



## IDENTIFICAÇÃO DA MICROBIOTA SALIVAR MAIS REPRESENTATIVA DE SEQUENCIAMENTOS DE *SHOTGUN* E *AMPLICON*

Rafaela Nishiyama<sup>\*1</sup>, Claudio Trigo<sup>1</sup>, Simone Oliveira<sup>1,2</sup>, Flavio Aguiar<sup>1</sup>, Rodrigo Jardim<sup>3</sup>

**1** Faculdade de Odontologia de Piracicaba, FOP/UNICAMP, Piracicaba, Brasil

**2** Faculdade de Odontologia da Universidade do Estado do Rio de Janeiro, FO/UERJ, Rio de Janeiro, Brasil

**3** Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, IOC/Fiocruz, Rio de Janeiro, Brasil

### RESUMO

A microbiota oral é considerada a segunda mais complexa do corpo humano e apresenta mais de 700 espécies de microrganismos. Alguns são utilizados como marcadores de doenças bucais ou sistêmicas. O objetivo deste estudo foi identificar a microbiota oral mais representativa a partir de dados metagenômicos da saliva humana. Foram identificados na plataforma MG-RAST dados metagenômicos de saliva humana, obtidos através de sequenciamento por *Amplicon* (16S) e DNA total (*Shotgun*). Através do *software* R, foi realizada a análise descritiva de 5 variáveis (*Failed*, *Unknown*, *Predicted*, *Alpha* e *Rarefaction*), além do estudo de correlação dessas variáveis por tipo de sequenciamento (*Amplicon* e *Shotgun*) e da análise de componentes principais da microbiota. Foram recuperadas 476 amostras (332 de *Amplicon* e 142 de *Shotgun*). Apenas a variável *Predicted* não apresentou diferença significativa entre os tipos de sequenciamento ( $p < 0.05$ ). Neste estudo foi encontrado que a microbiota representativa da saliva humana apresenta gêneros normalmente observados em doenças bucais, mas não se limita a eles. Gêneros recentes, patogênicos e ainda pouco conhecidos foram encontrados em amostras de *Shotgun*. A diversidade da microbiota em amostras salivares variou segundo a técnica de sequenciamento metagenômico, sendo que a diversidade do núcleo da microbiota obtida por *Shotgun* foi maior do que a observada em *Amplicon*.

**Palavras-chave:** Microbioma oral, Saliva, Metagenômica, Amplicon, Shotgun



<https://www.youtube.com/watch?v=YV3gj0628cA&feature=youtu.be>

### INTRODUÇÃO

A microbiota oral é considerada a segunda mais complexa do corpo humano e nela podem ser encontrados mais de 700 espécies de microrganismos. No entanto, cerca de 68% desses microrganismos são cultiváveis<sup>1</sup>. Interações estabelecidas entre comunidades de microrganismos e o hospedeiro permitem estabelecer os perfis microbiológicos de condições fisiológicas quanto patológicas. Nas últimas décadas o desenvolvimento de metodologias e análises para identificação e caracterização de microbiomas, possibilitou prever doenças associadas a mudanças no ambiente e seus reflexos na microbiota, particularmente as que compartilham o mesmo nicho (*core*)<sup>2</sup>. Os 10% dos organismos mais abundantes do *core* são considerados dominantes enquanto, que os 65% menos abundantes, raros. Alterações no ambiente oral podem acarretar disbiose deste microbioma e consequente o surgimento ou agravamento de doenças bucais como cárie, placa bacteriana e doenças periodontais<sup>3-5</sup>.

Desde 2008 bases de dados especializadas em dados biológicos têm permitido o depósito de dados brutos e dados analisados de projetos de metagenômica<sup>6,7</sup>. Um dos pioneiros em plataformas de armazenamento e análise de dados metagenômicos é o portal de acesso público MG-RAST<sup>7</sup>, que tem depositado projetos gerados a partir de diferentes tecnologias metagenômicas: *Amplicon*, *Shotgun* e metatranscriptômica. Os projetos de sequenciamento por *Amplicon* utilizam produto da reação em cadeia de polimerase (PCR) pela amplificação de genes marcadores (*primers*), como os genes 16S rRNA, 23S rRNA e 18S rRNA. Enquanto, que sequenciamentos por *Shotgun* (ou *Whole Genome Sequencing*, *WGS*) utilizam todo o DNA extraído de uma amostra e os projetos de metatranscriptômica sequenciam todo o RNA extraído de uma amostra.

Este estudo tem por objetivo identificar a microbiota salivar mais representativa de sequenciamentos por *Shotgun* e *Amplicon*, pelo estudo de metagenomas disponíveis no MG-RAST, publicados no período de 2014 a 2019.

### METODOLOGIA

A coleta de dados foi realizada no portal Metagenomics - Rapid Annotation using Subsystems Technology (MG-RAST) (<https://www.mg-rast.org>). Foram recuperados projetos contendo o termo

SALIVA na variável Material (material = 'SALIVA'), entre os anos de 2014 e 2019. Um *script in house* (programa) em Python (<http://www.python.org>) foi desenvolvido em colaboração com o Laboratório de Biologia Computacional e Sistemas (LBCS - IOC / FIOCRUZ), que permitiu a busca automática dos projetos e a tabulação de metadados selecionados para o estudo (Tabela 1). Apenas projetos de metagenômica foram considerados apenas projetos com ambas as tecnologias (*Amplicon* e *Shotgun*), com um número de metagenomas acima de 3. O software R (versão 3.6.1) <sup>8</sup> foi utilizado para as análises estatísticas: análise descritiva, normalidade (Shapiro-Wilk), homocedasticidade, teste de Mann-Whitney-Wilcoxon, correlação e análise de componentes principais (PCA). As análises descritivas e de correlação foram realizadas para as variáveis numéricas. Ambas as análises foram agrupadas por tipo de tecnologia. O estudo da taxonomia foi realizado por PCA para avaliar os organismos mais representativos nas amostras. A fim de padronizar o estudo taxonômico, este foi baseado no nível taxonômico do gênero.

Tabela 1: Variáveis capturadas no MG-RAST.

VARIÁVEL	DESCRIÇÃO	NOME NO MG-RA
ID	Identificação do metagenome	MG-RAST-ID
Project	Identificação do projeto	Project ID
Material	Material coletado	Material
Failed	<i>Reads</i> que falharam em qualidade	QC_failed
Unknown	<i>Reads</i> sem taxonomia	QC_unknown
Predicted	<i>Reads</i> com taxonomia	QC_predicted
Alpha	Diversidade Alpha do índice de Shannon	ALPHA
Rarefaction	Ponto máximo da curva de rarefação	RAREFACTION
Type	Tecnologia de sequenciamento	Sequence_type
Taxonomy	Taxonomia no nível de gênero	Taxonomy

Observação: Amarelo: variáveis numéricas (análise descritiva e de correlação); vermelho: tipo de tecnologia de sequenciamento; azul: nível taxonômico (análise de componentes principais - PCA).

## RESULTADOS

O *script* em Python recuperou 476 metagenomas, (332 de *Amplicon*, 142 de *Shotgun* e 2 de metatranscriptoma, que foram excluídos). Os 474 metagenomas restantes foram distribuídos em 12 projetos. Destes, apenas 4 projetos apresentavam metagenomas com ambas as tecnologias de sequenciamento (*Amplicon* e *Shotgun*). Um projeto que tinha apenas 1 metagenoma de cada tecnologia também foi descartado. Assim, o estudo foi realizado com 3 projetos, com 245 metagenomas obtidos por *Amplicon* e 164 obtidos por *Shotgun*. Todos os táxons para o nível de gênero foram recuperados em todos os metagenomas.

As representações gráficas das variáveis estudadas: *Failed*, *Predicted*, *Unknown*, *Alfa* e *Rarefaction* está apresentada na Figura 1. Em *Amplicon*, nenhuma sequência apresentou problemas de qualidade. No estudo de normalidade, apenas *Alpha*, em *Shotgun*, apresentou distribuição normal. Nas demais, a distribuição foi não normal. Todas as comparações das variáveis, segundo tipo de sequenciamento apresentaram diferenças significantes ( $p < 0,05$ ), exceto a variável *Predicted*.

O estudo de correlação apresentou divergências entre os tipos de sequenciamento (Figura 2). A técnica de *Shotgun* apresentou alta correlação negativa entre *Rarefaction* e *Failed* ( $r = -0,78$ ) e baixa associação, também negativa, entre *alpha* e *unknown* ( $r = -0,12$ ). Por outro lado, na técnica de *Amplicon*, as variáveis *rarefaction* e *unknown* ( $r = 0,63$ ) apresentaram a maior correlação. Diferentemente, *alpha* e *predicted* ( $r = -0,03$ ) apresentaram a menor associação.

No PCA para *Amplicon*, foi encontrado um core com menor diversidade de microrganismos, com maior representação de *Propionibacterium* e *Lactobacillus*. Em *Shotgun*, a diversidade do core de microrganismos foi maior, com maior representação de *Escherichia* e *Chitinophaga*. *Streptococcus* foi considerado altamente representativo em ambas as técnicas de sequenciamento (Figura 3).

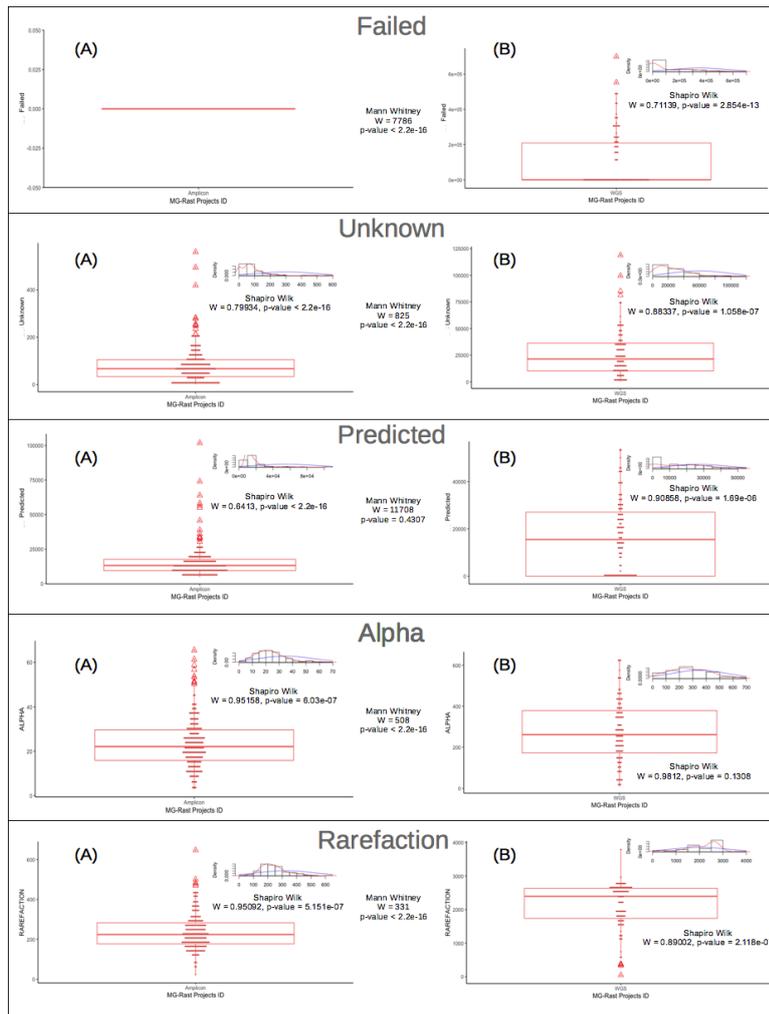


Figure 1: Representação gráfica da análise descritiva das variáveis *Failed*, *Predict*, *Unknown*, *Alfa* e *Rarefaction*, segundo a técnica de sequenciamento. (A): *Amplicon*; (B): *Shotgun*.

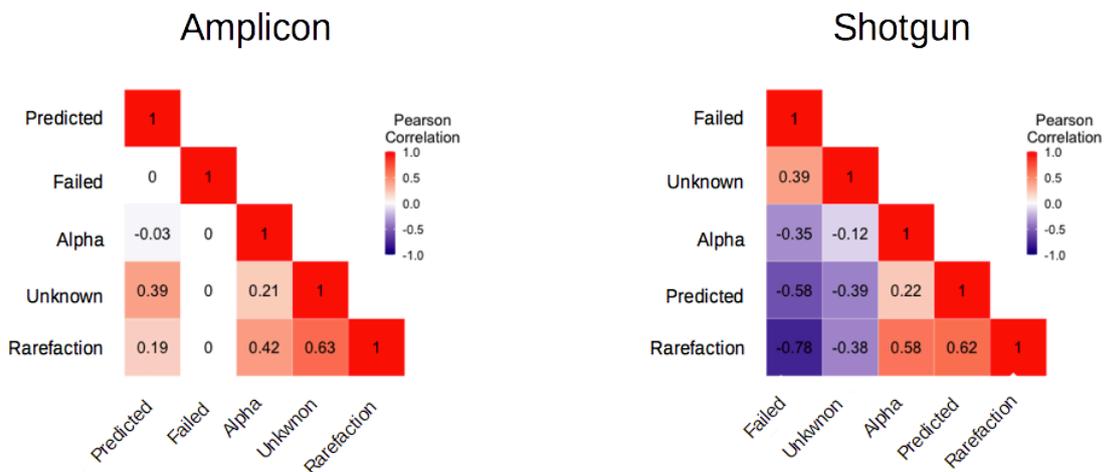


Figura 2: Estudo de Correlação por tipo de sequenciamento das variáveis: *Failed*, *Unknown*, *Predicted*, *Alpha* and *Rarefaction*.

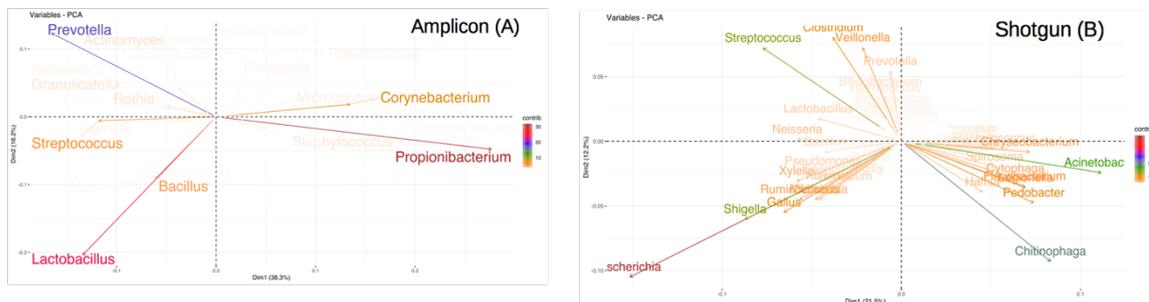


Figura 3. Cores obtidos por Análise de Componentes Principais (PCA) de *Amplicon* (a) e *Shotgun* (b).

## DISCUSSÃO

Nessa pesquisa foram selecionados metagenomas da saliva humana como material de análise de microbiomas. A obtenção de amostras salivares é simples, de fácil realização, não invasiva, não causa desconforto ao paciente, de baixo custo, e representativa do ambiente oral. A saliva, assim como outros fluidos biológicos contém DNA, RNA, proteínas e produtos metabólicos, além da microbiota. Apesar da facilidade de sua coleta, o conhecimento do comportamento da microbiota salivar e demais constituintes ainda é um desafio. Estudos que utilizam análises sofisticadas como o sequenciamento de última geração, permitem a investigação de diferenças no perfil bacteriano em pacientes com doenças bucais, como cárie e periodontite, comparados a indivíduos saudáveis <sup>9</sup>.

A saliva de indivíduos com alta experiência de cárie é associada à elevada abundância salivar de *Streptococcus* e inúmeras espécies de *Lactobacillus* além de outras bactérias capazes de degradar açúcares e formar polissacarídeos extracelulares <sup>10</sup>. Amostras de indivíduos saudáveis com baixa experiência de cárie foram associadas a maior abundância dos gêneros *Neisseria*, *Haemophilus* e *Fusobacterium*, das quais a maioria das espécies desse gênero fermenta apenas levemente o açúcar <sup>11</sup>. De acordo com Tanner e colaboradores (2018)<sup>10</sup> a composição da saliva na cavidade oral é um dos principais os fatores de risco associados à cárie. A disbiose do biofilme resulta no aumento de espécies acidogênicas e acidúricas, capazes de modular a composição do *core* no biofilme. Enquanto que em casos de gengivite, o aumento da quantidade de placa ao redor da margem gengival induz a resposta inflamatória no hospedeiro, levando ao aumento dos níveis de bactérias anaeróbias, incluindo espécies proteolíticas Gram-negativas, especialmente as pertencentes aos gêneros *Prevotella*, *Porphyromonas*, *Tannerella*, *Fusobacterium* e *Treponema* <sup>12</sup>.

As variáveis estudadas apresentaram comportamentos distintos segundo o tipo de técnica de sequenciamento. Contudo, o resultado mais expressivo o da variável *Predicted*, a única que não apresentou diferença estatística significativa entre as técnicas de sequenciamento, demonstrando que, em termos relativos, as duas técnicas são capazes de inferir de forma semelhante a taxonomia no nível de gênero. Estes resultados se assemelham ao observado por Laudadio e colaboradores (2018)<sup>13</sup>, que investigaram a composição microbiana do intestino humano. No entanto, os resultados obtidos por *Shotgun*, permitiram uma caracterização mais complexa do microbioma, com a identificação de maior diversidade e em nível taxonômico de espécie, quando comparado ao sequenciamento do 16S rDNA por *Amplicon*.

No estudo de correlação a técnica *Shotgun* se mostrou com dados mais confiáveis, uma vez que o comportamento esperado das variáveis foi observado apenas nas amostras de *Shotgun*. Conforme o observado entre *Predicted* e *Rarefaction*, que em *Shotgun* apresentou correlação positiva ( $r = 0,62$ ). Por sua vez, em *Amplicon* a associação foi de apenas 0,19. Cabe salientar que conforme o número sequencias conhecidas, maior será o número de *taxa* identificado. De forma semelhante, *Failed* e *Rarefaction* apresentaram alta correlação negativa em *Shotgun* ( $r = - 0,78$ ), como esperado. Em *Amplicon*, não foi observado associação entre estas variáveis. Isto pode ser explicado em razão de todos os metagenomas, de todos os projetos, terem zero como valor para *Failed*. Este estudo permitiu ainda comprovar que a técnica de *Shotgun* é capaz de identificar um número maior de microorganismos presentes na microbiota da saliva

humana, do que a técnica de *Amplicon*, evidenciando dessa forma, que a diversidade de taxa é técnica-dependente, semelhante ao observado por Laudadio e colaboradores (2018)<sup>13</sup>.

Em ambos os *cores* foram encontradas algumas das bactérias já descritas como pertencentes à microbiota bucal como, por exemplo, *Streptococcus*, *Lactobacillus* e *Prevotella*. Por outro lado, a presença de *Corynebacterium*, *Escherichia*, *Pseudomonas* e *Shigella*, por exemplo, evidenciam microorganismos patogênicos nas amostras de saliva. Assim como, também ressaltam a presença de novo taxa, *Chitinophaga*, observado nas amostras sequenciadas de *Shotgun*. Este gênero se apresentou com elevada representatividade nas amostras salivares, devendo ser objeto de investigação em novos estudos, que possam demonstrar sua relação e função junto aos demais taxa na microbiota bucal. Uma vez que estudos recentes relatam seu potencial patogênico<sup>14,15</sup>.

## CONCLUSÃO

A partir desse estudo, pode-se concluir que no *core* do microbioma salivar humano, podem ser encontrados gêneros de bactérias observados em doenças bucais, mas não se limitam a eles. Os resultados indicam que o estudo do *core* do microbioma salivar oral é técnica-dependente e que a técnica de sequenciamento influencia a diversidade da microbiota. Os resultados obtidos sugerem que *Shotgun* e *Amplicon* possuem sensibilidades similares de detecção no nível taxonômico de gênero, apesar da grande diferença no número de sequências geradas entre as técnicas, e que as associações observadas nas variáveis em *Shotgun* apresentam-se de forma mais previsível e esperada do que em *Amplicon*.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Caselli E, Fabbri C, D'Accolti M, Soffritti I, Bassi C, Mazzacane S, et al. Defining the oral microbiome by whole-genome sequencing and resistome analysis: The complexity of the healthy picture. *BMC Microbiol.* 2020;20(1):1–19.
2. Shade A, Handelsman J. Beyond the Venn diagram: The hunt for a core microbiome. *Environ Microbiol.* 2012;14(1):4–12.
3. Nibali L, Sousa V, Davrandi M, Spratt D, Alyahya Q, Dopico J, et al. Differences in the periodontal microbiome of successfully treated and persistent aggressive periodontitis. *J Clin Periodontol.* 2020;47(8):980–90.
4. Papapanou PN, Park H, Cheng B, Kokaras A, Paster B, Burkett S, et al. Subgingival microbiome and clinical periodontal status in an elderly cohort: The WHICAP ancillary study of oral health. *J Periodontol.* 2020;(March):1–12.
5. Nomura Y, Otsuka R, Hasegawa R, Hanada N. Oral microbiome of children living in an isolated area in myanmar. *Int J Environ Res Public Health.* 2020;17(11):1–12.
6. Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res.* 2011;39(SUPPL. 1):2010–2.
7. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:1–8.
8. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
9. Suárez Moya A. Microbioma y secuenciación masiva. *Spanish J Chemother.* 2017;30(5):305–11.
10. Tanner ACR, Kressirer CA, Rothmiller S, Johansson I, Chalmers NI. The Caries Microbiome: Implications for Reversing Dysbiosis. *Rev Adv Dent Res.* 2018;29(1):78–85.
11. Belstrøm D, Holmstrup P, Fiehn NE, Kirkby N, Kokaras A, Paster BJ, et al. Salivary microbiota in individuals with different levels of caries experience. *J Oral Microbiol.* 2017;9(1):1–8.
12. Willmann C, Mata X, Hanghoej K, Tonasso L, Tisseyre L, Jeziorski C, et al. Oral health status in historic population: Macroscopic and metagenomic evidence. *PLoS One.* 2018;13(5):1–18.
13. Laudadio I, Fulci V, Palone F, Stronati L, Cucchiara S, Carissimi C. Quantitative Assessment of Shotgun Metagenomics and 16S rDNA Amplicon Sequencing in the Study of Human Gut Microbiome. *Omi A J Integr Biol.* 2018;22(4):248–54.
14. Tran TLQ, Anani H, Trinh HT, Pham TPT, Dang VK, Ho VM, et al. *Chitinophaga vietnamensis* sp. Nov., a multi-drug resistant bacterium infecting humans. *Int J Syst Evol Microbiol.* 2020;70(3):1758–68.
15. Crémet L, Bemer P, Zambon O, Reynaud A, Caroff N, Corvec S. *Chitinophaga terrae* bacteremia in human. *Emerg Infect Dis.* 2009;15(7):1134–5.