited 242 volunteers (129 males, 113 females) and sampled tissues from 15 body sites in men and 18 body sites in women (Huttenhower et al., 2012). By incorporating several complementary techniques and analyses including 16S ribosomal RNA (rRNA) gene sequencing, whole-genome shotgun sequencing (WGS) and aligning the assembled sequences to the reference microbial genomes they were able to define the microbiome composition of each body site they sampled from (Huttenhower et al., 2012). Several studies has been conducted afterward in order to characterize the core microbiome that make up a status of symbiosis or in other words "microbiome in health" (Lloyd-Price et al., 2017). A healthy individual maintains a unique balance between the microbiome, immune system for protection from invading bacteria (Belkaid & Hand, 2014). However, in some cases pathogenic bacteria increases in numbers replacing the beneficial bacteria (microbial dysbiosis) leading to a significant impact on our health (Carding, Verbeke, Vipond, Corfe, & Owen, 2015).

2. Development of the microbiome: Inception from the early days of life

The process of microbiome colonization starts early in life even before birth

(Mueller, Bakacs, Combellick, Grigoryan, & Dominguez-Bello, 2015). The microbial contact during prenatal life and the inoculum transferred during birth and breastfeeding imprint the infant's microbiota and the immune system.

The mode of birth was shown to play an important role in the composition of the microbiome in the early days of life as babies born through a cesarean section carry a different microbiome as compared to those who were born naturally (Francavilla, Cristofori, Tripaldi, & Indrio, 2018). Moreover, the mode of feeding the neonates was also shown to play an important role in the seeding of their gut microbiome, as babies who were strictly breastfed had a different gut microbiome composition as compared to those who were mixed fed with formula milk (Backhed et al., 2015). By the end of the second year of age, the taxonomic composition of the gut microbiome stabilizes and converges towards a characteristic adult gut microbiome.

3. The salivary microbiome

Saliva is produced by multiple salivary glands, the major glands (major is referred to their size) which are the parotid glands and the submandibular gland (Humphrey & Williamson, 2001). The major glands accounts for 90% of salivary secretion (de Paula et al., 2017). The minor glands which are found in the tongue, cheeks and lower lip account for 10% of the produced saliva, these glands are responsible for the production of serous saliva (de Paula et al., 2017). The average amount normally produced by humans ranges from 1 to 1.5 liters of saliva daily (Humphrey & Williamson, 2001).

Saliva contains several components such as electrolytes, proteins, immunoglobulins, enzymes and microbes (Humphrey & Williamson, 2001). Electrolytes together with proteins for examples maintain the process of mineralization and remineralization (Humphrey & Williamson, 2001). Immunoglobulins help in the neutralization of harmful pathogens without eliciting an inflammation. Urea, bicarbonate and phosphatase maintain a steady salivary pH of 6 to 7 (Humphrey & Williamson, 2001).

One of the major components that promote the colonization of bacteria and fungi is mucin (Humphrey & Williamson, 2001). Mucins are non-immunologic glycosylated proteins that also act as lubricants, and provide protection from acid penetration to the cells by forming a barrier (Humphrey & Williamson, 2001). Being highly available, saliva is considered as an *easy to collect* sample that does not require hospitalization or special preparation (Kaczor-Urbanowicz et al., 2017).

Salivary microbiome is temporally stable as it does not fluctuate according to the circadian rhythm (Belstrøm et al., 2016). Collection of saliva can be achieved by several methods including spitting, swabbing, and the draining method (Priya & Prathibha, 2017). The method of collection exert a minimal effect on the extraction of microbial DNA (Y. Lim, Totsika, Morrison, & Punyadeera, 2017).

The healthy adult human mouth hosts a complex and resilient ecosystem of hundreds of different microbial species (Rosier, Marsh, & Mira, 2018). These microbes reside in different sites of the oral cavity, which is mainly composed of a soft mucosa that is constantly shedding, and the hard surface which comprises the teeth (He, Li, Cao, Xue, & Zhou, 2015). The saliva is a representing constituent of both locations but more profoundly the soft mucosa (He et al., 2015). According to

many studies, the most predominant phyla of the salivary microbiome are Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, and Fusobacteria (Ahn et al., 2011;(Zhou et al., 2016)).

Located at the opening of the gastrointestinal tract, the oral cavity provides a convenient, accessible site for collecting and analyzing microbial samples in the saliva (Kodukula et al., 2017). It is also worth noting that, the salivary microbiome mirrors the gut microbiome in terms of complexity and diversity (Kodukula et al., 2017).

4. The gut microbiome

The gut microbiome had been studied extensively for many reasons such as the immense populations of microbiota residing at this site, the easiness of obtaining the fecal matter as a representative of the microbiota of that site (Lynch & Pedersen, 2016). It is estimated that the number of gut microbiota may be more than 10^{14} and the genomic content of the microbiota is 100 times more than the human genome (Thursby & Juge, 2017). The gut microbiome was described as part of the MetaHIT cohorts and indicated that a large proportion of the microbial genes are shared between individuals of this cohort, and more than 99% of these genes are bacterial; representing 1000 to 1150 bacterial species(Dusko Ehrlich, 2010). A healthy gut has mainly five phyla which is mainly composed of strict anaerobes and dominated by 2 main phyla: Bacteroidetes and Firmicutes(Rinninella et al., 2019).

5. Factors that affect the salivary microbiome composition

Several pre and post-natal factors can significantly alter the composition of the

salivary microbiome including host genetics, the mode of delivery at birth; the method of infant feeding; teeth eruption, the use of medications, especially antibiotics; smoking, oral hygiene and diet among others.

The salivary microbiome in neonates was shown to be different in breast-fed or formula fed babies (Al-Shehri et al., 2016). The study had shown that there is a higher prevalence of *Bacteroidetes* in the mouths of formula-fed infants when compared to breast-fed infants, but in contrast *Actinobacteria* were more prevalent in breast-fed babies and *Proteobacteria* was more prevalent in the saliva of breast-fed babies when compared to the than in formula-fed neonates (Al-Shehri et al., 2016).

Moreover, the Human Oral Microbe Identification Microarray (HOMIM) was used to compare the oral microbiome of babies who were born via c-section versus those who were born naturally, and found the later has a greater number of taxa (Lif Holgerson, Harnevik, Hernell, Tanner, & Johansson, 2011). The newborn oral cavity initial colonizers are *Streptococcus salivarius*, since it has the ability to adhere to the epithelial surfaces and produce polymers that facilitate the adherence of other bacteria including *Actinomyces sp* (Sampaio-Maia & Monteiro-Silva, 2014).

When eruption of teeth starts, the colonization of *Streptococcus mutan*s begins, which is considered a cariogenic *Streptococcus sp* due their preference on colonizing hard tissue such as teeth (Sampaio-Maia & Monteiro-Silva, 2014). Although, it was also found that *S. mutans are also present in* edentulous infants (Cephas et al., 2011).

The major role that the diet plays in modulating the microbiome composition, promoted the potential for interventional studies using diet or dietary products to alter and improve the microbiome composition especially the gut microbiome in order to improve health (Wen & Duffy, 2017). A recent longitudinal study examined the effect

of the consumption of yogurt rich with probiotic versus using a standard yogurt on the composition of the salivary microbiome, an increase in the alpha diversity after a short term consumption of probiotics was observed (Dassi et al., 2018).

While genetics was shown to play a role in the microbiome composition (Julia K. Goodrich et al., 2014), a microbiome study conducted for Ashkenazi Jewish population by analyzing the salivary microbiome for pedigrees of families, concluded that the environment and shared household, plays a greater role in shaping the microbiome composition when compared to genetics (Shaw et al., 2017). While multifactorial, it is clear that the environment and host genetics together are major modulators of the microbiome composition (M. Y. Lim et al., 2014).

The use of broad-spectrum antibiotics has negative effects on the composition of the gut microbiota, which results in significant drops in taxonomic richness, diversity and evenness of the gut microbial communities (Thiemann, Smit, & Strowig, 2016). Another study has concluded that individuals exposed to a single antibiotic treatment, undergo considerable microbial shifts and enrichment in antibiotic resistance in their gut microbiome composition, while their salivary microbiome composition remains unexpectedly stable (Zaura et al., 2015). On the other hand, it was shown that children receiving antibiotics treatment for otitis media infection have a microbial shift in their salivary microbiome that was recovered three weeks after the treatment ended (Lazarevic et al., 2013).

5.1. Oral hygiene

Saliva flow along with good hygienic practices both aid in the detachment of biofilms that are known to cause proliferation of pathogenic species (Zarco, Vess, & Ginsburg, 2012). Moreover as a result of poor hygiene, acid-byproduct as a result of

sugar or carbohydrate product may aid in the formation of caries (Yadav & Prakash, 2016). A study demonstrated the role of poor oral hygiene status in children in association to salivary microbiome composition, found an increase of *Veillonella species* (Mashima et al., 2017).

5.2. Smoking

In a large American cohort study, J. Wu et al compared the salivary microbiome composition in both current smokers and non-smokers (J. Wu et al., 2016). They observed that the salivary microbiome of smokers reflected a decrease in the abundance of the phylum Proteobacteria, and in Capnocytophaga , Peptostreptococcus and Leptotrichia genera; while the genera Atopobium and Streptococcus were found to be elevated in smokers compared to non-smokers (J. Wu et al., 2016). Another study examined the oral microbiome of smokers and non-smoker in addition to the cytokine levels, where they found that smoking altered the cytokine levels and the salivary microbiome composition (Rodríguez-Rabassa et al., 2018).

6. Microbiome in disease: Microbial dysbiosis

A healthy individual maintains a unique balance between the microbiome, immune system for protection from invading bacteria (H.-J. Wu & Wu, 2012). Dysbiosis refers to the disturbance in the composition of microbiota at a particular site (Petersen & Round, 2014). Dysbiosis in gut microbiome for example, has been linked

to several diseases and conditions such as Crohn's disease and irritable bowel syndrome (Frank et al., 2007; Ni, Wu, Albenberg, & Tomov, 2017). Another example of dysbiosis is bacterial vaginosis caused by dysbiosis of the vaginal microbiota which usually triggered by *G. vaginalis* (Younes et al., 2018).

Dysbiosis of the oral microbiome has been implicated in various oral diseases such as periodontitis where it promotes pathogenic bacterial growth and enables the dissemination of the oral bacteria systemically (Li, Kolltveit, Tronstad, & Olsen, 2000).

Many studies were conducted in order to assess the dental and periodontal health in association with the microbiome composition: In a study that compared caries free individuals versus caries experienced individuals, they found that caries free individuals had a greater microbial diversity (Yasunaga et al., 2017). Similar finding was also observed by other studies that showed that healthy individuals have a greater microbial diversity and a greater abundance of *Neisseria*, *Haemophilus*, and *Fusobacterium*, in contrast to those who have dental caries where *Streptococcus* was the most abundant genus detected (Belstrøm et al., 2017).

Moreover, the oral microbiome dysbiosis have been associated with systemic diseases including obesity, diabetes, cancer, rheumatoid arthritis, Parkinson's disease and cardiovascular diseases among others (Acharya et al., 2017; Karpiński, 2019; Pereira et al., 2017; X. Zhang et al., 2015).

A recent study by Janem *et al* indicated that higher rates of type 2 diabetes (T2D) were observed in kids with improper oral health in comparison with lean and obese children without diabetes (Janem et al., 2017). The authors indicated that *Fretibacterium was only found in the diabetic group* but *Alloprevotella*, *Haemophilus,*

Lautropia and *Pseudomonas* were decreased in the diabetic group in comparison to healthy controls (Janem et al., 2017). A Thai study also support the previous in which the acid-tolerant bacteria, which are also associated with dental caries, are found more prevalent in T2D patients (Kampoo, Teanpaisan, Ledder, & McBain, 2014). Moreover, the salivary microbiome composition differs in people who are obese compared to the lean controls, the study also proposed that such microbial differences can help predict the susceptibility of people with obesity to T2D o(Y. Wu, Chi, Zhang, Chen, & Deng, 2018).

7. Tools used to assess the microbiome composition: 16S rRNA gene sequencing

The ribosomal RNA transcriptional machinery is an essential component of life(Wang & Qian, 2009). Thus, the gene that represents it has a conservative nature which led to an enormous opportunity for the exploration of many microbial communities by sequencing specific hypervariable regions of the 16S rRNA gene (Wang & Qian, 2009). Amplicons are generated using universal primers used to anneal with the conserved regions of the 16S rRNA gene sequences (Wang & Qian, 2009). Those hypervariable region sequences are highly polymorphic, therefore, allowing the classification of the bacterial taxa from phyla to species levels (Cox, Cookson, & Moffatt, 2013). The resulting sequencing data is then aligned to the 16S rRNA reference database to yield the operational taxonomy units (OTUs) of bacteria (Nguyen, Warnow, Pop, & White, 2016). Despite being efficient and relatively cheap, the 16S rRNA sequencing has its limitations, as it only covers microbiota of the bacterial kingdom, thus neglecting other important microbes belonging to Fungi, viruses and others, in addition to biases generated due to the variability of the 16S

gene copy number in different bacteria (Liu, Gibbons, Ghodsi, Treangen, & Pop, 2011).

The shotgun sequencing as the name infers, is the process of fragmenting the whole genome into short oligonucleotide bases followed by sequencing and assembly of the contigs (Jovel et al., 2016). This method allows a greater depth of sequencing data in comparison to the 16S sequencing approach discussed previously (Vincent et al., 2016). The shotgun sequencing provides information about the taxonomic profile, functional composition and gene abundance of the microbiota (Quince, Walker, Simpson, Loman, & Segata, 2017). It can also provide information up to species/strain level compared to 16S rRNA gene sequencing (Quince et al., 2017). The major disadvantages of the shotgun sequencing is the huge amount of data it generates, which requires a high-throughput sequencing instruments and thus a higher cost (Franzosa et al., 2015).

8. Aims of the project

The aim of this project is to profile the salivary microbiome of 100 saliva samples collected from Qatari participants and correlate the microbial composition with the clinical data in order to identify specific signatures associated with various pathological conditions with a focus on diabetes and obesity being the most prevalent non-communicable diseases and pathological conditions in Qatar.

CHAPTER 3: METHODOLOGY

The layout of the Methods used in this study in Figure 1.

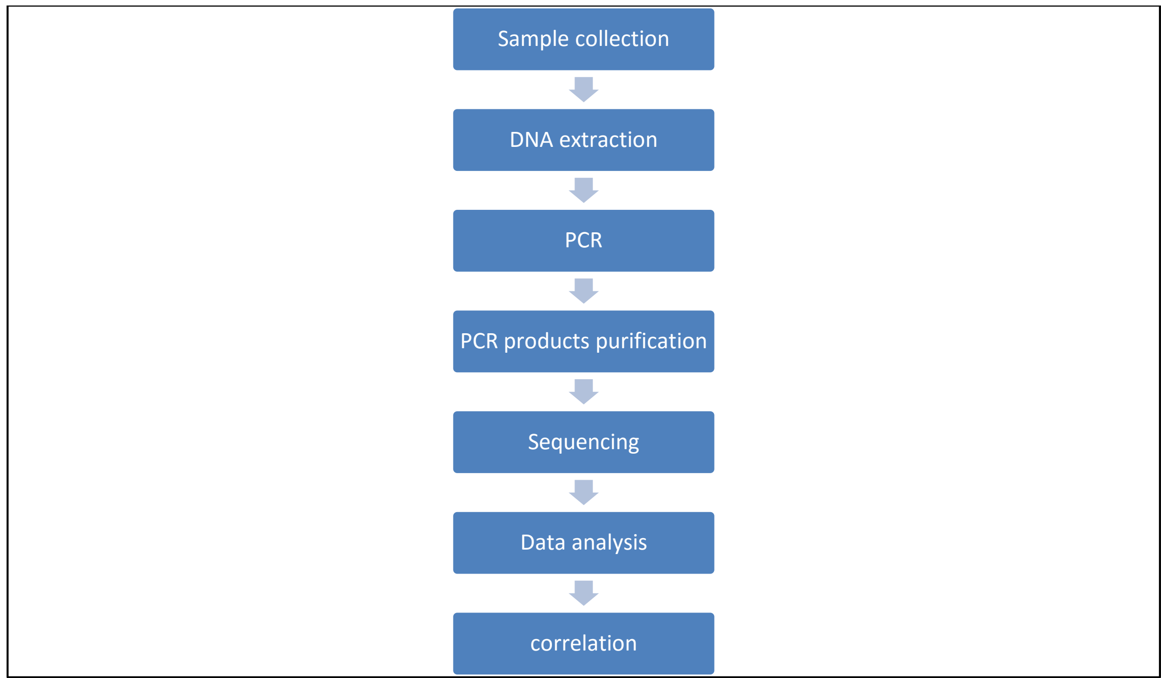


Figure 1:Summerized method workflow

1. Description of the study participants

A total of 100 participants were randomly selected from QBB, and included both females and males. Using the body mass index (BMI), samples were categorized in normal weight (BMI 18.5 to 24.9 kg/m^2), overweight: (BMI 25 to 29.9 kg/m^2) and obese (BMI is 30 kg/m^2 or more). Also, the study subjects were stratified into normal and diabetic according to their HbA1C levels (diabetic: HbA1c ≥6.5% (48 mmol/mol)).

2. Inclusion/ Exclusion criteria

The samples were collected by QBB from Qatari subjects participating in the Qatar Genome Project, who were eighteen years old and above. No exclusion criteria

were used in the subject's selection.

3. Sample collection

In this project, an agreement between QBB and Sidra Medicine was signed in order to collect de-identified salivary samples, phenotypic and clinical data from a total of 100 participants that were selected randomly.

Salivary samples were collected after obtaining the IRB (Institutional Review board) approval from QBB (E/2017/RES-ACC-0046/0003) and Sidra (IRB#1510001907). Samples were stored at – 80°C. Before processing, samples were kept at 4°C overnight. The first step before extraction was to incubate the salivary samples at 50°C for 2 hours.

4. DNA extraction from saliva using QIAsymphony

Automated extraction following Qiagen QIAsymphony protocol was carried on following the manufacturer's instructions. Briefly, the procedure is composed of four steps including lysis, binding, washing, and elution. As illustrated in Figure 1, the instrument utilizes binding of magnetic beads to the genomic DNA which is then later bound to a mechanical magnetic rod. One of the biggest advantages of the magnetic-particle technology is the isolation of quality DNA without impurities.

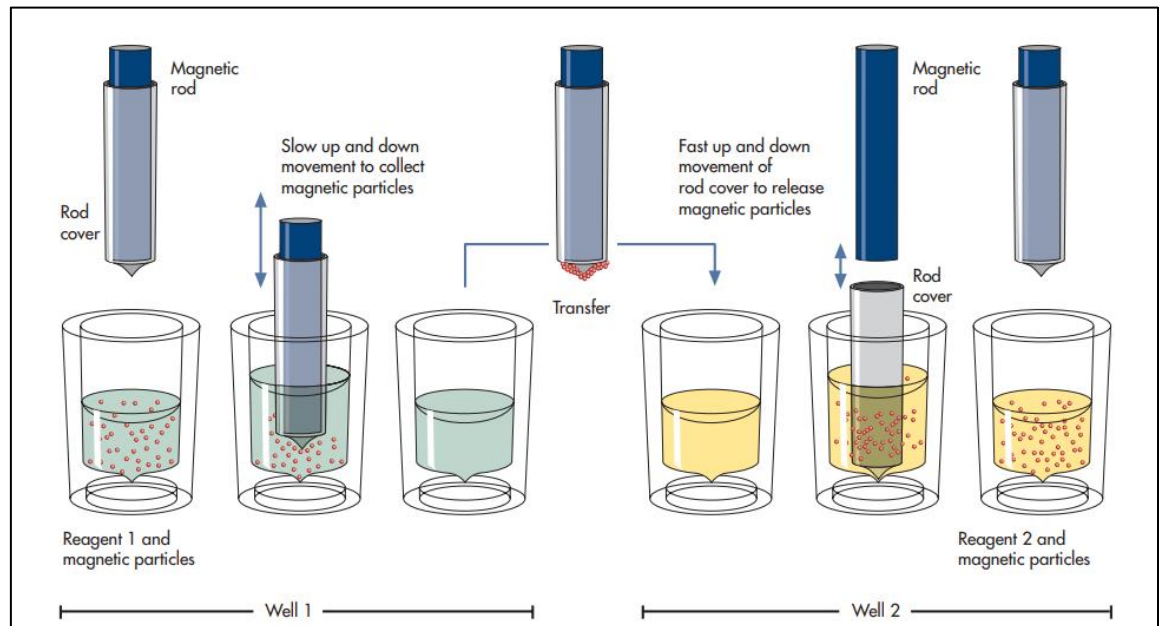


Figure 2: Schematic representation of the QIAsymphony SP protocol used for DNA extraction

Step 1: Lysis

Reagent 1 is added to the sample which causes lysis to the cell and release of the genomic DNA, then the magnetic beads are added to bind to genomic DNA.

Step 2: binding

The magnetic rod is inserted into the well holding the sample and this attracts the magnetic beads.

Step 3: washing

The magnetic rod is inserted into another well, and the magnetic beads are released.

Step 4: Elution

Another reagent is added to elute the genomic DNA from the magnetic beads. Then the magnetic rod will attract the remaining magnetic beads leaving the eluted volume

in the DNA stock solution.

5. DNA Quality control

The quality and quantity of the extracted DNA was checked using nanodrop (Thermo Scientific). The Nanodrop utilizes the concept of spectrophotometry that uses absorbance as a measurement of nucleic acid concentration.

6. Amplification of 16S rRNA gene fragment by polymerase chain reaction

Polymerase chain reaction (PCR) of the 16S rRNA gene was carried on for each sample. The designed 27-F forward primers are linked to a specific sequence that is used as a barcode to distinguish each sample for the sake of multiplexing in downstream procedures. The primers are also designed to align with the conserved region of the 16S rRNA gene (Figure 2). The reverse primer, which also align with a conserved region, was common across all sample's reactions. In this study, the targeted region as shown in Figure 2 is the V1 until V3 region which yield an amplicon size of 610 base pairs.

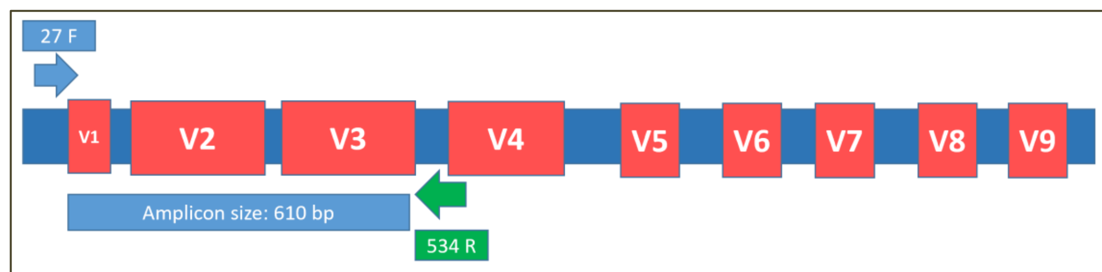


Figure 3: 16S rRNA gene: the red boxes represent the variable regions, whereas the blue areas in between the boxes represents the conserved region.

Table 1

The sequence of the forward and reverse primers to target V1V3 region of 16S rRNA gene

Primer	**Illumina adaptors**	**Barcode**	**Forward Primer Pad**	**Linker**	**Forward primer sequence**
27F1	AATGATACGGCGACCACCGAGATCTACACGCT	AGCCTTCGTCGC	TATGGTAATT	GT	AGAGTTTGATCMTGGCTCAG
534-R	CAAGCAGAAGACGGCATACGAGAT	-	AGTCAGCCAG	CC	ATTACCGCGGCTGCTGG

Table 2

PCR reaction mix

#	**Reagent**	**Volume**
1	PCR gradient, water	adjusted for each sample to reach a reaction volume of 50 µl
2	Phusion Master Mix catalogue #f531L	25 µl
3	27-F1 primer (F1-F96)	1 µl
4	534-R primer	1 µl
5	Samples DNA	Adjusted with water to reach 10 ng/µl
Total volume		50 µl

To set-up the PCR reaction each sample template concentration was adjusted with water to reach the optimum concentration of 10ng/50uL reaction. The forward primer that is designated for each sample was then added. Then followed by addition of the reverse primer and the master mix. The master mix is a readymade (cat# F531L) that includes all the essential PCR components such as dNTPs, DNA polymerase enzyme, Buffer, and MgCl2. . The final volume of the PCR reaction is 50 µl.

7. PCR thermal conditions and duration

The thermal cycling conditions used to amplify the 16S rRNA gene were set as following: 5 min of initial denaturation at 94 °C; 30 cycles of denaturation at 94 °C for 30 s, annealing at 62 °C for 30 s, and elongation at 72 °C for 30. Finally, after the 30 cycles at 72 °C the samples were incubated for additional 10 minutes.

8. Gel electrophoresis

The quality of the amplified PCR products was assessed using 2% agarose gel electrophoresis. 100 bp DNA ladder (Thermo Scientific) was used to check the size of the product.

Once the libraries of each sample were constructed, the samples were pooled in one tube. Since each sample has a unique sequence barcode, the sequencing analyzer can read and demultiplex each sample in a separate FASTAQ file. The optimum volume of each sample to be pooled is 5uL (Intense band), 10uL (Moderate band), 15uL (Faint band).

9. 16S library magnetic bead purification

This step (Figure 3) employs the use of magnetic beads that is combined with a carboxyl molecule that allows the separation of negatively charged DNA. This technique is termed solid phase reversible immobilization or SPRI (Hawkins, O'Connor-Morin, Roy, & Santillan, 1994). There are many advantages for this technique: easy to perform as it does not require centrifugation nor filtration and that it does favor the isolation of PCR products over excess primers or genomics DNA.

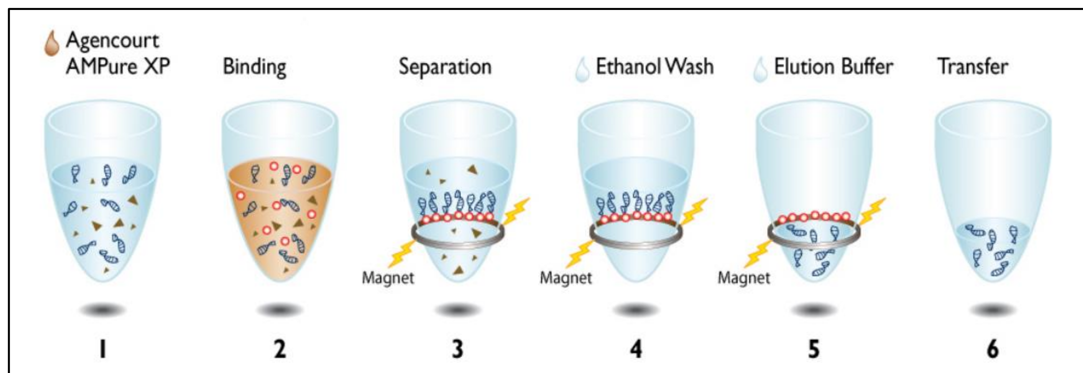


Figure 4: AMPure XP magnetic beads purification of the PCR product

10. AmpPure magnetic bead purification

We followed the procedure below:

1. Add 100 µl of PCR pool to 112 µl of Ampure beads.
2. Mix the solution thoroughly by pipetting up and down for 10 times.
3. Incubate the solution for 2 minutes at room temperature on the magnetic stand.
4. Discard the supernatant.
5. Add 200 µl of 80% freshly prepared ethanol.

6. Incubate the solution for 30 seconds at room temperature on the magnetic stand.
7. Discard the supernatant.
8. Repeat step 5-7 one more time.
9. Allow to air dry for 10 minutes.
10. Elute with 55 µl 10 mM Tris pH 8.5.

11. Assessment of the PCR product quality using Agilent bioanalyzer and Qubit fluorometer

11.1. Agilent High sensitivity kit

The High sensitivity DNA kit was used to assess the quality of the PCR constructed libraries. The Agilent bioanalyzer uses the concept of small capillary electrophoresis. The samples were run against a ladder and a marker. To start up the run, the following procedure was performed:

1. The High Sensitivity DNA dye concentrate (blue) and High Sensitivity DNA gel matrix (red) has to reach room temperature for 30 minutes before proceeding.
2. The blue vial with High Sensitivity DNA dye concentrate has to be vortexed for 10 seconds followed by centrifugation or spin down.
3. Pipette 15 µl of the blue dye concentrate (blue) into a High Sensitivity DNA gel matrix vial (red). Store the dye concentrate at 4 °C in the dark again.
4. Take the entire gel- dye mix to the top part of the spin filter.
5. Centrifuge the spin filter for 10 minutes at 6000 rpm at room temperature.
6. Discard the filter and label the tube with the preparation date.

7. Loading the Gel-Dye mix: place a new High Sensitivity DNA chip on priming station. Pipette 9.0 µl of the gel-mix at the well-labled.
8. Loading the marker: dispense 5 µL of green- High Sensitivity DNA marker (green) into the well labeled with the ladder figure and also into all sample wells.
9. Dispense 1 µl of the High Sensitivity DNA ladder vial (yellow) in the well-labeled with the ladder figure. In each of the 11 sample wells, dispense 1 µl of sample or 1 µL of marker for the unused wells.
10. Start the Chip Run.

11.2. Qubit Fluorometer

The Qubit™ 4 Fluorometer is a benchtop fluorometer that can be used for the quantitation of DNA. The measurement of the fluorescence in samples reflect the concentration of DNA. In this study, Qubit™ 4 Fluorometer was used to measure the concentration of the constructed libraries before proceeding to sequencing in order to ensure the optimal concentration needed for sequencing.

12. Sequencing

In order to sequence the constructed libraries, the Illumina Miseq platform was used in this study. The principle that Illumina follows is sequencing by synthesis and reversible termination. This principle is based on the addition of fluorescently labeled dNTPs that are reversible terminators. All the added terminator bases are added at the same time; thus, less competition of incorporation is seen. According to Illumina MiSeq protocol, the Illumina sequencing workflow is composed mainly of four steps:

samples preparation, cluster generation, sequencing and data analysis.

Sample preparation: the libraries are customized as mentioned in the PCR step above. Each forward primer is designed with a unique barcode and an illumina compatible adaptor. In addition, the reverse strands are also linked with illumina compatible adaptors.

Cluster generation: the flow cell contains oligo adaptors that are complementary to the adaptors of the library fragments. The pooled libraries are loaded to the flow cell. When hybridization of the forward strand with the flow cell adaptor occurs, the reverse strand is then synthesized. Next, when the forward strand is washed away, the reverse strand folds hybridizing to the neighboring adaptor. Then, a bridge amplification by the DNA polymerase follows. This process forms the clusters that will be later sequenced.

Sequencing: The four fluorescently labeled nucleotide are added at the same time in which they compete to bind to the template attached to the flow cell. The added nucleotide are called reversible terminators since they are blocked for further reaction at the 3-OH end (Ambardar, Gupta, Trakroo, Lal, & Vakhlu, 2016). The instrument captures the fluorescence of the designated base and records the read. After the cleavage of the terminating moiety and the fluorophore molecule, another cycle begins again. This cycle is repeated 300 times for the read 1 and the same number of cycles for read 2. Figure 4 illustrates the sequencing workflow step by step.

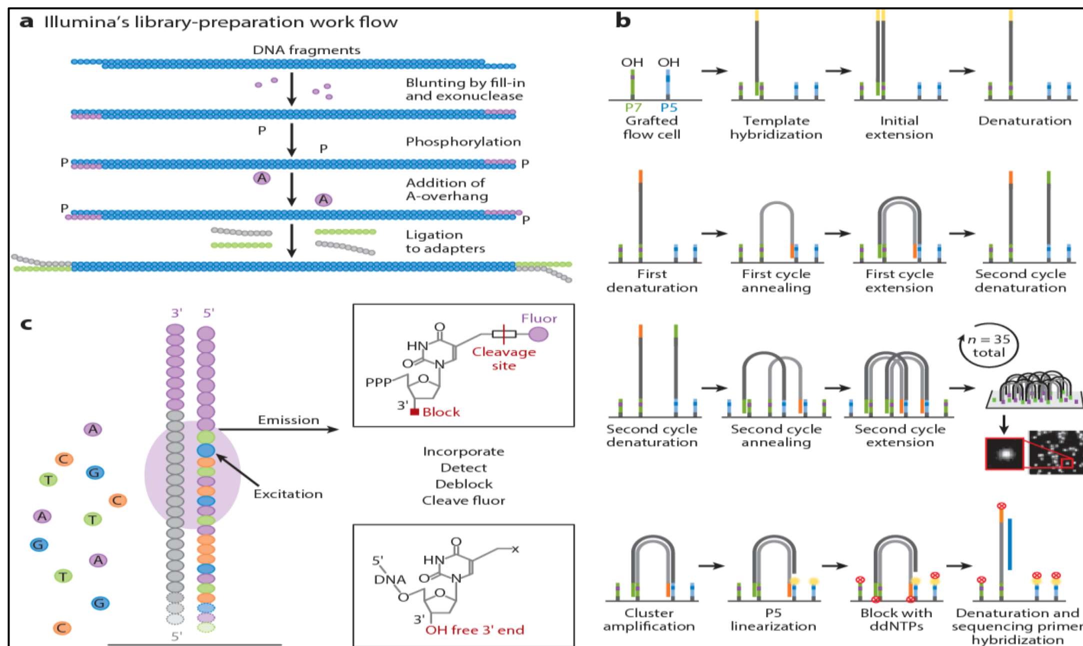


Figure 5: Sequencing by synthesis, Source:(Mardis, 2013)

13. Sequencing Procedure

Before loading the purified pooled libraries, dilution and denaturation following the protocol provided by Illumina (document # 15052877) should be followed. In the case of using customized primers, the primers also should be added to the reagent cartridge.

Sequencing was performed using MiSeq Reagent Kit v3 600 cycle kit which consists of a reagent cartridge, flow cell and PR2 reagent. The MiSeq System Guide for Local Run Manager (15027617 v04) illustrate the full procedure in setting up the instrument and loading the reagent cartridge. The procedure was performed according to the manufacturer's instruction and recommendations.

14. Data Analysis

Sequenced data were demultiplexed using MiSeq Control Software (MCS) in MiSeq (Illumina) sequencer. Generated demultiplexed data were revised for quality control using FastQC (Andrews S, 2012). Forward and reverse end sequences of respective samples were merged through PEAR tool (J. Zhang, Kobert, Flouri, & Stamatakis, 2014)) and sequence reads of quality score < 30 were discarded. All merged reads were trimmed to 160bp>Reads<500bp using Trimmomatic tool (Bolger, Lohse, & Usadel, 2014). Trimmed FASTQ files were converted into FASTA files. Demultiplexed FASTA files were analyzed using QIIME software v1.9.0 pipeline (Caporaso et al., 2010). QIIME is a an open-source software which analyzes raw data as an input to generate output data in the desired format. such as OTUs (Kuczynski et al., 2011) by aligning against the Human Oral Microbiome Database or HOMD (Escapa et al., 2018). Alpha diversity was calculated using phyloseq package in R platform (McMurdie & Holmes, 2013). Beta diversity was calculated using Unifrac distance method for all microbial communities of the sample through principal coordinates analysis (Chang, Luan, & Sun, 2011). Association between clinical metadata and salivary microbiome data were done using Hierarchical All-against-All significance testing or abbreviated HAllA (Gholamali Rahnavard). In order to assess the microbial changes that were statistically significant, LefSe analysis (linear discriminant analysis of effect size) was used (Segata et al., 2011). The predictive functional profiling of microbial communities in association with BMI categories and diabetic versus non-diabetic categories was analyzed using PiCrust (Langille et al., 2013).

CHAPTER 4: RESULTS

1. **Description of the study participants**

Table 3

Demographic data of Studied Qatari Population

	FEMALE	MALE	Total
AGE	40.63±11.44	38.76±10.04	
GENDER	48	52	100
BMI			
Normal	11 (22.92%)	11 (21.15%)	22
Overweight	14 (29.17%)	18 (34.62%)	32
Obese	23 (47.91%)	23 (44.23%)	46
Diabetes			
Non-Diabetic	34 (70.8%)	37 (71.15%)	71
Diabetic	14 (29.2%)	15 (28.85%)	29

Note-Age - Average ± Standard deviation

2. Microbial DNA Quality and Quantity

The quality and quantity of DNA were measured using spectrophotometery. The optimal amount of DNA needed for PCR is 10 ng . The absorbance ratio of A260/A280 of 1.7–2.0 is the optimum in terms of DNA quality.

3. PCR product visualization using agarose gel electrophoresis

PCR product from each sample was checked against a ladder using 2% agarose gel electrophoresis. Figure 5 is an example of amplified PCR libraries used for sequencing.

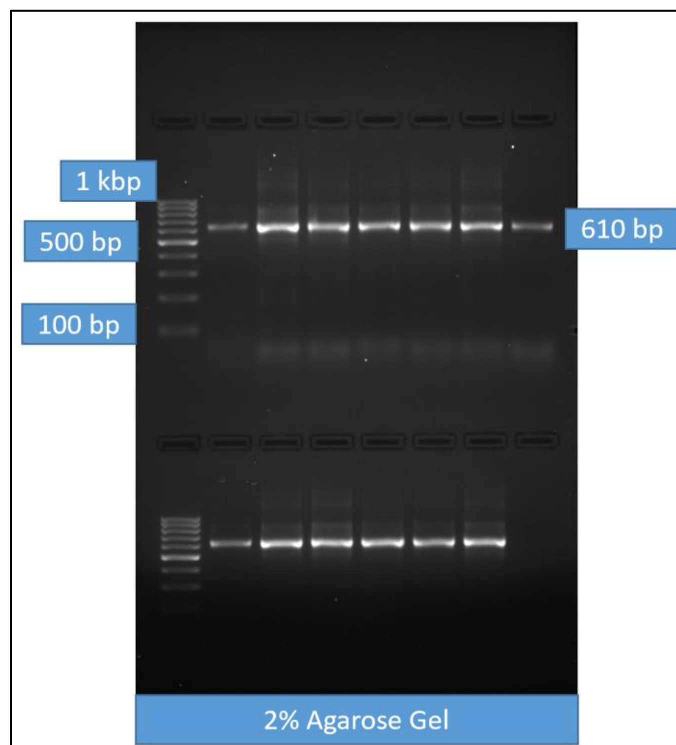


Figure 6: Gel electrophoresis for PCR products

4. Taxa summary

Characterization of the relative abundance of salivary microbiome at the phylum level in saliva samples collected from normal, overweight and obese Qatari participants revealed that *Bacteroidetes, Firmicutes and Proteobacteria* were the most common phyla detected (Figure 6). The results showed that *Firmicutes* was higher in normal subjects (35.4%) compared to overweight (25%) and obese (29.3%) participants.

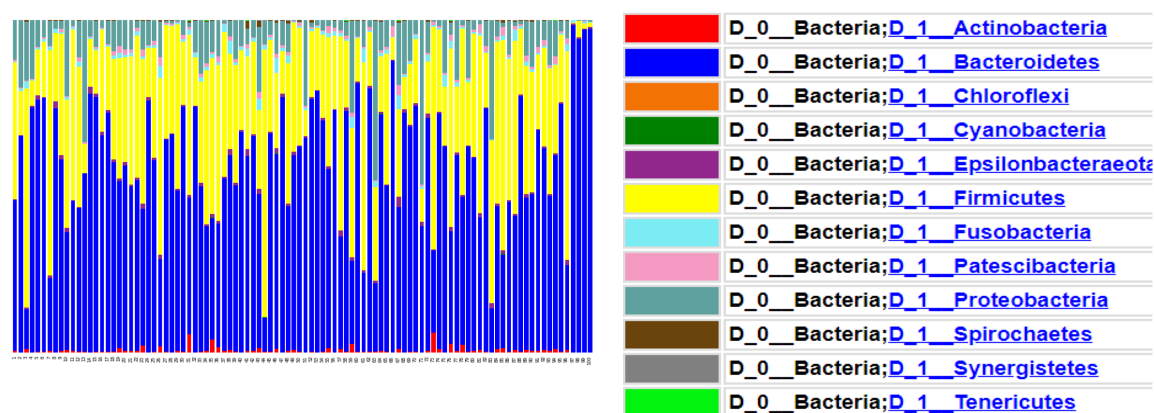
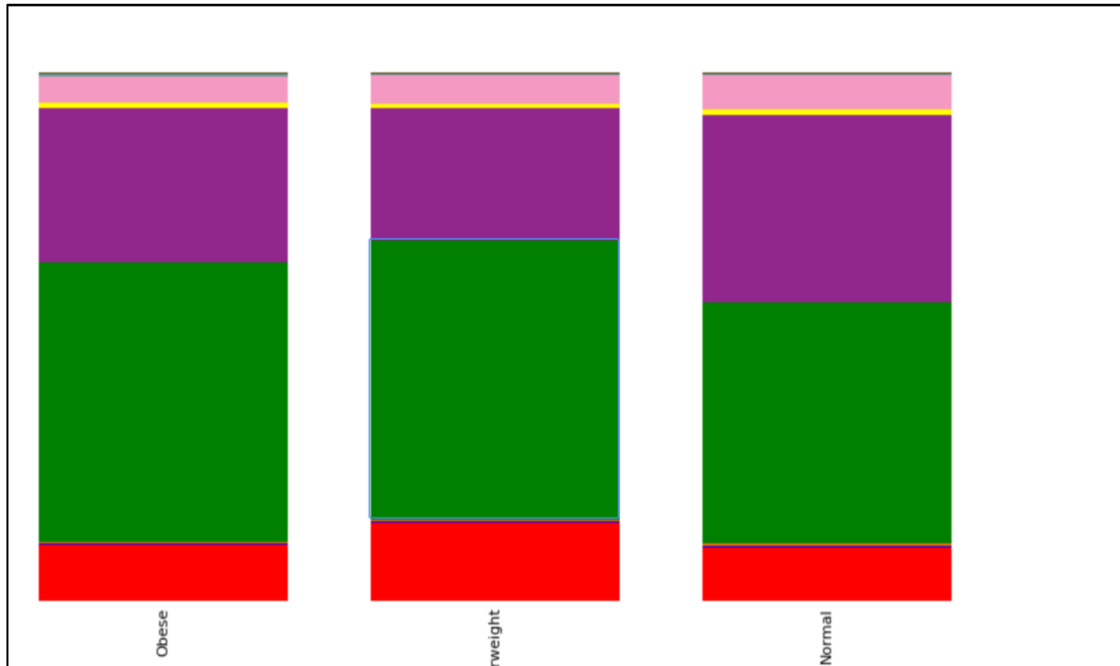


Figure 7: Relative abundance of bacterial phyla among salivary samples

On the other hand, *Bacteroidetes* was higher in overweight (52.8%) and obese subjects (52.8%) when compared to normal weight participants (45.8%) (Figure 7). Abundance of *Proteobacteria* was slightly higher in normal weight (6.6%) than overweight (5.3%) and obese (5.2%) participants (Figure 7).

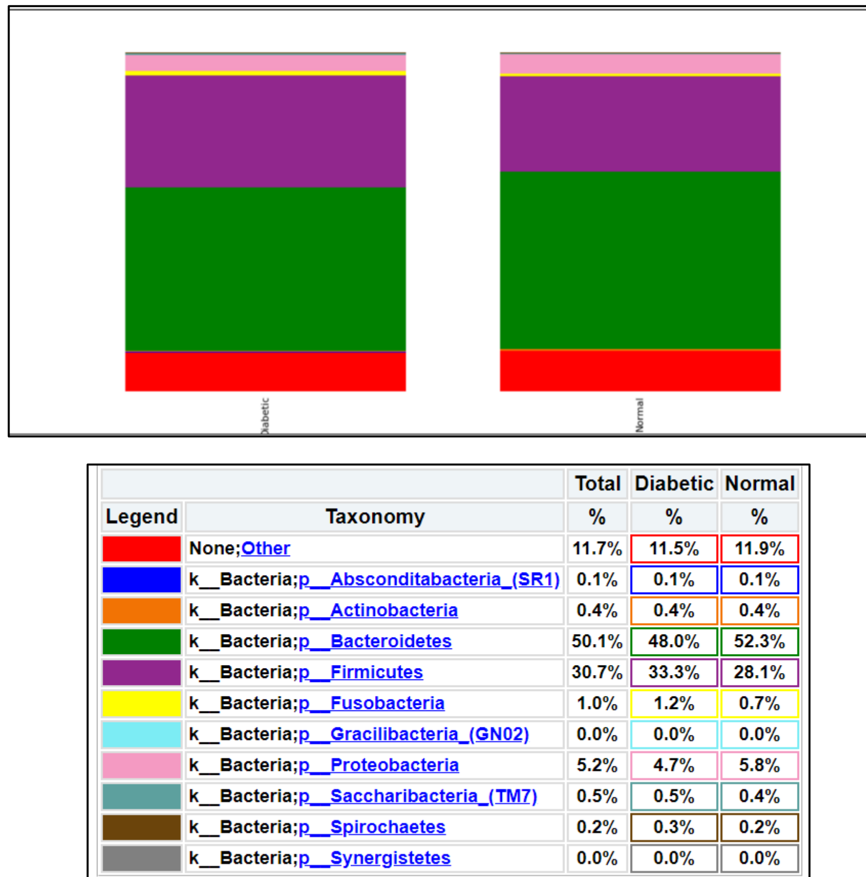
The relative abundance of salivary microbiome at phylum level between normal (non-diabetic) and diabetic group revealed that *Firmicutes* was higher in the diabetic group (33.3%) compared to the normal (non-diabetic) group (28.1%). On the other hand, *Bacteroidetes* was elevated in the normal group (52.3%) than diabetic group

(48%) (Figure 8). *Proteobacteria* was higher in the normal group (5.8%) compared to the diabetic group (4.7%). *Fusobacteria, Saccharibacteria*, *Abscondibacteria* and *Actinobacteria* were the least abundant phyla (Figure 8). Although there were many differences in the relative abundance at both the phyla level and genus level, statistical analyses using Kruskal Wallis test did not detect any significant differences (p value>0.05).



		Total	Obese	Overweight	Normal
Legend	**Taxonomy**	**%**	**%**	**%**	**%**
	None;Other	11.9%	10.7%	14.9%	10.2%
	k__Bacteria;p__Absconditabacteria_(SR1)	0.1%	0.1%	0.2%	0.1%
	k__Bacteria;p__Actinobacteria	0.4%	0.3%	0.4%	0.6%
	k__Bacteria;p__Bacteroidetes	50.4%	52.8%	52.8%	45.8%
	k__Bacteria;p__Firmicutes	29.9%	29.3%	25.0%	35.4%
	k__Bacteria;p__Fusobacteria	0.9%	0.8%	0.9%	0.9%
	k__Bacteria;p__Gracilibacteria_(GN02)	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Proteobacteria	5.7%	5.2%	5.3%	6.6%
	k__Bacteria;p__Saccharibacteria_(TM7)	0.4%	0.5%	0.5%	0.4%
	k__Bacteria;p__Spirochaetes	0.2%	0.2%	0.2%	0.2%
	k__Bacteria;p__Synergistetes	0.0%	0.0%	0.0%	0.0%

Figure 8: Comparison of the microbial taxa at the phylum level in normal weight, overweight and obese



		Total	Diabetic	Normal
Legend	Taxonomy	%	%	%
	None;Other	11.7%	11.5%	11.9%
	k__Bacteria;p__Absconditabacteria_(SR1)	0.1%	0.1%	0.1%
	k__Bacteria;p__Actinobacteria	0.4%	0.4%	0.4%
	k__Bacteria;p__Bacteroidetes	50.1%	48.0%	52.3%
	k__Bacteria;p__Firmicutes	30.7%	33.3%	28.1%
	k__Bacteria;p__Fusobacteria	1.0%	1.2%	0.7%
	k__Bacteria;p__Gracilibacteria_(GN02)	0.0%	0.0%	0.0%
	k__Bacteria;p__Proteobacteria	5.2%	4.7%	5.8%
	k__Bacteria;p__Saccharibacteria_(TM7)	0.5%	0.5%	0.4%
	k__Bacteria;p__Spirochaetes	0.2%	0.3%	0.2%
	k__Bacteria;p__Synergistetes	0.0%	0.0%	0.0%

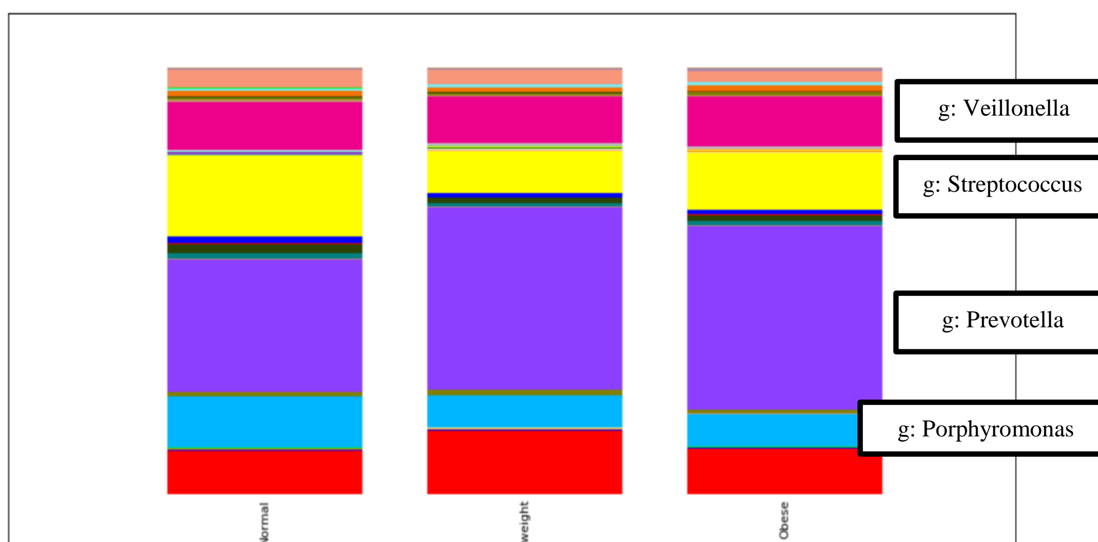
Figure 9: Taxa summary of the phylum level for Diabetic Versus Normal

5. Relative abundance of the salivary microbiome at the Genus level

Characterization of the relative abundance of the salivary microbiome at genus level in saliva samples collected from normal, overweight and obese Qatari participants revealed that Prevotella sp., Streptococcus sp., Veillonella sp., and Porphyromonas were the most abundant genera (Figure 9). The results showed that Streptococcus sp., was higher in the normal group (19%) compared to the overweight (9.9%) and obese (13.2%) participants. Likewise, Porphyromonas sp., was higher in normal (11.8%) compared to the overweight (7.5%) and obese participants (7.5%). On the other hand, Prevotella sp., was higher in overweight (42.8%) and obese

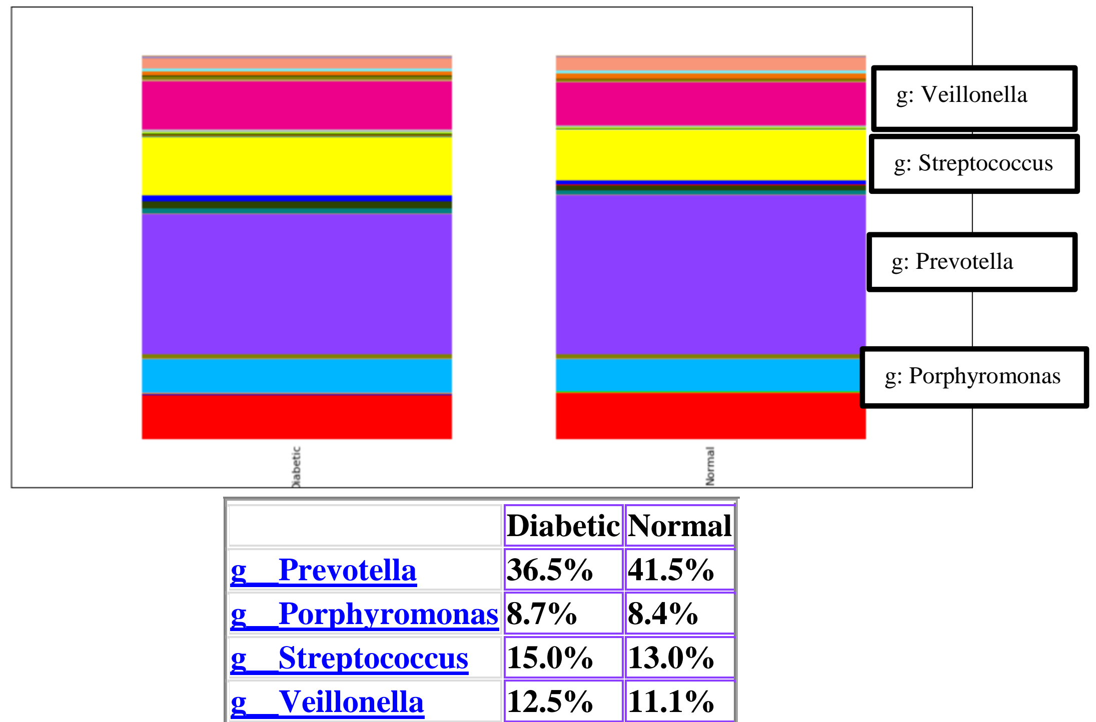
(42.8%) compared to the normal weight participants (30.9%) (Figure 9). Abundance of *Veillonella sp*. was slightly higher in obese (11.8%) compared to overweight (10.9%) and normal weight participants (11.2%) (Figure 9).

The relative abundance of salivary microbiome at genus level between normal and diabetic group revealed that *Streptococcus sp* was slightly higher in the diabetic group (15%) compared to the normal group (13%). On the other hand, *Prevotella sp*., was elevated in the normal group (41.5%) than diabetic group (36.5%) (Figure 10). Although there were many differences in the relative abundance at both the phyla level and genus level, statistical analyses using Kruskal Wallis test did not detect any significant differences (p value>0.05).



	Normal	Overweight	obese
g__Porphyromonas	11.8%	7.5%	7.5%
g__Prevotella	30.9%	42.8%	42.8%
g__Streptococcus	19.0%	9.9%	13.2%
g__Veillonella	11.2%	10.9%	11.8%

Figure 10: Taxa summary of the genus level for BMI categories with most abundant genera



	Diabetic	Normal
g__Prevotella	36.5%	41.5%
g__Porphyromonas	8.7%	8.4%
g__Streptococcus	15.0%	13.0%
g__Veillonella	12.5%	11.1%

Figure 11: Taxa summary of the genus level for Diabetic versus Normal with most abundant genera

6. Comparative relative abundance between various categories

The top common 25 bacterial genera were compared for normal weight participants, overweight and obese (figure 11). It can be noticed that the genus *Prevotella* is less abundant in normal weight individuals as compared to the overweight and obese groups similar to *Campylobacter, Leptotrichia, Saccharibacteria, Megashaera*. On the other hand, *Granulicatella, Gemella, Capnocytophaga, Actinomyces, Bergeyella, Fusobacterium* are the higher genera in the normal weight group when compared to the overweight and obese groups. Figure 12 shows the differential representation of the top 25 bacterial genera in the normal (non diabetic) group and diabetic groups.

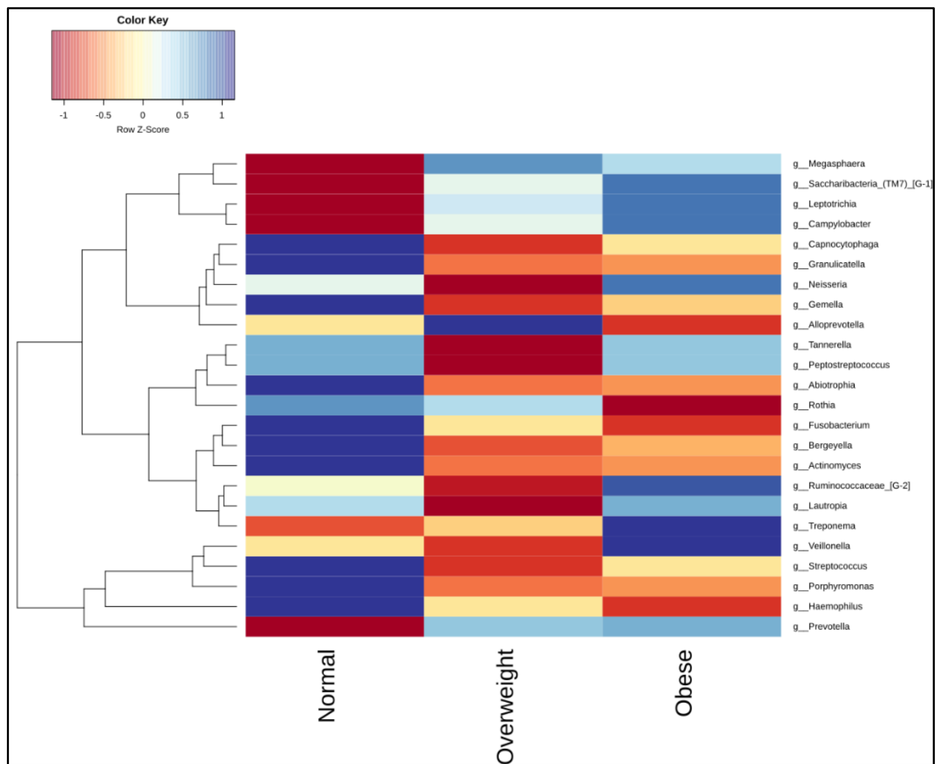


Figure 12: Hierarchical clustering heatmap of top 25 genera between Normal, Overweight and Obese group. Heat map generated using the relative abundance (percent) of top 25 abundant bacterial genera. The heat map was generated using the gplots package in R by clustering of diabetic and normal group relative abundance of bacterial genera. The heat map scale displays the row Z score (Z score = [actual relative abundances of a genus in relevant group − mean relative abundance of the same genus in the relevant group]/standard deviation)

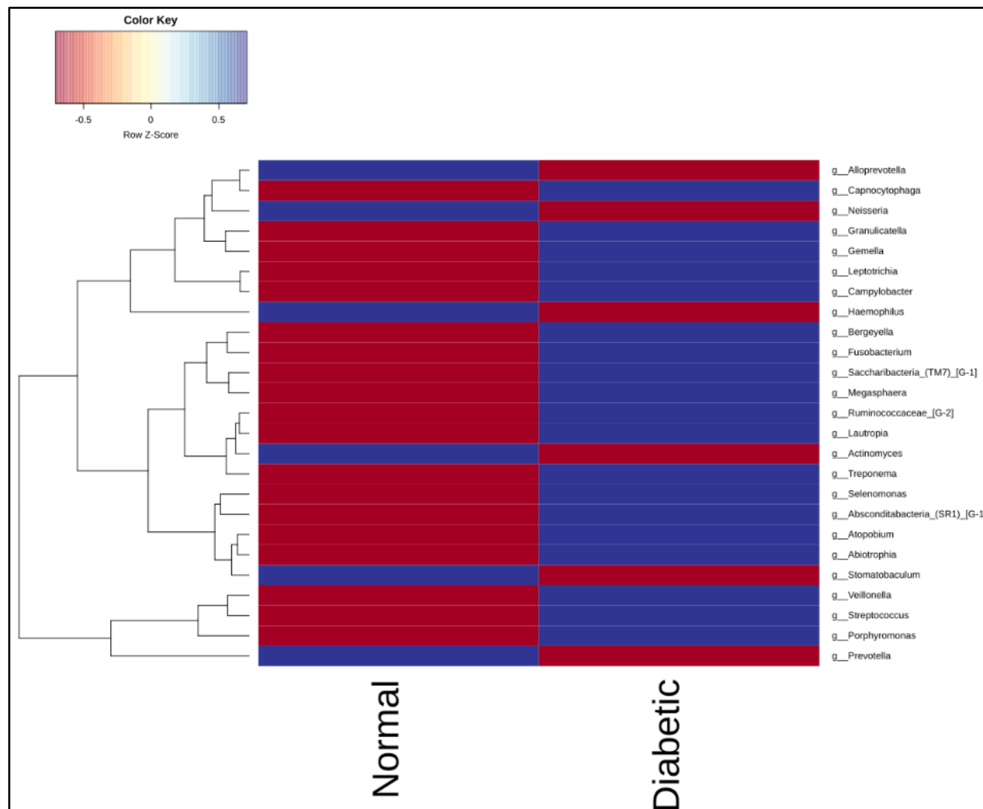


Figure 13: Hierarchical clustering heatmap of top 25 genera between diabetic and normal group. Heat map generated using the relative abundance (percent) of top 25 abundant bacterial genera. The heat map was generated using the gplots package in R by clustering of diabetic and normal group relative abundance of bacterial genera. The heat map scale displays the row Z score (Z score = [actual relative abundances of a genus in relevant group − mean relative abundance of the same genus in the relevant group]/standard deviation)

7. Interindividual and intraindividual variability: Alpha and beta-diversity

Microbiome measurements can be used as indicators of health and disease especially when performing a case-control study (J. K. Goodrich et al., 2014). Also, it can be an indicator of prognosis by conduction of longitudinal studies (Fukuyama et al., 2017) An alpha diversity measure refers to the mean species diversity or species richness in a given ecosystem or sample while a beta diversity score identifies whether two samples or communities are similar or not (Wagner et al., 2018). Interindividual

variability was assessed using measurement of the alpha diversity for each sample. Chao1 analysis was used to compare the alpha diversity between normal and diabetic group. (Figure 11) shows that the diabetic samples are less diverse than the normal. Also, when the BMI categories are compared (Figure 12) the normal weight has a greater diversity than the obese and overweight groups.

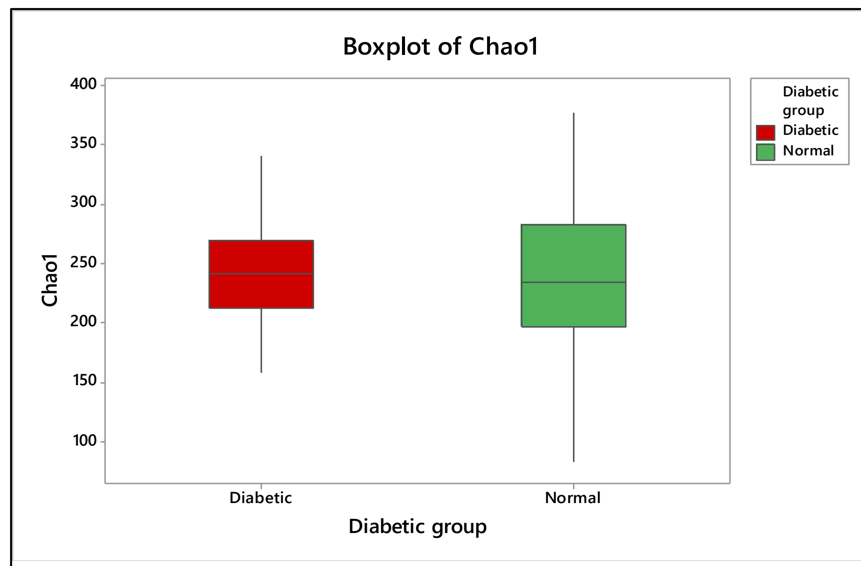


Figure 14: Alpha diversity of diabetic and normal (non-diabetic) subjects

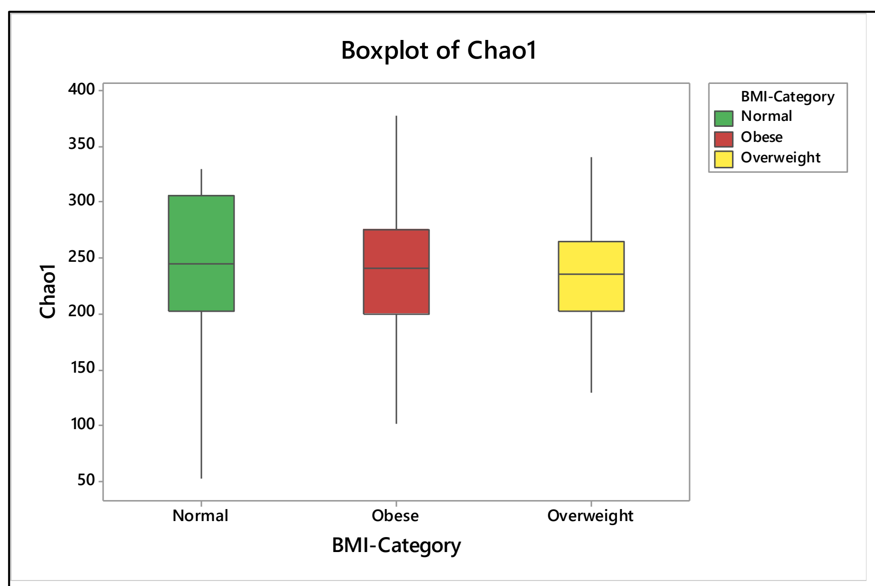


Figure 15: Chao1 analysis for alpha diversity of BMI categories

Beta diversity were also calculated, and the results were visualized using Principle of Coordinate Analysis (PCoA) by applying weighted UniFrac in which the similarities among samples can be illustrated. When Beta diversity was compared among the BMI categories, it did not show any form clustering according to the sample category, Figure 15. The same was noticed when diabetic group was compared to normal, Figure 16.

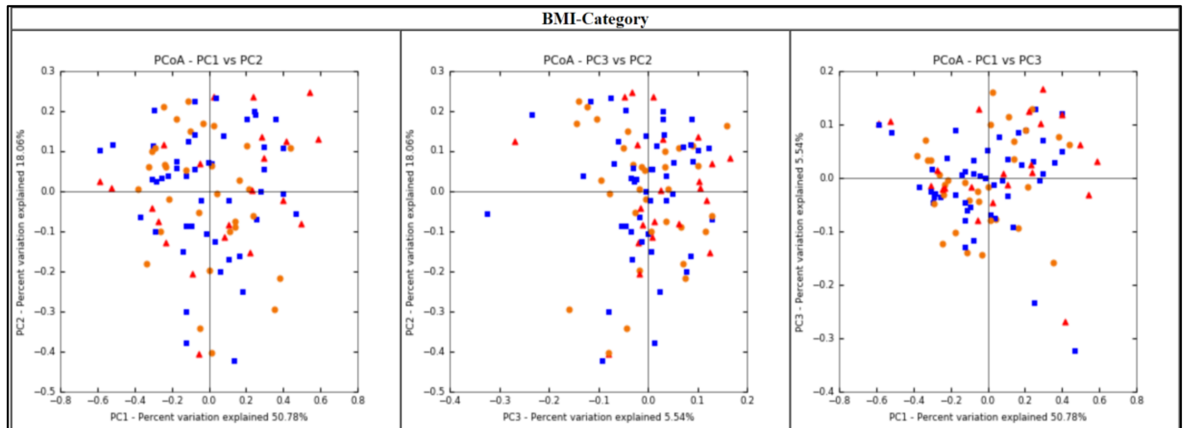


Figure 16: Principle of Coordinate Analysis (PCoA) by applying weighted UniFrac of BMI category; red triangle: normal, blue: obese, circle: overweight

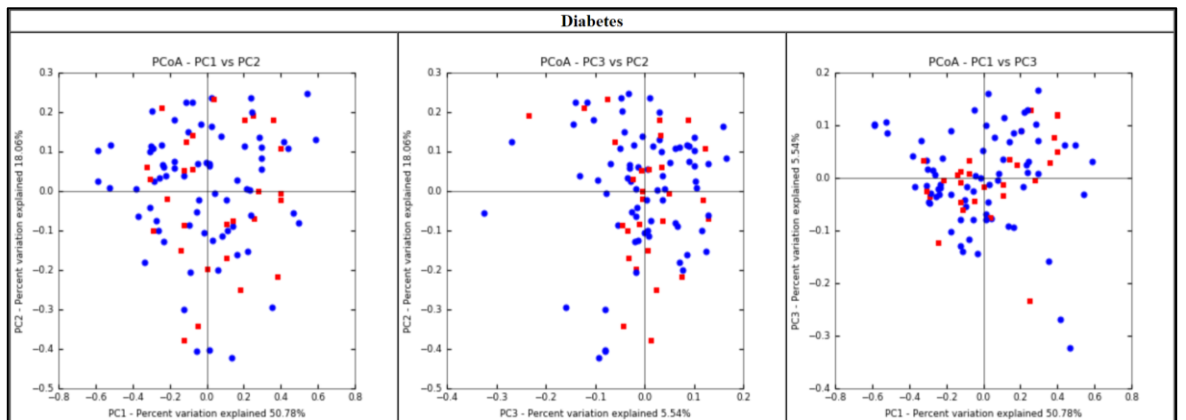


Figure 17: Principle of Coordinate Analysis (PCoA) by applying weighted UniFrac of diabetic and normal; blue: normal, red: diabetic

8. Correlation between clinical variables and microbial data: HALLA

Hierarchical All-against-All significance testing or abbreviated HAllA considers two data sets and tries to find positive and negative correlation between them. In this case, we considered the clinical parameters of all and each group in

relation to their microbial composition. When all the participants' data were tested, the genus *Catonella* and the genus *Peptostreptococcus* were positively correlated with parameters related to hypertension (Figure 17).

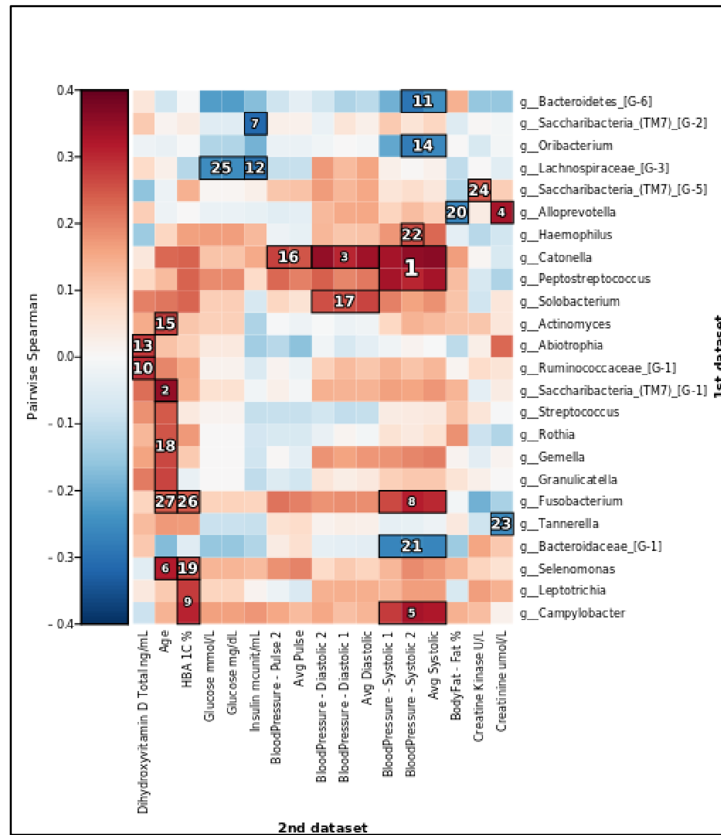


Figure 18: Associations between clinical parameters and microbial composition (genus level) for all participants

Figure 18 shows that the genera *Rothia*, *Lautropia*, *Corynebacterium*, *Pseudopropionibacterium* are positively correlated with parameters related to glucose such as glucose level and HbA1C.

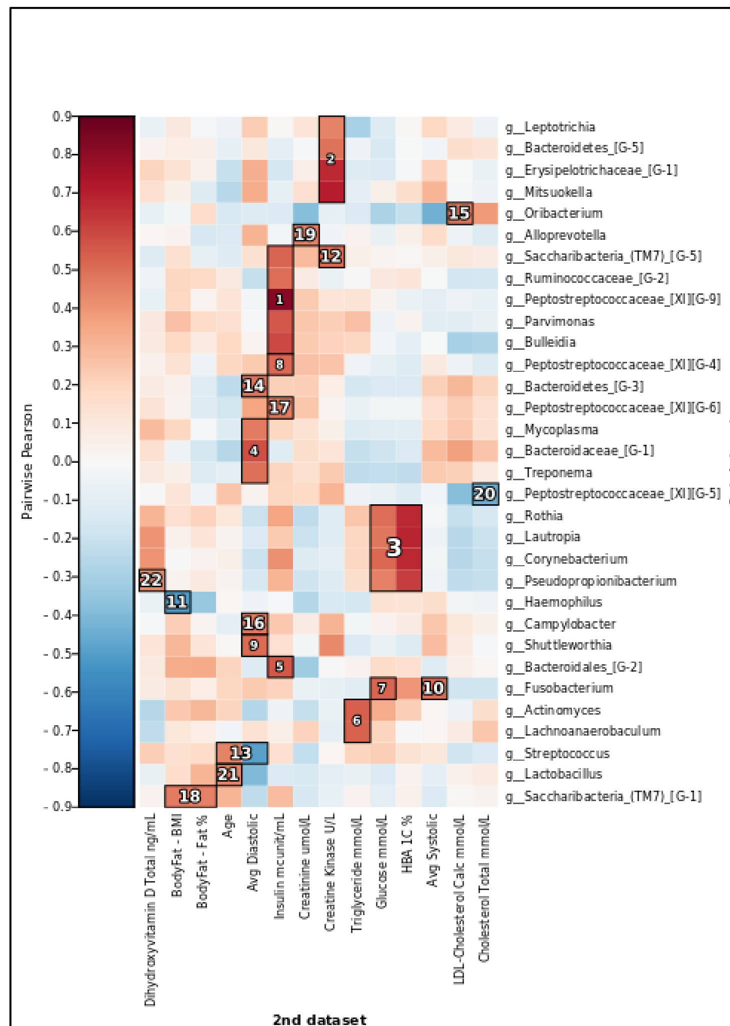


Figure 19: Associations between clinical parameters and microbial composition (genus level) for the Diabetic group

The association between participants group according to their BMI and microbial composition were tested. We found that *Pseudopropionibacterium* is positively correlated with glucose level in normal weight participants (Figure 19)., while the genus *Peptostreptococcacea* is positively correlated with the insulin level in obese group (Figure 20). The genera *Capnocytophaga, Fusobacterium, Butyrivibrio, Pseudopropionibacterium,* and *Lautropia* were positively correlated

with parameters related to blood glucose in overweight group in Figure 21.

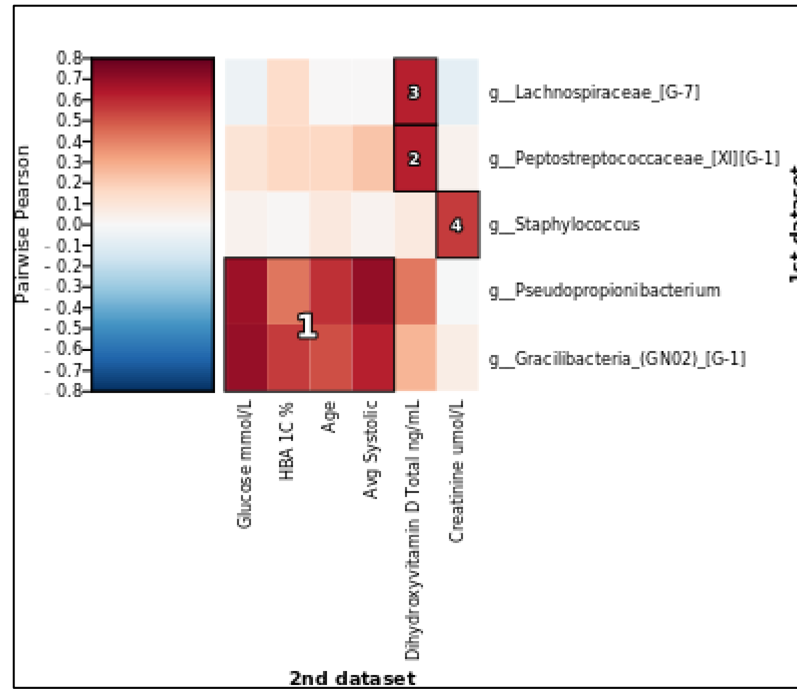


Figure 20: Associations between clinical parameters and microbial composition (genus level) for the normal weight group

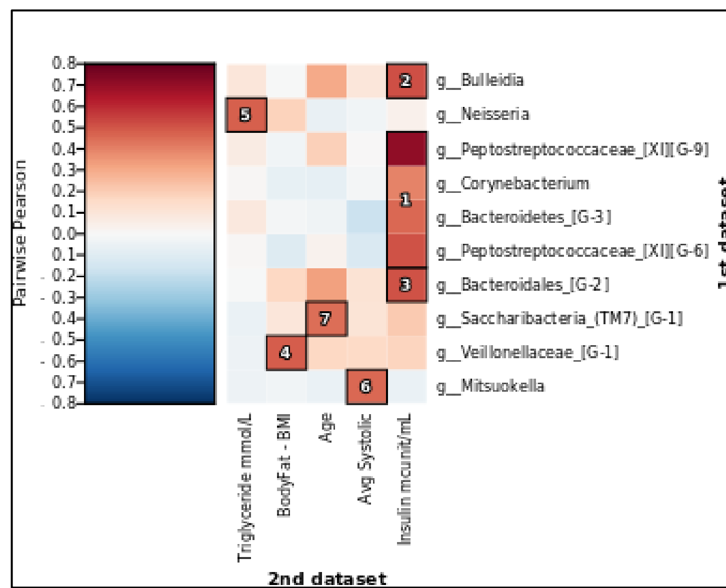


Figure 21: Associations between clinical parameters and microbial composition (genus level) for the obese group

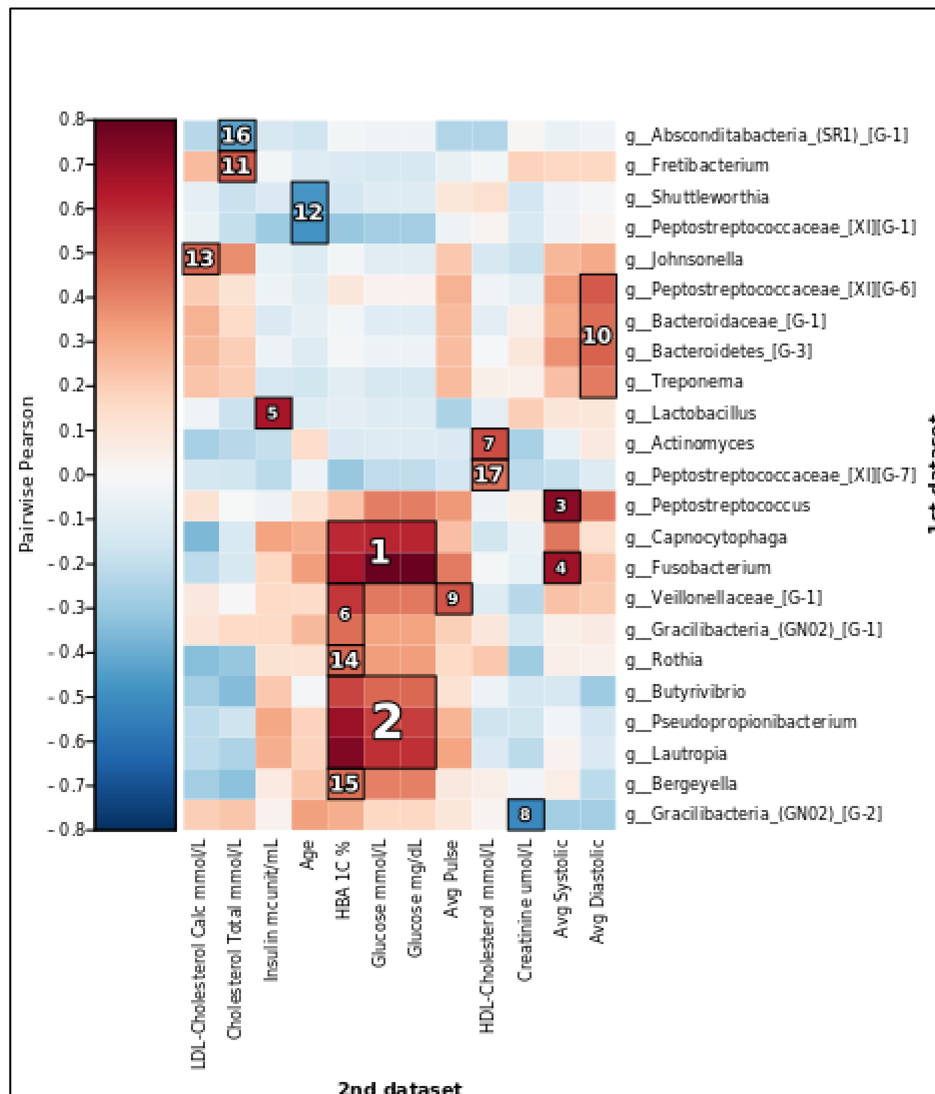


Figure 22: Associations between clinical parameters and microbial composition (genus Level) for Overweight group

9. Significant differences among clinical categories using LDA (Linear discriminant analysis) Effect Size (LEfSe)

In order to assess what microbial changes are statistically significant, we ran the LefSe analysis. The microbial composition of the BMI groups which include normal,

overweight and obese where checked. When comparing the three groups, we identified the genus *Catonella* as the most significantly associated with obesity (Figure 22). While when the normal weight group was compared to the overweight group, *Mogibacterium* and *Solobacterium* were significantly correlated with the BMI increase in addition to *Catonella* (Figure 22). Moreover, when the microbial composition of diabetic versus normal was compared, there was no significant genera associated with either conditions.

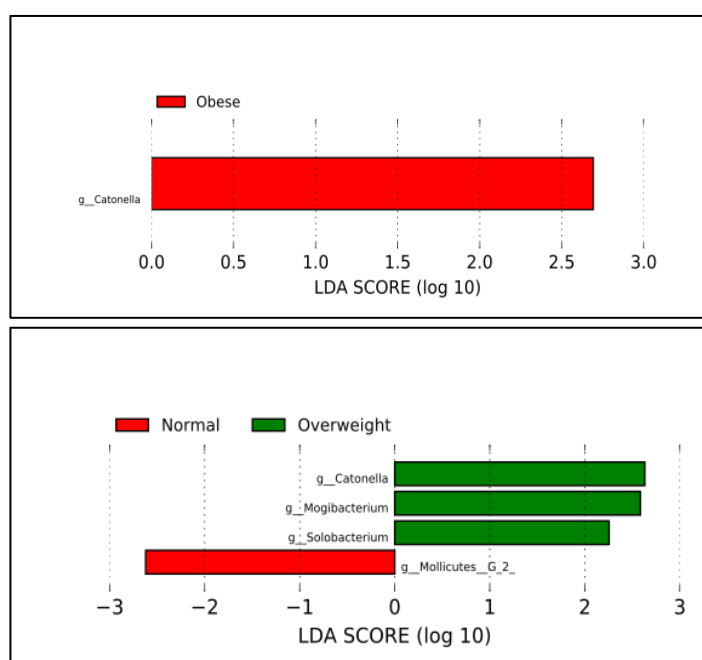


Figure 23: LefSe analysis of BMI groups and their microbial composition

10. Predictive functional profiling of microbial communities using PiCrust

Predictive functional microbial profiles of each taxa were studied against each subject category. We found that the N-glycan biosynthesis is highly associated with

diabetes and obesity (Figure 23 and 24). There was also not any other significant predictive functional profile associated with obesity specifically.

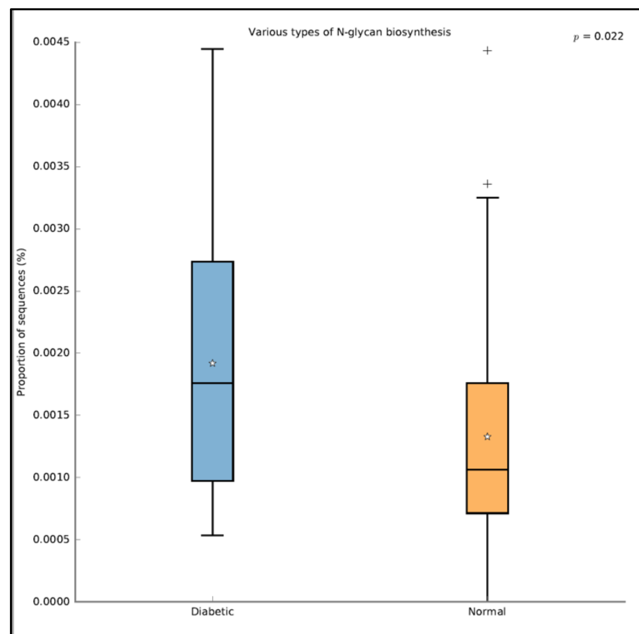


Figure 24: PiCrust analysis of Normal and diabetic in relation to N-glycan biosynthesis (significant difference was observed, p value <0.05)

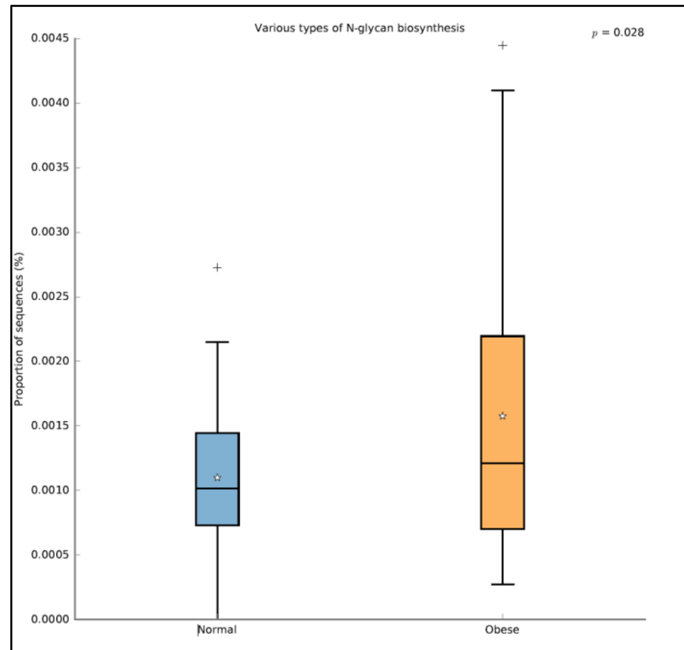


Figure 25: PiCrust analysis of normal weight and obese in relation to N-glycan biosynthesis (significant difference was observed, p value <0.05)

CHAPTER 5: DISCUSSION

The salivary microbiome can be a promising diagnostic biomarker as well as a forensic marker since the sample is easy to collect and the saliva contents are relatively stable. We observed that each individual saliva sample possesses a specific microbial signature. Through measurements of alpha and beta diversity, we were able to find the interindividual diversity within each sample and between all the samples tested as well. We found common shared phyla which resembles *Bacteroidetes*, *Firmicutes* and *Proteobacteria* respectively.

Although the salivary microbiome of many population remained unexamined, there are many studies that support the existence of differences in the microbial composition based on the ethnicity (Gupta, Paul, & Dutta, 2017).

When we examined the alpha diversity, samples from the diabetic subjects were less diverse than the non-diabetic subjects. In addition, when the BMI categories were compared, saliva samples from the normal weight subjects showed a greater microbial diversity as compared to the obese and overweight groups. A study conducted by (Lambeth et al., 2015) to examine the gut microbiome of prediabetic and diabetic microbiome composition did not find any relatedness between diversity and HbA1C level which may be due to their small sample size. Similar findings from another study that compared the alpha diversity between healthy controls, obese and diabetic individuals did not find any significant differences in the alpha diversity between the three groups (Janem et al., 2017). In contrast, a recent study examined the salivary microbiome of diabetic and non-diabetic subjects and stratified the samples according to their periodontal health (Sabharwal et al., 2019). In this study, the authors found

that in the diabetic group, microbial alpha diversity was decreased in association with periodontal disease.

We considered HAllA analysis to find association between clinical parameters and microbial composition. We found that the genus *Catonella* and the genus *Peptostreptococcus* were positively correlated with parameters related to hypertension. Moreover, the genera *Rothia*, *Lautropia*, *Corynobacterium*, *Pseudopropinibactrium* are positively correlated with parameters related to glucose such as glucose level and HbA1C.

In this study, we also investigated the effect of BMI in relation to the microbial composition, we show that in the normal weight group, a balanced abundance of *Bacteroidetes* and *Firmicutes* was observed, whereas, in the overweight and obese subjects, *Bacteroidetes* were almost double the abundance of *Firmicutes.* A study on a chinese population investigated the salivary microbiome composition in obese subjects and found significant differences between the obese group and normal weight group at a more specific taxonomic levels, mainly *Haemophilus* and *Cardiobacterium* were more abundant in normal weight group, while *Prevotella were more abundant in the obese group* (Y. Wu et al., 2018). When we compared the genus level of BMI categories, we found that *Prevotella* are more abundant in both the overweight and obese groups as compared to the normal weight group. Moreover, we show that *Streptococcus* and *Porphyromonas* were decreased in both the obese group and the overweight group.

When we tested the BMI categories in relation to microbiome profile using significant differences among clinical categories using LDA (Linear discriminant analysis) Effect Size or (LEfSe), we found that the two genera *Catonella* and

Solobacterium are constantly associated with obesity which is consistent with the findings of the above mentioned study (Y. Wu et al., 2018).

We have also studied the predictive functional profiling of microbial communities using PiCrust, in which it had revealed that N-Glycan biosynthesis is associated with the diabetic group and obese subjects. N-Glycan biosynthesis is a feature known to be associated with eukaryotes(Dell, Galadari, Sastre, & Hitchen, 2010). However, it is also shared by some bacterial species such as *Campylobacter jejuni* (Kelly et al., 2006).

In the case of assessing the diabetes and obesity there were some limitations to this study due to the small sample size. Moreover sequencing 16S rRNA has its limitations, therefore metagenomic or meta-transcriptomic analyses can further extend our knowledge in this regard.

Our study is the first to assess the salivary microbiome of the Qatari population. We show that the salivary microbiome is highly diverse between subjects and changes according to the health status. More in depth analysis is needed to understand how the identified microbial signatures described in our study, can contribute to disease. While the small sample size is a limiting factor of this study, promising results indicated that a specific salivary microbial signature can be detected in overweight or obese individuals. While using 16S rRNA sequencing technology is sufficient to assess the microbial composition in any body site, a deep understanding of the function of those microbes is still needed. Hence, applying more in-depth techniques such as shotgun metagenomics, metatranscriptomics, microbial metabolomics will enrich our understanding on the role of microbiome in various disease conditions including obesity and diabetes. Unfortunately, this was not possible in our study due to limited

funds allocated to this project.

Despite its limitations, this study will pave the way towards the possibility to find microbial biomarkers that can later be used to detect certain pathological conditions such as diabetes and obesity.

REFRENCES

Acharya, A., Chan, Y., Kheur, S., Jin, L. J., Watt, R. M., & Mattheos, N. (2017). Salivary microbiome in non-oral disease: A summary of evidence and commentary. *Arch Oral Biol, 83*, 169-173.

Al-Shehri, S., Sweeney, E., Cowley, D., Liley, H., Ranasinghe, P., Charles, B., et al. (2016). Deep sequencing of the 16S ribosomal RNA of the neonatal oral microbiome: a comparison of breast-fed and formula-fed infants. *Scientific reports, 6*, 38309.

Althani, A. A., Marei, H. E., Hamdi, W. S., Nasrallah, G. K., El Zowalaty, M. E., Al Khodor, S., et al. (2016). Human Microbiome and its Association With Health and Diseases. *J Cell Physiol, 231*(8), 1688-1694.

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J Microbiol, 56*(4), 394-404.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., et al. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe, 17*(5), 690-703.

Belkaid, Y., & Hand, T. W. (2014). Role of the microbiota in immunity

and inflammation. *Cell, 157*(1), 121-141.

Belstrøm, D., Holmstrup, P., Bardow, A., Kokaras, A., Fiehn, N.-E., & Paster, B. J. (2016). Temporal Stability of the Salivary Microbiota in Oral Health. *PLOS ONE, 11*(1), e0147472.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Cameron, S. J., Huws, S. A., Hegarty, M. J., Smith, D. P., & Mur, L. A. (2015). The human salivary microbiome exhibits temporal stability in bacterial diversity. *FEMS Microbiol Ecol, 91*(9), fiv091.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods, 7*(5), 335-336.

Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., & Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. *Microbial ecology in health and disease, 26*(1), 26191.

Cephas, K. D., Kim, J., Mathai, R. A., Barry, K. A., Dowd, S. E., Meline, B. S., et al. (2011). Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or

primary care givers using pyrosequencing. *PLoS One, 6*(8), e23503.

Chang, Q., Luan, Y., & Sun, F. (2011). Variance adjusted weighted UniFrac: a powerful beta diversity measure for comparing communities based on phylogeny. *BMC Bioinformatics, 12*, 118.

Consortium, H. M. P. (2012a). A framework for human microbiome research. *Nature, 486*(7402), 215-221.

Consortium, H. M. P. (2012b). Structure, function and diversity of the healthy human microbiome. *Nature, 486*(7402), 207-214.

Cox, M. J., Cookson, W. O., & Moffatt, M. F. (2013). Sequencing the human microbiome in health and disease. *Hum Mol Genet, 22*(R1), R88-94.

Dassi, E., Ferretti, P., Covello, G., Bertorelli, R., Denti, M. A., De Sanctis, V., et al. (2018). The short-term impact of probiotic consumption on the oral cavity microbiome. *Sci Rep, 8*(1), 10476.

de Paula, F., Teshima, T. H. N., Hsieh, R., Souza, M. M., Nico, M. M. S., & Lourenco, S. V. (2017). Overview of Human Salivary Glands: Highlights of Morphology and Developing Processes. *Anat Rec (Hoboken), 300*(7), 1180-1188.

Dell, A., Galadari, A., Sastre, F., & Hitchen, P. (2010). Similarities and

Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology, 2010*, 14.

Dusko Ehrlich, S. (2010). Metagenomics of the intestinal microbiota: potential applications. *Gastroentérologie Clinique et Biologique, 34*, S23-S28.

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., & Lemon, K. P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems, 3*(6), e00187-00118.

Francavilla, R., Cristofori, F., Tripaldi, M. E., & Indrio, F. (2018). Intervention for Dysbiosis in Children Born by C-Section. *Ann Nutr Metab, 73 Suppl 3*, 33-39.

Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences, 104*(34), 13780-13785.

Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., et al. (2015). Sequencing and beyond: integrating

molecular'omics' for microbial community profiling. *Nature Reviews Microbiology, 13*(6), 360.

Fukuyama, J., Rumker, L., Sankaran, K., Jeganathan, P., Dethlefsen, L., Relman, D. A., et al. (2017). Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. *PLoS computational biology, 13*(8), e1005706-e1005706.

Gholamali Rahnavard, E. A. F., Lauren J. McIver, Emma Schwager, Jason Lloyd-Price, George Weingart, Yo Sup Moon, Xochitl C. Morgan, Levi Waldron, Curtis Huttenhower. *High-sensitivity pattern discovery in large multi'omic datasets*.Unpublished manuscript.

Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., et al. (2014). Conducting a microbiome study. *Cell, 158*(2), 250-262.

Goodrich, Julia K., Waters, Jillian L., Poole, Angela C., Sutter, Jessica L., Koren, O., Blekhman, R., et al. (2014). Human Genetics Shape the Gut Microbiome. *Cell, 159*(4), 789-799.

Gupta, V. K., Paul, S., & Dutta, C. (2017). Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome

Composition and Diversity. [Review]. *Frontiers in Microbiology, 8*(1162).

Hawkins, T. L., O'Connor-Morin, T., Roy, A., & Santillan, C. (1994). DNA purification and isolation using a solid-phase. *Nucleic acids research, 22*(21), 4543-4544.

He, J., Li, Y., Cao, Y., Xue, J., & Zhou, X. (2015). The oral microbiome diversity and its relation to human diseases. *Folia Microbiol (Praha), 60*(1), 69-80.

Humphrey, S. P., & Williamson, R. T. (2001). A review of saliva: normal composition, flow, and function. *J Prosthet Dent, 85*(2), 162-169.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, function and diversity of the healthy human microbiome. *nature, 486*(7402), 207.

Janem, W. F., Scannapieco, F. A., Sabharwal, A., Tsompana, M., Berman, H. A., Haase, E. M., et al. (2017). Salivary inflammatory markers and microbiome in normoglycemic lean and obese children compared to obese children with type 2 diabetes. *PLoS One, 12*(3), e0172647.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the Gut Microbiome Using 16S or

Shotgun Metagenomics. [Technology Report]. *Frontiers in Microbiology, 7*(459).

Kaczor-Urbanowicz, K. E., Martin Carreras-Presas, C., Aro, K., Tu, M., Garcia-Godoy, F., & Wong, D. T. (2017). Saliva diagnostics - Current views and directions. *Exp Biol Med (Maywood), 242*(5), 459-472.

Kampoo, K., Teanpaisan, R., Ledder, R. G., & McBain, A. J. (2014). Oral bacterial communities in individuals with type 2 diabetes who live in southern Thailand. *Appl. Environ. Microbiol., 80*(2), 662-671.

Karpiński, T. M. (2019). Role of Oral Microbiota in Cancer Development. *Microorganisms, 7*(1), 20.

Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., et al. (2006). Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J Bacteriol, 188*(7), 2427-2434.

Kodukula, K., Faller, D. V., Harpp, D. N., Kanara, I., Pernokas, J., Pernokas, M., et al. (2017). Gut Microbiota and Salivary Diagnostics: The Mouth Is Salivating to Tell Us Something. *BioResearch open access, 6*(1), 123-132.

Kuczynski, J., Stombaugh, J., Walters, W. A., González, A., Caporaso, J.

G., & Knight, R. (2011). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current protocols in bioinformatics, Chapter 10*, Unit10.17-10.17.

Lambeth, S. M., Carson, T., Lowe, J., Ramaraj, T., Leff, J. W., Luo, L., et al. (2015). Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. *J Diabetes Obes, 2*(3), 1-7.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol, 31*(9), 814-821.

Lazarevic, V., Manzano, S., Gaïa, N., Girard, M., Whiteson, K., Hibbs, J., et al. (2013). Effects of amoxicillin treatment on the salivary microbiota in children with acute otitis media. *Clinical Microbiology and Infection, 19*(8), e335-e342.

Li, X., Kolltveit, K. M., Tronstad, L., & Olsen, I. (2000). Systemic diseases caused by oral infection. *Clinical microbiology reviews, 13*(4), 547-558.

Lif Holgerson, P., Harnevik, L., Hernell, O., Tanner, A. C. R., & Johansson, I. (2011). Mode of birth delivery affects oral microbiota

in infants. *Journal of dental research, 90*(10), 1183-1188.

Lim, M. Y., Rho, M., Song, Y. M., Lee, K., Sung, J., & Ko, G. (2014). Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and diet. *Sci Rep, 4*, 7348.

Lim, Y., Totsika, M., Morrison, M., & Punyadeera, C. (2017). The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols. *Scientific reports, 7*(1), 8523-8523.

Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., & Pop, M. (2011). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. [journal article]. *Genome Biology, 12*(1), P11.

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature, 550*(7674), 61-66.

Lynch, S. V., & Pedersen, O. (2016). The human intestinal microbiome in health and disease. *New England Journal of Medicine, 375*(24), 2369-2379.

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry, 6*(1), 287-303.

Mashima, I., Theodorea, C. F., Thaweboon, B., Thaweboon, S.,

Scannapieco, F. A., & Nakazawa, F. (2017). Exploring the salivary microbiome of children stratified by the oral hygiene index. *PloS one, 12*(9), e0185274-e0185274.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One, 8*(4), e61217.

Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z., & Dominguez-Bello, M. G. (2015). The infant microbiome development: mom matters. *Trends in molecular medicine, 21*(2), 109-117.

Nguyen, N.-P., Warnow, T., Pop, M., & White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. [Perspective]. *Npj Biofilms And Microbiomes, 2*, 16004.

Ni, J., Wu, G. D., Albenberg, L., & Tomov, V. T. (2017). Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol, 14*(10), 573-584.

Pereira, P. A. B., Aho, V. T. E., Paulin, L., Pekkonen, E., Auvinen, P., & Scheperjans, F. (2017). Oral and nasal microbiota in Parkinson's disease. *Parkinsonism & Related Disorders, 38*, 61-67.

Petersen, C., & Round, J. L. (2014). Defining dysbiosis and its influence on host immunity and disease. *Cellular microbiology, 16*(7), 1024-

1033.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., et al. (2009). The NIH Human Microbiome Project. *Genome Res, 19*(12), 2317-2323.

Priya, K. Y., & Prathibha, K. M. (2017). Methods of collection of saliva-a review. *International Journal of Oral Health Dentistry, 3*(3), 149-153.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol, 35*(9), 833-844.

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., et al. (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms, 7*(1), 14.

Rodríguez-Rabassa, M., López, P., Rodríguez-Santiago, R. E., Cases, A., Felici, M., Sánchez, R., et al. (2018). Cigarette Smoking Modulation of Saliva Microbial Composition and Cytokine Levels. *International journal of environmental research and public health, 15*(11), 2479.

Rosier, B. T., Marsh, P. D., & Mira, A. (2018). Resilience of the Oral

Microbiota in Health: Mechanisms That Prevent Dysbiosis. *Journal of Dental Research, 97*(4), 371-380.

Sabharwal, A., Ganley, K., Miecznikowski, J. C., Haase, E. M., Barnes, V., & Scannapieco, F. A. (2019). The salivary microbiome of diabetic and non-diabetic adults with periodontal disease. *J Periodontol, 90*(1), 26-34.

Sampaio-Maia, B., & Monteiro-Silva, F. (2014). Acquisition and maturation of oral microbiome throughout childhood: An update. *Dental research journal, 11*(3), 291-301.

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

Shaw, L., Ribeiro, A. L. R., Levine, A. P., Pontikos, N., Balloux, F., Segal, A. W., et al. (2017). The Human Salivary Microbiome Is Shaped by Shared Environment Rather than Genetics: Evidence from a Large Family of Closely Related Individuals. *MBio, 8*(5).

Thiemann, S., Smit, N., & Strowig, T. (2016). Antibiotics and the Intestinal Microbiome : Individual Responses, Resilience of the Ecosystem, and the Susceptibility to Infections. *Curr Top Microbiol Immunol, 398*, 123-146.

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochem J, 474*(11), 1823-1836.

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the human microbiome. *Nutr Rev, 70 Suppl 1*, S38-44.

Wagner, B. D., Grunwald, G. K., Zerbe, G. O., Mikulich-Gilbertson, S. K., Robertson, C. E., Zemanick, E. T., et al. (2018). On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Frontiers in microbiology, 9*, 1037-1037.

Wang, Y., & Qian, P.-Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS one, 4*(10), e7401-e7401.

Wen, L., & Duffy, A. (2017). Factors Influencing the Gut Microbiota, Inflammation, and Type 2 Diabetes. *J Nutr, 147*(7), 1468S-1475S.

Wu, H.-J., & Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. *Gut microbes, 3*(1), 4-14.

Wu, J., Peters, B. A., Dominianni, C., Zhang, Y., Pei, Z., Yang, L., et al. (2016). Cigarette smoking and the oral microbiome in a large study of American adults. [Original Article]. *The Isme Journal, 10*, 2435.

Wu, Y., Chi, X., Zhang, Q., Chen, F., & Deng, X. (2018). Characterization of the salivary microbiome in people with obesity.

PeerJ, 6, e4458.

Yadav, K., & Prakash, S. (2016). Dental caries: A review. *Asian Journal of Biomedical and Pharmaceutical Sciences, 6*(53), 01.

Younes, J. A., Lievens, E., Hummelen, R., van der Westen, R., Reid, G., & Petrova, M. I. (2018). Women and Their Microbes: The Unexpected Friendship. *Trends Microbiol, 26*(1), 16-32.

Zarco, M., Vess, T., & Ginsburg, G. (2012). The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Diseases, 18*(2), 109-120.

Zaura, E., Brandt, B. W., de Mattos, M. J. T., Buijs, M. J., Caspers, M. P., Rashid, M.-U., et al. (2015). Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *MBio, 6*(6), e01693-01615.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30*(5), 614-620.

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. [Article]. *Nature*

Medicine, 21, 895.

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