BACTERIAL DIVERSITY IN ORAL CAVITY AND ITS ASSOCIATION WITH ORAL CANCER

A THESIS

SUBMITTED TO

SVKM'S NMIMS (DEEMED-TO-BE-UNIVERSITY)

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOLOGICAL SCIENCES

BY

Ms. SHRIYA RAJENDRA SAWANT

UNDER THE SUPERVISION OF

Dr. HARINDER SINGH

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SVKM'S Narsee Monjee Institute of Management Studies

(Deemed-To-Be University)

V.L.Mehta Road, Vile-Parle (West), Mumbai – 400056

JANUARY, 2023

BACTERIAL DIVERSITY IN ORAL CAVITY AND ITS ASSOCIATION WITH ORAL CANCER

A THESIS

SUBMITTED TO

SVKM'S NMIMS (DEEMED-TO-BE-UNIVERSITY)

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOLOGICAL SCIENCES

BY

Ms. SHRIYA RAJENDRA SAWANT

Project Guide
Dr. Harinder Singh

Incharge Dean
Dr. Purvi Bhatt

SVKM'S NMIMS UNIVERSITY

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SVKM'S Narsee Monjee Institute of Management Studies

(Deemed-To-Be University)

V.L.Mehta Road, Vile-Parle (West), Mumbai – 400056

JANUARY, 2023

DECLARATION BY THE STUDENT

This is to certify work embodied in the thesis '**BACTERIAL DIVERSITY IN ORAL CAVITY AND ITS ASSOCIATION WITH ORAL CANCER**' for the award of the Degree of Doctor of Philosophy in Biological Sciences is my own contribution to the research work carried out under the supervision of 'Dr. Harinder Singh'. The work has not been submitted for the award of any other degree/to any other University. Wherever a reference has been made to earlier reported findings, it has been cited in the thesis. The thesis fulfils requirements of the ordinance relating to the award of the Ph.D. degree of the University.

Sawant

Place: Mumbai
Date: 10/1/23

Ms. Shriya Rajendra Sawant

Forwarded by

Harinder Singh

Dr. Harinder Singh
Assistant Professor, Biological Scinecs
Sunandan Divatia School of Science
SVKM's NMIMS (Deemed-to-be University)
V.L. Mehta Road, Vile Parle (West)
Mumbai-400056

CERTIFICATE

This is to certify that work described in this thesis entitled "**BACTERIAL DIVERSITY IN ORAL CAVITY AND ITS ASSOCIATION WITH ORAL CANCER**" has been carried out by Ms. Shriya Sawant under my supervision. I certify that this is her bonafide work. The work described is original and has not been submitted for any degree to this or any other University.

Dr. Purvi Bhatt
Incharge Dean
Date: 10/1/2023

Dr. Harinder Singh
Project Guide
Date: 10/1/23

SUNANDAN DIVATIA SCHOOL OF SCIENCE
SVKM's Narsee Monjee Institute of Management Studies
(Deemed-to-be University)
V. L. Mehta Road,
Vile-Parle (West), Mumbai – 400056.

This thesis is solely dedicated to

MY PARENTS

RAJENDRA SAWANT AND NAYNA SAWANT

&

MY SISTER RIYA

Acknowledgements

"Every success story, starts with a dream"

In this most important dream and chapter of my life, there have been people without whom this reality would still be a dream. There have been numerous people who have not only guided but also led me along the way, who have supported, channeled, and paved the way.

First and foremost, I would like to express my heartfelt sincere gratitude to my mentor, **Dr. Harinder Singh**, for conceptualizing this project and believing in me to see it through. He has always encouraged me to achieve my full potential and provided me invaluable expertise and support. He made me believe in myself at times when I couldn't and picked me up when I was at my lowest. He taught me to never give up and that patience and hard work always pays off. Without his support this journey would have been impossible. I couldn't have asked for a better mentor for this journey.

I would like to thank **Dr. Purvi Bhatt**, Incharge-Dean and HOD, Sunandan Divatia School of Science, for providing research amenities and permission to carry out this work. I would also like to thank **Dr. Aparna Khanna** (Ex-Dean) for her valuable inputs during the progress meetings. I would like to acknowledge **Dr**. **Dhananjaya Saranath**, Ex-Emeritus Professor, Sunandan Divatia School of Science for her help in initial designing of the project. She has always been inspiring and enthusiastic about science and full of life. I feel extremely fortunate to have met her in my early years of education and research.

Next I would like to extend thanks to my doctoral committee members, **Dr. Kayzad Nilgiriwala** (Assistant Director -Research, Foundation of Medical Research, Mumbai) and **Dr. Binoj Kutty** (Assistant Professor, St. Xavier's College, Mumbai), for their timely evaluation and insightful suggestions during my Ph.D journey which helped me shape this thesis into what it is today.

I am very grateful to my clinical collaborators **Dr. Jinesh Dugad** and **Dr. Deepak Parikh** (Somaiya Ayurvihar-Asian Cancer Institute) for accepting this project proposal and providing their help in

collection of samples from cancer patients. They have been extremely supportive and introduced me to the clinical aspect of the study. Their immense support and research interest has helped me complete this study. I would like to extend my special thanks to our collaborator, **Dr. Sathiyaraj Srinivasan** (Associate Professor, Seoul Women's University) for his help with sequencing all our samples. Without his help, it would have been extremely difficult to achieve the most important goal of this project.

I am thankful to all faculty and staff of Sunandan Divatia School of Science that includes **Dr. Ekta Khattar** for providing us with the materials whenever necessary for conducting the study as well as her constant guidance to improve cell culture experiments. She has always been encouraging and ready to help whenever possible without any hesitation. Thanks to **Dr. Brijesh Sukumaran** for his thoughtful insight and suggestions related to cell culture experiments. **Mrs. Madhuri Khanolkar** deserves a special recognition for helping us with our requirements for the experiments and making sure to procure everything we needed to complete our study on time. Throughout the journey, she has been standing, through the thick and thin, from listening to our complaints and disappointments, to celebrating our successes and joys. Also, **Mrs. Manasi Naik** and **Mrs. Karuna Dhamane** for their help related to all academic and administrative handling and they made sure to resolve any queries at the earliest.

I'd like to express my heartfelt gratitude to my friends, whom I now consider to be family. It would have been extremely difficult to complete this journey sanely without this close-knit team of seven of us. Firstly, **Mr. Mitesh Joshi**, who has been my pillar of strength, support, and my go-to-person throughout all these years. He is the most selfless and kind person I have known so far. He has never hesitated to help me in all ways that he could. **Ms. Asmita Kamble**, who has been like a sister, with whom I could laugh and cry out to about anything and everything. Her hugs have been a constant throughout and they have got me through painful, stressful as well as successful days of my life. **Ms. Ruchita Joshi** who has been a host to all our sleepovers which brought all of us closer. **Ms. Jasmeet Kaur Virdi**, with whom the journey began with an intellectual bond and evolved into a personal bond which I could never fathom. **Ms. Meenakshi Shukla** and **Mrs. Hiral Khakkar-Hindocha** for the perfect breakfast sessions we had to kick start

our days. This journey would have been mundane without all the trips, food outings, the coffee-breaks, Juhu visits, impromptu plans, and important life decisions and journeys together with all of them. My stressbusters, **Ms. Kavita Matange**, **Ms. Prachi Shah** and **Ms. Srinidhi Shetty**, with each of them pushing me whenever I felt low and making me believe “This too shall pass”. I am forever grateful for the friends I have, who have made me and my journey a better one by contributing in their own ways and providing support which I will never forget.

Of all, I would like to thank the most important people in my life, my family. My father, **Mr. Rajendra Sawant** and mother **Mrs. Nayna Sawant** for always believing in me, encouraging me, pushing me to do my best, and constantly reminding me that I can do this. In some way I feel they have lived their dream through me, and I am extremely fortunate to have them. My sister, **Riya Sawant**, whose unending love strength has got me through the worst. I couldn’t have been the person I am without my family, getting through each day, knowing I have them to go back to everyday has got me this far. The physical, emotional, mental support they provided, the tears they shed for me, with me, the way they handled me through all my ups and downs cannot be justified by words. They are my everything and I owe them my all.

Lastly, I would like to thank all the participants and patients from the bottom of my heart, without whom this project wouldn’t have been possible.

Shriya Sawant

Abstract

Abstract

The oral cavity harbours approximately 700 microbial species, collectively known as the oral microbiota, and maintaining the ecological balance of these organisms is critical for good health. Dysbiosis of the oral microbiome has been reported to play a key role in the genesis and progression of oral cancer (OC). Research has shown that there is an association between use of tobacco, a well-established risk factor for OC, and oral dysbiosis. With the advent of next-generation sequencing technology, identification of bacterial composition in the oral cavity has become much more achievable. The oral microbiota has been widely studied across the globe, but there is paucity of the oral cancer microbiome data in the Indian population that contributes to one-third of global oral cancer cases. Hence, the aim of the current work was to identify the bacterial composition in the oral cavity of Indian OC patients and long-term tobacco chewers and compare them with those of healthy individuals.

In this study, oral saline rinse samples were collected from 120 participants belonging to three study groups (consisting 40 participants each) namely, control group, long-term tobacco chewers and oral cancer patients. Genomic DNA was isolated from these samples, and they were subjected to paired-end V6-V8 16S rRNA sequencing using Illumina MiSeq. The raw data obtained after sequencing was analysed using Quantitative Insights Into Microbial Ecology (QIIME2) metagenomic pipeline. Quality of sequences was assessed using FASTQC and low-quality sequences were discarded. The sequences were clustered into operational taxonomic units (OTU) followed by assignment of taxonomy using GreenGenes database. The (α) alpha diversity was assessed using Chao1 index, ACE index, Goods coverage, pielou_e, Shannon, Simpson index and observed OTUs, which portrays highest alpha diversity in tobacco group followed by control group and lowest in OC group. The between groups (β) diversity was assessed using PERMANOVA, PERMDISP and ANOSIM along with 3D matrices which suggests more similarity and close clustering of control and tobacco samples together while OC group away from the cluster and less similar to control and tobacco chewers microbiota. The bacterial composition and abundance were further analysed. Identification of significant biomarkers for each group using LDA Effect Size (LEfSe) revealed *Leptotrichia, Treponema, Lautropia, Tannerella, Selenomonas, Filifactor, Campylobacter* and *Cardiobacterium* as potential biomarker for tobacco group. On the other hand, *Pseudomonas, Capnocytophaga, Mycoplasma, Bifidobacterium, Peptostreptococcus* and *Paludibacter* were associated as biomarkers for OC and bacteria belonging to genera *Rothia, Neisseria, Actinobacillus, Veillonella and Corynebacterium* were identified as potential biomarkers for control population. The functional pathways were predicted using PICRUSt2. The results demonstrated pathways related to lipid, fatty acid and coenzymeA synthesis were upregulated

in the OC group. The Reductive TCA cycle and pyrimidine biosynthesis pathway were upregulated in tobacco chewers.

Although NGS sequencing is a preferred approach to perform microbiome analysis, however, a simple and cost-effective method will be quite helpful, in small scale set-ups, in performing absolute quantification of specific bacterial populations in a particular environment/ecosystem. In most of the microbiome studies, 16S rRNA gene variable regions are used as the reference gene, which is a multicopy gene that can lead to inaccurate quantification of bacteria. So, we aimed to develop a qPCR based method for absolute quantitation of oral bacteria using species-specific single copy *rpoB* gene primers. Using *E. coli* as a model organism, we established a relationship between DNA and BGE (Bacterial Genome Equivalent)/ml. *E. coli* specific *rpoB* primers were used to obtain qPCR based standard curve, which was further used to calculate absolute abundance of bacteria using Ct values. Eight different bacteria reported in oral cancer, namely *Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula,* and *Streptococcus mutans,* was tested with this method. The abundance of *P. gingivalis, F. nucleatum, C. gingivalis, P. melaninogenica* and *V. parvula* depicted increase in OC samples as compared to control, whereas the patter was opposite *for R. mucilaginosa, H. parainfluenzae* and *S. mutans* . Parameters associated with the study participants were also assessed and the pattern of abundances concur with previously published reports, that helped in validating the developed method.

To extend our understanding of the oral microbiome and its interaction with the host, a preliminary host-microbe study was carried out, using *Streptococcus mutans*, a causative agent of dental caries. This will help us to gain knowledge about the mechanism of infection and its potential contribution to the development of cancer. Following the bacterial infection with human oral cell lines, initial cytotoxicity assay, microscopic observation of infection, parameters for gene expression studies were standardized. Differential expression analysis of genes related to various hallmarks of cancer development were considered for the analysis. Overexpression of Stathmin, CyclinD1 and BCl2 was observed at 3 and 7 hrs of infection as compared to uninfected cells. The result indicated that cancer pathways related to the overexpressed genes could be triggered/dysregulated due to infection, however, this work needs further investigations for further understanding.

To conclude our study, we identified the bacterial population in the oral cavity of oral cancer patients and tobacco chewers and compared them with healthy individuals in the Indian population. Additionally, a simple, efficient, user and resource-friendly method was developed for absolute

quantification of bacteria in the oral cavity. Further, a preliminary evaluation of the gene expression was carried out to understand the host-microbe interaction.

Table of contents

List of abbreviations

α	Alpha
AAR	Age adjusted rate
ACE	Abundance-based coverage estimator
ANOSIM	Analysis of similarities
ANOVA	Analysis of variance
β	Beta
Bax	Bcl associated factor X
Bcl	B-cell lymphoma
BGE	Bacterial genomic equivalent
BHI	Brain heart infusion
bp	Base pair
C	Control population
°C	Degree celsius
CFU	Colony forming units
Ct	Cycle threshold
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribose nucleic acid
DPBS	Dulbecco's phosphate buffered saline
EDTA	Ethylene diaminetetraacetic acid
EtBr	Ethidium bromide
EtOH	Ethanol
FBS	Fetal bovine serum
g	Grams
HMP	Human microbiome project
HNSCC	Head and neck squamous cell carcinoma

hrs	Hours
IL	Interleukin
IPA	Isopropylalcohol
LEfSe	LDA effect size
min	Minutes
ml	Milliliter
MOI	Multiplicity of infection
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NaCl	Sodium chloride
NBM	Nil By Mouth
NGS	Next generation sequencing
OC	Oral cancer
O.D.	Optical density
OSCC	Oral squamous cell carcinoma
OTU	Operational taxonomic units
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCoA	Principal coordinate analysis
PERMANOVA	Permutational multivariate analysis of variance
PERMDISP	Permutational analysis of multivariate dispersions
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative real time polymerase chain reaction
rcf	Relative centrifugal force
RDP	Ribosome database project
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature

rpoB	β-subunit of RNA polymerase
SD	Standard deviation
STAMP	Statistical analysis of taxonomic and functional profiles
T	Long term tobacco chewers
TAE	Tris-acetic acid-EDTA
Tm	Melting temperature
TSA	Tryptone soya agar
UV	Ultraviolet
VBNC	Viable but non culturable

List of figures

List of Tables

CHAPTER 1

Introduction

Introduction

Carcinogenesis, also termed as Oncogenesis or Tumorigenesis, is a complex process of transformation of normal cells into cancer cells, characterized by accumulation of genetic and epigenetic events which lead to uncontrolled cell division (Vogelstein and Kinzler 1993). Like most cancers, oral cancers too progress through hyperplasia, dysplasia, carcinoma in situ, invasive carcinoma and metastasis to distant organs. Five percent of all tumors of the human body are accounted for by head and neck cancers, half of which are cancers of the oral cavity (Kademani 2007). Oral squamous cell carcinoma (OSCC), a subset of oral cancer accounts for approximately 90% of all oral cancers (Tandon et al. 2017). Oral cancers are classified based on the International Classification of Diseases (ICD-10) as C00 to C06. The parts of the oral cavity have been given different names, as shown in Fig. 1.1., and the Oral cancers include cancers of the lip (C00), tongue (C01-C02), gum (C03), floor of mouth (C04), palate (C05), retromolar trigone (C06.2) and unspecified parts of the mouth (C06) (www.ICD10Data.com). The histopathological classification of cancer lesions depends on the degree of tumor differentiation as well differentiated, moderately differentiated and poorly differentiated cancers (Warnakulasuriya et al. 2007).

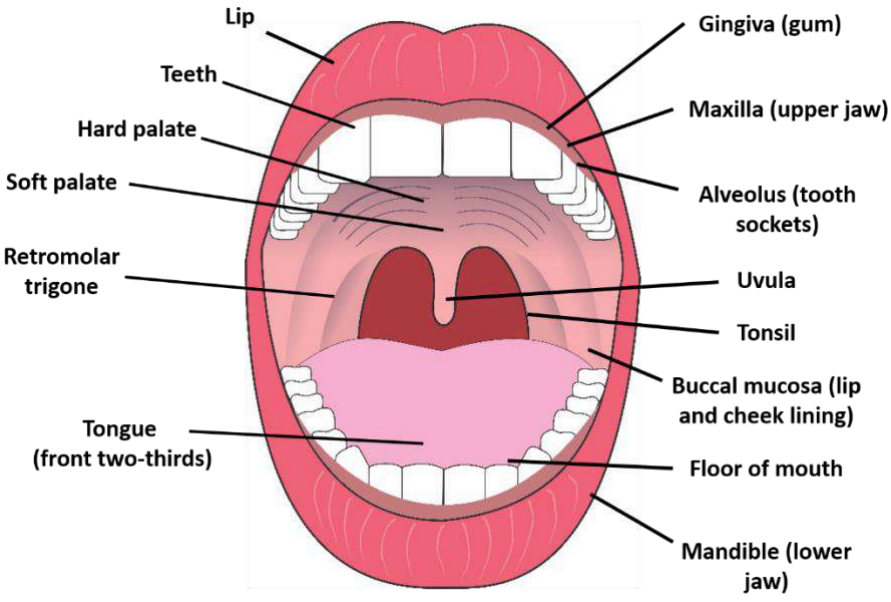
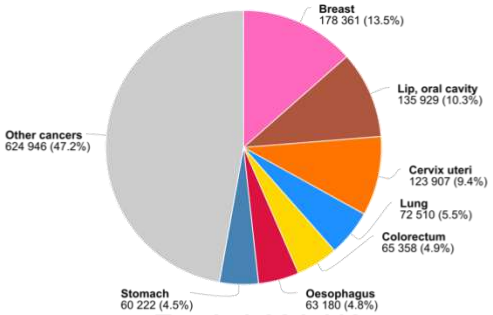


Fig. 1.1: Sites affected by oral cancer in the oral cavity

Oral cancer is the 16th most common cancer worldwide with an annual incidence of 377,713 new cases, mortality of 177,757 and five-year prevalence of 959,539. It is the 11th most common cancer in males and 18th most common cancer in females across the globe (https://gco.iarc.fr/today/). Oral cancer is one of the major health concerns in India, contributing 35.9% to the global burden of oral cancer (Fig. 1.2), with annual incidence of 135,929; mortality of 75,290 and 300,413 five-year prevalence. Oral cancer is the most common cancer in Indian males with an annual incidence of 104,661 and fourth most common cancer in Indian females with annual incidence of 31,268. (Globocan 2020, as accessed on June 03, 2021).

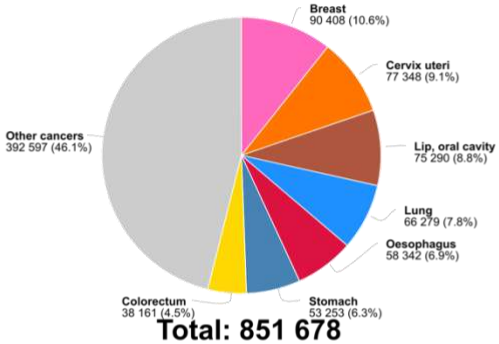
Oral cancer is a multifactorial disease and multiple risk factors attribute to the onset and progression of the disease. The major risk factors of oral cancer include tobacco use (smokeless and smoked), areca nut chewing, alcohol consumption, high risk oncogenic Human Papilloma Virus (HPV) (Warnakulasuriya 2009). Long term tobacco use (smoking and smokeless) contributes to approximately 75-80% towards causation of oral cancer. Similarly, arecanut use, alcohol use, and HPV 16/18 contributes 50%, 7-19% , and 5-10% respectively. Along with these factors, the probable risk of diet attributing to causation of oral cancer is 10-15% (Multani and Saranath 2016). Apart from these well-known risk factors, the new emerging risk factor associated with oral cancer is the involvement of bacterial flora/microbiota of an individual, also called the microbiome. With a huge progress and development in genome sequencing, especially, the next generation sequencing (NGS) technology, studies exploring the role of bacterial flora in health, diseases , and cancer have gained great interest.

(a) Estimated number of new cases in 2020, India, Both sexes



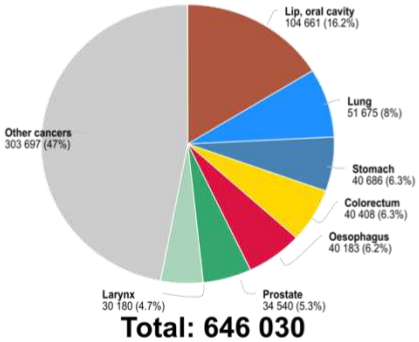
Total: 1 324 413

(b) Estimated number of deaths in 2020, India, Both sexes



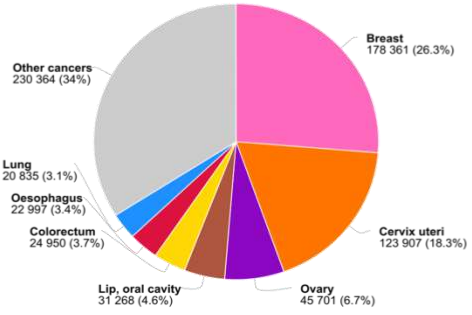
Total: 851 678

(c) Estimated number of new cases in 2020, India, Males



Total: 646 030

(d) Estimated number of new cases in 2020, India, Females



Total: 678 383

Fig.1.2 : Statistics of oral cancer in India a)new cases in both sexes; b) deaths in both cases; c) new cases (males); d) new cases (females).

A myriad of microorganisms inhabit different sites of the human body and are commonly called the human microbiota. These microorganisms play an essential role in maintaining the wellness of an individual. These inhabitants, including bacteria, fungi and viruses, continuously interact with the body cells and have an important role to play in health, disease and physiological maintenance (Cho and Blaser 2012). The involvement of human bacterial flora as a risk factor for causing cancers, including oral cancer, has been a topic of recent interest. The presence of certain bacterial species, or dysbiosis of the normal microflora, is known to increase the risk of initiation and progression towards cancer. The cause-effect relationship of bacteria and cancer is still under investigation and very less is known till date. There are more questions than answers when it comes to involvement of bacteria in causation of cancer. With the advancement in sequencing technology, it has become faster and more affordable to detect and identify the microbiome for many samples, environmental or clinical. With the help of NGS, the bacterial diversity can be unveiled, and that includes culturable as well as non-culturable bacteria, the ones that cannot be identified by conventional microbiological methods (Chelsie 2015). The most common oral pathogens identified, which have been correlated with oral diseases are *Porphyromonas gingivalis*, *Fusobacterium nucleatum, Prevotella, Capnocytophaga spp.* On the other hand, a few bacteria are shown to be associated with a healthy oral cavity, such as *Streptococcus, Rothia, Haemophilus* and *Lautropia* (Healy and Moran 2019). Although these few bacterial species have been shown to be associated with health or disease, the bacterial diversity in oral cavity is dynamic in nature in a sense that it depends on multiple factors such as age, ethnicity, diet, socioeconomic status, heredity, lifestyle and oral hygiene (Marsh and Percival 2006; Belstrøm et al. 2014).

To understand such microbial population diversity, a metagenomics approach, a molecular tool or method is generally used to study microbial communities from biological/environmental samples, without the need to isolate pure cultures, by using DNA obtained from the samples (Ghosh et al. 2018). For most metagenomic studies, the 16S rRNA gene acts as a molecular marker. The 16S rRNA is universally present in all prokaryotes, is not affected by lateral gene transfer, contains multiple hypervariable regions, along with which they are conserved across prokaryotes making it a useful marker for taxonomic classification (Lan et al. 2016). This 16S rRNA gene based bacterial classification has given rise to three major databases namely, Ribosomal Database Project (RDP), SILVA and GreenGenes

(McDonald et al. 2012; Quast et al. 2013; Cole et al. 2014). The 16S rRNA gene (Fig. 1.3) is approximately 1500bp in size and composed of nine hypervariable regions, which have conserved regions between them. The hypervariable regions named V1-V9 are used for taxonomic assessment. The conserved regions flanking the variable regions are used for designing primers specific for amplification (Yang et al. 2016). With varied regions used for classification, the depth and precision of assigning taxonomy changes too. The data obtained from the sequencing can be analysed using multiple pipelines such as MG-RAST, QIIME2, MOTHUR, GALAXY, *etc*. These pipelines analyse the raw data obtained from the sequencer depending on the quality of the sequences obtained, clustering of the sequences in operational taxonomic units, assigning taxonomy to the sequences as per chosen databases, and finally calculating abundances of bacteria. The taxonomy can be obtained across all phylogenetic levels, *i.e.*, kingdom, phyla, family, class, order, genus and species.

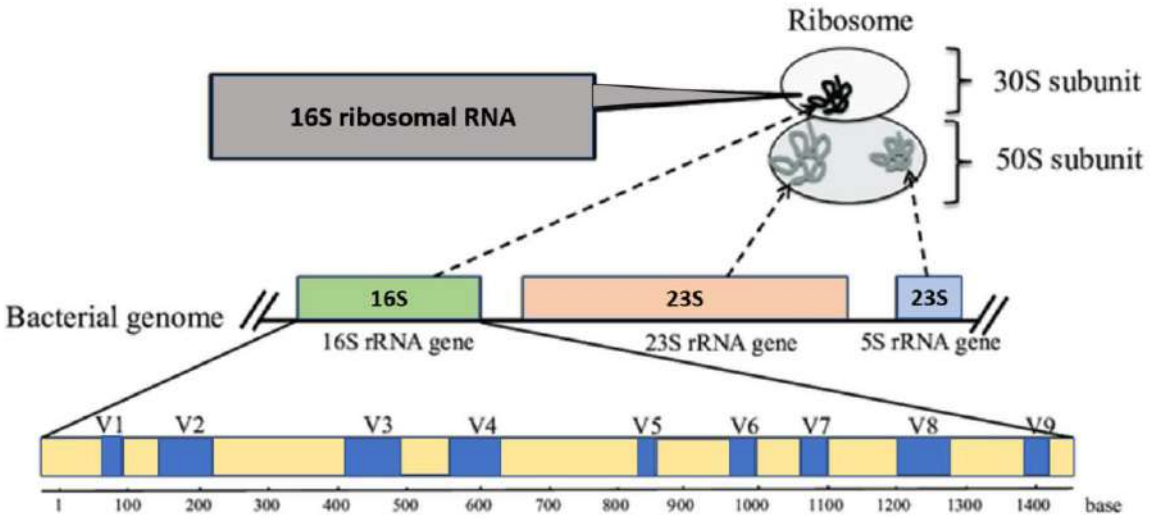


Fig.1.3: The 16S rRNA gene structure. Image source: (Fukuda et al. 2016)

Such metagenomic studies for the oral microbiome of a specific population has been widely studied across the globe. However, there is a dearth in the knowledge about healthy oral microbiome in the Indian population as well as the oral microbiome of the Indian oral cancer patients. India along with other Southeast Asian countries is also a hub of tobacco chewers, which is one of the most crucial risk factor associated with oral cancer. Therefore, NGS can be used to unveil, correlate and find association between the oral microbiome in the Indian healthy population, patients with oral cancer as well as long term tobacco chewers.

Although the metagenomics approach provides us with the detailed composition of the bacteria present in a given sample, which include viable as well as non-viable bacteria, there is a pitfall associated with this approach due to the use of 16S rRNA gene. The downfall of working with this gene is the multicopy characteristic, *i.e.*, the 16S rRNA gene is a multicopy gene which means there are variable copies of the gene in various organisms, which makes it difficult to carry out accurate quantification of the bacterial community using the 16S rRNA gene. To compensate for this issue with the 16S rRNA gene, researchers have resorted to various other ecology markers (Wu et al., 2013). One such gene rpoB gene, that has similar ecological and taxonomic characteristics across bacteria, and is also a single copy gene, *i.e.,* one gene per cell (Case et al. 2007). So, using such a single copy gene in a quantitation method like quantitative real time PCR (qPCR) can provide a better approach for absolute quantification of specific bacterial population from samples like the oral cavity. Such methods can also provide species level confirmation which is not always available in 16S rRNA gene metagenomics studies, as it mostly provides genus level accurate information (Adékambi et al., 2009).

Quantitative PCR techniques were developed as a modification of PCR for real time quantitation of DNA samples, and then later expanded for quantitation of RNA copies also that was named as qRTPCR (VanGuilder et al., 2008). Eventually, the application of this technique has spread across various areas of biology, including microbial ecology, to quantify the presence and abundance of DNA and RNA gene copies that is also useful for quantifying microbial species in any given environmental sample (Wang et al., 1989). The principle of a qPCR method is based on specific amplification of a target nucleic acid template that increases the intercalation of fluorescent SYBR dye (N,N-dimethyl-N'-[4-[(E)-(3-methyl-1,3-benzothiazol-2-ylidene)methyl]-1-phenylquinolin-1-ium-2-yl]-N'-propylpropane-1,3-diamine.) in double stranded DNA, and the increase in fluorescent signal can be correlated directly with amount of amplicon product formation in each cycle of the PCR. So, the number of amplicons detected is directly proportional to the initial target sequences copies in the source sample, and finding and comparing them between samples is usually the objective of such studies . Usage of fluorescence dyes offers higher sensitivity and enables better discrimination of gene copies, which is not the case in commonly used end-point PCR (Smith and Osborn 2009). The most common fluorescence detection chemistries used in qPCR are

TaqMan probe system and SYBR green assay (Fig. 1.4). A post-PCR dissociation (melting) curve analysis should be performed to ensure that the fluorescence signal is generated solely from target templates. During the dissociation curve, double stranded amplicon template is heated over a temperature range, which denatures the amplicon and leads to decrease in fluorescence signals due to dissociation of SYBR green dye from the double stranded product (Giglio et al. 2003; González-Escalona et al. 2006). A qPCR platform collects the increasing fluorescent data from every amplification cycle and plotted against cycle number, resulting in an amplification curve. The gene numbers are quantified using the Ct (cycle threshold) method. The Ct value for a particular amplification is reached when accumulation of fluorescence by amplified products is significantly greater than background fluorescence level unlike in initial stages of amplification where the background fluorescence is higher than that of amplified products. There is an inverse relation between the Ct value and initial target template amount. When the initial concentration of the target template is high, the Ct values will be lower or it will be reached early as compared to the case when the initial target template is low, then the Ct values will be high or it will be reached at a later amplification cycle (Livak & Schmittgen, 2001). Using these values, the qPCR method can be used to carry out absolute quantification of single copy genes like *rpoB*, that can be correlated with cell numbers of a significant bacteria and thus measure their abundance in different sample groups.

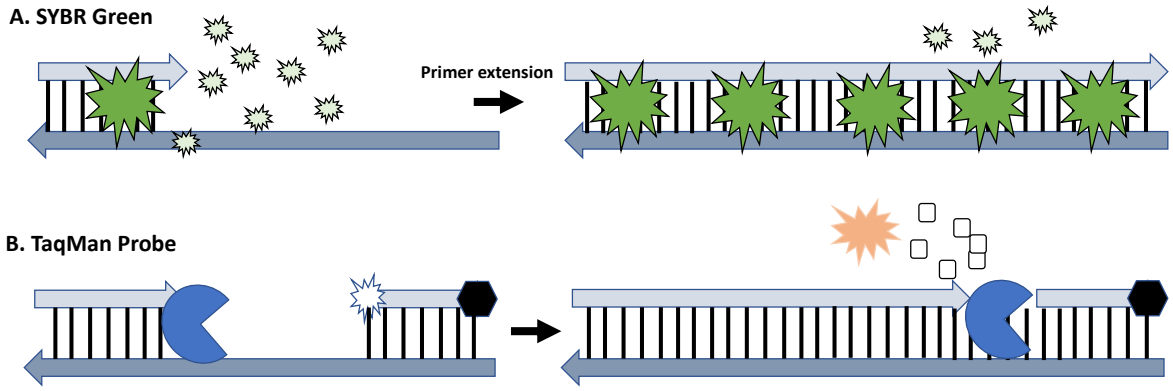


Fig.1.4: Mechanism of fluorescence-based quantitation in real time PCR.

The present study involves the process of identifying the bacterial community in the oral cavity and to quantitate them. Apart from these objectives, it will be worthwhile to investigate and understand the molecular mechanisms or changes by which certain bacterial species might be associated with oral cancer development or progression. Various mechanisms have been proposed for the bacterial involvement in carcinogenesis. Some of the proposed mechanisms have been stated below:

1. Certain bacteria can cause chronic infections, which can lead to inflammation. They also produce and release toxins which can alter the host cell cycle and result in altered cell growth patterns (Alistair Lax, 2005; Stone & Darlington, 2017).
2. Infection causing pathogenic bacteria destabilize host cell signalling pathways which in turn lead to enhanced chances of survival of pathogen. These signalling pathways are most commonly involved in tumor development or inhibition. Often the gross result of alteration of pathways are similar to tumor development which often reduces after antibiotic treatment and bacterial clearance (Beltran et al., 2021; Olsen & Yilmaz, 2019)
3. Cell proliferation and DNA replication are modulated by mitogen activated kinase (MAPK) and Cyclin D1 pathways. These pathways can be induced due to chronic infections as well, which can thereby increase chances of genetic mutations and thus enhance cellular replication and cellular transformation (Bennedsen et al., 2022; Hirata et al., 2001).
4. Escaping the immune surveillance by the bacteria and their accumulation over time can lead to suppression of apoptosis genes in the host cellular machinery. Along with evasion of bacterial elimination, this mechanism aids in skipping destruction of partially transformed host cells as well. As a result, partially transformed cells can avoid self-destruction and advance to a greater level of transformation, eventually becoming tumorigenic (Wong et al., 2019).
5. Metabolism of potentially carcinogenic substances by bacteria is another proposed mechanism for involvement of bacteria in carcinogenesis. Substances consumed by an individual could be transformed into carcinogenic products using the enzymes from specific bacteria present in the host. For example, ethanol which is consumed by a significant population across the globe is non-carcinogenic, although its derivative

acetaldehyde is a potent carcinogen. Bacteria such as *Streptococcus* spp. is capable of converting ethanol to acetaldehyde by enzymes such as alcohol dehydrogenases (ADH) produced by the bacteria. This ethanol derived acetaldehyde is capable of inducing DNA damage, mutagenesis and secondary hyperproliferation of the epithelium (Kurkivuori et al., 2007; Moritani et al., 2015; Pavlova et al., 2013).

Therefore, it is important to understand the underlying mechanism of action of any bacteria that is capable of inducing any process leading to carcinogenesis. Investigating such mechanisms will not only help understand the host-microbe interaction process that can progress towards cancer, but also opens more opportunities and possibilities in terms of treatment and diagnostic modalities.

In view of the recent research work conducted at global scale and lack of similar information in Indian population, an attempt was made to identify and understand the oral microbiota, using Next Generation Sequencing technology, of Indian health subjects, patients suffering from oral cancer and long-term tobacco chewers population, which is a major risk factor for oral cancer. Further, we developed a quantitative real time PCR method that uses a single-copy gene for absolute quantification of specific bacterial species present in our study groups. To extend the microbiome study, gene expression study was conducted to identify the molecular changes that might be associated with the progression or development of carcinogenesis, when the host cells were incubated with a specific bacterial cell population. Hence, the present work aims to find the role of bacteria as a risk factor in oral cancer, especially the Indian population.

CHAPTER 2

LITERATURE REVIEW

Literature Review

2.1 Oral Cancer

2.1.1 Epidemiology of oral cancer

Globally, oral cancer is 11th most common cancer in males and 18th most common in females with an annual incidence of 377,713 new cases (Sung et al. 2021). The incidence of oral cancer in developing countries (199,550) is twice as compared to that of developed countries(100,823) (Ferlay et al. 2015). According to Globocan 2020 statistics, largest contribution of oral cancer incidences are reported in Asia (65.7%), Europe (17.28%), North America (7.27%), Latin America and the Caribbean (4.73%), Africa (3.78%) and Oceania (1.17%). The highest mortality rate was observed in Asia (34.84 %) followed by Europe (6.50 %), Africa (2.14 %), Latin America and the Caribbean (1.31 %), North America (2.9%) and Oceania (0.25 %) (Sarode et al. 2020). The Asian continent contributes 248,360 new cases annually. Glancing at the countries in the Asian continent, India accounts for 135,929 cases, followed by China (30,117), Pakistan (16,959), Bangladesh (13,985), Japan (11,210). As high as 75,290 deaths have been observed in India, 14,785 in China, 10,617 in Pakistan and 8,137 in Bangladesh etc. The projected burden of oral cancer for both genders is estimated to increase among males by 59.5% and among females by 62.3% in less developed countries, compared to an estimated increase among males by 21.6% and females by 22.1% in more developed countries by the year 2030 (Gupta et al. 2017).

Table 2.1: Oral cancer incidence according to Globocan 2020

Population	Male	Female
World	264 211	113 502
Asia	178 396	69 964
India	104 661	31 268

(Estimated oral cancer incidence in 2020: Accessed 21st December 2022)

In India, oral cancer is the most common cancer in males and 4th most common cancer in females. Oral cancer contributes to approximately 35.9 % of global burden for oral cancer and 10.36% cancers in the country. It is estimated that the incidence of total oral cancer cases in India will increase from 1.0 million in 2012 to 1.7 million in 2035, also indicating a rise in

mortality from 680,000 to 1.2 million (Varshitha 2015). The National Cancer Registry Programme (NCRP), established by the Indian Council of Medical Research in 1981 provides data from 28 Population Based Cancer Registries (PBCRs) located in India. According to NCRP, age-adjusted rate (AAR) of cancer of tongue in males is observed highest in Aizawl which is as high as 10.2 per 100,000 followed by Bhopal (AAR 10.0), Gandhinagar (AAR 7.5), Ahmedabad (AAR 6.6), Delhi (AAR 5.6), Chennai (AAR 4.8) and Mumbai (AAR 4.5). Similarly, cancer of mouth excluding tongue is observed highest in Wardha, Maharashtra as high as AAR 14.1, followed by Kanyakumari (AAR 11.6), Thiruvananthapuram (AAR 10.9), Bhopal (AAR 8.6), Pune (AAR 6.8) and Ahmedabad (AAR 5.8). Similarly, AAR of mouth cancer, excluding tongue for females in India, is highest in Kolar, Karnataka (AAR 10.7), followed by Bangalore (AAR 8.9), Pondicherry (AAR 7.8) and Bhopal (AAR 6.4).

2.2 Risk factors of oral cancer

Just like any other cancer, oral cancer too is a multifactorial disease. These risk factors can be categorised into non-modifiable factors which include genetic predisposition and age; modifiable risk factors such as lifestyle habits including tobacco chewing, smoking, alcohol consumption, betel quid use and diet; and other risk factors such as HPV infection and immunosuppression (Meurman 2010; Warnakulasuriya 2009). Apart from these risk factors, there was an increase in oral cancer cases which could not be attributed to the known risk factors, thereby identifying other factors like microbes as a new and emerging risk factor for the same. Here, major risk factors of oral cancer are explained in the next few sub-sections. the.

2.2.1 Tobacco use

Smokeless tobacco has been identified with various proven carcinogenic constituents including tobacco specific N-nitrosamines, N’-nitrosonornicotine (NNN), polycyclic aromatic hydrocarbons (PAH) and 4(methylnitrosamino)-l-(3-pyridyl)-l-butanone (NNK). Tobacco smoking in the form of cigarettes and bidis have been identified with the presence of carbon monoxide, tar, phenols, nicotine, hydrogen cyanide benzo-[a]-pyrene etc (Xue, Yang, and Seng 2014). The metabolites of NNK and NNN are potent carcinogens, the metabolism pathways of which are catalysed by cytochrome P450 enzymes. Cytochrome P450 enzymes catalyse the α-hydroxylation of NNNs forming α-hydroxy-NNN resulting in DNA adducts.

Benzo[a]pyrene is metabolized by cytochrome P450 enzymes to epoxides resulting in DNA adducts (Hecht and Hoffmann 1988). The formation of DNA adducts can lead to miscoding in the DNA leading to mutations and thereby oncogene activation. Consecutively, tobacco specific nitrosamines can bind to nicotinic-acetylcholine receptor leading to activation of protein kinase A and serine-threonine kinase ,thereby decreasing cellular apoptosis, and increased cell differentiation and proliferation (Hecht 2003). Tobacco smoke causes oxidative harm due to the presence of free radicals, for example, nitric oxide and combinations of hydroquinones, semiquinones and quinones, which prompt redox cycling (Hecht 2011).

A dose dependent relationship has been established between tobacco use and the risk of developing cancer, with the frequency of tobacco use and chewing is directly proportional to the risk of developing cancer. In India, it is estimated that in the age range of 15-49 years, 57% men and 11% women use some form of tobacco. An increased risk of cancer of 9.2 times has been observed in women chewing tobacco more than 10 times a day as compared to non-chewers. Similarly, the risk for gutka consumption was 7.3 times, chewing tobacco was 5.3 times and consumption of supari was 4 times, whereas the risk for use of misri was comparatively lower (Varshitha A 2015).

2.2.2 Areca nut

There is high consumption of arecanut in South Asian countries in the form of paan. The common components of paan are tobacco, seeds, quenched/slaked lime, spices, and arecanut enfolded in betel quid. Areca nut has been known to increase collagen synthesis, ROS production due to oxidation of areca nut phenols as well as DNA and fibroblast damage (Niaz et al. 2017). The underlying mechanism of carcinogenesis due to arecanut is not yet well understood. Chronic inflammation and irritation of the oral cavity, resulting in the damage to epithelial cells has been observed in regular chewers. Pro-carcinogenic aberrations such as increased expression of heat-shock proteins and integrins, increased crosslinking of collagen and inhibition of collagenases, increased inflammatory cytokines and interleukins, modulation of matrix metalloproteinases, has been linked to the use of arecoline or arecanut (Hernandez et al. 2017). Areca nut induces cytotoxicity, DNA-protein adduct formation, DNA strand breakage and spontaneous DNA synthesis, further inhibiting DNA repair mechanisms and inducing terminal differentiation (Jeng, Chang, and Hahn 2001). Areca nut chewing is well associated with leukoplakia and oral submucous fibrosis and oral cancer.

2.2.3 Alcohol

Alcohol is a well-known risk factor for oral cancer, which can act individually as well as synergistically with tobacco. Alcohol is metabolised into its active carcinogen acetaldehyde by alcohol dehydrogenase (ADH), and subsequently by aldehyde dehydrogenase 2 to form acetate (Yu et al. 2010). Acetaldehyde is genotoxic in nature and reacts with DNA to form adduct which leads to blocking of DNA synthesis, along with which it also inhibits DNA repair mechanism and causes free radical damage such as mutations (Brooks and Theruvathu 2005). Ethanol consumption can cause local as well as systemic damage to an individual consuming alcohol. Alcohol is known to increase penetration of carcinogens into the oral cavity through the extracellular lipid layer on the oral epithelium which generally acts as a barrier to harmful compounds (Wight and Ogden 1998). Chronic exposure to alcohol also causes epithelial atrophy and tissue damage in the oral cavity depending on the concentration of alcohol (Müller et al. 1983). Enlargement of the parotid gland may occur due to extended exposure to alcohol, which leads to atrophy and lipomatous transformation of parenchyma cells of the parotid gland resulting in impaired saliva flow, hyposalivation. Hyposalivation will lead to increased exposure of mucosal surface to active carcinogens thereby increasing cancer risk (Seitz et al. 2001).

2.2.4 Human Papillomavirus (HPV)

High risk HPV is the major known causative agent of cervical and anogenital cancers. The attributable risk of high risk HPV 16/18 to oral cancer is 20-25% (Chaturvedi and Chocolatewala 2009). The HPV virus has an ability to specifically insert E1/E2 sequence in the host genome, which eventually disrupts the p53 and pRb tumor suppressor genes and alter cellular proteins involved in carcinogenesis (Tran, Rose, and O'Brien 2007). The p16 and Akt/mTOR pathway is also disrupted by HPV, which acts as another mechanism for inducing carcinogenesis (Tran, Rose, and O'Brien 2007). Of the two high risk HPV strains, HPV16 was detected in 90% HPV positive oral cancer patients, on the other hand HPV18 was detected in 47% PV positive oral cancer patients (Pfister 1996; Kreimer et al. 2005). The prognosis of HPV positive oral cancer cases is better than HPV negative oral cancer cases, along with better response and higher sensitivity to radiotherapy (Ritchie et al. 2003).

2.2.5 Microbiota

In the past decade due to advancement in the next generation sequencing technology (NGS), the importance of role of bacteria in human health and disease have been recognized. The host and microbiota co-evolve as a "super-organism" sharing a symbiotic relationship. However, this relationship can also lead to various diseases, including cancer when disturbed due to intrinsic/extrinsic factors. For example, the role of *Helicobacter pylori* in gastric cancer is extensively studied (Schwabe and Jobin 2013). Similarly, a few oral bacteria such as *Porphyromonas gingivalis* and *Fusobacterium nucleatum* are known to play a role in oral cancer. Working with well characterized cultivable bacteria using traditional or conventional microbiology techniques is a straightforward approach, however, with the emerging knowledge about the role of Viable but Non Culturable (VBNC) bacteria, it has become quite important to identify and study the VBNCs as well. This can be carried out using the NGS sequencing approach, which helps to identify the bacterial composition of a specific microbial sample.

2.3 Next Generation Sequencing

Identifying the exact sequence of nucleotides present in a DNA or RNA molecule is referred to as Nucleic acid sequencing. The completion of the Human Genome Project using Sanger Sequencing in 2003 marked the milestone event in this field of sequencing, which cost ~$3 billion and took approximately 13 years to complete. The most common sequencing method is using the Sanger sequencing method (Grada & Weinbrecht, 2013). For longer and big sequences like genomic DNA and high-throughput experiments, the sanger sequencing is an expensive and time-consuming technique, and that leads to an increased demand for a cheaper and affordable high-throughput sequencing method. Ultimately, this resulted in the development of NGS methods and various companies developed platforms for performing high-throughput sequencing (Ansorge, 2009). NGS platforms can perform massive parallel high throughput sequencing where millions of DNA fragments from a single sample are sequenced at once. Smaller genomes can be sequenced within a day with the help of NGS platforms (Grada and Weinbrecht 2013). The use of the NGS platform is not limited to just knowing the sequence of DNA or RNA. NGS is used in clinical practices to screen for mutations in an individual's genome that could be linked to a particular disease, identify novel mutations, detect mosaic mutations, in oncology to diagnose cancers and also personalised

medicine, used in microbiology to screen for pathogenic bacteria and pathogenic genes as well as to decipher and unravel the human microbiome etc (Behjati and Tarpey 2013).

If we observe the progress in the sequencing method, the Sanger Sequencing is considered as the 1st generation sequencing platform. With the need of better sequencers, the 2nd generation sequencing platforms were developed. The first commercialised 2nd generation sequencer was Roche 454 in 2005, followed by AB SOLiD system in 2006, followed by Solexa Genome Analyzer (now referred to as Illumina) and Ion Torrent in 2010. Each of these systems have their own advantages and disadvantages and depending on the output/aim/need of the study, the sequencing platform can be chosen. With constant upgrading the third generation of sequencing platforms were developed, some of which are PacBio RS and Nanopore sequencers (Liu et al. 2014). Of the above-mentioned platforms, Illumina is the most commonly used sequencing platform used in metagenomic studies.

2.3.1 Illumina Sequencing

Illumina is the current market leader for short base-read sequencing (SBS). The first step in sequencing is library preparation, where sequences complementary to adaptors are added to DNA fragments, following which the template gets adhered to a glass slide fixed with complimentary adaptors. Once the single strands are attached to the glass slide, the DNA fragments are PCR amplified by bridge amplification (Single end or Paired end), resulting in simultaneous amplification of billions of clonal DNA templates which can be sequenced. In each cycle, a cocktail of fluorescent labelled and 3'-blocked dNTPs are added. The complimentary dNTP to each strand gets added and the unbound dNTPs are removed. The base incorporation signals are recorded by sensitive fluorescence by high resolution optical imaging. This process is carried out for read lengths of 75-300bp (Liu et al. 2014; Goodwin, McPherson, and McCombie 2016). Illumina has a wide range of sequencers including iSeq, MiniSeq, MiSeq, NextSeq, HiSeq and NovaSeq. Each of these has their own specifications, such as variable maximum reads per run, maximum output, read lengths, sequencing run time. The applications include small genome sequencing, targeted gene sequencing, long range PRC, amplicon sequencing, 16S metagenomic sequencing, mRNA-Seq gene expression profiling, small RNA analysis, Exome sequencing and whole transcriptome sequencing (Kumar, Cowley, and Davis 2019). The advantages of Illumina sequencers include high accuracy, low cost per base, diverse applications and high output. Some pitfalls include

challenged identification of genetic variants, such as structural variants and inability to resolve repetitive regions of the genome. Illumina sequencers have been used in cancer research, microbial genomics, agrigenomics, complex disease genomics, reproductive health, genetic and rare diseases, and genomics in drug development. (https://sapac.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html)

Table 2.2: Types and Specifications of different Illumina platforms

Platform	Read length (bp)	Throughput	Reads	Runtime	Error	Cost per Gb (US$) (approx)
Illumina MiniSeq Mid output	150 (SE)	2.1-2.4Gb	14-16M	17h	<1%	$200-300
Illumina MiniSeq High output	75 (SE)-150 (PE)	1.6-7.5Gb	22-50M	7-24h	<1%	$200-300
Illumina MiSeq v2	36 (SE)-250 (PE)	540Mb-8.5Gb	12-30M	4-39h	<0.1%	$150-900
Illumina MiSeq v3	75 (PE)-300 (PE)	3.3-15Gb	44-50M	21-56h	<0.1%	$100-250
Illumina HiSeq 2500 v2 Rapidrun	36 (SE)-250 (PE)	9-150Gb	300M-600M	7-60h	<0.1%	$40-230
Illumina HiSeq 2500v3	36 (SE)-100 (PE)	47-300Gb	1.5B- 3B	2d-11d	<0.1%	$45-180
Illumina HiSeq 3000/4000	50 (SE)-150 (PE)	105-750Gb	2.5B	1-3.5d	<0.1%	

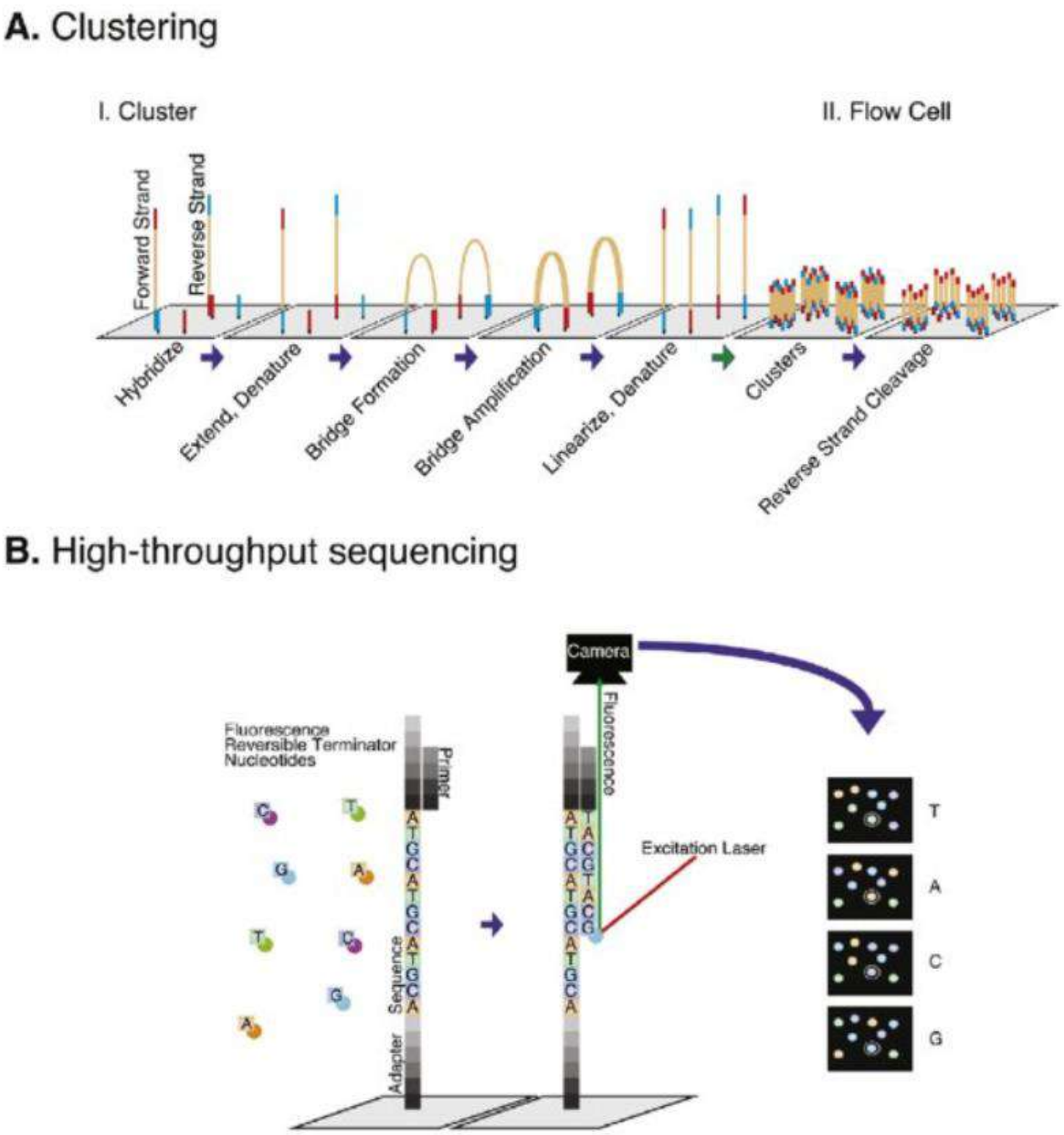


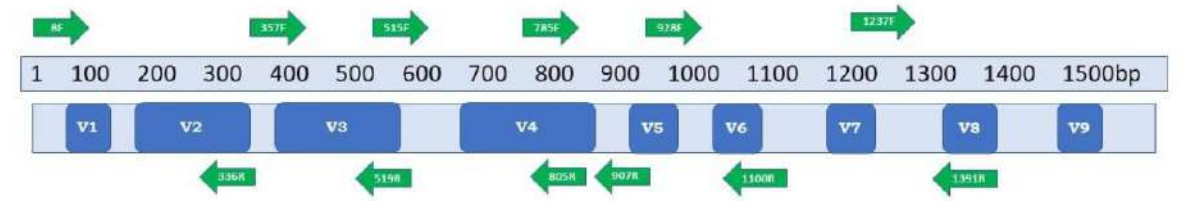
Fig.2.1: Representative image displaying Illumina sequencing by synthesis.

2.3.2 The 16S rRNA gene

The 16S ribosomal RNA (16S rRNA) is the component of the 30S subunit of prokaryotic ribosome (SSU). The 30S is made up of 5S, 16S and 23S rRNAs organized into an operon (Fig.2.3). The 16S rRNA has a structural role and acts as a scaffold to define positions of ribosomal proteins. The 3' end of 16S rRNA contains an anti-Shine-Dalgarno sequence which binds to the Shine-Dalgarno sequence upstream of the start codon of the mRNA. It is a

multicopy gene, *i.e.,* each bacterial species can possess 1-15 operons of the 16S rRNA gene (Rainey et al. 1996; Czernilofsky, Kurland, and Stöffler 1975).

The sequence of 16S rRNA gene has been used widely for phylogenetic analysis and identification of bacteria. The gene is particularly made up of 9 hypervariable regions (V1-V9) and 9 conserved regions. The hypervariable regions consist of sequence diversity among bacterial species which aid in species identification. The variable regions of the gene are generally flanked by conserved regions which can be used for PCR amplification using universal primers. Even though the hypervariable regions have sequence diversity, no single region is proficient to distinguish bacterial species. Due to this, minimum 2-3 variable regions are used to obtain species-level bacterial taxonomy. Of all the regions, V1-V3, V3-V4/V5, V6-V8 are the most commonly used region pairs for assigning taxonomy using NGS (Klappenbach, Dunbar, and Schmidt 2000; Pei et al. 2010; Chakravorty et al. 2007; Case et al. 2007). For amplification of these regions, degenerative primer pairs are generally used; the details are shown in Fig.2.2 below.



Name of primer	Sequence	Name of primer	Sequence
8F	AGAGTTTGATCCTGGCTCAG	336R	ACTGCTGCSYCCCGTAGGAGTCT
357F	CTCCTACGGGAGGCAGCAG	519R	GTATTACCGCGGCTGCTGG
515F	GTGYCAGCMGCCGCGGTAA	805R	GACTACHVGGGTATCTAATCC
785F	GGATTAGATACCCTGGTA	907R	CCGTCAATTCCTTTRAGTTT
928F	TAAAACTYAAAKGAATTGACGGG	1100R	GGGTTGCGCTCGTTG
1237F	GGGCTACACACGYGCWAC	1391R	GACGGGCGGTGTGTRCA

Fig.2.2: Location and sequences of universal PCR primers used to amplify subregions of 16S rRNA gene (Kamble, Sawant, and Singh 2020).

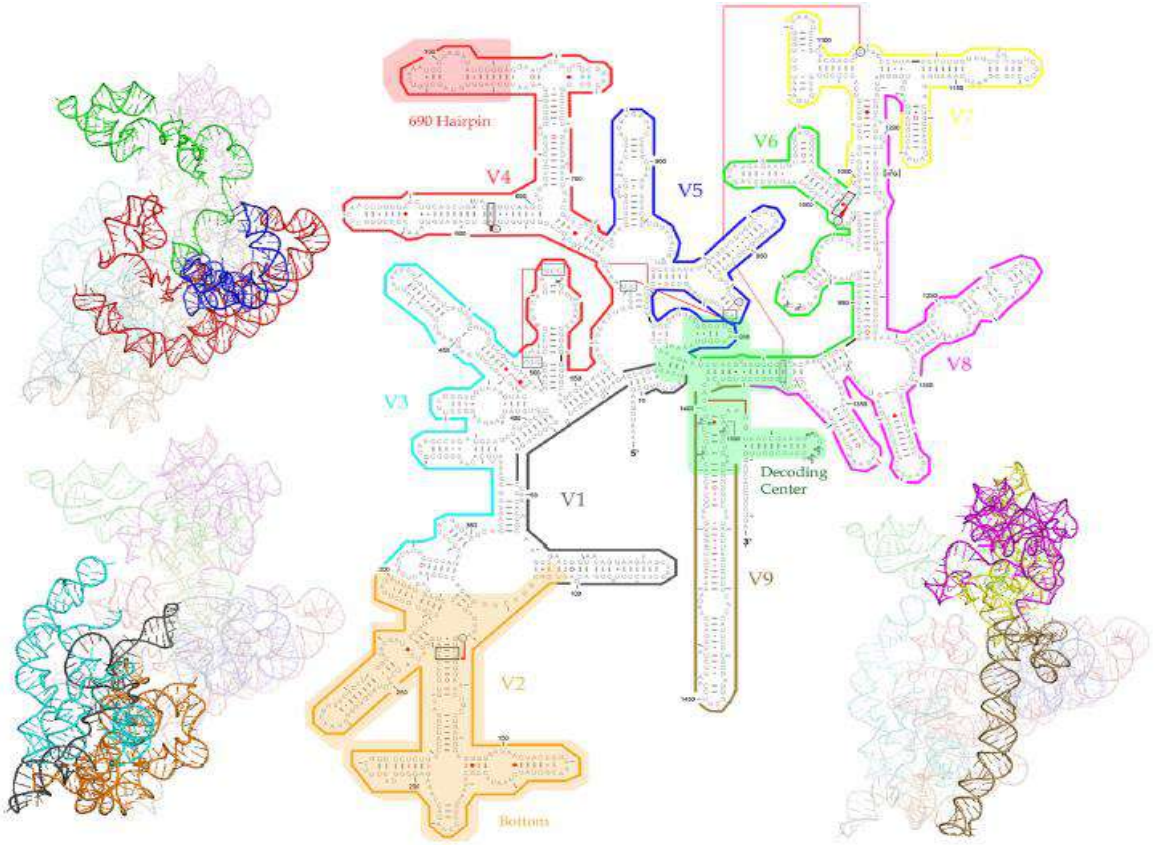


Fig.2.3: The 2D-3D structures of the 16S rRNA gene. Individual regions are identified by the same color in both the 2D and 3D structures(B. Yang, Wang, and Qian 2016)

Since the inception of the HMP (Human Microbiome Project) and other microbiome projects, large numbers of bacterial 16S rRNA sequences have been deposited in various databases dedicated to 16S rRNA gene sequences. The major databases are SILVA (https://www.arb-silva.de), Ribosomal Database Project (RDP) (https://rdp.cme.msu.edu/), GreenGenes (http://greengenes.secondgenome.com) and EzBioCloud (https://www.ezbiocloud.net). The RDP and GreenGenes consist of data for small subunit rRNA for Archaea and Bacteria, whereas SILVA also includes Eukaryota in addition to the two mentioned domains. RDP database uses Bergeys taxonomy which is the standard for microbiological taxonomies, whereas GreenGenes and SILVA use taxonomies developed by their own developers. GreenGenes database is based on *de novo* phylogenetic tree whereas SILVA database is based on hidden Markov model-based rRNA gene prediction (Quast et al. 2013; McDonald et al. 2012). The choice of database selection is the researcher's personal choice depending on the type of analysis expected out of a study. The SILVA database is the largest database but not well curated, which increases the chance of taxonomy assessment but has chances of errors.

On the other hand, RDP is the smallest database although it is based on the standard curated manual of microbiology which increases the confidence in results obtained but increases the chances of losing out the novel bacteria, which cannot be assigned taxonomy due to smaller database size.

2.3.3 QIIME2

The data obtained from NGS sequencers need to be analysed using pipelines designed to process the raw data and result in taxonomic classification of the sequences. One of the pipelines used in this study is QIIME2, which is an open source and developed for easy processing, analysing, and visualizing microbiome data, and can be used in multiple ways depending on the data analysis and experiment goals. QIIME2 can indeed process data from other microbiome indicators such as 18S rRNA, internal transcribed spacers (ITS), and cytochrome oxidase I (COI), as well as shotgun omics and untargeted metabolomics.

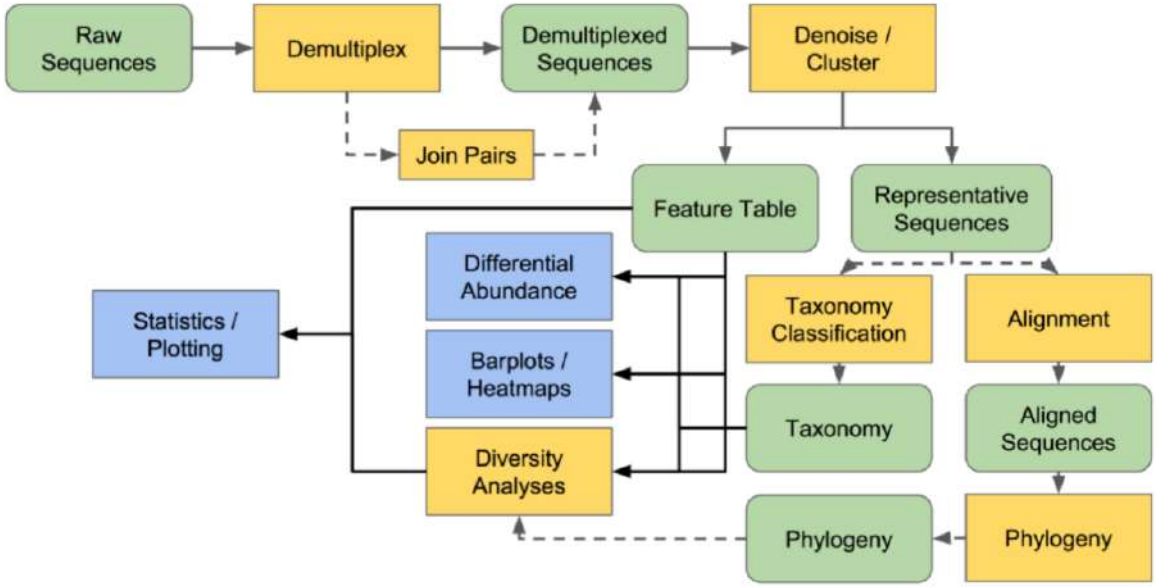


Fig.2.4: The QIIME2 workflow

QIIME2 can be installed on any computer system including MacOS, Linux or Windows which provides user flexibility for the use of the pipeline (Fig. 2.4). The additional data and variables related to study called the metadata such as sample characteristics, collection site, sample processing can be included for better analyses. The sequenced DNA data has to be imported into QIIME2, the command line for which depends on multiple factors such as if the files are single-end or paired-end or multiplexed or demultiplexed, barcode files etc. Once the data

has been imported, the quality of sequencing is assessed using a XY plot which describes the base pairs on X axis and Q (Phred) score on Y axis as well as average sequence length for the sample. In general, Q< 20 depicts minimum 99% base call accuracy and any Q value above 20 is acceptable for further downstream analysis. Followed by visualization of quality of base pairs, the quality control is performed by trimming and filtering sequences based on the Q score followed by clustering of sequences into the operational taxonomic units (OTUs) based on similarity (~97%) called as denoising. This can be performed in QIIME2 using 2 plugins namely DADA2 and Deblur, although not much difference has been reported in both the methods. The output obtained is a descriptive table and a file containing the representative sequences associated with a feature ID, which will be used for taxonomic assessment. The table output describes the mean and median number of features present in each sample as well as total number of features in the entire dataset, the number of samples in which a particular feature is present, and the number of features present per samples.

Once all the sequences are obtained, the next step is to classify these sequences using a taxonomy classifier by choosing a database against which the sequences generated will be matched against. The most commonly used databases are GreenGenes, RDP (Ribosomal Database Project) and SILVA. The choice of database varies from user to user. This results in each sequence being annotated by a bacterium depending on the similarity of the region of 16S rRNA amplified. Once the sequences are mapped against one of the databases and the taxonomy is being allotted, the abundance of each of the bacteria in each sample can be analysed and visualized by the taxa bar plots.

On the other hand, the table output generated after denoising can be used for generating alpha and beta diversity plots as well as phylogenetic trees. The alpha diversity plots include the rarefaction plots, where the sampling depth can be studied for the sample's sequences using indices such as Shannon index, OTUs generated, Simpson index, and ACE index. Similarly, the data can also be analysed to obtain beta matrices which include Bray-Curtis matrix, Jaccard matrix, Unweighted and Weighted UniFrac distance matrices.

2.4 Oral microbiota

2.4.1 The healthy oral microbiota

The human body is colonized by 10-100 trillion of symbiotic microbial cells, called the human microbiota. Please note that microbiota refers to the bacterial collection, whereas the microbiome represents the collective genome sequence of all microbes together in a sample (Weinstock, 2012). This microbiota includes bacteria, fungi, virus, archaea and protozoa. Most of these microbes typically live in harmony with the human host cells sharing a symbiotic relationship (Olsen 2015). With the increasing importance of human microbiota being discovered, it has even been termed as an 'essential organ' (O'Hara and Shanahan 2006). The microbiota is crucial to human health via several mechanisms. The microbiota has potential for increased energy extraction from food consumed. There are approximately 150 times more genes that microbiota contribute to the host which aid the host with special enzymes and biochemical pathways. The host microbiota also provide physical barrier as a protection against pathogenic microorganisms (B. Wang et al. 2017). The gut and oral microbiota are the most widely studied till date.

The NIH Human Microbiome Project (HMP) marked the initiation mark towards microbiome studies. NIH HMP initiated in 2007 aimed to identify the core microbiome of a healthy human being. Various individuals across the globe participated in the study and four major body organs were targeted for the study which include the gut, skin, oral cavity and the vagina. Although this was a huge breakthrough, often the taxonomic composition of the samples did not correlate with the host phenotype. This gave rise to the advanced version of HMP, termed as Integrative HMP which took into account the microbial composition as well as to gain a more holistic view of host-microbe interactions over time, including immunity, metabolism, and dynamic molecular activity (Proctor et al. 2019). This detailed analysis of the human microbiota composition was possible only due to the advent of next generation sequencing (NGS) technology. NGS technology made it easy and possible to identify the microbial composition in a sample which consisted not only of the viable and culturable microbes, but viable and non-culturable microorganisms as well.

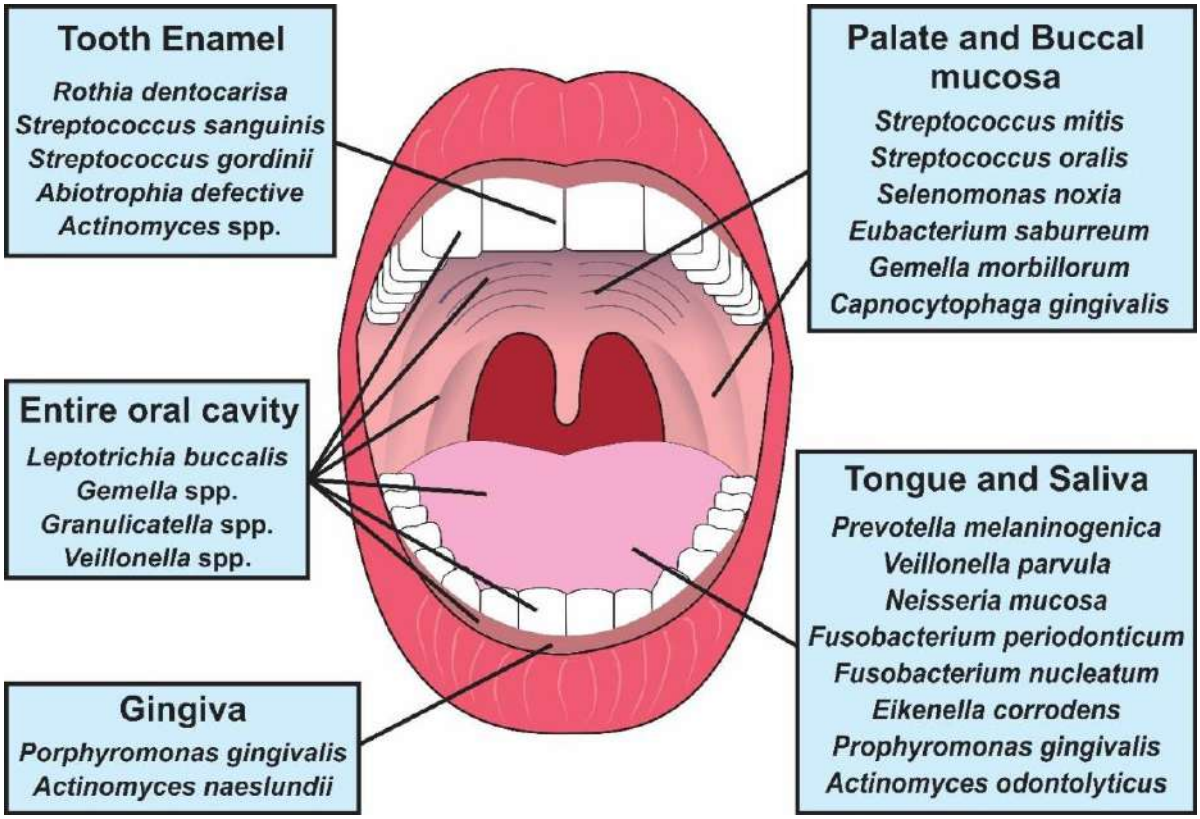


Fig.2.5: Dominant microbiota in different parts of the oral cavity

Someone has well said “Unlike Vegas, what happens in the oral cavity, does not stay in the oral cavity”. The oral cavity is a crucial part of the digestive system composed of both hard as well as soft tissues and the teeth. The entire oral cavity is covered by oral mucosa consisting mainly of a layer of squamous epithelium. The entire oral cavity consists of a wide array of microorganisms including the bacteria, virus, and fungi (Deo & Deshmukh, 2019). Massive numbers of these microbes are known to form biofilms in the oral cavity and the variation of surfaces in the oral cavity aid in the biofilm formation process. Because of the biofilm protection, these bacteria show resistance to antibiotic treatments, as well as mechanical stress caused by daily activities of an individual.

Even within the oral cavity, different parts have displayed a variety in the microbiome composition (Fig. 2.5). One of the early studies published by Simón-Soro et al in 2013 explored the variations of the oral microbiome based on the specific location of samples collected within the oral cavity including tooth surface, gingiva, saliva and tongue(Simón-Soro et al. 2013).

Table 2.3: Dominant microbiota based on parts of the oral cavity

Part of oral cavity	Dominant microorganism
Hard palate	*Streptococcus, Uncl.Pasteurellaceae, Veillonella, Prevotella, Uncl.Lactobacillales*
Tongue dorsum	*Streptococcus, Veillonella, Prevotella, Uncl. Pasteurellaceae, Actinomyces*
Saliva	*Prevotella, Streptococcus, Veillonella, Uncl. Pasteurellaceae*
Palatine tonsils	*Streptococcus, Veillonella, Prevotella, Uncl. Pasteurellaceae, Fusobacterium*
Throat	*Streptococcus, Veillonella, Prevotella, Uncl. Pasteurellaceae, Actinomyces, Fusobacterium, Uncl. Lactobacillale*
Buccal mucosa	*Streptococcus, Uncl. Pasteurellaceae, Gemella*
Keratinised gingiva	*Streptococcus, Uncl. Pasteurellaceae*
Supragingival plaque	*Streptococcus, Capnocytophaga, Corynebacterium, Uncl. Pasteurellaceae, Uncl. Neisseriaceae*
Subgingival plaque	*Streptococcus, Fusobacterium, Capnocytophaga, Prevotella, Corynebacterium*
Dentures	*Staphylococcus epidermidis, Streptococcus*
Lips	*Streptococcus, Candida albicans*

(Adapted from (Jia et al. 2018))

Interestingly, the oral microbiome of a healthy individual depends on multiple factors and therefore the exact oral microbiome composition for a population cannot be defined or standardized, however a pattern can be found based on major populations and specific changes The factors affecting an individual's oral microbiome composition include intrinsic factors such as hormonal changes, alteration of oral mucosa physiology/anatomy, decrease of salivary antibodies, deficiency of cell-mediated immunity, age and extrinsic factors such as denture wearing, medication, diet, systemic diseases, socioeconomic background, and lifestyle factors such as tobacco chewing, smoking, and alcohol consumption (Samaranayake and Matsubara 2017). The type of diet, micronutrients, sugars, and vitamins consumed can cause a shift in the oral microbiome. Increased abundance of *Streptococcus mutans* and

Fusobacterium nucleatum has been observed in cases of childhood caries pertaining to consumption of high sugar-containing snacks. Vitamin intake such as Vit B/E/C has been linked to increased Fusobacteria population. A moderately healthy diet is advised to maintain oral ecological balance (Murshid 2014; Tanner et al. 2011). Smoking can alter the oral microbiota in unforeseen ways. The cigarette itself contains bacteria such as *Bacillus* spp and *Clostridium* spp carried from the environment which have the potential to survive high temperatures from smoking. Smoking also promotes creation of anaerobic conditions which enhances the growth of facultative/strict anaerobes such as *Streptococcus, Veillonella and Actinomyces*(J.-H. Moon, Lee, and Lee 2015; Bizzarro et al. 2013). Consumption of alcohol by an individual contributes distinctly to the microbial population. Oral bacteria are known to metabolize alcohol into acetaldehyde, a known potent carcinogen to humans, using bacterial enzymes like alcohol dehydrogenase (Kurkivuori et al. 2007). Studies report that these lifestyle differences were reflected inthe bacterial profiles of saliva. *Megasphaera micronuciformis, Veillonella atypical, Veillonella parvula, Rothia mucilaginosa, Prevotella histicola, Fusobacterium periodontium, Granulicatella adiacens* and *Tannerella forsythia* were abundant in the high socioeconomic status group, while *Aggregatibacter segnis, Achromobacter xylosoxidans* and *Neisseria* cluster ii were abundant in the low socioeconomic group (Belstrøm et al. 2014).

Although, at the genus and species level, the microbial diversity changes depending on the above mentioned factors, the major phyla of most individuals oral microbiota composition remains more or less same,which includes Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria and Proteobacteria (Schmidt et al. 2014). At genus level the most common genera observed are *Prevotella, Neisseria, Streptococcus, Fusobacterium, Haemophilus, Capnocytophaga, Veillonella, Alloprevotella, Porphyromonas, Leptotrichia, Aggregatibacter, Selenomonas, Campylobacter, Granulicatella, Actinomyces, Gemella, Lachnoanaerobaculum,* and *Bergeyella* (Zhao et al. 2017b). As mentioned above, the abundances of the mentioned phyla and genera is multi-factor dependent and changes based on an individual/population/lifestyle habit.

2.4.2 Oral microbiota in Oral cancer

One of the first reports published related to the oral microbiome and oral cancer was by Nagy et al. who analysed the biofilm present on the surface of OSCC tissue and healthy contagious

mucosa from 21 patients. They observed an increase in median colony forming units (CFU/ml) of anaerobic as well as aerobic organisms at the tumor sites. The tumor sample harboured increased numbers of anaerobes such as *Veillonella, Fusobacterium, Prevotella, Porphyromonas, Actinomyces* and *Clostridium*, and aerobes such as *Haemophilus, Serratia liquefaciens, Klebsiella pneumoniae Citrobacter freundii, Enterobacteriaceae* and *Streptococcus β-haemolyticus*. The presence of *Candida albicans* was also reported in 8 of 21 tumor samples (Nagy et al. 1998). *S. anginosus* was quantified in the oral cancer tissues using real time Polymerase chain reaction (PCR). 13% of the total oral cancer and 44% of the total esophageal cancer samples showed detectable levels of *S. anginosus* as compared to healthy controls, depicting an increased correlation between *S. anginosus* and the tumor presence (Morita et al. 2003). Sasaki et al. (2005) analysed the link between *S. anginosus* and tumor formation in 46 oral cancer samples. *S. anginosus* was detected in 45% of oral cancer tissue specimens when detected using species-specific PCR assay and pulsed field gel electrophoresis techniques. Interestingly, this genotype was also present in the dental plaque of the same patients, suggesting dental plaque to be a dominant reservoir of *S. anginosus* (Sasaki et al. 2005).

Salivary counts of 40 common oral bacteria were estimated in 45 OSCC samples using DNA-DNA hybridization. Counts of 3 of 40 tested species, namely *Capnocytophaga gingivalis, Prevotella melaninogenica* and *Streptococcus mitis* were elevated in OSCC samples ($p < 0.001$). These elevated counts were observed in 80% of OSCC cases suggesting they could act as indicators for OSCC diagnosis. On the other hand, *C. ochracea, L. buccalis* and *E. saburreum* counts were elevated in OSCC free samples but no further analysis was carried out to confirm the output of the study (Mager et al. 2005). Hooper and colleagues analysed 20 OSCC tumor samples using 16S rRNA sequencing. *Exiguobacterium oxidotolerans, Prevotella melaninogenica, Staphylococcus aureus, Veillonella parvula*, and species of bacteria and micrococcus were isolated only from tumor samples (Hooper et al. 2006). In another study, 52 unique phylotypes were identified from tumorous tissues, most of which were saccharolytic and aciduric species. Organisms *Clavibacter michiganensis subsp. tessellarius, Fusobacterium naviforme* and *Ralstonia insidiosa* were found in more than 30% tumorous samples (Hooper et al. 2007).

In a pilot study by Pushalkar et al. (2011), Denaturing gradient gel electrophoresis (DGGE) and microbiome analyses was performed using 16S rRNA sequencing of saliva samples collected from three OSCC cases and compared to two matched controls. OSCC samples predominantly contained organisms from the genera *Streptococcus, Gemella, Rothia, Peptostreptococcus, Porphyromonas,* and *Lactobacillus* (Pushalkar et al. 2011). Later the researchers elucidated the oral microbiota from ten OSCC tumor samples OSCC using the same sequencing technique (Pushalkar et al. 2012). They found phylum Firmicutes, Class Bacilli consisting te Order *Lactobacillales* (54.8%) and *Bacillales* (11.8%) to be predominant in the tumor tissues. The tumor library was predominated by *Streptococcus* (50.8%), *Gemella* (11.6%), *Parvimonas* (4.6%), *Peptostreptococcus* (2.8%), *Xanthomonas* (2.4%), *Johnsonella* (1.6%), *Solobacterium* (1.6%), *Atopobium* (1.2%). *Eubacterium*[11][G-3], C*ampylobacter* and *Catonella* were exclusive to the tumor library which is a new panel of organisms to be associated with OSCC from the yet discussed studies, and therefore would need further analyses to eliminate any bias emerging from the samples.

Another microbiological study to observe the difference in microflora of OSCC swab samples as compared to oral mucosa lesions was performed by Cankovic et al. (2013). The microbiological assays depicted increased pathological flora and decreased normal flora in samples of the OSCC group as compared to the control group, a common observation found in most of the studies reported (Čanković et al. 2013a). Schmidt and colleagues studied the bacterial abundance in oral cancer samples (Schmidt et al. 2014). In both the studies, presence of Firmicutes and Actinobacteria decreased in the oral cancer samples as compared to healthy counterparts. Further, pre-cancerous samples also showed a similar decrease. Such studies suggest that alterations in microbiota may occur very early on in cancer development during pre-cancer stages. At genus level, significant reduction in the abundance of *Streptococcus* and *Rothia,* and a corresponding increase in *Fusobacterium* was reported in OSCC samples compared to their matched controls. Thus, such organisms can be used as microbial biomarkers to study the progression of cancers.

Bolz *et al.* performed a microbiological study to isolate oral bacteria from oral cancer, high risk individuals and control patients. The ratio of aerobes to anaerobes was found to be~1:1.6 in control samples, equal ratio (1:1) in high risk patients whereas the tumor patients displayed a reversal of ratio (0.5:1). This depicts an increase in abundance of anaerobic bacteria in tumor

samples (Bolz et al. 2014). This could be due to dysbiosis and physiological changes during tumorigenesis which favours the dominance of anaerobes. A similar microbiological study reported an increased median CFU/ml in saliva and swab of the oral cavity obtained from OSCC patients as compared to healthy controls. Increased abundance of *Candida* spp, particularly *Candida albicans, was* isolated from the carcinoma site (Metgud *et al.* 2014). *Candida albicans* under host immunocompromised state is capable of nitrosation to form nitrosamines which are capable of initiation of carcinogenesis and further development of oral carcinoma (Sanjaya et al. 2011).

In a pilot study, three OSCC samples obtained from biopsies were subjected to NGS. The samples were predominated with bacteria belonging to genera *Haemophilus, Neisseria, Prevotella, Fusobacteria, Streptococcus, Porphyromonas, Leptotrichia,* and *Aggregatibacter*. At species level, *Prevotella oris, Neisseria flava, N. flavescens/subflava, Fusobacterium nucleatum ss polymorphum, Aggregatibacter segnis, Streptococcus mitis, Fusobacterium periodonticum, Neisseria elongata, Porphyromonas sp. oral taxon 279,* and *Alloprevotella tannerae* were the most common species identified from the samples (Al-Hebshi *et al.* 2015). Although the sample number was low, this was one of the first reports studying the bacterial diversity in oral carcinoma using NGS technology and reporting the highest number of species-level taxa. Further, the diversity of oral microbiota in HNSCC was reported by Rafael et al. In HNSCC, phyla *Firmicutes* (67%) were predominant as compared to phyla *Firmicutes* (47.1%), *Bacteroidetes* (21.2%) and *Proteobacteria* (22.7%) in control samples. *Streptococcus, Peptostreptococcus* and *Tanerella* were higher in abundance in tumor samples as compared to *Agregatibacter, Lautropia, Haemophillus, Neisseria* and *Leptotricha* which had higher abundance in control samples. Using LefSe, significant increase in *Lactobacillus, Streptococcus , Staphylococcus* and *Parvimonas* was observed in HNSCC samples, whereas *Haemopilus, Neisseria, Gemellaceae* or *Aggregatibacter* was increased in control (Guerrero-Preston et al. 2016). Overall, the data suggests increased abundance of *Lactobacillus* and decreased abundance of *Haemopilus, Neisseria, Gemellaceae* or *Aggregatibacter* in saliva, although the number of samples were very low to categorise these bacteria as biomarkers for diagnosis of HNSCC. A similar study with a larger sample number would be helpful in understanding the abundance and role of the mentioned bacteria towards OSCC.

Mok *et al.* (2017) studied the oral microbiome diversity on oral premalignant potential disorders (OPMD) and OSCC. The relative abundance of *Streptococcus* and *Veillonella* was higher in control samples as compared to *Neisseria, Gemella* and *Granulicatella* was higher in-patient group. In this study, the researchers identified oral microbes associated with OPMD. Microbes such as *Megasphaera micronuciformis, Prevotella melaninogenica* and *P. veroralis* were identified as potential biomarkers for progression towards oral cancer. These organisms have been identified and isolated from various other sources related to the oral cavity, but first time in OPMD samples. The species richness was highest in the OPMD group as compared to both cancer and healthy samples (Mok et al. 2017). The higher species richness could be observed because OPMD being a transitional phase between healthy and oral cancer, the microbiome population in these individuals would be a blend between the two groups leading to an increased bacterial diversity. *Streptococcus infantis* and *Streptococcus anginosus* also depicted increased abundance in OPMD-betelnut chewers which could also act as biomarkers for the same (Hernandez et al. 2017). In a similar study, genus *Alistipes, Bacteroides, Blautia, Clostridium, Dorea, Escherichia, Faecalibacterium, Megamonas,* and *Phascolarctobacterium* displayed a positive correlation in oral lesions and oral cancer group. Similarly, significant differences were also observed with increased abundance of genera *Anaerostipes, Barnesiella, Butyricimonas, Cloacibacillus, Cloacibacterium, Collinsella, Coprococcus, Eggerthia, Erysipelothrix, Gemmiger, Klebsiella*, *Oscillospira,* and *Roseburia* in cancer and lesional group (Lee et al. 2017). Since these genera are not very commonly known and newly associated with diseased groups, it is necessary to carry out more similar studies to confirm the role of these genera and their functional significance in oral cancer pathogenesis.

In 2018, Yang *et al.* studied the dynamics of oral microbiota in OSCC staging. At phyla level, the abundance of *Fusobacteria* significantly increased from controls (2.98%), to OSCC stage 1(4.35%) to stage 4 (7.92%), whereas abundance of phyla *Bacteroidetes* and *Actinobacteria* significantly decreased in OSCC samples. The abundance of 5 most predominant genera in OSCC stage 4 were *Streptococcus* (28.21%), *Veillonella* (11.01%), *Neisseria* (9.65%), *Haemophilus* (9.37%), and *Rothia* (4.42%) as compared to *Streptococcus* (35.57%), *Haemophilus* (12.30%), *Veillonella* (11.56%), *Neisseria* (8.39%), and *Rothia* (5.12%) in control samples. Using LEfSe analysis, *Haemophilus parainfluenzae, Porphyromonas pasteri,*

Veillonella parvula, and *Actinomyces odontolyticus* were associated with the healthy controls as compared to *Fusobacterium periodonticum, Parvimonas micra, Streptococcus constellatus, Haemophilus influenza*, and *Filifactor alocis* in OSCC Stage 4. The decrease in abundance of genus *Streptococcus,* which is associated with commensal oral bacteria in OSCC, could be because of an increase in colonization of other opportunistic and pathogenic bacterial species in the resulting oral conditions (Yang *et al.* 2018). Similarly, the oral microbiome of HPV positive OSCC was recorded. The study displayed positive correlation of *Haemophilus* and *Gemella* with HPV infection. This study also concurs with the previous findings of various studies which suggest *Actinobacillus, Actinomyces, Aggregatibacter, Capnocytophaga, Fusobacterium, Oribacterium, Rothia, Haemophilus, Leptotrichia, Neisseria, Porphyromonas* and *Veillonella* could be potential biomarkers for OSCC. This study reported the decrease in abundance of *Paludibacter* and *Corynebacterium* in OSCC (Lim et al. 2018). Although, the study supports previously stated data from multiple studies, the lack of larger dataset can be a limitation for confidence in this study. Reports published in 2021 reveal the changes in microbiota in oral cancer and oral leukoplakia as compared to control samples. Their major findings support the increased abundance of genus *Granulicatella* and *Porphyromonas* in oral cancer group as has been stated before. Moreover, the data describes the association of *Salmonella, unclassified Enterobacteriae, Prevotella* and *Megasphaera* with the oral leukoplakia group. This study can provide a link to the association of oral dysbiosis in oral leukoplakia and oral cancer in the Indian population (Gopinath et al. 2021). Major findings of all the above mentioned studies are compiled and shown in a tabular format in Table 2.4.

Table 2.4 : List of oral cancer microbiota studies and their major findings

Study	**Number of Oral Cancer samples**	**Technology used**	**Sample used**	**Key bacteria associated with oral cancer**
(K. N. Nagy et al. 1998)	21	Microbiological	Tumor lesion	*Veillonella, Fusobacterium, Prevotella, Porphyromonas, Actinomyces, Clostridium, Haemophilus, Serratia liquefaciens, Klebsiella pneumoniae Citrobacter*

				freundii, Streptococcus β-*haemolyticus*, *Candida albicans*
(Morita et al. 2003)	19	16S rDNA qPCR	Tumor tissues	*S. anginosus*
(Sasaki et al. 2005)	49	Species specific PCR and Pulsed Field Gel Electrophoresis	Tissue biopsy, Dental plaque and saliva	
(Mager et al. 2005)	45	Checkerboard DNA-DNA hybridization	Unstimulated saliva	*C. gingivalis, P. melaninogenica, S. mitis may*
(Samuel J. Hooper et al. 2006)	20	16S rRNA sequencing (ABI PRISM 3100 Genetic Analyzer; Applied Biosystems) , Microbiological assay	Tumor tissues	*Exiguobacterium oxidotolerans, Prevotella melaninogenica, Staphylococcus aureus, Veillonella parvula, Micrococcus, S. anginosus,*
(S J Hooper et al. 2007)	10	FISH, 16S rRNA sequencing (ABI PRISM 3100 Genetic Analyzer; Applied Biosystems)	Tumor tissues	*Clavibacter michiganensis subsp. tessellarius, Fusobacterium naviforme* and *Ralstonia insidiosa*
(Pushalkar et al. 2011)	3	DGGE, 16S rRNA sequencing (Genome Sequencer GS FLX system)	Saliva	*Streptococcus, Gemella, Rothia, Peptostreptococcus, Porphyromonas, Lactobacillus*
(Pushalkar et al. 2012)	10	DGGE, 16 rRNA sequencing (Sanger sequencing)	Tumor tissues	*Streptococcus, Gemella*, *Parvimonas*, *Peptostreptococcus*, *Xanthomonas*, *Johnsonella*, *Solobacterium*, *Atopobium, Eubacterium[11][G-3], Campylobacter, Catonella*
(Čanković et al. 2013)	30	Microbiological	Swab samples	Decreased normal flora, increased pathogenic flora

(Metgud, Gupta, and Gupta 2014)	30	Microbiological	Swab samples	Higher median CFU/ml in cancer samples
(Schmidt et al. 2014)	15	16S rRNA sequencing (Roche 454 Life Science Sequencer)	Swab samples	*Fusobacterium* spp.
(Bolz et al. 2014)	30	Microbiological	Swab samples	Increased anaerobic bacteria
(Al-Hebshi et al. 2015)	3	16S rRNA sequencing (Roche 454 GS FLX-B)	Tumor biopsy	*Prevotella oris, Neisseria flava, N. flavescens/subflava, Fusobacterium nucleatum ss polymorphum, Aggregatibacter segnis, Streptococcus mitis, Fusobacterium periodonticum, Neisseria elongata, Porphyromonas sp. oral taxon 279, Alloprevotella tannerae.*
(Guerrero-Preston et al. 2016)	19	qPCR, Roche/454 GS Junior pyrosequencing platform	Tumor biopsy, Saline wash	*Lactobacillus, Streptococcus , Staphylococcus, Parvimonas, Tanerella*
(Mok et al. 2017)	9	DGGE, 16S rRNA sequencing	Swab sample	*Neisseria, Gemella* and *Granulicatella*
(Zhao et al. 2017a)		16S rRNA sequencing (Illumina MiSeq)	Swab sample	*Fusobacterium, Dialister, Peptostreptococcus, Filifactor, Peptococcus, Catonella, Parvimonas*
(Lee et al. 2017)	125	16S rRNA sequencing (Illumina MiSeq)	Saliva sample	*Anaerostipes, Barnesiella, Butyricimonas, Cloacibacillus, Cloacibacterium, Collinsella, Coprococcus, Eggerthia, Erysipelothrix, Gemmiger, Klebsiella, Oscillospira,* and *Roseburia*
(Lim et al. 2018)	31	16S rRNA sequencing (Illumina MiSeq)	Oral rinse	*Actinobacillus, Actinomyces, Aggregatibacter, Capnocytophaga,*

				Fusobacterium, Oribacterium, Rothia, Haemophilus, Leptotrichia, Neisseria, Porphyromonas, Veillonella
(C. Y. Yang et al. 2018)	197	16S rRNA sequencing (Illumina MiSeq)	Oral rinse	*Fusobacterium periodonticum, Parvimonas micra, Streptococcus constellatus, Haemophilus influenza, Filifactor alocis*
(Gopinath et al. 2021)	NA	16S rRNA sequencing (Illumina MiSeq)	Unstimulated whole mouth fluid	*Megasphaera, unclassified Enterobacteria, Salmonella, Prevotella, Porphyromonas*

2.4.3 Oral microbiota in tobacco chewers

There are limited reports published about the oral microbiome composition in smokeless tobacco users. A recent study has published, during the period of present study, by Srivastava *et al.* that have identified that Proteobacteria (25.836%) dominates over Firmicutes (23.844%) among SLT users followed by Bacteroidetes (19.664%), Fusobacteria (3.222%), Actinobacteria (5.392 %), Candidate division _GN02 (0.081 %), Spirochaetes (0.363 %), Candidate division _SR1 (0.3059 %), and Candidate division _TM7 (0.197 %). At genus level, the dominant genera were *Neisseria, Streptococcus* and *Prevotella* followed by *Hemophilus, Porphyromonas, Rothia, Granulicatella, Actinobacillus , Veillonella,* and *Fusobacterium*. On comparing microbiome of long term tobacco chewers with healthy individuals, it was found that genus *Fusobcaterium, Porphyromonas, Parvimonas, Gordonia, Moryella, Catonella, Microbacterium, Desulfobulbus, Enterococcus* were abundant in smokeless tobacco users (Srivastava, Mishra, and Verma 2021). In another study performed, the effect of shammah (a smokeless tobacco form used mainly in the Middle East) was explored. Six species were found to be enriched in shammah users namely, *R. mucilaginosa*, *Streptococcus sp. oral taxon 66, A. meyeri, S. vestibularis, S. sanguinis and a potentially novel Veillonella species* (Halboub et al. 2020). Yet another report published on the use of Snus (a smokeless tobacco product used in Sweden and Latvia) using traditional microbiological and RT-PCR method. Bacterial

concentrations of *Aggregatibacter actinomycetemcomitans* (>104), *Tannerella forsythia* (>105), *Treponema denticola* (>104) in periodontal pockets were higher than in the control group. Bacterial concentration of *Porphyromonas gingivalis* (< 105)(p = 0.42) was acceptable in periodontal pockets. Presence of *Prevotella intermedia* was negative in both the groups (Miļuna et al. 2017).

2.5 Real time quantitative Polymerase chain reaction for absolute quantification (qPCR)

2.5.1 Quantitative PCR using SYBR green chemistry

A real-time PCR is the continuous collection of fluorescent signals from one or more polymerase chain reactions over a range of cycles. Quantitative real-time PCR (qPCR) is the conversion of the fluorescent signals from each reaction into a numerical value for each sample. A qPCR is widely used in microbial ecology to determine gene copies/transcripts in a sample. The accuracy of the data obtained depends on the selection of target genes , a crucial part for quantitation which. Once the gene is selected, the primers specific to the gene are designed (VanGuilder, Vrana, and Freeman 2008). The primers determine the target specificity and sensitivity in a PCR reaction. The choice of target gene depends on the study objective, - whether to amplify a single organism or a group of organisms. If a community or entire load of microbial population has to be amplified and quantitated, a gene that is more conserved throughout the community is used for quantitation keeping in mind that gene would be universally present in all organisms, for example, the 16S rRNA gene. On the other hand, if a specific organism has to be identified, a gene that is organism specific can be used to design primers, for example, the *rpoB* gene (Case et al. 2007).

In a qPCR reaction, a range of fluorescent technologies can be used that include Taqman, Lightcycler, LUX, Molecular Beacons, and SYBR Green, out of which, Taqman and SYBR green, both utilise different technologies for fluorescent signalling, are the most commonly used methods. The SYBR green technology is a model that utilises a fluorescent intercalating dye. The dye fluoresces on binding to PCR amplified double-stranded DNA and the fluorescent signal can be easily detected. Intercalating dyes are often inexpensive and can be used in any reaction and do not need sequence specific probes that are comparatively expensive. The SYBR dye binds to double-stranded DNA in a non-selective fashion, including primer dimers (Zipper *et al.* 2004) , so the SYBR green chemistry requires a post amplification dissociation melt curve analysis. In the melt-curve dissociation step, the amplified products are subjected

to an increasing temperature gradient, where the products dissociate and result in the formation of a melt curve that can be used to confirm the specific product is amplified. Each amplified product has a single melting temperature (Tm) and any extra peak(s) identified in a melt curve step hints towards unwanted or non-specific amplification products.

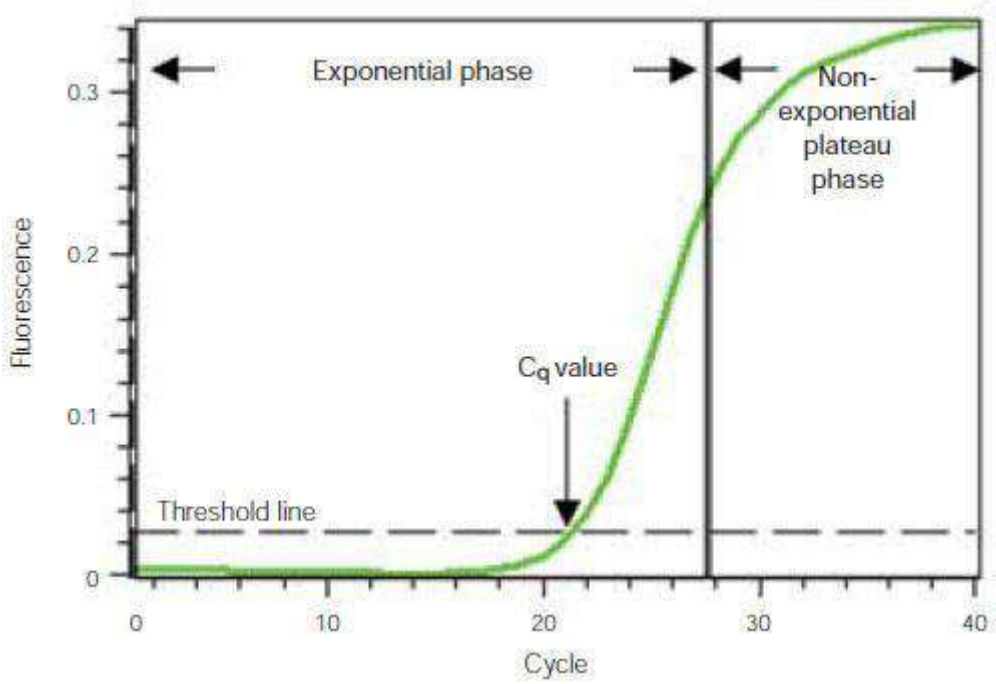


Fig.2.6: Phases of real time qPCR amplification curve

The most important step in the data analysis of qPCR assay utilizes the knowledge of crossing threshold (Ct) of a sample obtained. The Ct is the number of cycles that it takes each reaction to reach an arbitrary amount of fluorescence which can be detected by the machine (Fig. 2.6). This means, more the initial amount template of the DNA to be amplified, lesser is the number of amplification cycles required to reach the threshold fluorescence; *i.e.* higher amount of template in the initial reaction, lower would be the Ct and vice versa. This Ct value helps to identify the original template copies of the gene of interest in the sample (VanGuilder, Vrana, and Freeman 2008).

2.5.2 qPCR studies in microbial ecology

Till date, numerous qPCR assays have been performed to understand and quantitative microbial ecology of a sample including environmental, industrial as well as clinical samples. Most of these studies utilise the well-known and well characterised 16S rRNA gene which is an accepted taxonomic marker in the bacterial domain (Haugland et al., 2010; Ritalahti et al.,

2006). As mentioned earlier, the 16S rRNA gene consists of hypervariable as well as conserved regions that can be explored for such investigations. For example, in studies determining the total bacterial load in a sample, the conserved region of the gene can be used to design primers. This enables all the bacterial communities present in the sample to be amplified, since the conserved region is conserved in all bacterial species. Alternatively, if specific bacterial species have to be amplified, the hypervariable regions can be utilized to design primers. This is possible if the annotated sequence of the 16S rRNA for the specific species of interest is available. For example, in a study conducted by Lyons *et al.* in 2000, the dental plaque was examined to study the proportion of *P. gingivalis* as compared to the total bacteria present in the sample. For amplification of all bacteria present in the sample, the authors used universal primers 785F and 422R which target the conserved region of 16S rRNA that aids in amplification of all bacterial species, on the other hand, for specific amplification of *P. gingivalis*, the authors primer PG13, which is specific for *P. gingivalis* hypervariable region (Lyons, Griffen, and Leys 2000).

However, the limitation of using the 16S rRNA gene is also well known. The 16S rRNA is a multicopy gene possessing 1-15 copies of the gene in various bacterial species. This can result in inaccurate quantification, especially in cases where total bacterial load has to be accounted for. In studies, where total bacterial load has to be measured, the numbers obtained will be calculated based on the multiple copies of 16S rRNA, a much higher than the original count will be calculated. On the other hand, studies where individual bacterial genes have to be amplified for quantitation, the entire 16S rRNA sequence has to be annotated for identifying the bacteria, and the cell culture population needs to be calculated using conventional microbiological methods of plating and colony count, or obtaining microscopic cell count. (Poretsky et al. 2014). To tackle the problem of overestimation, a simple alternative could be to select and work with a single copy gene that will surely result in absolute quantification of target bacteria.

2.5.3 Improving real time quantitation using the *rpoB* gene

To overcome the limitations of 16S rRNA gene in microbial quantitation, an array of essential housekeeping genes has been suggested to be used. A few examples of housekeeping genes which can be used for microbial ecology purposes are rpoB, amoA, pmoA, nirS, nirK, nosZ, and pufM (Lan et al., 2016). From the list, a lot of studies has been reported with the rpoB as

the target gene (Vos et al. 2012, Rocha, Santos, and Pacheco 2015). The rpoB gene encodes for β-subunit of RNA polymerase, a subunit of RNA polymerase (RNAP), an enzyme crucial for transcription. The majority of the catalytic function of RNAP, which includes synthesizing mRNA, rRNA and tRNA, is carried out by the β-subunit (DJ and CA 1989). The size of this monocopy (single-copy) rpoB gene ranges between approximately 3411bp (*Staphylococcus aureus*)- 4185bp (*Neisseria meningitidis*). As mentioned earlier, the *rpoB* gene has been used for bacterial identification purposes . Using NGS, sequencing of *rpoB* gene can be utilized, where only 300-750bp is required for identification of bacterial isolates. Similar to the 16S rRNA gene, the *rpoB* gene is quite conserved and also contains both conserved and hypervariable regions, which can be used for identification purposes. The region between 2300bp and 3300bp containing the hypervariable region is most suitable for PCR-sequencing primer pairs used for identification and assignment of taxonomy. Apart from species identification and taxonomy assignment, this gene is also used for microbial quantitation, owing to the properties of being a universal gene in bacteria, a single-copy gene that can aid in accurate bacterial population quantitation (Adékambi, Drancourt, and Raoult 2009).

In a study performed for detection of *P. gingivalis* in periodontitis, *rpoB* gene was targeted as a marker for quantitation. SYBR green real time PCR assay was used and the assay using primers specific for *rpoB* gene could detect as little as 4fg of *P. gingivalis* chromosomal DNA (S. N. Park, Park, and Kook 2011). In another study, *in-silico* methods for phylogenetic tree reconstruction, DGGE was used to compare the efficiency of 16S rRNA gene as compared to *rpoB* gene as molecular markers. The genes were compared at various taxonomic levels to test their ability to resolve bacterial taxonomy and phylogeny which suggested *rpoB* gene provided better phylogenetic resolution in more than 53% of tested datasets as compared to 16S rRNA gene and close to 30% equal resolution to 16S rRNA gene. It was suggested that rpoB gene can be used as a universal gold standard molecular marker as it is capable of deciphering fine-scale phylogenetic relationships that go undetected using the 16S rRNA gene (Case et al. 2007). The species-specificity of rpoB gene was also tested in detection and quantitation of possible VBNC *Campylobacter jejuni* under stress-induced laboratory conditions, which can be used for application in poultry, environment and agri-food production systems (Lv et al. 2020). Moreover, wide application of *rpoB* gene has been observed in the food industry for quality check purposes as well as quantitative purposes. In

the winemaking industry, the monitoring and identification of Lactic acid bacterial species was carried out using culture independent *rpoB* based PCR-DGGE assay (Renouf et al. 2006). The use of *rpoB* gene avoided the interspecies heterogeneity problem caused by the use of the 16S rRNA gene (Renouf et al. 2006). In another study, *rpoB* based assay was developed for identification and enumeration of 19 *Cronobacter* spp. in food samples such as dry cereal, and milk products. The products contaminated by *Cronobacter* spp. were successfully identified and no cross reactivity was found using *rpoB* gene specific primers (Li et al. 2015). The reports published support the fact that the *rpoB* gene is a well-suited gene for absolute quantification of bacterial species, including VBNCs. The approach does have a wide array of applications, including industrial, environmental and clinical applications.

Keeping these points as the basis, we aimed to design a SYBR green dye based real-time quantitative PCR assay using *rpoB* gene for absolute quantification of significant bacteria involved in oral cancer. The aim for developing this method was to design a quick, simple and inexpensive assay which could be used for preliminary detection of pathogens (microbial markers) and early diagnosis of oral cancer. The technique also have applications in other areas as mentioned above, where the objective is to quantify the microbial population in a given sample.

2.6 Host-Microbe interaction at molecular level

There is enough evidence that suggests the process of carcinogenesis in humans is a multistep process (Hanahan & Weinberg, 2011; Spurgers et al., 2006). This multistep process reflects various genetic alterations/mutations that happen in normal human DNA/chromosomes that drive the progression of normal healthy human cells to malignant derivatives, also called as cancerous cells. Cancer cells have defects in regulatory circuits that govern normal cell proliferation and homeostasis. The vast genotypes of cancer cells is thought to be due to six essential alterations in cell physiology due to pathways responsible for the same (D Hanahan and Weinberg 2000). These are called the hallmarks of cancer, first proposed by Hanahan and Weinberg in 2000. 10 years later, two new emerging hallmarks were suggested, which were **deregulating cellular energetics** and **avoiding immune destruction,** as well two emerging characteristics were noted which were **genomic instability and mutation** and **tumor-promoting inflammation** (Fig. 2.7) (Douglas Hanahan and Weinberg 2011)**.**

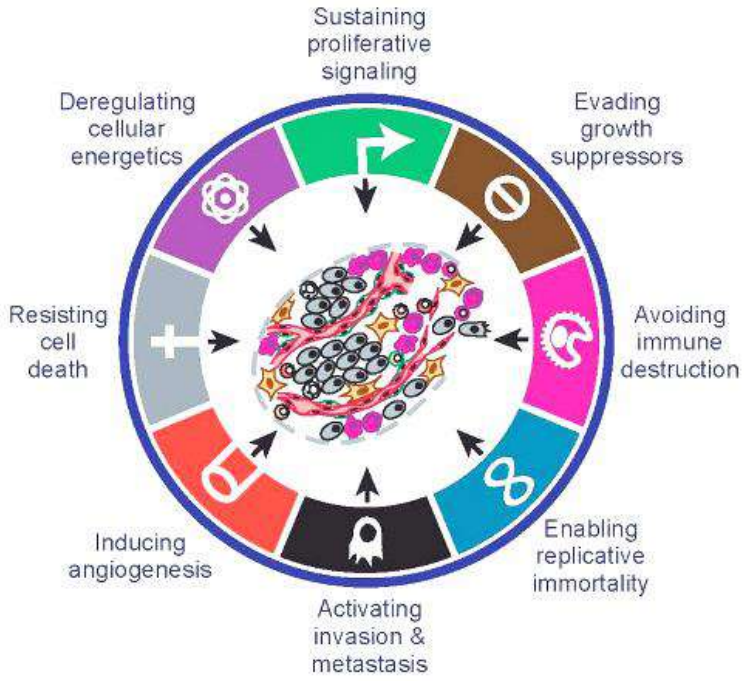


Fig.2.7: The hallmarks of cancer (Hanahan and Weinberg 2011)

2.6.1 Pathways involved in cancer/ hallmarks of cancer

The six major hallmarks of cancer are (i) Sustaining proliferative signalling, (ii) Evading growth suppressors, (iii) Resisting cell death, (iv) Enabling replicative immortality, (v) Inducing angiogenesis and (vi) Activating invasion and metastasis. These six hallmarks provide a solid foundation towards understanding tumor biology.

The first hallmarks of cancer to be identified is **sustaining proliferative signalling**. Cancer cells have the ability to survive and maintain chronic proliferation which is lacking in normal cells. Normal cells and tissues maintain homeostasis by controlling the production and release of growth signals and control cell growth and division maintaining the normal tissue architecture and function. The capacity of a cancer cell to survive proliferative signals can be obtained in a number of ways. They may generate their growth factor ligands, to which they might react by expressing the appropriate receptors, resulting in autocrine proliferative activation. Conversely, cancer cells could emit signals that activate normal cells in the tumor's stroma. (Cheng et al. 2008; Bhowmick, Neilson, and Moses 2004). Constitutive activation of constituents of signal transduction able to operate downstream of these receptors could also lead in neurotrophic autonomy, eradicating the need to induce these pathways via ligand-

mediated receptor stimulation. Through high-throughput DNA sequencing it has also been found that somatic mutations in certain cancers can be used to predict constitutive activation of signalling circuits, such as the mutations in B-Raf protein results in constitutive signalling through MAP-kinase pathway (Davies and Samuels 2010).

Apart from inducing and sustaining positive growth signals, cancer cells also **evade cell growth and proliferation suppressors** , namely the tumor suppressor genes which limit cell growth. The two major tumor suppressor proteins, RB (retinoblastoma associated) and TP-53, act as major control points for cell regulation pathways that govern proliferation, apoptosis and senescence. RB protein gathers signals from various cellular pathways and decides the fate of a cell. In cancer cells, the RB pathway is distorted leading to loss of gatekeeping function leading to never-ending cell proliferation. Along with the above mentioned mechanism, normal cells possess characteristic of contact inhibition where cell-to-cell contact by population of healthy cells, suppress further cell proliferation leading to formation of only monolayers of confluent cells. This property of contact inhibition is abolished by cancer cells.

Programmed cell death via apoptosis serves as a natural barrier to cancer development, which is often deregulated in cancer (Lowe, Cepero, and Evan 2004; Adams and Cory 2007). The apoptotic machinery is made up of both upstream and downstream regulators and effectors.. Although the cellular conditions that provoke apoptosis are still unknown, several abnormality sensors that play a critical role in tumour development have been identified. The most remarkable is a DNA-damage sensor that works through the TP53 tumour suppressor. TP53 induces apoptosis in response to significant levels of DNA breaks and other chromosomal abnormalities (Junttila and Evan 2009). To restrict or avert apoptosis, tumor cells employ a variety of strategies. The loss of the TP53 tumor suppressor element, which eliminates this crucial damage sensor from the apoptosis-inducing circuitry, is the most prevalent. Tumors may achieve similar outcomes by raising anti-apoptotic regulators (Bcl-2, Bcl-xL) or survival signals (Igf1/2), lowering pro-apoptotic factors (Bax, Bim, Puma), or circumventing the external ligand-induced death route. In such cases, any mutation or dysregulation in the genes in apoptosis or autophagy pathway can result in loss of the function, thereby resulting in cells that cannot undergo apoptosis, which is one of the properties of cancerous cells.

Another characteristic, i.e. **replicative immortality** is associated with cancer. Senescence is defined as an irrevocable transformation into a non-proliferative but viable state which eventually leads to cell death. As a consequence, once cells are propagated in culture, repeated cycles of mitosis induce senescence and the cell population enters a crisis process wherein the immense majority of cells die. Multiple pieces of evidence suggest that telomeres, play a critical role in the ability to proliferate indefinitely. The length of a cell's telomeric DNA defines how many successive cell generations its progeny can pass through before telomeres are primarily stripped away and, as a consequence, end up losing their protective effects, resulting in a crisis. Telomerase, the specific DNA polymerase that adds telomeric repetitive sections to the ends of telomeric DNA, is largely absent in immortalized cells but present at operationally elevated amounts in the vast majority of rapidly immortalized cells, including human cancer cells. (Shay and Wright 2000; Blasco 2005).

Tumors, like normal tissues, require nourishment in the form of nutrition and oxygen, along with the ability to expel metabolic wastes and CO2. The tumor-associated neovasculature is formed during the **angiogenesis** process. Angiogenesis is activated in adults as part of physiologic processes such as wound healing and female reproductive cycling, but only temporarily. During tumour progression, on the other hand, a ‘‘angiogenic switch" is almost always activated and remains active, causing normally quiescent vasculature to constantly begin to form new vessels that help support expanding neoplastic growths (Douglas Hanahan and Folkman 1996). A few of these angiogenic regulators are signalling proteins that bind to vascular endothelial cells' stimulatory or inhibitory cell-surface receptors. Vascular endothelial growth factor-A (VEGF-A) and thrombo-spondin-1 (TSP-1) are well-known angiogenesis inducers and inhibitors, respectively. When the expression of other proangiogenic indicators, such as members of the fibroblast growth factor (FGF) family, is severely enhanced, it has been associated to the maintenance of tumor angiogenesis (Baeriswyl and Christofori 2009). Blood vessels formed within tumours as a result of chronically active angiogenesis and an unbalanced mix of proangiogenic signals are typically abnormal. Precocious capillary sprouting, convoluted and excessive vessel branching, distorted and expanded vessels, irregular blood flow, micro hemorrhaging, leakiness, and unusual levels of endothelial cell proliferation and apoptosis classify tumour neovasculature (Baluk, Hashizume, and McDonald 2005; J. A. Nagy et al. 2010).

Cancers arising from the epithelial typically change structure in addition to their attachment to other cells and the extracellular matrix (ECM) as they progress to higher pathological grades of malignancy. The most well-studied change involved carcinoma cells losing E-cadherin, a crucial cell-to-cell adherence molecule. By trying to form adherens junctions with adjacent epithelial cells, E-cadherin aids in the forming of epithelial cell sheets as well as the maintenance of cell quiescence in these sheets (Berx and van Roy 2009). Furthermore, the expression level gene encodes other cell-to-cell and cell-to-ECM adhesion molecules and is modified in some highly aggressive carcinomas, with those favouring cytostasis typically downregulated. Adhesion molecules that are normally associated with cell migrations that occur during embryogenesis and inflammation, on the other hand, are frequently upregulated. N-cadherin, which is typically expressed in migrating neurons and mesenchymal cells during organogenesis, for example, is upregulated in many invasive carcinoma cells (Cavallaro and Christofori 2004). In addition to these factors, the cancer cells also produce proangiogenic and/or pro invasive matrix-degrading enzymes, including MMP-9 and other matrix metalloproteinases, cysteine cathepsin proteases, and heparanase which aid in movement and metastasis (Qian and Pollard 2010).

2.6.2 Molecular changes due to bacterial infection leading towards carcinogenesis

Various bacteria have been proposed to induce tumorigenesis by inducing chronic inflammation, interfering with eukaryotic cell cycle and signalling pathways, or resulting in mutagenesis through the metabolism of potentially carcinogenic substances such as acetaldehyde. *Exiguobacterium oxidotolerans, Prevotella melaninogenica, Staphylococcus aureus,* and *Veillonella parvula* were found to be specific for tumorigenic tissues in studies of isolated bacterial taxa from oral cancer tissue specimens and controls (Chocolatewala, Chaturvedi, and Desale 2010). The majority of the isolates are saccharolytic and acid tolerant. There is growing evidence that bacterial infections are dependent on precise interactions between pathogens and components of the host cell regulatory systems, such as given below:

1. Several bacteria have been shown to cause chronic infections or to produce toxins that disrupt the cell cycle and lead to altered cell growth(Han 2015).
2. Chronic infections stimulate cell proliferation and DNA replication by activating mitogen-activated kinase (MAPK) pathways and cyclin D1, as well as increasing the

incidence of cell transformation and tumour development by increasing the rate of genetic mutation (Coussens and Werb 2002).

3. Several infections result in pathogen intracellular accumulation, which suppresses apoptosis primarily through modulation of the expression of Bcl-2 family proteins or inactivation of the retinoblastoma protein, pRb. This strategy creates a niche in which the intracellular pathogen can survive despite the host immune system's attempts to destroy infected cells through apoptosis. As a result, partially transformed cells can avoid self-destruction and progress to a higher level of transformation, eventually becoming tumorigenic(Lara-Tejero and Galan 2000; Nougayrède et al. 2005).
4. Numerous bacterial pathogens that cause chronic infection via intracellular access disrupt host cell signalling pathways, enhancing pathogen survival. The control of these signalling factors is critical to the development or prevention of tumour formation. Such infections can resemble some of the gross effects seen in carcinogenesis, and the precancerous lesion formed as a result of such infections can regress with antibiotic treatment and bacterial clearance(Lax 2005).
5. Another potential explanation is the bacteria's metabolism of potentially carcinogenic substances. This is important in the oral cavity, where pre-existing local microflora may facilitate tumorigenesis by converting ethanol into its carcinogenic derivative, acetaldehyde, at levels capable of inducing DNA damage, mutagenesis, and secondary epithelial hyperproliferation. This concept is also supported by increased levels of microbial acetaldehyde production in heavy drinkers and smokers (Seitz et al. 2001).

2.6.3 Carcinogenesis mechanisms of specific bacteria

2.6.3.1 Fusobacterium nucleatum

Fusobacterium nucleatum is a gram negative, anaerobic bacterium, a dominant resident of periodonticum and dental plaque biofilms. It has also been previously isolated from extraoral sites like blood, brain, chest, lung, liver, joint, abdominal, obstetrical and gynaecological infections and abscesses. The bacteria exist in the oral cavity when an individual suffers from various periodontal diseases like mild gingivitis, chronic periodontitis, and aggressive periodontitis, (Griffen et al. 2012; N.-Y. Yang et al. 2014). With increase in severity of disease, inflammation and pocket depth, the abundance of *F. nucleatum* increases (J. H. Moon and Lee 2016). The abundance of F. nucleatum also depends on external factors like smoking habits and periodontal infections, when bacterial abundance increases (Rickard et al. 2003). Similar to *P. gingivalis, F. nucleatum* also has various virulence factors using which the bacteria invade the host cell and escapes immune surveillance. These virulence factors can be categorized into two, namely colonization and dissemination and induction of host responses. *F. nucleatum* is an adherent bacterium that co-aggregates with other bacterial species to form biofilms or dental plaques. For this purpose, *F. nucleatum* produces many adhesins like FadA, Fap2, RadD, aid1 of which FadA is the most studied and best characterized virulence factor which helps the bacteria to bind to the host cell. *F. nucleatum* binds to a variety of host cells like monocytes, erythrocytes, PMNs as well as a lot of host molecules such as cadherins, proteins, extracellular matrices, etc. FadA is required for adhesion as well as invasion. FadA binds to cell-junction molecules, *i.e.,* cadherins, namely VE-cadherin on endothelial cells and E-cadherin in epithelial cells leading to increase in permeability of the cells due to migration of cell-cell junction to intracellular compartments. *F. nucleatum* also plays a role in eliciting host immune response by various pathways. Human β-defensin 2 is induced from oral epithelial cells via Fad-I (Fardini et al. 2011), inflammatory cytokines IL-6, IL-8 and TNFα are stimulated (S.-R. Park et al. 2014) as well as NK cells mediated inflammatory responses are stimulated in periodontal diseases (Chaushu et al. 2012). In healthy periodontal condition, the inflammatory factors are kept in check. In dysbiosis, *F. nucleatum* induces aggravated inflammation and becomes a pathogen. *F. nucleatum* has also been linked to various other systemic diseases like atherosclerosis, stillbirths, and colorectal carcinoma.

2.6.3.2 Porphyromonas gingivalis

P. gingivalis, a gram negative facultative anaerobic periodontal pathogen, is known to damage the tissues supporting the teeth, which might eventually lead to tooth loss (Yilmaz, Watanabe, and Lamont 2002*). P. gingivalis* i) attaches to the epithelial cells of oral cavity ii) cause inflammation iii) evade host defence mechanism by utilizing numerous virulence factors, and iv) cause deregulation of innate immunity and inflammatory responses (M. Wang et al. 2007). The internalisation of *P. gingivalis* occurs in fimbriae dependent manner which binds to β-1 integrin on host cells, leading to rearrangement of actin fibres, thus allowing internalisation (Nakhjiri et al. 2001). Once the bacteria has invaded intracellular, the host cells do not undergo necrosis or apoptosis. The bacteria is also known to invade macrophages, the replication is less active, leading to better evasion of immune surveillance. Once intracellular, the bacteria secrete ATP-hydrolysing enzyme, thereby suppressing ATP-dependent apoptosis, allowing its survival in the host cell. Simultaneously, the bacteria infect neighbouring cells through actin-cytoskeleton bridges, thereby avoiding immune surveillance (Henry et al. 2012). It alters the host-cell cycle pathways, leading to increased proliferation of epithelial cells, which leads to a characteristic feature of periodontitis, i.e. enlarged periodontal pocket (P.-L. Wang and Ohura 2002). *P. gingivalis* can tolerate harsh conditions such as oxidative stress that exist in the periodontal pocket (Bainbridge and Darveau 2001). Various virulence factors like lipopolysaccharide (LPS), capsule/K-antigen, fimbriae and the gingipains, play a key role in the process. The variation in the composition of LPS and affinity towards host pattern receptors, i.e., TLRs and CD14 will determine the downstream events in the host cell and help in discriminating between commensal and pathogenic strains of the bacteria. The LPS is a major factor in development of periodontitis, as it activates the host inflammatory pathways and innate defense response (Sims et al. 2001; Laine and van Winkelhoff 1998). The capsule, also called the K-antigen, is also a major virulence factor of the bacteria. It was observed that encapsulated *P. gingivalis* were highly invasive as compared to non-capsulated, more resistant to phagocytosis, and can generate systemic IgG responses (Amano et al. 2004; Hajishengallis et al. 2007).

The fimbriae of *P. gingivalis* is a thin, long, filamentous protrusion which helps in invasion, adherence and attachment of the bacteria to host cells and also to other bacteria to form biofilms. There are two types of fimbriae known, the major [I], which helps in colonization

and invasion and the minor [II], which possesses higher pro-inflammatory properties (Hajishengallis, Wang, and Liang 2009). Each type of fimbriae triggers a different downstream pathway. The major fimbriae works through TR2 signalling to interact with CR3, which activates CR3 and allows internalization of *P. gingivalis* into the macrophages escaping the host immune system (Kadowaki et al. 2000). On the other hand, minor fimbriae mediates production of IL-6 and TNFα which release pro-inflammatory cytokines (Curtis, Aduse-Opoku, and Rangarajan 2001).

P. gingivalis also produces arginine- and lysine-specific cysteine proteinases called Gingipains, which account for about 85% of total proteolytic activity in the bacteria (Fagundes et al. 2011). Based on the substrate specificity, gingipains can be classified into 2 classes. The first, Arg-specific which degrades extracellular matrix components like complement and integrin-fibronectin. This class is further divided into 2 subclasses RgpA, which contains a proteolytic and an adhesion domain, and RgpB that contains only the proteolytic domain. Second class of gingipains Lys-X gingipain, Kgp, which contains both a proteolytic and an adhesion domain. There are sequence similarities between the adhesion domains of Kgp and RgpA (Lourbakos et al. 2001). These gingipains have various functions including i) cleavage of TCR [CD2, CD4, CD8], thereby hampering immunity, ii) stimulating expression of protease receptors on neutrophils, which are crucial for induction of cytokine response and inflammation, iii) stimulating IL-6 production in oral epithelial cells, and iv) enhancing inflammatory responses (Moore and Moore 1994).

2.6.3.3 Oral Streptococcal species

The oral biofilm production begins with the initial adhesion of early colonizers of which 80% are Streptococcal species which can be divided into 5 types namely Mutans group [*S. mutans, S. sobrinus*], Salivarius group [*S. salivarius*], Anginosus group [*S. anginosus, S. intermedius*], Sanguinis group [S*. sanguinis, S. gordonii*], and Mitis group [*S. mitis, S. oralis*] (Rosan and Lamont 2000; Facklam 2002). In the early colonization, *S. sanguinis* and *S. gordonii* were the first to adhere to the tooth surface with the help of specific adhesins. Dental caries is polymicrobial in nature, although with *S. mutans* as a dominant bacterium along with some *Lactobacilli* sp. Studies have suggested that with the increase in abundance of *S. mutans* in dental caries, abundance of *S. sanguinis* significantly decreases and vice versa (Becker et al. 2002). All three *Streptococcal* sps. are known to produce specific antimicrobial peptides, such

as *S. mutans* produces Mutacin I and Mutacin IV, *S. gordonii* produces Streptocins and *S. sanguinis* produces Sanguicin (Schlegel and Slade 1972, 1973; Deng et al. 2004). Differences lie in the specificity of the bacteriocins towards the target bacteria. Bacteriocins produced by *S.mutans* act against and inhibit growth of *S. sanguinis* and *S. gordonii*, whereas the bacteriocins produced by the two do not inhibit the growth of *S. mutans*. The bacteriocins produced by *S. mutans* also target the cell membrane of susceptible species leading to leakage of the cellular components thereby causing cell death. In defence, *S. gordonii* produces H2O2 as well as a protease which interferes with bacteriocin-competence regulation cascade (Kreth et al. 2005; B.-Y. Wang and Kuramitsu 2005). Oral *Streptococcus* spp. also interacts with other early colonizers like *Veillonella, Actinomyces* and late colonizers like *Fusobacterium* and *Porphyromonas*. *S. mutans* and *Veillonella* spp. are seen to have a symbiotic relationship. *Veillonella* sp. cannot utilize carbohydrates as a carbon/energy source and uses lactate instead, which is supplied by *S. mutans*. In return, *Veillonella* sp. detoxifies the environment for *S. mutans* (Mikx and Van der Hoeven 1975). It has also been seen that apart from nutritional relationship, dual species biofilm of S. mutans and V. parvula are less susceptible to antimicrobial treatment which helps in their colonization in dental diseases suggesting a more intrinsic metabolic complementation (Kara, Luppens, and Cate 2006).

In the case of late colonizer *Fusobacterium*, which is known to adhere and invade gingival cells, *S. cristasu*s is non-invasive bacteria. They are known to form co-aggregates where *S. cristasus* adheres to *Fusobacterium* which is mediated by an arginine-sensitive interaction and enables *S. cristasus* to invade human epithelial cells (Andrew M Edwards, Grossman, and Rudney 2006; A M Edwards, Grossman, and Rudney 2007). *Streptococcal spp*. not only adhere but also invade various cells of the oral cavity including hard tissue of the tooth and oral buccal epithelial cells, due to which, they are likely to have a close interaction with the host immune system. The host immune system interacts differently with commensal and pathogenic *Streptococcal* bacteria. For example, *S. cristasus* which is associated with *Fusobacterium* on invasion into the oral epithelium cells led to the increased IL-8 expression (Zhang, Chen, and Rudney 2008). Similarly, exposure of *Fusobacterium* led to increased expression of IL-6 and IL-8 but *S. gordonii* failed to elicit immune response (Hasegawa et al. 2007). Detailed study of such interactions is still required to understand the mechanism of differential responses by the host immune system to different bacteria. *Streptococcal* spp. is one of the best examples

to show how different species from the same genus act as commensal and pathogenic in the same environment.

2.7 Lacunae: The move ahead..

There is ever increasing evidence about the role and involvement of microbiota in oral cancer and other cancers in general. Although Next Generation Sequencing platforms have made it easier to identify and characterize the oral microbiota, there is not enough data about oral microbiome in the Indian population, especially in long term tobacco chewers and patients with oral cancer, and that became the motivation for the present work. Considering the increasing thirst for composition about oral microbiome, we aimed to decipher the oral microbiome using NGS in the Indian population suffering from oral cancer as well as long term tobacco chewers, since it is a prime risk factor for OC.

Hence, the present work deals with finding the association of oral microbiota in Indian oral cancer patients as well as long term tobacco chewers. Also, a qPCR method was developed to carry out absolute quantification of oral bacteria which could help in early diagnosis of patients with oral cancer. Alongside, a preliminary host-pathogen interaction studies were performed with human oral cell line and significant bacteria in oral cancer to look upon cellular pathways altered towards oral cancer due to bacterial infection.

CHAPTER 3

Rationale, Aim, and Objectives

Rationale, Aim and Objectives

3.1 Rationale

The microbiome flora in various sites in humans is important for maintaining good health. It has gained immense importance as it has been associated with cancer development, progression and response of patients to treatment. Most of the microorganisms on or in the human body are beneficial, however, a change in this set of microflora and occurrence of certain bacterial species are associated with diseases and certain cancers, such as colon, liver, breast and lung cancer. Similarly, reports have correlated the dysbiosis in the oral cavity with oral cancer cases. Reports, primarily from various countries other than India; like US, Hungary, UK, South Korea, and Japan; on the microbiome of the oral cavity have identified a large diversity of bacteria. However, there is a dearth of information on the microbiome in Indian groups of subjects, and the microflora in the oral cavity of healthy individuals and oral cancer patients has not been investigated. In India, 21.4% of adults, 29.6% of men and 12.8% of women, are tobacco chewers, and more than 0.35 million Indians die as a consequence of this habit each year. 66% of these tobacco chewers eventually develop oral cancer. Considering the heterogeneity of the population depending on their lifestyle habits, especially tobacco chewing, which is a major risk factor of oral cancer, diversity in oral microflora is expected to play a role in oral cancer development.

It is also not clearly known as to how the microbe interaction affects the host cells at a molecular level. As it is highly possible that changes in oral bacterial species may lead to Oral Cancer, the present project involves a preliminary level host-bacterial interaction study to understand its role in carcinogenesis. Based on these points, the aim and objectives of our study are given below.

3.2 Aim

To study bacterial diversity in oral cavity and its implications in Oral Cancer

3.3 Objectives

a. To identify the bacterial diversity in the oral cavity of oral cancer patients and tobacco chewers.

b. To find the correlation of oral bacterial diversity and oral cancer

c. To study host-microbe interaction at molecular level.

CHAPTER 4

Identification of bacterial diversity in oral cavity of oral cancer patients and tobacco chewers in Indian population

Identification of bacterial diversity in oral cavity of oral cancer patients and tobacco chewers in Indian population

4.1 Introduction

This chapter deals with the 16S rRNA gene metagenomics study that was performed to identify bacterial diversity in the oral cavity of three study groups namely, healthy controls (C), long term tobacco chewers (T), and oral cancer patients (OC). Following the sample collection from participants, genomic DNA was isolated using a silica column-based method for efficient isolation, and 16S rRNA targeted Next Generation Sequencing was performed using Illumina Miseq. The raw data obtained was analysed using QIIME2™ pipeline for identification and assessment of bacterial taxonomy, followed by diversity analysis, biomarker identification and pathway prediction.

4.1.1 Illumina MiSeq

Illumina has dominated the sequencing market in recent years, using a sequencing-by-synthesis approach on clonally amplified DNA templates attached to an acrylamide coating on the surface of a glass flow cell and fluorescently tagged reversible-terminator nucleotides (Ambardar et al., 2016). The Illumina Genome Analyzer and, more subsequently, the HiSeq 2000 set the bar for high-throughput massively parallel sequencing. However, in 2011, Illumina unveiled the MiSeq, a lower-throughput fast-turnaround equipment geared for smaller labs and the clinical diagnostic sector (Quail *et al.*, 2012). This is referred to as 4-channel sequencing, as all fluorescent tags have to be imaged for sequencing. The process of sequencing by Illumina is one of the fastest, and can perform automated clonal amplification, sequencing, and quality scored base calling in as few as 5-6 hours on the MiSeq System. The MiSeq System offers a cost-effective alternative to sequencing by capillary electrophoresis, and qPCR for applications, such as targeted resequencing, clone checking, amplicon sequencing, RNA expression, small genome sequencing, 16S metagenomics, RNA sequencing, targeted resequencing, and pre-implantation genetic screening.

4.1.2 QIIME2

With the advent of next generation sequencing techniques, a steep increase in the pipelines for analysis of NGS data has been observed. Various data analysis pipelines such as Mothur,

QIIME1/QIIME2, MG-RAST, *etc.*, are well known (Mukherjee et al., 2023). We used QIIME2 pipeline to analyse our NGS data. QIIME2 is an open-source microbiome data science platform available for analysis of metagenomic data. The de-multiplexed raw data needs to be imported in QIIME2 pipeline followed by quality assessment using FASTQC. The forward and reverse reads were joined using vsearch. The quality of joined reads was assessed and the reads were trimmed up to 300bp using Deblur to obtain a table output and representative sequences. The table output was used to analyse the number of unique features and distribution of the features across all the samples, whereas the representative sequence output was used to assign taxonomy to the sequences. There are different databases available for assigning taxonomy using 16S rRNA gene sequence, like GreenGenes, SILVA and RDP. In our study, we have used GreenGenes database to assign taxonomy to our sequences using QIIME2 inbuilt feature classifier. After Taxonomy assignment, taxa bar-plots were obtained, and the data extracted was used to visualize abundance of various genera in each study group. Alpha indices and Beta diversity matrices were also obtained using QIIME2 pipeline, that was used to study the bacterial diversity within and between different subject groups.

4.2 Materials and Methods

4.2.1 Materials (Chemicals, Kits, Reagents, Solutions, plasticwares)

A saline solution (0.85%; 0.85g NaCl in 100ml MQ water) was prepared and autoclaved, and aliquots were stored in 50 ml sterile falcon tubes, which were used for oral cavity sample collection. Invitrogen PureLink™ Genomic DNA kit (Cat. No. K182001) was procured from ThermoFisher Scientific, India. Molecular grade agarose powder was purchased from MP Biomedicals, USA. All the plasticwares used during experiments (including 15ml and 50ml falcons tubes) were purchased from SPL Life Sciences Pvt. Ltd., India.

4.2.2 Subject recruitment and information collection

Ethical approvals for the recruitment of healthy, long term tobacco chewing volunteers and oral cancer patients were taken from SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and The Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai (dt. 12/09/2017). Volunteers for the study were divided into three groups, namely Control (C), Long-term tobacco-chewers (T) and patients diagnosed with Oral Cancer (OC). A minimum of 40 individuals belonging to each study group were identified who

satisfied the study selection criteria (Table 4.1) and the samples were collected for further analysis. Exclusion criteria included individuals under the age of 18, medically compromised/unfit to give consent, subjects who were completely edentulous and those who received oncotherapy earlier. The healthy controls were defined as non-smokers and without any diagnosed diseases in the oral cavity, which was confirmed by prior clinical examination. Long term tobacco chewers were individuals chewing various forms of tobacco for 5 years or more and without any oral or periodontal disease. Informed written consent was obtained prior to the sample collection from all the participants. All diagnoses for patients with oral cancer were confirmed by biopsy and histopathological findings. At the time of sample collection, participants were devoid of any antibiotic treatment. Subjects were asked to refrain from eating, drinking, or oral hygiene procedures for at least one hour before sample collection. The samples were collected in the period of December 2017 to January 2020. Data on participants' medical history, age, gender, occupation, habits of tobacco ingestion, alcohol intake, and general oral hygiene questions were recorded.

Table 4.1: Participant Selection Criteria

Study group	Inclusion criteria	Exclusion criteria
All study groups	Minimum age: 18 years	Lactating/Pregnant women Infected with HIV
Healthy controls	Non smokers Non tobacco chewers Non-alcohol consumers	Any gingival/oral health issues such as ulcers, periodontitis, caries, pre malignant lesions
Oral cancer	Histopathologically confirmed oral cancer patients	
Tobacco chewers	Tobacco chewers (5-10 years)	Any gingival/oral health issues such as ulcers, periodontitis, caries, pre malignant lesions

4.2.2.1 Healthy controls

During the period of December 2017 to January 2020 period, samples from 40 healthy individuals with no history of tobacco use, no medical history of cancer, absence of premalignant lesion, genetically unrelated to the oral cancer cases and agreed to voluntary

participation were collected and included as a Control group in our study. These participants were recruited from various locations across the Mumbai city. Written informed consent was procured from the participants for voluntary participation (Annexure I and II). Demographic details were noted by response of the participants to a formatted questionnaire (Annexure IV).

4.2.2.2 Long term tobacco chewers

During the same period (mentioned in 4.2.2.1), 40 healthy individuals with long term tobacco use of > 5 years, no medical history of cancer, absence of premalignant lesion, genetically unrelated to the oral cancer cases, and agreed to voluntary participation were collected as tobacco chewers. These participants were recruited from various locations across the city. Written informed consent was procured from the participants for voluntary participation (Annexure I and II). Demographic details and tobacco exposure information were noted by response of the participants to a formatted questionnaire (Annexure III).

4.2.2.3 Oral cancer patients

Subjects suffering from oral cancer were recruited after histopathological confirmation of the disease at Asian Cancer Institute-Somaiya Ayurvihar hospital located in Mumbai, India between the period of December 2017 to January 2020. The subjects received complete information regarding the objectives and procedures of our study and provided written informed consent prior to sample collection (Annexure I and II). Inclusion criteria for oral cancer patients were histopathologically confirmed oral cancer, aged 18 years and above, and voluntary participation. The exclusion criteria considered were - history of any other malignancy, other concurrent disease, lactating/pregnant women, aged 18 years and below and unable or unwilling to comply. Patients with complete clinical records (site of cancer, histopathology report, TNM staging) were included in the study. Demographic and clinicopathological details of the patients were collected as per Annexure III

4.2.3 Sample collection

Participants rinsed their mouths with 20 ml of sterile normal saline (0.85%) for 30 secs and spit it into a 50-ml sterile falcon tube. Salivary samples of patients were taken at the pre-operative stage while the patient was NBM (Nil By Mouth). Subjects from healthy & tobacco groups were asked to refrain from eating, drinking, or oral hygiene procedures for at least

one hour before sample collection. Salivary samples were collected in well-labelled sterile Falcon tubes and stored at 4°C and processed within 48 hours.

4.2.4 DNA Isolation

The DNA isolation protocol was as per the manufacturer's recommendation of Invitrogen PureLink™ Genomic DNA kit (ThermoFisher Scientific, Waltham, Massachusetts, USA) (Cat No K182002), with slight modifications. First, the collected oral rinse samples (40 samples from each group) were centrifuged at 14000 rpm for 8 minutes at room temperature. The supernatant was discarded and the pellet was used for DNA Isolation by using Invitrogen PureLink™ Genomic DNA kit). The samples were assigned unique IDs as OC1 to OC40 for oral cancer patients, T1 to T40 for tobacco chewers and C1 to C40 for healthy control for the purpose of anonymity. .

After centrifugation, the supernatant was discarded and 180 µl Lysozyme digestion buffer (20mg/ml) was added to the pellet, vortexed well and incubated at 37°C for 30 mins. After incubation, 20 µl RNase A was added, mixed well and incubated at room temperature for 2 mins. Following this, 20µL Proteinase K was added, vortexed and incubated at room temperature for 2 mins. 200 µl binding buffer was added to the tube and vortexed to obtain a homogenous solution followed by incubation at 55°C for 30 mins to promote protein digestion. 200 µl chilled 100% ethanol was added to the lysate and vortexed. The lysate (~650 µl) was loaded onto the Spin Column and centrifuged at 8,000 rpm for 1 min at room temperature a. The collection tube was discarded and the column was placed into a clean collection tube. 500 µl wash buffer 1 was added to the column followed by centrifugation at 8,000 rpm for 1 min at room temperature. The collection tube was again discarded and the column was placed into a clean collection tube. 500 µl wash buffer 2 was added to the column followed by centrifugation of the column at maximum speed (14,000 rpm) for 3 mins at room temperature followed by discarding the collection tube.

The spin column was placed in a sterile 1.5 ml microcentrifuge tube and 100 µl of Elution Buffer was added to the column and was incubated at room temperature for 1 min. The assembly was centrifuged at maximum speed (14,000 rpm) for 1 min at room temperature. The spin column was discarded and the microcentrifuge tube containing purified DNA was stored at -20°C till further processing.

4.2.5 Assessment of quality and quantity of isolated DNA

DNA sample yield and integrity was assessed by electrophoresis on 1 % agarose gel. 1.0 g agarose powder was added in 100 ml of 1X Tris-Acetic acid-EDTA (TAE) buffer in a clean dry conical flask. The mixture was heated in the microwave and cooled for 5-7 mins followed by addition of 5 µl ethidium bromide (1µg/ml stock solution). The gel was poured in the gel casting unit containing comb and allowed to polymerize for 20-30 mins. After solidification, comb and casting gates were removed and gel was transferred to the horizontal electrophoresis tank. 350 ml 1X TAE buffer was poured in the electrophoresis tank. DNA samples were loaded after mixing with 6X loading dye (10 µl DNA sample + 2 µl 6X loading dye) followed by electrophorese at 100 V for 30-45 mins. The gel was visualised using the ChemiDOC XRS+ system (Bio-Rad, USA).

The purity and yield of DNA were also assessed spectrophotometrically using BioTek EPOCH2 microplate reader(Agilent, USA). The purity of DNA was determined by the ratio of absorbance at 260 nm and 280 nm (A260/A280).

4.2.6 Next Generation Sequencing using Illumina Miseq

The sequencing procedure was performed at Department of Bio & Environmental Technology, College of Natural Science, Seoul Women's University, Seoul with the collaborative help of Dr. Sathiyaraj Srinivasan. 40 samples from each study group were subjected to NGS sequencing using Illumina Miseq (Illumina, San Diego, CA). The library was prepared by standard library construction protocol (https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html) using Nextera XT kit (Illumina, San Diego, CA), following the manufacturer's instructions. The isolated DNA sample was diluted to obtain a concentration of 100 ng (10µl of 10ng/µl) and was used for the sequencing protocol. The PCR amplification of bacterial 16S rRNA hypervariable region V6-V8 was carried out using primers B969F (ACG CGH NRA ACC TTA CC) and BA1406R (ACG GGC RGT GWG TRC AA). The PCR reaction was carried out by 30 sec initial denaturation at 98 °C, 30 cycles of 10 sec denaturation at 98 °C, 30 sec annealing at 55 °C, 30 sec elongation at 72 °C, and a 5 min final extension at 72 °C. The specific amplicons for V6-V8 region were quantified in each reaction mixture along with Illumina sequence adapter, and index primers (Nextera XT Index kit) were utilized to generate amplicon libraries

followed by PCR clean up. MiSeq libraries were quantified and then subjected to 300-nucleotide paired-end multiplex sequencing on Illumina MiSeq sequencer.

4.2.7 Hardware and software requirements

The sequencing raw data was obtained and all the sequence analysis studies described herein were performed at SDSOS, NMIMS University, Mumbai on a HP laptop (Intel® Core™ i5-7200U CPU @ 2.50 GHz, RAM 16 GB) running Windows 10 64-bit Home Basic Operating System. To run QIIME2 on Windows host, Oracle VM VirtualBox (Version 6.1.22) was installed followed by importing and installing QIIME2 (Version 2021.2).

4.2.8 Taxonomic assessment using QIIME2

The demultiplexed paired-end V6-V8 sequenced data was obtained in Casava One Eight Single Lane Per Sample format in folder 'oralmicrobes'.

- This data was imported in QIIME2 pipeline using the following command line:

```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path oralmicrobes \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path demux-paired-end.qza
qiime demux summarize \
  --i-data demux-paired-end.qza \
  --o-visualization demux.qzv
```

The (.qza) files are QIIME2 artifacts, which can be visualised after converting them into a (.qzv) file.

- The forward and reverse reads were joined using "vsearch" using the following command line:

```
qiime vsearch join-pairs \
  --i-demultiplexed-seqs demux.qza \
  --o-joined-sequences demux-joined.qza
qiime demux summarize \
  --i-data demux-joined.qza \
  --o-visualization demux-joined.qzv
```

- The quality control of sequences was carried out to discard low quality reads using the following command line:

```
qiime quality-filter q-score-joined \
  --i-demux demux-joined.qza \
  --o-filtered-sequences demux-joined-filtered.qza \
  --o-filter-stats demux-joined-filter-stats.qza
```

- The next denoising step was carried using “Deblur” and the sequences were trimmed resulting in 300 nucleotide sequences since best quality reads (Q Phred score>30) were observed up to 300bp using following command line:

```
qiime deblur denoise-16S \
  --i-demultiplexed-seqs demux-joined-filtered.qza \
  --p-trim-length 300 \
  --p-sample-stats \
  --o-representative-sequences rep-seqs.qza \
  --o-table table.qza \
  --o-stats deblur-stats.qza
qiime feature-table summarize \
  --i-table table.qza \
  --o-visualization table.qzv
--m-sample-metadata-file NGSmetadata-Sheet1.csv
qiime feature-table tabulate-seqs \
  --i-data rep-seqs.qza \
  --o-visualization rep-seqs.qzv
```

- The resulting outputs i.e. table.qza (Feature table (frequency)) and rep-seqs.qza (Feature data sequences) are important outputs, which were used for further down processing. The result of both of these methods will be a FeatureTable[Frequency] QIIME 2 artifact, which contains counts (frequencies) of each unique sequence in each sample in the dataset, and a FeatureData[Sequence] QIIME 2 artifact that maps feature identifiers in the FeatureTable to the sequences they represent.
- The metadata file (NGSmetadata-Sheet1.csv) was included in the analysis which consists of participant information such as sample ID, sample barcode, group and age of the participant, habits etc.
- The taxonomy assessment of the sequences was carried out using Naive Bayes classifier trained against GreenGenes (13-8-99 version) database using the following command line:

```
qiime feature-classifier classify-sklearn \
  --i-classifier gg-13-8-99-nb-classifier.qza \
  --i-reads rep-seqs.qza \
  --o-classification taxonomy.qza
qiime metadata tabulate \
  --m-input-file taxonomy.qza \
  --o-visualization taxonomy.qzv
```

- The taxonomic composition of samples with interactive bar plots was viewed using following commands:

```
qiime taxa barplot
```

```
 --i-table table-deblur.qza
 --i-taxonomy taxonomy.qza
 --m-metadata-file Final-metadata-NGS.tsv
 --o-visualization taxa-bar-plots.qzv
```

The results were extracted in (.csv) files and analysed in MS Excel.

4.2.9 Alpha and Beta diversity analysis by QIIME2

- To analyse phylogenetic distances between the samples, a rooted and unrooted phylogenetic tree was obtained using the MAFFT multiple alignment program, as given below.

```
qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences rep-seqs-deblur.qza \
  --o-alignment aligned-rep-seqs.qza \
  --o-masked-alignment masked-aligned-rep-seqs.qza \
  --o-tree unrooted-tree.qza \
  --o-rooted-tree rooted-tree.qza
```

- The resulting rooted and unrooted trees were used to analyse alpha and beta diversity matrices. An important parameter of sampling depth needs to be set to assure even sampling of all samples for diversity analyses.

```
qiime diversity core-metrics-phylogenetic \
 --i-phylogeny rooted-tree.qza \
 --i-table table-deblur.qza \
 --p-sampling-depth 1325 \
 --m-metadata-file Final-metadata-NGS.tsv \
 --output-dir core-metrics-results
```

- The core metrics directory results in outputs, such as faith pd vector, Observed features vector, Shannon vector evenness vector, weighted and unweighted unifrac distance matrices, Bray Curtis and Jaccard matrices.
- Each of the above outputs can be visualised by converting them into visualization files (.qzv).
- Further alpha diversity indices (Chao1/ACE/goodscoverage/observed otus/shannon/simpson/pielou_e) were obtained using the feature table and using following commands:

```
qiime diversity alpha \
  --i-table table.qza \
  --p-metric observed_otus \
  --o-alpha-diversity observed_otus_vector.qza
```

The --p-metric was replaced with the desired index to be studied.

- For beta diversity indices, (Weighted and Unweighted PERMANOVA/PERMDISP/ANOSIM), following command line was used:

```
qiime diversity beta-group-significance   \
--i-distance-matrix                                core-metrics-
results/unweighted_unifrac_distance_matrix.qza \
 --m-metadata-file Final-metadata-NGS.tsv \
 --m-metadata-column group \
  --o-visualization    core-metrics-results/unweighted-unifrac-
permdisp.qzv \
 --p-pairwise \
 --p-method permdisp
```

The –p method was replaced depending on the index to be tested.

- The beta diversity matrices (Bray-Curtis, Jaccard, Weighted Unifrac and Unweighted unifrac) were exported and used for visualization purposes in PhyloToast to obtain PCoA and Linear Disciminant Analysis (LDA) plots using the following command line:

```
PCoA.py -i (distance_matrix) -m (metadata_file) -g (column of
group of comparison in metadata file) -c (column depicting color
in metadata file) -s 250 –pc_order 1 2 –ggplot2_style -o
(output_filename.svg)

LDA.py -dm (distance_matrix) -m (metadata_file) -g (column of
group of comparison in metadata file) -c (column depicting color
in     metadata     file)     -s     250     –ggplot2_style     -o
(lda_output_filename.svg) –save_lda_input (lda_filename.csv)
```

4.2.10 Identification of Microbial biomarker using LEfSe

The identification of microbial biomarkers was performed using Linear discriminant analysis Effect Size (LEfSe) (Segata *et*. al., 2011). It determines the features (organisms, clades, operational taxonomic units, genes, or functions) that are most likely to explain differences between classes by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance. We performed LEfSe analyses on Huttenhower Galaxy Server (https://huttenhower.sph.harvard.edu/galaxy/).

To provide input for LEfSe, the OTU feature table was converted from absolute counts to relative frequency after collapsing the taxonomy at the highest level of testing, in our case genus level (level-6). Following set of commands were used:

- To collapse to level 6:

```
qiime taxa collapse \
```

```
  --i-table feature-table.qza \
  --i-taxonomy taxonomy.qza \
  --p-level 6 \
  --o-collapsed-table phyla-table.qza
```

- To convert the data to relative frequency:

```
qiime feature-table relative-frequency \
  --i-table phyla-table.qza \
  --o-relative-frequency-table rel-phyla-table.qza
```

- To export from .qza format to tsv format (input format for LEfSe):

```
qiime tools export \
  --input-path rel-phyla-table.qza \
  --output-path rel table
biom convert -i feature-table.biom -o rel-phyla-table.tsv --to-tsv
```

- The relative abundance is then transformed to LEfSe format followed by performing LDA Effect Size on formatted dataset using Kruskal-Wallis and Wilcoxon test to test between classes ($p < 0.05$) and the threshold on logarithmic LDA score for discriminative features 2.0.
- The results obtained with discriminative features were plotted using 'Plot LEfSe Results' and 'Plot Differential Features'
- The cladogram for differential features was obtained using 'Plot Cladogram'

4.2.11 Functional prediction using PICRUSt2

We predicted functional abundances based on marker gene sequences (16S rRNA) using software PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Reference). We installed and activated picrust2-2.4.1 followed by running the picrust2 pipeline using the following command line:

```
picrust2_pipeline.py -s rep-seqs.fna -i table.biom -o picrust2_out_pipeline -p 1 -t sepp -e 0
```

where, data input is representative sequences in FASTA format (.fna) and feature tables in BIOM format (.biom). The key outputs produced were:

- EC_metagenome_out (Enzyme Commission) - Folder containing unstratified EC number metagenome predictions (pred_metagenome_unstrat.tsv.gz), sequence table normalized by predicted 16S copy number abundances (seqtab_norm.tsv.gz),

and the per-sample NSTI values weighted by the abundance of each ASV (weighted_nsti.tsv.gz).

- KO_metagenome_out (KEGG Orthologs)- As EC_metagenome_out above, but for KO metagenomes.
- pathways_out - Folder containing predicted pathway abundances and coverages per-sample, based on predicted EC number abundances.

Further, descriptions were added to each functional id in the output abundance tables using following commands:

```
add_descriptions.py -i EC_metagenome_out/pred_metagenome_unstrat.tsv.gz -m EC -o EC_metagenome_out/pred_metagenome_unstrat_descrip.tsv.gz

add_descriptions.py -i pathways_out/path_abun_unstrat.tsv.gz -m METACYC -o pathways_out/path_abun_unstrat_descrip.tsv.gz

add_descriptions.py -i KO_metagenome_out/pred_metagenome_unstrat.tsv.gz -m KO -o KO_metagenome_out/pred_metagenome_unstrat_descrip.tsv.gz
```

The outputs obtained along with the metadata file were visualised using STAMP (version 2.1.3) after testing using Student's t-test followed by Bonferroni correction. In all mentioned tests, a p-value < 0.05 was considered statistically significant. The results were plotted as post-hoc plots.

4.3 Results

4.3.1 Demographic and Clinicopathological data of participants

The Demographic and Clinicopathological data of participants from three group were documented and shown in Table 4.2 here. For healthy volunteers without any tobacco chewing or smoking habit, and without any history of oral cancer; the demographic data of 40 Controls was documented with 67.5% males and 32.5% females and mean age was 40 ± 10 years.

For Long Term Tobacco Chewers (T) group, tobacco chewing/smoking habit was a common demographic character. The demographic data of 40 tobacco chewers with no history of oral cancer, and no premalignant lesions was documented with 87.5% males and 12.5 % females and mean age was 43 ± 10 years. Of these, 75% individuals were only tobacco chewers, 15% were alcohol consumers along with tobacco chewing, and 5% were cigarettes smokers or have smoked and consumed alcohol along with tobacco chewing.

The cohort of Indian oral cancer patients consisted of 40 individuals with 87.5 % males and 12.5% females, with mean age 54 ± 11 years. The major cancer sites included buccal mucosa (42.5%), tongue (22.5%), mandible (17.5%), and cancer of retromolar trigone (2.5%), floor of mouth (5%), lip (2.5%) and maxilla/gingiva/alveolus (7.5%). The patients were diagnosed and categorised in grade II (75%) and grade III (25%) group. From the total oral cancer population, 40% were tobacco chewers, 5% were tobacco chewers and smokers, and 2.5% were tobacco chewers and alcohol consumers. The remaining 52.5% did not have any tobacco/alcohol/smoking related lifestyle habits .

4.3.2 Power analysis

We performed power analysis (http://www.openepi.com/Power/PowerCC.htm) to improve the confidence in the population statistics for the study groups in our study. While studying the variable as tobacco habit, the percentage of tobacco chewers in control group, tobacco group and OC group is 0%, 100% and 40% respectively. Using these values, the power based on normal approximation is >99% when each group is compared with the other, stating high confidence in the participants selected for the data.

Table 4.2: Participant Demographic Information

Table 4.2 (a): Gender distribution			
Sex	Control	Tobacco chewers	Oral Cancer patients
Male	27	35	35
Female	13	5	5
Table 4.2 (b): Age distribution			
Mean age (yrs)	Control	Tobacco chewers	Oral Cancer patients
	41±10	43±10	54±11
Table 4.2 (c): Lifestyle habits			
Habits	Control	Tobacco chewers	Oral Cancer patients
Tobacco chewing	NA	30	16
Tobacco + Smoking		2	2
Tobacco + Alcohol		6	1
Tobacco + Smoking + Alcohol		2	0

Table 4.2 (d): Primary site of cancer		
Site	Number of individuals	Percentage
Buccal mucosa	17	42.5
Tongue	9	22.5
Mandible	7	17.5
Retromolar trigone	1	2.5
Lip	1	2.5
Floor of mouth	2	5.0
Maxilla	1	2.5
Gingiva	1	2.5
Alveolus	1	2.5

Table 4.2 (e): Staging		
Grade	Number of individuals	Percentage
Grade II	30	75
Grade III	10	25

4.3.3 Qualitative and quantitative assessment of isolated DNA

The DNA extracted from 20 ml of oral wash was 5.0-227.0 ng/µl with 260/280 ratio range of 1.8-2.0.

4.3.4 Taxonomic assessment of sequences

Out of 120 samples that were subjected to V6-V8 region sequencing, eight samples failed the sequencing procedure, three samples belonged to control, one belonged to tobacco and four were from oral cancer group. The remaining 112 samples resulted in a total of 62,96,186 forward and reverse reads. The frequency of reads in different samples and sequence quality scores are shown in Fig. 4.1 and 4.2. An average of 56,215 reads per sample were observed with a maximum of 155,742 reads and minimum 5458 reads.

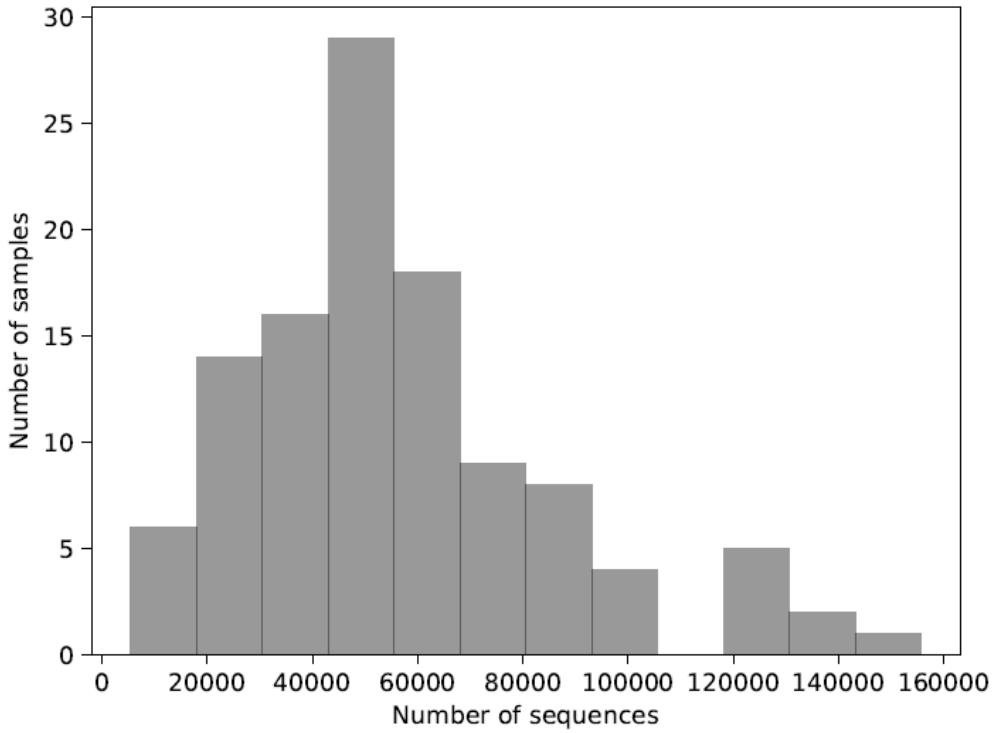


Fig. 4.1: Histogram depicting frequency of reads in samples.

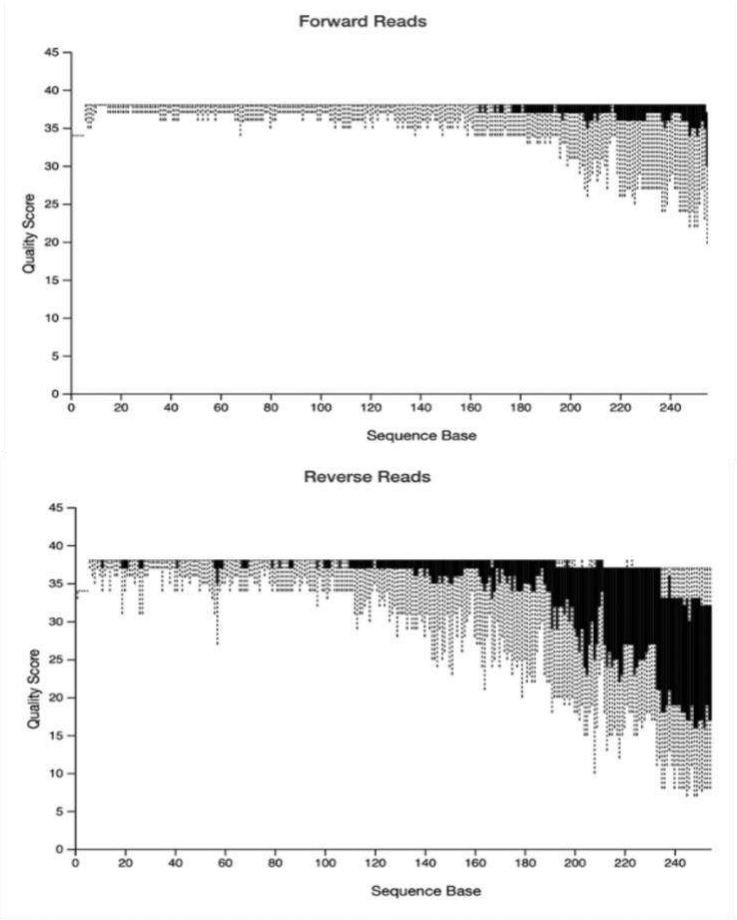


Fig. 4.2: Plots depicting sequence quality scores in forward and reverse reads.

The forward and reverse reads were joined using v-search (Fig. 4.3) and 54,07,163 reads were retained after obtaining approximately 430bp reads.

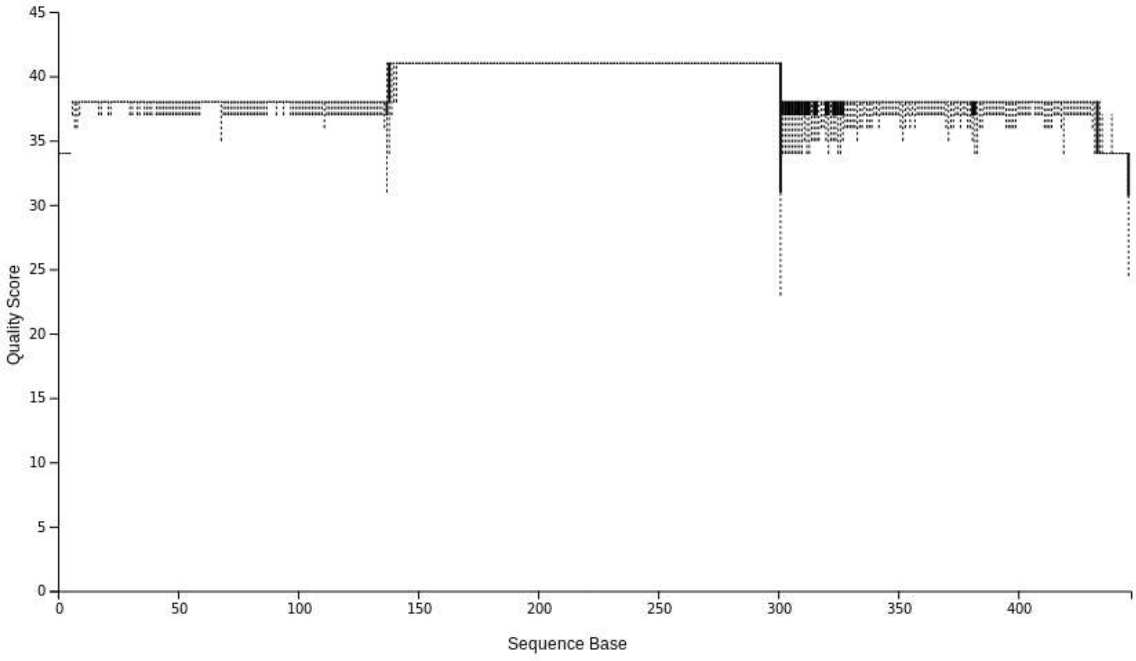


Fig. 4.3: Plots depicting sequence quality scores in joined reads.

After filtering reads, they were clustered into a total of 6,733 OTU's belonging to 9 phyla, 17 classes, 30 orders, and 55 families, and 94 genera. The bacteria belonged to phyla Bacteroidetes, Proteobacteria, Firmicutes, Actinobacteria, Fusobacteria, Spirochaetes, Cyanobacteria, Tenericutes and one unknown phyla (Fig 4.4). Out of total numbers,~approximately 99% bacteria belonged to five phyla namely, Bacteroidetes, Proteobacteria, Firmicutes, Actinobacteria, Fusobacteria. Phyla Bacteroidetes dominated all the study groups, followed by phyla Proteobacteria and Firmicutes.

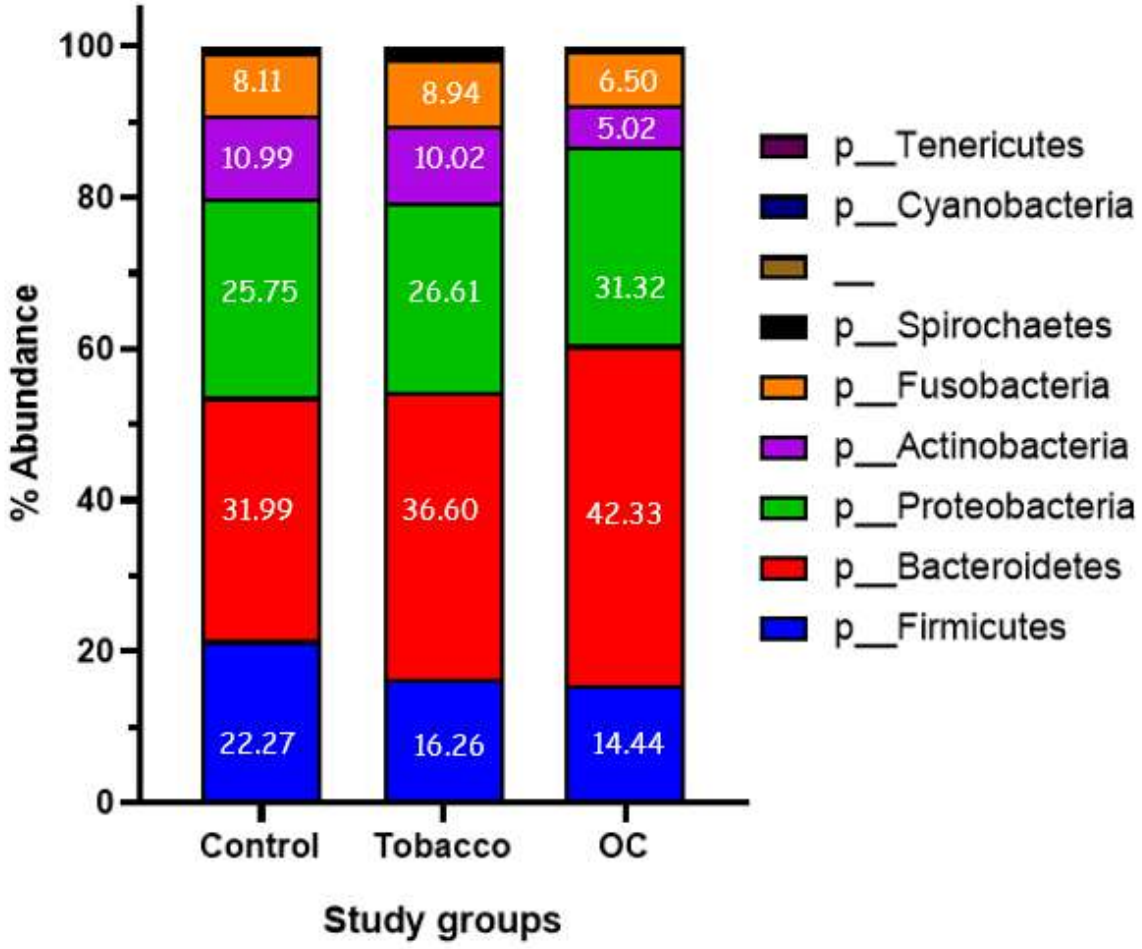


Fig. 4.4: Stacked bar chart depicting relative abundance of phylum-level bacterial composition in the oral microbiome of the control, tobacco chewing and oral cancer Indian population.

Also, out of the 94 genera identified, 15 genera take up about~ 85% of the total bacterial composition (Fig 4.5). In the control population, *Prevotella* (16.4%), *Neisseria* (13.5%), *Streptococcus* (10.9%), *Rothia* (7.85%) *and Porphyromonas* (7.4%) were the top most

abundant genera. In the tobacco chewing study group, *Prevotella* (20.6%), *Neisseria* (12.3%), *Porphyromonas* (7.11%), *Streptococcus* (7.4 %) and *Rothia* (6.81%) were predominant. On the other hand, *Prevotella* (23.6%), *Neisseria* (11.3%), *Paraprevotella* (9.1%), *Porphyromonas* (7.1%), *Streptococcus* (5.5%) were five most abundant genera observed in oral cancer study group.

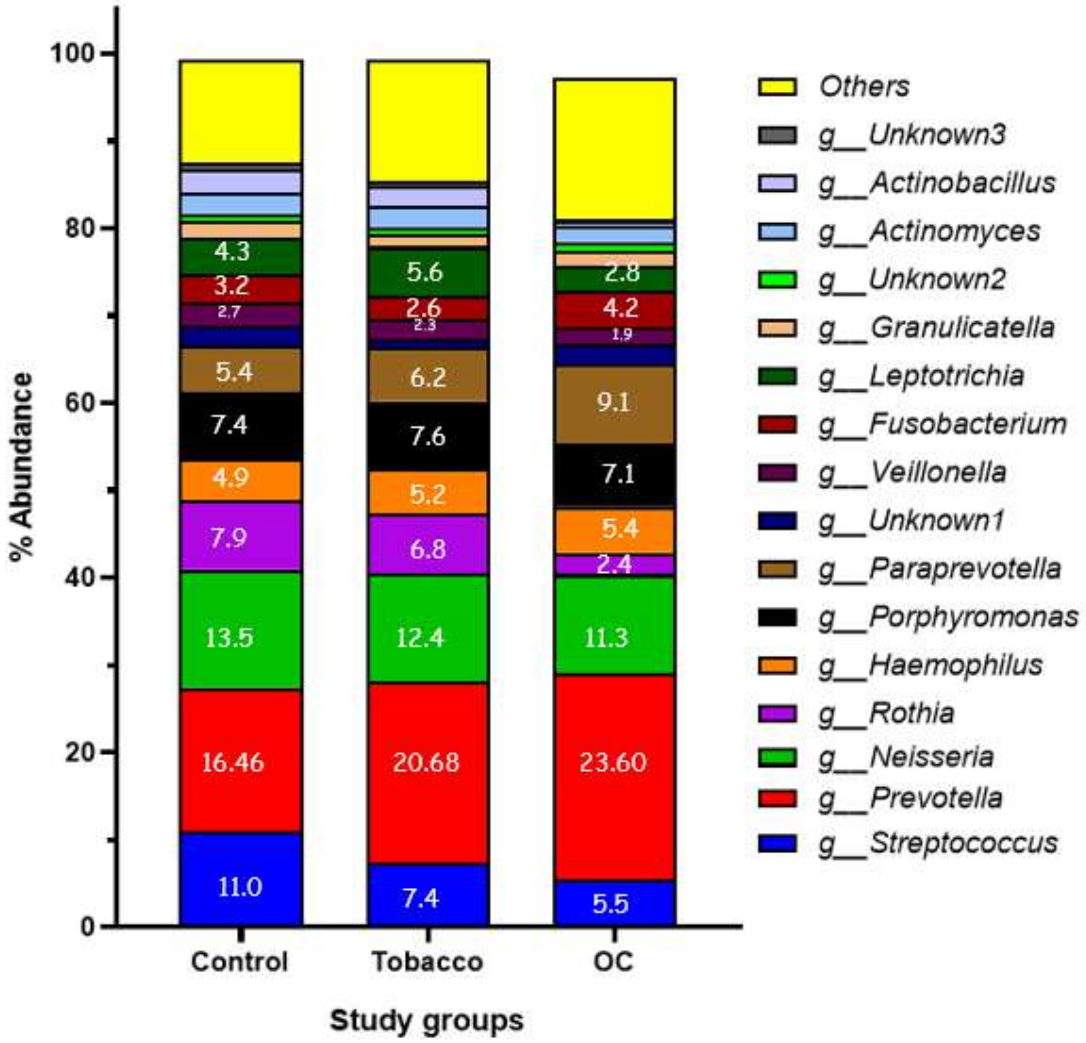


Fig. 4.5: Stacked bar chart depicting relative abundance of genus-level bacterial composition in the oral microbiome of the control, tobacco chewing and oral cancer Indian population.

When further analysis was performed for significant bacterial populations (Fig 4.6), we found that the abundance of genus *Streptococcus, Neisseria, Rothia* and *Veillonella* were highest in the control population, and decreased in the Tobacco group, followed by the OC group. On

the other hand, abundance of genera *Prevotella, Fusobacterium* and *Porphyromonas* were lowest in the control group and highest in the OC group, with tobacco group showing abundance in between that of control and OC group. The abundance of genus *Haemophilus* was highest in the tobacco group as compared to control and OC groups.

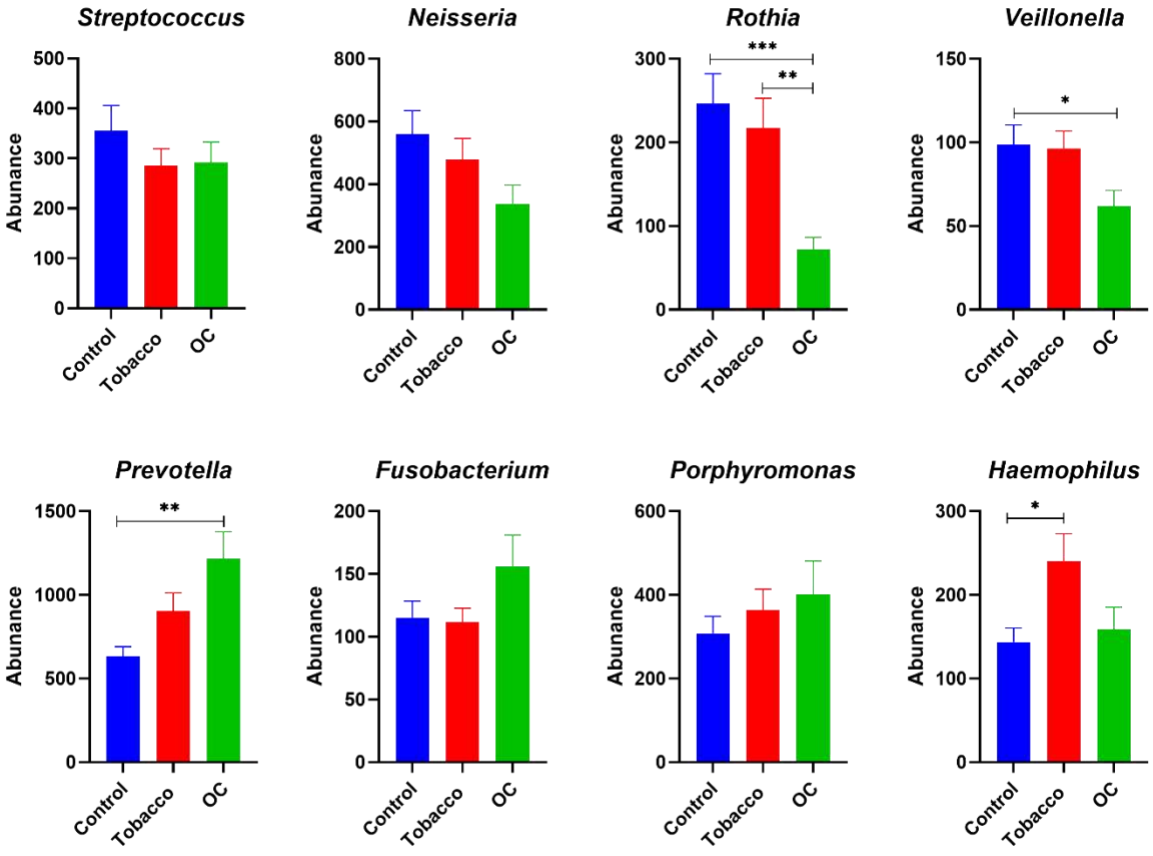


Fig. 4.6: Bacterial abundances in control, tobacco and OC study group (a) *Streptococcus*, (b) *Neisseria*, (c) *Rothia*, (d) *Veillonella*, (e) *Prevotella*, (f) *Fusobacterium*, (g) *Porphyromonas* and (h) *Haemophilus*. Data depicted is mean ± SD. Asterisk above columns indicate statistically significant difference compared between the three study groups (∗∗∗ =p <0.001, ∗∗ =p <0.01, ∗ =p <0.05) on GraphPad Prism software (version 8.0.2).

4.3.5 Microbial biomarkers

The unique bacterial community composition associated with the oral rinse was investigated using LEfSe analysis to compare the relative abundance of taxa across the C, T and OC groups. A total of 27 bacterial genera were observed to be differential in the three study groups (Fig. 4.7). *Leptotrichia, Treponema, Lautropia, Tannerella, Selenomonas, Filifactor, Campylobacter* and *Cardiobacterium* were identified as potential biomarkers for the tobacco group. On the other hand, *Pseudomonas, Capnocytophaga, Mycoplasma, Bifidobacterium, Peptostreptococcus* and *Paludibacter* were associated as biomarkers for the OC group. Bacteria belonging to genera *Rothia*, *Neisseria, Actinobacillus, Veillonella* and *Corynebacterium* along with four unknown genera were identified as potential biomarkers for control population.

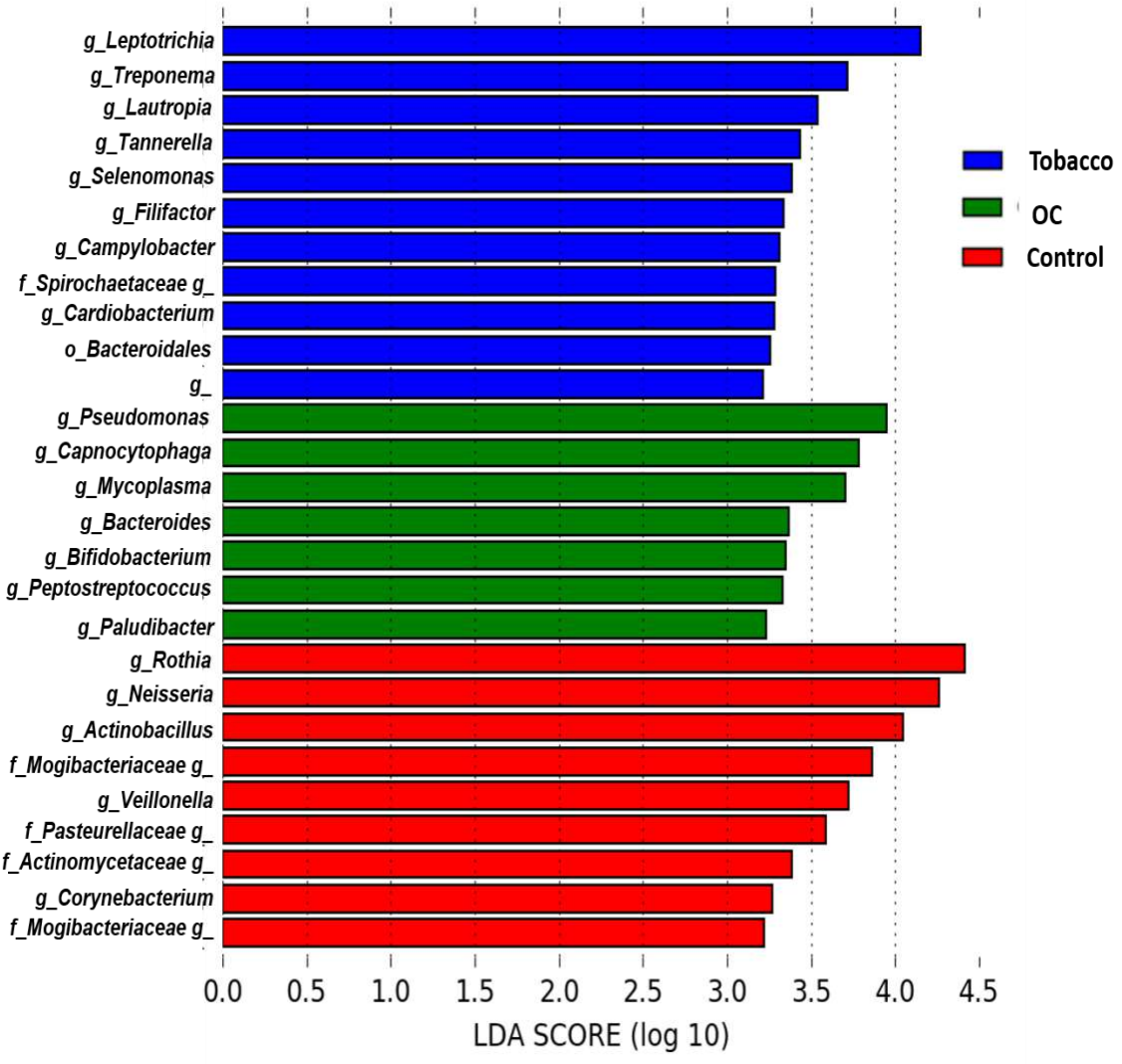


Fig. 4.7: Bacterial biomarkers identified for tobacco, OC and control study groups by LEfSe analysis.

Relative abundance of differential features for each sample for tobacco group (Fig 4.8), OC group (Fig 4.9) and control group (Fig 4.10) was plotted in a bar graph to observe the distribution of the biomarker in every sample. The graph displays the mean and median value of each column.

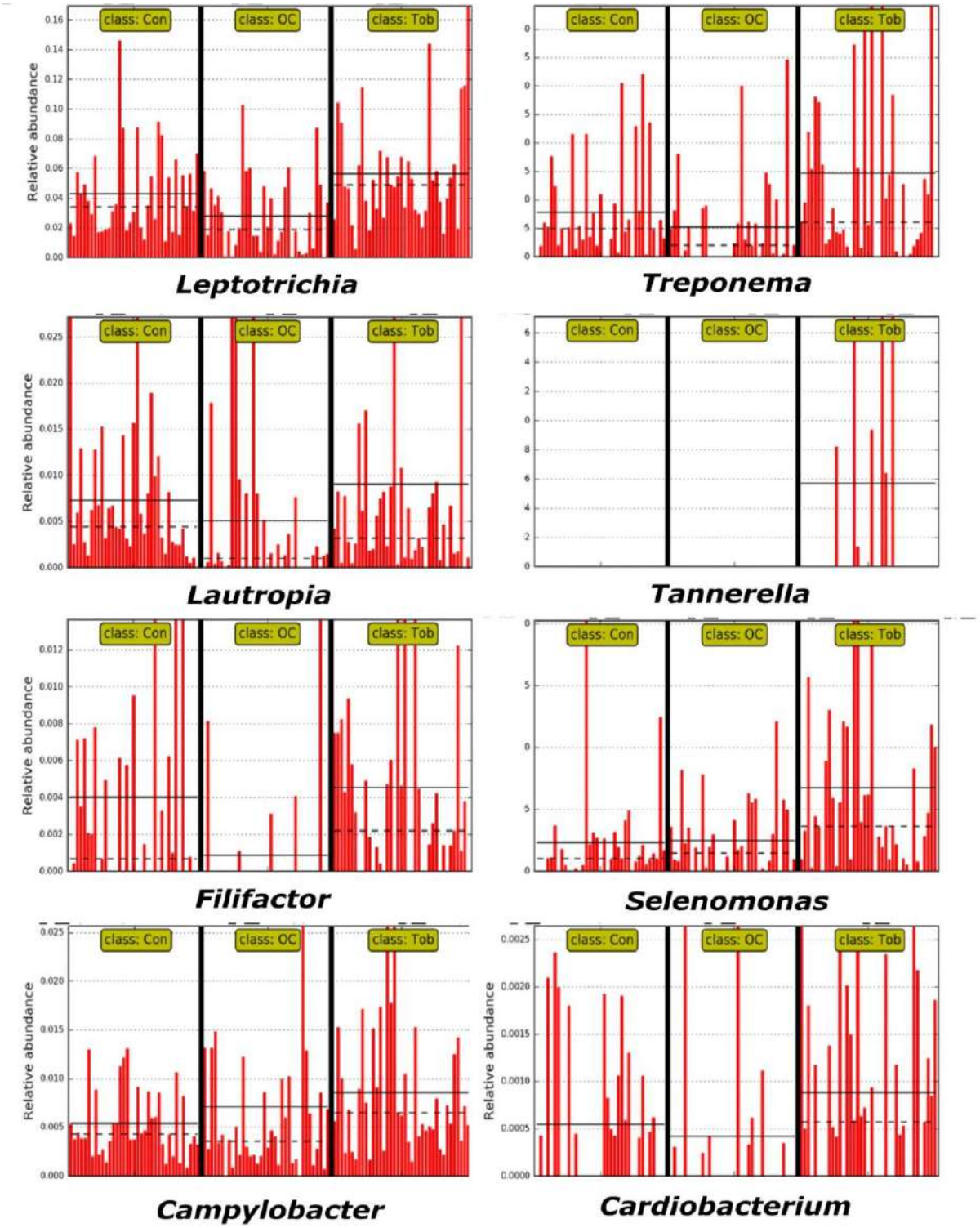


Fig. 4.8: Biomarkers identified for tobacco group (a) *Leptotrichia*, (b) *Treponema*, (c) *Lautropia*, (d) *Tannerella*, (e) *Filifactor*, (f) *Selenomonas*, (g) *Campylobacter* and (h) *Cardiobacterium*. Straight line represents mean of study group, dotted line represents median of study group.

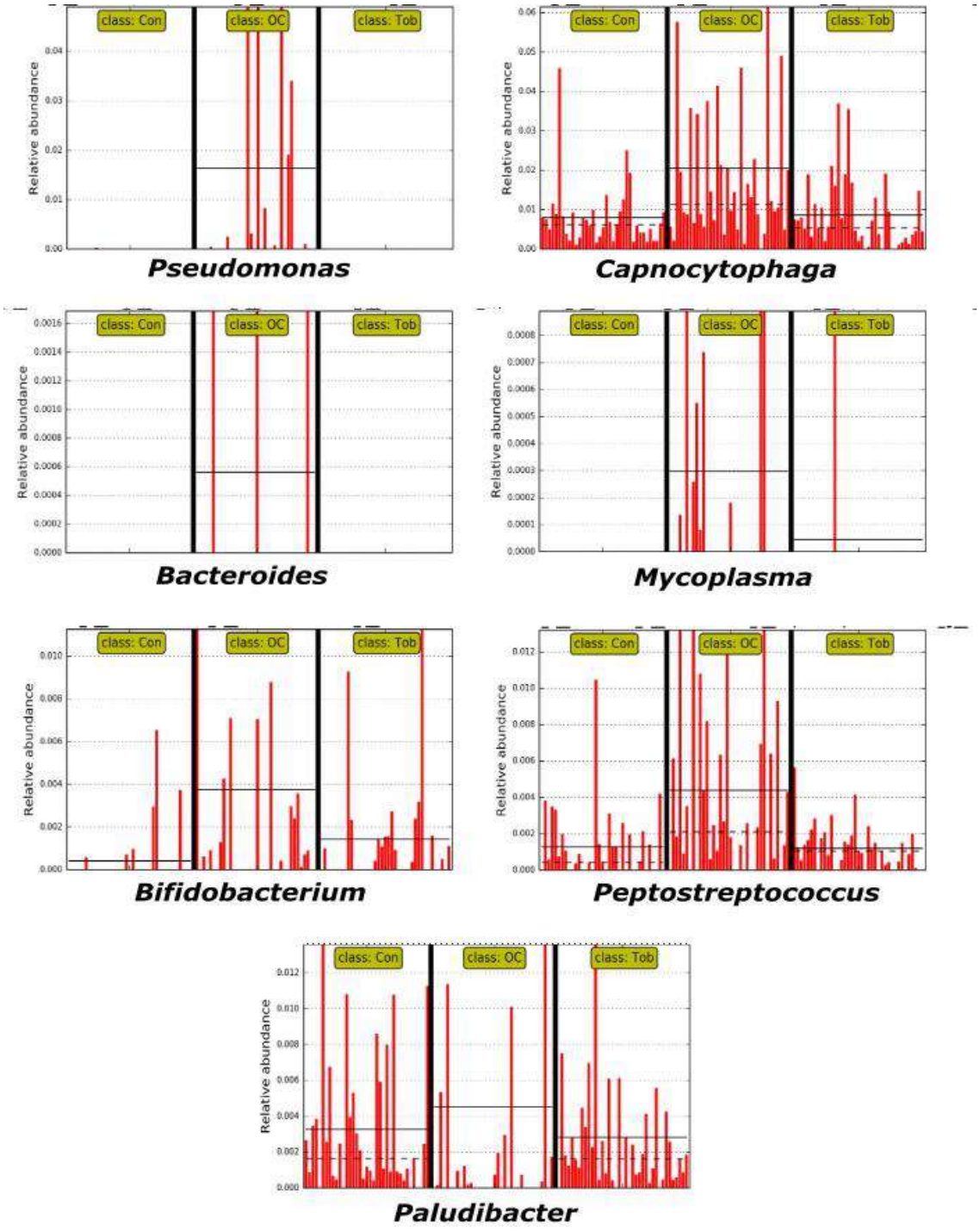


Fig.4.9: Biomarkers identified for OC group (a) *Pseudomonas*, (b) *Capnocytophaga*, (c) *Bacteroides*, (d) *Mycoplasma*, (e) *Bifidobacterium*, (f) *Peptostreptococcus* and (g) *Paludibacter*. Straight line represents mean of study group, dotted line represents median of study group.

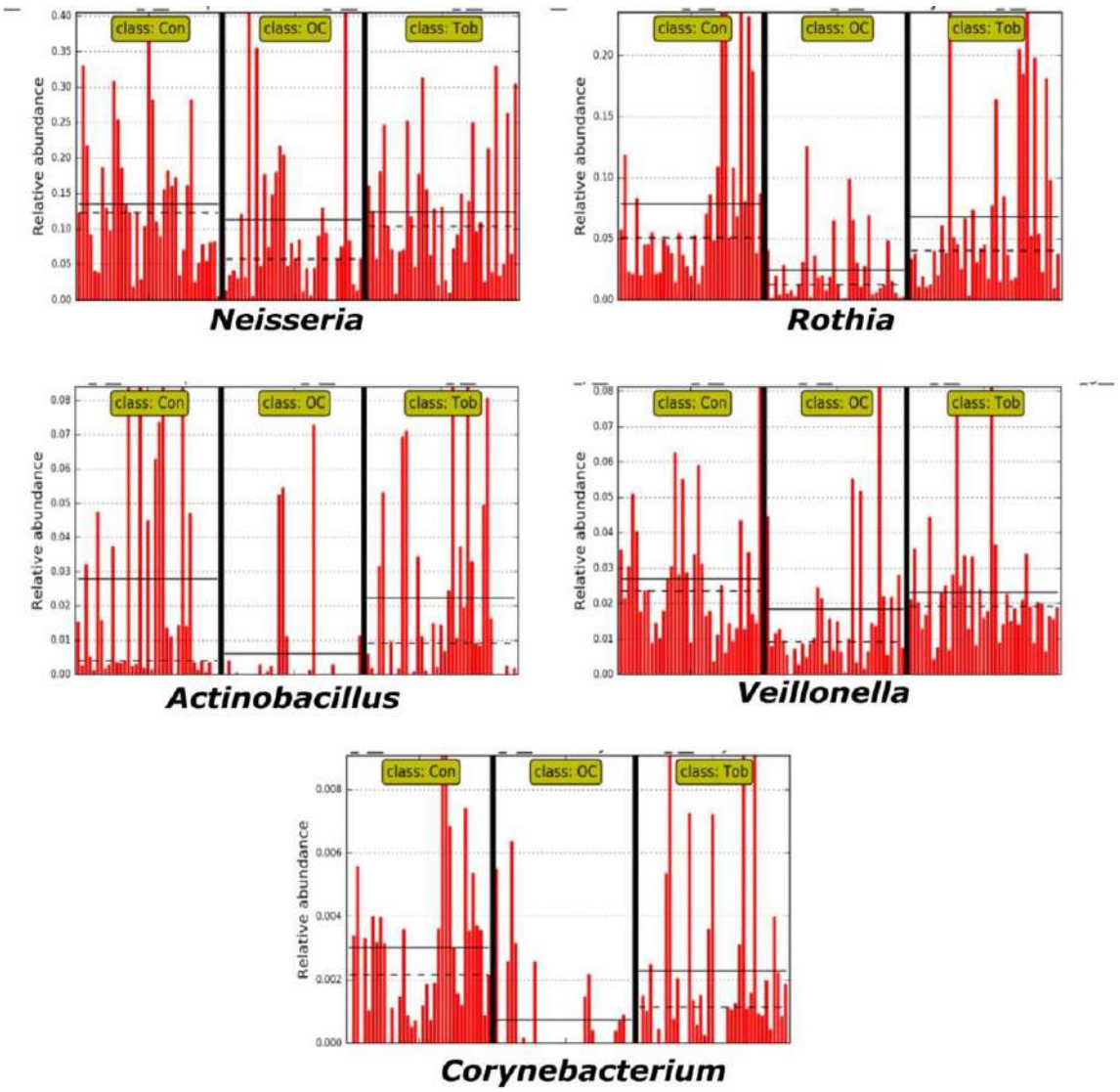


Fig. 4.10: Biomarkers identified for Control group (a) *Neisseria*, (b) *Rothia*, (c) *Actinobacillus*, (d)*Veillonella* and (e) *Corynebacterium*. Straight line represents mean of study group, dotted line represents median of study group.

A cladogram (Fig 4.11) was plotted to identify phylogenetic relationship between all Operational Taxonomic Units (OTU) identified in the study population, with highlighting the biomarkers identified for each study group. The nodes in yellow belong to nonsignificant bacteria whereas the ones in colour, belong to each biomarker identified in corresponding study group.

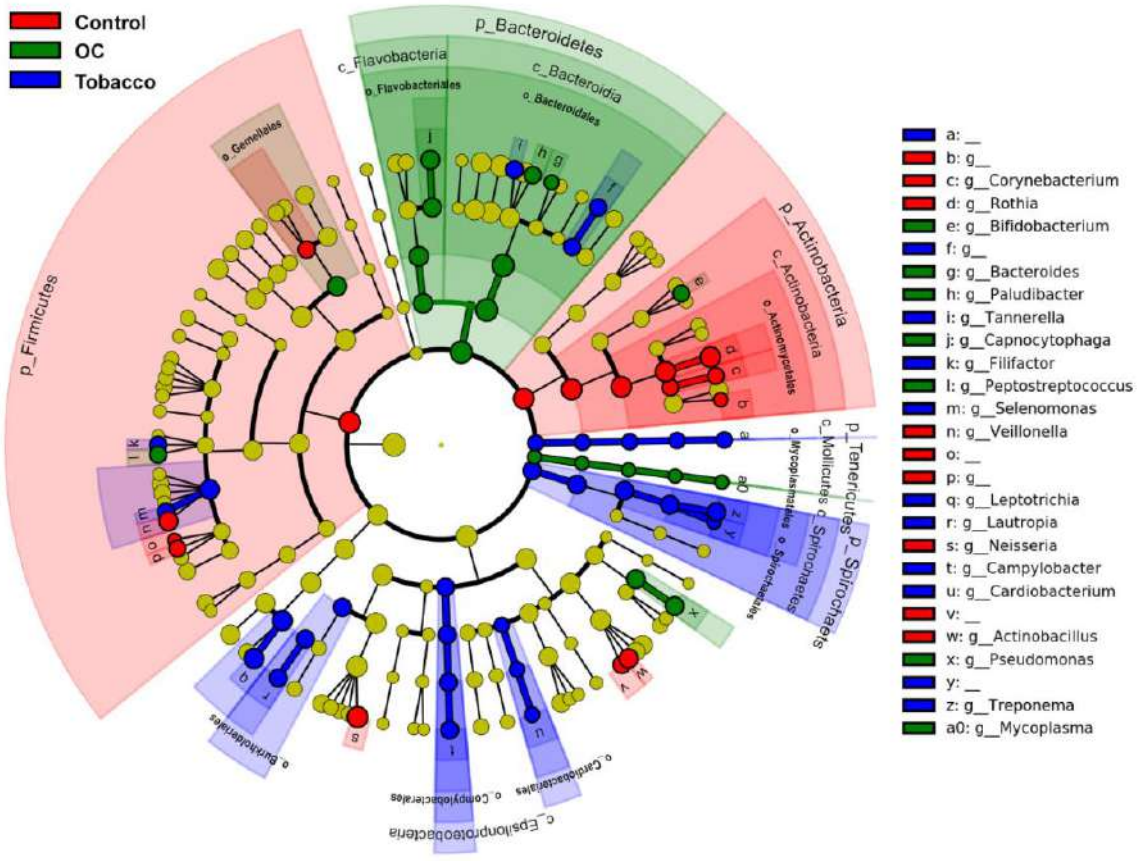


Fig. 4.11: The cladogram shows the microbial species with significant differences in the three groups, with the species classification at the level of phylum, class, order, family, and genus shown from the inside to the outside.

4.3.6 Alpha diversity

The alpha diversity indices were analysed to assess the bacterial diversity within the study groups (Fig. 4.12). Alpha diversity (α-diversity) is defined as the mean diversity of species in different sites or habitats within a local scale. Firstly, rarefaction curves were studied for estimating species richness. The number of OTU's are plotted against sequencing depth. This helps in confirming that sufficient sequence depth were achieved per sample to capture entire sample richness and diversity.

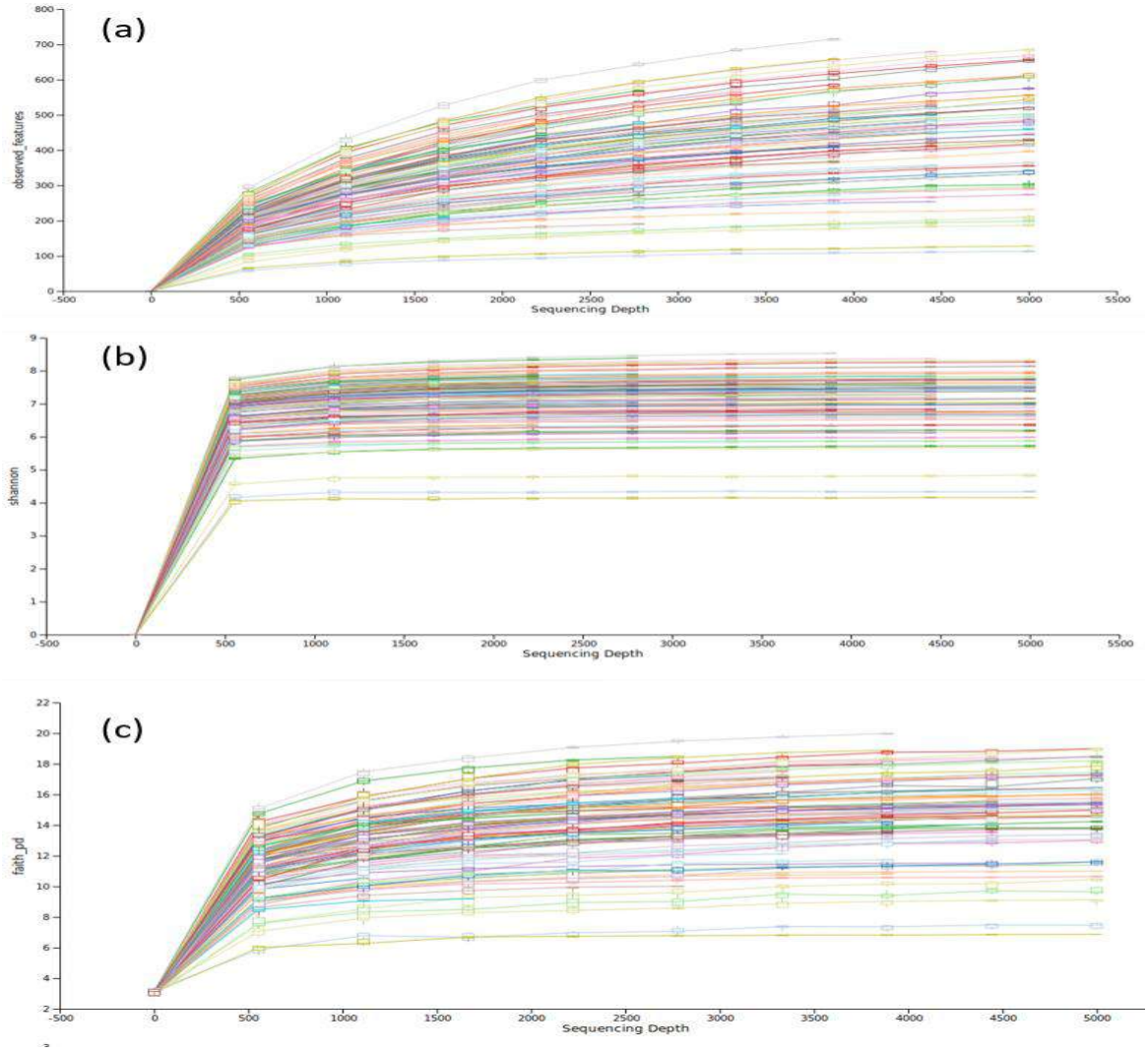


Fig.4.12: Rarefaction curves up to sampling depth 4500 sequences (a) Observed OTU's, (b) Shannon, (c) Faith_pd curves. Each curve represents a single sample.

Alpha (within-group) diversity indices like observed OTUs, Ace index, Chao1, Goods coverage, Shannon and Simpson indices, and Pielou_e were used to understand the species richness and diversity of samples (Fig. 4.13 and Table 4.3). Good's coverage was >96% for sequences

in all the study groups, indicating that the sequences measured in each sample represented almost all the bacterial sequences in the sample. Significantly higher numbers of mean observed OTUs were observed in tobacco chewers and control population (486 ± 127 ,464 ± 111 respectively) as compared to the OC group (355 ± 140). Other alpha diversity indices like those of species richness (ACE/Chao1) and diversity index (Shannon index) also depict statistically higher alpha diversity observed in tobacco chewers and control population as compared to OC group, thereby indicating lowest alpha diversity in OC group.

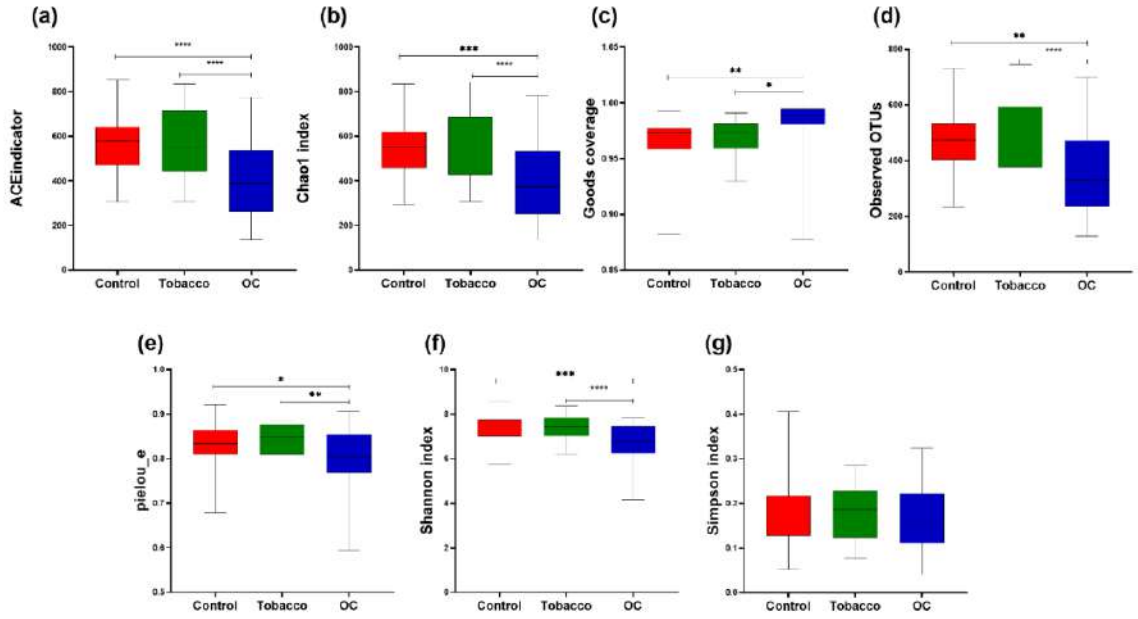


Fig. 4.13: Alpha diversity indices for the study groups; (a) ACE indicator; (b) Chao1 index; (c) Goods coverage; (d) Observed OTU's; (e) pielou_e (f) Shannon index; (g) Simpson index. Data plotted is mean ± SD. Asterisk above columns indicate statistically significant difference compared between the three study groups (*** =p <0.001, ** =p <0.01, * =p <0.05) on GraphPad Prism software (version 8.0.2).

Table 4.3: Alpha diversity indices assessed in study groups.

		Observed OTUs	ACE index	chao1	goods coverage	shannon	simpson	pielou_e
Control	Mean	464	557	539	0.966	7.344	0.1713	0.8323
	SD	111	132	130	0.020	0.627	0.077	0.051
Tobacco	Mean	486	578	562	0.969	7.450	0.1774	0.8388
	SD	127	145	147	0.016	0.524	0.060	0.039
OC	Mean	355	401	394	0.982	6.681	0.1633	0.7976
	SD	140	156	156	0.022	0.903	0.075	0.074

4.3.7 Beta diversity

Beta diversity (between-group) was studied using various parameters depicted in Table 4.4 and Fig.4.14. PERMANOVA allows you to assess whether the centres (centroids) of a cluster of samples differ between study groups statistically, whereas PERMDISP assesses whether the dispersion of samples within study groups are statistically significant. ANOSIM tests whether distances between groups are greater than within groups. When all three statistical tests are examined, PERMANOVA shows a significant difference in the mean composition of all three study groups, but PERMDISP and ANOSIM show that C-OC (p=0.003, 0.001) and T-OC (p=0.013, 0.001) have higher compositional diversity than C-T. To advocate beta-diversity results obtained to assess community dissimilarity, Bray-Curtis matrix, Jaccard matrix, Weighted and Unweighted Unifrac matrices were compared (Fig. 4.15). All beta-diversity matrices affirm the bacterial communities in the OC group and the controls-tobacco group clustered discreetly, suggesting the overall structures of the bacterial communities in the groups were significantly different.

Table: 4.4: Beta diversity indices assessed in study groups.

Comparison groups			PERMANOVA		PERMDISP		ANOSIM	
			pseudo-F	p-value	F-value	p-value	R	p-value
Control	OC	Unweighted	5.28	0.001	21.24	0.001	0.249	0.001
		Weighted	5.63	0.002	10.93	0.003	0.167	0.001
Control	Tobacco	Unweighted	1.91	0.003	3.69	0.079	0.0481	0.01
		Weighted	2.36	0.042	1.48	0.208	0.009	0.208
OC	Tobacco	Unweighted	5.58	0.001	4.76	0.035	0.236	0.001
		Weighted	2.96	0.022	6.30	0.013	0.101	0.001

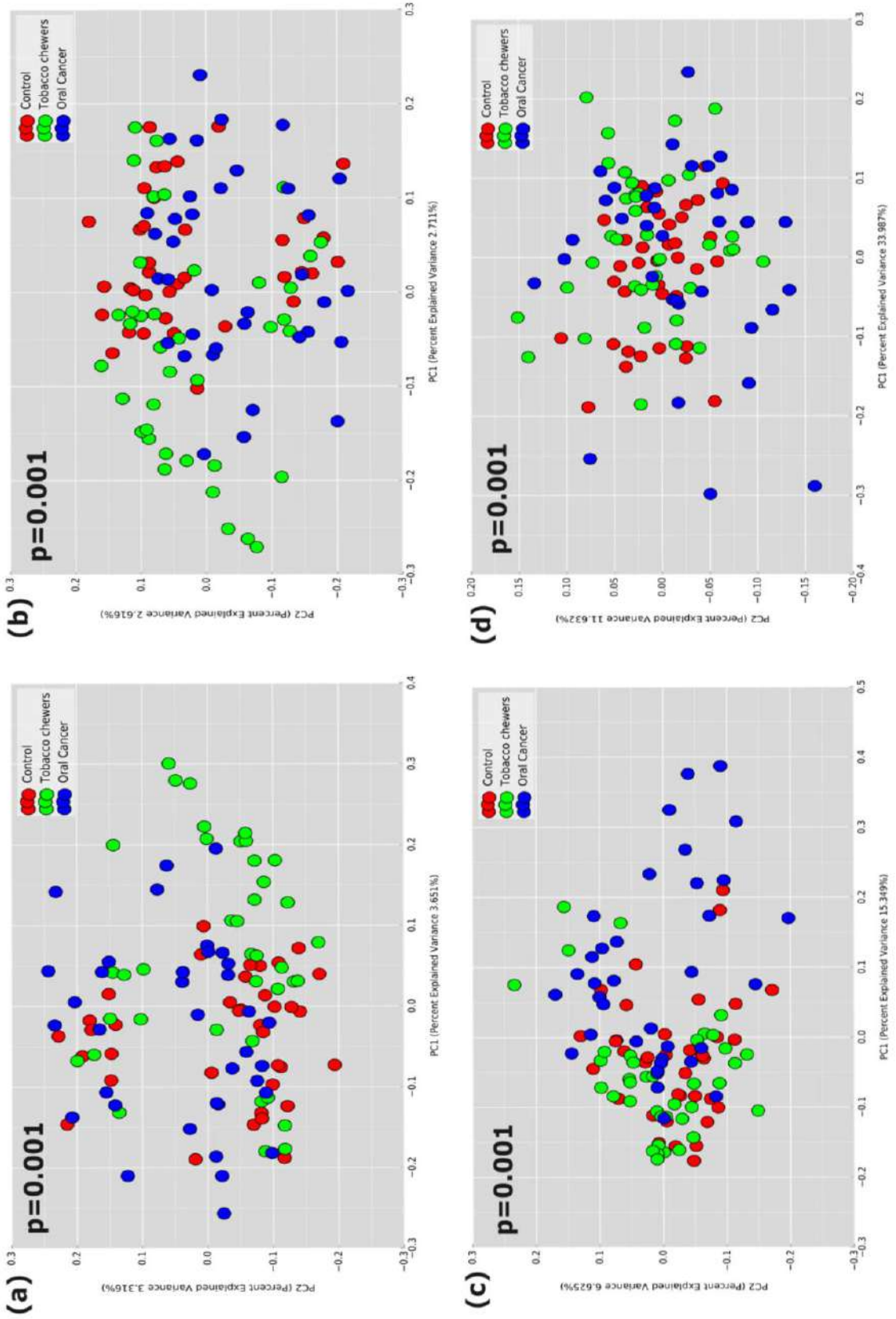


Fig.4.14: Beta diversity PCoA plots depicting sample diversity between groups; (a) Bray- Curtis plot; (b) Unweighted Unifrac matrix; (c) Jaccard matrix; (d) Weighted Unifrac matrix.

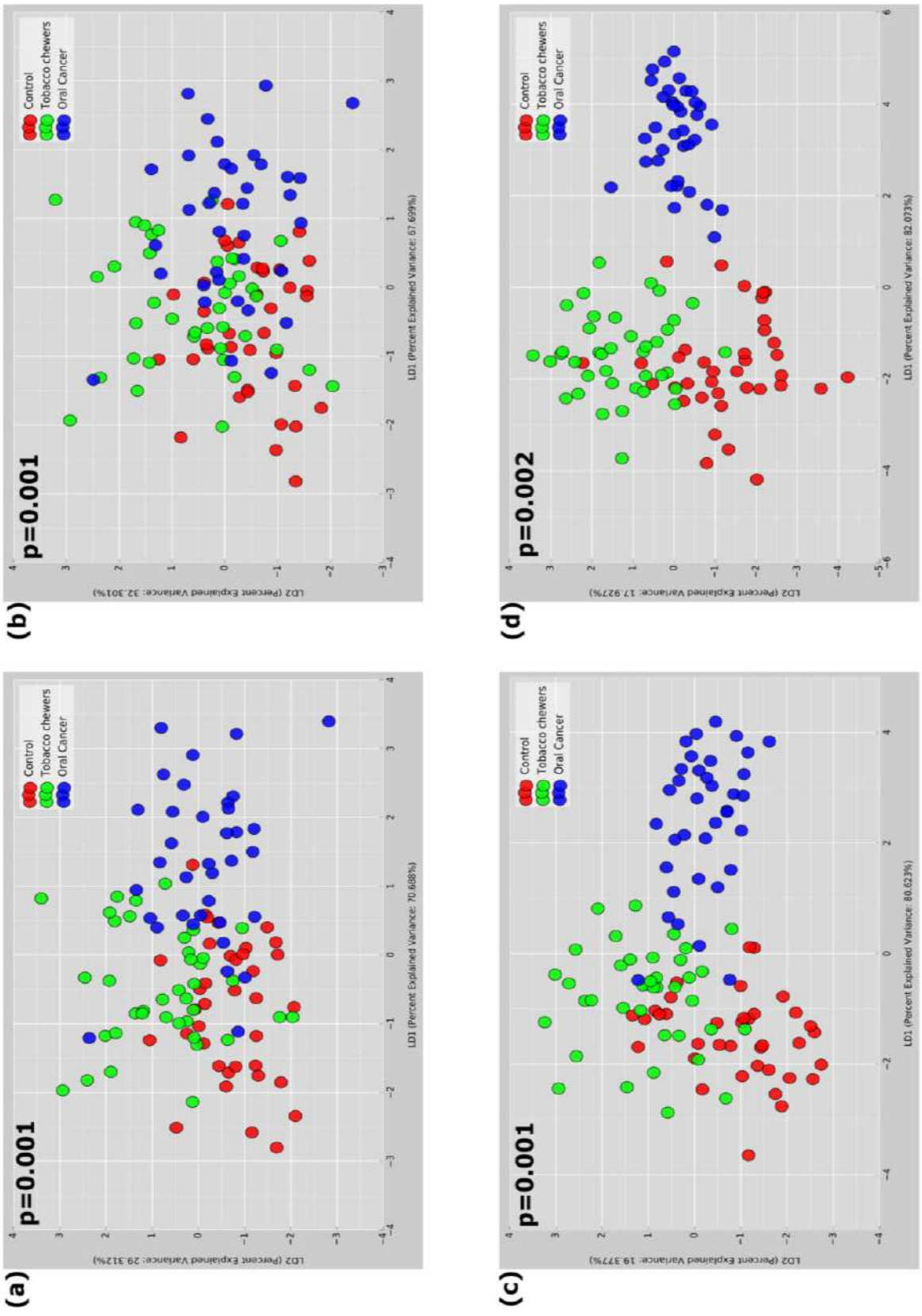


Fig.4.15: Beta diversity LDA plots depicting sample diversity between groups; (a) Bray- Curtis plot; (b) Unweighted Unifrac matrix; (c) Jaccard matrix; (d) Weighted Unifrac matrix.

4.3.8 Metagenome prediction

We used the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) method to envisage oral microbial roles linked to the formation of OSCC, and MetaCyc pathways were constructed for the study groups. PICRUSt2 estimates the gene families that are present, using an extended ancestral-state reconstruction technique, and then joins gene families to provide a comprehensive metagenome of the data. Significantly upregulated pathways, related to amino acid biosynthesis (aspartate, lysine, methionine, threonine, isoleucine, valine), sugar-fermentation (glycolysis, Entner-Doudoroff, pyruvate) and pyrimidine salvage and biosynthesis were detected in healthy controls as compared to OC group (Fig. 4.16). Conversely, pathways related to co-enzymeA (p= 0.024), aspartate, asparagine (p=0.023) and lipid biosynthesis (p=0.042), along with fatty acid elongation (p= 0.038) were upregulated in OC group as compared to controls (Fig. 4.17). Tobacco group revealed upregulated pathways related to reductive TCA cycle (p= 0.010) and pyrimidine biosynthesis (p= 7.95e-3), as compared to OC group (Fig. 4.18).

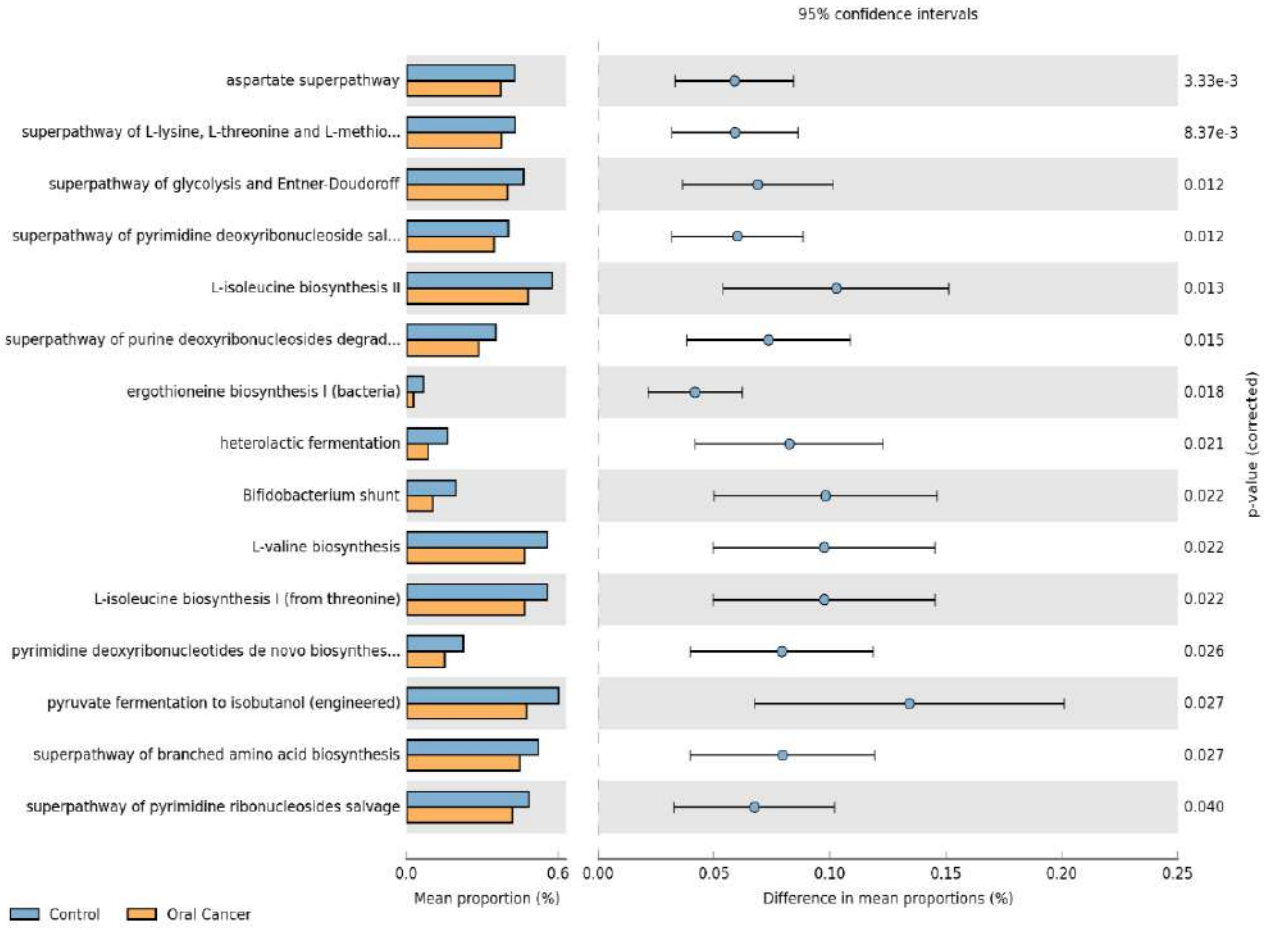


Fig. 4.16: Post-hoc plot depicting pathways enhanced in mean proportion of control population as compared to OC population.

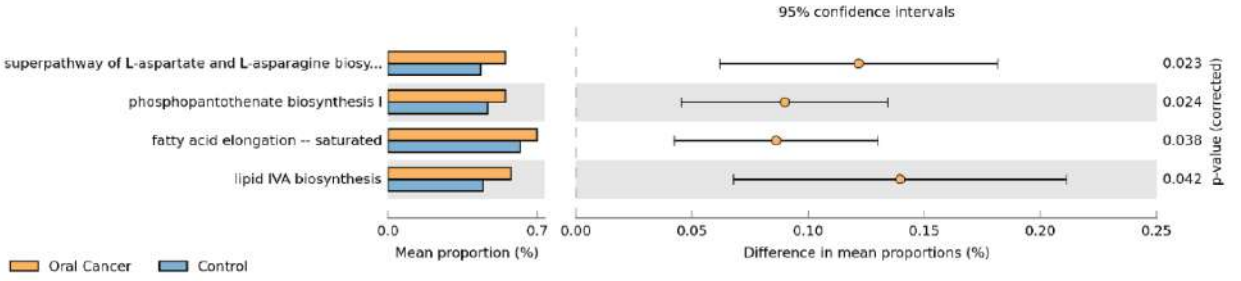


Fig. 4.17: Post-hoc plot depicting pathways enhanced in mean proportion of OC population as compared to controls.

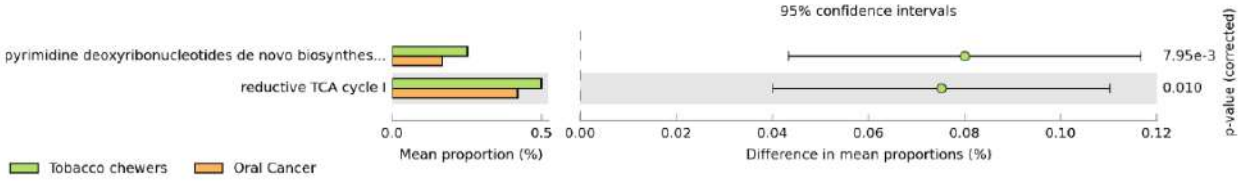


Fig. 4.18: Post-hoc plot depicting pathways enhanced in mean proportion of tobacco chewing population as compared to OC population.

4.3.9 Bacterial abundance based on variables

4.3.9.1 Shared and unshared bacterial abundance in study groups

On analysing the bacterial population in all three study groups, few genera were found only in either one or two study groups. These genera were identified and listed in Table 4.5 and shown as a Venn plot below in Fig. 4.18 .

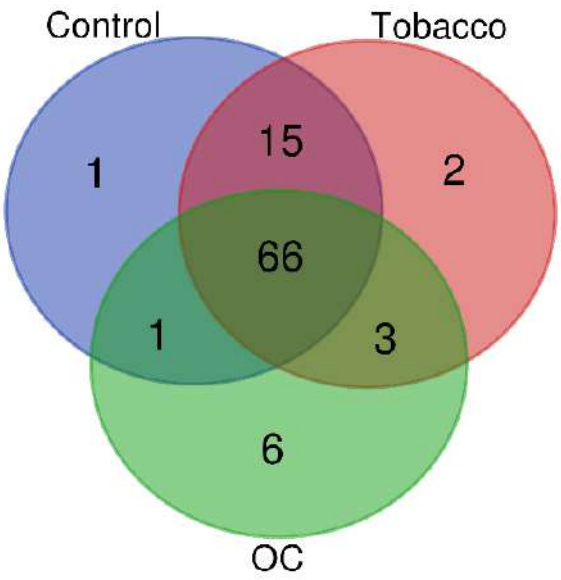


Fig. 4.19: Venn diagram depicting common and unshared bacteria in each study group.

Overall, 66 genera were shared between all three study groups. Out of 66, 15 genera including *Schwartzia, Desulfobulbus, Halomonas,* and 12 other genera were shared between control and tobacco group. Genus *Bilophilia* was the only genus common between control and OC. On the other hand, Genera *Acinetobacter, Mycoplasma, Desulfovibrio* were shared between OC and tobacco groups. The Genera *Pseudomonas, Morganella, Alloscardovia, Aeromonas, Bacteroides, Propionibacterium* were only present in the OC group.

Table: 4.5: List of shared and unique bacteria in study groups.

Study group	Number of genera	Genera
Control-OC-Tobacco	66	*Cardiobacterium, Oribacterium, Campylobacter, Scardovia, Slackia, Bifidobacterium, Veillonella, Granulicatella, Tannerella, Capnocytophaga, Abiotropia, Neisseria, Klebsiella, Atopobium, Actinomyces, Fusobacterium, Actinobacillus, Hameophilus, Porphyromonas, Treponema, Parvimonas, Lautropia, Sharpea, Filifactor, Mogibacterium, Lactobacillus, Kingella, Peptostreptococcus, Shuttleworthia, Prevotella, Selenomonas, Corynebacterium, Lachnoanaerobaculum, Catonella, Paludibacter, Leptotrichia, Megasphaera, Bulledia, Rothia, Paraprevotella, Peptococcus, Aggregatibacter, Streptococcus, Eikenella, Dialister, Moryella, Gemella, 19 unidentified genera*
Control-Tobacco	15	*Schwartzia, Desulfobulbus, Halomonas, 12 unidentified genera*
Control-OC	1	*Bilophilia*
OC-Tobacco	3	*Acinetobacter, Mycoplasma, Desulfovibrio*
Control	1	*Unidentified genus*
Tobacco	2	*Unidentified genera*
OC	6	*Pseudomonas, Morganella, Alloscardovia, Aeromonas, Bacteroides, Propionibacterium*

4.3.9.2 Bacterial diversity based on habits of tobacco chewers

The relative bacterial abundance was analysed with respect to various habits of tobacco chewers to find any relevant pattern for their abundance (Fig. 4.20). The abundance of *Prevotella* was highest in individuals involved in chewing tobacco and smoking, whereas that of *Porphyromonas*, *Rothia* and *Haemophilus* was lowest in this group.

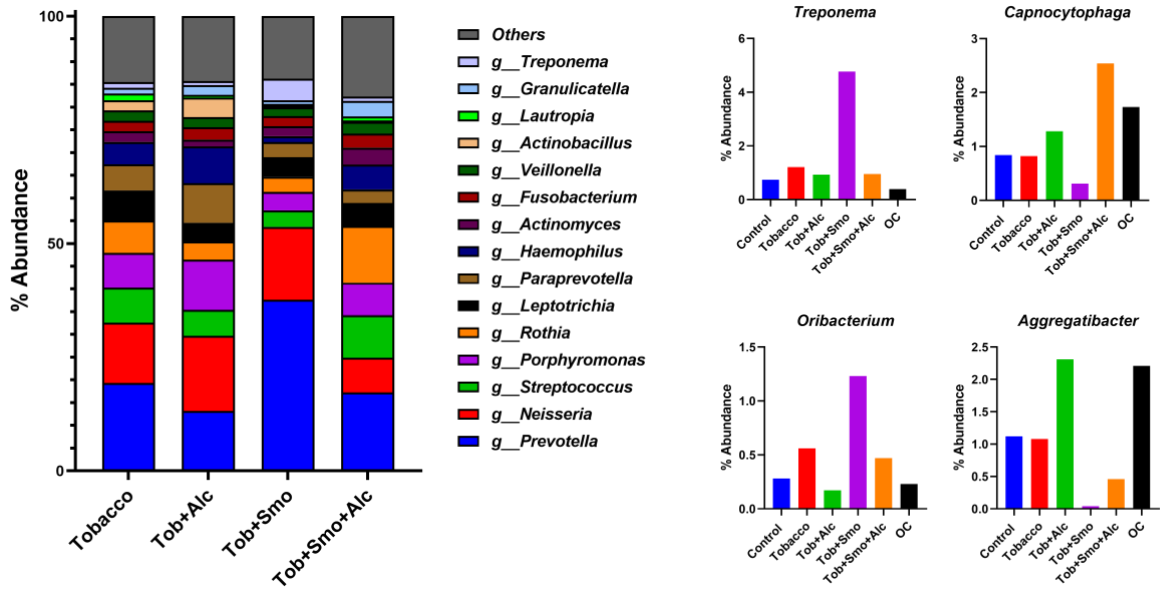


Fig.4.20: Abundance of bacteria in tobacco chewers depending on varying habits associated with tobacco chewing, alcohol consumption and smoking.

Important genera reported in literature such as *Treponema, Capnocytophaga, Oribacterium* and *Aggregatibacter* were studied in more detail. It was observed that abundance of *Treponema* was highest in individuals who were indulged in smoking along with tobacco chewing habit,. Since *Treponema* is a part of the red complex formation during periodontitis, it is possible that individuals who smoke and chew tobacco might have an increased risk of periodontitis. Similarly, the abundance of *Capnocytophaga* was highest in tobacco chewers who consumed alcohol and smoked . Since *Capnocytophaga* is a biomarker identified in our population for OC, the individuals with multiple risk factor exposure (tobacco + smoking + alcohol), might be at a higher risk of oral cancer. The abundance of *Oribacterium* was also highest in tobacco chewers and smokers as compared to other groups. The abundance of genus *Aggregatibacter* was highest in tobacco chewers and alcohol consumers. This level was similar to that observed in OC patients hinting towards increased risk of tobacco and alcohol consumers to progress towards oral cancer.

4.3.9.3 Bacterial diversity based on habits of oral cancer patients

We assessed the oral cancer patients based on their lifestyle habits like smoking, tobacco and alcohol consumption and results are shown below in Fig. 4.21.

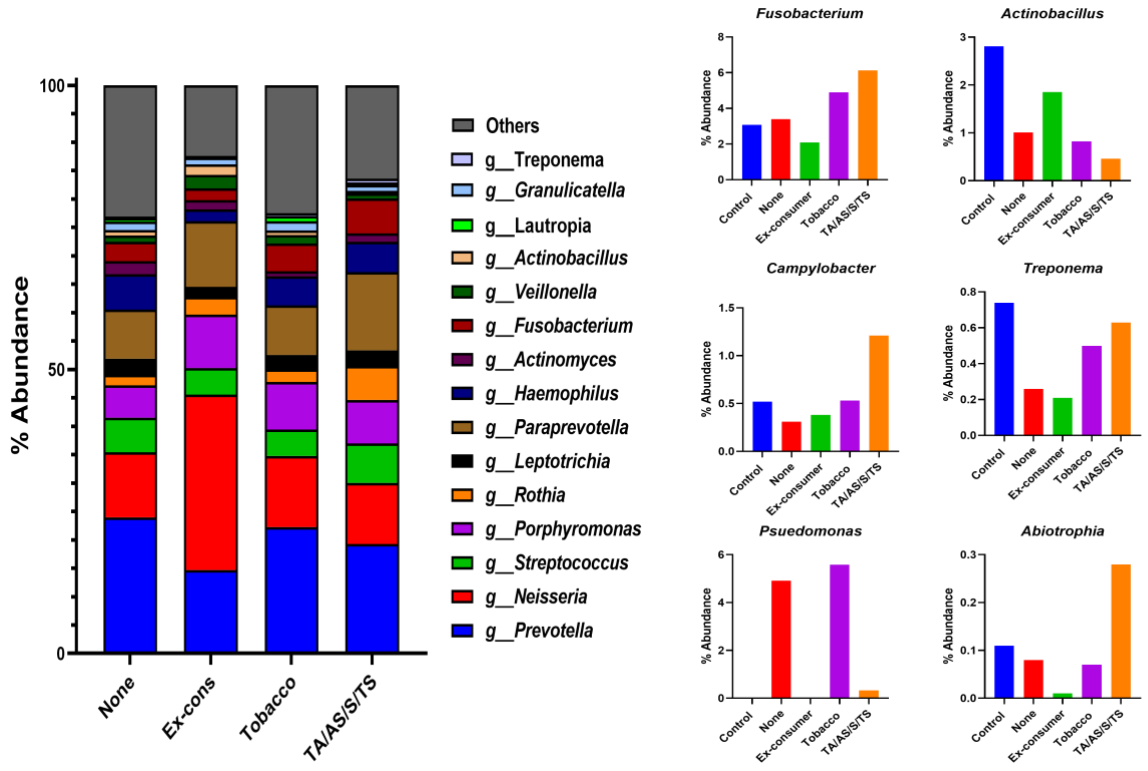


Fig. 4.21: Abundance of bacteria in patients with oral cancer depending on varying habits associated with tobacco chewing, alcohol consumption and smoking.

On inspecting the bacterial abundances, we found the proportion of bacteria in ex-tobacco consumers was similar to that of healthy controls. For example, the abundance of major genus *Neisseria* was highest in ex-tobacco consumers as compared to other groups. Neisseria is a biomarker for healthy participants in our study. We found that the abundance of *Fusobacterium, Campylobacter, Treponema, Abiotrophia* were highest in individuals with two or more risk factors combined (tobacco chewing, alcohol consumption, smoking). These are significant bacteria known to be present in oral cancer microbiota, and therefore their presence in individuals with multiple risk factors is evident. On the other hand*, Actinobacillus*, a biomarker for healthy individuals seems to be lowest in abundance in individuals associated with multiple risk factors which depicts high dysbiosis and loss of commensal healthy bacteria. *Pseudomonas* was found only in the oral cancer population and specifically in only tobacco chewers or individuals with no lifestyle risk factor.

4.3.9.4 Bacterial diversity based on site of cancer

Next, we analysed the bacterial abundance in oral cancer patients based on the site of cancer.

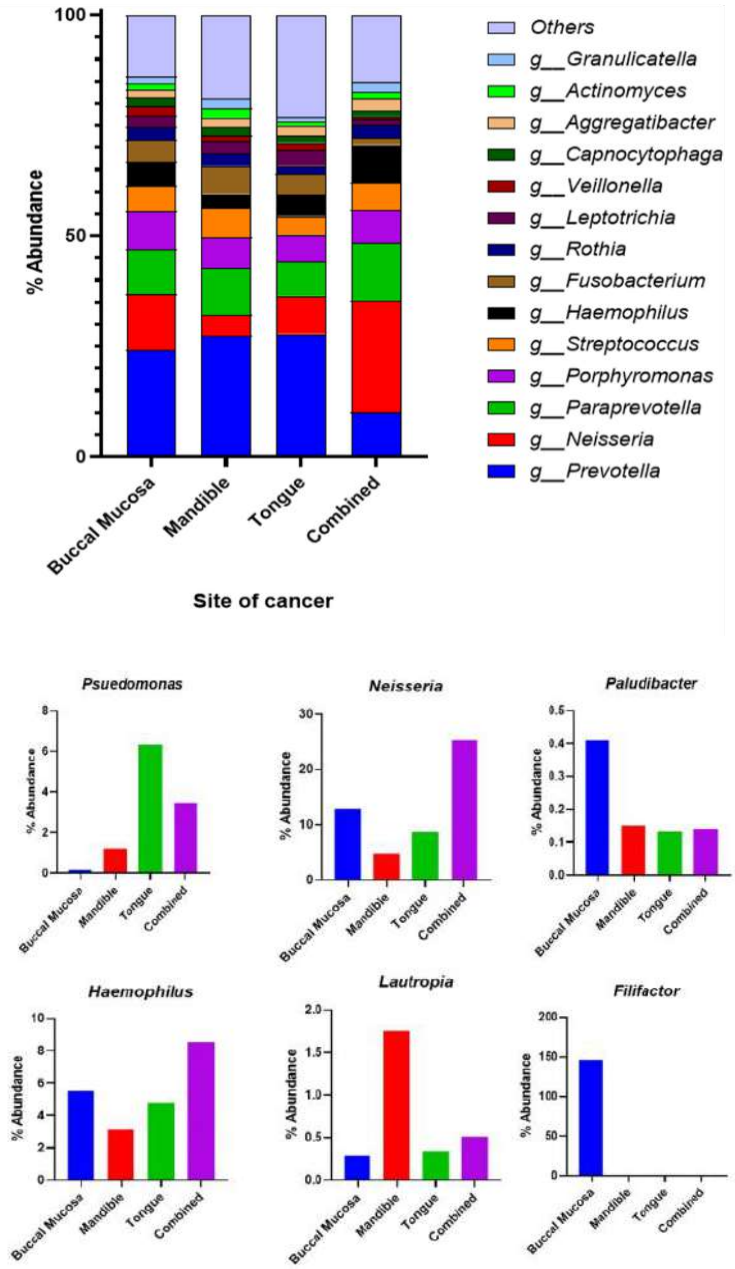


Fig. 4.22: Abundance of bacteria in patients with oral cancer depending on varying sites of cancer. (*combined = Floor of Mouth +Gingiva +Lip +Maxilla +Retromolar Trigone)

We found the abundance of *Pseudomonas, Neisseria* and *Haemophilus* was higher in individuals suffering from tongue, floor of mouth/gingiva/lip/maxilla/retromolar trigone cancer subjects (Fig. 4.22). The abundance of *Lautropia* was highest in cancer of the mandible, whereas *Paludibacter* and *Filifactor* were highest in cancer of the buccal mucosa.

4.3.9.5 Bacterial diversity based on grade of cancer

Study participants belonging to the oral cancer group can be segregated into either grade II or grade III cancer. On assessing the abundance data, the bacterial abundance of a major proportion of bacteria was higher in grade III cancer as compared to grade II cancer (Fig. 4.23). This could be evidently observed specially in periodontal pathogens and those that have been previously reported to be associated with oral cancer.

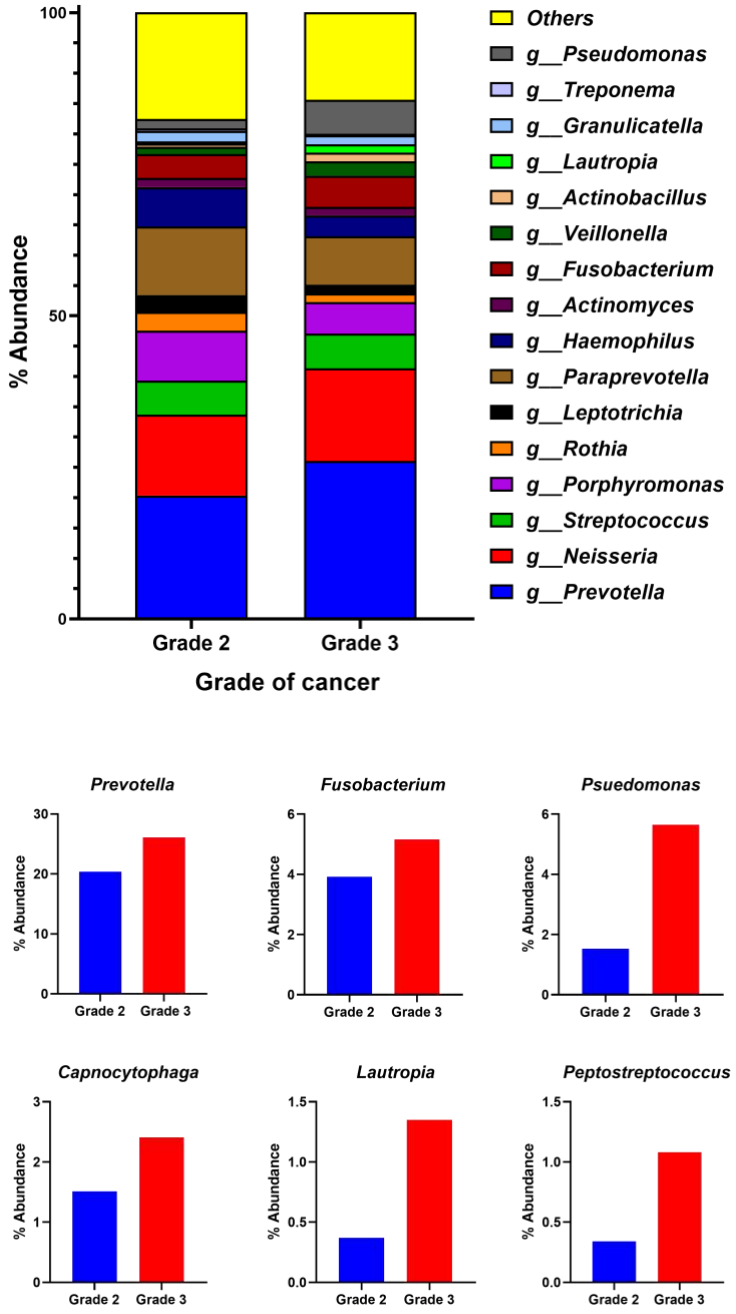


Fig. 4.23: Abundance of bacteria in patients with oral cancer depending on grade II vs grade III oral cancer.

4.3.9.6 Bacterial diversity based on diet of individuals

We identified individuals as those consuming vegetarian versus non-vegetarian diets, and tried to find if diet has any effect on the bacterial composition. In our study, there was no significant differences between vegetarian and non-vegetarian consumers in all study groups (Fig. 4.24). Therefore, we can infer that diet does not play a significant role in modulating the oral microbiome in our study population.

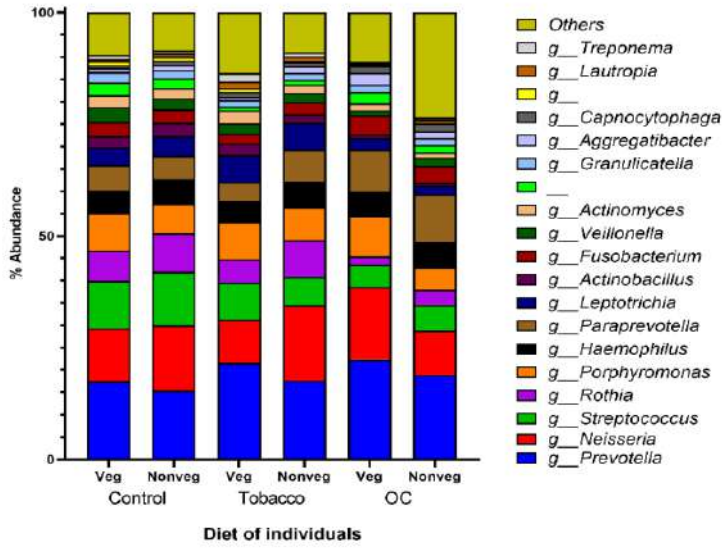


Fig. 4.24: Abundance of bacteria in participants based on diet of individuals.

4.3.9.7 Bacterial diversity based on gender of individuals

Similar to diet, gender of individuals did not show any significant effect on the oral microbiome in the three study groups (Fig. 4.25).

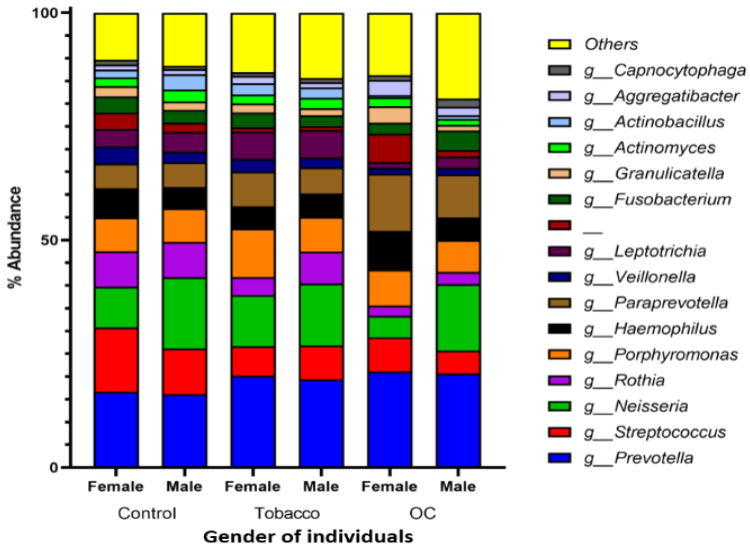


Fig. 4.25: Abundance of bacteria in participants based on gender of individuals.

4.4 Discussion

India is the world's second most populated country, with a population of 1.39 billion people, and it has been observed that oral health is largely ignored (Gambhir *et al.*, 2016; Gambhir and Gupta, 2016). Microorganisms play an increasingly important role in maintaining oral health in both healthy individuals and diseased conditions. Throughout the world, numerous oral microbiome-based studies have been conducted to enhance our understanding of bacterial dynamics in the context of various external factors as well as disease conditions, primarily cancer (Bik *et al.*, 2010; Amer *et al.*, 2017; Ling *et al.*, 2010). The Indian subcontinent's population, on the other hand, is extremely diverse in terms of ethnicity, culture, lifestyle, geographic location, and cuisine. The Indian population is exposed to a variety of lifestyle factors such as tobacco chewing, smoking, and alcohol consumption, and it ranks first in the world in terms of males suffering from oral cancer (OC). As a result of the high incidence rate and exposure to risk factors, the Indian population is an important demographic for studying the OC microbiome (Poddar *et al.*, 2018). There have been a few studies on the Indian oral microbiome, but the present study is the first report to compare the oral microbiomes of healthy controls, tobacco chewers, and OC patients.

In this study, we identified and reported that the oral microbiota, especially the bacteria residing in the oral cavity of healthy individuals, individuals chewing forms of tobacco for more than five years, and oral cancer patients in the Indian population. We have used oral saline rinse as a sample collection method that offers a non-invasive technique for better patient compliance and captures bacterial diversity throughout the oral cavity, as compared to tumor tissue samples. The power analyses increased our confidence in the study participants selected for the study. The samples were collected in the form of oral rinse in sterile saline to maintain the osmolarity of the sample followed by storage under low temperature conditions until DNA isolation was performed to ensure maximum integrity of the sample.

The 16S rRNA gene metagenome sequencing was performed at Seoul Women's University, Korea under the collaborative aid of Dr. Sathiyaraj Srinivasan using Illumina Miseq (Illumina, San Diego, CA). We aimed at obtaining 40,000 reads/sample and obtained a higher average of 56,215 reads/sample. According to the findings, five phyla account for 99% of the total oral microbiome composition (Fig 4.4), while 23 genera account for roughly 90% (Fig.4.5). In

comparison to the other two study groups, the OC group had the highest abundance of Bacteroidetes and Proteobacteria. These phyla predominantly contained gram-negative bacteria, namely Proteobacteria that was largely composed of gram-negative pathogenic bacteria. Apart from the five major phyla, tobacco chewers had the highest abundance of phylum Spirochaetes, and only tobacco and OC chewers had the presence of phylum Tenericutes, which could be associated with the presence of periodontal pathogens in tobacco chewers and diseased conditions in OC. At genus level, the abundance of *Prevotella* and *Neisseria* was highest in all the study groups. However, earlier reports have shown the highest abundance of *Streptococcus* in the oral cavity (Bik *et al.*, 2010; Gopinath *et al.*, 2021). The abundance of the genera *Streptococcus, Rothia, Veillonella*, and *Neisseria* was found to be lower in OC patients compared to healthy controls, which could be due to the fact that all of the above-mentioned genera are found in the healthy oral microbiota of humans. As a result, their high abundance in controls is due to a healthy status of the oral cavity, whereas their low abundance in OC and tobacco chewers could be due to oral dysbiosis (Lloyd-Price *et al.*, 2016). *Prevotella, Haemophilus*, and *Fusobacterium*, on the other hand, are known oral pathogens, which illustrates why their counts are higher in OC and tobacco chewing. Tobacco chewers have higher levels of *Prevotella* and *Fusobacterium*, which could lead to a synergistic activity of the bacterial toxins and nicotine, which can have negative health effects.

Next, we identified microbial biomarkers for each study group. We identified genera *Leptotrichia, Treponema, Lautropia, Tannerella, Selenomonas, Filifactor, Campylobacter* and *Cardiobacterium* as biomarkers for tobacco chewers (Fig. 4.7). Except *Cardiobacterium*, all the other genera have been reported to be associated with tobacco habits (Uehara et al., 2021; Xu et al., 2022). We report the presence of *Cardiobacterium* for the first time as a biomarker in individuals with tobacco chewing habits. On assessing the OC microbiota in our study population, *Pseudomonas, Capnocytophaga, Mycoplasma, Bacteroides, Bifidobacterium, Peptostreptococcus,* and *Paludibacter* were identified as biomarkers of oral cancer population. *Pseudomonas* and *Bacteroides* were found solely in the OC patients that can be used for clinical early diagnosis. This can be developed as a simple, specific, and non-invasive diagnostic method. Similarly, the healthy Indian population oral microbiota could be distinguished with increased abundance of genus *Rothia, Neisseria, Actinobacillus, Veillonella* and *Corynebacterium*. Aside from the biomarkers discovered, only a few genera have been

discovered in specific groups. *Acinetobacter, Mycoplasma*, and *Desulfovibrio*, for example, have only been found in tobacco and OC populations (Fig 4.19). Bacteria belonging to the genus *Morganella, Alloscardovia, Aeromonas, Propionibacterium*, along with *Pseudomonas* and *Bacteroides* have been identified only in the OC population. For the first time in our study, *Aeromonas, Alloscardovia*, and *Morganella* were identified as part of the OC oral microbiota.

Alpha and beta diversity matrices and indices were assessed to analyse the diversity within groups and between the groups respectively. Alpha (within-group) diversity indices such as ACE, Chao1 and Shannon signifies species richness and diversity within the sample group, Good's coverage confirms capture of diversity ,Pielou_e determines even distribution within the group, whereas Observed OTU's depicts number of features per sample. Collectively all alpha indices are used to assess the diversity within the study groups. Our results indicate maximum within group bacterial diversity in tobacco chewers, followed by healthy controls, and least diversity within oral cancer patients' group (Fig. 4.13). Similarly, beta diversity indices and matrices indicated close clustering of oral samples belonging to control and tobacco groups, whereas oral microbiota of individuals suffering from oral cancer was different and clustered away from that of control and tobacco chewers (Fig. 4.14 and Fig. 4.15). When the abundances of bacteria in all three study groups are considered, as well as the diversity indices, the oral microbiome of tobacco chewers appears to be somewhere between that of the control and that of the oral cancer population. PERMANOVA depicts statistically significant differences between all three study groups (Table 4.4). Although on analysing PERMDISP, the differences between control and tobacco chewers as compared to OC group was statistically significant, whereas the difference between control and tobacco chewers was insignificant and similar results were observed in ANOSIM. All three statistical tests along with the 3D plots support the similarity between control and tobacco group as compared to that of OC group clustering away.

The functional pathways in each study group predicted amino acid biosynthesis and sugar fermentation pathways being upregulated in control samples as compared to fatty acid and lipid biosynthesis pathways in OC samples. This can be a result of higher abundance of proteolytic/amino acid-degrading bacteria in the OC group, such as *Prevotella, Fusobacterium* and *Porphyromonas* species that degrade proteins and peptides into amino acids, which are then degraded further via specific pathways to produce short-chain fatty acids, ammonia,

sulfur compounds, and indole/skatole, all of which act as virulent and modifying factors in periodontitis, oral malodor and other oral health issues. Short-chain fatty acids and lipids are known to be cytotoxic, causing tissue inflammation and promoting apoptosis by modulating immune responses (Takahashi.,2015).

The effect of variables such as gender, age, lifestyle habits, grade and site of cancer, on oral microbiota from participant demographics was assessed. An individual's lifestyle habits (tobacco/alcohol), grade and site of cancer all demonstrated to alter the oral microbiome, in contrast to diet and gender, which did not significantly alter the oral flora.

Based on the subject demographic, the study may have certain limitations. Male participants are more numerous in the oral cancer and tobacco group than female participants. This is partly because males are more likely than females to use tobacco products, and more men than women are diagnosed with mouth cancer. Additionally, because age is a confounding risk factor for malignancies, including oral cancer, the study population includes participants in the OC group who are older than those in the control and tobacco chewing groups, potentially creating an age-related bias. The study may be limited by the inability to control the aforementioned variables, thus this should be taken into account.

4.5 Conclusion

In conclusion, the oral microbiota for the Indian population belonging to the three study groups (control, tobacco chewers and oral cancer patients) was identified and reported in this study. The pattern of abundance of bacterial genera was evaluated and biomarkers for each group was identified, diversity indices were studied, and study variables were assessed. Further, we aimed to develop a simple, effective and novel method for absolute quantification purposes using the qPCR method. This method would provide absolute quantification which could surpass faulty quantification due to multi-copy 16S rRNA and provide species specificity, and be used towards early diagnosis of oral cancer using the biomarkers identified for specific populations.

CHAPTER 5

Absolute Quantitation of Oral Bacteria Involved in Oral Cancer and Tobacco Chewers by Real-Time PCR

Absolute Quantitation of Oral Bacteria Involved in Oral Cancer and Tobacco Chewers by Real-Time PCR

5.1 Introduction

In this chapter we have developed quantitative PCR (qPCR) method, which is widely used for quantitation of gene expression at mRNA levels as well as for microbial ecology purposes, to determine the microbial load and diversity in any environmental sample (Nadkarni *et al.*, 2002; Abbott *et al.*, 1988). Microbial qPCR methods are intended to target highly conserved 16S rRNA genes to measure overall bacterial load in a sample, whereas hypervariable regions of this gene can be targeted for selective amplification and quantification of a specific bacterial population (Smith and Osborn, 2009a). Because every bacterial species contains varied copies (1–15 copies) of 16S rRNA, this technique has a disadvantage that must be noted, *i.e.* 16S rRNA gene copies cannot be accurately associated or correlated with cell counts or numbers (Klappenbach *et al.*, 2000). There are genes present as single copy per cell in bacterial species, such as nusA, pyrG, tsf, rplS, frr, which can be used instead of the 16S rRNA gene for quantitation of cell populations (Wu et al., 2013). One of such genes is the housekeeping gene ,RNA Polymerase-subunit gene *(rpoB)*, which is present as a single copy per cell in bacterial cells. Several studies have found that the hypervariable region of the rpoB gene is best suited for species and subspecies identification and phylogenetic discrimination (Adékambi et al., 2009), which is not always possible with the 16sRNA gene. We aimed to develop a reliable method for absolute quantification of bacteria using the *rpoB* gene (Case *et al.*, 2007). In the present study, assessment of the microbial population dynamics e in healthy subjects versus diseased conditions can help in understanding the dysbiosis and abundance of certain bacterial populations that could lead to the development of cancer.

5.1.1 qPCR in microbial ecology

The development of culture-independent techniques in microbial ecology has relied heavily on the use of PCR in conjunction with the extraction of nucleic acids (DNA and RNA) from environmental samples. These tools have been used to analyse the microbial communities existing within environmental systems since the early 1990s, and have significantly enhanced our understanding of microbial community composition, structure, and diversity under different environmental niches (VanGuilder *et al.*, 2008). Combining different techniques like

i) whole genome isolation ii) PCR amplification of taxonomic (i.e. rRNA) and functional gene markers, iii) DNA fingerprinting and sequencing-based analyses, have helped in identifying many of previously uncharacterized microorganisms. This has resulted in the discovery of new microbial lineages, description of genetic diversity in a variety of functional gene markers, and enrichment of different databases. In microbial ecology, quantitative PCR, or qPCR, is now commonly employed to assess the number of genes and/or transcripts present in environmental samples. The design of the primers and/or probe are the key factors that determines the target specificity of any qPCR experiment, allowing measurement of taxonomic or functional gene markers, present within a mixed population from the domain level down to particular species or phylotypes. Under different environmental or experimental settings, qPCR has been shown to be a robust, highly reproducible, and sensitive tool for quantitatively tracking phylogenetic and functional gene changes across temporal and spatial dimensions (Smith and Osborn, 2009b).

5.1.2 The rpoB gene

The 16S rRNA gene has become the most widely used molecular marker in microbial ecology due to several qualities, including its vital function, ubiquity, and evolutionary traits. However, one feature that has been ignored is that a single bacteria might have numerous copies of this gene. Because the sequences of these intragenomic copies might differ, several ribotypes for a single organism can be identified (Dahllof *et al.*, 2000) . Although the 16S rRNA gene is by far the most commonly utilised gene in molecular microbial ecology, this gene is not the only one that may be employed. The RNA polymerase subunit gene (rpoB) is an alternative core single copy housekeeping gene that can be used for community analysis in microbial ecology, as it can allow for accurate measurement of diversity and phylogenetic relationships, avoiding phylogenetic resolution loss and diversity measurement biases due to intragenomic heterogeneity(Vos *et al.*, 2012).

Keeping these points in mind, we developed a *rpoB* gene-based qPCR method which would aid in absolute quantification of bacteria, resulting in user friendly, accurate, sensitive, and specific quantification of test organisms.

5.2 Materials and Methods

5.2.1 Materials

All the primers and nuclease free water were purchased from Sigma-Aldrich. *E. coli* MG1655 bacterial culture was maintained in Sunandan Divatia School of Science, NMIMS's culture collection. Tryptone Soya Agar and Tryptone Soya Broth was purchased from HiMedia Laboratories Pvt Ltd. Invitrogen PureLink™ Genomic DNA kit and Applied Biosystems™ PowerUp™ SYBR™ Green Master Mix was purchased from ThermoFisher Scientific, USA. Molecular grade agarose powder was purchased from MP Biomedicals.

5.2.2 Selection of significant bacteria in oral cancer

Thorough literature review was performed to select the significant candidate organisms whose abundance has to be quantitated for method development. Microbiological as well as NGS-metagenomic studies were referred and eight bacteria reported to be linked with oral cavity and cancer research were shortlisted. Bacterial species that were chosen and evaluated for their either increased or decreased abundances, included *Porphyromonas gingivalis (P. gingivalis)*, *Fusobacterium nucleatum (F. nucleatum), Capnocytophaga gingivalis (C. gingivalis), Prevotella melaninogenica (P. melaninogenica), Haemophilus parainfluenzae (H. parainfluenzae), Rothia mucilaginosa (R. mucilaginosa), Veillonella parvula (V. parvula),* and *Streptococcus mutans (S. mutans)*.

5.2.3 rpoB Primer designing

5.2.3.1 Primer Selection

All primers were designed using Genscript® using rpoB gene sequences for selected organisms from the NCBI database (Table 5.1). The parameters for all primer designing were as follows:

PCR amplicon range	100-400 bp
Primer Tm minimum	52°C
Primer Tm maximum	60°C

5.2.3.2 Preparation of Primer stock

The primers were dissolved in sterile MQ water to make up a final concentration of 100 mM/µl as per manufacturer's instructions. The mixture was vortexed and spun thoroughly. To make a working stock solution of 10 mM/µl, 20 µl of 100 mM/µl primer was added to 180 µl of sterile MQ water, vortexed thoroughly and spun and stored at -20°C.

5.2.3.3 Primer Standardization

The working stock of the primers was used to perform a PCR annealing temperature standardization for primers. All the primers were tested for optimal amplification at temperatures ranging between 52°C to 64°C followed by performing agarose gel electrophoresis to determine the best temperature for amplification.

Table 5.1: Primer sequences of organisms under investigation

Organism	**Primer Sequences**
Escherichia coli MG1655	**F**: 5'- GTTGACTCCGGTGTAACTGC-3'
	R: 5'- ACGGGTGTATTTGGTCAGGT -3'
Porphyromonas gingivalis	**F**: 5'- TCGACCTGATGGACGTTTCGC-3'
	R: 5'- ACGGTTGGCATCGTCGTGTT-3'
Fusobacterium nucleatum	**F**: 5'GGTTCAGAAGTAGGACCGGGAGA3'
	R: 5'-ACTCCCTTAGAGCCATGAGGCAT-3'
Capnocytophaga gingivalis	**F**: 5'- TGGATGCCATCGGAGCCAAC-3'
	R: 5'- GGCATCGAGGGTACGGGAGA-3'
Haemophilus parainfluenzae	**F**: 5'- AGCGAGTACGGAACACGCAA-3'
	R: 5'- TGCAGTTCCAATTCCCGATCCA-3'
Prevotella melaninogenica	**F**: 5'- GTGCTCGTGTTGAGCCAGGT-3'
	R: 5'- TCACCAGCCTTGTCACCGAA-3
Rothia mucilaginosa	**F**: 5'- AGTCTGAGGCTCCCGTGGTT-3'

	R: 5'- AACCACACCGGGCTTCTTCG-3'
Veillonella parvula	**F**: 5'- GCGCGAACATGCAACGTCAA-3'
	R: 5'- CACGCGCCAATACGCAAACA-3'
Streptococcus mutans	**F**: 5'- CGTCATGGGACGTCATCAAGGG-3'
	R: 5'-AGCTCCCATAAGGGCACGGT-3'

5.2.4 Establishing standard curve using *E. coli* MG1655

E. coli MG1655 strain was used as a model organism to define the relationship between DNA concentration and CFU/ml. 16-18hrs old culture in Tryptone Soya Broth (TSB) was centrifuged at 10000 rpm for 5 mins at RT. The supernatant was discarded and the cell pellet was resuspended using standard saline. The O.Ds (0.1,0.2,0.3, 0.4) were adjusted at 600 nm using Biotek EPOCH2 microplate reader. The CFU/ml for the set O.Ds were obtained using spread plate technique on BHI agar plates. From the same O.D. adjusted stock, 1ml culture was processed for DNA Isolation using the Invitrogen PureLink™ Genomic DNA kit protocol discussed in section 4.2.4. The isolated DNA was quantified and assessed for purity using Biotek EPOCH2 microplate reader. The DNA quantification values were used to establish a relationship between O.D.-CFU/ml-DNA concentration. Based on the correlation between DNA yield and O.D., the DNA isolated from 0.1 O.D. at 600nm was selected as a reference for comparison with other bacteria of interest.

DNA concentration of 0.1OD culture was noted and 10-fold dilutions ranging from 3560pg/µL to 35.6fg/µL was used to construct the standard curves for rpoB gene amplification. The reaction mixture was prepared using PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) in a total volume of 10µL as per following set-up (Table 5.2).

Table 5.2 Reaction setup for *E.coli* standard curve

PowerUp™ SYBR® Green Master mix	5.0µL
Forward primer	0.5µL
Reverse primer	0.5µL
DNA template (3560pg-35.6fg)	1.0µL
DEPC treated water	3.0µL

For *E. coli* MG1655 standardization, 6-serial dilutions of DNA isolated from 0.1 O.D. *E. coli* MG1655 were prepared and used to obtain a standard graph. All real time PCR runs were performed on Applied Biosystems™ StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific).

The qPCR reaction was set up as follows (**Table 5.3)**

Stage	Temperature	Time	Cycles
Initial denaturation	95°C	2 minutes	
Denaturation	95°C	15 seconds	40 cycles
Annealing	58°C	30 seconds	
Elongation	72°C	60 seconds	
Melt curve	60°C-95°C	0.5°C s-1	

All real-time PCR runs were performed in triplicates and standard curves of Ct values were plotted against the log of DNA concentration to find a linear relationship between them. The linear equation of the graph was further used for all calculations. Tm of the amplified product was noted to ensure amplification of the same product in every run. The amplified products were resolved on agarose gel electrophoresis gel to confirm the expected size of the amplified product.

5.2.5 Real time PCR for significant bacteria

The *rpoB* gene primers for selected eight bacterial species (*Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula* and *Streptococcus mutans)* were chosen for the study, i.e and standardized for optimum PCR annealing temperature. For the PCR, reactions were set up (as mentioned in Table 5.2) for a total volume of 10µL, where 30-40 ng of template DNA was used.

The reaction was set up as follows (**Table 5.4)**

Stage	Temperature	Time	Cycles
Initial denaturation	95°C	2 minutes	
Denaturation	95°C	15 seconds	40 cycles
Annealing	58°C/59°C	30 seconds	
Elongation	72°C	60 seconds	
Melt curve	60°C-95°C	0.5°C s-1	

The Tm obtained by amplification of rpoB of all bacterial species was noted and the amplified products were resolved on 2% agarose gel for the size of amplified products as mentioned above. All real-time PCR runs were performed in triplicates and Ct values were recorded. The calculations were performed on Excel using the standard curve equation obtained using *E. coli* to obtain the absolute count in terms of bacterial genomic equivalent per ml (BGE/ml) (Sundin *et al.*, 2017).

5.2.6 Statistical analysis

Statistical analysis was performed for the absolute counts (BGE/ml) obtained using GraphPad Prism (version. 8.0.2). Outliers from the data were removed to prevent any bias using the ROUT method (Q=0.5%) to maintain stringency. The cleaned data was subjected to One-Way Anova followed by Tukey's multiple comparison test to test for statistical significance where $p<0.05$ was considered significant.

5.3 Results

5.3.1 Primer standardization

The parameters Such as annealing temperature, melting temperature were standardized for all the primers, and the values obtained along with the amplicon size are tabulated in Table 5.2 below.

Table 5.5: Standardized primer parameters

Organism	Primer Sequences	Annealing temp (°C)	Tm obtained (°C)	Size of product (bp)
Escherichia coli MG1655	**F**: 5'- GTTGACTCCGGTGTAACTGC-3’	58	81.38	147
	R: 5'- ACGGGTGTATTTGGTCAGGT -3’			
Porphyromonas gingivalis	**F**: 5'- TCGACCTGATGGACGTTTCGC-3'	58	80.02	86
	R: 5'- ACGGTTGGCATCGTCGTGTT-3’			
Fusobacterium nucleatum	**F**: 5'GGTTCAGAAGTAGGACCGGGAGA3'	59	77.02	161
	R: 5'-ACTCCCTTAGAGCCATGAGGCAT-3’			
Capnocytophaga gingivalis	**F**: 5'- TGGATGCCATCGGAGCCAAC-3'	58	83.48	104
	R: 5'- GGCATCGAGGGTACGGGAGA-3'			
Haemophilus parainfluenzae	**F**: 5'- AGCGAGTACGGAACACGCAA-3'	58	79.53	113
	R: 5'- TGCAGTTCCAATTCCCGATCCA-3'			
Prevotella melaninogenica	**F**: 5'- GTGCTCGTGTTGAGCCAGGT-3'	58	84.4	115
	R: 5'- TCACCAGCCTTGTCACCGAA-3			
Rothia muciloginosa	**F**: 5'- AGTCTGAGGCTCCCGTGGTT-3'	58	84.8	92
	R: 5'- AACCACACCGGGCTTCTTCG-3'			
Veillonella parvula	**F**: 5'- GCGCGAACATGCAACGTCAA-3'	58	82.75	114
	R: 5'- CACGCGCCAATACGCAAACA-3'			
Streptococcus mutans	**F**: 5'- CGTCATGGGACGTCATCAAGGG-3'	59	81.46	151
	R: 5'-AGCTCCCATAAGGGCACGGT-3'			

5.3.2 *E. coli* MG1655 standard curve

Maintenance and enumeration of *E. coli* MG 1655 was performed using Standard Tryptone Soya Broth and sterile Tryptone Soya Agar. Overnight grown (16-18 hrs old) culture was used to obtain a relationship between optical density (O.D.) and colony forming unit (CFU/ml). Table (5.3) shows CFU/ml obtained at various O.D.600. The same O.D. culture was used for isolating DNA followed by quantitation at 260/280 nm, also depicted in Table (5.3).

Table 5.6: Relation between Optical density, CFU/ml and DNA concentration

Optical Density (600 nm)	CFU/ml	DNA concentration (ng/µl)
0.1	3.856 (±1.5) x 10^7	3.560
0.2	2.671 (±0.5) x 10^8	10.092
0.3	2.075 (±0.8) x 10^9	24.492
0.4	2.548 (±2.0) x 10^{10}	31.065

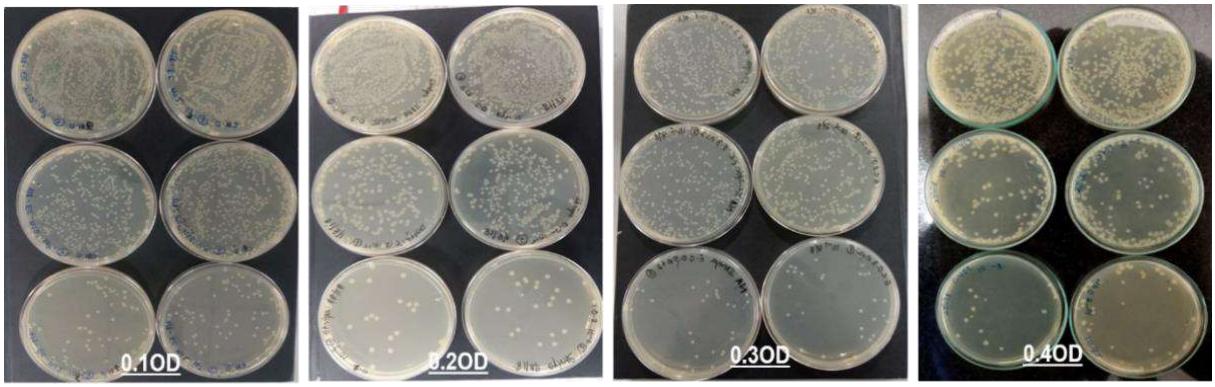


Fig. 5.1 : Colony forming assay (CFU) results performed at various O.Ds, namely 0.1, 0.2, 0.3 and 0.4 @600nm (left to right).

DNA concentration of 0.1 O.D. culture was selected and 10-fold dilutions ranging from 3560pg/µL to 35.6fg/µL was used to construct the standard curves for *rpoB* gene amplification. The melt-curve analysis showed that Tm was constant, i.e. 81.75 in all runs confirming the amplification of a single product. The standard graph for Ct plotted against concentration range (3560pg to 35.6fg) was obtained and the curve was linear (R^2> 0.99). The slope of the standard curve for rpoB was -3.3. This also confirms the PCR efficiency is close to 100 percent. The final equation of the standard graph obtained (Y= -3.304X + 38.556) was

used for all further calculations where, Y is the Ct value for a sample and X is the log of DNA concentration.

According to the calculations, 3.56ng DNA corresponds to 3.856 *10^7 cells, or 3.56fg DNA = 3.856 cells. Using this correlation data, absolute count of the listed bacterial species was obtained for the oral cavity sample, after standardizing the PCR conditions for each of the bacterial species and analysing their melt peaks and amplicon sizes.

Table 5.7: Ct values obtained for *E. coli* MG1655 Standard curve

Concentration of DNA (fg)	Log DNA Concentration	Ct value
3560000 (A)	6.55145	16.76 ± 0.90
356000(B)	5.55145	20.02 ± 1.02
35600(C)	4.55145	23.87 ± 0.08
3560	3.55145	27.03 ± 0.22
356	2.55145	30.18 ± 0.08
35.6	1.55145	33.16 ± 0.26

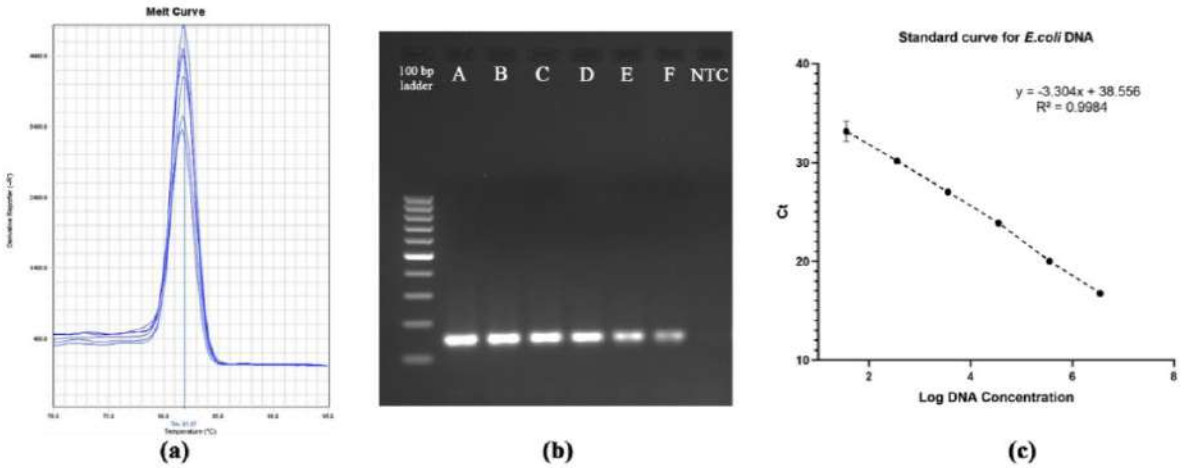


Fig. 5.2: Standard curve for *E.coli* (a) Melt peaks observed for standard curve of E.coli depicting steady increase in amplification. (b) Agarose gel electrophoresis of standard curve products obtained after amplification of *E.coli* using Real time PCR. (c) Standard curve linear graph Ct vs Log DNA concentration depicting R2 > 0.998.

5.3.3 Absolute enumeration for significant bacteria between three study groups

The amplification peaks for every organism tested are shown below in Fig. 5.3. The single peaks depict single product amplification necessary for specific amplification of the target.

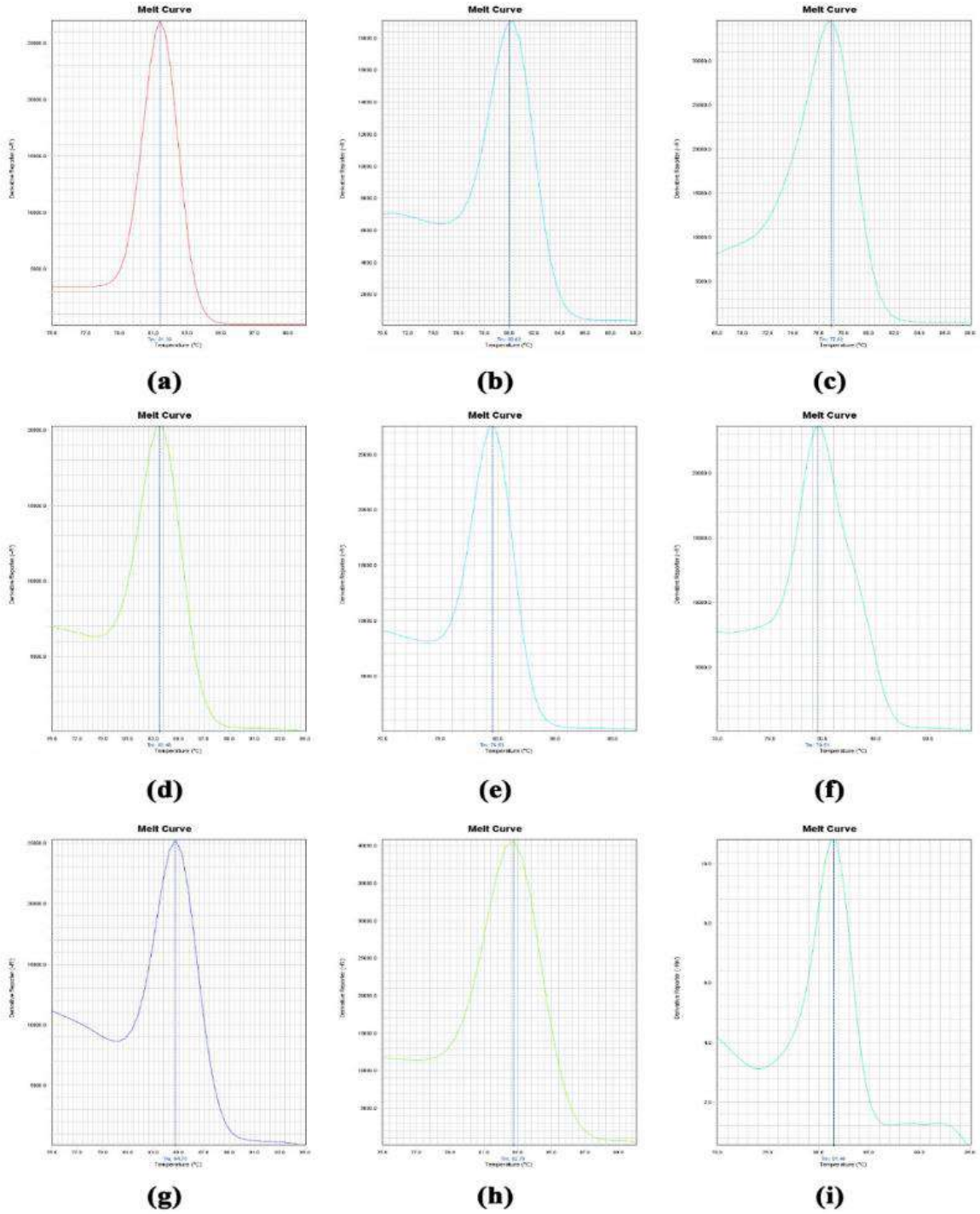


Fig. 5.3: Melt peaks for rpoB gene amplification of nine bacteria (a) *Escherichia coli*, (b) *Porphyromonas gingivalis*, (c) *Fusobacterium nucleatum*, (d) *Capnocytophaga gingivalis*, (e) *Haemophilus parainfluenzae*, (f) *Prevotella melaninogenica*, (g) *Rothia mucilaginosa*, (h) *Veillonella parvula*, (i) *Streptococcus mutans*. Single peaks and constant melting temperature (Tm) are obtained in all the runs demonstrating amplification of a single product.

The results obtained after performing qPCR using 40 samples of each study group (Control, Tobacco and Oral cancer) after calculations have been depicted in Figure 5.4 below.

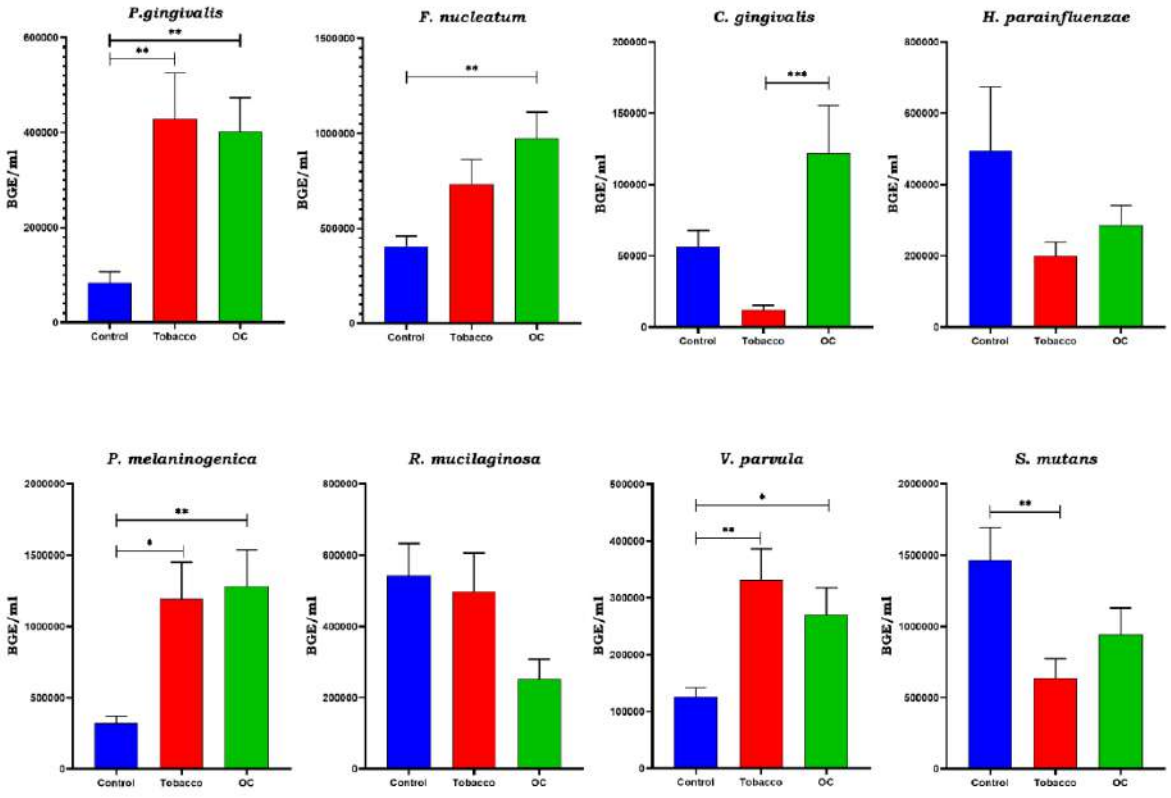


Fig. 5.4: Bar graphs displaying difference in abundance of eight tested bacteria in 3 different study groups. (a) *Porphyromonas gingivalis*, (b) *Fusobacterium nucleatum*, (c) *Capnocytophaga gingivalis*, (d) *Haemophilus parainfluenzae*, (e) *Prevotella melaninogenica*, (f) *Rothia mucilaginosa*, (g) *Veillonella parvula*, (h) *Streptococcus mutans*.

Table 5.8: Final bacterial counts of bacteria in three study groups

Organism name	Control	TC	OSCC patients
P. gingivalis	8.45 X 10^4	4.28 X 10^5	4.01 X 10^5
F. nucleatum	4.06 X 10^5	7.32 X 10^5	9.73 X 10^5
C. gingivalis	5.65 X 10^4	1.22 X 10^4	1.21 X 10^5
H. parainfluenzae	4.94 X 10^5	1.99 X 10^5	2.85 X 10^5
P. melaninogenica	3.21 X 10^5	1.19 X 10^6	1.28 X 10^6
R. mucilaginosa	5.42 X 10^5	4.98 X 10^5	2.53 X 10^5
V. parvula	1.25 X 10^5	3.31 X 10^5	2.70 X 10^5
S. mutans	1.46 X 10^6	6.37 X 10^5	9.42 X 10^5

The abundance of bacteria in three study groups were plotted and shown in Fig. 5.5. The total abundance of bacteria enumerated in the OC group was higher as compared to tobacco chewers and healthy individuals (Fig. 5.5). This is pattern was observed in studies comparing oral cancer microbiome and control population (Marttila et al., 2013). Similar to the OC group, the abundance of bacteria in the tobacco chewing group was comparatively higher as compared to control population, but lower than that of OC group.

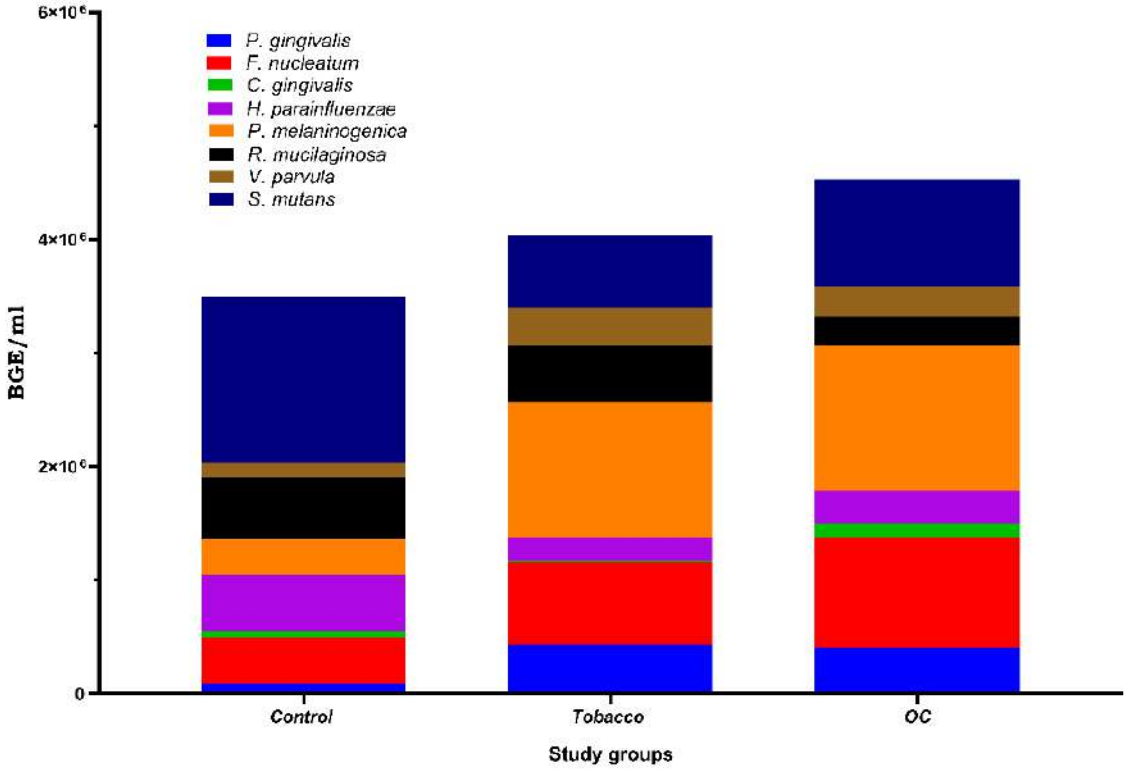


Fig. 5.5: Stacked bar plots depicting total bacterial abundance in the study groups.

The graph also clearly depicts the proportions of the bacteria in these three study groups as shown in Fig 5.5 and Table 5.9. *S. mutans* constitute the highest percentage in control population, *i.e.,* approximately 41.8 % of total control population, whereas its proportion is approximately 15.8% and 20.8% of total tobacco and OC population. *S. mutans* is a known commensal organism present in the oral cavity of any healthy individual in a high proportion. Although it is a known commensal organism, due to dysbiosis, it can sometimes act as an opportunistic pathogen and contribute towards various oral diseases such as dental caries. The abundance of *S. mutans* decreasing in Tobacco and OC group depict dysbiosis where the abundance of a commensal bacteria has decreased. This could be attributed to various

reasons such as unfavourable chemicals and metabolites in tobacco, increased proportion of pathogenic bacteria, oral complications and diseases due to lifestyle habits, stress.

On the other hand, the organism that constitutes the highest proportion in tobacco chewers and oral cancer patients is *Prevotella melaninogenica* i.e 29.5% and 28.3% respectively, which constitutes only approx. 9.2% in the control population.

Table 5.9: Percent population of investigated bacteria in study groups

Bacterial species	Control (%)	Tobacco (%)	Oral cancer (%)
S. mutans	41.85	15.82	20.82
R. mucilaginosa	15.53	12.37	5.59
H. parainfluenzae	14.16	4.94	6.30
F. nucleatum	11.64	18.18	21.50
P. melaninogenica	9.20	29.55	28.29
V. parvula	3.58	8.22	5.97
P. gingivalis	2.42	10.63	8.86
C. gingivalis	1.62	0.30	2.67

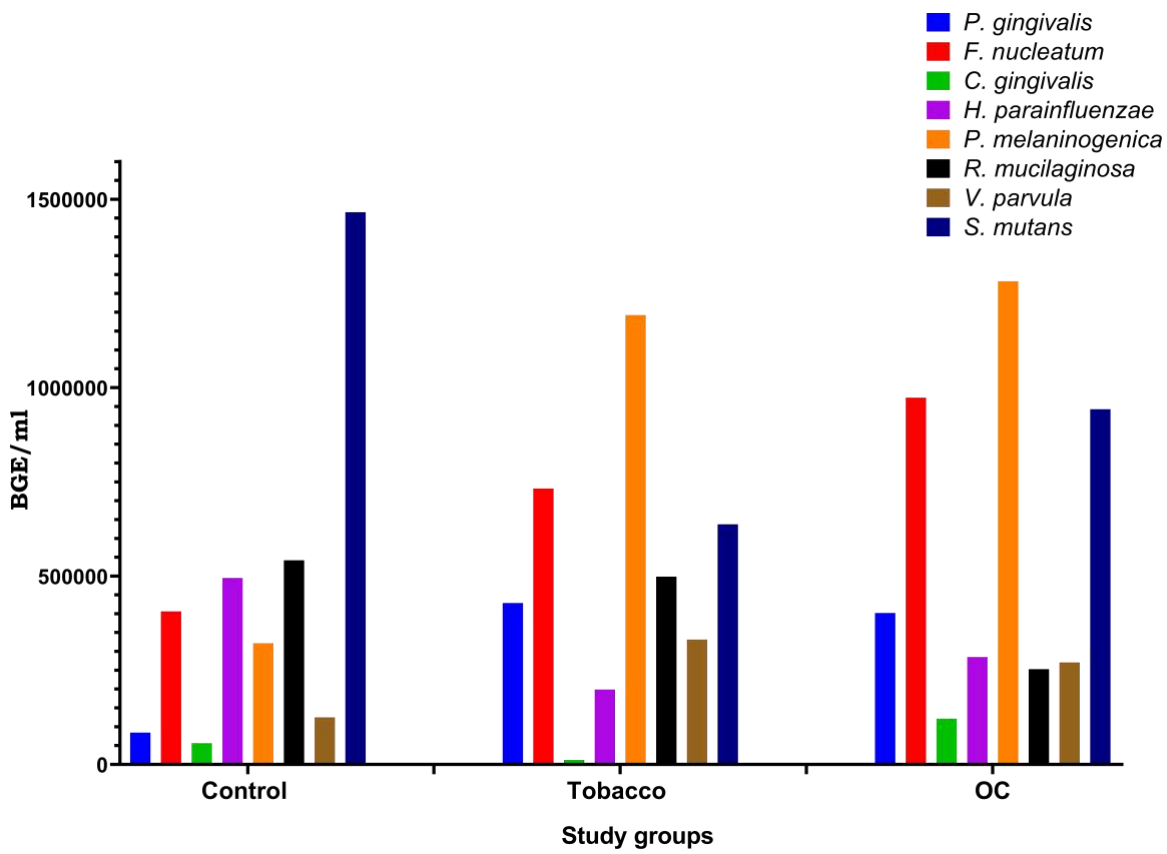


Fig. 5.6: Comparison of significant bacterial abundance in three study groups.

Significant differences in the absolute count were obtained between control and long-term tobacco chewing population for *P. gingivalis* (~ 8.45 X 10^4 vs~ 4.28 X 10^5 BGE/ml) as well as between control and oral cancer patients (~ 8.45 X 10^4 vs~ 4.01 X 10^5 BGE/ml). For *F. nucleatum*, a significant difference was observed in absolute count between control and oral cancer patients (~ 4.06 X 10^5 vs~ 9.73 X 10^5 BGE/ml). For both these bacterial species, there was an increase in the abundance of respective bacterial populations on comparing the control group with tobacco chewers as well as oral cancer patients.

C. gingivalis, a significant difference was observed on comparing bacterial count in oral cancer patients (~ 1.21 X 10^5 BGE/ml) with long term tobacco chewing population (~ 1.2 X 10^4 BGE/ml). The absolute abundance of *P. melaninogenica* was significantly different between control population and tobacco chewers as well as oral cancer patients. Increased abundance was observed in tobacco chewers and oral cancer patients (1 .19 X 10^6 and~ 1.28 X 10^6 BGE/ml respectively) BGE/ml, as compared to the control population (3 .2 X 10^5 BGE/ml). Similarly, in the case of *V. parvula*, significant increase in the abundance of bacteria was observed from

control population (~1.25 X 10^5 BGE/ml) to tobacco chewers and patients with oral cancer (~ 3.31 X 10^5 and~ 2.70 X 10^5 BGE/ml). For *S. mutans*, the pattern observed was opposite, there was a significant decrease in absolute counts of *S. mutans* in tobacco chewing population (~ 6.37 X 10^5 BGE/ml) as compared to control population (~ 1.46 X 10^6 BGE/ml). A non-significant decrease was observed in absolute bacterial counts in patients with oral cancer as compared to control population.

For *H. parainfluenzae* and *R. mucilaginosa*, the absolute abundance of these bacteria in oral samples depicts a pattern, but no significant difference was observed. The pattern of abundance of bacteria mentioned in this study correlates well with earlier reports, thereby supporting the results obtained. The abundance of *H. parainfluenzae* was observed to be higher in the control population as compared to OSCC patients, suggested by NGS data. On the contrary, the population of *P. melaninogenica* was reported higher in OSCC patients as compared to healthy individuals. Increased abundance *P. gingivalis* was observed in Tobacco chewers and oral cancer patients as compared to healthy controls.

5.3.4 Considering variables in bacterial abundance

5.3.4.1 Effect of Grade of Cancer on Microbiota

The effect of grade of cancer on the selected bacteria are shown in the Fig. 5.7 below. All the organisms showed an increased abundance in individuals suffering from grade III as compared to grade II oral cancer. Although the increase is not significant in all bacteria, only *R. mucilaginosa* and *V. parvula* showed a significant increase from grade II to grade III.

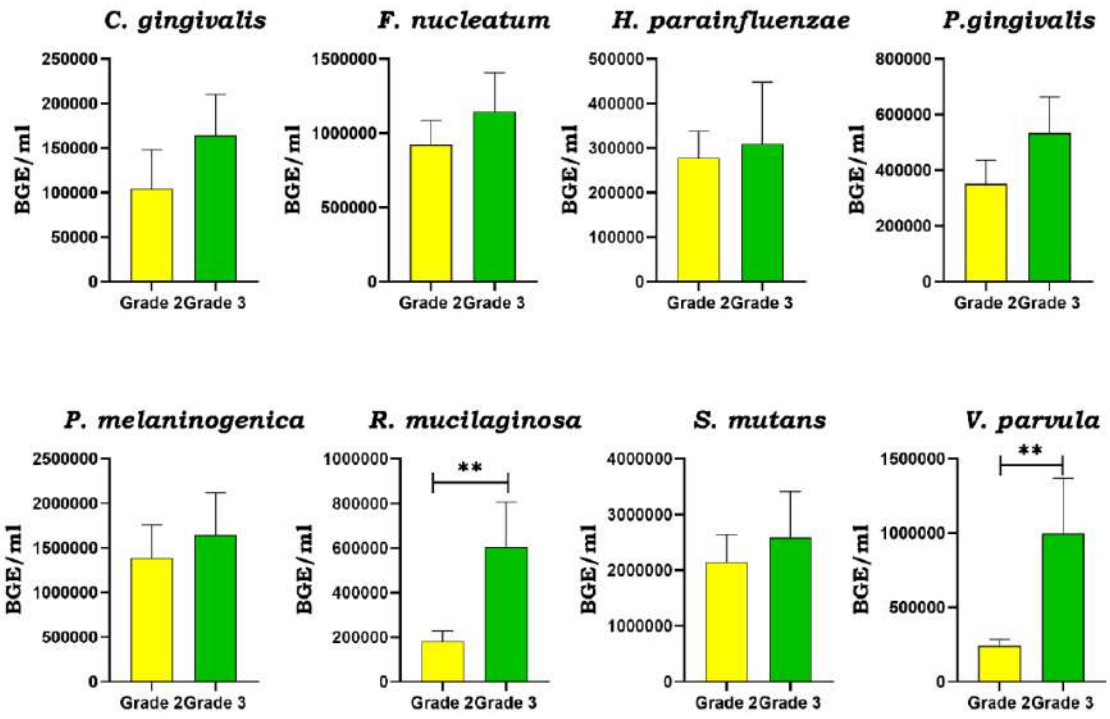
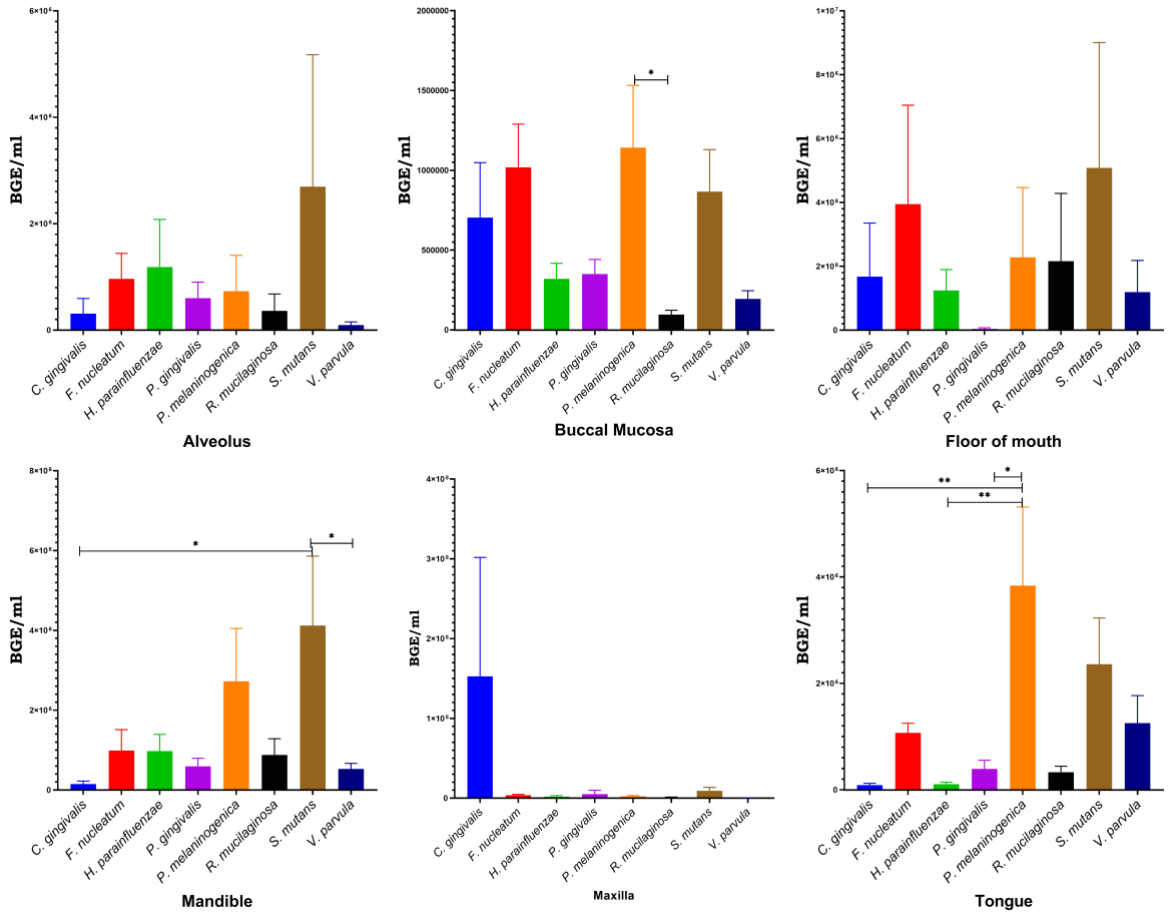


Fig. 5.7: Bar graphs depicting bacterial quantitation based on grade of cancer. Data plotted is mean ± SD after performing student's t-test. P<0.05 was considered statistically significant.

5.3.4.2 Effect of site of Cancer on Microbiota

The relative abundance of each bacterium in individuals suffering from oral cancer is depicted in Fig.5.8. Each bar represents a site affected by cancer in the oral cavity. In individuals suffering from cancer of the alveolus, the abundance of *S. mutans* was higher than any other test organism, though the difference was non-significant. In individuals suffering from buccal mucosa, abundance of *P. melaninogenica* was observed to be highest, whereas *R. mucilaginosa* was lowest. Apart from *P. melaninogenica*, *C. gingivalis, F. nucleatum* and *S. mutans* were also depicted in high abundance. In individuals suffering from cancer of the mandible, abundance of *S. mutans* was significantly highest, whereas that of *C. gingivalis* and *V. parvula* was significantly lower. The abundance of *C. gingivalis* was highest in individuals suffering from cancer of the maxilla as compared to other bacteria tested, whereas in those suffering from cancer of the tongue, *P. melaninogenica* was significantly higher as compared to other tested bacteria, and that of *C. gingivalis, H. parainfluenzae* and *P. gingivalis* was significantly lower.



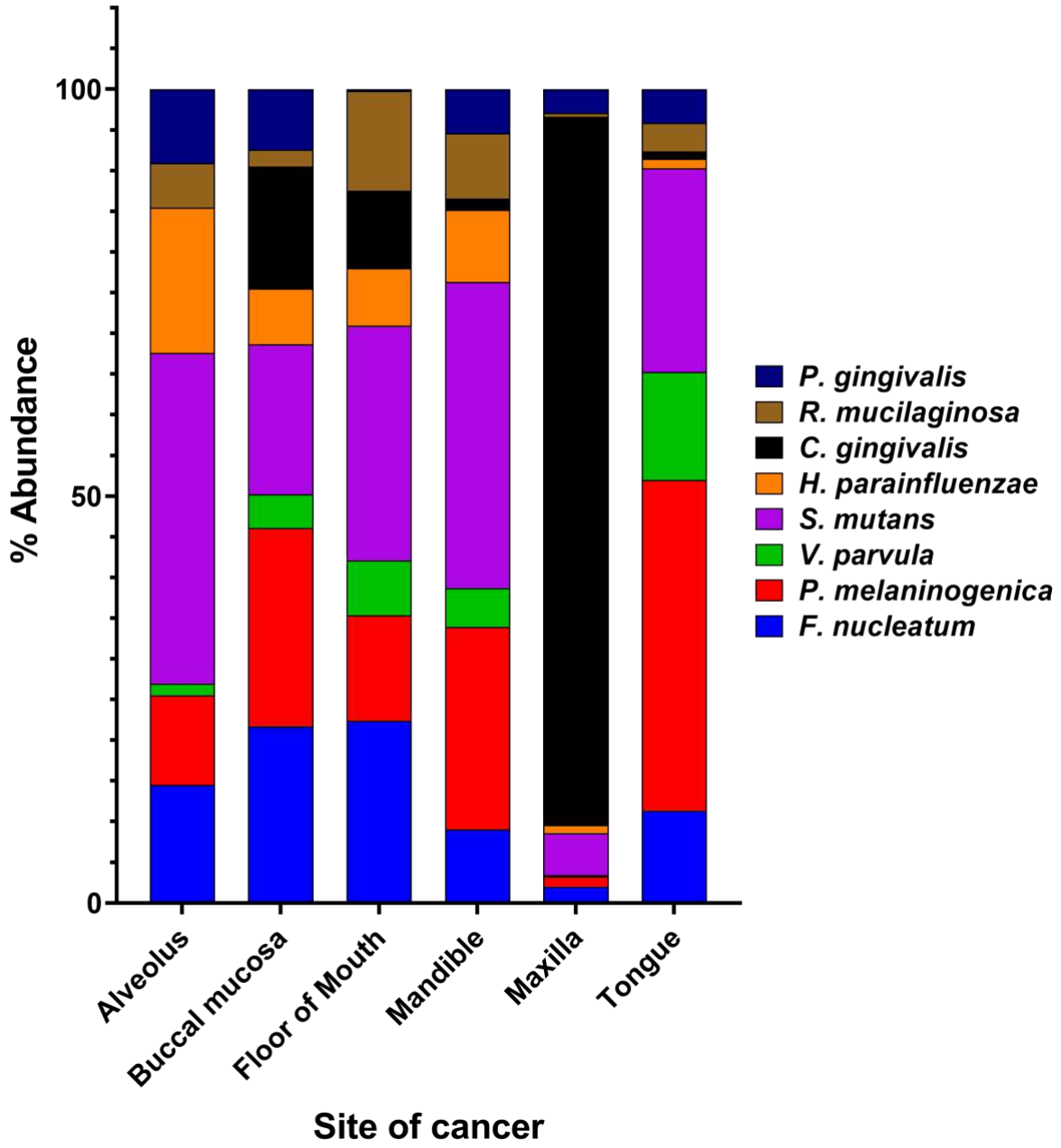
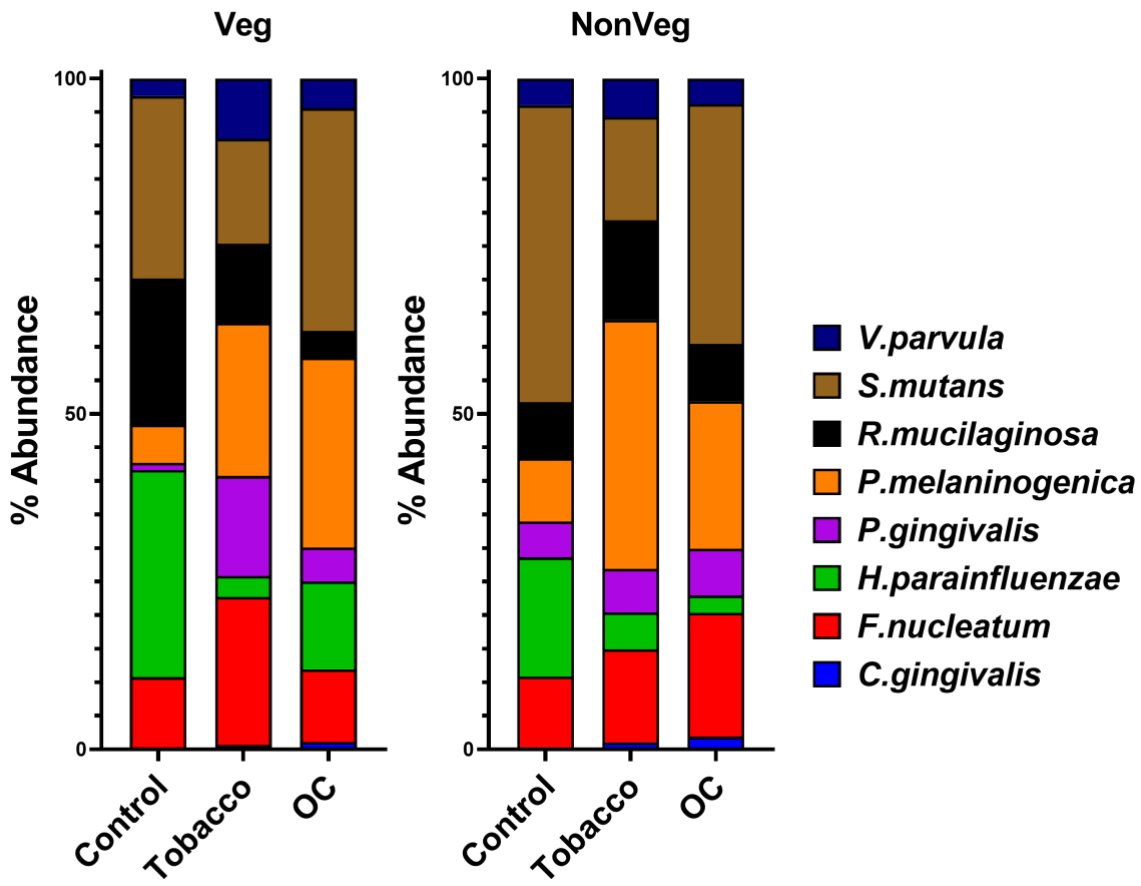
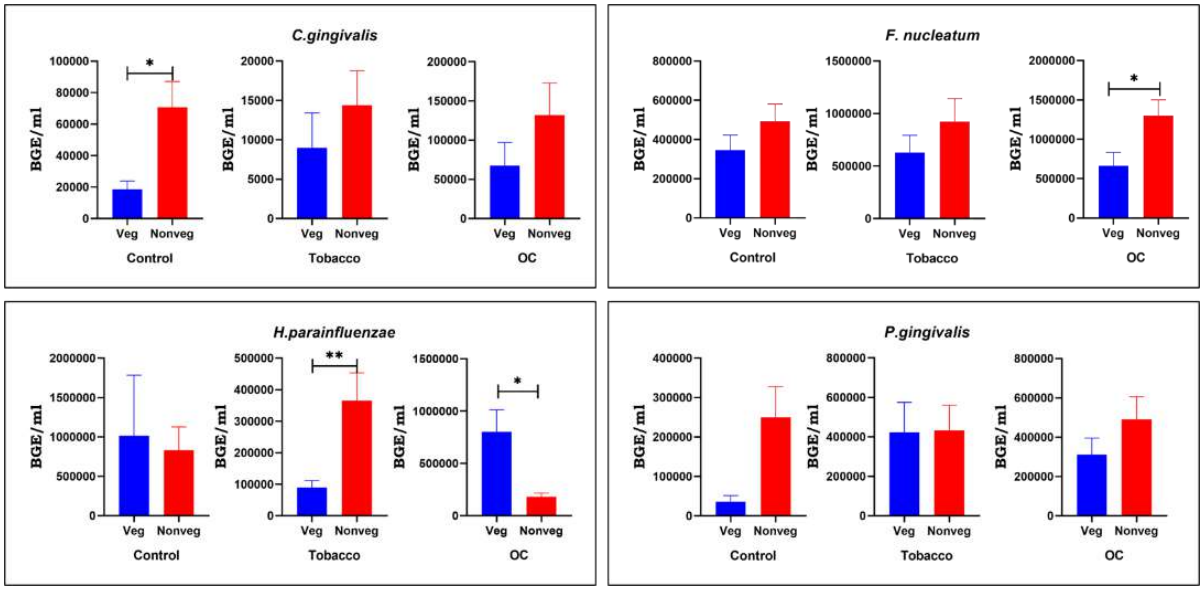


Fig. 5.8: Bar graphs and stacked bar graph depicting bacterial quantitation based on the site of cancer. Data plotted is mean ± SD after performing one-way Anova followed by Tukey's multiple comparison test. $P<0.05$ was considered statistically significant.

5.3.4.3 Effect of diet on Microbiota





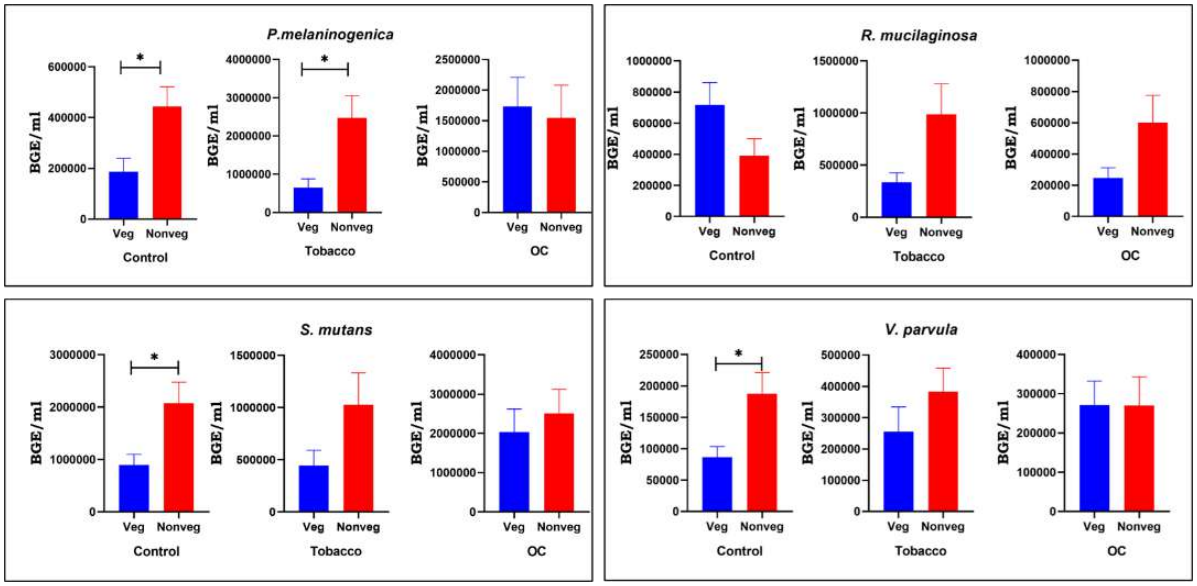
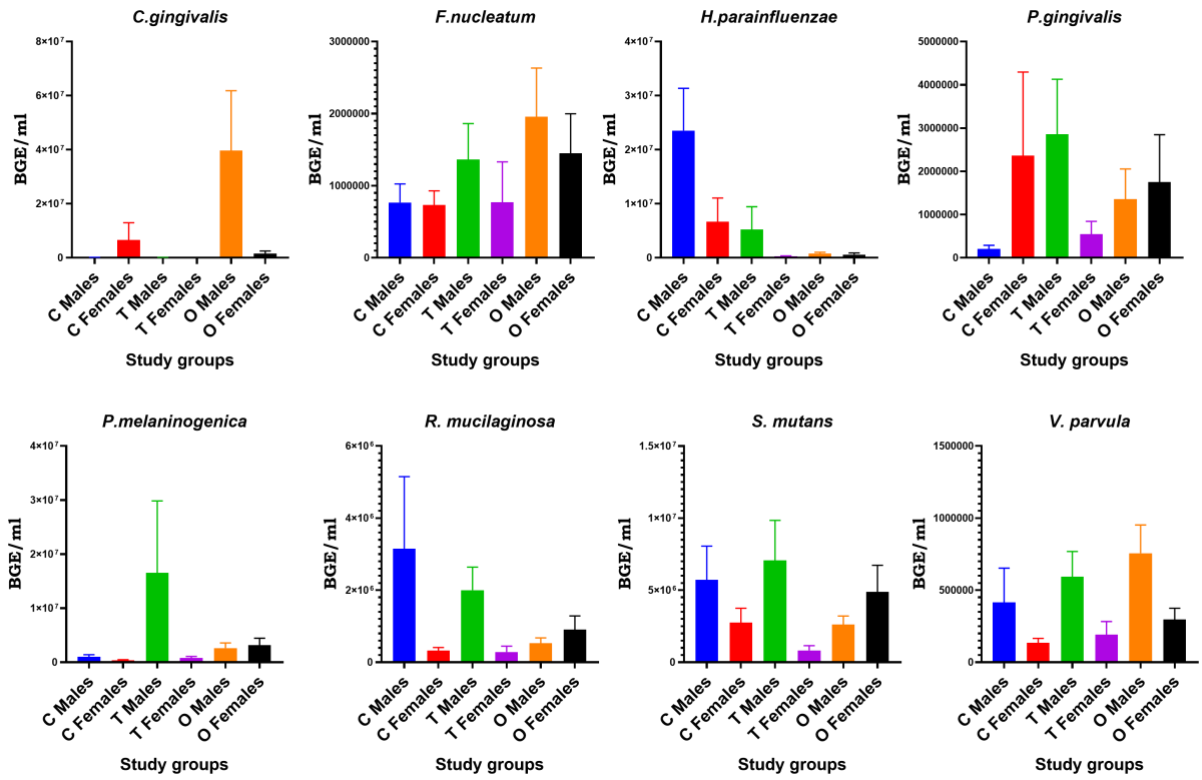


Fig. 5.9: Bar graph and stacked bar graph depicting bacterial quantitation based on diet of individuals. Data plotted is mean ± SD after performing student's t-test. $p<0.05$ was considered statistically significant.

The study population in three study groups were segregated based on the diet of individuals, namely vegetarians and non-vegetarians. A common trend was observed in almost all bacterial quantification across all study groups on assessing the abundance of bacteria based on the diet of individuals. In most cases, the bacterial abundance was found higher in individuals consuming non-vegetarian diets as compared to those consuming vegetarian diets. In *C. gingivalis*, higher bacterial count was observed in non-vegetarians compared to vegetarians, though this difference was significant only in the control population. Similarly, in *F. nucleatum* higher bacterial count was observed in non-vegetarians compared to vegetarians, though this difference was significant only in the OC population. In *H. parainfluenzae*, a significantly higher count was observed in non-vegetarians in the tobacco group and vegetarians in the OC group. In case of *P. melaninogenica, S. mutans* and *V. parvula*, significantly higher abundance was observed in non-vegetarians belonging to the control group, whereas in other study groups, this difference was non-significant.

5.3.4.4 Effect of gender on Microbiota

After comparing the gender of study participants belonging to the three study groups, no significant difference was found between male and female population belonging to control, tobacco and OC population. Therefore, the gender of individuals does not appear to play a role in determining the abundance of bacteria in the oral cavity.



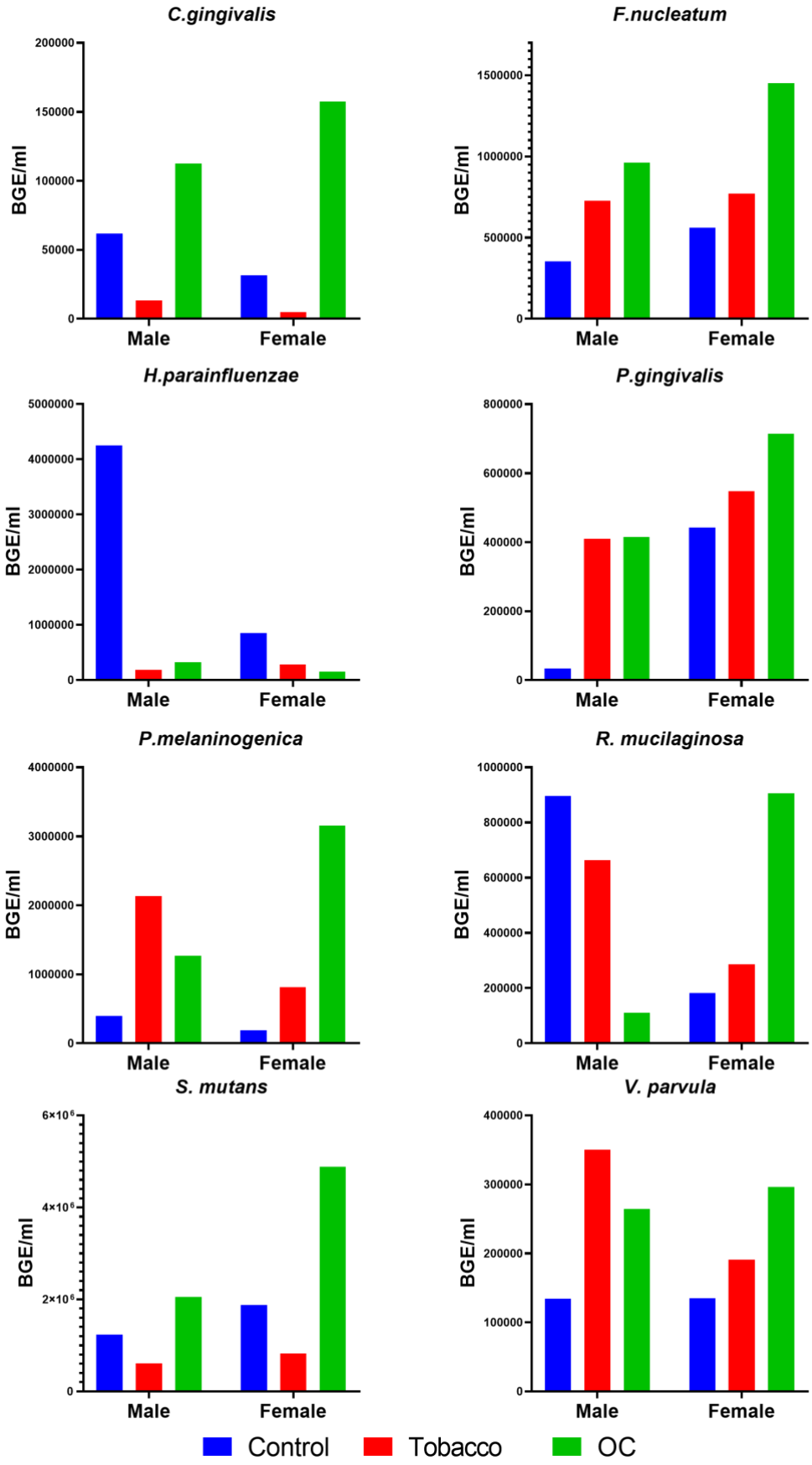


Fig. 5.10: Bacterial quantitation based on gender of individuals

5.3.4.5 Effect of age on Microbiota

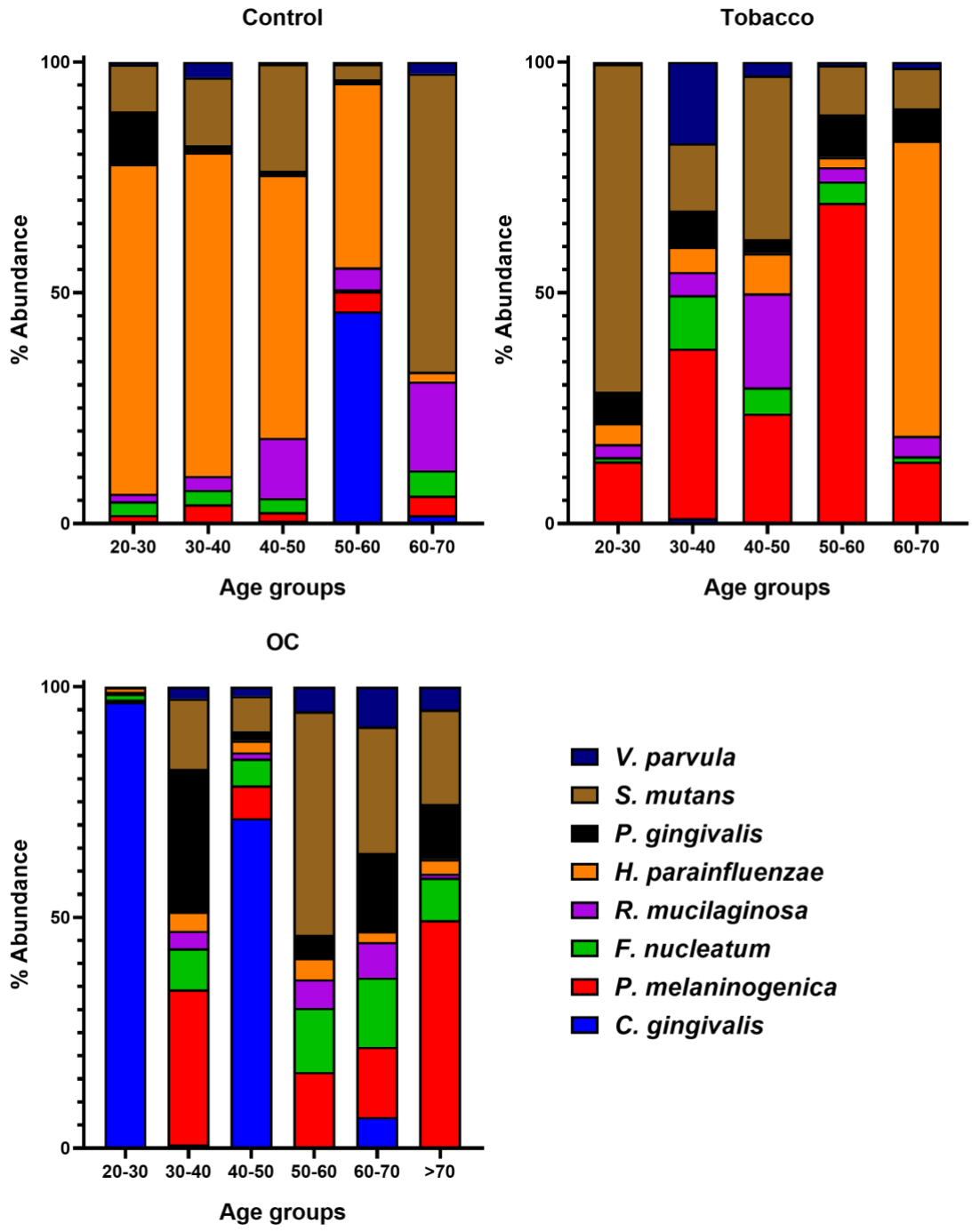


Fig. 5.11: Bacterial quantitation based on age of individuals in control, tobacco and OC group

The effect of age on bacterial abundance was also analysed and shown in Fig. 5.11 - 13. In case of *C. gingivalis*, significantly higher abundance was observed in ages 50-60yrs and 20-30yrs in control and OC respectively, as compared to 20-30 yrs and 60-70 yrs in tobacco group. For *F. nucleatum*, individuals aged 50-60 yrs in the tobacco group and aged 20-30 in the OC

group showed significantly higher abundance as compared to other age groups in the study population. *H. parainfluenzae* displayed significantly higher abundance in the control population aged 40-50 yrs as compared to 20-30 yrs in the OC group. In the case of tobacco group decreasing abundance of *H. parainfluenzae* with increasing age was observed but this difference was non-significant. In case of *P. gingivalis*, the bacterial abundance was observed to increase in all study groups with increasing age. In the case of *R. mucilaginosa*, the bacterial abundance increased with age in control and tobacco chewing individuals, but the trend was found to be reversed in the OC population. In tobacco chewers, abundance of *S. mutans* was significantly higher in individuals belonging to the group 20-30 yrs and that of *V. parvula* was significantly higher in the age group 60-70 yrs.

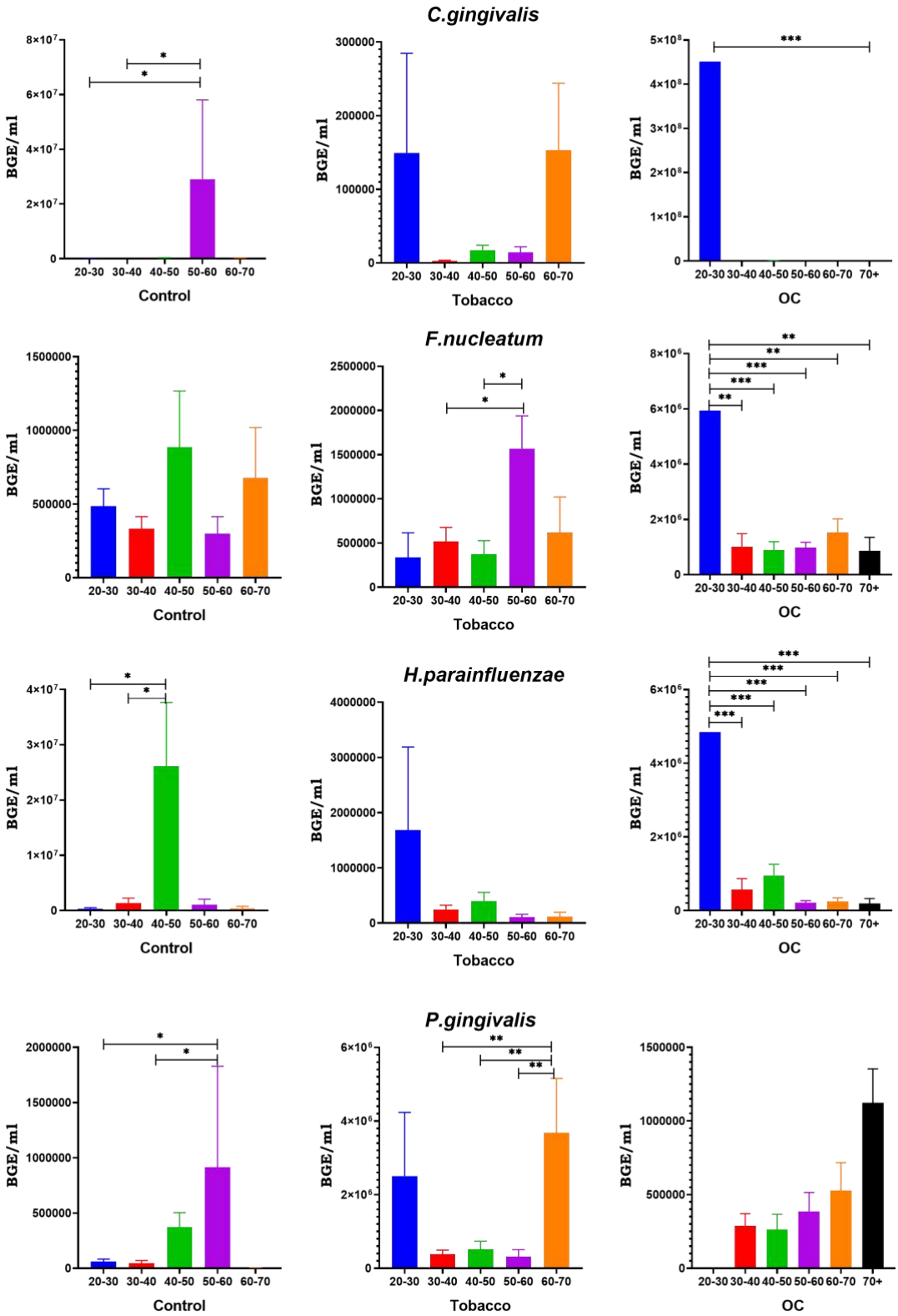


Fig. 5.12: Bar graph depicting bacterial abundance based on age of individuals in control, tobacco and OC group in *C. gingivalis, F. nucleatum, H. parainfluenzae* and *P. gingivalis.* Data represents mean ± SD after performing one-way Anova followed by Tukey's multiple comparison test. P<0.05 is considered statistically significant.

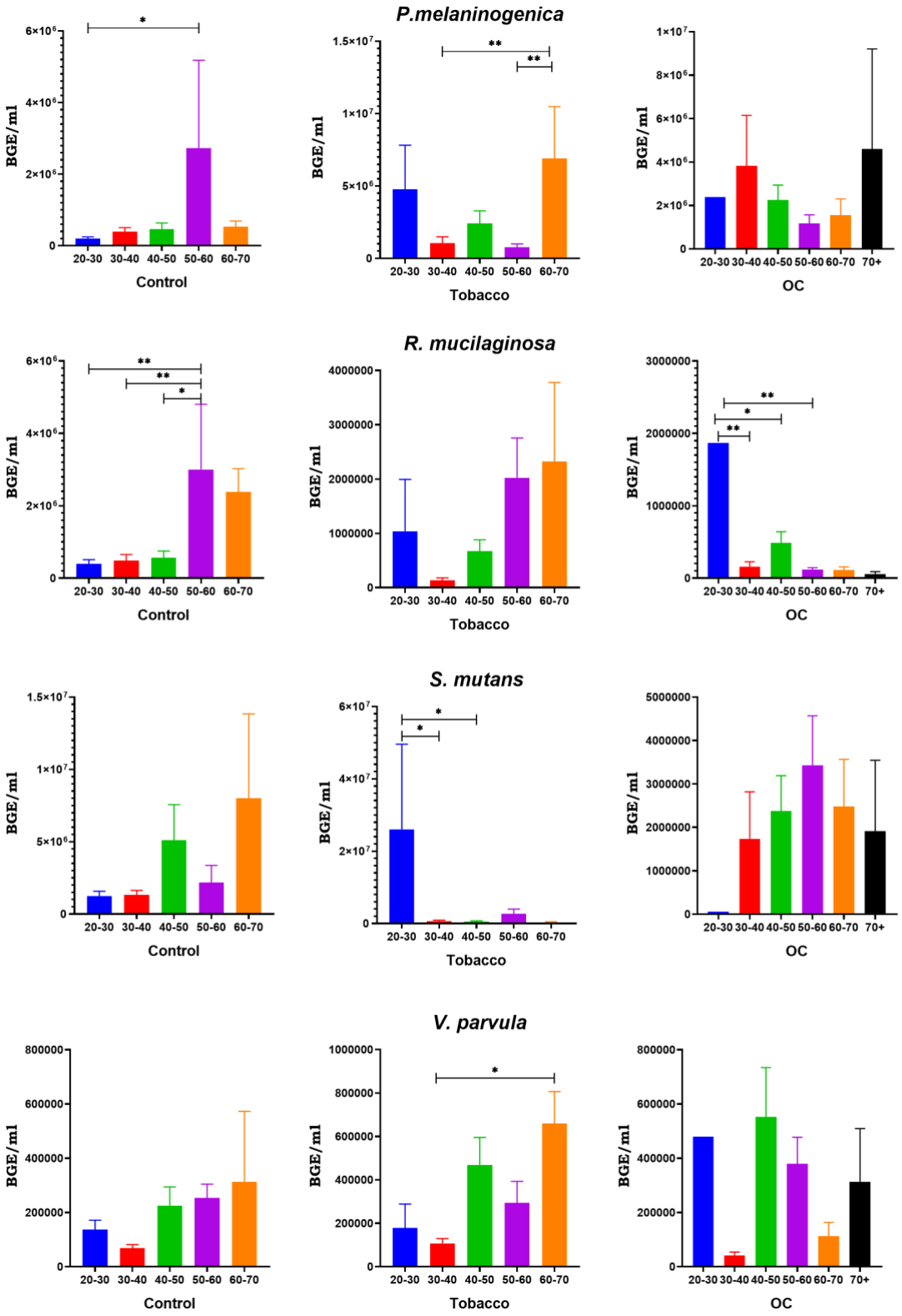


Fig. 5.13: Bar graph depicting bacterial abundance based on age of individuals in control, tobacco and OC group in *P. melaninogenica, R. mucilaginosa, S. mutans* and *V. parvula.* Data represents mean ± SD after performing one-way Anova followed by Tukey's multiple comparison test. $p < 0.05$ is considered statistically significant.

5.4 Discussion

Although NGS sequencing is the method of choice for microbiome study, a simple and inexpensive technique would be very useful for undertaking small scale studies for absolute quantification of a particular bacterial population in a given environment or ecosystem. Most microbiome research study employ the multicopy 16S rRNA gene variable regions as the reference gene, which might result in non-accurate bacterial quantification. Using species-specific single copy *rpoB* genes, our goal was to create a qPCR method for the absolute quantification of oral bacteria. We found a correlation between DNA and BGE (Bacterial Genome Equivalent)/ml using *E. coli* as a model organism. To create a qPCR-based standard curve, *E. coli*-specific rpoB primers were utilised. The Ct values from this standard curve were then used to determine the absolute abundance of the bacterium.

This approach was used to examine eight distinct oral cancer-causing bacteria, including *Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula,* and *Streptococcus mutans*. The abundance of *P. gingivalis, F. nucleatum, C. gingivalis, P. melaninogenica* and *V. parvula* depicted increase in OC samples as compared to control, whereas for *R. mucilaginosa, H. parainfluenzae* and *S. mutans* it was vice-versa. *P. gingivalis* and *F. nucleatum* are known periodontal pathogens as well as *C. gingivalis, P. melaninogenica* and *V. parvula* have well established their role in promotion of oral cancer (Karpiński, 2019). Therefore, their increased numbers in the oral cavity of patients with oral cancer and tobacco chewers can be attributed to unhealthy oral conditions as well as dysbiosis leading to colonization of pathogenic bacteria. Similarly, decrease in the abundance of *R. mucilaginosa, H. parainfluenzae* and *S. mutans in* oral cancer suggests a decrease in the normal oral microflora of healthy individuals, suggesting dysbiosis in the oral cavity.

We further analysed bacterial abundance considering various parameters contributing to the metadata of participants. On analysing the bacterial quantitation depending on grade of cancer, it was found that individuals with grade III OC had higher abundance of all bacteria tested as compared to individuals with grade II OC and similarly in most cases, an increased number of bacteria was observed in individuals who consumed non-vegetarian diet as compared to vegetarian diet. The increased number of oral microflora observed in grade III individuals could be resulting from increased unhealthy oral cavity as compared to grade II

OC. With increasing grade and spread of cancer, the nutrients inflow to the tumor to support its growth, leading to more nutrient availability for the bacteria on the surface to strive through. Similarly, higher abundance of oral pathogens have been found in meat consumers leading to a high risk for oral diseases (Lu *et al.*, 2019).

It was concluded on analysing the age and gender of individuals in our study that these parameters do not showed any effect on the bacterial composition in the oral cavity of our study population. Although numerous studies have been performed to understand and identify the composition of oral microbiota in various conditions, it really is a complex ecosystem. Therefore, understanding the effect of variables such as age, diet in oral microbiome would require a more controlled dataset with higher sample numbers.

5.5 Conclusion

In conclusion, this chapter deals with development of quantitative real time PCR based methods for absolute quantification of oral bacteria from samples of human source. The development of this method helps to resolve the multi-copy gene issue present in the 16S rRNA gene based microbial quantification. Our method was further supported by the data collected from oral cavity samples for the bacteria that were important to the oral cavity populations and the findings were correlating with observations given in the previous reports. We believe that the method is not limited to bacterial enumeration of the oral cavity, but can be used for various natural and man-made microbial samples, like agricultural soil, waste water sample, sewage sample, faecal samples, or any microbial ecosystems where one aims to identify the bacterial abundance, composition or diversity.

CHAPTER 6

Assessment of Host mRNA expression during Host-Microbe Interaction

Assessment of Host mRNA Expression during Host-Microbe Interaction

6.1 Introduction

This chapter deals with assessing gene expression alterations due to host-bacterial interaction. We investigated *Streptococcus mutans*, a known resident of the oral cavity, for its effect on gene expression (related to cancer development) in OC cells. Cavities, also known as tooth decay, are brought on by certain bacteria that create acid and other decay-causing substances, causing the dentin and enamel of the tooth to erode (Mark 2018). *S. mutans* is the bacteria extensively linked to tooth decay; and the mechanism of action in tooth decay is well studied (Loesche 1986; Ranganathan and Akhila 2019). The metabolites of *S. mutans* alter the oral flora and environment to promote colonisation and the development of dental plaque. It also exploits three virulence factors linked to the carcinogenicity once well-established in dental cavities.

6.1.1 Virulence factors of *S. mutans*

Glycan synthesis, the capacity to become more acid tolerant, and lactic acid generation are the three pathogenicity factors associated with *S. mutans* (Banas 2004). Glycans are reported to play a part in the growth and spread of tumours. Through altering the amount of functional E-cadherin at the cell-cell border, glycans influence host cell-cell adhesion and speed up the invasion and dissociation of tumour cells. Numerous studies have shown that the N-glycosylation of cell surface glycoproteins is changed in cancer and that there are many complex N-glycans present as the tumour progresses (Li et al. 2016). Increased lactate is linked to a bad prognosis in a number of human cancers, and elevated glycolysis is a metabolic emblem of cancer. Although lactic acid is not the only sign of the glycolytic flow, it does directly influence the development and growth of tumours (Niu et al. 2021).

Although *S. mutans* virulence and pathogenesis are well-known, there is a lack of knowledge regarding the involvement of *S. mutans* in oral carcinogenesis. With the present knowledge of pathogenic factors produced by *S. mutans,* which possibly can contribute to cancer, we performed a preliminary investigation of the role of *S. mutans* in deregulation of genes involved in cancer.

6.2 Materials and Methods

6.2.1 Materials

AW13516 oral cancer cell line was a gift from Dr. S.V. Chiplunkar (ACTREC, Navi Mumbai). Brain Heart Infusion Agar and Brain Heart Infusion Broth were purchased from HiMedia Laboratories Pvt Ltd. Dulbecco's Modified Eagle's Medium (DMEM), foetal bovine serum (FBS), Dulbecco's Phosphate Buffered Saline (DPBS), 0.25% Trypsin-EDTA with phenol red indicator were procured from Gibco, USA. Penicillin and streptomycin were purchased from Sigma, USA. 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was purchased from Sigma-Aldrich, India. Dimethyl sulfoxide (DMSO) was purchased from MP Biomedicals, India. All reagents, media, buffers, chemicals and plastic ware (SPL Life Science, Korea) used for cell studies were of cell culture grade.

6.2.2 Oral cell line procurement and maintenance

The oral cancer cell line (AW13516) used in this study was procured from Tata Memorial Centre- Advanced Centre for Treatment, Research and Education (ACTREC), Navi Mumbai. The cell line was isolated and established in 1990 by Takate et al. (Tatake et al. 1990). The cell line was isolated from a 35 year old Indian male suffering from Squamous cell carcinoma of the oral cavity (tongue) (T4N1M0). It is a fast-growing cell line with doubling time 35.5hrs (Tatake et al. 1990).

The cell line was originally obtained and grown in Iscove's Modified Dulbecco's Medium (IMDM) media supplemented with L-glutamine. Following which the cells were adapted to Dulbecco's Modified Essential Medium (DMEM) supplemented with 10% Fetal bovine serum (FBS). For cryopreservation purposes, the freezing mixture was prepared with FBS: DMSO (9:1) with a density of 1.5 X 10^6 cells/ml.

6.2.3 Growth and maintenance of *S. mutans*

Streptococcus mutans (*S. mutans*) culture was grown and maintained in St. Brain Heart Infusion (BHI broth and BHI agar). The culture was freshly grown at 37°C under anaerobic conditions using anaerobic jar for 48 hrs.

6.2.4 Standardization of *S. mutans* cell density

The culture density was standardized using BioTek EPOCH2 microplate reader at 600 nm. The initial objective was to optimize the optical density (bacterial cell density) to be greater than 10^8. This was required for multiplicity of infection (MOI) for infection purposes that will range from 1-1000. Different O.D.s were taken and CFU/ml was determined after plating dilutions of *S. mutans* culture on St. BHI agar and incubating in anaerobic jar for 48hrs. Based on the results, a culture density adjusted to 10^8 was used in the as 1:10 serial dilution for infection purposes to obtain various MOIs.

6.2.5 Cytotoxicity assay (MTT) for standardization of parameters

Actively growing AW13516 culture was grown in complete DMEM media until minimum 85% confluency was reached. The cells were harvested by washing the flask containing cells followed by trypsinization for 1-2 mins. The harvested cells were centrifuged at 1200rpm for 6 minutes at RT. The supernatant was discarded and the cell pellet was resuspended in 1ml complete DMEM media. The cell density was determined using the haemocytometer counting method using trypan blue as a staining dye.

Three variable parameters were used in the assay: the cell density (Table 6.1) (10000 cells/well), time of infection (0hr.1hr, 3hr, 5hr, 24hr) and MOI (1,10,100,1000).

After seeding, the cells were incubated for 24hrs for attachment and proliferation. After 24hrs, the cells were infected with 100µl actively growing culture of *S. mutans* resulting in different MOIs.

Table 6.1: Bacterial cell densities used for infection at variable MOIs

MOI	Bacterial cell density for 10000 cells/well
1	10^4
10	10^5
100	10^6
1000	10^7

After infection, at specified time points, 20ul 5mg/ml MTT was added to all the wells and incubated for 2hrs at 37ºC. After incubation, the media from all wells was removed completely and 100µl DMSO was added to dissolve formazan. The plate was incubated at RT

under dark conditions on a rocker for 15 mins followed by measuring optical density on ELISA microplate reader at 570 nm and 650nm. Similarly, MTT was performed on only bacterial cells with all bacterial cell densities which were used as bacterial control to adjust the blank. This was performed since bacterial cells also have the ability to reduce MTT thereby contributing to the absorbance. The assay was performed in triplicates.

- Following equation was used for absorbance calculation:

  Actual absorbance= (Absorbance at 570nm - Absorbance at 650nm)

  Percent (%) Viability= (Actual absorbance X 100) / Absorbance of (Cell control + Bacterial control)

6.2.6 Microscopic examination of infection

Sterile coverslips were placed in each well of 6-well plate for culturing. AW13516 cells were harvested at minimum 80% confluency as described above (Section 6.2.5). The cell density was adjusted to seed approximately 0.3 X 10^6 cells per well. The plates were incubated for 24 hrs for the cells to attach and proliferate. After 24hrs, the cells were infected with OD adjusted *S. mutans* cells grown in St. BHI broth under anaerobic conditions. As the MOI of infection was set as 100, the *S. mutans* cells of density 1 X 10^7 cells/ml were used for infection purposes. The infected cells were incubated at 37°C for 1hr, 3hr, 7hr and 24hrs. After the set time of infection was complete, the coverslips were washed in PBS twice to get rid of excess unadhered bacteria and media. After washing, the slide was dehydrated and fixed using 10% formaldehyde followed by staining with methylene blue. The cells and bacteria stained were observed under light microscope for morphological changes in cells.

6.2.7 Infecting oral cell line with *S. mutans*, RNA isolation and cDNA synthesis

The AW13516 cells were grown until a minimum 80% is reached and cells are harvested as mentioned above (6.2.5). For RNA isolation purposes, the cells were seeding in sterile 60mm dishes at cell density of 1.0 X 10^6 cells. These cells were incubated at 37°C for 24hrs for attachment and proliferation. After 24hrs, these cells were infected with bacterial cells (*S. mutans*) at an MOI of 100, i.e., 1.0 X 10^8 cells of *S. mutans* were used for infection. The infected cells were incubated at 37°C for predetermined time points (3hr, 7hr and 24hr). Uninfected cells were used as control. After the determined time period, cells were harvested for RNA isolation. All the media from the 60mm dish was removed and the cells were washed with

sterile DPBS under sterile conditions. 1ml TRI reagent was added to the culture dish and swirled around to ensure exposure to all the cells attached. The suspension was triturated forcefully till the contents dissolved completely. The cell suspension was collected in a sterile eppendorf tube. 100µl chloroform was added to 1ml suspension containing cells dissolved in TRI reagent and shaken vigorously for 30-40 seconds followed by incubation at RT for 2-3 mins. The contents were centrifuged at 12000rcf for 15 minutes at 4°C. The aqueous layer was carefully collected into a fresh eppendorf and 500µl 100% isopropyl alcohol (IPA) was added followed by incubation at RT for 10 minutes. The contents were centrifuged again at 12000rcf for 10 mins at 4°C. A pellet was formed at the bottom. The eppendorf was incubated at -20°C for 2hrs or overnight. The contents were centrifuged again at 12000rcf for 10 mins at 4°C. After centrifugation, the supernatant was discarded and 500µl 75% EtOH (ethanol) was added and centrifuged at 7500 rcf for 7 mins at 4°C. This step was repeated again. The EtOH was removed and 1ml chilled 100% EtOH was added to the pellet followed by centrifugation at 7500rcf for 7 mins at 4°C. The EtOH supernatant was discarded completely and the tube was kept at RT for 10 mins for drying. Care was taken to not dry the pellet too much to avoid difficulty in dissolving. 40µl DEPC water was added to the pellet and triturated till pellet dissolved. Isolated RNA was stored at -20°C until further use. The concentration and purity of RNA isolated was checked using a spectrophotometer.

The isolated RNA was converted to cDNA using GeneSure First Strand cDNA Synthesis kit as per manufacturer's protocol. Briefly, 1µl Oligo dT primer and Random hexamer primer was added to 1µg template RNA and the volume was made up to 12µl using nuclease free water. The reaction was heated at 65°C for 5 mins followed by addition of 4.0µl 5X reaction buffer, 1.0µl RNAse inhibitor, 2.0µl 10mM dNTP mix, 1.0µl M-MuLV reverse transcriptase to make the total volume up to 20.0µl. The reaction was set at 25 °C for 5 mins followed by 42°C for 60 mins and termination at 70 °C for 5 mins.

6.2.8 Gene expression study using real time quantitative PCR

At least one gene belonging to every hallmark of cancer was targeted in the study protocol (Table 6.2). The primers required for assessing the gene expression were a kind gift from Dr. Ekta Khattar (Associate Professor, Sunandan Divatia School of Science, NMIMS University, Mumbai). GAPDH was used as a housekeeping gene. Cyclin D1 was used to assess proliferation, Bax and BCl2 for apoptosis, Interleukins for inflammation, Stathmin for invasion,

MMP9 for metastasis and bFGF for angiogenesis. Primers were standardized to obtain their optimal annealing temperature and standard curve was performed to assess PCR efficiency. The primer sequences used for qPCR are mentioned in Table 6.2.

Table 6.2: List of genes and primer sequences used for qPCR.

Pathway	**Gene name**	**Primer sequence**
Housekeeping	Glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)	F: 5'- GTCAGTGGTGGACCTGACCT-3'
		R: 5'- CACCACCCTGTTGCTGTAGC -3'
Proliferation	CyclinD1	F: 5'- GTGCTGCGAAGTGGAAACCATC-3'
		R: 5'-GACCTCCTTCTGCACACATTTGA-3'
Survival/ Apoptosis	B-cell lymphoma-2 (Bcl-2)	F: 5'- CCTGTGGATGACTGAGTACCTG-3'
		R: 5'- CTTGTGGCCCAGATAGGCAC-3'
	BCl-associated X protein (Bax)	F: 5'- GAACTGGACAGTAACATGGAGC-3'
		R: 5'- GTCCAGCCCATGATGGTTCT-3'
Inflammation	Human interleukin-6 (hIL-6)	F: 5'- AGAGTAGTGAGGAACAAGCCAGAG-3'
		R: 5'- CATTTGCCGAAGAGCCCTCAG-3'
	Interleukin-1β (IL-1β)	F: 5'- TCGTTATCCCATGTGTCGAA-3'
		R: 5'- GCTGAGGAAGATGCTGGTTC-3'
Invasion	Stathmin	F: 5'- GTGAAAGAACTGGAGAAGCGTG-3'
		R: 5'- CTTCATGGGACTTGCGTCTTT-3'
Metastasis	Matrix Metalloproteinase-9 (MMP9)	F: 5'-GGGCAGATTCCAAACCTTTGAG-3'
		R: 5'- CACCAAACTGGATGACGATGTC-3'
Angiogenesis	b-Fibroblast Growth Factors-5 (FGF5)	F: 5'- ACTTCTATGTATTGCTGAGGCAT-3'
		R: 5'- CGAGGAGTTTTCAGCAACAAATT-3'

The real time PCR setup was as mentioned **(Table 6.3)**

PowerUp™ SYBR® Green Master mix	5.0 µL
Forward primer	0.5µL (final concentration 0.5µM)
Reverse primer	0.5µL (final concentration 0.5µM)
cDNA	25ng
DEPC water	Make up volume to 10 µL

The PCR program was set as follows **(Table 6.4)**

Initial denaturation	95°C for 3 minutes
Denaturation	95°C for 15 sec
Annealing	54/56/60/62 (°C) for 30secs
Elongation	72°C for 30 secs
Meltpeak	60 to 95°C at 0.5 °C s-1

All real time PCR runs were performed on Applied Biosystems™ StepOnePlus™ Real-Time PCR System (ThermoFisher Scientific). All the experiments and runs were performed in triplicates to ensure biological and technical replicates.

6.2.9 Statistical Analysis

For gene expression analysis, ΔΔCt method was used followed by One-way ANOVA was performed and Dunnett multiple comparison test using GraphPad Prism (version. 8.0.2).

6.3 Results

6.3.1 Cell line was successfully propagated and maintained

The cell line AW13516 was successfully cultured and maintained in our lab set-up. Figure 6.1 shows inverted phase contrast microscopic images of the adherent cell line. The characteristic squamous cells were seen and the cell line grew as a monolayer.

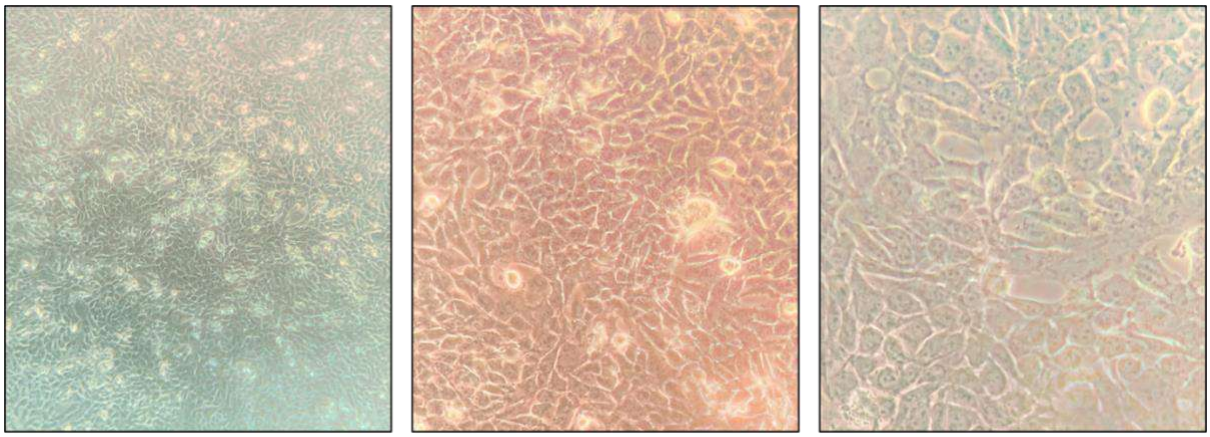


Fig. 6.1 : Phase contrast microscopy images of the cell line as observed under 10X, 20X and 40X magnifications.

6.3.2 Parameters standardized using cytotoxicity assay

The results of the MTT assay denoting cytotoxicity are shown in Figure 6.2. No significant cytotoxicity has been observed at all MOIs tested (1,10,100,1000) at time periods 1hr, 3hr and 5hr. Close to 100% host cell viability was observed at time points (1hr, 3hr, 5hr). At 24hrs after infection, the cell viability is found to be reduced (~50%). This depicts 50% cell death which is not favourable in the study since a different array of cell signalling pathways is activated during cell apoptosis/necrosis, and this is not expected for the study of interest that concentrates on cancer pathways. Since the aim of the study is to identify the effect of bacterial infection on pathways involved in cancer and not death, 24hrs

sample was not processed further.

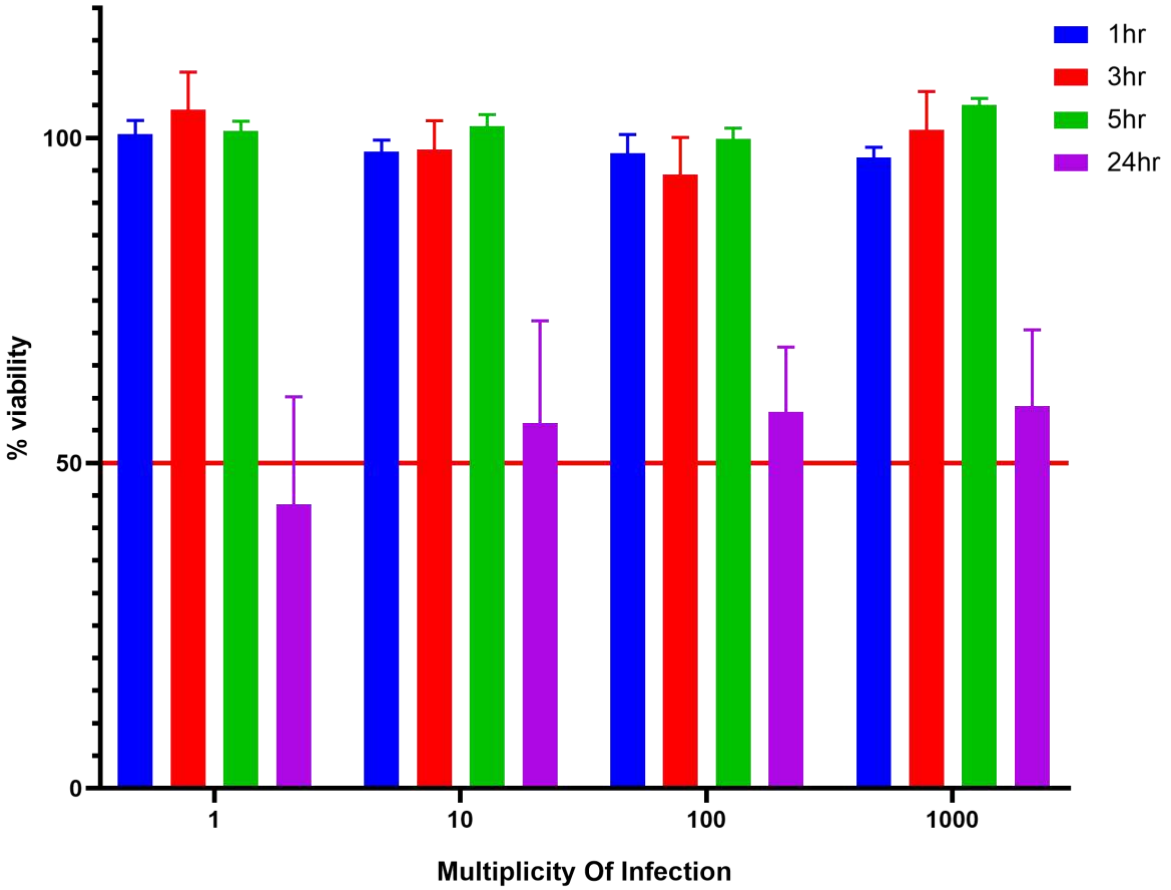


Fig. 6.2: Effect of bacterial infection at various MOIs (1,10,100,1000) on mitochondrial function in AW13516. Cell viability was determined by MTT assay after 1hr, 3hr,5hr, 24hr. Data are mean ± SD. The graph was plotted on GraphPad Prism software (version 8.0.2). The 50% viability of cells is represented by the red line.

6.3.3 Microscopic observation of infecting AW13516 cells with *S. mutans*

To confirm the results by visualization, a methylene blue staining assay was performed. Figure 6.3 and 6.4 below depicts the results obtained on staining cells with methylene blue after infection with *S. mutans.* Normal squamous cell morphology was observed in control uninfected cells. In the slide infected with *S. mutans* for 1hr, along with the host cell, bacterial cells could be observed as well. Although the number of bacterial cells observed were few and no apparent change in the host-cell morphology was observed. The bacterial cells observed were distant from host cells with no evident attachment or interaction.

At 3hrs of infection, a greater number of bacterial chains of Streptococci could be observed as compared to 1hr. These chains of *S. mutans* seemed to have clustered near cells at multiple locations, but no change in host-cell morphology could be observed. At 5hrs of infection much more bacterial cells could be observed all over the slide and more bacterial cells were seen to have interacted with host cells. Similar results were observed at 7hrs of infection and the number of infected host-cells and bacteria were much greater as compared to any other time points. At this time point, some changes in the morphology of few host cells could be observed. At 24hrs of infection, a much lower number of host cells were observed. This could be due to cell detachment after death. Although the remaining cells attached showed loss of characteristic cellular, nuclear morphology depicted their progress towards cell death. The observed chains of Streptococci were enormous clustered forming microcolonies.

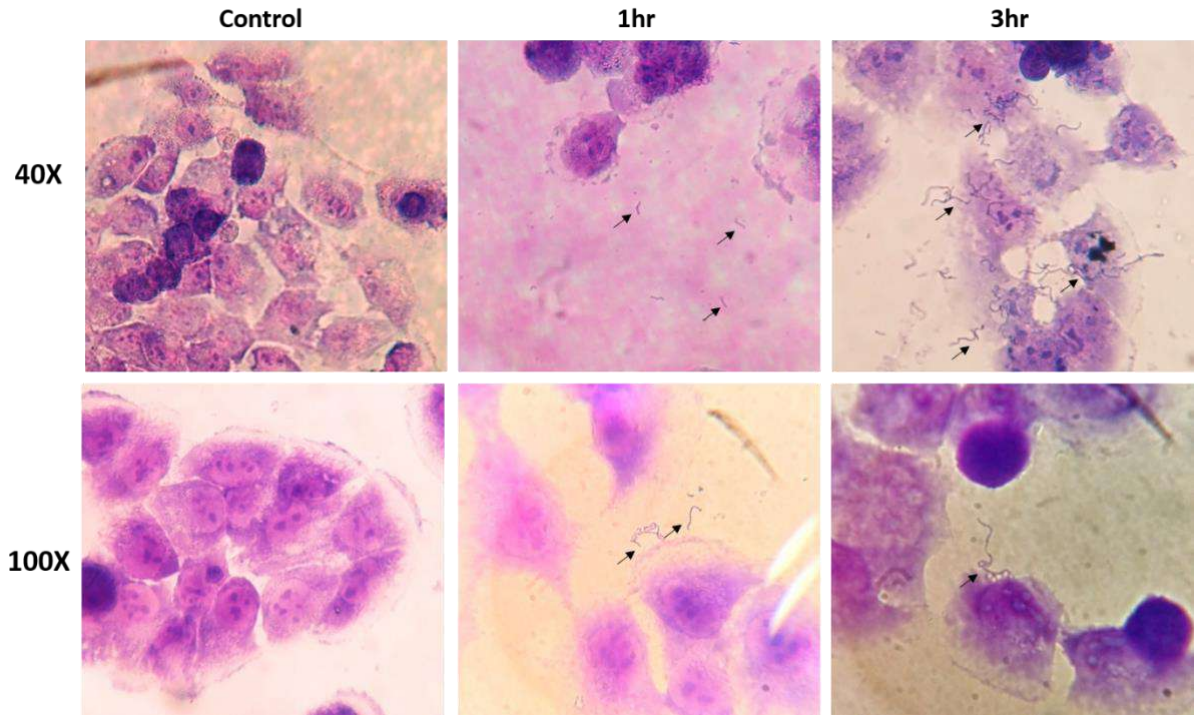


Figure 6.3: Methylene blue stained AW13516 cells infected with *S. mutans* at 1hr and 3hrs, with uninfected cell control.

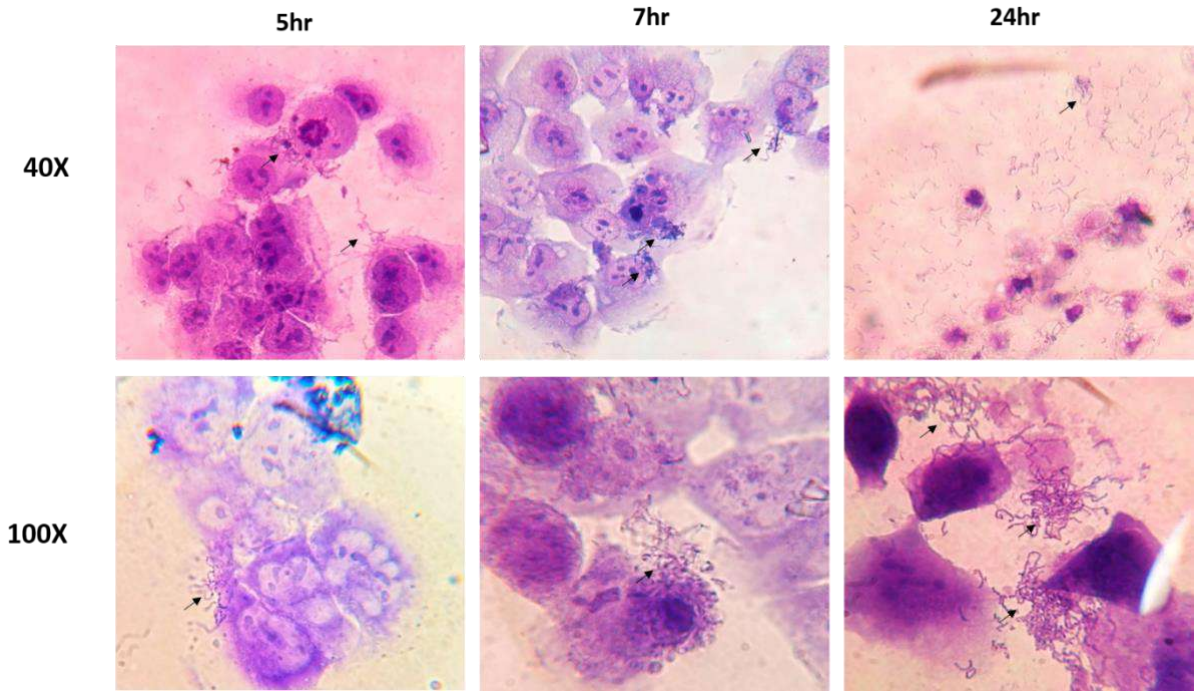


Fig. 6.4: Methylene blue stained AW13516 cells infected with *S. mutans* at 5hr, 7hr and 24hrs.

6.3.4 Gene expression analysis post infection with *S. mutans*

The primers obtained were standardized prior to evaluating gene expression by optimizing the annealing temperature and evaluating primer efficiency. The results obtained are tabulated below in Table 6.5. The primer efficiency ranged between 90-110%.

Table 6.5: List of primers, annealing temperature, standard curve equation and PCR efficiency associated with the primers used for qPCR.

Pathway	Gene name	Annealing temp (°C)	Standard curve	PCR efficiency
Housekeeping	GAPDH	60	-3.2105x + 32.095	104.87%
Proliferation	CyclinD1	60	-3.4059x + 35.801	96.61%
Survival/ Apoptosis	BCl-2	60	-3.1505x + 40.296	107.69%
	Bax	62	-3.111x + 38.355	107%
Inflammation	hIL-6	62	-3.4011x + 39.672	96.80%
	IL-1b	60	-3.1639x + 40.123	107.05%
Invasion	Stathmin	60	-3.2257x + 35.273	104.18%
Metastasis	MMP9	54	-3.1683x + 36.778	106.84%
Angiogenesis	bFGF5	56	-3.1797x + 40.87	106.30%

The expression of targeted genes at various time points, namely 3hrs, 7hrs and uninfected control systems has been depicted in Fig 6.5. Out of the eight genes considered for the study, statistically significant upregulation of genes Stathmin, CyclinD1, BCL2 were observed with increase in expression from 3hrs to 7hrs as compared to control uninfected cells. On the other hand, expression of FGF5 was found to be downregulated with increasing time of infection. The expression of IL1b and MMP9 was decreased with increase in time of infection, whereas that of hIL6 was increased but these changes were non-significant. The upregulation of Stathmin, Cyclin D1 and BCl2 suggest an activation of inflammation, survival/apoptosis and invasion related hallmarks of cancer.

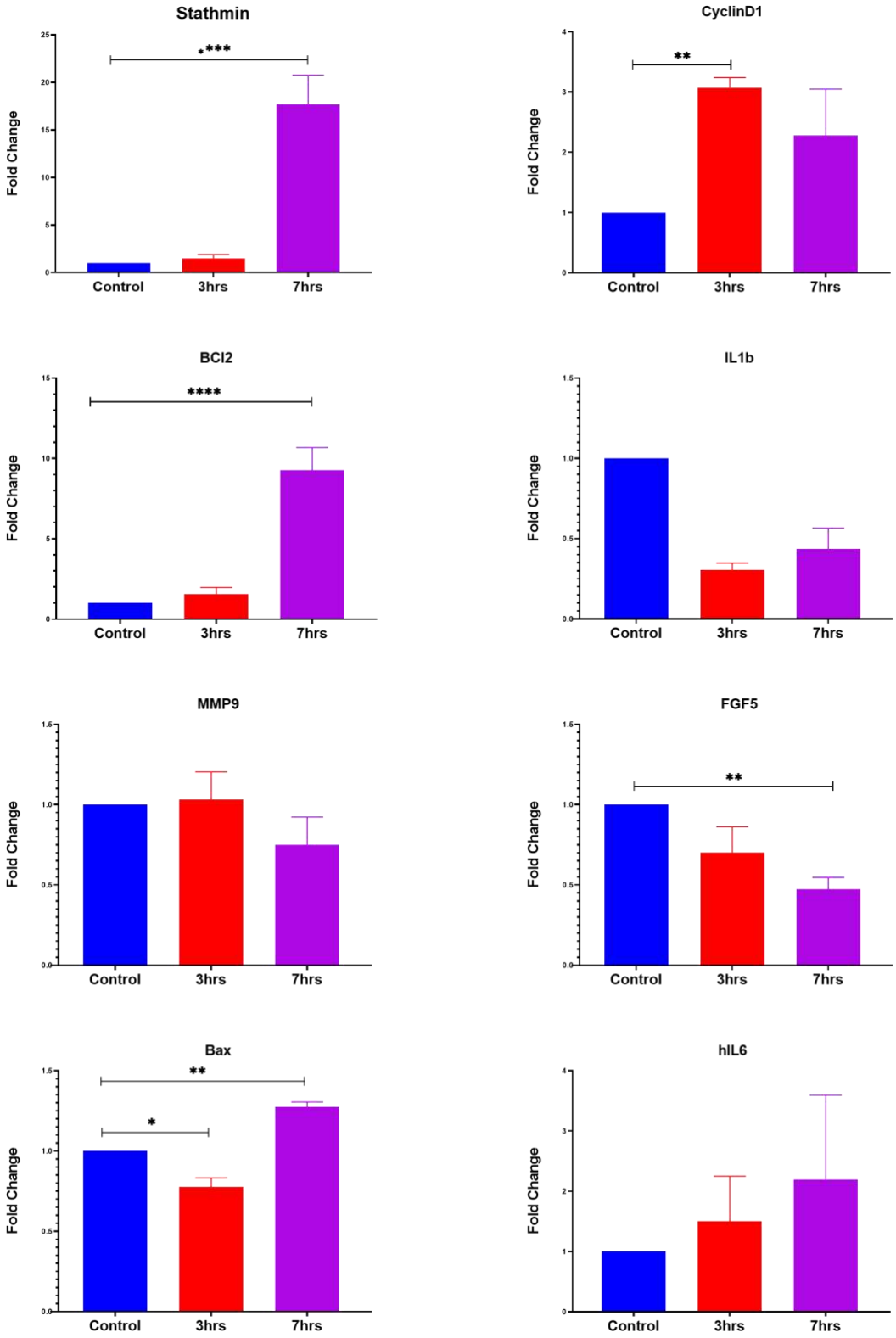


Fig.6.5: mRNA expression of AW13516 cells infected with *S. mutans* at 3hr, 7hr as compared to uninfected control. Data are mean ± SD. The graph was plotted on GraphPad Prism software (version 8.0.2).

6.4 Discussion

A variety of microbes thrive in the optimal ecological niche of the human oral cavity, and some of these pathogens are capable of causing serious clinical disorders. These clinical problems may appear in forms that could have extremely negative effects. Certain bacterial species predominate in the oral cavity, with the Streptococcus species, especially *S. mutans* being one of the main invaders. This bacterial colonization results in production of glycan and lactic acid that has a significant impact on the infection. It is known that glycan, a hydrolytic proteoglycan with the capacity to stick to epithelial cells, plays a pathogenic role. The production of glycan and lactic acid are known virulence factors of *S. mutans* along with high acid tolerance. The increased exposure to glycan and lactic acid have been proven to contribute towards oral carcinogenesis.

In the current study we aimed to analyse certain host gene expressions due to the effect of infection of oral cells with *S. mutans*. The genes under investigation were chosen such that each of the six hallmarks of cancer could be targeted by at least one gene for preliminary assessment (Sinevici and O'sullivan 2016). The multiplicity and time of infection are crucial variables of the study which determine the fate of the host cells which were standardised using MTT cytotoxicity assay. Since pathways altered in cancer were of interest, the experiment's goal was to select MOI and a time of infection where host cells did not experience apoptosis and necrosis. Hence the time point of 24hrs post-infection was eliminated from further course of assays since approximately 50% cell death was observed at the time point. Further, to confirm and visualize the evidence of infection and attachment, cell staining and microscopic observations was performed.

The visualization results suggested time dependent attachment and replication of *S. mutans* after adhering to the oral host cells. The attachment was observed at 3 hrs of infection, after which the bacteria formed microcolonies at the site of attachment. *S. mutans* are known to bind to cells in the oral cavity by various mechanisms. The expression of collagen-binding proteins (CBPs), which *S. mutans* can interact with, gives this oral pathogen an alternative to the sucrose-dependent mode of colonisation that is typically linked to the development of caries. Due to collagen's prevalence and distribution throughout the human body, *S. mutans* is given the chance to infect several host sites due to its strict adhesion to this molecule. It has been demonstrated that surface proteins such SpaP, WapA, Cnm, and Cbm bind collagen

in vitro, and it has been hypothesised that these molecules are involved in the colonisation of oral and extra-oral tissues (Avilés-Reyes et al. 2017). Since no attachment of the bacteria to the host cells was observed at 1hr of infection, the possibility of any change in the gene expression would be negligible. The 24hr infection observations support results obtained by MTT assay depicting host cell death. Therefore, further gene expression analysis study was performed at time points lying between 1hr and 24hrs, i.e 3hr and 7hr.

The genes under study confirmed upregulation of Cyclin D1, Stathmin and BCl2. Increased expression of CyclinD1 is associated with uncontrolled proliferation of cells (Inoue and Fry 2015). Similarly, overexpression of stathmin is linked to many human malignancies, thereby also called Oncoprotein 18 (Op18). Stathmin directly causes microtubule catastrophe or sequesters free tubulin dimers to depolymerize microtubules at the molecular level. Since both stathmin overexpression and downregulation cause failure in the proper completion of cell division, stathmin has been suggested to play a vital function in the regulation of mitosis (Belletti and Baldassarre 2011). Similarly, BCl2, a known antiapoptotic oncogene was found overexpressed on infection with *S. mutans* in our study. Overexpression of BCl2 and Bax:BCL2 ratio is often investigated to confirm the prevention of apoptosis in host cells. Our findings support the anti-apoptotic nature of host cells, thereby suggesting activation of cancer pathways.

6.5 Conclusion

In this chapter, we conclude that *S. mutans*, being the causative agent of tooth decay and dental caries, have pathogenic factors which can trigger oral carcinogenesis. These streptococci adhere to host oral cells followed by deregulation of genes involved in pathways of cancer. In our study, we observed the streptococci adhered to host cells after 3 hrs of infection and led to deregulation of Stathmin, CyclinD1, Bax, Bcl2 and FGF5 over time. Since this was a preliminary assessment, detailed host-microbe investigation at gene and protein expression would be required to confirm the pathways deregulated. Furthermore, drugs for controlling *S. mutans* populations, targeting attachment of *S. mutans* to host cells can be designed which could alter/prevent attachment and thereby prevent deregulation of genes leading to carcinogenesis.

Chapter 8

Summary and Conclusion

In present study, the Indian healthy, tobacco chewer and oral cancer oral microbiota was characterized using a 16S rRNA metagenomics approach with Illumina MiSeq Sequencing, followed by analysis with the QIIME2™ pipeline. Majority of the oral bacterial composition belonged to five major phyla, which include Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria. A total of 16 out of the 94 genera identified contributed up to 85% of total bacterial makeup, out of which, *Streptococcus, Neisseria, Rothia* and *Veillonella* depicted higher abundance in control population, whereas *Prevotella, Fusobacterium* and *Porphyromonas* displayed higher abundance in OC patients. The abundance of major bacterial genera in tobacco chewers was in between that of control and oral cancer patients. Further, Linear discriminant analysis of effect size (LEfSe) was performed and distinguished microbial biomarkers for each study group were identified. In the current study, we identified *Cardiobacterium* and *Bifidobacterium* as biomarkers for Tobacco chewing population and oral cancer patients respectively for the first time. The within-group (alpha) and between-group (beta) diversity were identified using various indices and matrices respectively. Maximum alpha diversity was observed in tobacco chewers followed by healthy and oral cancer patients. On the other hand, beta diversity using PCoA plots demonstrated maximum diversity between control and OC as well as tobacco and OC study groups, whereas diversity between control and tobacco study groups was lower and these groups were like each other. Following diversity analysis, the functionally differential pathways expressed were predicted using PICRUSt2, which suggested pathways related to amino acid synthesis were comparatively overexpressed in control population, those related to reductive TCA overexpressed in tobacco groups, whereas pathways for lipid biosynthesis and fatty acid elongation were overexpressed in OC population.

To compensate for the disadvantages of 16S rRNA gene-based metagenomics approach used for enumeration purpose (*i.e.,* presence of multiple copies of 16S rRNA in every bacterial cell) in microbial ecology, a simple, fast, cost-effective, sensitive, and specific method where qPCR assays were used against a single copy *rpoB* gene, for absolute enumeration of bacteria in oral samples. Using this method, we performed enumeration of eight significant bacteria involved in OC in our samples. The abundance of P*. gingivalis, F. nucleatum, C. gingivalis, P. melaninogenica* and *V. parvula* was significantly higher in the oral cancer group as compared to healthy individuals,

whereas that of *S. mutans* was significantly higher in the control group. The abundance of each bacteria in all three study groups correlated well with the reports published earlier, thereby providing confidence in the method reliability. For the enumeration of bacteria, additional characteristics such as sex, age, dietary habits, cancer stage, and cancer locations were also evaluated.

In order to investigate host-bacterial interactions, related to OC *in vitro*, *S. mutans (*a well-known cariogenic oral bacterium) cells were incubated with an OC cell-line. The cariogenic feature is owed to its capacity to adhere to the tooth surface by generating sticky extracellular polysaccharides from sucrose, as well as fermenting sucrose and other sugars to acids that can damage the tooth enamel. Microscopic observations indicated interaction/adherence between oral cells and bacteria between 1hr to 7hrs post-infection. Time dependent gene expressions were studied which depicted significantly increased expression of Stathmin, CyclinD1, BCl2, and Bax: Bcl2 ratio and decreased expression of FGF5 with time. The results indicate dysregulation of genes involved in cancer related pathways.

To summarize, the current study has helped in deciphering the composition and differences in the oral microbiota of the healthy Indian population and subjects suffering from oral cancer, as well as long term tobacco chewers population. The study also led to the identification of group specific bacterial biomarkers. Another part of the study resulted in the development of a simple qPCR method for absolute quantification of significant bacteria involved in oral cancer. Such an approach can be used for early diagnosis of oral microbial biomarkers. The method developed can be further utilized in different area applications to quantify bacteria from a given sample. The last part of the study was a preliminary investigation to understand the role of *S. mutans* in oral cancer, and that has resulted in identification of deregulated genes involved in cancer. This can be further explored and followed up by examining more genes related to deregulated pathways and performing a multiomics approach to understand the molecular mechanism of host-microbe interactions. The study paved way for more investigations and can be further extended in a few directions like validating oral cavity biomarkers for clinical application, using

the qPCR assays for other important microbial populations studies, proteomics, transcriptomics, and metabolomics approaches to understand the changes in oral cavity of different subject group, which can be correlated with microbial composition and diversity changes between different groups, and how it could be linked with development of cancer..

Bibliography

Bibliography

- Abbott, Mark A., et al. "Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified in vitro." *Journal of Infectious Diseases* 158.6 (1988): 1158-1169.
- Adams, Jerry M., and Suzanne Cory. "The Bcl-2 apoptotic switch in cancer development and therapy." *Oncogene* 26.9 (2007): 1324-1337.
- Adékambi, Toïdi, Michel Drancourt, and Didier Raoult. "The rpoB gene as a tool for clinical microbiologists." *Trends in microbiology* 17.1 (2009): 37-45.
- Al-Hebshi, Nezar Noor, et al. "Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples." *Journal of oral microbiology* 7.1 (2015): 28934.
- Amano, Atsuo, et al. "Variations of Porphyromonas gingivalis fimbriae in relation to microbial pathogenesis." *Journal of periodontal research* 39.2 (2004): 136-142.
- Ambardar, Sheetal, et al. "High throughput sequencing: an overview of sequencing chemistry." *Indian journal of microbiology* 56.4 (2016): 394-404.
- Amer, Abdrazak, et al. "The microbiome of potentially malignant oral leukoplakia exhibits enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia species." *Frontiers in microbiology* 8 (2017): 2391.
- Ansorge, Wilhelm J. "Next-generation DNA sequencing techniques." *New biotechnology* 25.4 (2009): 195-203.
- Avilés-Reyes, Alejandro, et al. "Collagen-binding proteins of Streptococcus mutans and related streptococci." *Molecular oral microbiology* 32.2 (2017): 89-106.
- Baeriswyl, Vanessa, and Gerhard Christofori. "The angiogenic switch in carcinogenesis." *Seminars in cancer biology*. Vol. 19. No. 5. Academic Press, 2009.
- Bainbridge, Brian W., and Richard P. Darveau. "Porphyromonas gingivalis lipopolysaccharide: an unusual pattern recognition receptor ligand for the innate host defense system." *Acta Odontologica Scandinavica* 59.3 (2001): 131-138.

- Baluk, Peter, Hiroya Hashizume, and Donald M. McDonald. "Cellular abnormalities of blood vessels as targets in cancer." *Current opinion in genetics & development* 15.1 (2005): 102-111.
- Banas, Jeffrey A. "Virulence properties of Streptococcus mutans." *Frontiers in Bioscience-Landmark* 9.2 (2004): 1267-1277.
- Becker, Mitzi R., et al. "Molecular analysis of bacterial species associated with childhood caries." *Journal of clinical microbiology* 40.3 (2002): 1001-1009.
- Belletti, Barbara, and Gustavo Baldassarre. "Stathmin: a protein with many tasks. New biomarker and potential target in cancer." *Expert opinion on therapeutic targets* 15.11 (2011): 1249-1266.
- Behjati, Sam, and Patrick S. Tarpey. "What is next generation sequencing?." *Archives of Disease in Childhood-Education and Practice* 98.6 (2013): 236-238.
- Belstrøm, Daniel, et al. "Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status." *Journal of oral microbiology* 6.1 (2014): 23609.
- Bennedsen, Astrid LB, et al. "The gut microbiota can orchestrate the signaling pathways in colorectal cancer." *APMIS* 130.3 (2022): 121-139.
- Berx, Geert, and Frans Van Roy. "Involvement of members of the cadherin superfamily in cancer." *Cold Spring Harbor perspectives in biology* 1.6 (2009): a003129.
- Bhowmick, Neil A., Eric G. Neilson, and Harold L. Moses. "Stromal fibroblasts in cancer initiation and progression." *Nature* 432.7015 (2004): 332-337.
- Bik, Elisabeth M., et al. "Bacterial diversity in the oral cavity of 10 healthy individuals." *The ISME journal* 4.8 (2010): 962-974.
- Bizzarro, Sergio, et al. "Subgingival microbiome in smokers and non-smokers in periodontitis: an exploratory study using traditional targeted techniques and a next-generation sequencing." *Journal of clinical periodontology* 40.5 (2013): 483-492.
- Blasco, Maria A. "Telomeres and human disease: ageing, cancer and beyond." *Nature Reviews Genetics* 6.8 (2005): 611-622.
- Bolz, J., et al. "Bacterial colonization of microbial biofilms in oral squamous cell carcinoma." *Clinical oral investigations* 18.2 (2014): 409-414.

- Brooks, Philip J., and Jacob A. Theruvathu. "DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis." *Alcohol* 35.3 (2005): 187-193.
- Case, Rebecca J., et al. "Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies." *Applied and environmental microbiology* 73.1 (2007): 278-288.
- Cankovic, Milos, et al. "The histological characteristics of clinically normal mucosa adjacent to oral cancer." *Journal of cancer research and therapeutics* 9.2 (2013): 240.
- Čanković, Miloš, et al. "Bacterial flora on the surface of oral squamous cell carcinoma." *Archive of Oncology* 21.2 (2013): 62-64.
- Cavallaro, Ugo, and Gerhard Christofori. "Cell adhesion and signalling by cadherins and Ig-CAMs in cancer." *Nature Reviews Cancer* 4.2 (2004): 118-132.
- Chakravorty, Soumitesh, et al. "A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria." *Journal of microbiological methods* 69.2 (2007): 330-339.
- Chaushu, Stella, et al. "Direct recognition of Fusobacterium nucleatum by the NK cell natural cytotoxicity receptor NKp46 aggravates periodontal disease." *PLoS pathogens* 8.3 (2012): e1002601.
- Cheng, Nikki, et al. "Transforming growth factor-β signaling–deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion." *Molecular Cancer Research* 6.10 (2008): 1521-1533.
- Cho, Ilseung, and Martin J. Blaser. "The human microbiome: at the interface of health and disease." *Nature Reviews Genetics* 13.4 (2012): 260-270.
- Chocolatewala, Noureen M., and Pankaj Chaturvedi. "Role of human papilloma virus in the oral carcinogenesis: an Indian perspective." *Journal of cancer research and therapeutics* 5.2 (2009): 71.
- Chocolatewala, Noureen, Pankaj Chaturvedi, and Rushikesh Desale. "The role of bacteria in oral cancer." *Indian journal of medical and paediatric oncology* 31.04 (2010): 126-131.
- Cole, James R., et al. "Ribosomal Database Project: data and tools for high throughput rRNA analysis." *Nucleic acids research* 42.D1 (2014): D633-D642.

- Coussens, Lisa M., and Zena Werb. "Inflammation and cancer." *Nature* 420.6917 (2002): 860-867.
- Curtis, M. A., J. Aduse-Opoku, and M. Rangarajan. "Cysteine proteases of Porphyromonas gingivalis." *Critical Reviews in Oral Biology & Medicine* 12.3 (2001): 192-216.
- Czernilofsky, A. P., C. G. Kurland, and G. Stöffler. "30S ribosomal proteins associated with the 3'-terminus of 16S RNA." *FEBS letters* 58.1-2 (1975): 281-284.
- Dahllöf, Ingela, Harriet Baillie, and Staffan Kjelleberg. "rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity." *Applied and environmental microbiology* 66.8 (2000): 3376-3380.
- Davies, M. A., and Y. Samuels. "Analysis of the genome to personalize therapy for melanoma." *Oncogene* 29.41 (2010): 5545-5555.
- Deo, Priya Nimish, and Revati Deshmukh. "Oral microbiome: Unveiling the fundamentals." *Journal of oral and maxillofacial pathology: JOMFP* 23.1 (2019): 122.
- Deng, D. M., and J. M. Ten Cate. "Demineralization of dentin by Streptococcus mutans biofilms grown in the constant depth film fermentor." *Caries Research* 38.1 (2004): 54-61.
- Edwards, Andrew M., Tracy J. Grossman, and Joel D. Rudney. "Fusobacterium nucleatum transports noninvasive Streptococcus cristatus into human epithelial cells." *Infection and immunity* 74.1 (2006): 654-662.
- Edwards, A. M., T. J. Grossman, and J. D. Rudney. "Association of a high-molecular weight arginine-binding protein of Fusobacterium nucleatum ATCC 10953 with adhesion to secretory immunoglobulin A and coaggregation with Streptococcus cristatus." *Oral microbiology and immunology* 22.4 (2007): 217-224.
- Estaki, Mehrbod, et al. "QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data." *Current protocols in bioinformatics* 70.1 (2020): e100.
- Facklam, Richard. "What happened to the streptococci: overview of taxonomic and nomenclature changes." *Clinical microbiology reviews* 15.4 (2002): 613-630.

- Fagundes, José Américo G., et al. "Porphyromonas gingivalis is associated with protease-activated receptor-2 upregulation in chronic periodontitis." *Journal of periodontology* 82.11 (2011): 1596-1601.
- Fardini, Yann, et al. "Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity." *Molecular microbiology* 82.6 (2011): 1468-1480.
- Ferlay, Jacques, et al. "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012." *International journal of cancer* 136.5 (2015): E359-E386.
- Gambhir, Ramandeep Singh, et al. "Dental public health in India: An insight." *Journal of family medicine and primary care* 5.4 (2016): 747.
- Gambhir, R. S., and T. Gupta. "Need for oral health policy in India." *Annals of medical and health sciences research* 6.1 (2016): 50-55.
- Ghosh, Shreya, and Alok Prasad Das. "Metagenomic insights into the microbial diversity in manganese-contaminated mine tailings and their role in biogeochemical cycling of manganese." *Scientific reports* 8.1 (2018): 1-12.
- Giglio, Steven, Paul T. Monis, and Christopher P. Saint. "Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR." *Nucleic acids research* 31.22 (2003): e136-e136.
- González-Escalona, Narjol, et al. "Quantitative reverse transcription polymerase chain reaction analysis of Vibrio cholerae cells entering the viable but non-culturable state and starvation in response to cold shock." *Environmental microbiology* 8.4 (2006): 658-666.
- Goodwin, Sara, John D. McPherson, and W. Richard McCombie. "Coming of age: ten years of next-generation sequencing technologies." *Nature Reviews Genetics* 17.6 (2016): 333-351.
- Gopinath, Divya, et al. "Salivary bacterial shifts in oral leukoplakia resemble the dysbiotic oral cancer bacteriome." *Journal of oral microbiology* 13.1 (2021): 1857998.
- Grada, Ayman, and Kate Weinbrecht. "Next-generation sequencing: methodology and application." *The Journal of investigative dermatology* 133.8 (2013): e11.

- Griffen, Ann L., et al. "Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing." *The ISME journal* 6.6 (2012): 1176-1185.
- Guerrero-Preston, Rafael, et al. "16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment." *Oncotarget* 7.32 (2016): 51320.
- Gupta, Bhawna, et al. "Associations between oral hygiene habits, diet, tobacco and alcohol and risk of oral cancer: a case–control study from India." *Cancer epidemiology* 51 (2017): 7-14.
- Hajishengallis, George, et al. "Complement receptor 3 blockade promotes IL-12-mediated clearance of Porphyromonas gingivalis and negates its virulence in vivo." *The Journal of Immunology* 179.4 (2007): 2359-2367.
- Hajishengallis, George, Min Wang, and Shuang Liang. "Induction of distinct TLR2-mediated proinflammatory and proadhesive signaling pathways in response to Porphyromonas gingivalis fimbriae." *The journal of immunology* 182.11 (2009): 6690-6696.
- Halboub, Esam, et al. "Tongue microbiome of smokeless tobacco users." *BMC microbiology* 20.1 (2020): 1-9.
- Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of cancer: the next generation." *cell* 144.5 (2011): 646-674.
- Hanahan, Douglas, and Robert A. Weinberg. "The hallmarks of cancer." *cell* 100.1 (2000): 57-70.
- Hanahan, Douglas, and Judah Folkman. "Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis." *cell* 86.3 (1996): 353-364.
- Han, Jee Eun, et al. "Photorhabdus insect-related (Pir) toxin-like genes in a plasmid of Vibrio parahaemolyticus, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp." *Diseases of aquatic organisms* 113.1 (2015): 33-40.
- Hasegawa, Yoshiaki, et al. "Gingival epithelial cell transcriptional responses to commensal and opportunistic oral microbial species." *Infection and immunity* 75.5 (2007): 2540-2547.

- Haugland, Richard A., et al. "Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR." *Systematic and Applied Microbiology* 33.6 (2010): 348-357.
- Healy, Claire M., and Gary P. Moran. "The microbiome and oral cancer: More questions than answers." *Oral oncology* 89 (2019): 30-33.
- Hecht, Stephen S. "Tobacco carcinogens, their biomarkers and tobacco-induced cancer." *Nature Reviews Cancer* 3.10 (2003): 733-744.
- Hecht, Stephen S. "Research opportunities related to establishing standards for tobacco products under the Family Smoking Prevention and Tobacco Control Act." *Nicotine & Tobacco Research* 14.1 (2011): 18-28.
- Henry, Leroy G., et al. "Oxidative stress resistance in Porphyromonas gingivalis." *Future microbiology* 7.4 (2012): 497-512.
- Hernandez, Brenda Y., et al. "Betel nut chewing, oral premalignant lesions, and the oral microbiome." *PLoS One* 12.2 (2017): e0172196.
- Hirata, Yoshihiro, et al. "Helicobacter pylori activates the cyclin D1 gene through mitogen-activated protein kinase pathway in gastric cancer cells." *Infection and immunity* 69.6 (2001): 3965-3971.
- Hoffmann, D. H. S. S., and S. S. Hecht. "Advances in tobacco carcinogenesis." *Chemical carcinogenesis and mutagenesis I*. Springer, Berlin, Heidelberg, 1990. 63-102.
- Hooper, Samuel J., et al. "Viable bacteria present within oral squamous cell carcinoma tissue." *Journal of clinical microbiology* 44.5 (2006): 1719-1725.
- Hooper, Samuel J., et al. "A molecular analysis of the bacteria present within oral squamous cell carcinoma." *Journal of medical microbiology* 56.12 (2007): 1651-1659.
- Inoue, Kazushi, and Elizabeth A. Fry. "Aberrant expression of cyclin D1 in cancer." *Signal transduction insights* 4 (2015): STI-S30306.
- Jeng, JIIANG-HUEI, M. C. Chang, and L. J. Hahn. "Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives." *Oral oncology* 37.6 (2001): 477-492.

- Junttila, Melissa R., and Gerard I. Evan. "p53—a Jack of all trades but master of none." *Nature reviews cancer* 9.11 (2009): 821-829.
- Kademani, Deepak. "Oral cancer." *Mayo Clinic Proceedings*. Vol. 82. No. 7. Elsevier, 2007.
- Kadowaki, Tomoko, et al. "Porphyromonas gingivalis proteinases as virulence determinants in progression of periodontal diseases." *The Journal of Biochemistry* 128.2 (2000): 153-159.
- Kara, Duygu, Suzanne BI Luppens, and Jacob M. ten Cate. "Differences between single- and dual-species biofilms of Streptococcus mutans and Veillonella parvula in growth, acidogenicity and susceptibility to chlorhexidine." *European journal of oral sciences* 114.1 (2006): 58-63.
- Kamble, Asmita, Shriya Sawant, and Harinder Singh. "16S ribosomal RNA gene-based metagenomics: A review." *Biomedical Research Journal* 7.1 (2020): 5.
- Klappenbach, Joel A., John M. Dunbar, and Thomas M. Schmidt. "rRNA operon copy number reflects ecological strategies of bacteria." *Applied and environmental microbiology* 66.4 (2000): 1328-1333.
- Kreimer, Aimee R., et al. "Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review." *Cancer Epidemiology Biomarkers & Prevention* 14.2 (2005): 467-475.
- Kreth, Jens, et al. "Competition and coexistence between Streptococcus mutans and Streptococcus sanguinis in the dental biofilm." *Journal of bacteriology* 187.21 (2005): 7193-7203.
- Kumar, Kishore R., Mark J. Cowley, and Ryan L. Davis. "Next-generation sequencing and emerging technologies." *Seminars in thrombosis and hemostasis*. Vol. 45. No. 07. Thieme Medical Publishers, 2019.
- Kurkivuori, Johanna, et al. "Acetaldehyde production from ethanol by oral streptococci." *Oral oncology* 43.2 (2007): 181-186.
- Laine, M. L., and A. J. Van Winkelhoff. "Virulence of six capsular serotypes of Porphyromonas gingivalis in a mouse model." *Oral microbiology and immunology* 13.5 (1998): 322-325.

- Lan, Yemin, Gail Rosen, and Ruth Hershberg. "Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains." *Microbiome* 4.1 (2016): 1-13.
- Lara-Tejero, Maria, and Jorge E. Galán. "A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein." *Science* 290.5490 (2000): 354-357.
- Lax, Alistair J. "Bacterial toxins and cancer—a case to answer?." *Nature Reviews Microbiology* 3.4 (2005): 343-349.
- Lax, Alistair J., ed. *Bacterial protein toxins: Role in the interference with cell growth regulation*. Vol. 7. Cambridge University Press, 2005.
- Lee, Wei-Hsiang, et al. "Bacterial alterations in salivary microbiota and their association in oral cancer." *Scientific reports* 7.1 (2017): 1-11.
- Li, Yuanhong, et al. "Novel development of a qPCR assay based on the rpoB gene for rapid detection of Cronobacter spp." *Current Microbiology* 72.4 (2016): 436-443.
- Li, Xiang, et al. "Role of glycans in cancer cells undergoing epithelial–mesenchymal transition." *Frontiers in oncology* 6 (2016): 33.
- Lim, Yenkai, et al. "The performance of an oral microbiome biomarker panel in predicting oral cavity and oropharyngeal cancers." *Frontiers in cellular and infection microbiology* 8 (2018): 267.
- Ling, Zongxin, et al. "Analysis of oral microbiota in children with dental caries by PCR-DGGE and barcoded pyrosequencing." *Microbial ecology* 60.3 (2010): 677-690.
- Liu, Lin, et al. "Comparison of next-generation sequencing systems." *J Biomed Biotechnol* 2012.251364 (2012): 251364.
- Livak, Kenneth J., and Thomas D. Schmittgen. "Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method." *methods* 25.4 (2001): 402-408.
- Lloyd-Price, Jason, Galeb Abu-Ali, and Curtis Huttenhower. "The healthy human microbiome." *Genome medicine* 8.1 (2016): 1-11.
- Loesche, Walter J. "Role of Streptococcus mutans in human dental decay." *Microbiological reviews* 50.4 (1986): 353-380.

- Lourbakos, Afrodite, et al. "Arginine-specific protease from Porphyromonas gingivalis activates protease-activated receptors on human oral epithelial cells and induces interleukin-6 secretion." *Infection and immunity* 69.8 (2001): 5121-5130.
- Lowe, Scott W., Enrique Cepero, and Gerard Evan. "Intrinsic tumour suppression." *Nature* 432.7015 (2004): 307-315.
- Lu, Maoyang, Songyu Xuan, and Zhao Wang. "Oral microbiota: A new view of body health." *Food Science and Human Wellness* 8.1 (2019): 8-15.
- Lv, Ruiling, et al. "Detection and Quantification of Viable but Non-culturable Campylobacter jejuni." *Frontiers in microbiology* 10 (2020): 2920.
- Lyons, Sharon R., Ann L. Griffen, and Eugene J. Leys. "Quantitative real-time PCR for Porphyromonas gingivalis and total bacteria." *Journal of clinical microbiology* 38.6 (2000): 2362-2365.
- Mager, D. L., et al. "The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects." *Journal of translational medicine* 3.1 (2005): 1-8.
- Marsh, P. D., and R. S. Percival. "The oral microflora—friend or foe? Can we decide?." *International dental journal* 56 (2006): 233-239.
- Mark, Anita M. "Options for dealing with tooth decay." *The Journal of the American Dental Association* 149.10 (2018): 927-928.
- McDonald, Daniel, et al. "An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea." *The ISME journal* 6.3 (2012): 610-618.
- Metgud, R., K. Gupta, and J. Gupta. "Exploring bacterial flora in oral squamous cell carcinoma: a microbiological study." *Biotechnic & Histochemistry* 89.2 (2014): 153-159.
- Meurman, Jukka H. "Oral microbiota and cancer." *Journal of oral microbiology* 2.1 (2010): 5195.
- Miļuna, Sintija, et al. "Mutes mikrobioma īpatnības bezdūmu tabakas lietotājiem." *Zinātniskie raksti: 2016. gada medicīnas nozares pētnieciskā darba publikācijas*. Rīgas Stradiņa universitāte, 2017. 284-291.

- Mikx, F. H. M., and J. S. Van der Hoeven. "Symbiosis of Streptococcus mutans and Veillonella alcalescens in mixed continuous cultures." *Archives of Oral Biology* 20.7 (1975): 407-410.
- Mok, Shao Feng, et al. "The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity." *The Malaysian journal of pathology* 39.1 (2017): 1-15.
- Moritani, K., et al. "Acetaldehyde production by major oral microbes." *Oral diseases* 21.6 (2015): 748-754.
- Morita, Eiko, et al. "Different frequencies of Streptococcus anginosus infection in oral cancer and esophageal cancer." *Cancer science* 94.6 (2003): 492-496.
- Moon, J-H., J-H. Lee, and J-Y. Lee. "Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients." *Molecular oral microbiology* 30.3 (2015): 227-241.
- Müller, P., et al. "Tissue damage in the rabbit oral mucosa by acute and chronic direct toxic action of different alcohol concentrations." *Experimental Pathology* 24.2-3 (1983): 171-181.
- Multania, S., and D. Saranath. "Single nucleotide polymorphisms and risk of oral cancer: Indian Case-Control Study." *J Clin Cell Immunol* 7.448 (2016): 2.
- Mukherjee, Arghya, et al. "Metagenomic approaches for understanding microbial communities in contaminated environments: Bioinformatic tools, case studies and future outlook." *Metagenomics to Bioremediation*. Academic Press, 2023. 103-156.
- Murshid, Ebtissam Z. "Diet, oral hygiene practices and dental health in autistic children in Riyadh, Saudi Arabia." *Oral Health Dent Manag* 13.1 (2014): 91-6.
- Nadkarni, Mangala A., et al. "Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set." *Microbiology* 148.1 (2002): 257-266.
- Nagy, K. N., et al. "The microflora associated with human oral carcinomas." *Oral oncology* 34.4 (1998): 304-308.
- Nagy, Janice A., et al. "Heterogeneity of the tumor vasculature." *Seminars in thrombosis and hemostasis*. Vol. 36. No. 03. © Thieme Medical Publishers, 2010.

- Niaz, Kamal, et al. "Smokeless tobacco (paan and gutkha) consumption, prevalence, and contribution to oral cancer." *Epidemiology and health* 39 (2017).
- Niu, Dun, et al. "Lactic acid in tumor invasion." *Clinica Chimica Acta* 522 (2021): 61-69.
- Nougayrède, Jean-Philippe, et al. "Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle." *Trends in microbiology* 13.3 (2005): 103-110.
- O'Hara, Ann M., and Fergus Shanahan. "The gut flora as a forgotten organ." *EMBO reports* 7.7 (2006): 688-693.
- Olsen, Ingar, and Özlem Yilmaz. "Possible role of Porphyromonas gingivalis in orodigestive cancers." *Journal of oral microbiology* 11.1 (2019): 1563410.
- Olsen, Ingar. "The oral microbiome in health and disease." *Oral infections and general health*. Springer, Cham, 2016. 97-114.
- Park, Soon-Nang, Jae-Yoon Park, and Joong-Ki Kook. "Development of Porphyromonas gingivalis-specific quantitative real-time PCR primers based on the nucleotide sequence of rpoB." *The Journal of Microbiology* 49.2 (2011): 315-319.Pavlova, Sylvia I., et al. "Multiple alcohol dehydrogenases but no functional acetaldehyde dehydrogenase causing excessive acetaldehyde production from ethanol by oral streptococci." *Microbiology* 159.Pt 7 (2013): 1437.
- Park, Se-Ra, et al. "Diverse Toll-like receptors mediate cytokine production by Fusobacterium nucleatum and Aggregatibacter actinomycetemcomitans in macrophages." *Infection and immunity* 82.5 (2014): 1914-1920.
- Pei, Anna Y., et al. "Diversity of 16S rRNA genes within individual prokaryotic genomes." *Applied and environmental microbiology* 76.12 (2010): 3886-3897.
- Pfister, Herbert. "The role of human papillomavirus in anogenital cancer." *Obstetrics and gynecology clinics of North America* 23.3 (1996): 579-595.
- Poretsky, Rachel, et al. "Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics." *PloS one* 9.4 (2014): e93827.
- Poddar, Aayush, et al. "Head and neck cancer risk factors in India: protocol for systematic review and meta-analysis." *Bmj Open* 8.8 (2018): e020014.

- Proctor, Lita. "Priorities for the next 10 years of human microbiome research." (2019): 623-625.
- Pushalkar, Smruti, et al. "Microbial diversity in saliva of oral squamous cell carcinoma." *FEMS Immunology & Medical Microbiology* 61.3 (2011): 269-277.
- Pushalkar, Smruti, et al. "Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma." *BMC microbiology* 12.1 (2012): 1-15.
- Quast, Christian, et al. "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." *Nucleic acids research* 41.D1 (2012): D590-D596.
- Quail, Michael A., et al. "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers." *BMC genomics* 13.1 (2012): 1-13.
- Rainey, Fred A., et al. "Clostridium paradoxum DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences." *Microbiology* 142.8 (1996): 2087-2095.
- Ranganathan, V., and C. H. Akhila. "Streptococcus mutans: has it become prime perpetrator for oral manifestations." *J Microbiol Exp* 7.4 (2019): 207-213.
- Renouf, Vincent, et al. "Lactic acid bacteria evolution during winemaking: use of rpoB gene as a target for PCR-DGGE analysis." *Food Microbiology* 23.2 (2006): 136-145.
- Rickard, Alexander H., et al. "Bacterial coaggregation: an integral process in the development of multi-species biofilms." *Trends in microbiology* 11.2 (2003): 94-100.
- Ritalahti, Kirsti M., et al. "Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains." *Applied and environmental microbiology* 72.4 (2006): 2765-2774.
- Ritchie, Justine M., et al. "Human papillomavirus infection as a prognostic factor in carcinomas of the oral cavity and oropharynx." *International journal of cancer* 104.3 (2003): 336-344.
- Rocha, Danilo JP, Carolina S. Santos, and Luis GC Pacheco. "Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis." *Antonie Van Leeuwenhoek* 108.3 (2015): 685-693.

- Rodríguez-Beltrán, Jerónimo, et al. "Beyond horizontal gene transfer: the role of plasmids in bacterial evolution." *Nature Reviews Microbiology* 19.6 (2021): 347-359.
- Rosan, Burton, and Richard J. Lamont. "Dental plaque formation." *Microbes and infection* 2.13 (2000): 1599-1607.
- Samaranayake, Lakshman, and Victor H. Matsubara. "Normal oral flora and the oral ecosystem." *Dental Clinics* 61.2 (2017): 199-215.
- Sanjaya, P. R., et al. "Candida in oral pre-cancer and oral cancer." *Medical hypotheses* 77.6 (2011): 1125-1128.
- Sarode, Gargi, et al. "Epidemiologic aspects of oral cancer." *Disease-a-Month* 66.12 (2020): 100988.
- Sasaki, M., et al. "Streptococcus anginosus infection in oral cancer and its infection route." *Oral diseases* 11.3 (2005): 151-156.
- Schlegel, Richard, and Hutton D. Slade. "Bacteriocin production by transformable group H streptococci." *Journal of Bacteriology* 112.2 (1972): 824-829.
- Schmidt, Brian L., et al. "Changes in abundance of oral microbiota associated with oral cancer." *PloS one* 9.6 (2014): e98741.
- Schwabe, Robert F., and Christian Jobin. "The microbiome and cancer." *Nature Reviews Cancer* 13.11 (2013): 800-812.
- Seitz, Helmut K., et al. "Alcohol and cancer." *Alcoholism: Clinical and Experimental Research* 25 (2001): 137S-143S.
- Shay, Jerry W., and Woodring E. Wright. "Hayflick, his limit, and cellular ageing." *Nature reviews Molecular cell biology* 1.1 (2000): 72-76.
- Simon-Soro, A., et al. "Microbial geography of the oral cavity." *Journal of dental research* 92.7 (2013): 616-621.
- Sims, T. J., et al. "Immunoglobulin G response of periodontitis patients to Porphyromonas gingivalis capsular carbohydrate and lipopolysaccharide antigens." *Oral microbiology and immunology* 16.4 (2001): 193-201.

- Smith, Cindy J., and A. Mark Osborn. "Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology." *FEMS microbiology ecology* 67.1 (2009): 6-20.
- Spurgers, K. B., et al. "Molecular mediators of cell death in multistep carcinogenesis: a path to targeted therapy." *Cell Death & Differentiation* 13.8 (2006): 1360-1370.
- Srivastava, Ankita, et al. "Comparative and analytical characterization of the oral bacteriome of smokeless tobacco users with oral squamous cell carcinoma." *Applied Microbiology and Biotechnology* (2022): 1-14.
- Stone, Trevor W., and L. Gail Darlington. "Microbial carcinogenic toxins and dietary anti-cancer protectants." *Cellular and Molecular Life Sciences* 74.14 (2017): 2627-2643.
- Sundin, Olof H., et al. "The human jejunum has an endogenous microbiota that differs from those in the oral cavity and colon." *BMC microbiology* 17.1 (2017): 1-17.
- Sung, Hyuna, et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 71.3 (2021): 209-249.
- Takahashi, N. "Oral microbiome metabolism: from “who are they?” to “what are they doing?”." *Journal of dental research* 94.12 (2015): 1628-1637.
- Tatake, Revati J., et al. "Establishment and characterization of four new squamous cell carcinoma cell lines derived from oral tumors." *Journal of cancer research and clinical oncology* 116.2 (1990): 179-186.
- Tandon, Parul, et al. "The prevalence of squamous cell carcinoma in different sites of oral cavity at our Rural Health Care Centre in Loni, Maharashtra–a retrospective 10-year study." *Contemporary Oncology/Współczesna Onkologia* 21.2 (2017): 178-183.
- Tanner, A. C. R., et al. "Microbiota of severe early childhood caries before and after therapy." *Journal of dental research* 90.11 (2011): 1298-1305.
- Tran, Nham, Barbara R. Rose, and Christopher J. O'Brien. "Role of human papillomavirus in the etiology of head and neck cancer." *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck* 29.1 (2007): 64-70.

- Uehara, Osamu, et al. "Alteration of oral flora in betel quid chewers in Sri Lanka." *Journal of Microbiology, Immunology and Infection* 54.6 (2021): 1159-1166.
- VanGuilder, Heather D., Kent E. Vrana, and Willard M. Freeman. "Twenty-five years of quantitative PCR for gene expression analysis." *Biotechniques* 44.5 (2008): 619-626.
- Varshitha, A. "Prevalence of oral cancer in India." *Journal of Pharmaceutical Sciences and Research* 7.10 (2015): 845.
- Vogelstein, Bert, and Kenneth W. Kinzler. "The multistep nature of cancer." *Trends in genetics* 9.4 (1993): 138-141.
- Vos, Michiel, et al. "A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity." *PloS one* 7.2 (2012): e30600.
- Wang, Alice M., Michael V. Doyle, and David F. Mark. "Quantitation of mRNA by the polymerase chain reaction." *Proceedings of the National Academy of Sciences* 86.24 (1989): 9717-9721.
- Wang, Baohong, et al. "The human microbiota in health and disease." *Engineering* 3.1 (2017): 71-82.
- Wang, Bing-Yan, and Howard K. Kuramitsu. "Interactions between oral bacteria: inhibition of Streptococcus mutans bacteriocin production by Streptococcus gordonii." *Applied and environmental microbiology* 71.1 (2005): 354-362.
- Wang, Min, et al. "Fimbrial proteins of Porphyromonas gingivalis mediate in vivo virulence and exploit TLR2 and complement receptor 3 to persist in macrophages." *The Journal of Immunology* 179.4 (2007): 2349-2358.
- Wang, P-L., and K. Ohura. "Porphyromonas gingivalis lipopolysaccharide signaling in gingival fibroblasts–CD14 and Toll-like receptors." *Critical Reviews in Oral Biology & Medicine* 13.2 (2002): 132-142.
- Wang, Pao-Li, et al. "Effect of anti-CD14 antibody on experimental periodontitis induced by Porphyromonas gingivalis lipopolysaccharide." *The Japanese Journal of Pharmacology* 89.2 (2002): 176-183.

- Warnakulasuriya, Saman, Newell W. Johnson, and I. Van der Waal. "Nomenclature and classification of potentially malignant disorders of the oral mucosa." *Journal of oral pathology & medicine* 36.10 (2007): 575-580.
- Warnakulasuriya, Saman. "Global epidemiology of oral and oropharyngeal cancer." *Oral oncology* 45.4-5 (2009): 309-316.
- Weinstock, George M. "Genomic approaches to studying the human microbiota." *Nature* 489.7415 (2012): 250-256.
- Wight, A. J., and G. R. Ogden. "Possible mechanisms by which alcohol may influence the development of oral cancer—a review." *Oral oncology* 34.6 (1998): 441-447.
- Wu, Dongying, Guillaume Jospin, and Jonathan A. Eisen. "Systematic identification of gene families for use as “markers” for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups." *PloS one* 8.10 (2013): e77033.
- Xu, Biyang, and Yiping W. Han. "Oral bacteria, oral health, and adverse pregnancy outcomes." *Periodontology 2000* 89.1 (2022): 181-189.
- Xue, Jiaping, Suping Yang, and Seyha Seng. "Mechanisms of cancer induction by tobacco-specific NNK and NNN." *Cancers* 6.2 (2014): 1138-1156.
- Yang, Bo, Yong Wang, and Pei-Yuan Qian. "Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis." *BMC bioinformatics* 17.1 (2016): 1-8.
- Yang, Shun-Fa, et al. "Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer." *Oral oncology* 77 (2018): 1-8.
- Yang, Ning-Yan, et al. "Progression of periodontal inflammation in adolescents is associated with increased number of Porphyromonas gingivalis, Prevotella intermedia, Tannerella forsythensis, and Fusobacterium nucleatum." *International journal of paediatric dentistry* 24.3 (2014): 226-233.
- Yilmaz, Özlem, Kiyoko Watanabe, and Richard J. Lamont. "Involvement of integrins in fimbriae-mediated binding and invasion by Porphyromonas gingivalis." *Cellular microbiology* 4.5 (2002): 305-314.

- Yu, Hsu-Sheng, et al. "Formation of acetaldehyde-derived DNA adducts due to alcohol exposure." *Chemico-biological interactions* 188.3 (2010): 367-375.
- Zhao, Hongsen, et al. "Variations in oral microbiota associated with oral cancer." *Scientific reports* 7.1 (2017): 1-10.
- Zhang, G., R. Chen, and J. D. Rudney. "Streptococcus cristatus attenuates Fusobacterium nucleatum-induced interleukin-8 expression in oral epithelial cells." *Journal of periodontal research* 43.4 (2008): 408-416.
- Zipper, Hubert, et al. "Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications." *Nucleic acids research* 32.12 (2004): e103-e103.

Annexures

Annexure I: Participant Information Sheet

PROJECT - INFORMATION SHEET

Title: Identification of the Microbiome in Oral Cavity of Oral cancer patients and clinical implications

Brief Introduction:

The microbiome is a collection of microorganisms present in different areas of humans like skin, oral cavity and human gut. Most of the microorganisms in/on human body are beneficial. However, certain organisms have been associated with different diseases and specific cancers. Identifying microbiome organisms in oral cavity will be helpful in analysis of the number and type of organisms and its association with oral cancer and tobacco habits.

You are requested to participate in the study voluntarily. The Principal Investigator (PI) is Dr. Harinder Singh, Asst. Professor, Biological Sciences, Sunandan Divatia School Of Science, NMIMS (Deemed-to-be) University; and Co-Investigator (Co-I) is Dr. Deepak Parikh, Prof. and Chief, Dept. of Head & Neck Surgery, Somaiya Ayurvihar - Asian Cancer Institute, Off Eastern Express Highway, Behind Everard nagar, Somaiya Ayurvihar, Sion (East), Mumbai.

Aim of the Project

The aim of the project is to investigate the differences in oral cavity microbial population between the oral cancer patients, long term tobacco users, and healthy individuals with no tobacco habit/oral cancer. The study involves identifying the oral microbial diversity and its association with tobacco use and oral cancer.

Duration of the study and number of participants

The duration of study is three year
The number of participants in the study will be minimum 10.
You will be one of the individuals who will participate in this study.

What is requested of you? And what does it involve?

You will be requested to give 1-3 ml of mouth rinse on rinsing with sterile physiological saline in a container. The sterile saline solution is a mild salt solution (0.85% salt solution).

You will also be requested to answer a questionnaire relevant to the project, to get information on your contact details (address, phone number, e-mail address), demographic data such as age, sex, tobacco habits and alcohol habits.

If you are diagnosed as an oral cancer patient, relevant clinico-pathological information will be noted from your medical file. This information will be helpful in our project, where we need to identify the microbial diversity in oral cavity and correlate it with lifestyle habits and oral cancer.

Risk and Benefits of the participants

There is no risk involved to you as a result of your participation in the project and donating sputum/saliva/oral rinse, as an oral rinse is absolutely safe and is similar to a routine activity like rinsing your mouth with water.

There is no direct personal benefit to you for the voluntary participation on the project. However, the research project will provide valuable information, which may benefit the society and science by identifying the oral microbial diversity in Indian population and its association with tobacco usage and oral cancer.

Confidentiality

Your name and all the information which you provide will be strictly confidential. Your identity will be kept confidential with a unique identity number for each participant, which will not be available to other than the PI/Co-I on the project.

Voluntary Participation

Your consent and participation on the project is strictly voluntary, with no coercion from any of the associated medical, research or social workers on the study. You may ask questions on the

project for any clarification on the project. You reserve the right to withdraw from the project at any time, and this will not cause any consequence in your treatment and patient medical care, and your association with any personnel on the project.

If you do not understand any aspect explained to you, the coordinator will explain it to any relative/friend/member who accompanies you. The information will also be given to you as a written sheet.

If you agree to participate in the project voluntarily, you will sign an informed consent form provided to you. You are requested to go through the same and sign for voluntary participation. If you do not understand the written informed consent or are not literate but would like to participate in the project, your relative accompanying you or legal heir (a person who is legally authorized to sign on your behalf) will be explained the project and your voluntary participation by the coordinator PI/Co-I of the project in the project.

Signed consent form

The informed consent form will be given to you for your signature, after the project details have been explained to you in the language you understand. Besides, the patient information sheet will be given to you for your reference, forms are available in English or other regional languages.

Contact for further information

If you have any query regarding the research study at any time you can contact the Principal Investigator Dr. Harinder Singh, Assistant Professor, Sunandan Divatia School of Science, NMIMS (Deemed-to-be) University, at +91 22 4235 5938 between 10:00 am to 5:00 pm, during week days.

Name and signature of the PI:

Name and signature of Co-I/ Clinician:

Name of Hospital/ Institute:

Date: Place:

Annexure II: Informed Consent Form

INFORMED CONSENT FORM

Project: 'Identification of the Microbiome in Oral Cavity of Oral cancer patients and clinical implications'

PI: Dr. Harinder Singh

Co-Investigator: Dr. Deepak Parikh

Informed Consent

I have been invited to participate in a research study on '**Identification of the Microbiome in Oral Cavity of Oral cancer patients and clinical implications'.** I have been explained that the project involves examining the differences in the microbial population between patients with oral cancer, long term tobacco users, and healthy individuals.

On the project, I will be required to donate 1-3 ml oral rinse/sputum/saliva, and requested to answer a questionnaire relevant to the project, including my contact details, demographic data such as age, sex, tobacco habits, alcohol habit, and clinico-pathological cancer related information from my medical file. The oral rinse will be given by using sterile saline (0.85% Salt solution).

It has been explained to me that the donation of the sample (1-3 ml oral saline rinse) has no associated risk and will not produce any harm as it is a simple activity like rinsing mouth with low quantity of salt in water.

In confirmation, I provide my signature below as proof of my acceptance to participate in the study. My signature below indicates that I have read the patient information sheet, understand its meaning, have had a chance to ask questions, have had these questions answered to my satisfaction, and consent to my participation in this program. I give my consent to have 1-3 ml oral rinse/sputum/saliva collected from me, and the necessary information on the project, for the research study; '**Identification of the Microbiome in Oral Cavity of Oral cancer patients and**

clinical implications'. I have also been explained about possible risk and benefits of this research work and I am willing to participate in this study.

My participation on the project is voluntary with no coercion from any of the associated medical, research or social workers on the study. I understand that I reserve the right to withdraw my participation on the project, and will suffer no consequence of my withdrawal on my medical care or any other relevant aspect. I will be given a signed copy of this consent form.

I hereby make a donation of 1-3 ml oral rinse for the purpose of the mentioned research/study/education. I have been explained the entire procedure and purpose of this study in the language known to me. I, hereby certify that I have signed this agreement freely and voluntarily, without any threat or coercion and am willing to participate in this research study, with a clear understanding about the likely benefits and the risks involved. My signature also indicates that I have received a copy of the consent form. I also understand that in the event of my not participating in this study or withdrawal of consent at any time, my treatment will not be affected any way.

Name of the participant:

Age of the participant:

Signature or thumb impression of donor:

Witnessed by:

Unrelated to the donor:

Date of Consent: Place

Signature of Co-I/Clinician:

Name of Hospital/ Institute:

Date: Place:

Annexure III: Patient Data Sheet

PATIENT DATA SHEET FOR CASES

***Please fill the applicable data**

O **Oral Cancer Case** O Long Term Tobacco User

Name: _____

Date of Sample Collection (DD/MM/YYYY): _____

Registration No.:_____

Name of Oncologist/Clinician:

_____

Name of Hospital:

_____

Date of Diagnosis (DD/MM/YYYY): _____ Age

(in years): _____ Date of Birth (DD/MM/YYYY): _____

Residential Address, e/mail address, and Phone No.:

_____

_____

***Please tick the correct option given below**

Sex: Male/Female

Habits:

Habit	Frequency (No. of times/day) or Quantity (Nos./sachets/packets)	Duration (months/years)
O None	_____	_____
O Tobacco chewing	_____times/day	_____
O Areca nut chewing	_____times/day	_____
O Pan masala	_____times/day	_____
O Gutka	_____Sachet/day	_____
O Mawa	_____times/day	_____
O Cigarette smoking	_____Cig./day	_____
O Bidi smoking	_____bidis/day	_____
O Alcohol	_____quantity/week	_____

Pre-malignant condition:

O None

O Leukoplakia

Homogenous: Simple / Smooth, Plaque like / Corrugated / Fine Lines / Wrinkled / Papillomatous

Non-Homogenous: Speckled /Nodular / Verrucous / Exophytic / Irregular Flat

O Erythroplakia

O Melanoplakia

O Submucous fibrosis

Any other information

Primary Cancer Site:

- O Buccal Mucosa (left/right)
- O Tongue
- O Floor of the Mouth
- O Alveolus and Gingiva (upper/lower)
- O Hard Palate
- O Lip / Labial mucosa
- O Retromolar Trigone

Tumor size as per TNM classification

Not Known / Tis / T1 / T2/ T3 / T4a / T4b / Tx

Nodal Status:

Not Known / N0/ N1/ N1/ N2a / 2b / 2c / N3 / NX

Distant Metastasis:

Not Known / M0/ M1 / NX

Histopathological Report:

Squamous cell carcinoma / Verrucous carcinoma

Well differentiated / Moderately differentiated / Poorly differentiated

Stage:

Stage 0 / I / II / III / IVA / IVB / IVC

Annexure IV: Data Sheet for Controls

DATA SHEET FOR CONTROLS

***Please fill the applicable data**

O **Control**

Name: _____

Date of Sample Collection (DD/MM/YYYY): _____

Registration No.:_____

Age (in years): _____ Date of Birth (DD/MM/YYYY): _____

Participant's Address and Phone No.:

_____

_____

***Please tick the correct option given below**

Sex: Male/Female

Habits:

Habit	Frequency (No. of times/day) or Quantity (Nos./sachets/packets)	Duration (months/years)
O None	_____	_____
O Tobacco chewing	_____times/day	_____
O Areca nut chewing	_____times/day	_____
O Pan masala	_____times/day	_____
O Gutka	_____Sachet/day	_____
O Mawa	_____times/day	_____
O Cigarette smoking	_____Cig./day	_____
O Bidi smoking	_____bidis/day	_____
O Alcohol	_____quantity/week	_____

Dental hygiene related questionnaire

Name:

1. Do you clean your teeth?
   Yes No

2. How do you clean your teeth?
   i. Toothbrush and toothpaste
   ii. Toothbrush and powder
   iii. Electronic toothbrush
   iv. Mouthwash
   v. Others (Datun/Finger/charcoal powder)

3. How often do you brush your teeth each day?

   i. Once
   ii. Twice
   iii. More than twice
   iv. Sometimes

4. How long do you normally take to brush your teeth?

   i. About 30 seconds
   ii. About 1 minute
   iii. About 2 minutes
   iv. More than 2 minutes

5. What type of brushing method do you employ?

   i. Vertical
   ii. Horizontal
   iii. Combined

6. Which secondary methods to control plaque do you use?

   i. Dental floss
   ii. Interdental brushes
   iii. Toothpicks
   iv. None

7. Do you clean your tongue?

Yes No

8. How often do you visit the dentist?

i. Only in problem
ii. Once in 3 months
iii. Once in 6 months
iv. Between 1 and 2 years

9. Have you suffered from any dental issues in the past 6 months?

Yes No

If yes, specify._ _____

10. Have you consumed any antibiotics in the past 1 months? If yes, please specify.

Ethics Committee Approval

1. SVKM's Institutional Ethics Committee Approval

SVKM'S NMIMS Deemed to be UNIVERSITY

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SVKM's INSTITUTIONAL ETHICS COMMITTEE (IEC)

30th August, 2016

To
Principal Investigator
Sunandan Divatia School of Science,
NMIMS

Sub: Project proposal No: NMIMS/IEC/008/2016

The Committee has reviewed and approved the proposal entitled "Identification of the microbiome in oral cavity of oral cancer patients and clinical implications" in its meeting held on 4th July 2016 for a period of three years. The progress report and final report of the proposal should be submitted to the Committee.

Signature of the Member-Secretary
(Dr. Brijesh Sukumaran)

Signature of the Chairman
(Dr. Rabindranath Mukhopadhyay)

Copy forwarded to protocol applicant:
Copy in File

3rd Floor, Bhaidas Sabhagriha Building, Vile Parle (West), Mumbai - 400 056.
Tel: (91-22) 42355958 / 59 | Fax: (91-22) 26114512 | Email: admissions.sos@nmims.edu | Web: www.nmims.edu

2. K.J.Somaiya Medical College and Hospital Ethics Committee Approval.

K. J. SOMAIYA MEDICAL COLLEGE & HOSPITAL, MUMBAI
INSTITUTIONAL SCIENTIFIC COMMITTEE
[ESTABLISHED 1997]
Reg No. ECR/138/Inst/MH/2013
Tel.: 24090253, Extn.2215, Fax: 022 24091855

Date: 12th September 2017

To,
Dr. Harinder Singh / Dr. Deepak Parikh
Asian Cancer Institute
K. J. Somaiya Medical College,
Sion, Mumbai.

Subject: Scientific Committee Approval

The Institutional Scientific Committee of K. J. Somaiya Medical College & Hospital in its meeting held on 11th September 2017, has reviewed and discussed your project titled – 'Identification of the microbiome in oral cavity of oral cancer patients and clinical implications'.

The following members of the Scientific Committee were present at the meeting:

Dr. V. A. Sabnis	Chairperson	Dean & Prof of Pathology
Dr. Sunil Kowli	Member	Senior Consultant (External)
Dr. Nilay Chakrabarti	Member	HOD Surgery
Dr. Niharika Gill	Member	HOD Medicine
Dr. Usha V. Nayak	Member	HOD Pharmacology
Mr. Prashant Shah	Member	Statistician

The Institutional Scientific Committee approves the project submitted by you.
You are requested to submit a final report of your project at the end of the study.

Dr. V. A. Sabnis
Chairperson

List of Conferences/Workshops/ Seminars/Awards

List of Conferences/Workshops/Seminars Attended

Conferences and Poster presentations

1. Poster presentation at 18th Asia Pacific Congress of Clinical Microbiology and Infection, Singapore (Nov. 2021)
2. Poster presentation at 1st Student Research Congress at BNCP and University of Mumbai, Mumbai, India (Sep. 2020)
3. Poster presentation at 60th Annual Conference of AMI & International Symposium on Microbial Technologies in Sustainable Development of Energy, Environment, Agriculture and Health, Haryana, India (Nov. 2019)
4. Attended conference conducted by NMIMS (deemed-to-be) University on Advances in Material Sciences & Applied Biology, Mumbai, India (Jan. 2019)

Workshops/Seminars attended

1. "From Biology to Omics", ACTREC, Mumbai (Mar 10-12th,2021)
2. "Molecular Docking in Rational drug design", Ramnarain Ruia College, Mumbai (Feb 4-6th,2021)
3. "Computational Pharmaceutics: Can it help understand drug targeting", IWSA-NMIMS, Mumbai (Jan 7th, 2021)
4. "Understanding what and how of publishing ethics", Springer Nature (Oct 14th, 2020)
5. "Using MS Word effectively for scientific report writing", Bombay College of Pharmacy, Mumbai (May 27-29th, 2020)
6. "Technical Training in Drug Designing & NGS Data Analysis", GN Khalsa College, Mumbai (Mar. 2020)

Awards

1. Women's graduate union scholarship recipient (2020-2021)

List of Publications

List of Publications

1. **Sawant, Shriya**, Jinesh Dugad, Deepak Parikh, Sathiyaraj Srinivasan, and **Harinder Singh**. "Oral Microbial Signatures of Tobacco Chewers and Oral Cancer Patients in India." *Pathogens* 12, no. 1 (2023): 78. https://doi.org/10.3390/pathogens12010078 **(I.F:4.531)**

2. **Sawant, Shriya**, Jinesh Dugad, Deepak Parikh, Sathiyaraj Srinivasan, and **Harinder Singh**. "Identification & correlation of bacterial diversity in oral cancer and long-term tobacco chewers-A case-control pilot study." *Journal of Medical Microbiology* 70, no. 9 (2021): 001417. https://doi.org/10.1099/jmm.0.001417 **(I.F:3.196)**

3. **Sawant, Shriya**, Jinesh Dugad, Deepak Parikh, and **Harinder Singh**. "Absolute quantitation of oral bacteria involved in oral cancer by real-time PCR." *Medicine in Microecology* 7 (2021): 100034. https://doi.org/10.1016/j.medmic.2021.100034

4. Kamble, Asmita, **Shriya Sawant**, and **Harinder Singh**. "16S ribosomal RNA gene-based metagenomics: A review." *Biomedical Research Journal* 7, no. 1 (2020): 5. https://doi.org/10.4103/BMRJ.BMRJ_4_20

5. Phanse, Shirish K., **Shriya Sawant**, **Harinder Singh**, and Sudeshna Chandra. "Physico-Chemical and Antimicrobial Efficacy of Encapsulated Dhavana Oil: Evaluation of Release and Stability Profile from Base Matrices." *Molecules* 27, no. 22 (2022): 7679. https://doi.org/10.3390/molecules27227679 **(I.F: 4.927)**

pathogens

MDPI

Article

Oral Microbial Signatures of Tobacco Chewers and Oral Cancer Patients in India

Shriya Sawant [1], Jinesh Dugad [2], Deepak Parikh [2], Sathiyaraj Srinivasan [3,4,*] and Harinder Singh [1,*]

[1] Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed-to-be University, Vile Parle (West), Mumbai 400056, India
[2] Somaiya Ayurvihar-Asian Cancer Institute, Off Eastern Express Highway, Behind Everard Nagar, Somaiya Ayurvihar, Sion (East), Mumbai 400022, India
[3] Department of Bio & Environmental Technology, College of Natural Science, Seoul Women's University, Seoul 139-774, Republic of Korea
[4] Gene Strand Technologies Pvt. Ltd., Chennai 600056, India
* Correspondence: drsrini@swu.ac.kr (S.S.); harinder.singh@nmims.edu (H.S.)

Abstract: Dysbiosis of the oral microbiome has been found to play a key role in the genesis and progression of oral cancer (OC). Tobacco chewing, a risk factor for oral cancer, is also associated with oral dysbiosis. Since tobacco chewing is a lifestyle habit in the South Asian subcontinent, including India, and contributes to one-third of the global oral cancer burden; we aimed to identify the oral bacterial diversity of Indian oral cancer patients and tobacco chewers. We used 16S rRNA amplicon sequencing to study the composition of oral microbiota in OC patients and tobacco chewers in India and compared it with healthy controls. The abundance of predominant phyla, Firmicutes, and Bacteroidetes varied between the study groups. Our study identified Leptotrichia, Treponema, Lautropia, and Cardiobacterium as significantly enriched in tobacco chewers, whereas genera Pseudomonas, Capnocytophaga, and Mycoplasma were enriched in oral cancer, which could be potential biomarkers for the Indian population. Furthermore, the functional prediction revealed that genes involved in lipid biosynthesis and fatty acid elongation were upregulated in the oral cancer group, whereas those for the reductive TCA cycle were upregulated in the tobacco group. As the role of bacteria in oral cancer is becoming more evident, identification of bacterial diversity and biomarkers for tobacco chewers and OC patients can aid in the early diagnosis of OC in high-risk individuals.

Keywords: oral cancer; dysbiosis; tobacco; biomarker; diagnosis; 16S rRNA

Citation: Sawant, S.; Dugad, J.; Parikh, D.; Srinivasan, S.; Singh, H. Oral Microbial Signatures of Tobacco Chewers and Oral Cancer Patients in India. *Pathogens* **2023**, *12*, 78. https://doi.org/10.3390/pathogens12010078

Academic Editor: Biao Ren

Received: 14 December 2022
Revised: 29 December 2022
Accepted: 30 December 2022
Published: 3 January 2023

1. Introduction

Cancer of the oral cavity is one of the most common malignancies, especially in Asia, where it contributes to approximately 66% of the global oral cancer (OC) burden, with an estimated 248,360 new cases and 131,610 deaths every year. The incidence of OC appears to be increasing worldwide, and this common cancer is most prevalent among males in India [1]. Despite advances in surgical methods, adjuvant radiation, and chemotherapy, the overall 5-year survival rate of OC patients is approximately 50–60%. OC treatment success rates can be improved by early identification and interdisciplinary therapy [2].

The most well-established risk factors associated with OC include chewing tobacco, betel quid, smoking cigarettes, alcohol consumption, and HPV-16/18 [3–5]. Over 90% of the worldwide smokeless tobacco usage burden is believed to be in Southeast Asia, with over 100 million individuals using smokeless tobacco in India and Pakistan alone [6]. Moreover, it was shown that the combination of smoking, drinking alcohol, and poor oral hygiene increases the risk of OC onset due to chronic inflammation and infection, which are the main factors in cancer pathogenesis, influencing the resident microbiota involved in the oral environment's homeostasis [7,8]. Several metagenomic investigations of the microbiome have revealed microbial pattern changes in OC, which further vary depending

on the stage of OC, malignant lesions, and diseases of the oral cavity, according to reports published [9].

Along with dysbiosis in OC, reports also suggest bacterial alterations due to tobacco chewing, thereby making an individual prone to bacterial infections by inducing bacterial virulence, deregulation of host immune functions, and physiological and structural changes in the human oral cavity [10]. Increased abundance of pathogenic bacterial genera such as Fusobacterium, Cardiobacterium, Synergistes, Selenomonas, Haemophilus, and Pseudomonas has been observed in tobacco users, depicting early acquisition and colonization of pathogens in oral biofilms due to tobacco exposure [11]. Even though most OC cases arise from the Indian subcontinent and tobacco chewing is a common lifestyle habit associated with OC in the population, there is a dearth of information on the microbiome in Indian groups of subjects. Especially, the microflora in the oral cavity of healthy individuals, tobacco chewers, and oral cancer patients has not been investigated.

In the current study, we aim to identify the bacterial diversity in the oral cavity of OC patients and long-term tobacco chewers from India. We hypothesize that the variations in oral microbiomes between tobacco chewers, OC patients, and healthy people are expressed in oral rinse samples, which may be detected by 16S rRNA gene amplicon sequencing. These variations might subsequently be linked to cancer development and exploited as a biomarker panel to predict tobacco chewers with a high risk of OC in the Indian population in a clinical setting with effective diagnostic accuracy.

2. Materials and Methods

2.1. Subject Recruitment

A total of 120 participants in the study were divided into three study groups, consisting of 40 participants each healthy controls (C), patients suffering from oral squamous cell carcinoma (OC), and long-term tobacco chewers (T). The sample size was calculated using power analysis. Individuals without any documented disorders in the oral cavity, as determined by earlier clinical evaluation, were considered healthy controls. Participants categorized as long-term tobacco chewers were those who had been chewing tobacco for at least 5 years. Biopsy and pathology results validated all diagnoses among OC participants. The clinical examination of the participants' oral cavities was performed by a maxillofacial prosthodontist and a surgical oncologist. At the time of sample collection, the participants were devoid of any antibiotic treatment for a week prior to sample collection. Exclusion criteria included individuals under the age of 18, those medically compromised/unfit to give consent, subjects who were completely edentulous, and those who received oncotherapy earlier. All the samples were collected in the period from January 2018 to March 2020, in Mumbai, India. The oral cancer samples were collected from patients admitted at Somaiya Ayurvihar-Asian Cancer Institute, Mumbai.

The work described has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Medical history, age, gender, employment, cigarette and alcohol consumption habits, and general oral hygiene questions were all documented for participating individuals. All individuals provided written, informed permission prior to the sample collection. For the study, ethics approval was obtained from SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and the Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai.

2.2. Sample Collection, DNA Isolation and Sequencing

Oral rinse samples were collected from study participants as mentioned earlier [12,13]. Briefly, during the sample collection procedure, patients were asked to rinse their mouths for 30 s with sterile normal saline and spit into a sterile tube, of 50 mL. Participants were advised to abstain from eating, drinking, and doing oral hygiene procedures for at least one hour before sample collection. Salivary samples were collected in well-labeled sterile falcon tubes, stored at 4 °C, and processed within 48 h.

DNA extraction, V6–V8 hypervariable region amplification, sequencing, and processing of reads have been carried out as mentioned in our recent publication [13]. DNA was isolated using the Invitrogen PureLink™ Genomic DNA Kit (Cat no. K182002), according to the manufacturer's recommendation. The PCR amplification of bacterial 16S rRNA hypervariable region V6-V8 was carried out using primers B969F (ACG CGH NRA ACC TTA CC) and BA1406R (ACG GGC RGT GWG TRC AA). The whole sequencing process was performed using Illumina (Illumina, San Diego, CA, USA), and MiSeq libraries were quantified and then subjected to 300-nucleotide paired-end multiplex sequencing on an Illumina MiSeq sequencer.

2.3. OTU Assignment and Diversity Analyses

The quality of the reads from the sequencer was assessed using FASTQC. The resulting pairs of reads from each sample were merged to obtain longer reads to improve the quality of reads (Phred score Q > 30) and taxonomy classification using VSEARCH. The standard QIIME2 (v. 2021.2) pipeline was used to analyze microbial diversity [14]. A closed reference-based OTU selecting technique, with 97% sequence similarity to the Greengenes database (gg_13_5), was utilized to cluster readings into operational taxonomic units (OTUs) and assign taxonomy to the OTUs at different taxonomic levels.

QIIME2 was used to assess alpha and beta diversity indices. Alpha diversity was assessed by indices such as ACE indicator, Chao1 index, Goods coverage, observed OTUs, pielou_e, Shannon index, and Simpson index. Whereas, beta diversity was assessed using phylogenetic (weighted and unweighted) and non-phylogenetic (Bray-curtis and Jaccard) Linear Discriminant Analysis (LDA) matrices and plots created using PhyloToAST [15].

2.4. Identification of Biomarkers and Prediction of Metagenomes

In order to identify the potential biomarker, LDA effect size (LEfSe) (https://huttenhower.sph.harvard.edu/galaxy/) (accessed on 1 May 2022) was performed to find out the differentially enriched taxa among the groups. The threshold for discriminative features was set to 2.0, and the results were displayed in a cladogram and histogram. The functional prediction of microbiota was performed with PICRUSt2 to obtain MetaCyc pathway abundances between the study groups.

2.5. Statistical Analyses

The relative abundance of bacteria and alpha diversity indices were compared and displayed using GraphPad Prism 8.0.2 (GraphPad Software, Inc., La Jolla, CA, USA). A one-way ANOVA followed by Tukey's multiple comparison test was performed to evaluate the significance of alpha diversity indices. MANOVA/Wilks lambda was used to test for the significance of LDA clustering. The Lda Effect Size (LEfSe) was analyzed using the Kruskal Wallis test. The statistical analysis of predicted pathways obtained after PICRUSt2 in between the groups revealed significant findings using STAMP (version 2.1.3) after testing using Student's t-test followed by Bonferroni correction. In all mentioned tests, a p-value < 0.05 was considered statistically significant.

3. Results

3.1. Characterization of Study Participants

The study cohort was composed of 40 participants belonging to each study group, i.e., healthy controls (C), long-term tobacco chewers (T), and histopathologically confirmed oral squamous cell carcinoma patients (OC). The clinical characteristics of the participants are included in Table S1 (Supplementary Materials). Of these 120 samples, 3 samples from the control, 1 sample from the tobacco group, and 4 samples from the OC group failed the sequencing procedure, and therefore their data is not included in the results below. The 16S rRNA amplicon sequencing data from this study have been deposited in the NCBI BioProject under accession number PRJNA751046.

3.2. OTU Assignment and Taxonomic Analyses of Bacterial Diversit

A total of 6,296,186 sequencing reads, ranging from 5458 to 155,742 per sample, were generated from the V6-V8 hypervariable region of the 16S rRNA gene. After strict quality and size filtering, 5,407,163 reads were retained, with an average of 48,278 reads per sample, and assigned to 6733 OTUs using the Greengenes database (gg_13_5). Rarefaction curves demonstrate that a species richness plateau (up to 500 OTUs) was reached in approximately 5000 readings per sample. To minimize sample variability, approximately 5000 reads were chosen as the minimum sampling depth to estimate diversity. Furthermore, the shape of the species accumulation curve derived from our dataset indicates that the community was well sampled because the specimens we gathered held significant information regarding total species richness.

Overall, these OTUs were assigned to 9 phyla, 17 classes, 30 orders, 55 families, and 94 genera. Among the 5 most abundant phyla, Bacteroidetes dominated in all three groups, followed by Proteobacteria (Figure 1a). The next dominant phyla were Firmicutes, followed by Actinobacteria and Fusobacteria. The abundance of phyla consisting of Gram-negative organisms (Bacteroidetes and Proteobacteria) was higher in OC and tobacco samples than in healthy individuals, whereas that of Gram-positive organisms (Firmicutes and Actinobacteria). The five most abundant genera observed in all groups were Streptococcus, Prevotella, Neisseria, Rothia, and Haemophilus, which constituted up to 50% of total abundance at the genus level in all 3 study groups. The abundance of major genera Streptococcus, Neisseria, Rothia, Veillonella, and Leptotrichia was higher in the control population, followed by the tobacco group, and least in the OC group, whereas that of Prevotella, Haemophilus, Fusobacterium, Capnocytophaga, and Aggregatibacter was higher in OC and decreased in the control population (Figure 1b). When examined closely, the genera Pseudomonas, Morganella, Alloscardovia, Aeromonas, Bacteroides, and Propionibacterium were found only in the OC group (Figure 2).

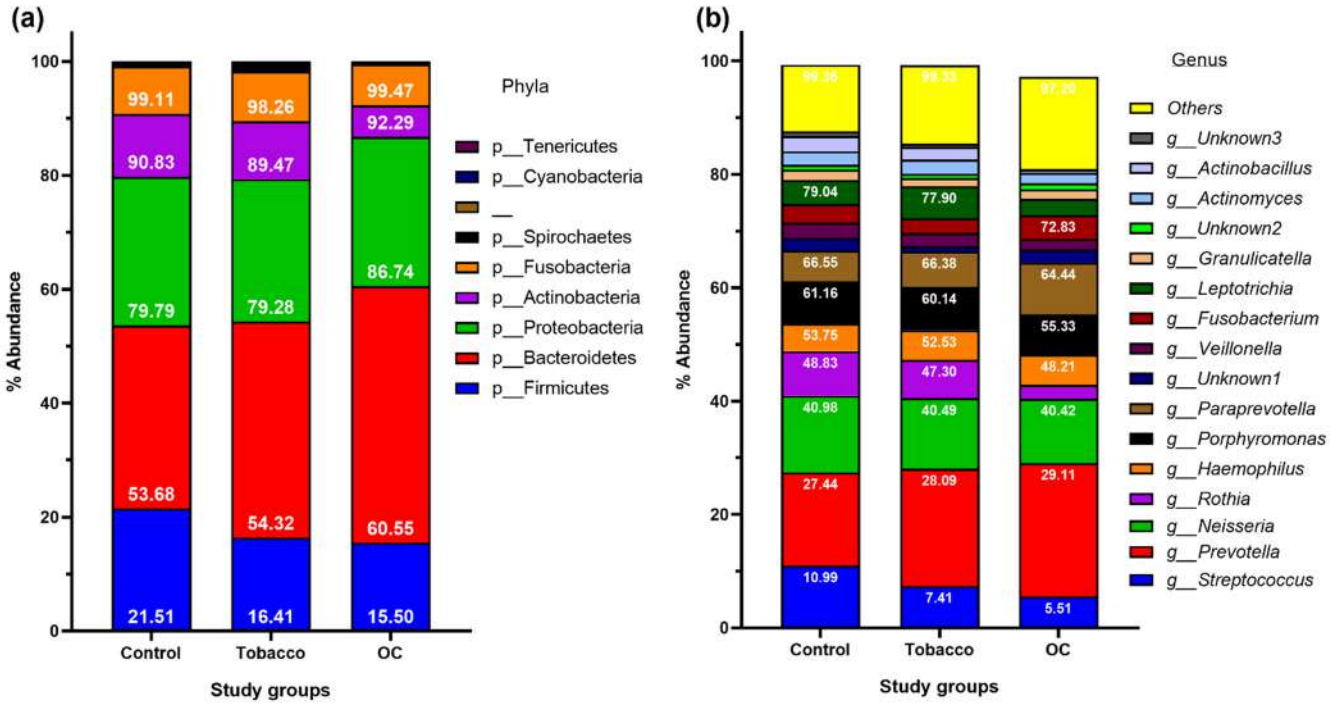
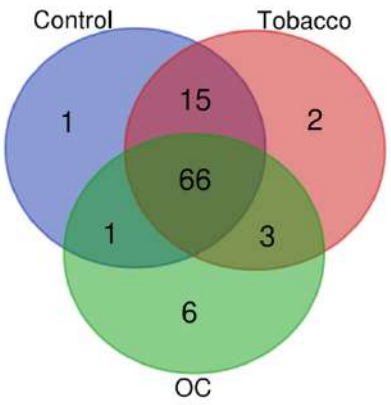


Figure 1. Oral bacterial profiles of healthy, tobacco chewing, and oral cancer patients in India at (**a**) Phyla-level and (**b**) Genus-level.



Study group	Number of genera	Genera
Control-OC-Tobacco	66	*Cardiobacterium, Oribacterium, Campylobacter, Scardovia, Slackia, Bifidobacterium, Veillonella, Granulicatella, Tannerella, Capnocytophaga, Abiotropia, Neisseria, Klebsiella, Atopobium, Actinomyces, Fusobacterium, Actinobacillus, Hameophilus, Porphyromonas, Treponema, Parvimonas, Lautropia, Sharpea, Filifactor, Mogibacterium, Lactobacillus, Kingella, Peptostreptococcus, Shuttleworthia, Prevotella, Selenomonas, Corynebacterium, Lachnoanaerobaculum, Catonella, Paludibacter, Leptotrichia, Megasphaera, Bulledia, Rothia, Paraprevotella, Peptococcus, Aggregatibacter, Streptococcus, Eikenella, Dialister, Moryella, Gemella, 19 unidentified genera*
Control-Tobacco	15	*Schwartzia, Desulfobulbus, Halomonas, 12 unidentified genera*
Control-OC	1	*Bilophilia*
OC-Tobacco	3	*Acinetobacter, Mycoplasma, Desulfovibrio*
Control	1	*Unidentified genus*
Tobacco	2	*Unidentified genus*
OC	6	*Pseudomonas, Morganella, Alloscardovia, Aeromonas, Bacteroides, Propionibacterium*

Figure 2. Venn diagram depicting the common and unique number of bacteria.

3.3. Microbial Biomarkers in Control, Tobacco and OC Individuals

The unique bacterial community composition associated with the oral rinse was investigated using LEfSe analysis to compare the relative abundance of taxa across the C, T, and OC groups (Figure 3a). A total of 27 bacterial genera were observed to be different in the 3 study groups. Leptotrichia, Treponema, Lautropia, Tannerella, Selenomonas, Filifactor, Campylobacter, and Cardiobacterium were identified as potential biomarkers for the tobacco group. On the other hand, Pseudomonas, Capnocytophaga, Mycoplasma, Bifidobacterium, Peptostreptococcus, and Paludibacter were associated as biomarkers for OC. Bacteria belonging to the genera Rothia, Neisseria, Actinobacillus, Veillonella, and Corynebacterium were identified as potential biomarkers for the control population. Furthermore, the cladogram could be used to determine the branch evolution connection, which also depicts the biomarkers identified in the OC group mainly belonging to phyla Bacteroidetes (Figure 3b).

3.4. Diversity of Microbiota Associated with Tobacco Chewing and OC

Alpha diversity matrices were generated using observed OTUs, the Ace index, Chao1, Goods coverage, Shannon and Simpson indices, and Pielou_e to understand the species richness and diversity of the samples (Figure 4). Good's coverage was >96% for sequences in all the study groups, indicating that the sequences measured in each sample represented almost all the bacterial sequences in the sample. A significantly higher number of mean OTUs was observed in tobacco chewers and control populations compared to the OC group. Other alpha diversity indices, such as hose of species richness (ACE/Chao1) and diversity index (Shannon index) also depict statistically higher alpha diversity observed in tobacco chewers and control populations as compared to the OC group, thereby indicating the lowest alpha diversity in the OC group. Beta diversity was studied using various parameters depicted in Figure 5. To advocate for the beta-diversity results obtained to assess community dissimilarity, the Bray-Curtis matrix, the Jaccard matrix, and the Weighted and Unweighted Unifrac matrices were compared (Figure 5). All beta-diversity matrices affirm the bacterial communities in the OC group and the controls-tobacco group clustered discretely, suggesting the overall structures of the bacterial communities in the groups were significantly different.

3.5. Functional Prediction of Bacterial Communities Related Tobacco Chewing and OC

We used the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) method to envisage oral microbial roles linked to the formation of OSCC, and MetaCyc pathways were constructed for the study groups. PICRUSt2 estimates which gene families are present using an extended ancestral-state reconstruction technique, and then joins gene families to provide a comprehensive metagenome of the data. Significantly upregulated pathways related to amino acid biosynthesis (aspartate,

lysine, methionine, threonine, isoleucine, valine), sugar fermentation (glycolysis, Entner-Doudoroff, pyruvate), and pyrimidine salvage and biosynthesis were detected in healthy controls as compared to the OC group (Figure 6). Conversely, pathways related to Co-enzyme A ($p = 0.024$), aspartate, asparagine ($p = 0.023$), lipid biosynthesis ($p = 0.042$), and fatty acid elongation ($p = 0.038$) were upregulated in the OC group as compared to controls. The tobacco group revealed upregulated pathways related to the reductive TCA cycle ($p = 0.010$) and pyrimidine biosynthesis ($p = 7.95 \times 10^{-3}$) as compared to the OC group.

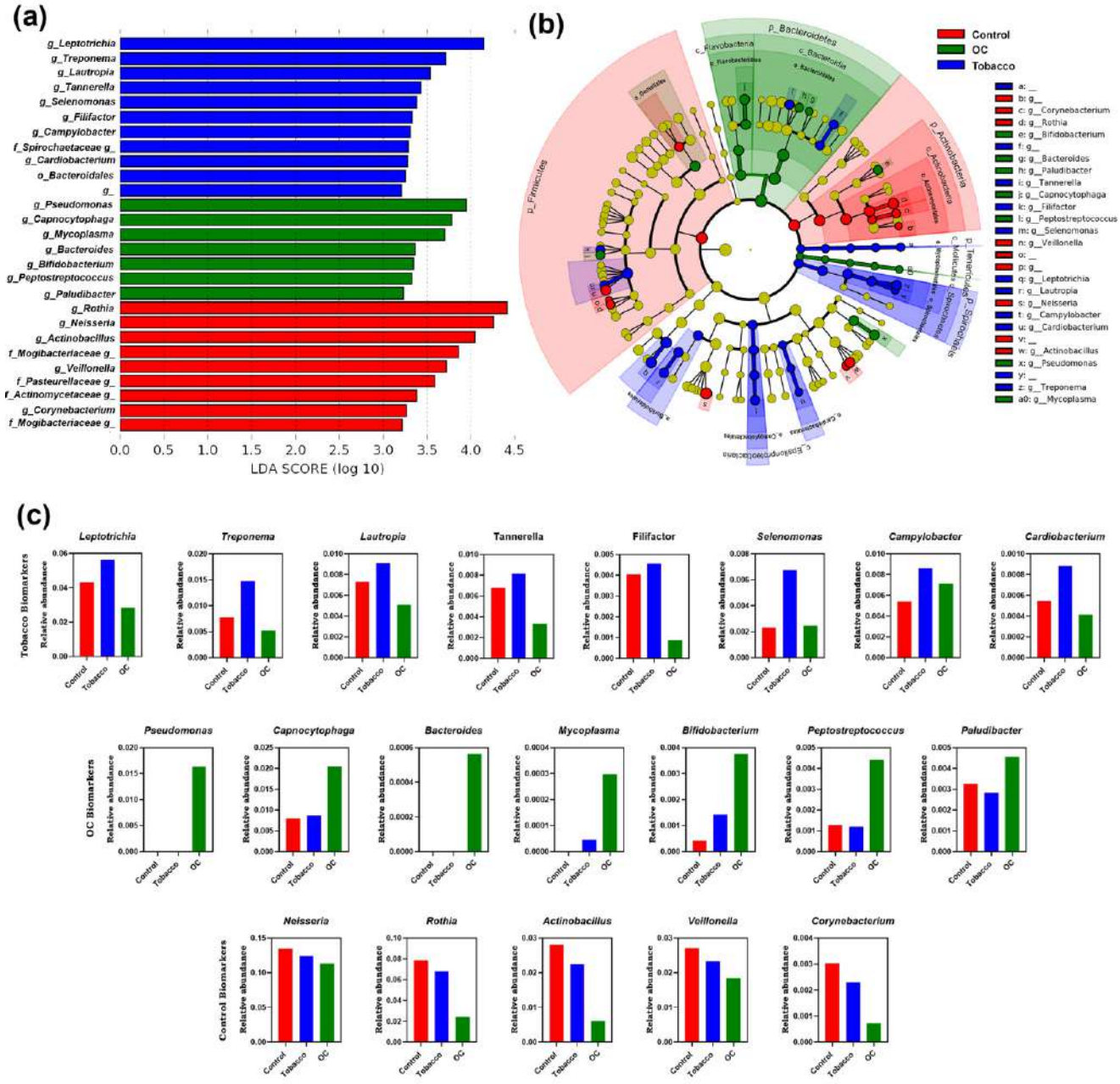


Figure 3. Distinct taxa were identified in the C, T, and OC groups using LEfSe analysis. (**a**) LDA scores showed significant bacterial differences within groups at the genus level; (**b**) a Cladogram was constructed using the LEfSe method to indicate the phylogenetic distribution of bacteria that were remarkably enriched in the control, tobacco, and OC groups; (**c**) and the mean relative abundance of biomarker taxon across all study groups.

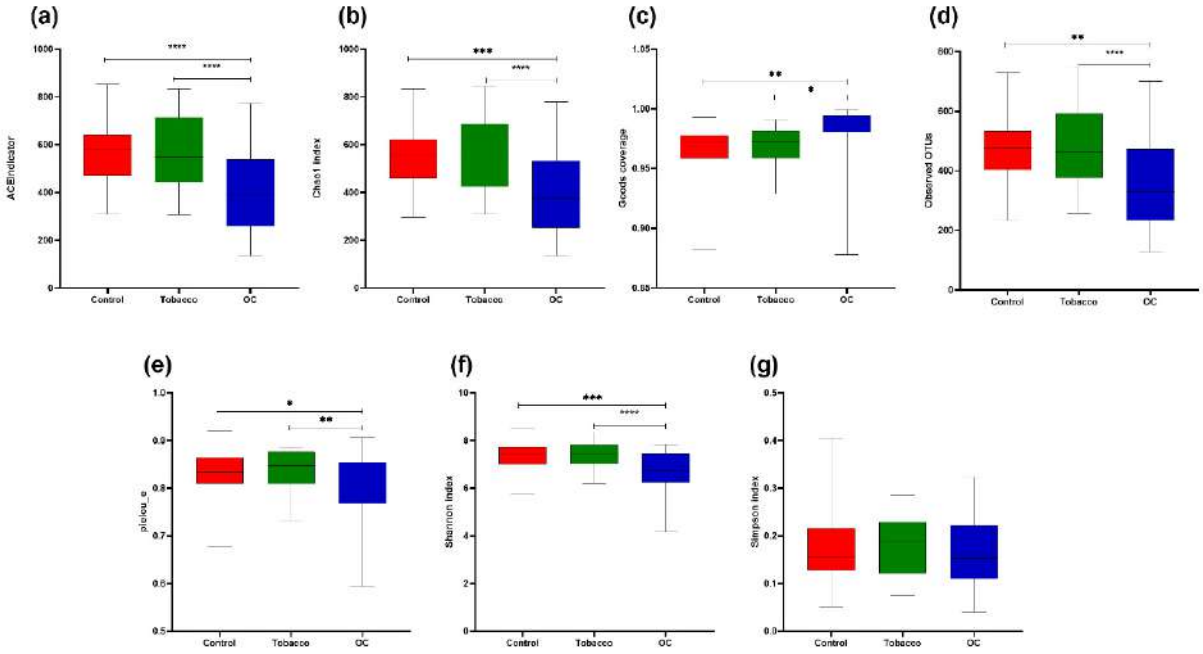


Figure 4. Alpha diversity indices for the study groups; (**a**) ACE indicator; (**b**) Chao1 index; (**c**) Goods coverage; (**d**) Observed OUT's; (**e**) pielou_e; (**f**) Shannon index; (**g**) Simpson index. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **** $p < 0.0001$).

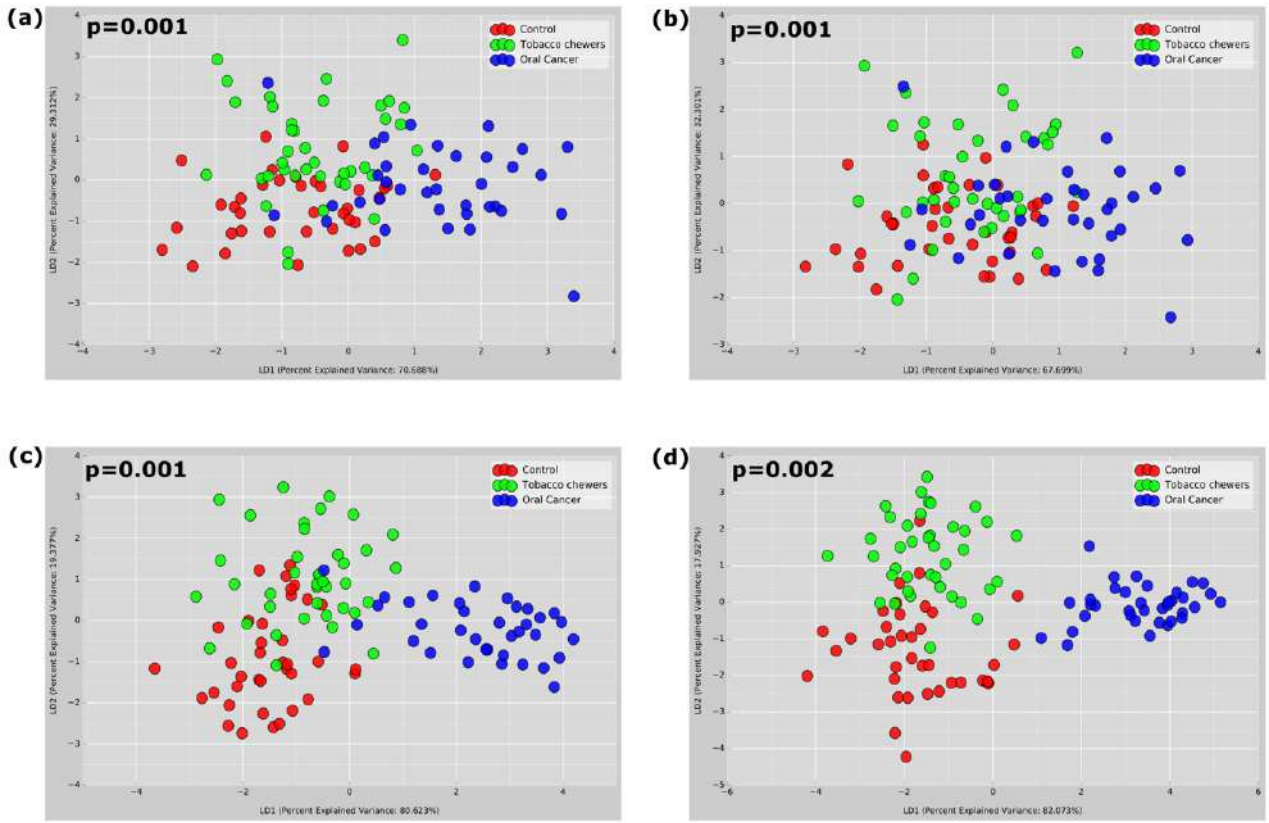


Figure 5. Beta diversity LDA plots depicting sample diversity between groups; (**a**) Bray- Curtis plot; (**b**) Jaccard matrix; (**c**) Unweighted Unifrac matrix; (**d**) Weighted Unifrac matrix.

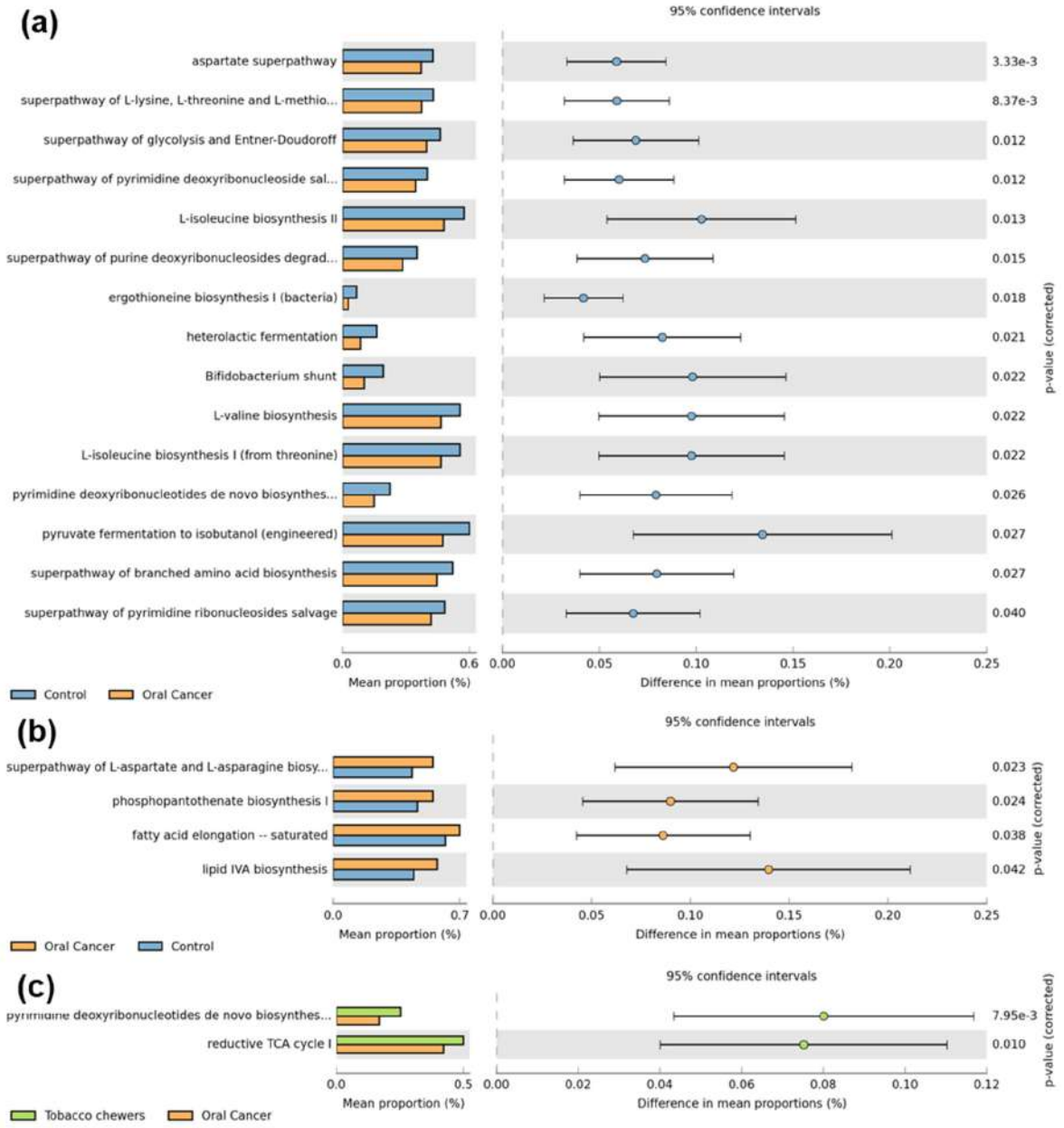


Figure 6. Prediction of microbial gene functions among distinct mutational signature clusters of OC. (**a**) depicts significant pathways upregulated in healthy control samples as compared to oral cancer samples, which includes amino acid biosynthesis (aspartate, lysine, methionine, threonine, isoleucine and valine), sugar fermentation (glycolysis, Entner Doudoroff, pyruvate) and pyrimidine salvage and biosynthesis pathways. (**b**) depicts pathways significantly upregulated in oral cancer samples as compared to healthy control samples, which include co-enzyme A, aspartate, asparagine and lipid biosynthesis pathways. (**c**) depicts pathways upregulated in the tobacco group as compared to the oral cancer group which include reductive TCA and pyrimidine biosynthesis pathways.

4. Discussion

Numerous oral microbiome-based research studies have been conducted throughout the world to better understand bacterial dynamics in the context of diverse external factors and diseases, mainly cancer. However, the population of the Indian sub-continent is highly diverse in terms of ethnicity, culture, lifestyle, geographic location, and food. The Indian population is exposed to a wide variety of lifestyle factors, including tobacco chewing, smoking, and alcohol consumption, and ranks first in the incidence of males suffering from

OC across the globe [16]. Thereby, the Indian population acts as a suitable demography to study the OC microbiome due to the high incidence rate as well as exposure to the risk factors. Handful studies regarding the Indian oral microbiome have been published, but this is the first report comparing the oral microbiome of healthy controls and tobacco chewers with OC patients.

In this study, we have analyzed a larger population group for accurate information regarding the study population compared to an earlier published report [13]. The present study reveals five phyla and 23 genera contributing to approximately 90% of the total oral microbiome composition. The abundance of Bacteroidetes and Proteobacteria observed was highest in the OC group as compared to the other two study groups. These phyla are composed mainly of Gram-negative bacteria, whereas Proteobacteria includes mainly Gram negative pathogenic bacteria [17]. The increased presence of Gram-negative bacteria in the oral microbiota of OC patients has been previously reported [18]. Apart from the five major phyla, the abundance of phyla Spirochaetes was highest in the tobacco group, and the presence of phyla Tenericutes was observed in tobacco and OC groups only, which could be attributed to the presence of periodontal pathogens in tobacco chewers and diseased conditions in OC [19]. The abundance of genera Streptococcus, Rothia, Veillonella, and Neisseria was found to decrease in individuals suffering from OC compared to those in healthy controls, owing to the mentioned genera being part of the healthy oral microbiota in humans. Therefore, their high abundance in controls is due to a healthy oral cavity, whereas their decreased abundance in OC and tobacco chewers could be due to dysbiosis in the oral cavity of the said individuals. On the other hand, genera such as Prevotella, Haemophilus, and Fusobacterium are known pathogens of the oral cavity, thereby justifying their increased counts in the OC and tobacco chewing [20]. The increased abundance of Prevotella and Fusobacterium in tobacco chewers leads to the synergistic activity of toxins from the bacteria and nicotine, thereby leading to detrimental health effects [10]. Fusobacterium spp. Has been linked to cell adhesion, tumorigenesis, epithelial-to-mesenchymal transition, inflammasomes, the cell cycle, and other aspects of oral cancer [21,22].

Along with being residents of the human body, some microorganisms can also cause host damage. Any kind of damage can cause inflammation, which is a defense mechanism to eliminate harmful metabolites and damaged tissues and is followed by the initiation of wound healing [23]. The use of smokeless tobacco is another source of tissue damage that can disrupt the wound-healing process. Recent studies have provided a hypothesis that human immunity has emerged as an entity that can control the damage exerted to host tissues by the inflammation process and also manage the microbes present inside and around the human body for nutrition [24]. Tobacco chewing, along with microbial dysbiosis, can lead to chronic inflammation that can initiate and progress toward the development of oral cancer. Because the role of microbiota and lifestyle habits such as tobacco chewing are linked to inflammation, identifying microbial biomarkers can help in the recognition of inflammation markers and related molecular pathways. Genus Leptotrichia, a biomarker for tobacco chewers, has been previously linked to tobacco chewing habits [25]. Leptotrichia and Campylobacter have been linked to the core OC microbiota [26]. Similarly, genera Treponema and Tannerella are well-known periodontal pathogens that play a crucial role in the formation of a red complex periodontitis [27]. Therefore, the increased abundance of these bacteria in the tobacco-chewing population can be attributed to a higher risk of periodontitis development in these individuals. The presence of other tobacco biomarkers, such as Lautropia, Filifactor, and Selenomonas, has been linked to the occurrence of OC in different populations [9]. Filifactor bacteria have been shown to secrete proinflammatory cytokines, activate specific oncogenes, and maintain an inflammatory state [28]. Although the genus Cardiobacterium has been associated with endocarditis and oral mucositis [29], we report a significant increase in the genus Cardiobacterium in tobacco chewers for the first time.

On assessing the OC microbiome, Pseudomonas and Bacteroides were found solely in the OC patients, previously reported as a part of OC microbiota [30] can be used for early clinical diagnosis by using simple, specific, non-invasive methods for identification purposes [12]. Pseudomonas can convert salivary nitrite to nitric oxide (NO), which modulates various cancer-related appearances such as apoptosis, cell cycle, angiogenesis, invasion, and metastasis [31]. Similarly, concurring with the previous study, Capnocytophaga and Peptostreptococcus were enriched in OC patients, whereas the abundance of Bifidobacterium is upregulated in our study in the OC group as opposed to previous reports [32]. Similar to increased abundances of tobacco and OC biomarkers, decreased populations of healthy control biomarkers can also be used to diagnose dysbiosis, thereby predisposing individuals to diseased conditions.

Apart from the biomarkers identified, a few genera, such as Acinetobacter, Mycoplasma, and Desulfovibrio, have been found only in the tobacco and OC populations. The proportions of Mycoplasma and Desulfovibrio were observed to be higher in oral cancer patients as compared to tobacco chewers, respectively, and are well reported [33]. Since Mycoplasma is already identified as a biomarker, and Desulfovibrio also shows similar patterns of existence, these can be an important choice of bacteria to monitor the initiation and progression of oral cancer in tobacco chewers. Similarly, bacteria belonging to the genus Morganella, Alloscardovia, Aeromonas, and Propionibacterium, along with Pseudomonas and Bacteroides have been identified only in the OC population. For the first time in our study, Aeromonas, Alloscardovia, and Morganella have been identified as part of the OC oral microbiota and therefore need to be studied in more detail. Furthermore, the predicted functions enriched in the OC samples depict increased lipid and fatty acid synthesis. These molecules have inflammatory functions and have been reported to initiate and aggravate oral cancer [34].

When the bacterial makeup of the three research groups is compared, it is observed that the abundance of major genera in tobacco chewers lies in between that of the control and OC populations. Similarly, beta diversity plots display the clustering of control and tobacco samples together, compared to OC samples that cluster away. Considering all of the parameters, it can be concluded that the composition of tobacco chewers is comparable to both the control and OC populations in several aspects, indicating the transitional phase of the tobacco chewers' oral microbiota. Apart from OC, there are a few reports on the oral microbiota of oral potentially malignant disorders (OPMD). OPMD progresses to oral cancers through a series of histopathological stages, beginning with hyperkeratosis/hyperplasia and progressing to various degrees of dysplasia. Similar to the microbiota of tobacco chewers and the oral cancer identified in this study, the abundance of the phyla Bacteroidetes and Proteobacteria were higher, whereas that of Firmicutes was lower in the OPMD group. At the genus level, Alloprevotella, Leptotrichia, Fusobacterium, Campylobacter, Neisseria, Gemella, and Granulicatella were found in higher abundance in OPMD, similar to tobacco chewers and OC patients as compared to the control group reported in this study [26,35].

Based on the subject demographic, the study may have certain limitations. Male participants are more numerous in oral cancer and tobacco group than female participants. This is partly because males are more likely than females to use tobacco products, and more men than women are diagnosed with mouth cancer. Additionally, because age is a confounding risk factor for malignancies, including oral cancer, the study population includes participants in the OC group who are older than those in the control and tobacco chewing groups, potentially creating an age-related bias. The study may be limited by the inability to control the aforementioned variables; thus, this should be taken into account.

In conclusion, a compositionally distinct microbiota is identified using oral saline rinse in healthy, tobacco-chewing, and OC patients in the Indian population. Oral cancer is frequently thought to be a complicated illness caused by a number of interdependent host–environment interactions. As a result, using a single biomarker to identify oral cancer is exceedingly improbable. The present study used a non-invasive method for sample collection and NGS analysis for the identification of an array of oral microbial biomarkers,

which can be useful for the early diagnosis of OC, especially in individuals susceptible to OC due to lifestyle habits such as tobacco chewing. Since the present study focused on the Indian population, where such information is scarce, this can serve as a reference and basis for future microbiome analysis and oral microbial biomarker studies related to oral cancer. This study provides the first epidemiological evidence for the association of Cardiobacterium in tobacco chewers and Aeromonas, Alloscardovia, and Morganella with OC. In addition to the presented data, it is necessary to investigate the role of differentially abundant taxa and discovered pathways in the development and progression of OC.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pathogens12010078/s1, Table S1: Clinical characteristics of participants.

Author Contributions: Conceptualization, S.S. (Shriya Sawant). and H.S.; methodology, S.S. (Shriya Sawant), J.D., S.S. (Sathiyaraj Srinivasan) and H.S.; software, S.S. (Shriya Sawant) and H.S.; validation, S.S. (Shriya Sawant) and H.S.; formal analysis, S.S. (Shriya Sawant) and H.S.; investigation, S.S. (Shriya Sawant), J.D. and H.S.; resources, S.S. (Shriya Sawant), J.D., H.S. and S.S. (Sathiyaraj Srinivasan); data curation, S.S. (Shriya Sawant) and H.S.; writing—original draft preparation, S.S. (Shriya Sawant) and H.S.; writing—review and editing, S.S. (Shriya Sawant)., J.D., D.P., H.S. and S.S. (Sathiyaraj Srinivasan); visualization, S.S. (Shriya Sawant) and H.S.; supervision, J.D. and H.S.; project administration, H.S.; funding acquisition, H.S. and S.S. (Sathiyaraj Srinivasan). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the research grant from Seoul Women's University (2023).

Institutional Review Board Statement: For the study, ethics approval was obtained from SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and The Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The 16S rRNA amplicon sequencing data from this study have been deposited in the NCBI BioProject under accession number PRJNA751046.

Acknowledgments: The authors would like to thank the working staff of Somaiya Ayurvihar-Asian Cancer Institute (K. J. Somaiya Medical College and Hospital, Mumbai) for providing help with oral cancer sample collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Yang, C.Y.; Yeh, Y.M.; Yu, H.Y.; Chin, C.Y.; Hsu, C.W.; Liu, H.; Huang, P.J.; Hu, S.N.; Liao, C.T.; Chang, K.P.; et al. Oral Microbiota Community Dynamics Associated with Oral Squamous Cell Carcinoma Staging. *Front. Microbiol.* **2018**, *9*, 862. [CrossRef] [PubMed]
3. Tenore, G.; Nuvoli, A.; Mohsen, A.; Cassoni, A.; Battisti, A.; Terenzi, V.; Della Monaca, M.; Raponi, I.; Brauner, E.; De Felice, F.; et al. Tobacco, Alcohol and Family History of Cancer as Risk Factors of Oral Squamous Cell Carcinoma: Case-Control Retrospective Study. *Appl. Sci.* **2020**, *10*, 3896. [CrossRef]
4. Rao, S.V.K.; Mejia, G.; Roberts-Thomson, K.; Logan, R. Epidemiology of Oral Cancer in Asia in the Past Decade—An Update (2000–2012). *Asian Pac. J. Cancer Prev.* **2013**, *14*, 5567–5577. [CrossRef]
5. Chaturvedi, P.; Singh, A.; Chien, C.Y.; Warnakulasuriya, S. Tobacco Related Oral Cancer. *BMJ* **2019**, *365*, l2142. [CrossRef]
6. Imam, S.Z.; Nawaz, H.; Sepah, Y.J.; Pabaney, A.H.; Ilyas, M.; Ghaffar, S. Use of Smokeless Tobacco among Groups of Pakistani Medical Students—A Cross Sectional Study. *BMC Public Health* **2007**, *7*, 231. [CrossRef]
7. Ganly, I.; Yang, L.; Giese, R.A.; Hao, Y.; Nossa, C.W.; Morris, L.G.T.; Rosenthal, M.; Migliacci, J.; Kelly, D.; Tseng, W.; et al. Periodontal Pathogens Are a Risk Factor of Oral Cavity Squamous Cell Carcinoma, Independent of Tobacco and Alcohol and Human Papillomavirus. *Int. J. Cancer* **2019**, *145*, 775–784. [CrossRef] [PubMed]
8. O'Grady, I.; Anderson, A.; O'Sullivan, J. The Interplay of the Oral Microbiome and Alcohol Consumption in Oral Squamous Cell Carcinomas. *Oral Oncol.* **2020**, *110*, 105011. [CrossRef]
9. Zhao, H.; Chu, M.; Huang, Z.; Yang, X.; Ran, S.; Hu, B.; Zhang, C.; Liang, J. Variations in Oral Microbiota Associated with Oral Cancer. *Sci. Rep.* **2017**, *7*, 11773. [CrossRef]

10. Bagaitkar, J.; Demuth, D.R.; Scott, D.A. Tobacco Use Increases Susceptibility to Bacterial Infection. *Tob. Induc. Dis.* **2008**, *4*, 12. [CrossRef]
11. Kumar, P.S.; Matthews, C.R.; Joshi, V.; de Jager, M.; Aspiras, M. Tobacco Smoking Affects Bacterial Acquisition and Colonization in Oral Biofilms. *Infect. Immun.* **2011**, *79*, 4730–4738. [CrossRef]
12. Sawant, S.; Dugad, J.; Parikh, D.; Singh, H. Absolute Quantitation of Oral Bacteria Involved in Oral Cancer by Real-Time PCR. *Med. Microecol.* **2021**, *7*, 100034. [CrossRef]
13. Sawant, S.; Dugad, J.; Parikh, D.; Srinivasan, S.; Singh, H. Identification & Correlation of Bacterial Diversity in Oral Cancer and Long-Term Tobacco Chewers- A Case-Control Pilot Study. *J. Med. Microbiol.* **2021**, *70*, 001417. [CrossRef]
14. Estaki, M.; Jiang, L.; Bokulich, N.A.; McDonald, D.; González, A.; Kosciolek, T.; Martino, C.; Zhu, Q.; Birmingham, A.; Vázquez-Baeza, Y.; et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Curr. Protoc. Bioinform.* **2020**, *70*, e100. [CrossRef]
15. Dabdoub, S.M.; Fellows, M.L.; Paropkari, A.D.; Mason, M.R.; Huja, S.S.; Tsigarida, A.A.; Kumar, P.S. PhyloToAST: Bioinformatics Tools for Species-Level Analysis and Visualization of Complex Microbial Datasets. *Sci. Rep.* **2016**, *6*, 29123. [CrossRef]
16. Poddar, A.; Aranha, R.R.; Muthukaliannan, G.K.; Nachimuthu, R.; Jayaraj, R. Head and Neck Cancer Risk Factors in India: Protocol for Systematic Review and Meta-Analysis. *BMJ Open.* **2018**, *8*, e020014. [CrossRef]
17. Rizzatti, G.; Lopetuso, L.R.; Gibiino, G.; Binda, C.; Gasbarrini, A. Proteobacteria: A Common Factor in Human Diseases. *BioMed. Res. Int.* **2017**, *2017*, 9351507. [CrossRef] [PubMed]
18. Bolz, J.; Dosá, E.; Schubert, J.; Eckert, A.W. Bacterial Colonization of Microbial Biofilms in Oral Squamous Cell Carcinoma. *Clin. Oral Investig.* **2014**, *18*, 409–414. [CrossRef]
19. Mukherjee, P.K.; Wang, H.; Retuerto, M.; Zhang, H.; Burkey, B.; Ghannoum, M.A.; Eng, C. Bacteriome and Mycobiome Associations in Oral Tongue Cancer. *Oncotarget* **2017**, *8*, 97273. [CrossRef] [PubMed]
20. Lim, Y.; Fukuma, N.; Totsika, M.; Kenny, L.; Morrison, M.; Punyadeera, C. The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 267. [CrossRef]
21. Pignatelli, P.; Romei, F.M.; Bondi, D.; Giuliani, M.; Piattelli, A.; Curia, M.C. Microbiota and Oral Cancer as A Complex and Dynamic Microenvironment: A Narrative Review from Etiology to Prognosis. *Int. J. Mol. Sci.* **2022**, *23*, 8323. [CrossRef]
22. Fujiwara, N.; Kitamura, N.; Yoshida, K.; Yamamoto, T.; Ozaki, K.; Kudo, Y. Involvement of Fusobacterium Species in Oral Cancer Progression: A Literature Review Including Other Types of Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 6207. [CrossRef]
23. Costantini, S.; Sharma, A.; Colonna, G. The Value of the Cytokinome Profile. In *Inflammatory Diseases—A Modern Perspective*; Nagal, A., Ed.; BoD—Books on Demand: Norderstedt, Germany, 2011.
24. Broderick, N.A. A Common Origin for Immunity and Digestion. *Front. Immunol.* **2015**, *6*, 72. [CrossRef]
25. Halboub, E.; Al-Ak'hali, M.S.; Alamir, A.H.; Homeida, H.E.; Baraniya, D.; Chen, T.; Al-Hebshi, N.N. Tongue Microbiome of Smokeless Tobacco Users. *BMC Microbiol.* **2020**, *20*, 1–9. [CrossRef]
26. Amer, A.; Galvin, S.; Healy, C.M.; Moran, G.P. The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species. *Front. Microbiol.* **2017**, *8*, 2391. [CrossRef]
27. Hajishengallis, G.; Darveau, R.P.; Curtis, M.A. The Keystone-Pathogen Hypothesis. *Nat. Rev. Microbiol.* **2012**, *10*, 717–725. [CrossRef]
28. Vyhnalova, T.; Danek, Z.; Gachova, D.; Linhartova, P.B. The Role of the Oral Microbiota in the Etiopathogenesis of Oral Squamous Cell Carcinoma. *Microorganisms* **2021**, *9*, 1549. [CrossRef] [PubMed]
29. Reyes-Gibby, C.C.; Wang, J.; Zhang, L.; Peterson, C.B.; Do, K.A.; Jenq, R.R.; Shelburne, S.; Shah, D.P.; Chambers, M.S.; Hanna, E.Y.; et al. Oral Microbiome and Onset of Oral Mucositis in Patients with Squamous Cell Carcinoma of the Head and Neck. *Cancer* **2020**, *126*, 5124–5136. [CrossRef] [PubMed]
30. Perera, M.; Al-hebshi, N.N.; Perera, I.; Ipe, D.; Ulett, G.C.; Speicher, D.J.; Chen, T.; Johnson, N.W. Inflammatory Bacteriome and Oral Squamous Cell Carcinoma. *J. Dent. Res.* **2018**, *97*, 725–732. [CrossRef] [PubMed]
31. Kakabadze, M.Z.; Paresishvili, T.; Karalashvili, L.; Chakhunashvili, D.; Kakabadze, Z. Oral Microbiota and Oral Cancer: Review. *Oncol. Rev.* **2020**, *14*, 476. [CrossRef]
32. Sarkar, P.; Malik, S.; Laha, S.; Das, S.; Bunk, S.; Ray, J.G.; Chatterjee, R.; Saha, A. Dysbiosis of Oral Microbiota During Oral Squamous Cell Carcinoma Development. *Front. Oncol.* **2021**, *11*, 614448. [CrossRef] [PubMed]
33. Chattopadhyay, I.; Verma, M.; Panda, M. Role of Oral Microbiome Signatures in Diagnosis and Prognosis of Oral Cancer. *Technol. Cancer Res. Treat.* **2019**, *18*, 1533033819867354. [CrossRef] [PubMed]
34. Wu, J.S.; Zheng, M.; Zhang, M.; Pang, X.; Li, L.; Wang, S.S.; Yang, X.; Wu, J.B.; Tang, Y.J.; Tang, Y.L.; et al. Porphyromonas Gingivalis Promotes 4-Nitroquinoline-1-Oxide-Induced Oral Carcinogenesis with an Alteration of Fatty Acid Metabolism. *Front. Microbiol.* **2018**, *9*, 2081. [CrossRef] [PubMed]
35. Mok, S.F.; Karunthan, C.; Cheah, Y.K.; Ngeow, W.G.; Rosnah, B.Z.; Yap, S.F.; Alan, O.H.K. The Oral Microbiome Community Variations Associated with Normal, Potentially Malignant Disorders and Malignant Lesions of the Oral Cavity. *Malays. J. Pathol.* **2017**, *39*, 1–15.

JOURNAL OF MEDICAL MICROBIOLOGY

RESEARCH ARTICLE

Sawant *et al.*, *Journal of Medical Microbiology* 2021;70:001417

DOI 10.1099/jmm.0.001417

MICROBIOLOGY SOCIETY

Identification & correlation of bacterial diversity in oral cancer and long-term tobacco chewers- A case-control pilot study

Shriya Sawant[1], Jinesh Dugad[2], Deepak Parikh[2], Sathiyaraj Srinivasan[3,4,*] and Harinder Singh[1,*]

Abstract

Introduction. Squamous cell carcinoma is a highly aggressive type of oral cancer (OC). It is the most common cancer among men, and accounts for almost 90% of all oral cancers in India. Consumption of tobacco is a leading factor contributing to maximum oral cancer incidences as per the WHO.

Hypothesis/Gap statement. Researchers reported a direct association of microorganisms with dysbiosis in various oral lesions including oral cancer. However, there is a dearth of information related to compositional changes in the oral microbiome in long-term tobacco chewers and the Indian oral cancer population.

Aim. The aim of this study was to identify and correlate the bacterial diversity in the oral cavity of tobacco chewers, patients with oral cancer and healthy subjects in the Indian population.

Methods. Oral rinse samples were collected for ten subjects in each group followed by DNA extraction. The variable regions of the bacterial 16S rRNA gene (V6-V8) were amplified, sequenced, processed, and analysed using QIIME2 platform to assess alpha and beta diversity between the study groups.

Results. This pilot study showed genus *Streptococcus* dominated the control group (18.54%), and the abundance decreased in tobacco and OC group (9.63 and 5.45% respectively); whereas genus *Prevotella* dominated the tobacco and OC group (21.01 and 26.03% respectively). A shift in abundance of microbiome was observed from control population to oral cancer via the tobacco chewing population. Maximum alpha diversity of oral microbiome was found in Indian tobacco chewers. Beta diversity of tobacco chewers was similar to both the healthy population as well as oral cancer patients suggesting transitioning of the oral microbiome from healthy to oral cancer microbiome via the tobacco chewers microbiome.

Conclusion. The data provides evidence of oral bacterial dysbiosis due to tobacco chewing habits that can further lead to progression towards cancer.

INTRODUCTION

The human body is a complex ecosystem of human cells and microorganisms. A huge diversity of microbial populations are found in the human body and called the microbiota. This microbiota provides healthy associations and supports good health [1]. The oral cavity, with its various hard and soft anatomic surfaces, makes the oral environment conducive for the growth of the different types of microbes and harbours more than 700 species of bacteria [2]. But, continuous exposure to external agents, a sedentary lifestyle and dietary habits can affect the human microbiota in terms of its diverse profile, causing dysbiosis which refers to the alterations in the bacterial composition [3, 4]. Dysbiosis in the oral cavity has been linked to various diseases, including cancers [5–8].

The risk factors for oral cancer include consumption of tobacco and tobacco products, smoking, alcohol consumption, poor oral hygiene and viral infections like HPV (Human papilloma virus) [9, 10]. In South Asia, 250 million people use

Received 15 February 2021; Accepted 28 July 2021; Published 23 September 2021

Author affiliations: [1]Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS University, Vile Parle (West), Mumbai, India; [2]Somaiya Ayurvihar -Asian Cancer Institute, Off Eastern Express Highway, Behind Everard Nagar, Somaiya Ayurvihar, Sion (East), Mumbai, India; [3]Department of Bio & Environmental Technology, College of Natural Science, Seoul Women's University, Seoul 01797, Republic of Korea; [4]Gene Strand Technologies Pvt. Ltd., Chennai, India.

***Correspondence:** Sathiyaraj Srinivasan, drsrini@swu.ac.kr; Harinder Singh, Harinder.Singh@nmims.edu

Keywords: Dysbiosis; Microbiome; Oral cancer; Tobacco.

Abbreviations: OC, oral cancer; OTU, operational taxonomic unit; QIIME, quantitative insights into microbial ecology.

One supplementary table is available with the online version of this article.

smokeless tobacco on a regular basis and tobacco chewing has been linked to increased risk of oral cancer by a factor of six, thereby making it the most common risk factor for oral cancer in Asia [11, 12]. Apart from these, 15% of oral cancer cases cannot be linked to the above-mentioned risk factors, which has increased the need to detect more risk factors [13]. There is an increasing evidence of the role of bacteria in the metabolism of carcinogens, such as tobacco-specific-N-nitrosamines which emerge from tobacco products [14].

Oral cancer continues to be one of the leading cancers with an ever-increasing incidence as well as mortality rate. According to the Globocan 2018 data, India contributes to about 33.81% of global cases and is the second most common cancer in the country. Oral Squamous Cell Carcinoma (OSCC) makes up for 90% of oral cancer cases that are the most commonly found in Indian males (https://gco.iarc.fr/today/home). With the steep rise in the oral cancer cases in the Indian subcontinent, and the associated involvement of tobacco usage, it is necessary to evaluate the role of the oral microbiome in these cases [9, 15, 16].

Along with these factors, the role of the oral microbiome in oral cancer has been unveiled by the Human Microbiome Project (HMP) [17, 18]. With the help of next-generation sequencing (NGS) technology, the identification and quantification of the microbiome including bacteria, viruses, and fungi in healthy or diseased individuals, have become effortless [19, 20]. Indeed, numerous metagenomics studies of the microbiome have highlighted microbial pattern modifications in various types of cancer and viral infections. Reports have also been published depicting diversity in the oral microbial flora depending on the stage of oral cancer [21], the malignant lesions and disorders of the oral cavity [22], and mutational changes in oral cancer [23]. However, there is a dearth of genomic study on the bacterial community composition in the oral cavity and its comprehensive profile in cancer in the Indian population.

The purpose of this study is to analyse the typical bacterial composition and diversity of the oral microbiome in patients with oral cancer and long-term tobacco chewers in the Indian population, using the next generation sequencing (NGS) technology. We hypothesized that the oral microbiota community could dynamically change in OSCC patients. To the best of our knowledge, no previous reports have been published describing the oral microbiome associated with oral cancer patients and the tobacco chewers in India.

METHODS

Subject recruitment and data collection

Thirty subjects in the study were classified into three groups - Healthy (control, *n*=10), OSCC (OC, *n*=10) and long-term tobacco chewers (tobacco, *n*=10), each consisting of ten individuals. All the participants enrolled in the study were 18 years or older. The healthy controls were defined as non-smokers and without any diagnosed diseases in the oral cavity, which was confirmed by prior clinical examination. Long-term tobacco chewers were individuals chewing various forms of tobacco for 5 years or more and without any oral or periodontal disease. All diagnoses for patients with OSCC were confirmed by biopsy and pathological findings. All control and long-term tobacco chewing participants were devoid of any antibiotic treatment at the time of sample collection. The samples were collected in the period of January 2018 to March 2020.

Data on participant's medical history, age, gender, occupation, habits of tobacco chewing, alcohol intake, and general oral hygiene questions were recorded. Informed written consent was obtained prior to the sample collection from all the participants. Ethical approvals for the recruitment of healthy, long-term tobacco chewing volunteers and oral cancer patients were granted by SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and The Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai.

Sample collection

Patients rinsed their mouths with 20 ml of sterile normal saline for 30 s and spit into a 50 ml sterile tube. Salivary samples from patients were taken at the pre-operative stage while the patient was NBM (Nill By Mouth). Subjects from healthy and tobacco groups were asked to refrain from eating, drinking, or oral hygiene procedures for at least 1 h before sample collection. Salivary samples were collected in well-labelled sterile falcon tubes and stored at 4 °C and processed within 48 h.

DNA Extraction and quantitation

The collected samples were centrifuged at 14000 r.p.m. for 8 min at room temperature. The supernatant was discarded and the pellet was used for DNA isolation by using Invitrogen PureLink Genomic DNA kit (Cat no. K182002) according to the manufacturer's recommendation. DNA was eluted using 100 µl elution buffer and was stored at −20 °C until PCR amplification. DNA sample integrity was assessed by electrophoresis on 1% agarose gel. The purity and yield of DNA were assessed spectrophotometrically using BioTek EPOCH2 microplate reader. The purity of DNA was determined by the ratio of absorbance at 260 nm and 280 nm (A260/A280).

Next generation sequencing

The PCR amplification of bacterial 16S rRNA hypervariable region V6-V8 was carried out using primers B969F (ACG CGH NRA ACC TTA CC) and BA1406R (ACG GGC RGT GWG TRC AA) [24, 25]. The V6-V8 region has been accepted as a low error-prone region for taxonomic assignment and community clustering [26]. The PCR reaction was carried out by 30 s initial denaturation at 98 °C, 30 cycles of 10 s denaturation at 98 °C, 30 s annealing at 55 °C, 30 s elongation at 72 °C, and a 5 min final extension at 72 °C. The whole sequencing process was done using Illumina (Illumina, San Diego, CA). The library was prepared by standard library construction protocol (https://support.illumina.com/downloads/

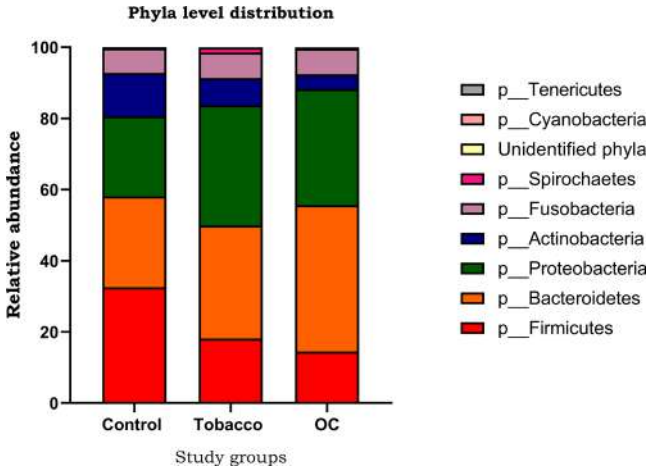


Fig. 1. Phyla-level distribution of bacteria within groups.

16s_metagenomic_sequencing_library_preparation.html) using Nextera XT kit (Illumina, San Diego, CA), following the manufacturer's instructions. The specific amplicons for V6-V8 region were quantified in each reaction mixture along with Illumina sequence adapter, and index primers (Nextera XT Index kit) were utilized in emulsion PCR to generate amplicon libraries followed by PCR clean up. MiSeq libraries were quantified and then subjected to 300-nucleotide paired-end multiplex sequencing on Illumina MiSeq.

Bioinformatic analysis

Demultiplexed raw reads (V6-V8) from Illumina sequencer were obtained and imported into Quantitative Insights into Microbial Ecology 2 (QIIME2) pipeline for further analysis. The quality of the reads from the sequencer were assessed using FASTQC. The reads for each sample were paired using VSEARCH and the quality was assessed again using FastQC. Low quality reads (Phred score Q<30) were filtered out followed by denoising and assignment into Operational Taxonomic Units (OTUs) by Deblur. Taxonomy was assigned using feature-classifier-sklearn against the Greengenes database (gg_13_5, https://greengenes.secondgenome.com/?prefix=downloads/greengenes_database/gg_13_5/) at different taxonomic levels. Bacterial richness and diversity were analysed by alpha and beta diversity matrices and indices such as Observed OTUs, Shannon index, Chao index, ACE index, Rarefaction curves, Weighted and Unweighted Unifrac distance matrices, and PCoA plots.

Observed OTUs exist based on sequence data, i.e. similar sequences are clustered into a single OTU. Based on OTUs, Shannon and Simpson diversity indices are measured to estimate bacterial diversity in a sample and provide more information about the bacterial composition as compared to simple species richness or evenness. On the other hand, Rarefaction, ACE and Chao1 indices estimate richness of a sample [27]. Pielou_e indicates an evenness measure which describes how evenly bacterial communities are distributed in the samples and Good's coverage index depicts sufficient sampling for the data [28, 29].

Statistical analysis

The differential abundance of bacterial genera was analysed using LEfSe (https://huttenhower.sph.harvard.edu/galaxy/). The Lda Effect Size was analysed using Kruskal Wallis test (threshold=0.05) and Pairwise Wilcoxon test (threshold=0.05). To analyse the beta diversity, PERMANOVA, ANOSIM and PERMDISP was performed using QIIME2 pipeline. Graphpad Prism 8 was used to carry out statistical analysis of the results obtained.

RESULTS

Clinical data

Ten volunteers from each group (healthy volunteers, long-term tobacco chewers, and patients with oral cancer) were recruited in the study. Of these ten, three samples, one from each of the study groups failed the sequencing procedure and therefore their data is not included in the results below. The demographic data of the volunteers is tabulated in Table S1 (available in the online version of this article).

Microbial richness and diversity

In the study V6-V8 hypervariable regions of 16S rRNA were sequenced using the Illumina MiSeq system. A total of 1444088 raw reads of 300 bp length were generated. After joining the read pairs and filtering out low-quality sequences (Phred score Q<30), 1231511 high-quality sequences were retained. After joining the forward and reverse reads, sequences of 430 bp length were obtained which were then trimmed to 300 bp length due to drop in quality of the sequences beyond 300 bp with an average number of reads 45611 reads per sample which were classified into 1719 OTUs using Greengenes database (version gg_13_5).

After assigning taxonomy, nine phyla, 15 classes, 22 orders, 44 families, 66 genera, and 101 species were identified. The top five phyla observed were Firmicutes, Proteobacteria, Bacteroidetes, Fusobacteria, and Actinobacteria. Maximum abundance of phyla Firmicutes was observed in the control group (32.56%), phyla Proteobacteria in tobacco (33.86%), and phyla Bacteroidetes in the OC group (41.13%) (Fig. 1). The ten most abundant genera observed in all groups were *Streptococcus, Prevotella, Rothia, Neisseria, Haemophilus, Porphyromonas, Veillonela, Leptotrichia, Fusobacterium,* and *Granulicatella*. Of these, genus *Streptococcus* was most abundantly present in control samples (18.544%), whereas the abundance was decreased in the tobacco chewing group and OC group to 9.634 and 5.455% respectively (Fig. 2). Genus *Prevotella* dominated the tobacco (21.009%) and OC group (26.034%). Genus *Porphyromonas,* which is a known major oral periodontal pathogen was found in increased abundance in OC patients (11.621%) as opposed to tobacco chewers (7.696%) and control population (5.610%). The relative abundance of the ten most abundant genera is shown in Fig. 3. The other oral pathogens such as those belonging to genus

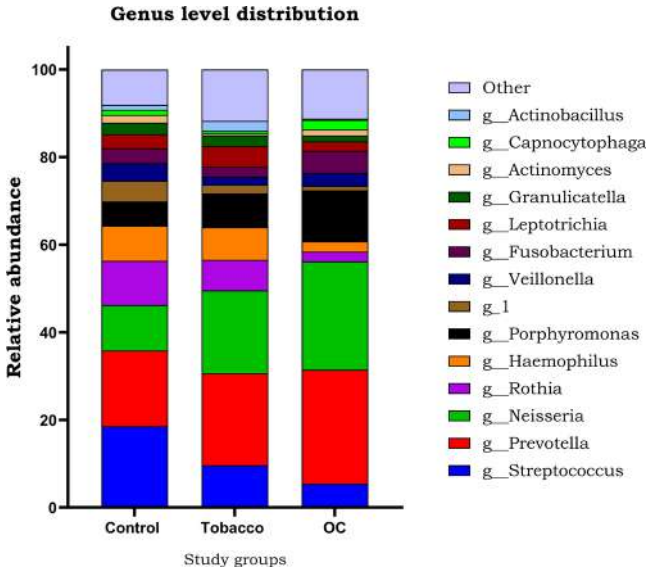


Fig. 2. Genus-level distribution of bacteria within groups. 'Other' contains the following genera grouped together: *Aggregatibacter, Paludibacter, Lautropia, Camplylobacter, Tannerella, Lactobacillus, Gemella, Treponema, Dialsiter, Atopobium, Oribacterium, Parvimonas, Corynebacterium, Filifactor, Peptostreptococcus, Abiotrophia, Megasphaera, Bulleidia, Selenomonas, Halomonas, Eikenella, Scardovia, Bifidobacterium, Moryella, Shuttleworthia, Catonella, Mycoplasma, Lachnoanaerobaculum, Pseudomonas, Propionibacterium, Bacteroidetes, Morganella, Alloscardovia, Peptococcus, Klebsiella,* g_2, g_3, g_4, g_5, g_6, g_7, g_8, g_9, g_10, g_11, g_12, g_13, g_14, g_15,g_16, g_17, g_18,

Aggregatibacter, Prevotella, Parvimonas, Campylobacter, and *Peptostreptococcus* were also found in increased abundance in OC patients or tobacco chewers as compared to control population.

On comparing the diversity at the genus level, few genera were found to be explicitly present in one or two of the three groups (Fig. 4). Genera *Abiotrophia* along with four unidentified genera were present in control and tobacco group; *Shuttleworthia* and *Bifidobacterium* were present in samples of control and OC group, whereas genera *Catonella, Lachnoanaerobaculum* and *Mycoplasma* was exclusively present in OC and tobacco group. Genera *Pseudomonas*, *Peptococcus, Morganella, Alloscardovia, Klebsiella, Bacteroides,* and *Propionibacterium* were present in samples from oral cancer patients only. Since species level taxonomy was not well attained, the data has not been included.

Alpha and beta diversity comparisons

To understand the species richness and diversity of samples, alpha (within-sample) diversity matrices were calculated using observed OTUs, Ace index, Chao1, Goods coverage, Shannon and Simpson indices, and Pielou_e. The Good's coverage for all the groups was greater than 98%. Rarefaction curves displayed plateau for species richness (up to 350 OTUs) was reached at approximately 6000 reads per sample. The alpha diversity was slightly higher in the tobacco chewing group as compared to the control group, although this result was non-significant. In contrast, alpha diversity is comparatively lower in the oral cancer group as compared to both the control group and the tobacco chewing group. As compared to the tobacco group, the OC group had lower alpha diversity based on Observed OTUs (mean number of OTUs 277 and 180 respectively, P=0.0071), ACE index (P=0.0022) as well as Shannon index (P=0.0326) (Fig. 5) (Table 1).

Beta diversity (between-group) was studied using various parameters as mentioned in Table 2 and shows significant differences between the three groups. Unweighted Unifrac distances take into account the qualitative differences/diversity between the groups of study whereas Weighted Unifrac distances account for the diversity between groups quantitatively. Between control and OC groups, unweighted unifrac distances significantly differed based on PERMANOVA (P=0.017), ANOSIM (P=0.011), and PERMDISP (P=0.016). Significant differences in unweighted unifrac were also observed between control and tobacco groups based on PERMANOVA (P=0.009) and ANOSIM (P=0.018) along with significant differences between OC and tobacco group based on PERMANOVA (P=0.003), ANOSIM (P=0.008) and PERMDISP (P=0.012).

PCoA plots show significantly differential clustering of OC samples as compared to tobacco and control samples which have clustered together (unweighted unifrac-qualitative). Similarly, this kind of clustering together of control and tobacco samples is observed in weighted unifrac (quantitative) PCoA plots, Bray-Curtis matrix (quantitative) as well as Jaccard matrix (qualitative) (Fig. 6). The unifrac PCoA plots take into account the phylogenetic distances between the OTUs present in the sample, whereas Bray-Curtis and Jaccard matrices take into account the community dissimilarity in the samples, without looking at the phylogenetic distances between them. The presence of similar bacterial species in control and tobacco groups as compared to the OC group can be inferred from the PCoA plots, thereby suggesting the transformation of bacterial diversity from control to tobacco to the OC group. This observation positively correlates and supports the fact that tobacco chewing is a very important risk factor for oral cancer. No significant differences were found in the weighted unifrac based on PERMANOVA, ANOSIM, and PERMDISP on comparing all three study groups to each other.

LEfSe analysis carried using Kruskal-Wallis test and Wilcoxon test displayed no differentially abundant taxa which could act as a biomarker for this study.

DISCUSSION

Numerous reports have been published displaying the diversity and variation in oral microflora in various conditions including oral cancer. Most of these studies have reported an increase in the abundance of pathogenic bacteria in the oral cavity of oral cancer patients as compared to healthy individuals. The oral microbiome of tobacco chewers and its

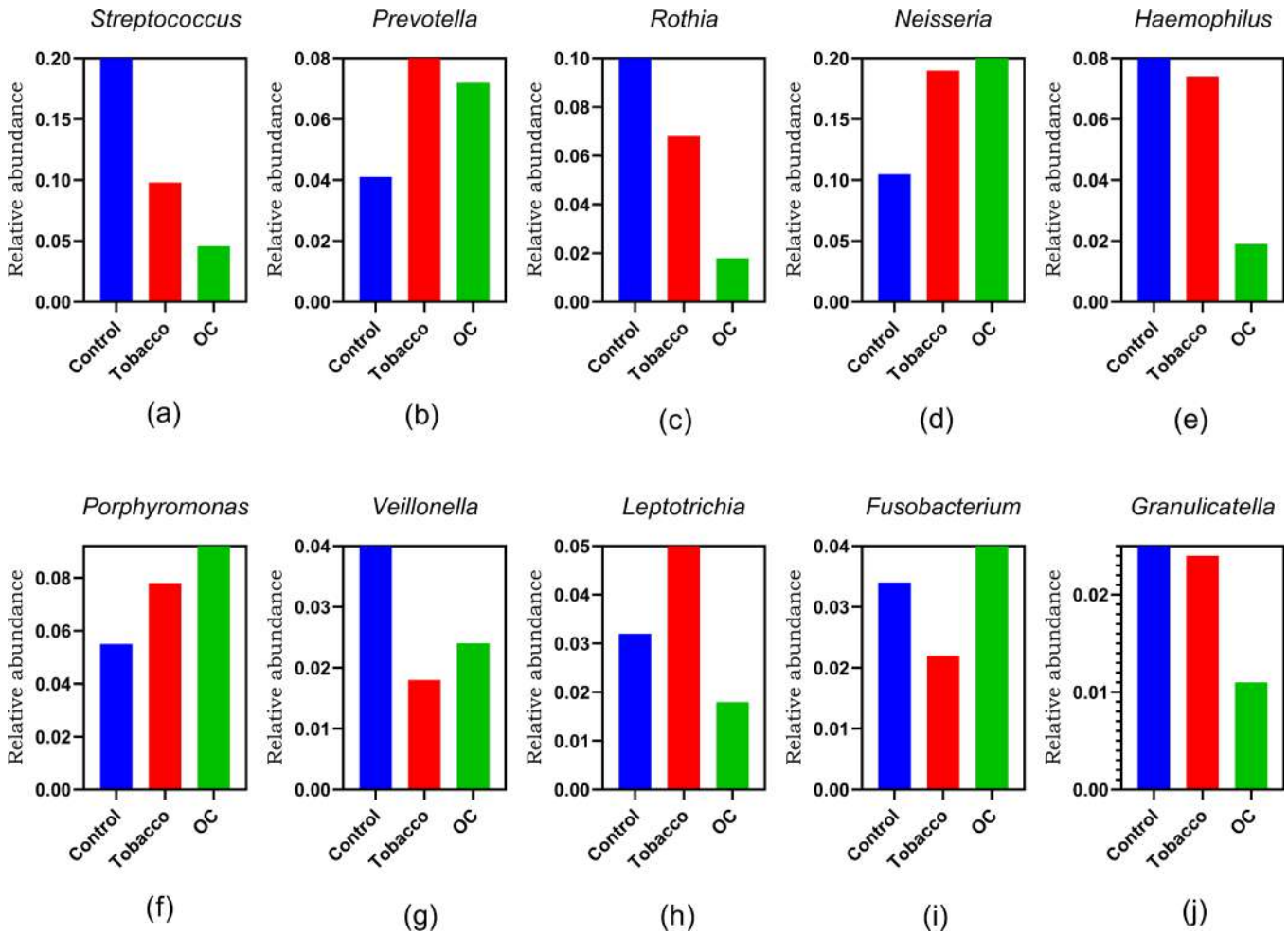
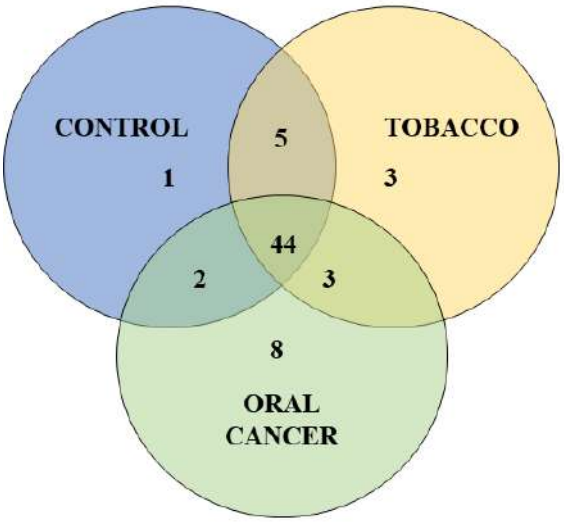


Fig. 3. Relative abundance of ten most abundant genera found in the study between the three study groups; (a) *Streptococcus*, (b) *Prevotella*, (c) *Rothia*, (d) *Neisseria*, (e) *Haemophilus*, (f) *Porphyromonas*, (g) *Veillonella*, (h) *Leptotrichia*, (i) *Fusobacterium*, (j) *Granuicatella*.

comparison with oral cancer microbiome has been scarcely studied. We, therefore, hypothesized that the oral microbiome of tobacco chewers must act as an intermediate stage between control and oral cancer microbiome.

In this pilot study, we have provided a comprehensive profile of oral microbiome of patients having oral cancer and long-term tobacco chewers as compared to the control group of an Indian population. The predominant phyla observed were *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Fusobacteria*, and *Actinobacteria*. Phyla *Firmicutes* was most abundantly observed in the oral microbiome of control population, whereas it was third most abundant phyla in tobacco and OC group, which differed from previously published data for other studies where *Firmicutes* makes up maximum microflora for both healthy as well as OC patients [21, 22]. The most abundant flora found in tobacco chewers and OC patients was *Proteobacteria* followed by phyla *Bacteroidetes*, and for OC group was phyla *Bacteroidetes* followed by phyla *Proteobacteria*, which is not in concordance with data published before. This depicts the diversity and variation in the Indian oral microbiome, which could be attributed to many factors including lifestyle, food habits, etc. At the genus level, the abundance of *Streptococcus* decreased in OC and tobacco as compared to control. Genus *Streptococcus* accounts for a major population of commensal bacteria in the oral microbiome, which in this case has clearly declined to depict dysbiosis and possible cause for a rise in pathogenic bacteria in the non-control group [30]. Although oral *Streptococci* is not novel to the Indian population and is a common oral bacterium found in various populations across the globe, the high abundance in the healthy population could be attributed to the carbohydrate-rich diet of the Indian population. Sugars are metabolized to acids by supragingival saccharolytic bacteria, including *Streptococcus*. The contribution of carbohydrates from the host diet to microbial growth in the oral cavity is not well understood, but there is evidence that carbohydrate is a growth-limiting substrate for oral *Streptococci* [31–33]. Another genus *Neisseria* was found to have increased abundance in OC and tobacco groups as compared to the control group, which has been reported previously [22, 30]. Studies



Groups	Number of genera	Genera
Control-OC-Tobacco	44	*Streptococcus, Prevotella, Neisseria, Rothia, Haemophilus, Porphyromonas, g__1, Veillonella, Fusobacterium, Leptotrichia, Granulicatella, Actinomyces, Capnocytophaga, Actinobacillus, g__2, g__3, Aggregatibacter, Paludibacter, Lautropia, Campylobacter, Tannerella, Lactobacillus, Gemella, Treponema, g__6, Dialister, Atopobium, Oribacterium, Parvimonas, g__8, Corynebacterium, g__9, Filifactor, Peptostreptococcus, Megasphaera, Bulleidia, g__10, Selenomonas, Halomonas, Eikenella, g__12, g__13, Scardovia, Moryella*
Control-Tobacco	5	*g__5, g__7, Abiotrophia, g__11, g__14*
Tobacco-OC	3	*Catonella, Mycoplasma, Lachnoanaerobaculum*
OC-Control	2	*Bifidobacterium, Shuttleworthia*
Control only	1	*g__4*
Tobacco only	3	*g__16, g__17, g__18*
OC only	8	*Pseudomonas, Propionibacterium, g__15, Bacteroides, Morganella, Alloscardovia, Peptococcus, Klebsiella*

Genera labelled g__1 to g__18 are unidentified bacterial genus

Fig. 4. Venn diagram depicting common and unique number of bacteria.

have also reported the involvement of *Neisseria* species in alcohol metabolism, a known risk factor of oral cancer as it produces acetaldehyde, a known carcinogen [34]. Our results also depict the increase in abundance of genus *Porphyromonas* in OC and tobacco groups as compared to healthy individuals. *Porphyromonas gingivalis* is a well-studied and known periodontal pathogen [35, 36]. *P. gingivalis* attaches to the epithelial cells of the oral cavity, causes inflammation, evades host defence mechanisms by utilizing numerous virulence factors, and causes deregulation of innate immunity and inflammatory responses [37]. *P. gingivalis* along with *Treponema denticola* and *Tannerella forsythia* is a part of a red-complex which causes periodontitis [38].

Genus *Prevotella* was present in an increased proportion in OC and tobacco samples as compared to the control group. A few of the known species belonging to genus *Prevotella* that are associated with dental diseases and oral cancer include *P. melaninogenica* and *P.veroralis,* which have been reported to be potential biomarkers of the pre-cancerous stage [22]. *P. intermedia, P. nigresans, P. denticola* and *P. histicola* are few other oral pathogens associated with subgingival plaque, chronic periodontitis, alcohol metabolism etc. [39–41]. Our study also showed increased proportion of genus *Aggregatibacter* in OC and tobacco compared to control. This bacterium induces bone resorption using virulence factors such as Leukotoxin, LPS and enzymes in periodontal diseases [42]. The differential abundance of specific bacteria between the three groups depicts a transitional shift in the microbiome of the control to the OC via the tobacco chewing stage.

Genus *Rothia* showed a change in its abundance within the three study groups, with 10.06% in control and 6.96% in tobacco chewers whereas only 2.20% in the OC group. These results obtained are supported by a few previous studies where a significant reduction in the abundance of *Rothia* was observed in the cancer population as compared to healthy individuals [8, 30, 43]. Although a few other reports suggest the opposite, i.e. strong association with OC samples [44–46]. Therefore, the involvement of genus *Rothia* in oral cancer remains debateable.

Along with the well-known oral bacteria, a few uncommon bacterial genera were also identified in this study which showed differential abundances in their proportion in the three study groups. The presence of genus *Catonella* was observed in this study in the tobacco group and OC group. Although the abundance of this genus is low, it has been previously reported in a few studies. *Catonella morbi* is a Gram-negative bacillus which resides in the oral cavity as a commensal, but can act as an opportunistic pathogen and has a role in endodontic infections [47]. The increased abundance of this bacteria in OSCC cases as compared to healthy individuals has been previously reported [22, 48]. Another such genus is *Lachnoanaerobaculum*, which was reported in the present study in lower abundance but has been identified as

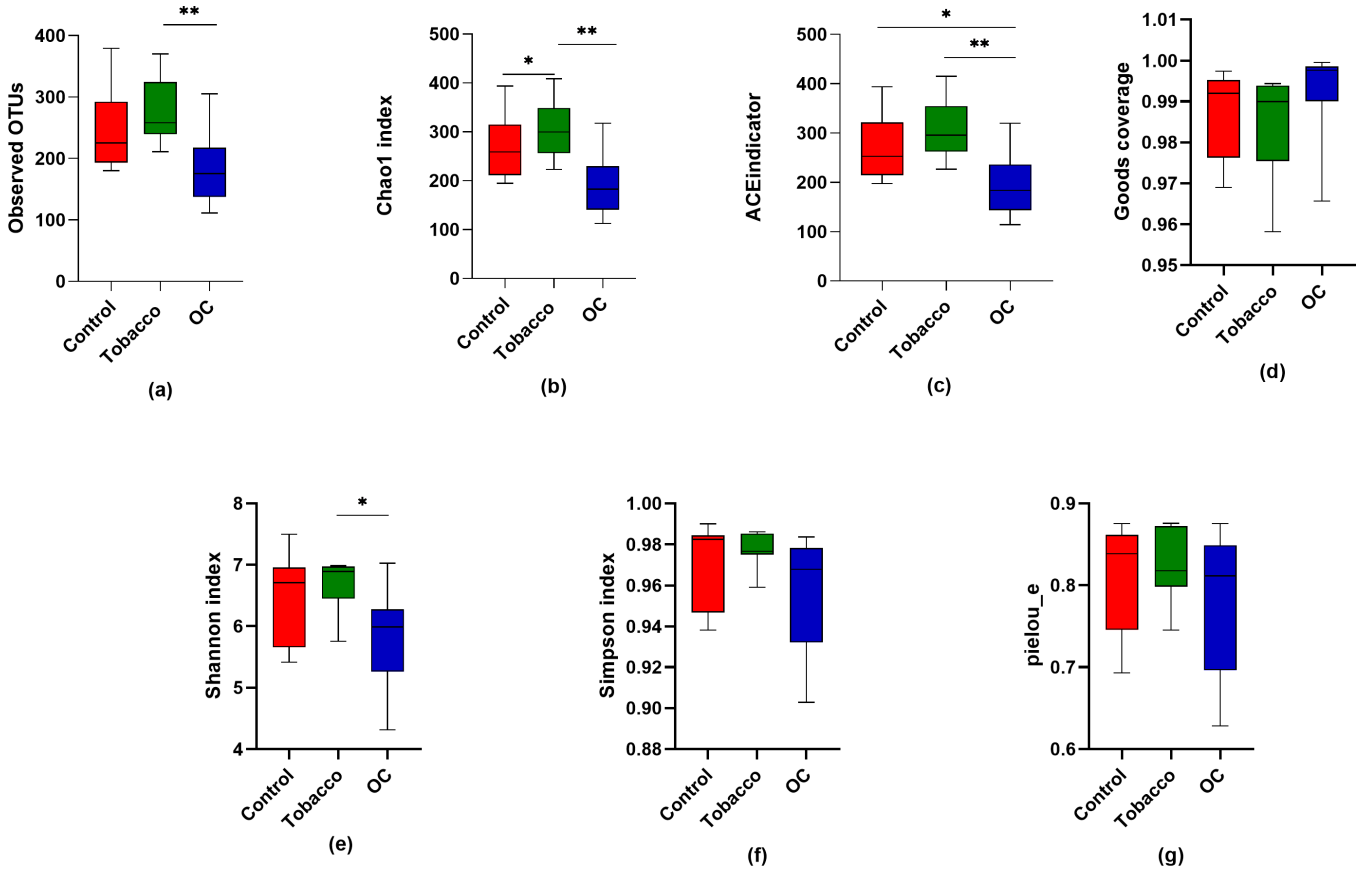


Fig. 5. Alpha diversity indices for the study groups; (a) Observed OTUs; (b) Chao1 index; (c) ACE indicator; (d) Goods coverage; (e) Shannon index; (f) Simpson index; (g) pielou_e.

a part of the OSCC microbiome in reported studies. Multiple species belonging to this genus have been associated with the oral microbiome such as *L. umeaense, L. saburreum, L. orale* and *L.* sp. *OBRC5-5* [22, 49, 50]. The genus has also been reported as a part of the core Indian salivary oral microbiome [51]. In this study, we observed the presence of the genus *Lachnoanaerobaculum* in tobacco and OC groups only.

The results obtained in this study are in parallel to most of the oral cancer metagenomic studies published previously. This signifies that the Indian oral cancer microbiome is similar to that in other parts of the world. Although, the additional study group i.e. the tobacco-chewing group shows an interesting result. The oral microbiome of long-term tobacco chewers was found to be similar to both the oral cancer microbiome as well the healthy microbiome. This pattern hints towards a transition of the oral microbiome from a healthy to an OC group, through the tobacco-chewing group, where the tobacco group shows intermediate (in-between) microbial composition, when comparing the alpha diversity indices and the bacterial composition. On the other hand, on observing the beta diversity plots, the diversity of tobacco chewers displays a similarity to control group. Hence no final judgement

Table 1. Comparison of alpha diversity matrices

Study group		Observed OTUs	ACE index	Chao1	Goods coverage	Shannon	Simpson	Pielou_e
Control	Mean	247.666	269.828	267.245	0.9872	6.43000	0.9699	0.81140
	SD	65.34	64.9824	64.9424	0.01060	0.71710	0.0205	0.06581
Tobacco	Mean	277.222	307.595	304.457	0.98384	6.66871	0.97777	0.82422
	SD	55.59	62.2552	61.2431	0.01267	0.41120	0.00841	0.04337
OC	Mean	180.777	191.744	188.896	0.9925	5.81041	0.95670	0.78061
	SD	61.49	64.6472	64.5031	0.01075	0.82789	0.03106	0.09107

Table 2. Comparison of beta diversity based on unifrac distances

Comparison groups			Permanova		Anosim		Permdisp	
			pseudo-F	*P*-value	R	*P*-value	F-value	*P*-value
Control	OC	Unweighted	1.7941	0.017	0.180384	0.011	9.105808	0.016
		Weighted	2.0164	0.093	0.21845	0.019	1.88538	0.123
Control	Tobacco	Unweighted	2.071999	0.009	0.216393	0.018	0.865104	0.339
		Weighted	1.535968	0.174	0.044925	0.213	0.2354	0.511
OC	Tobacco	Unweighted	2.378046	0.003	0.211591	0.008	15.12506	0.012
		Weighted	0.750181	0.534	0.038752	0.209	1.272984	0.256

can be made about the role of oral microbiome in tobacco chewers and their involvement in oral cancer development. Although the role of tobacco chewing in oral cancer is already a well-known phenomenon, exploring and investigating the microbial composition of tobacco chewers is necessary to support the role of microbiota along with tobacco-associated carcinogens towards the onset of oral cancer.

The demographic location, eating habits, and other lifestyle changes have been shown to play an important role in the

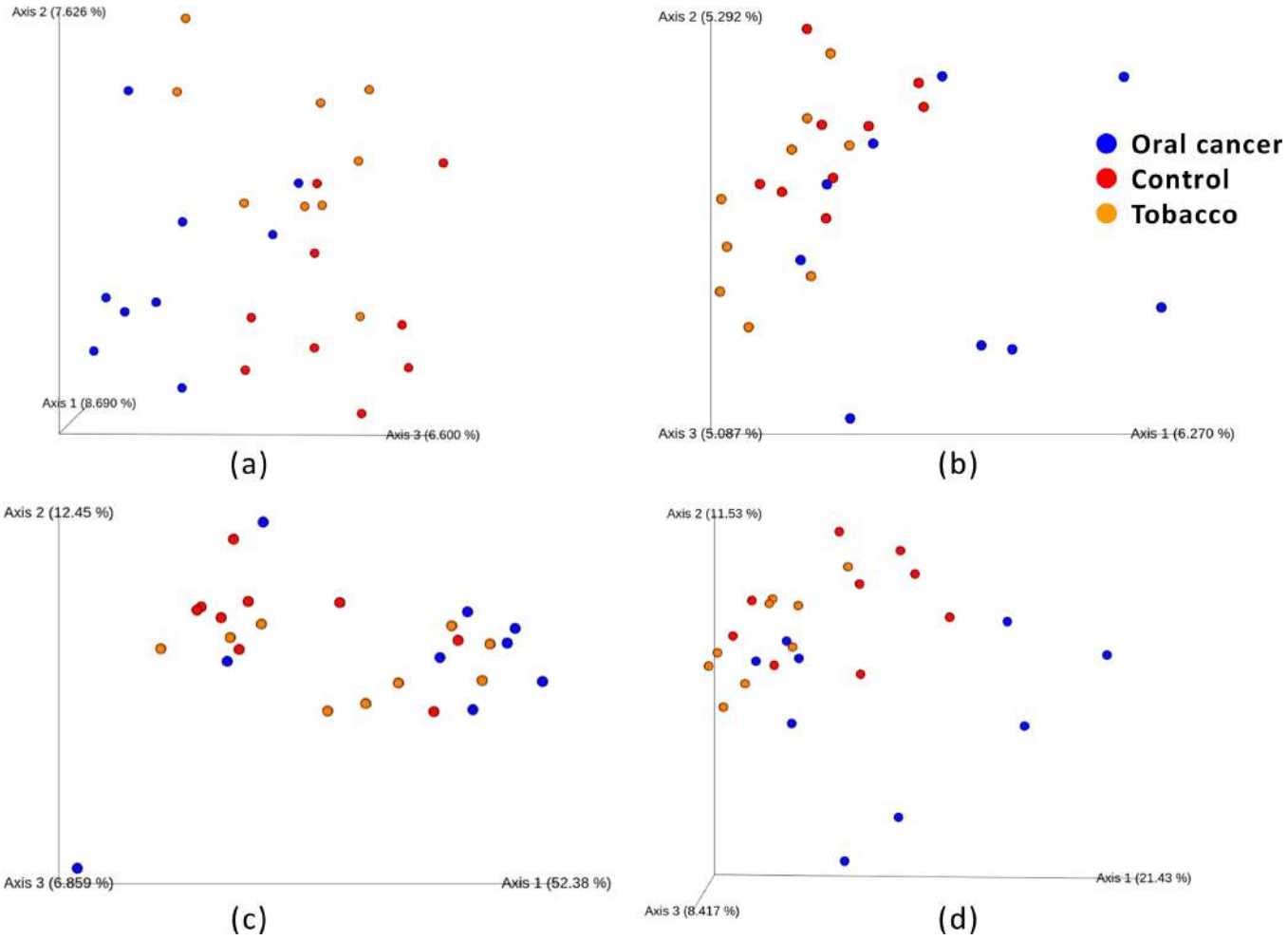


Fig. 6. Beta diversity PCoA plots depicting sample diversity between groups; (a) Bray-Curtis plot; (b) Jaccard matrix; (c) Weighted Unifrac matrix; (d) Unweighted Unifrac matrix.

dynamics of the oral microbiome [52]. The use of smokeless tobacco and areca nut, a regular habit in the United States and South Asia has been reported to play a role in oral cancer development. The tobacco-specific nitrosamines, namely N0-nitroso-nornicotine, 4-(Methylnitrosamino)−1-(3-pyridyl)−1-butanone) and the metabolite 4-(methylnitrosamino)−1-(3-pyridyl)−1-butanol are the strongest carcinogens in smokeless tobacco products and are labelled 'carcinogenic to humans' [53]. In India and Southeast Asia, tobacco is consumed in various forms such as paan, mawa, gutka, each of them possessing a varied risk of progression towards oral cancer. Gutka, which is sweet in taste is often used as a mouth freshener, but the sweetness leads to aggregation of various bacteria in the mouth leading to tooth damage. Similarly, consumption of tobacco in the form of paan which contains tobacco seeds, quenched lime, spices and areca nut folded in a betel leaf leads to pathophysiological and cellular changes in the oral cavity. Areca nut causes ROS production and hampers collagen synthesis and degradation pathways leading to increased collagen synthesis and inhibition of collagen degradation thereby causing oral submucous fibrosis, a precancerous disorder. Tobacco causes increased gingival blood flow and fibroblast damage whereas slaked lime causes cellular turnover, ROS production and chromosomal DNA damage [54, 55]. These ingredients not only cause host cell damage but also lead to dysbiosis in the oral microbiota. The abundance of oral bacteria like *Eubacterium nodatum, Peptostreptococcus micros, Streptococcus anginosus, Streptococcus constellatus, Gemella morbillorum, S. sanguinis, Veillonella parvula, P. micros, F. nucleatum, Fusobacterium periodonticum, and G. morbilorum, Treponema socranskii* and *Porphyromonas gingivalis* has been observed to increase after their interaction with tobacco extracts [14]. These bacteria are either known periodontal pathogens or have shown an increased abundance in oral cavity of oral cancer cases as well.

Recent research developments have pointed out that indeed, the microbiome population has the potential to affect or alter the human health and immune response and thus induce diseases and certain cancers [56–59]. However, the precise role of specific species and the mode/pathway of action of the microbiome leading to the development of cancer is still not clearly deciphered and understood. The reported data articulates that there exists diversity between cancer and non-cancer population from different community individuals. Beyond this, however, further understanding is required to gain strong insights into the functional characterization and mechanisms involved in the establishment of different microbial communities, understand the interspecies interaction and microbe-host interactions, which would eventually help in developing preventive or treatment therapeutics. Typically, in cancer patients, the oral flora diversity and load is affected by various factors like the disease and malignancy stage, antibiotic use, surgery or chemo-radiation treatment, age and overall health status. There is a dearth of reports that give information about microbiome analysis and its effect on oral cancer progression, treatment or morbidity.

Conclusion

Current studies on oral microbiome can help in developing microbe screening tests using simple saliva sample swabs from patients. This can be an attractive option as oral samples are easier to take, are usually non-invasive and is an inexpensive way of collecting samples as opposed to invasive methods. Even though immense research is being conducted worldwide to understand oral cancer and its nuances, there is still an enormous need to fully understand the aetiology and the role of microbes, understand the interspecies interactions and host-microbe interaction which could eventually help in diagnostics, prevention, and treatment therapies. The microbiome analysis of oral cancer patients and its correlation with premalignant conditions and oral cancer is anticipated to be useful in predictive diagnostics, progression and therapeutic regimen, or identification of therapeutic targets in the precancerous stages and oral cancers. This database for the Indian population generated by the study can help in developing a microbiome-based treatment strategy in oral cavity related complications and diseases.

Since this is a pilot study, limited information could be deciphered from current work. Although the results obtained in this study are in concordance with the previously published reports, along with an insight about the microbiota of tobacco chewers, future studies are required with more number of samples for better confidence in the results.

Data availability

The 16S rRNA amplicon sequencing data from this study have been deposited in the NCBI BioProject under accession number PRJNA751046.

Funding information
This research work was supported by research grant from NMIMS University seed grant (University Seed Grant 2016) and Seoul Women's University (2021).

Acknowledgements
The authors would like to thank Dr Suresh Shettigar, BARC Hospital, Mumbai for providing initial guidance with NGS analysis. The authors would also like to thank the working staff of Somaiya Ayurvihar-Asian Cancer Institute (K. J. Somaiya Medical College and Hospital, Mumbai) for providing help with oral cancer sample collection.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
Informed written consent was obtained prior to the sample collection from all the participants. Ethical approvals for the recruitment of healthy, long-term tobacco chewing volunteers and oral cancer patients were granted by SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016, dt. 4/7/2016) and K. J. Somaiya Medical College and Hospital, Mumbai Institutional Ethics Committee (dt. 14/9/2017).

References

1. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet* 2012;13:260–270.
2. Kilian M, Chapple I, Hannig M, Marsh PD, Meuric V, *et al*. The oral microbiome - An update for oral healthcare professionals. *Br Dent J* 2016;221:657–666.

3. Belstrøm D, Holmstrup P, Nielsen CH, Kirkby N, Twetman S, *et al*. Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. *J Oral Microbiol* 2014;6.
4. Gupta B, Bray F, Kumar N, Johnson NW. Associations between oral hygiene habits, diet, tobacco and alcohol and risk of oral cancer: A case–control study from India. *Cancer Epidemiol* 2017;51:7–14.
5. Hernandez BY, Zhu X, Goodman MT, Gatewood R, Mendiola P, *et al*. Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLoS One* 2017;12:e0172196.
6. Čanković M, Bokor-Bratić M, Lončar J, Marinoski J, Ilić MP. Bacterial flora on the surface of oral squamous cell carcinoma. *Arch Oncol* 2013;21:62–64.
7. Gholizadeh P, Eslami H, Yousefi M, Asgharzadeh M, Aghazadeh M, *et al*. Role of oral microbiome on oral cancers, a review. *Biomed Pharmacother* 2016;84:552–558.
8. Verma D, Garg PK, Dubey AK. Insights into the human oral microbiome. *Arch Microbiol* 2018;200:525–540.
9. Muwonge R, Ramadas K, Sankila R, Thara S, Thomas G, *et al*. Role of tobacco smoking, chewing and alcohol drinking in the risk of oral cancer in Trivandrum, India: A nested case-control design using incident cancer cases. *Oral Oncol* 2008;44:446–454.
10. Mahapatra S, Kamath R, Shetty BK, Binu VS. Risk of oral cancer associated with gutka and other tobacco products: a hospital-based case-control study. *J Cancer Res Ther* 2015;11:199–203.
11. Ali J, Sabiha B, Jan HU, Haider SA, Khan AA, *et al*. Genetic etiology of oral cancer. *Oral Oncol* 2017;70:23–28.
12. Tenore G, Nuvoli A, Mohsen A, Cassoni A, Battisti A, *et al*. Tobacco, alcohol and family history of cancer as risk factors of oral squamous cell carcinoma: case-control retrospective study. *Appl Sci* 2020;10:1–11.
13. Mager D. Bacteria and cancer: cause, coincidence or cure? A review. *JTranslMed* 2006;4:14.
14. Liu M, Jin J, Pan H, Feng J, Cerniglia CE, *et al*. Effect of smokeless tobacco products on human oral bacteria growth and viability. *Anaerobe* 2016;42:152–161.
15. Rao SVK, Mejia G, Roberts-Thomson K, Logan R. Epidemiology of oral cancer in Asia in the past decade - An update (2000-2012). *Asian Pacific J Cancer Prev* 2013;14:5567–5577.
16. Naveen-Kumar B, Tatapudi R, Sudhakara-Reddy R, Alapati S, Pavani K, *et al*. Various forms of tobacco usage and its associated oral mucosal lesions. *J Clin Exp Dent* 2016;8:e172-7.
17. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, *et al*. The human microbiome project. *Nature* 2007;449:804–810.
18. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, *et al*. The human microbiome project: A community resource for the healthy human microbiome. *PLoS Biol* 2012;10:e1001377.
19. Lecuit M, Eloit M. The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening. *Front Cell Infect Microbiol* 2014;4:25.
20. Bzhalava D. Bioinformatics for Viral Metagenomics. *J Data Mining Genomics Proteomics* 2013;4.
21. Yang CY, Yeh YM, Yu H-Y, Chin CY, Hsu CW, *et al*. Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. *Front Microbiol* 2018;9:862.
22. Mok SF, Karuthan C, Cheah YK, Ngeow WC, Rosnah Z, *et al*. The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. *Malays J Pathol* 2017;39:1–15.
23. Yang SF, Huang H-D, Fan WL, Jong YJ, Chen MK, *et al*. Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer. *Oral Oncol* 2018;77:1–8.
24. Michl SC, Beyer M, Ratten JM, Hasler M, LaRoche J, *et al*. A diet-change modulates the previously established bacterial gut community in juvenile brown trout (Salmo trutta). *Sci Rep* 2019;9:2339.
25. Michl SC, Ratten JM, Beyer M, Hasler M, LaRoche J, *et al*. The malleable gut microbiome of juvenile rainbow trout (*Oncorhynchus mykiss*): Dietdependent shifts of bacterial community structures. *PLoS One* 2017;12:e0177735.
26. Guo J, Cole JR, Zhang Q, Brown T, Tiedje JM. Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. *Appl Environ Microbiol* 2016;82:157–166.
27. Kim DW, Shin J, Guevarra RB, Lee JH, Kim DW, *et al*. Deciphering diversity indices for a better understanding of microbial communities. *J Microbiol Biotechnol* 2017;27:2089–2093.
28. Jost L. The relation between evenness and diversity. *Diversity (Basel)* 2010;2:207–232.
29. Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, *et al*. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* 2009;4:e6669.
30. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, *et al*. Changes in abundance of oral microbiota associated with oral cancer. *PLoS One* 2014;9:e98741.
31. Takahashi N. Oral microbiome metabolism. *J Dent Res* 2015;94:1628–1637.
32. De Jong MH, Van Der Hoeven JS, Van Os JH, Olijve JH. Growth of oral *Streptococcus* species and *Actinomyces viscosus* in human saliva. *Appl Environ Microbiol* 1984;47:901–904.
33. Tandon S, Prabhakar AR, Udupa S. Effect of diet on caries-inducing *Streptococcus* mutans growth. *Food Chem* 1990;38:289–296.
34. Muto M, Hitomi Y, Ohtsu A, Shimada H, Kashiwase Y, *et al*. Acetaldehyde production by non-pathogenic Neisseria in human oral microflora: Implications for carcinogenesis in upper aerodigestive tract. *Int J Cancer* 2000;88:342–350.
35. Khajuria N, Metgud R. Role of bacteria in oral carcinogenesis. 2015;6:37–43.
36. Ferguson B, Barber S, Asher I, Wood C. Role of oral microbial infections in oral cancer. *Dent Clin North Am* 2017;61:425–434.
37. Nakhjiri SF, Park Y, Yilmaz O, Chung WO, Watanabe K, *et al*. Inhibition of epithelial cell apoptosis by Porphyromonas gingivalis. *FEMS Microbiol Lett* 2001;200:145–149.
38. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. *Nat Rev Microbiol* 2012;10:717–725.
39. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. *J Clin Periodontol* 1998;25:134–144.
40. Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, *et al*. New bacterial species associated with chronic periodontitis. *J Dent Res* 2003;82:338–344.
41. Marttila E, Uittamo J, Rusanen P, Lindqvist C, Salaspuro M, *et al*. Acetaldehyde production and microbial colonization in oral squamous cell carcinoma and oral lichenoid disease. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013;116:61–68.
42. Gholizadeh P, Pormohammad A, Eslami H, Shokouhi B, Fakhrzadeh V, *et al*. Oral pathogenesis of Aggregatibacter actinomycetemcomitans. *Microb Pathog* 2017;113:303–311.
43. Al-Hebshi NN, Borgnakke WS, Johnson NW. The microbiome of oral squamous cell carcinomas: a functional perspective. *Curr Oral Heal Reports* 2019;6:145–160.
44. Amer A, Galvin S, Healy CM, Moran GP. The microbiome of potentially malignant oral leukoplakia exhibits enrichment for fusobacterium, leptotrichia, campylobacter, and rothia species. *Front Microbiol* 2017;8:2391.
45. Hooper SJ, Crean SJ, Fardy MJ, Lewis MA, Spratt DA, *et al*. A molecular analysis of the bacteria present within oral squamous cell carcinoma. *J Med Microbiol* 2007;56:1651–1659.
46. Banerjee S, Tian T, Wei Z, Peck KN, Shih N, *et al*. Microbial signatures associated with oropharyngeal and oral squamous cell carcinomas. *Sci Rep* 2017;7:4036.
47. Sakamoto M, Rôças IN, Siqueira JF, Benno Y. Molecular analysis of bacteria in asymptomatic and symptomatic endodontic infections. *Oral Microbiol Immunol* 2006;21:112–122.

48. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R, *et al*. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol* 2012;12:144.
49. Lee WH, Chen HM, Yang SF, Liang C, Peng CY, *et al*. Bacterial alterations in salivary microbiota and their association in oral cancer. *Sci Rep* 2017;7:1–11.
50. Zhao H, Chu M, Huang Z, Yang X, Ran S, *et al*. Variations in oral microbiota associated with oral cancer. *Sci Rep* 2017;7:1–10.
51. Sarkar A, Stoneking M, Nandineni MR. Unraveling the human salivary microbiome diversity in Indian populations. *PLoS One* 2017;12:e0184515.
52. Warnakulasuriya S. Causes of oral cancer – an appraisal of controversies. *Br Dent J* 2009;207:471–475.
53. Hecht SS. *Metabolism and DNA Adduct Formation of Carcinogenic Tobacco-specific Nitrosamines Found in Smokeless Tobacco Products*. Elsevier Inc, 2020.
54. Niaz K, Maqbool F, Khan F, Bahadar H, Ismail Hassan F, *et al*. Smokeless tobacco (paan and gutkha) consumption, prevalence, and contribution to oral cancer. *Epidemiol Health* 2017;39:e2017009.
55. Shah G, Chaturvedi P, Vaishampayan S. Arecanut as an emerging etiology of oral cancers in India. *Indian J Med Paediatr Oncol* 2012;33:71–79.
56. Scannapieco FA. The oral microbiome: Its role in health and in oral and systemic infections. *Clin Microbiol Newsl* 2013;35:163–169.
57. Yamashita Y, Takeshita T. The oral microbiome and human health. *J Oral Sci* 2017;59:201–206.
58. Le Bars P, Metamoros S, Montassier E, Le Vacon F, Potel G, *et al*. The oral cavity microbiota: between health, oral disease and cancers of the aerodigestive tract. *Can J Microbiol* 2017;492:cjm-2016-0603.
59. Marsh PD. In Sickness and in Health-What Does the Oral Microbiome Mean to Us? An Ecological Perspective. *Adv Dent Res* 2018;29:60–65.

Medicine in Microecology 7 (2021) 100034

Contents lists available at ScienceDirect

Medicine in Microecology

journal homepage: www.journals.elsevier.com/medicine-in-microecology

Absolute quantitation of oral bacteria involved in oral cancer by real-time PCR

Shriya Sawant [a], Jinesh Dugad [b], Deepak Parikh [b], Harinder Singh [a,*]

[a] Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS University, Vile Parle (West), Mumbai, India
[b] Somaiya Ayurvihar -Asian Cancer Institute, Off Eastern Express Highway, Behind Everard Nagar, Somaiya Ayurvihar, Sion (East), Mumbai, India

ARTICLE INFO

Keywords:
Realtime-PCR
Absolute quantification
Specificity
Oral cancer
Tobacco chewing

ABSTRACT

India carries approximately 34% of the global oral cancer burden. Along with tobacco chewing habit, there are other factors that are being studied globally like the involvement of bacteria in oral cancer. NGS studies have identified few bacterial species and the changes in their abundance in healthy and diseased populations based on 16S rRNA metagenomics. The present study shows a method for absolute bacterial quantification from oral cavity rinse samples. The method is a real-time qPCR assay and based on the fact that certain genes are present in one copy per cell (*rpoB* gene) and we can correlate the copy numbers of these genes with cell numbers in a sample. This method is more accurate than the NGS 16S rRNA gene-based approach which is multicopy gene. Linear correlation between qPCR assay and cell numbers of a model system was established. Consequently, the assay was performed on oral rinse samples of oral cancer/tobacco chewers and healthy subjects to quantitate significant oral bacterial species. The obtained bacterial quantification correlated well with the previous reports. The developed qPCR method is an efficient, faster and resource-friendly method and can be used to quantify bacterial population in cancer/diseased subjects, and can have application in determining the susceptibility of an individual towards a specific disease.

1. Introduction

Quantitative PCR is an adaptation of PCR, which is widely used for quantitation of gene expression at mRNA levels [1]. Endpoint PCR requires the detection and quantification of amplified products by gel electrophoresis, whereas real time qPCR technology allows the quantification of amplified products in real time, *i.e.* at the end of each cycle, it provides a quantitative measurement of the amplified products accumulated during the previous reaction [2]. Quantitative PCR is extensively used for microbial ecology purposes to determine the microbial load and diversity in any environmental sample [3]. The specificity of any qPCR depends on the design of the primers, which determines which genes will be amplified. It also provides the quantification of taxonomic or functional gene markers from the domain level to the quantification of individual species or phylotypes within a mixed community. Microbial qPCR assays are designed to target highly conserved 16S rRNA genes to quantify total bacterial load within a sample, whereas hypervariable regions within this gene may be targeted for specific amplification and quantification of specific bacterial population [4]. A pitfall of this approach, which must be noted, is that 16S rRNA gene copies cannot be directly correlated with cell numbers accurately, since every bacterial species has variable copies (1–15 copies) of 16S rRNA [5]. However, an alternative to 16S rRNA gene is to use a housekeeping gene, RNA Polymerase β-subunit gene (*rpoB*). The gene *rpoB* is a single copy gene that enables microbial communities to be quantified accurately. *rpoB* is widely present in all prokaryotes, has a housekeeping function, which makes it less vulnerable to certain types of lateral gene transfer. It has well-defined hypervariable areas at genus and species level, improving phylogenetic resolution comparable to 16S rRNA gene [6]. A few reports have been published, where *rpoB* gene was used for specific amplification and quantification of bacterial populations, such as *Porphyromonas gingivalis* and *Veillonella* sps [7–9].

The correlation of certain bacteria with different diseases including cancers is well known and reported. The role of bacteria in some cancers was elucidated in recent work, and for many others, it still remains unknown [10–12]. With the advancement in contemporary techniques like the invention of Next Generation Sequencing systems, researchers have established the changes in the microbial composition (microbiome) of affected locations during cancer. Assessing the dynamics of the microbial population in diseased conditions can help in determining the

* Corresponding author.
E-mail address: Harinder.Singh@nmims.edu (H. Singh).

https://doi.org/10.1016/j.medmic.2021.100034
Received 18 October 2020; Received in revised form 5 February 2021; Accepted 24 March 2021
Available online 31 March 2021

mechanisms, which bacteria can use to trigger or develop the cancer. In this study, a method is developed for absolute quantification of eight different bacterial species, associated with oral cancer using quantitative real time PCR method. Three study groups were used, namely oral cancer patients, long term tobacco chewers, and healthy controls. Since tobacco chewing is one of the most important risk factors of oral cancer, tobacco user group can help in finding any population changing pattern.

2. Material and methods

2.1. Sample collection and data collection

A total of 120 subjects in the study were classified into three groups - Healthy (control, n-40), Oral Squamous Cell Carcinoma (OSCC) stage 2 and 3, (OC n-40) and long-term tobacco chewers (Tobacco, n-40), each consisting of 40 individuals. The healthy controls were defined as individuals without any diagnosed diseases in the oral cavity, which was confirmed by prior clinical examination. Long term tobacco chewers were individuals chewing various forms of tobacco for 5 years or more. All diagnoses of OSCC were confirmed by biopsy and pathological findings. All participants (control and long-term tobacco chewers) were not undergoing antibiotics treatment at the time of sample collection.

Data on participant's medical history, age, gender, occupation, habits of tobacco ingestion, alcohol intake, and general oral hygiene questions were recorded. Informed written consent was obtained prior to the sample collection from all the participants. Ethical approvals for the recruitment of healthy, long term tobacco chewing volunteers and oral cancer patients were granted by SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and The Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai.

2.2. Sample collection

The sample was collected in the form of oral rinse. The volunteers were asked to rinse their oral cavity with sterile saline for 30 s and the sample was collected in a sterile 50 mL falcon tube. Salivary samples of patients who were taken the pre-operative stage while the patient was NBM (Nill By Mouth). Subjects from healthy & tobacco groups were asked to refrain from eating, drinking, or oral hygiene procedures for at least 1 h before sample collection. Salivary samples were collected in well-labelled sterile Falcon tubes and stored at 4 °C and processed within 48 h.

2.3. DNA extraction and quantitation

The collected samples were centrifuged at 14000 rpm for 8 min at room temperature. The supernatant was discarded and the pellet was used for DNA isolation by using Invitrogen PureLink™ Genomic DNA kit (Cat no. K182002) according to the manufacturer's recommendation. DNA samples were eluted using 100 μL Elution buffer. DNA was kept frozen at −20 °C until PCR amplification. The purity and yield of DNA were assessed spectrophotometrically using BioTek EPOCH2 microplate reader. The purity of DNA was determined by the ratio of absorbance at 260 and 280 nm (A260/A280).

2.4. Selection of candidate organism and primer designing

Eight bacterial species i.e., *Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula and Streptococcus mutans* were chosen based on previous reports on prevalent and significant oral cavity bacterial species linked with dysbiosis in oral cancer [13,14]. For absolute quantification, *rpoB* gene specific primers were designed using GenScript Real-time PCR Primer Design software. Along with the eight oral bacteria, primers were designed for *Escherichia coli* specific *rpoB* gene as well since it was the model organism for the study. The primer sequences are listed in Table 1.

2.5. Establishing a standard curve for E.coli

E. coli MG1655 was used as a model organism to establish a correlation between threshold cycle (C_t) of amplification and Colony Forming Units (CFU/mL). A 16–18hr old *E. coli* MG1655 culture was inoculated in Tryptone Soya Broth, and used to perform CFU assay. Optical density (OD_{600}) of the culture was adjusted to 0.1OD. From this sample, 1 mL aliquots were serially diluted and spread plated for CFU estimation. Parallelly, from 1 mL aliquot, DNA isolation was carried out using Invitrogen PureLink™ Genomic DNA kit. The purity and yield of DNA were assessed spectrophotometrically using BioTek EPOCH2 microplate reader. The purity of DNA was determined by the ratio of absorbance at 260 and 280 nm (A260/A280). The correlation between OD, CFU/mL and DNA concentration (fg/μL) was established.

To construct the standard curve, DNA concentration of 0.1OD culture was noted and 10-fold dilutions were used to construct the standard curves for *rpoB* gene amplification. The reaction mixture was prepared using PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) in a total volume of 10 μL: 0.5 μL each primer (final concentration 0.5 μM), 5.0 μL PowerUp™ SYBR® Green Master mix. For *E. coli* MG1655 standardization, 6-serial dilutions of DNA isolated from 0.1OD *E. coli* MG1655 were prepared and used to obtain standard graph. All real-time PCR runs were performed in triplicates and C_t values were plotted against the log of DNA concentration to find equation of standard curve which was used for all further calculations.

2.6. Real-time qPCR for significant bacteria

For absolute quantification, primers designed for *rpoB* gene of *Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula and Streptococcus mutans* were used. For the PCR, reactions were set up for a total volume of 10 μL, containing 0.5 μL each primer (final concentration 0.5 μM), 5.0 μL PowerUp™ SYBR® Green Master mix, 30–40 ng sample DNA was used, and the rest of the volume was DNAase free water. The thermal cycling protocol used for amplification was as follows: initial denaturation for 10 min at 95 °C followed by 40 cycles of 15 s at 95 °C, 30 s at 58 °C, and 1 min at 72 °C. The fluorescence signal was measured at the end of each extension step at 72 °C. After the amplification, a melting peak analysis with a temperature gradient of 0.5 °C s−1 from 60 to 95 °C was performed to confirm that only the specific products were amplified. Finally, the samples were cooled down to 40 °C for 30 s. The Tm obtained by amplification of *rpoB* of all bacteria was noted and the amplified products were checked on 2% agarose gel for the size of amplified products.

Real time PCR was carried using SYBR green reagent for 40 samples from each group (Control, Tobacco chewers, Cancer) for all 8 significant bacteria. The C_t values were recorded and calculations were performed to obtain the absolute count in terms of Bacterial Genomic Equivalent per mL (BGE/mL). All the real-time PCR runs for all samples were performed in triplicates and a non-template control was maintained for each run.

3. Results and discussion

After performing the CFU assay and carrying out DNA isolation using *E.coli* as the model organism, the correlation between OD, CFU/mL and DNA concentration (ng/μL) was established where 0.1OD @ 600 nm corresponded to 3.856 (±1.5) X 10^7 CFU/mL and the DNA isolated from this culture was 3.560 ng/μL. DNA concentration of 0.1OD culture was noted and 10-fold dilutions ranging from 3560 pg/μL to 35.6 fg/μL was used to construct the standard curves for *rpoB* gene amplification. The melt-curve analysis showed that Tm was constant, i.e. 81.75 °C in all runs confirming the amplification of single product. The standard graph for C_t plotted against concentration range (3560 pg–35.6 fg) was obtained and

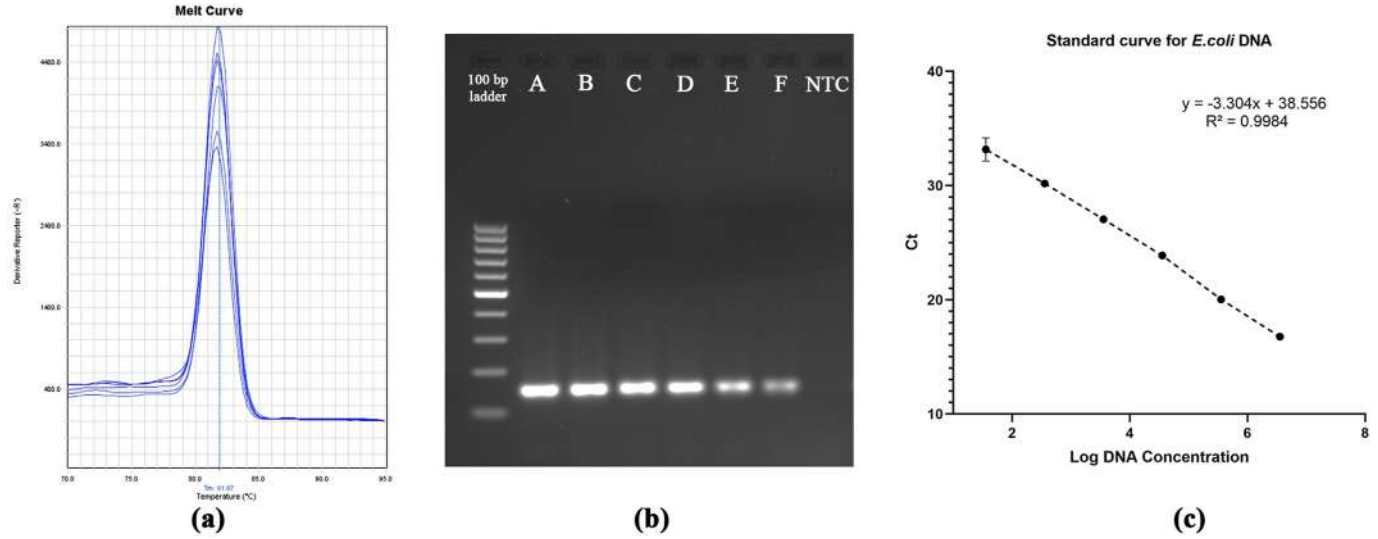
Table 1
Primer sequences, specific Tm values and size of product obtained for bacteria of interest.

Organism	Primer Sequences	Tm obtained (°C)	Size of product (bp)
Escherichia coli MG1655	**F**: 5′- GTTGACTCCGGTGTAACTGC-3′ **R**: 5′- ACGGGTGTATTTGGTCAGGT -3′	81.38	147
Porphyromonas gingivalis	**F**: 5′- TCGACCTGATGGACGTTTCGC-3′ **R**: 5′- ACGGTTGGCATCGTCGTGTT-3′	80.02	86
Fusobacterium nucleatum	**F**: 5′GGTTCAGAAGTAGGACCGGGAGA3′ **R**: 5′-ACTCCCTTAGAGCCATGAGGCAT-3′	77.02	161
Capnocytophaga gingivalis	**F**: 5′- TGGATGCCATCGGAGCCAAC-3′ **R**: 5′- GGCATCGAGGGTACGGGAGA-3′	83.48	104
Haemophilus parainfluenzae	**F**: 5′- AGCGAGTACGGAACACGCAA-3′ **R**: 5′- TGCAGTTCCAATTCCCGATCCA-3′	79.53	113
Prevotella melaninogenica	**F**: 5′- GTGCTCGTGTTGAGCCAGGT-3′ **R**: 5′- TCACCAGCCTTGTCACCGAA-3	84.4	115
Rothia mucilaginosa	**F**: 5′- AGTCTGAGGCTCCCGTGGTT-3′ **R**: 5′- AACCACACCGGGCTTCTTCG-3′	84.8	92
Veillonella parvula	**F**: 5′- GCGCGAACATGCAACGTCAA-3′ **R**: 5′- CACGCGCCAATACGCAAACA-3′	82.75	114
Streptococcus mutans	**F**: 5′- CGTCATGGGACGTCATCAAGGG-3′ **R**: 5′-AGCTCCCATAAGGGCACGGT-3′	81.46	151

the curve was highly linear ($R^2 > 0.99$). The slope of the standard curve for *rpoB* was −3.3. This also confirms the PCR efficiency close to 100%. The final equation of standard graph obtained ($Y = -3.304X + 38.556$) was used for all further calculations where, Y is the C_t value for a sample and X is the log of DNA concentration (Fig. 1).

The standard curve was used to calculate the DNA concentration of a sample. The equation used is $Y = -3.304X + 38.556$, where Y is the C_t value and X is the log of DNA concentration. According to the established data, 3.56 ng DNA = 3.856 *10^7 cells, so 3.56 pg DNA = 3.856*10^4 cells and 3.56 fg DNA = 3.856 cells. Using this correlation data, absolute count of the listed bacterial species was obtained for oral cavity sample, after standardizing the PCR conditions for each of the bacterial species and analysing their melt peaks (Fig. 2).

As seen in Fig. 3, significant differences in the absolute count were obtained between control and long-term tobacco chewing population for *P. gingivalis* (~ 8.45 X 10^4 vs.~ 4.28 X 10^5 BGE/ml) as well as between control and oral cancer patients (~ 8.45 X 10^4 vs.~ 4.01 X 10^5 BGE/ml). For *F. nucleatum,* a significant difference was observed in absolute count between control and oral cancer patients (~ 4.06 X 10^5 vs.~ 9.73 X 10^5 BGE/ml). For both these bacterial species, there was an increase in the abundance of respective bacterial population on comparing the control group with tobacco chewers as well as oral cancer patients. For



(a) (b) (c)

Concentration of DNA (fg)	Log DNA Concentration	Ct value
3560000 (A)	6.55145	16.76 ± 0.90
356000 (B)	5.55145	20.02 ± 1.02
35600 (C)	4.55145	23.87 ± 0.08
3560 (D)	3.55145	27.03 ± 0.22
356 (E)	2.55145	30.18 ± 0.08
35.6 (F)	1.55145	33.16 ± 0.26

(d)

Fig. 1. Standard curve for *E.coli* (a) Melt peaks observed for standard curve of *E.coli* depicting steady increase in amplification. (b) Agarose gel electrophoresis of standard curve products obtained after amplification of *E.coli* using Real time PCR. NTC: No Template Control (c) Standard curve linear graph C_t vs Log DNA concentration depicting $R^2 > 0.998$.

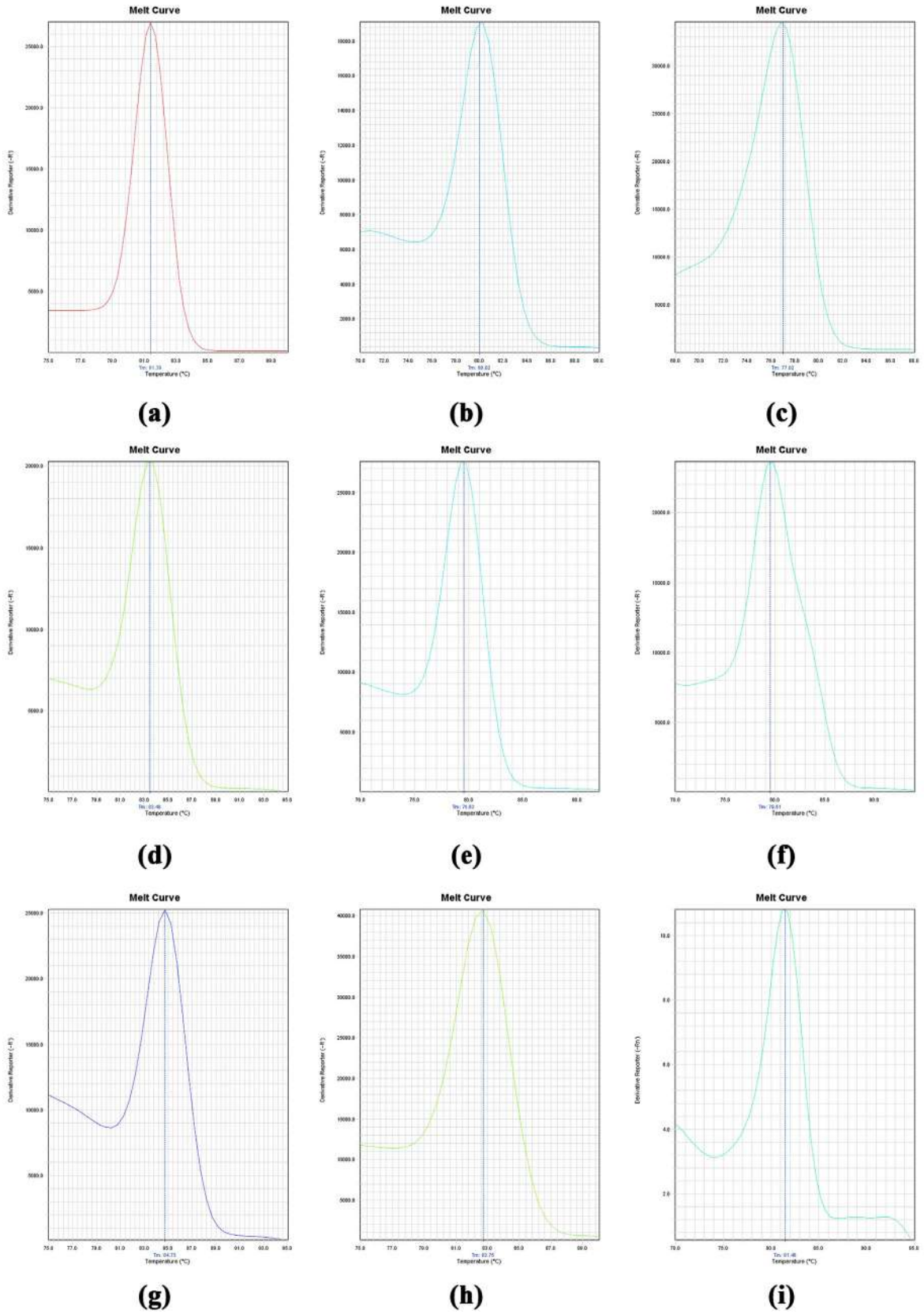


Fig. 2. Melt peaks for *rpoB* gene amplification of nine bacteria (a) *Escherichia coli*, (b) *Porphyromonas gingivalis*, (c) *Fusobacterium nucleatum*, (d) *Capnocytophaga gingivalis*, (e) *Haemophilus parainfluenzae*, (f) *Prevotella melaninogenica*, (g) *Rothia mucilaginosa*, (h) *Veillonella parvula*, (i) *Streptococcus mutans*. Single peaks and constant melting temperature (Tm) are obtained in all the runs demonstrating amplification of single product.

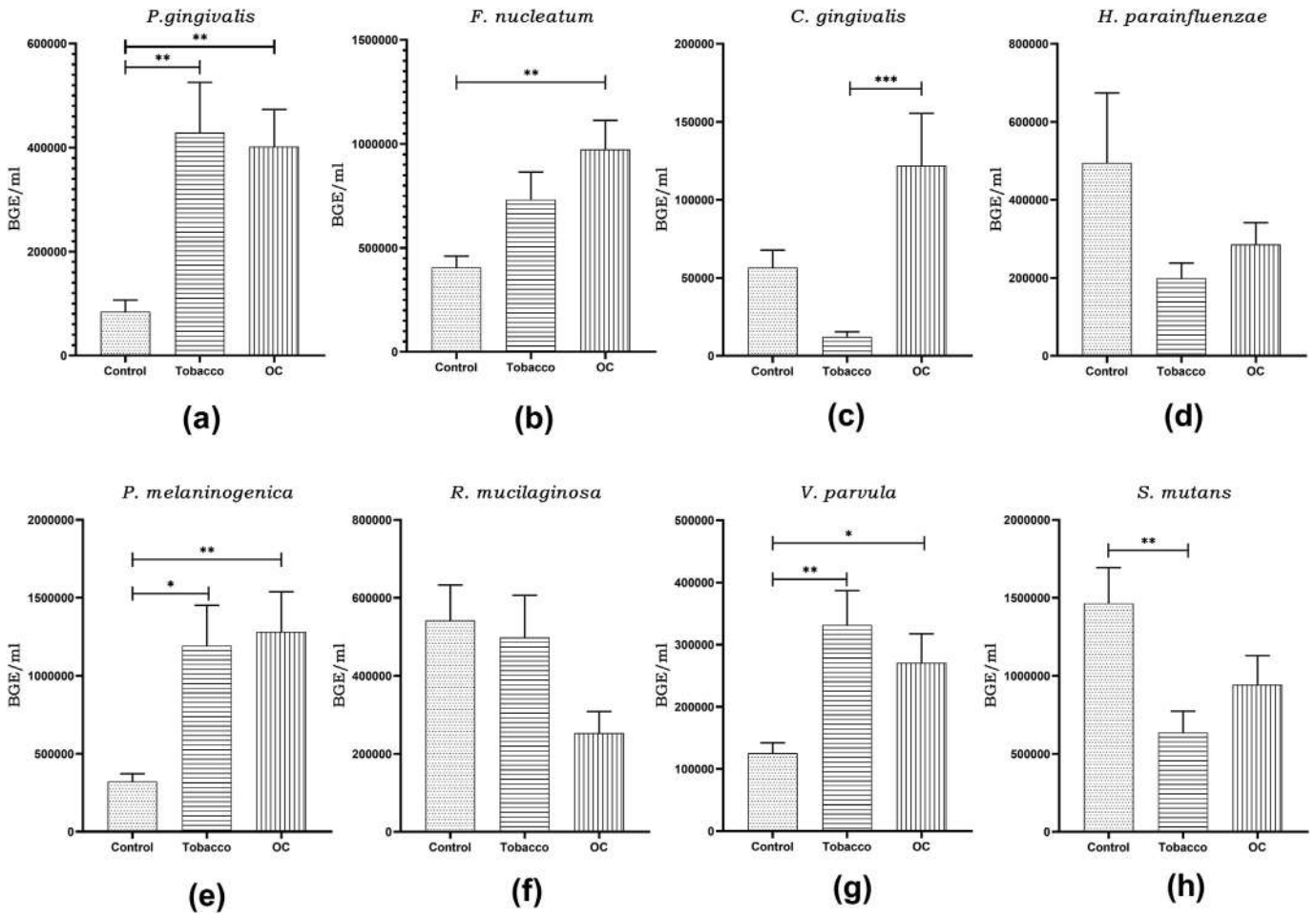


Fig. 3. Bar graphs displaying difference in abundance of eight tested bacteria in 3 different study groups. (a) *Porphyromonas gingivalis*, (b) *Fusobacterium nucleatum*, (c) *Capnocytophaga gingivalis*, (d) *Haemophilus parainfluenzae*, (e) *Prevotella melaninogenica*, (f) *Rothia mucilaginosa*, (g) *Veillonella parvula*, (h) *Streptococcus mutans*.

C. gingivalis, a significant difference was observed on comparing bacterial count in oral cancer patients (~ 1.21 X 10^5 BGE/ml) with long term tobacco chewing population (~ 1.2 X 10^4 BGE/ml). The absolute abundance of *P. melaninogenica* was significantly different between control population and tobacco chewers as well as oral cancer patients. Increased abundance was observed in tobacco chewers and oral cancer patients (~1.19 X 10^6 and~ 1.28 X 10^6 BGE/ml respectively) BGE/ml, as compared to the control population (~3.2 X 10^5 BGE/ml). Similarly, in the case of *V. parvula*, significant increase in the abundance of bacteria was observed from control population (~ 1.25 X 10^5 BGE/ml) to tobacco chewers and patients with oral cancer (~ 3.31 X 10^5 and~ 2.70 X 10^5 BGE/ml). For *S. mutans*, the pattern observed was opposite, there was a significant decrease in absolute counts of *S. mutans* in tobacco chewing population (~ 6.37 X 10^5 BGE/ml) as compared to control population (~ 1.46 X 10^6 BGE/ml). A non-significant decrease was observed in absolute bacterial counts in patient with oral cancer as compared to control population.

For *H. parainfluenzae* and *R. mucilaginosa*, the absolute abundance of these bacteria in oral samples depicts a pattern, but no significant difference was observed. The pattern of abundance of bacteria mentioned in this study correlates well with earlier reports, thereby supporting the results obtained. The abundance of *H. parainfluenzae* was observed to be higher in control population as compared to OSCC patients, suggested by NGS data [15]. On the contrary, population of *P. melaninogenica* was reported higher in OSCC patients as compared to healthy individuals [13–16].

Increased abundance *P. gingivalis* was observed in Tobacco chewers and Oral cancer patients as compared to healthy controls [17–19]. Similarly, increased levels of *F. nucleatum* was found in oral cancer patients as compared to healthy individuals [20,21]. The increased abundance of *C. gingivalis, P. melaninogenica* and *V. parvula* in oral cancer patients as compared to healthy individuals was seen in present study, and has been previously reported [13–15]. Similarly increased abundance of *H. parainfluenzae* and *R. mucilaginosa* with healthy controls has been recorded [16,22].

Vast amount of information is published about the role and possible mechanism of action of periodontal pathogens such as *P. gingivalis* and *F. nucleatum* in causation of oral diseases, including oral cancer. The most common mechanisms of action of bacteria include chronic infection and inflammation, nitrosation of compounds and metabolism of potentially carcinogenic substances [16]. *P. gingivalis* is also known to be a keystone pathogen of disease provoking periodontal microbiota [23]. In-vitro cell line interactions have also shown a dysregulation caused to oral cells on exposure to *P. gingivalis*. Upregulation in expression of B7–H1 and B7-DC receptors involved majorly in supressing immune system in cancer have been detected on exposure of various oral cell lines to *P.gingivalis* [24,25]. Similarly, increased expression of MMPs was observed in oral cells after incubation with *P. gingivalis*. MMPs are matrix metalloproteinases which play a major role in metastasis [18]. *Fusobacterium nucleatum* is a gram-negative adhesive oral commensal, which turns into a pathogen under certain conditions. The most common adhesin/virulence factor studied is FadA, which is also used as a diagnostic marker for the detection of *F.nucleatum* and mediates direct invasion into host cells [26]. *F.nucleatum* is also known to stimulate high secretion of IL-8, which has been characterised in cancer cells [27].

Our results showed that the real-time PCR based assay can be used for absolute quantification of bacteria. Using *rpoB* gene as the target is better

for absolute quantification, as it is a single copy gene per cell, as compared to 16s rRNA, which has variable copy number per cell. Also, *rpoB* is a species-specific gene, thus leading to a specific amplification. By using this method, we have successfully calculated absolute abundance of eight different oral bacteria which have been reported to show variable abundance in control population as compared to oral cancer patient microbiome. To the best of our knowledge, this is the first report which involves absolute quantification of bacterial abundance in long term tobacco chewing population and patients with oral cancer using qPCR approach.

4. Conclusion

In conclusion, we have shown a SYBR green based real-time PCR assay for the absolute quantification of bacteria from different sample sources. The assay was used to quantitate eight oral bacteria in oral cavity rinse samples of Oral cancer patients, Long term tobacco chewers and controls. The assay shown here is highly sensitive, specific, reproducible, and consistent method to quantify bacterial population as well can be used as a simple method to understand the progression and prognosis of an individual towards cancer.

Credit author statement

Harinder Singh: Conceptualization, Methodology, Supervision, Writing-Reviewing and editing, Shriya Sawant: Methodology, Data curation, Writing – original draft, Jinesh Dugad: Methodology, Data curation, Writing-Reviewing and editing, Deepak Parikh: Methodology, Writing-Reviewing and editing.

Funding

This research was supported by NMIMS University seed grant (University Seed Grant 2016). This research did not receive specific grant from any other funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that there is no conflict of interests.

References

[1] Abbott MA, Poiesz BJ, Byrne BC, Kwok S, Sninsky JJ, Ehrlich GD, et al. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral dna amplified in vitro. J Infect Dis 1988;158:1158–69. https://doi.org/10.1093/infdis/158.6.1158.

[2] VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 2008. https://doi.org/10.2144/000112776.

[3] Nadkarni M, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad range (universal) probe and primer set. Microbiology 2002;148:257–66. https://doi.org/10.1128/JCM.40.5.1698.

[4] Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol 2009;67:6–20. https://doi.org/10.1111/j.1574-6941.2008.00629.x.

[5] Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol 2000;66:1328–33. https://doi.org/10.1128/AEM.66.4.1328-1333.2000.

[6] Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 2007;73:278–88. https://doi.org/10.1128/AEM.01177-06.

[7] Beighton D, Clark D, Hanakuka B, Gilbert S, Do T. The predominant cultivable Veillonella spp. of the tongue of healthy adults identified using *rpoB* sequencing. Oral Microbiol Immunol 2008;23:344–7. https://doi.org/10.1111/j.1399-302X.2007.00424.x.

[8] Mashima I, Kamaguchi A, Nakazawa F. The distribution and frequency of oral veillonella spp. in the Tongue biofilm of healthy young adults. Curr Microbiol 2011;63:403–7. https://doi.org/10.1007/s00284-011-9993-2.

[9] Park SN, Park JY, Kook JK. Development of Porphyromonas gingivalis-specific quantitative real-time PCR primers based on the nucleotide sequence of *rpoB*. J Microbiol 2011;49:315–9. https://doi.org/10.1007/s12275-011-1028-y.

[10] Lim Y, Totsika M, Morrison M, Punyadeera C. Oral microbiome: a new biomarker reservoir for oral and oropharyngeal cancers. Theranostics 2017;7:4313–21. https://doi.org/10.7150/thno.21804.

[11] Kerr AR. The oral microbiome and cancer. J Dent Hyg 2015;89(Suppl 1):20–3.

[12] Mason MR, Nagaraja HN, Camerlengo T, Joshi V, Kumar PS. Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. PloS One 2013;8:1–7. https://doi.org/10.1371/journal.pone.0077287.

[13] Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med 2005;3:27. https://doi.org/10.1186/1479-5876-3-27.

[14] Zhao H, Chu M, Huang Z, Yang X, Ran S, Hu B, et al. Variations in oral microbiota associated with oral cancer. Sci Rep 2017;7:1–10. https://doi.org/10.1038/s41598-017-11779-9.

[15] Al-hebshi NN, Nasher AT, Maryoud MY, Homeida HE, Chen T, Idris AM, et al. Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. Sci Rep 2017;7:1834. https://doi.org/10.1038/s41598-017-02079-3.

[16] Chocolatewala N, Chaturvedi P, Desale R. The role of bacteria in oral cancer. Indian J Med Paediatr Oncol 2010;31:126. https://doi.org/10.4103/0971-5851.76195.

[17] Wang H, Zhou H, Duan X, Jotwani R, Vuddaraju H, Liang S, et al. Porphyromonas gingivalis-induced reactive oxygen species activate JAK2 and regulate production of inflammatory cytokines through c-Jun. Infect Immun 2014;82:4118–26. https://doi.org/10.1128/IAI.02000-14.

[18] Ha NH, Park DG, Woo BH, Kim DJ, Choi J Il, Park BS, et al. Porphyromonas gingivalis increases the invasiveness of oral cancer cells by upregulating IL-8 and MMPs. Cytokine 2016;86:64–72. https://doi.org/10.1016/j.cyto.2016.07.013.

[19] Sayehmiri F, Sayehmiri K, Asadollahi K, Soroush S, Bogdanovic L, Azizi Jalilian F, et al. The prevalence rate of Porphyromonas gingivalis and its association with cancer: a systematic review and meta-analysis. Int J Immunopathol Pharmacol 2015;28:160–7. https://doi.org/10.1177/0394632015586144.

[20] Meurman JH. Oral microbiota and cancer. J Oral Microbiol 2010;2:5195. https://doi.org/10.3402/jom.v2i0.5195.

[21] Gholizadeh P, Eslami H, Kafil HS. Carcinogenesis mechanisms of Fusobacterium nucleatum. Biomed Pharmacother 2017;89:918–25. https://doi.org/10.1016/j.biopha.2017.02.102.

[22] Yang CY, Yeh YM, Yu HY, Chin CY, Hsu CW, Liu H, et al. Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. Front Microbiol 2018;9. https://doi.org/10.3389/fmicb.2018.00862.

[23] Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. Nat Rev Microbiol 2012;10:717–25. https://doi.org/10.1038/nrmicro2873.

[24] Groeger S, Domann E, Gonzales JR, Chakraborty T, Meyle J. B7-H1 and B7-DC receptors of oral squamous carcinoma cells are upregulated by Porphyromonas gingivalis. Immunobiology 2011;216:1302–10. https://doi.org/10.1016/j.imbio.2011.05.005.

[25] Groeger S, Jarzina F, Mamat U, Meyle J. Induction of B7-H1 receptor by bacterial cells fractions of Porphyromonas gingivalis on human oral epithelial cells. Immunobiology 2016;222:1–11. https://doi.org/10.1016/j.imbio.2016.10.011.

[26] Fardini Y, Wang X, Témoin S, Nithianantham S, Lee D, Shoham M, et al. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. Mol Microbiol 2011;82:1468–80. https://doi.org/10.1111/j.1365-2958.2011.07905.x.

[27] Han YW, Shi W, Huang GTJ, Kinder Haake S, Park NH, Kuramitsu H, et al. Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. Infect Immun 2000;68:3140–6. https://doi.org/10.1128/IAI.68.6.3140-3146.2000.

Synopsis

SYNOPSIS OF THE THESIS TO BE SUBMITTED
TO THE UNIVERSITY OF NMIMS (DEEMED-TO-BE-UNIVERSITY)
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (BIOLOGICAL SCIENCES)

NAME OF CANDIDATE:	**Shriya Rajendra Sawant**
DOCTORAL COMMITTEE:	**(i) Dr. Kayzad Nilgiriwala,** Research Officer, Foundation of Medical Research, Mumbai **(ii) Dr. Binoj Kutty,** Assistant Professor, St. Xavier's College, Mumbai
NAME AND DESIGNATION OF MENTOR:	**Dr. Harinder Singh** Assistant Professor, Sunandan Divatia School of Science, SVKM's NMIMS University, Vile Parle (W), Mumbai – 400 056
TITLE OF THESIS:	Bacterial diversity in oral cavity and its association with oral cancer
PLACE OF RESEARCH:	Sunandan Divatia School of Science, SVKM's NMIMS University, Vile Parle (W), Mumbai- 400 056
NUMBER AND DATE OF REGISTRATION:	75102150004/26th September 2017
DATE OF SUBMISSION OF SYNOPSIS:	14th July'2022
SIGNATURE OF THE CANDIDATE:	Sawant
SIGNATURE OF THE MENTOR:	Harindersingh

INDEX

Chapter 1: Introduction

In India, oral cancer (OC) is the second most frequent malignancy, and survival rates have not improved in recent decades. Oral cancer is the most prevalent cancer in men in India, with 104,661 cases reported. It is also the fourth most common cancer in women, with 31,268 cases reported, accounting for nearly 36% of the worldwide burden of oral cancer (Sung et al., 2021). The buccal mucosa is the most afflicted area, followed by the lateral surface of the tongue, base of the tongue, lips, and hard palate. Oral squamous cell carcinoma (OSCC) is the term used to describe a cancer that arises from the mucosal lining of the mouth which contributes to about 90% OC cases (Warnakulasuriya, 2018).

The high prevalence of OC in India is related to tobacco/areca nut chewing, alcohol intake, and Human Papillomavirus types HR-HPV16/18, with tobacco use beginning at the age of 912 years on average (Coelho, 2012). Tobacco is extensively used in Asian countries, both in smoking and non-smoking forms. Cigarettes, bidis, and hookah are all forms of smoking, with bidi smoking having the highest risk of OC when compared to cigarettes and hookah. In India, smokeless tobacco has been introduced in a variety of forms, including pan masala and gutka, which are tobacco, areca nut, lime, and catechu mixes that are highly carcinogenic and genotoxic (Chaturvedi et al. 2019). Alcohol usage, which includes the frequency of consumption, the kind of alcohol consumed, and the consumer's geographic location, all contribute to the risk of oral cancer. Tobacco chewing and smoking have a synergistic impact with alcohol towards progression to OC (Ghantous et al. 2018).

Another risk factor for OC is diet, which is influenced by lifestyle, socioeconomic level, and culture. OC is also connected to poor oral hygiene, leading to different infections and periodontal illnesses, which involves role of bacterial in causing inflammation and eventually instigate carcinogenesis (Rodríguez-Molinero et al. 2021). Recent findings in cancer research suggest a link between the presence of specific types of bacteria and the development of cancer in the human body, the reports are based on genome sequencing. Dysbiosis, or microbial imbalance, is characterised by a decrease of microbial diversity, the loss of helpful bacteria, and the increase of pathogenic germs. These factors can contribute to the development of cancer (Kakabadze et al. 2020).

In the current work, the oral microbiota of the Indian OC population and tobacco chewers was identified and compared to healthy individuals. From the microbiome data, we also identified

the microbial biomarkers for three studied groups. A simple qPCR-based method was developed that would aid in early diagnosis of OC in high-risk individuals. A preliminary investigation was performed to understand the host-pathogen interactions at the cellular and molecular level. The resulting data would help to distinguish the oral microbiota of Indians from that of the rest of the globe, as well as add to the global human oral microbiota knowledge pool. It will also improve our understanding of the complexities of bacterial pathogenesis in oral cancer.

Chapter 2: Review of Literature

2.1 Bacteria in Oral Cancer

The oral microbiota is composed of around 600 different bacterial species. The majority among these species, however, are uncultivable. The advent of modern sequencing technologies allows the discovery of bacterial populations that live in the mouth and are involved in human health (Dewhirst et al. 2010; Proctor et al. 2019). The most prevalent schemas proposed to describe the pathophysiology involving microbiota in the development of cancer (la Rosa et al. 2020) are:

i) Chronic inflammation and immune responses that can promote carcinogenesis and tumour growth
ii) Metabolic activity changes that lead to increased production of toxic and carcinogenic metabolites
iii) Alteration of epithelial barrier integrity and epigenetic modulations induced by microbiota.

Periodontal disorders and chronic inflammation are linked to anaerobic oral bacteria such as *Fusobacterium, Porphyromonas*, and *Prevotella* species. Inflammatory mediators released by bacteria have the ability to interact with cells in several tissues, causing dispersed inflammation. Expression of Interleukins (IL-1, IL-6, IL-17, IL-23), Tumour Necrosis Factor (TNF), and Matrix Metallo-Proteinases (MMP-8, -9 and -13) are all influenced by periodontal bacteria (Szkaradkiewicz and Karpiński 2013; Lamont et al. 2018). Sulphur compounds, acids, and free radicals, primarily nitric and oxygen reactive species, are produced by bacterial metabolism and can cause genetic damage. Moreover, certain bacteria have an alcoholic metabolism that produces acetaldehyde, which is necessary for neoplastic transformation (Karpiński 2019).

2.2 Oral Cancer Microbiome Studies

One of the earliest reports published on the oral microbiome and OC was by Nagy et al., and they analysed the biofilm present on the surface of OSCC tissue and healthy contagious mucosa. The study showed increased numbers of anaerobes such as *Veillonella, Fusobacterium, Prevotella, Porphyromonas, Actinomyces* and *Clostridium*, and aerobes such

as *Haemophilus, Serratia liquefaciens, Klebsiella pneumoniae Citrobacter freundii, Enterobacteriaceae* and *Streptococcus β-haemolyticus* (Nagy et al. 1998). In another pilot study reported by Pushalkar et al., microbiome analyses were performed using 16S rRNA sequencing of saliva samples collected from three OSCC cases and Denaturing Gradient Gel Electrophoresis (DGGE) was used to compare the profile with two matched controls. OSCC samples predominantly contained organisms from the genera *Streptococcus, Gemella, Rothia, Peptostreptococcus, Porphyromonas*, and *Lactobacillus* (Pushalkar et al. 2011). Later the researchers reported the oral microbiota from ten OSCC tumor samples using the same sequencing technique (Pushalkar et al. 2012). They found phylae Firmicutes, Class *Bacilli* consisting order *Lactobacillales* (54.8%) and *Bacillales* (11.8%) to be predominant in the tumor tissues. The tumor library was predominated by *Streptococcus* (50.8%), *Gemella* (11.6%), *Parvimonas* (4.6%), *Peptostreptococcus* (2.8%), *Xanthomona*s (2.4%), *Johnsonella* (1.6%), *Solobacterium* (1.6%), *Atopobium* (1.2%). *Eubacterium [11] [G-3], Campylobacter* and *Catonella* were exclusive to tumor library, which was a new panel of organisms to be associated with OSCC from the yet discussed studies, and required further validations. This may also suggest that the change in oral health of an individual can cause dysbiosis in the oral cavity that can turn a commensal bacterial population to pathogenic.

Another microbiological study, performed by Cankovic et al. (2013), was aimed to observe the difference in microflora of OSCC swab samples when compared to oral mucosa lesions. The microbiological assays depicted increased pathological flora and decreased normal flora in samples of the OSCC group, when compared to the control group, a common observation found in most of the studies reported (Čanković et al. 2013). Schmidt and colleagues studied the bacterial abundance in OC samples (Schmidt et al. 2014). In both the studies, presence of *Firmicutes* and *Actinobacteria* decreased in the OC samples as compared to healthy counterparts. Further, pre-cancerous samples also showed a similar decrease. Such studies suggest alterations in microbiota may occur very early on in cancer development during precancer stages. At genus level, significant reduction in the abundance of *Streptococcus* and *Rothia* and a corresponding increase in *Fusobacterium* was reported in OSCC samples compared to their matched controls. Thus, such organisms can be used as microbial biomarkers to study the progression of cancers.

Bolz et al. performed a microbiological study to isolate oral bacteria from oral cancer, highrisk individuals and control patients. The ratio of aerobes to anaerobes was found to be 1:1.6 in control samples, equal ratio (1:1) in high-risk patients whereas the tumor patients displayed a reversal of ratio (0.5:1). This depicts an increase in abundance of anaerobic bacteria in tumor samples (Bolz et al. 2014). This could be due to dysbiosis and physiological changes during tumorigenesis, which favours the dominance of anaerobes. A similar microbiological study reported an increased median CFU/ml in saliva and swab of the oral cavity obtained from OSCC patients as compared to healthy controls. Increased abundance of *Candida* spp, particularly *Candida albicans* was isolated from the carcinoma site (Metgud et al. 2014). OC samples had more bacterial species than healthy persons, according to Zhao et al. The most overexpressed bacteria in OSCC samples with past periodontitis are *Catonella, Dialister, Filifactor, Fusobacterium, Parvimonas, Peptococcus,* and *Peptostreptococcus*. The examination of *Fusobacterium* bacteria, in particular, has been considered as a diagnostic criterion for OSCC (Zhao et al. 2017). The number of Fusobacteria was considerably greater in OC patients who were progressing, especially in stage-IV patients (7.92%) compared to healthy controls (2.98%) (Yang et al. 2018). Lim et al (181) analysed oral rinse to evaluate the microbiome variations and the results indicated *Capnocytophaga gingivalis*, *Peptostreptococcus sp., Porphyromonas gingivalis, Prevotella sp*., and *Streptococcus* sp. as the oral microorganism mostly associated with OSCCs (Lim et al. 2018).

2.3 Effect of Tobacco on Oral Microflora

It is a well-established fact that use of tobacco (smoking or chewing) makes an individual more susceptible to chronic health conditions including pulmonary diseases, various cancers, periodontal diseases, stroke etc. However, use of tobacco also makes an individual more prone to bacterial infections. This occurs due to effect of tobacco, making the bacterial population in the organ affected relatively unstable and diverse as compared to healthy nontobacco using individuals (Bagaitkar et al. 2008). Comparison of the oral biofilms between tobacco users and non-users showed increased pathogenic bacterial species such as *Fusobacterium, Cardiobacterium, Synergistes, Selenomonas, Haemophilus* and *Pseudomonas* in the tobacco users. This depicts the use of tobacco favours early acquisition and colonization of pathogens in oral biofilms (Kumar et al. 2011).

There are three mechanisms by which tobacco can increase risk of infection by pathogenic/opportunistic bacteria:

i) Tobacco induced physiological and structural changes in humans

ii) Tobacco induced increase in bacterial virulence iii) Tobacco induced deregulation of immune function

Hence, the microbiota alteration due to habits such as tobacco chewing can also lead to progression towards carcinogenesis. The microbial diversity of the oral cavity in Indian population of long-term tobacco chewers and OC patients is unexplored. Microbiome research is still in its early stages. A lot of research is being done, and data is being uploaded on a regular basis. Studies with a bigger sample size and several sites in health and illness are required to be discovered to have consistent trends and provide actual results. This will help to find distinct biomarkers and aid in targeted medicines and customised medicine in clinical practise for improved patient management.

Chapter 3: Lacunae, Rationale, Aim and Objectives

3.1 Lacunae/Rationale

Oral cancer (OC) the most common cancer observed in Indian males. Risk factors of OC include major lifestyle factors such as tobacco chewing, alcohol consumption, smoking which are widely observed in the Indian and South Asian subcontinent. Apart from these, there has been a rise in the OC cases in individuals who do not follow the said lifestyle habits. This led to the curiosity for identification of newer risk factors associated with OC, where microbiota was found to be a potential risk factor. Several studies have demonstrated the necessity of a normal, healthy microbiota in supporting good health, and dysbiosis can lead to issues such as cancer. The human body's healthy and altered microbiota, including the oral microbiome and its role in disorders including periodontitis, dental caries, and oral cancer; is being studied extensively across the world, however there is a lack of understanding about the oral microbiome in the Indian population.

The Indian population offers a wide scope of individuals following wide variety of lifestyle factors including tobacco chewing, smoking, alcohol consumption, etc. Thereby the Indian population acts as a suitable demography to study the OC microbiome due to the high incidence rate as well as exposure to the risk factors. This would aid us in identifying the oral microbiome of Indian OC patients, tobacco chewing population and compare with the healthy individuals. The study would also help us identify microbial biomarkers, if any found in the Indian population for the study groups. This data can eventually help in identification of risk factor for diagnosis as well as treatment strategies. The biomarkers identified in the OC group can help in early diagnosis of individuals at high risk of developing OC, such as tobacco chewers/alcohol consumers/smokers. Along with identification, it is essential to unveil the mechanism of action of significant bacteria which causes gene dysregulation and thereby progression to oral cancer.

3.2 Aim and Objectives

3.2.1 Aim

To identify bacterial diversity in oral cavity and its association with Oral Cancer.

3.2.2 Objectives

1. To identify the bacterial diversity in oral cavity of oral cancer patients and tobacco chewers.
2. To find the correlation of oral bacterial diversity and oral cancer.
3. To study host-microbe interaction at molecular level.

Chapter 4: Identification of Bacteria in Oral Cavity of Oral Cancer Patients and Tobacco Chewers in Indian Population

This chapter deals with the 16S rRNA metagenomics study performed for identification of bacterial diversity in oral cavity in three study groups namely, healthy controls, long term tobacco chewers and oral cancer patients. Following sample collection from participants, DNA was isolated using silica column-based method for efficient isolation and 16S rRNA targeted Next Generation Sequencing was performed. The raw data obtained was subjected to analysis using QIIME2™ pipeline for identification and assessment of bacterial taxonomy.

4.1 Study subjects

The study subjects in the study were divided into three study groups consisting of 40 individuals in each group, namely histopathologically confirmed as patients suffering from Oral Cancer (OC), long term tobacco chewers (T) and healthy control population (C). Informed consent was obtained from the participants after explaining them about the study. Data on participant's medical history, age, gender, occupation, habits of tobacco chewing, alcohol intake, and general oral hygiene questions were recorded. Ethical approvals for the recruitment of healthy, long-term tobacco chewing volunteers and OC patients were obtained from SVKM's Institutional Ethics Committee (NMIMS/IEC/008/2016) and The Ethics Committee of K. J. Somaiya Medical College and Hospital, Mumbai.

4.2 Sample collection and DNA isolation

Patients rinsed their mouths with 20 ml of sterile saline for 30 s and collected into a 50 ml sterile tube (Al-Hebshi et al. 2019; la Rosa et al. 2020). Salivary samples from patients were taken at the pre-operative stage while the patient was NBM (Nil by Mouth). Subjects from healthy and tobacco groups were asked to refrain from eating, drinking, or oral hygiene procedures for at least 1 hr before sample collection. Salivary samples were collected in welllabelled sterile falcon tubes and stored at 4 °C and processed within 48 hrs. The DNA was isolated using Invitrogen DNA Isolation Kit and quantified using BioTek EPOCH2 microplate reader. The quality of the isolated DNA was assessed by performing agarose gel electrophoresis using 1% agarose.

4.3 Next Generation Sequencing using Illumina Miseq

The NGS was performed at Seoul Women's University, South Korea with the collaborative help of Dr. Sathiyaraj Srinivasan. The PCR amplification of bacterial 16S rRNA hypervariable region V6-V8 was carried out using primers B969F (ACGCGH NRA ACC TTA CC) and BA1406R (ACG GGC RGTGWG TRC AA). The whole sequencing process was done using Illumina (Illumina, San Diego, CA). MiSeq libraries were quantified and then subjected to 300nucleotide paired-end multiplex sequencing(Lalande et al. 2014; Guo et al. 2016).

A total of 6,296,186 forward and reverse sequences of 300-bp were obtained with a mean of 56,219 and median of 51,278 sequences per sample.

4.4 Taxonomic assessment using QIIME2™

Demultiplexed raw reads (V6-V8) from Illumina sequencer were obtained and imported into Quantitative Insights into Microbial Ecology 2 (QIIME™2) pipeline for further analysis (Estaki et al. 2020). The quality of the reads from the sequencer were assessed using FASTQC followed by filtering out low quality reads by denoising using Deblur and assignment into Operational Taxonomic Units (OTUs). Taxonomy was assigned using the GreenGenes database at different taxonomic levels and compared between the study groups. The differential abundance of bacterial genera was analysed using LEfSe (https:// huttenhower. sph. harvard. edu/ galaxy/)(Wang et al. 2019). The Lda Effect Size was analysed using Kruskal Wallis test (threshold=0.05) and Pairwise Wilcoxon test (threshold=0.05).

The reads were classified into 6,733 OTU's using Greengenes database which resulted in identification of 9 phyla, 17 classes, 30 orders, 55 families, 94 genera, and 148 species. The top five phyla observed were Firmicutes, Proteobacteria, Bacteroidetes, Fusobacteria, and Actinobacteria. The ten most abundant genera observed in all groups were *Streptococcus, Prevotella, Neisseria, Rothia, Haemophilus, Porphyromonas, Paraprevotella, Veillonella, Fusobacterium* and *Leptotrichia.* Of these, *Prevotella* was the most dominant genus in all 3 study groups followed by *Neisseria* and *Streptococcus*. The abundance of *Streptococcus, Neisseria, Rothia* and *Veillonella* was higher in control population as compared to OC patients, whereas abundance of *Prevotella, Porphyromonas* and *Fusobacterium* was higher in OC patients as compared to healthy individuals. 11 genera for tobacco chewing population, 7

genera for OC and nine genera for healthy individuals were identified as microbial biomarkers after performing LEfSe analysis.

4.5 Alpha diversity analysis

Bacterial richness and diversity were analysed by alpha and beta diversity matrices and indices such as Observed OTUs, Shannon index, Simpson index, Chao index, ACE index and Rarefaction curves. The results have been expressed as mean ± SD. Statistical analysis for differences in mean values was performed using one way Analysis of Variance (ANOVA) followed by Tukey's multiple comparisons test using GraphPad Prism software (version 8.0.2). p-value < 0.05 was considered statistically significant.

The alpha rarefaction curves depict adequate sampling to perform further analysis. Alpha diversity indices portray highest alpha diversity in tobacco group followed by control group and lowest in OC group. Statistically significant differences were observed between control and OC and tobacco and OC group by alpha diversity indices.

4.6 Beta diversity analysis

To analyse the beta diversity, PERMANOVA, ANOSIM and PERMDISP was performed using QIIME2™ pipeline. Bray Curtis matrix, Jaccard matrix, Weighted and Unweighted Unifrac distance matrices were plotted to assess the distribution of sample compositions in 3D space.

PERMANOVA and ANOSIM demonstrate significant beta diversity between all three study groups suggesting difference in the centroids of the groups. Further, PERMDISP suggests significant difference between control-OC and tobacco-OC groups taking into account the dispersion between the groups. PCA plots show clustering of control and tobacco samples whereas OC samples stretch out away from the C-T cluster. These analyses confirm the bacterial composition of control population is similar to tobacco chewers, whereas that of OC patients is varied from control and tobacco chewers.

4.7 Functional prediction of metabolic pathways

The functional MetaCyc pathways were predicted using PICRUSt2. The results have been expressed as a post-hoc plot using STAMP software. Statistical analysis for significance was

performed using students t-test between groups was followed by Bonferroni correction (Zhang et al. 2020). p value < 0.05 was considered statistically significant.

Pathways related to amino acid synthesis, glucose and sugar utilization were upregulated in control samples as compared to OC group. On the contrary, pathways related to lipid, fatty acid and coenzyme A synthesis were upregulated in OC group. Reductive TCA cycle and pyrimidine biosynthesis pathway were upregulated in tobacco chewers.

Chapter 5: Absolute Quantitation of Oral Bacteria Involved in Oral Cancer and Tobacco Chewers by Real-Time PCR

In this chapter we have exploited quantitative PCR which is an adaptation of PCR, widely used for quantitation of gene expression at mRNA levels as well as for microbial ecology purposes to determine the microbial load and diversity in any environmental sample (Abbott et al. 1988; Nadkarni et al. 2002). Microbial qPCR methods are intended to target highly conserved 16S rRNA genes to measure overall bacterial load in a sample, whereas hypervariable regions of this gene can be targeted for selective amplification and quantification of a specific bacterial population (Smith and Osborn 2009). Because every bacterial species contains varied copies (1–15 copies) of 16S rRNA, this technique has a flaw that must be noted, *i.e.* 16S rRNA gene copies cannot be correctly associated with cell counts (Klappenbach et al. 2000). A housekeeping gene such as RNA Polymerase-subunit gene *(rpoB)*, can be used instead of the 16S rRNA gene. Therefore, we aimed to develop a reliable method for absolute quantification of bacteria using the *rpoB* gene(Case et al. 2007). Assessing the dynamics of the microbial population in diseased conditions can help in determining the mechanisms, as well as abundance of bacteria, which can trigger or develop the cancer.

5.1 Selection of candidate organism and primer designing

Porphyromonas gingivalis, Fusobacterium nucleatum, Capnocytophaga gingivalis, Haemophilus parainfluenzae, Prevotella melaninogenica, Rothia mucilaginosa, Veillonella parvula, and *Streptococcus mutans* were chosen, based on previous reports on oral cavity bacterial species linked to dysbiosis in OC and *rpoB* gene specific primers were used for absolute quantitation(Gholizadeh et al. 2016; Yang et al. 2018). *rpoB* primers were designed against *E. coli* MG1655 as well which was chosen as model organism.

5.2 Establishing standard curve for *E. coli* MG1655

Relation between Ct and CFU/ml was established using *E. coli* as model organism. In brief, 0.1OD culture of *E. coli* was used to obtain CFU/ml using spread plate technique. Parallelly, the culture was used for DNA isolation followed by quantitation and quality assessment of the DNA isolated. The isolated DNA was then serially diluted and subjected to qPCR using SYBR

green chemistry to obtain Ct values, which were used to plot standard curve and equation obtained was used for all further calculations.

The correlation between OD, CFU/mL and DNA concentration (ng/µL) was established where 0.1OD @ 600 nm corresponded to 3.856 (±1.5) X 107 CFU/mL and the DNA isolated from this culture was 3.560 ng/µL. DNA concentration of 0.1OD culture was noted and 6-10-fold dilutions were used to construct the standard curves for *rpoB* gene amplification. The final equation of standard graph obtained was used for all further calculations where, Y is the Ct value for a sample and X is the log of DNA concentration. The abundance is calculated in terms of BGE/ml (Bacterial genomic equivalent/ml)(Sundin et al. 2017).

5.3 PCR analysis for significant bacteria in Oral Cancer

For absolute quantification, primers designed for *rpoB* gene of *P. gingivalis, F. nucleatum, C. gingivalis, H. parainfluenzae, P. melaninogenica, R. mucilaginosa, V. parvula* and *S. mutans* were used. For the PCR, reactions were set up for a total volume of 10 µL followed by melt curve analysis to confirm that only the specific products were amplified. The amplified products were checked on 2% agarose gel for the size of amplified products. All the studies were carried out thrice in triplicates; the results have been expressed as mean ± standard deviation (SD). Statistical analysis for differences in mean values was performed using one way Analysis of Variance (ANOVA) followed by Tukey's multiple comparisons test using GraphPad Prism software (version 8.0.2). p value < 0.05 was considered statistically significant.

Significant increase in abundance of *P. gingivalis, F. nucleatum, C. gingivalis, P. melaninogenica* and *V. parvula* in OC patients as compared to healthy individuals. Whereas *H. parainfluenzae, R. mucilaginosa* and *S. mutans* showed higher abundance in healthy controls as compared to OC patients. Variable factors such as age, gender, food, site of cancer, stages of cancer were also assessed.

Chapter 6: Assessment of Host mRNA Expression during HostMicrobe Interaction

This chapter discusses preliminary study performed to assess the host mRNA expression on infection with *S. mutans*, which is a known resident of the oral cavity. Previous reports have demonstrated the role of chronic inflammation and immunological responses in promoting carcinogenesis and tumour growth due to pathogenic bacteria such as *P. gingivalis* and *F. nucleatum* in the development of oral cancer. These studies have not been conducted on the many bacterial species found in the oral cavity, which are essential for optimal oral health. *S. mutans* is an essential organism in maintaining healthy oral cavity, but hardly explored for its effect on expression of genes linked with cancer development.

6.1 Oral Cancer Cell Line procurement and maintenance

The oral cancer cell line (AW13516) used in this study was procured from Tata Memorial Centre- Advanced Centre for Treatment, Research and Education (ACTREC), Navi Mumbai. The cell line was maintained in DMEM containing antibiotics supplemented with 10% FBS. For cryopreservation purposes, the freezing mixture was prepared with FBS: DMSO (9:1).

6.2 Growth and maintenance of *S. mutans*

S. mutans was cultured and maintained in St. Brain Heart Infusion (BHI broth). The culture was freshly grown at 37ºC under anaerobic conditions using anaerobic jar for 48hrs. The culture density was standardized at 0.35OD @ 600nm resulting in 10^8 cells/ml. This culture density and 1:10 serial dilutions are used for infection purposes**.**

6.3 Cytotoxicity assessment for standardization of MOI and time of infection

The extent of cytotoxicity due to bacterial infection was evaluated by MTT assay. After 24hrs of cell seeding, the host cells were infected with *S. mutans* at variable MOI (1,10,100,1000) for variable time periods (0hr.1hr, 3hr, 5hr, 24hr). After infection, at specified time points, 20ul 5mg/ml MTT was added to all the wells and incubated for 2hrs at 37ºC. After incubation, the media from all wells was removed completely and 100µl DMSO was added and the optical density on ELISA microplate reader at 570nm and 650nm (Aggarwal et al. 2019).

The results of cytotoxicity assay showed ~50% cell death at 24hrs infection at all MOIs, but no significant difference was observed at other MOI and times of infection, suggesting the bacterial interaction doesn't hamper host cell viability up to 5hrs of infection.

6.4 Microscopic visualization of infection

Sterile coverslips were placed in each well of 6-well plate for culturing. Host cells were infected with *S. mutans* at MOI of 100 followed by incubation for 1hr, 3hr, 5hr,7hr and 24hrs at 37°C. After the set time of infection was complete, the coverslips were washed in PBS followed by 10% formaldehyde treatment and staining with methylene blue(Brijesh et al. 2006).

The microscopic observation revealed no interaction between host cells and *S. mutans* at 1hr after infection. At 3hrs, the bacteria adhered to the host cell, at 5hrs and 7hrs, the bacteria started to form microcolonies and their density increased. At 24hrs, the host cell death was observed attributed to infection of increased population of *S. mutans,* suggesting the expected interaction between host cells and bacteria must happen between 1hr to 7hrs of infection.

6.5 Evaluation of mRNA expression post infection with *S. mutans*

We next examined the level of mRNA expression in oral cell line after infection with *S. mutans*. We aimed to target at least one gene belonging to pathways of the six-hall marks of cancer. GAPDH was selected as the housekeeping gene for expression analysis. All the studies were carried out thrice in triplicates; the results have been expressed as mean ± standard deviation (SD). Statistical analysis for differences in mean values was performed using one way Analysis of Variance (ANOVA) followed by Tukey's multiple comparisons test using GraphPad Prism software (version 8.0.2). p value < 0.05 was considered statistically significant.

mRNA upregulation was observed for Stathmin, CyclinD1, BCL2 whereas downregulation was observed in FGF5 with increased infection duration. IL1b, MMP9, Bax and hIL6 did not show any difference in their expression as the time of infection progressed. This suggests an activation of pathway associated with carcinogenesis.

Chapter 7: Discussion

Microbiome is defined as the collection of microbes present in a given sample, location or ecosystem. The microbiome studies have seen huge progress in the last decade. A famous example is the Human Microbiome Project (HMP), initiated by the National Institutes of Health (NIH) to identify the complete healthy human microbiome in order to appreciate the diversity and complexity of the microbial communities. Similar research initiatives are taken by various countries across the globe like Australia, Canada, France, Germany, Japan, Korea, and India (Dubilier et al. 2015). With the help of newer next generation sequencing technologies, the role and composition of microbiota has been identified in individuals in health and disease, for example cancer.

The oral microbiota in healthy individuals and OC patients has been deciphered in many countries. A few genera predominate the healthy oral microbiota such as *Streptococcus, Gemella, Parvimonas, Peptostreptococcus, Neisseria, Rothia*, *Veillonella*, *Haemophilus, Treponema, Lactobacterium, Eikenella, Leptotrichia, Capnocytophaga, Selenomonas, Granulicatella, Prevotella.* These microbes are normal commensal and generally beneficial, however, some of them can become pathogenic and can aid in progression of healthy cells towards OC under compromised conditions (Aas et al. 2005; Bik et al. 2010; Pushalkar et al. 2012).

In present study, we unearthed the composition of oral microbiota in the healthy Indian population and compared it with the composition of OC patients and long-term tobacco chewers. Indian population belongs to varied demographics and therefore can act as an ideal study population. OC is the most common cancer in Indian males and therefore study of the role of microbiota in Indian OC patients is crucial. Along with this, the use of tobacco in various forms in the Indian population is known to contribute towards high incidence of oral cancer. Therefore, exploring the potential of oral microbiota in tobacco chewing population and their role in OC is of prime importance.

Similar to various reports, we identified Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria as the most common phyla present in the oral cavity of all individuals (Sampaio-Maia et al. 2016; Al-hebshi et al. 2017). The abundance of gram-negative phyla *i.e.,* Bacteroidetes and Proteobacteria were enriched in OC and tobacco chewing

microbiota as compared to gram positive enriched phyla *i.e.,* Firmicutes and Actinobacteria in healthy individuals. The increased abundance of gram negative bacteria in OC has been reported earlier (Bolz et al. 2014). At genus level*, Streptococcus, Prevotella, Neisseria, Rothia, Haemophilus, Porphyromonas, Veillonella, Fusobacterium, Leptotrichia* and *Granulicatella* were the top 10 genera observed contributing to approx. 80% of total bacterial composition. Of these, the abundance of *Streptococcus, Neisseria, Rothia, Veillonella* and *Leptotrichia* was higher in healthy controls. Being commensals, these organisms are known to dominate a healthy oral cavity and maintain healthy functioning (Zaura et al. 2009). On the other hand, increased proportion of bacteria belonging to genera *Prevotella, Haemophilus, Fusobacterium and Porphyromonas* in tobacco chewers and OC patients can be attributed to dysbiosis due to unhealthy oral cavity leading to increased pathogenic bacteria thereby contributing towards the progression towards cancer. The mechanisms of pathogen such as *F. nucleatum* and *P. gingivalis* in OC progression is well known and therefore their increased abundance in unhealthy oral cavity is palpable. *P. gingivalis* is known to increase invasiveness of OC cells by upregulating MMPs and IL8 as well as initiates mesenchymal transition in gingival epithelial cells through ZEB1 (Ha et al. 2016; Sztukowska et al. 2016). On the other hand*, F. nucleatum* uses its adhesin FadA as well as upregulates IL-8 expressions in oral epithelial cells (Gholizadeh et al. 2017).

Although these organisms show significant differences in their abundance in all three study groups, not all could be called as biomarkers when we analyse the entire bacterial composition. Using LEfSe analysis, which along with statistical significance takes into account biological consistency, microbial biomarkers for the study groups were identified. In the tobacco chewing study group, *Leptotrichia, Treponema, Tannerella, Filifactor, Selenomonas* and *Campylobacter* were identified as biomarkers, which are known organisms abundantly known to be present in microbiota of tobacco chewers or OC (Wu et al. 2016; Amer et al. 2017). *Tannerella* and *Treponema* are a part of the red complex which also leads to initiation of periodontitis (Bodet et al. 2006). *Cardiobacterium* was newly identified biomarker for tobacco chewers in our study population.

Similarly, for OC population, *Capnocytophaga, Mycoplasma, Peptostreptococcus* which are identified as biomarker in our study population, are well known bacteria associated with OC (Bik et al. 2010; Mok et al. 2017). Our study population also reports the presence of *Pseudomonas* as a biomarker for OC which has been reported only once before (Al-hebshi et

al. 2017). *Paludibacter* and *Bifidobacterium* have also been identified as a microbial biomarker in our oral cancer study population for the first time, although *Bifidobacterium* has been used to deliver anti-tumor vaccines using various novel techniques (Shirakawa and Kitagawa 2018). Conversely, the genera observed in higher abundance in control population, namely *Neisseria, Veillonella, Rothia* and *Actinobacillus* along with *Corynebacterium* were identified as biomarkers for control population. The higher abundance of these bacteria in healthy oral cavity is well known and reported. In addition, cladogram obtained confirms the presence of greater number of gram-negative genera and phyla as biomarker for oral cancer as compared to gram-positive for control and tobacco chewing population (Bolz et al. 2014).

The assessment of various diversity indices helped us understand within-group (alpha) and between-group (beta) diversity. Various alpha indices were assessed such as ACE and Chao1 to evaluate species richness, Observed OTU's for evaluation of number of OTUs, Goods coverage for evaluation of coverage of sequencing, Pielou_e to understand evenness index and Shannon and Simpson to evaluate diversity within group. All the alpha diversity indices depicted highest diversity within tobacco group, followed by control group and lowest diversity within OC group. On the other hand, beta diversity analysis was performed using PERMANOVA, PERMDISP and ANOSIM. The three indices concluded significant beta diversity between control and OC group and tobacco and OC group. PERMANOVA demonstrated significant beta diversity between control and tobacco group as well, but PERMDISP did not concur with the result, suggesting the centroids of the two groups are significantly different, but the dispersion caused due to samples makes them biologically non-significant and hence not as diverse. The PCA Jaccard, Bray Curtis, and Unifrac plots also display the same results where tobacco and control samples cluster together whereas the OC samples are away from this cluster. This suggests the diversity between control and tobacco microbiota is similar whereas that of OC group is diverse than the other two groups.

The functional prediction depicted enriched pathways related to amino acid biosynthesis and sugar fermentation in healthy controls, as has been reported earlier (Al-Hebshi et al. 2019). On the other hand, fatty acid and lipid biosynthesis pathways were enriched in diseased samples (Liu et al. 2012). The tobacco group demonstrated enriched pathways related to reductive TCA cycle.

The abundances of bacteria in the three study groups, the biomarkers, alpha and beta diversity suggest significant diversity between control and OC group, whereas the Tobacco group lies in between control and OC group, suggesting a transition of microbiota from healthy control to OC through tobacco group.

Simultaneously, we developed a new method for absolute quantification of oral bacteria, to eliminate the drawbacks of 16S rRNA in quantitation approach. The 16S rRNA gene is widely acceptable for all microbial ecology purposes due to it conserved and variable regions, which help distinguish bacteria up to species level. One major drawback of this gene is presence of varying number of multiple copies throughout the microbial kingdom which leads to inaccuracy while enumeration (Buchan and Ledeboer 2014). The *rpoB* gene was an alternative to 16S rRNA, with similar capability to distinguish bacteria up to species level, but also being a single copy gene, aids in accurate enumeration. The housekeeping function ensures presence in every bacterial cell therefore acting as an ideal candidate (Case et al. 2007).

Increased abundance *P. gingivalis* was observed in Tobacco chewers and Oral cancer patients as compared to healthy controls (Wang et al. 2014; Ha et al. 2016). Similarly, increased levels of *F. nucleatum* was found in oral cancer patients as compared to healthy individuals. The increased abundance of *C. gingivalis, P. melaninogenica* and *V. parvula* in oral cancer patients as compared to healthy individuals was seen in present study, and has been previously reported (Mager et al. 2005; Al-hebshi et al. 2017; Zhao et al. 2017). Similarly increased abundance of *H. parainfluenzae* and *R. mucilaginosa* with healthy controls has been recorded (Chocolatewala et al. 2010; Yang et al. 2018).

Vast amount of information is published about the role and possible mechanism of action of periodontal pathogens such as *P. gingivalis* and *F. nucleatum* in causation of oral diseases, including oral cancer. The most common mechanisms of action of bacteria include chronic infection and inflammation, nitrosation of compounds and metabolism of potentially carcinogenic substances (Chocolatewala et al. 2010). *In-vitro* cell line interactions have also shown a dysregulation caused to oral cells on exposure to *P. gingivalis*. Upregulation in expression of B7–H1 and B7-DC receptors involved majorly in supressing immune system in cancer have been detected on exposure of various oral cell lines to *P.gingivalis* (Groeger et al. 2016, 2017). *Fusobacterium nucleatum* is a gram-negative adhesive oral commensal, which turns into a pathogen under certain conditions. The most common adhesin/virulence factor

studied is FadA, which is also used as a diagnostic marker for the detection of *F. nucleatum* and mediates direct invasion into host cells (Fardini et al. 2011). *F. nucleatum* is also known to stimulate high secretion of IL-8, which has been characterised in cancer cells (Han et al. 2000).

For *H. parainfluenzae* and *R. mucilaginosa*, the absolute abundance of these bacteria in oral samples depicts a pattern, but no significant difference was observed. The pattern of abundance of bacteria mentioned in this study correlates well with earlier reports, thereby supporting the results obtained. The abundance of *H. parainfluenzae* was observed to be higher in control population as compared to OSCC patients, suggested by NGS data (Perera et al. 2018). On the contrary, population of *P. melaninogenica* was reported higher in OSCC patients as compared to healthy individuals (Mager et al. 2005; Zhao et al. 2017). Our results showed that the real-time PCR based assay can be used for absolute quantification of bacteria. Using *rpoB* gene as the target is better for absolute quantification, as it is a single copy gene per cell, as compared to 16s rRNA, which has variable copy number per cell. In addition, *rpoB* is a species-specific gene, thus leading to a specific amplification. By using this method, we have successfully calculated absolute abundance of eight different oral bacteria, which have been reported to show variable abundance in control population as compared to oral cancer patient microbiota.

As per literature, *S. mutans* is an important bacterium of the healthy oral cavity, yet it is associated with dental caries or in individuals using dental prosthesis. It exploits three virulence factors related with carcinogenicity once well established in dental cavities or surrounding the dental prosthesis. The first factor is synthesis glycans, ability to become more acid tolerant, and production of lactic acid (Alanazi et al. 2018). Glycans influence host cell–cell adhesion via changing the amount of functional E-cadherin at the cell–cell barrier, facilitating cancer cell separation and invasion. Various investigations have found that cancer patients have altered N-glycosylation of cell surface glycoproteins and that complex N-glycans are prevalent during tumour growth (Pinho and Reis 2015). We started out to investigate the altered gene expression in oral cell line after infection with *S. mutans*. Cytotoxicity assay depicted approx. 100% host cell viability after infection with *S. mutans* for up to 5hrs. 24hrs infected cells depicted up to 50% cell death. Previous reports studying host-microbe interaction have reported similar cell death at longer incubation periods and therefore smaller incubation periods with MOI 100 are often chosen to study gene expression (Cho et

al. 2014; Johnson et al. 2015). Further, microscopic examination demonstrated interaction of *S. mutans* with host cells only after 3hrs of infection followed by microcolony formation depicting active replication of bacterial cells. Similar to cytotoxicity assay, 24hrs infection depicted host cells undergoing apoptosis/necrosis. Further gene expression study depicted significant upregulation of Stathmin, CyclinD1 and BCl2, genes related to invasion, proliferation and survival respectively. These genes are intermediates of pathways associated with one of the hallmarks of cancer (Hanahan and Weinberg 2011). Thereby, in this preliminary study, we suggest the activation of pathways involved in cancer due to the infection with *S mutans*.

Chapter 8: Summary and Conclusion

In summary, the Indian oral microbiota was identified using 16S rRNA metagenomics approach using Illumina Mi-Seq Sequencing followed by analysis using QIIME2™ pipeline. Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria were the major phyla observed in all study subjects. A total of 16 genera contributed up to 85% of total bacterial diversity of which *Streptococcus, Neisseria, Rothia* and *Veillonella* depicted higher abundance in control population, whereas *Prevotella, Fusobacterium* and *Porphyromonas* displayed higher abundance in OC patients. Microbial biomarkers were identified for each study group using LEfSe analysis. On assessing the alpha diversity in study groups, maximum alpha diversity was observed in tobacco chewers, followed by control and least in OC patients. On the other hand, beta diversity demonstrated maximum diversity between control and OC as well as tobacco and OC study groups, whereas diversity between control and tobacco study groups was lower and these groups were similar to each other. Following diversity analysis, the functionally differential pathways expressed were predicted using PICRUSt2, which suggested Pathways related to amino acid synthesis were comparatively overexpressed in control population, those related to reductive TCA overexpressed in tobacco groups, whereas pathways for lipid biosynthesis and fatty acid elongation were overexpressed in OC population.

Further, to overcome the drawback of use of 16S rRNA in microbial ecology for enumeration purpose, *i.e.,* presence of multiple copies of 16S rRNA in every bacterial cell, we developed a simple, cost effective, sensitive and specific method using single copy *rpoB* gene for absolute enumeration of bacteria in oral samples. Using this method, we performed enumeration of eight significant bacteria involved in OC in our samples. The abundance of each bacteria in all three study groups correlated well with the reports published earlier, thereby providing confidence in the method reliability. For the enumeration of bacteria, additional characteristics such as sex, age, dietary habits, cancer stage, and cancer locations were also evaluated.

Further, to explore host-bacterial interaction *in-vitro*, we infected OC cell line with a known oral bacterium, *S. mutans*. Microscopic visualization depicted interaction/adherence between oral cells and bacteria between 1hr to 7hrs. mRNA expressions depicted significantly

increased expression of Stathmin, CyclinD1, BCl2, and Bax:Bcl2 ratio and decreased expression of FGF5 which gave an indication towards involvement of cancer related pathways.

Hence, our study helped identifying oral bacterial diversity and microbial biomarkers in OC and tobacco chewers, which can be further validated before implementing in clinical settings. Also, the role of genes identified to be upregulated in oral cell line on infection with *S. mutans* can be elucidated and confirmed by protein expression studies.

Chapter 9: Significance of the Study

Our study brings to light the composition of oral microbiota of healthy, long-term tobacco chewers and OC patients in the Indian population. With the ever-increasing cases of OC and related mortality in India and the unveiling of role of microbiota in oral cancer, our study would help in understanding the diversity in the oral microbiota and aid in comparing it with the global reports. The microbial biomarkers identified along with the qPCR method developed would also aid in early diagnosis and prediction of individuals at high risk of developing oral cancer, especially long-term tobacco chewers. At the same time, the preliminary assessment of bacterial infection in the oral cell line opens new doors to explore the mechanism by which bacteria can promote tumorigenesis.

Publications and Presentations

Publications

1. **Sawant S**, Dugad J, Parikh D, Srinivasan S**, Singh H**. (2021) Identification & correlation of bacterial diversity in oral cancer and long- term tobacco chewers- A case- control pilot study. *Journal of Medical Microbiology*. 70(9), 001417. (**IF:2.35)**
2. **Sawant S,** Dugad J, Parikh D, **Singh H.** (2021) Absolute quantitation of oral bacteria involved in oral cancer by real-time PCR. *Medicine in Microecology*. 7, 100034.
3. Kamble A, **Sawant S**, **Singh H**. (2020) 16S ribosomal RNA gene-based metagenomics: A review. *Biomedical Research Journal* 7.1: 5.

Conferences and poster presentations

1. The **60th Annual Conference of Association of Microbiologists of India (AMI-2019)** and International Symposium on "Microbial Technologies in Sustainable Development of Energy, Environment, Agriculture and Health", Haryana, India (15th-18th November'19): presented poster entitled "Absolute quantitation of oral bacteria involved in oral cancer by real-time PCR"
2. **18th Asia Pacific Congress of Clinical Microbiology and Infection, Singapore (11th13th November'21)** presented poster entitled "Identifying oral microbiome diversity of Indian tobacco chewing population and patients with oral cancer using next generation sequencing."
3. Attended "Advances in Materials Science & Applied Biology" 2019, NMIMS, Mumbai.

Scholarship

1. Women's Graduate Union Scholarship 2020-21

References

1. Aas, J. a, Paster, B.J., Stokes, L.N., Olsen, I., and Dewhirst, F.E. 2005. Defining the Normal Bacterial Flora of the Oral Cavity Defining the Normal Bacterial Flora of the Oral Cavity. J. Clin. Microbiol. **43**(11): 5721–5732. doi:10.1128/JCM.43.11.5721.
2. Abbott, M.A., Poiesz, B.J., Byrne, B.C., Kwok, S., Sninsky, J.J., Ehrlich, G.D., and Ehrlich, G.D. 1988. Enzymatic gene amplification: Qualitative and quantitative methods for detecting proviral dna amplified in vitro. J. Infect. Dis. **158**(6): 1158–1169. Oxford Academic. doi:10.1093/infdis/158.6.1158.
3. Aggarwal, S., John, S., Sapra, L., Sharma, S.C., and Das, S.N. 2019. Targeted disruption of PI3K/Akt/mTOR signaling pathway, via PI3K inhibitors, promotes growth inhibitory effects in oral cancer cells. Cancer Chemother. Pharmacol. **83**(3): 451–461. Springer Berlin Heidelberg. doi:10.1007/s00280-018-3746-x.
4. Al-Hebshi, N.N., Borgnakke, W.S., and Johnson, N.W. 2019. The Microbiome of Oral Squamous Cell Carcinomas: a Functional Perspective. Curr. Oral Heal. Reports **6**(2): 145–160. Current Oral Health Reports. doi:10.1007/s40496-019-0215-5.
5. Al-hebshi, N.N., Nasher, A.T., Maryoud, M.Y., Homeida, H.E., Chen, T., Idris, A.M., and Johnson, N.W. 2017. Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. Sci. Rep. **7**(1): 1834. Springer US. doi:10.1038/s41598-017-02079-3.
6. Alanazi, S.A.S., Alduaiji, K.T.A., Shetty, B., Alrashedi, H.A., Acharya, B.L.G., Vellappally, S., and Divakar, D.D. 2018. Pathogenic features of Streptococcus mutans isolated from dental prosthesis patients and diagnosed cancer patients with dental prosthesis. Microb. Pathog. **116**(December 2017): 356–361. doi:10.1016/j.micpath.2018.01.037.
7. Amer, A., Galvin, S., Healy, C.M., and Moran, G.P. 2017. The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species. Front. Microbiol. **8**(DEC): 2391. Frontiers Media S.A. doi:10.3389/fmicb.2017.02391.
8. Bagaitkar, J., Demuth, D.R., and Scott, D. a. 2008. Tobacco use increases susceptibility to bacterial infection. Tob. Induc. Dis. **4**: 12. doi:10.1186/1617-9625-4-12.
9. Bik, E.M., Long, C.D., Armitage, G.C., Loomer, P., Emerson, J., Mongodin, E.F., Nelson, K.E., Gill, S.R., Fraser-Liggett, C.M., and Relman, D.A. 2010. Bacterial diversity in the oral cavity of 10 healthy individuals. ISME J. **4**(8): 962–974. Nature Publishing Group. doi:10.1038/ismej.2010.30.
10. Bodet, C., Chandad, F., and Grenier, D. 2006. [Pathogenic potential of Porphyromonas gingivalis, Treponema denticola and Tannerella forsythia, the red bacterial complex associated with periodontitis]. Pathol. Biol. (Paris). **55**(3–4): 154–162. doi:10.1016/J.PATBIO.2006.07.045.
11. Bolz, J., Dosá, E., Schubert, J., and Eckert, A.W. 2014. Bacterial colonization of microbial biofilms in oral squamous cell carcinoma. Clin. Oral Investig. **18**(2): 409–414. doi:10.1007/s00784-013-1007-2.
12. Brijesh, S., Daswani, P.G., Tetali, P., Rojatkar, S.R., Antia, N.H., and Birdi, T.J. 2006. Studies on Pongamia pinnata (L.) Pierre leaves: understanding the mechanism(s) of action in infectious diarrhea. J. Zhejiang Univ. Sci. B. **7**(8): 665–674. doi:10.1631/jzus.2006.B0665.

13. Buchan, B.W., and Ledeboer, N.A. 2014. Emerging technologies for the clinical microbiology laboratory. Clin. Microbiol. Rev. **27**(4): 783–822. doi:10.1128/CMR.00003-14.
14. Čanković, M., Bokor-Bratić, M., Lončar, J., Marinoski, J., and Ilić, M.P. 2013. Bacterial flora on the surface of oral squamous cell carcinoma. Arch. Oncol. **21**(2): 62–64. doi:10.2298/AOO1302062C.
15. Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., and Kjelleberg, S. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl. Environ. Microbiol. **73**(1): 278–288. doi:10.1128/AEM.01177-06.
16. Chaturvedi, P., Singh, A., Chien, C.Y., and Warnakulasuriya, S. 2019, June 5. Tobacco related oral cancer. British Medical Journal Publishing Group. doi:10.1136/bmj.l2142. 17. Cho, T.J., Wee, S.W., Woo, V.H., Choi, J. Il, Kim, S.J., Shin, H.I., Lee, J.H., and Park, H.R. 2014. Porphyromonas gingivalis-induced autophagy suppresses cell proliferation through G1 arrest in oral cancer cells. Arch. Oral Biol. **59**(4): 370–378. Elsevier Ltd. doi:10.1016/j.archoralbio.2014.01.001.
18. Chocolatewala, N., Chaturvedi, P., and Desale, R. 2010. The role of bacteria in oral cancer. Indian J. Med. Paediatr. Oncol. **31**(4): 126. doi:10.4103/0971-5851.76195.
19. Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C.R., Yu, W.H., Lakshmanan, A., and Wade, W.G. 2010. The human oral microbiome. J. Bacteriol. **192**(19): 5002– 5017. doi:10.1128/JB.00542-10.
20. Dubilier, N., Mcfall-Ngai, M., and Zhao, L. 2015, October 28. Microbiology: Create a global microbiome effort. Nature Publishing Group. doi:10.1038/526631a.
21. Estaki, M., Jiang, L., Bokulich, N.A., McDonald, D., González, A., Kosciolek, T., Martino, C., Zhu, Q., Birmingham, A., Vázquez-Baeza, Y., Dillon, M.R., Bolyen, E., Caporaso, J.G., and Knight, R. 2020. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr. Protoc. Bioinforma. **70**(1): 1–46. doi:10.1002/cpbi.100.
22. Fardini, Y., Wang, X., Témoin, S., Nithianantham, S., Lee, D., Shoham, M., and Han, Y.W. 2011. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. Mol. Microbiol. **82**(6): 1468–80. NIH Public Access. doi:10.1111/j.1365-2958.2011.07905.x.
23. Ghantous, Y., Schussel, J.L., and Brait, M. 2018. Tobacco and Alcohol induced Epigenetic changes in Oral Carcinoma. Curr. Opin. Oncol. **30**(3): 152. NIH Public Access. doi:10.1097/CCO.0000000000000444.
24. Gholizadeh, P., Eslami, H., and Kafil, H.S. 2017. Carcinogenesis mechanisms of Fusobacterium nucleatum. Biomed. Pharmacother. **89**: 918–925. doi:10.1016/j.biopha.2017.02.102.
25. Gholizadeh, P., Eslami, H., Yousefi, M., Asgharzadeh, M., Aghazadeh, M., and Kafil, H.S. 2016. Role of oral microbiome on oral cancers, a review. Biomed. Pharmacother. **84**: 552–558. doi:10.1016/j.biopha.2016.09.082.
26. Groeger, S., Howaldt, H.P., Raifer, H., Gattenloehner, S., Chakraborty, T., and Meyle, J. 2017. Oral Squamous Carcinoma Cells Express B7-H1 and B7-DC Receptors in Vivo. Pathol. Oncol. Res. **23**(1): 99–110. Pathology & Oncology Research. doi:10.1007/s12253-016-0100-7.

27. Groeger, S., Jarzina, F., Mamat, U., and Meyle, J. 2016. Induction of B7-H1 receptor by bacterial cells fractions of Porphyromonas gingivalis on human oral epithelial cells. Immunobiology **222**(2): 1–11. Elsevier GmbH. doi:10.1016/j.imbio.2016.10.011.
28. Guo, J., Cole, J.R., Zhang, Q., Brown, T., and Tiedje, J.M. 2016. Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. Appl. Environ. Microbiol. **82**(1): 157–166. American Society for Microbiology. doi:10.1128/AEM.02772-15.
29. Ha, N.H., Park, D.G., Woo, B.H., Kim, D.J., Choi, J. Il, Park, B.S., Kim, Y.D., Lee, J.H., and Park, H.R. 2016. Porphyromonas gingivalis increases the invasiveness of oral cancer cells by upregulating IL-8 and MMPs. Cytokine **86**: 64–72. Elsevier Ltd. doi:10.1016/j.cyto.2016.07.013.
30. Han, Y.W., Shi, W., Huang, G.T.J.J., Kinder Haake, S., Park, N.H., Kuramitsu, H., and Genco, R.J. 2000. Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. Infect. Immun. **68**(6): 3140–3146. doi:10.1128/IAI.68.6.3140-3146.2000.
31. Hanahan, D., and Weinberg, R.A. 2011. Hallmarks of cancer: The next generation. Cell **144**(5): 646–674. Elsevier Inc. doi:10.1016/j.cell.2011.02.013.
32. Johnson, L., Atanasova, K.R., Bui, P.Q., Lee, J., Hung, S.C., Yilmaz, Ö., and Ojcius, D.M. 2015. Porphyromonas gingivalis attenuates ATP-mediated inflammasome activation and HMGB1 release through expression of a nucleoside-diphosphate kinase. Microbes Infect. **17**(5): 369–377. doi:10.1016/j.micinf.2015.03.010.
33. Kakabadze, M.Z., Paresishvili, T., Karalashvili, L., Chakhunashvili, D., and Kakabadze, Z. 2020, July 6. Oral microbiota and oral cancer: Review. PAGEPress. doi:10.4081/ONCOL.2020.476.
34. Karpiński, T.M. 2019, January 13. Role of oral microbiota in cancer development. Multidisciplinary Digital Publishing Institute. doi:10.3390/microorganisms7010020.
35. Klappenbach, J.A., Dunbar, J.M., and Schmidt, T.M. 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. **66**(4): 1328–1333. American Society for Microbiology. doi:10.1128/AEM.66.4.1328-1333.2000.
36. Kumar, P.S., Matthews, C.R., Joshi, V., de Jager, M., and Aspiras, M. 2011. Tobacco smoking affects bacterial acquisition and colonization in oral biofilms. Infect. Immun. **79**(11): 4730–4738. doi:10.1128/IAI.05371-11.
37. Lalande, V., Barnabé, S., and Côté, J.-C. 2014. Comparison of the Bacterial Microbiota in a Bale of Collected Cardboard Determined by 454 Pyrosequencing and Clone Library. Adv. Microbiol. **04**(12): 754–760. doi:10.4236/aim.2014.412082.
38. Lamont, R.J., Koo, H., and Hajishengallis, G. 2018. The oral microbiota: dynamic communities and host interactions. Nat. Rev. Microbiol. **16**(12): 745–759. Springer US. doi:10.1038/s41579-018-0089-x.
39. Lim, Y., Fukuma, N., Totsika, M., Kenny, L., Morrison, M., and Punyadeera, C. 2018. The performance of an oral microbiome biomarker panel in predicting oral cavity and oropharyngeal cancers. Front. Cell. Infect. Microbiol. **8**(AUG): 267. Frontiers Media S.A. doi:10.3389/fcimb.2018.00267.
40. Liu, B., Faller, L.L., Klitgord, N., Mazumdar, V., Ghodsi, M., Sommer, D.D., Gibbons, T.R., Treangen, T.J., Chang, Y.-C., Li, S., Stine, O.C., Hasturk, H., Kasif, S., Segrè, D., Pop, M.,

and Amar, S. 2012. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. PLoS One **7**(6): e37919. Public Library of Science. doi:10.1371/journal.pone.0037919.

41. Mager, D.L., Haffajee, A.D., Devlin, P.M., Norris, C.M., Posner, M.R., and Goodson, J.M. 2005. The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med **3**(1): 27. doi:10.1186/1479-5876-3-27.
42. Metgud, R., Gupta, K., and Gupta, J. 2014. Exploring bacterial flora in oral squamous cell carcinoma: a microbiological study. Biotech. Histochem. **89**(2): 153–159. doi:10.3109/10520295.2013.831120.
43. Mok, S.F., Karuthan, C., Cheah, Y.K., Ngeow, W.C., Rosnah, Z., Yap, S.F., and Ong, H.K.A. 2017. The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. Malays. J. Pathol. **39**(1): 1–15. Available from http://www.ncbi.nlm.nih.gov/pubmed/28413200.
44. Nadkarni, M., Martin, F.E., Jacques, N.A., and Hunter, N. 2002. Determination of bacterial load by real-time PCR using a broad range (universal) probe and primer set. Microbiology **148**: 257–266. doi:10.1128/JCM.40.5.1698.
45. Nagy, K.N., Sonkodi, I., Szöke, I., Nagy, E., and Newman, H.N. 1998. The microflora associated with human oral carcinomas. Oral Oncol. **34**(4): 304–308. doi:10.1016/S1368-8375(98)00008-6.
46. Perera, M., Al-hebshi, N.N., Perera, I., Ipe, D., Ulett, G.C., Speicher, D.J., Chen, T., and Johnson, N.W. 2018. Inflammatory Bacteriome and Oral Squamous Cell Carcinoma. doi:10.1177/0022034518767118.
47. Pinho, S.S., and Reis, C.A. 2015, September 25. Glycosylation in cancer: Mechanisms and clinical implications. Nat Rev Cancer. doi:10.1038/nrc3982.
48. Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., Buck, G.A., Snyder, M.P., Strauss, J.F., Weinstock, G.M., White, O., and Huttenhower, C. 2019. The Integrative Human Microbiome Project. Nature **569**(7758): 641–648. doi:10.1038/s41586-019-1238-8.
49. Pushalkar, S., Ji, X., Li, Y., Estilo, C., Yegnanarayana, R., and Singh, B. 2012. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma.
50. Pushalkar, S., Mane, S.P., Ji, X., Li, Y., Evans, C., Crasta, O.R., Morse, D., Meagher, R., Singh, A., and Saxena, D. 2011. Microbial diversity in saliva of oral squamous cell carcinoma. FEMS Immunol. Med. Microbiol. **61**(3): 269–277. doi:10.1111/j.1574695X.2010.00773.x.
51. Rodríguez-Molinero, J., Migueláñez-Medrán, B.D.C., Puente-Gutiérrez, C., DelgadoSomolinos, E., Carreras-Presas, C.M., Fernández-Farhall, J., and López-Sánchez, A.F. 2021, April 15. Association between oral cancer and diet: An update. Multidisciplinary Digital Publishing Institute. doi:10.3390/nu13041299.
52. la Rosa, G.R.M., Gattuso, G., Pedullà, E., Rapisarda, E., Nicolosi, D., and Salmeri, M. 2020, April 1. Association of oral dysbiosis with oral cancer development (Review). Spandidos Publications. doi:10.3892/ol.2020.11441.

53. Sampaio-Maia, B., Caldas, I.M., Pereira, M.L., Pérez-Mongiovi, D., and Araujo, R. 2016. The Oral Microbiome in Health and Its Implication in Oral and Systemic Diseases. *In* Advances in Applied Microbiology. Elsevier Ltd. doi:10.1016/bs.aambs.2016.08.002.
54. Schmidt, B.L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P.M., Queiroz, E.L.S., Nightingale, K., Kerr, A.R., DeLacure, M.D., Veeramachaneni, R., Olshen, A.B., and Albertson, D.G. 2014. Changes in abundance of oral microbiota associated with oral cancer. PLoS One **9**(6). doi:10.1371/journal.pone.0098741.
55. Shirakawa, T., and Kitagawa, K. 2018. Antitumor effect of oral cancer vaccine with Bifidobacterium delivering WT1 protein to gut immune system is superior to WT1 peptide vaccine. Hum. Vaccines Immunother. **14**(1): 159–162. Taylor & Francis. doi:10.1080/21645515.2017.1382787.
56. Smith, C.J., and Osborn, A.M. 2009. Advantages and limitations of quantitative PCR (QPCR)-based approaches in microbial ecology. FEMS Microbiol. Ecol. **67**(1): 6–20. Oxford University Press. doi:10.1111/j.1574-6941.2008.00629.x.
57. Sundin, O.H., Mendoza-Ladd, A., Zeng, M., Diaz-Arévalo, D., Morales, E., Fagan, B.M., Ordoñez, J., Velez, P., Antony, N., and McCallum, R.W. 2017. The human jejunum has an endogenous microbiota that differs from those in the oral cavity and colon. BMC Microbiol. **17**(1): 1–17. BMC Microbiology. doi:10.1186/s12866-017-1059-6.
58. Szkaradkiewicz, A.K., and Karpiński, T.M. 2013. Microbiology of Chronic Periodontitis. Rev. J. Biol. Earth Sci. **3**(1): 14–20. Available from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1019.3177&rep=rep1&type=pdf [accessed 10 May 2022].
59. Sztukowska, M.N., Ojo, A., Ahmed, S., Carenbauer, A.L., Wang, Q., Shumway, B., Jenkinson, H.F., Wang, H., Darling, D.S., and Lamont, R.J. 2016. Porphyromonas gingivalis initiates a mesenchymal-like transition through ZEB1 in gingival epithelial cells. Cell. Microbiol. **18**(6): 844–858. doi:10.1111/cmi.12554.
60. Wang, H., Zhou, H., Duan, X., Jotwani, R., Vuddaraju, H., Liang, S., Scott, D.A., and Lamont, R.J. 2014. Porphyromonas gingivalis-induced reactive oxygen species activate JAK2 and regulate production of inflammatory cytokines through c-Jun. Infect. Immun. **82**(10): 4118–4126. doi:10.1128/IAI.02000-14.
61. Wang, Q., Rao, Y., Guo, X., Liu, N., Liu, S., Wen, P., Li, S., and Li, Y. 2019. Oral Microbiome in Patients with Oesophageal Squamous Cell Carcinoma. Sci. Rep. **9**(1): 1–9. Nature Research. doi:10.1038/s41598-019-55667-w.
62. Wu, J., Peters, B.A., Dominianni, C., Zhang, Y., Pei, Z., Yang, L., Ma, Y., Purdue, M.P., Jacobs, E.J., Gapstur, S.M., Li, H., Alekseyenko, A. V., Hayes, R.B., and Ahn, J. 2016. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J. **10**(10): 2435–2446. Nature Publishing Group. doi:10.1038/ismej.2016.37.
63. Yang, C.Y., Yeh, Y.M., Yu, H.Y., Chin, C.Y., Hsu, C.W., Liu, H., Huang, P.J., Hu, S.N., Liao, C.T., Chang, K.P., and Chang, Y.L. 2018. Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. Front. Microbiol. **9**(MAY). doi:10.3389/fmicb.2018.00862.
64. Zaura, E., Keijser, B.J., Huse, S.M., and Crielaard, W. 2009. Defining the healthy "core microbiome" of oral microbial communities. BMC Microbiol. **9**(1): 1–12. BioMed Central. doi:10.1186/1471-2180-9-259.

65. Zhang, L., Liu, Y., Zheng, H.J., and Zhang, C.P. 2020. The Oral Microbiota May Have Influence on Oral Cancer. Front. Cell. Infect. Microbiol. **9**: 476. Frontiers Media S.A. doi:10.3389/fcimb.2019.00476.
66. Zhao, H., Chu, M., Huang, Z., Yang, X., Ran, S., Hu, B., Zhang, C., and Liang, J. 2017. Variations in oral microbiota associated with oral cancer. Sci. Rep. **7**(1): 1–10. Springer US. doi:10.1038/s41598-017-11779-9.