

**UNIVERSIDADE DO ESTADO DO AMAZONAS – UEA  
ESCOLA DE CIÊNCIAS DA SAÚDE – ESA  
DOUTORADO EM BIOQUÍMICA E BIOLOGIA MOLECULAR**

**ANDERSON NOGUEIRA BARBOSA**

**ANÁLISE METAGENÔMICA DA MICROBIOTA ASSOCIADA À  
DOENÇA PERIODONTAL**

**MANAUS – 2019**

**ANDERSON NOGUEIRA BARBOSA**

**ANÁLISE METAGENÔMICA DA MICROBIOTA ASSOCIADA À  
DOENÇA PERIODONTAL**

Tese apresentada ao Programa de Pós-graduação em Bioquímica e Biologia Molecular (PMBqBM) da Universidade do Estado do Amazonas (UEA) como parte dos requisitos para obtenção do Título de Doutor em Bioquímica e Biologia Molecular.

Orientador: Prof. Dr. Daniel Saito

Coorientadora: Profa. Dra. Aline Maria da Silva

**MANAUS – 2019**

## **Ficha Catalográfica**

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).  
**Sistema Integrado de Bibliotecas da Universidade do Estado do Amazonas.**

B238a    Barbosa, Anderson Nogueira  
          Análise metagenômica da microbiota associada à doença  
          periodontal. / Anderson Nogueira Barbosa. Manaus : [s.n],  
          2019.  
          107 f.: color.; 30 cm.

          Tese - PGSS - Doutorado em Bioquímica e Biologia  
          Molecular (Doutorado) - Universidade do Estado do  
          Amazonas, Manaus, 2019.  
          Inclui bibliografia  
          Orientador: Saito, Daniel  
          Coorientador: Silva, Aline Maria da

          1. Periodontite.    2. Microbiota.    3. Metagenômica. I.  
          Saito, Daniel (Orient.). II. Silva, Aline Maria da  
          (Coorient.). III. Universidade do Estado do Amazonas. IV.  
          Análise metagenômica da microbiota associada à doença  
          periodontal.

**Elaborado por Jeane Macelino Galves - CRB-11/463**



**ANDERSON NOGUEIRA BARBOSA**

**“ANÁLISE METAGENÔMICA DA MICROBIOTA ASSOCIADA À  
DOENÇA PERIODONTAL”.**

TESE DE DOUTORADO DO ALUNO  
**ANDERSON NOGUEIRA BARBOSA** EM  
BIOQUÍMICA E BIOLOGIA MOLECULAR  
PMBqBM-UEA.

Orientador (a): Prof. Dr. Daniel Saito  
Coorientador (a): Profa. Dra. Aline Maria da Silva

**Banca examinadora:**

---

**Prof. Dr. Daniel Saito**  
Orientador – Presidente da banca

---

**Prof. Dr. Cleiton Fantin Rezende**  
Membro Interno

---

**Prof. Dr. Rudi Emerson de Lima Procópio**  
Membro Externo

---

**Prof. Dr. Edson Júnior do Carmo**  
Membro Externo

---

**Prof. Dr. Gemilson Soares Pontes**  
Membro Externo

## AGRADECIMENTOS

Agradeço primordialmente a Deus por estar sempre presente em minha vida. É ele o principal responsável por ainda ter forças para continuar esta caminhada e superar os momentos em que cheguei a abrir mão de tudo.

É também indescritível o apoio que tive dos meus familiares. É impossível descrever o quanto dói ver seus pais tentando ajudá-lo mesmo sem ter conhecimento do seu trabalho. É nesse momento que uma merenda da tarde, uma carona, um passeio pelo Rio Amazonas entre pai e filho, sem motivo algum, em um dia útil qualquer, faz toda a diferença. Obrigado meu pai, Aldair Kleber Rodrigues Barbosa, pelos conselhos que me deu, foram importantíssimos para que mantivesse minha mente sã. Minha mãe, Alcilene de André Nogueira, pelos conselhos e exemplo de ser a mulher que dedicou a vida para que os filhos fossem o mais longe possível em seus sonhos. Minha irmã, Ana Paula Nogueira Barbosa, que me ajudou consideravelmente com a língua inglesa, além dos momentos de descontração, brincadeiras e inspiração. Minha companheira, Sabrina Moura Parente, que desde a graduação tem apoiado minha caminhada e sido um braço direito rumo à realização dos meus sonhos.

Existem pessoas que de longa data me ajudaram de forma indireta e torceram positivamente para que este momento chegasse. Agradeço pelo apoio e sentimento de amizade mútuos. Este agradecimento vai para as Sras. Ivete Lima Moreira, Maria das Neves Rodrigues Rufino, Helineia Maria Nunes Meireles, Maria de Fátima Arruda Reis, Maria Núbia Rodrigues Barbosa, Angela Maria da Silva Moura e Irmã Rosário Rodrigues Barbosa.

Ao meu orientador Daniel Saito, agradeço pelo apoio durante os anos que estivemos trabalhando juntos na realização desta pesquisa. Não há palavras que descrevam o quanto sou imensamente grato pela paciência e compreensão das adversidades que se fizeram presente durante essa caminhada. É indiscutível o quanto foi importante sua pessoa e o quanto pude crescer profissionalmente tendo-o como exemplo.

A Profa. Tsai Siu Mui e o pesquisador Leandro Lemos do Centro de Energia Nuclear na Agricultura (CENA) da Universidade do Estado de São Paulo (USP), agradeço pelo apoio dado ao trabalho.

A Profa. Cristiane Pereira Borges Saito e ao aluno Rodrigo Oliveira, agradeço pela coleta do material biológico realizado durante as fases iniciais do trabalho.

Ao Emerson Bacellar Máximo, digo: “Não havia amigo melhor para encontrar no setor em que realizei meu trabalho”. Certamente um amigo inestimável para os diferentes

momentos e fases que tive durante esta formação. Sábios foram aqueles que o conheceram e puderam aproveitar seus conselhos e conhecimentos como fiz.

Devo agradecer ao André Miasato Higa que acompanhou toda a minha caminhada durante o curso. Certamente foi a pessoa com a qual tirei muitas dúvidas além de aprender diversas práticas laboratoriais paralelas ao projeto e que contribuíram imensamente para meu crescimento como pesquisador.

Uma das pessoas que admiro é Fabio Marinho Machado. Considero-o um guerreiro frente as adversidades que nos afligiram durante este período. De fato, um amigo certo para o momento. Este sabe o quanto lutamos juntos para conquistar mutuamente nosso título de pós-graduados. Vindo da história que conheço, agradeço por ser uma das inspirações que me mantiveram firme nessa jornada até este momento.

Amiga desde a graduação, Alessandra Silva e Silva também merece seu destaque. Obrigado pelos momentos que tivemos de descontração, conversas e apoio mútuo. Pode parecer simples, mas o fato de ainda nos encontrarmos juntos reforça a necessidade de união. Muita coisa precisa ser mudada e muito caminho há para atingirmos nosso sonho de graduados – unidos espero que consigamos.

Embora a maioria tenha deixado a universidade bem antes de mim, agradeço por fazerem parte desta etapa da minha formação acadêmica. Muitos foram os momentos que tivemos juntos, sejam bons ou ruins. Muito obrigado por tudo, Nazaré Messias, Rafael de S. Rodrigues, Rodrigo Maciel, Jéssica Varela, Gilmar Seixas da Silva, Vanderlei Saboia, Paola Ortiz, Ana Tana, Marcelo Capistana, Giovana Caetano, Aldiane Oliveira, Ana Cláudia Nunes, André Marcos e Remígio Cenepo.

Finalmente, agradeço a Universidade do Estado do Amazonas (UEA) pelo apoio concedido. Agradeço também a Coordenação de Aperfeiçoamento de Pessoa de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) pelo auxílio financeiro.

*“Posso não ter feito o melhor, mas lutei para que o melhor fosse feito. Posso não ter me tornado o que deveria ser, mas certamente não sou o que era antes.”*

***Martin Luther King***

## RESUMO

A periodontite é uma doença infecciosa que afeta os tecidos que envolvem e suportam os dentes. Embora algumas bactérias tenham sido associadas ao desenvolvimento dessa doença, pesquisas baseadas em técnicas moleculares podem trazer informações importantes, uma vez que uma parcela significativa da microbiota oral não pode ser detectada pelo cultivo em laboratório. Este estudo prospectivo de caso-controle objetiva avaliar a filiação taxonômica e os processos funcionais envolvidos na periodontite, com base na análise metagenômica de portadores de periodontite e indivíduos saudáveis. Amostras de saliva não-estimulada foram coletadas de indivíduos saudáveis (grupo controle, n = 13) e com periodontite (grupo caso, n = 14) atendidos na Policlínica Odontológica da UEA. O DNA total foi extraído das amostras, quantificado e submetido a sequenciamento de última geração na plataforma Illumina HiSeq2500. Após a limpeza dos dados, as sequências nucleotídicas foram submetidas ao emparelhamento pelo software PEAR, previsão taxonômica pelos programas Metaphlan 2.0 e MG-RAST (banco de dados RefSeq), aferição dos índices de diversidade alfa (Shannon-Wiener e Simpson) e beta (Whittaker) e anotação nos bancos de dados funcionais KEGG e CAZy. A diferença de frequência dos táxons e categorias funcionais entre os grupos foi avaliada pelo teste de Kruskal-Wallis ( $\alpha = 0,05$ ). Um total de 104.004.730 sequências foram submetidas à anotação taxonômica, sendo 3.267.049 (3,23%) e 49.391.119 (47,49%) anotadas pelos Metaphlan 2.0 e MG-RAST, respectivamente. As análises estatísticas foram realizadas pelos programas STAMP e R v3.4.3. Com base nos resultados, os filos mais abundantes foram *Firmicutes* (29,52%), *Actinobacteria* (27,83%), *Bacteroidetes* (20,43%), *Proteobacteria* (17,94%), *Fusobacteria* (2,97%), *Candidatus Saccharibacteria* (0,45%), *Spirochaetes* (0,33%) e *Tenericutes* (0,03%). A análise comparativa entre os grupos de estudo evidenciou a existência de variações na abundância ou presença/ausência de táxons que podem estar associadas à doença periodontal. De fato, a constatação de frequências aumentadas de representantes dos filos *Bacteroidetes*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Synergistes* no grupo de portadores de periodontite sugere um potencial papel patogênico. Ademais, a análise de correlação indica que os gêneros *Eubacterium*, *Leptotrichia*, *Corynebacterium*, *Treponema*, *Parvimonas*, *Fusobacterium*, *Filifactor*, *Capnocytophaga*, *Tannerella*, *Bifidobacterium*, *Peptostreptococcus*, *Dialister*, *Catonella*, *Selenomonas*, *Campylobacter* e *Porphyromonas* constituem um grupo com alto potencial periodontopatogênico. A análise funcional mostrou que a categoria “Metabolismo” do banco de dados KEGG foi predominante na cavidade oral, enquanto que a categoria “Processamento de informação genética” foi positivamente relacionada à doença periodontal. Outras categorias que demonstraram maior abundância no grupo doente ( $p < 0,05$ ) foram “Biossíntese de aminoácil-RNAt”, “Metabolismo do piruvato” e “Vias de fixação de carbono em procariontes”. Além disso, a análise funcional revelou 11 potenciais genes bioindicadores da doença periodontal, sendo eles: *rpoC*, *rpoB*, *carB*, *secA*, *uvrA*, *pflD*, *thrS*, *pheT*, *ntpA*, *cobQ* e *aspS*. Conclui-se que a cavidade oral apresenta uma comunidade microbiana diversificada, característica e com alta variabilidade interindividual. No entanto, existem grupos seletos de bactérias que parecem estar associadas a condições de saúde e/ou doença. Essa microbiota é capaz de conviver em sinergismo, modulando sua atividade funcional e complexando ainda mais o mecanismo da doença. Por fim, esse estudo reitera que a cavidade oral abriga quantidade significativa de bactérias desconhecidas, assim como genes com potencial envolvimento na doença periodontal, e que o emprego de técnicas de sequenciamento de DNA de alto rendimento constitui importante ferramenta no estudo do microbioma oral.



## ABSTRACT

Periodontitis is an infectious disease that affects the tissues involving and supporting teeth. Although specific bacteria have been implicated with periodontitis, application of molecular-based research can bring additional information, since a significant parcel of the oral microbiota cannot be detected by traditional cultivation methods. This prospective case-control study aims to assess the taxonomic affiliation and functional processes involved in periodontitis, based on the metagenomic analysis of periodontitis carriers and healthy subjects. Saliva samples were collected from healthy (control group, n = 13) and periodontitis subjects (case group, n = 14) attended at the UEA Dental Clinic. Total DNA was extracted from the samples, quantified, and submitted to next-generation sequencing in Illumina HiSeq2500 platform. After data cleaning, the nucleotide sequences were subjected to pairing by the PEAR software, taxonomic prediction by Metaphlan 2.0 and MG-RAST (RefSeq data), measurement of alpha (Shannon-Wiener and Simpson) and beta diversity indexes (Whittaker), and functional annotation with KEGG and CAZy molecular database tools. The difference in the frequency of taxa and functional categories between groups was assessed by the Kruskal-Wallis test ( $\alpha = 0.05$ ). A total of 104,004,730 sequences was submitted to annotation, of which 3,267,049 (3.23%) and 49,391,119 (47.49%) were annotated by Metaphlan 2.0 and MG-RAST, respectively. The statistical analyzes were performed by STAMP and R v3.4.3 softwares. Based on the results, the most abundant phyla were *Firmicutes* (29.52%), *Actinobacteria* (27.83%), *Bacteroidetes* (20.43%), *Proteobacteria* (17.94%), *Fusobacteria* (2.97%), *Candidatus Saccharibacteria* (0.45%), *Spirochaetes* (0.33%) and *Tenericutes* (0.03%). Comparative analysis revealed variations in the abundance or presence/absence of taxa between the study groups that suggests an intricate association with periodontal disease. Moreover, the increased frequencies of representatives of *Bacteroidetes*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* and *Synergistes* in the periodontitis group suggest potential pathogenic roles. Moreover, correlation analysis indicated that the genera *Eubacterium*, *Leptotrichia*, *Corynebacterium*, *Treponema*, *Parvimonas*, *Fusobacterium*, *Filifactor*, *Capnocytophaga*, *Tannerella*, *Bifidobacterium*, *Peptostreptococcus*, *Dialister*, *Catonella*, *Selenomonas*, *Campylobacter* and *Porphyromonas* embody a potential periodontopathogenic group. Functional analysis showed that KEGG's "Metabolism" category was predominant in the oral cavity. Meanwhile, the category "Genetic Information Processing" was positively related to periodontal disease. Other categories that demonstrated higher abundance ( $p < 0.05$ ) for the sick group were the Aminoacyl-tRNA biosynthesis, pyruvate metabolism, and carbon fixation pathways in prokaryotes. In addition, functional analysis resulted in 11 potential bioindicator genes of periodontal disease such as: *rpoC*, *rpoB*, *carB*, *secA*, *uvrA*, *pflD*, *thrS*, *pheT*, *ntpA*, *cobQ* e *aspS*. It can be concluded that the oral cavity presents a diverse microbial community with high inter-individual variability. However, there are select groups of bacteria that appear to be associated with health and disease conditions. This microbiota is able to live in synergism by modulating its functional activity and by complicating even more the mechanism of the disease. Finally, this study reiterates that the oral cavity houses a significant number of unknown bacteria, as well as genes potentially involved in periodontal disease, and that high throughput DNA sequencing constitutes an important tool in the study of the oral microbiome.

## LISTA DE FIGURAS

- Figura 1. Anatomia do periodonto. Adaptado de Lindhe, Karring, Lang (2005)..... 21
- Figura 2. Complexos de Socransky&Haffajee. Adaptado de Socransky, Haffajee (2005)..... 23
- Figura 3. Procedimento de sondagem para obtenção de índice periodontal comunitário (CPI). Adaptado de Fitzpatrick, Wijeyewickrema, Pike (2009). ..... 24
- Figura 4. Perfil taxonômico de todas as mostras envolvidas no estudo. Os perfis foram construídos através da abundância relativa dos principais filos obtidos durante a previsão taxonômica pelo programa Metaphlan 2.0. .... 43
- Figura 5. *Boxplot* demonstrando a distribuição na proporção do filo *Actinobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0..... 43
- Figura 6. *Boxplot* demonstrando a distribuição na proporção do filo *Bacteroidetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0..... 44
- Figura 7. *Boxplot* demonstrando a distribuição na proporção do filo *Candidatus Saccharibacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos da previsão taxonômica pelo programa Metaphlan 2.0. .... 44
- Figura 8. *Boxplot* demonstrando a distribuição na proporção do filo *Fusobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0..... 45
- Figura 9. *Boxplot* demonstrando a distribuição na proporção do filo *Proteobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0..... 45
- Figura 10. *Boxplot* demonstrando a distribuição na proporção do filo *Spirochaetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0..... 46

Figura 11. <i>Boxplot</i> demonstrando a distribuição na proporção do filo <i>Synergistetes</i> com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os <i>outliers</i> por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.....	46
Figura 12. <i>Heatmap</i> de todos os gêneros que apresentaram abundância com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.....	49
Figura 13. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa das espécies dos grupos doente (azul) e saudável (laranja) obtidos durante análise pelo programa Metaphlan 2.0. ....	49
Figura 14. <i>Heatmap</i> das 25 espécies com maiores abundâncias e com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.....	50
Figura 15. Análise de Discriminantes Lineares (LDA) sobre os gêneros resultantes da anotação taxonômica através do programa Metaphlan 2.0.....	50
Figura 16. Análise de Discriminantes Lineares (LDA) sobre as espécies resultantes da anotação taxonômica através do programa Metaphlan 2.0.....	51
Figura 17. Matriz de correlações de Spearman, para os grupos saudável (A) e doente (B), construída a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste estatístico não paramétrico Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0. ....	54
Figura 18. Curva de rarefação gerada a partir das sequências anotadas pelo servidor MG-RAST para grupo doente. As curvas representam a média do número de diferentes sequências anotadas para cada amostra do grupo. ....	57
Figura 19. Curva de rarefação gerada a partir das sequências anotadas pelo servidor MG-RAST para grupo saudável. As curvas representam a média do número de diferentes sequências anotadas para cada amostra do grupo.....	57
Figura 20. Perfil taxonômico de todas as mostras envolvidas no estudo, exceto a amostra 1. Os perfis foram construídos através da abundância relativa dos principais filios obtidos durante a previsão taxonômica pelo servidor MG-RAST. ....	59
Figura 21. <i>Boxplot</i> demonstrando a distribuição na proporção do filo <i>Actinobacteria</i> com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os <i>outliers</i> por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST. ....	60

- Figura 22. *Boxplot* demonstrando a distribuição na proporção do filo *Bacteroidetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST. .... 60
- Figura 23. *Boxplot* demonstrando a distribuição na proporção do filo *Fusobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST. .... 61
- Figura 24. *Boxplot* demonstrando a distribuição na proporção do filo *Spirochaetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST. .... 61
- Figura 25. *Heatmap* dos gêneros com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST. .... 62
- Figura 26. Diagrama de Venn da riqueza de espécies encontrada entre os grupos doente (vermelho) e saudável (azul) fazendo uso dos dados de abundância obtidos do programa Metaphlan 2.0. .... 62
- Figura 27. Estimativa de riqueza de espécies em amostras do grupo saudável. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0. .... 63
- Figura 28. Estimativa de riqueza de espécies em amostras do grupo doente. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0. .... 63
- Figura 29. *Boxplot* demonstrando a distribuição do índice de Shannon com diferença significativa para o teste estatístico Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa e os *outliers* por círculos. .... 65
- Figura 30. *Boxplot* demonstrando a distribuição do índice de Simpson com diferença significativa para o teste estatístico Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa e os *outliers* por círculos. .... 65
- Figura 31. Representação filogenética em forma de cladograma dos táxons resultantes da análise taxonômica pelo programa Metaphlan 2.0 do grupo saudável. O tamanho dos círculos internos assim como das barras externas demonstra, de modo proporcional, a abundância do nível taxonômico e das espécies, respectivamente. .... 67
- Figura 32. Representação filogenética em forma de cladograma dos táxons resultantes da análise taxonômica pelo programa Metaphlan 2.0 do grupo doente. O tamanho dos círculos

internos assim como das barras externas demonstra, de modo proporcional, a abundância do nível taxonômico que se esteja observando e das espécies, respectivamente. .... 68

Figura 33. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 1 (KEGG) dos grupos doente (azul) e saudável (laranja)..... 70

Figura 34. *Boxplot* demonstrando a distribuição na proporção categoria “Processamento de informação genética” com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos da análise funcional a partir do banco de dados KEGG, nível hierárquico 1, realizada pelo servidor MG-RAST. .... 71

Figura 35. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 2 (KEGG) dos grupos doente (azul) e saudável (laranja)..... 72

Figura 36. Perfil funcional de cada indivíduo envolvido no estudo com base no nível hierárquico 2 obtido do banco de dados KEGG. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. .... 73

Figura 37. *Boxplot* demonstrando a distribuição na proporção categoria “Transcrição” com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos da análise funcional a partir do banco de dados KEGG, nível hierárquico 2, realizada pelo servidor MG-RAST. .... 74

Figura 38. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 3 (KEGG) dos grupos doente (azul) e saudável (laranja)..... 75

Figura 39. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico “função” (KEGG) dos grupos doente (azul) e saudável (laranja). .... 76

Figura 40. Análise de Discriminantes Lineares (LDA) aplicada aos dados de abundância relativa dos resultados obtidos para o nível hierárquico “função” (KEGG) com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. Em vermelho, todos os genes resultantes para o grupo doente. Em verde, apenas os 11 genes que obtiverem as melhores pontuações para o grupo saudável..... 77

Figura 41. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para a categoria “Família” (CAZy) dos grupos doente (azul) e saudável (laranja)..... 79

Figura 42. *Boxplot* demonstrando a distribuição na proporção da família Polissacarídeo Lases (PL) com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruces. Dados de abundância obtidos a partir da análise funcional a partir do banco de dados CAZy, nível de família, realizada pelo programa DIAMOND (Blastp)..... 80

## LISTA DE TABELAS

Tabela 1. Quantificação de DNA total, realizada em Qubit 3.0, para cada amostra envolvida no estudo.....	38
Tabela 2. Número de sequências e Mpb resultantes em cada amostra envolvida no estudo após o pré-processamento dos dados. ....	39
Tabela 3. Quantidade de sequências emparelhadas, pares de bases e seus respectivos tamanhos médios, para cada amostra envolvida no estudo, gerados durante a etapa de montagem. O símbolo “*” indica as amostras que apresentaram apenas um arquivo contendo <i>reads</i> durante o sequenciamento.....	40
Tabela 4. Número de sequências submetidas e anotadas, para cada amostra envolvida no estudo, durante a previsão taxonômica através do programa Metaphlan 2.0. ....	41
Tabela 5. Abundância relativa geral, de todos os filios encontrados, para cada amostra envolvida no estudo, obtida durante a previsão taxonômica através do programa Metaphlan 2.0. ....	42
Tabela 6. Abundância relativa, de todos os filios encontrados, para cada grupo de estudo, obtida durante a previsão taxonômica através do programa Metaphlan 2.0. O símbolo “*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ).....	42
Tabela 7. Abundâncias relativas dos 10 gêneros mais abundantes no grupo doente, obtidas durante a previsão taxonômica através do programa Metaphlan 2.0. ....	47
Tabela 8. Abundâncias relativas dos 10 gêneros mais abundantes no grupo saudável, obtidas durante a previsão taxonômica através do programa Metaphlan 2.0. ....	47
Tabela 9. Abundância relativa, para ambos os grupos de estudo, de todos os gêneros que apresentaram diferença estatística significativa para o teste não paramétrico Kruskal-Wallis ( $p < 0,05$ ), obtida durante a previsão taxonômica através do programa Metaphlan 2.0.....	48
Tabela 10. Correlações de Spearman fortes ( $r < -0,7$ e $r > +0,7$ ), para o grupo saudável, obtidas a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0.....	52
Tabela 11. Correlações de Spearman fortes ( $r < -0,7$ e $r > +0,7$ ), para o grupo doente, obtidas a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0. ....	53
Tabela 12. Número de sequências, pares de base (pb) e tamanho médio, para cada amostra envolvida no estudo, submetida a análise pelo servidor MG-RAST.....	55

Tabela 13. Número de sequências anotadas, para cada amostra envolvida no estudo, durante a previsão taxonômica através do servidor MG-RAST. *Apresentou apenas um arquivo contendo <i>reads</i> durante o sequenciamento.....	56
Tabela 14. Abundância absoluta e relativa geral dos principais filós encontrados na cavidade oral durante a previsão taxonômica através do servidor MG-RAST.....	58
Tabela 15. Abundância relativa dos principais filós encontrados para os grupos doente e saudável durante a previsão taxonômica através do servidor MG-RAST. O símbolo “**” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ). .....	59
Tabela 16. Índices de diversidade de Shannon e Simpson (1-D), e riqueza de cada amostra. Foram utilizados os dados de abundância obtidos do programa Metaphlan 2.0. ....	64
Tabela 17. <i>Open Read Frames</i> (ORFs) anotadas em banco de dados KEGG através do servidor MG-RAST. ....	69
Tabela 18. Abundância relativa para o nível hierárquico 1 obtido do banco de dados KEGG, para cada grupo de estudo. O símbolo “*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. ....	70
Tabela 19. Abundância relativa das categorias obtidas do nível hierárquico 3 do banco de dados KEGG que apresentaram diferença estatística significativa entre os grupos de estudo para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. ....	75
Tabela 20. <i>Open Read Frames</i> (ORFs) anotadas em banco de dados CAZy através do programa DIAMOND. ....	78
Tabela 21. Abundância relativa das Famílias CAZy obtidas do banco de dados CAZy para cada grupo de estudo. O símbolo “**” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. .	79



## LISTA DE ABREVIATURAS E SIGLAS

µg – Micrograma  
µL – Microlitro  
AA – Atividades Auxiliares  
CAZy – Carbohydrate-Active Enzyme  
CBM – Módulos de Ligação à Carboidratos  
CE – Esterases de Carboidratos  
CPI – Índice periodontal comunitário  
DNA – Ácido desoxirribonucleico  
EggNOG – Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups  
ESA – Escola Superior de Ciências da Saúde  
EUA – Estados Unidos da América  
Ex. – Exemplo  
GH – Glicosídeos Hidrolases  
GT – Glicosiltransferases  
KEGG – Kyoto Encyclopedia of Genes and Genomes  
KO – KEGG Orthology  
LDA – Análise de discriminantes lineares  
MB – Megabites  
Mpb – Mega pares de bases  
NCBI – National Center for Biotechnology Information  
ng – Nanograma  
ORF – Open Reading Frame  
pb – Pares de bases  
PCA – Análise de Componentes Principais  
PCR – Reação em Cadeia da Polimerase  
PEAR – Paired-end read merger  
PGH – Projeto Genoma Humano  
PL – Polissacarídeo Liases  
RefSeq – Reference Sequence Database  
RNA – Ácido ribonucleico  
rRNA – Ácido ribonucleico ribossomal  
STAMP – Statistical analysis of taxonomic and functional profiles  
tSMS - True Single Molecule Sequencing  
UEA – Universidade do Estado do Amazonas  
UV – Ultravioleta

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>19</b>
1.1 MICROBIOTA ORAL .....	19
1.2 PERIODONTITE .....	20
1.3 SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS) .....	25
1.4 METAGENÔMICA .....	26
<b>2 JUSTIFICATIVA .....</b>	<b>29</b>
<b>3 OBJETIVOS .....</b>	<b>30</b>
3.1 OBJETIVOS GERAIS .....	30
3.2 OBJETIVOS ESPECÍFICOS .....	30
<b>4 MATERIAL E MÉTODOS .....</b>	<b>31</b>
4.1 CONSIDERAÇÕES ÉTICAS .....	31
4.2 POPULAÇÃO DE ESTUDO .....	31
4.3 CRITÉRIOS DE INCLUSÃO .....	31
4.4 SELEÇÃO DE PACIENTES .....	31
4.5 COLETA .....	32
4.6 EXTRAÇÃO DE DNA TOTAL E QUANTIFICAÇÃO .....	32
4.7 CONSTRUÇÃO DA BIBLIOTECA E SEQUENCIAMENTO .....	32
4.8 PRÉ-PROCESSAMENTO DOS DADOS E EMPARELHAMENTO DE <i>READS</i> .....	33
4.9 ANOTAÇÃO TAXONÔMICA VIA METAPHLAN 2.0.....	33
4.10 ANOTAÇÃO TAXONÔMICA VIA MG-RAST .....	34
4.11 RELAÇÕES FILOGENÉTICAS, RIQUEZA E DIVERSIDADE .....	35
4.12 ANOTAÇÃO FUNCIONAL VIA KEGG .....	35
4.13 ANOTAÇÃO FUNCIONAL VIA CAZY.....	35
4.14 ANÁLISE ESTATÍSTICA.....	36
<b>5 RESULTADOS .....</b>	<b>38</b>
5.1 EXTRAÇÃO DE DNA TOTAL E QUANTIFICAÇÃO .....	38
5.2 PRÉ-PROCESSAMENTO DOS DADOS .....	39
5.3 EMPARELHAMENTO DE <i>READS</i> .....	40
5.4 ANOTAÇÃO TAXONÔMICA VIA METAPHLAN 2.0.....	41
5.5 ANOTAÇÃO TAXONÔMICA VIA MG-RAST .....	55
5.6 RELAÇÕES FILOGENÉTICAS, RIQUEZA E DIVERSIDADE.....	62
5.7 ANOTAÇÃO FUNCIONAL VIA KEEG .....	69
5.8 ANOTAÇÃO FUNCIONAL VIA CAZy .....	77
<b>6 DISCUSSÃO .....</b>	<b>81</b>
<b>7 CONCLUSÃO.....</b>	<b>95</b>
<b>8 REFERÊNCIAS .....</b>	<b>96</b>
<b>ANEXOS .....</b>	<b>104</b>

# 1 INTRODUÇÃO

## 1.1 MICROBIOTA ORAL

A cavidade oral representa um dos mais complexos microbiomas já descritos, sendo composto, principalmente, por uma ampla variedade de espécies bacterianas, tanto comensais, quanto transientes e/ou patogênicas. Nesse microbioma, é possível encontrar mais de 600 espécies bacterianas ou filotipos, os quais coexistem principalmente na forma de biofilme banhado por saliva, aderido comumente aos dentes, gengiva e língua. Essa forma de convivência em biofilmes foi refinada no decorrer do processo coevolutivo, de modo que muitos dos indivíduos perderam a capacidade de sobreviver fora da comunidade, devido à uma estreita interdependência metabólica (DEWHIRST et al. 2010; JENKINSON, LAMONT, 2005). A ausência de condições ambientais nas quais se torna possível a sobrevivência dos microrganismos fora do biofilme resulta na incapacidade de cultivá-los *in vitro*, constituindo-se o principal fator complicador no estudo da microbiota oral (PASTER et al. 2006; PENNISI, 2005).

Embora o cultivo bacteriano tenha sido inicialmente utilizado para a compreensão da microbiota oral, este método tem sido gradativamente substituído por outros, denominados “independentes de cultivo”. Dentre estes, destaca-se a clonagem molecular do gene ribossômico 16S (16S rDNA), por meio da qual é possível tanto revelar a identidade de organismos passíveis de cultivo, quanto a dos não-cultiváveis. O avanço e disseminação de técnicas moleculares independentes de cultivo têm demonstrado diversos resultados que se somam àqueles obtidos através de métodos limitados à cultura. Sabe-se, agora, que a microbiota oral é composta principalmente por bactérias pertencentes aos filos *Firmicutes*, *Actinobacteria*, *Proteobacteria*, *Bacteroidetes*, *Fusobacteria*, *Treponema* e *Candidatus Saccharibacteria*. Além disso, a composição das comunidades microbianas orais pode variar significativamente entre indivíduos, dependendo do sítio analisado (língua, dentes, periodonto, palato, mucosa jugal etc.) e da condição clínica do indivíduo (saúde ou doença), (DEWHIRST et al. 2010; AAS et al. 2005).

Embora composta também por bactérias patogênicas, a microbiota oral pode conviver em harmonia com o hospedeiro sem ocasionar um estado de doença. Entretanto, essa comunidade polimicrobiana pode responder a diferentes condições ambientais, que atuam como agentes moduladores da composição de espécies e tamanho da população, dentre elas: variações de temperatura, pH, potencial redox, imunoglobulinas, capacidade tampão e fluxo

salivar. Uma vez perturbada a condição ambiental na qual a comunidade está inserida, pode haver mudanças fisiológicas nesta, podendo acarretar no sobressalto de atividades patogênicas e, conseqüentemente, um estado de infecção (AVILA, OJCIUS, YILMAZ, 2009).

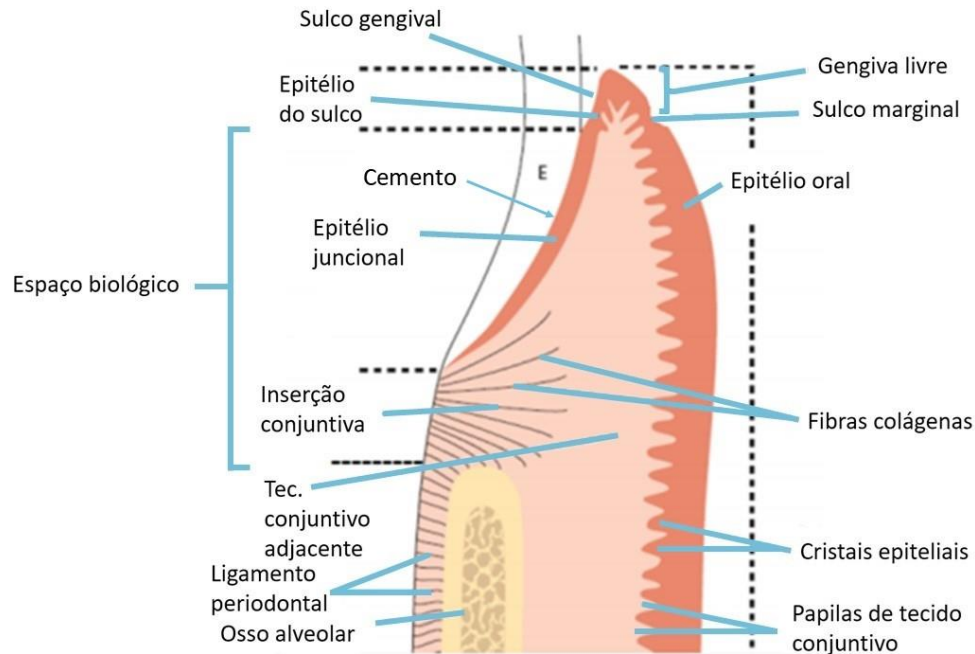
## 1.2 PERIODONTITE

Qualquer que seja a doença, hereditária ou adquirida, que afete os tecidos que envolvem e suportam os dentes (periodonto), pode ser definida como uma doença periodontal. Entretanto, esse termo se refere normalmente às doenças inflamatórias “gengivite” e “periodontite” (PIHLSTRON, MICHALOWICZ, JOHNSON, 2005), ambas causadas por microrganismos.

O periodonto (Figura 1) corresponde a um sistema de estruturas teciduais que promovem a sustentação do dente na cavidade oral otimizando a fixação e a manutenção da estrutura anatômica bucal. Seus componentes estruturais são: sulco gengival, epitélio juncional, gengiva, inserções conjuntivas, mucosa alveolar, osso alveolar, cimento e ligamento periodontal. A mucosa alveolar fica localizada após a gengiva, a qual, por sua vez, divide-se em gengiva inserida e gengiva livre, a primeira se encontra fortemente anexada ao dente enquanto a gengiva livre se apresenta menos fixada a estrutura dentária o que permite a formação de um sulco gengival onde há significativa atividade bacteriana. O epitélio juncional é uma estrutura localizada no ângulo formado pela gengiva inserida e o dente ao fim do sulco gengival, essa estrutura possui características próprias tais quais maior permeabilidade relativa em comparação com os tecidos epiteliais constitutivos do sulco gengival e da cavidade oral. As inserções conjuntivas, por sua vez, encontram-se logo abaixo do epitélio juncional seguidas dos ligamentos periodontais, os quais estão fixados no osso alveolar para garantir maior sustentação e imobilidade do dente durante as funções normais. Nesse contexto, o cimento é o tecido que se encontra entre o dente e o ligamento periodontal e atua tanto na fixação quanto na proteção do dente na cavidade oral (PONTES, 2016).

A periodontite é definida como uma doença em que ocorre bolsa periodontal de profundidade maior ou igual a 4 mm e a resposta inflamatória do hospedeiro está associada à perda óssea alveolar. Por outro lado, define-se gengivite como sendo uma doença periodontal não associada à perda de massa óssea, geralmente limitada a sulcos gengivais com profundidade menores ou iguais a 3 mm (HAJISHENGALLIS, 2015).

Figura 1. Anatomia do periodonto. Adaptado de Lindhe, Karring, Lang (2005).



A periodontite é classificada conforme o sistema definido no Workshop Internacional de Classificação de Condições e Doenças Periodontais, em 1999. Esse sistema de classificação é bastante detalhado e complexo. De forma geral, ele define a periodontite como crônica (localizada ou generalizada); agressiva (localizada ou generalizada); com manifestação de doença sistêmica; necrosante; abscessos do periodonto; e associada a lesões endodônticas (WIEBE, PUTNINS, 2000; ARMITAGE, 2004; HIGHFIELD, 2009; MITTAL et al. 2013).

A doença apresenta como resultado a destruição dos tecidos periodontais, sendo a causa mais comum de perda de dentes em todo o mundo (DARVEAU, 2010). De acordo com Caffesse (2015), diversos estudos apontam a periodontite como um fator de risco significativo para várias condições sistêmicas, incluindo: doenças cardiovasculares (BAHEKAR, 2007) doenças degenerativas como: artrite reumatoide (DE PABLO, DIETRICH, McALINDON, 2008); doenças pulmonares como: doença pulmonar obstrutiva crônica (PETER et al. 2013); efeitos adversos na gravidez como: baixo peso fetal, parto prematuro, aborto e/ou natimorto (SANZ, KORNMAN, 2013).

De acordo com um estudo longitudinal de base populacional realizado por Oppermann et al. (2015), foram executados no Brasil três inquéritos nacionais de saúde oral (em 1986, 2002-2003 e 2010). Em 1986 e entre 2002-2003, a prevalência de indivíduos com periodontite foi semelhante, sendo aproximadamente 10% de indivíduos entre 35-44 anos de

idade e 6% em indivíduos entre 65-74 anos. Entretanto, no inquérito de 2010, houve um aumento de 19,4% de indivíduos com periodontite nos grupos mencionados nos inquéritos anteriores. Caffesse (2015), por sua vez, afirma que o número de casos de periodontite é mais elevado em indivíduos adultos, com uma maior frequência no sexo masculino, fumantes, obesos e indivíduos de baixo nível socioeconômico.

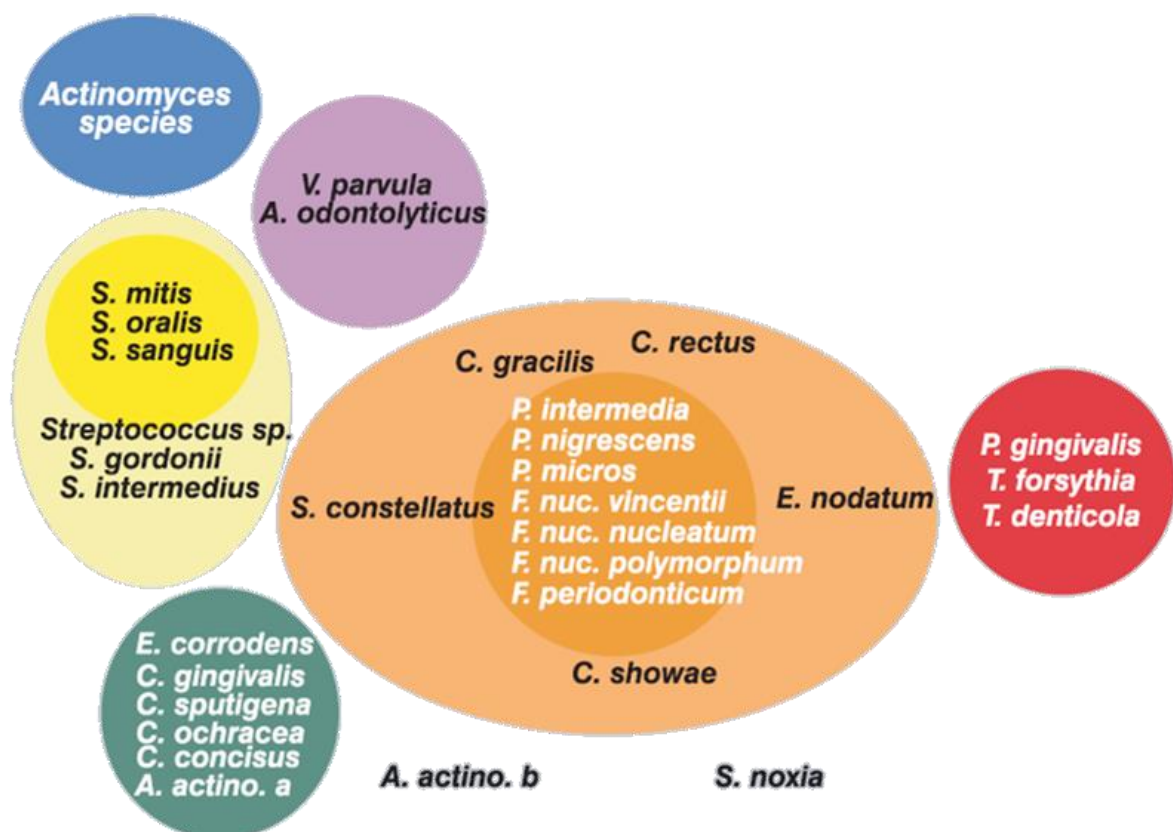
Por um período de tempo relativo, acreditou-se que os agentes etiológicos da periodontite fossem representados por organismos específicos, como: *Porphyromonas gingivalis*, *Treponema denticola* e *Tannarella forsythia*. Isso ocorreu devido à superestimação dos métodos de diagnósticos baseados em cultura, utilizados com o objetivo de definir os agentes etiológicos da doença. Pelo fato de grande parte dos microrganismos presentes na mucosa oral ser refratária ao cultivo, esse paradigma perdurou até os avanços de técnicas moleculares independentes de cultivo, as quais sugeriram que a etiologia da doença envolvia sinergia polimicrobiana e disbiose (HAJISHENGALLIS, 2014; DARVEAU, 2010). A natureza polimicrobiana da periodontite pode ser bem observada no estudo de identificação molecular de bactérias realizado por Sakamoto et al. (2003).

A microbiota oral contém uma gama de espécies bacterianas aeróbias e anaeróbias, tanto patogênicas quanto comensais. Nesse contexto, a periodontite pode ocorrer devido à capacidade desses microrganismos se agruparem em biofilmes, como comunidades microbianas complexas e interdependentes. Esses biofilmes se formam nas mais diversas regiões anatômicas da cavidade oral e, quando presentes sobre os dentes de forma densa e compacta, denomina-se placa dentária. A placa dentária pode permanecer em harmonia com o hospedeiro. Entretanto, quando presente na região inferior do dente (subgingival), a placa pode induzir um processo inflamatório importante, levando à periodontite. Isso ocorre principalmente por um aumento na quantidade de bactérias gram-negativas anaeróbias, muitas vezes causada por uma disbiose provocada por variações ambientais diversas (PIHLSTRON, MICHALOWICZ, JOHNSON, 2005; DARVEAU, 2010; HAJISHENGALLIS, 2015).

O trabalho realizado por Socransky, Haffajee (2005), é extensamente utilizado na literatura como referência no estudo da microbiota associada a doença periodontal. Socransky, Haffajee (2005) descreveu táxons microbianos que poderiam estar relacionados com a periodontite em qualquer grau. Além disso, descreveu complexos de microrganismos que poderiam estar desempenhando papéis específicos dentro da comunidade microbiana formadora da placa periodontal. Os complexos de Socransky, Haffajee (2005) são divididos em 6 cores distintas: azul, amarelo, verde, roxo, laranja e vermelho (Figura 2). Os complexos azul, amarelo, verde e roxo são capazes de se aderir a região gengival ou dental e estabelecer

relações entre si para a formação do biofilme. Enquanto isso, espécies presentes no complexo laranja são capazes de se aderir aos demais complexos atuando como colonizadores secundários. Estes por sua vez, são capazes de produzir um efeito patogênico. Contudo, um papel importante do complexo laranja para o estabelecimento da periodontite está na interação que o mesmo apresenta com o complexo vermelho (KOLENBRANDER et al. 2006). O complexo vermelho é composto por espécies que tem forte relação com o complexo laranja e são as principais espécies relacionadas a periodontite. Estas espécies, são especializadas na produção de fatores de virulência que levam a evasão do sistema imunológico, além de atuarem mais ativamente na destruição e penetração dos tecidos periodontais (LUX, 2001; YILMAZ, 2008).

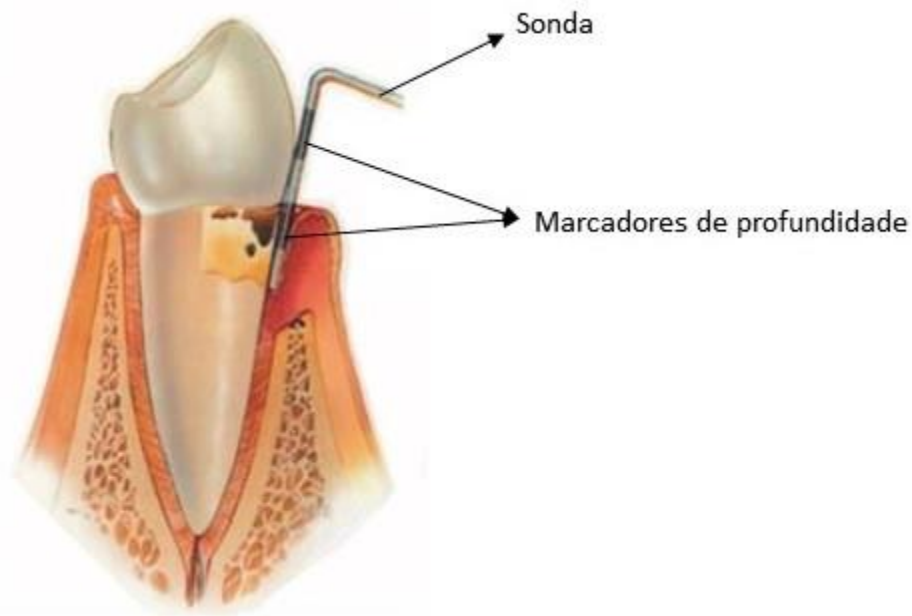
Figura 2. Complexos de Socransky&Haffajee. Adaptado de Socransky, Haffajee (2005).



O diagnóstico clínico da periodontite é realizado por um dentista através de técnicas convencionais de exame clínico, as quais incluem: inspeção visual dos tecidos gengivais; medição da profundidade da bolsa e perda de inserção periodontal; exame radiográfico; teste de mobilidade dentária; observação de sangramento à sondagem, sendo este último o mais

informativo quanto à presença da doença (ARMITAGE, 2004; GIANNOBILE et al. 2009). Além disso, embora raramente utilizada na prática clínica, também pode ser feito o uso de radiografia de subtração para detectar mudanças diminutas na quantidade de cálcio do osso alveolar (GIANNOBILE et al. 2009). O índice periodontal comunitário (CPI) é a forma mais rápida e simples de avaliar a saúde periodontal através do sangramento gengival ocasionado pela presença de bolsas periodontais. Durante a realização do exame, são utilizadas sondas periodontais marcadas. As sondas são inseridas em toda a extensão do sulco ou bolsa gengival do dente (Figura 3). Dessa forma, observa-se a profundidade do sulco gengival e a presença de sangramento (CHALUB, PÉRET, 2010).

Figura 3. Procedimento de sondagem para obtenção de índice periodontal comunitário (CPI). Adaptado de Fitzpatrick, Wijeyewickrema, Pike (2009).



Quanto ao uso de outras técnicas de diagnóstico complementares, esforços vêm sendo tomados na última década quanto à utilização de biomarcadores orais para medidas mais precisas da situação atual da doença, monitoramento do tratamento e indicadores de prognóstico. Esses estudos têm se focado no tecido ósseo e em proteínas do hospedeiro (enzimas e imunoglobulinas), células do hospedeiro, hormônios, bactérias e produtos bacterianos (RAM et al. 2015; KINNEY, RAMSEIER, GIANNBILE, 2007). Esses estudos têm principalmente a saliva como objeto de coleta, pois é de acesso não invasivo, abundante, além de conter mediadores locais e sistêmicos da doença (TABA et al. 2005; KINNEY, RAMSEIER, GIANNBILE, 2007; NABORS, McGLENNEN, THOMPSON, 2010; PATIL, PATIL, 2011; SHIMAZAKI et al. 2011; GUPTA, GOVILA, SAINI, 2015).



De acordo com Chapple et al. (2015), o combate à periodontite deve ser iniciado com a prevenção da formação da placa dentária por meio da escovação dos dentes, fazendo-se uso de dentífrico fluoretado duas vezes ao dia por pelo menos 2 minutos e limpeza inter-dental. Quando o paciente já se encontra com gengivite, além dos métodos de prevenção contra a formação da placa, ainda se recomenda o uso de adjuvantes químicos. É importante ressaltar que, para pacientes com periodontite já estabelecida, 2 minutos de escovação são insuficientes.

O tratamento da doença visa controlar principalmente o biofilme bacteriano. O tratamento não cirúrgico realizado por um dentista inicia-se tipicamente com a remoção da placa dentária por raspagem e aplainamento radicular através de vários instrumentos manuais ou motorizados. É dada uma maior atenção ao desbridamento do biofilme presente nas bolsas periodontais. Além da raspagem, recomenda-se o uso de quimioterápicos locais e antibióticos locais e sistêmicos. Espera-se que, com esse tratamento inicial, ocorra redução da inflamação e melhora na condição global do periodonto em um curto período de tempo. A partir de então, segue-se o tratamento com a higienização pessoal do paciente, com previsão de cura em cerca de 1 a 2 meses (PIHLSTRON, MICHALOWICZ, JOHNSON, 2005; LAUDENBACH, SIMON, 2014). Para pacientes com doença avançada, há diversos procedimentos cirúrgicos que podem ser aplicados como estratégias de tratamento, os quais costumam objetivar a redução das bolsas periodontais, ganho de acesso para o desbridamento do cálculo dentário e placa residual e estímulo à regeneração do suporte periodontal (PIHLSTRON, MICHALOWICZ, JOHNSON, 2005).

### 1.3 SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS)

Sanger, Nicklen e Coulson (1977) publicaram o método comumente conhecido como “sequenciamento de Sanger” que consiste em uma metodologia de sequenciamento baseada na terminação de cadeia. Em pouco tempo, essa metodologia passou a ser o modo de sequenciamento mais utilizado no período. O Projeto Genoma Humano (PGH), uma iniciativa internacional que objetivou o sequenciamento completo do genoma humano, fez uso do sequenciamento de Sanger, sendo que cerca de 13 anos foram necessários para que seu objetivo fosse atingido (NHGRI, 2016). Devido às limitações inerentes à técnica, quando aplicada em estudos de alta demanda, houve a necessidade de se buscarem novas estratégias que obtivessem resultados em menor tempo e custo.

Assim, novas plataformas de sequenciamento surgiram no mercado, a partir do ano

de 2005, denominadas conjuntamente de sequenciamento de Nova Geração (*Next Generation Sequencing*, NGS). Destacam-se, entre elas: 454 FLX (Roche), Solexa (Illumina), SOLiD (Applied Biosystems), Ion Torrent (Life Technologies) e Heliscope (tSMS). Essas plataformas apresentam como principais características o sequenciamento em grande escala, podendo gerar dados correspondentes a milhões de pares de nucleotídeos em uma única corrida, além de serem rápidas e de baixo custo em relação às tecnologias aplicadas anteriormente (MARDIS, 2011).

A técnica de sequenciamento cíclico em matriz está entre as mais utilizadas nos dias de hoje. Essa abordagem, de modo geral, faz uso de reações enzimáticas controladas sobre uma matriz sólida onde o DNA alvo está localizado. Essas reações acontecem em inúmeros ciclos em que cada um gera dados com a decodificação de poucos pares de bases da sequência alvo, com uma capacidade final de decodificação na escala de milhares de nucleotídeos em um curto período de tempo (SHENDURE, JI, 2008).

O mercado de sequenciadores permanece em constante crescimento e evolução. Atualmente, como exemplo, a empresa Illumina conta com diferentes sistemas de sequenciamento como: sistemas de escala de produção, NextSeq, HiSeq e NovaSeq 6000; e sistemas de bancada, iSeq 100, MiniSeq, MiSeq e NextSeq (ILLUMINA, 2018).

#### 1.4 METAGENÔMICA

A vida microbiana domina a Terra em diferentes aspectos e estima-se que  $5 \times 10^{30}$  células procarióticas habitam nosso planeta nos mais distintos ambientes possíveis (WHITMAN, COLEMAN, WIEBE, 1998). Tradicionalmente, o primeiro passo na investigação de processos microbianos consiste na obtenção de culturas puras. Entretanto, estima-se que as técnicas de culturas padrão consigam recuperar apenas 1% ou menos da diversidade microbiana presente em amostras ambientais. Com base nisso e no intuito de se compreender melhor a estrutura e processos microbianos envolvidos em amostras ambientais, um conjunto de técnicas independentes de cultivo é necessário para se buscar as espécies de organismos ainda não cultiváveis (RIESENFELD, SCHLOSS, HANDELSMAN, 2004).

Novas tecnologias de sequenciamento e a redução drástica no custo estão ajudando a superar esses limites. Atualmente é possível extrair informação genômica diretamente de comunidades microbianas em seus habitats naturais. De modo geral, ao invés de observar algumas espécies individualmente, pode-se agora estudar dezenas de milhares de espécies em conjunto. Devido ao fato de a informação ser obtida diretamente de genomas, tanto espécies

cultiváveis quanto não cultiváveis estarão presentes no conjunto de dados. Essa abordagem na qual os dados de genomas são tirados diretamente do ambiente foi denominada de “Metagenômica” (WOOLEY, GODZIK, FRIEDBERG, 2010). De acordo com Thomas, Gilbert e Meyer (2012), “Metagenômica” pode ser definida com a análise genética direta de genomas contidos em uma amostra ambiental.

Originalmente, a metagenômica baseava-se na extração de DNA total de amostras ambientais e posterior clonagem, seguida de sequenciamento. Com o surgimento das plataformas de sequenciamento de última geração e a evolução conjunta das ferramentas de análises de dados, os estudos de metagenômica ganharam maior robustez. Atualmente, as amostras ambientais não precisam necessariamente de clonagem e, portanto, após a extração do DNA, podem seguir diretamente ao sequenciamento, abordagem conhecida como “Shotgun”. Nessa abordagem, os dados gerados melhor representam toda a comunidade microbiana, possibilitando, por exemplo, a comparação entre comunidades de diferentes ambientes como evidenciado no trabalho de Quaiser et al. (2011).

As análises de dados metagenômicos é de fato um desafio devido ao enorme volume de dados gerados a partir de uma única amostra ambiental. Os dados gerados chegam facilmente a escala de centenas de gigabytes. Isso torna necessário um maquinário computacional robusto o suficiente para tratar os dados gerados além do avanço conjunto da bioinformática. As análises de cunho metagenômico ocorrem basicamente por meio de ferramentas computacionais, muitas das quais se encontram fortemente relacionadas a bancos de dados de domínio público, como: *National Center for Biotechnology Information* (NCBI, <http://www.ncbi.nlm.nih.gov/>) contendo *Reference Sequence Database* (RefSeq) e GenBank, *Kyoto Encyclopedia of Genes and Genomes* (KEGG, <http://www.genome.jp/kegg/>), *Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups* (EggNOG, <http://eggnogdb.embl.de/#/app/home>) e *Carbohydrate-Active Enzyme* (CAZy, <http://www.cazy.org/>) (QUAISER et al. 2011; BIKEL et al. 2015; THOMAS, GILBERT, MEYER, 2012).

Através de ferramentas de bioinformática aplicadas sobre dados metagenômicos, é possível realizar diferentes abordagens, dentre as quais se destacam: anotação taxonômica, estudo da diversidade microbiana e anotação funcional, ou determinação da diversidade funcional gênica (WOOLEY, GODZIK, FRIEDBERG, 2010; BIKEL et al. 2015). Entretanto, por necessitar de alto poder computacional ainda é dificultosa a análise de dados desse tipo. Desta forma, estão ganhando cada vez mais espaço neste tipo de análise servidores públicos que disponibilizam de poder computacional suficiente, a exemplo: MG-RAST. Esse servidor,

por sua vez, é capaz de receber (via *upload*) e processar dados metagenômicos em uma sequência programada de análises, denominada *pipeline*, contra diversos bancos de dados, desde análises taxonômicas até funcionais, facilitando dessa forma pesquisadores e até mesmo institutos que não possuem infraestrutura básica para a realização de tais análises (WILKE, 2015). Além disso, o desenvolvimento de novos programas que objetivam processar grandes volumes de dados sem usar de muito poder computacional vem aumentando expectativas no rumo da metagenômica. É possível, por exemplo, realizar montagem de sequências, perfilhamento taxonômico e predição gênica com os softwares PEAR, Metaphlan 2.0 e MetaGeneMark, respectivamente, sem necessariamente de fazer uso de um alto poder computacional (ZHANG et al. 2014; TRONG et al. 2015; ZHU, LOMSADZE, BORODOVSKY, 2010).

## **2 JUSTIFICATIVA**

Apesar do grande número de estudos já conduzidos sobre a periodontite, os fatores microbianos envolvidos no desenvolvimento desta doença ainda necessitam de investigações mais aprofundadas. Tal fato ganha maior relevância quando se considera que uma parcela significativa das bactérias orais não é passível de cultivo em laboratório. Nesse sentido, a análise metagenômica da microbiota oral por técnica de sequenciamento de última geração (NGS) pode proporcionar uma visão mais abrangente do microambiente oral, auxiliando na determinação de genes de origem bacteriana potencialmente envolvidos na doença periodontal, bem como na identificação de espécies cultiváveis e não-cultiváveis. Os resultados do estudo podem, em última instância, contribuir para um melhor entendimento dos mecanismos etiológicos da periodontite, bem como para a definição de marcadores moleculares para estudos futuros.

### 3 OBJETIVOS

#### 3.1 OBJETIVOS GERAIS

- Estabelecer a estrutura taxonômica e processos funcionais envolvidos na periodontite, com base na análise metagenômica da saliva de portadores de periodontite e voluntários saudáveis.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Realizar a identificação taxonômica dos táxons presentes em cada amostra.
- Estabelecer a abundância para cada táxon encontrado.
- Conduzir a anotação funcional em diferentes bancos de dados de proteínas para cada amostra.
- Correlacionar os resultados obtidos tanto de taxonomia quanto funcional por grupo.
- Identificar unidades taxonômicas, genes e/ou vias metabólicas potencialmente associadas à suscetibilidade ou proteção à periodontite.

## 4 MATERIAL E MÉTODOS

### 4.1 CONSIDERAÇÕES ÉTICAS

Este projeto foi devidamente avaliado e aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da Universidade do Estado do Amazonas (UEA), de acordo com as normas estabelecidas pelo SISNEP para pesquisas em indivíduos humanos sob registro C.A.A.E. 19521113.0.0000.5016 (Anexo A).

### 4.2 POPULAÇÃO DE ESTUDO

A população de estudo foi composta por indivíduos selecionados dentre os pacientes previamente indicados ao tratamento odontológico na Policlínica Odontológica da Universidade do Estado do Amazonas – UEA, Manaus, Amazonas.

### 4.3 CRITÉRIOS DE INCLUSÃO

Foram incluídos indivíduos de ambos os gêneros, com idade mínima de 18 anos. Indivíduos que receberam terapia com antibióticos em um período de 6 meses previamente à coleta não puderam participar do estudo.

### 4.4 SELEÇÃO DE PACIENTES

Todos os pacientes foram submetidos à anamnese, durante a qual foram colhidas informações sobre a história médica e odontológica, exame clínico e radiográfico. Os seguintes parâmetros clínicos foram observados e anotados para cada elemento dentário: índice de placa, nível clínico de inserção, profundidade de sondagem, sangramento à sondagem, sangramento gengival, presença de exsudato purulento e perda óssea radiográfica. Medidas de profundidade de sondagem e nível clínico de inserção foram tomadas das 6 faces (vestibular, lingual, méso-vestibular, disto-vestibular, méso-lingual e disto-lingual) de todos os elementos dentários, exceto dos terceiros molares. Foram utilizados dois grupos de estudo: Grupo I, representado por 14 pacientes portadores periodontite; e Grupo II, representado por 13 pacientes saudáveis. Como critério de classificação para o Grupo I, foram considerados pacientes apresentando um mínimo de 12 elementos dentários e 4 dentes com pelo menos 1

sítio com perda de inserção  $\geq 4$ mm. Como critério de classificação para o Grupo II, foram utilizados pacientes apresentando um mínimo de 20 elementos dentários e, no máximo, 4 sítios com perda de inserção e profundidade de sondagem de até 5mm (excluindo-se terceiros molares).

#### 4.5 COLETA

Amostras de saliva não estimulada foram coletadas em tubos Omnigene OM-505 Microbial DNA (Oragene) para conservação de ácidos nucleicos. Os tubos foram devidamente identificados, acondicionados em recipiente refrigerado e imediatamente transportados ao laboratório de Genética de Microrganismos da Escola Superior de Ciências da Saúde, ESA – UEA, sendo acondicionados a  $-20^{\circ}\text{C}$  até uso futuro.

#### 4.6 EXTRAÇÃO DE DNA TOTAL E QUANTIFICAÇÃO

No intuito de extrair o DNA genômico de bactérias gram-negativas e gram-positivas, foi utilizado o kit *MasterPure™ Complete DNA and RNA Purification* (Epicentre) associado às enzimas lisozima e proteinase K. Após o processo de extração, o DNA foi analisado em fluorímetro (Qubit 3.0, Invitrogen) para determinação da quantidade do DNA extraído.

#### 4.7 CONSTRUÇÃO DA BIBLIOTECA E SEQUENCIAMENTO

As amostras com concentração de DNA  $>20\text{ng}/\mu\text{L}$  foram enviadas para Instituto Novogene (Hong Kong) para a construção da biblioteca *shotgun* e sequenciamento. De acordo com o protocolo estabelecido pelo Instituto Novogene, para dar seguimento ao sequenciamento, as amostras foram checadas quanto: o grau de degradação do DNA e a contaminação potencial em géis de agarose a 1%; a pureza do DNA (OD260 / OD280, OD260 / OD230) usando o espectrofotômetro *NanoPhotometer®* (IMPLEN, CA, EUA), com valor de OD aceito entre 1,8~2,0; e concentração de DNA utilizando o kit *Qubit® dsDNA Assay* em Qubit® 2.0 *Flurometer* (Life Technologies, CA, EUA), com valor de concentração aceito  $>20\text{ng}/\mu\text{L}$ .

Uma total de  $1\mu\text{g}$  de DNA por amostra foi usado como material de entrada para a preparação das bibliotecas de sequenciamento. As bibliotecas de sequenciamento foram geradas usando o *NEBNext® Ultra™ DNA Library Prep Kit* para a Illumina (NEB, EUA),



seguindo as recomendações do fabricante e códigos de índice foram adicionados às sequências para cada amostra. A amostra de DNA foi fragmentada por sonicação até um tamanho de 300 pb. Em seguida, os fragmentos de DNA foram clivados em sua extremidade (cauda A) e ligados com adaptador seguido de amplificação por PCR. Por fim, os produtos de PCR foram purificados (sistema AMPure XP) e as bibliotecas foram analisadas quanto à distribuição de tamanho pelo *Agilent2100 Bioanalyzer*. A quantificação ocorreu por PCR em tempo real. A geração de *clusters* nas amostras codificadas por índice foi realizada em sistema cBot de acordo com as instruções do fabricante. Após a geração do cluster, as preparações da biblioteca foram sequenciadas em plataforma Illumina HiSeq 2500 (Illumina, EUA) com estratégia *paired-end*. Após o sequenciamento, todos os arquivos resultantes foram convertidos para o formato “.FASTQ” para análise da qualidade.

#### 4.8 PRÉ-PROCESSAMENTO DOS DADOS E EMPARELHAMENTO DE *READS*

Nesta etapa, os dados foram filtrados inicialmente com a remoção de sequências de adaptadores utilizadas durante o sequenciamento. Em seguida, foram removidas sequências cuja quantidade de nucleotídeos com  $Q \leq 38$  ultrapassou 40 pb, sequências cuja quantidade de nucleotídeos indeterminados (N) ultrapassou 10 pb e sequências que correspondiam a DNA humano. Por se tratar de uma estratégia de sequenciamento *paired-end*, o resultado gera dois arquivos para cada amostra, onde cada arquivo contém uma das direções de um dado fragmento sequenciado. Por tanto, optou-se por realizar o emparelhamento dos fragmentos, no intuito de garantir melhor qualidade nas demais análises. O emparelhamento das sequências ocorreu através do programa PEAR v. 0.9.6 com os parâmetros padrões (ZHANG et al. 2014). Após o emparelhamento, uniu-se, para cada amostra, o arquivo contendo *reads* emparelhadas junto ao arquivo contendo *reads* não emparelhadas de modo a não se perder informação durante o processo.

#### 4.9 ANOTAÇÃO TAXONÔMICA VIA METAPHLAN 2.0

A primeira etapa de anotação taxonômica ocorreu através do programa Metaphlan 2.0 associado ao programa Bowtie2 (LANGMEAD, SALZBERG, 2012). Esta ferramenta computacional é capaz de perfilar a composição de comunidades microbianas (Bactérias, Arqueas, Eucariotos e Vírus) a partir de dados provenientes de sequenciamento *shotgun* com resolução ao nível de espécie (TRUONG et al. 2015). Os arquivos gerados pelo

emparelhamento das *reads* foram analisados pelo programa Metaphlan 2.0, utilizando todo o banco de dados e parâmetros *default*. Os resultados foram gerados na forma de tabelas de espécies contendo suas respectivas abundâncias relativas para cada amostra. Os dados foram então utilizados na construção de tabelas de abundâncias absoluta e relativa contendo os táxons encontrados descritos do nível de domínio até espécie. Além disso, foram criados *heatmaps* dos melhores resultados ranqueados ao nível de gênero e espécie, de acordo com o protocolo do programa. Para isso, utilizou-se um programa intrínseco ao Metaphlan 2.0 denominado “metaphlan\_hclust\_heatmap.py”, com índice de similaridade de Bray-Curtis, o qual possibilita o agrupamento de amostras com perfis taxonômicos próximos entre si e também o agrupamento táxons que apresentam perfis de abundâncias próximas entre si.

#### 4.10 ANOTAÇÃO TAXONÔMICA VIA MG-RAST

Os arquivos resultantes do processo de emparelhamento para cada amostra foram submetidos a anotação através do servidor *Metagenomics Rapid Annotation* (MG-RAST, <<http://metagenomics.anl.gov/>>) versão 4.0.3 (WILKE et al. 2015). Os parâmetros iniciais para a submissão incluíram: remoção de DNA humano presente nos metagenoma (*H. sapiens*, NCBI v36); remoção de sequências com phred<30. Os perfis taxonômicos, de cada amostra, gerados por meio do MG-RAST, foram provenientes do banco de dados RefSeq. Para identificar resultados positivos durante a busca no banco de dados utilizou-se como parâmetros BlastX com um alinhamento mínimo de 50 pb, identidade de 90% e E-value de corte em  $E < 1 \times 10^{-5}$ . Como resultado, foram obtidas tabelas com valores absolutos para cada táxon encontrado. Com base na tabela de valores de abundância absolutos, construiu-se uma tabela de frequência normalizada contendo cada táxon encontrado para cada metagenoma. A metodologia empregada para a normalização da abundância absoluta foi a divisão da abundância absoluta obtida para um dado táxon pela soma total da abundância de todos os táxons para cada amostra individualmente. Após o sequenciamento, é comum que as amostras apresentem quantidades diferentes de sequências geradas, portanto, o processo de normalização dos dados procura evitar que essas diferenças quantitativas tenham influência na interpretação dos resultados.

#### 4.11 RELAÇÕES FILOGENÉTICAS, RIQUEZA E DIVERSIDADE

Para esta etapa foram utilizados os dados de abundâncias ao nível de espécie gerados pelo programa Metaphlan 2.0. Os cladogramas contendo as representações filogenéticas de cada grupo de estudo foram gerados utilizando o programa GraPhlAn (ASNICAR et al. 2015). Para visualização e comparação da riqueza entre os grupos foi construído um diagrama de Venn utilizando o programa FunRich v3.1.3 (PATHAN et al. 2015). Além disso, foram geradas curvas de rarefação utilizando o método Mao's Tau, através do programa PAST3 (HAMMER, DAVID, PAUL, 2001), para cada grupo de estudo. Com uso do programa EstimateS v9.1.0 (COLWELL, 2019) foram calculados os estimadores de riqueza Chao 2, Jackknife (1 e 2) e Bootstrap. Para a análise da diversidade alfa foram calculados os índices de Shannon-Wiener ( $H'$ ) e Simpson (1-D) para cada indivíduo envolvido na pesquisa. A diversidade beta foi avaliada de acordo com o índice de Whittaker (WHITTAKER, 1960). Todos os resultados de índices de diversidade mencionados foram obtidos através do programa PAST3 (HAMMER, DAVID, PAUL, 2001).

#### 4.12 ANOTAÇÃO FUNCIONAL VIA KEGG

Os dados dispostos no servidor MG-RAST foram anotados no banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Durante a anotação, utilizou-se como medidas de corte um e-value  $\leq 1e-5$  e identidade de 90%. Após a notação, foi gerada uma tabela geral de abundância absoluta contendo categorias KEGG encontradas em níveis hierárquicos distintos, sendo eles: nível 1, nível 2, nível 3 e função. Este último corresponde ao registro KO (KEGG Orthology). Os KO, assim como os demais níveis hierárquicos, foram utilizados na análise funcional e/ou obtenção de vias metabólicas envolvidas nos diferentes metagenomas. As tabelas de abundância absoluta foram convertidas em abundância relativa com o intuito preparar os dados para análise estatística. A metodologia aplicada durante a conversão foi a divisão do valor de cada *hit* obtido de uma dada amostra pelo total de *hits* da mesma.

#### 4.13 ANOTAÇÃO FUNCIONAL VIA CAZY

Uma vez que o servidor MG-RAST executa um *pipeline* com diferentes etapas durante a análise dos dados submetidos, porém não contempla análise funcional em banco de dados

*Carbohydrate-Active Enzyme* (CAZy), foram extraídos os arquivos que continham somente sequências de DNA preditas como ORFs (*Open Read Frames*) para cada amostra. Em seguida, utilizou-se essas sequências contra o banco de dados CAZy (CAZyDB.07202017.fa, atualizado em 20/07/2017, disponível em <http://csbl.bmb.uga.edu/>) fazendo uso da ferramenta BLASTP disposta no programa DIAMOND v.0.9.22 (BUCHFINK, XIE, HUSON, 2015). Os parâmetros utilizados durante o BLASTP foram: e-value  $\leq 1e-5$ ; e identidade de 90%. O uso do banco de dados CAZy, teve como intuito obter informações quanto ao metabolismo de carboidratos presentes nos metagenomas. Ao término da análise, os resultados gerados foram dispostos em tabelas de abundância absoluta contendo família e subfamílias CAZy para cada amostra. Essas tabelas foram convertidas em tabelas de abundância relativa. Utilizou-se como metodologia de conversão a divisão do valor de cada *hit* obtido de uma dada amostra pelo total de *hits* da mesma.

#### 4.14 ANÁLISE ESTATÍSTICA

Os dados de abundância relativa obtidos via Metaphlan 2.0 foram convertidos em formato compatível (.spf) e analisados através do programa *Statistical Analysis of Taxonomic and Functional Profiles* (STAMP). O STAMP corresponde a um pacote de programas para a análise de perfis metagenômicos tendo como foco tanto a anotação taxonômica quanto a anotação funcional (PARKS et al. 2014). A conversão das tabelas obtidas do Metaphlan 2.0 para o formato compatível com STAMP (.spf) foi realizada através do script em linguagem de programação Perl denominado “metaphlan\_to\_stamp.pl”, desenvolvido por Morgan Langille – obtido em <https://github.com/LangilleLab>.

Os demais passos seguem o mesmo padrão de análise tanto para os dados provenientes do programa Metaphlan 2.0 quanto para anotação taxonômica pelo servidor MG-RAST e análise funcional através dos bancos de dados KEGG e CAZy, sempre partindo das tabelas de abundância relativa. Inicialmente, no programa STAMP, foi realizada a Análise de Componentes Principais (PCA) buscando observar a variância entre as amostras de ambos os grupos. Em seguida, para verificar a diferença significativa de cada variável, aplicou-se o teste estatístico não paramétrico Kruskal-Wallis com  $p < 0,05$ . Com base nos resultados, gráficos foram construídos no intuito de gerar uma melhor interpretação dos resultados.

Uma vez que a resolução dos resultados do Metaphlan 2.0 ocorre ao nível de espécie, realizou-se análise de correlação entre as espécies obtidas para esses dados. Para isso, foram calculados todos os possíveis coeficientes de correlação de Spearman. Foram considerados

para isso apenas aquelas espécies que obtiveram diferença estatística significativa ( $p < 0,05$ ) e correlação forte ( $r < -0,7$  ou  $r > +0,7$ ). A análise de correlação de Spearman foi performada através do programa R v. 3.4.3, pacote “*R Stats*”, função “cor”.

Ainda sobre os dados de abundâncias ao nível de espécie gerados pelo Metaphlan 2.0, usando a plataforma Galaxy/Hutlab (<http://huttenhower.sph.harvard.edu/galaxy/>), foi realizada Análise de Discriminantes Lineares (LDA) com  $p < 0,05$  tanto para o teste fatorial Kruskal-Wallis entre classes, quanto para o teste de Wilcoxon entre subclasses. A análise foi realizada tanto nos dados taxonômicos de gênero quanto espécie, quanto funcionais a nível de gene.

Quanto a análise estatística dos dados de diversidades gerados, temos que, os índices de diversidade de Shannon-Wiener ( $H'$ ) e Simpson (1-D) de cada indivíduo foram agrupados de acordo com os grupos de estudo. Em seguida, foi aplicado o teste estatístico de Kruskal-Wallis com  $p < 0,05$  para checar se houve diferença significativa entre os grupos. Para o uso do índice de Whittaker, foi aferido a diversidade beta de cada grupo separadamente com base nos dados de abundância de espécies gerados pelo Metaphlan 2.0.

## 5 RESULTADOS

### 5.1 EXTRAÇÃO DE DNA TOTAL E QUANTIFICAÇÃO

A extração de DNA total das amostras demonstrou resultados bastante satisfatórios. A quantificação de DNA total revelou que todas as amostras tiveram rendimento superior a 20 ng/ $\mu$ L. Isso permitiu que todo o grupo amostral pudesse seguir para a etapa de sequenciamento. Os resultados individuais da quantificação do DNA total das amostras encontram-se descritos na Tabela 1.

Tabela 1. Quantificação de DNA total, realizada em Qubit 3.0, para cada amostra envolvida no estudo.

Amostra	Condição	DNA (ng/ $\mu$ L)
1	Saudável	112,6
2	Saudável	178,7
3	Saudável	171,6
4	Saudável	86,5
5	Saudável	211,2
6	Saudável	47
7	Saudável	93,9
8	Saudável	163
9	Saudável	39,5
10	Saudável	530,8
11	Saudável	59,7
12	Saudável	49,1
13	Saudável	57,8
14P	Doente	242,5
15P	Doente	281,5
16P	Doente	330,2
17P	Doente	169,8
18P	Doente	796
19P	Doente	207,6
20P	Doente	167,6
21P	Doente	228,4
22P	Doente	762,4
23P	Doente	124,2
24P	Doente	132,2
25P	Doente	167,3
26P	Doente	3661,9
27P	Doente	164,2

## 5.2 PRÉ-PROCESSAMENTO DOS DADOS

No total, 75.304,32 Mpb de dados brutos foram obtidos após o sequenciamento (2.789,05 MB por amostra, em média), enquanto que 74.430,59 Mpb de dados limpos permaneceram após o controle de qualidade, resultando em uma média de 2.756,69 Mpb por amostra, correspondendo a uma taxa de eficiência (dados limpos sobre dados brutos) de 98,84%. Após a remoção de sequências relacionadas a DNA humano restaram 35.770,13 MB (104.004.730 sequências, 15.761,62 Mbp) de dados com uma média de 1.324,81 MB (3.852.027,04 sequências, 583,76 Mbp) por amostra (Tabela 2)

Tabela 2. Número de sequências e Mpb resultantes em cada amostra envolvida no estudo após o pré-processamento dos dados.

Amostras	Megabytes	Sequências	Mpb	Tamanho médio (pb)
1	3.062.224	8.695.195	1.304,28	150
2	1.462.749	4.153.480	623,02	150
3	1.059.271	3.007.863	451,18	150
4	1.778.158	5.049.354	757,4	150
5	1.391.929	3.952.422	592,86	150
6	1.239.622	3.520.145	528,02	150
7	2.796.121	7.939.561	1.190,93	150
8	2.047.972	5.815.265	872,29	150
9	1.483.121	4.211.353	792,62	188
10	2.023.390	5.745.540	861,83	150
11	1.803.648	5.121.756	768,26	150
12	745.283	2.116.385	317,46	150
13	1.446.388	4.107.008	616,05	150
14P	1.193.067	3.387.744	508,16	150
15P	684.591	1.943.928	291,59	150
16P	934.632	2.653.911	398,09	150
17P	2.923.334	8.300.854	1.245,13	150
18P	466.516	1.324.687	198,7	150
19P	1.193.626	3.389.329	508,4	150
20P	1.871.169	5.313.239	796,99	150
21P	473.615	1.344.845	201,73	150
22P	284.815	808.758	121,31	150
23P	756.162	2.147.133	322,07	150
24P	816.606	2.314.931	347,24	150
25P	1.743.882	4.951.778	742,77	150
26P	235.311	668.183	100,23	150
27P	711.413	2.020.083	303,01	150

### 5.3 EMPARELHAMENDO DE *READS*

Um total de 35.720.993 *reads* (ou sequências) foram emparelhadas, com uma média de 1.322.999,74 por amostra, representando um total de 8.703.238,65 Mpb com uma média de 322.342,17 Mpb por amostra. O tamanho da menor sequência emparelhada foi de 50 pb enquanto a maior foi 290 pb. O tamanho médio das sequências emparelhadas variou entre 161 pb a 260,9 pb sendo que a maioria das amostras apresentou média > 200 pb (Tabela 3).

Tabela 3. Quantidade de sequências emparelhadas, pares de bases e seus respectivos tamanhos médios, para cada amostra envolvida no estudo, gerados durante a etapa de montagem. O símbolo “\*” indica as amostras que apresentaram apenas um arquivo contendo *reads* durante o sequenciamento.

Amostras	Sequências emparelhadas	Mpb	Tamanho médio (pb)
1*	0	1.304,28	150
2	511.341	106,81	209
3	980.438	240,86	246
4	2.679.045	564,99	211
5	454.852	90,4	199
6	2.551.726	532,79	209
7	5.766.033	1.477,97	256
8	1.052.661	257,88	245
9	2.718.484	568,69	209
10	445.383	98,6	221
11	3.896.568	797,59	205
12	1.591.082	323	203
13	3.635.293	894,77	246
14P	367.254	71,57	195
15P	453.771	79,58	175
16P	567.720	122,39	216
17P	3.036.046	792,28	261
18P	295.126	48,07	163
19P	499.583	106,61	213
20P	909.007	218,2	240
21P	321.685	52,62	164
22P	321.789	51,78	161
23P	565.895	114,27	202
24P	459.670	84,9	185
25P	933.431	222,95	239
26P	297.404	48,67	164
27P	409.706	75,43	184



#### 5.4 ANOTAÇÃO TAXONÔMICA VIA METAPHLAN 2.0

Um total de 104.004.730 de sequências foi submetido ao programa Metaphlan 2.0. Dessas sequências, 3.267.049 (3,23%) foram anotadas (Tabela 4).

Tabela 4. Número de sequências submetidas e anotadas, para cada amostra envolvida no estudo, durante a previsão taxonômica através do programa Metaphlan 2.0.

Amostras	Sequências		
	Submetidas	Anotadas	% Anotadas
1	8.695.195	212.368	2,44%
2	4.153.480	142.654	3,43%
3	3.007.863	85.913	2,86%
4	5.049.354	150.512	2,98%
5	3.952.422	142.622	3,61%
6	3.520.145	100.655	2,86%
7	7.939.561	409.592	5,16%
8	5.815.265	207.975	3,58%
9	4.211.353	128.591	3,05%
10	5.745.540	217.371	3,78%
11	5.121.756	161.550	3,15%
12	2.116.385	60.513	2,86%
13	4.107.008	123.609	3,01%
14P	3.387.744	116.355	3,43%
15P	1.943.928	52.010	2,68%
16P	2.653.911	84.230	3,17%
17P	8.300.854	258.986	3,12%
18P	1.324.687	28.749	2,17%
19P	3.389.329	108.981	3,22%
20P	5.313.239	150.418	2,83%
21P	1.344.845	27.002	2,01%
22P	808.758	13.475	1,67%
23P	2.147.133	51.398	2,39%
24P	2.314.931	69.514	3,00%
25P	4.951.778	189.420	3,83%
26P	668.183	9.946	1,49%
27P	2.020.083	50.412	2,50%

Os resultados obtidos mostram a presença dos domínios *Bacteria* (99,5%) e *Viruses* (0,5%). O filo *Firmicutes* representou a maior abundância nos resultados com 29,52% das sequências, seguido pelos filios *Actinobacteria* (27,83%), *Bacteroidetes* (20,43%) e *Proteobacteria* (17,94%), os demais filios podem ser visualizados na Tabela 5.

Tabela 5. Abundância relativa geral, de todos os filios encontrados, para cada amostra envolvida no estudo, obtida durante a previsão taxonômica através do programa Metaphlan 2.0.

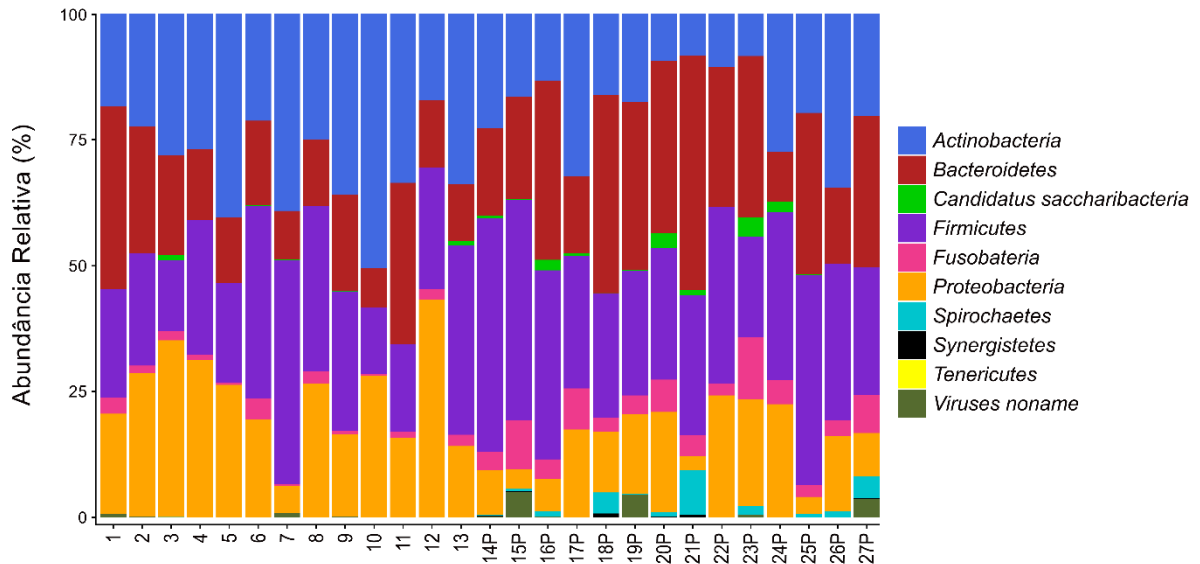
Domínio	Filo	% Geral
<i>Bacteria</i>	<i>Firmicutes</i>	29,52
<i>Bacteria</i>	<i>Actinobacteria</i>	27,83
<i>Bacteria</i>	<i>Bacteroidetes</i>	20,43
<i>Bacteria</i>	<i>Proteobacteria</i>	17,94
<i>Bacteria</i>	<i>Fusobacteria</i>	2,97
<i>Viruses</i>	<i>Viruses noname</i>	0,51
<i>Bacteria</i>	<i>Candidatus Saccharibacteria</i>	0,45
<i>Bacteria</i>	<i>Spirochaetes</i>	0,33
<i>Bacteria</i>	<i>Synergistetes</i>	0,03
<i>Bacteria</i>	<i>Tenericutes</i>	0,00079

Diferenças são encontradas quando se observa os dados para cada grupo de estudo separadamente. O resultado relacionado a presença de vírus ao nível de filo apresentou maiores abundâncias no grupo composto por portadores de periodontite em comparação ao grupo de pacientes saudáveis, porém não houve diferença estatística significativa. Além disso, ao observamos os filios bacterianos *Firmicutes*, *Bacteroidetes*, *Fusobacteria*, *Spirochaetes*, *Candidatus Saccharibacteria* e *Synergistetes* também notamos uma maior presença no grupo de portadores. Enquanto que os filios *Actinobacteria* e *Proteobacteria* estão mais presentes no grupo saudável (Tabela 6, Figura 4).

Tabela 6. Abundância relativa, de todos os filios encontrados, para cada grupo de estudo, obtida durante a previsão taxonômica através do programa Metaphlan 2.0. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ )

Reino	Filo	%Doentes	%Saudáveis	p-value
<i>Bacteria</i>	<i>Actinobacteria</i> *	20,24	32,12	3,60E-03
<i>Bacteria</i>	<i>Bacteroidetes</i> *	26,13	17,22	1,74E-02
<i>Bacteria</i>	<i>Candidatus Saccharibacteria</i> *	1,01	0,13	1,12E-02
<i>Bacteria</i>	<i>Firmicutes</i>	32,24	27,98	1,09E-01
<i>Bacteria</i>	<i>Fusobacteria</i> *	5,63	1,46	1,54E-04
<i>Bacteria</i>	<i>Proteobacteria</i> *	12,89	20,79	7,61E-03
<i>Bacteria</i>	<i>Spirochaetes</i> *	0,90	0,01	1,13E-04
<i>Bacteria</i>	<i>Synergistetes</i> *	0,09	0,00	7,17E-04
<i>Bacteria</i>	<i>Tenericutes</i>	0,00	0,0012	2,99E-01
<i>Viruses</i>	<i>Viruses noname</i>	0,88	0,30	5,09E-01

Figura 4. Perfil taxonômico de todas as mostras envolvidas no estudo. Os perfis foram construídos através da abundância relativa dos principais filos obtidos durante a previsão taxonômica pelo programa Metaphlan 2.0.



Os únicos filos que não apresentaram diferença significativa ( $p < 0,05$ ) para teste estatístico não paramétrico Kruskal-Wallis foram: *Firmicutes*, *Viruses noname* e *Tenericutes*. Os demais filos com diferença significativa podem ser visualizados da Figura 5 a Figura 11.

Figura 5. *Boxplot* demonstrando a distribuição na proporção do filo *Actinobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

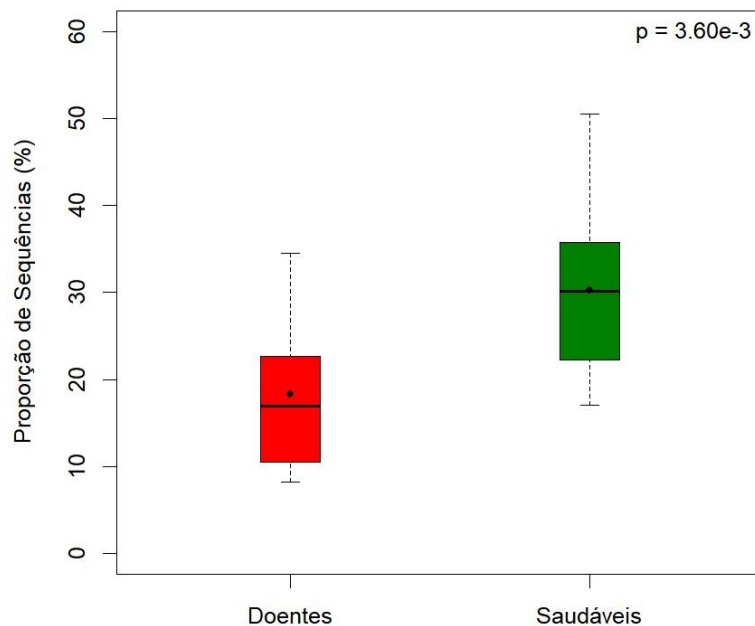


Figura 6. *Boxplot* demonstrando a distribuição na proporção do filo *Bacteroidetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

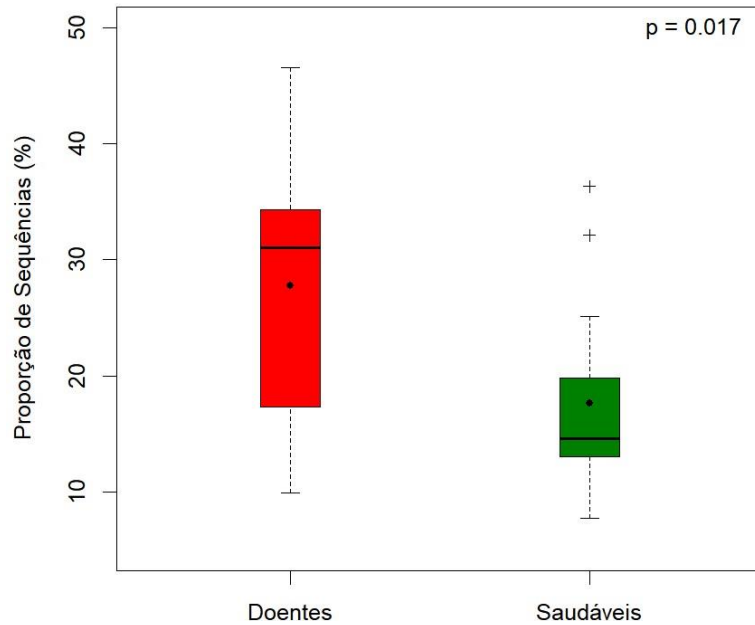


Figura 7. *Boxplot* demonstrando a distribuição na proporção do filo *Candidatus Saccharibacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos da previsão taxonômica pelo programa Metaphlan 2.0.

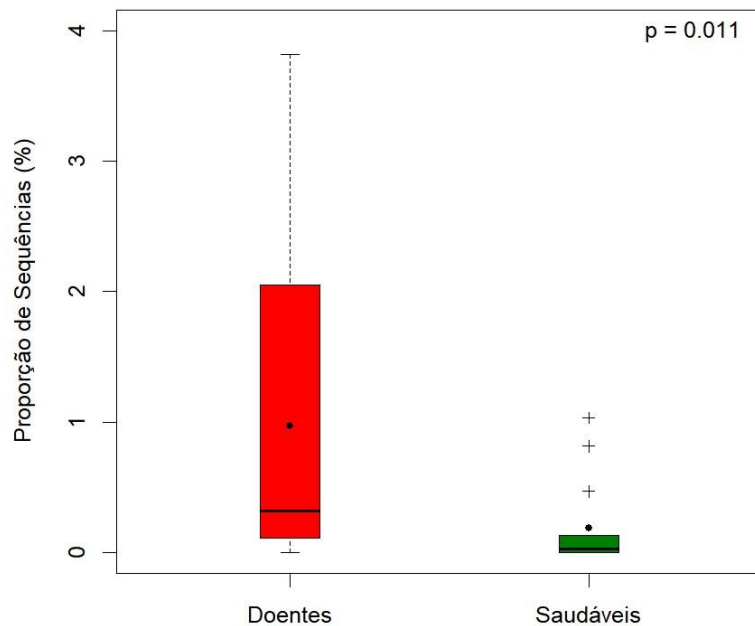


Figura 8. *Boxplot* demonstrando a distribuição na proporção do filo *Fusobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

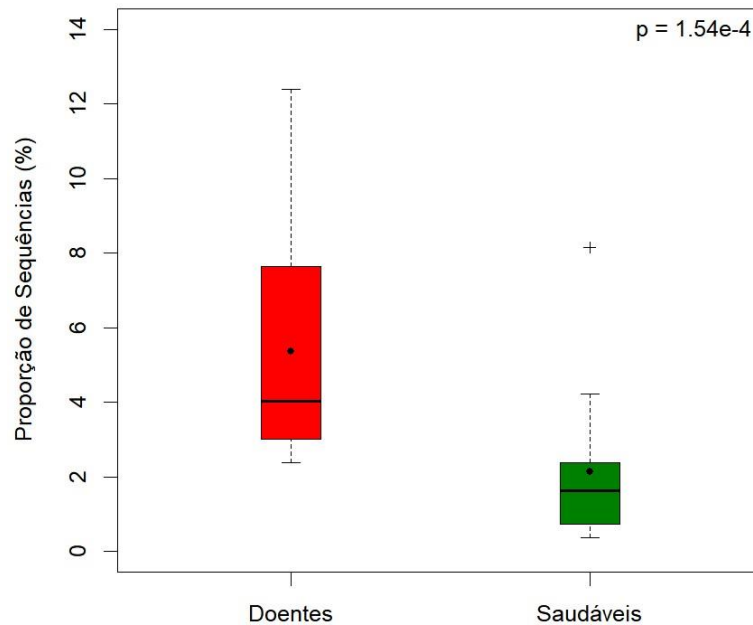


Figura 9. *Boxplot* demonstrando a distribuição na proporção do filo *Proteobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

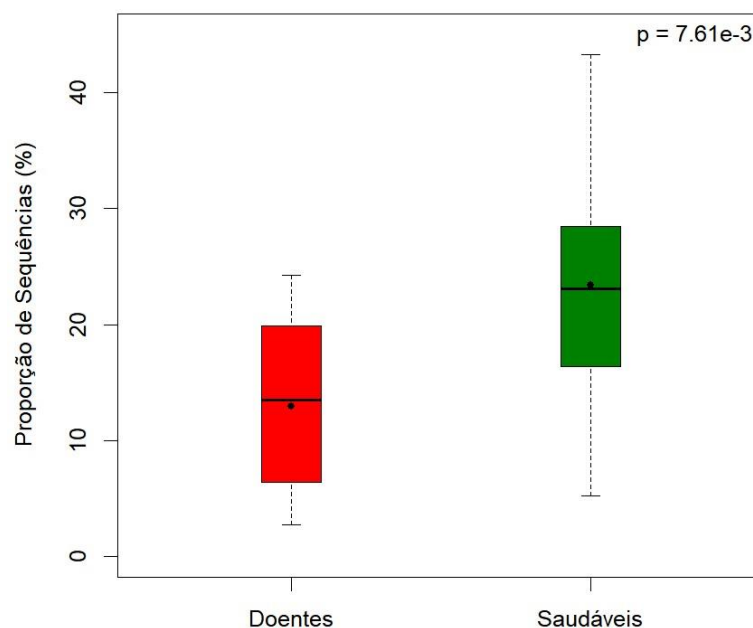


Figura 10. *Boxplot* demonstrando a distribuição na proporção do filo *Spirochaetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

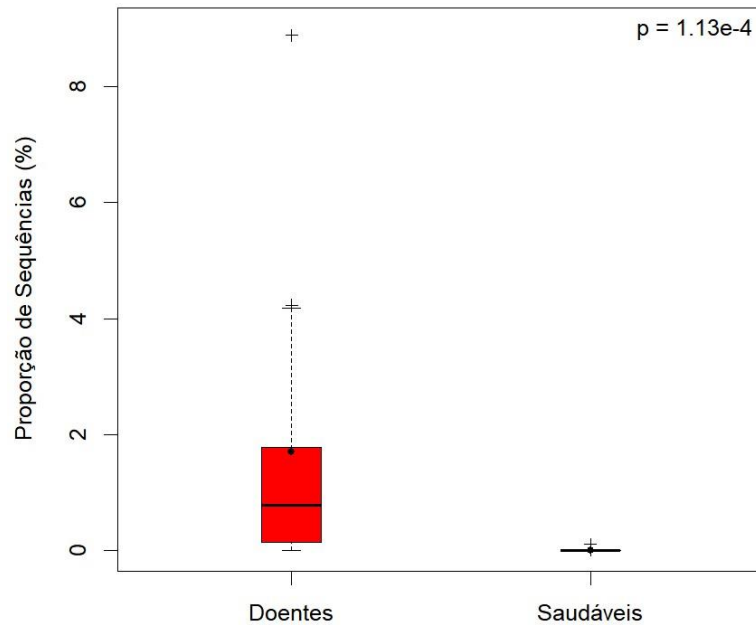
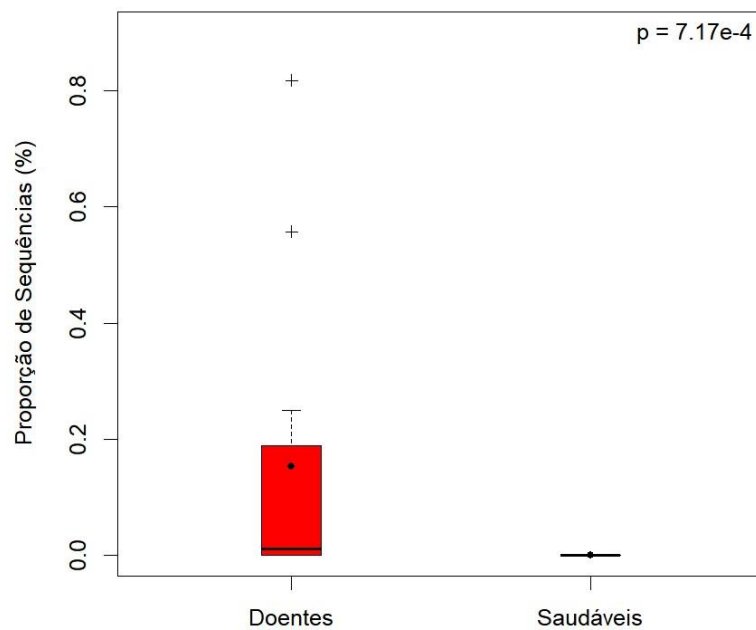


Figura 11. *Boxplot* demonstrando a distribuição na proporção do filo *Synergistetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.



Um total de 76 gêneros foi encontrado durante análise no programa Metaphlan 2.0. Na Tabela 7 e Tabela 8 podem ser observados os 10 gêneros com maiores abundâncias entre os grupos de estudo. Do total de gêneros, 25 apresentaram diferença significativa ( $p < 0,05$ ) para o teste estatístico não paramétrico Kruskal-Wallis (Tabela 9, Figura 12).

Tabela 7. Abundâncias relativas dos 10 gêneros mais abundantes no grupo doente, obtidas durante a previsão taxonômica através do programa Metaphlan 2.0.

Gênero	Abundância Relativa (%)
<i>Prevotella</i>	14,12
<i>Rothia</i>	13,27
<i>Streptococcus</i>	11,68
<i>Veillonella</i>	9,64
<i>Porphyromonas</i>	7,10
<i>Haemophilus</i>	5,48
<i>Neisseria</i>	4,13
<i>Corynebacterium</i>	3,54
<i>Leptotrichia</i>	2,62
<i>Fusobacterium</i>	2,14

Tabela 8. Abundâncias relativas dos 10 gêneros mais abundantes no grupo saudável, obtidas durante a previsão taxonômica através do programa Metaphlan 2.0.

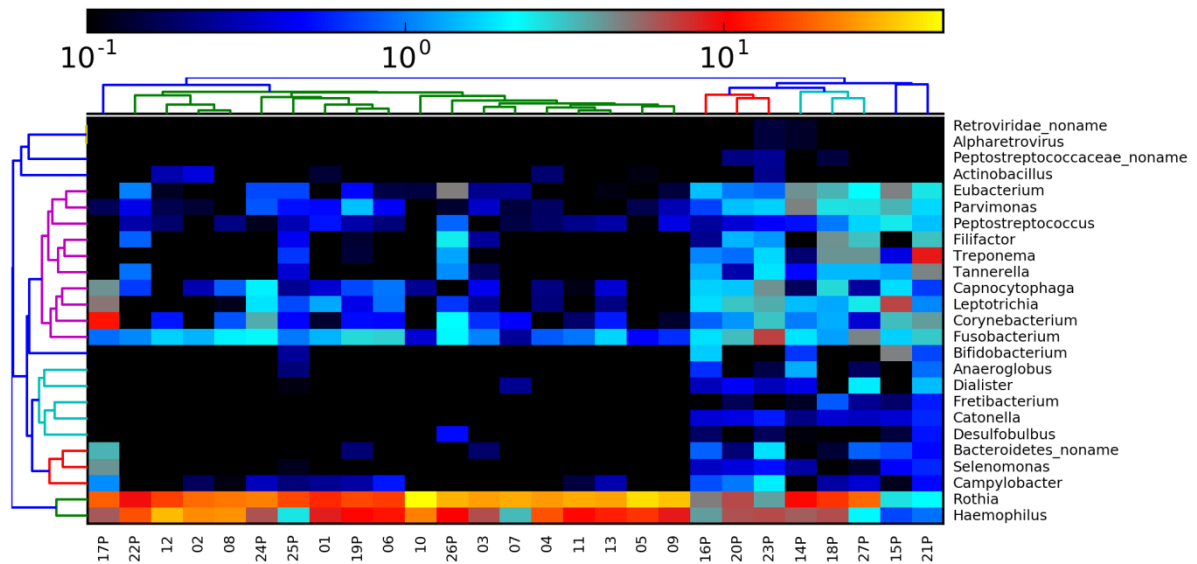
Gênero	Abundância Relativa (%)
<i>Rothia</i>	27,72
<i>Streptococcus</i>	13,72
<i>Haemophilus</i>	13,45
<i>Prevotella</i>	9,52
<i>Veillonella</i>	8,88
<i>Neisseria</i>	6,51
<i>Porphyromonas</i>	6,31
<i>Actinomyces</i>	2,57
<i>Granulicatella</i>	1,51
<i>Atopobium</i>	1,24

Tabela 9. Abundância relativa, para ambos os grupos de estudo, de todos os gêneros que apresentaram diferença estatística significativa para o teste não paramétrico Kruskal-Wallis ( $p < 0,05$ ), obtida durante a previsão taxonômica através do programa Metaphlan 2.0.

Domínio	Gênero	%Doente	%Saudável	p-value
<i>Bacteria</i>	<i>Rothia</i>	13,27	27,72	2,23E-03
<i>Bacteria</i>	<i>Haemophilus</i>	5,48	13,72	1,15E-03
<i>Bacteria</i>	<i>Corynebacterium</i>	3,54	0,31	2,06E-03
<i>Bacteria</i>	<i>Leptotrichia</i>	2,62	0,25	4,76E-04
<i>Bacteria</i>	<i>Fusobacterium</i>	2,14	1,04	8,78E-03
<i>Bacteria</i>	<i>Capnocytophaga</i>	1,84	0,29	1,24E-02
<i>Bacteria</i>	<i>Parvimonas</i>	1,44	0,19	2,73E-04
<i>Bacteria</i>	<i>Eubacterium</i>	1,37	0,13	1,39E-04
<i>Bacteria</i>	<i>Selenomonas</i>	1,08	0,007	8,78E-04
<i>Bacteria</i>	<i>Bacteroidetes noname</i>	1,01	0,024	2,56E-03
<i>Bacteria</i>	<i>Treponema</i>	0,90	0,006	1,13E-04
<i>Bacteria</i>	<i>Filifactor</i>	0,66	0,018	4,14E-03
<i>Bacteria</i>	<i>Campylobacter</i>	0,61	0,12	2,54E-02
<i>Bacteria</i>	<i>Tannerella</i>	0,60	0,010	4,71E-05
<i>Bacteria</i>	<i>Peptostreptococcus</i>	0,48	0,21	5,68E-03
<i>Bacteria</i>	<i>Bifidobacterium</i>	0,43	0,00	2,00E-02
<i>Bacteria</i>	<i>Dialister</i>	0,28	0,06	4,90E-03
<i>Bacteria</i>	<i>Anaeroglobus</i>	0,25	0,002	5,51E-03
<i>Bacteria</i>	<i>Catonella</i>	0,19	0,006	1,86E-03
<i>Bacteria</i>	<i>Fretibacterium</i>	0,086	0,00	7,17E-04
<i>Bacteria</i>	<i>Desulfobulbus</i>	0,067	0,00	7,17E-04
<i>Bacteria</i>	<i>Peptostreptococcaceae noname</i>	0,049	0,00	4,17E-03
<i>Bacteria</i>	<i>Actinobacillus</i>	0,018	0,08	2,18E-03
<i>Viruses</i>	<i>Alpharetrovirus</i>	0,033	0,00	4,13E-02
<i>Viruses</i>	<i>Retroviridae noname</i>	0,031	0,00	4,13E-02



Figura 12. *Heatmap* de todos os gêneros que apresentaram abundância com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.



No total, 201 espécies foram encontradas, das quais 70 apresentaram diferença significativa ( $p < 0,05$ ). A análise multivariada de componentes principais (PCA) para todas as espécies pode ser observada na Figura 13. As 25 espécies que obtiveram maiores abundâncias e diferença significativa entre os grupos de estudo, de acordo com metodologia aplicada pelo programa Metaphlan 2.0, podem ser visualizadas na Figura 14.

Figura 13. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa das espécies dos grupos doente (azul) e saudável (laranja) obtidos durante análise pelo programa Metaphlan 2.0.

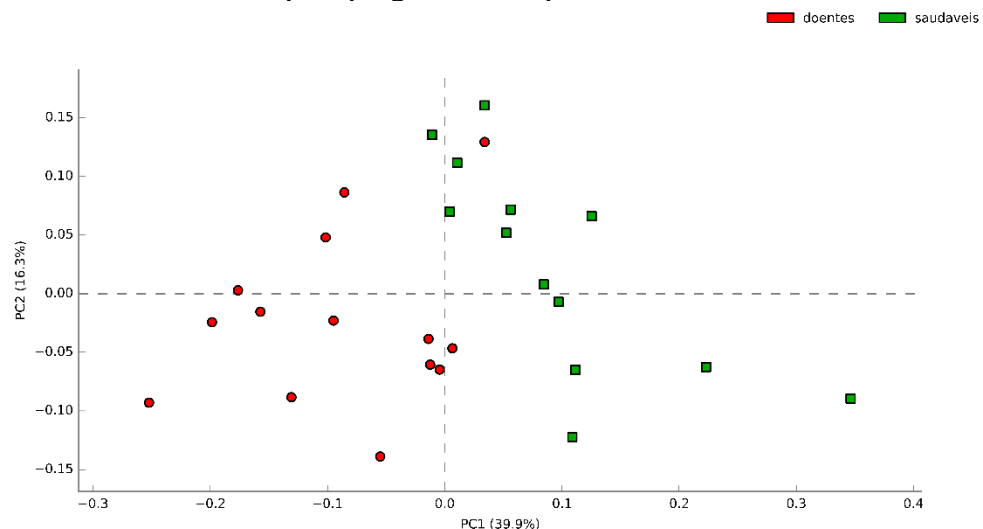
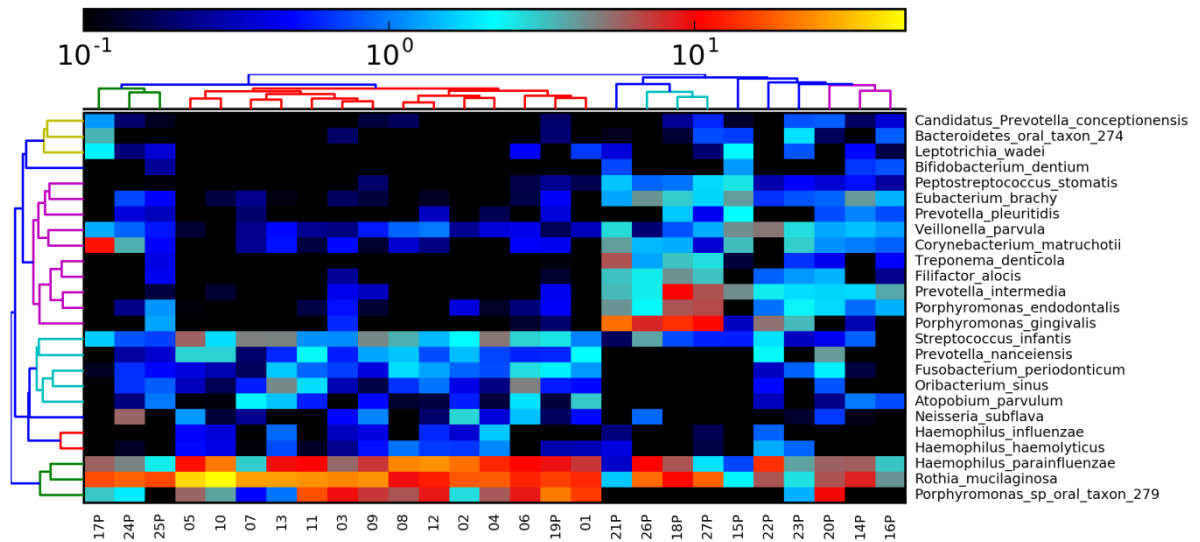


Figura 14. *Heatmap* das 25 espécies com maiores abundâncias e com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.



A Análise de Discriminantes Lineares (LDA) mostrou diferentes agrupamentos de filotipos entre os grupos de estudo tanto ao nível de gênero quanto ao nível de espécie ( Figura 15 e Figura 16).

Figura 15. Análise de Discriminantes Lineares (LDA) sobre os gêneros resultantes da anotação taxonômica através do programa Metaphlan 2.0.

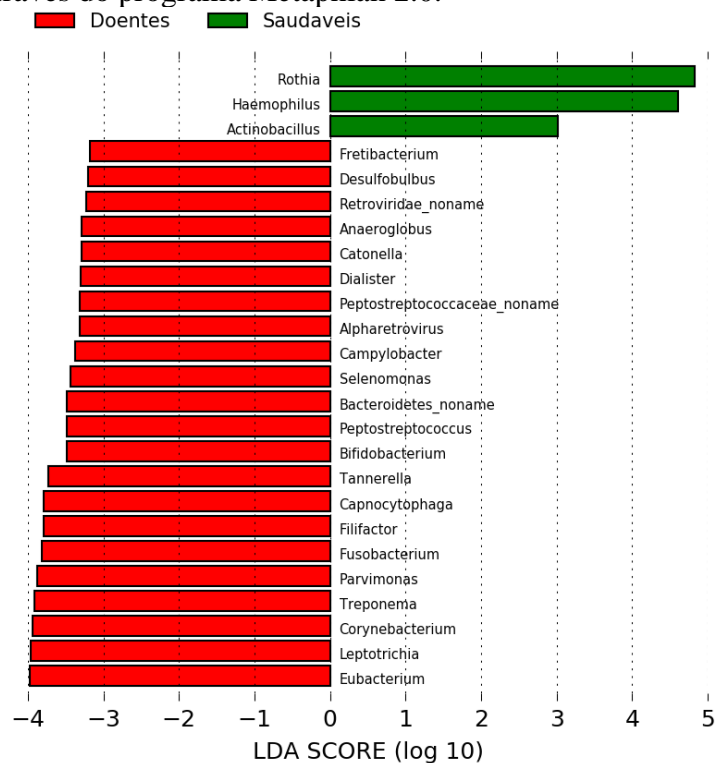
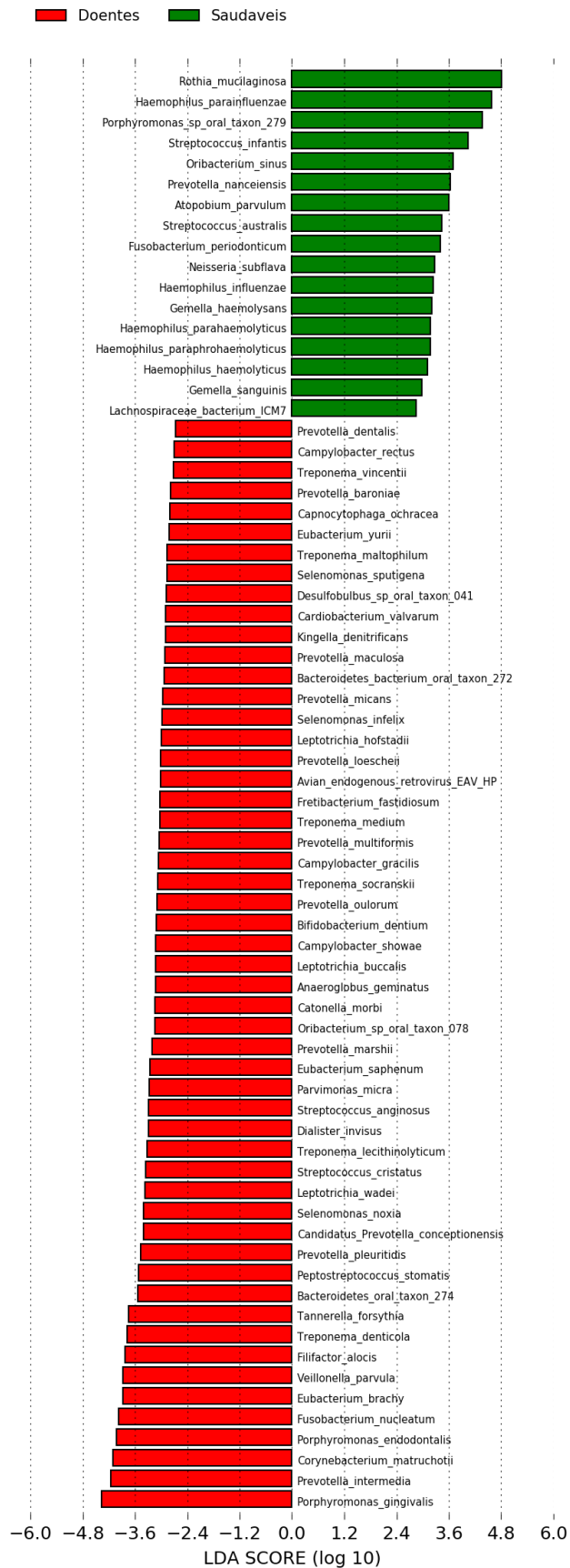


Figura 16. Análise de Discriminantes Lineares (LDA) sobre as espécies resultantes da anotação taxonômica através do programa Metaphlan 2.0.



Através da análise de correlação, foram obtidas tanto correlações positivas fortes ( $r > +0,7$ ) quanto correlações negativas fortes ( $r < -0,7$ ), para ambos os grupos de estudo. Foi possível correlacionar todas as 25 espécies no grupo doente, sendo a sua maior parte representada por correlações positivas. Para a análise de correlação, foram removidas, do grupo saudável, as espécies *Bifidobacterium dentium* e *Treponema denticola*, pois estas não apresentavam valores de abundância para este grupo. Os resultados das análises estatísticas para o grupo saudável seguem descritos na Tabela 10 e Figura 17A, enquanto para o grupo doente seguem descritos na Tabela 11 e na Figura 17B.

Tabela 10. Correlações de Spearman fortes ( $r < -0,7$  e  $r > +0,7$ ), para o grupo saudável, obtidas a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0.

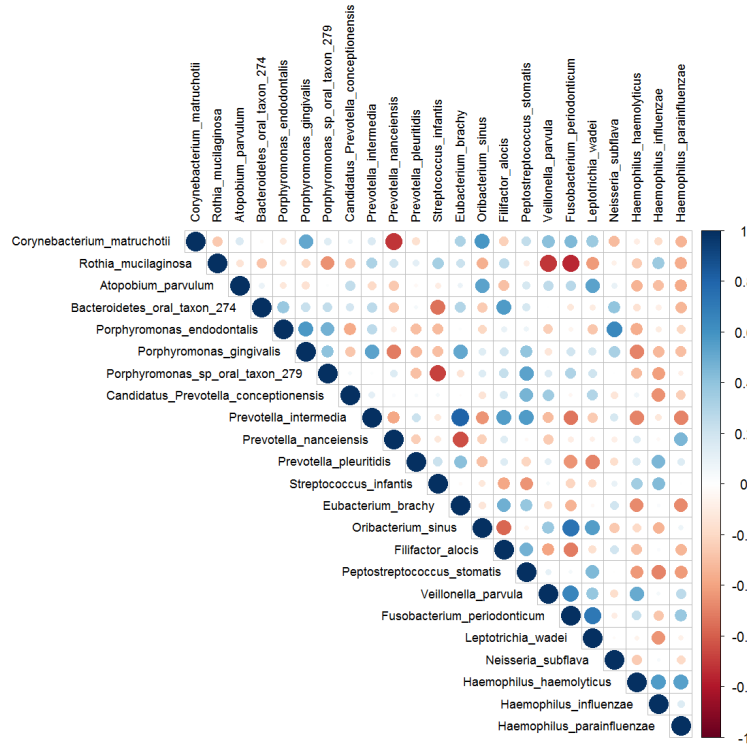
Espécie 1	Espécie 2	Valor de r
<i>Prevotella intermedia</i>	<i>Eubacterium brachy</i>	+0,80
<i>Oribacterium sinus</i>	<i>Fusobacterium periodonticum</i>	+0,73
<i>Fusobacterium periodonticum</i>	<i>Leptotrichia wadei</i>	+0,72
<i>Corynebacterium matruchotii</i>	<i>Prevotella nanceiensis</i>	-0,71
<i>Rothia mucilaginosa</i>	<i>Veillonella parvula</i>	-0,72
<i>Rothia mucilaginosa</i>	<i>Fusobacterium periodonticum</i>	-0,75

Tabela 11. Correlações de Spearman fortes ( $r < -0,7$  e  $r > +0,7$ ), para o grupo doente, obtidas a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0.

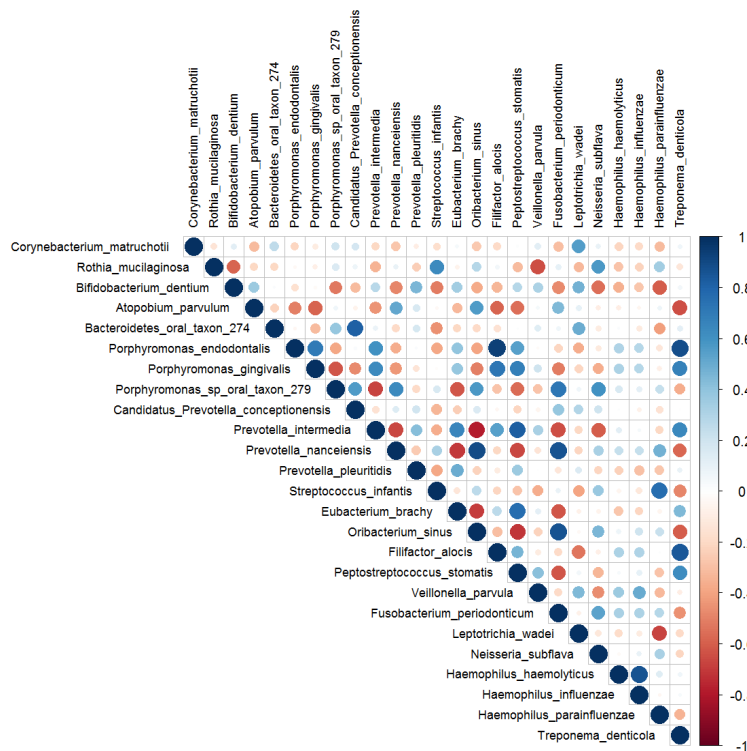
Espécie 1	Espécie 2	Valor de r
<i>Filifactor alocis</i>	<i>Porphyromonas endodontalis</i>	+0,94
<i>Oribacterium sinus</i>	<i>Prevotella nanceiensis</i>	+0,90
<i>Treponema denticola</i>	<i>Porphyromonas endodontalis</i>	+0,89
<i>Haemophilus influenzae</i>	<i>Haemophilus haemolyticus</i>	+0,87
<i>Fusobacterium periodonticum</i>	<i>Prevotella nanceiensis</i>	+0,86
<i>Fusobacterium periodonticum</i>	<i>Oribacterium sinus</i>	+0,86
<i>Treponema denticola</i>	<i>Filifactor alocis</i>	+0,85
<i>Peptostreptococcus stomatis</i>	<i>Prevotella intermedia</i>	+0,82
<i>Candidatus Prevotella conceptionensis</i>	<i>Bacteroidetes oral taxon 274</i>	+0,81
<i>Haemophilus parainfluenzae</i>	<i>Streptococcus infantis</i>	+0,77
<i>Peptostreptococcus stomatis</i>	<i>Eubacterium brachy</i>	+0,77
<i>Fusobacterium periodonticum</i>	<i>Porphyromonas sp oral taxon 279</i>	+0,73
<i>Filifactor alocis</i>	<i>Porphyromonas gingivalis</i>	+0,73
<i>Porphyromonas gingivalis</i>	<i>Porphyromonas endodontalis</i>	+0,71
<i>Eubacterium brachy</i>	<i>Prevotella nanceiensis</i>	-0,71
<i>Peptostreptococcus stomatis</i>	<i>Oribacterium sinus</i>	-0,71
<i>Oribacterium sinus</i>	<i>Prevotella intermedia</i>	-0,79

Figura 17. Matriz de correlações de Spearman, para os grupos saudável (A) e doente (B), construída a partir da abundância relativa das 25 espécies mais abundantes e que obtiveram diferença estatística significativa entre os grupos de estudo para o teste estatístico não paramétrico Kruskal-Wallis ( $p < 0,05$ ). Os valores de abundância foram obtidos durante a previsão taxonômica através do programa Metaphlan 2.0.

A)



B)



## 5.5 ANOTAÇÃO TAXONÔMICA VIA MG-RAST

Um total de 104.004.730 sequências foi submetido à análise no servidor MG-RAST, sendo 40.569.403 (39%) e 63.435.327 (61%) provenientes dos grupos doente e saudável, respectivamente (Tabela 12).

Tabela 12. Número de sequências, pares de base (pb) e tamanho médio, para cada amostra envolvida no estudo, submetida a análise pelo servidor MG-RAST.

Amostra	MG-RAST ID	N. de sequências	Mpb	Média±Desvio
1	mgm4763328.3	8.695.195	1304,28	150 ± 0
2	mgm4763336.3	4.153.480	653,13	157 ± 30
3	mgm4763326.3	3.007.863	544,97	181 ± 55
4	mgm4763321.3	5.049.354	920,53	182 ± 45
5	mgm4763323.3	3.952.422	615,03	156 ± 28
6	mgm4763345.3	3.520.145	678,06	193 ± 46
7	mgm4763314.3	7.939.561	1804	227 ± 56
8	mgm4763316.3	5.815.265	972,27	167 ± 44
9	mgm4763318.3	4.211.353	792,62	188 ± 47
10	mgm4763319.3	5.745.540	893,62	156 ± 27
11	mgm4763312.3	5.121.756	981,36	192 ± 45
12	mgm4763334.3	2.116.385	401,79	190 ± 46
13	mgm4763347.3	4.107.008	965,52	235 ± 44
14P	mgm4763325.3	3.387.744	524,65	155 ± 26
15P	mgm4763330.3	1.943.928	303,1	156 ± 33
16P	mgm4763313.3	2.653.911	435,32	164 ± 40
17P	mgm4763341.3	8.300.854	1582	191 ± 59
18P	mgm4763340.3	1.324.687	202,5	153 ± 30
19P	mgm4763337.3	3.389.329	540,07	159 ± 34
20P	mgm4763320.3	5.313.239	878,84	165 ± 42
21P	mgm4763344.3	1.344.845	206,1	153 ± 32
22P	mgm4763315.3	808.758	124,82	154 ± 37
23P	mgm4763317.3	2.147.133	351,46	164 ± 43
24P	mgm4763332.3	2.314.931	363,19	157 ± 32
25P	mgm4763322.3	4.951.778	825,7	167 ± 43
26P	mgm4763338.3	668.183	104,29	156 ± 37
27P	mgm4763335.3	2.020.083	316,99	157 ± 32

Do total de sequências, 49.391.119 (47,49%) puderam ser anotadas com base no banco de dados RefSeq. Os resultados da análise estatística referentes à anotação podem ser observados na Tabela 13.

Tabela 13. Número de sequências anotadas, para cada amostra envolvida no estudo, durante a previsão taxonômica através do servidor MG-RAST. \*Apresentou apenas um arquivo contendo *reads* durante o sequenciamento.

Amostra	Sequências		
	Submetidas	Anotadas	% Anotadas
1*	8.695.195	26	0,0003
2	4.153.480	1.987.368	47,85
3	3.007.863	1.891.843	62,9
4	5.049.354	3.409.479	67,52
5	3.952.422	2.233.107	56,5
6	3.520.145	1.571.744	44,65
7	7.939.561	6.074.337	76,51
8	5.815.265	4.037.280	69,43
9	4.211.353	3.090.220	73,38
10	5.745.540	3.556.651	61,9
11	5.121.756	2.706.694	52,85
12	2.116.385	1.106.807	52,3
13	4.107.008	2.366.448	57,62
14P	3.387.744	1.097.320	32,39
15P	1.943.928	323.802	16,66
16P	2.653.911	873.814	32,93
17P	8.300.854	4.960.473	59,76
18P	1.324.687	84.406	6,37
19P	3.389.329	1.401.829	41,36
20P	5.313.239	2.158.455	40,62
21P	1.344.845	60.015	4,46
22P	808.758	42.226	5,22
23P	2.147.133	514.070	23,94
24P	2.314.931	684.331	29,56
25P	4.951.778	2.670.320	53,93
26P	668.183	28.408	4,25
27P	2.020.083	459.646	22,75

A amostra 1, obteve um número de sequências anotadas quase nulo em comparação com as demais amostras. Esse fato parece ter sido influenciado pela ausência do segundo arquivo contendo *reads* durante a saída de resultados do sequenciamento. Portanto, essa amostra foi excluída quando referente as análises executadas através do servidor MG-RAST. Sendo assim, a maior parcela dos táxons anotados foi referente ao domínio *Bacteria* (99,77%), seguida pelo domínio *Eukaryota* (0,21%). Os demais domínios obtiveram valores consideravelmente inferiores, sendo eles: *Virus* (0,02%) e *Archaea* (0,0001%). Aqueles resultados que foram agrupados pelo servidor como ‘não classificadas’ ou ‘derivadas de’



foram considerados como ‘não classificados’. Na Figura 18 e Figura 19, é possível observar a curva de rarefação para cada amostra.

Figura 18. Curva de rarefação gerada a partir das sequências anotadas pelo servidos MG-RAST para grupo doente. As curvas representam a média do número de diferentes sequências anotadas para cada amostra do grupo.

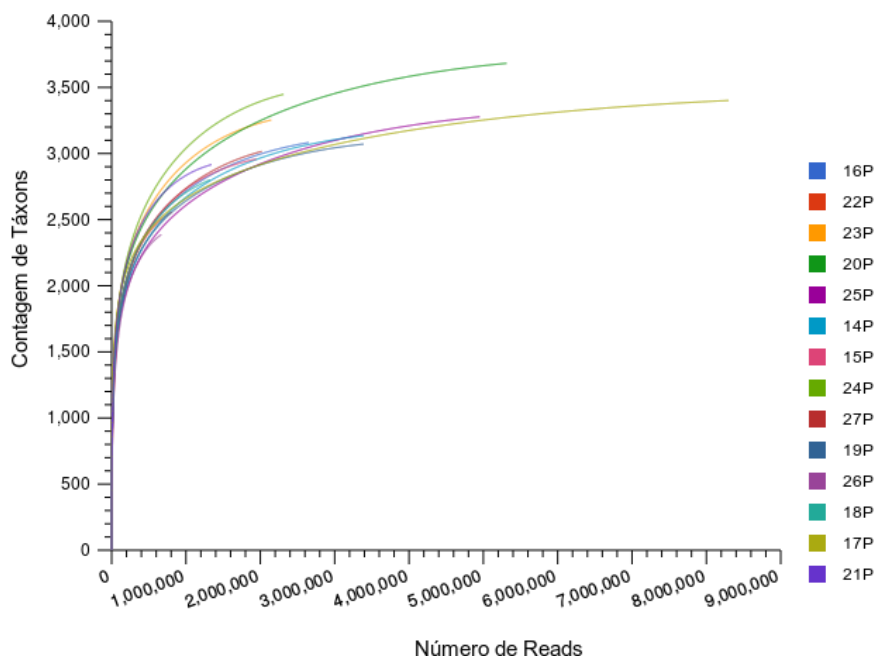
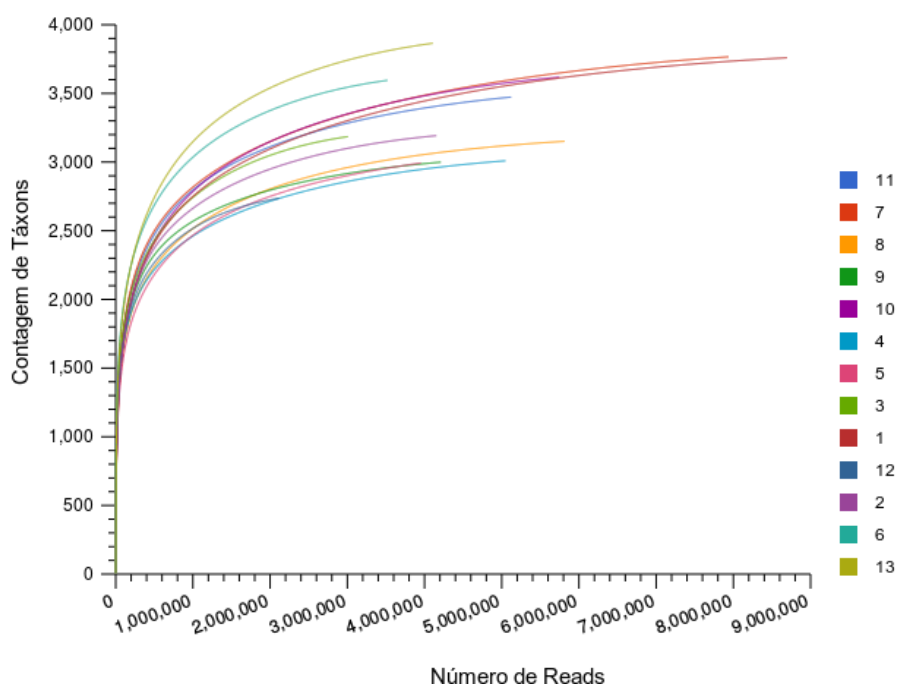


Figura 19. Curva de rarefação gerada a partir das sequências anotadas pelo servidos MG-RAST para grupo saudável. As curvas representam a média do número de diferentes sequências anotadas para cada amostra do grupo.



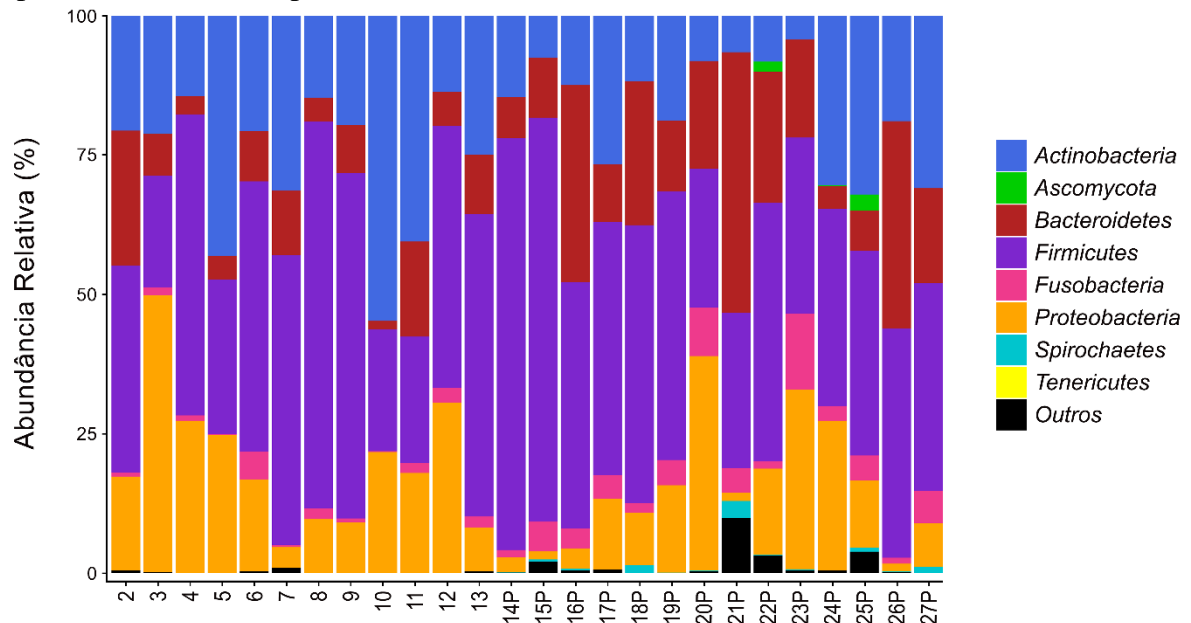
Um total de 42 filios foi obtido, tendo representantes dos domínios *Bacteria*, *Eukaryota* e *Archaea*. Embora o domínio *Virus* tenha apresentado resultados para filo, estes foram agrupados como ‘não classificados (derivados de *Viruses*)’. Isso significa que os resultados não puderam ser anotados em nível de filo, ou mesmo nos demais níveis taxonômicos. O filo *Firmicutes* destacou-se em abundância com 44,73%, seguido pelos filios *Actinobacteria* (25,14%), *Proteobacteria* (16,02%), *Bacteroidetes* (11,54%), *Fusobacteria* (2,17%) e *Spirochaetes* (0,038%). Os demais filios apresentaram abundâncias inferiores a 0,008%. Os filios obtidos com mais de 1.000 *hits*, os quais contemplaram conjuntamente > 99% dos resultados, encontram-se dispostos na Tabela 14.

Tabela 14. Abundância absoluta e relativa geral dos principais filios encontrados na cavidade oral durante a previsão taxonômica através do servidor MG-RAST.

Domínio	Filios	Abundância	
		Absoluta	Relativa (%)
<i>Bacteria</i>	<i>Firmicutes</i>	21884801	44,73
<i>Bacteria</i>	<i>Actinobacteria</i>	12301983	25,14
<i>Bacteria</i>	<i>Proteobacteria</i>	7840299	16,02
<i>Bacteria</i>	<i>Bacteroidetes</i>	5645873	11,54
<i>Bacteria</i>	<i>Fusobacteria</i>	1062820	2,17
<i>Bacteria</i>	<i>Spirochaetes</i>	18592	0,038
<i>Eukaryota</i>	<i>Ascomycota</i>	3785	0,008
<i>Bacteria</i>	<i>Tenericutes</i>	1543	0,003

O perfil, por amostra, dos principais filios encontrados nos grupos doente e saudável pode ser visualizado na Figura 20.

Figura 20. Perfil taxonômico de todas as mostras envolvidas no estudo, exceto a amostra 1. Os perfis foram construídos através da abundância relativa dos principais filos obtidos durante a previsão taxonômica pelo servidor MG-RAST.



Os filos *Proteobacteria*, *Fusobacteria*, *Spirochaetes*, *Ascomycota* e *Tenericutes* apresentaram maiores abundâncias no grupo doente. Enquanto que os filos *Actinobacteria* e *Bacteroidetes* obtiveram maiores valores de abundância no grupo saudável. O filo *Firmicutes*, por sua vez, obteve valores de abundância próximos para ambos os grupos de estudo (Tabela 15). Os filos *Actinobacteria*, *Bacteroidetes*, *Fusobacteria* e *Spirochaetes* apresentaram diferença estatística significativa para o teste estatístico não paramétrico Kruskal-Wallis com  $p < 0,05$  (Figura 21 a Figura 24).

Tabela 15. Abundância relativa dos principais filos encontrados para os grupos doente e saudável durante a previsão taxonômica através do servidor MG-RAST. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ).

Domínio	Filos	Abundância Relativa (%)		p-value
		Doentes	Saudáveis	
<i>Bacteria</i>	<i>Firmicutes</i>	43,55	45,24	8,77E-01
<i>Bacteria</i>	<i>Actinobacteria</i> *	18,91	27,87	2,70E-02
<i>Bacteria</i>	<i>Proteobacteria</i>	14,44	16,71	8,96E-02
<i>Bacteria</i>	<i>Bacteroidetes</i> *	18,14	8,65	1,80E-02
<i>Bacteria</i>	<i>Fusobacteria</i> *	4,34	1,223	4,67E-03
<i>Bacteria</i>	<i>Spirochaetes</i> *	0,109	0,007	5,69E-04
<i>Eukaryota</i>	<i>Ascomycota</i>	0,015	0,005	7,89E-01
<i>Bacteria</i>	<i>Tenericutes</i>	0,005	0,002	6,06E-01

Figura 21. *Boxplot* demonstrando a distribuição na proporção do filo *Actinobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST.

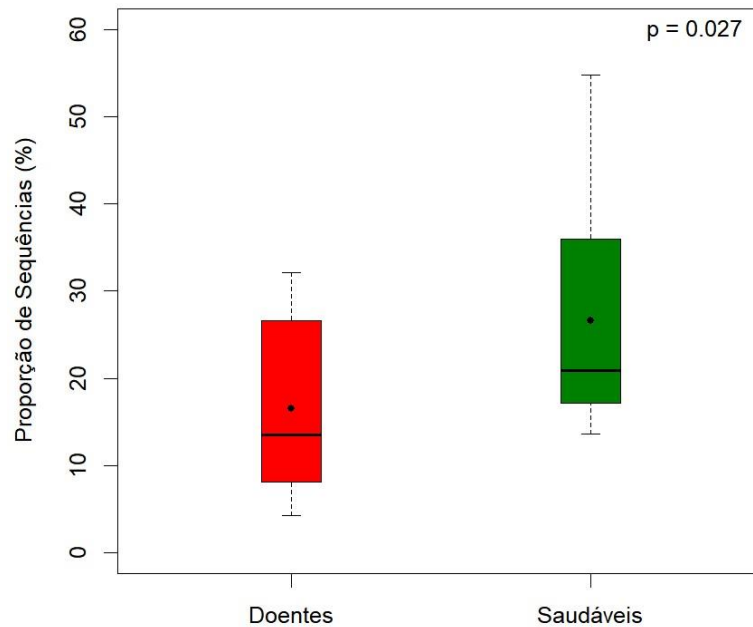


Figura 22. *Boxplot* demonstrando a distribuição na proporção do filo *Bacteroidetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST.

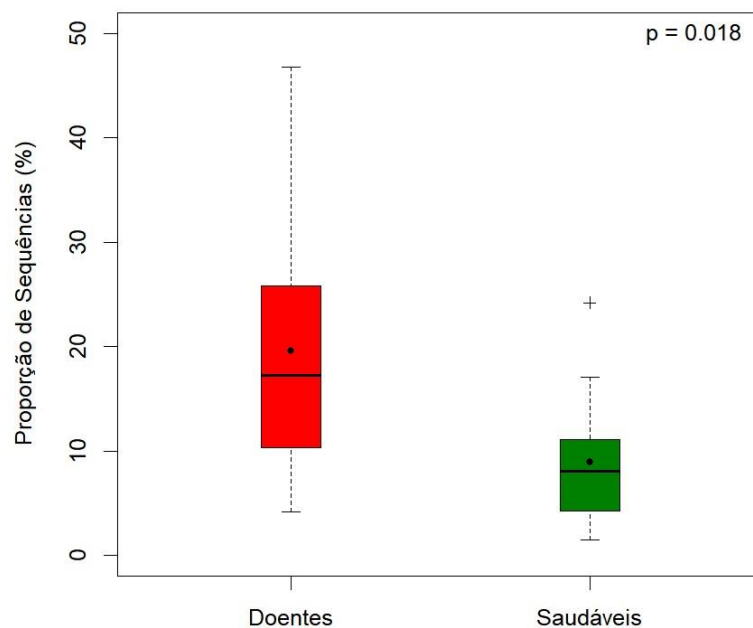


Figura 23. *Boxplot* demonstrando a distribuição na proporção do filo *Fusobacteria* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST.

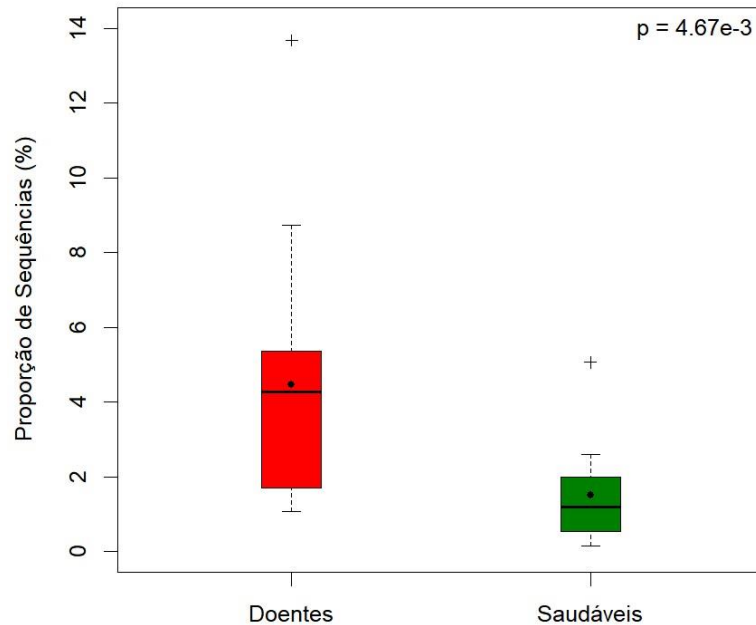
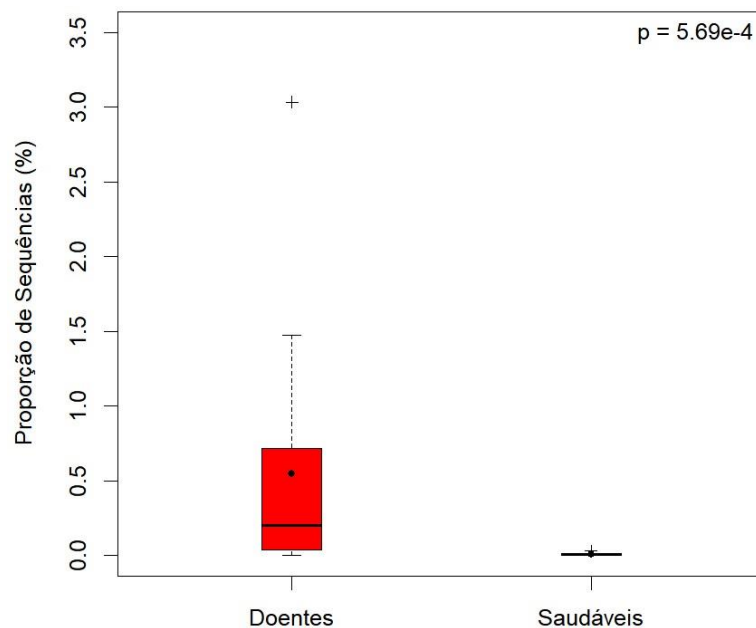
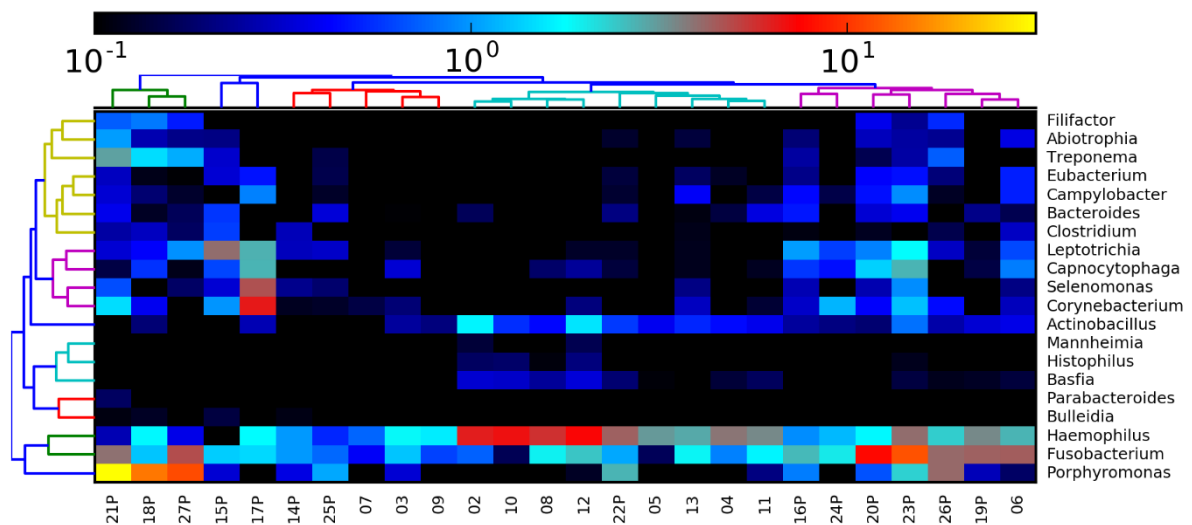


Figura 24. *Boxplot* demonstrando a distribuição na proporção do filo *Spirochaetes* com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST.



Um total de 557 gêneros foi obtido. Deste total, 98 gêneros apresentaram diferença estatística significativa para o teste estatístico não paramétrico Kruskal-Wallis com  $p < 0,05$ . Os 20 gêneros que apresentaram diferença estatística e obtiveram os maiores valores de abundâncias estão representados na Figura 25.

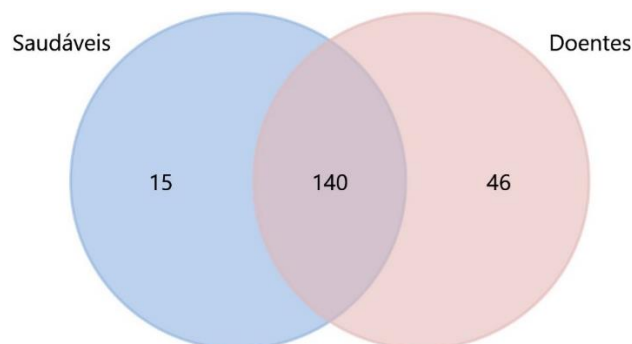
Figura 25. *Heatmap* dos gêneros com diferença estatística significativa entre os grupos de estudo para o teste Kruskal-Wallis ( $p < 0,05$ ). Utilizou-se para a construção das árvores uma matriz de similaridade de Bray-Curtis. Dados de abundância obtidos a partir da previsão taxonômica pelo servidor MG-RAST.



## 5.6 RELAÇÕES FILOGENÉTICAS, RIQUEZA E DIVERSIDADE

Com base em todas as espécies encontradas durante a análise feita pelo programa Metaphlan 2.0, foi observado um total de 155 espécies distintas no grupo saudável, enquanto o grupo doente apresentou 186 (Figura 26) demonstrando uma riqueza de espécies maior.

Figura 26. Diagrama de Venn da riqueza de espécies encontrada entre os grupos doente (vermelho) e saudável (azul) fazendo uso dos dados de abundância obtidos do programa Metaphlan 2.0.



Todos os estimadores de riqueza aplicados aos dados de cada amostra em ambos os grupos mostraram valores maiores de riqueza (Figura 27 e Figura 28) que variam entre 167,66-196,44 (grupo saudável) e 194,81-209,64 (grupo doente). Além da riqueza de espécies apresentar maior valor no grupo doente, os maiores valores quanto aos índices de diversidade de Shannon e Simpson (1-D) concentram-se também nas amostras do grupo doente. Os dados numéricos tanto da riqueza quanto dos índices de diversidade para cada amostra podem ser encontrados na Tabela 16.

Figura 27. Estimativa de riqueza de espécies em amostras do grupo saudável. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

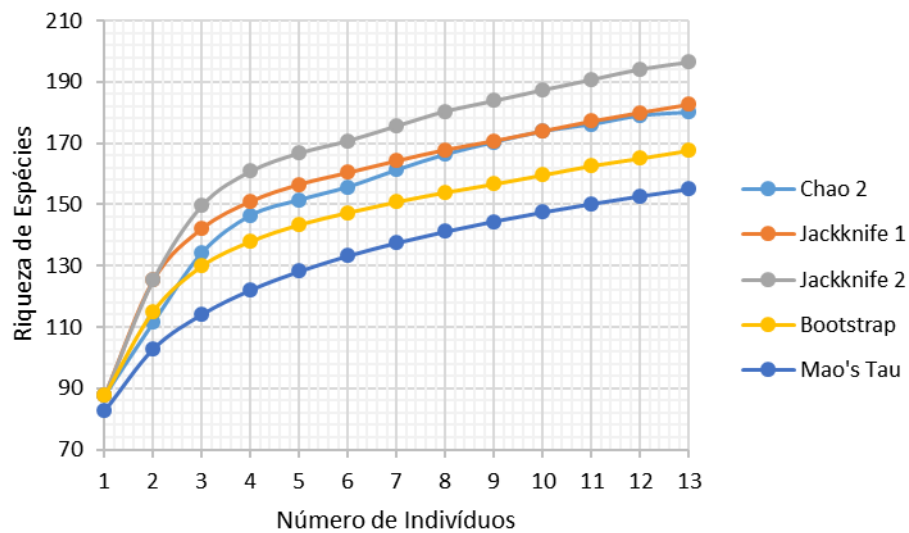


Figura 28. Estimativa de riqueza de espécies em amostras do grupo doente. Dados de abundância obtidos a partir da previsão taxonômica pelo programa Metaphlan 2.0.

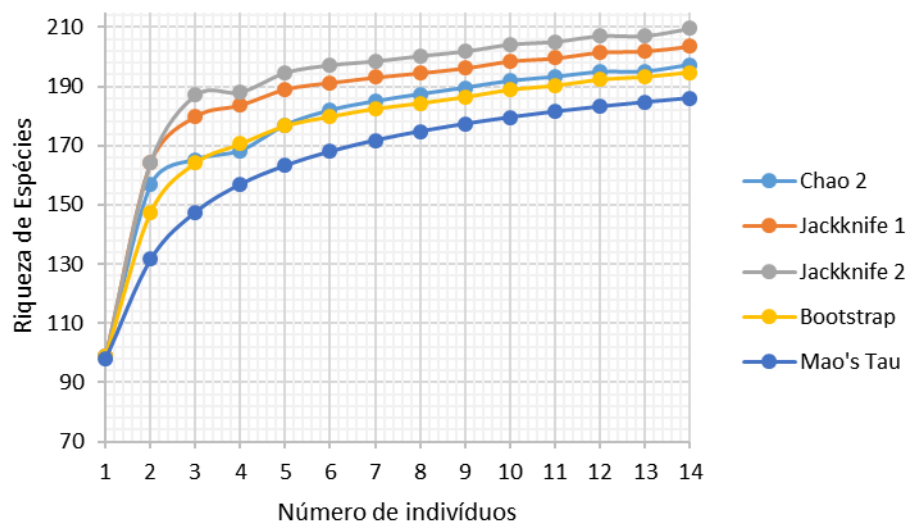


Tabela 16. Índices de diversidade de Shannon e Simpson (1-D), e riqueza de cada amostra. Foram utilizados os dados de abundância obtidos do programa Metaphlan 2.0.

Amostra	Riqueza	Shannon	Simpson (D-1)
01	98	3,053	0,913
02	77	2,812	0,884
03	97	2,838	0,874
04	78	2,876	0,907
05	70	2,348	0,800
06	81	3,274	0,934
07	119	2,942	0,865
08	89	2,904	0,896
09	81	2,939	0,893
10	64	1,720	0,662
11	74	2,602	0,860
12	65	2,727	0,874
13	83	3,018	0,883
14P	123	3,872	0,963
15P	96	3,794	0,967
16P	113	3,795	0,963
17P	88	3,178	0,923
18P	81	3,247	0,933
19P	117	3,005	0,888
20P	135	3,731	0,955
21P	81	3,385	0,931
22P	46	3,208	0,937
23P	127	4,087	0,974
24P	115	3,416	0,929
25P	117	3,166	0,914
26P	32	2,832	0,913
27P	102	3,208	0,922

Tanto os índices de diversidade de Shannon quanto o de Simpson (1-D) apresentaram diferença significativa para o teste estatístico Kruskal-Wallis para  $p < 0,05$  (Figura 29 e Figura 30). O índice de diversidade beta (Whittaker) resultou em 0,873 para o grupo saudável e 0,897 para o grupo doente.



Figura 29. *Boxplot* demonstrando a distribuição do índice de Shannon com diferença significativa para o teste estatístico Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa e os *outliers* por círculos.

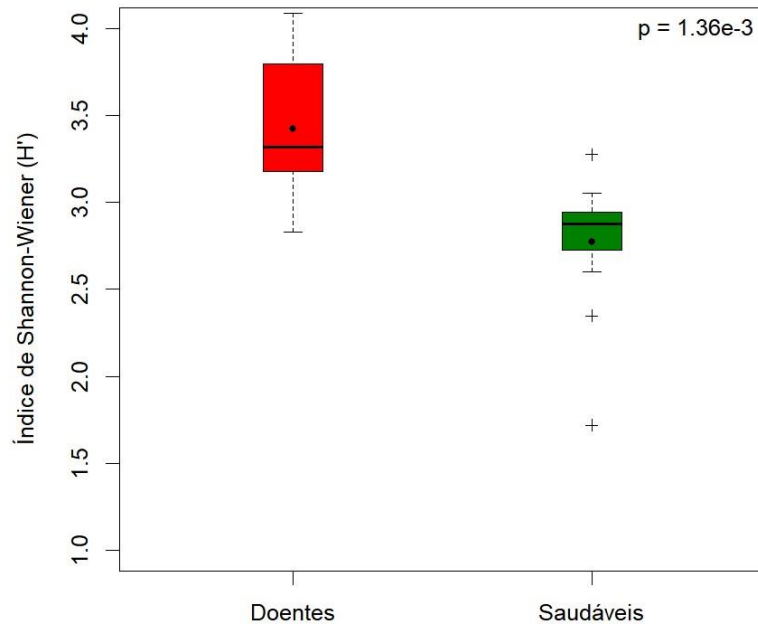
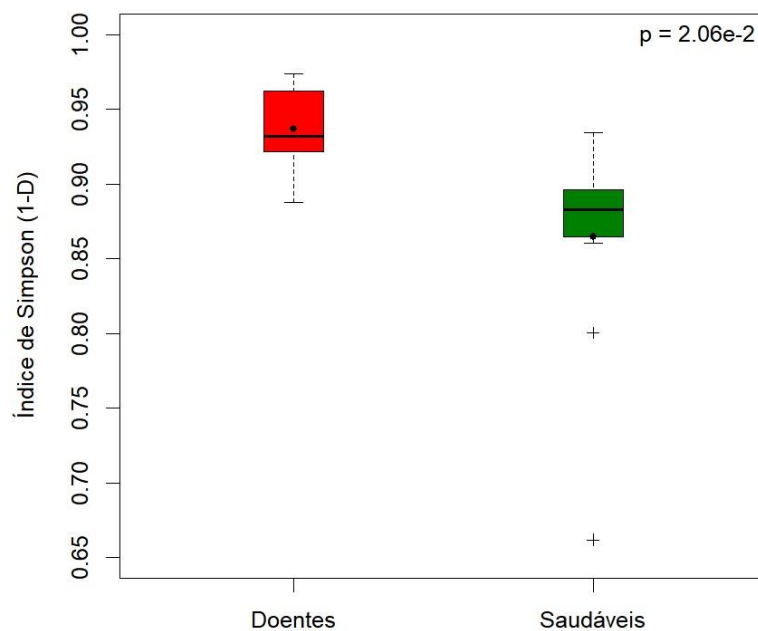


Figura 30. *Boxplot* demonstrando a distribuição do índice de Simpson com diferença significativa para o teste estatístico Kruskal-Wallis ( $p < 0,05$ ) entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa e os *outliers* por círculos.



As relações filogenéticas demonstradas através de cladogramas auxiliam na visualização dos dados de abundância encontrados pelo programa Metaphlan 2.0. Desta

forma, a representação filogenética das espécies bacterianas para ambos os grupos de estudo pode ser visualizada na Figura 31 e Figura 32.

Figura 31. Representação filogenética em forma de cladograma dos táxons resultantes da análise taxonômica pelo programa Metaphlan 2.0 do grupo saudável. O tamanho dos círculos internos assim como das barras externas demonstra, de modo proporcional, a abundância do nível taxonômico e das espécies, respectivamente.

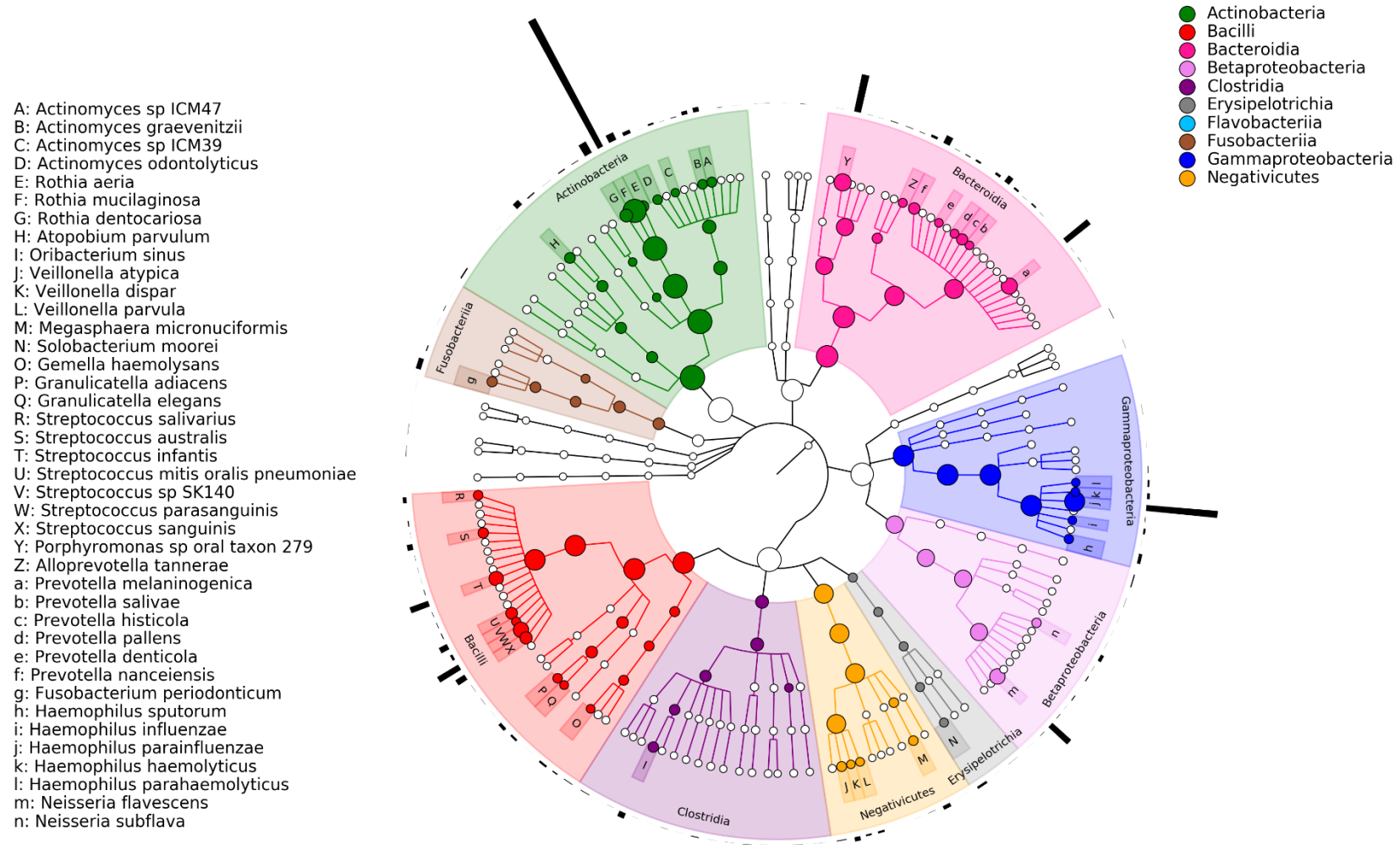
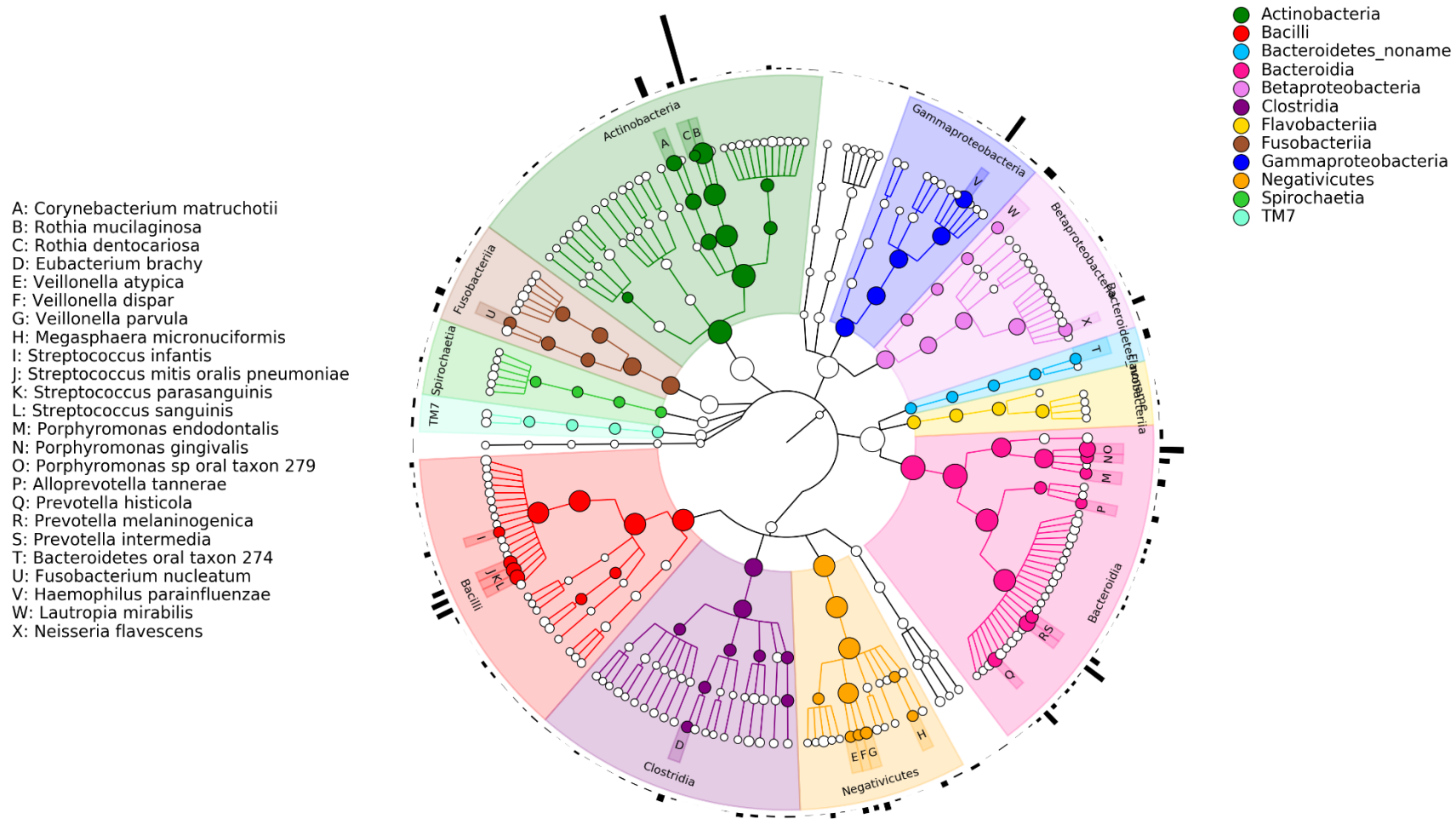


Figura 32. Representação filogenética em forma de cladograma dos táxons resultantes da análise taxonômica pelo programa Metaphlan 2.0 do grupo doente. O tamanho dos círculos internos assim como das barras externas demonstra, de modo proporcional, a abundância do nível taxonômico que se esteja observando e das espécies, respectivamente.



## 5.7 ANOTAÇÃO FUNCIONAL VIA KEEG

Durante a análise funcional realizada através do servidor MG-RAST com uso do banco de dados KEEG foi submetido um *pool* total de 78.815.223 ORFs, das quais foram anotadas 3.221.748 para o grupo doente e 7.581.449 para o grupo saudável (Tabela 17). Esses resultados estão representados por 1343 genes distintos e encontram-se distribuídos entre os grupos de estudo de acordo com a classificação de nível hierárquico 1, como: processos celulares, processamento de informação ambiental, processamento de informação genética, doenças humanas, metabolismo, sistemas organizacionais.

Tabela 17. *Open Read Frames* (ORFs) anotadas em banco de dados KEGG através do servidor MG-RAST.

<i>Open Read Frames</i>			
Amostra	Submetidos	Anotados	%Anotados
2	3.403.497	491.632	14,44
3	2.531.841	419.188	16,56
4	4.240.819	809.512	19,09
5	3.263.377	526.112	16,12
6	2.882.044	314.663	10,92
7	6.688.384	1.197.647	17,91
8	4.975.781	953.751	19,17
9	3.498.422	669.826	19,15
10	4.806.217	858.842	17,87
11	4.155.435	560.682	13,49
12	1.636.140	282.768	17,28
13	3.415.868	496.826	14,54
14P	2.860.341	252.942	8,84
15P	1.423.592	81.645	5,74
16P	2.158.242	175.666	8,14
17P	7.308.104	957.059	13,10
18P	994.862	20.676	2,08
19P	2.860.354	341.648	11,94
20P	4.570.297	434.529	9,51
21P	96.9250	13.162	13,58
22P	49.9960	11.577	23,16
23P	1.687.514	119.766	7,10
24P	1.825.933	170.403	9,33
25P	4.208.027	529.646	12,59
26P	384.511	6.775	1,76
27P	1.566.411	106.254	6,78

Com base nos dados de abundância relativa para todo o conjunto amostral, mais de 50% dos resultados obtidos estão contidos na categoria “Metabolismo” seguido por “Processamento de informação genética” e “Processamento de informação ambiental” (Tabela 18). Embora a Análise de Componentes Principais (PCA) realizada sobre os dados referentes ao nível hierárquico 1 não mostre separação entre os grupos de estudo (Figura 33), houve diferença estatística significativa para “Processamento de informação genética” com maior abundância para o grupo doente (Tabela 18 e Figura 34).

Tabela 18. Abundância relativa para o nível hierárquico 1 obtido do banco de dados KEGG, para cada grupo de estudo. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg.

Nível hierárquico 1	%Saudável	%Doente	p-value
Metabolismo	53,17	53,28	2,41E-01
Processamento de informação genética*	27,30	28,03	4,49E-02
Processamento de informação ambiental	14,69	13,88	2,72E-01
Processos celulares	2,76	2,65	8,17E-01
Doenças humanas	1,50	1,53	8,77E-01
Sistemas organizacionais	0,57	0,63	2,72E-01

Figura 33. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 1 (KEGG) dos grupos doente (azul) e saudável (laranja).

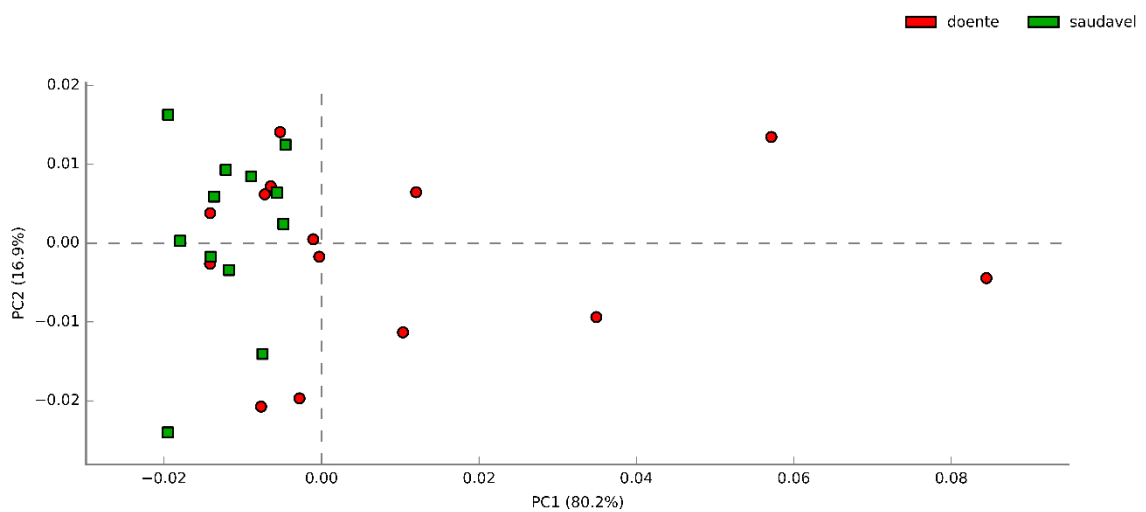
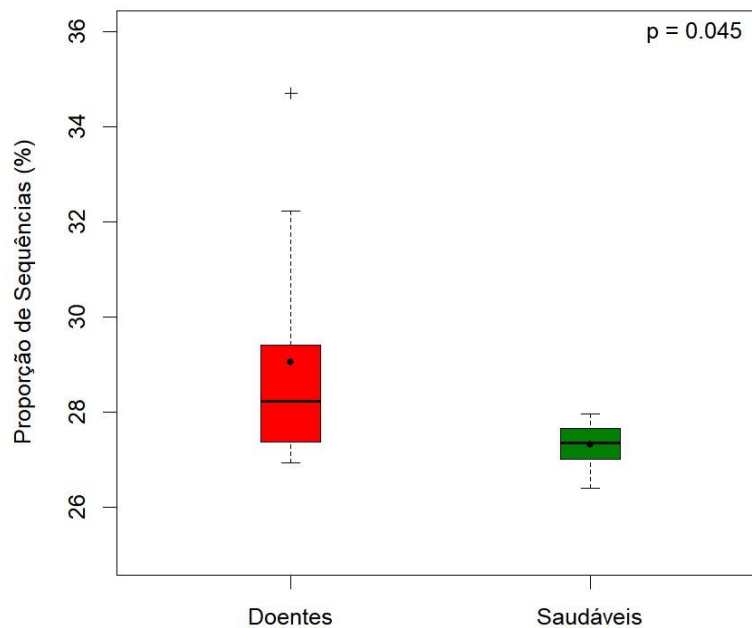


Figura 34. *Boxplot* demonstrando a distribuição na proporção categoria “Processamento de informação genética” com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos da análise funcional a partir do banco de dados KEGG, nível hierárquico 1, realizada pelo servidor MG-RAST.



Os resultados obtidos a partir do nível hierárquico 2 estão distribuídos em 38 categorias distintas. As categorias que se destacam por apresentar maiores valores de abundância, tanto em doentes quanto em saudáveis, respectivamente, são: Metabolismo de aminoácidos (17%; 19,95%), tradução (13,06%; 13,49%), metabolismo de carboidratos (11,79%; 12,21%) e transporte de membrana (11,23%; 10,53%).

A Análise de Componentes Principais (PCA) para o nível hierárquico 2 não demonstrou agrupamento de amostras do mesmo grupo de estudo, assim como não houve separação clara entre amostras dos grupos distintos (Figura 35). A análise exploratória presente na Figura 36, por sua vez, com base na abundância relativa das categorias em cada amostra, aponta para um perfil similar entre todas as amostras. Embora os resultados exploratórios tenham mostrado similaridade entre os grupos de estudo, ao realizar análise estatística pelo teste estatístico não paramétrico Kruskal-Wallis ( $p < 0,05$ ) foi possível encontrar diferença estatística significativa para a categoria “transcrição” com maior abundância no grupo doente (3,01%) em relação ao saudável (2,57%) (Figura 37).

Figura 35. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 2 (KEGG) dos grupos doente (azul) e saudável (laranja).

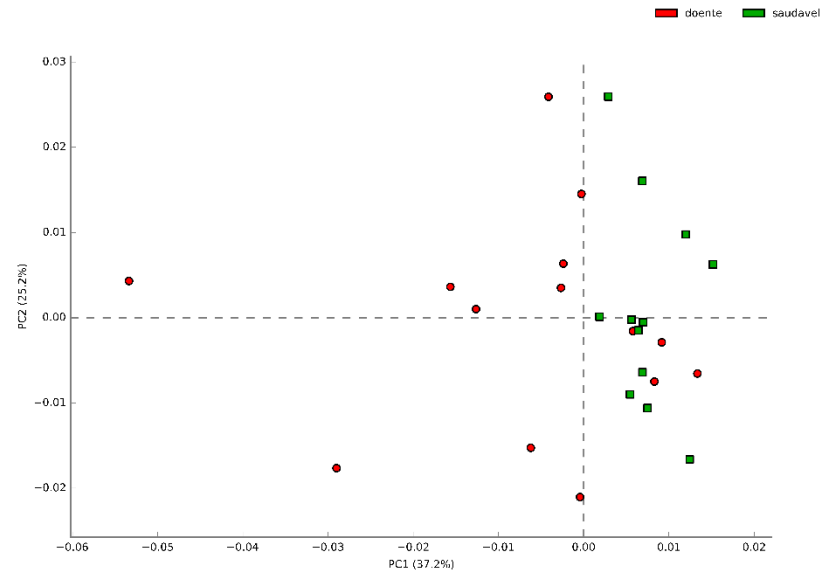




Figura 36. Perfil funcional de cada indivíduo envolvido no estudo com base no nível hierárquico 2 obtido do banco de dados KEGG. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg.

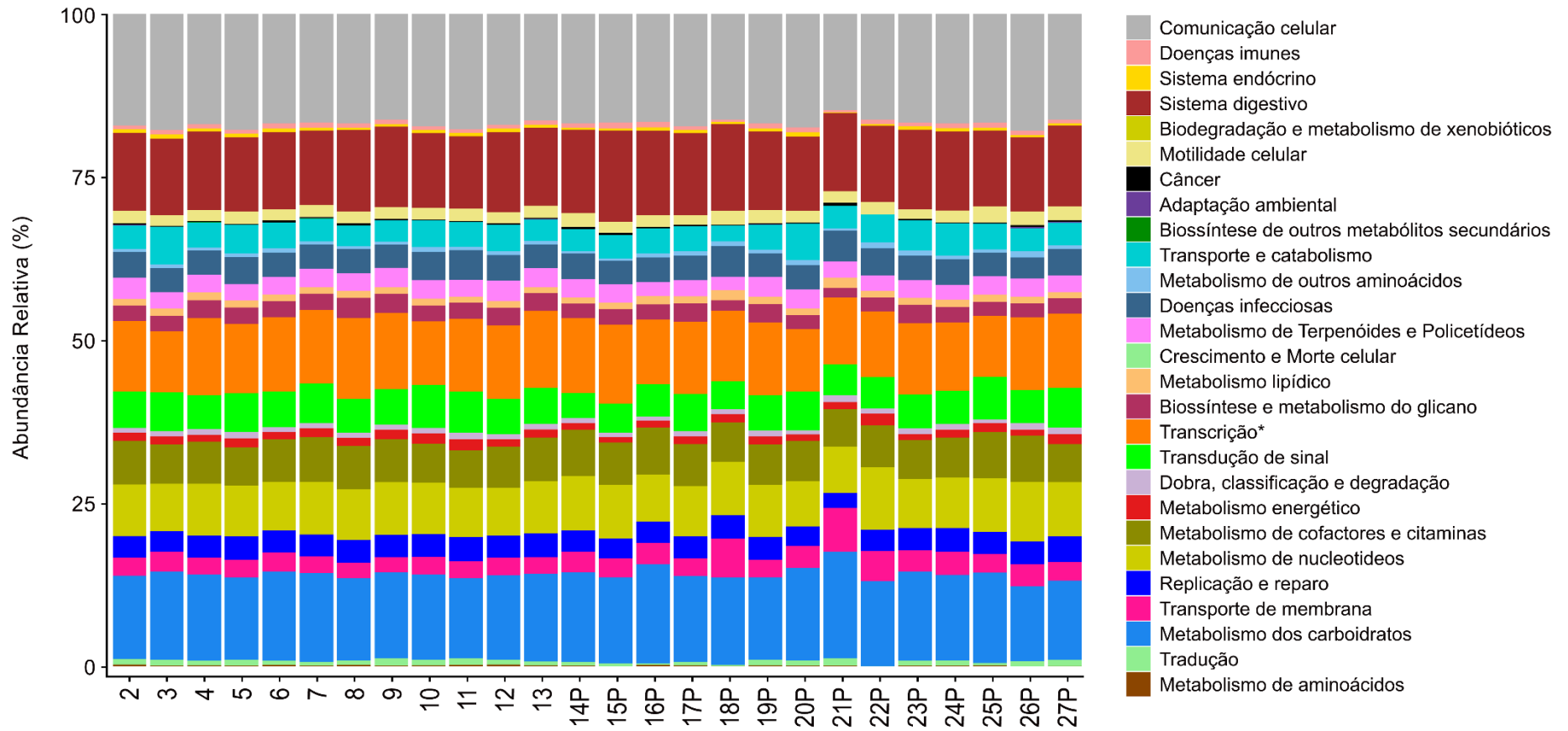
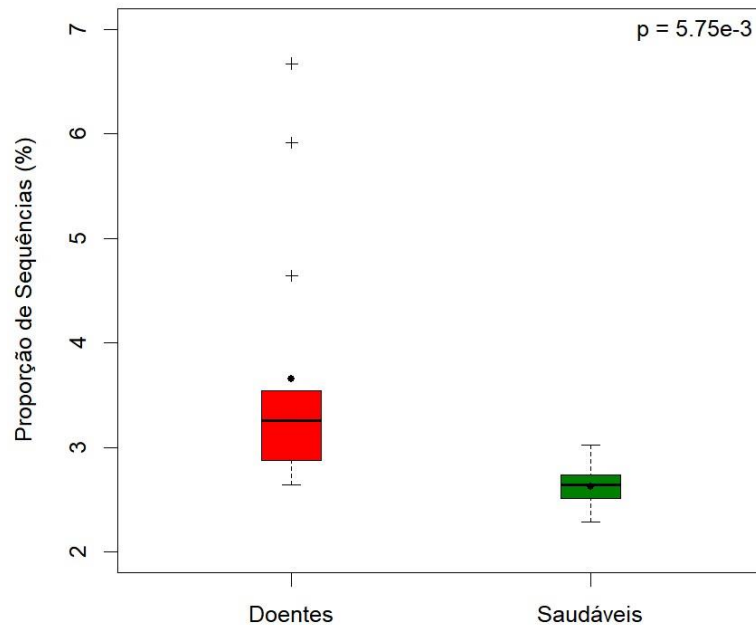


Figura 37. *Boxplot* demonstrando a distribuição na proporção categoria “Transcrição” com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzes. Dados de abundância obtidos da análise funcional a partir do banco de dados KEGG, nível hierárquico 2, realizada pelo servidor MG-RAST.



Os resultados obtidos a partir do nível hierárquico 3 apresentaram um total de 124 categorias distintas. A Análise de Componentes Principais (PCA), quando observado o PC1, demonstrou uma melhor separação dos grupos em comparação com as PCAs executadas nos níveis hierárquicos anteriores (Figura 38). Das categorias obtidas, destacam-se “transportadores ABC” e “biossíntese de aminoacil-RNA<sup>t</sup>”, com maiores abundâncias durante o estudo. Entretanto, somente 4 categorias obtiveram diferença significativa com maior abundância no grupo doente (Tabela 19).

Figura 38. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico 3 (KEGG) dos grupos doente (azul) e saudável (laranja).

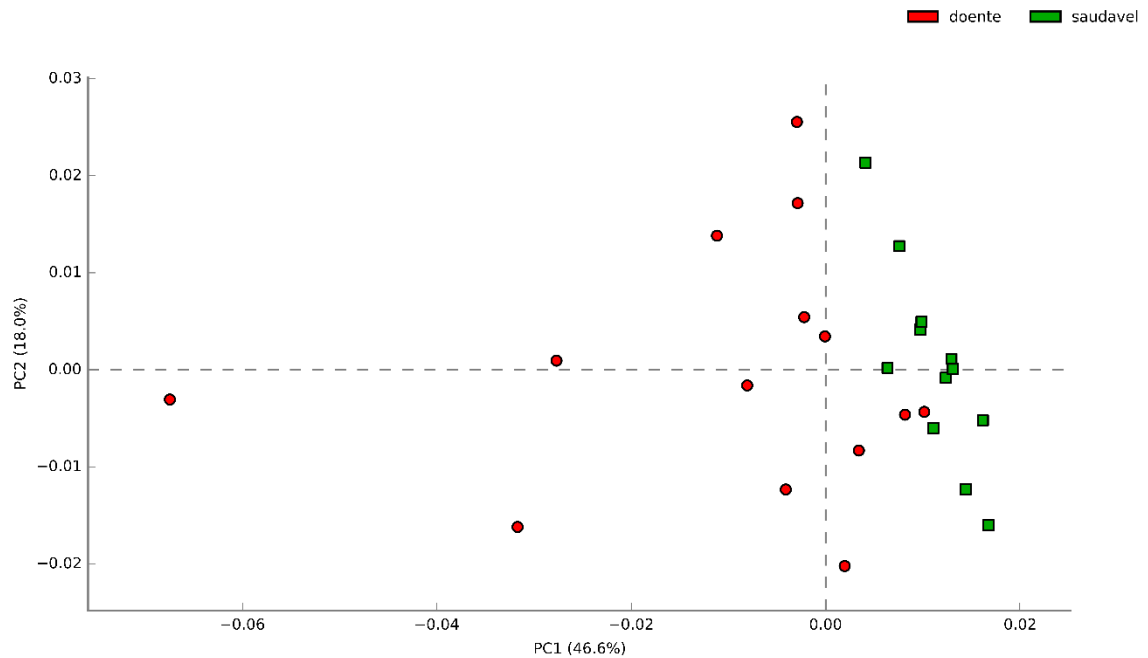
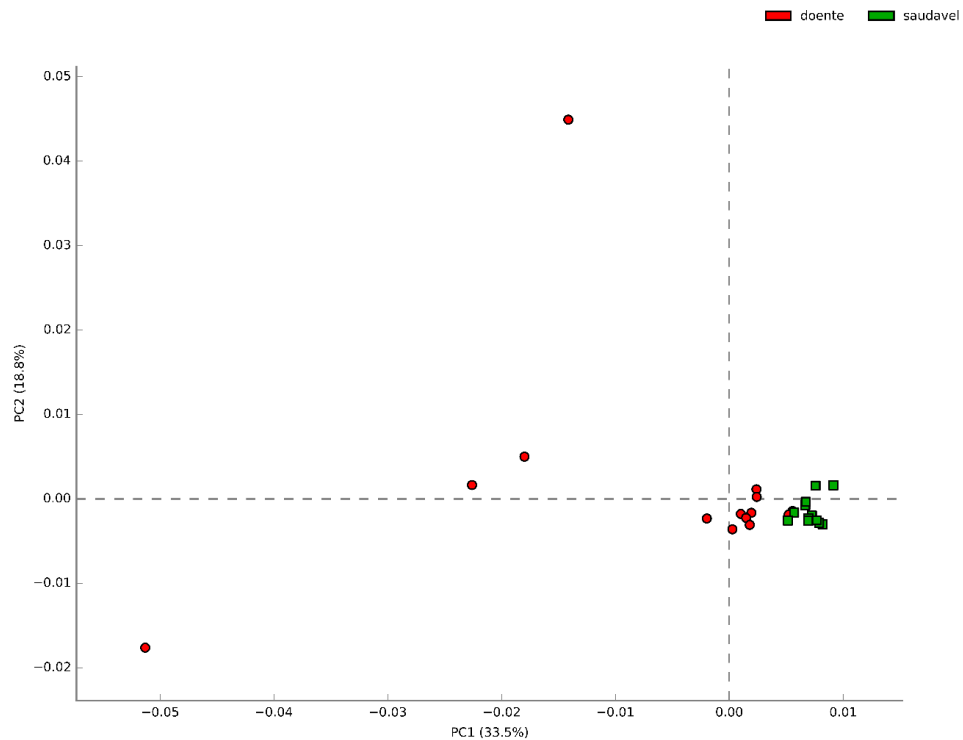


Tabela 19. Abundância relativa das categorias obtidas do nível hierárquico 3 do banco de dados KEGG que apresentaram diferença estatística significativa entre os grupos de estudo para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg.

Nível hierárquico 3	%Saudáveis	%Doentes	p-values
Biossíntese de aminoacil-RNAt	8,03	8,63	1,08E-02
RNA polimerase	2,57	3,01	8,80E-03
Metabolismo do piruvato	1,99	2,28	3,14E-02
Vias de fixação de carbono em procariontes	0,42	0,49	1,18E-02

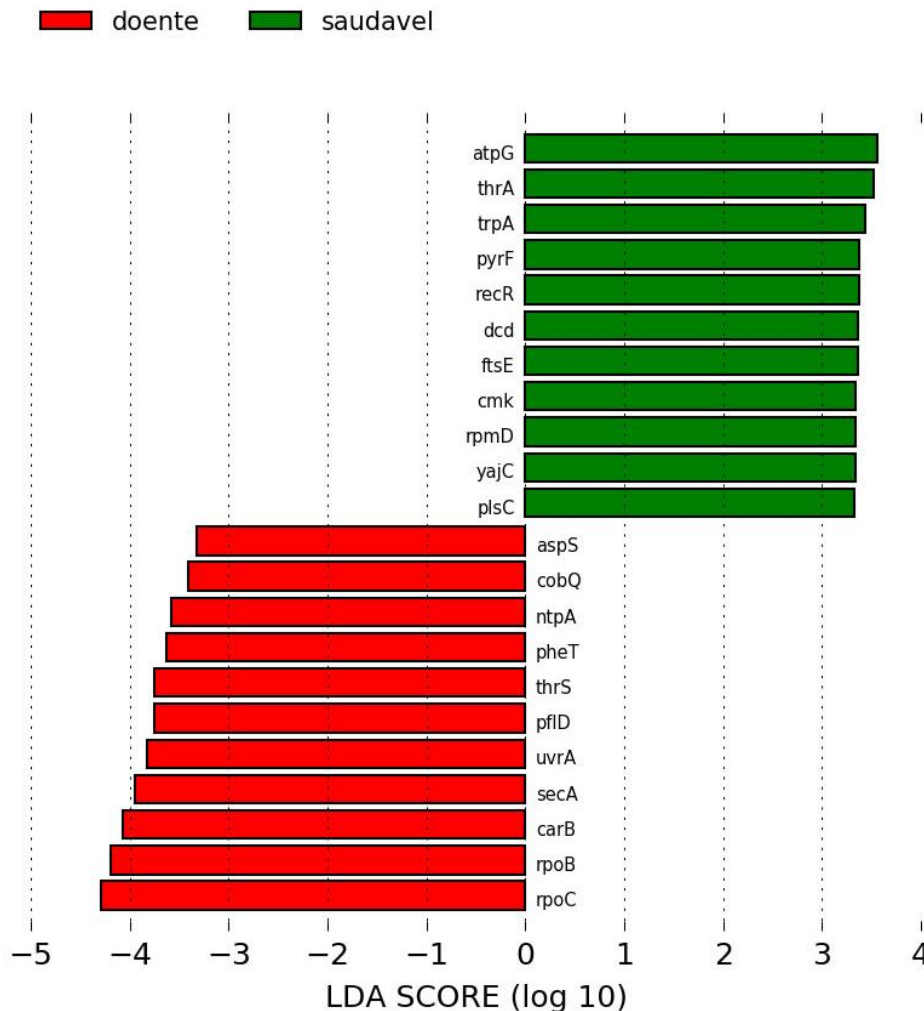
Um total de 1343 genes foram encontrados durante a análise através do banco de dados KEGG para o nível hierárquico “função”. A Análise de Componentes Principais (PCA) mostrou agrupamento de amostras provenientes do mesmo grupo, sugerindo distinção entre grupos de estudo (Figura 39).

Figura 39. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para o nível hierárquico “função” (KEGG) dos grupos doente (azul) e saudável (laranja).



A partir do total de genes obtidos, 98 apresentaram diferença estatística significativa para o teste Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. Após aplicar a Análise de Discriminantes Lineares (LDA), apenas 2 genes não passaram no teste, restando 96 genes. O resultado final, portanto, revelou 85 genes associados ao grupo saudável, enquanto 11 genes foram associados ao grupo doente (Figura 40).

Figura 40. Análise de Discriminantes Lineares (LDA) aplicada aos dados de abundância relativa dos resultados obtidos para o nível hierárquico “função” (KEGG) com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg. Em vermelho, todos os genes resultantes para o grupo doente. Em verde, apenas os 11 genes que obtiverem as melhores pontuações para o grupo saudável.



## 5.8 ANOTAÇÃO FUNCIONAL VIA CAZy

Através da análise realizada no banco de dados CAZy, obteve-se um total de 1.598.407 *hits* distribuídos em 569.575 para o grupo doente e 1.028.832 para o grupo saudável (Tabela 20). Esses resultados estão classificados entre famílias e subfamílias CAZy. As famílias onde os resultados encontram-se distribuídos são: Glicosídeos Hidrolases (GH), Glicosiltransferases (GT), Polissacarídeo Liases (PL), Módulos de Ligação a Carboidratos (CBM), Esterases de Carboidratos (CE) e Atividades Auxiliares (AA).

Tabela 20. *Open Read Frames* (ORFs) anotadas em banco de dados CAZy através do programa DIAMOND.

Amostra	<i>Open Read Frames</i>		
	Submetidos	Anotados	% Anotados
2	3.403.497	77.871	2,29
3	2.531.841	42.117	1,66
4	4.240.819	97.271	2,29
5	3.263.377	68.286	2,09
6	2.882.044	40.763	1,41
7	6.688.384	174.407	2,61
8	4.975.781	144.560	2,91
9	3.498.422	102.379	2,93
10	4.806.217	97.608	2,03
11	4.155.435	74.532	1,79
12	1.636.140	35.961	2,20
13	3.415.868	73.077	2,14
14P	2.860.341	66.025	2,31
15P	1.423.592	27.388	1,92
16P	2.158.242	34.806	1,61
17P	7.308.104	138.775	1,90
18P	994.862	15.789	1,59
19P	2.860.354	51.324	1,79
20P	4.570.297	56.471	1,24
21P	969.250	11.751	1,21
22P	499.960	6.692	1,34
23P	1.687.514	23.192	1,37
24P	1.825.933	31.354	1,72
25P	4.208.027	79.669	1,89
26P	384.511	4.424	1,15
27P	1.566.411	21.915	1,40

Ao observar as abundâncias entre os grupos de estudo, notou-se que as famílias GH, GT e CBM são detentoras de maior abundância em ambos os grupos. Embora a Análise de Componentes Principais (PCA) não mostre uma clara separação espacial entre os indivíduos dos diferentes grupos de estudo (Figura 41), observou-se mediante análise estatística que a família Polissacarídeo Lases (PL) apresentou diferença estatística significativa quando comparada as abundâncias entre os diferentes grupos de estudo (Tabela 21, Figura 42).

Tabela 21. Abundância relativa das Famílias CAZy obtidas do banco de dados CAZy para cada grupo de estudo. O símbolo “\*” indica os resultados com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg.

Família	%Saudáveis	%Doentes	p-values
Glicosídeos Hidrolases (GH)	40.14	42.15	1.52E-01
Glicosiltransferases (GT)	37.30	37.77	9.18E-01
Módulos de Ligação à Carboidratos (CBM)	17.90	15.21	6.16E-01
Esterases de Carboidratos (CE)	4.53	4.60	5.60E-01
Polissacarídeo Liases (PL)*	0.06	0.21	1.56E-03
Atividades Auxiliares (AA)	0.07	0.07	9.56E-01

Figura 41. Análise de Componentes Principais (PCA) realizada através do programa STAMP a partir dos dados de abundância relativa dos resultados obtidos para a categoria “Família” (CAZy) dos grupos doente (azul) e saudável (laranja).

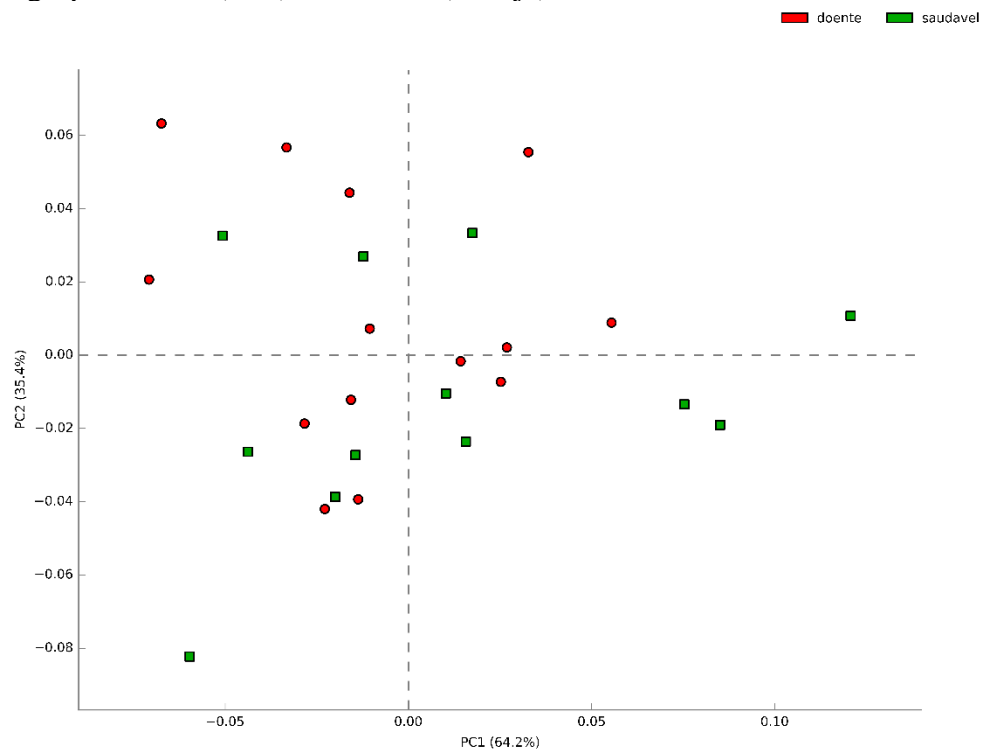
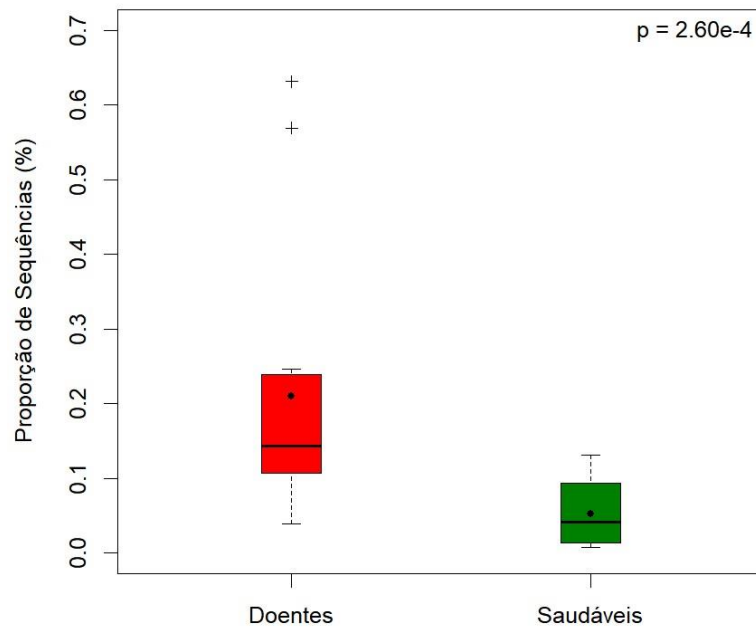


Figura 42. *Boxplot* demonstrando a distribuição na proporção da família Polissacarídeo Liasés (PL) com diferença significativa para o teste estatístico de Kruskal-Wallis ( $p < 0,05$ ), corrigido pelo método Benjamini-Hochberg, entre os grupos doente (vermelho) e saudável (verde). O valor da mediana é representado por uma linha no interior da caixa, a média por um ponto preto e os *outliers* por cruzeiros. Dados de abundância obtidos a partir da análise funcional a partir do banco de dados CAZy, nível de família, realizada pelo programa DIAMOND (Blastp).





## 6 DISCUSSÃO

O corpo humano é composto por diferentes microambientes capazes de abarcar microbiotas específicas que, comumente, possuem uma estreita relação adaptativa com os mesmos. A composição dessas comunidades microbianas no corpo humano pode variar entre indivíduos e ser observada em diferentes estudos (GRICE et al. 2008; LOZUPONE et al. 2012).

Os resultados encontrados no presente estudo sugerem que há uma composição taxonômica característica da cavidade oral. Tal composição apresenta uma predominância do domínio *Bacteria* com relativamente poucos representantes dos domínios *Archaea*, *Eukaryota* e *Virus*. De acordo com os resultados obtidos a partir do programa Metaphlan 2.0, mais de 99% do perfil taxonômico da cavidade oral poderia ser representado pelos filos bacterianos *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Tenericutes*. AAS et al. (2005) e DEWHIRST et al. (2010), analisando sequências do gene 16S rRNA obtidas a partir de cavidade oral em diferentes sítios e condições clínicas, obtiveram resultados similares, contribuindo para a afirmação de que o domínio *Bacteria*, assim como os filos obtidos neste estudo, constituem a maior parcela da microbiota presente na cavidade oral.

De modo geral, os resultados em nível de filo apontam para um perfil taxonômico similar quanto a presença e ausência de táxons, tanto para o estado saudável, quanto doente. Esses resultados são tema de discussões que envolvem a busca por um microbioma central, denominado na literatura como *core microbiome*. Acredita-se que exista um *core microbiome* composto por genes ou espécies que todos compartilhamos. Uma vez definido este *core microbiome*, seria possível compreender melhor o microbioma humano como um todo (HAMADY, KNIGHT, 2009).

O caminho em direção à busca pelo *core microbiome* oral, fazendo uso de seu conceito, foi iniciado por Zaura et al. (2009), que definiu com clareza a presença dos filos *Firmicutes*, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* e *Fusobacteria* como integrantes efetivos. Foi proposto que a abundância de filotipos pertencentes a estes filos estaria relacionada a estabilidade funcional e homeostase necessárias para que houvesse um ecossistema saudável. Em nosso estudo, foi observada a presença dos filos propostos por Zaura et al. (2009), porém com variações nas abundâncias nos estados doente e saudável, sugerindo que, embora constata-se um *core microbiome*, variações na abundância dos filotipos presentes podem interferir no ecossistema e refletir diretamente na homeostase oral.

Na análise realizada através do programa Metaphlan 2.0, nota-se que os filos *Actinobacteria* e *Proteobacteria* apresentaram uma maior abundância para o grupo saudável, 32,12% e 20,79%, respectivamente (Kruskal-Wallis,  $p < 0,05$ ) corroborando achados anteriores (CHEN et al. 2018). Por sua vez, o grupo doente apresenta a maioria dos filos obtidos com uma abundância maior que o grupo saudável tanto na análise via Metaphlan 2.0 (Tabela 6) quanto MG-RAST (Tabela 15). Destacam-se os filos *Bacteroidetes*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Synergistes* com abundâncias de 26,13%, 5,63%, 1,01%, 0,9%, 0,09%, respectivamente (Kruskal-Wallis,  $p < 0,05$ ).

Vários organismos do filo *Bacteroidetes* podem ser encontrados colonizando animais e seres humanos sem apresentar complicações. Esse filo é conhecido por serem bons em degradação de matéria orgânica polimérica, em especial carboidratos e proteínas. Já é descrito na literatura a presença do filo *Bacteroidetes* tanto na saliva quanto em placa dentária de indivíduos saudáveis ou doentes (KEIJSER et al. 2008; THOMAS et al. 2011). Contudo, é interessante salientar que este filo apresenta uma estreita relação com o estado de doença periodontal, uma vez que nele estão contidos vários gêneros gram-negativos que atuam na doença. *Prevotella*, *Capnocytophaga*, *Porphyromonas* e *Tannerella*, todos presentes no grupo doente em nosso trabalho, correspondem aos gêneros mais importantes descritos na literatura que atuam durante o estabelecimento da periodontite (PÉREZ-CHAPARRO, GONÇALVES, FIGUEIREDO, 2014).

Organismos do filo *Fusobacteria* também podem ser encontrados na cavidade oral tanto de indivíduos saudáveis quanto doentes. Para o estudo da cavidade oral, ganha destaque o gênero *Fusobacterium* pertencente à família *Fusobacteriaceae*, composta por indivíduos obrigatoriamente gram-negativos que habitam a cavidade oral atuando como patógenos oportunistas (HAN, 2015). Seu papel durante o estabelecimento das doenças orais está relacionado à capacidade de coagregação que seus membros têm com diferentes organismos que habitam a cavidade oral (KOLENBRANDER et al. 2006). Dessa forma, a característica de coagregação auxilia consideravelmente o crescimento de biofilme e, portanto, o estabelecimento de doenças orais como infecções endodônticas, gengivite, periodontite agressiva e periodontite crônica (SIQUEIRA et al. 2007; GRIFFEN et al. 2012).

O filo *Spirochaetes* é notável em infecções que causam doenças periodontais. Embora algumas espécies deste filo já tenham sido cultivadas, a maioria ainda permanece sem abordagens de cultivo estabelecidas, por serem organismos com alta dependência metabólica do hospedeiro. Durante o estabelecimento da periodontite, o gênero *Treponema* é o mais conhecido. Nele estão compreendidas espiroquetas orais que apresentam uma ampla

variedade de fatores de virulência e medeiam a penetração, destruição tecidual e evasão de respostas imunes do hospedeiro. Dentre as espécies do gênero, destaca-se a *Treponema denticola* que, além de apresentar as características já citadas, demonstra um grande potencial de coagregação com espécies do filo *Bacteroidetes* e *Fusobacteria*, também presentes no biofilme responsável pela periodontite (CHAN, MCLAUGHLIN, 2000; VISSER, ELLEN, 2011).

Os membros que compõem o filo *Synergistes* são pouco conhecidos. Contudo, sabe-se que esses organismos são caracteristicamente gram-negativos, anaeróbios, bons degradadores de aminoácidos e apresentam ampla distribuição na natureza. A identificação de táxons desse filo em amostras provenientes de pacientes portadores de infecção sugere um possível papel patogênico. A presença do filo *Synergistes* em doenças como periodontite, cáries e infecções endodônticas já foi relatado anteriormente. No biofilme periodontal, é possível encontrar microrganismos que degradam substratos proteicos complexos (por exemplo, filo *Bacteroidetes*), enquanto outros degradam aminoácidos disponibilizados (*Fusobacterium* e *Peptostreptococcus*). Neste contexto, é possível encontrar um nicho similar, onde os microrganismos do filo *Synergistes* possam atuar (SONIA et al. 2007).

De acordo com Bor et al. (2019) as bactérias do filo *Candidatus Saccharibacteria* (antigo TM7) estão comumente presentes na cavidade oral e são encontradas em baixa abundância. No entanto, um aumento de membros de *Candidatus Saccharibacteria* foi associado com a presença de gengivite e doença periodontal (CAMELO-CASTILLO et al. 2015). As espécies desse filo têm como principal característica o tamanho ultrapequeno, vivendo, portanto, em parasitose com sua bactéria hospedeira. Além disso, estudos genômicos revelam que há ausência de caminhos metabólicos para a biossíntese de nucleotídeos, lipídios e aminoácidos. Dessa forma, *Candidatus Saccharibacteria* consiste de táxons auxotróficos e metabolicamente dependentes de outros organismos. Seu papel na doença periodontal ainda não foi esclarecido; contudo, por ser uma bactéria que parasita outras bactérias, acredita-se que estes organismos possam atuar modulando a microbiota do biofilme (BOR et al. 2019).

Pérez-Chaparro, Gonçalves e Figueiredo (2014) citam o filo *Euryarcheota* proveniente do domínio *Archaea* e domínio *Eukaryota*, reino *Fungi* como atuantes na periodontite. O filo *Euryarcheota* foi obtido em nosso estudo através da análise pelo servidor MG-RAST com abundância relativa para os grupos doente e saudável nos seguintes valores: 0,00024% e 0,000062%, respectivamente. Ainda assim, não houve diferença estatística significativa.

Embora os filios sejam um importante ponto de partida para compreender melhor a microbiota envolvida na periodontite, é possível encontrar táxons presentes nesses filios que

necessariamente não estejam relacionadas a doença (AAS et al. 2005; ZAURA et al. 2009). Portanto, faz-se necessária a observação dos demais níveis taxonômicos. Ainda assim, é possível afirmar que, mediante os resultados observados, altos valores de abundâncias para os filos *Bacteroidetes*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Synergistes* podem atuar como indicativo ou sinal de alerta para o desenvolvimento de um quadro de periodontite. O domínio *Virus* e demais filos bacterianos, *Firmicutes* e *Tenericutes*, não apresentaram diferença estatística significativa ao nível de filo.

De acordo com Nazidze et al. (2009) e Ling et al. (2010), os gêneros *Streptococcus*, *Prevotella*, *Veillonella*, *Neisseria*, *Haemophilus*, *Rothia*, *Porphyromonas* e *Fusobacterium* juntos podem chegar a representar > 70% de todos os táxons presentes. Em nosso estudo, foi possível encontrar todos esses gêneros. Entretanto, ao observar a Tabela 7 e a Tabela 8, notamos que esses gêneros apresentaram um valor menor no grupo doente (67,56%), enquanto que no grupo saudável, tal valor correspondeu a 87,15%. Embora não representado na Tabela 8, o gênero *Fusobacterium* teve como abundância relativa 1,04% no grupo saudável.

É conhecido na literatura que os gêneros *Rothia*, *Porphyromonas*, *Prevotella*, *Capnocytophaga*, *Neisseria* e *Treponema* podem estar envolvidos na suscetibilidade de um indivíduo à doença periodontal (TAKESHITA et al. 2009; READY et al. 2008; YILMAZ, 2008). Todos estes gêneros foram encontrados em nosso estudo, no qual *Capnocytophaga* (1,84%) e *Treponema* (0,9%) apresentaram maior abundância para os doentes com diferença estatística significativa. Os demais que apresentaram maior abundância no grupo doente, porém sem diferença estatística significativa foram: *Prevotella* (14,12%) e *Porphyromonas* (7,1%). Curiosamente os gêneros *Rothia* e *Neisseria*, embora relacionados à susceptibilidade à doença periodontal apresentaram maior abundância no grupo saudável, respectivamente, 27,72% ( $p < 0,05$ ) e 6,51%. Quando se observa os resultados da Análise de Discriminantes Lineares (LDA) na

Figura 15, fica ainda mais claro a presença do gênero *Rothia* ao grupo saudável.

Os dados exploratórios e estatísticos com base nas diferentes análises realizadas possibilitaram a criação de um grupo forte de gêneros candidatos que possivelmente estejam envolvidos na periodontite, sendo eles: *Eubacterium*, *Leptotrichia*, *Corynebacterium*, *Treponema*, *Parvimonas*, *Fusobacterium*, *Filifactor*, *Capnocytophaga*, *Tannerella*, *Bifidobacterium*, *Peptostreptococcus*, *Dialister*, *Catonella*, *Selenomonas*, *Campylobacter* e *Porphyromonas*. Todos esses gêneros apresentam espécies descritas na literatura como

relacionadas com a periodontite (SOCRANSKY, 2005; CHAPARRO, GONÇALVERS E FIGUEIREDO, 2014).

A Análise de Discriminantes Lineares (LDA) na Figura 16, demonstrou uma separação entre os grupos de estudo no que diz respeito a composição e/ou abundância de espécies encontradas nos mesmos. daquelas espécies descritas na literatura como sendo estabeledoras da periodontite, foram encontradas em nosso trabalho *Porphyromonas gingivalis*, *Treponema denticola* e *Tannerella forsythia*. Os resultados para essas espécies sugerem que de fato há uma estrita relação com indivíduos portadores de periodontite. Por esse motivo, altas abundâncias dessas espécies em pacientes podem atuar como um forte indicador do estabelecimento da periodontite. Contudo, não podemos esquecer que uma cadeia de eventos é necessária até que essas espécies se estabeleçam. Socransky, Haffajee (2002, 2005), definiram complexos que orientam como espécies podem se comportar, assim como, qual a contribuição de cada complexo durante o estabelecimento da periodontite. Com base na Análise de Discriminantes Lineares (LDA) apresentada na Figura 16, mais de 50 espécies encontradas em nosso estudo podem estar envolvidas com a periodontite.

Baseando-se no diagrama elaborado por Socransky, Haffajee (2005), observado na Figura 2, é possível notar que as espécies *V. parvula*, *C. ochracea*, *C. gracilis*, *C. rectus*, *F. nucleatum*, *P. intermedia*, *P. gingivalis*, *T. forsythia* e *T. denticola* foram encontradas em nosso trabalho com maior abundância para o grupo doente (Kruskal-Wallis,  $p < 0,05$ ). Enquanto, *C. showae* e *F. nucleatum* foram encontradas com maior abundância para o grupo saudável (Kruskal-Wallis,  $p < 0,05$ ). Outras espécies que podem ser mescladas com o trabalho de Socransky, Haffajee (2005) e que estiveram presentes em nossas análises, porém sem diferença significativa entre os grupos, foram: *A. naeslundii*, *S. gosdonii*, *S. intermedius*, *S. sanguinis*, *C. gingivalis*, *E. corrodens*, *A. odontolyticus*, *P. nigrescens* e *S. constellatus*.

Durante o estudo dos complexos propostos por Socransky, Haffajee (2002, 2005), nota-se que a maior parcela das espécies responsáveis pela colonização primária é do gênero *Streptococcus*, o qual foi abundante tanto no grupo saudável, quando no doente não apresentando diferença estatística para a maioria das espécies desse gênero. Essa elevada abundância do gênero *Streptococcus* é esperada na cavidade oral, como observado nos estudos de AAS et al. (2005) e DEWHIRST et al. (2010).

As espécies do gênero *Veillonella* tem a capacidade de interagir com espécies do gênero *Streptococcus*. Essa interação ocorre através de adesão facilitando posteriormente a associação com outros microrganismos presentes na cavidade oral vindo a ocasionar o crescimento da placa dentária. Por conta dessa característica, é comum encontrarmos espécies

do gênero *Veillonella* em maior abundância em pacientes com periodontite estabelecida (MASHIMA, NAKAZAWA, 2015). Em nosso estudo foram encontradas *V. atypica* (0,98%), *V. díspar* (0,87%) e *V. parvula* (1,3%) em ambos os grupos, sendo esta abundância em destaque proveniente do grupo doente, o qual sempre apresentou os maiores valores. Apenas a espécie *Veillonella sp oral taxon 780* apresentou maior abundância no grupo saudável. Essa espécie em questão já foi encontrada previamente habitando a cavidade oral em condições que indicam a presença de cáries dentárias (JIANG et al. 2016). Contudo, o resultado apresentado para essa espécie no grupo saudável é proveniente de apenas uma amostra sendo, portanto, considerado um caso isolado.

As espécies do gênero *Capnocytophaga* fazem parte da microflora oral normal do homem e também de animais (KATO et al. 2011), considerados oportunistas, costumam estar associados a doenças periodontais podendo ocasionar até mesmo infecção sistêmica em hospedeiros imunocompetentes e imunocomprometidos (JOLIVET-GOUGEON et al. 2007). Algumas espécies do gênero *Capnocytophaga* são consideradas colonizadoras precoces durante a formação da placa dentária como é o caso de *C. ochracea* (0,07%) encontrada em nosso estudo com diferença estatística e maior abundância no grupo doente. Esta espécie é capaz de se agregar inicialmente a outras do gênero *Streptococcus* e *Actinomyces*. Além disso, colonizadores tardios ainda são capazes de se utilizar *C. ochracea* como ponto de adesão. Por esses motivos, *C. ochracea* é considerada importante durante a formação do biofilme dentário (KOLENBRANDER et al. 2006).

De acordo com Henne et al. 2014, *Campylobacter* spp. está presente em comunidades orais tanto em indivíduos saudáveis quanto portadores de periodontite. Entretanto, é possível observar aumento *C. gracilis* e *C. rectus* quando o indivíduo é portador de periodontite e menor abundância de *C. concisus* em indivíduos saudáveis. De fato, observamos a mesma característica em nosso trabalho tendo *C. gracilis* e *C. rectus* com 0,21% e 0,03% de abundância, respectivamente, para o grupo doente com diferença significativa. Enquanto *C. concisus* apresentou maior abundância para o grupo saudável (0,07%) e menor para o doente (0,05%). Embora descrita anteriormente na literatura como sendo relacionada a gengivite e periodontite, a espécie *C. showae* apareceu em nosso estudo com maior relação ao estado saudável com uma abundância de 0,03% ( $p < 0,05$ ) sendo que para o estado doente permaneceu em 0,02%.

*Fusobacterium nucleatum*, alocado no complexo laranja por Socransky, Haffajee (2005), pode ser considerado um organismo central para a relação entre colonizadores primários e colonizadores tardios durante a formação da placa dentária. Neste estudo, *F.*

*nucleatum* apresentou abundâncias de 1,49% para os doentes e 0,17% em saudáveis com  $p < 0,05$ . *F. nucleatum* é um microrganismo capaz de estabelecer relações de coagregação com diversos outros microrganismos durante a formação da placa dental (KOLENBRANDER et al. 2006; HAN, 2015).

Algumas espécies que já foram citadas na literatura estando relacionadas ao estado de doença periodontal foram mais prevalentes em saudáveis em nosso estudo, destacando-se: *Fusobacterium periodonticum*, *Rothia mucilaginosa* e *Porphyromonas sp oral táxon 279*. Os motivos que levaram a maior prevalência das mesmas no grupo saudável permanecem incógnitos. Pacientes com doença periodontal abrigam níveis elevados de microrganismos periodontopatógenos na saliva tendo uma elevação na área da bolsa periodontal. Contudo, é importante notarmos que a cavidade oral é composta por diferentes regiões e que cada uma é capaz de abarcar uma microbiota distinta. A saliva por sua vez está em contato com a maior parte das regiões presentes na cavidade oral. Portanto, a saliva abarca microrganismos que definem a microbiota oral juntamente com aqueles que seguem em desequilíbrio de abundância em estado de doença periodontal (SAYGUN et al. 2011). Nesse contexto, a presença pontual de microrganismos dos complexos distintos do vermelho, proposto por Socransky, Haffajee (2005), em indivíduos saudáveis, pode ocorrer em análises que façam uso da saliva uma vez que a mesma abarca o conjunto de microrganismos presente na cavidade oral.

A espécie não usual *Porphyromonas sp oral taxon 279*, cujo gênero é similar a de uma representante do complexo vermelho (*P. gingivalis*), também obteve maiores abundâncias no grupo saudável com  $p < 0,05$ . A literatura se demonstrou vaga quanto ao conhecimento gerado a respeito desta espécie em questão. Contudo, Yost et al. (2015) também encontrou *Porphyromonas sp oral táxon 279* relaciona a periodontite. O'Flynn et al. (2015) identificou que *Porphyromonas sp oral táxon 279* apresenta uma maior flexibilidade metabólica em comparação com *P. gingivalis*. Essa flexibilidade foi observada mediante a presença quase completa da via metabólica para a síntese de protoporfirina IX. As dez enzimas que compõem essa via trabalham na conversão do L-glutamato em protoheme. Enquanto *P. gingivalis* apresenta apenas três enzimas, *Porphyromonas sp oral táxon 279* apresenta oito.

Durante análise da riqueza é possível notar que o estado saudável e doente compartilham a grande maioria das espécies encontradas nesse estudo (Figura 26). Os resultados de riqueza e diversidade (índices de Shannon e Simpson) mostram que o grupo doente apresentou valores maiores para ambas as métricas em comparação com o grupo saudável. A diversidade beta calculada pelo índice Whittaker demonstrou que a

heterogeneidade na composição de espécies é muito parecida entre os grupos de estudo. Ainda assim, a heterogeneidade foi 0,02 maior para grupo doente com o uso do índice de Whittaker. Uma vez que comunidades microbianas que apresentam diversidade e riqueza baixas abrigam patógenos, simbioses e patobioses em frequências baixas e proporções que garantem o estado de saúde. Contudo, quando ocorrem perturbações ambientais nos tecidos periodontais ou o hospedeiro apresenta susceptibilidade genética, os patógenos-chaves podem provocar inflamação capaz de alterar a base de nutrientes da bolsa periodontal. Desta forma, ocorre a expansão proporcional de patobioses em relação aos simbioses, de forma que a inflamação causada leve à destruição do tecido conjuntivo e tecido ósseo. O resultado do processo, portanto, é um aumento de riqueza e diversidade em portadores de periodontite (COSTALONGA, HERTZBERG, 2014).

Uma espécie que se destacou durante a pesquisa foi *Filifactor alocis*. Essa espécie vem sendo descrita na literatura (ARUNI, 2015) como sendo uma nova contribuinte para a doença periodontal. Em nosso estudo encontramos uma correlação positiva forte desta espécie tanto com *Porphyromonas gingivalis* (+0,73) quanto com *Treponema denticola* (+0,86). Além dessa correlação, *F. alocis* ainda ficou entre as 10 espécies com maior relação ao grupo doente (Figura 16). Dessa forma, é possível que *F. alocis* esteja desenvolvendo um papel importante no estabelecimento das espécies do gênero *Porphyromonas*, em destaque *P. gingivalis*, assim como para a espécie *T. denticola*. As espécies *P. gingivalis* e *T. denticola* por sua vez apresentaram correlações positivas fortes, porém, observando-se os indivíduos doentes 20P e 22P (Figura 14), nota-se que a ausência de uma das espécies não é suficiente para que a periodontite não se estabeleça.

A análise funcional utilizando o banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG) apontou para uma similaridade nos perfis funcionais das amostras envolvidas no estudo (Figura 36). Essa similaridade reflete tanto quando comparadas amostras do mesmo grupo quanto entre os grupos ao observar níveis funcionais elevados. Essa similaridade em níveis funcionais elevados nos leva a interpretar que a cavidade oral apresenta um perfil funcional conservado. JORTH et al. (2017) também observou tal característica ao notar que, embora haja variação no perfil taxonômico em comunidades associadas à saúde e doença, a presença de determinadas vias metabólicas se repete em diferentes indivíduos, denotando uma característica de conservação. Contudo, na medida em que a análise atinge níveis funcionais mais profundos notamos que os grupos de estudos tendem a se distinguir, fato este que pode ser observado durante a análise exploratória PCA em nosso estudo.



De modo geral, a categoria “Metabolismo” ganhou destaque por ter apresentando o maior valor de abundância, correspondendo a >50% dos resultados obtidos. Esses dados, por sua vez, distribuem-se em três principais subcategorias: metabolismo de aminoácidos, metabolismo de carboidratos e metabolismo de nucleotídeo. É observado, portanto, que tanto no estado de doença quanto no estado saudável, a cavidade oral é predominada pela categoria funcional “Metabolismo”, uma vez que há uma abundância maior que as demais ao mesmo tempo em que não há diferença estatística significativa. Essa predominância é seguida pelas categorias “Processamento de informação genética” e “Processamento de informação ambiental”.

A categoria “Processamento de informação genética” foi a única de nível 1 que apresentou diferença estatística significativa. Além disso, sua abundância é maior no grupo doente. Nesta categoria estão inclusas: Dobra, classificação e degradação; Replicação e reparo; Transcrição; Tradução. Dentre estas subcategorias, “Transcrição” foi a única que apresentou maior abundância para o grupo doente com diferença estatística significativa. Ao observarmos os resultados da Tabela 19, nota-se que, em sincronia com a abundância no processo de transcrição, tem-se uma maior abundância nas categorias que envolvem biossíntese de aminoacil-RNAt e RNA polimerase também com diferença estatística significativa. Essa abundância acontece em grande parte pelos genes *rpoC* e *rpoE* codificantes das subunidades (beta e delta, respectivamente) da RNA polimerase bacteriana e *aspS*, *thrS* e *pheT* que codificam a sintetases de RNAt para aspartato, treonina e fenilalanina. Esses genes ainda se apresentaram fortemente relacionados a doença como mostra a análise LDA (Figura 40). Durante o processo e o estabelecimento do biofilme periodontal, ocorre disbiose e sinergismo entre os microrganismos (TAN et al. 2014). Dessa forma, o dogma central da biologia molecular pode ocorrer de forma intensificada por indução devido à comunicação entre as bactérias periodontais durante a infecção. A comunicação por sinergismo gera respostas transcriptômicas e proteômicas recíprocas capazes de promover a aquisição de nutrientes, processos metabólicos e fatores de virulência (KUBONIWA et al. 2012; HAJISHENGALLIS, 2015).

Embora a categoria “vias de fixação de carbono em procariontes” também tenha se apresentado com maior abundância no grupo doente e com diferença estatística significativa (Tabela 19), apenas 2 genes se mostraram presentes: *fhs* e *folD*. Esses genes correspondem, respectivamente, às enzimas “formate-tetrahydrofolate ligase” e “metilenetetrahydrofolate desidrogenase”. A via metabólica na qual esses genes atuam é conhecida como “Wood–Ljungdahl” ou “via redutora de acetil-coenzima A”. Essa via é utilizada por organismos

acetogênicos que, portanto, são capazes de produzir acetil-coenzima a partir de hidrogênio (H) e dióxido de carbono (CO<sub>2</sub>) (SCHUCHMANN, MÜLLER, 2016). Esse processo ocorre em meio anaeróbico, sendo a bolsa periodontal um ambiente amplamente propício. Em nosso trabalho, foram encontradas, na análise pelo servidor MG-RAST, sequências relacionadas taxonomicamente a gêneros conhecidos por apresentarem organismos acetogênicos, como: *Acetobacter*, *Acetohalobium*, *Clostridium*, *Enterobacter*, *Enterococcus*, *Pelobacter* (ABREU, 2007). O estudo de organismos acetogênicos na periodontite pode ser um ponto interessante, uma vez que poucos trabalhos foram realizados com esse foco. Alguns trabalhos, embora com pouca abundância, relataram a presença de *Archaeas* metanogênicas em bolsa periodontal (HORZ, CONRADS, 2011). Sabendo que as arqueias metanogênicas utilizam o acetato para produzir metano (ABREU, 2007), é possível que haja uma relação desses grupos na periodontite. Embora tenhamos encontrado sequências relacionadas taxonomicamente a alguns gêneros de arqueias metanogênicas e bactérias acetogênicas, a abundância foi diminuta e em indivíduos pontuais, demonstrando ser um fator limitante para a comparação entre os grupos.

No trabalho desenvolvido por JORTH et al. 2014 foi observado que vias de fermentação de lisina para butirato, catabolismo de histidina, biossíntese de nucleotídeos e fermentação de piruvato apresentaram expressão gênica melhorada em todos os sítios doentes. A presença dessas vias em diferentes amostras sugere fortemente a importância para a estabilidade das populações associadas à doença e provavelmente contribuem para o processo da doença. Contudo, nosso trabalho apresentou apenas o metabolismo do piruvato em comum com estudo de JORTH et al. 2014. O metabolismo do piruvato apresentou maior abundância no grupo doente ( $p < 0,05$ ) sendo o gene *pflD* aquele que apresentou a maior relação (Figura 40). O gene *pflD* é responsável por codificar a proteína formiato C-transferase. Esse gene costuma ser expresso em ambientes hipóxicos e seu papel é auxiliar na regulação do metabolismo anaeróbico da glicose, catalisando a reação reversível de piruvato e coenzima A em formiato e acetil-coA (WEIDNER, SAWERS, 1996; YESILKAYA et al. 2009). Em organismos periodonto- patogênicos, ainda não houve estudos que deram importância a este gene. Contudo, um estudo realizado por Yesilkaya et al. (2009) demonstrou que o metabolismo de açúcares pode aumentar a virulência em *pneumococos* e que os genes *pfl* são importantes contribuintes.

Wang et al. (2013), utilizando-se de amostras de placa subgengival, encontraram divergências funcionais ao estudar a saúde periodontal e doença. Foi encontrado uma maior presença nas categorias que envolvem transporte de membrana, transdução de sinais e

motilidade celular. Destas, apenas motilidade celular apresentou diferença em nosso trabalho, com maior abundância para o grupo doente. É conhecido que a presença de genes relacionados à motilidade celular pode ser um indício do estabelecimento da colonização de organismos patogênicos (BUTLER, CAMILLE, 2004). Na periodontite em particular, já se sabe que os genes envolvidos na montagem flagelar são considerados pontos de partida para o desenvolvimento da doença (SOLBIATI, FRIAS-LOPEZ, 2018). Os flagelos são considerados estruturas essenciais nos mecanismos de patogenicidade, sendo capazes de proporcionar mobilidade ou aumento da adesão (quando desempenham o papel de adesinas) das bactérias à mucosa oral.

Duran-Pinedo et al. (2014), encontraram uma superexpressão de *FlaA-B*, *FlaG*, *FleN*, *FlgB-I*, *FlgK-L*, *Fla-B*, *FlhF*, *FliD-I*, *FliK-N*, *Flip-S*, *MotA* e *MotB*. Enquanto Wang et al. (2013) encontraram maiores abundâncias com diferença para o grupo doente nos seguintes genes envolvidos na montagem flagelar: *fliD/C*, *flgL/K/E*, *fliE/F* e *flgB/G/H/I*. Nossos achados quanto a “motilidade celular” levam em sua composição genes que envolvem quimiotaxia bacteriana (*cheA*, *cheW*, *cheY*, *fliG*, *fliM*, *fliNY*, *malE*, *mcp*, *mglB*, *motB* e *rbsB*) e montagem flagelar (*flgE*, *flgG*, *flgK*, *flhA*, *fliC*, *fliD*, *fliF*, *fliH*, *fliL* e *fliP*).

Quimiotaxia e mobilidade celular andam em conjunto quando se discute a capacidade de penetração em camadas teciduais em hospedeiros com periodontite, principalmente quando se relaciona a infecção pelo gênero *Treponema*. Já foi observado que a presença do aparato flagelar não é o suficiente para que o processo de penetração tecidual aconteça em sua melhor condição (LUX, 2001).

Nenhum dos genes encontrados para quimiotaxia bacteriana durante a pesquisa apresentou diferença entre os grupos quando analisadas de modo individual. Contudo, observou-se que os genes relacionados a quimiotaxia bacteriana estiveram mais abundantes na pesquisa em comparação com aquelas de motilidade celular. Dentre estes genes relacionados à quimiotaxia bacteriana, destacam-se aqueles que se apresentaram com abundâncias individuais maiores para os doentes: *cheA*, *fliG*, *fliM*, *fliNY*, *malE* e *motB*. Uma expressão regulada positivamente para genes provenientes de quimiotaxia bacteriana em indivíduos com periodontite também foi encontrada no trabalho de Deng et al. (2017) com diferença significativa.

Assim como os genes encontrados para quimiotaxia bacteriana, aqueles relacionados para montagem flagelar não apresentaram diferença entre os grupos quando analisados de modo individual. Ainda assim, mesmo em pequena abundância, o gene *fliC* codificante da proteína flagelina foi aquele que apresentou maior abundância entre os que compõem a

formação do flagelo. Este apresentou abundância maior no grupo doente. Essa abundância maior do gene *fliC* no grupo acometido por periodontite já foi descrita e relacionada diretamente com a patogênese (SZAFRANSKI et al. 2015). Os flagelos possuem um grande potencial de expor a estrutura flagelar ao hospedeiro e induzir severas respostas do sistema imunológico, sendo considerada a principal responsável por essa indução a proteína flagelina (RAMOS, RUMBO, SIRARD, 2004). Essas respostas por sua vez são capazes somar aos eventos imunológicos necessários para o estabelecimento da doença periodontal (HAJISHENGALLIS, 2015).

Além da quimiotaxia e motilidade celular, também são discutidos como presentes durante o acontecimento da periodontite a presença de genes do sistema de secreção do tipo III, sistema CRISPR-Cas tipo III e sistema de dois componentes (DENG et al. 2017). Destas categorias funcionais, apenas se fez presente em nosso trabalho o “sistema de dois componentes”. Entretanto, não houve diferenças nem relações com o grupo doente. Esse sistema composto por vias de transdução de sinais é descrito como responsável por permitir os microrganismos “sentirem”, reagirem e adaptarem-se aos ambientes internos e externos, a fim de maximizar a sobrevivência (SKERKER et al. 2005).

A análise LDA realizada sobre os dados da categoria “função” do KEGG mostrou 11 genes com forte relação com estado de doença periodontal, sendo eles: *rpoC*, *rpoB*, *carB*, *secA*, *uvrA*, *pflD*, *thrS*, *pheT*, *ntpA*, *cobQ* e *aspS*. Os genes *rpoC* e *rpoB* já foram descritos acima como tendo relação com a categoria transcrição, por serem responsáveis pela síntese de subunidades da RNA polimerase procariota.

O gene *secA* faz parte de uma via de translocases denominada Sec. Essa via é responsável pelo transporte de proteínas do citosol para a membrana citoplasmática, membrana externa ou espaço aquoso entre o periplasma em procariotos. Esse processo ocorre antes mesmo da nova proteína atingir sua estrutura terciária estável (CRANEA, RANDALL, 2017). Durante o processo, sabe-se que a proteína codificada pelo gene *secA* permanece associada ao ribossomo recebendo a nova cadeia polipeptídica. Essa proteína, então, transporta a nova proteína não terciária e impulsiona sua translocação através de um canal denominado SecYEG, presente na membrana plasmática – o processo requer a hidrólise de ATP. Nesta situação, a nova proteína pode compor a membrana plasmática ou ser transportada por proteínas da via Sec, ou outras vias de excreção, e se tornar constituinte da membrana citoplasmática, membrana externa ou espaço aquoso entre o periplasma em procariotos (DENKS et al. 2014). Encontramos os genes *secA*, *B*, *D*, *DF*, *E*, *F*, *G* e *Y* em nosso estudo. Contudo, apenas *secA* ( $p = 4,7e-4$ ) e *secY* ( $p = 0,328$ ) foram mais abundantes

em doentes. O filo *Bacteroidetes* é composto por diferentes espécies gram-negativas que externam fatores de virulência utilizando o sistema Sec (DIEGO et al. 2016). Os principais patógenos periodontais, presentes no complexo vermelho, fazem uso do sistema Sec. A espécie *Porphyromonas gingivalis* é capaz de excretar proteases cisteínicas denominadas gingivinas que desempenham o importante papel na evasão ao sistema imunológico do hospedeiro e aquisição de nutrientes e heme. Essas gingivinas apresentam em sua composição N-terminal um trecho sinalizador típico tendo como alvo as proteínas do sistema Sec (FITZPATRICK, WIJEYEWICKREMA, PIKE, 2009; NEMOTO, OHARA-NEMOTO, 2016). Outras proteínas com papel semelhante em *Tanerella forsythia* e *Treponema denticola* fazem uso do sistema Sec de modo análogo (DIEGO et al. 2016).

Não foi descrito na literatura caso algum que demonstrasse a relação entre periodontite e abundância dos genes *uvrA*, *uvrB* e *uvrC*. Esses genes são responsáveis por produzir proteínas que atuam no reparo de excisão de nucleotídeos bacterianos causados por danos de radiação ultravioleta (UV) (CROWLEY et al. 2006). Contudo, o gene *uvrA* apresentou maior abundância em indivíduos doentes em nosso trabalho com  $p = 3,37e-3$ . Embora a correlação da abundância desse gene com a sua expressão em pacientes com periodontite ainda careça de estudos mais aprofundados, esse resultado pode contribuir como um alerta ao tratamento da doença. Foi teorizado por Song et al. 2013 que o tratamento por luz ultravioleta poderia ser eficaz em pacientes com periodontite. Sua teoria é embasada na incidência de luz ultravioleta contra patógenos específicos (*Aggregatibacter actinomycetemcomitans*, *Fusobacterium nucleatum* e *Porphyromonas gingivalis*) em estado planctônico. Contudo, na periodontite, temos a presença de um biofilme que pode atuar como atenuador da luz ultravioleta, além de contar com uma grande variedade de espécies distintas atuando em sinergismo. Nosso trabalho mostra que espécies dentro desse biofilme apresentam genes que promovem resistência aos danos causados por luz ultravioleta. Uma análise metatranscriptômica pode auxiliar na confirmação da expressão desses genes por microrganismos periodontopatogênicos, explicando eventuais casos de insucesso no tratamento.

O gene *cobQ* faz parte de um conjunto de genes que codificam proteínas que atuam na síntese de cobalamina (vitamina B12). A cobalamina existe como metilcobalamina, cianocobalamina e outros análogos que podem ser sintetizados apenas por bactérias, arqueobactérias e espécies limitadas de algas (XU et al. 2018). Em nosso estudo, foi encontrada uma forte relação entre a abundância do gene *cobQ* e a doença periodontal (Figura 40). De acordo com Degnan, Taga, Goodman (2014), a variação nos níveis de cobalamina no intestino humano é capaz de modular a microbiota e produtos expressos pela mesma. Yoste et

al. (2015) encontrou uma relação positiva entre indivíduos com periodontite e a presença de cobalamina. Em seu trabalho, notou-se que o aumento de cobalamina pode intensificar o crescimento de patógenos periodontais. Esse fato pode ocorrer pois alguns patógenos do complexo vermelho (ex. *P. gingivalis*) não são capazes de sintetizar a cobalamina por apresentarem a via de síntese incompleta (ROPER et al. 2000). Portanto, quando outros microrganismos contribuem com a produção de cobalamina para o meio, ocorre um aumento daqueles que não a produzem, podendo ocasionar a progressão da doença (YOSTE et al. 2015).

A análise realizada no banco de dados CAZy mostrou que a família de Glicosídeos Hidrolases (GH) apresenta-se com maior abundância na cavidade oral de indivíduos, tanto saudáveis (40,14%), quanto doentes (42,15%). Enquanto a família Polissacarídeos Liases (PL) foi única que apresentou diferença significativa, com maior abundância para o grupo doente (0,21%) frente aos saudáveis (0,06%). Glicosídeos Hidrolases (GH) e Polissacarídeos Liases (PL) são famílias de enzimas especializadas na degradação de polímeros de carboidratos (CANTAREL et al. 2009; THOMAS et al. 2011). Com respeito aos filos bacterianos, aquele que mais se sobressai na produção de GH e PL é o *Bacteroidetes*, com diversos membros que codificam uma variedade de genes para essas famílias (XU et al. 2003; KUWAHARA et al. 2004). Em nosso estudo, um dos filos com maior abundância no grupo doente foi *Bacteroidetes* (Tabela 6 e Tabela 15) podendo estar, portanto, contribuindo para o resultado positivo dessas famílias de enzima. Além disso, já foi descrito na literatura que a espécie *Tannerella forsythia* é capaz de utilizar enzimas da família GH como fatores de virulência durante a doença periodontal. E enzima categorizada como GH29 (TfFuc1) é uma  $\alpha(1,2)$ -fucosidase que atua no metabolismo de oligossacarídeos curtos, no periplasma do microrganismo, indiretamente proporcionando condições de virulência (MEGSON et al. 2015). Contudo, o estudo de enzimas Glicosídeos Hidrolases e Polissacarídeos Liases permanece escasso quando referente à doença periodontal.

## 7 CONCLUSÃO

- A cavidade oral é predominantemente habitada por organismos pertencentes ao domínio *Bacteria*, com poucos representantes dos domínios *Archaea*, *Eukaryota* e *Virus*.
- Até 99% da microbiota oral é representada pelos filos *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Tenericutes*.
- Os filos *Actinobacteria* e *Proteobacteria* são predominantes em indivíduos saudáveis, enquanto que *Bacteroidetes*, *Fusobacteria*, *Candidatus Saccharibacteria*, *Spirochaetes* e *Synergistes* estão mais presentes em indivíduos com periodontite.
- A Análise de Discriminantes Lineares (LDA) sugere que os gêneros *Eubacterium*, *Leptotrichia*, *Corynebacterium*, *Treponema*, *Parvimonas*, *Fusobacterium*, *Filifactor*, *Capnocytophaga*, *Tannerella*, *Bifidobacterium*, *Peptostreptococcus*, *Dialister*, *Catonella*, *Selenomonas*, *Campylobacter* e *Porphyromonas* formam um grupo potencialmente periodontopatogênico.
- A microbiota periodontal em estado de periodontite apresenta maior riqueza e diversidade, em relação à microbiota no estado saudável.
- A alta abundância das espécies *Porphyromonas gingivalis*, *Treponema denticola* e *Tannerella forsythia* (complexo vermelho) em pacientes portadores de periodontite reitera sua forte relação com o estabelecimento de doença.
- A espécie *Filifactor alocis* pode ser um forte contribuinte para o estabelecimento da periodontite, uma vez que a mesma apresentou forte correlação positiva com os periodontopatógenos *Porphyromonas gingivalis* e *Treponema denticola*.
- A categoria funcional KEGG "Metabolismo" é predominante na cavidade oral, tanto no estado de doença, como em saúde periodontal, contendo > 50% dos resultados funcionais. Nesse contexto, a distribuição de vias metabólicas observadas ressalta a importância de mais estudos sobre o envolvimento de microrganismos acetogênicos e metanogênicos na periodontite.
- A alta abundância de Glicosídeos Hidrolases (GH) e Polissacarídeos Liases (PL) na periodontite apontam para a necessidade de investigações mais aprofundadas sobre as enzimas envolvidas na degradação de carboidratos nesta doença.

## 8 REFERÊNCIAS

AAS, J. A. et al. Defining the normal bacterial flora of the oral cavity. **Journal of Clinical Microbiology**, v. 43, n. 11, p. 5721-5732, 2005.

ABREU, É.F. **Estudo da diversidade microbiana metanogênica em reatores uasb tratando esgoto sanitário**. 2007. 105p. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos), Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte.

ARMITAGE, C.G. Periodontal diagnoses and classification of periodontal diseases. **Periodontology 2000**, v. 34, p. 9-21, 2004.

ARUNI, W.A. et al. *Filifactor alocis* — a new emerging periodontal pathogen. **Microbes Infect.**, v. 17, n. 7, p. 517–530, 2015.

ASNICAR, F. et al. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. **PerrJ**, p. 1-17, 2015.

AVILA, M.; OJCIUS, D. M.; YILMAZ, O. The oral microbiota: living whit a permanent guest. **DNA and Cell Biology**, v. 28, n. 8, p. 405-411, 2009.

BAHEKAR, A.A. et al. The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. **Am Heart J**, v. 154, p. 830-837, 2007.

BIKEL, S. et al. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. **Computational and Structural Biotechnology Journal**, v. 13, p. 390-401, 2015.

BOR, B. et al. *Saccharibacteria* (TM7) in the Human Oral Microbiome. **Journal of Dental Research**, v. 98, n. 5, p. 500-509, 2019.

BUCHFINK, B; XIE, C; HUSON, D. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, n. 1, p. 59-60, 2015.

BUTLER, S.M.; CAMILLI, A. Both chemotaxis and net motility greatly influence the infectivity of *Vibrio cholerae*. **PNAS**, v. 101, n. 14, p. 5018–5023, 2004.

CAFFESSE, R.G.A Latin American perspective of periodontology. **Periodontology 2000**, v. 67, p. 7-12, 2015.

CAMELO-CASTILLO, A.J. et al. Subgingival microbiota in health compared to periodontitis and the influence of smoking. **Front. Microbiol.**, v. 6, p. 1-12, 2015.

CHAN, E.C.S.; MCLAUGHLIN, R. Taxonomy and virulence of oral Montreal *Spirochetes*. **Oral Microbiol Immunol**, v. 15. p. 1-9, 2000.



CHALUB, L.L.F.; PÉRET, A.C.A. Performance of the community periodontal index (CPI) on periodontal status determination: focus on partial recording. **Arqu bras odontol**, v. 6, n. 3, p. 155-162, 2010.

CHAPPLE, I.L.C. et al. Primary prevention of periodontitis: managing gingivitis. **J Clin Periodontol**, v. 42, p. 71-76, 2015.

CHEN, C. et al. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. **The ISME Journal**, v. 12, p. 1210-1224, 2018.

COLWELL, R.K. Estimates: statistical estimation of species richness and shared species from samples. **Version 8.0.b1. User's Guide and application**. Disponível em: <<http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstimateS.php>>. Acesso em: 15 jan. 2019.

COSTALONGA, M.; HERZBERG, M.C. The oral microbiome and the immunobiology of periodontal disease and caries. **Immunol Lett.**, v. 162, n. 200, p. 22–38, 2014.

CRANEA, J.C.; RANDALLA, L.L. The Sec System: Protein Export in *Escherichia coli*. **EcoSal Plus**, v. 7, n. 2, p. 1-73, 2017.

CROWLEY, D.J. et al. The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium sp.* NRC-1. **Saline Systems**, v. 2, n. 11, p. 1-13, 2006.

DARVEAU, R.P. Periodontitis: a polymicrobial disruption of host homeostasis. **Nature Reviews Microbiology**, v. 8, p. 481-490, 2010.

DE PABLO, P.; DIETRICH, T.; McALINDON, T. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. **The Journal of Rheumatology**, v. 35, n. 1, p. 70-76, 2008.

DEGNAN, P.H.; TAGA, M.E.; GOODMAN, A.L. Vitamin B12 as a modulator of gut microbial ecology. **Cell Metab**, v. 20, n. 5, p. 769-778, 2014.

DENG, Z. et al. Dysbiosis in chronic periodontitis: Key microbial players and interactions with the human host. **Scientific Reports**, v. 7, n. 3703, p. 1-13, 2017.

DENKS, K. et al. The Sec translocon mediated protein transport in prokaryotes and eukaryotes. **Mol Membr Biol**, v. 31, n. 3, p. 58–84, 2014.

DEWHIRST, F.E. et al. The human oral microbiome. **Journal of Bacteriology**, v. 192, n. 19, p. 5002-5017, 2010.

DIEGO, I. et al. The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal  $\beta$ -sandwich domain. **Scientific Reports**, v. 6, n. 23123, p. 1-17, 2016.

DURAN-PINEDO, A. et al. Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. **The ISME Journal**, v. 8, p. 1659–1672, 2014.

- FITZPATRICK, R.; WIJEYEWICKREMA, L.C.; PIKE, R.N. The gingipains: scissor and glue of the periodontal pathogen *Porphyromonas gingivalis*. **Future Microbiol.**, v. 4, n. 4, p. 471-487, 2009.
- GIANNOBILE, W.V. et al. Saliva as a diagnostic tool for periodontal disease: current state and future directions. **Periodontology** 2000, v. 50, p. 52-64, 2009.
- GRICE, E.A. et al. A diversity profile of the human skin microbiota. **Genome Research**, v. 18, p. 1043-1050, 2008.
- GRIFFEN, A.L. et al. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. **The ISME Journal**, v. 6, p. 1176-1185, 2012.
- GUPTA, A.; GOVILA, V.; SAINI, A. Proteomics The research frontier periodontics. **Journal of Oral Biology and Craniofacial Research**. v. 5, p. 46-52, 2015.
- HAIJSHENGALLIS, G. Immunomicrobial pathogenesis of periodontitis: keystones, pathobionts, and host response. **Trends in Immunology**, v. 35, n. 1, p. 3-11, 2014.
- HAIJSHENGALLIS, G. Periodontitis: from microbial immune subversion to systemic inflammation. **Nature Reviews Immunology**, v. 15, p. 30-44, 2015.
- HAMADY, M.; KNIGHT, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. **Genome Res.** v. 19, n. 7, p. 1141-1152, 2009.
- HAMMER, O.; HARPER, D.A.T.; RYAN, P.D. Past: Paleontological Statistics Software Package For Education And Data Analysis. **Palaeontologia Electronica**, v. 4, n. 1, 2001.
- HAN, Y.W. *Fusobacterium nucleatum*: a commensal-turned pathogen. **Current Opinion in Microbiology**, v. 23, p. 141-147, 2015.
- HENNE, K et al. Shifts in *Campylobacter* species abundance may reflect general microbial community in periodontitis progression. **Journal of Oral Microbiology**, v. 6, p. 1-6, 2014.
- HIGHFIELD, J. Diagnosis and classification of periodontal disease. **Australian Dental Journal**, v. 54, n. 1, p. 11-26, 2009.
- HORZ, H.; CONRADS, G. Methanogenic Archaea and oral infections – ways to unravel the black box. **Journal of Oral Microbiology**, v. 3, n. 5940, p. 1-9, 2011.
- ILLUMINA. Systems. Disponível em: <<http://www.illumina.com/systems.html>> Acesso em: 15/03/2018.
- JENKINSON, H. F.; LAMONT, R. J. Oral microbial communities in sickness and in health. **Trends in Microbiology**, v. 13, n. 12, p. 589-595, 2005.
- JIANG, S. et al. Salivary Microbiome Diversity in Caries-Free and Caries-Affected Children. **International Journal of Molecular Sciences**, v. 17, p. 1-13, 2016.

JOLIVET-GOUGEON, A. et al. **Antimicrobial treatment of Capnocytophaga infections. Int J Antimicrob Agents**, v. 29, n. 4, p. 367-377, 2007.

JORTH, P. et al. Metatranscriptomics of the Human Oral Microbiome during Health and Disease. **mBio**, v. 5, n. 2, p. 1-10, 2014.

KATO, Y. et al. Molecular detection of human periodontal pathogens in oral swab specimens from dogs in Japan. **J Vet Dent, Summer**, v. 28, n. 2, p. 84-89, 2011.

KEIJSER, B. J. et al. Pyrosequencing analysis of the oral microflora of healthy adults. **J. Dent. Res.**, v. 87, p. 1016–1020, 2008

KINNEY, J.S.; RAMSEIER, C. A.; GIANNOBILE, W. V. Oral fluid-base biomarkers of alveolar bone loss in periodontitis. **Annals of the New York Academy of Sciences**, v. 1098, p. 230-251, 2007.

KOLENBRANDER, P.E. et al. Bacterial interactions and successions during plaque development. **Periodontology 2000**, v. 42, p. 47-79, 2006.

KUBONIWA, M. et al. Insights into the virulence of oral biofilms: discoveries from proteomics. **Expert Rev Proteomics**, v. 9, n. 3, p. 311-323, 2012.

KUWAHARA, T. et al. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. **Proc Natl Acad Sci USA**, v. 101, p. 14919–14924, 2004.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, p. 357-359, 2012.

LAUDENBACH, J.M.; SIMON, Z. Common Dental and Periodontal Diseases Evaluation and Management. **Medical Clinics of North American**, v. 98, v. 1239-1260, 2014.

LINDHE, J.; KARRING, T.; LANG, N.P. Tratado de periodontia clinica e implantologia oral. 4. ed. Rio de Janeiro: Guanabara Koogan, 2005.

LING, Z. et al. Analysis of Oral Microbiota in Children with Dental Caries by PCR-DGGE and Barcoded Pyrosequencing. **Microb Ecol**, v. 30, n. 3, p. 677-690, 2010.

LOZUPONE, C.A. et al. Diversity, stability and resilience of the human gut microbiota. **Nature**, v. 489, p. 220-230, 2012.

LUX, R. et al. Motility and Chemotaxis in Tissue Penetration of Oral Epithelial Cell Layers by *Treponema denticola*. **Infection and Immunity**, v. 69, n. 10, p. 6276-6283, 2001.

MARDIS, E.R. A decades perspective on DNA sequencing technology. **Nature**, v. 470, p. 198-203, 2011.

MASHIMA, I.; NAKAZAWA, F. Interaction between *Streptococcus spp.* and *Veillonella tobetsuensis* in the Early Stages of Oral Biofilm Formation. **Journal of Bacteriology**, v. 197, n. 13, p. 2104-2111, 2015.

MEGSON, Z.A. et al. Characterization of an  $\alpha$ -l-fucosidase from the periodontal pathogen *Tannerella forsythia*. **Virulence**, v. 6, n. 3, p. 282-292, 2015.

MITTAL, V. et al. A placticable approach for periodontal classification. **Dental Research Journal**, v. 10, n. 6, p. 697-703, 2013.

NABORS, T.W.; McGLENNEN, R. C.; THOMPSON, D. Salivary testing for periodontal disease diagnosis and treatment. **Dentistry Today**, v.29, n. 6, p. 53-54, 2010.

**NATIONAL HUMAN GENOME RESEARCH INSTITUTE (NHGRI)**. The Human Genome Project Completion: Frequently Asked Questions. Disponível em: <<https://www.genome.gov/11006943/>> Acesso em: 07/08/2016.

NEMOTO, Y.K.; OHARA-NEMOTO, Y. Exopeptidases and gingipains in *Porphyromonas gingivalis* as prerequisites for its amino acid metabolism. Japanese **Dental Science Review**, v. 52, 22—29, 2016.

O'FLYNN, C. et al. Comparative Genomics of the Genus *Porphyromonas* Identifies Adaptations for Heme Synthesis within the Prevalent Canine Oral Species *Porphyromonas cangingivalis*. **Genome Biol. Evol.** v.7, n. 12, p. 3397–3413, 2015.

OPPERMANN, R. V. et al. Epidemiology of periodontal diseases in adults from Latin American. **Periodontology 2000**, v. 67, p. 13-33, 2015.

PARKS, D. et al. STAMP: statistical analysis of taxonomic and functional profiles. **Bioinformatics**, v. 30, n. 21, p. 3123-3124, 2014.

PASTER, B.J. et al. The breadth of bacterial diversity in human periodontal pocket and other oral sites. **Periodontology 2000**, v. 42, p. 80-87, 2006.

PATHAN, M. et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. **Proteomics**, v. 15, p. 2597–2601, 2015.

PATIL, P.B.; PATIL, B. R. Saliva: A diagnostic biomarker of periodontal diseases. **Journal of Indian Society of Periodontology**, v. 15, n. 4, p. 310-317, 2011.

PENNISI, E. A mouthful of microbes. **Science**, v. 307, n. 5717, p. 1899-1901, 2005.

PÉREZ-CHAPARRO, P.J.; GONÇALVES, C.; FIGUEIREDO, L.C. Newly Identified Pathogens Associated with Periodontitis. **J Dent Res**, V. 93, N. 9, p. 846-858, 2014.

PETER, K.P. et al. Association between periodontal disease and chronic obstructive pulmonary disease: a reality or just a dogma?. **Journal of Periodontology**, v. 84, n. 12, p. 1717-1723, 2013.

PIHLSTRON, B.L.; MICHALOWICZ, B.; JOHNSON, N.W. **Periodontal diseases**. The Lancet, v. 366, p. 1809-1820, 2005.

PONTES, C.S. **Avaliação da prevalência e do perfil de virulência de *Candida spp.* bucais isoladas de pacientes portadores de doença periodontal e diabetes tipo 2**. 2016. 175p.

Dissertação (Mestrado em Saúde, Sociedade e Endemias na Amazônia, Instituto Leônidas e Maria Deane-Fiocruz, Universidade Federal do Amazonas, Manaus.

QUAISER, A. et al. Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. **The ISME Journal**, v. 5, p. 285-304, 2011.

RAM, V.S. et al. Bonebiomarkers in periodontal disease: a review article. **Journal of Clinical and Diagnostic Research**, v. 9, n. 1, p. 7-10, 2015.

RAMOS, H.C.; RUMBO, M.; SIRARD, J. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. **Trends in Microbiology**, v. 12, n. 11, p. 509-517, 2004.

READY, D. et al. Disease severity associated with presence in subgingival plaque of *Porphyromonas gingivalis*, *Aggregatibacter actinomycetemcomitans*, and *Tannerella forsythia*, singly or in combination, as detected by nested multiplex PCR. **J Clin Microbiol**, v. 46, p. 3380–3383, 2008.

RIESENFELD, C.S.; SCHLOSS, P.D.; HANDELSMAN, J. Metagenomics: Genomic Analysis of Microbial Communities. **Annu. Rev. Genet.**, v. 38, p. 525-552, 2004.

ROPER, J.M. The Enigma of Cobalamin (Vitamin B12) Biosynthesis in *Porphyromonas gingivalis*. *The Journal of Biological Chemistry*, v. 275, n. 51, p. 40316–40323, 2000.

SAKAMOTO, M. et al. Application of terminal RFLP analysis to characterize oral bacterial flora in saliva of healthy subjects and patients with periodontitis. **Journal of Medical Microbiology**, v. 52, p. 79-89, 2003.

SANGER, F.; NICKLEN, S.; COULSON, R. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. USA**. v. 74, n. 12, p. 5463-5467, 1997.

SANZ, M.; KORNMAN, K. Periodontitis and adverse pregnancy outcome: consensus report of the joint EFP/AAP Workshop on periodontitis and systemic diseases. **Journal of Periodontology**, v. 84, n. 4, p. 164-169, 2013.

SAYGUN, I. et al. Salivary infectious agents and periodontal disease status. **J Periodont Res**, v. 46, p. 235-239, 2011.

SCHUCHMANN, K.; MÜLLER, V. Energetics and Application of Heterotrophy in Acetogenic Bacteria. **Applied and Environmental Microbiology**, v. 82, n. 14, p. 4056-4069, 2016.

SHENDURE, J.; HANLEE, J. Next-generation DNA sequencing. **Nature biotechnology**, v. 26, n. 10, p. 1135-1144, 2008.

SHIMAZAKI, Y. et al. Effectiveness of the salivary occult blood test as a screening method for periodontal status. **Journal of Periodontology**, v. 82, n. 4, p. 581-587, 2011.

SIQUEIRA, J.F. et al. Cultivable bacteria in infected root canals as identified by 16S rRNA gene sequencing. **Oral Microbiology Immunology**, v. 22, p. 266–271, 2007.

- SKERKER, J.M. et al. Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. **PLoS Biology**, v. 3, n. 10, p. 1771-1788, 2005.
- SOCRANSKY, S.S., HAFFAJEE A.D. Dental biofilms: difficult therapeutic targets. **Periodontology** **2000**, v. 28, p. 12-55, 2002.
- SOCRANSKY, S.S., HAFFAJEE A.D. Periodontal microbial ecology. **Periodontology** **2000**, v. 38, p. 135-187, 2005.
- SOLBIATI, J.; FRIAS-LOPES, J. Metatranscriptome of the Oral Microbiome in Health and Disease. **Journal of Dental Research**, v. 97, n. 5, p. 492-500, 2018.
- SONG, H. et al. Phototoxic effect of blue light on the planktonic and biofilm state of anaerobic periodontal pathogens. **J Periodontal Implant Sci**, v. 43, n. 72-78, 2013.
- SONIA, R. et al. The division 'Synergistes'. **Anaerobe**, v. 13, p. 99–106, 2007.
- SZAFRANSKI, S.P. et al. Functional biomarkers for chronic periodontitis and insights into the roles of *Prevotella nigrescens* and *Fusobacterium nucleatum*; a metatranscriptome analysis. **npj Biofilms and Microbiomes**, v. 15017, p. 1-7, 2015.
- TABA, M. et al. Diagnostic biomarkers for oral and periodontal diseases. **Dental Clinics of North America**, v. 49, n. 3, p. 551-vi, 2005.
- TAKESHITA, T. et al. The ecological proportion of indigenous bacterial populations in saliva is correlated with oral health status. **ISME J**, v. 3, p. 65–78, 2009.
- TAN, K.H. et al. Porphyromonas gingivalis and Treponema denticola Exhibit Metabolic Symbioses. **PLoS Pathog**, v. 10, n. 3, p. 1-11, 2014.
- THOMAS, F. et al. Environmental and gut Bacteroidetes: the food connection. **Cellular and Infection Microbiology**, v. 2, p. 1-17, 2011.
- THOMAS, T.; GILBERT, J.; MEYER, F. Metagenomics a guide from sampling to data analysis. **Microbial Informatics and Experimentation**, v. 2, n. 3, p. 1-12, 2012.
- TRUONG, D.T. et al. Metaphlan 2.0 for enhanced metagenomic taxonomic profiling. **Nature Methods**, v. 12, p. 902-903, 2015.
- VISSER, M.B.; ELLEN, R.P. New insights into the emerging role of oral *Spirochaetes* in periodontal disease. **Clinical Microbiology and Infection**, v. 17, p. 502-512, 2011.
- WANG, J. et al. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. **Scientific Reports**, v. 3, n. 1843, p. 1-10, 2013.
- WEIDNER, G.; SAWERS, G. Molecular Characterization of the Genes Encoding Pyruvate Formate-Lyase and Its Activating Enzyme of *Clostridium pasteurianum*. **Journal of Bacteriology**, v. 178, n. 8, p. 2440–2444, 1996.

WHITMAN, W.B.; COLEMAN, D.C.; WIEBE, W.J. Prokaryotes: The unseen majority. **Proc. Natl. Acad. Sci. USA**, v. 95, p. 6578-6583, 1998.

WHITTAKER, R.H. Vegetation of the Siskiyou Mountains, Oregon and California. **Ecological Monographs**, v. 30, p. 279-338, 1960.

WIEBE, C.B.; PUTNINS, E.E. The periodontal disease classification system of the American Academy of Periodontology an update. **JADC**, v. 66, n. 11, p. 594-597, 2000.

WILKE, A. et al. A RESTful API for accessing microbial community data for MG-RAST. **PLoS Comput Biol**, v. 11, n. 1, 2015.

WOOLEY, J.C.; GODZIK, A.; FRIEDBERG, I. A primer on metagenomics. **Plos Computational Biology**, v. 6, n. 2, p. 1-13, 2010.

XU, J. et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. **Science**, v. 299, p. 2074–2076, 2003.

XU, Y. et al. Cobalamin (Vitamin B12) Induced a Shift in Microbial Composition and Metabolic Activity in an in vitro Colon Simulation. **Frontiers in Microbiology**, v. 9, p. 1-9, 2018.

YESILKAYA, H. et al. Pyruvate Formate Lyase Is Required for *Pneumococcal* Fermentative Metabolism and Virulence. **Infection and Immunity**, v. 77, n. 12, p. 5418–5427, 2009.

YILMAZ, O. The chronicles of *Porphyromonas gingivalis*: the microbium, the human oral epithelium and their interplay. **Microbiology**, v. 154, p. 2897–2903, 2008.

YOST, S. et al. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. **Genome Medicine**, n. 7, n. 27, p. 1-19, 2015.

ZAURA, E. et al. Defining the healthy 'core microbiome' of oral microbial. **BMC Microbiol**, v 9, n 259, p. 1-12, 2009.

ZHANG, J. et al. PEAR: a fast and accurate Illumina Paired-End reAd merger. **Bioinformatics**, n. 30, v. 30, n. 5, p. 614-620, 2014.

ZHU, W.; LOMSADZE, A.; BORODOVSKY, M. Ab initio gene identification in metagenomic sequences. **Nucleic Acids Research**, v. 38, n. 12, p. e132, 2010.

**ANEXOS**



**Anexo A – Parecer do Conselho de Ética e Pesquisa**

UNIVERSIDADE DO ESTADO  
DO AMAZONAS - UEA

**PARECER DO COLEGIADO****DADOS DO PROJETO DE PESQUISA**

**Título da Pesquisa:** Metagenômica da cavidade oral

**Pesquisador:** Daniel Saito

**Área Temática:**

**Versão:** 2

**CAAE:** 19521113.0.0000.5016

**Instituição Proponente:** UNIVERSIDADE DO ESTADO DO AMAZONAS

**Patrocinador Principal:** UNIVERSIDADE DO ESTADO DO AMAZONAS

**DADOS DO PARECER**

**Número do Parecer:** 445.302

**Data da Relatoria:** 01/11/2013

**Apresentação do Projeto:**

A periodontite constitui, ainda hoje, um sério problema de saúde pública. Muito embora algumas espécies de bactérias já tenham sido correlacionadas ao desenvolvimento da periodontite, a grande diversidade de microrganismos encontrada na cavidade oral aponta para a necessidade de estudos microbiológicos mais aprofundados. Nesse sentido, abordagens de caracterização molecular são de grande utilidade, especialmente porque uma parcela significativa das bactérias orais não é passível de cultivo em laboratório. As análises de metagenomas e metatranscriptomas constituem novas vertentes na pesquisa microbiológica, as quais podem trazer informações importantes sobre a estrutura genética, taxonômica e funcional de comunidades microbianas complexas. Similarmente, a análise de 16S rDNA tem contribuído enormemente para o estudo do microbioma oral, permitindo a detecção de bactérias nunca antes descritas. Estas análises, associadas à técnica de seqüenciamento paralelo de DNA, possuem alta aplicabilidade no estudo da microbiota em diversos ambientes e, particularmente, nas comunidades bacterianas da cavidade oral, em saúde e em doença. A presente proposta visa a análise metagenômica, metatranscriptômica e de diversidade bacteriana da microbiota oral, com ênfase na bioprospeção de genes e espécies bacterianas com alto potencial de virulência em periodontite. Espera-se que a identificação e a

**Endereço:** Av. Djalma.Batista, nº 3578, Chapada

**Bairro:** chapada

**CEP:** 69.050-030

**UF:** AM

**Município:** MANAUS

**Telefone:** (92)3878-4368

**Fax:** (92)3878-4368

**E-mail:** cep.uea@gmail.com



UNIVERSIDADE DO ESTADO  
DO AMAZONAS - UEA



Continuação do Parecer: 445.302

caracterização

metabólica de novos genes de virulência e espécies possa contribuir para uma melhor compreensão dos mecanismos etiológicos da periodontite, para o

Apresentação do Projeto: Patrocinador Principal: UNIVERSIDADE DO ESTADO DO AMAZONAS

**Objetivo da Pesquisa:**

Objetivo Primário:

O projeto tem como objetivo primário a análise metagenômica e metatranscriptômica da microbiota oral, com auxílio da técnica de seqüenciamento

paralelo de DNA (piroseqüenciamento). O estudo buscará informações genéticas e funcionais em alto rendimento (high throughput) sobre as comunidades microbianas da cavidade oral e, em uma segunda instância, a identificação (bioprospecção) de genes com expressão diferenciada em periodontite, com ênfase nas principais vias metabólicas ativas em doença, proporcionando conhecimento mais abrangente dos fatores de virulência dos microrganismos causadores da periodontite.

Objetivo Secundário:

O estudo terá como objetivo secundário a detecção das bactérias com diferentes prevalências nos grupos de estudo, buscando-se identificar as espécies potencialmente patogênicas.

**Avaliação dos Riscos e Benefícios:**

Riscos:

O estudo apresenta não apresenta risco previsível, visto que a coleta de amostras será realizada com curetas gengivais, cotonetes e alças de plástico descartáveis. Este procedimento não implica em riscos e desconfortos, uma vez que os instrumentos de coleta se encontrarão devidamente esterilizados e serão utilizados com máxima cautela, sem o uso de força ou pressão excessiva.

Benefícios:

Não existe previsão de benefício direto aos sujeitos da pesquisa. Porém, aqueles que requisitarem seus resultados, poderão ter conhecimento das espécies de bactérias que habitam a sua cavidade bucal. Os sujeitos da pesquisa que se enquadram no grupo dos portadores de periodontite terão seu tratamento periodontal normalmente conduzido, sem detrimento ou atrasos devido à participação no presente estudo.

**Comentários e Considerações sobre a Pesquisa:**

Pesquisa de relevância na área de odontologia, uma vez que prevê o conhecimento da microbiota que causa a periodontite, o que ajudará em futuros tratamentos.

Endereço: Av. Djalma Batista, nº 3578, Chapada

Bairro: chapada

CEP: 69.050-030

UF: AM

Município: MANAUS

Telefone: (92)3878-4368

Fax: (92)3878-4368

E-mail: cep.uea@gmail.com



UNIVERSIDADE DO ESTADO  
DO AMAZONAS - UEA



Continuação do Parecer: 445.302

**Considerações sobre os Termos de apresentação obrigatória:**

Todos os termos obrigatórios para a pesquisa foram apresentados.

**Recomendações:**

Recomenda-se que o resultado da pesquisa seja apresentada aos alunos de odontologia para conhecimento e aplicação prática profissional.

**Conclusões ou Pendências e Lista de Inadequações:**

Sou de parecer favorável a execução do presente projeto, porém para que este seja aprovado por este CEP necessita dos seguintes ajustes:

- Inserir o cálculo amostral justificando o número da amostra que será atendida no projeto;
- Rever o cronograma de execução, ajustando o início do projeto ao período de aprovação do mesmo por este CEP;
- Incluir o endereço residencial dos pesquisadores no TCLE;
- Inserir o termo de anuência. O termo apresentando não é de anuência e sim uma ciência do local em que serão coletadas as amostras;

**Situação do Parecer:**

Aprovado

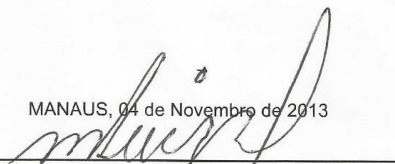
**Necessita Apreciação da CONEP:**

Não

**Considerações Finais a critério do CEP:**

Parecer aprovado.

MANAUS, 04 de Novembro de 2013

  
 Assinado por:  
 Prof. M. G. Manoel Luiz Neto  
 (Coordenador)  
 Comitê de Ética em Pesquisa  
 ESA/UEA

Endereço: Av. Djalma Batista, nº 3578, Chapada

Bairro: chapada

CEP: 69.050-030

UF: AM

Município: MANAUS

Telefone: (92)3878-4368

Fax: (92)3878-4368

E-mail: cep.uea@gmail.com