

ANÁLISE DE POPULAÇÕES BACTERIANAS NO CÂNCER DA CAVIDADE ORAL POR METAGENÔMICA

ANDREW MALTEZ THOMAS

**Dissertação apresentada à Fundação Antônio
Prudente para obtenção do Título de Mestre em
Ciências**

Área de concentração: Oncologia

Orientador: Dr. Emmanuel Dias-Neto

São Paulo

2012

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Thomas, Andrew Maltez

Análise de populações bacterianas no câncer da cavidade oral por metagenômica / Andrew Maltez Thomas – São Paulo, 2012.

106p.

Dissertação (Mestrado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Emmanuel Dias-Neto

Descritores: 1. CAVIDADE ORAL. 2. CARCINOMA EPIDERMÓIDE. 3. METAGENÔMICA. 4. RNA RIBOSSÔMICO 16S. 5. CÂNCER DE CABEÇA E PESCOÇO. 6. SEQUENCIAMENTO DE NUCLEOTÍDEOS EM GRANDE ESCALA.

“Seja a mudança que você quer ver no mundo.”

Mahatma Gandhi

DEDICATÓRIA

À minha mãe, Patrícia, por todo seu incentivo, amor incondicional e exemplo de perseverança.

Ao meu pai, Stephen, pelo apoio e incentivo.

Obrigado por tudo!

AGRADECIMENTOS

Aos meus pais, pelo grande incentivo e ao apoio recebido durante cada desafio conquistado. Sem vocês nada seria possível.

À minha grande família (avós, tios e primos), pelas alegrias e boas energias que vocês trazem para minha vida.

À meu orientador, Prof. Dr. Emmanuel Dias-Neto, pela oportunidade de me apresentar o mundo da pesquisa e poder realizar o meu Mestrado, pelo auxílio, ensinamentos importantes e confiança no meu trabalho. Um exemplo a ser seguido como cientista e profissional.

À Dra. Diana Noronha Nunes, pelo auxílio, ensinamentos importantes e toda a sua ajuda.

Ao Prof. Dr. Brentani, pelo auxílio e ensinamentos importantes, um exemplo a ser seguido.

Aos meus amigos, Aldo, Alex, Anderson, Juan e Marco, pela nova amizade que nasceu em meio desta caminhada, que a mesma permaneça por toda vida.

Aos colegas do Laboratório de Genômica Médica Gabriela, Iago, Juliana, Luiza, Maria e em especial ao Frederico, pelas novas descobertas profissionais e pessoais, pela colaboração, pela ajuda nos momentos difíceis, pelos momentos de descontração e principalmente pelos momentos de alegria e companheirismo que vivemos durante estes dois anos.

Aos colegas do Laboratório de Bioinformática e Estatística, Renato, Leandro, Prof. Dr. Carlos Pereira e Aline, pela ajuda nas análises.

Aos membros da banca de qualificação, Dr. Raphael Parmigiani e Profa. Dra. Eloiza Tajara pelas sugestões e comentários na confecção deste trabalho.

Ao Dr. Luiz Paulo Kowalski e à colega Ana Lúcia Francisco, pelo auxílio, pela ajuda na coleta de amostras, além de toda sua atenção.

Aos Profs. Drs. Marcos Brasilino de Carvalho, Arthur Guerra, André Malbergier e equipes, pela ajuda na coleta das amostras.

Ao Dr. José Eduardo Levi, pelo auxílio na genotipagem dos pacientes.

Às funcionárias da Pós-Graduação, Ana, Luciana e Vanuza pelo auxílio oferecido no decorrer do Mestrado.

À toda a equipe da biblioteca, principalmente a bibliotecária Suelly Francisco, pela ajuda durante as pesquisas bibliográficas e pela revisão deste trabalho.

À CNPq e à CAPES, pelo suporte financeiro indispensável para o desenvolvimento deste projeto.

RESUMO

Thomas AM. **Análise de populações bacterianas no câncer da cavidade oral por metagenômica**. São Paulo; 2012. [Dissertação de Mestrado-Fundação Antônio Prudente]

Álcool e tabaco são os agentes externos mais importantes para o surgimento do câncer da cavidade oral. Outro elemento exógeno que talvez seja importante na composição deste cenário é a microbiota da cavidade oral, que atua de forma marcante contribuindo para gerar agentes potencialmente carcinogênicos, tais como o acetaldeído. Neste projeto, avaliamos a composição da microbiota oral bacteriana utilizando sequenciamento em larga escala, um método que independe do crescimento ou cultivo de bactérias *in vitro*, e permite a identificação e quantificação de táxons bacterianos após análises de bioinformática. Para avaliar as populações bacterianas presentes em biofilmes de pacientes com câncer da cavidade oral, três grupos foram estudados: 1) 10 indivíduos sem câncer de cavidade oral e que não usam álcool ou tabaco; 2) 5 indivíduos sem câncer de cavidade oral e que fazem uso diário de tabaco e álcool e 3) 11 indivíduos portadores de câncer oral e que fazem uso destas duas drogas. Para tal, amplificamos regiões polimórficas do gene 16S-rDNA, marcamos as sequências oriundas de cada indivíduo com *tags* específicas e sequenciamos os *amplicons* na plataforma *Ion Torrent*TM. As sequências obtidas foram anotadas utilizando bancos de dados públicos. Os dados obtidos sugerem uma maior diversidade bacteriana nos pacientes com câncer oral, quando comparados aos grupos-controle sem câncer, sendo esta ainda mais elevada entre os pacientes com estadiamento tumoral mais avançado (T3-T4). Bactérias do gênero *Actinomyces* se mostraram mais prevalentes nos pacientes e controles que fumam e bebem. Os resultados obtidos devem auxiliar na compreensão da composição da microbiota bacteriana oral de pacientes com câncer, representando o primeiro passo para uma avaliação do papel de carcinógenos comuns sobre microbiota e o papel de ambos no desenvolvimento do câncer de cavidade oral.

SUMMARY

Thomas AM. [A metagenomic survey of bacterial populations in oral cavity cancer]. São Paulo; 2012. [Dissertação de Mestrado-Fundação Antônio Prudente]

Alcohol and tobacco are the two most important external agents that cause oral squamous cell carcinomas (OSCC). The oral cavity's microbiota is also another exogenous element that may have an important role in this scenario, generating potentially carcinogenic agents, such as acetaldehyde. In this project, the oral cavity's bacterial microbiota composition was evaluated using metagenomic large-scale sequencing, a method that doesn't rely on the *in vivo* growth or cultivation of bacteria, and allows their identification and quantification after *in silico* analysis. In order to evaluate the bacterial populations present in the biofilms of patients with OSCC, three groups were studied: 1) 10 individuals without OSCC that do not smoke nor drink, 2) 5 individuals without OSCC that smoke and drink on a daily basis, 3) 11 individuals that have OSCC and use these two drugs. To do so, polymorphic regions of the 16S-rDNA gene were amplified by PCR and the resulting amplicons were barcoded with specific sequences for each individual and were then sequenced on the Ion Torrent™ platform. The resulting sequences were compared to those present in public databases. The data suggests a higher bacterial diversity in patients with OSCC, when compared to the group without cancer, the group without cancer that smoke and drank and even more so in patients with advanced tumours (T3-T4). The bacterial genus *Actinomyces* was found to be more abundant in patients with OSSC then both control groups. The results obtained in this study should help understand the bacterial composition of patients with oral cancer, a first step in evaluating the role that common carcinogens play on the microbiota and oral cavity cancer.

LISTA DE FIGURAS

Figura 1	Modelo da carcinogênese oral associada ao álcool.....	6
Figura 2	Estrutura secundária e terciária da molécula 16S-RNA.....	11
Figura 3	Representação esquemática do gene 16S-rDNA bacteriano.....	12
Figura 4	Principais etapas de um estudo metagenômico baseado na amplificação do gene 16S-rDNA.....	14
Figura 5	Visualização dos produtos de amplificação da região V1 do gene 16S-rDNA em gel de poliacrilamida 8% corado com sais de prata.....	36
Figura 6	Avaliação das amostras em chip de DNA <i>High Sensitivity</i> no <i>Bioanalyzer</i>	37
Figura 7	Resultado da busca de HPV em algumas amostras deste estudo.....	43
Figura 8	Curvas de saturação/rarefação obtidas a partir da análise individual para os três grupos de estudo.....	51
Figura 9	Curvas de saturação/rarefação obtidas a partir da análise agrupada dos indivíduos dos três grupos de estudo.....	53
Figura 10	Curvas de saturação/rarefação obtidas a partir da análise de sub-grupos de pacientes agrupados conforme tamanho da lesão tumoral.....	55

Figura 11	Modelo hipotético do efeito do álcool e do tabaco na diversidade da microbiota oral e seu aumento durante o aparecimento de tumores menos e mais avançados.....	57
Figura 12	Filos bacterianos mais abundantes nos três grupos de estudo.....	63
Figura 13	Gráfico mostrando o número de core OTUs observado e calculado para o grupo CFB, considerando diferentes quantidades de amostras avaliadas.....	71
Figura 14	Gráfico de barras mostrando o número de core OTUs observadas em 100% das amostras nos 3 grupos, considerando-se 10 amostras para o grupo CFB.....	72
Figura 15	Alinhamento por BLAST da OTU 298928 com a sequência de <i>Actinomyces graevenitzi</i> pelo banco online HOMD.....	79
Figura 16	Diagram de Venn com as core OTUs encontradas em 100, 90 e 80% das amostras dos três grupos.....	80
Figura 17	Gráfico de grupos de frequência de OTUs observadas para cada amostra dos três grupos de estudo.....	85
Figura 18	Gráfico de grupos de frequência de OTUs observadas para os três grupos de estudo.....	86
Figura 19	Gráfico de grupos de frequência de OTUs observadas para os grupos de estudo, com ênfase no estadiamento tumoral para o grupo dos pacientes.....	87
Figura 20	Número de espécies esperadas para os grupos P e C gerados pelo Catchall.....	88

Figura 21	Número de espécies observadas para os grupos P e C.....	89
Figura 22	Número de espécies observadas para o grupo C e os sub-grupos P T1-T2 e P T3-T4.....	90

LISTA DE QUADROS

Quadro 1	Grupos de estudo deste trabalho.....	29
Quadro 2	Detalhamento dos grupos e critérios de inclusão.....	30
Quadro 3	Características gerais de indivíduos que compõe o Grupo de Controles (C).....	31
Quadro 4	Características gerais de indivíduos que compõe o Grupo de Controles que Fumam e Bebem (CFB).....	31
Quadro 5	Características gerais dos indivíduos do Grupo de Pacientes (P) diagnosticados com OSCC e avaliados neste estudo.....	32

LISTA DE TABELAS

Tabela 1	Características gerais de indivíduos dos grupos-controle (C e CFB) e pacientes (P) avaliados neste estudo.....	42
Tabela 2	Filtragem inicial das sequências geradas.....	45
Tabela 3	Número de sequências obtidas para as amostras do estudo (antes e após a aplicação dos filtros), e mediana do tamanho.....	46
Tabela 4	Análise qualitativa e quantitativa dos principais filos bacterianos encontrados na cavidade oral, de acordo com a região amostrada do gene 16S-rDNA e a plataforma de sequenciamento utilizada....	49
Tabela 5	<i>OTUs</i> mais abundantes no grupo dos controles.....	58
Tabela 6	<i>OTUs</i> mais abundantes no grupo dos controles quem fumam e bebem.....	59
Tabela 7	<i>OTUs</i> mais abundantes no grupo dos pacientes.....	60
Tabela 8	Prevalência dos filos bacterianos mais abundantes em diversos estudos voltados para a avaliação da microbiota oral.....	64
Tabela 9	Número de <i>core OTUs</i> encontrados em 100, 90 e 80% das amostras dos três grupos.....	67
Tabela 10	Número de <i>core OTUs</i> encontrados em 100, 90 e 80% das amostras dos três grupos com a metade do <i>n</i> amostral.....	68
Tabela 11	Número de <i>core OTUs</i> encontrados em 100% das amostras dos sub-grupos de indivíduos do grupo CFB.....	70

Tabela 12	<i>Core OTUs</i> encontrados em 100% das amostras do grupo dos controles (C).....	73
Tabela 13	<i>Core OTUs</i> encontrados em 100% das amostras do grupo dos controles que fumam e bebem (CFB).....	74
Tabela 14	<i>Core OTUs</i> encontrados em 100% das amostras do grupo dos pacientes (P).....	76
Tabela 15	Comparação do número de <i>reads</i> normalizados para a <i>OTU</i> correspondente ao gênero <i>Saccharomonospora</i> nos três grupos.....	83

LISTA DE SIGLAS E ABREVIATURAS

16S-rDNA	Ácido Desoxirribonucleíco ribossomal 16S
16S-rRNA	Ácido Ribonucleíco ribossomal 16S
DNA	Ácido Desoxirribonucleíco
ePCR	Reação em cadeia da polimerase em emulsão
FISH	<i>Fluorescent in situ Hybridization</i> - Hibridização Fluorescente <i>in situ</i>
HIGK	Células de queratinócitos gengivais humanos imortalizados
HNSCC	<i>Head and Neck Squamous Cell Carcinoma</i> - Câncer Epidermóide da Cabeça e Pescoço
HOMD	<i>Human Oral Microbiome Database</i>
HPV	<i>Human Papilloma Virus</i> - Vírus do Papiloma Humano
INCA	Instituto Nacional do Câncer
mRNA	Ácido Ribonucleíco mensageiro
nt	Nucleotídeo
OSCC	<i>Oral Squamous Cell Carcinoma</i> - Câncer epidermóide da cavidade oral
OUT	<i>Operational Taxonomic Unit</i> - Unidade Taxonômica Operacional
PCR	<i>Polymerase Chain Reaction</i> - Reação em cadeia da polimerase
qPCR	Reação em cadeia da polimerase quantitativa
RDP	<i>Ribosomal Database Project</i>
RNA	Ácido Ribonucleíco
rRNA	Ácido Ribonucleíco ribossomal
TNF-α	<i>Tumor necrosis factor alpha</i> - Fator de necrose tumoral alfa

ÍNDICE

1	INTRODUÇÃO.....	1
1.1	Tumores da Cabeça e Pescoço	1
1.1.1	Epidemiologia	1
1.1.2	Características clínicas do câncer oral	2
1.1.3	Fatores de risco	3
1.1.4	Álcool e Tabaco- Influência no câncer da cavidade oral	4
1.2	Metagenômica Bacteriana	6
1.2.1	Avaliando a importância da microbiota bacteriana no câncer oral: definição, abordagens técnicas e utilização de estudos de metagenômica	6
1.2.2	Abordagens de estudos de metagenômica bacteriana	9
1.2.3	Os genes 16S-rDNA	10
1.2.4	Vantagens e limitações do uso da metagenômica baseada em 16S-rDNA.....	13
1.3	Microbiota humana	15
1.3.1	Microbiota e doenças	16
1.3.2	A microbiota e câncer	17
1.4	A microbiota oral	20
1.4.1	Biofilmes e saliva.....	20
1.4.2	Alterações da Microbiota Oral na Presença de Álcool e/ou Tabaco.....	24
1.4.3	Alterações da Microbiota Oral em Pacientes com Câncer Oral	26
2	HIPÓTESE E OBJETIVOS.....	28
2.1	Objetivo Geral.....	28
2.2	Objetivos Específicos.....	28
3	MATERIAL E MÉTODOS	29
3.1	Amostras	29
3.2	Crerios de inclusão/exclusão	29
3.3	Coleta de amostras do biofilme.....	33
3.3.1	Obtenção do DNA.....	33

3.3.2	Eluição do DNA.....	34
3.4	Genotipagem de HPV dos pacientes	34
3.5	Amplificação da região V1 do gene 16S-rDNA	35
3.6	Ligação dos adaptadores, ePCR e sequenciamento	36
3.7	Análise taxonômica.....	37
3.8	Estimando a riqueza de espécies bacterianas	39
4	RESULTADOS E DISCUSSÃO	41
4.1	Características gerais dos indivíduos selecionados para o estudo	41
4.1.1	Análise da presença de HPV nas amostras de pacientes com OSCC.....	42
4.2	Sequenciamento	45
4.2.1	Aspectos do sequenciamento	47
4.3	Análises das sequências obtidas.....	50
4.3.1	Avaliação da cobertura obtida e inferências sobre a diversidade bacteriana nos grupos estudados	50
4.4	Táxons bacterianos presentes nos grupos avaliados	62
4.5	Análises de um possível <i>Core Microbiome</i>	66
4.6	Discrepâncias entre os <i>Core Microbiomes</i> de controles e pacientes	80
4.7	Avaliação do efeito do maior número de sequências para o grupo CFB.....	85
4.8	Estimativa da riqueza da diversidade microbiana usando o <i>Catchall</i>	88
5	CONCLUSÕES.....	91
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	93

ANEXOS

Anexo 1 A plataforma *Ion Torrent™* da *Applied Biosystems*

Anexo 2 Sequenciamento de amostras do grupo de controles (C) na plataforma 454-Roche

Anexo 3 Termo de Consentimento Livre e Esclarecido

Anexo 4 Questionário aplicado aos indivíduos deste estudo

1 INTRODUÇÃO

1.1 TUMORES DA CABEÇA E PESCOÇO

1.1.1 Epidemiologia

Os tumores de cabeça e pescoço ocorrem em vários sítios do trato aéreo-digestivo superior; mais de 90% dos casos são carcinomas epidermóides e ocorrem predominantemente na cavidade oral, orofaringe, hipofaringe e laringe. No ano 2000, os tumores de cabeça e pescoço eram a oitava causa mais comum de câncer no mundo (RAGIN et al. 2006). Aproximadamente 480 mil novos casos são registrados a cada ano, dos quais 320 mil resultam em morte dos pacientes em um intervalo inferior a 5 anos (SHIBUYA et al. 2002). Na América do Sul e no Caribe, o câncer da cavidade oral e o câncer da faringe chegam a ocupar o quinto lugar entre as neoplasias mais frequentes nos homens e o sexto entre as mulheres (WARNAKULASURIYA 2009). Na população brasileira, o carcinoma epidermóide da cavidade oral (*Oral Squamous Cell Carcinoma, OSCC*) é a sétima neoplasia mais comum (WARNAKULASURIYA 2009), sendo que a população masculina possui um alto risco de desenvolver esta doença, situando-se mundialmente atrás apenas da França e da Índia (WUNSCH-FIHO e DE CAMARGO 2001).

A incidência destes tumores está crescendo em São Paulo e em outras áreas do mundo, provavelmente como consequência do uso excessivo e crescente de álcool e/ou de tabaco em certas regiões (COLEMAN et al. 1993). No Brasil, a maior incidência do câncer da cavidade oral e da orofaringe é vista na cidade de São Paulo,

com 25,3 casos por 100 mil habitantes (WUNSCH-FIHO 2002); nesta cidade, 11% de todos os casos de câncer registrados em homens são classificados como Câncer Epidermóide da Cabeça e Pescoço (*Head and Neck Squamous Cell Carcinoma, HNSCC*) (CURADO et al. 2007). No ano de 2005, de acordo com os dados do Instituto Nacional de Câncer-INCA, o câncer da cavidade oral e orofaringe era a 6^a neoplasia mais comum em homens sendo que este estado possui ainda a maior estimativa de incidência de casos de *OSCC* para o ano de 2010 de todo Brasil, chegando a 4.120 indivíduos afetados apenas neste ano (Ministério da Saúde 2011).

1.1.2 Características clínicas do câncer oral

Nos dias de hoje, a taxa total da sobrevida global aos 5 anos para *OSCC* é de 58%, enquanto que a média de sobrevida livre de doença é de 56% (BELL et al. 2007). Tais taxas não mudaram muito nos últimos anos, e o diagnóstico precoce permanece como a melhor maneira de combater esta doença com eficácia.

A dor local é o principal sintoma de *OSCC*, porém quando a doença chega neste estágio, a lesão já é grande e o tratamento é menos efetivo. A maior parte dos portadores de *OSCC* não apresentam sintomas, porém quando estes ocorrem podem incluir: dor no ouvido, disfagia, sangramento, dificuldade na respiração, dificuldade na fala e mobilidade alterada dos dentes (HAYA-FERNÁNDEZ et al. 2004). O *OSCC* pode ocorrer em qualquer local da boca, porém se apresenta mais frequentemente na língua e no assoalho da boca (MASHBERG et al. 1989). Outras áreas afetadas incluem a mucosa bucal, área retro-molar, gengiva e palato mole (BAGAN et al. 2010). Lesões pré-malignas que precedem o *OSCC* geralmente ocorrem na forma de leucoplasias, lesões teciduais extremamente bem demarcadas,

com áreas vermelhas e brancas, observadas na mucosa oral (BAGAN et al. 2010). As lesões malignas em estágios avançados possuem duas principais características; a ulceração e nodulação, onde são encontrados tumores exofíticos, invasivos e com regiões mal delimitadas (BAGAN et al. 2010).

1.1.3 Fatores de risco

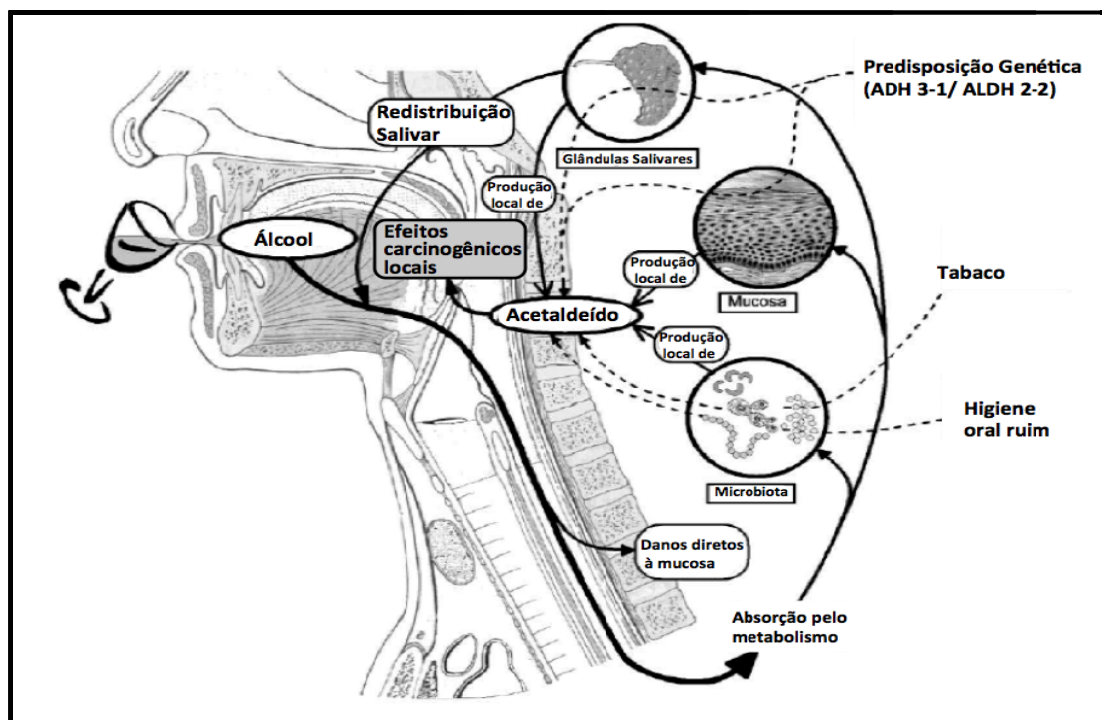
Os agentes etiológicos do *OSCC* incluem tabaco, álcool e também traumas causados por próteses dentárias, que demonstraram possuir alguma associação com o seu desenvolvimento. Outros agentes também já demonstraram possuir alguma associação com o desenvolvimento do *OSCC*, incluindo o chimarrão e o betel (mascado em regiões do oriente). Um estudo realizado por DE STEFANI et al. (1988) mostrou que indivíduos que consomem mate possuíam um risco 3 vezes maior de desenvolver câncer da laringe, quando comparados a controles normalizados para o uso de tabaco e álcool, provavelmente causado pela alta temperatura do mate quando consumido. OREGGIA et al. (1991) observaram que homens uruguaios que faziam uso de chimarrão possuíam um risco relativo 2,5 vezes maior para o desenvolvimento do câncer de língua comparados a controles. Surpreendentemente, o consumo de café mostrou ter uma associação inversa com o desenvolvimento do câncer oral e da faringe. GALEONE et al. (2010) encontraram uma associação oposta entre o consumo de café cafeinado e o risco de desenvolvimento de câncer da cavidade oral e faringe. Os autores explicam que isto pode ocorrer porque o café possui centenas de substâncias, incluindo algumas com atividades anti-oxidantes e anti-mutagênicas (demonstrado em modelos animais e em cultura de células), que podem proteger a mucosa oral.

1.1.4 Álcool e tabaco - influência no desenvolvimento do câncer da cavidade oral

Mais de 70% dos casos de *HNSCC* poderiam ser evitados com mudanças no estilo de vida que levassem a uma redução na exposição aos fatores de risco mais importantes: o uso de tabaco e de álcool (SOERJOMATARAM et al. 2007). Apenas em 1957, o tabagismo foi apontado como causa importante para o desenvolvimento de câncer oral e de orofaringe (WYNDER e BOSS 1957). Posteriormente, o uso de tabaco e de álcool foram apontados como os dois fatores de risco mais importantes para o desenvolvimento destes tipos de câncer (revisto em FRANCESCHI et al. 1999). Desde então, diversos artigos têm mostrado que o uso destas duas drogas aumenta o risco de desenvolvimento destes tumores de modo diretamente proporcional à dose utilizada (OLSEN et al. 1985; BLOT et al. 1988; LEWIN et al. 1998; TALAMINI et al. 2002; CASTELLAGUE et al. 2004). Enquanto o risco relativo de desenvolvimento de *HNSCC* para o uso isolado destas duas drogas parece ser similar (HASHIBE et al. 2007), o uso simultâneo de ambas desencadeia um efeito sinérgico, aumentando enormemente o risco de desenvolvimento da doença (TALAMINI et al. 2002). Sendo assim, fumantes habituais possuem de 5 a 25 vezes mais chances de desenvolver tumores da cabeça e pescoço, do que indivíduos que não fumam. Se estes mesmos indivíduos ingerirem o equivalente a cinco doses de bebidas alcoólicas por dia, este risco sobe para cerca de 40 vezes (MARUR e FORASTIERE 2008).

Aparentemente, o etanol não é um agente carcinogênico *per se*, mas sob certas condições experimentais pode agir como um co-carcinógeno e/ou promotor da tumorigênese, através de mecanismos que incluem a geração de espécies reativas de

oxigênio, mudanças no metabolismo do folato, e pela sua capacidade de atuar como solvente de agentes carcinogênicos presentes no tabaco (HOMANN 2001). Evidências sugerem que seu papel na tumorigênese se dê principalmente através dos efeitos genotóxicos de seus metabólitos, principalmente o acetaldeído (CH_3CHO) (revisto por BOFFETTA e HASHIBE 2006). O acetaldeído - produto da oxidação do etanol, reação catalizada pelas enzimas álcool desidrogenases teciduais, ou por enzimas similares produzidas por bactérias da flora normal - é um composto carcinogênico, com atividades genotóxicas e mutagênicas, que se liga ao DNA e proteínas e causa aberrações cromossômicas (BROOKS e THERUVATHU 2005). Em associação com outros componentes encontrados na fumaça do tabaco, que por si só já é rica em acetaldeído (HOFFMANN e HOFFMANN 2001), o acetaldeído produzido endogenamente, ou ingerido via dieta pode levar a uma maior proliferação celular. Independentemente de sua origem, o acetaldeído parece agir como um carcinógeno acumulativo no trato digestivo superior humano (SALASPURO 2009). O uso de tabaco e álcool, em conjunto com uma higiene oral inadequada, pode alterar a flora oral e contribuir para o aumento da produção de acetaldeído (HOMANN 2001). Em 2009 a Organização Mundial da Saúde-OMS elevou a classificação do acetaldeído para o nível de carcinógeno tipo I (International Agency for Research of Cancer-IARC 2012), reforçado por evidências que indicam que o acetaldeído seja o principal causador da carcinogênese associada ao álcool (revisto em PÖSCHL e SEITZ 2004). Um modelo da carcinogênese oral associada ao álcool, salientando o papel central do acetaldeído, é apresentado na **Figura 1**.



Legenda - A figura mostra um modelo de carcinogênese oral tendo como elementos centrais a produção de acetaldeído devido ao consumo de álcool e de tabaco, associado a uma higiene oral ruim, predisposição genética e a microbiota oral.

Fonte: Adaptado de HOMANN (2001).

Figura 1 – Modelo da carcinogênese oral associada ao uso de álcool

1.2 METAGENÔMICA BACTERIANA

1.2.1 Avaliando a importância da microbiota bacteriana no câncer oral:

Definição, abordagens técnicas e utilização de estudos de metagenômica

Inicialmente o estudo de comunidades bacterianas complexas era totalmente baseado em técnicas de cultivo. Esta abordagem, usada ainda hoje, se baseia em uma análise fenotípica das colônias formadas em placas de cultivo, dependendo das condições necessárias para o crescimento bacteriano (meio aeróbico ou anaeróbico). No entanto, apesar do desenvolvimento de novos métodos de cultivo, a principal desvantagem persiste: o método é limitado pelo fato de que poucas espécies

bacterianas conseguem crescer nos meios de cultura usados. Deste modo, métodos baseados em cultivo não permitem uma identificação exaustiva da microbiota, sendo que esta técnica fica limitada a análises clínicas para a identificação de certos táxons bacterianos.

O reconhecimento de que a maioria dos microorganismos não é cultivável estimulou o desenvolvimento de uma nova abordagem de estudo, a metagenômica. O estudo de uma mistura complexa de genomas de bactérias não-cultiváveis foi possível graças a métodos baseados no isolamento de DNA das bactérias presentes em um dado ambiente, seguido de seu sequenciamento, permitindo a identificação acurada. O termo *metagenomic* foi empregado pela primeira vez por RONDON et al. (2000). Diversas regiões do DNA bacteriano podem ser avaliadas, cada uma com seu grau de informatividade específico, dependente do seu nível de variabilidade inter-específico. Inicialmente o método de Sanger foi o mais amplamente usado, tendo como alvo principal as regiões polimórficas do gene 16S-rDNA. No entanto, com o desenvolvimento das abordagens de sequenciamento, hoje é claro que métodos mais vantajosos podem ser usados permitindo superar limitações do método anterior. Tais limitações se referem principalmente a requerimentos técnicos (necessidade de clonagem de *amplicons*) e financeiros, que até poucos anos atrás impediam a geração das dezenas ou centenas de milhares de sequências necessárias para uma cobertura ampla da diversidade microbiana, a ponto de saturar uma dada microbiota.

O estudo de comunidades microbianas utilizando metagenômica permite a exploração de todos os genomas microbianos presentes em determinado nichos ecológicos, seja em plantas ou em hospedeiros animais (revisto em RIESENFELD et al. 2004). Exemplos de estudos metagenômicos incluem as análises de comunidades

bacterianas, virais ou mesmo de eucariotos encontrados nos mais diversos nichos ecológicos, incluindo águas de oceanos profundos, cavernas, ou mesmo encontrados na pele humana ou outras áreas corporais (revisto por PETROSINO et al. 2009).

Estimativas indicam que de 20 a 60% da microbiota humana não é cultivável (PEI et al. 2004; ZHOU et al. 2004). Deste modo, o uso da metagenômica permite identificar um maior número de táxons quando comparado a métodos dependentes de cultivo. Nos últimos 20 anos, novos levantamentos utilizando regiões das subunidades de RNAs ribossomais revelaram que a diversidade microbiana global é bem maior do que as ~5 mil espécies descritas anteriormente utilizando características fenotípicas. Tornou-se evidente que em alguns casos apenas cerca de 1% dos diferentes organismos unicelulares da maioria das comunidades microbianas é cultivável (HUSE et al. 2008). Um estudo complexo da microbiota oral utilizando o método clássico de cultura de bactérias necessitaria de técnicas novas e tecnologias de cultivo cujo desenvolvimento é laborioso e intrinsicamente limitado (LAZAREVIC et al. 2010).

Os métodos atualmente disponíveis para sequenciamento em larga-escala são rápidos e de menor custo, e por isto causaram uma grande revolução em diversas áreas, incluindo a metagenômica. No caso de projetos de metagenômica baseados em sequenciamento de ultra-larga escala do gene 16S-rDNA, um tamanho de leitura de fragmento (*read-length*) maior é importante, pois permite identificar com maior acurácia a presença de uma dada bactéria (HAMADY e NIGHT 2009). A incapacidade de produção de leituras mais longas é o principal fator limitante para a inexistência, até os dias de hoje, de estudos de metagenômica baseados no gene 16S-rDNA usando a plataforma SOLiD™ (Applied Biosystems/Life Technologies).

Atualmente, considera-se que um bom estudo de metagenômica baseado em 16S-rDNA na plataforma 454™, deva ter uma cobertura de pelo menos 1000 sequências por amostra (HAMADY e KNIGHT 2009). Esta cobertura permite inferir a frequência de espécies que estão em 1% de abundância na amostra com uma acurácia razoável, porém isto causa uma perda na identificação de inúmeras espécies menos abundantes (<1%), além de obviamente ser dependente da complexidade do microbioma avaliado, e da capacidade dos iniciadores de amplificar e representar fielmente esta complexidade. Atualmente outras plataformas de sequenciamento tem sido cada vez mais utilizadas, demonstrando resultados consistentes capazes de contribuir significativamente para o conhecimento detalhado da microbiota de tecidos de interesse. Em um estudo de LIU et al. (2007), foi demonstrado que sequências de 100 a 200 nt, tais como as produzidas pelo pirosequenciamento, podem proporcionar a mesma *clusterização* que sequências de tamanho maiores (como o gene 16S-rDNA inteiro) e que sequências mais curtas permitem uma resolução substancial de similaridades e diferenças biologicamente significantes entre amostras de comunidades microbianas (LIU et al. 2007). Após o sequenciamento, uma identificação mais acurada dos táxons bacterianos presentes é determinada após análises de bioinformática subsequentes.

1.2.2 Abordagens de estudos de metagenômica bacteriana

Duas abordagens clássicas podem ser empregadas em estudos metagenômicos de comunidades bacterianas: 1) amplificação por PCR do gene 16S-rDNA dos organismos presentes na amostra e 2) Através do isolamento e sequenciamento massivo do DNA total da amostra, sem etapas prévias de

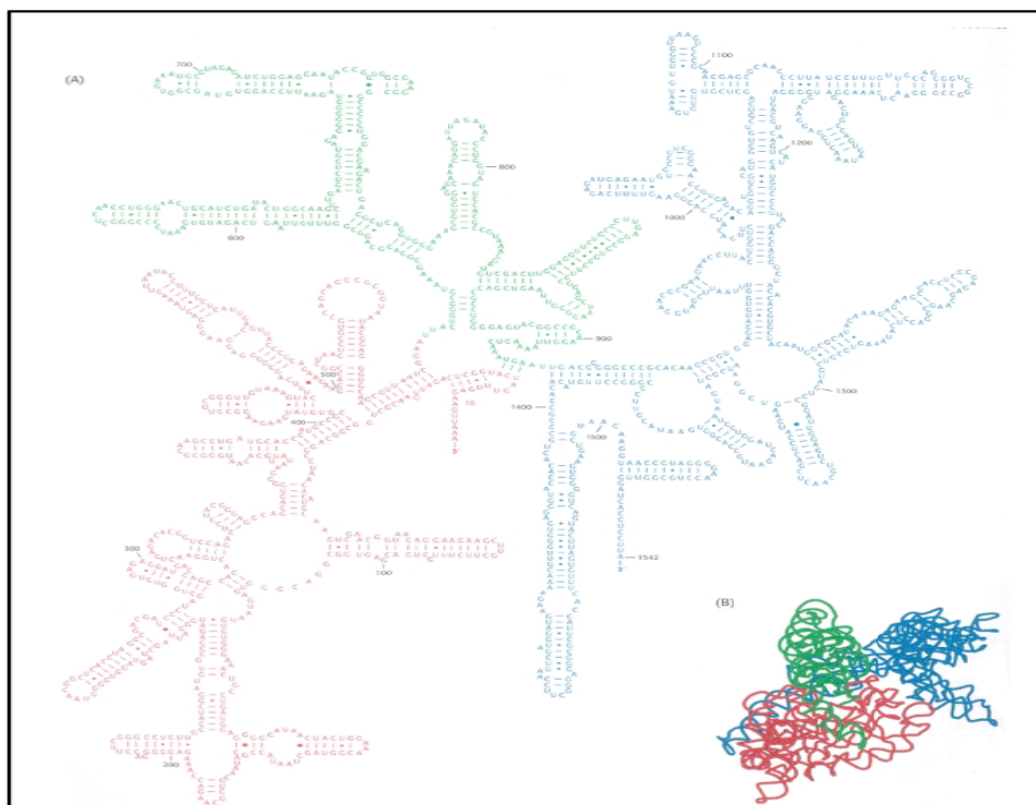
amplificação, o chamado *shotgun sequencing*. Para a determinação do perfil da comunidade microbiana de um local, o uso da amplificação do gene 16S-rDNA é o mais recomendado pela literatura, já que foca o estudo na análise de genes bacterianos específicos, sem necessidade do sequenciamento simultâneo do genoma do hospedeiro, ou de outros organismos presentes na amostra. Além disso, os custos e o tempo necessário são reduzidos de modo significativo.

1.2.3 Os genes 16S-rDNA

Os genes codificadores dos RNAs ribossomais 16S (16S-rDNA) são essenciais em procariotos, onde estão presentes em pelo menos uma cópia por genoma (ACINAS et al. 2004). A universalidade destes genes os tornam alvos ideais para estudos filogenéticos e classificações taxonômicas (WOESE 1987). Este gene possui regiões nucleotídicas conservadas nas bactérias. No entanto, o mesmo também possui regiões altamente variáveis, específicas a uma ou mais classes de bactérias (VAN DER PEER et al. 1996). Uma das razões pelas quais a molécula 16S-rDNA é um marcador filogenético tão adequado e usado em diversos trabalhos recentes, se deve à sua propriedade de *molecular clock* bem calibrado (WOESE 1987). Isto se dá pois, durante a evolução, ocorreram mutações em diferentes regiões deste gene, e em velocidades diferentes, permitindo que a maioria de suas relações filogenéticas (incluindo as mais distantes) possam ser mensuradas (WOESE 1987). O gene codificador do 16S-rRNA possui tamanho e quantidade de polimorfismos interespecíficos suficientes para permitir medidas de distância acuradas, englobando vários domínios da molécula (CLARRIDGE 2004). Há aproximadamente 50 ramos

helicoidais na estrutura secundária da molécula 16S-rRNA (GUTELL et al. 1985) (ilustrado na **Figura 2**).

Os genes que codificam para os 16S-rRNA são de fundamental importância pois codificam uma das proteínas necessárias para a formação de ribossomos nos procariotos, unindo-se a proteína 30S que auxilia na tradução de RNAs mensageiros (mRNA). Um ribossomo pode ser separado em duas subunidades; a subunidade grande (50S) e a subunidade pequena (30S). Estas duas subunidades podem ser divididas em mais proteínas. A subunidade 30S contém 21 proteínas diferentes e uma molécula de 16S-RNA. Os três RNAs presentes: 5S-, 16S- e 23S- são críticos para a arquitetura e função do ribossomo (BERG et al. 2002).

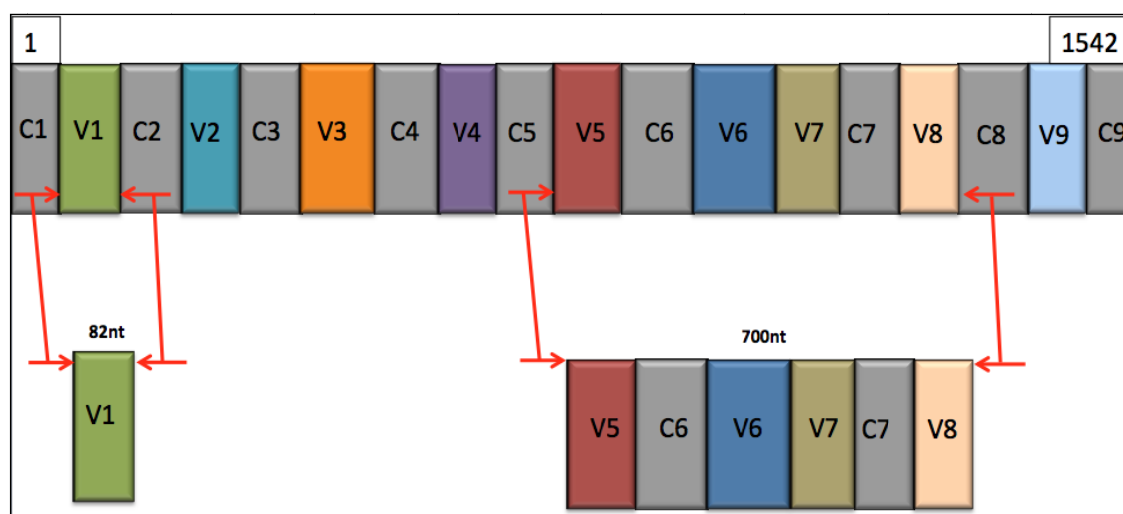


Legenda - A figura mostra a estrutura secundária do RNA ribossomal 16S (A) estrutura terciária do 16S RNA obtido através de cristalografia por raio-x (B).

Fonte: BERG et al. (2002).

Figura 2 - Estrutura secundária e terciária da molécula 16S-rRNA

Estudos diversos têm usado o gene 16S-rDNA como um marcador metagenômico para a cavidade oral (FAVERI et al. 2008), vagina (HYMAN et al. 2005) e pele (GAO et al. 2007). Nestes estudos, as regiões variáveis V1-V9 (todas ou apenas algumas delas) são amplificadas e sequenciadas, permitindo a identificação de táxons. Deste modo, os iniciadores escolhidos se ligam à regiões conservadas do gene 16S-rDNA, flanqueando regiões variáveis espécie-específicas, permitindo sua amplificação pela reação em cadeia da polimerase (PCR). A seguir os *amplicons* gerados são sequenciados e é feita a atribuição taxonômica. Uma representação do gene 16S-rDNA hipotético, com suas regiões conservadas (C) e variáveis (V) é apresentada a seguir na **Figura 3**.



Legenda - A figura mostra as regiões conservadas (designadas com a letra C) e as regiões hipervariáveis (indicadas com um V) de um gene 16S-rDNA típico (com 1542 nt). Exemplos de *amplicons* de duas regiões variáveis (V1, com 82 pares de bases e V5-V8, com 700 pares de bases) são apresentados em destaque. As setas vermelhas indicam a localização dos iniciadores.

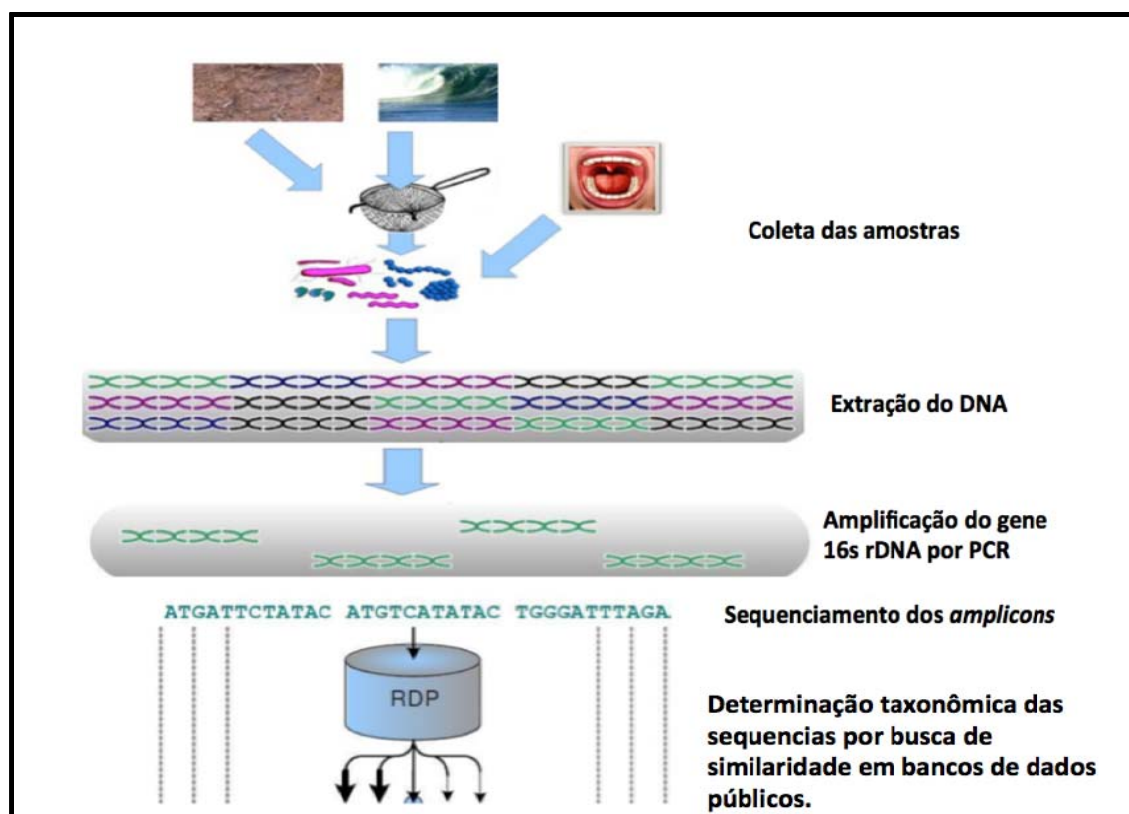
Fonte: Adaptado de PETROSINO et al. (2009).

Figura 3 - Representação esquemática do gene 16S-rDNA bacteriano

1.2.4 Vantagens e limitações da metagenômica baseada no gene 16S-rDNA

Estudos prévios demonstraram que o uso dos genes 16S-rDNA para identificação de táxons bacterianos através de uma etapa de amplificação por PCR pode levar a um viés no estudo causado pelo tamanho do *amplicon* avaliado e pela sequência dos iniciadores usados. Em um estudo de ENGELBREKTSON et al. (2010), foi avaliado o impacto do uso de *amplicons* de diferentes tamanhos cobrindo regiões distintas do gene 16S-rDNA bacteriano e como essa abordagem poderia influenciar na representação da diversidade e uniformidade das populações bacterianas durante a sua atribuição taxonômica. Os autores concluíram que *amplicons* mais curtos (<400pb) produzem uma maior estimativa de riqueza microbiana, e que uma diferença de tamanho nos *amplicons* de uma mesma região do gene 16S-rDNA não interfere muito na uniformidade da representação bacteriana, quanto uma comparação entre regiões diferentes. WU et al. (2010) mostraram os efeitos da PCR em relação ao número de ciclos, da polimerase usada e do efeito da diluição do DNA molde na atribuição taxonômica das sequências geradas pela plataforma de sequenciamento em larga-escala Illumina™ (Illumina). Como era de se esperar, uma maior diluição do DNA molde levou a uma diminuição artefactual da riqueza taxonômica observada. Além disto, um número maior de ciclos de amplificação aumentou também o número de artefatos de amplificação observados após o sequenciamento, porém sem afetar de modo significativo a estrutura da comunidade bacteriana estudada. Interessantemente, os autores constataram que a DNA polimerase empregada foi a principal causa de flutuações irreais nas populações bacterianas avaliadas. Isto se deve ao fato de uma das DNA polimerases de alta fidelidade testadas ter gerado um alto número de artefatos de PCR

(nucleotídeos que não eram complementares à fita molde foram erroneamente incorporados), aumentando o número de *OTUs* (*Operational Taxonomics Units* - Unidades Taxonômicas Operacionais) gerados, o que afetou significativamente a análise taxonômica das comunidades microbianas (WU et al. 2010). Uma representação esquemática das etapas envolvendo um estudo metagenômico centrado no gene 16S-rDNA bacteriano é apresentada a seguir, na **Figura 4**.



Legenda - A figura mostra as principais etapas da avaliação da microbiota de diversas fontes utilizando a amplificação do gene 16S-rDNA bacteriano.

Figura 4 – Principais etapas de um estudo metagenômico baseado na amplificação do gene 16S-rDNA

1.3 MICROBIOTA HUMANA

A microbiota humana, ou seja, a coleção de microorganismos que vivem dentro e fora de nós, consiste em 100 trilhões de células que excedem nossas células humanas na proporção de 10:1 (SAVAGE 1997), e nos fornecem capacidades metabólicas que não possuímos em nosso genoma (GILL et al. 2006). Um estudo recentemente publicado pelo “The Human Microbiome Project Consortium” em 2012, avaliou a estrutura, função e diversidade da microbiota humana de 15 a 18 locais diferentes do corpo de 300 indivíduos saudáveis, avaliadas em dois intervalos de tempo diferentes. A função metabólica, os filos e espécies bacterianas mais abundantes, a estrutura microbiana, a variação temporal dos filos, entre outros dados, foram avaliados para determinar qual a microbiota existente em um ser humano saudável. Avaliando a mucosa oral, os filos mais abundantes encontrados em ordem decrescente foram: *Firmicutes*, *Proteobacteria*, *Bacteroidetes* e *Actinobacteria*.

Inúmeros benefícios são proporcionados pela flora endógena a seus organismos hospedeiros. Para citar alguns poucos exemplos, temos a síntese de certas vitaminas (como o complexo B e K), a prevenção contra a infecção por patógenos (através da competição direta por nichos, ou pela imunidade induzida por espécies comensais) e seu impacto no desenvolvimento normal do sistema imune (GILL et al. 2006). Bactérias da pele parecem ter papel crucial na prevenção da infecção por espécies patogênicas (revisto por KONG e SEGRE 2012), enquanto que bactérias da boca e do trato gastro-intestinal tem um importante papel na fisiologia digestiva, estando envolvidas com a quebra de carboidratos e proteínas (revisto em MACFARLANE e MACFARLANE 2012) podendo tanto prevenir, quanto causar infecções

(LAZAREVIC et al. 2010). O leite materno humano contém açúcares importantes para a manutenção de uma flora intestinal saudável em recém-nascidos (ZIVKOVIC et al. 2011) e também carrega uma grande diversidade de bactérias (pelo menos 600 espécies distintas) com papel na formação da microbiota intestinal (HUNT et al. 2011). Neste mesmo artigo, ao avaliarem 16 mulheres, cada uma fornecendo 3 amostras de leite em diferentes semanas, HUNT et al. (2011) definiram um *core milk microbiome* composto por 9 espécies, presentes em todas as amostras de todas as mulheres avaliadas.

1.3.1 Microbiota e Doenças

Estudos recentes apontam para uma relação causal entre os perfis da microbiota oral e doenças específicas, como a periodontite (FAVERI et al. 2008) ou entre a microbiota do sistema digestivo e a síndrome do cólon irritável (*Irritable bowel syndrome*- IBD), a diabetes e a obesidade (FRANK et al. 2007; TURNBAUGH et al. 2006). Evidências iniciais que apontavam um envolvimento da microbiota intestinal no desenvolvimento da IBD foram obtidas ao se observar uma melhora no curso da síndrome após o uso de certos antibióticos. Ao mesmo tempo, em modelos animais, foi observado que uma colonização prévia por bactérias causadoras de inflamação era necessária para a manifestação desta síndrome (BARNICH et al. 2007). Quando comparados a controles, pacientes com IBD apresentaram uma redução na diversidade dos filos *Firmicutes* e *Bacteroidetes* e um aumento de bactérias invasoras da mucosa, como *E. coli* (FRANK et al. 2007).

Já foi demonstrado por TURNBAUGH et al. (2006) que a microbiota intestinal de indivíduos obesos possui uma abundância relativa diferente dos filos

Firmicutes e *Bacteroidetes* quando comparada com a de indivíduos não-obesos, sendo que os indivíduos obesos possuem uma maior capacidade de absorção intestinal, mediada pela microbiota bacteriana. Estes autores também demonstraram que uma colonização por uma “microbiota obesa” em camundongos axênicos levou à um aumento significativo da gordura corpórea quando comparados a camundongos colonizados por uma “microbiota magra”.

1.3.2 A Microbiota e o Câncer

Evidências apontam que algumas bactérias são responsáveis pelo desenvolvimento do câncer. A mais bem aceita é a de que a bactéria *Helicobacter pylori* está associada ao desenvolvimento do câncer gástrico. Isto porque no primeiro estágio do câncer gástrico a mucosa normal se torna inflamada causando uma gastrite, que então pode progredir para uma atrofia, metaplasia, displasia e por fim pode resultar em um adenocarcinoma (HOOPER et al. 2009). A infecção por *H. pylori* está fortemente associada à gastrite superficial crônica (MARSHALL e WARREN 1984; DOOLEY et al. 1989; BLASER 1990), dando indícios de que poderia estar envolvida na iniciação e estimulação do adenocarcinoma gástrico em vários pacientes. Após vários testes clínicos e estudos em animais, esta hipótese foi confirmada. Deste modo, em 1994, a OMS definiu que uma infecção pela bactéria *Helicobacter pylori* era uma causa consolidada de câncer gástrico em humanos.

Dois grupos independentes recentemente observaram uma maior prevalência do gênero bacteriano *Fusobacterium* (KOSTIC et al. 2012) ou da espécie bacteriana *Fusobacterium nucleatum* (CASTELLARIN et al. 2012) em pacientes portadores de câncer colorretal. No estudo de CASTELLARIN et al. (2012), amostras de 11

tumores colorretais e 11 controles foram coletadas e foi realizado o *shotgun sequencing* do DNA na plataforma Illumina™. Para 9 das 11 amostras tumorais sequenciadas foi observado o dobro de sequências da espécie *Fusobacterium nucleatum*, quando comparadas aos controles. Posteriormente, uma maior prevalência desta bactéria foi confirmada em tumores colorretais por qPCR através do estudo de 99 casos câncer e do mesmo número de controles.

Diversas evidências, tanto biológicas quanto epidemiológicas, indicam que a presença de determinados tipos de vírus do papiloma humano (HPV) seja um evento precursor na gênese de neoplasias do cérvix, pênis e ânus (FRANCO 1991, 1992 e 1995). Porém, determinados tipos de HPV podem também possuir um papel importante em outros tipos de câncer. O papel do HPV no câncer oral e orofaríngeo foi primeiramente discutido por SYRJÄNEN et al. (1983). Estudos demonstraram a presença de DNA de HPV em biópsias de tumores localizados na cabeça e pescoço e, mais recentemente, estudos de larga escala da prevalência de DNA do HPV na mucosa oral têm demonstrado que o HPV pode ser considerado um fator de risco independente para *HNSCC* (HANSSON et al. 2005). Estudos mostraram que cerca de 25% dos casos de câncer oral estão associados com o HPV (BRANDWEIN et al. 1994). *HPVs* de alto risco, como os tipos 16, 18 e 31, são notórios por causarem a gênese de tumores em células epiteliais humanas (MARUR e FORASTIERE 2008). As oncoproteínas virais E6 e E7, provindas dos *HPVs* de alto risco, promovem a progressão de tumores inativando TP53 (GILLISON 2004).

HOOPER et al. (2006) estudaram as bactérias presentes no interior do tecido tumoral de 10 pacientes com *HNSCC* através da técnica de FISH (hibridização *in situ* fluorescente). Usando sondas de oligonucleotídeos universal para eubactérias, o

autores identificaram a presença de bactérias dentro do tumor, evidenciando que sua sobrevivência no microambiente é possível. Eles também avaliaram as bactérias presentes no tumor e no tecido normal adjacente por metagenômica, através da clonagem de produtos de PCR gerados por iniciadores que permitiram a amplificação de todo o gene 16S-rDNA e o seu sequenciamento utilizando o método Sanger. Setenta táxons bacterianos distintos foram encontrados, sendo 52 nos tecidos tumorais e 37 nos tecidos normais adjacentes. Destes, 4 táxons estavam presentes em pelo menos 30% dos tecidos controles, mas estavam ausentes nos tecidos tumorais: *Granulicatella adjacens*, *Porphyromonas gingivalis*, *Sphingomonas sp* e *Streptococcus mits/oralis*. Outros 3 táxons estavam presentes em pelo menos 30% dos tecidos tumorais, mas ausentes nos controles: *Fusobacterium naviforme*, *Ralstonia insidiosa* e *Clavibacter michiganensis*. As espécies encontradas nos tecidos tumorais são sacarolíticas e acidúricas.

Em um trabalho realizado por NAGY et al. (1998) bactérias presentes em 10 amostras de biofilme retiradas da superfície de tumores da cavidade oral e 10 amostras de biofilme retiradas da mucosa saudável dos mesmos pacientes, foram avaliadas através do método de cultivo. Os gêneros bacterianos mais abundantes encontrados nos biofilmes tumorais foram; *Veillonella*, *Fusobacterium*, *Prevotella*, *Porphyromonas*, *Actinomyces*, *Clostridium*, *Haemophilus*, *Enterobacteriaceae* e *Streptococcus*.

1.4 A MICROBIOTA ORAL

1.4.1 Biofilmes e Saliva

A cavidade oral possui uma grande diversidade de flora bacteriana, com importante variação quantitativa e qualitativa entre indivíduos. Usando métodos tradicionais e técnicas moleculares, mais de 700 espécies bacterianas já foram identificadas na cavidade oral humana (PASTER et al. 2006). Mais de 400 destas espécies foram encontradas na bolsa periodontal e as 300 remanescentes foram identificadas de outras áreas da cavidade oral como a língua, mucosas orais, lesões causadas por cáries e lesões endodônticas. PASTER et al. (2001) já sugeriam a presença de pelo menos 100-200 espécies de bactérias na cavidade de qualquer indivíduo normal e sugeriam grande diversidade. Estudos iniciais evidenciaram restrita estruturação geográfica da comunidade microbiana oral, porém gêneros como *Serratia* e *Enterobacter* variam significativamente em localizações geográficas diferentes (NASIDZE et al. 2009a). Estas variações potencialmente incluem bactérias importantes no desenvolvimento de carcinomas orais. Uma comparação entre as floras bacterianas encontradas em cavidades orais saudáveis e doentes sugere a existência de um perfil padrão para a microbiota oral em cavidades orais saudáveis (AAS et al. 2005), e que uma correlação entre a estrutura e função da microbiota oral e o câncer possa existir (MEURMAN 2010; FLEMMIG e BEIKLER 2011).

Biofilmes são colônias de microrganismos complexas, que se encontram predominantemente em superfícies tanto duras, quanto mucosas, da cavidade oral (FLEMMIG e BEIKLER 2011). Em condições saudáveis, um balanço entre a composição microbiana e sua atividade mantém os biofilmes saudáveis e estáveis,

porém a separação e eliminação dos biofilmes da mucosa e das superfícies duras é necessária para que espécies patogênicas não comecem a proliferar e causar doenças (FILOCHE et al. 2010). Se espécies bacterianas potencialmente patogênicas se encontrarem em números mais elevados do que as espécies comensais nos biofilmes, elas são eliminadas através do deslocamento dos biofilmes, para que um balanço saudável possa novamente ser estabelecido. Uma boa higiene oral e o fluxo salivar são necessários para que esse deslocamento ocorra, porém a maioria dos pacientes portadores de câncer oral possuem uma higiene bucal ruim (MEURMAN 2010), o que pode contribuir para a manifestação de espécies patogênicas em seus biofilmes.

Um estudo realizado por MANS et al. (2009), mostrou que uma mono infecção por *Porphyromonas gingivalis* em uma cultura de células de queratinócitos gengivais humanos imortalizados (HIGK) alterava seu transcriptoma de maneira significativa, quando comparado com uma co/mono infecção com *Streptococcus gordonii* (espécie comensal/comum da flora bacteriana). Eles também mostraram que 45 genes envolvidos no ciclo celular estavam diferencialmente expressos nas diferentes infecções (mono/co). Além disso, os autores também demonstraram que genes associados com ao câncer foram diferentemente impactados com uma mono infecção por *P. gingivalis*.

No estudo de LAZAREVIC et al. (2009), amostras de saliva de 3 indivíduos foram coletadas e a região V5 do gene 16S-rDNA foi amplificada por PCR. Os produtos foram sequenciados na plataforma Illumina™ gerando 1,4 milhões de sequências que, após análise bioinformática, permitiram a identificação de 135 gêneros de bactérias, sendo predominantes os filos *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Fusobacteria*, *TM7* e *Spirochaetes*.

Um outro estudo realizado por LAZAREVIC et al. (2010) estudou as diferenças nas comunidades microbianas da saliva de indivíduos normais, verificando a existência de variações intra- ou inter-indivíduos. Eles estudaram também se há uma mudança significativa na microbiota salivar ao longo do tempo. Através do uso da metagenômica e amplificação das regiões variáveis V1-V3 do gene 16S-rDNA, eles coletaram amostras de saliva de 5 indivíduos em três momentos ao longo de um período de 29 dias, fazendo o sequenciamento na plataforma 454™(Roche, EUA). Foram obtidas 50.708 sequências, que após filtragem geraram um *set* de 31.170 de sequências úteis. Os filos mais abundantes entre todas as amostras eram *Firmicutes* e *Proteobacteria*. Um total de 81 gêneros foram identificados, sendo que 19 não haviam sido identificados anteriormente na microbiota oral, nem estavam listados no *Human Oral Microbiome Database* (HOMD). Eles concluíram que a comunidade microbiana da saliva é estável por pelo menos 5 dias, e que amostras colhidas do mesmo indivíduo em intervalos de tempo mais próximos não eram necessariamente mais similares, em termos de comunidade microbiana, do que amostras colhidas em intervalos de tempo maiores. Eles concluem que isto indica que há uma persistência de táxons específicos de um dado sujeito e cuja frequência pode flutuar entre determinados intervalos de tempo.

Em um estudo de NASIDZE et al. (2009b), amostras de saliva de 120 indivíduos de diferentes continentes foram coletadas e uma região do gene 16S-rDNA foi sequenciada após a clonagem dos fragmentos, gerando um total de 14.691 sequências (122 sequências por amostra). Análises deste limitado *set* de sequências sugeriram que cada indivíduo albergava de 6 a 30 diferentes gêneros de bactérias, cenário que hoje se mostra claramente sub-estimado. NASIDZE et al. (2009b)

demonstraram haver uma grande diversidade microbiana entre indivíduos, porém esta demonstrou pequena estruturação geográfica.

BELDA-FERRE et al. (2012) estudaram a microbiota bacteriana da cavidade oral em condições “saudáveis” e na presença de cáries. Eles sequenciaram 8 amostras de placa dentária de indivíduos com diferentes *status* de saúde oral, utilizando o *shotgun sequencing*. O sequenciamento na plataforma 454™ gerou aproximadamente 1,9 milhões de sequências e eles observaram que nos indivíduos sem cáries havia uma maior frequência de bactérias aeróbicas (táxons pertencentes as classes *Bacilli* e *Gammaproteobacteria*) e nos indivíduos com cáries bactérias anaeróbicas, tais como as pertencentes ao filo *Bacteroidetes* e à ordem *Clostridiales*, eram mais frequentes.

No estudo de COSTELLO et al. (2009), a microbiota da cavidade oral, do intestino, do canal auditivo, interior das narinas, cabelo e superfícies da pele foi estudada em 3 dias diferentes para observar se havia uma variação significativa entre os indivíduos estudados, entre os locais amostrados e também se havia uma variação da microbiota existente no mesmo indivíduo em intervalos de tempos diferentes. Os filos mais abundantes encontrados, quando todos os locais amostrados foram agrupados, foram: *Actinobacteria* (36.6%), *Firmicutes* (34.3%), *Proteobacteria* (11.9%) e *Bacteroidetes* (9.5%). Eles observaram que cada local abrigava uma microbiota característica e os táxons mais abundantes permaneciam estáveis entre diferentes indivíduos e em intervalos de tempo diferentes. Também foi observado que as comunidades bacterianas de uma dada topografia possuíam uma maior similaridade entre si, em indivíduos distintos, do mesmo sexo, ou do mesmo intervalo de tempo, indicando que há uma especificidade na colonização de bactérias para cada região do corpo humano. As comunidades microbianas se agruparam

primeiramente por local amostrado, depois por indivíduo e por último, pelo tempo da amostragem. O local do corpo que possui a menor variação na comunidade microbiana tanto entre indivíduos, quanto no mesmo indivíduo em intervalos diferentes, era a cavidade oral, sugerindo uma maior especificidade e uma menor variação na comunidade bacteriana oral.

KEIJSER et al. (2008) estudaram a microbiota salivar de 71 indivíduos saudáveis e também a microbiota encontrada em placas dentárias de 98 indivíduos saudáveis. Eles amplificaram a região V6 do gene 16S-rDNA bacteriano por PCR e sequenciaram os *amplicons* na plataforma 454™. Eles obtiveram no total 197.600 sequências e os filos bacterianos mais abundantes encontrados na saliva foram; *Firmicutes* (41%), *Bacteroidetes* (27%), *Proteobacteria* (22%), *Actinobacteria* (6%) e *Fusobacteria* (6%). Os filos mais abundantes encontrados na placa dentária foram; *Firmicutes* (32%), *Actinobacteria* (25%), *Proteobacteria* (16%), *Bacteroidetes* (15%) e *Fusobacteria* (9%). Eles observaram que os filos *Actinobacteria* e *Fusobacteria* eram mais representativos nas amostras de placas dentárias, enquanto que os filos *Bacteroidetes*, *Firmicutes* e *Proteobacteria* eram mais abundantes nas amostras da saliva, sugerindo diferenças significativas entre esta e os biofilmes dos dentes.

1.4.2 Alterações da Microbiota Oral na Presença de Álcool e/ou Tabaco

Evidências recentes, baseadas na análise de algumas poucas espécies bacterianas, demonstraram que o uso de cigarros leva à seleção do crescimento de algumas espécies. Um exemplo é o achado de que algumas variantes de *Staphylococcus aureus* não crescem na presença de condensados provindos da

fumaça do cigarro (HOOPER et al. 2009). Alguns isolados de *S. aureus* resistentes ao alcatrão foram capazes de induzir mediadores inflamatórios, como o fator de necrose tumoral alfa (TNF- α) *in vitro* (FUJIKI et al. 2004). O aumento da produção de TNF- α tem um importante papel no processo de inflamação e têm sido bastante implicado na carcinogênese oral (JENG et al. 2003).

O consumo de álcool também pode mudar a composição da microbiota oral. Um estudo recente demonstrou que os níveis de *Streptococcus anginosus* na saliva se encontram significativamente elevados em alcoólatras, quando comparados com não-alcoólatras (MORITA et al. 2005). Esta variação da composição da microbiota bacteriana pode influenciar positiva ou negativamente as bactérias que produzem acetaldeído, o primeiro metabólito do etanol, que tem propriedades carcinogênicas tanto em modelos animais quanto *in vitro* (HOOPER et al. 2009). Já foi demonstrado que algumas bactérias encontradas na microbiota oral produzem acetaldeído a partir do etanol, tais como os *Streptococci*, algumas bactérias gram-positivas aeróbicas e leveduras (SALASPURO 2003). O acetaldeído é capaz de produzir efeitos mutagênicos como *cross-linking* do DNA, aneuploidias e aberrações cromossômicas (HOMANN et al. 1997), de uma maneira dependente de enzimas do epitélio e também produzidas pela microbiota oral (HOOPER et al. 2009).

Algumas espécies de *Neisseria* presentes na boca possuem altos níveis de desidrogenases, além de serem capazes de produzir quantidades significantes de acetaldeído na presença de etanol (MUTO et al. 2000). Alguns estudos demonstraram que o nível de acetaldeído encontrado em fumantes expostos ao etanol é sete vezes maior quando comparado a não fumantes (SALASPURO e SALASPURO 2004). Portanto, o acetaldeído é um importante fator na etiologia do

câncer oral e como a microbiota oral possui um papel importante na produção de acetaldeído, é de suma importância que estudemos os microrganismos encontrados na cavidade oral para melhor compreender a patogênese desta doença.

No trabalho de HOMANN et al. (2000), os níveis de acetaldeído e a flora bacteriana de 26 pacientes com câncer da cavidade oral, 64 alcoólatras e 114 indivíduos controles foram investigados para encontrar alguma associação entre altas taxas de produção de acetadeído e bactérias presentes na saliva das amostras. Os autores observaram que fumar e beber era um forte indicador da produção de acetaldeído a partir da flora bucal. O uso de álcool demonstrou forte associação com a produção de acetaldeído. Analisando a microbiota, através do uso do método de cultivo, de 10 indivíduos com uma alta produção de acetaldeído e 10 com uma baixa produção, os autores observaram que bactérias aeróbicas estavam mais presentes na saliva de indivíduos com um alta produção de acetaldeído e as bactéria aeróbicas mais comumente encontradas foram do gênero *Streptococcus*.

1.4.3 Alterações da Microbiota Oral em Pacientes com Câncer Pral

O trabalho de MAGER et al. (2005) investigou se a presença de certas espécies bacterianas encontradas na cavidade oral poderiam ser indicadoras de OSCC. Este grupo coletou saliva de 45 pacientes com OSCC e 229 controles, a fim de avaliarem a abundância de 40 espécies bacterianas comumente encontradas na cavidade oral através da hibridização de DNA da saliva com sondas de regiões variáveis do gene 16S-rDNA específicas destas espécies como alvo. Os resultados indicaram 3 espécies bacterianas significativamente elevadas nos pacientes com OSCC: *Prevotella melaninogenica*, *Streptococcus mitis* e *Capnocytophaga*

gingivalis. As limitações deste interessante estudo se referem à escala das análises que se restringiram a 40 espécies bacterianas, não permitindo uma análise mais aprofundada da microbiota oral.

PUSHALKAR et al. (2011) avaliaram a microbiota salivar de 5 indivíduos com OSCC e de 2 controles, usando iniciadores que cobriam a região V4-V5 do gene 16S-rDNA e geraram 54 mil sequências. No total, foram encontrados 8 filos bacterianos com predominância de *Firmicutes* (45%) e *Bacteroidetes* (25%) em ambos os grupos. Os outros filos encontrados nos dois grupos foram; *Actinobacteria* (14%), *Proteobacteria* (10%), *Fusobacteria* (5%), *SR1* (0.6%), *Spirochaetes* (0.2%) e *TM7* (0.2%). Os gêneros mais abundantes encontrados nos pacientes foram: *Gemella*, *Streptococcus*, *Rothia*, *Porphyromonas* e *Lactobacillus*. Nos controles, os gêneros bacterianos mais abundantes foram: *Prevotella*, *Neisseria*, *Leptotrichia*, *Capnocytophaga*, *Actinobacillus* e *Oribacterium*.

No entanto, vemos que até hoje os estudos que avaliaram alterações da microbiota no câncer da cavidade oral possuem algumas limitações. Em nosso trabalho procuramos contribuir com esta nova e importante área de conhecimento.

2 HIPÓTESE E OBJETIVOS

A microbiota da mucosa oral pode estar alterada pelo uso frequente de álcool e tabaco, e estas alterações podem influenciar no desenvolvimento de câncer da cavidade oral.

2.1 OBJETIVO GERAL

Avaliar a composição da flora bacteriana da cavidade oral utilizando abordagem metagenômica, buscando investigar diferenças presentes em pacientes com OSCC.

2.2 OBJETIVOS ESPECÍFICOS

- Padronizar a amplificação e o sequenciamento de porções informativas do gene 16S-rDNA bacteriano por PCR, a partir de amostras da mucosa oral.
- Avaliar as alterações quantitativas e qualitativas da flora oral nos três grupos de estudo, utilizando técnicas de metagenômica baseadas na sequência hipervariável do gene codificador do 16S-rRNA.

3 MATERIAIS E MÉTODOS

3.1 AMOSTRAS

Foi aplicado um questionário a cada um dos indivíduos participantes, a fim de determinar seus hábitos alimentares, histórico e frequência de uso de bebidas alcoólicas e tabaco, qual o tipo de higienização bucal comumente realizada, sua frequência e se o indivíduo fazia ou não uso de próteses dentárias. Este estudo foi feito com amostras obtidas de um total de 26 indivíduos distribuídos em três grupos, conforme **Quadro 1**, apresentado abaixo.

Quadro 1 – Grupos de estudo deste trabalho

Grupo	Características das Amostras
Controles (C)	Grupo controle composto por 10 indivíduos que não bebem nem fumam e não possuem câncer da cavidade oral
Controles que fumam e bebem (CFB)	Grupo controle composto por 5 indivíduos que bebem diariamente, fumam mas não possuem câncer da cavidade oral
Pacientes (P)	11 pacientes que fumam, bebem e possuem câncer de cavidade oral.

3.2 CRITÉRIOS DE INCLUSÃO/EXCLUSÃO

Os critérios de inclusão e exclusão, assim como as características dos indivíduos de cada grupo são descritos em maior detalhe a seguir, no **Quadro 2**:

Quadro 2 – Detalhamento dos grupos e critérios de inclusão

Categorias	Grupo 1- Controles I	Grupo 2- Controles II	Grupo 2- Pacientes
Uso de Álcool e de Tabaco	Nunca fumou; ingere menos de 1 dose de bebida alcoólica por dia.	Fuma 20 cigarros/dia por pelo menos 10 anos; ingere pelo menos 1 dose de bebida alcoólica por dia.	Fuma 20 cigarros/dia por pelo menos 10 anos; ingere pelo menos 1 dose de bebida alcoólica por dia.
Doença	Sem evidência de câncer e sem histórico familiar de câncer de cabeça e pescoço. Ausência de outros tumores	Sem evidência de câncer e sem histórico familiar de câncer de cabeça e pescoço. Ausência de outros tumores	Presença de câncer da cavidade oral (exceto lábio). Ausência de outros tumores. Ainda não tratado (quimioterapia, radioterapia ou cirurgia).
Etnia	Não Asiático	Não Asiático	Não Asiático
Moradia	São Paulo (últimos 5 anos)	São Paulo (últimos 5 anos)	São Paulo (últimos 5 anos)
Hábitos de Higiene Oral	Não faz uso diário de enxaguante bucal com álcool	Indiferente	Indiferente
Uso de Antibióticos	Não faz uso há pelo menos 3 meses	Não faz uso há pelo menos 3 meses	Não faz uso há pelo menos 3 meses
Idade	Maior que 40 anos	Maior que 40 anos	Maior que 40 anos

Notas – Critérios foram adotados para reduzir a influência de alguns fatores na composição da flora oral. Por exemplo, decidimos excluir indivíduos de origem Asiática, pois este grupo por vezes tem hábitos alimentares distintos que poderiam causar alterações mais marcantes na microbiota oral, levando a um fator de confusão. Do mesmo modo, incluímos apenas indivíduos que residem em SP pelos últimos 5 anos, a fim de reduzir a possibilidade de uma variação da microbiota devida a efeitos da culinária regional.

As características individuais de cada um dos pacientes recrutados e avaliados por sequenciamento neste estudo são apresentados nos **Quadros 3, 4 e 5**.

Quadro 3 - Características gerais de indivíduos que compõe o Grupo de Controles (C)

Amostra	Idade (anos)	Sexo	Cor da pele	Peso (Kg)	Altura (cm)	Uso de tabaco	Uso de álcool	Uso de enxaguante bucal
C01	60	M	Branca	84	167	Nunca	Sim e continua: 350ml cerveja ou 50ml de vinho/mês	Dia-sim/Dia-não (sem álcool)
C04	57	M	Branca	85	188	Nunca	Sim e continua: bebe 50ml de vinho/mês desde os 17 anos de idade	Nunca
C06	61	M	Branca	74	164	Nunca	Nunca	Nunca
C07	71	M	Branca	-	-	Nunca	Nunca	Nunca
C09	70	M	Parda	54	163	Nunca	Nunca	Nunca
C15	50	M	Branca	80	174	Nunca	Sim e continua: 350ml de cerveja/mês	Nunca
C16	48	F	Branca	82	160	Nunca	Nunca	Nunca
C19	51	M	Branca	81	180	Nunca	Nunca	Nunca
C20	53	F	Branca	63	156	Nunca	Sim e continua: 100ml 2x/mês de vinho desde os 25 anos.	<1x/semana (sem álcool)
C21	55	M	Branca	100	171	Nunca	Sim e continua: 400ml de vinho 3x/semana há 3 anos	Nunca

Quadro 4 - Características gerais de indivíduos que compõe o Grupo de Controles que Fumam e Bebem (CFB)

Amostra	Idade (anos)	Sexo	Cor da pele	Peso (Kg)	Altura (cm)	Uso de tabaco	Uso de álcool	Uso de enxaguante bucal
CFB10	59	M	Branca	93	180	Sim e continua: 20-30 cigarros/dia desde os 47 anos.	Sim e continua: 700ml de cerveja/dia desde os 20 anos de idade	Nunca
CFB18	47	M	Branca	85	179	Sim e continua: 20 cigarros/dia desde os 20 anos de idade	Sim e continua 700ml de cerveja/dia, 1 litro de vinho/mês e 1 litro destilados/mês	Nunca
CFB29	67	M	Branca	61	163	Sim e continua: 6 cigarros/dia desde os 15 anos de idade	Sim e continua: 1.2 litros de cerveja/dia desde os 16 anos de idade	2x/dia
CFB30	54	M	Branca	78	176	Sim e continua: 30-40 cigarros/dia desde os 12 anos de idade	Sim e continua: Até 3 meses atrás: 1 litro de destilados/dia, hoje bebe 1 litro de destilados/semana desde os 14 anos de idade	1x/dia
CFB31	55	F	Branca	68	165	Sim e continua: 10 cigarros/dia desde os 17 anos de idade	Sim, parou há 1 mês: 1 litro de cerveja/dia e 300ml de vinho/semana desde os 17 anos de idade.	2x/semana

Quadro 5 - Características gerais dos indivíduos do Grupo de Pacientes (P) diagnosticados com OSCC e avaliados neste estudo

Amostra	Idade	Sexo	Cor da pele	Peso (Kg)	Altura (m)	Uso de tabaco	Uso de álcool	Estadiamento tumoral	Uso de enxaguante bucal
P01	68	M	Branca	76	1,72	Sim e continua: 20 cigarros/dia há 55 anos	Sim e continua: Semanalmente (Cerveja e Destilados)	T3N0M0	Dia Sim/Dia Não
P03	73	M	Branca	-	-	Sim e continua: 20 cigarros/dia há 66 anos	Sim e continua: Diariamente Cerveja e destilados há 58 anos	T3N1M0	Menos de 1x/semana
P04[§]	57	M	Branca	81	1,75	Sim e continua: 20 cigarros/dia há 44 anos	Sim e continua: Mensalmente 1 ou 2 cervejas por 5 anos.	T1N0M0	Menos de 1x/semana
P08	60	M	Branca	62	1,66	Sim e continua: fuma há 46 anos	Sim e continua: 1 litro de destilados/semana, 350ml de cerveja/semana e 300ml de vinho/mês	T4N2Mx	Nunca
P09	46	M	Branca	69	1,76	Sim e continua: 20 cigarros/dia há 31 anos	Sim e continua: 1.5 litros de cerveja/dia, quatro doses de destilados/dia	T4N1M0	Nunca
P11	68	M	Parda	66	1,65	Sim e continua: 20-30 cigarros/dia desde os 17 anos	Sim e continua : 1.5 litros de cerveja/dia e três doses de destilados/semana	T2N0M0	Nunca
P23	50	M	Branca	74	1,70	Sim e continua: fuma 20 cigarros/dia desde os 7 anos de idade	Sim e continua: 330 ml de destilados/dia desde os 7 anos de idade	T2N0M0	Nunca
P27	52	M	Branca	95	1,89	Sim e continua: Fuma 40 cigarros/dia desde os 23 anos de idade	Sim e continua: 1.5 litros de cerveja/dia e 400ml de vinho/semana desde os 16 anos de idade	T2N0M0	Nunca
P33	52	F	Branca	52	1,57	Sim e continua: Fuma 30 cigarros/dia dos 37 aos 52 anos de idade	Sim e continua: 1.5 litros de cerveja/dia	T4N2M0	Nunca
P36*	43	M	Branca	46	1,65	Sim, parou há 3 meses: fumava 20 cigarros/dia dos 17 aos 43 anos de idade	Sim, parou há 3 meses: 4-5 doses de destilados/dia e 600ml de cerveja/dia desde os 17 anos de idade	T2N2M0	1x/dia
P46*	53	M	Parda	63	1,72	Sim e continua: Fuma 40 cigarros/dia desde os 15 anos de idade	Sim, parou há 3 meses: 5-6 doses/dia de destilados e 12 latinhas/semana de cerveja desde os 15 anos de idade	T1N1M0	1x/dia

§ - indica paciente positivo para HPV; * - indica os pacientes derivados do Hospital Heliópolis, São Paulo, SP.

3.3 COLETA DE AMOSTRAS DO BIOFILME

3.3.1 Obtenção do DNA

A obtenção das amostras de DNA coletadas na cavidade oral foi feita através de um *swab* bucal de cada indivíduo. Este *swab* (Whatman) possui uma esponja coletora que foi passada em regiões pré-estabelecidas da cavidade oral dos pacientes e controles. As regiões amostradas foram: base da língua, assoalho da boca, lateral da língua, gengiva inferior e porção interna das bochechas, de tal modo a coletar o máximo de amostras de biofilme e também obter uma boa representação das bactérias existentes nos locais onde o carcinoma epidermóide é mais frequente. O material coletado no *swab* foi transferido para um cartão FTA Elute (Whatman) nos dois círculos superiores do cartão, deixando os dois círculos inferiores para serem usados como controles negativos da reação. Este cartão possui propriedade químicas patenteadas que causam a lise imediata das células e protegem o DNA extraído da ação de nucleases e contaminação por microrganismos presentes no local da coleta. As amostras foram coletadas somente após a assinatura de um termo de consentimento livre e esclarecido (Anexo 3), previamente aprovado pelos comitês de ética em pesquisas do Hospital AC Camargo (protocolo 1459/10), pelo comitê de ética do Hospital Heliópolis (protocolo CEPHH-817) e pelo comitê de ética em pesquisa do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (protocolo 0895/11). Vinte e três controles que não fumavam ou bebiam foram coletados, dos quais 20 atingiram os critérios de inclusão. Dentre os controles que fumam e bebem 34 indivíduos foram coletados mas apenas 5 atingiam os critérios de inclusão previamente estabelecidos. Para os pacientes portadores de câncer da

cavidade oral, 62 indivíduos tiveram amostras coletadas mas apenas 20 atingiram os critérios de inclusão previamente estabelecidos.

3.3.2 Eluição do DNA

Quatro discos de 3mm foram retirados dos cartões FTA, lavados em H₂O estéril e incubados a 95°C por 25 minutos em 30µl de H₂O estéril. Após agitação vigorosa em vórtex, os discos foram retirados do eluato que continha o DNA extraído das amostras. Este DNA foi usado para as reações de PCR subsequentes.

3.4 GENOTIPAGEM DO HPV DOS PACIENTES

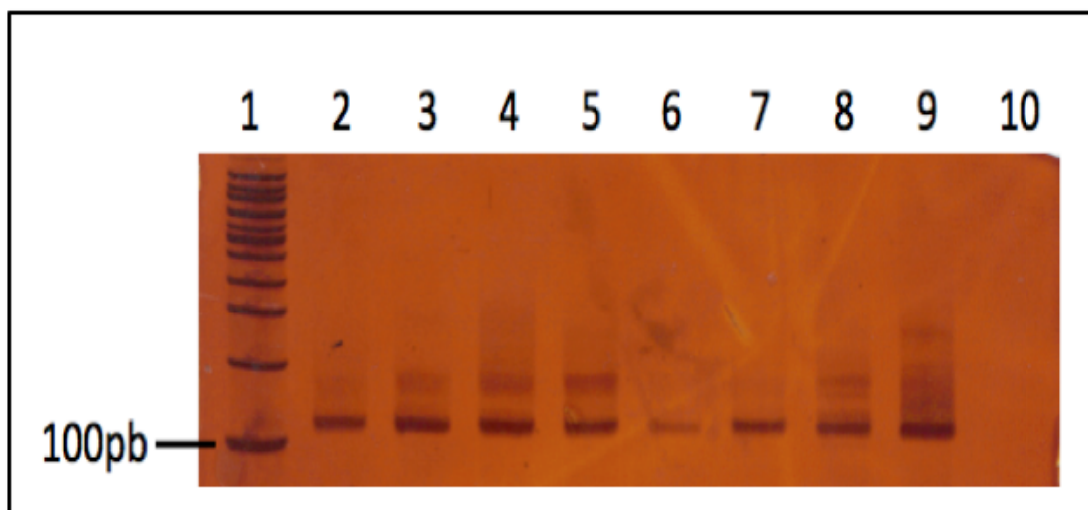
O kit Inno-Lipa HPV Genotyping *Extra* (Innogenetics, Bélgica) foi usado para investigar a presença HPV em amostras de pacientes. O status de positividade de HPV dos 11 pacientes foi avaliado utilizando o DNA eluído do cartão *FTA elute*. Este kit permite a identificação de 54 tipos de HPV, e apresenta uma maior robustez, permitindo trabalhar com amostras de qualidade questionável (tais como DNA extraído de tecido parafinado). Este kit consiste na amplificação por PCR do DNA do vírus, eventualmente presente na amostra, com iniciadores específicos que cobrem as regiões GP5 e GP6, amplificando um fragmento de 65 nt. Após a PCR, os fragmentos amplificados são desnaturados e hibridados à sondas que são complementares às regiões amplificadas do vírus encontradas em fitas. Posteriormente, um ensaio imunoenzimático é realizado, assim se houver a presença do DNA do HPV a sonda fica corada pois houve hibridização, indicando a positividade para o vírus para aquela determinada amostra. Iniciadores específicos para a beta-globina também são

adicionados à reação como controles positivos da amostra avaliada. Os resultados foram obtidos com o auxílio de nossos colaboradores do Instituto de Medicina Tropical da Faculdade de Medicina da Universidade de São Paulo, sob a coordenação do Prof. Dr. José Eduardo Levy.

3.5 AMPLIFICAÇÃO DA REGIÃO V1 DO GENE 16S-RDNA

Foram usados iniciadores que cobriam a região hiper-variável 1 do gene 16S-rDNA bacteriano, descritos por SUNDQUIST et al. (2007) com um tamanho médio de *amplicon* de 113 nt (iniciadores: V1F-5' AGAGTTTGATCMTGGCTCAG- 3' e V1R-5' TTACTCACCCGTICGCCRCT-3'). Tais iniciadores foram escolhidos por já terem sido usados com sucesso em outros estudos de avaliação da microbiota, e por produzirem *amplicons* de tamanho adequado para o sequenciamento na plataforma aqui utilizada. Nas amplificações por PCR, usamos o sistema *master mix* com uma enzima de alta fidelidade (*High-Fidelity*, Roche), empregada para reduzir manipulações durante o preparo da reação, e para ajudar a reduzir possíveis erros de incorporação de nt pela DNA polimerase. No total, amostras da microbiota de 26 indivíduos foram usadas nas amplificações, sendo: 10 controles sem neoplasia, sem histórico de uso de tabaco e com registro de uso de álcool inferior a 1 vez por mês (Grupo C); 5 controles sem neoplasia, mas com uso diário de tabaco e álcool (Grupo CFB) e 11 pacientes acometidos com OSCC, todos com uso diário de tabaco e álcool (Grupo P). Para cada amostra foram feitas seis reações de amplificação, em volume final de 10µl. Posteriormente 2µl de cada reação foram avaliados em géis de poliacrilamida 8% corados com prata (SANGUINETTI et al. 1994) como ilustrado

na **Figura 5**. *Amplicons* de um mesmo indivíduo foram agrupados e quantificados no fluorímetro Qubit (Invitrogen, Carlsbad, EUA).

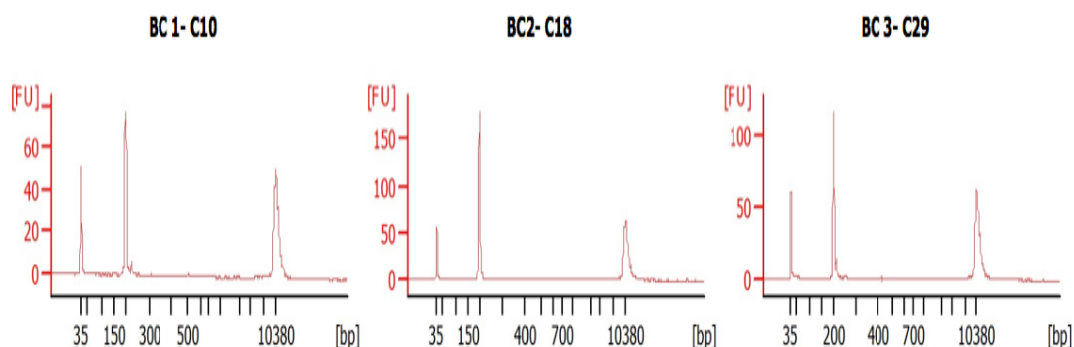


Legenda - A figura mostra um gel de poliacrilamida 8% corado com prata, contendo os *amplicons* gerados para cada amostra. **1**-Ladder 100pb; **2**- C06; **3**-C07; **4**-C09; **5**-C16; **6**-C19; **7**-C21; **8**-P03; **9**- Controle Positivo (DNA extraído de *E. coli*); **10**- Controle Negativo

Figura 5 – Visualização dos produtos de amplificação da região V1 do gene 16S-rDNA em gel de poliacrilamida 8% corado com sais de prata.

3.6 LIGAÇÃO DOS ADAPTADORES, ePCR E SEQUENCIAMENTO

Antes da ligação dos adaptadores para o sequenciamento, as amostras foram purificadas com as *Ampure XP Beads* (Beckman Coulter, Indianapolis, EUA), para a remoção de remanescentes de iniciadores, enzima, DNA e proteínas. Após a purificação, os adaptadores para o sequenciamento na plataforma *Ion Torrent™* foram ligados usando kits da Applied Biosystems, seguindo os protocolos padrão. Posteriormente, *barcodes* específicos de 10 nts foram adicionados a cada amostra e a concentração dos *amplicons* foi novamente determinada com o fluorímetro Qubit. A **Figura 6** mostra os tamanhos dos *amplicons* obtidos após a adição dos *barcodes* e dos iniciadores de sequenciamento.



Legenda - A figura mostra os perfis dos tamanhos dos *amplicons* apresentados depois da adição dos *barcodes* e iniciadores de sequenciamento utilizando 1µl de amostra.

Figura 6 – Avaliação de amostras em chip de DNA *High Sensitivity* no *Bioanalyzer*

No total foram feitas 6 corridas de sequenciamento em 5 chips tipo 314 independentes e também em um único chip 316. As amostras foram distribuídas da seguinte maneira: **Corrida 1** - 2 pacientes e 2 controles I; **Corrida 2** - 5 pacientes; **Corrida 3** - 2 controles I e 2 pacientes; **Corrida 4** - 4 controles I; **Corrida 5** - 2 controles I e 2 pacientes; **Corrida 6**- 5 controles II (**Tabela 3**).

Nas seis corridas, após a adição dos *barcodes* e a determinação das concentrações das amostras, quantidades equimolares de cada amostra foram combinadas para formar um *pool* para a ePCR. A emulsão foi quebrada e as *beads* que continham os adaptadores e *amplicons* foram recuperadas, enriquecidas e posteriormente depositadas nos chips.

3.7 ANÁLISE TAXONÔMICA

Para a análise das sequências foi utilizado o programa *Qiime* (*Quantitative Insights Into Microbial Ecology* - CAPORASO et al. 2010b). Para isto, avaliamos

cada etapa da análise, e alteramos alguns parâmetros originais, buscando adequá-los às sequências geradas. Com isto definimos o seguinte *pipeline* de análise:

1. ***Picking OTUs (Original Taxonomic Units) pick_otus.py*** - **Script usado para a formação dos *clusters*.** Método: Uclust (Robert Edgar, unpublished, 2009): Cria sequências chamadas “sementes” que formarão um determinado grupo (*cluster*) de acordo com um mínimo de identidade (97%). Este *cluster* é denominado *OTU (Operational Taxonomic Unit)*, e será usado para atribuições taxonômicas nos próximos passos.
2. ***pick_rep_set.py*** – **Script usado na escolha de uma sequência representativa de cada *cluster*.** Método: A sequência mais abundante dentro de um determinado *cluster* é usada como representativa deste *cluster*, sendo usada nas análises subsequentes. A informação do tamanho do *cluster* (número de sequências que o compõe) é mantida e indica a abundância da respectiva *OTU* na amostra.
3. ***align_seqs.py*** - **Alinha as sequências contra um banco de dados pré-determinado.** Método: PYNAST (CAPORASO et al. 2010a): Um alinhador do tipo NAST (DESANTIS et al. 2006b). Tamanho mínimo na região do alinhamento: 50 nt. Porcentagem mínima de identidade com o banco de dados: 75%. O banco de dados aqui usado foi o *Greengenes* (Um banco de dados com 408.135 sequências alinhadas do gene 16S-rDNA bacteriano de alta qualidade que foram filtradas do banco do *National Center for Biotechnology Information (NCBI)* (DESANTIS et al. 2006a). Sequências que não alinham contra este banco com a porcentagem mínima de identidade determinada foram excluídas das consequentes análises.

4. ***assign_taxonomy.py*** - Indica a determinação taxonômica de uma determinada sequência. Método: BLAST: Porcentagem de identidade entre as sequências do banco *Ribosomal Database Project* (um banco de dados que contém 1.379.424 de sequências do gene 16S-rDNA de bactérias e arqueobactérias) com as sequências representativas dos clusters: 90%.
5. ***make_otu_table.py*** – Script que faz uma tabela com cada OTU, com sua respectiva taxonomia e indica sua presença em cada uma das amostras, assim como a frequência encontrada- Somente OTUs que possuem um mínimo de 3 sequências foram mantidas para as análises posteriores.
6. ***summarize_taxa_through_plots.py*** - Script que produz gráficos de área e de barras para cada amostra com a taxonomia encontrada.
7. ***alpha_diversity.py*** - Métrica: Espécies Observadas - Calcula o número de espécies únicas observadas com um determinado número de sequências.

3.8 ESTIMANDO A RIQUEZA DE ESPÉCIES BACTERIANAS

Para estimar se o número de espécies que estávamos observando nas nossas amostras estava coerente entre elas e com os dados da literatura, utilizamos o programa Catchall (BUNGE et al. 2012) para estimar a riqueza das espécies em nossas amostras. O arquivo de entrada para este programa é uma contagem da frequência em que as espécies (no caso deste trabalho utilizamos os OTUs) aparecem em uma determinada amostra. O programa testa diversos modelos matemáticos, tais como *Poisson* e *Chao*, baseados na frequência e distribuição das espécies, afim de determinar qual o modelo que melhor estima a riqueza, ou seja, o número de espécies

esperadas de uma determinada amostra. Ele gera então uma tabela aonde mostra qual foi o melhor modelo matemático adequado para aquela determinada amostra, o número total de espécies observadas e esperadas, e os resultados da riqueza de espécies obtidos com os outros modelos matemáticos. Neste trabalho, utilizamos o número de espécies esperadas que foi gerado pelo melhor modelo matemático escolhido pelo programa. Estes dados foram então plotados em gráficos e comparados.

4 RESULTADOS E DISCUSSÃO

4.1 CARACTERÍSTICAS GERAIS DOS INDIVÍDUOS SELECIONADOS PARA O ESTUDO

Após a coleta das amostras, de acordo com os critérios de inclusão e exclusão determinados, foram selecionados 26 indivíduos, cujas amostras da microbiota bacteriana foram sequenciadas na plataforma *Ion Torrent™* (Life Technologies/Applied Biosystems). Os grupos foram divididos como:

- **Grupo C**, composto por 10 indivíduos-controle saudáveis que não possuíam nenhuma lesão maligna visível em sua mucosa e não faziam uso de tabaco e álcool. Todas estas amostras foram coletadas no Hospital AC Camargo, durante as campanhas de prevenção ao câncer realizadas por este hospital;
- **Grupo CFB**, segundo grupo-controle, intermediário, composto por cinco amostras de indivíduos com mucosa oral saudável, sem lesões aparentes, mas que faziam uso diário de tabaco e álcool. Todas as amostras deste grupo foram obtidas graças à colaboração com o Prof. Dr. Arthur de Andrade Guerra, supervisor geral do Grupo Interdisciplinar de Estudos de Álcool e Drogas (GREDA), do Instituto de Psiquiatria, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.
- **Grupo P**, composto por um total de onze pacientes portadores de OSCC. A maioria dos pacientes (9/11 – 82%) vieram do Departamento de Cabeça e Pescoço do Hospital AC Camargo - colaboração com Dr. Luiz Paulo Kowalski. As outras duas amostras foram obtidas de pacientes do Hospital

Heliópolis, fruto de colaboração com o Dr. Marcos Brasilino de Carvalho (Serviço de Cirurgia de Cabeça e Pescoço). Todos faziam uso diário de tabaco e álcool.

As características gerais de todos os indivíduos podem ser vistas na **Tabela 1**.

Tabela 1 - Características gerais de indivíduos dos grupos-controle (C e CFB) e pacientes (P) avaliados neste estudo.

	C	CFB	P
Masculino	8 (80%)	4 (80%)	10 (91%)
Feminino	2 (20%)	1 (20%)	1 (9%)
Idade Média	57,6 anos $\pm$ 7,9	56,4 anos $\pm$ 7,3	56,5 anos $\pm$ 9,6
Altura Média	1,69m $\pm$ 0,1m	1,72m $\pm$ 0,08m	1,70m $\pm$ 0,15m
Peso Médio	78,1 kg $\pm$ 27,7kg	77 kg $\pm$ 12,8kg	68,4 kg $\pm$ 24,6kg

4.1.1 Análise da Presença de HPV nas Amostras de Pacientes com OSCC

De acordo com os objetivos do estudo, um dos interesses era investigar se bactérias poderiam estar relacionadas com o desenvolvimento do câncer oral. Deste modo, diante da importância do HPV como agente causador de uma parcela destes tumores, era importante determinar o *status* de positividade dos pacientes selecionados para o estudo. A presença de HPV foi determinada em todas as amostras coletadas de pacientes com OSCC, incluindo diversos casos que não foram selecionados para as análises da sua microbiota. O diagnóstico da presença de HPV foi feito de acordo com o protocolo apresentado na seção de Materiais e Métodos e um resultado representativo é mostrado na **Figura 7**. Dentre os pacientes que tiveram sua microbiota sequenciada, apenas o paciente P04 foi positivo para HPV, com resultados demonstrando a presença do HPV-16.

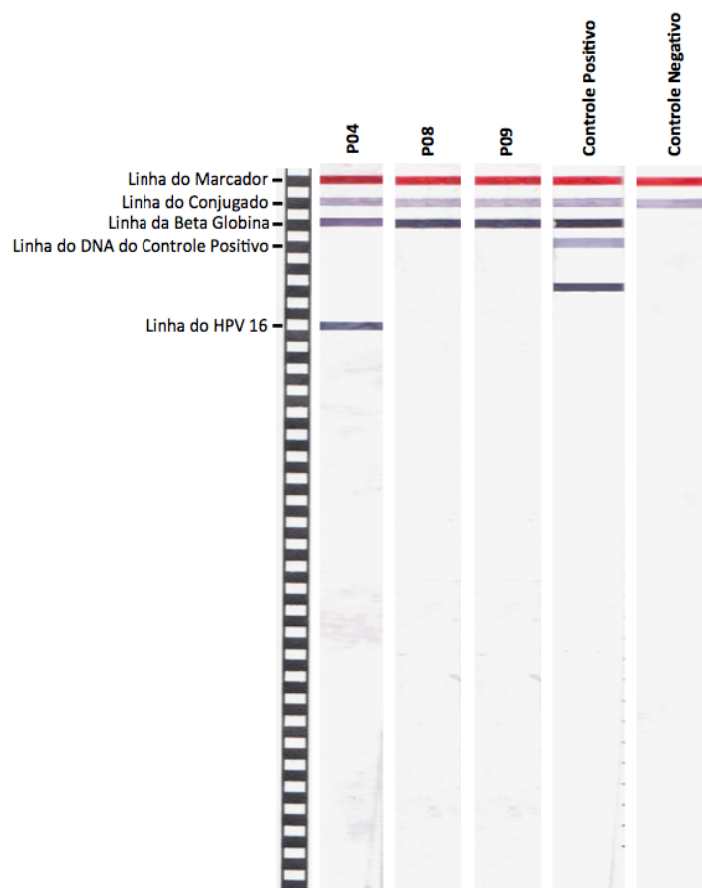


Figura 7 – Resultado da busca de HPV em algumas amostras deste estudo.

Em análise recentemente publicada sobre a presença de HPV, feita em 114 pacientes do Hospital AC Camargo portadores de OSCC (KAMINAGAKURA et al. 2012), foi observada uma prevalência global do HPV16 em 19.2% dos casos avaliados. Entre os pacientes avaliados por estes autores, pertencentes a faixa etária compatível com nosso estudo, a presença do vírus pôde ser demonstrada em 7/67 (10,4%). Tal prevalência é similar à encontrada em nosso trabalho (9,1%).

É sabido que as proteínas virais carcinogênicas encontradas em HPV são capazes de induzir a produção e liberação de diversas citocinas pró-inflamatórias (revisto em AGGARWAL et al. 2006), o que talvez possa ter algum papel na modulação da microbiota. Neste sentido, TEZAL et al. (2012), encontraram uma

relação positiva entre a presença de HPV e periodontite em pacientes portadores de *HNSCC*. No entanto, esta associação foi maior para câncer de orofaringe do que câncer da laringe ou da cavidade oral. Este fato, aliado ao encontro de apenas um paciente HPV+ no estudo, sugere que a microbiota aqui observada não deve ter sofrido significativo impacto da presença de citocinas inflamatórias induzidas pela presença de HPV na cavidade oral.

Por outro lado, seria interessante termos avaliado o status de positividade de HPV também para os indivíduos dos dois grupos-controle, pois resta a possibilidade de que a microbiota dos controles possa ter sido influenciada pela presença de HPV em alguns dos indivíduos estudados. No entanto, temos que observar dois fatores relevantes: 1) a prevalência de HPV na cavidade oral de pacientes sem câncer parece ser baixa. HERRERO et al. (2003), reportaram o encontro de DNA de HPV em células esfoliadas da cavidade oral em 6,9% de 1732 indivíduos-controle avaliados. 2) Nosso projeto original não previa a avaliação do status de HPV em nossos grupos-controle e a obtenção desta informação poderia levantar questionamentos éticos relevantes. O que fazer com um dado de positividade para HPV, fora do contexto de doença, agravado pelo fato de que não foi solicitado consentimento para esta avaliação? De qualquer modo, diante da prevalência relativamente baixa de HPV observada aqui, não acreditamos que diferenças entre os grupos possam ter sido devidas à maior presença de HPV em algum dos grupos estudados.

4.2 SEQUENCIAMENTO

No total foram feitas 6 corridas de sequenciamento: 5 corridas em chips tipo 314 e uma corrida em um chip 316. As corridas foram distribuídas da seguinte maneira: Corrida 1 - 2 pacientes e 2 controles negativos; Corrida 2 - 5 pacientes; Corrida 3 - 2 controles negativos e 2 pacientes; Corrida 4 - 4 controles negativos; Corrida 5 - 2 controles negativos e 2 pacientes; Corrida 6 (chip 316) - 5 controles (CFB).

No total, geramos 2.332.475 sequências, com um tamanho médio de 113nt. A mediana de sequências por amostra foi de 31.535 e a média foi de 92.735.

Os filtros aplicados, assim como a quantidade de sequências obtidas e remanescentes e seu tamanho final são apresentados nas **Tabelas 2 e 3**, a seguir.

Tabela 2 - Filtragem inicial das sequências geradas.

Filtros	Atributo	Sequências Excluídas (%)
Bases Ambíguas (N)	>2	0 (0)
Tamanho Mínimo	40	4.187 (0,17)
Número Máximo de Homopolímeros	6	7.133 (0,3)
Total		11.320 (0,48)

Ao final da etapa de sequenciamento, obtivemos um total de 2,3 milhões de sequências úteis. Em média, cada amostra teve uma cobertura de 89 mil sequências (variando de 9.190 a 385.593).

Tabela 3 - Número de sequências obtidas para as amostras do estudo (antes e após a aplicação dos filtros), e mediana do tamanho.

Amostra	Número de sequências antes dos filtros (dados brutos)	Sequências remanescentes após os filtros		Tamanho médio (nt)
		(N)	(%)	
Corrida 1 - Chip 314				
P33	20.058	19.885	99,1	104
P27	10.350	10.281	99,3	107
C15	12.455	12.368	99,3	104
C20	9.324	9.190	98,5	109
Total	52.187	51.724	99,1	106
Corrida 2 - Chip 314				
P01	73.204	72.798	99,4	114
P08	74.453	74.249	99,7	116
P09	122.432	122.118	99,7	114
P23	43.986	43.960	99,9	113
P46	54.146	54.100	99,9	113
Total	368.221	367.225	99,7	114
Corrida 3 - Chip 314				
C01	25.253	24.161	95,6	118
C04	18.811	18.006	95,7	118
P11	19.413	19.347	99,6	116
P36	15.798	15.732	99,5	118
Total	79.125	77.246	97,6	118
Corrida 4 - Chip 314				
C06	21.234	20.996	98,8	121
C07	32.566	32.326	99,2	120
C09	21.568	20.132	93,3	125
C16	27.670	27.522	99,4	123
Total	103.038	100.976	97,9	122
Corrida 5 - Chip 314				
C19	12.744	12.659	99,3	125
C21	20.201	20.066	99,3	125
P03	16.159	16.053	99,3	126
P04	18.403	18.302	99,4	124
Total	67.507	67.080	99,3	125
Corrida 6 - Chip 316				
CFB10	386.556	385.593	99,8	108
CFB18	309.436	307.783	99,5	113
CFB29	361.268	360.272	99,7	111
CFB30	334.125	333.205	99,7	112
CFB31	271.012	270.051	99,6	112
Total	1.662.397	1.656.904	99,7	111
Total Geral	2.332.475	2.321.155	99,5	116

4.2.1 Aspectos do Sequenciamento

Quando contrastamos os números obtidos no sequenciamento das amostras deste estudo, com os dados já gerados por outros grupos na avaliação da microbiota de indivíduos portadores de OSCC ou mesmo outros estudos da microbiota oral, podemos mensurar mais adequadamente a contribuição relativa do nosso estudo. Um aspecto negativo do nosso estudo é o tamanho relativamente pequeno dos fragmentos que sequenciamos, apenas 113 nt, contra *amplicons* maiores dos demais estudos, sequenciados na plataforma 454. Outro aspecto desfavorável é o uso da plataforma *Ion Torrent*TM que, a princípio, possui uma taxa de erro superior a aquela observada em outras plataformas. Em comparação recentemente publicada, a taxa de erro da plataforma *Ion Torrent*TM foi de 1,5%/100-bases, enquanto as taxas de erro da plataforma *Illumina*TM foi de 0,1%/100-bases (LOMAN et al. 2012) e a taxa de erro da plataforma 454TM (Roche) foi de 1,07%/101-bases (GILLES et al. 2011). Tanto a plataforma *Ion Torrent*TM quanto a plataforma 454-Roche tem uma taxa de erros significativamente alta (0,4 a 1,5%) para regiões homopoliméricas. No entanto, como pôde ser visto anteriormente, a estratégia adotada de definição de OTUs como grupos de sequências com 97% de identidade, contribui para reduzir o impacto desta deficiência. Em relação aos aspectos favoráveis do nosso estudo, o primeiro que podemos destacar é o tamanho da casuística, já que avaliamos mais indivíduos portadores de OSCC do que os demais estudos já publicados. Como exemplo, no estudo de PUSHALKAR et al. (2011) somente 3 pacientes portadores de OSCC e 2 controles sem neoplasia foram avaliados. De modo geral, o universo de indivíduos avaliados e a média de sequências obtidas por amostra foram significativamente superiores a aquelas dos demais estudos. Outro aspecto que devemos ressaltar é o

fato de termos nos preocupado em avaliar a microbiota da mucosa oral, enquanto que os demais estudos em *OSCC* avaliaram microbiota salivar. Diante das diferenças encontradas entre as populações de bactérias de diferentes partes do corpo, incluindo diferenças entre a composição da microbiota da saliva e da mucosa (HUSE et al. 2012), se buscamos bactérias associadas ao desenvolvimento do câncer oral, acreditamos ser mais provável que estas se encontrem fisicamente associadas ao tecido que sofre o processo de malignização – o biofilme da mucosa oral. Tais comparações podem ser melhor avaliadas na **Tabela 4**.

Tabela 4 – Análise qualitativa e quantitativa dos principais filos bacterianos encontrados na cavidade oral, de acordo com a região amostrada do gene 16S-rDNA e a plataforma de sequenciamento utilizada.

Região avaliada no gene 16S-rDNA	V1	V1	V1	V5-V8	V2	V6	V6	V4-V5	V4-V5	V1-V3	V5
Plataforma de sequenciamento usada	Ion Torrent	Ion Torrent	Ion Torrent	454-Roche	454-Roche	454-Roche	454-Roche	454-Roche	454-Roche	454-Roche	Illumina
Tamanho médio do amplicon (nt)	113	113	113	700	311	~150	~150	392	300	486	96
Amostra	Biofilme oral	Biofilme oral	Biofilme oral	Biofilme oral	Saliva	Saliva	Placa dentária sub-gengival	Saliva	Saliva	Saliva	Saliva
Indivíduos avaliados (N)	Saudáveis (10) Grupo C	Fumantes e etilistas (5) Grupo CFB	Câncer oral (11) Grupo P	Indivíduos saudáveis (10) Grupo C	Indivíduos saudáveis (13)	Indivíduos saudáveis (71)	Indivíduos saudáveis (98)	Saudáveis (26) ou com cáries (19)	Câncer oral (3) e controles (2)	Indivíduos saudáveis (5)	Indivíduos saudáveis (3)
Número de sequências geradas	197.426	1.662.397	466.825	107.307	1.070.000	197.600		425.623	58.000	50.708	1.237.319
Média de sequências por amostra	19.743	332.479	42.438	10.730	1.315	1.169		9.458	11.600	10.141	412.439
<i>Firmicutes</i>	50,1%	46,9%	40,5%	48,6%	35%	41%	32%	32,7%	45%	52%	39,3%
<i>Actinobacteria</i>	26,1%	30,1%	34%	22,2%	10%	6%	25%	<4%	14%	9%	2,8%
<i>Proteobacteria</i>	19,1%	19,5%	18,2%	15,4%	30%	22%	16%	17%	10%	27%	27,8%
<i>Bacteroidetes</i>	1,7%	1,5%	2,7%	9,3%	15%	27%	15%	31,3%	25%	8%	0,02%
<i>Fusobacteria</i>	0,6%	0,2%	0,1%	2,4%	5%	6%	9%	12,9%	5%	2%	1,6%
Referência	A	A	A	A*	B	C	C	D	E	F	G

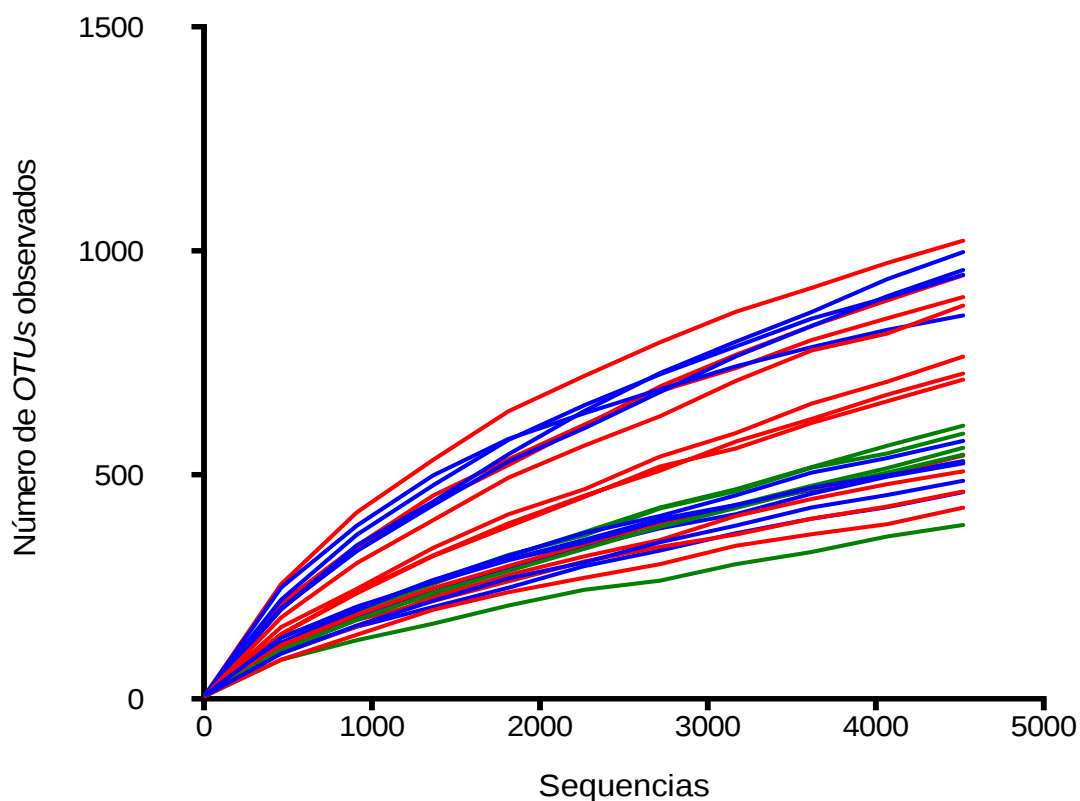
Referências: **A** – THOMAS et al. este estudo, artigo em preparação. **B** – COSTELLO et al. (2009); **C** – KEIJSER et al. (2008); **D** - YANG et al. (2012); **E** – PUSHALKAR et al. (2011); **F** – LAZAREVIC et al. (2010); **G** – LAZAREVIC et al. (2009). * - Dados apresentados no Anexo 2 deste trabalho.

4.3 ANÁLISES DAS SEQUÊNCIAS OBTIDAS

4.3.1 Avaliação da cobertura obtida e inferências sobre a diversidade bacteriana nos grupos estudados

Buscando avaliar a diversidade da microbiota oral, as sequências obtidas foram *clusterizadas* utilizando o programa *uclust* com uma porcentagem de identidade de 97% entre sequências do mesmo *cluster*. A seguir, conforme descrito na seção de materiais e métodos, sequências representativas de cada um dos *clusters* criados foram usadas em um BLAST contra o banco de dados 16S-rDNA *Ribosomal Database Project* (que conta com 1.379.424 de sequências deste gene oriundas de bactérias e arqueobactérias das mais diversas fontes - COLE et al. 2009). Um *hit* foi considerado apenas quando uma identidade de pelo menos 90% era obtido entre as sequências do banco e as sequências representativas do *cluster*. Cerca de 33% das sequências obtidas não puderam ser atribuídas a nenhum táxon por análises de BLAST. Tais sequências certamente contém artefatos e erros de sequenciamento, mas provavelmente também contêm espécies novas, as quais não foram avaliadas neste trabalho. Se um dado *cluster* possuía um mínimo de 3 sequências ele era mantido, se não, era descartado. Cabe notar no entanto que, se por um lado esta estratégia evita a criação de possíveis artefatos devido a erros de sequenciamento e/ou baixa qualidade das bases das sequências, este tipo de restrição por frequência, certamente sub-estima a diversidade da microbiota, ao excluir as espécies pouco frequentes. Posteriormente, fizemos uma análise de saturação/rarefação afim de determinar qual a cobertura aproximada foi obtida através dos *OTUs* para a microbiota das amostras individuais de cada grupo (**Figura 8**).

Curva de rarefação por grupos de amostras



Legenda – A figura mostra o número de *OTUs* observadas para cada amostra individual dos três grupos; pacientes (Grupo P - linhas em vermelho), controles que fumam e bebem (Grupo CFB - linhas em verde) e controles que não fumam e não bebem (Grupo C - linhas em azul).

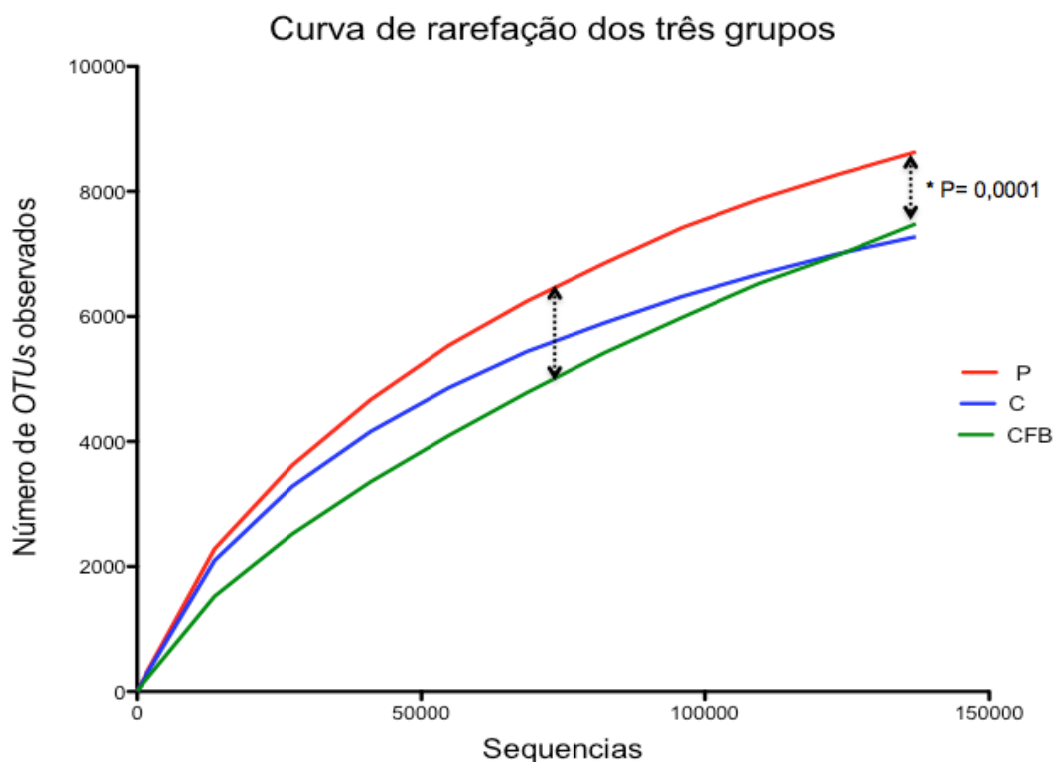
Figura 8 – Curvas de saturação/rarefação obtidas a partir da análise individual para os três grupos de estudo

Estes dados permitem estimar que, de acordo com os níveis de cobertura obtidos e utilizando um mesmo número de *reads* para cada amostra, os indivíduos deste estudo albergam em sua mucosa oral, de 300 a 1000 *OTUs* diferentes. Este número é um pouco mais elevado do que outros apresentados na literatura. Talvez estes dados sejam fruto de um artefato causado pelo uso de fragmentos menores, pois no estudo de ENGELBREKTSON et al. (2010), foi sugerido que fragmentos <400pb erroneamente indicavam uma maior diversidade de espécies bacterianas do que fragmentos maiores que 400pb. De modo similar, no estudo de LAZAREVIC et al.

(2009), que utilizaram a plataforma *Illumina™* para sequenciar *amplicons* que cobriam a região variável 5 do gene 16S-rDNA, com um tamanho médio de 96 nt, foram encontrados cerca de 6000 *OTUs* nas amostras de saliva de apenas 3 indivíduos. Ambos estudos (este trabalho e LAZAREVIC et al. 2009) se basearam na análise de fragmentos relativamente curtos (113nt), e ambos encontraram uma diversidade aparentemente maior que outros estudos, como exemplo, em outro estudo de LAZAREVIC et al. (2010) foram encontradas cerca de 50-200 *OTUs* por indivíduo, em 5 amostras salivares sequenciadas na plataforma 454 com fragmentos maiores. Não podemos deixar de observar que este maior número de *OTUs* aqui observado também pode ser fruto de uma maior taxa de erros gerada pelo sequenciamento no *Ion Torrent™*, como comentado anteriormente (LOMAN et al. 2012). No entanto, empregamos estratégias para minimizar este possível aumento artefactual do número de *OTUs* devido a erros de sequenciamento gerados pelo *Ion Torrent™*, tais como a *clusterização* das sequências com 97% de identidade entre si ou a consideração de *OTUs* que possuem pelo menos 3 sequências e que foram atribuídas a táxons bacterianos utilizando BLAST. Como todas as amostras neste estudo foram amplificadas com o mesmo par de iniciadores e o mesmo número de ciclos de PCR, portanto, artefatos devidos a estes aspectos não influenciaram nos resultados finais.

Tendo feito estas observações, cabe notar que todos os indivíduos do grupo CFB apresentaram menos de 500 *OTUs*/indivíduo neste nível de cobertura, enquanto que a maioria dos pacientes (linhas em vermelho) apresentaram mais de 500 *OTUs*/indivíduo. Buscando avaliar com maior cuidado as diferenças inter-grupos,

analisamos as curvas de saturação obtidas quando os indivíduos de cada grupo foram avaliados em conjunto (**Figura 9**), visando reduzir o efeito de variações individuais.



Legenda – A figura mostra o número de OTUs observadas nos três grupos, após agruparmos os indivíduos de cada grupo.

Figura 9 – Curvas de saturação/rarefação obtidas a partir da análise agrupada dos indivíduos dos três grupos de estudo

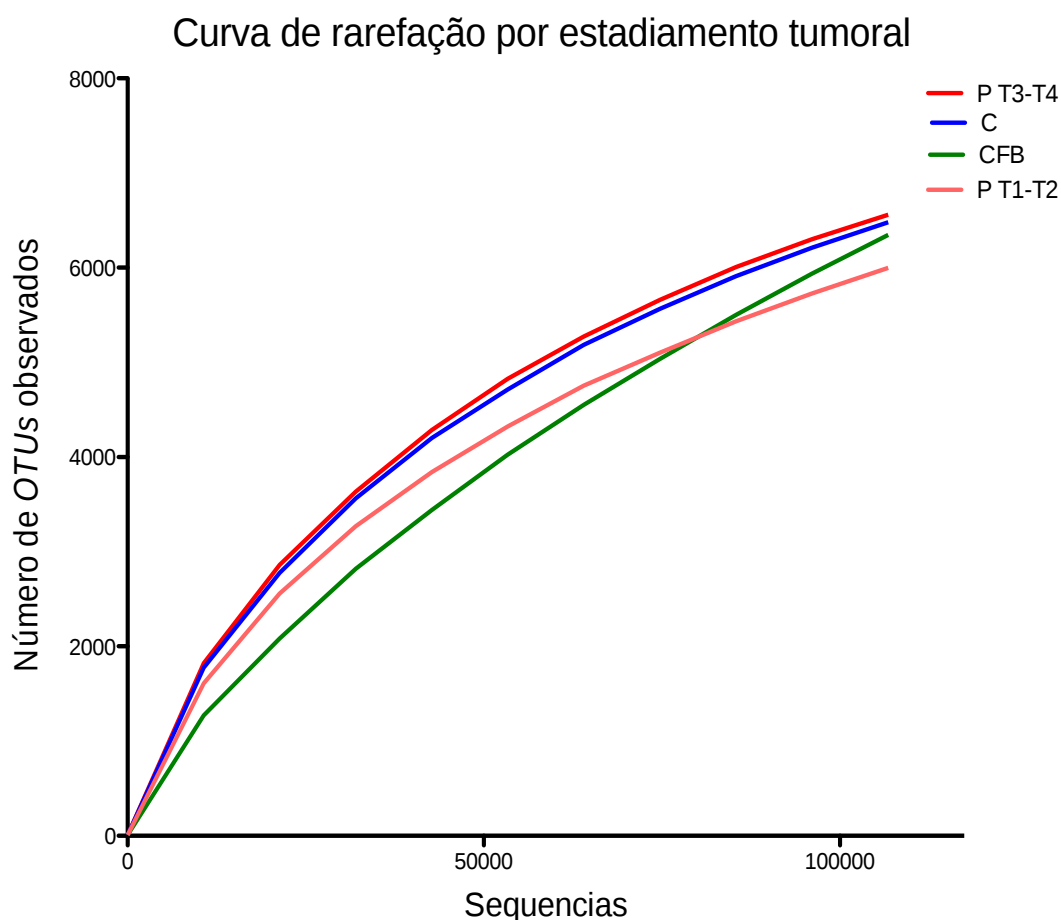
Após aumentarmos o número de sequências usadas para estimar a diversidade, o que foi possível graças ao agrupamento das sequências de cada indivíduo dentro do mesmo núcleo (Grupos C, CFB e P), as diferenças entre os grupos tornaram-se mais claras e consistentes. Observamos uma clara diferença em termos de diversidade para o grupo P, enquanto que os grupos C e CFB aparentam uma diversidade similar e mais reduzida (dentro do N de sequências utilizadas, ao redor de 140.000/grupo). Quando comparamos dois pontos da curva (indicado com as setas) entre o grupo dos

pacientes (P) e grupo dos controles que fumam e bebem (CFB), vemos uma maior diversidade entre os pacientes ($P=0,0001$) usando o teste estatístico T student.

No entanto, não podemos deixar de comentar a respeito do perfil das curvas dos grupos C e P, que tendem a atingir um *plateau*, enquanto que a curva do grupo CFB aparentemente tende a continuar subindo. Tal perfil é sugestivo da presença de uma diversidade significativa no grupo CFB, mas que talvez seja permeada por algumas *OTUs* de alta frequência, que torne necessária a geração de uma maior quantidade de sequências para que a diversidade global possa ser plenamente contemplada.

Com este nível de cobertura temos indícios de uma maior diversidade bacteriana no grupo dos pacientes portadores de OSCC quando comparado aos demais grupos, indicativo de um desequilíbrio da microbiota, que leva à presença (ainda que passageira) de um maior número de espécies nos indivíduos com câncer. Este achado pode ser um indício da presença de bactérias oportunistas na cavidade oral, resultado do efeito combinado da reduzida imunidade dos pacientes e de uma maior dificuldade de higienização devido à presença da lesão. Se esta segunda situação for verdadeira, deveríamos observar uma diversidade bacteriana ainda maior nos pacientes com tumores mais avançados, quando comparados a pacientes com tumores menores.

Buscando investigar esta possibilidade, dividimos o grupo de pacientes (P) em dois sub-grupos, conforme o tamanho do tumor. O grupo P T1-T2 contendo pacientes com tumores menores ($N=6$) ou o grupo P T3-T4, composto por pacientes com tumores mais avançados ($N=5$). Os dados são apresentados a seguir, na **Figura 10**.



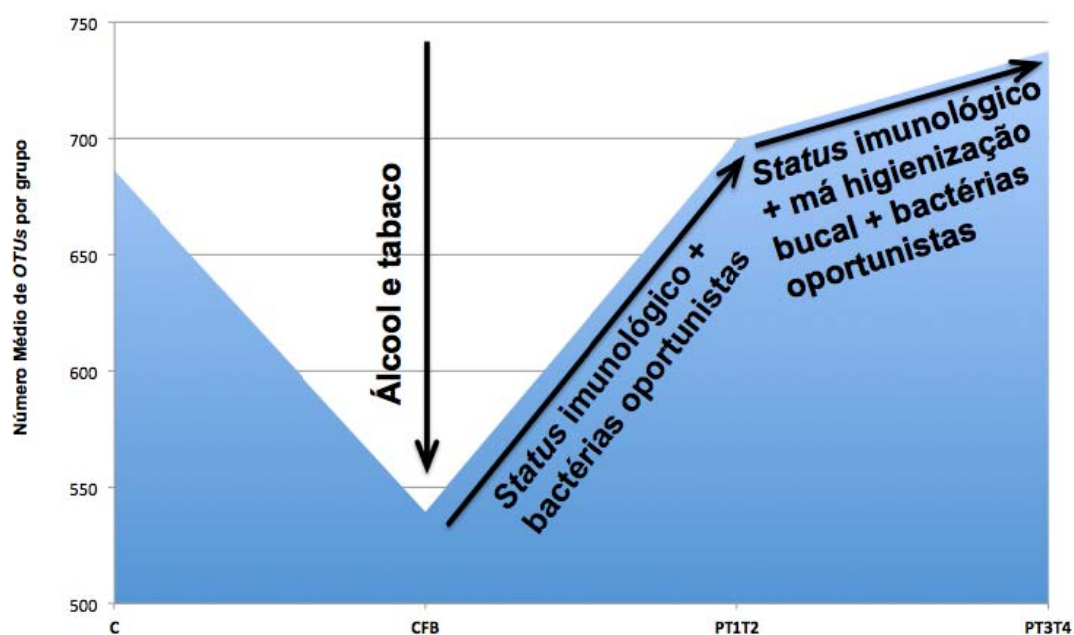
Legenda – A figura mostra o número de OTUs observados nos 4 grupos; **P T3-T4**: Pacientes com estadiamento tumoral T3-T4 (composto por 5 amostras: P01, P03, P08, P09 e P33), **C**: Controles que não fumam e nem bebem (composto por 10 amostras), **CFB**: Controles que fumam e bebem (composto por 5 amostras) e **P T1-T2**: Pacientes com estadiamento tumoral T1-T2 (composto por 6 amostras: P04, P11, P23, P27, P36 e P46).

Figura 10 – Curvas de saturação/rarefação obtidas a partir da análise de sub-grupos de pacientes agrupados conforme tamanho da lesão tumoral.

Interessantemente, observamos que os pacientes portadores de OSCC com estadiamentos tumorais mais avançados (P T3-T4), possuem um maior número de OTUs do que pacientes portadores de OSCC com estadiamentos tumorais mais baixos (P T1-T2), e ambos se encontram em pontos opostos da curva de diversidade microbiana aqui observada, mesmo apresentando uma menor quantidade de amostras no seu grupo (5 indivíduos em PT3-T4 versus 6 indivíduos em PTI-T2). Tal achado

aponta para uma maior diversidade bacteriana em pacientes com tumores avançados, compatível com déficits imunológicos na mucosa oral e/ou maior dificuldade de higienização.

Neste ponto, seria tentador supor um modelo no qual a diversidade bacteriana normal, é afetada negativamente pelo uso contínuo de álcool e tabaco, e este quadro se agravaria ainda mais nos pacientes já com tumores iniciais, sendo posteriormente exacerbado por uma colonização por bactérias oportunistas nos pacientes com tumores avançados. Este modelo (ilustrado na **Figura 11**) poderia ser proposto se tomamos apenas os pontos finais da curva mostrada na **Figura 10**, no entanto, temos que observar que a tendência da curva do grupo CFB é totalmente diferente das demais. Dois aspectos devem ser considerados: i) o grupo CFB foi sequenciado mais extensivamente que os demais (cerca de 10X mais), o que certamente induz a descoberta de *OTUs* mais raras e ii) esta diversidade maior que pode ser inferida pela tendência da curva de CFB aparentemente se dá pela colonização com táxons diversos, de frequência muito baixa, mas tem o efeito de sua abundância reduzido pela presença de algumas presentes em grande quantidade, o que se refletiria na menor diversidade do grupo como um todo. Tal efeito foi recentemente demonstrado por SEGATA et al. (2012), onde a abundância de *Streptococcus* causava um efeito reduzido da diversidade microbiana no que este grupo chamou de grupo G1 (mucosa bucal, gengiva queratinizada e palato duro) e a abundância de *Bacteroidetes* reduziu a diversidade da microbiota intestinal (SEGATA et al. 2012).



Legenda – A figura mostra o número médio de *OTUs* observados para cada amostra dos três grupos, usando o mesmo número de sequências para cada amostra de cada grupo.

Figura 11 – Modelo hipotético do efeito do álcool e do tabaco na diversidade da microbiota oral e seu aumento durante o aparecimento de tumores menos e mais avançados.

A seguir as **Tabelas 5, 6 e 7** mostram as *OTUs* mais abundantes encontradas nos 3 grupos, C, CFB e P.

Tabela 5 - OTUs mais abundantes no grupo dos controles

#OTU ID	Denominação consenso da OTU	Média	C01	C04	C06	C07	C09	C15	C16	C19	C20	C21	CFB ¹	P ²
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	5047	8133	4315	7840	913	11494	1536	3386	5308	348	7191	6310	3937
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	4978	940	900	8411	1104	11976	720	8219	6572	87	10849	8799	3293
184308	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	2495	50	178	962	18954	1008	137	1181	869	305	1306	3074	2257
310184	"Proteobacteria"; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria	2109	956	3204	700	136	1063	218	9607	2243	588	2372	459	1805
71322	"Proteobacteria"; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria	1990	145	733	429	87	661	81	14094	1311	98	2258	454	1611
412169	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	1981	522	339	1681	1228	750	1504	1966	3144	272	8407	4435	1300
88717	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	1844	377	622	624	13095	725	32	981	1122	229	638	2039	1483
338554	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	1823	844	1883	1372	74	3169	1205	2529	3420	544	3189	2424	589
120655	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	1566	2235	2982	1548	408	2454	429	1857	2180	11	1555	1926	1493
249042	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	1437	406	966	1086	77	2513	809	3303	2733	359	2113	2566	761
412166	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	1211	418	428	676	260	1147	3735	2489	1098	65	1794	3902	1876
160926	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	1205	1784	1016	1019	690	442	534	1054	1777	109	3628	2620	1004
52726	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	1055	145	156	1586	377	1604	849	2456	1398	11	1964	2512	1161
381866	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	1016	29	50	152	7152	472	40	1166	719	163	219	506	287
198535	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	915	1573	305	2748	677	30	170	105	3136	0	404	1610	3155
177278	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	884	203	1860	3439	668	924	8	33	316	65	1326	1339	93
314198	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	830	17	205	4082	563	1108	24	7	348	54	1889	1782	55
200268	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	821	335	17	2124	647	864	332	280	3065	0	548	2690	2801
209284	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	810	459	766	486	186	894	1949	1312	1074	33	942	2710	1420
223532	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	741	83	122	119	5782	283	0	607	269	0	150	286	162

Todos os números apresentados se referem ao número de *reads* normalizado por 100.000 sequências.

¹: Média de *reads* normalizados por 100.000 encontrados no grupo CFB para esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo P para esta OTU

Tabela 6 - OTUs mais abundantes no grupo dos controles que fumam e bebem

#OTU ID	Denominação consenso da OTU	Média	CFB10	CFB18	CFB29	CFB30	CFB31	C ¹	P ²
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	8799	1281	11823	7305	15199	8387	4978	3293
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	6310	1085	9112	6203	8448	6700	5047	3937
338555	"Proteobacteria"; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	4796	23962	1	11	0	7	2	0
184308	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	3074	7474	2121	2174	1803	1800	2495	2257
209284	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	2710	1899	1705	2933	1619	5395	810	1420
200268	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	2690	25	576	5437	6326	1085	821	2801
80819	"Proteobacteria"; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	2644	13209	0	6	0	4	2	0
160926	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	2620	1183	3454	4859	1846	1757	1205	1004
249042	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	2566	2844	3159	3307	85	3434	1437	761
52726	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	2512	340	3512	1207	2444	5059	1055	1161
338554	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	2424	2685	2713	3450	74	3199	1823	589
88717	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	2039	4650	1406	1910	1128	1103	1844	1483
120655	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	1926	215	2610	1265	1408	4135	1566	1493
314198	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	1782	82	1384	40	6334	1068	830	55
198535	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	1610	15	89	3821	3876	248	915	3155
177278	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	1339	89	1348	67	4200	991	884	93
298928	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	755	102	289	1164	1577	642	105	739
248938	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	673	84	276	1087	1342	574	158	692
250405	"Firmicutes"; "Bacilli"; Bacillales; "Staphylococcaceae"; Gemella	506	840	347	434	622	288	136	80
381866	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	506	250	659	799	508	313	1016	287

Todos os números apresentados se referem ao número de *reads* normalizado por 100.000 sequências.

¹: Média de *reads* normalizados por 100.000 encontrados no grupo C para esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo P para esta OTU

Tabela 7 - OTUs mais abundantes no grupo dos pacientes

#OTU ID	Denominação consenso da OTU	Média	P01	P03	P04	P08	P09	P11	P23	P27	P33	P36	P46	C ¹	CFB ²
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	3937	2692	354 5	5054	250 0	456 0	838 9	134	159 5	342	654 1	7952	504 7	6310
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	3293	4131	297 8	5803	292 1	464 2	868 2	298	642	106	172 3	1210 9	497 8	8799
198535	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	3155	2467	684 6	1121 7	183	664 4	472 9	177	224	50	215 5	9	915	1610
200268	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	2801	4073	446 0	1177 5	387	832 7	708	255	136	45	165	479	821	2690
184308	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	2257	1720	173 8	4775	161 1	126 0	78	1039 8	185	900	95	2067	249 5	3074
412166	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	1876	2179	122 7	1098	585 279	124 1	2186	507 7	654	197	5919	121 1	3902	
310184	"Proteobacteria"; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria	1805	2146	718 9	612	712 9	155	26	66	29	196	0	129	210 9	459
71322	"Proteobacteria"; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria	1611	3289	325 8	949	0	180	0	118	10	146	585	235	199 0	454
120655	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	1493	2740	124 6	1798	120 0	131 3	295 1	91	111 9	228	2	1665	156 6	1926
88717	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	1483	962	168 8	3273	118 1	952	346	5808	185	236	458	1220	184 4	2039
209284	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	1420	1312	110 3	945	102 170	342	0	1381	321 0	427	178	4351	810	2710
412169	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	1300	1897	105 132	262	5 120	9	367	2675	350	1	305	4116	198 1	4435
52726	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	1161	3089	132 1	2005	111 3	274	134	170	2	65	305	1564	105 5	2512
422556	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified_Pasteurellaceae	1057	163	1074 2	118	0	280	9	0	0	292	19	0	238	170
160926	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	1004	1040	129 4	237	339	706	6	1497	136	297	998	2475	120 5	2620
249042	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	761	556	202 6	903	49	387	9	222	5462	136	5	470	143 7	2566
298928	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	739	2507	106 5	19	546	191	720	398	34	982	15	540	105	755
248938	"Actinobacteria"; Actinobacteria; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomyces	692	1378	102 1	37	481	155	743	5	30	1	5	985	158	673
85018	"Bacteroidetes"; Flavobacteria; "Flavobacteriales"; Flavobacteriaceae; unclassified_Flavobacteriaceae	608	25	985 1715	6	5	322	12	0	5505	0	810	0	261	7
232192	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified_Pasteurellaceae	546	59	0	143	0	203	13	5	5089	0	473	19	465	166

Todos os números apresentados se referem ao número de *reads* normalizado por 100.000 seqüências.

¹: Média de *reads* normalizados por 100.000 encontrados no grupo C para esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo CFB para esta OTU

Pela **Tabela 5** vemos que há duas *OTUs* atribuídas ao gênero *Haemophilus*, que chamam atenção por sua presença em maior abundância nos grupos sem câncer (C e CFB) em relação ao grupo dos pacientes (P). Usando o teste estatístico de Mann-Whitney, estas duas *OTUs* se mostraram significativamente mais elevadas em ambos os grupos controle (C e CFB) e diminuídas no grupo dos pacientes (*Haemophilus* 338554 com $P=0,019$ e *Haemophilus* 249042 com $P=0,036$).

Já na **Tabela 6** observamos 4 *OTUs* no grupo CFB que possuem frequência de interesse, correspondendo aos gêneros *Ganulicatella*, *Haemophilus*, *Saccharomonospora* e *Veillonella*. As mesmas possuem o dobro da quantidade de *reads* normalizados que os outros dois grupos. A *OTU* correspondente ao gênero *Veillonella*, que possui uma grande quantidade de *reads* normalizados em todos os grupos, porém no grupo CFB possui um quantidade duas vezes maior. Isto pode causar o efeito que vimos anteriormente de uma menor diversidade encontrada no grupo dos CFBs, causada pela presença de alguns táxons em grande quantidade, mascarando aqueles de menor abundância, gerando uma menor saturação, comparado aos outros dois grupos, e um número limitado de espécies únicas e diferentes.

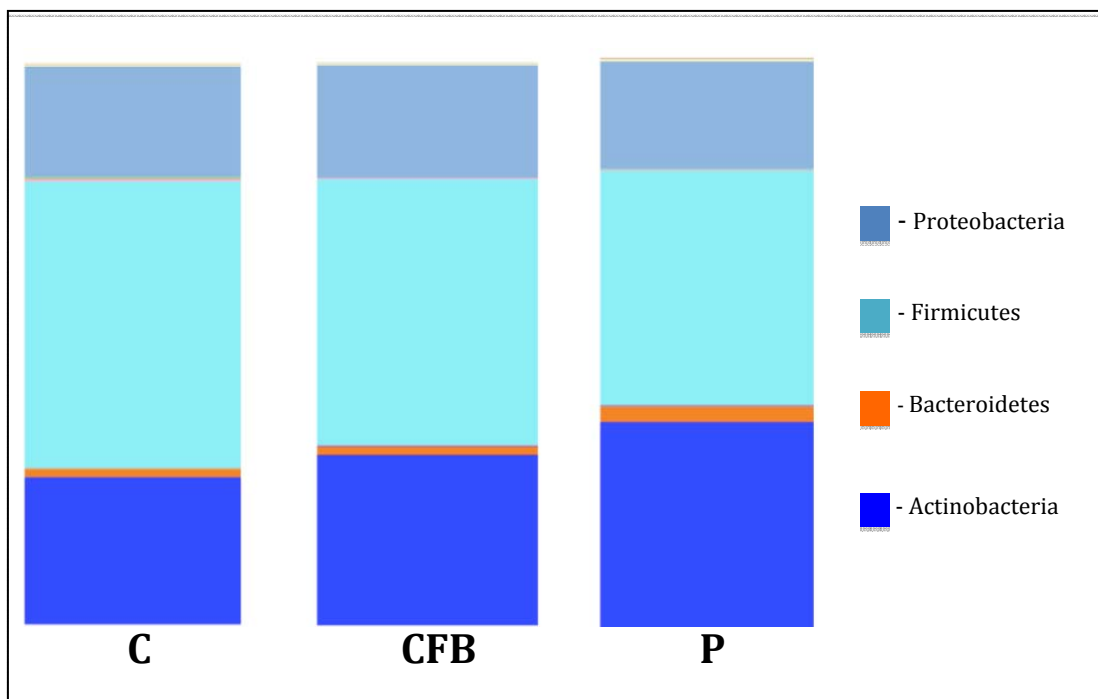
Na **Tabela 7** vemos que há uma *OTU* com comportamento interessante e que é atribuída ao gênero *Actinomyces*, ela é muito abundante nos grupos P e CFB e pouco abundantes no grupo dos controles.

As **Tabelas 5, 6 e 7**, que mostram as *OTUs* mais abundantes, normalizadas por 100.000 sequências, podemos observar que a *OTU* mais abundante encontrada nos três grupos é aquela que corresponde ao gênero *Veillonella*. No grupo dos pacientes, podemos observar que a segunda *OTU* mais abundante pertence ao gênero

Actinomyces, no grupo dos controles que fumam e bebem pertence a *Pseudomonas* e no grupo dos controles que não fumam e não bebem pertence a *Streptococcus*. Isso sugere uma variação quantitativa e qualitativa entre os três grupos mesmo no que se refere às *OTUs* mais abundantes, algo positivo quando vemos que a cobertura de sequenciamento foi heterogênea entre os 3 grupos. Deste modo, avaliamos os demais táxons observados, levando em consideração a sua frequência nas amostras de cada grupo, e também nos dois sub-grupos de pacientes (T1/T2 x T3/T4).

4.4 TÁXONS BACTERIANOS PRESENTES NOS GRUPOS AVALIADOS

A princípio, investigamos a frequência dos táxons bacterianos mais prevalentes nos diferentes grupos avaliados. Para tal usamos a atribuição taxonômica determinada pelo programa *Qiime* (CAPORASO et al. 2010b) e os dados são apresentados na **Figura 12** e na **Tabela 8**.



Nota – A Figura apresenta os 4 táxons bacterianos mais abundantes entre os controles (C), controles que fumam e bebem (CFB) e pacientes (P).

Figura 12 – Filos bacterianos mais abundantes nos três grupos de estudo

No que se refere à frequência dos táxons mais abundantes, nenhuma diferença significativa pôde ser observada entre os grupos. Conforme comentado anteriormente, cabe novamente ressaltar que, quando comparamos nossos dados com os demais dados da literatura, vemos algumas diferenças interessantes nestes grupos, que podem ser sugestivas da origem das amostras usada para o estudo: saliva versus biofilme. Para facilitar as comparações, a seguir apresentamos a **Tabela 8**, que consiste em uma modificação da **Tabela 4**, re-organizada para permitir contrastar os achados entre a microbiota da mucosa oral (biofilme) e da saliva, e a média de alguns estudos.

Tabela 8 - Prevalência dos filos bacterianos mais abundantes em diversos estudos voltados para a avaliação da microbiota oral.

Estudo	1	2	3	Média estudos biofilme (1 a 3)	4	Média estudos saliva (6 a 11)	5	6	7	8	9	10	11
	Ion Torrent	Ion Torrent	Ion Torrent		454-Roche		Placa dentária sub-gengival	Saliva	Saliva	Saliva	Saliva	Saliva	Saliva
Amostra	Biofilme oral (C)	Biofilme oral (CFB)	Biofilme oral (P)		Biofilme oral (C)								
<i>Firmicutes</i>	50,1%	46,9%	40,5%	45,8%	48,6%	40,9%	32%	35%	41%	32,7%	45%	52%	39,3%
<i>Actinobacteria</i>	26,1%	30,1%	34%	30,1%	22,2%	7,6%	25%	10%	6%	<4%	14%	9%	2,8%
<i>Proteobacteria</i>	19,1%	19,5%	18,2%	18,9%	15,4%	22,3%	16%	30%	22%	17%	10%	27%	27,8%
<i>Bacteroidetes</i>	1,7%	1,5%	2,7%	2,0%	9,3%	17,7%	15%	15%	27%	31,3%	25%	8%	0,02%
<i>Fusobacteria</i>	0,6%	0,2%	0,1%	0,3%	2,4%	5,4%	9%	5%	6%	12,9%	5%	2%	1,6%
Referência	1	1	1		1*		4	3	2	5	6	7	8

Referências: 1-THOMAS et al. em preparação; 2- KEIJSER et al. (2008); 3- COSTELLO et al. (2009); 4- KEIJSER et al. (2008); 5- YANG et al. (2012); 6- PUSHALKAR et al. (2011); 7- LAZAREVIC et al. (2010); 8- LAZAREVIC et al. (2009). * Dados referentes ao Anexo 2 deste trabalho.

Na **Tabela 8**, vemos que todos os nossos resultados referentes à microbiota oral são consistentes em atribuir *Actinobacteria* como o segundo filo mais abundante. Tal resultado foi corroborado inclusive pelo sequenciamento de uma segunda região do gene 16S-rDNA, realizado em outra plataforma (estudo 4 da **Tabela 5**). Nos demais estudos da microbiota oral o único que apresenta *Actinobacteria* como segundo filo mais abundante é o estudo de KEIJSER et al. 2008 (estudo 5 da **Tabela 5**), derivado do biofilme de uma placa dentária sub-gengival. Além destas diferenças, vemos reduções significativas na frequência de *Bacteroidetes* e *Fusobacteria* nos biofilmes da mucosa, de modo consistente entre os grupos (estudos 1 a 4). Apesar de não termos, em nosso estudo, investigado a microbiota da saliva e da mucosa de um mesmo indivíduo (pois este objetivo fugia dos objetivos propostos), nossos dados sugerem que a microbiota associada à mucosa oral, presente nos biofilmes seja distinta da microbiota salivar, em todos os grupos estudados.

Cabe notar no entanto que o filo *Actinobacteria*, que mostrou prevalência aumentada em nosso estudo, é composto predominantemente por bactérias Gram-positivas, que possuem uma parede celular menos resistente e portanto, mais facilmente rompida. O sistema de coleta e extração de DNA que usamos se baseia na lise celular ocorrendo no cartão *FTA*, seguido de sua liberação no meio aquoso após a fervura de fragmentos obtidos do cartão. Testes específicos não foram feitos para avaliar se a parede celular de bactérias Gram-negativas é efetivamente lisada neste sistema, e se isto pode ter levado a algum tipo de viés de representação dos filos bacterianos. Porém vemos que um outro estudo da microbiota de biofilme, derivado de placa dentária sub-gengival (KEIJSER et al. 2008) obteve resultados semelhantes aos nossos, com *Actinobacteria* estando em segundo lugar como filo bacteriano mais

abundante. Posteriormente pretendemos investigar melhor as diferenças entre biofilme da mucosa e saliva, mas a princípio nosso estudo ressalta que, diante da potencial diferença do microbioma entre estes ambientes, quando se investiga a microbiota relacionada ao câncer, deve ser dada uma ênfase preferencial à investigação das bactérias associadas à mucosa e não daquelas encontradas livres na saliva. Durante a redação final deste documento, SEGATA et al. (2012), publicaram um estudo comparativo da microbiota humana distribuída ao longo de sete regiões distintas do trato digestivo. Importantes diferenças entre a microbiota da mucosa oral e da saliva também são apontadas por estes autores.

4.5 ANÁLISES DE UM POSSÍVEL *CORE MICROBIOME*

No trabalho de HUSE et al. (2012) foram analisados quais *OTUs* estariam presentes em todas as amostras de um grande número de indivíduos saudáveis, e que deveriam compor um possível *core microbiome* saudável. Tal abordagem tem aplicações interessantes, pois um desbalanço deste *core* pode ser um biomarcador indireto de um distúrbio fisiológico relevante e, sua detecção poderia indicar que medidas devem ser tomadas para a recomposição do *core microbiome*. Buscando identificar este *core*, HUSE et al. (2012) sequenciaram em larga-escala o gene 16S-rDNA bacteriano (duas regiões diferentes V1-V3 e/ou V3-V5) de amostras de diversos indivíduos, e buscaram quais as *OTUs* estariam presentes em todos os indivíduos avaliados.

Em nosso estudo, buscamos avaliar a mesma questão, e investigamos se haveria um *core microbiome* representativo dos nossos grupos. Assim como no

estudo de HUSE et al. (2012) investigamos as *OTUs* que estavam presentes em 100% das amostras de cada grupo, ou em apenas ~90% das amostras (9 indivíduos para o grupo dos controles, 9 indivíduos para grupo dos pacientes e 4 indivíduos para o grupo dos controles que fumam e bebem) ou ~80% das amostras (8 indivíduos para o grupo dos controles, 8 indivíduos para grupo dos pacientes e 3 indivíduos para o grupo dos controles que fumam e bebem). Os resultados são mostrados na **Tabela 9**, a seguir. Para efeitos de comparação apresentamos também os dados obtidos por HUSE et al. (2012) apenas para a mucosa oral de indivíduos controle saudáveis. Dados sobre uso de álcool e tabaco referentes a estes indivíduos não foram disponibilizados pelos autores deste artigo.

Tabela 9 - Número de *core OTUs* encontrados em 100, 90 e 80% das amostras dos três grupos

Porcentagem das amostras contendo o <i>core microbiome</i>	Número de <i>core OTUs</i> por grupo			
	C (n= 10)	CFB (n=5)	P (n= 11)	C (n=114)*
100%	38	1.036	32	3
90%	82	2.656	115	15
80%[#]	148	4.684 [#]	229	25**

* - Dados de HUSE et al. (2012), baseados no sequenciamento da região V1-V3. ** - valor obtido para *OTUs* presentes em 75% das amostras. # - *OTUs* presentes em 60% do grupo (3/5 amostras).

A princípio é clara a discrepância entre os dados do nosso grupo CFB e os demais. Isto certamente se deve à cobertura mais elevada da microbiota do grupo CFB, devido a dois fatores: i) o menor N-amostral deste grupo (apenas 5 indivíduos) e ii) sobretudo devido ao sequenciamento destas amostras em um chip de maior capacidade, que gera cerca de 10X mais dados que os chips usados nas demais corridas. Em conjunto, estes fatores fizeram com que o grupo CFB tivesse uma

média de sequências/amostra equivalente a 332.479, o que foi mais de 10X mais que o obtido para os outros grupos do estudo (C= 19.426/amostra e P=42.438/amostra). Buscando avaliar o peso de cada um destes fatores, reavaliamos nossos dados, investigando, a princípio, a importância do tamanho amostral na composição do *microbiome core*. Investigamos qual o efeito da redução do N-amostral dos grupos C e P, avaliando apenas 5 indivíduos em cada grupo, número igual ao estudado no grupo CFB. Como pode ser visto na **Tabela 10**, a redução do N de cada grupo, artificialmente infla o *core microbiome* de modo significativo.

Tabela 10 – Número de *core OTUs* encontrados em 100, 90 e 80% das amostras dos três grupos com a metade do *n* amostral

Porcentagem das amostras	Número de <i>core OTUs</i> por grupo				
	C (n= 10)	C (n= 5)	CFB (n=5)	P (n= 11)	P (n=5)
100%	38	54	1.036	32	117
90%	82	178	2.656	115	390
80%[#]	148	429 [#]	4.684 [#]	229	452 [#]

[#] - *OTUs* presentes em 60% do grupo (3/5 amostras).

Vemos no trabalho de HUSE et al. (2012) um efeito exatamente oposto ao que observamos aqui. Por terem feito um estudo de um número muito maior de amostras (114 contra 10), que tiveram sua microbiota coberta por uma quantidade média de sequências muito menor (34.843 de nosso estudo, contra 6.168 – já excluindo CFB onde nossa cobertura foi mais de 10X maior), acreditamos que o *core microbiome* da mucosa oral sugerido por estes autores (apenas 3 com 100% das amostras) seja bastante sub-estimado.

Afim de usar de modo mais adequado os dados referentes ao grupo CFB e P, para fins da avaliação do *core microbiome*, fizemos os seguintes ajustes:

- 1) Correção para a maior quantidade de sequências – Como para o grupo menos sequenciado de nosso estudo (Grupo C), tivemos uma média de 19.426 sequências/indivíduo, avaliamos, para cada um dos 5 componentes do grupo CFB, quantos *reads* foram obtidos e usamos isto como um fator para considerarmos a presença de uma *OTU* no *core microbiome*. Sendo assim se um dado CFB foi sequenciado 13X mais que a média dos controles, apenas consideramos a presença de *OTUs* no *core* se estas tiverem pelo menos 13 *reads*. Deste modo, após implementarmos este ajuste, o *core microbiome* 100% deste grupo caiu de 1.036 para 107. Avaliamos também, para cada um dos componentes do grupo P, quantos *reads* foram obtidos e usamos isto como um fator para considerarmos a presença de uma *OTU* no *core microbiome*. Sendo assim se um dado P foi sequenciado 1X, 2X ou 3X mais que a média dos controles, apenas consideramos a presença de *OTUs* no *core* se estas tiverem pelo menos 1, 2 ou 3 *reads*. Deste modo, após implementarmos este ajuste, o *core microbiome* 100% deste grupo caiu de 32 para 30, o *core microbiome* 90% caiu de 115 para 47 e o *core microbiome* 80% caiu de 229 para 74.
- 2) Correção para o tamanho menor do grupo – Para isto, usamos os valores de corte para cada indivíduo, conforme calculado acima, considerando apenas as *OTUs* acima de uma dada frequência. A seguir, calculamos o *core microbiome* para 4 níveis de sub-grupos: todos os possíveis pares, todos os possíveis trios, os dois quartetos e para todos os indivíduos simultaneamente.

Os valores obtidos para estes microbiomas, conforme o tamanho dos sub-grupos comparados são apresentados a seguir, na **Tabela 11**.

Tabela 11 – Número de *core OTUs* encontrados em 100% das amostras dos sub-grupos de indivíduos do grupo CFB.

Sub-grupos	Média de amostras do <i>core microbiome</i>
<i>Core microbiome</i> de todos os possíveis pares	287,1 +/- 30.1
<i>Core microbiome</i> de todos os possíveis trios	189,4 +/- 20.2
<i>Core microbiome</i> dos dois quartetos possíveis.	139,6 +/- 19.6
<i>Core microbiome</i> de todos os 5 indivíduos CFB	107

De posse destes valores, testamos diversos modelos de regressão, buscando qual ofereceria o menor erro (calculado pelo método *Root Mean Square Error* – *RMSE*). As funções avaliadas e seus respectivos RMSEs calculados foram: Função linear (17,089); Função quadrática (12,712); Função exponencial (8,577); Função logarítmica (8,195); Função recíproca (2,408); Função de potência (1,094). Deste modo, selecionamos a função de menor erro – Função de potência - e derivamos uma equação de regressão não-linear capaz de estimar o tamanho do *core microbiome* do grupo CFB.

A equação derivada, seguindo a Lei de Potência ($Y=ax^b$) foi:

$$y = e^{(6.41091234116 + -1.07119716283 * \ln(x))}$$

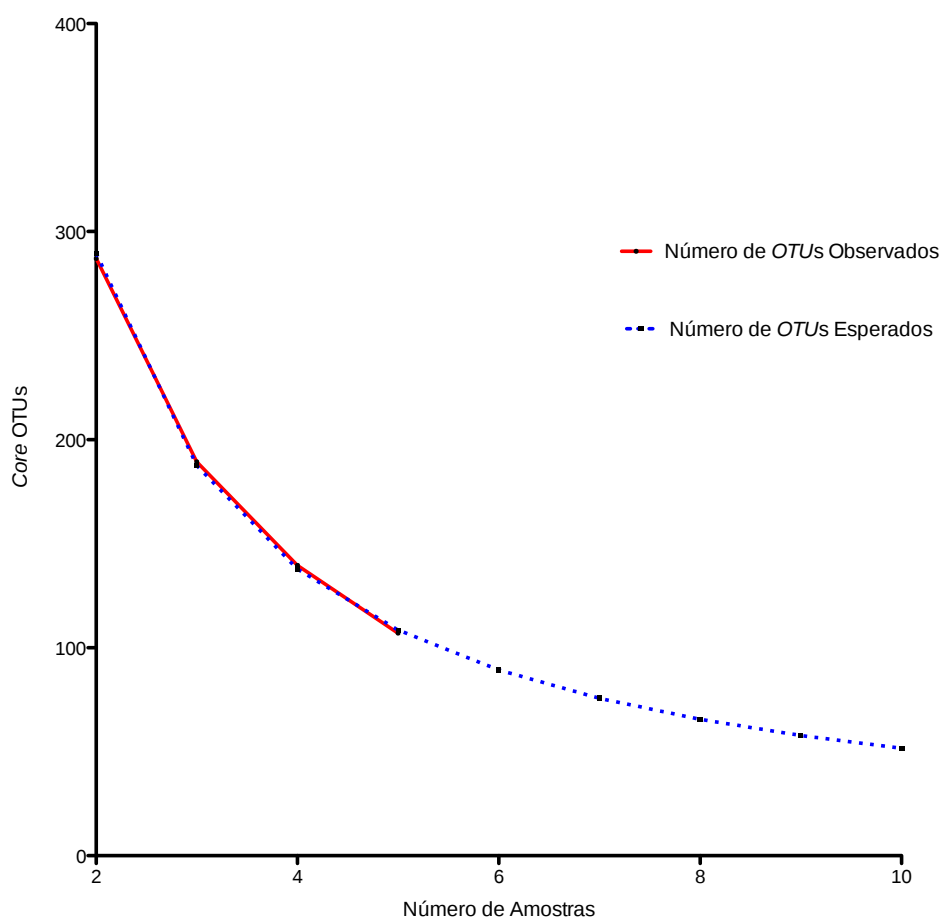
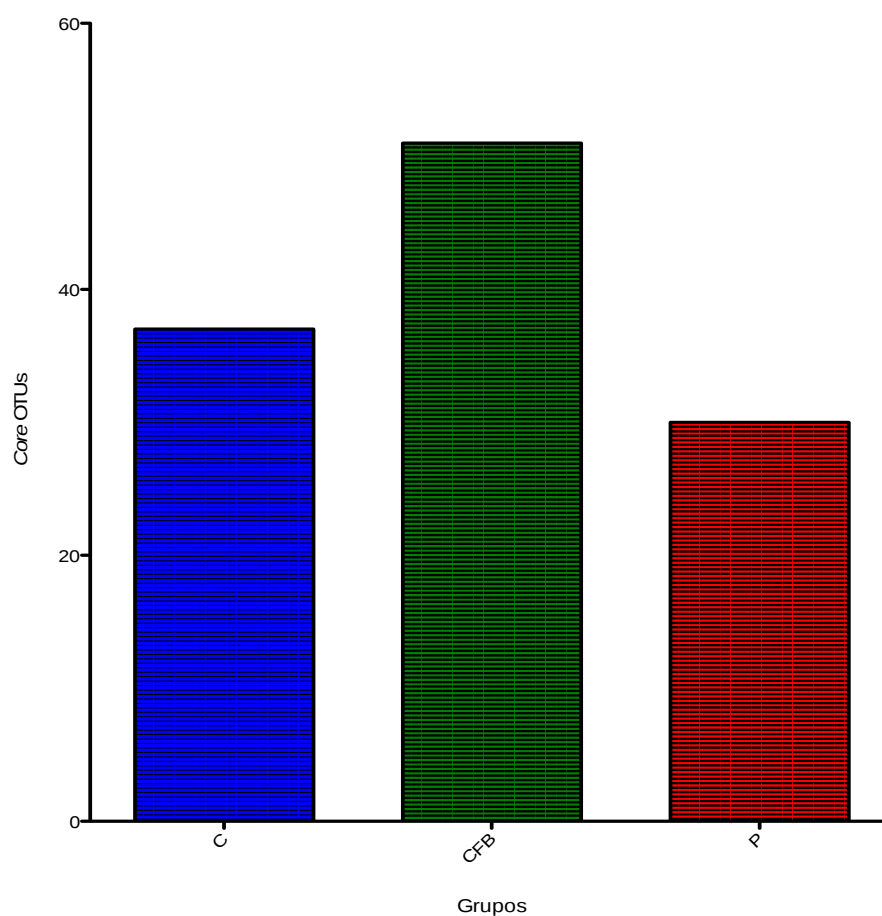


Figura 13 – Gráfico mostrando o número de *core OTUs* observado e calculado para o grupo CFB, considerando diferentes quantidades de amostras avaliadas.

Aplicando esta equação, temos que para um $N=10$ amostras do grupo CFB é calculado um *core microbiome* de 51.4 *OTUs* enquanto um $N=11$ levaria a um *core microbiome* estimado de 47.8 *OTUs*. Deste modo, a **Figura 14**, abaixo, mostra uma representação gráfica do *core microbiome* calculado para todos os 3 grupos de estudo.



Nota – para o grupo CFB o valor plotado foi calculado de acordo com a fórmula apresentada na página 70, utilizando um N=10 amostras.

Figura 14 – Gráfico de barras mostrando o número de *core OTUs* observadas em 100% das amostras nos 3 grupos, considerando-se 10 amostras para o grupo CFB

A seguir, nas **Tabelas 12, 13 e 14**, relacionamos as *OTUs* presentes em 100% das amostras do nosso grupo de controles, do nosso grupo de controles que fumam e bebem e do nosso grupo de pacientes, com sua respectiva taxonomia atribuída, além do número de sequências normalizadas por 100.000.

Tabela 12 – Core OTUs encontrados em 100% das amostras do grupo dos controles (C)

#OTU ID	Consensus Lineage	C01	C04	C06	C07	C09	C15	C16	C19	C20	C21	Mediana	Média	P ¹	CFB ²
30807	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	50	33	105	1327	154	24	69	71	11	85	70	193	125 (100%)	115 (100%)
42061	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	17	56	71	1680	318	57	552	182	11	249	127	319	217 (100%)	269 (100%)
43059	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	25	167	110	517	119	24	563	150	22	224	135	192	356 (90%)	98 (100%)
52726	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	145	156	1586	377	1604	849	2456	1398	11	1964	1124	1055	1161 (100%)	2512 (100%)
71322	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Neisseria	145	733	429	87	661	81	14094	1311	98	2258	545	1990	1611 (90%)	454 (100%)
81646	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	161	106	43	9	89	49	44	39	22	20	43	58	55 (90%)	23 (100%)
88717	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	377	622	624	13095	725	32	981	1122	229	638	631	1844	1843 (100%)	2039 (100%)
95915	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	66	50	24	155	60	24	7	16	44	10	34	46	120 (100%)	64 (100%)
107429	"Proteobacteria"; "Epsilonproteobacteria"; Campylobacteriales; Helicobacteraceae; Sulfurovum	54	78	95	9	194	89	149	142	65	219	92	109	47 (80%)	44 (100%)
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	940	900	8411	1104	11976	720	8219	6572	87	10849	3838	4978	3293 (100%)	8799 (100%)
120655	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	2235	2982	1548	408	2454	429	1857	2180	11	1555	1706	1566	1493 (100%)	1926 (100%)
123281	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococccineae; Intrasporangiaceae; Serinicoccus	17	50	681	3	15	437	18	79	250	189	64	174	11 (55%)	11 (90%)
157240	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	29	111	919	28	94	49	4	403	11	588	71	224	5 (36%)	8 (100%)
160926	"Firmicutes"; "Bacilli"; Lactobacillales; Carnobacteriaceae; Granulicatella	1784	1016	1019	690	442	534	1054	1777	109	3628	1018	1205	1004 (100%)	2620 (100%)
166220	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	12	50	810	46	124	8	18	355	163	35	48	162	71 (55%)	190 (100%)
177278	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	203	1860	3439	668	924	8	33	316	65	1326	492	884	93 (100%)	1339 (100%)
184308	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	50	178	962	18954	1008	137	1181	869	305	1306	916	2495	2257 (100%)	3074 (100%)
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	8133	4315	7840	913	11494	1536	3386	5308	348	7191	4812	5047	3937 (100%)	6310 (100%)
209284	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococccineae; Micrococccaceae; Rothia	459	766	486	186	894	1949	1312	1074	33	942	830	810	1420 (100%)	2710 (100%)
233445	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Neisseria	83	100	48	3	60	65	273	95	98	115	89	94	86 (90%)	4 (90%)
236642	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	339	106	57	22	253	73	65	142	11	100	86	117	141 (100%)	142 (100%)
237369	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Neisseria	8	28	29	3	10	16	254	47	11	95	22	50	41 (60%)	5 (100%)
241855	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	145	50	81	3	40	121	44	8	11	105	47	61	38 (90%)	31 (100%)
249042	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	406	966	1086	77	2513	809	3303	2733	359	2113	1026	1437	761 (100%)	2566 (100%)
259581	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Kingella	4	56	48	6	268	8	15	213	98	30	39	75	14 (70%)	3 (90%)
277502	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Kingella	8	44	52	6	243	16	33	134	33	25	33	60	14 (60%)	6 (100%)
295297	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	4	17	48	12	10	8	7	24	33	5	11	17	23 (80%)	7 (100%)
296426	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	4	39	110	461	99	49	752	95	44	164	97	182	438 (90%)	106 (100%)
310184	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Neisseria	956	3204	700	136	1063	218	9607	2243	588	2372	1010	2109	1805 (100%)	459 (100%)
314198	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	17	205	4082	563	1108	24	7	348	54	1889	277	830	55 (90%)	1782 (100%)
323914	"Firmicutes"; "Clostridia"; Clostridiales; Lachnospiraceae; unclassified_Lachnospiraceae	25	28	52	34	104	8	73	103	11	20	31	46	25 (90%)	63 (90%)
327425	"Bacteroidetes"; "Flavobacteria"; Flavobacteriales; Flavobacteriaceae; unclassified_Flavobacteriaceae	46	489	129	65	139	32	218	324	718	75	134	223	102 (90%)	168 (100%)
338554	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	844	1883	1372	74	3169	1205	2529	3420	544	3189	1627	1823	589 (100%)	2424 (100%)
363086	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	95	1244	800	62	124	24	11	379	207	15	110	296	66 (55%)	143 (100%)
381866	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	29	50	152	7152	472	40	1166	719	163	219	191	1016	287 (100%)	506 (100%)
412166	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococccineae; Micrococccaceae; Rothia	418	428	676	260	1147	3735	2489	1098	65	1794	887	1211	1876 (100%)	3902 (100%)
412169	"Firmicutes"; "Bacilli"; Lactobacillales; Carnobacteriaceae; Granulicatella	522	339	1681	1228	750	1504	1966	3144	272	8407	1366	1981	1300 (100%)	4435 (100%)
413823	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	41	605	476	201	20	16	11	150	229	154	152	190	300 (100%)	450 (100%)

¹: Média de *reads* normalizados por 100.000 encontrados no grupo P para esta OTU e em parênteses quantas amostras apresentaram esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo CFB para esta OTU e em parênteses quantas amostras apresentaram esta OTU

Tabela 13 – Core OTUs encontrados em 100% das amostras do grupo dos controles que fumam e bebem (CFB)

#OTU ID	Consensus Lineage	CFB10	CFB18	CFB29	CFB30	CFB31	Mediana	Média	P ¹	C ²
834	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	4	38	39	38	33	38	30	1 (20%)	1 (20%)
1822	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	4	15	20	4	7	7	10	3 (70%)	4 (50%)
13458	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	64	12	11	5	16	12	22	51 (80%)	73 (70%)
14512	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	25	634	41	11	211	41	184	23 (70%)	502 (80%)
16542	"Firmicutes"; "Bacilli"; Bacillales; Staphylococcaceae; Gemella	394	152	225	301	115	225	237	53 (90%)	121 (80%)
17104	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	10	39	69	23	26	26	34	8 (60%)	11 (70%)
17710	"Proteobacteria"; Neisseriales; Neisseriaceae; Bergeriella	16	20	33	20	60	20	30	12 (60%)	5 (50%)
30807	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	226	86	102	64	94	94	115	125 (100%)	193 (100%)
32388	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	8	22	25	11	11	11	15	10 (60%)	10 (50%)
33478	"Firmicutes"; "Bacilli"; Bacillales; Staphylococcaceae; Gemella	13	9	13	15	9	13	12	1 (10%)	0 (0%)
39271	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	5	17	22	15	6	15	13	12 (70%)	14 (70%)
42061	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	412	357	242	81	255	255	269	217 (100%)	319 (100%)
43057	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	84	34	37	10	23	34	38	51 (100%)	81 (90%)
43059	"Proteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	137	215	13	14	113	113	98	356 (90%)	192 (100%)
52726	"Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	340	3512	1207	2444	5059	2444	2512	1161 (100%)	1055 (100%)
67463	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	8	11	22	12	29	12	17	0 (0%)	0 (0%)
70808	"Firmicutes"; "Bacilli"; Bacillales; Staphylococcaceae; Gemella	28	15	16	23	13	16	19	1 (20%)	3 (40%)
71322	"Proteobacteria"; BetaproteoNeisseriales; Neisseriaceae; Neisseria	1199	287	651	9	122	287	454	1611 (90%)	1990 (100%)
72200	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	17	249	22	73	51	51	82	33 (80%)	36 (80%)
76340	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	22	759	36	23	226	36	213	16 (40%)	186 (60%)
77021	"Proteobacteria"; Campylobacterales; Campylobacteraceae; Campylobacter	21	13	40	40	19	21	26	78 (70%)	20 (90%)
88717	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	4650	1406	1910	1128	1103	1406	2039	1483 (100%)	1844 (100%)
89405	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	7	9	11	6	21	9	11	19 (80%)	17 (60%)
95915	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	48	137	7	35	94	48	64	120 (100%)	46 (100%)
108789	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	7	13	16	8	28	13	14	2 (60%)	2 (40%)
112988	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	11	48	27	66	110	48	52	60 (100%)	32 (90%)
118095	"Firmicutes"; "Bacilli"; Bacillales; Staphylococcaceae; Staphylococcus	114	68	100	5	7	68	59	20 (70%)	197 (90%)
118712	"Actinobacteria"; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium	15	29	9	10	36	15	20	17 (50%)	74 (90%)
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	1281	11823	7305	15199	8387	8387	8799	3293 (100%)	4978 (100%)
118915	"Actinobacteria"; Actinobacteridae; Actinomycetales; Propionibacterineae; Propionibacteriaceae; Propionibacterium	206	92	128	85	17	92	106	43 (90%)	276 (80%)
120655	"Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	215	2610	1265	1408	4135	1408	1926	1493 (100%)	1566 (100%)
122093	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	18	9	4	165	85	18	56	26 (60%)	50 (70%)
126815	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	10	5	11	7	20	10	10	7 (60%)	5 (80%)
136222	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinobaculum	6	262	11	12	6	11	59	10 (20%)	125 (30%)
136478	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	6	13	19	16	49	16	20	3 (30%)	16 (70%)
144956	"Firmicutes"; "Clostridia"; Clostridiales; Lachnospiraceae; Oribacterium	17	47	364	122	177	122	146	200 (70%)	165 (80%)
155362	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	4	43	32	54	27	32	32	52 (90%)	66 (90%)
158798	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	10	58	21	58	100	58	49	74 (100%)	19 (90%)
160926	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	1183	3454	4859	1846	1757	1846	2620	1004 (100%)	1205 (100%)
163489	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	58	5	22	10	5	10	20	4 (70%)	5 (40%)
166220	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	63	751	35	10	91	63	190	71 (55%)	162 (100%)
169589	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	175	279	277	91	127	175	190	67 (80%)	176 (80%)
177278	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	89	1348	67	4200	991	991	1339	93 (100%)	884 (100%)
177530	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	20	17	31	5	10	17	17	5 (70%)	15 (60%)
184308	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	7474	2121	2174	1803	1800	2121	3074	2257 (100%)	2495 (100%)
184741	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	8	8	11	6	16	8	10	11 (70%)	8 (50%)
189574	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	72	209	169	72	325	169	169	71 (90%)	77 (90%)
198535	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	15	89	3821	3876	248	248	1610	3155 (100%)	915 (90%)
200268	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	25	576	5437	6326	1085	1085	2690	2801 (100%)	821 (90%)
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	1085	9112	6203	8448	6700	6700	6310	3937 (100%)	5047 (100%)
209284	"Actinobacteria"; ActinoActinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	1899	1705	2933	1619	5395	1899	2710	1420 (100%)	810 (100%)
212418	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	9	97	61	97	58	61	64	108 (80%)	95 (80%)
214208	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Megaspheara	5	95	71	66	54	66	58	0 (0%)	2 (30%)

223532	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	214	385	281	349	201	281	286	162 (80%)	741 (80%)
232192	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified_Pasteurellaceae	244	13	39	56	477	56	166	546 (70%)	465 (80%)
232195	"Proteobacteria"; Epsilonproteobacteria; Campylobacteriales; Campylobacteriaceae; Campylobacter	21	15	53	41	21	21	30	123 (90%)	37 (90%)
236120	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	4	17	20	7	6	7	11	8 (50%)	12 (80%)
236642	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	20	220	152	168	151	152	142	141 (100%)	117 (100%)
239464	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	16	13	19	21	54	19	25	32 (90%)	30 (70%)
241855	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	7	48	29	45	28	29	31	38 (90%)	61 (100%)
244578	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	5	7	12	9	32	9	13	3 (60%)	5 (50%)
245648	"Firmicutes"; "Bacilli"; Bacillales; "Staphylococcaceae"; Gemella	33	17	27	28	16	27	24	6 (40%)	5 (40%)
248880	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	11	264	26	42	34	34	75	25 (60%)	289 (90%)
248938	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	84	276	1087	1342	574	574	673	692 (100%)	158 (80%)
249042	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	2844	3159	3307	85	3434	3159	2566	761 (100%)	1437 (100%)
250405	"Firmicutes"; "Bacilli"; Bacillales; "Staphylococcaceae"; Gemella	840	347	434	622	288	434	506	80 (80%)	136 (90%)
254605	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	11	83	49	68	51	51	52	48 (70%)	63 (80%)
269970	"Firmicutes"; "Clostridia"; Clostridiales; "Ruminococcaceae"; Oscillibacter	6	9	22	9	30	9	15	0 (0%)	0 (0%)
279866	"Firmicutes"; "Clostridia"; Clostridiales; "Lachnospiraceae"; Oribacterium	9	74	8	31	30	30	31	172 (80%)	82 (80%)
280528	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	116	353	218	109	469	218	253	86 (100%)	109 (90%)
291965	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	8	7	12	10	27	10	13	22 (70%)	30 (70%)
295250	"Firmicutes"; "Bacilli"; Bacillales; "Staphylococcaceae"; Gemella	14	9	11	15	7	11	11	2 (30%)	15 (50%)
296426	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus	134	245	14	16	122	122	106	438 (90%)	182 (100%)
298308	"Fusobacteria"; "Fusobacteria"; "Fusobacteriales"; "Leptotrichiaceae"; Leptotrichia	23	239	17	9	121	23	82	13 (60%)	157 (80%)
298928	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	102	289	1164	1577	642	642	755	739 (100%)	105 (90%)
302885	"Firmicutes"; "Clostridia"; Clostridiales; "Lachnospiraceae"; Oribacterium	31	93	436	196	250	196	201	207 (80%)	156 (90%)
304516	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	4	8	4	25	29	8	14	21 (80%)	7 (60%)
310184	"Proteobacteria"; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria	1221	236	706	7	127	236	459	1805 (100%)	2109 (100%)
314198	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	82	1384	40	6334	1068	1068	1782	55 (90%)	830 (100%)
326731	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	5	267	222	86	455	222	207	314 (80%)	385 (80%)
327425	"Bacteroidetes"; Flavobacteria; "Flavobacteriales"; Flavobacteriaceae; unclassified_Flavobacteriaceae	226	71	439	10	92	92	168	102 (90%)	223 (100%)
328269	"Fusobacteria"; "Fusobacteria"; "Fusobacteriales"; "Leptotrichiaceae"; Leptotrichia	22	165	14	9	96	22	61	10 (50%)	161 (90%)
331851	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	41	43	72	34	135	43	65	25 (70%)	8 (50%)
338554	"Proteobacteria"; Gammaproteobacteria Pasteurellales; Pasteurellaceae; Haemophilus	2685	2713	3450	74	3199	2713	2424	589 (100%)	1823 (100%)
341135	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	4	11	10	35	21	11	16	42 (90%)	20 (70%)
347100	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	10	13	12	9	31	12	15	15 (70%)	19 (70%)
349212	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	35	84	4	25	47	35	39	78 (100%)	28 (70%)
363086	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	48	554	34	8	70	48	143	66 (55%)	269 (100%)
366238	"Actinobacteria"; Actinobacteridae; Actinomycetales; Propionibacterineae; Propionibacteriaceae; Propionibacterium	191	88	123	59	9	88	94	48 (90%)	353 (90%)
369892	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	8	8	8	4	13	8	8	25 (80%)	17 (60%)
371151	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	20	15	13	20	17	17	17	15 (80%)	14 (50%)
376707	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	4	44	26	76	36	36	37	20 (80%)	17 (90%)
380962	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	15	797	47	878	98	98	367	167 (80%)	56 (80%)
381866	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	250	659	799	508	313	508	506	287 (100%)	1016 (100%)
387090	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	6	17	25	12	11	12	14	6 (50%)	9 (70%)
388312	"Firmicutes"; "Bacilli"; "Lactobacillales"; Streptococcaceae; Streptococcus	19	10	13	11	8	11	12	10 (90%)	17 (60%)
388851	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	5	16	24	12	11	12	14	19 (70%)	7 (80%)
391798	"Bacteroidetes"; Flavobacteria; "Flavobacteriales"; Flavobacteriaceae; unclassified_Flavobacteriaceae	392	108	673	14	129	129	263	56 (70%)	209 (90%)
392210	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	5	72	41	83	53	53	51	16 (70%)	13 (90%)
401714	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	7	7	15	5	24	7	12	24 (70%)	8 (60%)
411103	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	4	80	56	30	164	56	67	12 (70%)	21 (70%)
412166	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	2597	2554	3970	2653	7734	2653	3902	1876 (100%)	1211 (100%)
412169	"Firmicutes"; "Bacilli"; "Lactobacillales"; "Carnobacteriaceae"; Granulicatella	2184	5730	8218	2883	3162	3162	4435	1300 (100%)	1981 (100%)
413823	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	16	900	78	1070	186	186	450	300 (100%)	190 (100%)
415141	"Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	5	7	11	10	27	10	12	33 (60%)	39 (60%)
421117	"Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	11	213	23	26	30	26	61	28 (60%)	184 (80%)
422556	"Proteobacteria"; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified_Pasteurellaceae	224	11	45	80	489	80	170	1057 (70%)	238 (70%)

¹: Média de *reads* normalizados por 100.000 encontrados no grupo P para esta OTU e em parênteses quantas amostras apresentaram esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo C para esta OTU e em parênteses quantas amostras apresentaram esta OTU

Tabela 14 – Core OTUs encontrados em 100% das amostras do grupo dos pacientes (P)

#OTU ID	Consensus Lineage	P01	P03	P04	P08	P09	P11	P23	P27	P33	P36	P46	Mediana	Média	C ¹	CFB ²
30807	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	73	12	568	66	61	57	216	39	75	64	144	66	125	193 (100%)	115 (100%)
42061	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	177	37	699	197	121	57	369	19	101	57	549	121	217	319 (100%)	269 (100%)
43057	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	27	6	76	46	8	16	9	58	221	64	28	28	51	81 (90%)	38 (100%)
52726	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	3089	1321	2005	1203	1113	274	134	1702	65	305	1564	1203	1161	1055 (100%)	2512 (100%)
88717	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	962	1688	3273	1181	952	346	5808	185	236	458	1220	962	1483	1844 (100%)	2039 (100%)
95915	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	73	6	322	38	197	31	16	49	50	57	477	50	120	46 (100%)	64 (100%)
112988	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	73	100	98	34	74	31	5	146	5	38	59	59	60	32 (90%)	52 (100%)
118912	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	4131	2978	5803	2921	4642	868	298	642	106	1723	12109	2921	3293	4978 (100%)	8799 (100%)
120655	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	2740	1246	1798	1200	1313	2951	91	1119	20	2282	1665	1313	1493	1566 (100%)	1926 (100%)
158798	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	209	31	49	71	148	26	2	117	5	32	120	49	74	19 (90%)	49 (100%)
160926	"Firmicutes"; "Bacilli"; Lactobacillales; Carnobacteriaceae; Granulicatella	1040	237	339	706	1294	2026	1497	136	297	998	2475	998	1004	1205 (100%)	2620 (100%)
177278	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	56	156	120	35	135	98	14	233	70	89	11	89	93	884 (100%)	1339 (100%)
183901	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharomonospora	37	62	49	20	15	31	2	126	5	25	54	31	39	40 (80%)	27 (100%)
184308	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	1720	1738	4775	1611	1260	78	10398	185	900	95	2067	1611	2257	2495 (100%)	3074 (100%)
198535	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	2467	6846	11217	183	6644	4729	177	224	50	2155	9	2155	3155	915 (90%)	1610 (100%)
200268	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	4073	4460	11775	387	8327	708	255	136	45	165	479	479	2801	821 (90%)	2690 (100%)
209283	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	2692	3545	5054	2500	4560	8389	134	1595	342	6541	7952	3545	3937	5047 (100%)	6310 (100%)
209284	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	1312	1103	945	529	342	1840	1381	3210	427	178	4351	1103	1420	810 (100%)	2710 (100%)
236642	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Veillonella	125	25	120	127	170	408	2	49	10	267	244	125	141	117 (100%)	142 (100%)
248938	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	1378	37	481	155	743	1065	30	1021	5	985	1715	743	692	158 (80%)	673 (100%)
249042	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	556	903	49	387	9	222	5462	136	5	470	176	222	761	1437 (100%)	2566 (100%)
280528	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	74	56	93	70	42	26	11	49	55	6	466	55	86	109 (90%)	253 (100%)
298928	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes	2507	19	546	191	720	398	34	982	15	540	2176	540	739	105 (90%)	755 (100%)
310184	"Proteobacteria"; "Betaproteobacteria"; Neisseriales; Neisseriaceae; Neisseria	2146	7189	612	7129	155	26	66	29	196	2180	129	196	1805	2109 (100%)	459 (100%)
338554	"Proteobacteria"; "Gammaproteobacteria"; Pasteurellales; Pasteurellaceae; Haemophilus	218	1246	82	268	4	610	2625	331	5	992	104	268	589	1823 (100%)	2424 (100%)
349212	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	60	6	153	9	66	16	16	126	191	6	209	60	78	28 (70%)	39 (100%)
381866	"Firmicutes"; "Bacilli"; Lactobacillales; Streptococcaceae; Streptococcus	250	12	956	94	270	16	250	117	272	89	828	250	287	1016 (100%)	506 (100%)
412166	"Actinobacteria"; "Actinobacteria"; Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Rothia	2179	1227	1098	585	279	1241	2186	5077	654	197	5919	1227	1876	1211 (100%)	3902 (100%)
412169	"Firmicutes"; "Bacilli"; Lactobacillales; Carnobacteriaceae; Granulicatella	1897	262	541	1025	1709	367	2675	350	1051	305	4116	1025	1300	1981 (100%)	4435 (100%)
426865	"Firmicutes"; "Clostridia"; Clostridiales; Veillonellaceae; Megaspheara	26	56	44	15	53	16	2	1741	453	25	20	26	223	194 (80%)	32 (100%)

¹: Média de *reads* normalizados por 100.000 encontrados no grupo C para esta OTU e em parênteses quantas amostras apresentaram esta OTU

²: Média de *reads* normalizados por 100.000 encontrados no grupo CFB para esta OTU e em parênteses quantas amostras apresentaram esta OTU

Vemos pelas tabelas que *OTUs* distintas, mas com a mesma atribuição taxonômica, são frequentemente encontradas – indicando possivelmente a ocorrência de diferentes espécies do mesmo gênero bacteriano. A ocorrência destas bactérias na mucosa oral de todos os indivíduos também encontra suporte na literatura pois gêneros como *Streptococcus*, *Neisseria*, *Actinomyces* e *Granulicatella* são comumente achados e descritos na flora bucal.

No *core OTUs* do grupo dos pacientes podemos observar duas *OTUs* que possuem um comportamento interessante. A *OTU* 298928 foi atribuída ao gênero *Actinomyces*, e possui um número elevado de sequências normalizadas por 100.000, comparado ao grupo de controles. Esta *OTU* se encontra em todas as amostras do grupo dos pacientes e do grupo dos CFBs e possui tanto uma média quanto uma mediana de sequências maior que o grupo dos controles (média de 739 e 755, respectivamente vs 105 em 90% das amostras do grupo C). Outro aspecto interessante desta *OTU* é que ela parece estar diminuída no grupo C e não se encontra em todas as amostras, depois sua presença parece aumentar significativamente no grupo CFB em todas as amostras (apesar de se tratarem de apenas cinco) e sua quantidade permanece aumentada no grupo dos pacientes, além de ser encontrada em todas as amostras. Fizemos um BLAST com a sequência representativa desta *OTU* contra o banco *HOMD* (CHEN et al. 2010) *online* para ver se seria possível distinguir que espécie de *Actinomyces* seria esta (pois o banco *RDP* não possui informações à nível de espécies), e encontramos que ela parece representar a espécie *Actinomyces graevenitzi* com uma porcentagem de identidade de 97.2% (**Figura 15**). Este gênero está envolvido em diversas infecções bacterianas e processos inflamatórios e já foi demonstrado que uma infecção por algumas

espécies de *Actinomyces* causa uma lesão semelhante, em termos clínicos, à duas lesões neoplásicas, o câncer de ovário (DRIESSEN et al. 2012) e o câncer de pulmão (SCHWEIGERT et al. 2012). Certas espécies do gênero *Actinomyces* causam infecções bacterianas graves em indivíduos immuno-comprometidos, corroborando os dados de que os pacientes realmente possuem uma maior diversidade bacteriana e que esta maior diversidade pode ser causada por um estado imuno deprimido, permitindo uma maior colonização de espécies bacterianas, em especial às pertencentes ao gênero *Actinomyces*. Um fato interessante é de que este gênero também se encontra mais elevado no grupo de controles que fumam e bebem.

A redução da competência imune na mucosa de pacientes com câncer da cavidade oral parece ser um aspecto relevante neste sentido. Um estudo brasileiro publicado em 2003, avaliou os níveis séricos e salivares de IgA em 34 pacientes com câncer (boca ou orofaringe) e em 34 controles pareados. Enquanto alterações séricas não foram observadas entre os grupos, o estudo detectou que os pacientes apresentavam menos da metade dos níveis salivares de IgA quando comparados aos controles (7.2 +/- 5.0 mg/dl para casos e 17.0 +/- 10.4 mg/dl para controles) (DE SOUZA et al. 2003). Posteriormente, SHPITZER et al. (2007) compararam a composição da saliva entre 25 pacientes portadores de OSCC e 25 controles. Dentre os diversos parâmetros avaliados, os autores observaram que os níveis de IgA secretória estavam reduzidos em 45% nos pacientes, quando comparados aos controles ($P < 0,001$), corroborando o estudo brasileiro. Mais recentemente, AVŞAR et al. (2009) avaliaram os níveis salivares de diversos fatores (proteínas, eletrólitos, imunoglobulinas e outros) em um grupo de 180 crianças, sendo que metade era composta por fumantes passivos.

O principal achado do estudo foi que as crianças do grupo fumante-passivo apresentaram níveis salivares de IgA significativamente menores que as crianças do grupo controle ($P < 0,05$). Sendo assim, o estado imune da mucosa oral parece ser afetado pelo fumo, e está significativamente reduzido nos pacientes com câncer oral.

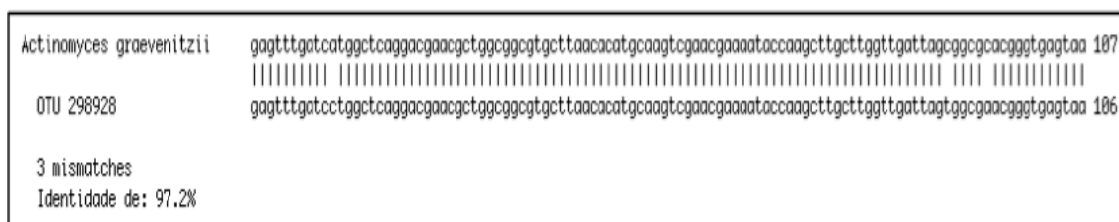


Figura 15 – Alinhamento por BLAST da OTU 298928 com a sequência de *Actinomyces graevenitzii* pelo banco online HOMB

Outra OTU que possui um comportamento interessante é a 52726, que foi atribuída ao gênero *Saccharomonospora*. Ela também se encontra em todas as amostras do grupo dos pacientes e do grupo dos CFBs e possui um número tanto médio quanto mediano maior de sequências que o grupo dos controles C (médias de 1161 e 2512, respectivamente, vs 1055 em 100% das amostras do grupo C). Esta OTU tem níveis basais no grupo C, e está elevada nos grupos P e CFB. Foi demonstrado por MALONEY et al. (2009) que uma espécie de *Saccharomonospora* produzia metabolitos secundários, como o alcalóide lodopiridona, que eram citotóxicos à cultura de células de câncer de cólon. De novo, fizemos um BLAST para determinar qual espécie era correspondente à esta OTU, porém não obtivemos sucesso em atribuir essa sequência à uma espécie específica.

Diante da ausência de estudos na literatura que tenham investigado um elevado número de amostras, com uma cobertura bastante elevada nas mesmas, ainda

é cedo para estimarmos o tamanho ou mesmo determinarmos a existência de um *core microbiome* da cavidade oral humana. No entanto, dentro do universo aqui sequenciado, estimamos que este esteja ao redor de 100 a 200 *OTUs*.

4.6 DISCREPÂNCIAS ENTRE OS CORE MICROBIOMES DE CONTROLES E PACIENTES

Diante das listas de *OTUs* mais consistentemente encontradas em cada grupo, passamos então a investigar se estas seriam capazes de revelar diferenças mais consistentes entre os grupos. Como já descrito, nosso foco foi centrado na comparação entre os grupos C e P. No entanto, permanecemos com os dados de CFBs para mostrar o comportamento das *OTUs* de interesse também neste grupo. A **Figura 16** mostra os *OTUs* que compõem o núcleo de 100, 90 e 80% das amostras de cada grupo e se estes *OTUs* estão presentes nos outros grupos através de um diagrama de *Venn*.

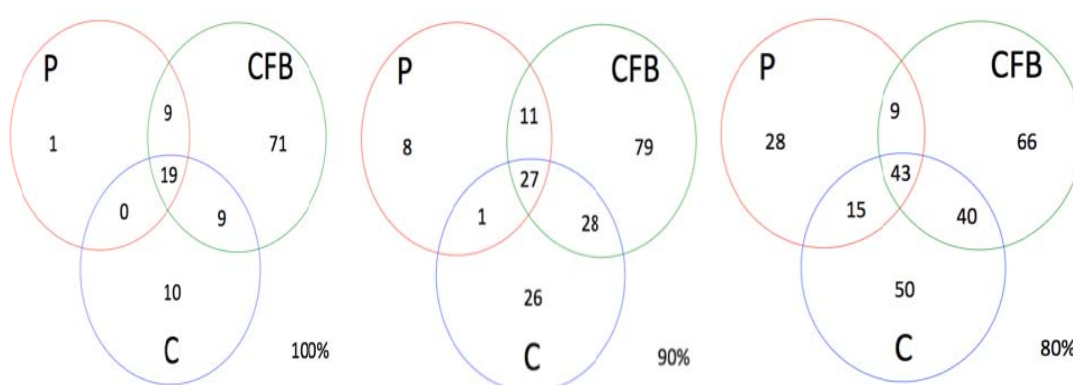


Figura 16 - Diagram de *Venn* com as *core OTUs* encontradas em 100, 90 e 80% das amostras do três grupos.

Como pode se observar pelo diagrama de *Venn* e pelas tabelas anteriores, o núcleo comum de espécies bacterianas presentes em todas as amostras de todos os grupos é composto por 19 *OTUs*, distribuídas em um número limitado de gêneros: *Streptococcus*, *Saccharomonospora*, *Veillonella*, *Granulicatella*, *Actinomyces*, *Rothia*, *Haemophilus* e *Neisseria*.

Na **Figura 16**, vemos que uma *OTU* é encontrada unicamente no *core microbiome* de pacientes, quando 100% das amostras são consideradas, esta *OTU* corresponde ao gênero *Saccharomonospora*. Na **Tabela 15** vemos a comparação do número de sequências normalizadas desta *OTU* nos outros grupos.

Dentre as *OTUs* que aparecem somente nos pacientes e controles que fumam e bebem em todas as amostras temos somente dois gêneros *Streptococcus* e *Actinomyces*. O gênero *Streptococcus* apareceu com 3 *OTUs* somente nestes dois grupos e o gênero *Actinomyces* apareceu com 6 *OTUs*.

Observando o diagrama de *Venn* com 90% das amostras, observamos que nove *OTUs* estão presentes somente no grupo dos pacientes, estas correspondem aos gêneros *Actinomyces* (3 *OTUs*), *Veillonella* (3 *OTUs*), *Neisseria* (1 *OTU*) e *Streptococcus* (2 *OTUs*).

Observando o diagrama de *Venn* com 80% das amostras, observamos que 28 *OTUs* estão presentes somente no grupo dos pacientes, estas correspondem principalmente aos gêneros *Actinomyces*, *Veillonella* e *Streptococcus*, entre outros.

Em um estudo de KURKIVUORI et al. (2007) foi observado que certas espécies de *Streptococcus* produziam significativamente mais acetaldeído, a partir de etanol, do que outras. Os autores encontraram que as espécies *Streptococcus mitis*, *Streptococcus intermedius* e *Streptococcus salivarius* eram as que mais produziam

acetaldeído. Poderíamos postular, já que observamos *OTUs* diferentes que correspondem ao gênero *Streptococcus* e que são somente encontradas em 100%, em 90% e em 80% das amostras do grupo do CFB e pacientes, que existem espécies diferentes de *Streptococcus* entre estes dois grupos e o grupo de controles, e que estas espécies poderiam ser as que mais produzem acetaldeído, devido ao uso de álcool e tabaco. Tentamos discriminar qual seria a espécie correspondente a algumas destas *OTUs* utilizando um BLAST *online* com o banco *HOMD*, porém não obtivemos sucesso em distinguir uma espécie da outra, permanecendo como *Streptococcus sp.* Ainda é cedo para fazer tal afirmação, mas é um primeiro indício de que este gênero pode possuir alguma associação com o surgimento de neoplasias da mucosa oral influenciadas pelo consumo de álcool e tabaco.

Tabela 15 – Comparação do número de *reads* normalizados para a *OTU* correspondente ao gênero *Saccharomonospora* nos três grupos

	Número de reads/100k <i>OTU</i>
Amostra	<i>Saccharomonospora</i>
C01	8
C04	0
C06	10
C07	6
C09	40
C15	97
C16	94
C19	32
C20	0
C21	110
Mediana do Grupo C	21
Porcentagem de amostras que apresentam <i>reads</i>	80%
CFB10	3
CFB18	37
CFB29	9
CFB30	19
CFB31	65
Mediana do Grupo CFB	19
Porcentagem de amostras que apresentam <i>reads</i>	100%
P.01	37
P.03	62
P.04	49
P.08	20
P.09	15
P.11	31
P.23	2
P.27	126
P.33	5
P.36	25
P.46	54
Mediana do Grupo P	31
Porcentagem de amostras que apresentam <i>reads</i>	100%

Vemos pela tabela que realmente há uma maior quantidade *reads* normalizados para esta *OTU* no grupo dos pacientes do que nos controles. PUSHALKAR et al. (2011) avaliaram a microbiota salivar de 5 indivíduos com OSCC e de 2 controles, usando iniciadores que cobriam a região V4-V5 do gene 16S- rDNA (300pb) e geraram 54 mil sequências. Os gêneros mais abundantes encontrados nos pacientes foram; *Gemella*, *Streptococcus*, *Rothia*, *Porphyromonas* e *Lactobacillus*. Nos controles, os gêneros bacterianos mais abundantes foram:

Prevotella, *Neisseria*, *Leptotrichia*, *Capnocytophaga*, *Actinobacillus* e *Oribacterium*. Os autores encontraram 218 *OTUs* que eram comuns entre os dois grupos (controles e pacientes), porém não especificaram quais eram suas respectivas taxonomias. Os autores encontraram duas *OTUs* correspondentes ao gênero *Actinomyces* que estavam somente presentes no grupo dos pacientes, mostrando uma concordância entre os estudos. Em trabalho realizado por NAGY et al. (1998) bactérias presentes em 10 amostras de biofilme retiradas da superfície de tumores da cavidade oral e 10 amostras de biofilme retiradas da mucosa saudável dos mesmos pacientes, foram avaliadas através do método de cultivo. Os gêneros bacterianos mais abundantes encontrados nos biofilmes tumorais foram; *Veillonella*, *Fusobacterium*, *Prevotella*, *Porphyromonas*, *Actinomyces*, *Clostridium*, *Haemophilus*, *Enterobacteriaceae* e *Streptococcus*. No estudo de NAGY et al. (1998) uma maior quantidade do gênero bacteriano *Actinomyces* foi encontrado nos biofilmes tumorais, corroborando os dados achados neste estudo. Neste estudo achamos uma maior quantidade de *OTUs* no grupo dos pacientes e no grupo do controles que fumavam e bebiam pertencentes ao gênero *Veillonella*, o que também corrobora os achados de NAGY et al. (1998). Estas *OTUs* eram as mais abundantes nos dois grupos, e também compunham o *core microbiome* em ambos os grupos, demonstrando possuir alguma importância, seja ela em decorrência ou uma causa do processo de malignização do tecido mucoso oral.

4.7 AVALIAÇÃO DO EFEITO DO MAIOR NÚMERO DE SEQUÊNCIAS PARA O GRUPO CFB

Investigamos também se a quantidade muito maior de sequências geradas para o grupo CFB, fruto do uso de um chip diferente no sequenciamento, alterava as estimativas do número total de *OTUs* de modo significativo, não permitindo assim um análise comparativa entre os três grupos. Assim buscamos um conjunto de 23 mil sequências selecionadas ao acaso (número próximo à média de sequências geradas para o grupo C – aquele para o qual geramos menor quantidade de sequências). Na **Figura 17** observamos o número de *OTUs* gerados para cada amostra dos três grupos, mesmo aqueles que possuíam somente uma sequência em seu *cluster*.

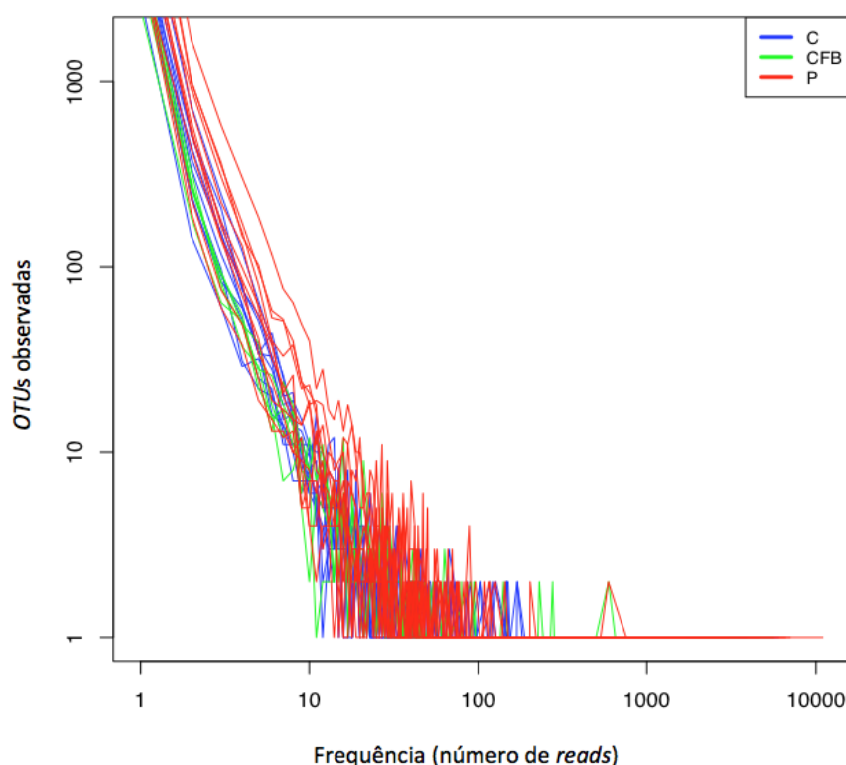


Figura 17 - Gráfico de grupos de frequência de *OTUs* observadas para cada amostra dos três grupos de estudo.

Observamos pelo gráfico que o número de *OTUs* gerados com qualquer frequência pelo grupo P foi maior do que nos outros, e que este é condizente com os dados de rarefação gerados anteriormente. Observamos ainda que a maioria das amostras do grupo dos pacientes possui um maior número de *OTUs* com menor frequência (<20X) do que os demais grupos. Fizemos ainda uma análise conjunta, agrupando os dados de todos os indivíduos de um mesmo grupo (**Figura 18**).

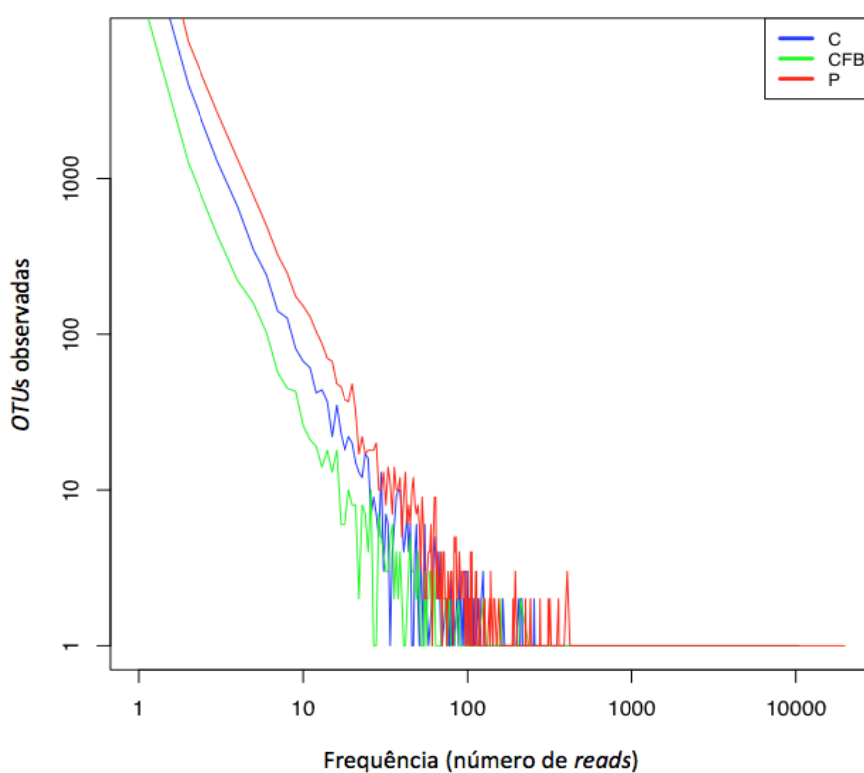


Figura 18 - Gráfico de grupos de frequência de *OTUs* observadas para os três grupos de estudo

Como se pode observar pelo gráfico, os pacientes ainda possuem uma maior quantidade de *OTUs* do que os dois grupos de controles, na maioria das classes de frequência. De modo oposto, o grupo CFB possui uma menor quantidade de *OTUs* para a maioria das classes de frequência, corroborando os gráficos de

saturação/rarefação. Prosseguimos com estas análises, desta vez dividindo o grupo de pacientes conforme estadiamento tumoral, como descrito anteriormente (**Figura 19**).

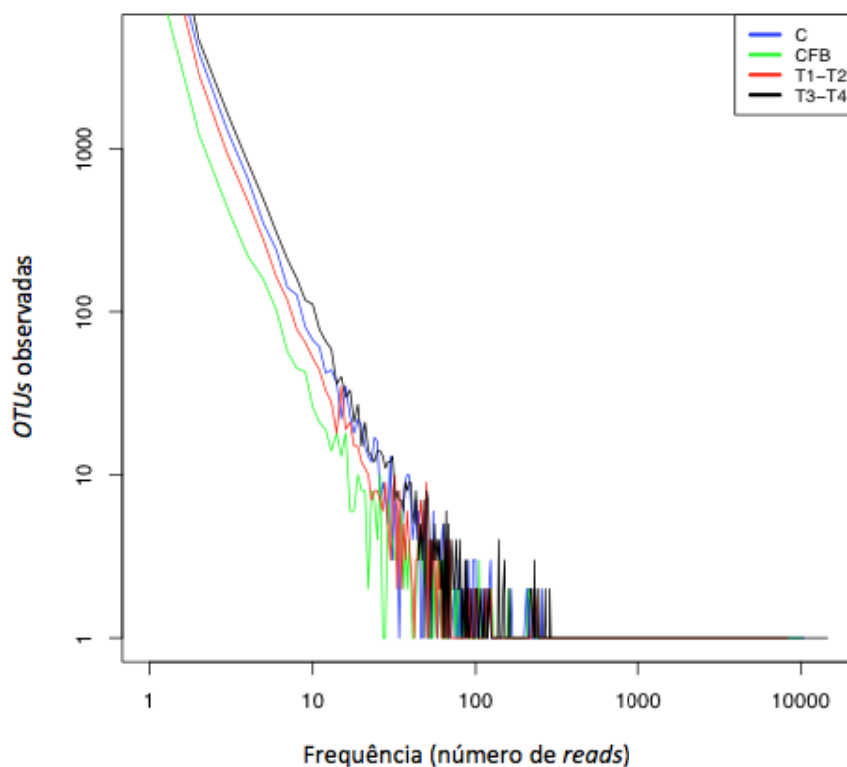
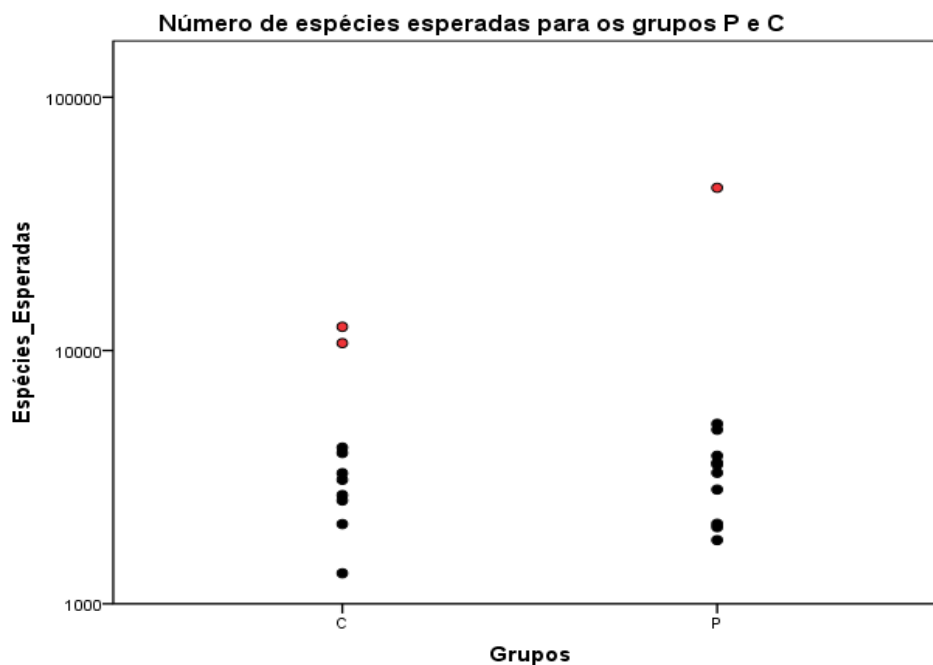


Figura 19 - Gráfico de grupos de frequência de *OTUs* observadas para os grupos de estudo, com ênfase no estadiamento tumoral para o grupo dos pacientes.

Observando o gráfico, o número de *OTUs* com frequência única é maior agora no grupo com pacientes com tumores T3 e T4, seguido pelo grupo dos controles e depois pelos pacientes T1 e T2, por último encontramos o grupo dos CFB. Deve se levar em consideração que o grupo dos controles possui um n amostral maior do que do grupo dos pacientes T3 e T4 (10 vs 6), e mesmo assim observamos uma maior quantidade de *OTUs* com frequência única e com uma frequência maior, demonstrando sua maior diversidade. Os dados apresentados são condizentes com as curvas de saturação/rarefação mostradas anteriormente.

4.8 ESTIMATIVA DA RIQUEZA DA DIVERSIDADE MICROBIANA USANDO O CATCHALL

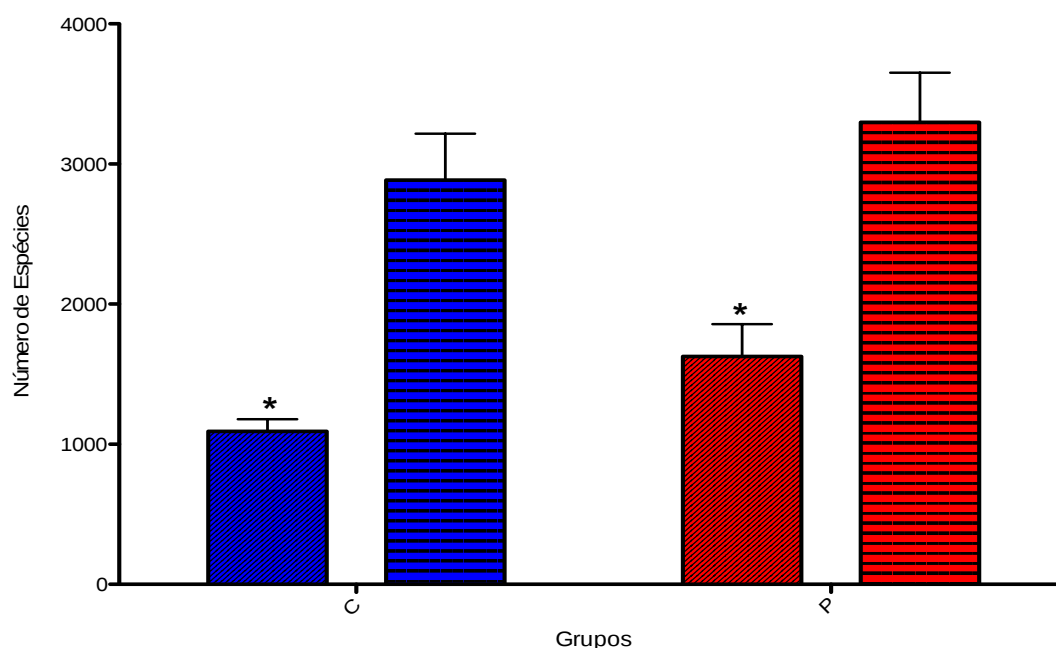
A seguir, buscamos investigar melhor a questão da suposta maior diversidade bacteriana observada no grupo dos pacientes (P), quando comparada ao grupo controle (C). Para isto, utilizamos uma estimativa de diversidade bacteriana gerada pelo programa Catchall (BUNGE et al. 2012). Primeiramente queríamos saber qual a estimativa da riqueza nos dois grupos, observada na **Figura 20**. Exluímos o grupo CFB destas análises pois o número de *reads* gerados para este grupo foi muito maior do que os gerado para os outros grupos e não conseguimos usar o programa Catchall de forma adequada com um *sub-set* das sequências do grupo CFB para estas análises.



Legenda – Temos o número de espécies esperadas para cada amostras dos dois grupos, controles e pacientes, conforme calculado pelo programa Catchall. As amostras que estão em vermelho, consideradas como *outliers*, foram excluídas do cálculo da média de espécies esperadas.

Figura 20 – Número de espécies esperadas para os grupos P e C gerados pelo Catchall

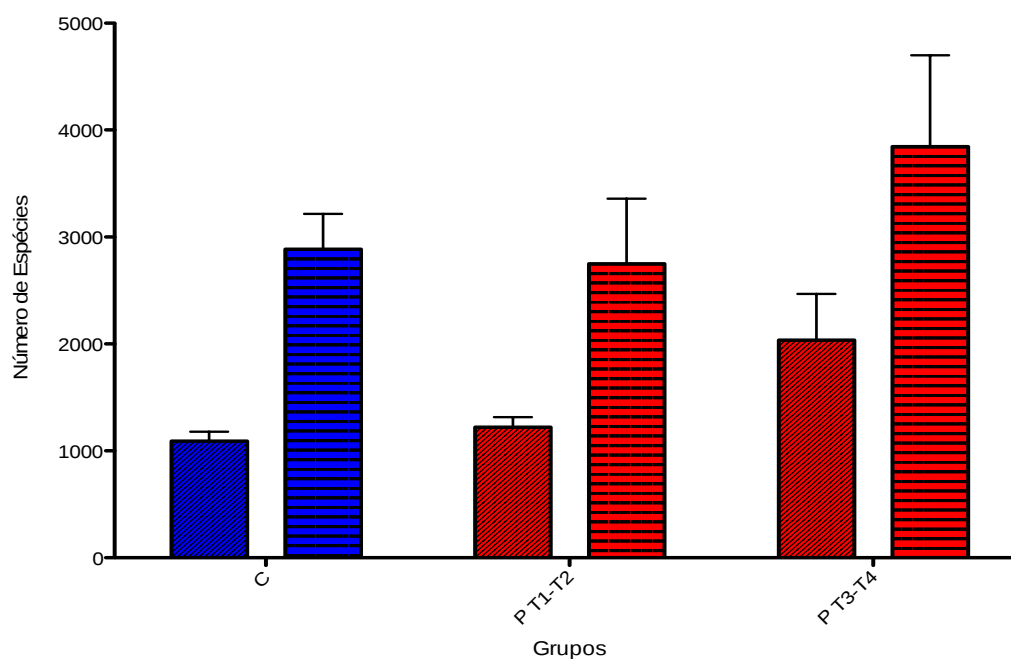
Após isto avaliamos, se o número de *OTUs* observadas nos pacientes era maior que dos controles, como ilustra a **Figura 21**.



Legenda – Em azul achurado temos o número médio de espécies observadas para o grupo dos controles, em azul com linhas temos o número médio de espécies esperadas geradas pelo programa Catchall. Em vermelho achurado temos o número médio de espécies observadas para o grupo dos pacientes, em vermelho com linhas temos o número médio de espécies esperadas. Observamos diferença estatisticamente significativa (teste T não-pareado) quando comparamos o número de espécies observadas entre os grupos P e C ($P=0.0195$). O número de espécies esperadas não se mostrou diferente entre os grupos ($P=0.879$).

Figura 21 – Número de espécies observadas para os grupos P e C

Podemos observar que o número de espécies observadas, aqui consideradas como *OTUs*, é significativamente maior no grupo dos pacientes, comparados aos controles, o que reforça os dados vistos e analisados anteriormente. Fizemos a mesma análise agora separando o grupo dos pacientes por seu estadiamento tumoral (**Figura 22**), no entanto, neste caso não observamos diferenças significativas ($P>0.05$).



Legenda – Em azul achurado temos o número médio de espécies observadas para o grupo dos controles, em azul com linhas temos o número médio de espécies esperadas geradas pelo programa Catchall. Em vermelho achurado temos o número médio de espécies observadas para o grupo dos pacientes (T1-T2 e T3-T4), em vermelho com linhas temos o número médio de espécies esperadas para o grupo dos pacientes (T1-T2 e T3-T4).

Figura 22 – Número de espécies observadas para o grupo C e os sub-grupos P T1-T2 e P T3-T4.

Podemos observar que o número de espécies observadas, aqui consideradas como *OTUs*, é maior no grupo de pacientes com tumores mais avançados, como T3 e T4, do que os pacientes com tumores menos avançados, como T1 e T2. Ainda vemos que este grupo T3-T4 possui uma maior quantidade de espécies do que o grupo dos controles, mesmo com uma menor quantidade de amostras compondo esse grupo, reforçando dos dados vistos e analisado anteriormente. Também é possível observar que os pacientes T1 e T2 possuem um maior número de espécies observadas do que o grupo de controles analisados anteriormente.

5 CONCLUSÕES

- Pacientes possui uma maior diversidade bacteriana do que ambos os grupos de controles. Foi demonstrado neste trabalho tanto por curvas de rarefação de amostras individuais e de uma análise como grupos, quanto por gráficos de frequência de *OTUs* que os pacientes possuem uma maior quantidade de *OTUs* do que ambos os grupos de controles.
- Pacientes com tumores mais avançados possuem uma maior diversidade bacteriana quando comparados a pacientes com tumores T1-T2 ou a ambos os grupos-controle. Foi demonstrado neste trabalho pelas curvas de rarefação e pelos gráficos de frequência de *OTUs* que pacientes com tumores mais avançados possuem uma maior diversidade bacteriana que os outros grupos analisados.
- Neste estudo identificamos um *core microbiome* composto por 19 *OTUs* presentes na cavidade oral de todos os indivíduos avaliados em todos os 3 grupos de estudo. Dentre estas *OTUs* houve uma prevalência de bactérias dos gêneros *Streptococcus*, *Saccharomonospora*, *Veilonella*, *Granulicatella*, *Actinomyces*, *Rothia*, *Haemophilus* e *Neisseria*. Dentre as *OTUs* que aparecem somente nos pacientes e controles que fumam e bebem em todas as amostras temos somente dois gêneros *Streptococcus* e *Actinomyces*. O gênero *Streptococcus* apareceu com 3 *OTUs* somente nestes dois grupos e o gênero *Actinomyces* apareceu com 6 *OTUs*.

- *OTUs* correspondentes a bactérias do gênero *Saccharomonospora* e ao gênero *Actinomyces* se encontraram elevadas nos pacientes, comparados aos controles, havendo um aumento gradual quando se passava dos grupos C para CFB e para P, indicando que estas poderiam estar envolvidas no processo de malignização do tecido mucoso da cavidade oral.
- O gênero *Actinomyces* se encontra mais elevado no grupo dos pacientes do que em ambos os grupos controle.

6 REFERÊNCIAS BIBLIOGRÁFICAS

Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. **J Clin Microbiol** 2005; 43:5721-32.

Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. **J Bacteriol** 2004; 186:2629-35.

Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G. Inflammation and cancer: how hot is the link? **Biochem Pharmacol** 2006; 72:1605-21.

Avşar A, Darka O, Bodrumlu EH, Bek Y. Evaluation of the relationship between passive smoking and salivary electrolytes, protein, secretory IgA, sialic acid and amylase in young children. **Arch Oral Biol** 2009; 54:457-63.

Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. **Oral Oncol** 2010; 46:414-7.

Barnich N, Carvalho FA, Glasser AL, et al. CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. **J Clin Invest** 2007; 117:1566-74.

Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, et al. The oral metagenome in health and disease. **ISME J** 2012; 6:46-56.

Bell RB, Kademani D, Homer L, Dierks EJ, Potter BE. Tongue cancer: is there a difference in survival compared with other subsites in the oral cavity?. **J Oral Maxillofac Surg** 2007; 65:229-36.

Berg JM, Tymoczko JL, Stryer L. **Biochemistry**. 5th ed. Nova Iorque: WH Freeman; 2002. Section 29.3, A Ribosome Is a Ribonucleoprotein Particle (70S) Made of a Small (30S) and a Large (50S) Subunit. Available from: <URL:<http://www.ncbi.nlm.nih.gov/books/NBK22335>> [2012 jul 12]

Blaser MJ. *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. **J Infect Dis** 1990; 161:626-33.

Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. **Cancer Res** 1988; 48:3282-7.

Boffetta P; Hashibe M. Alcohol and cancer. **Lancet Oncol** 2006; 7:149-56.

Brandwein M, Zeitlin J, Nuovo GJ, et al. HPV detection using "hot start" polymerase chain reaction in patients with oral cancer: a clinicopathological study of 64 patients. **Mod Pathol** 1994; 7:720-7.

Brooks PJ, Theruvathu JA. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. **Alcohol** 2005; 35:187-93.

Bunge J, Woodard L, Böhning D, Foster JA, Connolly S, Allen HK. Estimating population diversity with CatchAll. **Bioinformatics** 2012; 28:1045-7.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. **Bioinformatics** 2010a; 26:266-7.

Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. **Nat Methods** 2010b; 7:335-6.

Castellague X, Quintana MJ, Martinez MC et al. The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. **Int J Cancer** 2004; 108:741-9.

Castellarin M, Warren RL, Freeman JD, et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. **Genome Res** 2012; 22:299-306.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. **Science** 2009; 326:1694-7.

Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. **Database** 2010; 2010:baq013.

Clarridge JE 3rd. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on clinical microbiology and infectious disease. **Clin Microbiol Rev** 2004; 17:840-62, table of contents.

Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. **Nucleic Acids Res** 2009; 37(Database issue):D141-5.

Coleman MP, Estève J, Damiecki P, Arslan A, Renard H. Trends in cancer incidence and mortality. **IARC Sci Publ** 1993; (121):1-806.

Curado MP, Edwards B, Shin HR, et al. **Cancer incidence in five continents, Vol. IX**. Lyon: IARC; 2007. (IARC Scientific Publications, n.160).

DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. **Appl Environ Microbiol** 2006a; 72:5069-72.

DeSantis TZ Jr, Hugenholtz P, Keller K, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. **Nucleic Acids Res** 2006b; 34(Web Server issue):W394-9.

De Stefani E, Correa P, Oreggia F, et al. Black tobacco, wine and mate in oropharyngeal cancer. A case-control study from Uruguay. **Rev Epidemiol Sante Publique** 1988; 36:389-94.

de Souza RM, Lehn CN, Denardin OV. Serum and salivary immunoglobulin A levels in patients with cancer of the mouth and oropharynx. **Rev Assoc Med Bras** 2003; 49:40-4.

Driessen SR, den Hartog WC, Peters AA. Actinomycosis resembling an ovarian malignancy. **Ned Tijdschr Geneesk** 2012; 156:A4511.

Dooley CP, Cohen H, Fttzgibbons PL, et al. (1989). Prevalence of *Helicobacter pylori* infection and histologic gastritis in asymptomatic persons. **N Engl J Med** 1989; 321:1562-6.

Engelbrektson A, Kunin V, Wrighton KC, et al. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. **ISME J** 2010; 4:642-7.

Faveri M, Mayer MP, Feres M, de Figueiredo LC, Dewhirst FE, Paster BJ. Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. **Oral Microbiol Immunol** 2008; 23:112-8.

Filoche S, Wong L, Sissons CH. Oral biofilms: emerging concepts in microbial ecology. **J Dent Res** 2010; 89:8-18.

Franceschi S, Levi F, La Vecchia C, et al. Comparison of the effect of smoking and alcohol drinking between oral and pharyngeal cancer. **Int J Cancer** 1999; 83:1-4.

Franco EL. Viral etiology of cervical cancer: a critique of the evidence. **Rev Infect Dis** 1991; 13:1195-206.

Franco EL. Summary and discussion: epidemiology of HPV and anogenital neoplasms. In: Muñoz N, Bosch FX, Shah KV, Maheus A, editors. **The epidemiology of human papillomavirus and cervical cancer**. Oxford: Oxford University Press; 1992. p.283-4.

Franco EL. Cancer causes revisited: human Papillomavirus and cervical neoplasia. **J Natl Cancer Inst** 1995; 87:779-80.

Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. **Proc Natl Acad Sci U S A** 2007; 104:13780-5.

Flemmig TF, Beikler T. Control of oral biofilms. **Periodontol** 2000 2011; 55:9-15.

Fujiki H, Takeuchi H, Nishitani N, et al. Carcinogenic potential of tobacco tar-resistant *Staphylococcus aureus* in buccal cavity. **J Cancer Res Clin Oncol** 2004; 130:301-5.

Galeone C, Tavani A, Pelucchi C, et al. Coffee and tea intake and risk of head and neck cancer: pooled analysis in the international head and neck cancer epidemiology consortium. **Cancer Epidemiol Biomarkers Prev** 2010; 19:1723-36.

Gao Z, Tseng CH, Pei Z, Blaser MJ. Molecular analysis of human forearm superficial skin bacterial biota. **Proc Natl Acad Sci U S A** 2007; 104:2927-32.

Gilles A, Megléc E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. **BMC Genomics** 2011; 12:245.

Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. **Science** 2006; 312:1355-9.

Gillison ML. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. **Semin Oncol** 2004; 31:744-54.

Gutell RR, Weiser B, Woese CR, Noller HF. Comparative anatomy of 16S-like ribosomal RNA. **Prog Nucleic Acid Res Mol Biol** 1985; 32:155-216.

Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. **Genome Res** 2009; 19:1141-52.

Hansson BG, Rosenquist K, Antonsson A, et al. Strong association between infection with human papillomavirus and oral and oropharyngeal squamous cell carcinoma: a population-based case-control study in southern Sweden. **Acta Otolaryngol** 2005; 125:1337-44.

Hashibe M, Brennan P, Benhamou S, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. **J Natl Cancer Inst** 2007; 99:777-89.

Haya-Fernández MC, Bagán JV, Murillo-Cortés J, Poveda-Roda R, Calabuig C. The prevalence of oral leukoplakia in 138 patients with oral squamous cell carcinoma. **Oral Dis** 2004; 10:346-8.

Herrero R, Castellsagué X, Pawlita M, et al. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. **J Natl Cancer Inst** 2003; 95:1772-83.

Hoffmann D, Hoffmann I. The changing cigarette: chemical studies and bioassays. In: Shopland DR, editor. **Risks associated with smoking cigarettes with low machine-measured yields of tar and nicotine**. Washington, DC: National Cancer Institute; 2001. (Monograph 13), p.159-84.

Homann N, Jousimies-Somer H, Jokelainen K, Heine R, Salaspuro M. High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. **Carcinogenesis** 1997; 18:1739-43.

Homann N, Tillonen J, Meurman JH, et al. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. **Carcinogenesis** 2000; 21:663-8.

Homann N. Alcohol and upper gastrointestinal tract cancer: the role of local acetaldehyde production. **Addict Biol** 2001; 6:309-23.

Hooper SJ, Crean SJ, Wilson MJ, et al. Viable bacteria present within oral squamous cell carcinoma tissue. **J Clin Microbiol** 2006; 44:1719-25.

Hooper SJ, Wilson MJ, Crean SJ. Exploring the link between Microorganisms and Oral Cancer: a systematic review of the literature. **Head Neck** 2009; 31:1228-39.

Human Microbiome Consortium Project. Structure, function and diversity of the healthy human microbiome. **Nature** 2012; 486:207-14.

Hunt KM, Foster JA, Forney LJ, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. **PLoS ONE** 2011; 6:e21313.

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. **PLoS Genet** 2008; 4:e1000255.

Huse SM, Ye Y, Zhou Y, Fodor AA. A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. **PLoS ONE** 2012; 7:e34242.

Hyman RW, Fukushima M, Diamond L, Kumm J, Giudice LC, Davis RW. Microbes on the human vaginal epithelium. **Proc Natl Acad Sci U S A** 2005; 102:7952-7.

[IARC] International Agency for Research of Cancer. **Agents Classified by the IARC Monographs, Volumes 1–105.** Available from: <URL:<http://monographs.iarc.fr/ENG/Classification/ClassificationsAlphaOrder.pdf>> [2012 jul 12]

Jeng JH, Wang YJ, Chiang BL, et al. Roles of keratinocyte inflammation in oral cancer: regulating the prostaglandin E2, interleukin-6 and TNF-alpha production of oral epithelial cells by areca nut extract and arecoline. **Carcinogenesis** 2003; 24:1301-15.

Kaminagakura E, Villa LL, Andreoli MA, et al. High-risk human papillomavirus in oral squamous cell carcinoma of young patients. **Int J Cancer** 2012; 130:1726-32.

Keijser BJF, Zaura E, Huse SM, et al. Pyrosequencing analysis of the Oral Microflora of healthy adults. **J Dent Res** 2008; 87:1016-20.

Kong HH, Segre JA. Skin microbiome: looking back to move forward. **J Invest Dermatol** 2012; 132:933-9.

Kostic AD, Gevers D, Pedamallu CS, et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. **Genome Res** 2012; 22:292-8.

Kurkivuori J, Salaspuro V, Kaihovaara P, et al. Acetaldehyde production from ethanol by oral streptococci. **Oral Oncol** 2007; 43:181-6.

Lazarevic V, Whiteson K, Huse S, et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. **J Microbiol Methods** 2009; 79:266-71.

Lazarevic V, Whiteson K, Hernandez D, François P, Schrenzel J. Study of inter- and intra-individual variations in the salivary microbiota. **BMC Genomics** 2010; 11:523.

Lewin F, Norell SE, Johansson H, et al. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: a population-based case-referent study in Sweden. **Cancer** 1998; 82:1367-75.

Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. **Nucleic Acids** 2007; 35:e120.

Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. **Nat Biotechnol** 2012; 30:562.

Macfarlane GT, Macfarlane S. Bacteria, colonic fermentation, and gastrointestinal health. **J AOAC Int** 2012; 95:50-60.

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.

Maloney KN, Macmillan JB, Kauffman CA, et al. Lodopyridone, a structurally unprecedented alkaloid from a marine actinomycete. **Org Lett** 2009; 11:5422-4.

Mans JJ, von Lackum K, Dorsey C, et al. The degree of microbiome complexity influences the epithelial response to infection. **BMC Genomics** 2009; 10:380.

Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. **Lancet** 1984; 1:1311-5.

Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis and treatment. **Mayo Clin Proc** 2008; 83:489-501.

Mashberg A, Merletti F, Boffetta P, et al. Appearance, site of occurrence, and physical and clinical characteristics of oral carcinoma in Torino, Italy. **Cancer** 1989; 63:2522-7.

Meurman JH. Oral microbiota and cancer. **J Oral Microbiol** 2010; 2.

Ministério da Saúde. Instituto Nacional do Cancer. **Estimativa/2012 Incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2011.

Morita E, Narikiyo M, Yokoyama A, et al. Predominant presence of *Streptococcus anginosus* in the saliva of alcoholics. **Oral Microbiol Immunol** 2005; 20:362-5.

Muto M, Hitomi Y, Ohtsu A, et al. Acetaldehyde production by non-pathogenic *Neisseria* in human oral microflora: implications for carcinogenesis in upper aerodigestive tract. **Int J Cancer** 2000; 88:342-50.

Nagy KN, Sonkodi I, Szoke I, Nagy E, Newman HN. The microflora associated with human oral carcinomas. **Oral Oncology** 1998; 34:304-8.

Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. **Genome Res** 2009a; 19:636-43.

Nasidze I, Quinque D, Li J, Li M, Tang K, Stoneking M. Comparative analysis of human saliva microbiome diversity by barcoded pyrosequencing and cloning approaches. **Anal Biochem** 2009b; 391:64-8.

Olsen J, Sabreo S, Fasting U. Interaction of alcohol and tobacco as risk factors in cancer of the laryngeal region. **J Epidemiol Community Health** 1985; 39:165-8.

Oreggia F, De Stefani E, Correa P, Fierro L. Risk factors for cancer of the tongue in Uruguay. **Cancer** 1991; 67:180-3.

Paster BJ, Boches SK, Galvin JL, et al. Bacterial diversity in human sub- gingival plaque. **J Bacteriol** 2001; 183:3770-83.

Paster BJ, Olsen I, Aas JA, Dewhirst FE. The breadth of bacterial diversity in the human periodontal pocket and other oral sites. **Periodontol** 2006; 42:80-7.

Pei Z, Bini EJ, Yang L, Zhou M, Francois F, Blaser MJ. Bacterial biota in the human distal esophagus. **Proc Natl Acad Sci USA** 2004; 101:4250-5.

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J. Metagenomic pyrosequencing and microbial identification. **Clin Chem** 2009; 55:856-66.

Pöschl G, Seitz HK. Alcohol and cancer. **Alcohol Alcohol** 2004; 39:155-65.

Pushalkar S, Mane SP, Ji X, et al. Microbial Diversity in saliva of oral squamous cell carcinoma. **FEMSI Immunol Med Microbiol** 2011; 61:269-77.

Ragin CCR, Modugno F, Gollin SM. The epidemiology and risk factors of head and neck cancer a focus on human papillomavirus. **J Dent Res** 2007; 86:104-14.

Riesenfeld CS, Schloss PD, Handelsman J. METAGENOMICS: Genomic Analysis of Microbial Communities. **Ann Rev Genet** 2004; 38:525-52.

Rondon MR, August PR, Bettermann AD, et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. **Appl Environ Microbiol** 2000; 66:2541-7.

Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. **Nature** 2011; 475:348-52.

Salaspuro MP. Acetaldehyde, microbes, and cancer of the digestive tract. **Crit Rev Clin Lab Sci** 2003; 40:183-208.

Salaspuro V, Salaspuro M. Synergistic effect of alcohol drinking and smoking on in vivo acetaldehyde concentration in saliva. **Int J Cancer** 2004; 111:480-3.

Salaspuro M. Acetaldehyde as a common denominator and cumulative carcinogen in digestive tract cancers. **Scand J Gastroenterol** 2009; 44:912-25.

Sanguinetti CJ, Dias Neto E, Simpson AJ. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. **Biotechniques** 1994; 17:914-21.

Savage DC. Microbial ecology of the gastrointestinal tract. **Annu Rev Microbiol** 1977; 31:107-33.

Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. **Genome Biol** 2012; 13:R42.

Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJ. Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease 2000. **BMC Cancer** 2002; 2:37.

Shpitzer T, Bahar G, Feinmesser R, Nagler RM. A comprehensive salivary analysis for oral cancer diagnosis. **J Cancer Res Clin Oncol** 2007; 133:613-7.

Sundquist A, Bigdeli S, Jalili R, et al. Bacterial flora-typing with targeted, chip-based Pyrosequencing. **BMC Microbiol** 2007; 7:108.

Soerjomataram I, deVries E, Pukkala E, Coebergh JW. Excess of cancers in Europe: a study of eleven major cancers amenable to lifestyle change. **Int J Cancer** 2007; 120:1336-43.

Syrjänen KJ, Pyrhönen S, Syrjänen SM, Lamberg MA. Immunohistochemical demonstration of human papilloma virus (HPV) antigens in oral squamous cell lesions. **Br J Oral Surg** 1983; 21:147-53.

Schweigert M, Dubecz A, Beron M, Ofner D, Stein HJ. Pulmonary infections imitating lung cancer: clinical presentation and therapeutical approach. **Ir J Med Sci** 2012 May 17. [Epub ahead of print].

Talamini R, Bosetti C, La Vecchia C, et al. Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. **Cancer Causes Control** 2002; 13:957-64.

Tezal M, Scannapieco FA, Wactawski-Wende J, et al. Local inflammation and human papillomavirus status of head and neck cancers. **Arch Otolaryngol Head Neck Surg** 2012; 138:669-75.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. **Nature** 2006; 444; 1027-31.

Van der Peer Y, Chapelle S, De Wachter R. A quantitative map of nucleotide substitution rates in bacterial rRNA. **Nucl Acids Res** 1996; 24:3381-91.

Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. **Oral Oncol** 2009; 45:309-16.

Woese CR. Archaeobacteria and cellular origins: an overview [abstract]. **Zentralbl Bakteriell Parasitenkd Infektionskr Hyg** 1982; 3:1-17.

Woese CR. Bacterial evolution. **Microbiol Rev** 1987; 51:221-71.

Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F, Zhou HW. Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. **BMC Microbiol** 2010; 10:255.

Wunsch-Fiho V, de Camargo EA. The burden of mouth cancer in Latin America and the Caribbean: epidemiologic issues. **Semin Oncol** 2001; 28:158-68.

Wunsch-Fiho V. The epidemiology of oral and pharynx cancer in Brazil. **Oral Oncol** 2002; 38:737-46.

Wynder EL, Bross IJ. Aetiological factors in mouth cancer: an approach to its prevention. **Br Med J** 1957; 1:1137-43.

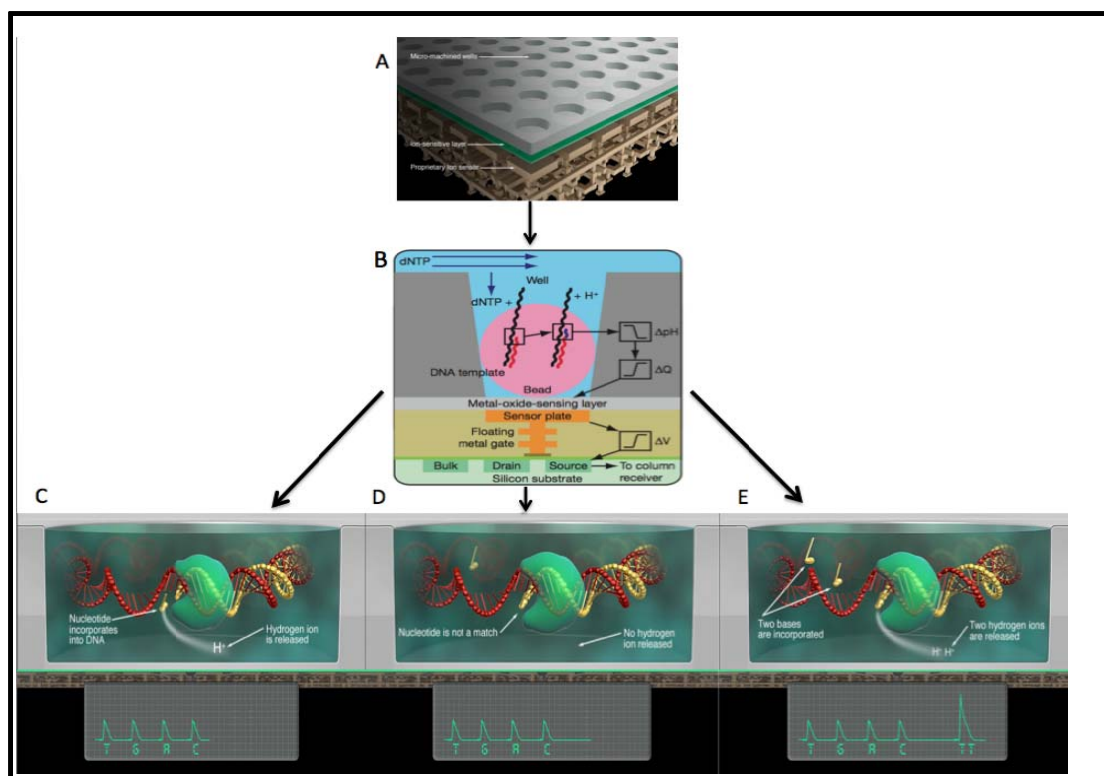
Yang F, Zeng X, Ning K, et al. Saliva microbiomes distinguish caries-active from healthy human populations. **ISME J** 2012; 6:1-10.

Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. **Microbiology** 2004; 150:2565-73.

Zivkovic AM, German JB, Lebrilla CB, Mills DA. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. **Proc Natl Acad Sci USA** 2012; 108(Suppl 1):4653-8.

Anexo 1 - A plataforma *Ion Torrent™* da *Applied Biosystems*

A química de sequenciamento do *Ion Torrent™* (*Applied Biosystems*) é diferente das demais plataformas de sequenciamento de nova geração criada até agora. Esta plataforma baseia-se em semicondutores que detectam uma diferença de potencial elétrico causada pela mudança de pH quando um nucleotídeo é incorporado à fita de DNA nascente pela ação de uma DNA polimerase. A liberação de um único átomo de hidrogênio não é suficiente para alterar drasticamente o pH do meio, porém, como uma única fita de DNA está anelada a apenas uma *bead* e ela é amplificada mais ou menos um milhão de vezes através de uma PCR em emulsão (ePCR), há uma drástica mudança de pH no meio devido à grande quantidade de átomos de hidrogênio sendo liberados simultaneamente. Assim, uma corrente elétrica passa pelos poços aonde se encontram as *beads* e quando há uma mudança de pH gerada pela incorporação de uma base, este potencial muda, indicando a incorporação de uma base. Deste modo, cada um dos nucleotídeos é adicionado de uma vez, um a um, o que permite associar a mudança de potencial elétrico com a incorporação da respectiva base (como ilustrado na **Figura A1**). Atualmente, esta plataforma é capaz de gerar sequências de aproximadamente 200 nt, porém com alterações no seu número de ciclos, este número pode ser aumentado. Com o chip 314 (o primeiro a ser lançado), é possível obter 100,000 sequências de 100 nt, ou seja, 10MB (Mega Bases). Com o chip 316 (o segundo a ser lançado), é possível obter 10 vezes mais sequências que o 314 (100MB), enquanto que o chip 318 gera ao redor de 1GB de dados em uma única corrida.



A figura mostra os milhões de *wells* no chip do *Ion torrent* com as camadas sensíveis à uma mudança de potencial da corrente elétrica (A); um desenho esquemático de um *well* do chip do *Ion torrent* com um *bead* e a adição de um nucleotídeo alterando o pH do meio e consequentemente o potencial da corrente elétrica, podendo então ser detectado pelo sensor (ROTHBERG et al. 2011) (B); a adição de um nucleotídeo complementar à fita de DNA pela DNA polimerase causando a liberação de um hidrogênio, mudando o pH e consequentemente o nucleotídeo é identificado (C); quando o nucleotídeo não é complementar à fita, portanto não há uma mudança no pH e não causa uma mudança de potencial da corrente elétrica (D); dois nucleotídeos são incorporados à fita de DNA complementar, causando uma liberação dobrada de hidrogênio e uma mudança do potencial elétrico duas vezes maior (E).

Figura A1.1 – Química de sequenciamento da plataforma Ion Torrent™, Applied Biosystems

Anexo 2 - Sequenciamento de 10 controles na plataforma 454-Roche

Escolha dos Iniciadores

Primeiramente foi feita uma seleção de iniciadores que permitiriam: i) a amplificação do maior número possível de táxons e ii) a amplificação de segmentos polimórficos do gene 16S-rDNA que permitam diferenciar os táxons amplificados. A fim de identificar os iniciadores mais adequados, foram avaliados aqueles que permitissem atingir os objetivos i e ii acima descritos. Foi feita uma busca exaustiva na literatura e em bancos de dados (tais como <http://bioinfo.unice.fr/454/>), para identificar a grande maioria dos iniciadores já descritos cujo alvo é a amplificação do gene 16S-rDNA de Eubactérias. Para todos os iniciadores foi feita uma análise da sequência original, e de seu respectivo reverso-complementar. A partir deste *set*, foi desenvolvido um *pipeline* de bioinformática onde estes iniciadores foram avaliados contra os bancos *Human Oral Microbiome Database* (HOMD) e *Ribosomal Database Project* (RDP). Foi anotado quantos *hits* cada um destes iniciadores era capaz de gerar em cada um destes bancos, e aqueles que eram encontrados em percentuais muito baixos eram descartados.

Tendo definido os melhores iniciadores individuais, fizemos um estudo das combinações que permitissem amplificar a maioria dos táxons gerando a maior frequência relativa de *amplicons* únicos (táxon-específicos, gerados pelas combinações dos melhores iniciadores individuais). Deste modo, o melhor par de iniciadores seria aquele que permitisse amplificar a maior parcela dos diferentes táxons, gerando a maior quantidade de *amplicons* únicos.

As 20 melhores combinações de iniciadores de cada banco foram colocadas em uma tabela e seu ranking foi somado junto com o ranking que esta mesma combinação possuía no banco oposto. Por exemplo, um dos melhores pares de iniciadores para a amplificação dos táxons do banco RDP possui *ranking* 3 para este banco, mas aparece como o 10º melhor par no banco do HOMD, deste modo a sua soma de *rankings* é equivalente a 13. As combinações com a menor somatória de *rankings* foram escolhidas como os iniciadores que seriam usados.

- ***Fusion Primers***

Com o uso da plataforma de sequenciamento 454, que atualmente gera cerca de 400-500 nucleotídeos úteis, não é possível sequenciar os *amplicons* gerados em sua totalidade. Deste modo, para a construção dos *fusion-primers* adequados à amplificação da região de interesse seguida do sequenciamento no sistema 454, foi determinado qual extremidade do *amplicon* seria mais informativa no sequenciamento. Deste modo, uma análise *in silico* foi feita, avaliando a informatividade dos 400-500 nt mais próximos de cada um dos iniciadores -F ou R – selecionados na etapa anterior. Nestes fragmentos foram investigadas quantas sequências únicas poderiam ser geradas em cada banco de dados.

Fusion Primer Reverse

(Iniciador de Sequenciamento B 454 + 16s rDNA Forward):

E343F- CCTATCCCCTGTGTGCCTTGGCAGTCTCAG TACGGRAGGCAGCAG

Fusion Primers Forward:

(Iniciador de Sequenciamento A 454 + Tag + Iniciador reverse 16s rDNA):

TAG 01- CCATCTCATCCCTGCGTGTCTCCGACTCAG ACACGT CRRACGAGCTGICGAC
TAG 03- CCATCTCATCCCTGCGTGTCTCCGACTCAG AGAGAG CRRACGAGCTGICGAC
TAG 05- CCATCTCATCCCTGCGTGTCTCCGACTCAG AGCATG CRRACGAGCTGICGAC
TAG 10- CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGCAG CRRACGAGCTGICGAC
TAG 13- CCATCTCATCCCTGCGTGTCTCCGACTCAG CAGATG CRRACGAGCTGICGAC
TAG 14- CCATCTCATCCCTGCGTGTCTCCGACTCAG CAGCAG CRRACGAGCTGICGAC
TAG 16- CCATCTCATCCCTGCGTGTCTCCGACTCAG CATCTG CRRACGAGCTGICGAC
TAG 17- CCATCTCATCCCTGCGTGTCTCCGACTCAG CATGAG CRRACGAGCTGICGAC
TAG 18- CCATCTCATCCCTGCGTGTCTCCGACTCAG CTCATG CRRACGAGCTGICGAC
TAG 20- CCATCTCATCCCTGCGTGTCTCCGACTCAG CTGCTG CRRACGAGCTGICGAC

Sequenciamento no 454

PCR

Foram realizadas 6 reações de 10µl para cada amostra, contendo 3µl de DNA eluído, 5µl de *Kapa HiFidelity Master Mix* (2x) e 1µl de cada iniciador *fusion* (2pmol/µl). As condições da termociclagem foram 95°C- 3 minutos; 98°C- 20 segundos, 60°C- 30 segundos, 72°C- 1 minuto (40 vezes); 72°C- 5 minutos. Os produtos foram analisados em um gel de poliacrilamida 8% corado com prata.

Ligação dos Adaptadores, emPCR e Sequenciamento

Como os adaptadores já estavam ligados devido aos *fusion primers*, as amostras foram quantificadas utilizando o fluorímetro *Qubit*, e seus perfis de tamanho foram analisados usando um chip *high sensitivity* de DNA no *Bioanalyzer*. Foi usado uma cópia de *amplicon* por *bead* de sequenciamento na emPCR. A PCR em emulsão foi quebrada e as *beads* recuperadas e enriquecidas para usar somente as que continham os adaptadores A e B de sequenciamento. Foram depositadas 1,8 milhões de *beads* na placa de sequenciamento.

Análise Taxonômica

A análise das sequências para a identificação dos *barcodes* assim como a análise taxonômica em si foram feitas utilizando o software *Qiime* (CAPORASO et al. 2010b).

Os iniciadores selecionados para a plataforma 454 são apresentados a seguir, na **Tabela A2.1**:

Pares de iniciadores	Forward	Reverse	Em relação ao banco RDP (Total de 1.182.620 sequências)				Em relação ao banco HOMD (Total de 753 sequências)			
			Match primers (%)	Av. amplic size (pb)	Ident. taxon	Rel. Total (%)	Match primers (%)	Av. amplic size (pb)	Ident. taxon	Rel. Total (%)
E341F-Unice NYR	CCTACGGG[N]GGC[N]GCA	GTCTG[N]CAGCTCGTG[YY]G	552404 (46,71)	693,72	346719	29,32	712 (94,56)	695,22	685	90,97
E341F-Unice YR	CCTACGGG[N]GGC[N]GCA	GTCTGTCAGCTCGTG[YY]G	550980 (46,59)	693,73	345852	29,24	692 (91,90)	695,66	665	88,31
343F-Unice NYR	TACGG[R]AGGCAGCAG	GTCTG[N]CAGCTCGTG[YY]G	541841 (45,82)	692,63	338187	28,60	730 (96,95)	694,08	703	93,36
343F-Unice YR	TACGG[R]AGGCAGCAG	GTCTGTCAGCTCGTG[YY]G	540368 (45,69)	692,64	337295	28,52	707 (93,89)	694,57	680	90,31
SV3F-Unice NYR	ACTCCTACGGGAGGCAGCAG	GTCTG[N]CAGCTCGTG[YY]G	531143 (44,91)	692,65	330303	27,93	709 (94,16)	694,21	682	90,57
SV3F-Unice YR	ACTCCTACGGGAGGCAGCAG	GTCTGTCAGCTCGTG[YY]G	529764 (44,80)	692,66	329473	27,86	689 (91,50)	694,65	662	87,92

Tabela A2.1: Combinações dos melhores iniciadores nos dois bancos. **Y:** C ou T; **R:** A ou G; **N:** A, C, T ou G. **Match primers:** Indica o total de sequências amplificáveis e seu percentual em relação ao total do respectivo banco; **Av. amplic size:** indica o tamanho médio dos *amplicons* produzidos com cada par de iniciadores em cada um dos bancos; **Ident. taxon:** Indica o total de táxons com *amplicons* de sequências distintas, portanto identificáveis com os iniciadores usados; **Rel. total:** Indica o percentual de táxons identificáveis em relação ao banco total usado.

Os resultados obtidos a partir do HOMD ou do RDP são mostrados a seguir, nas Tabelas **A2.2** e **A2.3**:

Tabela A2.2 – Potencial de identificação de táxons no banco HOMD, por cada um dos iniciadores selecionados.

--> Forward 500pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	753	730 (96,95%)	677	89,91%
SundquistV3F-Unice.Fr NYY	753	709 (94,16%)	656	87,12%
<-- Reverse 500pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	753	730 (96,95%)	661	87,78%
SundquistV3F_Unice. Fr NYY	753	709 (94,16%)	640	84,99%
--> Forward 400pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	753	730 (96,95%)	661	87,78%
SundquistV3F_Unice. Fr NYY	753	709 (94,16%)	642	85,26%
<-- Reverse 400pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	753	730 (96,95%)	626	83,13%
SundquistV3F_Unice. Fr NYY	753	709 (94,16%)	607	80,61%

Tabela A2.3 – Potencial de identificação de táxons no banco RDP, por cada um dos iniciadores selecionados.

--> Forward 500pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	1182620	541841 (45,82%)	274412	23,20%
SundquistV3F-Unice.Fr NYY	1182620	531143 (44,91%)	267548	22,62%
<-- Reverse 500pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	1182620	541841 (45,82%)	279944	23,67%
SundquistV3F_Unice. Fr NYY	1182620	531143 (44,91%)	273252	23,11%
--> Forward 400pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	1182620	541841 (45,82%)	229821	19,43%
SundquistV3F_Unice. Fr NYY	1182620	531143 (44,91%)	223627	18,91%
<-- Reverse 400pb				
Iniciadores	Total sequences	Primer Match	Identified	(Rel. Total)
343F_Unice.Fr NYY	1182620	541841 (45,82%)	235466	19,91%
SundquistV3F_Unice. Fr NYY	1182620	531143 (44,91%)	229555	19,41%

Características das sequências dos controles geradas pela plataforma Roche-454:

Input de Sequências: 387.577

Filtros	Números	Sequências Excluídas	Porcentagem
Máximo de bases Ambíguas	>2	889	0,2%
Tamanho Máximo	1000	279.628	72,1%
Tamanho Mínimo	80		
Erros na <i>Tag</i>	>2	216	0,06%
Erros no Iniciador (17nt)	>11	1.913	0,5%
Total		282.646	72,9%

Amostra	Tag	Número de Sequências	Tamanho das sequências antes dos filtros (nt)	
C15	CAGCAG	20.506	Mínimo	40
C23	CTGCTG	19.448	Máximo	1.196
C09	ATGCAG	14.068	Média	176
C20	CATGAG	11.498	Tamanho das sequências depois dos filtros (nt)	
C13	CAGATG	9.676	Mínimo	80
C06	AGCATG	8.516	Máximo	953
C19	CATCTG	9.116	Média	511
C04	AGAGAG	8.516	Sequências/Amostra	
C01	ACACGT	4.605		
C21	CTCATG	1.358		
Total		107.307		
Percentage m		27,7%	Mínimo	1.358
			Máximo	20.506
			Média	10.479

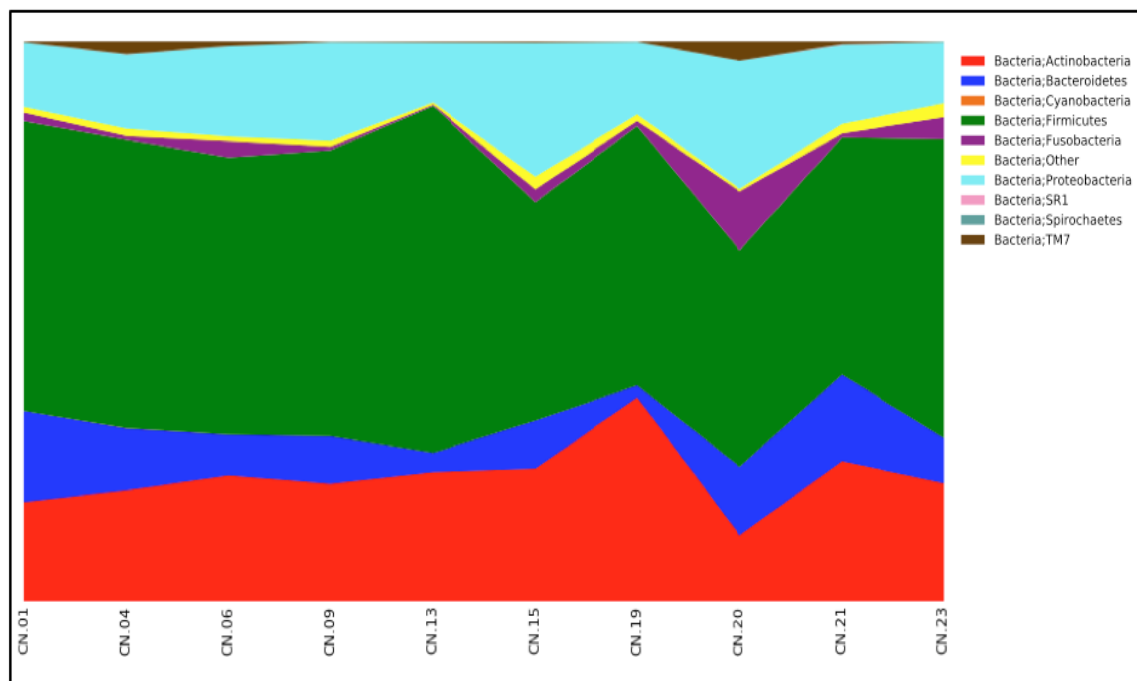


Figura A2.1 – Gráfico de área mostrando os filos bacterianos mais abundantes nos controles (C) sequenciados na plataforma 454-Roche

Anexo 3 - Termo de Consentimento Livre e Esclarecido



Centro de Tratamento e Pesquisa Hospital do Câncer A. C. Camargo

Fundação Antonio Prudente

Departamento de Cirurgia de Cabeça e Pescoço e Otorrinolaringologia

Rua Professor Antônio Prudente, 211 - 01509-900 - São Paulo - SP

Fone: 2189-5172

Termo de Consentimento Livre e Esclarecido

DADOS DE IDENTIFICAÇÃO DO PACIENTE OU RESPONSÁVEL LEGAL

Nome do Paciente: _____

Documento de identidade nº: _____ Sexo: M/F

Nascimento: __/__/____

Endereço: _____ nº _____ Compl. _____ CEP _____

Cidade: _____ Estado: _____ Telefone: _____

DADOS SOBRE O ESTUDO

O presente estudo, intitulado “Metagenômica da Cavidade Oral”, um projeto de pesquisas do Centro de Pesquisas do Hospital AC Camargo e do Departamento de Cirurgia de Cabeça e Pescoço e Otorrinolaringologia do mesmo Hospital.

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Tratamento e Pesquisa Hospital do Câncer A. C. Camargo, número do Protocolo **1457/10**, em **29/09/2010** e teve início em **04/10/2010**

EXPLICAÇÕES DO PESQUISADOR AO PACIENTE OU SEU REPRESENTANTE LEGAL

Justificativas da pesquisa

Mais de 70% dos casos de câncer de boca podem ser evitados com uma mudança no estilo de vida, como a diminuição do contato com os fatores de risco mais importantes para esta doença: o uso de tabaco e de bebidas alcoólicas. Evidências sugerem que o uso de tabaco e álcool causa importantes modificações na boca e com isto aumenta o risco de desenvolver câncer. É possível que o consumo destas substâncias leve a um desequilíbrio nas bactérias comumente encontradas na boca. Este estudo tentará esclarecer a relação entre o consumo de bebidas alcoólicas e tabaco, a população de bactérias encontradas na boca e a evolução deste tipo de câncer.

Objetivos da pesquisa

Avaliar as bactérias presentes na boca e a sua variação na presença ou ausência de câncer de boca em indivíduos que fazem ou não o uso de tabaco e de álcool.

Procedimentos que serão utilizados e propósitos

Este é um estudo sobre a relação dos fatores de risco para lesões de câncer de boca e as espécies de bactéria encontradas na cavidade oral.

Nesse estudo iremos obter informações sobre as lesões e sobre hábitos comuns, através de questionários aplicados a 20 pacientes com lesões de câncer de boca e faringe e 40 indivíduos que não possuem lesões de câncer de boca e faringe que serão selecionados através de seus hábitos comuns e que se apresentarem no Departamento de Cirurgia de Cabeça e Pescoço e Otorrinolaringologia do Centro de Tratamento de Pesquisa do Hospital do Câncer AC Camargo. Será feita uma coleta de células presentes na mucosa oral através de um *swab* bucal.

Para participar do estudo o paciente deverá:

- Ter idade maior que 40 anos;

Não poderão participar pacientes portadores de doenças mentais e aqueles com dificuldades de entendimento dos procedimentos necessários.

O tempo previsto de duração da entrevista é de 50 minutos. Após a entrevista, um indivíduo treinado fará a coleta das células encontradas na mucosa, usando uma pequena escova a qual será esfregada na face interna nas bochechas, língua, soalho bucal, gengivas e demais

mucosas da cavidade oral. Esse procedimento tem duração de menos de 5 minutos e não há necessidade de uso de qualquer tipo de anestesia. Nenhum desconforto é esperado para este procedimento. Se desconfortos forem percebidos o indivíduo participante pode exigir a interrupção do procedimento.

Os pacientes e seus familiares serão informados das razões da pesquisa, bem como dos eventuais benefícios que possam ser obtidos com o seu resultado. Será explicado detalhadamente o objetivo do estudo e o conteúdo e duração da entrevista. Será esclarecida a absoluta liberdade do paciente de participar ou não do estudo sem que isto acarrete qualquer comprometimento ou alteração em seu tratamento. Também será assegurado ao participante um completo sigilo das informações obtidas. Em caso de concordância, essa será uma colaboração voluntária, não havendo nenhuma forma de remuneração. O paciente pode desistir de participar, mesmo após a assinatura deste consentimento, sem que isso ocasione qualquer prejuízo ao tratamento do mesmo.. Após concordância de participar, este Termo de Consentimento deverá ser assinado em 2 vias, sendo que uma via será entregue ao sujeito da pesquisa e a outra permanecerá com o pesquisador. Os pacientes serão entrevistados em uma sala particular por um profissional previamente treinado.

Complicações e riscos esperados

Não existe nenhum risco ou complicação associado com a participação neste estudo. É pouco provável que a coleta de células presentes na mucosa oral cause qualquer desconforto. Porém, se isto ocorrer, deve ser bastante leve e totalmente suportável. Todas as perguntas e o método de coleta de células que será utilizada neste estudo foram previamente empregados em outros estudos. Todos os dados levantados serão usados apenas para pesquisa científica, visando a melhor compreensão do câncer de boca. Todos os dados serão mantidos confidenciais.

Benefícios que poderão ser obtidos

Os pacientes não se beneficiaram diretamente com a participação nesse estudo. Contudo, o estudo contribuirá para um melhor entendimento dos fatores que podem causar aparecimentos e evolução de lesões orais e faríngeas, desta forma ajudando o desenvolvimento de ações de prevenção.

6. Por favor, responda sim ou não às questões a seguir:

Foi garantido ao participante o direito de receber resposta a qualquer pergunta ou esclarecimento acerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa e o tratamento?

SIM NÃO

Ficou esclarecida ao participante a liberdade de retirar seu consentimento a qualquer momento, sem que isso traga prejuízo quanto à continuidade do tratamento no Hospital do Câncer A. C. Camargo?

SIM NÃO

Ficou esclarecido que não haverá ressarcimento ou qualquer tipo de pagamento referente à realização deste estudo?

SIM NÃO

Ficou esclarecido que não haverá indenização além das previstas pela lei, em reparação a dano imediato ou tardio, causado pela pesquisa em questão?

SIM NÃO

Ficou esclarecido a segurança de que a identidade do participante será preservada, mantendo-se todas as informações em caráter confidencial, sob guarda do pesquisador principal, e que os dados serão utilizados somente para análise estatística e posteriormente para publicações científicas, sem qualquer informação que possa identificar o participante?

SIM NÃO

CONSENTIMENTO PÓS-INFORMADO

Declaro que, após ter sido convenientemente esclarecido dos riscos e benefícios deste estudo, conforme definido nos itens 1 a 9, do inciso III, consinto em participar, na qualidade de paciente, do Projeto de Pesquisa referido no inciso II.

São Paulo, de de 20__

Assinatura do paciente ou responsável legal

Assinatura do pesquisador

Em casos de dúvidas, contatar Dr. Luiz P. Kowalski no Departamento de Cirurgia de Cabeça e Pescoço e Otorrinolaringologia do Hospital do Câncer A C Camargo pelo telefone 2189-5172. Se o pesquisador principal não fornecer as informações e/ou esclarecimentos suficientes, por favor, entre em contato com o Coordenador do Comitê de Ética do Hospital A.C. Camargo – SP, pelo Telefone 2189-5020.

Anexo 4- Questionário aplicado aos indivíduos deste estudo

QUESTIONÁRIO INICIAL

Projeto **Metagenômica da Cavidade Oral**

Número de Identificação.....

Nota: o número de identificação é o mesmo do registro geral hospitalar (RGH)

Nome do Participante

Endereço

.....

.....

.....

.....

.....

Telefone

Número de identificação.....|_|_|-|_|_|-|_|_|_|_|

INFORMAÇÕES GERAIS

B2 Sexo: (1) Masculino; (2) Feminino|_|

B3 Grupo étnico (deve ser informado pelo entrevistador):
(01) Branco; (02) Mulato; (03) Negro; (04) Asiático; (5) Indígena|_|_|

B4 Qual é a sua idade?|_|_|

B5 Qual é sua data de nascimento?.....|_|_|-|_|_|-|_|_|
Dia Mês Ano

B6 Em qual cidade/distrito você mora? [CL]|_|_|_|_|_|_|_|
Especificar -

B7 Há quanto tempo você mora nessa cidade?|_|_|
(Se for menos de 1 ano, pôr 00)

B8 Em que cidade/distrito você nasceu? [CL]|_|_|_|_|_|_|_|
Especificar -

B9 Qual é a sua religião?
(01) Católica (04) Judaica (07) Hinduísta
(02) Muçulmana (05) Budista (08) Nenhuma
(03) Protestante (06) Animista/Voodooista 09) Outra|_|_|

B10 Quando foi à última vez que você fez uso de antibióticos?|_|_|

B11 Se foi recentemente (últimos 3 meses), quais? E por quanto tempo?

Especificar-

RGH

EDUCAÇÃO E OCUPAÇÃO

C1 *Você foi à escola?* (1) Sim; (2) Não | |
[Se “não”, ir para C4)]

C2 *Por quantos anos você foi à escola?*| | |

C3 *Com que idade parou de estudar?*| | |

C4 *Qual é/foi seu emprego de maior duração?* [CL]...../ / /

Especifique -

RGH

USO DE FUMO DO TABACO

D1 *Você fuma ou já fumou diariamente por pelo menos 1 ano?*|_|

(1) Sim, e continua; (2) Nunca; (3) Sim, no passado [se “nunca” ir para D6]

Por favor, descreva os períodos da sua vida durante os quais você fumou cigarro, charuto, cachimbo ou cigarro de palha (bidi), a quantidade fumada e outros detalhes sobre o produto usado. Tente resumir as mudanças mais importantes da sua vida quanto à quantidade e tipo de produto. Ignore as mudanças que aconteceram por menos de 1 ano.

Entrevistador: Evite sobreposição de anos para o mesmo produto ou tipo de cigarro i.e. escreva 30-40, 41-45 ao invés de 30-40, 40-45.

D2 *Você fuma ou já fumou cigarro?* [Se “não” ir para D3]

	De idade?	Até idade?	Tabaco tipo	Filtro (b) (c)	Marca	Número por dia
(a)	_	_	_	_	_	_
_	_	_	_	_	_	_
	_	_	_	_	_	_
	_	_	_	_	_	_
	_	_	_	_	_	_

D3 *Você fuma ou já fumou charuto?* [Se “não” ir para D4]

	De idade?	Até idade?	Marca	Número por dia
(a)	_	_	_	_
_	_	_	_	_
	_	_	_	_

D4 *Você fuma ou já fumou cachimbo?* [Se “não” ir para D5]

	De idade?	Até idade?	Marca	Número por dia
(a)	_	_	_	_
_	_	_	_	_
	_	_	_	_

(a) (1) Sim; (2) Não

(b) (1) Claro; (2) Escuro; (9) Não sabe

(c) (1) Filtro; (2) Sem filtro; (9) Não sabe

D5 *Você fuma ou já fumou cigarro de palha (bidi)?* Se “não” ir para D6]

(a) (1) Sim; (2) Não |_|

Cigarro palha	(a)	De idade?	Até idade?	Marca	Número de por dia
	_	_	_	_	_
	_	_	_	_	_

D6 Você já mascou diariamente tabaco, *betel quid*, *noz de areca*, *pan massala* por pelo menos 1 ano?||

(1) Sim e continua; (2) Nunca; (3) Sim, mas parou desde a última visita [se “nunca” ir para D8]

Por favor, descreva os períodos da sua vida durante os quais vc mascou, a quantidade mascarada e outros detalhes sobre o produto usado. Tente resumir as mudanças mais importantes da sua vida quanto a quantidade e tipo de produto. Ignore as mudanças que aconteceram por menos de 1 ano.

Produto: (1) Tabaco (4) Noz de Areca com tabaco
 (2) Betel quid com tabaco (5) Noz de Areca sem tabaco
 (3) Betel quid sem tabacco (6) Pan massala
 (7) Outro (especifique) _____

D7 Produto	Da idade	até idade?	Quantas vezes por dia?

D8 Você já cheirou tabaco diariamente por pelo menos 1 ano?||

(1) Sim, e continua; (2) nunca; (3) No passado [se “nunca” ir para próxima página]

Por favor, descreva os períodos da sua vida durante os quais você cheirou, a quantidade cheirada e outros detalhes sobre o produto usado. Tente resumir as mudanças mais importantes da sua vida quanto a quantidade e tipo de produto. Ignore as mudanças que aconteceram por menos de 1 ano.

Da idade?	Até idade?	Quantas vezes por dia?

RGH

HÁBITOS DIETÉTICOS

Com que frequência você consumiu os seguintes alimentos e bebidas?

Unidade **Item** *Quantas vezes por semana?*
(se menos de 1x/semana, escrever 98; se nunca por 0)

E1	1 copo	Leite	
E2	1 pote	iogurte	
E3	1 porção	Pão	
E4	1 porção	Macarrão ou arroz	
E5	1 porção	Prato à base de milho	
E6	1 porção	Carne	
E7	1 porção	Peixe	
E8	1 porção	Presunto, salame, linguiça	
E9	1	Ovo	
E10	1 porção	Queijo	
E11	1 (média)	Batata	
E12	1 porção	Vegetais verdes crus e salada	
E13	1 porção	Crucíferos (brócoli, repolho, couve flor)	
E14	1 (média)	Cenoura	
E15	1 (média)	Tomates frescos	
E16	1 porção	Leguminosas (ervilha, feijão, lentilha, etc)	

E17 1 porção *Resumindo, com que frequência você come qualquer Tipo de vegetais (excluindo batatas)?*

E18	1 copo	Suco de frutas frescas	
E19	1 (média)	Maçã ou pera	
E20	1 (média)	Frutas cítricas (laranja, limão)	
E21	1 (média)	Bananas	
E22	1 (média)	<i>Resumindo, com que frequência você come qualquer tipo de frutas (incluindo salada de frutas)</i>	
	1 pedaço ou Porção	Bolo ou doces	

E23 *Que tipo de gordura você usa geralmente:*

- | | | |
|--------------------------|-----------------------------|--------------------------|
| (1) azeite de oliva | (4) óleo de outras sementes | (7) não usa gordura |
| (2) azeite de dendê/coco | (5) manteiga | (8) outra gordura animal |
| (3) óleo vegetal | (6) margarina | (9) não sabe |

E24 *para temperar vegetais etc.?*

E25 *para cozinhar?*

E26 Nos últimos 2 anos, você usou suplementos vitamínicos?

(1) Sim (2) Não (9) não sabe [se "não" ir para E29]

E27 Com que frequência você tem tomado suplementos vitamínicos?.....|_|

- (1) Diariamente
- (2) Ao menos 1x/semana
- (3) Ao menos 1x/mês
- (4) Ocasionalmente
- (5) Nunca

E28 Com que idade você começou a tomar vitaminas (na idade adulta)? |_|_|

E29 Qual o seu peso (Kg)?..... |_|_|_|

E30 Qual é sua altura (cm)?..... |_|_|_|

E31 Você nasceu de:

A) Parto Normal

B) Cesariana

RGH

USO DE ÁLCOOL

F1 *Você já tomou bebidas alcoólicas, por pelo menos 1x/mês.....* |

(1) Sim (2) Nunca (3) No passado [se “nunca” ir para próxima página]

F2 *Quando você bebe?*

(1) junto com as refeições; (2) entre refeições; (3) ambas

Por favor, descreva os períodos da sua vida durante os quais você bebeu, a quantidade e outros detalhes sobre o produto usado. Tente resumir as mudanças mais importantes da sua vida quanto a quantidade e tipo de bebida. Ignore as mudanças que aconteceram por menos de 1 ano ou beber alguma bebida ocasionalmente.

Entrevistador: Evite sobreposição de anos para o mesmo produto ou tipo de cigarro. (b,c), i.e. escreva 30-40, 41-45 ao invés de 30-40, 40-45.

F3 *Bebida (a)* *Da idade?* *Para idade?*

Unidade (b)	Consumo Quantidade?	Por (c)
-------------	------------------------	---------

[illegible]

- | | | |
|--|--|-------------|
| (a) (1) Vinho | (b) (1) Copo pequeno (50 ml/1.5 oz) | (c) (1) Dia |
| (2) Cerveja e cidra | (2) Copo médio (100 ml/3 oz) | (2) Semana |
| (3) Destiladas ($\geq 35^\circ$) | (3) Copo grande (250 ml) | (3) Mês |
| (uísque, conhaque, vodca, marc, gin, cachaça, pinga) | (4) $\frac{1}{2}$ garrafa pequena (330 ml) | |
| (4) Cocktail/Aperitivo ($<35^\circ$) | (5) Garrafa (700-750 ml) | |
| (Martini, porto, vermut) | | |
| (5) Outros (especificar) | | |

RGH

HISTÓRICO CIVIL

G1 *Você já foi casado(a) ou viveu como casado(a)?*|_|
(1) Sim (2) Não [se “não” ir para G5]

G2 *Você continua casado(a) ou vivendo como casado(a)?*|_|
(1) Sim (2) Não, Separado(a)/divorciado(a); (3) Não, Viúvo(a)

G3 *Quantas vezes você casou ou viveu como casado(a)?*|_|_|

G4 *Que idade você tinha quando casou ou viveu como casado(a) pela primeira vez?*
|_|_|

G5 *Quantos filhos você tem ou teve no total?*|_|_|

RGH

HISTÓRICO DE DOENÇAS DIVERSAS

H1 *Você já teve verrugas de pele?*|_|
(1) Sim (2) Não; (9) Não sabe [se “não” ir para H6]

Se “sim”, em que lugar? (1) Sim (2) Não

H2 Mãos|_|
H3 Pés.....|_|
H4 Cabeça e pescoço.....|_|
H5 Outro (especificar)|_|

H6 *Durante sua vida adulta, você já teve candidíase/sapinho:*.....|_|
(1) Sim (2) Não; (9) Não sabe [se “não” ir para H10]

Se “sim”, em que lugar? (1) Sim; (2) Não

H7 Genitais|_|
H8 Boca.....|_|
H9 Outro (especificar)|_|

H10 *Você já teve lesões de herpes?*|_|
(1) Sim (2) Não; (9) Não sabe [se “não” ir para H14]

Se “sim”, em que lugar? (1) Sim; (2) Não

H11 Lábio|_|
H12 Genitais.....|_|
H13 Outro (especificar)|_|

H14 *Você já teve doenças sexualmente transmissíveis?*|_|
(1) Sim (2) Não; (9) Não sabe [se “não” ir para J1]

Se “sim”, quais? (1) Sim (2) Não; (9) Não sabe

H15 Sífilis/úlceras|_|
H16 Gonorróea/corrimento|_|
H17 Condiloma/verruga.....|_|
H18 HIV/AIDS|_|

RGH

SAÚDE ORAL

J1 Com que frequência você limpa os dentes?.....|_|

- (0) Nunca (5) 2 x/dia
(1) < de 1x/semana (6) 3 x/dia
(2) 1-2 de x/semana (7) > 3 x/dia
(3) Dia sim, dia não (8) Nenhuma das anteriores [ir para J5]
(4) 1 x/dia

J2 O quê você usa para limpar seus dentes?..... |_|

- (1) Escova de dentes
(2) Dedo
(3) Palitos
(4) Outro _____ (especificar)

J3 Qual material você usa para limpeza dos dentes?|_|

- (1) Nada
(2) Pasta de dente
(3) Outro _____ (especificar)

J4 Sua gengiva sangra quando você limpa os dentes?.....|_|

- (1) Não (2) Às vezes; (3) Sempre ou quase sempre

J5 Com que frequência você usa desaguatório bucal?.....|_|

- (0) Nunca (5) 2 x/dia
(1) < de 1x/semana (6) 3 x/dia
(2) 1-2 x/semana (7) > 3 x/dia
(3) Dia sim, dia não (8) Nenhuma das anteriores
(4) 1 x/dia

J6 Você usa dentadura?|_|

- (1) Sim (2) Não [se "não" ir para J9]

J7 É uma dentadura total?.....|_|

- (1) Sim (2) Não

J8 Com que idade você começou a usar dentadura?|_|

J9 Nos últimos 20 anos, com que frequência você foi ao dentista?.....|_|

- (1) Todo ano (3) A cada 5 anos
(2) 1 x dentro de 2-5 anos (4) Nunca

J10 Você já fez biópsia oral?|_|

- (1) Sim (2) Não [se "não" ir para J13]

J11 Com qual idade?|_|

J12 O quê ela mostrou? _____ |_|
(especificar)