

## 4.02.99 - Odontologia

## DADOS PÚBLICOS PODEM REVELAR O NÚCLEO DO MICROBIOMA DA SALIVA HUMANA

Rafaela R. Nishiyama<sup>1\*</sup>, Simone G. Oliveira<sup>1,2</sup>, Claudio A.C. Trigo<sup>1</sup>, Rodrigo Jardim<sup>3</sup> e “Flavio H.B. Aguiar”<sup>1</sup>

1. Faculdade de Odontologia de Piracicaba, UNICAMP, Piracicaba, Brasil

2. Faculdade de Odontologia, UERJ, Rio de Janeiro, Brasil

3. Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brasil

**Resumo**

Sequenciamentos de Nova Geração são utilizados no estudo da diversidade da microbiota da saliva, fluido representativo do ambiente intra-oral. O objetivo deste trabalho foi identificar o núcleo (*core*) desse microbioma a partir de dados públicos depositados no MG-RAST. Foram avaliadas variáveis que inferem na qualidade, quantidade, diversidade e taxonomia das amostras, além da abordagem de sequenciamento (AS): *Amplicon* e *Shotgun*. O software R foi utilizado para as análises estatísticas ( $p < 0,05$ ) *Shapiro-Wilk*, homocedasticidade, *Mann-Whitney-Wilcoxon*, *Pearson* e Análise de Componentes Principais. Foram analisados 409 metagenomas. *Predicted* foi a única variável sem diferença entre as abordagens. A associação entre as variáveis foi divergente segundo a AS. *Shotgun* apresentou o *core* com maior diversidade, com os gêneros *Escherichia* e *Chitinophaga* como os mais representativos. A identificação do *core* é dependente da AS, que é capaz de influenciar a diversidade da microbiota estudada.

**Palavras-chave:** Microbioma Oral; Fluido Salivar; MG-RAST

**Apoio financeiro:** Programa Institucional de Bolsas de Iniciação Científica (PIBIC).

**Trabalho selecionado para a JNIC:** UNICAMP

**Introdução**

Perfis fisiológicos e patológicos do microbioma oral humano têm sido estabelecidos<sup>1-4</sup> e demonstram que a microbiota oral está intimamente associada a doenças bucais como cárie, placa bacteriana e doenças periodontais<sup>5,6</sup>, o que torna importante a identificação dos microrganismos mais representativos do microbioma oral para a predição de doenças, mudanças no ambiente oral e o comportamento de microrganismos pertencentes ou não ao núcleo (*core*)<sup>7</sup> desses microbiomas.

Apesar da microbiota da cavidade oral poder abrigar cerca de 700 *táxon* bacterianos<sup>4</sup>, um terço deles são descritos como não cultiváveis<sup>2,3</sup>. Buscando superar essa barreira os Sequenciamentos de Nova Geração (NGS) têm sido utilizados no estudo da diversidade de microrganismos<sup>8-10</sup> em amostras de diferentes ambientes, como a saliva, um fluido de fácil coleta, não invasivo e representativo do ambiente intra-oral<sup>11</sup>.

Atualmente existem duas abordagens no uso de NGS: *Amplicon* e *Shotgun*. O sequenciamento por *Amplicon* utiliza marcadores genéticos, que identificam de forma específica as sequências de microrganismos presentes nas amostras. Por sua vez, *Shotgun* é capaz de identificar os microrganismos de uma amostra, analisando diretamente o DNA total da amostra sem precisar do uso de marcadores específicos. Uma característica importante dessas abordagens é a elevada produção de dados gerados a serem tratados e analisados<sup>8,10,12</sup>.

Dados gerados por sequenciamento e publicados normalmente são depositados em bases de dados de domínio público<sup>5,12</sup>. Um dos pioneiros em plataformas de armazenamento e análise de dados metagenômicos é o portal Metagenomics - Rapid Annotation using Subsystems Technology (MG-RAST), lançado em 2008 e que disponibiliza gratuitamente dados metagenômicos e informações de características funcionais de sequências metagenômicas<sup>13</sup>. Pelo MG-RAST podem ser capturados dados de forma agrupada, por projetos ou tipo de amostras, dados da diversidade de microrganismos, de qualidade das sequências e sequências já descritas em outros bancos de dados, permitindo assim reconhecer nas amostras os microrganismos presentes<sup>14</sup>.

O objetivo deste trabalho foi identificar o *core* (núcleo) de microrganismos da microbiota salivar a partir de dados metagenômicos obtidos por *Amplicon* e *Shotgun* no MG-RAST.

**Metodologia**

A partir do MG-RAST foram recuperados dados de projetos de metagenômica que utilizaram ambas as abordagens: *Amplicon* e *Shotgun*. Foram analisados projetos depositados entre os anos de 2014 e 2019 e cujo material sequenciado fora **saliva**. Através de um programa escrito pelo grupo (*in-house*), foi possível recuperar esses dados de forma automática e disponibilizá-los em um arquivo do tipo *Coma Separated Values* (CSV).

Afim de padronizar os dados recuperados, só foram analisados os projetos de metagenomas que satisfizeram os seguintes critérios: (i) a abordagem de sequenciamento precisava ser *Amplicon* ou *Shotgun*; (ii) os projetos precisavam ter as duas abordagens e (iii) os projetos precisavam ter pelo menos 3 metagenomas por cada

abordagem.

Os dados foram analisados no software R (versão 3.6.1) e todos os testes estatísticos foram analisados com 95% de confiança ( $p\text{-value} < 0,05$ ).

Para este estudo foram utilizadas variáveis que descrevem a qualidade e a quantidade de sequências (*Failed*, *Unknown* e *Predicted*), variáveis que inferem a diversidade de cada metagenoma (*Alpha* e *Rarefaction*) e a taxonomia inferida, além da variável que descreve a abordagem do estudo (*Amplicon* e *Shotgun*).

Inicialmente os dados foram analisados quanto a sua distribuição, utilizando os testes *Shapiro-Wilk* e de homocedasticidade. As diferenças entre as abordagens de sequenciamento foram analisadas através do teste *Mann-Whitney-Wilcoxon*. Os dados foram ainda analisados segundo suas correlações, utilizando o índice de *Pearson*. Por fim, foi utilizada a Análise de Componentes Principais (PCA) para a inferência do *core* (núcleo) do microbioma para cada uma das abordagens, por ser uma técnica estatística de análise multivariada que permite a redução de um grande número de variáveis para um número menor ou igual de variáveis denominadas **componentes principais** (CP), mantendo a variância dos dados, de forma que o primeiro CP possua a maior variância, o segundo CP, a segunda maior variância e assim até o último CP.

## Resultados e Discussão

Foram obtidos pelo *script in house* 13 projetos que continham 476 metagenomas sendo, 332 de *Amplicon*, 142 de *Shotgun* e 2 de Metatranscriptoma. Os metagenomas de metatranscriptoma foram excluídos. Os 474 metagenomas restantes estavam distribuídos em 12 projetos. Destes, apenas 4 projetos apresentavam metagenomas com ambas as abordagens de sequenciamento (*Amplicon* e *Shotgun*). Um projeto que tinha apenas 1 metagenoma de cada abordagem 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 selecionados (Tabela 1).

Tabela 1: Número de metagenomas segundo a abordagem de sequenciamento

SEQUENCIAMENTO	PROJETO	NÚMERO DE METAGENOMAS
Amplicon	mgp3474	95
	mgp4843	97
	mgp7236	53
Shotgun	mgp3474	8
	mgp4843	73
	mgp7236	83

Na análise de normalidade, a maioria das variáveis apresentou distribuição não normal (Tabela 2), implicando na realização do teste não paramétrico de *Mann-Whitney-Wilcoxon*, para a avaliação da diferença estatística das variáveis entre os tipos de sequenciamento. Essas comparações apresentaram-se todas com diferenças significativas ( $p < 0,05$ ), com exceção de *Predicted* ( $p = 0,4307$ ). Em *Amplicon* não houve registro de sequências com problemas de qualidade (Tabela 2).

O estudo de correlação apresentou divergências entre as abordagens de sequenciamento. A associação mais elevada em *Amplicon* foi observada, entre as variáveis *Rarefaction* e *Unknown* ( $r = 0,63$ ) enquanto, que *Alpha* e *Predicted* foram as variáveis com menor associação ( $r = -0,03$ ) em *Amplicon*. Em *Shotgun*, *Rarefaction* e *Failed* apresentaram elevada correlação negativa ( $r = -0,78$ ) e *Alfa* e *Unknown* ( $r = -0,12$ ), a menor associação, também negativa.

O *core* com menor diversidade foi observado em *Amplicon*, tendo os gêneros mais representativos *Propionibacterium* e *Lactobacillus*. Em *Shotgun* os gêneros mais representativos foram *Escherichia* e *Chitinophaga*. *Streptococcus* foi altamente representativo em ambas as abordagens de sequenciamento (Figura 2).

Este estudo optou pela seleção dos estudos que escolheram a saliva como o material de análises dos microbiomas a fim de obter maior padronização de resultados. Segundo Tanner e colaboradores (2018)<sup>1</sup> a saliva está associada a um dos principais fatores de risco à cárie, pela predominância de bactérias acidogênicas e acidúricas, que modulam a composição do *core* do biofilme. A saliva é um fluido biológico facilmente coletado para análises microbianas sofisticadas, que utilizam sequenciamento de última geração, permitindo a realização de estudos capazes de identificar diferenças no perfil bacteriano em pacientes com doenças bucais, como cárie e periodontite, quando comparados a indivíduos saudáveis via oral<sup>11,15-17</sup>.

As variáveis estudadas apresentaram comportamentos distintos segundo a abordagem de sequenciamento. Contudo, o resultado mais expressivo foi o de *Predicted*, a única variável que não apresentou diferença estatística significativa entre as abordagens, demonstrando que, em termos relativos, *Amplicon* e *Shotgun* são capazes de inferir de forma semelhante a taxonomia no nível de gênero. Resultados obtidos por *Shotgun* permitem a caracterização mais detalhada do microbioma, com a identificação de maior diversidade e em nível taxonômico de espécie, quando comparado ao sequenciamento do 16S rDNA por *Amplicon*<sup>12</sup>.

Tabela 2: Análise descritiva dos dados.

Abordagem	Projeto	n	Variável	Min	Max	Média	Desvio-padrão	Shapiro Wilk P-value	
Amplicon	mgp3474	95	Failed	0	0	0,00	0,00	N/D	
			Unknown	20	495	101,00	62,30	Não	1,19E-010
			Predicted	5.478	55.077	13.525,81	6.499,75	Não	1,23E-010
			Alpha	4	65	23,36	12,35	Não	5,28E-006
			Rarefaction	140	647	276,86	96,42	Não	4,07E-005
	mgp4843	97	Failed	0	0	0,00	0,00	N/D	
			Unknown	15	560	98,30	86,10	Não	3,76E-012
			Predicted	6.557	101.899	15.410,12	12.713,13	Não	1,03E-014
			Alpha	7	54	24,65	10,59	Não	6,26E-003
			Rarefaction	139	429	244,01	57,93	Não	1,79E-003
	mgp7236	53	Failed	0	0	0,00	0,00	N/D	
			Unknown	0	0	0,00	0,00	N/D	
			Predicted	4.737	74.043	19.545,26	14.527,79	Não	6,56E-007
			Alpha	3	51	22,49	10,84	Sim	7,42E-001
			Rarefaction	24	469	154,14	64,52	Não	1,37E-006
Shotgun	mgp3474	8	Failed	205.803	483.518	339.000,00	95.079,01	Sim	8,99E-001
			Unknown	12.277	81.370	34.670,00	21.703,00	Sim	9,69E-002
			Predicted	0	35.033	15.083,00	14.503,80	Sim	1,38E-001
			Alpha	33	621	353,45	213,34	Sim	6,09E-001
			Rarefaction	1.068	1.984	1.618,42	315,52	Sim	1,31E-001
	mgp4843	73	Failed	0	0	0,00	0,00	N/D	
			Unknown	0	119.005	18.798,00	20.850,00	Não	8,10E-010
			Predicted	4.149	54.027	24.943,34	11.759,37	Não	1,92E-002
			Alpha	21	628	317,60	129,58	Sim	7,51E-001
			Rarefaction	1.769	3.791	2.546,36	296,34	Não	1,12E-006
	mgp7236	83	Failed	105.729	699.114	275.000,00	135.928,36	Não	3,39E-003
			Unknown	11.292	85.228	40.146,00	17.748,00	Sim	2,86E-001
			Predicted	0	4.821	243,50	922,79	Não	2,91E-011
			Alpha	9	481	168,42	104,75	Não	1,57E-003
			Rarefaction	50	1.993	1.327,83	584,53	Não	1,05E-003

Observação: n: número de seqüências por projeto; Min: valor mínimo e Máx: valor máximo

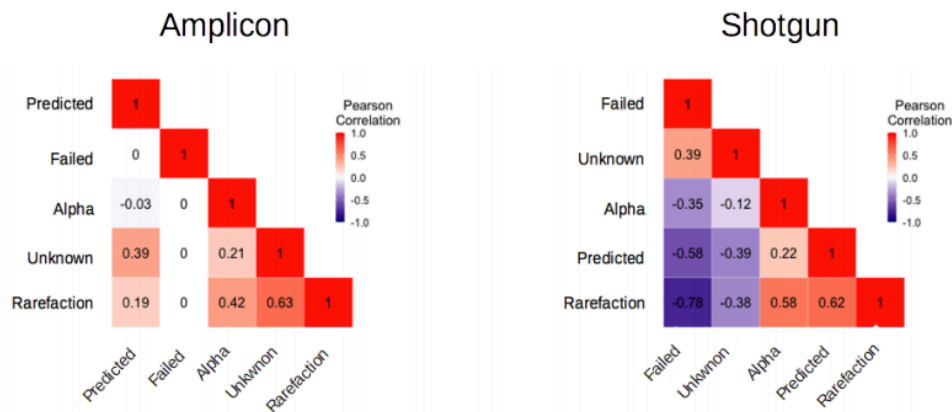


Figura 1: Estudo de Correlação por tipo de sequenciamento das variáveis: Failed, Unknown, Predicted, Alpha e Rarefaction.

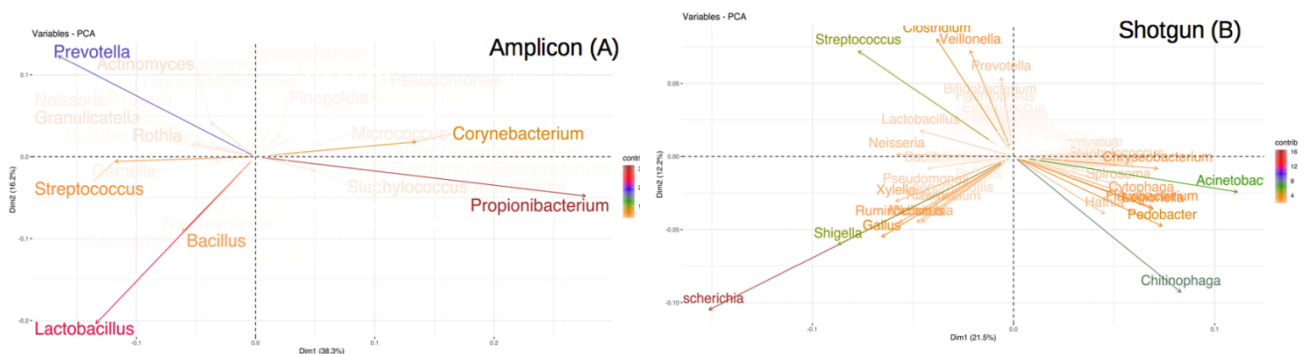


Figura 2: Gráfico da Análise de Componentes Principais (PCA) dos sequenciamentos de Amplicon e Shotgun.

No estudo de correlação, *Shotgun* se mostrou com dados mais confiáveis, uma vez que o comportamento esperado das variáveis foi observado apenas nas amostras desta abordagem. 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 sequências conhecidas, maior será o número de taxa identificado. De forma semelhante, *Failed e Rarefaction* apresentaram elevada 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 microorganismos depende da abordagem de sequenciamento escolhida.

### Conclusões

Nos *cores* da microbiota da saliva humana podem ser encontrados como gêneros mais representativos os associados a doenças bucais como cárie e doença periodontal, mas também gêneros que, até o momento, não são associados a estes problemas de saúde bucal.

Ambas as abordagens de sequenciamento apresentam a mesma sensibilidade de detecção da microbiota no nível taxonômico de gênero, apesar da grande diferença no número de sequências geradas entre as abordagens. Sequenciamentos realizados por *Shotgun* permitem a identificação de maior diversidade da microbiota em comparação ao sequenciamento por *Amplicon*.

A identificação do *core* é dependente da abordagem de sequenciamento e esta, por sua vez, é capaz de influenciar a diversidade da microbiota estudada.

Os dados estudados mostraram que a técnica *Shotgun* produz dados mais confiáveis do que a técnica de *Amplicon*.

A escolha da NGS empregada em estudos de microbioma deve considerar as características e as limitações das abordagens de sequenciamento. Sequenciamentos por *Shotgun* podem fornecer mais informações sobre a composição do microbioma e permitir a identificação de perfis de condições de saúde ou de doença. Sequenciamentos por *Amplicon* são mais indicados para a investigação de microorganismos de interesse que já seja conhecido ou definido ou para a investigar resultados previamente obtidos por *Shotgun*.

### Referências bibliográficas

1. Tanner ACR, Kressler CA, Rothmiller S, Johansson I, Chalmers NI. The Caries Microbiome: Implications for Reversing Dysbiosis. *Rev Adv Dent Res.* 2018;29(1):78–85.
2. Hurley E, Barrett MPJ, Kinirons M, Whelton H, Ryan CA, Stanton C, et al. Comparison of the salivary and dentinal microbiome of children with severe-early childhood caries to the salivary microbiome of caries-free children. *BMC Oral Health.* 2019;19(1):1–14.
3. 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.
4. Fakhruddin KS, Ngo HC, Samaranayake LP. Cariogenic microbiome and microbiota of the early primary dentition: A contemporary overview. *Oral Dis.* 2019;25(4):982–95.
5. 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.
6. 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.
7. Shade A, Handelsman J. Beyond the Venn diagram: The hunt for a core microbiome. *Environ Microbiol.* 2012;14(1):4–12.
8. Gupta S, Mortensen MS, Schjørring S, Trivedi U, Vestergaard G, Stokholm J, et al. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Commun Biol [Internet].* 2019;2(1):1–7. Available from: <http://dx.doi.org/10.1038/s42003-019-0540-1>
9. Willis JR, Gabaldón T. The human oral microbiome in health and disease: From sequences to ecosystems. *Microorganisms.* 2020;8(2):1–28.
10. Liu G, Wu C, Abrams WR, Li Y. Structural and Functional Characteristics of the Microbiome in Deep-Dentin Caries. *J Dent Res.* 2020;99(6):713–20.
11. Dawes C, Wong DTW. Role of Saliva and Salivary Diagnostics in the Advancement of Oral Health. *J Dent Res.* 2019;98(2):133–41.
12. Boers SA, Jansen R, Hays JP. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *Eur J Clin Microbiol Infect Dis.* 2019;38(6):1059–70.
13. 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.
14. Paczian T, Trimble WL, Gerlach W, Harrison T, Wilke A, Meyer F. The MG-RAST API explorer: an on-ramp for RESTful query composition. *BMC Bioinformatics.* 2019;20(1):1–8.
15. Belibasakis GN, Bostanci N, Marsh PD, Zaura E. Applications of the oral microbiome in personalized dentistry. *Arch Oral Biol.* 2019;104:7–12.
16. Rosier BT, Marsh PD, Mira A. Resilience of the oral microbiota in health - mechanisms that prevent dysbiosis. *J Dent Res.* 2018;97(4):371–80.
17. Zhang Y, Sun J, Lin CC, Abemayor E, Wang MB, Wong DTW. The emerging landscape of salivary diagnostics. *Periodontol 2000.* 2016;70(1):38–52.