

AVALIAÇÃO DO VIROMA E DO BACTERIOMA EM CARCINOMAS DE PÊNIS POR SEQUENCIAMENTO DE NOVA GERAÇÃO

ELIZABETH CRISTINA MIOLA

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

Área de concentração: Oncologia

Orientadora: Dra. Diana Noronha Nunes

Co-orientadora: Dra. Thais Fernanda Bartelli

São Paulo

2021

FICHA CATALOGRÁFICA

Miola, Elizabeth Cristina

Avaliação do viroma e do bacterioma em carcinomas de pênis por sequenciamento de nova geração/ Elizabeth Cristina Miola - São Paulo, 2021.
72p.

Dissertação (Mestrado)-Fundação Antônio Prudente.

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

Orientadora: Dra. Diana Noronha Nunes

Descritores: 1. Câncer de pênis 2. Sequenciamento 3. Viroma 4. Bacterioma
5. Ancestralidade

*Todos os direitos reservados à FAP. A violação dos direitos autorais constitui crime, previsto no art. 184 do Código Penal, sem prejuízo de indenizações cabíveis, nos termos da Lei nº 9.610/08.

DEDICATÓRIA

*Aos meus pais Evana e Gilson e meu irmão Matheus
por todo apoio incondicional, carinho e dedicação.
Agradeço imensamente!*

AGRADECIMENTOS

Quero primeiramente agradecer às três pessoas mais importantes da minha vida, minha mãe Evana que sempre ouviu sobre meu dia e sempre me apoiou, ao meu pai Gilson que sempre foi um exemplo de pessoa e profissional para mim, e ao meu irmão Matheus que me fez sorrir mesmo nos momentos difíceis. Sou extremamente grata! À minha Vó Elizabeth, que sempre me apoiou e orava por mim, para que tudo desse certo ao longo do dia. À toda minha família que sempre me incentivaram e foram meu suporte emocional.

Faço aqui também um agradecimento à Dra. Diana Noronha Nunes por ter me orientado e por todo aprendizado e a minha co-orientadora Dra. Thais Bartelli por ter me ajudado muito com o projeto. À Dra. Maria Galli de Amorim pelos ensinamentos e auxílio. Ao Dr. Emmanuel Dias Neto por ter me recebido no Laboratório de Genômica Médica e aos meus colegas de laboratório Gabriela, Bruno, Isabella e Luana por todo auxílio e colaboração durante a pesquisa.

À Dra. Valquíria Martins e a Dra. Kátia Torres da Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON) por cederem parte das amostras utilizadas nesta pesquisa. Aos membros do laboratório de bioinformática do A.C.Camargo Cancer Center (ACCCC) o Dr. Israel Tojal, Renan Valieris e Alexandre Defelicibus e do laboratório de bioinformática da USP de Ribeirão o Dr. Wilson Araújo e seu aluno Rafael Bezerra agradeço muito pela contribuição ao projeto.

Ao Dr. Stênio Zequi e a Dra. Stephania Bezerra por todo auxílio com os dados clínicos dos pacientes e as enfermeiras de pesquisa Fernanda e Cassia que ajudaram na coleta dos dados. Ao biobanco, biblioteca e secretaria de pós-graduação do ACCCC, meus agradecimentos. Aos membros da banca examinadora de qualificação e defesa pelas contribuições e sugestões.

Ao Programa Nacional de Apoio à Atenção Oncológica (PRONON), Instituto Nacional de Ciência e Tecnologia em Oncogenômica e Inovação Terapêutica (INCiTO-INOTE) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela contribuição financeira ao projeto. À Coordenação de Aperfeiçoamento de Pessoa de Nível Superior (CAPES) pela bolsa de estudos.

Agradeço imensamente a todos!

RESUMO

Miola EC. **Avaliação do viroma e do bacterioma em carcinomas de pênis por sequenciamento de nova geração.** [Dissertação]. São Paulo: Fundação Antônio Prudente; 2021.

O câncer de pênis é uma neoplasia rara, responsável por 2% de todos os tumores que atingem os homens no Brasil. Sua incidência é maior nas regiões Norte e Nordeste e em homens com idade superior a 50 anos, tendo como fatores de risco a má higiene, fimose, e infecção pelo vírus HPV. Um estudo multicêntrico nacional pioneiro do A.C.Camargo Cancer Center (ACCCC) demonstrou a associação entre sexo com animais (SWA) e o câncer de pênis, levantando a hipótese de que alguns microrganismos presentes em animais possam estar envolvidos nesta carcinogênese. Neste sentido, este estudo teve como objetivo caracterizar o viroma e o bacterioma tumoral de indivíduos diagnosticados com câncer de pênis visando a melhor compreensão da microbiota associada à esta neoplasia. As amostras foram coletadas de indivíduos que foram tratados no ACCCC e na Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON). O viroma foi avaliado em 23 amostras negativas para a presença dos vírus EBV e HPV por meio do enriquecimento de partículas virais extracelulares antes da extração do DNA, seguido por sequenciamento de nova geração com alta cobertura na plataforma NextSeq 500. Identificamos genomas virais em 9/23 amostras, com predomínio dos gêneros *Circovirus* e *Parvovirus*. No entanto, muitas sequências ainda permaneceram não anotadas devido à escassez de sequências anotadas disponíveis nos bancos de dados virais. Através de análises das sequências humanas *off-target* do sequenciamento do viroma, avaliamos a ancestralidade molecular dos 24 pacientes, que mostrou que o componente ameríndios foi mais enriquecido nas amostras do estado do Amazonas, enquanto que as amostras da coorte de São Paulo não tiveram predominância de nenhuma ancestralidade específica. Para caracterização do bacterioma, as regiões V3-V4 do gene 16S rRNA foram amplificadas e sequenciadas de 53 amostras tumorais e 6 tecidos não neoplásicos de borda de lesão, independente do seu status de EBV ou HPV. De maneira geral, houve predomínio dos gêneros *Fusobacterium*, *Corynebacterium* e *Pseudomonas*. A diversidade bacteriana foi maior nas amostras coletadas na FCECON, e a composição bacteriana variou significativamente de maneira quantitativa e qualitativa quando comparadas com as amostras do ACCCC, sendo que o gênero *Alcaligenes* foi mais abundante na coorte de São Paulo e *Pseudomonas* nos pacientes do Amazonas. Pacientes sem fimose apresentaram maior riqueza e diversidade bacteriana em

comparação ao grupo com fimose, sendo que o gênero *Enterococcus* (anaeróbios facultativos) foi mais abundante nos pacientes com fimose e o gênero *Oligella* (aeróbio) mais enriquecido nos pacientes sem fimose. O gênero *Alcaligenes* é um possível marcador para presença tanto de invasão perineural, quanto de metástase linfonodal na coorte de São Paulo, sugerindo que sua presença possa estar associada a um pior prognóstico na evolução do câncer de pênis. Os resultados preliminares obtidos neste estudo serão validados em uma coorte maior de pacientes, incluindo dados de sequenciamento *shogun* que permitirão obtermos classificação das bactérias a nível de espécie e cepa, além de permitir caracterização de outros microrganismos, como fungos.

Palavras-chaves: Câncer de pênis, sequenciamento de nova geração, viroma, bacterioma, ancestralidade.

SUMMARY

Miola EC. **[Virome and Bacteriome analysis from penile carcinoma by Next Generation Sequencing]**. [Dissertação]. São Paulo: Fundação Antônio Prudente; 2021.

Penile cancer is a rare neoplasm, responsible for 2% of all tumors that affect men in Brazil. Its incidence is higher in the North and Northeast regions and in men over 50 years of age, with risk factors such as poor hygiene, phimosis, and HPV virus infection. A pioneering national multicenter study by the A.C.Camargo Cancer Center (ACCCC) demonstrated the association between sex with animals (SWA) and penile cancer, raising the hypothesis that some microorganisms present in animals may be involved in this carcinogenesis. In this sense, this study aimed to characterize the tumor virome and bacteriome of individuals diagnosed with penile cancer, in order to better understand the microbiota associated with this neoplasm. Samples were collected from individuals who were treated at the ACCCC and the Amazonas State Oncology Control Center Foundation (FCECON). Virome was obtained for 23 samples negative for the presence of EBV and HPV viruses by enriching extracellular viral particles before DNA extraction, followed by next-generation sequencing with high coverage in the NextSeq 500 platform. We identified viral genomes on 9/ 23 samples, with a predominance of *Circovirus* and *Parvovirus*. However, many sequences remained unclassified due to the scarcity of annotated sequences available in viral databases today. Through analysis of human off-target sequences from the viroma sequencing, we assessed the molecular ancestry of the 24 patients, which showed that the Amerindian component was more enriched in samples from the state of Amazonas, while samples from the São Paulo cohort did not have a predominance of any specific ancestry. To characterize the bacteriome, we sequenced the V3-V4 regions of the 16S rRNA gene from 53 tumor samples and 6 non-neoplastic tumor lesion margins, regardless of their EBV or HPV status. In general, there was a predominance of the genera *Fusobacterium*, *Corynebacterium* and *Pseudomonas*. Bacterial diversity was greater in the samples collected at FCECON, and the bacterial composition varied significantly quantitatively and qualitatively when compared to the ACCCC samples, with *Alcaligenes* and *Pseudomonas* being differentially abundant between the two cohorts. In addition, individuals who not had phimosis showed greater bacterial diversity, although we did not observe significant differences in their composition compared with those who did have this condition (beta diversity). The preliminary results obtained in this study will be validated in a larger cohort of patients, including shogun

sequencing data that will allow us to obtain classification of bacteria at the level of species and strain, besides allowing the characterization of other microorganisms, such as fungi.

Keywords: Penile cancer, sequencing, virome, bacteriome, ancestry.

LISTA DE FIGURAS

Figura 1 - Anatomia do pênis.	2
Figura 2 - Metodologias utilizadas para estudo do microbioma.	5
Figura 3 - Proporção de população por regiões incluídas no estudo (Brasil).....	10
Figura 4 - Casuística incluída no estudo de acordo com a análise realizada (viroma ou bacterioma).	14
Figura 5 - Modelo de protocolo de isolamento de partículas virais.	16
Figura 6 -Análise de amplificação das regiões V3 e V4 do rRNA 16S em gel de agarose 1% de amostras de pacientes do Amazonas e de São Paulo.	22
Figura 7 - Avaliação de HPV por PCR nas amostras de câncer de pênis dos pacientes do ACCCC.....	32
Figura 8 - Análise de <i>reads</i> dos viromas dos pacientes com câncer de pênis. Sequências alinhadas com hg19	34
Figura 9 - Composição do viroma em nove amostras de tumores de pênis.	37
Figura 10 - Ancestralidade molecular de 24 pacientes das coortes do ACCCC (SP) e FCECON (AM), que tiveram seus viromas estudados	38
Figura 11 - Curvas de rarefação de <i>amplicons</i> do gene rRNA 16S derivado das amostras de câncer de pênis (OTUs e número de <i>reads</i>).	39
Figura 12 - Abundância relativa das principais famílias de bactérias identificadas em amostras de pênis com e sem neoplasia.....	40
Figura 13 - Abundância relativa dos principais gêneros de bactérias identificados em amostras de pênis com e sem neoplasia.....	41

Figura 14 - Gêneros bacterianos com frequência relativa acima de 1% em amostras de câncer de pênis.....	42
Figura 15 - Correlação da abundância relativa (log10) dos gêneros bacterianos entre as amostras do Amazonas e São Paulo.	43
Figura 16 - Gêneros bacterianos com frequência relativa acima de 1% em amostras do estado de São Paulo.	44
Figura 17 - Gêneros bacterianos com frequência relativa acima de 1% em amostras do estado do Amazonas.	44
Figura 18 – Medidas de diversidades alfa do bacterioma de carcinomas de pênis em relação a origem das amostras de SP e AM.....	45
Figura 19 – Medidas de diversidade Beta do microbioma de carcinomas de pênis em relação a origem das amostras de SP e AM.....	46
Figura 20 - Análise dos gêneros bacterianos diferencialmente abundantes (LefSe) entre as amostras tumorais do Amazonas e São Paulo..	47
Figura 21 - Frequência relativa e distribuição dos gêneros <i>Alcaligenes</i> e <i>Pseudomonas</i> entre as amostras tumorais de São Paulo e Amazonas.	47
Figura 22 – Medidas de diversidade alfa entre as amostras pareadas de tecido tumoral e tecido sem neoplasia (normal).....	48
Figura 23 – Medidas diversidade Beta das amostras pareadas de tecido tumoral e tecido sem neoplasia (normal).....	49
Figura 24 - Análise LefSe e abundância relativa (LefSe) do gênero <i>Lactobacillus</i> nas amostras de tecidos normais e tumorais..	50
Figura 25 – Medidas de diversidades alfa em relação ao status de fimose.....	51
Figura 26 – Medidas de diversidades beta em relação ao status de fimose.	51

Figura 27 - Análise dos gêneros bacterianos diferencialmente abundantes (Lefse) entre os grupos de pacientes com câncer de pênis com e sem fimose.	52
Figura 28 - Medidas de diversidades alfa em relação a ancestralidade molecular.....	53
Figura 29 - Medidas de diversidades beta em relação a ancestralidade molecular.	54
Figura 30 - Análise em LefSe e abundância relativa dos gêneros bacterianos em relação a ancestralidade molecular..	54
Figura 31 – Medidas de diversidade alfa em relação ao status dos vírus HPV e EBV em amostras tumorais.	55
Figura 32 – Medidas de diversidade beta em relação ao status dos vírus HPV e EBV em amostras tumorais	56
Figura 33 – Medidas de diversidade alfa em relação a invasão perineural nos casos de câncer de pênis de São Paulo.	57
Figura 34 – Medidas de diversidade beta em relação a invasão perineural nos casos de câncer de pênis de São Paulo.	57
Figura 35 - Análise de LefSe entre os grupos de pacientes com câncer de pênis do estado de SP com e sem invasão perineural.	58
Figura 36 – Medidas de diversidade alfa em relação a metástase linfonodal nos casos de carcinoma de pênis de SP e AM.....	59
Figura 37 - Medidas de diversidade beta em relação a metástase linfonodal nos casos de câncer de pênis de SP e AM.	59
Figura 38 - Análise LefSe das amostras de câncer de pênis com e sem metástase linfonodal..	60
Figura 39 - Medidas de diversidade alfa em relação ao tabagismo nos casos de câncer de pênis..	61

Figura 40 - Medidas de diversidade beta em relação ao tabagismo nos casos de câncer de pênis.

LISTA DE TABELAS

Tabela 1 -Estadiamento do câncer utilizando o sistema da 8ª. Edição TNM da AJCC - American Joint Committee on Cancer.	3
Tabela 2 - Características gerais dos pacientes selecionados para avaliação do viroma.....	26
Tabela 3 - Aspectos histopatológicos dos pacientes que tiveram os viromas sequenciados ...	27
Tabela 4 - Parâmetros clínicos dos pacientes selecionados para avaliação do viroma	28
Tabela 5 - Características clínicas dos pacientes que tiveram o bacterioma avaliado.....	29
Tabela 6 - Aspectos histopatológicos dos pacientes que tiveram bacterioma tumoral avaliado	30
Tabela 7 - Parâmetros clínicos dos pacientes que tiveram o bacterioma avaliado	31
Tabela 8 - Amostras positivas para HPV e/ou EBV dos pacientes de São Paulo (ACCCC) ..	32
Tabela 9 - Amostras selecionadas para avaliação do viroma.	33
Tabela 10 – Anotação das sequências do viroma	34
Tabela 11 - Amostras que apresentaram anotação para sequências virais.	35

LISTA DE SIGLAS E ABREVIATURAS

°C – Celsius

µl – Microlitro

µm – micrômetro

ACCCC – A.C Camargo Cancer Center

AJCC – *American Joint Committee on Cancer* (Comitê Conjunto Americano sobre Câncer)

bp – Pares de base

EBV – *Epstein-Barr virus* (vírus Epstein-Barr)

CaCl₂ – Cloreto de Cálcio

CEP – Comitê de Ética em Pesquisa

DNA – *Deoxyribonucleic acid* (Ácido desoxirribonucleico)

FCECON – Fundação Centro de Controle de Oncologia do Estado do Amazonas

FMRP – Faculdade de Medicina de Ribeirão Preto

GATAK – *Genome Analysis Toolkit*

HPV – *Human papiloma-virus* (papilomavírus humano)

h – Hora

INCiTO-INOTE - Instituto Nacional de Ciência e Tecnologia em Oncogenômica e Inovação Terapêutica.

LEFSE - Análise Discriminante Linear (LDA) associada a tamanho de efeito (SE)

MgCl₂ – Cloreto de Magnésio

min – Minuto

mg - Miligrama

mL – Mililitro

M – Molar

mM – Milimolar

NGS – *Next Generation Sequencing* (Sequenciamento de Nova Geração)

NCBI – *National Center for Biotechnology Information* (Centro Nacional de Informações sobre Biotecnologia)

ng – Nanograma

OTU – *Operational Taxonomic Unit* (Unidade taxonômica operacional)

PCR – *Polymerase Chain Reaction* (Reação em cadeia da polimerase)

PIN – *Penile Intraepithelial Neoplasia* (lesões intra-epiteliais penianas)

Qiime – *Quantitative Insights Into Microbial Ecology*

qPCR - Quantitative Polymerase Chain Reaction

(Reação em cadeia da polimerase quantitativa)

RNA – Ribonucleic acid (Ácido ribonucleico)

Rpm – Rotações por minuto

SUS – Sistema Único de Saúde

SEER - *Surveillance, Epidemiology, and End Results* (Programa de Vigilância, Epidemiologia e Resultados)

SNPs – *Single Nucleotide Polymorphism* (polimorfismo de nucleotídeo único)

SWA – *sex with animals* (sexo com animais)

seg – Segundos

TNM – *Tumor-Node-Metastase* (Tumor-Nódulo-Metástase)

TCGA – The Cancer Genome Atlas (Atlas do genoma do câncer)

UV – Ultravioleta

VB – Vaginose bacteriana

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Neoplasia maligna de pênis	1
1.2 Bacterioma e Viroma	4
1.3 Vírus associados ao câncer de pênis: EBV e HPV	8
1.4 Ancestralidade genética	9
 2. OBJETIVOS	 12
 3. PACIENTES E MATERIAIS E MÉTODOS	 13
3.1 Aspectos éticos e população estudada	13
3.2 Extração de ácidos nucleicos	14
3.3 Detecção de HPV e EBV	14
3.4 Isolamento de partículas virais	15
3.5 Homogeneização do tecido tumoral	16
3.6 Enriquecimento de partículas virais.....	17
3.7 Extração e quantificação de DNA/RNA	17
3.8 Amplificação do DNA/RNA e preparo da biblioteca	18
3.9 Sequenciamento das amostras.....	19
3.10 Análise de bioinformática do viroma.....	19
3.11 Ancestralidade genética	20
3.12 Extração de DNA das amostras selecionadas para o bacterioma	21
3.13 Geração dos <i>amplicons</i> das regiões V3-V4 do gene rRNA 16S.....	21
3.14 Análise de bioinformática do bacterioma	23
 4. RESULTADOS E DISCUSSÃO	 25
4.1 Informações clínicas dos pacientes incluídos no estudo do viroma	25
4.2 Informações clínicas dos pacientes incluídos no estudo do bacterioma	28
4.3 Extração e triagem para HPV e EBV.....	31
4.4 Avaliação dos viromas tumorais.....	33
4.5 Análise de ancestralidade molecular.....	37

4.6	Caracterização dos bacteriomas das amostras de câncer de pênis	38
4.7	Avaliação entre os bacteriomas e parâmetros clínicos-patológicos.....	42
4.7.1	Correlação entre os bacteriomas de São Paulo (ACCCC) e Amazonas (FCECON) .	42
4.7.2	Análise do bacterioma entre tecido tumoral versus normal	48
4.7.3	Avaliação do bacterioma entre amostras de carcinomas de pênis com e sem fimose	50
4.7.4	Avaliação do bacterioma dos pacientes em relação a ancestralidade molecular	53
4.7.5	Análise do bacterioma dos pacientes em relação ao status de HPV e EBV	55
4.7.6	Análise do bacterioma dos carcinomas de pênis em relação a invasão perineural e metástase linfonodal	56
4.7.7	Análise do bacterioma dos carcinomas de pênis em relação ao tabagismo	60
4.7.8	Avaliação de associação dos dados de bacterioma com outros parâmetros clínicos .	62
5.	CONCLUSÕES	63
6.	REFERÊNCIAS BIBLIOGRÁFICAS.....	65

ANEXOS

Anexo 1 Carta de aprovação do Comitê de Ética em Pesquisa (CEP)

Anexo 2 Carta de anuência da Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON)

1. INTRODUÇÃO

1.1 NEOPLASIA MALIGNA DE PÊNIS

Quando combinados, os tumores raros representam aproximadamente 25% de todos os cânceres e seu estudo é uma das linhas estratégicas de pesquisa do A.C.Camargo Cancer Center (ACCCC) avaliadas no projeto Instituto Nacional de Ciência e Tecnologia em Oncogenômica e Inovação Terapêutica (INCiTO-INOTE). Quando comparado a outros tipos mais frequentes, os tumores raros tendem a ter um pior prognóstico, em parte por serem menos estudados e apresentarem um leque limitado de terapias-alvo adequadas (1). Neste sentido, o ACCCC possui uma casuística relevante de tumores raros, que se apresenta muito valiosa para a pesquisa científica e com potencial de beneficiar futuros pacientes.

A neoplasia maligna de pênis é um tipo de tumor raro que atinge por volta de 22.000 homens por ano no mundo (2). No Brasil foram contabilizados 458 óbitos em 2019, sendo sua maior incidência nas regiões do Norte e Nordeste, com o estado do Maranhão apresentando a maior incidência de câncer de pênis no Brasil e no Mundo (3). Esse tipo de câncer afeta homens principalmente a partir dos 50 anos de idade (4) e está associado a fatores de risco como a infecção por HPV, fimose, baixo nível socioeconômico, más condições de higiene e tabagismo (5,6). Além desses fatores, um estudo multicêntrico nacional pioneiro do ACCCC demonstrou a associação entre sexo com animais (SWA) e câncer de pênis, levantando a hipótese de que microorganismos infecciosos pudessem estar envolvidos em sua carcinogênese (7). Por serem raros e mais incidentes em países não-desenvolvidos (8), estudos sobre a neoplasia maligna de pênis em nosso país são de grande importância, pois permitem um melhor entendimento da etiologia e desenvolvimento desta neoplasia rara, podendo oferecer caminhos futuros para o desenvolvimento de terapias mais efetivas.

Assim como na maioria dos tumores, o diagnóstico precoce está associado a um melhor prognóstico na neoplasia maligna de pênis. Segundo dados americanos do *Surveillance, Epidemiology, and End Results* (SEER), pacientes diagnosticados com tumores localizados apresentam uma taxa de sobrevida relativa em 5 anos de 82%, enquanto que pacientes que apresentam tumores com disseminação regional, tem sua sobrevida média reduzida para cerca de 50% (9).

O câncer de pênis pode ser confirmado ou descartado através de uma biópsia, podendo ser uma biópsia excisional ou por aspiração, dependendo da extensão e localização da lesão,

enquanto exames de imagens podem ser pedidos de forma complementar e para acompanhamento do tratamento (10). Em 95% dos casos o tipo histológico mais comum encontrado é o carcinoma epidermoide, porém podemos encontrar outras variantes histológicas como verrucoso, basocelular, adenocarcinoma, melanoma e lesões intra-epiteliais penianas (PIN) de alto grau que podem ser chamadas de doença de Bowen ou eritroplasia de Queyrat dependendo da localização do tumor (10,11). Infelizmente, após o surgimento das lesões iniciais, que são muitas vezes indolores, os pacientes demoram até um ano para buscar avaliação médica, fazendo com que a maioria dos casos seja diagnosticado quando existe a presença de lesões mais avançadas (12). No início a neoplasia maligna do pênis se apresenta nas camadas mais superficiais, mas quando progride a lesão se desenvolve para o corpo e prepúcio do pênis, acometendo a uretra, corpos cavernosos, podendo atingir também os linfonodos (13) (**Figura 1**).

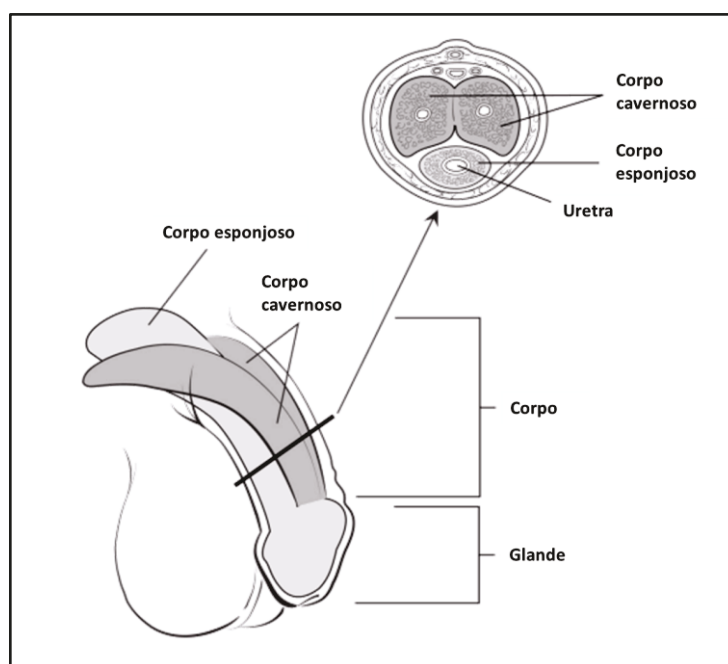


Figura 1 - Anatomia do pênis. **Fonte:** Figura adaptada de <https://www.cancer.org/cancer/penile-cancer/about/what-is-penile-cancer.html> (14).

Uma vez confirmado o diagnóstico anatomopatológico do câncer, o estadiamento clínico (TNM) do paciente é realizado para estabelecer a melhor forma de tratamento, oferecendo informações como tamanho do tumor (T), nódulos acometidos (N) e metástase (M) (15) (**Tabela 1**). Normalmente se o tumor é pequeno e não se espalhou o tratamento utilizado pode ser a biópsia excisional ou laserterapia. Em tumores avançados como ocorre na maior parte dos casos, pode ser realizada a cirurgia de penectomia parcial ou total, considerado o tratamento

mais eficaz para o controle da doença. Em casos muito avançados é possível que o paciente tenha que fazer a remoção do escroto e testículo, sendo a linfadenectomia realizada em pacientes com evidências de linfonodos inguinais comprometidos. Outros métodos que também podem ser utilizados em recidiva, ou mesmo paliativamente, são a radioterapia e quimioterapia, para carcinomas de células escamosas (PSCC) avançado. Neste caso a cisplatina é utilizada, porém as taxas de resposta são baixas (15-55%), sendo que a sobrevida geral não é superior a 12 meses (10–12). Segundo estatísticas do A.C.Camargo Cancer Center (ACCCC), 90% dos casos de câncer de pênis atendidos nas unidades são de estágio avançado quando a amputação é praticamente inevitável (10). O seguimento dos casos normalmente ocorre dentro de um período de 5 anos, que é quando ocorre aproximadamente 92% das recorrências (16).

Tabela 1 -Estadiamento do câncer utilizando o sistema da 8ª. Edição TNM da AJCC - *American Joint Committee on Cancer*.

Estágio	Agrupamento	Descrição do estágio
0	Tis ou Ta N0, M0	O tumor atingiu apenas a camada superficial da pele, também chamado de carcinoma in situ.
I	T1a, M0, N0	O tumor cresceu abaixo das camadas superficiais da pele, porém não atingiu os vasos sanguíneos ou linfáticos, nem linfonodos próximos ou áreas distantes.
IIA	T1b, N0, M0	Tumor cresceu em todo tecido abaixo da camada superficial da pele, cresceu em vasos sanguíneos, linfáticos próximos, porém não se espalhou para linfonodos próximos ou outras áreas distantes.
IIA	T2, N0, M0	O câncer se espalhou para o corpo esponjoso. Não se disseminando para outras áreas do corpo.
IIB	T3,N0,M0	O câncer se espalhou para corpo cavernoso. Não se disseminando para outras áreas do corpo.
IIIA	T1-T3, N1, M0	O tumor atingiu o tecido abaixo da camada superior da pele e pode ter alcançado corpo cavernoso e/ou corpo esponjoso. O tumor se espalhou para um único gânglio linfático, não se espalhando para partes mais distantes.
IIIB	T1-T3, N2, M0	O câncer se desenvolveu nos tecidos do pênis e pode ter alcançado o corpo esponjoso e/ou corpo cavernoso ou uretra, se espalhando para três ou mais linfonodos inguinais. Não se espalhando para partes mais distante.
IV	T4, qualquer N, M0 ou	O tumor atingiu estruturas próximas como escroto, próstata ou osso púbico.
	qualquer T, N3, M0 ou	O tumor pode ou não ter crescido em camadas mais profundas do pênis ou se disseminado para estruturas mais próximas.
	qualquer T, qualquer N, M1	O tumor pode ou não ter crescido em camadas mais profundas do pênis ou estruturas próximas. O câncer pode ou não ter se espalhado para nódulos linfáticos próximos. O câncer se disseminou para partes distantes do corpo.

Fonte: Adaptada de <https://www.cancer.org/cancer/penile-cancer/detection-diagnosis-staging/staging.html> (17) e <https://accamargo.org.br/tipos-de-cancer/penis> (18).

1.2 BACTERIOMA E VIROMA

O microbioma bacteriano, ou bacterioma, é o conjunto dos genomas das bactérias presentes em um determinado ambiente ou amostra. As bactérias são microorganismos procariotos e que são tradicionalmente classificadas de acordo com sua forma (19). O DNA bacteriano pode ser avaliado por meio do método de sequenciamento de produtos de PCR das regiões conservadas do gene codificador do RNA ribossomal 16S (rRNA 16S), ou pela metodologia de sequenciamento *shotgun*, tendo cada método suas vantagens e desvantagens dependendo do objetivo da pesquisa.

O gene 16S rRNA é um marcador genético universal das bactérias que apresenta regiões hipervariáveis e específicas para cada bactéria. O sequenciamento destas regiões hipervariáveis é utilizado em grande parte das pesquisas de microbioma que tem como objetivo estudar as apenas as bactérias, evitando assim o sequenciamento de outros microrganismos, ou mesmo do material genético do hospedeiro. Tal estratégia é importante especialmente em amostras que contém muito DNA humano (como biópsias teciduais) além de ser um método mais econômico e que proporciona a identificação até nível de gênero em mais de 90% dos casos. Dentre suas desvantagens está o fato de esta abordagem não permitir a determinação taxonômica à nível de espécie e cepas e por ser dependente de oligonucleotídeos, que levam a uma amplificação seletiva de algumas regiões gênicas, podendo não detectar alguns microrganismos (20,21). Em um estudo de Engelbrektson *et al*, 2010, em que foi avaliado o sequenciamento de *amplicons* por 16S, foi evidenciado a relação entre a riqueza e uniformidade dos taxa encontrados com o tamanho dos *amplicons* e os iniciadores utilizados (22). Isso ressalta a importância da escolha dos iniciadores e da região hipervariável do gene 16S rRNA a ser sequenciada para cada estudo. Por outro lado, o sequenciamento *shotgun* fornece informações mais completas sobre os microrganismos, pois capta todo material genético presente na amostra e está menos sujeito a viés na amplificação por não depender da escolha de oligonucleotídeos. Tal abordagem pode proporcionar a identificação a nível de espécie e cepa, no entanto, é uma metodologia bem mais cara (**Figura 2**) (23,24). Em nosso estudo, optamos por avaliar o gene 16S rRNA devido ao custo e porque tínhamos interesse em avaliar o bacterioma associado ao tecido.

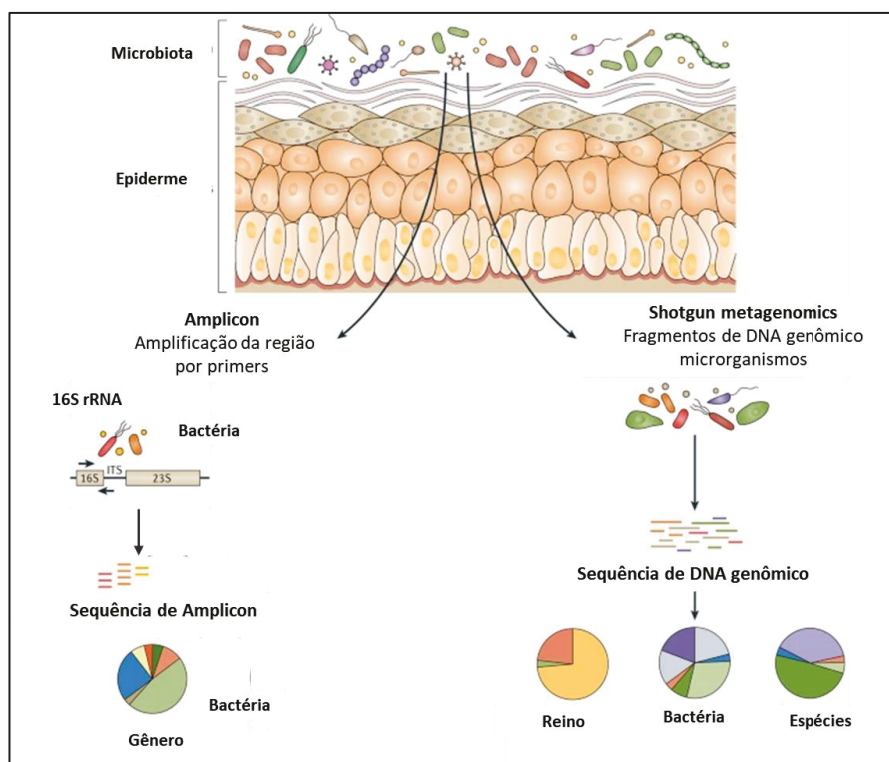


Figura 2 - Metodologias utilizadas para estudo do microbioma **Legenda:** O material genético extraído pode ser sequenciado a partir da amplificação de regiões hipervariáveis do gene 16S rRNA, utilizando iniciadores específicos, ou pela metodologia de *Shotgun* que permite o sequenciamento do DNA de todos os microorganismos presentes na amostra. **Fonte:** Figura adaptada de Byrd *et al*, 2011 (21).

Estudos da microbiota peniana são raros, porém os poucos avaliados foram voltados para identificação do bacterioma associada ao tecido. Tais estudos usaram as regiões hipervariáveis do gene 16S rRNA para a identificação taxonômica das bactérias presentes no tecido peniano. Um deles foi um estudo caso-controle realizado em homens não circuncidados infectados por HIV (casos) e não infectados (controle), para avaliação das bactérias anaeróbias, citocinas e HIV. O estudo averiguou se a abundância das bactérias anaeróbias, que aumentava diretamente o risco de infecção por HIV ao induzir uma resposta inflamatória, foi reduzido quando realizada a postectomia, cirurgia que leva a redução do número de bactérias anaeróbias. Neste estudo as regiões V3-V4 do gene 16S rRNA foram avaliadas por sequenciamento de nova geração utilizando amostras de *swab* peniano. Foi observado principalmente um aumento de abundância das bactérias anaeróbicas *Provatella*, *Dialister*, *Finegoldia* e *Peptoniphilus*, associadas aos casos HIV positivos, sugerindo que talvez as bactérias anaeróbias possam criar um ambiente de maior risco para infecção por HIV (25). Outro estudo que também observou a diferença da microbiota de homens com e sem infecções por HIV, incluindo também a avaliação de infecção por HPV, foi realizado na África do Sul por Onywera *et al*, 2020. Foram avaliadas 238 amostras de esfregaço de glândula, sulco-coronal e corpo do pênis e utilizando sequenciamento das regiões

V3-V4 do gene rRNA 16S. Os gêneros *Corynebacterium* e *Prevotella* foram as mais abundantes e diferenças significativas na composição da microbiota de homens com e sem infecção foram identificadas. A maior abundância relativa de *Staphylococcus*, e de bactérias associadas a vaginose bacteriana (*Prevotella*, *Peptinophilus* e *Dialister*), foi correlacionado com infecções por HIV e HPV, respectivamente (26).

Foi descrito em outros estudos a associação entre as microbiotas peniana e vaginal. Em Mehta *et al*, 2020 foram recrutados homens e suas parceiras sexuais para avaliar o potencial da microbiota peniana como fator de predição de vaginose bacteriana (VB) incidente, através do sequenciamento do gene rRNA 16S (regiões V3-V4 avaliadas) de amostras de esfregaço vaginal e esfregaços das regiões meatal, glândula e coronal do pênis. Foi observado que a microbiota peniana previu com precisão a incidência de vaginose bacteriana (VB), sendo que a modulação do microbioma peniano pode contribuir para o microbioma vaginal (27).

As bactérias que fazem parte da microbiota peniana têm perfis distintos, podendo ser aeróbias, anaeróbias facultativas e anaeróbias obrigatórias, além das microaerófilas, que são aquelas sensíveis a altas concentrações de oxigênio, mas ainda capazes de crescer na presença de mesmo. A circuncisão por exemplo, está associada a uma diminuição de bactérias anaeróbias (28). Em um estudo de Price e colaboradores de 2010, foram avaliadas as microbiotas de amostras de esfregaço do pênis de homens negativos para HIV. Através da amplificação e sequenciamento das regiões V3-V4 do gene rRNA 16S, foi observada uma diminuição de anaeróbios e uma mudança significativa na microbiota foi associada à circuncisão. Duas bactérias eram mais abundantes antes da circuncisão, as da ordem *Clostridiales* e família *Prevotellaceae*, enquanto após a circuncisão observou-se uma diminuição de bactérias anaeróbias (29).

Um dos grupos componentes da microbiota é o viroma, que é definido como o conjunto dos genomas de todos os vírus presentes em um determinado nicho (30). Os vírus variam de acordo com o tipo de material genético que eles possuem (DNA ou RNA), se esse material é simples ou dupla fita (ssDNA, ssRNA, dsDNA ou dsRNA), sua estrutura (linear ou circular), seu método de replicação (lítico ou lisogênico), e o tipo de célula que eles infectam (procarioto ou eucarioto). Com o avanço das técnicas de genômica, principalmente de sequenciamento de nova geração, ou NGS (*next generation sequencing*), hoje é possível analisar o material genético (DNA ou RNA) de todos os vírus presentes em uma amostra, de modo abrangente e a um custo relativamente baixo. O NGS possibilita uma melhor compreensão da relação entre os vírus e hospedeiros, devido à enorme quantidade de sequências genômicas que são geradas. Os dados gerados em estudos de metagenômica por NGS podem ser utilizados para identificação

de ácidos nucleicos virais, assim como para a caracterização de genes do hospedeiro, e também alterações na expressão gênica (se o RNA for avaliado simultaneamente) através da metatranscriptômica, durante a infecção do hospedeiro pelo vírus (31).

Apesar da recente e constante melhoria nas técnicas de sequenciamento e análise metagenômica, os bancos de dados de referências virais ainda são incompletos pois a maioria dos estudos ainda é focado na caracterização de bactérias (32). Além disso, devido a diversidade e complexidade viral, não existem marcadores universais que poderiam auxiliar na sua atribuição taxonômica, como é o caso do gene 16S rRNA para bactérias. O *National Center for Biotechnology Information* (NCBI) mantém um dos principais bancos de dados de sequências virais, que hoje inclui mais de 3 milhões de sequências nucleotídicas, contendo 15 mil genomas virais completos. Também há bancos de dados específicos para análise de variação viral (*Virus Variation Resource*), para os vírus influenza, dengue, ebola, rotavírus e Zika, dentre outros vírus relevantes.

Alguns estudos de viroma focam na avaliação das partículas virais extracelulares, que estão relacionadas com infecção ativa, enquanto outros são realizados com base em sequências virais integradas no genoma humano. Aproximadamente 8% do genoma humano consiste de elementos retrovirais endógenos, que não causam infecção (33). Cantalupo e colaboradores analisaram a presença de sequências virais em tecidos tumorais (34), utilizando dados do *The Cancer Genome Atlas* (TCGA), um banco de dados de sequenciamento de DNA composto de milhares de amostras de 33 tipos de câncer (35). Dentre os achados estão a detecção de sequências de vírus do papiloma humano (HPV) inseridas no genoma de praticamente todos os carcinomas cervicais, e também em alguns tumores de cabeça e pescoço e de bexiga; presença de vírus da hepatite B e hepatite C em alguns tumores de fígado; e de EBV em alguns tumores de estômago (34).

O microbiologista Jeffrey Gordon e seus colegas isolaram e caracterizaram o viroma de gêmeos idênticos adultos supostamente saudáveis e de suas mães utilizando amostras de fezes coletadas três vezes no decorrer de um ano. O número de vírus identificados nas amostras variou consideravelmente de 52 a 2773 em média por amostra, porém interessante o perfil viral de cada pessoa se manteve relativamente constante ao longo do tempo, demonstrando que pessoas saudáveis “estão cheias de vírus”.

Estudos como o realizado por Charles Langelier *et al*, 2018 demonstram a utilidade do NGS no manejo clínico (36). Foram sequenciados e analisados aspirados traqueais de 92 pacientes entubados na UTI com insuficiência respiratória aguda, infecções que são de difícil identificação em razão das limitações dos testes microbiológicos atualmente realizados (36).

Combinando dados de patógenos, microbiota comensal e transcriptoma do hospedeiro humano, foi possível identificar os agentes etiológicos (bacterianos e virais) das infecções em pacientes, cujos testes microbiológicos tradicionais foram negativos (36). A técnica de NGS também vêm sendo utilizada para identificação de agentes etiológicos em líquido cefalorraquidiano e tecido cerebral, buscando a elucidação das causas de doenças como meningite e encefalite e para identificação de casos de infecções oculares, incluindo crônica causada por rubéola intra-ocular (37).

Já muito caracterizados são os vírus de DNA oncogênicos, neste conjunto estão incluídos o *Herpesvirus* que está associado ao sarcoma de Kaposi, EBV, *poliomavírus humano* e o HPV. Esses agentes estão relacionados a 20% de todos os cânceres (38).

1.3 VÍRUS ASSOCIADOS AO CÂNCER DE PÊNIS: EBV E HPV

O HPV (*Human papiloma-virus*) é um vírus com grande potencial oncogênico, pertencente à família do *Papillomaviridae* e relacionado a vários tipos de carcinomas (39). Já existem hoje mais de 200 tipos de HPV identificados, que são organizados de acordo com seu risco oncogênico baixo ou alto e de acordo com os tipos de lesões que eles estão mais associados (39). O HPV de alto risco pode causar câncer nas mucosas onde infecta, como por exemplo no câncer de pênis, onde as células escamosas são infectadas. Dentre os vários tipos, os mais encontrados e de alto risco são os tipos 18 e 16, sendo este último o mais comum e considerado um carcinógeno tipo 1 para câncer de pênis (40). Neste sentido, estudos demonstraram que mais de 60% das neoplasias malignas de pênis estão associadas a infecções pelo HPV (41). O mecanismo de carcinogênese do HPV está relacionado principalmente a superexpressão das oncoproteínas virais E6 e E7, que induzem a desregulação do ciclo celular por meio da inativação de p53 e pRb, levando a proliferação celular descontrolada (42).

Dentre os vírus oncogênicos mais estudados, o EBV (*Epstein-Barr virus*) é um vírus da família *Herpesviridae* (43), cujo potencial oncogênico foi confirmado devido à sua associação com linfoma de Burkitt, carcinoma nasofaríngeal e alguns adenocarcinomas gástricos (44). Porém sua relação com o câncer de pênis ainda não é bem esclarecida, embora alguns estudos tenham sugerido que a expressão desajustadas de proteínas latentes de membrana LMP-1, LMP-2A e LMP-2B, atuando nos receptores de células B, expressas durante a latência de vírus podem levar ao desenvolvimento deste tipo tumoral (45,46). A infecção por EBV atinge principalmente os países mais pobres e está relacionado a aproximadamente 200 mil novos casos de câncer/ano no mundo (44). Alguns estudos têm buscado determinar a relação entre o

câncer de pênis e o EBV, como neste realizado no Brasil, que detectou a presença do DNA de EBV em 20/21 amostras de tumores de câncer peniano e foi identificada a proteína LMP-1 em duas amostras com tumores e positivas para EBV (47). Em outro estudo, os pesquisadores avaliaram 135 tecidos tumorais de pênis de homens brasileiros e encontraram 46,7% de prevalência para EBV, e 26,7% de co-infecção de HPV e EBV, sendo que a prevalência de HPV era de 60,7 %, e destes 29,7% apresentaram HPV 16 e 5,5% HPV 18 (48). Portanto, é possível que existam agentes etiológicos ainda não descritos, associados a esta neoplasia.

1.4 ANCESTRALIDADE GENÉTICA

A população brasileira é caracterizada predominantemente por três componentes genéticos de ancestralidade, europeia, africana e ameríndia. Para determinar a ancestralidade genética de indivíduos são utilizados Marcadores Informativos de Ancestralidade (AIMs) que exibem diferenças de frequências entre as populações, como os polimorfismo de nucleotídeos únicos (SNPs) que são diferenças em um único nucleotídeo, ou inserções-deleções (INDEls) que são diferenças em dois ou mais nucleotídeos que podem estar inseridos ou ausentes no DNA (49,50). Em uma meta-análise de 2015 sobre a população brasileira, os autores observaram que a ancestralidade genética europeia (0,62) é a mais abundante na nossa população, seguida pela africana (0,21) e ameríndia (0,17) (51). Entretanto esta proporção pode variar significativamente dependendo da região geográfica (51). Em 2011, Pena e colaboradores identificaram os componentes da ancestralidade molecular através de polimorfismos de DNA de brasileiros das quatro regiões mais populosas. Foi observado que a ancestralidade europeia foi a mais abundante em todas as regiões brasileiras, enquanto que a africana foi mais predominante nas regiões Nordeste e Sudeste. Já o componente ameríndio, como esperado, foi maior no Norte (52) (**Figura 3**).

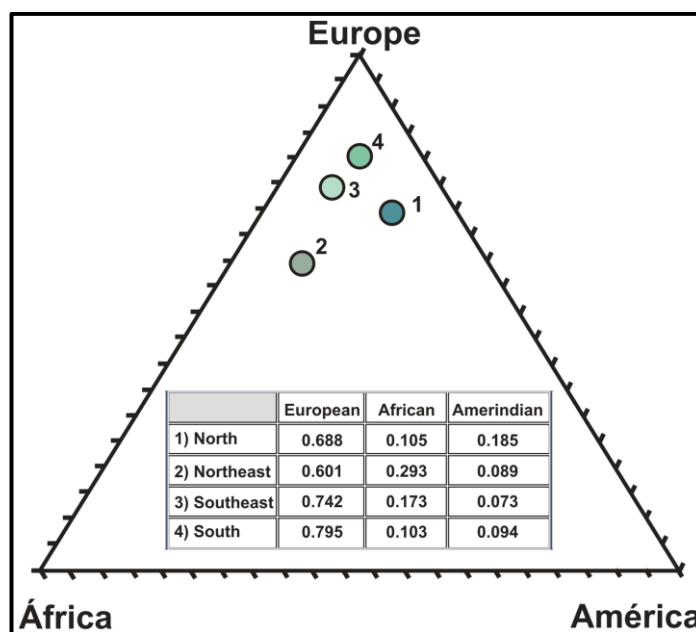


Figura 3 - Proporção de população por regiões incluídas no estudo (Brasil). **Legenda:** 1- Norte 2- Nordeste 3- Sudeste 4- Sul. Índice de componente genético identificado nos pacientes em relação a região geográfica. **Fonte:** Figura retirada de Pena et al, 2011 (52)

O Instituto Brasileiro de Geografia e Estatística (IBGE) classifica as etnias em cinco categorias: branca, parda, negra, amarela e indígena. Em 2008 foi observado que 190 milhões de pessoas no Brasil se autodeclararam em sua maioria como brancos (48,4%), pardos (43,8%), ou negros (6,8%) (52). Um estudo realizado em 2016 por Slopnick e colaboradores, avaliou a disparidade racial no tratamento e diagnóstico de 12.090 homens com carcinoma de células escamosas de pênis utilizando o *Nacional Cancer Database* (NCDB) de 1998 a 2012, estratificando os dados por raça/etnia e dados clínicos, tais como os indivíduos que foram submetidos a cirurgia, tipo de cirurgia, estágio da doença, tratamento e sobrevida. Foram identificados 76,8% indivíduos brancos, 10,2% afro-americanos e 8,7% hispânicos. Eles concluíram que os hispânicos com câncer de pênis têm maior probabilidade de desenvolver uma doença de alto risco em relação aos indivíduos brancos, com sobrevida comparável. Enquanto que os afro-americanos têm uma menor probabilidade de realizar um tratamento cirúrgico, porém apresentam altas taxas de mortalidade (53).

A subestrutura étnica e os diversos ancestrais genéticos são fatores que podem contribuir para uma predisposição ao desenvolvimento do câncer, assim como para o prognóstico da doença, principalmente em populações miscigenadas, como a brasileira (54). Alguns estudos vem demonstrando essa relação como o realizado por Leal e colaboradores, que avaliou ancestralidade em 427 pacientes brasileiros com adenocarcinoma de pulmão e encontrou uma associação entre mutações nos genes EGFR e KRAS, com uma maior e menor proporção de

ancestralidade asiática, respectivamente (55). Um outro estudo de Silva e colaboradores avaliou 137 pacientes diagnosticadas com adenocarcinoma gástrico e 262 indivíduos sem câncer, e avaliou a influência da ancestralidade genética na susceptibilidade a estes tumores. Os resultados apontam que uma maior contribuição da ancestralidade europeia pode ser um elemento de proteção para o desenvolvimento do câncer gástrico na população Amazônica avaliada (54). Em muitos estudos relacionados ao câncer de próstata, por exemplo, estes tumores parecem afetar principalmente homens de ascendência africana, que possuem um risco maior de seu desenvolvimento e um quadro clínico mais agressivo (56,57).

Tais estudos demonstram a importância de discutir sobre a ancestralidade genética correlacionando com os componentes étnicos, fatores ambientais e condições clínicas para compreender melhor o câncer em populações específicas, reduzindo as disparidades ao tratamento e visando o desenvolvimento de uma medicina personalizada. Principalmente em relação ao Ca de pênis em que os pacientes acabam procurando atendimento médico tardio, quando o tumor já está muito desenvolvido, em grande parte são de países em desenvolvimento e com baixo nível socioeconômico. No entanto, ainda não há estudos de ancestralidade de pacientes com câncer de pênis e um dos objetivos de nosso estudo é exatamente avaliar se este componente pode influenciar no desenvolvimento e agressividade destes tumores.

2. OBJETIVOS

Objetivo geral

Caracterizar o viroma e bacterioma de amostras tumorais de carcinomas de pênis coletados no ACCCC em São Paulo e na FCECON no Amazonas.

Objetivos específicos

- Avaliar por NGS o perfil de partículas virais em amostras de câncer de pênis (sem infecção por HPV e EBV) de duas instituições brasileiras: A.C. Camargo Cancer Center em São Paulo - SP e Fundação Centro de Controle de Oncologia do Estado do Amazonas em Manaus – AM;
- Determinar o bacterioma de tumores de pênis, por sequenciamento de nova geração de *amplicons* do gene 16S rRNA;
- Comparar o perfil do viroma e bacterioma identificadas nas amostras dos estados, de São Paulo e Amazonas, correlacionando-as com dados anatomopatológicos e clínicos dos pacientes;
- Utilizar sequências *off-target* do genoma humano para determinar a ancestralidade molecular dos pacientes incluídos no estudo

3. PACIENTES E MATERIAIS E MÉTODOS

3.1 ASPECTOS ÉTICOS E POPULAÇÃO ESTUDADA

Este é um estudo retrospectivo que foi aprovado pelo Comitê de Ética em Pesquisa – CEP do ACCCC através do número 2689/19 (**Anexo 1**) e pelo Comitê de Ética da Fundação Centro de Controle de Oncologia do Estado do Amazonas pelo parecer 2.230.007, dentro do estudo intitulado através do estudo: “Análise metagenômica do viroma das neoplasias penianas” (**Anexo 2**).

Foram incluídas no presente estudo todas as 15 amostras de tecido tumoral congelado de pacientes com CID C60 (Neoplasia maligna do pênis) disponíveis no biobanco do ACCCC, e foram coletadas no hospital entre Junho de 2006 a Março de 2018. As amostras foram disponibilizadas seguindo as Diretrizes Nacionais para Biorrepositório e Biobanco de Material Biológico Humano com Finalidade de Pesquisa.

Adicionalmente, foram utilizadas neste estudo 40 amostras tumorais de indivíduos com câncer de pênis recrutados pela FCECON no período de Outubro de 2012 a Junho de 2018. Tais amostras foram obtidas através das colaboradoras Kátia Luz Torres Silva e Valquíria do Carmo Alves Martins.

Para o sequenciamento do viroma foram selecionadas amostras negativas para HPV e EBV ($23/55 = 42\%$), dado que o objetivo era identificar novos vírus potencialmente oncogênicos nestes tumores. Amostras positivas para estes vírus foram usadas no estudo do bacterioma e um número menor foi utilizado no estudo do viroma, como controle positivo. Para o bacterioma foram utilizadas 37 amostras da FCECON e 19 amostras do ACCCC de tecido congelado com câncer de pênis, além de 8 amostras de tecidos congelados de pênis de margem cirúrgica sem neoplasia (**Figura 4**).

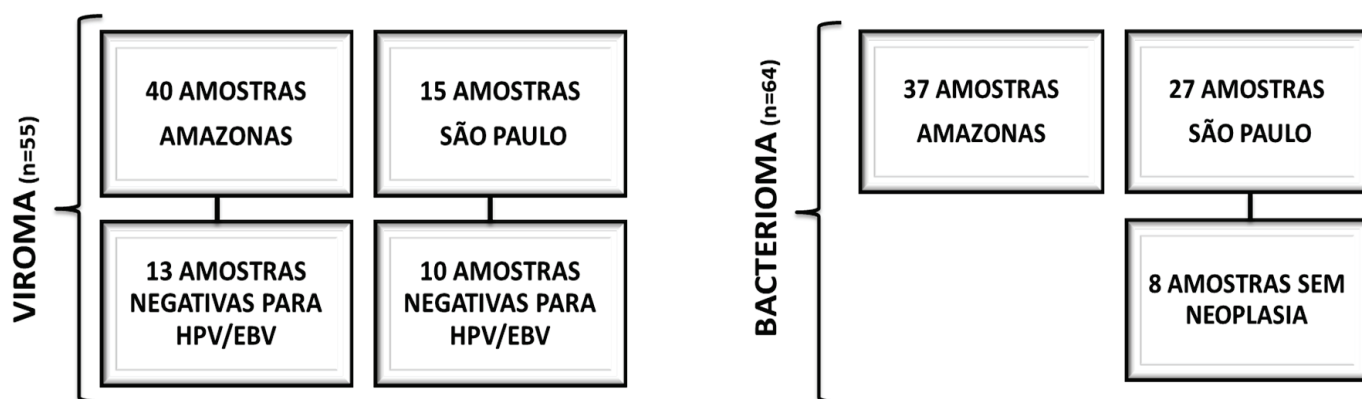


Figura 4 - Casuística incluída no estudo de acordo com a análise realizada (viroma ou bacterioma).

3.2 EXTRAÇÃO DE ÁCIDOS NUCLEICOS

A extração de DNA foi realizada a partir do tecido tumoral obtido na ressecção cirúrgica. Foi utilizado o equipamento Precellys (Bertin Technologies) para homogeneizar o material seguido de sua extração com o QIAamp DNA Mini Kit (Qiagen, EUA). Ao homogeneizado, foram adicionados 180µl de Buffer ATL e 20µl de proteinase- K. As amostras foram incubadas por 2h a 56°C. 200µl de Buffer AL e 200µl de etanol absoluto foram adicionados ao lisado. Na coluna QIAamp Mini foi adicionado 630µl da mistura, a qual foi centrifugada durante 1 min a 6.000g. Para purificação do DNA, foram adicionados 500µl de Buffer AW1 à coluna que foi centrifugada durante 1 min, adicionando depois 500µl do Buffer AW2, e centrifugando a coluna 3 min a 20.000g. Para eluição do DNA, foram adicionados 30µl de Buffer AVE, incubado por 1 min em temperatura ambiente e posteriormente centrifugado por 1 min a 6.000g, por duas vezes.

3.3 DETECÇÃO DE HPV E EBV

As amostras utilizadas no estudo do viroma foram avaliadas quanto à presença dos vírus HPV e EBV por PCR antes do seu sequenciamento. Para HPV, foi realizado PCR convencional utilizando um par de iniciadores (PGMY09/11) que se anela em regiões conservadas do genoma do vírus, permitindo a amplificação dos vários tipos de HPV (58). O mix da PCR continha: 12.5µl de *GoTaq® Green 2X Mastermix*, 10µl de 0.5pmol/µl de cada um dos iniciadores PGMY (59), 1.5µl água e 1µl DNA (58). A ciclagem no termociclador consistiu em uma desnaturação inicial de 95°C 5 min, seguida de 40 ciclos de 95°C 1 min, 55°C 1 min, 72°C 1

min, e uma extensão final de 72°C 10 min. A PCR foi avaliada por eletroforese em gel de agarose 1%, utilizando como controle positivo a linhagem de células Caski - DNA de carcinoma epidermoide cervical positivo para HPV 16. O tamanho de banda esperado para o produto de amplificação do HPV foi de 450 pb correspondente a região do gene L1.

A triagem do EBV foi realizada utilizando a técnica de PCR quantitativo em tempo real (qPCR) no qual foi preparado um mix contendo 10µl de *TaqMan™ Fast Universal PCR Master Mix*, 1µl de *EBVW Assay* (20X) (F: 5'-GCAGCCGCCCAGTCTCT-3' ; R: 5'-ACAGACAGTGCACAGGAGCCT-3'), 1µl de *ACTB3 Assay* (20X) (F: 5'-CCATCTACGAGGGGTATGC-3' ; R: 5'-GGTGAGGATCTTCATGAGGTA-3'), 7µl de água e 1µl de DNA. A reação é um multiplex com sonda de hibridização para o EBV marcada com fluoróforo FAM e sonda de hibridização para o controle humano β-actina marcada com o fluoróforo VIC. A qPCR foi realizada no aparelho de *real time StepOne Plus* (Thermo Fisher, EUA) com a ciclagem: 95°C 20 seg, seguida de 40 ciclos de 95°C 3 seg, 60°C 30 seg, e finalmente 60°C 30 seg (60).

3.4 ISOLAMENTO DE PARTÍCULAS VIRAIS

O protocolo de isolamento de partículas virais que foi utilizado neste projeto foi o descrito por Conceição-Neto e colaboradores (61). Esse protocolo é baseado no enriquecimento de partículas virais extracelulares por centrifugação, filtração e tratamento com nucleases antes da extração dos ácidos nucleicos (DNA e RNA) contidos dentro dos capsídeos virais (**Figura 5**). Nosso intuito ao utilizar esta metodologia foi identificar vírus que tivessem em infecção ativa e pudessem estar putativamente associados a neoplasia maligna de pênis.

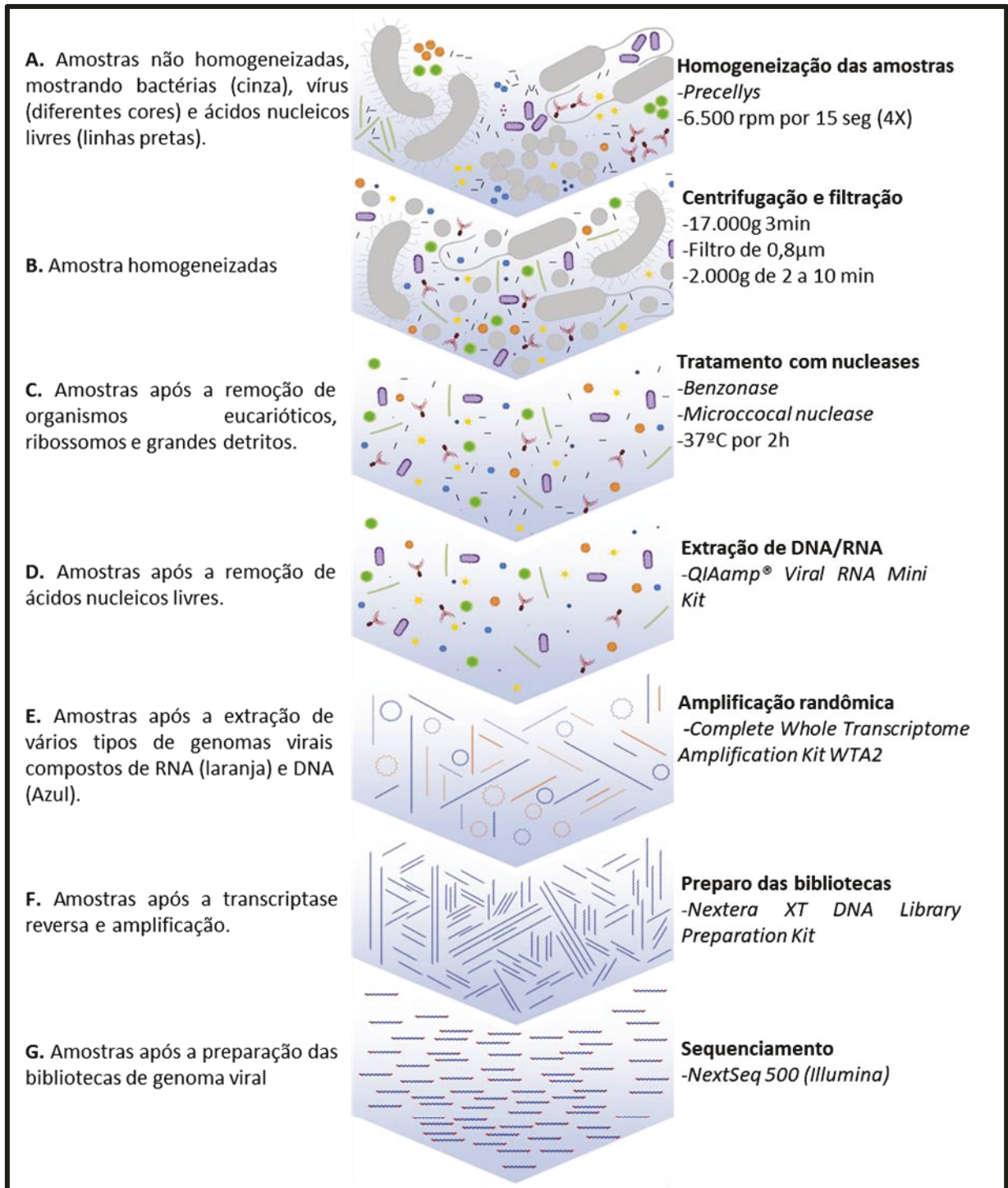


Figura 5 - Modelo de protocolo de isolamento de partículas virais. **Fonte** – Esquema adaptado do protocolo descrito por Conceição-Neto *et al*, 2015 (61).

3.5 HOMOGENEIZAÇÃO DO TECIDO TUMORAL

As amostras de tecido (massa entre 50-150mg) foram cortadas em nitrogênio líquido para remoção da resina e obtenção do tecido tumoral. A homogeneização foi realizada em tubos

contendo esferas cerâmicas, que permitem uma dissociação mecânica do tecido quando o tubo é agitado no equipamento *Precellys*. Foi adicionado 1mL de PBS filtrado aos tubos com as amostras, os quais foram agitados no *Precellys* a 6.500 rpm por 4 vezes durante 15 seg.

3.6 ENRIQUECIMENTO DE PARTÍCULAS VIRAIS

Após a homogeneização dos tecidos tumorais foi realizada uma centrifugação a 17.000g durante 3 min para retirar células humanas e partículas maiores. Todo o sobrenadante foi recuperado e 500µl filtrado a 2.000g de 2 a 10 min em filtro de 0,8µm, para eliminação de bactérias. Ao volume de 130µl do filtrado foi adicionado 8,4µl do Tampão 20X (Tris 1M, CaCl₂ 100mM e MgCl₂ 30mM), 1µl de microccocal nuclease (NEB) e 2µl de benzonase nuclease (Millipore) e incubado no thermomixer a 37°C por 2h. Este tratamento com nucleases foi realizado para a digestão de ácidos nucleicos livres (Fora das partículas virais). Para término da reação enzimática foi adicionado 7ul de EDTA 0,2M.

3.7 EXTRAÇÃO E QUANTIFICAÇÃO DE DNA/RNA

Para extração dos ácidos nucleicos (DNA/RNA) contidos dentro dos capsídeos virais foi utilizado *QIAamp® Viral RNA Mini Kit* (Qiagen, EUA). Foi inserido 560µl de *lysis Buffer* (AVL) utilizando o vortex para misturar por 15 seg e logo após incubado por 10 min a temperatura ambiente. 560µl de etanol absoluto foi adicionado nas amostras com nuclease e misturado novamente no vortex por 15 seg em seguida foi adicionado 630µl da mistura à coluna *QIAamp Mini* a qual foi centrifugada durante 1 min a 6.000g. Todo o tubo contendo o filtrado foi descartado e o procedimento foi repetido até que todo lisado tivesse sido conduzido na coluna. Para purificação das amostras foi adicionado 500µl de *Buffer AW1* e centrifugado a 6.000g por 1 min (sendo o tubo contendo o filtrado descartado e a coluna transportada para um novo tubo). Foram acrescentados 500µl do *Buffer AW2*, realizando uma centrifugação por 3 min a 20.000g e 14.000rpm (o processo de descarte do tubo foi repetido) e centrifugado por mais 1 min. A coluna *QIAamp Mini* foi transportada para um novo tubo de 1,5mL (filtrado descartado), onde adicionamos 2 vezes 30µl do *Buffer AVE* de modo que não se aproximasse muito da membrana do filtro, que foi incubada por 1 min em temperatura ambiente e posteriormente centrifugada por 1 min 6.000g e 8.000rpm. O DNA e RNA eluídos foram

quantificados por fluorimetria, com os kits *Qubit RNA HS Assay* e *Qubit dsDNA HS Assay* (Thermo Fisher, EUA).

3.8 AMPLIFICAÇÃO DO DNA/RNA E PREPARO DA BIBLIOTECA

A etapa de amplificação do DNA/RNA é fundamental para adquirirmos massa suficiente para fazer o sequenciamento. O *Complete Whole Transcriptome Amplification Kit WTA2* (Sigma-Aldrich, EUA), se baseia na transcrição reversa de RNA para cDNA através de oligonuclotídeos randômicos, contendo uma sequência universal que é utilizada subsequentemente para amplificação por PCR. A cada 2,82µl da amostra extraída (5-25ng de RNA) foi adicionada 0,5µl de *Library Synthesis Solution*, que foi então incubada no termociclador por 2 min a 95°C. Em seguida foi adicionado 0,5µl de *Library Synthesis Buffer*, 0,78µl de água isenta de RNase e 0,4µl de *Library Synthesis Enzyme*, seguido de incubação no termociclador (18°C por 10 min, 25°C por 10 min, 37°C por 30 min, 42°C por 10 min, 70°C por 20 min e resfriando até 4°C). A esses 5µl de biblioteca foi adicionado 7,5µl de *Amplification Mix*, 60,2µl de água, 1,58µl de *WTA dDNTTP mix*, 0,75µl de *Amplification enzyme*, seguido de incubação no termociclador por: 94°C por 2 min; seguido por 20 ciclos de 94°C por 30 seg; e 70°C por 5 min, e resfriando até 4°C.

O *QIAquick PCR Purification Kit* (Qiagen, EUA) foi utilizado para a purificação das amostras amplificadas. Aos 75µl de produto amplificado foi adicionado 375µl de *Buffer PB*, transferido para a coluna *QIAquick*, e centrifugado por 17.900g por 1 min. O eluato foi então descartado e a coluna lavada com 750µl de *Buffer PE*. O DNA foi eluído da coluna com 50µl de água. Com o fluorímetro *Qubit* (Thermo Fisher, EUA) o DNA foi quantificado.

Foi utilizado 3ng do DNA amplificado para a construção da biblioteca com o *Nextera XT DNA Library Preparation Kit* (Illumina, EUA). Foi preparado um mix contendo 5µl de *Tagment Buffer*, 2,5µl de *Amplicon Tagment Mix* e 2,5µl de DNA amplificado (diluído para 1,2ng/µl). Essa reação foi então incubada no *Hybex Microsample Incubator* (SciGene) a 55°C por 4 min para o processo de tagmentação. Foi adicionado 2,5µl de *Neutralize Tagment Buffer* para parar o processo. Na sequência, foram adicionados então os componentes: 2,5µl de iniciadores 1 (i7, N70X diferente para cada amostra) 2,5µl de iniciadores 2 (i5, S50X diferente para cada amostra) e 7,5µl de *Nextera PCR master mix*. Em seguida a reação foi incubada no termociclador a: 72°C por 3 min; 15 ciclos de 95°C por 30 seg; seguido por 15 ciclos de 95°C por 10 seg; 72°C por 15 seg e resfriando até 4°C.

As bibliotecas foram purificadas com *AMPure magnetic beads* (Thermo Fisher, EUA). Foi adicionado 15µl das esferas magnéticas a 25µl de produto de PCR, que foi incubado a temperatura ambiente por 5 min e posicionado no suporte magnético por 2 min para ligação das esferas ao ímã. O sobrenadante foi removido e as esferas lavadas utilizando 200µl de etanol 80%. O etanol foi removido, as *beads* secas e as bibliotecas eluídas das esferas utilizando 26,25µl de *Resuspension Buffer*. Após ligação das esferas ao ímã, 25µl do sobrenadante, contendo as bibliotecas purificadas, foi transferido para novos tubos. A concentração das bibliotecas foi determinada pelo *Qubit* (Thermo Fisher, EUA) e a distribuição de tamanhos dos fragmentos de DNA determinada pelo *Bioanalyzer* (Agilent, EUA). As bibliotecas foram quantificadas por qPCR com o *KAPA Library Quantification Kit* (Illumina, EUA).

Diluições de 1:500 e 1:5000 das bibliotecas foram preparadas para que ficassem dentro da curva padrão fornecida no Kit. Foi adicionado a 4µl de amostra diluída 6µl de uma mistura de 2X *KAPA SYBR FAST qPCR Master Mix*/10X Iniciadores *Premix/ROX High*. O qPCR foi realizado no equipamento *StepOne Plus* (Thermo Fisher, EUA) com o programa: 95°C 5 min, seguido por 35 ciclos de 95°C 30 seg e 60°C 45 seg. Cada biblioteca foi diluída para uma concentração final de 4nM, utilizando para tal a concentração obtida no qPCR e o tamanho médio obtido no *Bioanalyzer*. Uma mistura equimolar das bibliotecas foi então preparada para o sequenciamento.

3.9 SEQUENCIAMENTO DAS AMOSTRAS

O sequenciamento foi realizado na plataforma NextSeq 500 (Illumina, EUA) com leitura *pair-end* de 150bp, utilizando o *High-output flow cell*, com capacidade esperada de gerar até 400 milhões de sequências. Nosso objetivo foi obter aproximadamente 20 milhões de leituras por amostras.

3.10 ANÁLISE DE BIOINFORMÁTICA DO VIROMA

As análises dos viromas foram primeiramente feitas utilizando um *pipeline* montado pelo grupo de Bioinformática do Dr. Israel Tojal, do ACCCC. O programa metaSPAdes (59) foi utilizado para montagem “*de novo*” das sequências e geração de *contig* anotados com o programa Kraken2 (62), utilizando o banco de dados de nucleotídeos de vírus do NCBI (nt) (63). Em colaboração com o Dr. Wilson Araújo da Silva Jr e seu aluno de mestrado Rafael dos

Santos Bezerra, nossos dados foram também analisados com um *pipeline* alternativo utilizado pelo grupo do Departamento de Genética da Faculdade de Medicina de Ribeirão da USP. Primeiramente, os dados brutos do sequenciamento foram submetidos ao controle de qualidade utilizando o software FastQC (64). Após avaliar a qualidade das sequências foi realizada uma seleção daquelas com a melhor qualidade (Phred score >30), sendo que o corte para as sequências de baixa qualidade foi executado de duas maneiras distintas: pelo uso do Trimmomatic (65), para remover sequências de baixa qualidade, duplicatas e bases ambíguas; e uso do Cutadapt (66) para remoção de adaptadores Illumina. Para as análises subsequentes foram utilizadas apenas sequências com uma qualidade acima de 30 pontos. Uma segunda etapa de corte de sequências inadequadas foi feita usando o software AfterQC (67), que faz a remoção de caudas poli-X e bases ambíguas foi realizada. Para fazer a classificação taxonômica das sequências, foi utilizado o programa Kraken 2 (68). A montagem “*de novo*” foi realizada usando o programa SPAdes (69) e finalmente, a fim de identificar a similaridade de nucleotídeos e aminoácidos, realizou-se respectivamente o BLASTn e BLASTx com o uso do Diamond (70). O Diamond utiliza uma estratégia baseada na tradução das bases nucleotídicas para proteínas. Desta forma ele minimiza o esforço computacional e contrapõe essas proteínas contra um banco de dados específico, o que inclusive acelera muito o processo de análise. No caso utilizamos o banco de dados completo de proteínas disponível no NCBI (63). Por fim, foi utilizada uma opção de virtualização de máquina com o Docker (71). Mantendo assim a estabilidade do *script* e dos programas bases utilizados.

3.11 ANCESTRALIDADE GENÉTICA

A ancestralidade genética foi feita a partir das sequências *off-target* provenientes do sequenciamento dos viromas de todos os pacientes. Estas sequências mapearam no genoma humano e faziam parte de um conjunto total de 120 mil SNPs indicativos de ancestralidade, obtidos a partir de nove populações-referência (72). Esta abordagem já havia sido realizada com sucesso no nosso laboratório recuperando em média mais de 400 marcadores informativos de ancestralidade, numa coorte de câncer gástrico com sequenciamento de exomas ou painéis gênicos (73). Foram utilizadas as boas práticas do GATK (74), e realizada a chamada de variantes utilizando o método *HaplotypeCaller* (75). Por se tratar de regiões *off target*, uma vez que o objetivo principal do sequenciamento era identificação de vírus, optamos por fazer a

inferência por amostra, filtrando os SNPs presentes em bancos de dados e que tinham uma cobertura mínima de 3 *reads*.

3.12 EXTRAÇÃO DE DNA DAS AMOSTRAS SELECIONADAS PARA O BACTERIOMA

Avaliamos as amostras selecionadas para o bacterioma independente do status de HPV e EBV. Todas as amostras congeladas foram cortadas em nitrogênio líquido, para obtenção de 25 a 30mg de tecido. Primeiramente foi adicionado 1mL de SLX-Mlus *Buffer* do kit da *Omega E.Z.N.A.®Universal Pathogen* (Omega, EUA) em tubos com *beads* de vidro, para auxiliar na dissociação do tecido, que foi em seguida centrifugado a 2000g por 15 sec. Foram adicionados às amostras 40µl de proteinase- K e 72µl de DS *Buffer*, que foram incubadas no banho maria por 3h a 60°C. Logo após, o lisado foi centrifugado a 12.000g por 5 min e todo o sobrenadante foi recuperado. 600µl de RBB *Buffer* foram adicionados ao sobrenadante e mantido em temperatura ambiente por 5 min. Na coluna *MicroElute LE DNA* foi transferido 700µl da mistura, a qual foi centrifugada a 14.000rpm durante 1 min. Após o descarte do filtrado foram adicionados 500µl de HBC *Buffer* à coluna, que foi então centrifugada por 30 seg a 14.000rpm. 700µl de DNA *Wash Buffer* foram adicionados as novas colunas, que foram centrifugadas, e em seguida para remoção de qualquer traço de etanol, foi realizada mais uma centrifugação de 14.000rpm por 30 seg. Para eluição do DNA, foram adicionados 50µl de *Elution Buffer*, seguido de incubação por 2 min em temperatura ambiente e posteriormente centrifugada por 1 min a 14.000rpm. Com o mesmo volume foi realizada uma segunda eluição nas colunas. O DNA foi quantificado por fluorimetria com o kit *Qubit dsDNA HS Assay Kit* (Thermo Fisher, EUA).

3.13 GERAÇÃO DOS AMPLICONS DAS REGIÕES V3-V4 DO GENE rRNA 16S

Foram usados oligonucleotídeos para a amplificação das regiões hiper-variáveis V3 e V4 do gene rRNA 16S. Para gerar os *amplicons* uma reação de PCR convencional foi realizada padronizando 50ng como massa de DNA a ser amplificado. Foi preparado um mix contendo 17,5µl da enzima *Kapa 2G Robust Hotstart Ready Mix*, 3,5µl de 5µM de cada um dos iniciadores

U341F (5'-
ACTCTTTCCCTACACGACGCTCTTCCGATCT**CCTACGGG**RS**GCAGCAG**-3') e 806R
 (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**GGACTACH**VGGG
TWTCTAAT-3') (as sequências dos adaptadores Illumina estão sublinhadas e as sequências complementares ao gene bacteriano em negrito); e 10,5µl de DNA diluído em água para completar o volume final de 35µl. A ciclagem foi realizada no termociclador por: 95°C 2 min, seguida de 32 ciclos de 95°C 20 sec, 55°C 15 sec, e extensão final de 72°C 5 min, resfriando até 4°C. Os *amplicons* foram analisados por uma eletroforese em gel de agarose 1% corado com *SYBR Safe* (Thermo Fisher, EUA), utilizando 4µl de produto de PCR amplificado e como controle positivo o DNA de *E.coli*. As bandas foram visualizadas em luz ultravioleta (UV) conforme o observado na **Figura 6**.



Figura 6 -Análise de amplificação das regiões V3 e V4 do rRNA 16S em gel de agarose 1% de amostras de pacientes do Amazonas e de São Paulo. **Legenda:** 1- 27A 2- 30A 3- 37A 4- 41A 5- 43A 6- 46A 7- PA45T 8- PA54T 9- PA56T 10- PA60T 11- PA65N 12- PA55N 13- 52A 14- 63A 15- Controle positivo (*E.coli*) 16 – Controle negativo (sem DNA).

Em seguida, os *amplicons* gerados foram purificados com *AMPure magnetic beads XP* (Thermo Fisher, EUA), através da adição de 90µl das esferas magnéticas a 50µl de volume contendo água mais produto de PCR. Os tubos foram incubados por 5min a temperatura ambiente e em seguida foram colocados na raque magnética por 3min. Todo o sobrenadante foi descartado e 200µl de etanol 70% foi adicionado para lavar as *beads*. Logo em seguida as amostras foram incubadas em temperatura ambiente por 30 sec. O etanol foi removido e uma segunda lavagem com etanol 70% foi realizada. Após a remoção de todo etanol, as *beads* ficaram na raque magnética por aproximadamente 5min a temperatura ambiente para secarem. O DNA foi eluído das *beads* em 15µl de água. A concentração dos *amplicons* foi determinada por leitura no Qubit (Thermo Fisher, EUA).

As amostras seguiram para a construção de bibliotecas e sequenciamento em um serviço terceirizado prestado pela empresa *Neopropecta Microbiome Technologies* (Florianópolis,

Santa Catarina -Brasil). A construção das bibliotecas foi feita em triplicata, utilizando 5ng dos *amplicons* purificados por cada reação, mais enzima *Taq Platinum* (Invitrogen, EUA) e *barcodes* específicos para cada amostras. As triplicatas foram combinadas e purificadas com *Ampure XP Beads* e em seguida foi realizado o *pool* de bibliotecas utilizando uma mistura equimolar de cada amostra (76). As bibliotecas foram quantificadas por qPCR com o *KAPA Library Quantification Kit* (Kapa Biosystems) e o sequenciamento foi realizado na plataforma MiSeq com o kit MiSeq Reagent v2 (500 ciclos) na plataforma MiSeq (Illumina), visando gerar ≥ 50.000 *reads* por amostra.

3.14 ANÁLISE DE BIOINFORMÁTICA DO BACTERIOMA

A análise das sequências foram realizadas com um *pipeline* desenvolvido em parceria com o grupo de Bioinformática do ACCCC, e utilizou o programa Qiime2 (*Quantitative Insights Into Microbial Ecology*) para o processamento dos dados brutos (77,78). As sequências obtidas foram *clusterizadas* com porcentagem de 97% de identidade, para então realização da classificação taxonômica das *Operational taxonomic Units* ou *Unidades Taxonômicas Operacionais* (OTUs) contra o banco de dados SILVA (79). As análises de alfa e beta diversidade foram realizadas utilizando o pacote *phyloseq* no programa *RStudios* (80,81). Na diversidade Alfa foram avaliados os seguintes índices: *Observed*, onde foi avaliada o número de OTUs presente nas amostras; *Shannon* que considera a abundância e homogeneidade da distribuição das OTUs encontradas e o índice de *Simpson*, que mede a probabilidade das OTUs pertencerem ao mesmo taxa (variando de 0 a 1, sendo 1 maior diversidade) (82). Para comparação dos grupos foi utilizado o teste não paramétrico de *Mann-Whitney* para amostras independentes e estabelecido o nível de significância com 5% ($p < 0,05$) (82). Para diversidade Beta avaliamos a similaridade ou diferença na composição de bactérias entre os grupos de amostras, utilizando os índices de *Bray-curtis*, *Unifrac unweighted* e *Unifrac weighted*. Com o índice de *Bray-curtis* é possível avaliar a abundância dos diferentes taxa entre as amostras, enquanto com *Unifrac unweighted*, avaliamos a composição do bacterioma de maneira quantitativa, levando em conta a distância filogenética entre as OTUs, ou seja, quanto mais distantes há menor similaridade entre os grupos. Por outro lado, o índice de *Unifrac weighted* baseia-se também na abundância dos taxa entre as amostras, dando informação quantitativa e qualitativa sobre a composição bacteriana das amostras (82). A análise estatística destas métricas foi realizada pelo teste não paramétrico ANOSIM (*analysis of similarities*), também, com o nível de significância de 5% ($p < 0,05$). O teste LefSe (*Linear discriminant analysis effect*

size) foi aplicado para avaliar gêneros mais enriquecidos e diferencialmente abundantes entre os grupos estudados. Esta ferramenta identifica características entre os grupos avaliados, considerando sua abundância relativa, acoplado a testes estatísticos padrão e outros testes que avaliam a consistência biológica e relevância do achado (83).

4. RESULTADOS E DISCUSSÃO

4.1 INFORMAÇÕES CLÍNICAS DOS PACIENTES INCLUÍDOS NO ESTUDO DO VIROMA

A coleta dos dados clínicos dos pacientes recrutados no ACCCC foi realizada através da consulta dos seus prontuários médicos, com o auxílio das enfermeiras de pesquisa Fernanda Pintor e Cássia da Silva. Após o início das análises foi observado que duas amostras de São Paulo eram de um mesmo paciente (PA54T/PA56T), devido a isto há avaliação dos dados clínicos de 23 pacientes das 24 amostras sequenciadas.

Dentre as amostras tumorais analisadas, 39% (n=9) são de pacientes do ACCCC e a maior parte 61% (n=14) da FCECON. A idade do diagnóstico dos indivíduos incluídos no estudo variou entre 20 e 85 anos, com mediana de 52 anos e média de 50,87 anos, sendo que a maior parte dos pacientes 52% (n=12) tinham idade acima de 50 anos. A maior parte dos pacientes 48% (n=11) se declararam pardos; 31% (n=7) brancos; 13% (n=3) negros e 4% (n=1) indígena. Eram casados 39% (n=9) dos pacientes e em 87% (n=20) dos casos o tipo de serviço de atendimento utilizado foi através do SUS (Sistema Único de Saúde) (**Tabela 2**).

Tabela 2 - Características gerais dos pacientes selecionados para avaliação do viroma

PARÂMETRO	Número	Porcentagem
Origem das amostras		
São Paulo	9	39
Amazonas	14	61
Idade do diagnóstico (média – 50,87)		
< 30	2	9
30 a 50	9	39
> 50	12	52
Etnia		
Branca	7	31
Parda	11	48
Indígena	1	4
Preta	3	13
Não responderam	1	4
Estado civil		
Casado	9	39
Solteiro	6	26
União estável	3	13
Viúvo	3	13
Não responderam	2	9
Tipo de serviço hospitalar		
Convênio	1	4
Particular	2	9
SUS	20	87

Observando os dados dos aspectos histopatológicos, há uma maior predominância dos tumores localizados em glândula e prepúcio 35% (n=8), sendo que 22% (n= 5) atingem somente a glândula. A maior parte dos pacientes 70% (n=16) apresenta grau histológico II, sendo que pacientes com grau I, III e IV atingem uma porcentagem igualitária dos pacientes, apenas 4% (n=1). Contudo, há 18% (n=4) dos laudos sem especificações do local de acometimento do tumor. Quanto ao tipo histológico, 74% dos casos (n=17) apresentaram carcinoma epidermoide, o tipo mais comum entre os cânceres de pênis. A maior parte dos casos 57% (n=13) não apresentavam metástase linfonodal. Em relação a invasão perineural, 43% (n=10) não apresentaram invasão perineural e 56% (n=13) não apresentavam invasão vasolinfática. Infelizmente não conseguimos obter em 35% (n=8) dos casos informações para invasão perineural e para invasão vasolinfática dos dados clínicos dos pacientes da FCECON (**Tabela 3**).

Tabela 3 - Aspectos histopatológicos dos pacientes que tiveram os viromas sequenciados

PARÂMETROS	Número	Porcentagem
Local do tumor		
Prepúcio	2	9
Glande	5	22
Glande e prepúcio	8	35
Corpo cavernoso e esponjoso	1	4
Outras especificações	4	17
Sem especificação	3	13
Grau histopatológico		
Grau I	1	4
Grau II	16	70
Grau III	1	4
Grau IV	1	4
Sem especificação	4	18
Tipo histológico		
Carcinoma epidermoide	17	74
Carcinoma verrucoso	1	4
Outros tipos histológicos	2	9
Sem especificação	3	13
Metástase linfonodal		
Presente	9	39
Ausente	13	57
Sem especificação	1	4
Invasão perineural		
Presente	5	22
Ausente	10	43
Sem especificação	8	35
Invasão vasolinfática		
Presente	2	9
Ausente	13	56
Sem especificação	8	35

Dentre os parâmetros clínicos dos pacientes das coortes do ACCCC e da FCECON, como ocorrência de doenças pré- neoplásicas, observamos que 13% (n=3) dos pacientes apresentam Eritroplasia de Queyrat, uma patologia que pode comprometer mucosas e semimucosas do pênis e predispor ao carcinoma epidermoide. Em relação à fimose, 52% dos pacientes (n=12) já apresentaram fimose ao diagnóstico. Apenas 13% (n=3) dos pacientes disseram fazer uso de tabaco, enquanto que a maior parte dos pacientes 52% (n=12) negaram fazer uso de tabaco (**Tabela 4**).

Tabela 4 - Parâmetros clínicos dos pacientes selecionados para avaliação do viroma

PARÂMETROS	Número	Porcentagem
Doenças pré-neoplásicas		
Presente	3	13
Ausente	20	87
Fimose		
Sim	12	52
Não	4	17
Sem especificação	7	31
Tabagismo		
Sim	3	13
Não	12	52
Ex-tabagista	6	26
Não responderam	2	9

4.2 INFORMAÇÕES CLÍNICAS DOS PACIENTES INCLUÍDOS NO ESTUDO DO BACTERIOMA

O bacterioma de 64 pacientes, sendo 27 casos de SP (ACCCC) e 37 do AM (FCECON), foram avaliados através do sequenciamento de *amplicons* das regiões V3 e V4 do gene rRNA 16S (**Figura 4**), sendo averiguado os dados clínicos de 63 pacientes. A idade dos pacientes variou de 20 a 88 anos, com média de 56 anos e mediana de 57 anos. Em relação a etnia das duas coortes de casos de São Paulo e do Amazonas: a maior parte 53% (n=33) se declarou pardo; 33% (n=21) brancos; 8% (n=5) negros e 3% (n=2) se declararam indígenas, ou não responderam. Em 46% (n=29) dos casos os pacientes eram casados e 86% (n=54) foram atendidos pelo SUS (Sistema Único de Saúde) (**Tabela 5**).

Tabela 5 - Características clínicas dos pacientes que tiveram o bacterioma avaliado

PARÂMETRO	Número	Porcentagem
Etnia		
Branca	21	33
Parda	33	53
Indígena	2	3
Preta	5	8
Não responderam	2	3
Estado civil		
Casado	29	46
Divorciado	2	3
Não responderam	2	3
Solteiro	15	24
União estável	6	10
Viúvo	9	14
Tipo de serviço hospitalar		
Convênio	5	8
Particular	3	5
SUS	54	86
Sem informação	1	1

Os aspectos histopatológicos mostram uma maior predominância dos tumores localizados na glândula em 41% dos casos (n=26), seguido de prepúcio com 6,5% (n=4), assim como corpo cavernoso e/ou esponjoso. A maioria dos pacientes 54% (n=34) apresentou grau histológico II, sendo que pacientes com grau I perfizeram 24% (n=15). Casos de graus III e IV somaram apenas 3% (n=2) para cada. Quanto ao tipo histológico 86% (n=54) apresentam carcinoma epidermoide e 2% (n=1) carcinomas verrucosos. 56% (n=35) não apresentavam metástase linfonodal, enquanto que 38% (n=24) apresentavam metástase. Em sua maioria os pacientes 62% (n=39) não apresentavam invasão perineural e nem invasão vasolinfática 64% (n=40). Infelizmente não conseguimos obter informações de 13 casos sobre invasão perineural e vasolinfática dos dados clínicos dos pacientes da FCECON e apenas para um caso do ACCCC (Tabela 6).

Tabela 6 - Aspectos histopatológicos dos pacientes que tiveram bacterioma tumoral avaliado

PARÂMETROS	Número	Porcentagem
Local do tumor		
Prepúcio	4	6,5
Glande	26	41
Glande e prepúcio	7	11
Corpo cavernoso e/ou esponjoso	4	6,5
Outras especificações	15	24
Sem especificação	7	11
Grau histopatológico		
Grau I	15	24
Grau II	34	54
Grau III	2	3
Grau IV	2	3
Sem especificação	10	16
Tipo histológico		
Carcinoma epidermoide	54	86
Carcinoma verrucoso	1	2
Outros tipos histológicos	4	6
Sem especificação	4	6
Metástase linfonodal		
Presente	24	38
Ausente	35	56
Sem especificação	4	6
Invasão perineural		
Presente	9	14
Ausente	39	62
Duvidosa	1	2
Sem especificação	14	22
Invasão vasolinfática		
Presente	9	14
Ausente	40	64
Sem especificação	14	22

Conforme os dados clínicos da **Tabela 7**, observamos que 21% (n=13) dos pacientes apresentavam doenças pré- neoplásicas, dentre elas: Eritroplasia de Queyrat, Balanite, Condiloma acuminado e Leucoplasias. Em relação a fimose, 51% (n=32) apresentaram esta condição e a maior parte dos pacientes 43% (n=27) negaram fazer uso de tabaco, porém 38% (n=24) são tabagistas.

Tabela 7 - Parâmetros clínicos dos pacientes que tiveram o bacterioma avaliado

PARÂMETROS	Número	Porcentagem
Doenças pré-neoplásicas		
Presente	13	21
Ausente	50	79
Fimose		
Sim	32	51
Não	14	22
Sem especificação	17	27
Tabagismo		
Sim	24	38
Não	27	43
Ex-tabagista	8	13
Não responderam	4	6

4.3 EXTRAÇÃO E TRIAGEM PARA HPV E EBV

Após a macrodissecção das amostras, cada tecido foi adicionado a um tubo com 1mL PBS filtrado, para lise e homogeneização no equipamento do *Precellys®*. Parte do homogeneizado seguiu para a extração de DNA e RNA com o protocolo para avaliação das partículas virais, e a outra parte foi extraído DNA para avaliação de HPV e EBV. Para posterior utilização das amostras negativas para HPV e EBV no protocolo de isolamento de partículas virais 1mL do sobrenadante foi transferido para um novo tubo e armazenado a -20°C. Foi realizada a avaliação da presença de DNA de HPV por PCR convencional nas 15 amostras derivadas do ACCCC. Conforme podemos observar na **Figura 7**, quatro delas foram positivas para HPV, além do controle positivo de DNA da *CaSki cell* (58).

Na triagem para EBV por qPCR, apenas uma amostra de um paciente do ACCCC foi positiva, sendo esta mesma amostra negativa para HPV (**Tabela 8**) (60). As 40 amostras da Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON) foram previamente analisadas para HPV e EBV com a mesma metodologia empregada nas amostras dos pacientes do ACCCC.

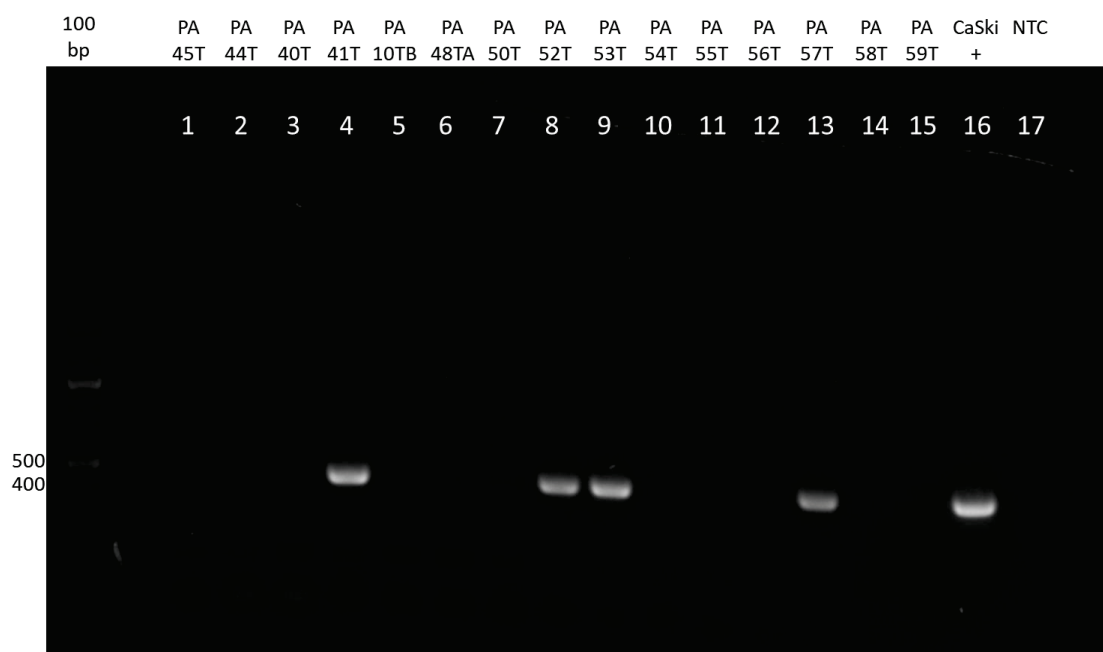


Figura 7 - Avaliação de HPV por PCR nas amostras de câncer de pênis dos pacientes do ACCCC. **Legenda:** Amostras do ACCCC 1-PA45T; 2-PA44T; 3-PA40T; 4-PA41T; 5-PA10TB; 6-PA48T; 7-PA50T; 8-PA52T; 9-PA53T; 10-PA54T; 11-PA55T; 12-PA56T; 13-PA57T; 14-PA58T; 15-PA59T; 16- controle positivo *CaSki cell* e 17- controle negativo da PCR.

Tabela 8 - Amostras positivas para HPV e/ou EBV dos pacientes de São Paulo (ACCCC)

AMOSTRAS	HPV	EBV
PA 41T	Positivo	Negativo
PA 52T	Positivo	Negativo
PA 53T	Positivo	Negativo
PA 57T	Positivo	Negativo
PA 59T	Negativo	Positivo

Seguiram para o sequenciamento do viroma no total 23 amostras negativas para ambos HPV/EBV, além da amostra controle positivo para HPV e EBV (24_39A) (**Tabela 9**).

Tabela 9 - Amostras selecionadas para avaliação do viroma.

INDEX	SAMPLE ID	ORIGEM
1	10_PA58T	SP
2	9_PA56T	SP
3	22_74A	AM
4	19_56A	AM
5	1_PA45T	SP
6	7_PA54T	SP
7	16_49A	AM
8	15_46A	AM
9	5_PA48TA	SP
10	2_PA44T	SP
11	12_32A	AM
12	11_30A	AM
13	8_PA55T	SP
14	4_PA10TB	SP
15	14_43A	AM
16	13_41A	AM
17	3_PA40T	SP
18	23_75A	AM
19	21_71A	AM
20	24_39A	AM
21	20_62A	AM
22	18_55A	AM
23	17_51A	AM
24	6_PA50T	SP

4.4 AVALIAÇÃO DOS VIROMAS TUMORAIS

Após a realização do protocolo escolhido para o enriquecimento de partículas virais (**Figura 8**), as amostras foram sequenciadas na plataforma *NextSeq* 500 (Illumina, EUA) e o sequenciamento gerou 115G de bases com 83% de alta qualidade (*Phred quality score* 30), uma média de 28 milhões de sequências por amostra e com um total de *reads* gerados de 674,921,012. Foram mapeados 8,180,640 (13%) sendo 7,619,616 (12,7%) mapeado *proper pair*, o que significa que ambos os *reads* foram mapeados na orientação e distância esperados. As análises de bioinformáticas foram feitas no ACCCC e foi observado que aproximadamente 99% das sequências mapeiam no genoma humano. Porém cinco amostras mostraram um enriquecimento mais expressivo para sequências não-humanas, conforme observado nas barras azuis (*unmpaped*) da **Figura 8**.

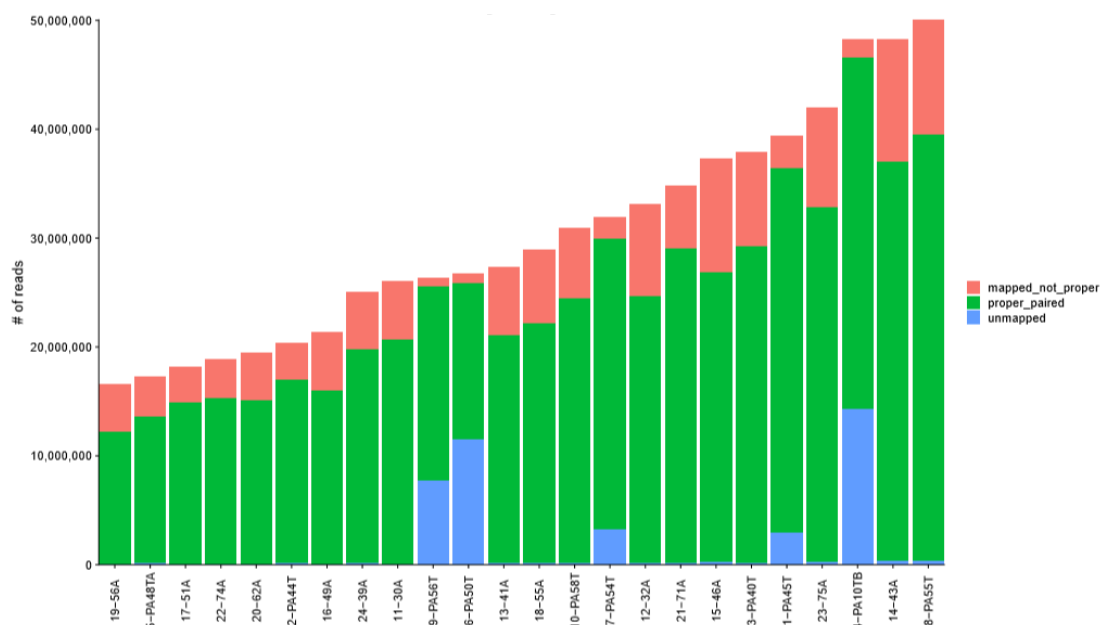


Figura 8 - Análise de *reads* dos viromas dos pacientes com câncer de pênis. Sequências alinhadas com hg19

Para estas cinco amostras, as sequências analisadas mapearam entre 442.598 a 2.432.037 com o genoma de bactérias (1,3 a 5,6%); 4.133.983 a 16.411.909 sequências de eucariotos (17,5 a 48,8%), e 30.955 a 285.717 (0,1 a 1%) com genomas virais (**Tabela 10**). Os bancos de genomas virais ainda são incompletos e muitas sequências de alta qualidade aqui obtidas (37,4-74,6%) permaneceram como não classificadas, mesmo após anotação com um dos mais abrangentes bancos do *National Center for Biotechnology Information* (NCBI).

Tabela 10 – Anotação das sequências do viroma

AMOSTRA	NÚMERO DE SEQUÊNCIAS (%)				
	NÃO CLASSIFICADO	BACTÉRIA/ARQUEA	EUCARIOTO	VÍRUS	TOTAL
7-PA54T	10.622.168 (37.4)	1.351.929 (4.8)	16.175.042 (56.9)	285.717 (1.0)	28.434.856
4-PA10TB	28.390.883 (65.9)	2.432.037 (5.6)	12.093.490 (28.1)	182.670 (0.4)	43.099.080
6-PA50T	17.660.181 (74.6)	1.791.428 (7.6)	4.133.983 (17.5)	101.738 (0.4)	23.687.330
9-PA56T	16.178.908 (69.8)	1.123.363 (4.8)	5.795.768 (25.0)	89.073 (0.4)	23.187.112
1-PA45T	16.777.398 (49.8)	442.598 (1.3)	16.411.909 (48.8)	30.955 (0.1)	33.662.860

Legenda: Anotação das sequências foi feita no banco de dados do <https://www.ncbi.nlm.nih.gov/nucleotide> (63)

De modo a aproveitar o máximo de informações geradas e visando melhor caracterização dos viromas destas amostras, buscamos colaboração com o grupo do Dr. Wilson Araújo da Silva Junior, especialmente o aluno de mestrado Rafael Bezerra (FMRP/ USP), que desenvolveu um *pipeline* específico para anotação de sequências virais (Páginas 25-26). Através destas análises nove amostras tiveram anotação de sequências virais, incluindo as cinco amostras já observadas nas análises anteriores realizadas pelo grupo de bioinformática do A.C. Camargo, sendo oito dessas amostras de pacientes do ACCCC e uma de um paciente da FCECON (**Tabela 11**).

Tabela 11 - Amostras que apresentaram anotação para sequências virais.

AMOSTRA	ANÁLISES	
	ACCCC	FMRP - USP
1-PA45T	X	X
2- PA44T		X
4-PA10TB	X	X
5-PA48T		X
6-PA50T	X	X
7-PA54T	X	X
8-PA55T		X
9-PA56T	X	X
15-46A		X

Legenda: Amostras de 1 a 9 derivadas do ACCCC/SP e amostra 15 (46A) do FCECON/AM.

Observamos uma maior quantidade de sequências derivadas de diferentes *Circovirus* em pelo menos sete amostras (PA45T- PA44T- PA10TB - PA50T-PA54T-PA55T-PA56T) (**Figura 9**). Estes vírus possuem DNA genômico de fita simples, circular, sem envelope e pertencem à família *Circoviridae*. São conhecidos por infectar porcos, pássaros, tendo sido previamente identificados em amostras de fezes humanas e de chipanzés e em tecido muscular de animais de fazenda, (84). Curiosamente, Liu *et al*, 2019 relatou que o vírus *Porcine circovirus 2* (PCV2), que é o patógeno principal de doenças causadas por *Circovirus* em porcos,

possui propriedades imunossupressoras, o que pode aumentar a susceptibilidade de infecções oportunistas. Desta forma, é possível que a co-infecção de PCV2 com outros patógenos aconteça em hospedeiros humanos. Essa informação abre um precedente interessante sobre a possibilidade de infecções putativas de *Circovirus* nos tecidos tumorais de carcinoma de pênis estarem associados a causa, ou piora da doença em si (85). Além disso, algumas sequências de *Parvovirus* e *Circovirus* foram identificadas em um estudo de doenças febris tropicais utilizando sequenciamento de nova geração, para caracterização de vírus presentes em soro de pacientes negativos para vírus da dengue (86). Também identificamos sequências de *Parvovirus* em mais de uma amostra (PA45T- PA10TB- PA50T- PA54T- PA56T) (**Figura 9**). Este vírus pertence à família *Parvoviridae* que infectam diferentes espécies de animais e estão relacionados a doenças como o eritema infeccioso (87). O *parvovirus NIH-CQV* foi descrito em amostras de pacientes com hepatite soronegativo (*non A e non E*) em 2013 por Xu *et al*, como um vírus híbrido de DNA linear pequeno, cuja forma circular do genoma foi denominada *Parvovirus NIH-CQV-co* (88). Em 2013, Smuts *et al* também descreveram a contaminação em Kits de coluna de sílica incluindo a do *QIAamp viral RNA Mini Kit* (Qiagen) pelo vírus *Parvovirus NIH-CQV* (89). Conforme descrito na extração do viroma, utilizamos esse kit (Páginas 22-23) para extração de ácidos nucleicos, o que nos alerta sobre o risco desta contaminação. Outros vírus amplamente relacionados a animais como *Bovine adenovirus* e *Porcine feces associated IAS virus like virus DNA*, também foram observados com quantidade menor de sequências em duas das amostras (PA45T-PA56T). Assim, de acordo com a **Figura 9**, foram identificados quatro tipos de *papilomavirus*: *unclassified Gammapapilomavirus*, *Human papilomavirus type 19*, *Human papilomavirus isolate ICB2*, e *Betapapilomavirus 4* (46A-PA54T-PA50T-PA45T). Também identificamos o *Merkel cell polyomavirus* (PA50T) um vírus de DNA de dupla fita, sem envelope, encontrado em tecido saudável, porém encontrado com alta carga viral em pacientes imunossuprimidos. Este vírus possui potencial oncogênico em aproximadamente 80% dos tumores de carcinoma de células de *Merkel* (90), um tumor agressivo e de ruim prognóstico, sendo pouco descrito no trato genital masculino (91,92).

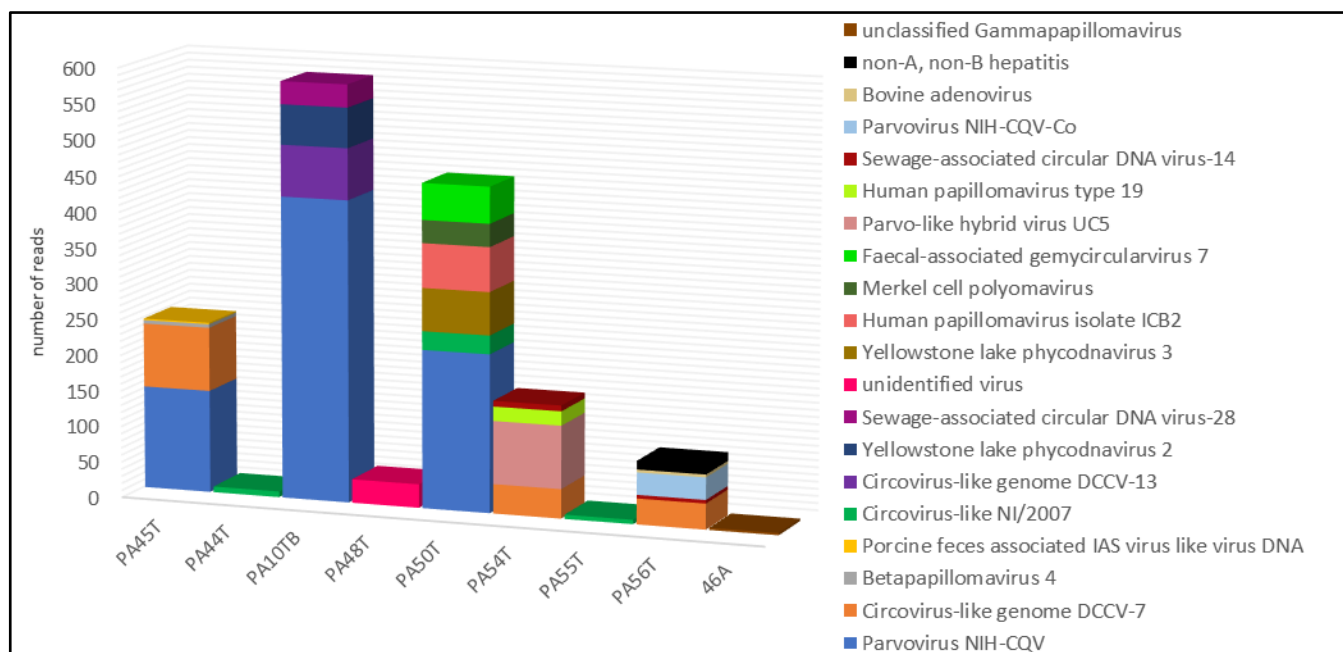


Figura 9 - Composição do viroma em nove amostras de tumores de pênis. **Legenda:** O eixo y representa o número de *reads* por amostra.

Alguns destes vírus identificados serão selecionados para posterior validação por qPCR, em outras amostras de carcinoma de pênis, principalmente os *Circovirus*, *Sewage-associated circular DNA vírus-14*, *Gemycircularvirus*, conforme pode ser observado na **Figura 9**, pois se mostraram mais frequentes nos tecidos tumorais de alguns pacientes avaliados neste estudo. Neste sentido, estamos aumentando nosso biobanco de amostras de câncer de pênis com colaboração de grupos do Ceará e Maranhão, e iremos avaliar também o bacterioma destes casos em um projeto PRONON já aprovado. Para algumas destas amostras faremos também *Shotgun sequencing*, como previsto no PRONON, o que nos permitirá avaliar não só partículas virais, mas também vírus inseridos no DNA humano. Além disso, iremos comparar os resultados do viroma das partículas virais descritos nessa tese, com os dados de *Shotgun Sequencing* que serão gerados das mesmas amostras.

4.5 ANÁLISE DE ANCESTRALIDADE MOLECULAR

Determinarmos a ancestralidade molecular de 24 pacientes das coortes de São Paulo e Amazonas, a partir de sequências *off-target* que mapearam no genoma humano, em regiões de marcadores informativos de ancestralidade (AIMs- “*Ancestry Informative Markers*”), em uma abordagem que classifica em nove populações referências “*NorthEast Asian, Mediterranean, South African, SouthWest Asian, Subsaharan African, Oceanian, SouthEast Asian, Northern*

European e Native American”(72). Foi utilizado o pacote de programas GATK (74), e realizada a chamada de variantes utilizando o método *HaplotypeCaller* (75). Por se tratar de regiões *off target*, uma vez que o objetivo principal do sequenciamento era identificação de vírus, optamos por fazer a inferência por amostra, filtrando os AIMs presentes em bancos de dados e que tinham uma cobertura mínima de 3 *reads*. Assim obtivemos uma média de aproximadamente 1500 AIMs, variando de 200 a 6100 por amostra. Conforme podemos observar na **Figura 10**, nas amostras do Amazonas (AM) o componente “*Native American*” foi bem mais abundante que nas amostras de São Paulo (SP), o que corrobora com os dados de etnia reportados pelos próprios pacientes. Verificamos que 65% dos pacientes do Amazonas (9 de 14) se declararam pardos, e o paciente que teve o resultado de 100% de componente étnico ameríndio foi também o único que se autodeclarou indígena (**Tabela 2**). Enquanto isso, nas amostras de São Paulo não houve predominância de nenhuma ancestralidade específica e 60% dos pacientes se auto declararam brancos.

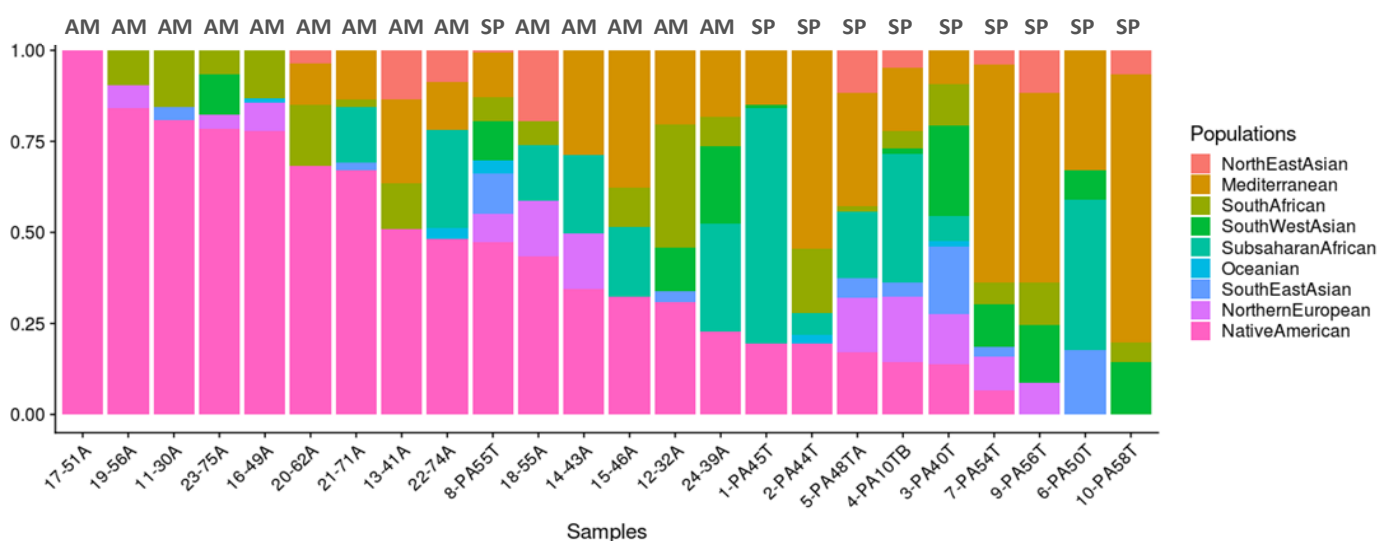


Figura 10- Ancestralidade molecular de 24 pacientes das coortes do ACCCC (SP) e FCECON (AM), que tiveram seus viromas estudados

4.6 CARACTERIZAÇÃO DOS BACTERIOMAS DAS AMOSTRAS DE CÂNCER DE PÊNIS

O sequenciamento das regiões V3-V4 do gene 16S rRNA gerou um total de 3.818.556 *reads*, com uma média de 59.665 *reads* por amostra, sendo que obtivemos um mínimo de 35 e um máximo de 95.736 *reads brutos*. Após processamento das *reads* brutas no Qiime2 e a

aplicação dos filtros de qualidade descritos na seção de materiais e métodos, realizamos curvas de rarefação buscando avaliar se a cobertura do sequenciamento foi suficiente para saturar toda a diversidade de bactérias presentes nas amostras avaliadas (**Figura 11**). Como podemos observar, a maior parte das amostras atingiu um platô, indicando que saturaram e toda a diversidade do bacterioma foi atingida com a quantidade de *reads* obtidas. As três amostras que tiveram poucos *reads* e não saturaram (**Figura 11B**) não foram incluídas nas análises subsequentes. Como observamos nos dados obtidos do viroma, as amostras de tecido de pênis tendem a apresentar grande quantidade de DNA humano, com pouca quantidade de bactérias anotadas, o que pode nos ajudar a explicar esse resultado. Com relação à quantidade de OTUs, observamos uma grande variação em seu número entre as amostras, com uma média de 91 e um mínimo e máximo de 17 e 177, respectivamente (**Figura 11**).

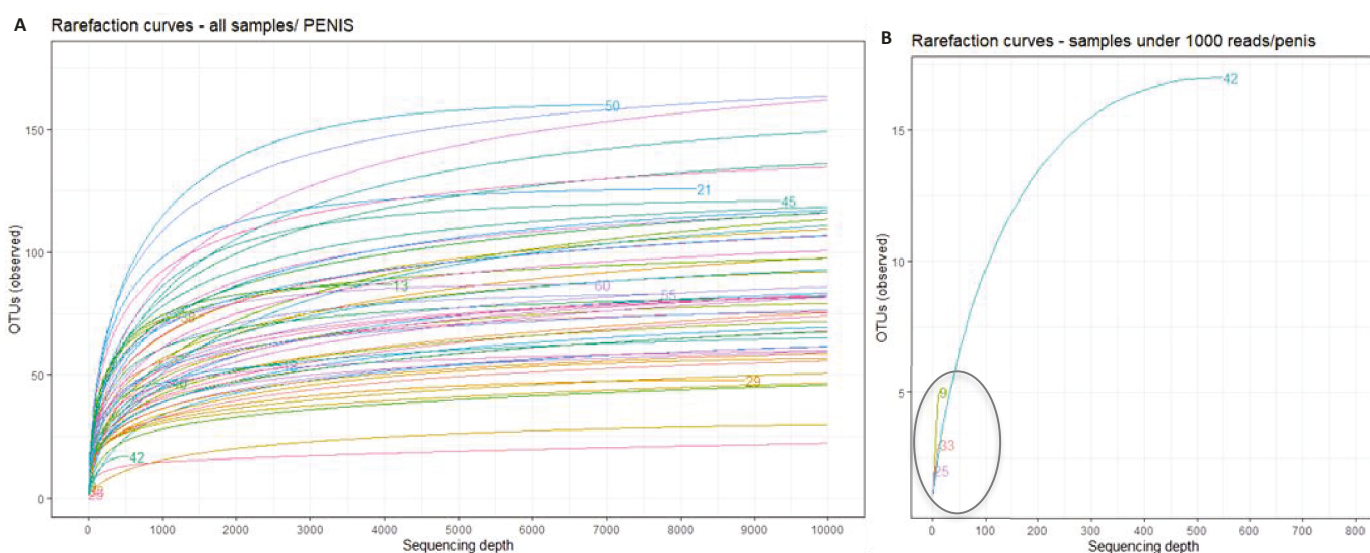


Figura 11 - Curvas de rarefação de *amplicons* do gene rRNA 16S derivado das amostras de câncer de pênis (OTUs e número de *reads*). **Legenda:** **A** - Curva de rarefação das 63 amostras de pênis, cada curva representa um paciente avaliado. **B** – Curva de rarefação indicando as 3 amostras que não saturaram (9-PA53T) (25-71A) (33-PA44N).

Para realização do processamento dos *reads* foram aplicados filtros que consideraram amostras com pelo menos 100 *reads*, e mínimo de 3 *reads* para ser considerado um OTU. Assim, 59 amostras seguiram para as análises subsequentes, com média de *reads* por amostra de aproximadamente 22 mil e mediana de 21.346 *reads*, sendo que a amostra com menor número teve 549 *reads*, e a amostra com o maior número 59.641 *reads*. Seis amostras de borda de lesão que foram incluídas no sequenciamento seguiram para as análises subsequentes, sendo

que quatro delas são correspondentes as amostras de tecido tumoral dos mesmos pacientes, que também foram sequenciados.

Um total de 1082 OTUs foram anotadas através do banco de dados SILVA (79), sendo que destas 6,1% foram classificadas como *unknown*, *uncultured* ou NA a nível de gênero. A classificação taxonômica das amostras a nível de família e gênero, está representada nas **Figuras 12 e 13**, respectivamente. Podemos observar em alguns indivíduos o predomínio de determinadas famílias e gêneros em suas amostras tumorais, como *Enterobacteriaceae*, *Corynebacteriaceae* e *Bacteroidaceae*, indicando uma baixa diversidade bacteriana nestas amostras. Quando observamos a composição da microbiota peniana encontrada nas amostras com e sem neoplasia, pudemos observar uma abundância maior de bactérias da família *Moraxellaceae* em 2 de 6 amostras de tecido normal, assim como do seu gênero *Actinetobacter* (2 de 6) (**Figuras 12 e 13**).

O *Actinetobacter* é um gênero de bactérias aeróbias, sendo frequentemente encontrado em tecidos de indivíduos saudáveis. Porém em alguns casos podem causar graves infecções incluindo no trato urinário quando presente em ambiente hospitalar, devido a sua capacidade de formação de biofilmes (93).

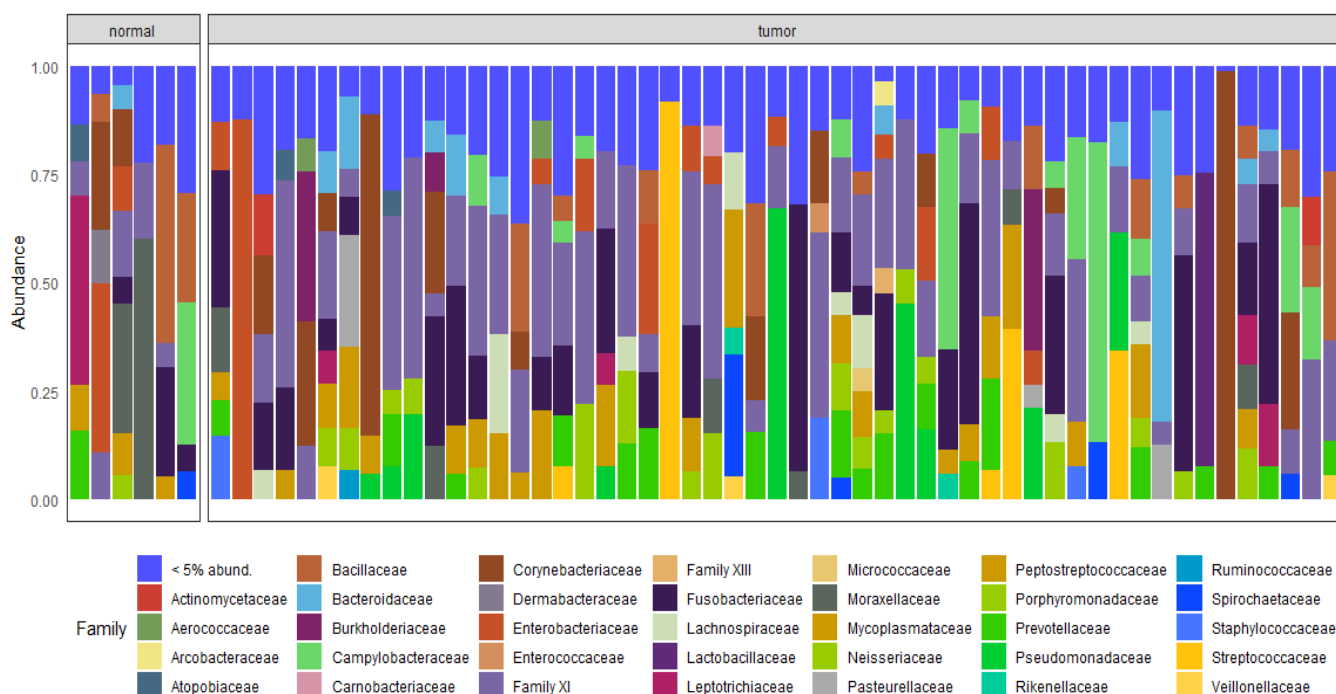


Figura 12 - Abundância relativa das principais famílias de bactérias identificadas em amostras de pênis com e sem neoplasia.

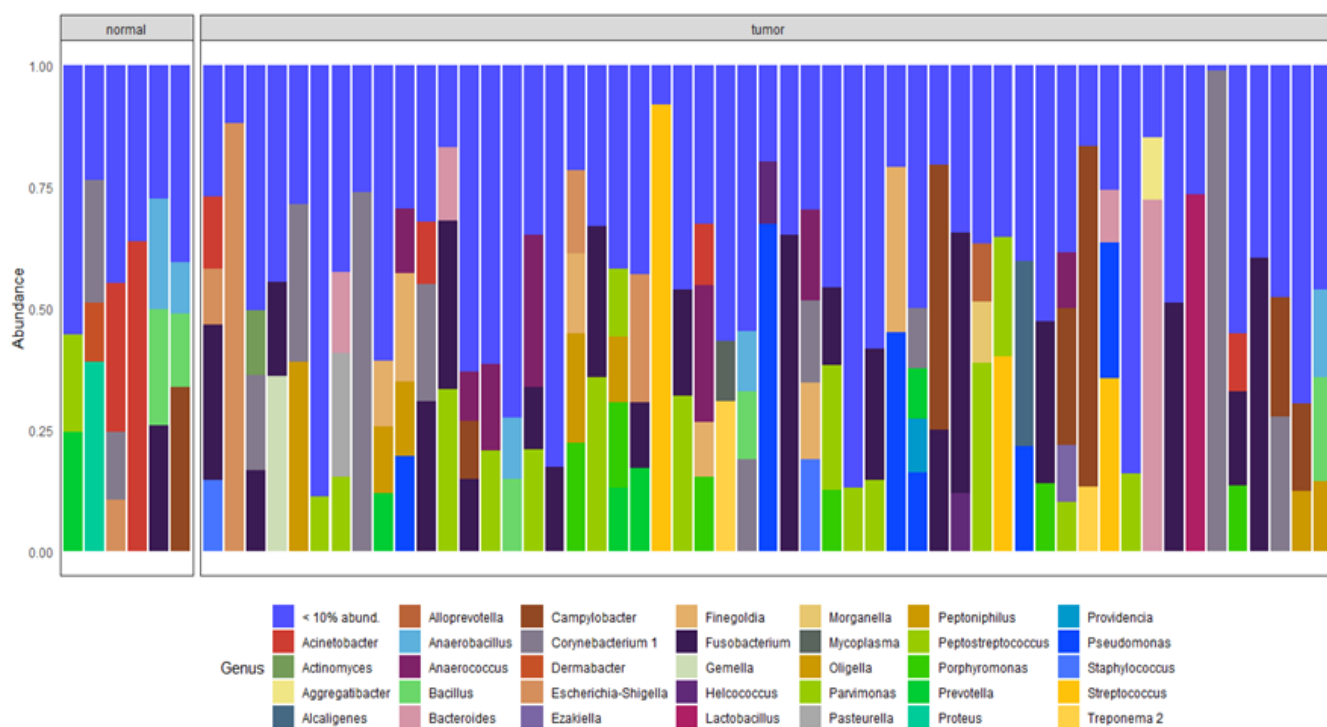


Figura 13 - Abundância relativa dos principais gêneros de bactérias identificados em amostras de pênis com e sem neoplasia.

Ao avaliarmos os gêneros com frequência maior ou igual a 1% nas 59 amostras de pênis (**Figura 14**), podemos observar que o gênero mais frequente foi o *Fusobacterium* (14,6%) seguido da *Corynebacterium* (6,3%), *Pseudomonas* (5,7%), *Preptostreptococcus* (5,5%) e *Bacterioides* (5,3%). *Fusobacterium* se trata de um gênero de bactérias anaeróbias, já descrito como membro da microbiota vaginal, oral e intestinal em disbiose, (94), além de estar fortemente associado a periodontite e a vaginose bacteriana (95). Um estudo realizado por Komiya *et al* 2018, sugere que *Fusobacterium* iniciaria a disbiose na cavidade oral, e posteriormente no colón. Em pacientes com câncer de pênis, talvez a disbiose possa ocorrer desta mesma maneira, porém isso teria que ser avaliado em estudos futuros da microbiota oral e intestinal em pacientes com câncer de pênis (96). Já o *Corynebacterium* é um gênero de bactérias aeróbias ou anaeróbias facultativas, que podem ser encontrados tanto em animais como em humanos, sendo que algumas espécies podem estar ocasionalmente envolvidas em infecções e outras fazem parte do microbioma normal. Seu papel no microbioma associado ao trato geniturinário ainda é pouco conhecido, no entanto foi relatado um aumento dessa bactéria após à circuncisão de pacientes (97).

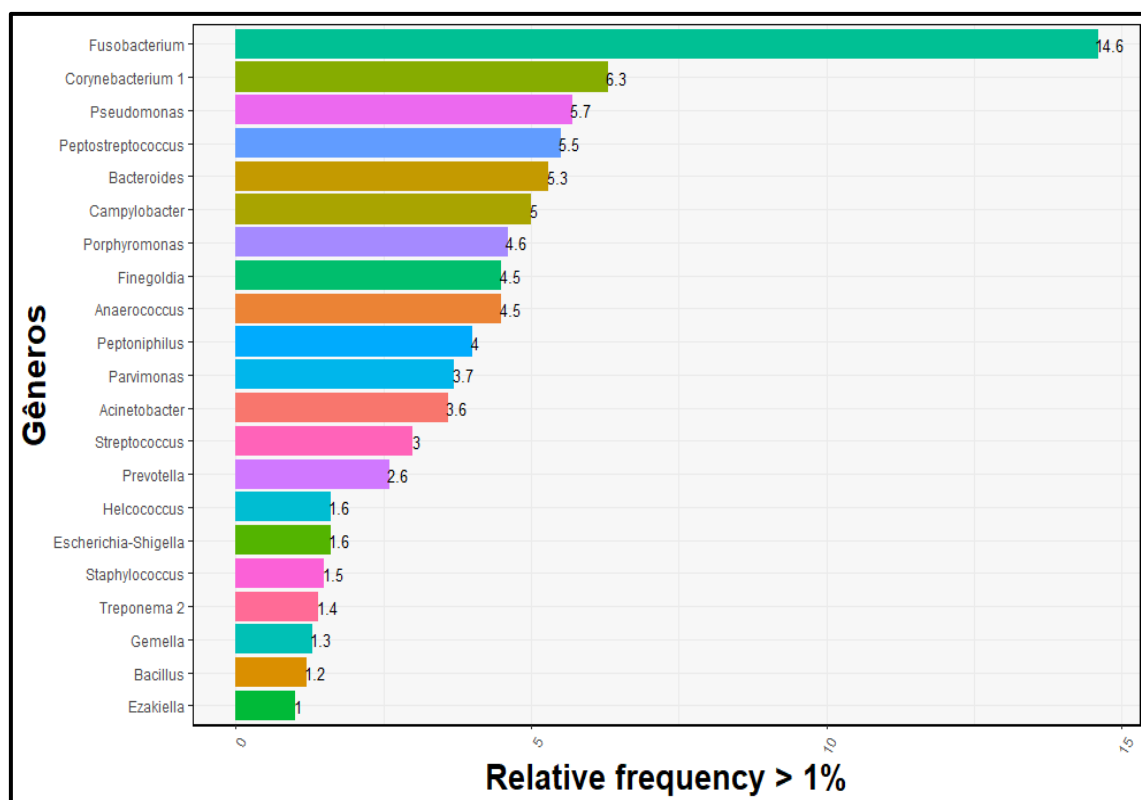


Figura 14 - Gêneros bacterianos com frequência relativa acima de 1% em amostras de câncer de pênis.

4.7 AVALIAÇÃO ENTRE OS BACTERIOMAS E PARÂMETROS CLÍNICOS-PATOLÓGICOS

4.7.1 Correlação entre os bacteriomas de São Paulo (ACCCC) e Amazonas (FCECON)

De posse dos dados gerados dos bacteriomas dos pacientes com neoplasia de pênis, buscamos correlacioná-los com alguns dados clínico-patológicos que nos pareceram mais relevantes.

Iniciamos comparando dados da composição dos gêneros de bactérias entre os grupos dos pacientes do Amazonas e de São Paulo, sendo que uma correlação positiva mediana foi observada, com r de 0,61 e p de 2.2×10^{-16} , entre os grupos (**Figura 1**). Desta maneira, podemos observar a presença de bactérias que são abundantes nas amostras do Amazonas e pouco frequentes ou ausentes nas de São Paulo, e vice-versa. A comparação da microbiota de ambos locais e sua associação com Ca de pênis se torna importante principalmente devido à grande diferença cultural e econômica entre as duas regiões.

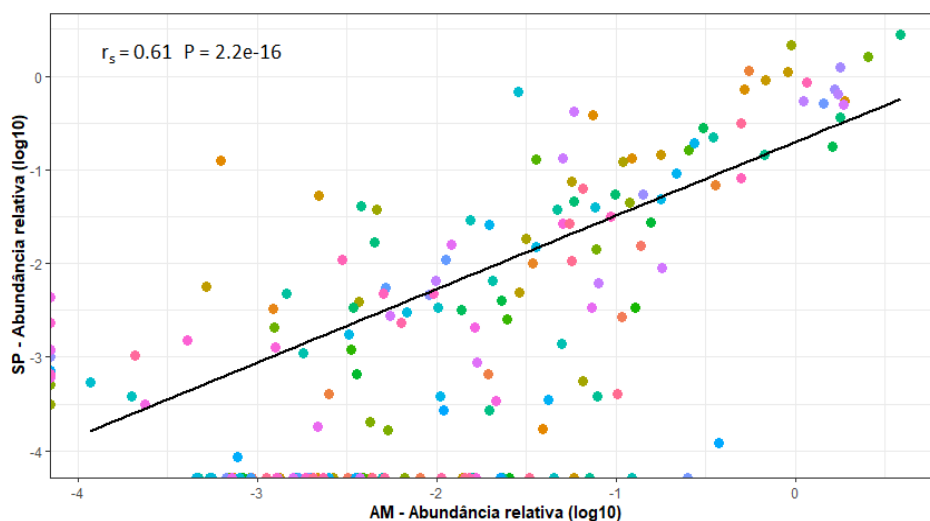


Figura 15 - Correlação da abundância relativa (log10) dos gêneros bacterianos entre as amostras do Amazonas e São Paulo. **Legenda:** Teste estatístico utilizado *Spearman* com diferença estatística significativa com $p < 0,05$ e r_s representando a correlação entre variáveis. Cada círculo colorido representa um gênero bacteriano.

Dos gêneros bacterianos com frequência superior a um por cento nas amostras de São Paulo, *Fusobacterium* se destaca com mais do dobro da frequência do que (18,6%) *Campylobacter* (9,5%) e *Bacteroides* (8,9%). Já no grupo de pacientes do Amazonas, os gêneros de bactérias *Fusobacterium* (12%), *Pseudomonas* (8,4%) e *Corynebacterium 1* (7,2%) foram os mais frequentes. Em ambos os estados o gênero *Fusobacterium* (SP:18.6%, versus AM:12%) foi o mais predominante (**Figura 16 e 17**). Em relação aos outros gêneros mais frequentes em São Paulo, o *Campylobacter* foi identificado como causador de gastroenterite em humanos, sendo comumente encontrada no trato intestinal e genital de animais e podendo levar a Campilobacteriose bovina (98,99). Quando observada a frequência nos dois grupos em conjunto os gêneros *Peptostreptococcus*, *Corynebacterium 1*, *Porphyromonas*, *Peptoniphilus* também se destacam. Essas bactérias são frequentemente encontradas na microbiota peniana, sendo que alguns estudos relatam haver um maior ou menor dessas bactérias associado a circuncisão e ao sítio atômico coletado (28).

Podemos observar que alguns gêneros com frequência superior a 1% aparecem em maior frequência apenas nas amostras dos pacientes de São Paulo (**Figura 16**), como *Campylobacter*, *Lactobacillus*, *Ezakiella*, *Aggregatibacter* e *Anaerobacillus*; e outros unicamente amostras de pacientes do Amazonas (**Figura 17**), como *Staphylococcus*, *Gamella* e *Actinomyces*.

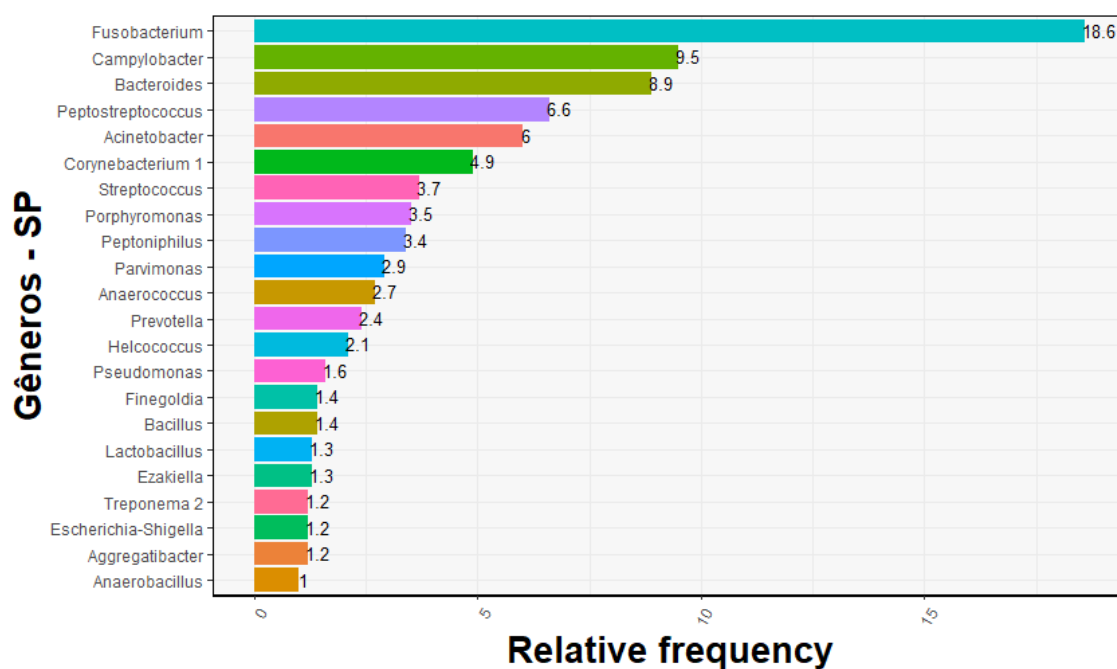


Figura 16 - Gêneros bacterianos com frequência relativa acima de 1% em amostras do estado de São Paulo.

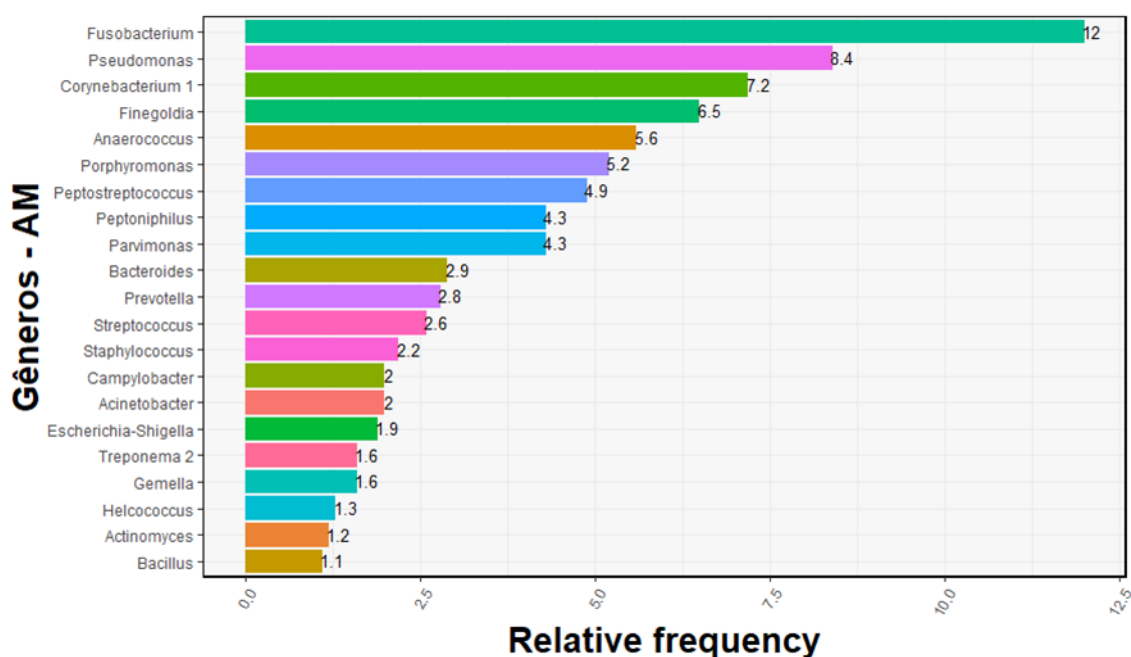


Figura 17 - Gêneros bacterianos com frequência relativa acima de 1% em amostras do estado do Amazonas.

Com intuito de comparar a diversidade do microbioma dentro de um grupo, ou na comparação entre os dois grupos, foram utilizados os índices de diversidade alfa e beta, respectivamente. Em relação ao local de origem das amostras, observamos uma diferença

estatisticamente significativa ($p < 0,05$) nas três métricas utilizadas para diversidade Alfa (**Figura 18**) em Amazonas e São Paulo. A diversidade alfa observada (*observed*) avalia o número de espécies presente em determinada amostra, enquanto que o índice *Shannon* considera a riqueza e homogeneidade da sua distribuição; e finalmente o índice Simpson mensura a diversidade, considerando o diferente número de OTUs presentes na amostra. É interessante observar como as amostras do Amazonas apresentam maior diversidade bacteriana, independente da métrica utilizada.

As três métricas avaliadas da diversidade Beta (**Figura 19**) também se mostraram estatisticamente significativa ($p < 0,05$), indicando diferenças quantitativas e qualitativas na composição bacteriana entre as duas coortes. Resumidamente, o índice *Bray-Curtis* avalia a dissimilaridade entre as duas populações avaliadas, com os valores da dissimilaridade variando entre 0 e 1, considerando 0 como populações idênticas e 1 como completamente diferentes. Já a métrica *Unweighted Unifrac* considera as relações filogenéticas entre as bactérias nas duas amostras avaliadas. Assim, duas bactérias do mesmo gênero são consideradas mais similares entre si, do que se fossem de gêneros distintos. E finalmente a métrica *Weighted Unifrac*, mede além da proximidade filogenética entre os grupos, a abundância dos gêneros avaliados.

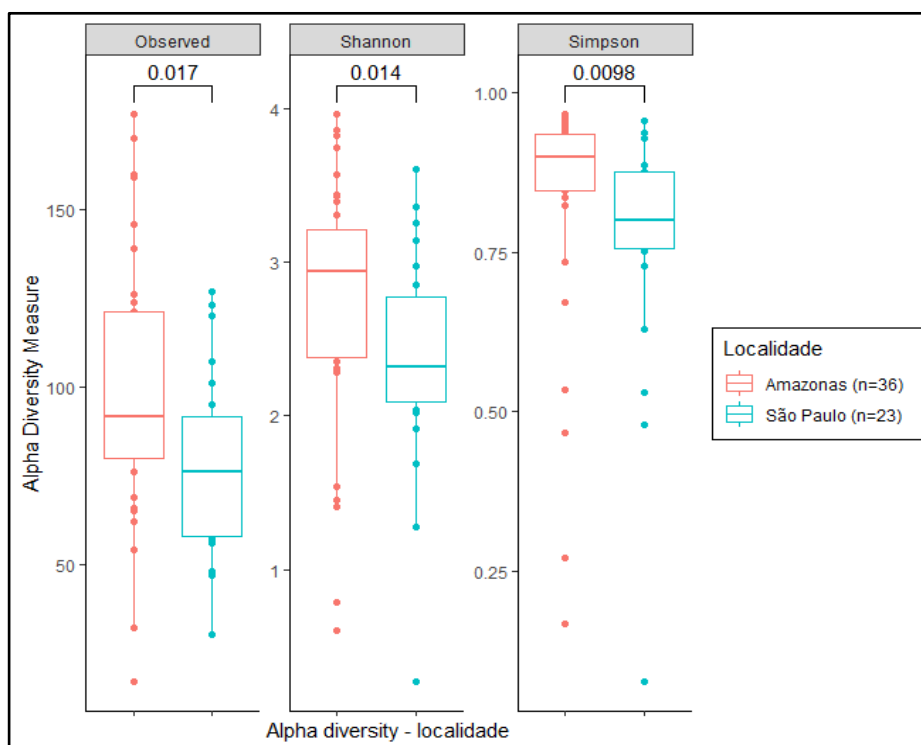


Figura 18 – Medidas de diversidades alfa do bacterioma de carcinomas de pênis em relação a origem das amostras de SP e AM. **Legenda:** Teste estatístico de alfa diversidade – Teste de Mann Whitney, $p < 0,05$.

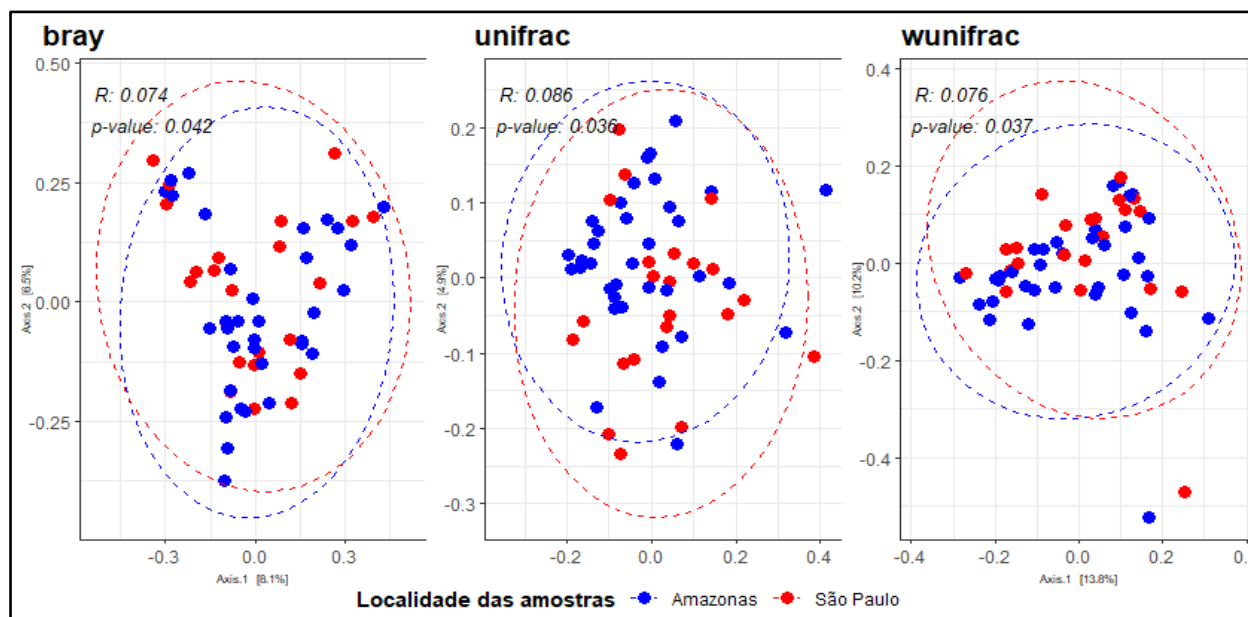


Figura 19 – Medidas de diversidade Beta do microbioma de carcinomas de pênis em relação a origem das amostras de SP e AM. **Legenda:** Teste estatístico de Beta diversidade– ANOSIM, $p < 0.05$ e R-coeficiente e correlação.

Utilizamos também a análise LefSe (*Linear discriminant analysis effect size*), aplicada para avaliar gêneros diferencialmente abundantes entre as duas coortes. Esta ferramenta identifica características entre os dois grupos avaliados, considerando sua abundância relativa, acoplado a testes estatísticos padrão e outros testes que avaliam a consistência biológica e relevância do achado. Os gêneros *Alcaligenes*, para as amostras de São Paulo, e *Pseudomonas* para as amostras do Amazonas, foram aqueles com maior significância estatística (LDA score ≥ 4) na comparação entre os dois grupos (**Figura 20**). Apesar de não haver trabalhos na literatura associando o gênero *Alcaligenes* a microbiota tumoral, existem relatos de pacientes imunossuprimidos que tiveram infecções oportunistas fatais por *Alcaligenes faecalis* (100). Já o gênero *Pseudomonas*, já foi associado a infecções bacterianas e vários estudos correlacionaram o mesmo a infecções no trato urinário e até mesmo estudos relacionados ao pênis (29).

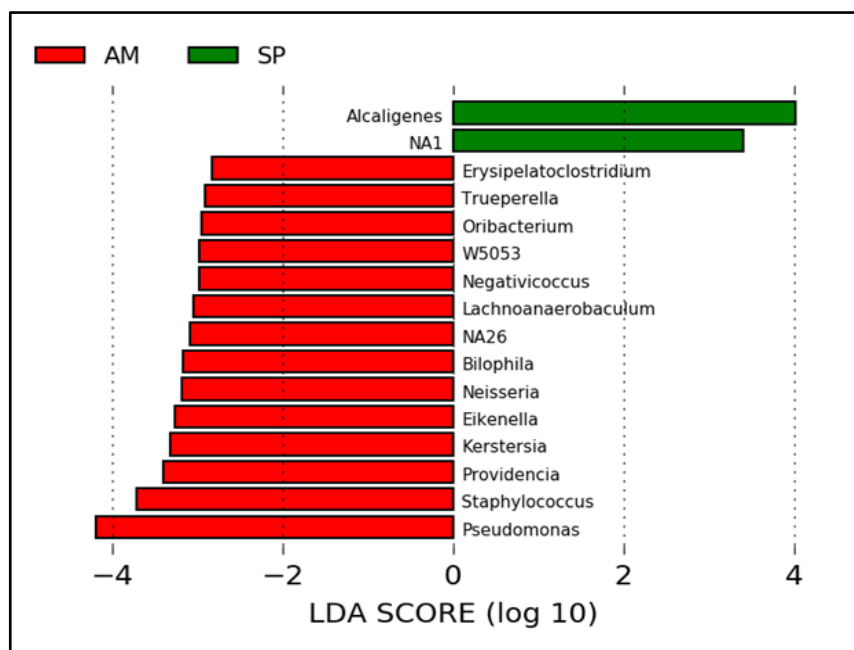


Figura 20 - Análise dos gêneros bacterianos diferencialmente abundantes (LefSe) entre as amostras tumorais do Amazonas e São Paulo. **Legenda:** LDA score ≥ 4 , $p \leq 0.05$.

O gênero *Alcaligenes* apresentou sua frequência relativa variando entre 0,01% a 1,6% nas amostras do Amazonas, estando presente em 15/36 (41,6%) destes pacientes, enquanto que para as amostras de São Paulo, *Alcaligenes* teve frequência de 0,01% a 36,9%, e foi encontrada em 3/23 (13%) destes indivíduos (**Figura 21**). Para o gênero *Pseudomonas*, a frequência variou entre 0,01% a 67% em 32/36 (88,9%) dos pacientes do Amazonas, tendo frequência relativa de 0,01% a 27,1% em 15/23 (62,2%) dos pacientes de São Paulo.

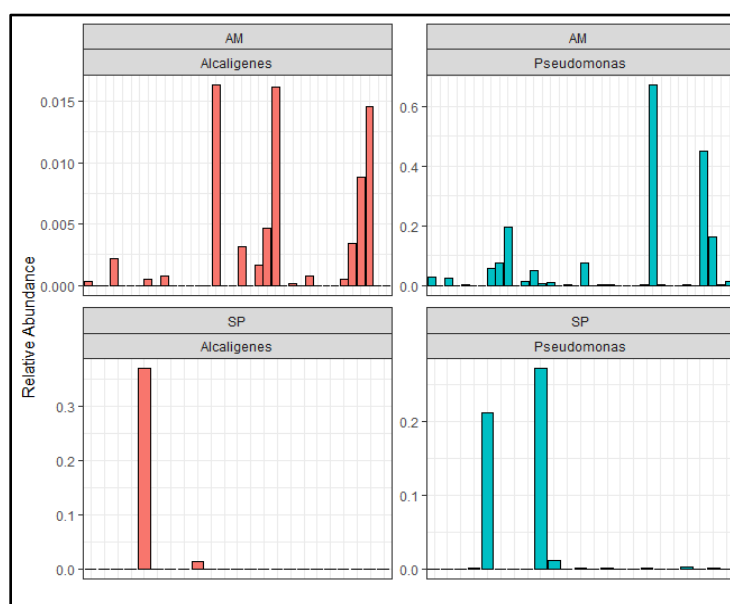


Figura 21 - Frequência relativa e distribuição dos gêneros *Alcaligenes* e *Pseudomonas* entre as amostras tumorais de São Paulo e Amazonas.

4.7.2 Análise do bacterioma entre tecido tumoral versus normal

As análises entre as amostras pareadas, com e sem neoplasia (4 pacientes), não mostraram diferenças estatísticas significativas ($p < 0,05$) nas diversidades alfa e beta. No entanto, o pequeno número de amostras avaliado pode ter influenciado este resultado (**Figura 22 e 23**). Embora o número amostral seja pequeno, observamos uma tendência de existir maior diversidade bacteriana nas amostras tumorais.

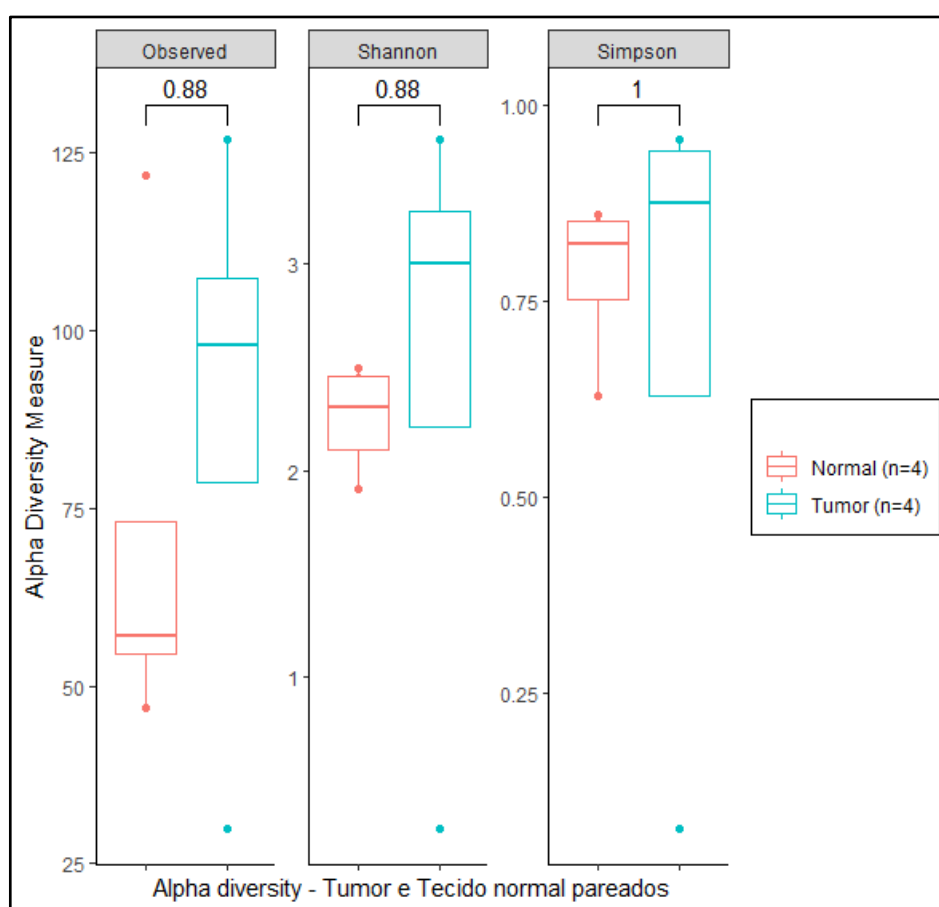


Figura 22 – Medidas de diversidade alfa entre as amostras pareadas de tecido tumoral e tecido sem neoplasia (normal). **Legenda:** Teste estatístico de alfa diversidade – Teste de Mann Whitney, $p < 0.05$.

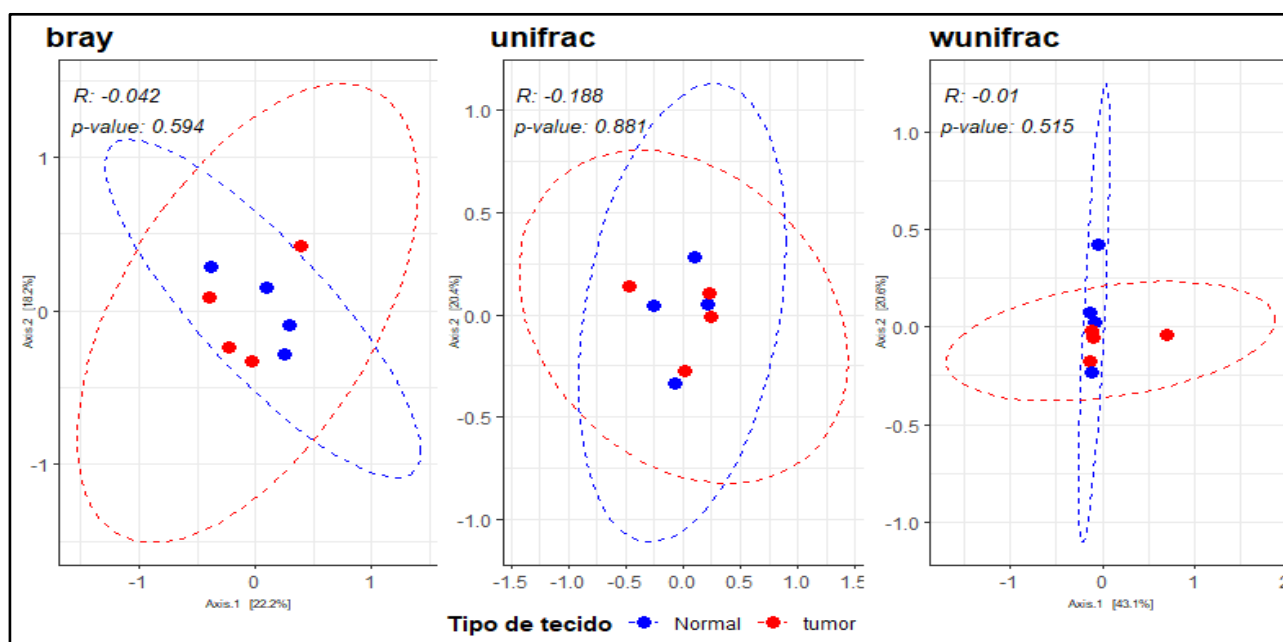


Figura 23 – Medidas diversidade Beta das amostras pareadas de tecido tumoral e tecido sem neoplasia (normal). **Legenda:** Teste estatístico de Beta diversidade– ANOSIM, $p < 0.05$.

A análise Lefse ($LDA \geq 4$) mostrou apenas o gênero *Lactobacillus* como diferencialmente abundante e mais frequente nos tumores, porém, com $LDA < 4$ (**Figura 24**). Geralmente bactérias deste gênero possuem papel anti-tumoral, o que confronta nossos resultados, porém a espécie *Lactobacillus iners* foi previamente associada com a bacteriose vaginal causada pelo microbioma peniano do parceiro (27). Ainda assim é importante reafirmar que o baixo número de amostras avaliadas pode influenciar neste resultado, e que será importante no futuro avaliar mais amostras normais e tumorais pareadas. Além disso, este gênero apareceu em baixíssima frequência ($\leq 0,4\%$) nas amostras tumorais, indicando que sua importância biológica deve ser pequena nesses casos.

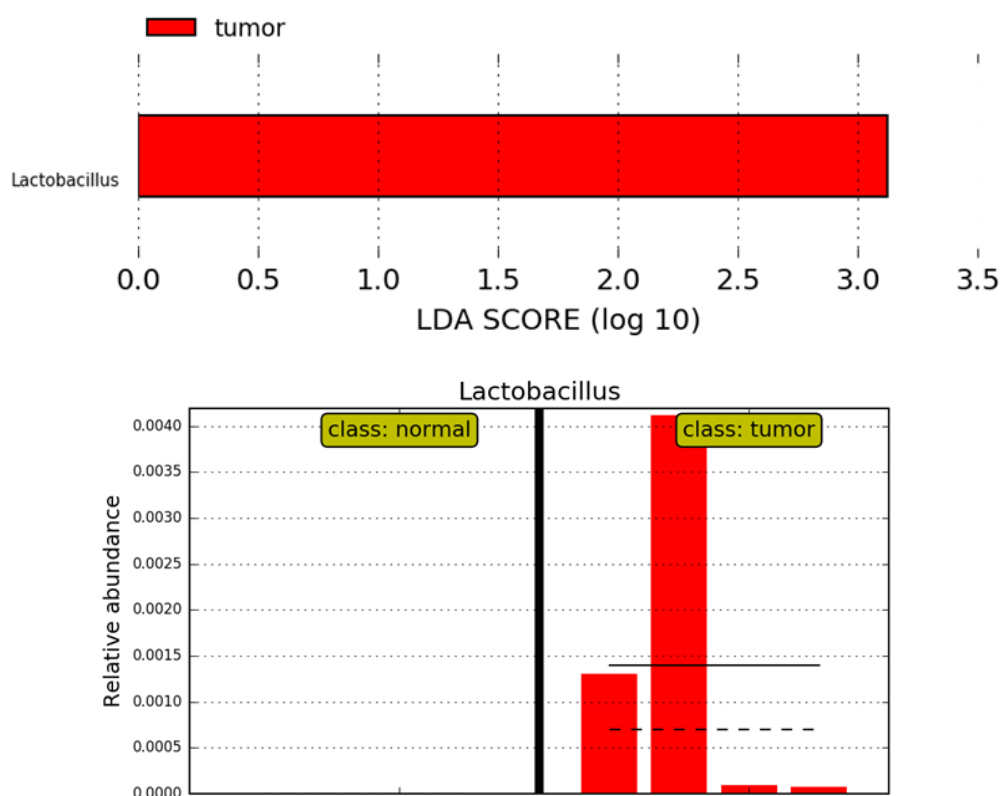


Figura 24 - Análise LefSe e abundância relativa (LefSe) do gênero *Lactobacillus* nas amostras de tecidos normais e tumorais. **Legenda:** LDA score ≥ 4 ; $p \leq 0.05$.

4.7.3 Avaliação do bacterioma entre amostras de carcinomas de pênis com e sem fimose

A fimose é um fator de alto risco para o desenvolvimento do câncer de pênis, devido a má higiene e inflamação crônica do prepúcio (5). Uma metanálise evidenciou que a circuncisão protege contra o desenvolvimento de carcinoma de pênis avançado (101).

A avaliação da diversidade alfa comparando os casos de câncer de pênis com e sem fimose mostram que há uma diferença estatística significativa, nas 3 métricas avaliadas. É possível observar uma maior riqueza e abundância de bactérias em pacientes que não tiveram fimose (**Figura 25**). Porém nenhuma das métricas das análises da diversidade beta mostraram uma composição distinta nos gêneros de bactérias entre os dois grupos (**Figura 26**). Normalmente os pacientes que apresentaram fimose realizam a cirurgia de postectomia, essa cirurgia pode ser realizada na infância quando é considerada um fator protetor contra o Ca de pênis, na idade adulta quando é mais comum, ou até mesmo durante a cirurgia oncológica. A circuncisão é um dos fatores mais que podem influenciar na composição da microbiota peniana e alguns estudos relatam que a riqueza de espécies da microbiota peniana antes da cirurgia de

postectomia é maior que após o paciente realizar a cirurgia, ocorrendo uma diminuição das bactérias anaeróbias e um aumento das bactérias anaeróbias facultativas (28,29).

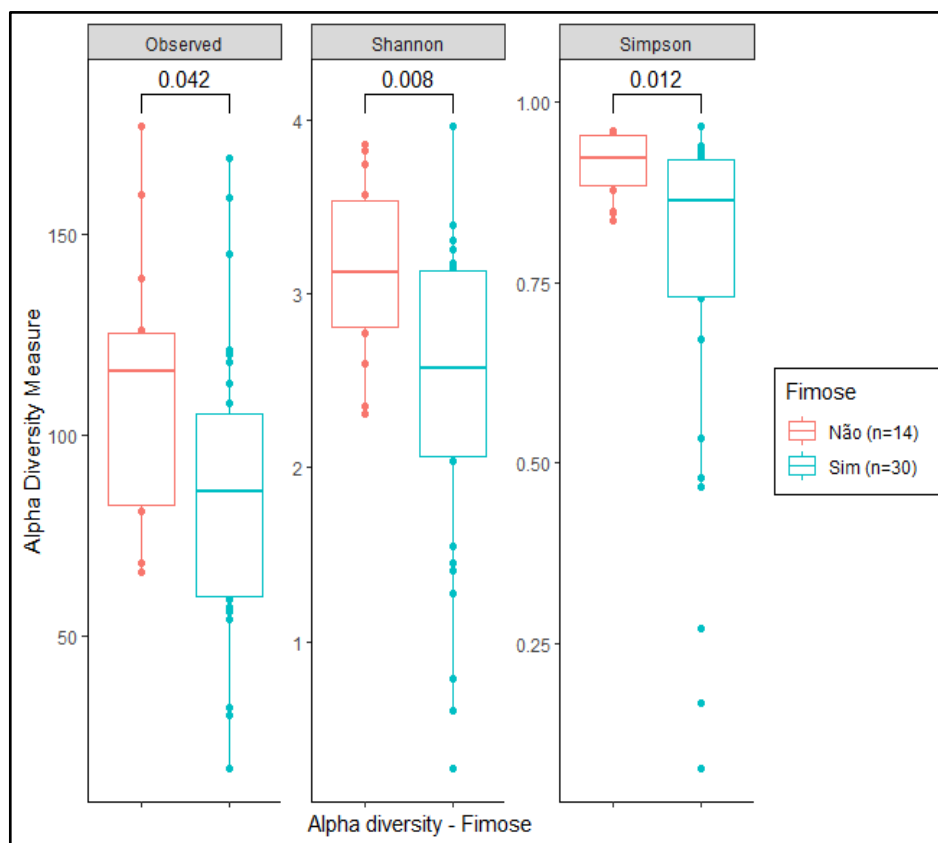


Figura 25 – Medidas de diversidades alfa em relação ao status de fimose. **Legenda:** Teste estatístico de alfa diversidade – Teste de Mann Whitney, $p < 0.05$

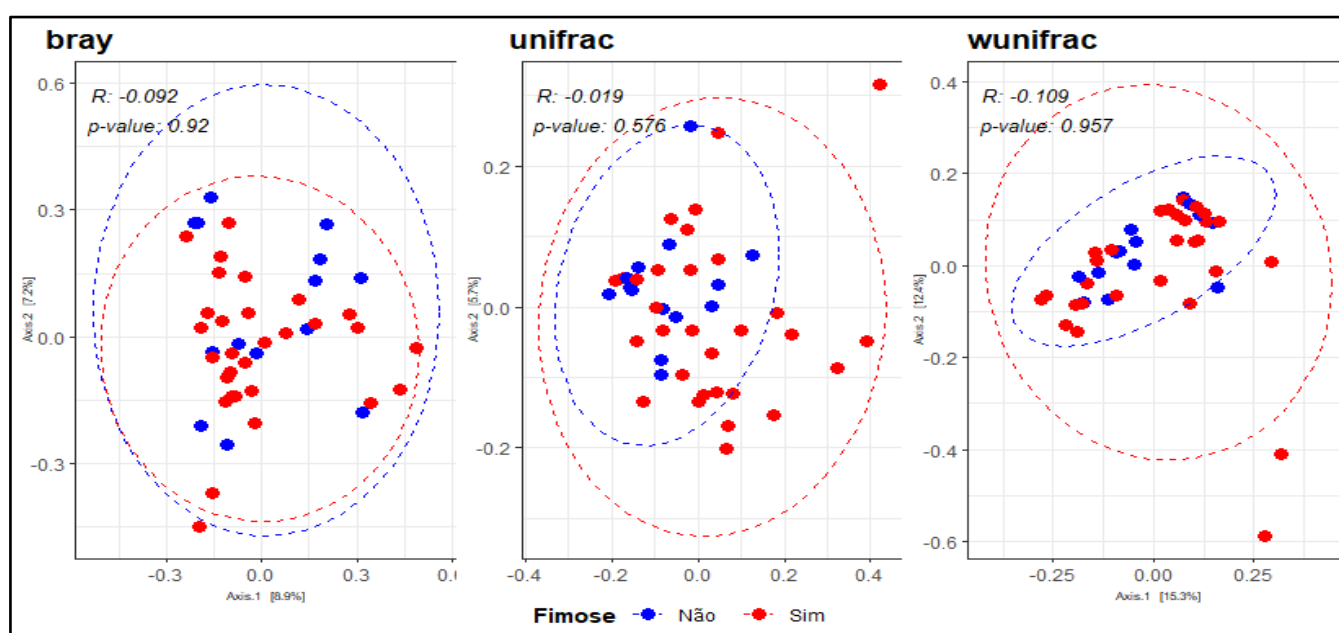


Figura 26 – Medidas de diversidades beta em relação ao status de fimose. **Legenda:** Teste estatístico de Beta diversidade – ANOSIM, $p < 0.05$

A análise Lefse identificou um enriquecimento de bactérias do gênero *Enterococcus* nos pacientes com fimose ($LDA \geq 4$) (Figura 22). São bactérias anaeróbias facultativas presentes nos intestinos de muitas espécies, atuando inclusive como um probiótico. No entanto, muitas delas são patogênicas, como *Enterococcus faecalis*, que está relacionado ao câncer colorectal. A disbiose do microbioma intestinal pode levar a translocação de *Enterococcus* para outros órgãos, ou até mesmo a bacteremia (102,103) (Figura 27).

Já nos pacientes sem fimose, o gênero *Oligella* ($LDA \geq 4$) se mostrou enriquecido (Figura 27). São bactérias aeróbias, que podem causar infecções raras. São mais frequentes em infecções urológicas, principalmente em pacientes imunocomprometidos e pode ser encontrada na urina de pacientes com cateteres urinários (104). Alguns outros gêneros como *Pasteurella* e *Mobiluncus* não chegaram a ter um LDA score ≥ 4 , mas possuem valores próximos ao *cut off* utilizado nesta análise, e serão futuramente avaliados em uma coorte maior de casos.

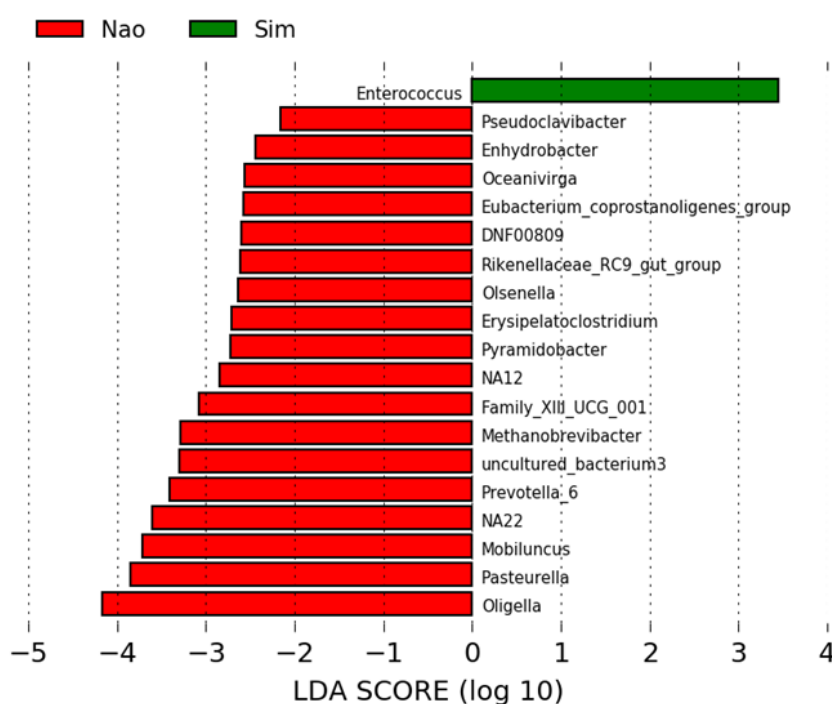


Figura 27 - Análise dos gêneros bacterianos diferencialmente abundantes (Lefse) entre os grupos de pacientes com câncer de pênis com e sem fimose. **Legenda:** LDA score ≥ 4 ; $p \leq 0.05$.

4.7.4 Avaliação do bacterioma dos pacientes em relação a ancestralidade molecular

Associamos os dados de ancestralidade obtidos a partir de sequências *off-target* do sequenciamento do viroma as análises do bacterioma. Não há até o momento estudos relacionado câncer de pênis e ancestralidade, esperamos que esses dados possam contribuir para futuros estudos e investigações relacionadas ao tumor de pênis.

Dos pacientes sequenciados 19 possuíam resultados de ancestralidade sendo que; 5 pacientes todos do estado do Amazonas apresentavam o componente amerindio, 4 pacientes todos do estado de São Paulo apresentavam o componente mediterrâneo, enquanto que 10 pacientes apresentavam uma mistura das populações utilizadas como referência. Na avaliação de diversidade alfa e beta não houve diferença estatisticamente significativa em nenhuma das métricas utilizadas (**Figura 28 e 29**). Nas análises em Lefse foi identificado como diferencialmente abundante para o componente Amerindio o gênero *Neisseria*, enquanto que para os pacientes com componente Mediterrâneo foi observado o gênero *Prevotella* 7 (**Figura 30**). Conforme o observado os resultados apresentaram um viés, visto que todos os pacientes nativos americanos eram do estado do Amazonas, e os mediterrâneos do estado de São Paulo, além disso as bactérias identificadas como enriquecidas apareceram em uma frequência pouco relevante nos pacientes (**Figura 30**).

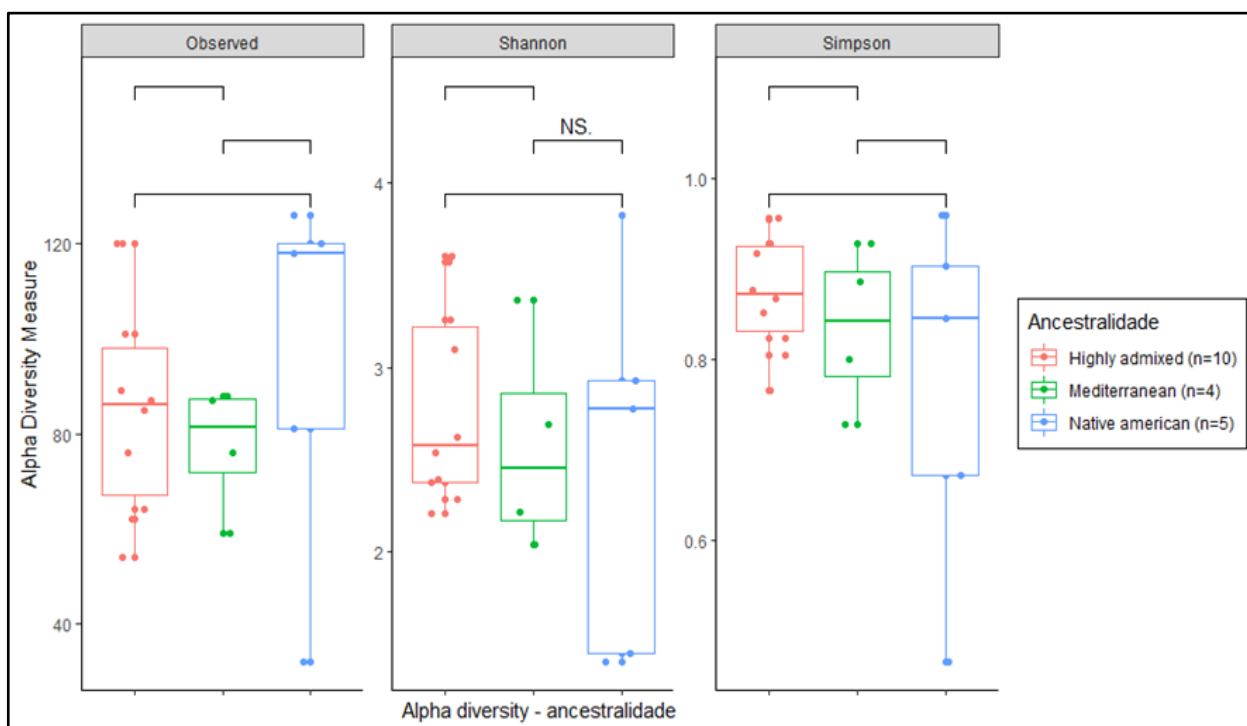


Figura 28 - Medidas de diversidades alfa em relação a ancestralidade molecular. **Legenda:** Teste estatístico de alfa diversidade – Teste de Mann Whitney, $p < 0.05$.

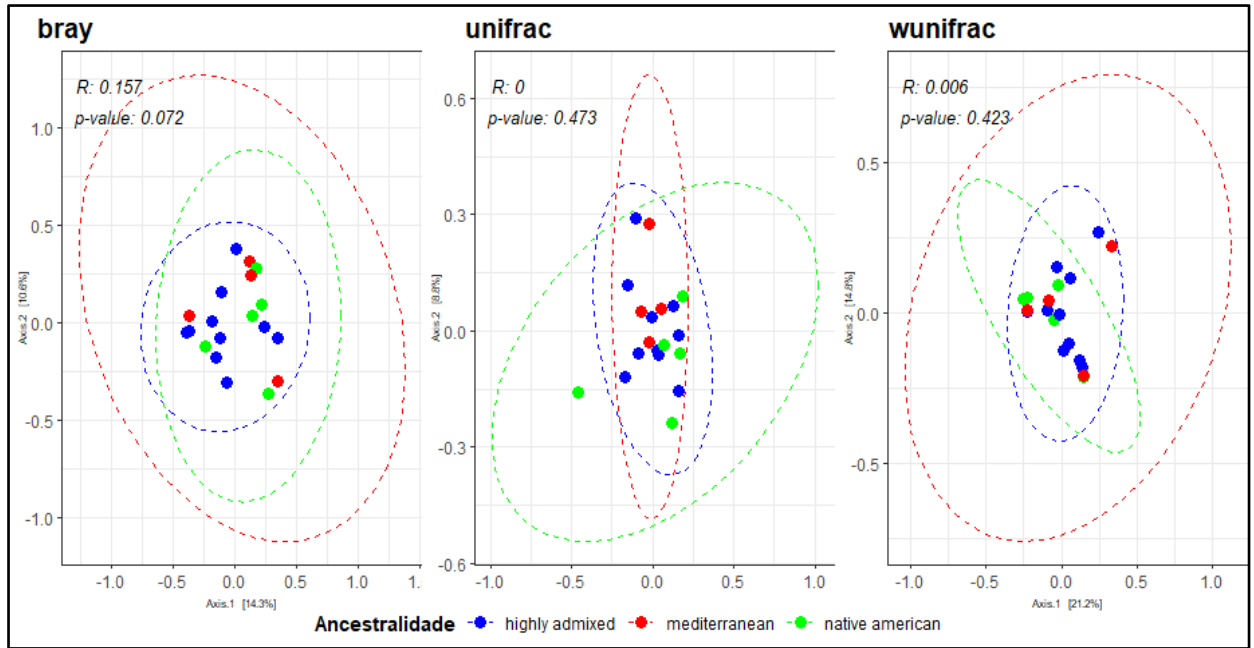


Figura 29 - Medidas de diversidades beta em relação a ancestralidade molecular. **Legenda:** Teste estatístico de Beta diversidade- ANOSIM, $p < 0.05$

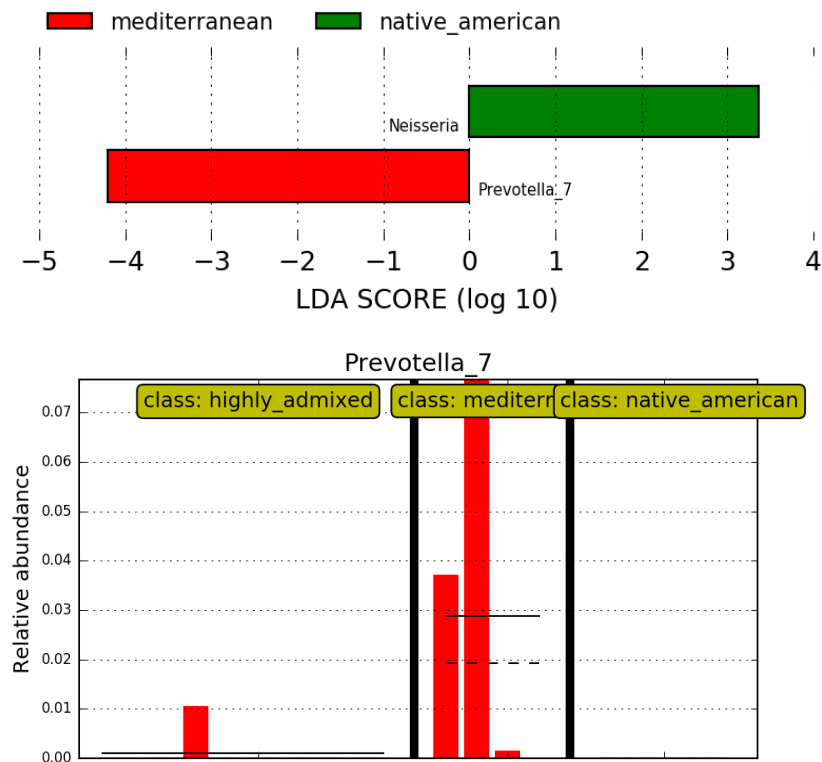


Figura 30 - Análise em LefSe e abundância relativa dos gêneros bacterianos em relação a ancestralidade molecular. **Legenda:** LDA score ≥ 4 ; $p \leq 0.05$.

4.7.5 Análise do bacterioma dos pacientes em relação ao status de HPV e EBV

Como descrito na introdução (**páginas 8 e 9**), ambos os vírus HPV e EBV são associados com o desenvolvimento do câncer de pênis. Neste sentido, avaliamos o status de HPV e EBV nas análises de diversidade Alfa e Beta, separando as amostras nos seguintes grupos: EBV-/HPV+ (n=17); EBV- / HPV- (n= 22); EBV+ / HPV- (n= 5) e EBV- / HPV+ (n=17). Estas análises não mostraram resultados estatisticamente significantes em nenhum dos grupos comparados (**Figuras 31 e 32**). No entanto, cabe salientar que o número de casos é baixo e muito variável entre os grupos, o que pode justificar em parte os resultados obtidos.

Novas análises serão feitas, buscando aumentar o número de casos nos grupos. Além disso, faremos novas análises cruzando dados clínicos com o status de HPV e EBV das amostras.

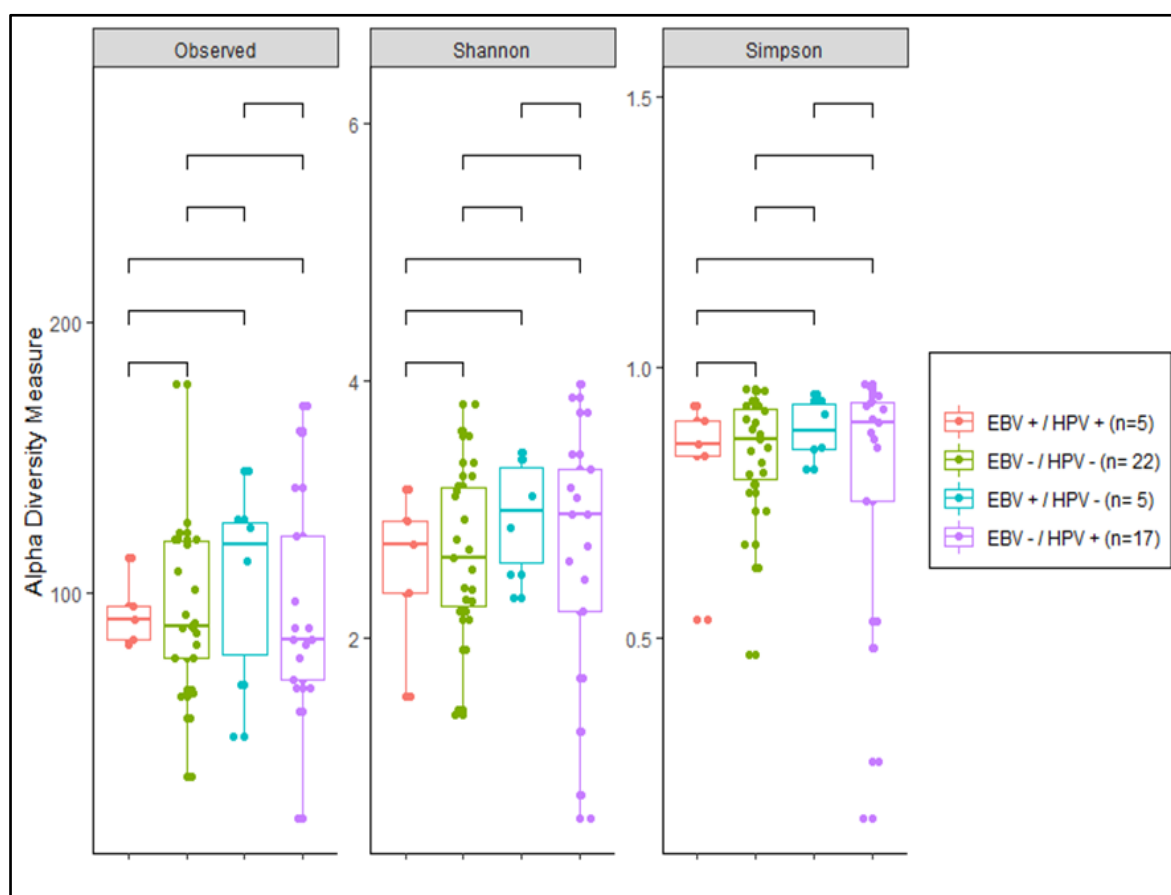


Figura 31 – Medidas de diversidade alfa em relação ao status dos vírus HPV e EBV em amostras tumorais. **Legenda:** Mann Whitney, $p \leq 0.05$ significante

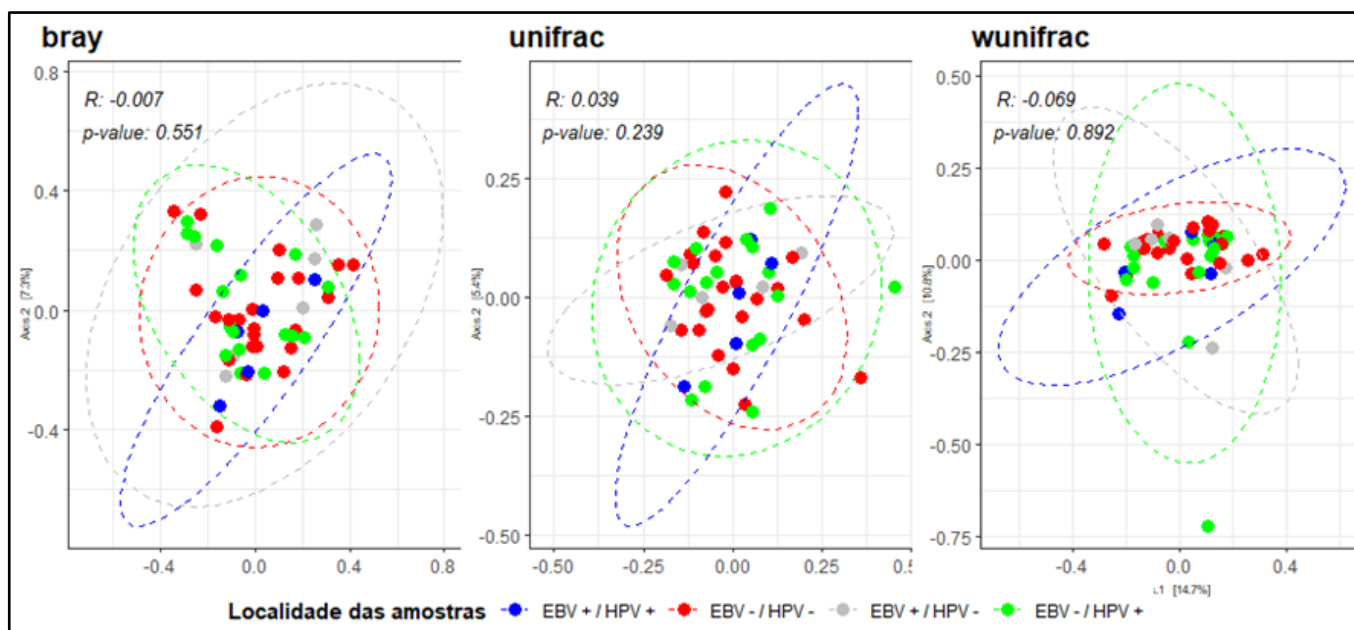


Figura 32 – Medidas de diversidade beta em relação ao status dos vírus HPV e EBV em amostras tumorais . **Legenda:** Teste estatístico de beta diversidade– ANOSIM, $p < 0.05$

4.7.6 Análise do bacterioma dos carcinomas de pênis em relação a invasão perineural e metástase linfonodal

Avaliamos os dados de invasão perineural dos casos dos pacientes de São Paulo e metástase linfonodal das duas coortes, correlacionando-os com os dados de bacterioma. Dados como invasão perineural e outros dados clínicos dos pacientes foram pedidos aos colaboradores do Amazonas, que estão reavaliando e atualizando essas informações. Em relação a ambas as análises de diversidades alfa e beta dos dados de invasão perineural, nenhum dos resultados foi estatisticamente significativo (**Figura 33 e 34**). Mas novamente, cabe notar o número pequeno e dispar de pacientes entre os grupos.

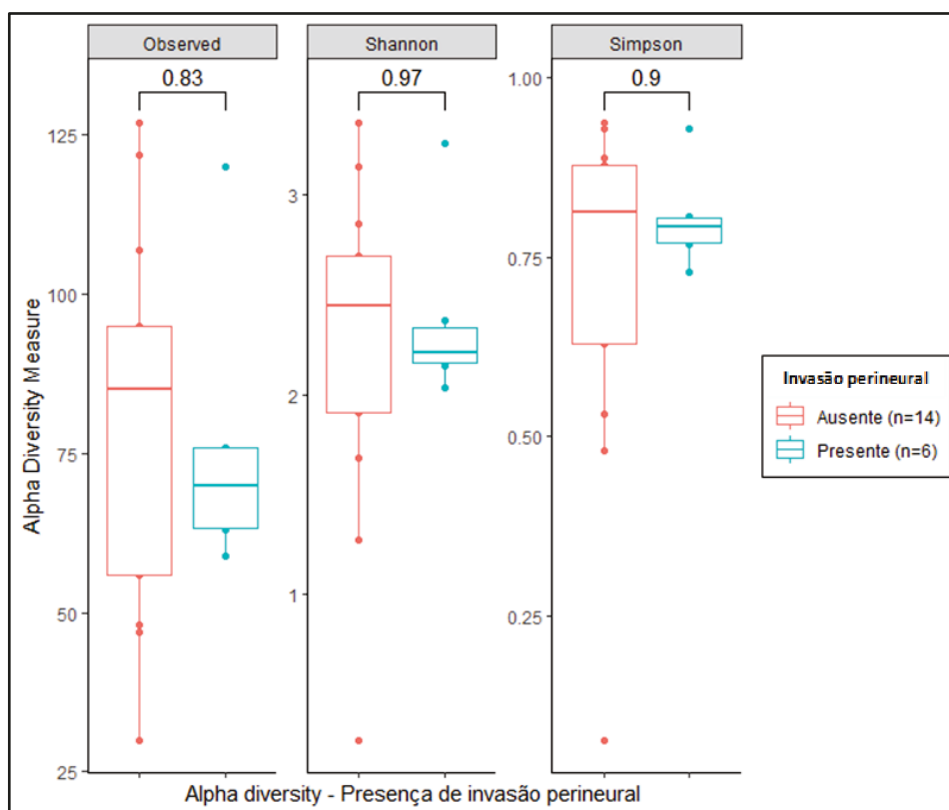


Figura 33 – Medidas de diversidade alfa em relação a invasão perineural nos casos de câncer de pênis de São Paulo. **Legenda:** Teste estatístico utilizado Mann Whitney, $p \leq 0.05$

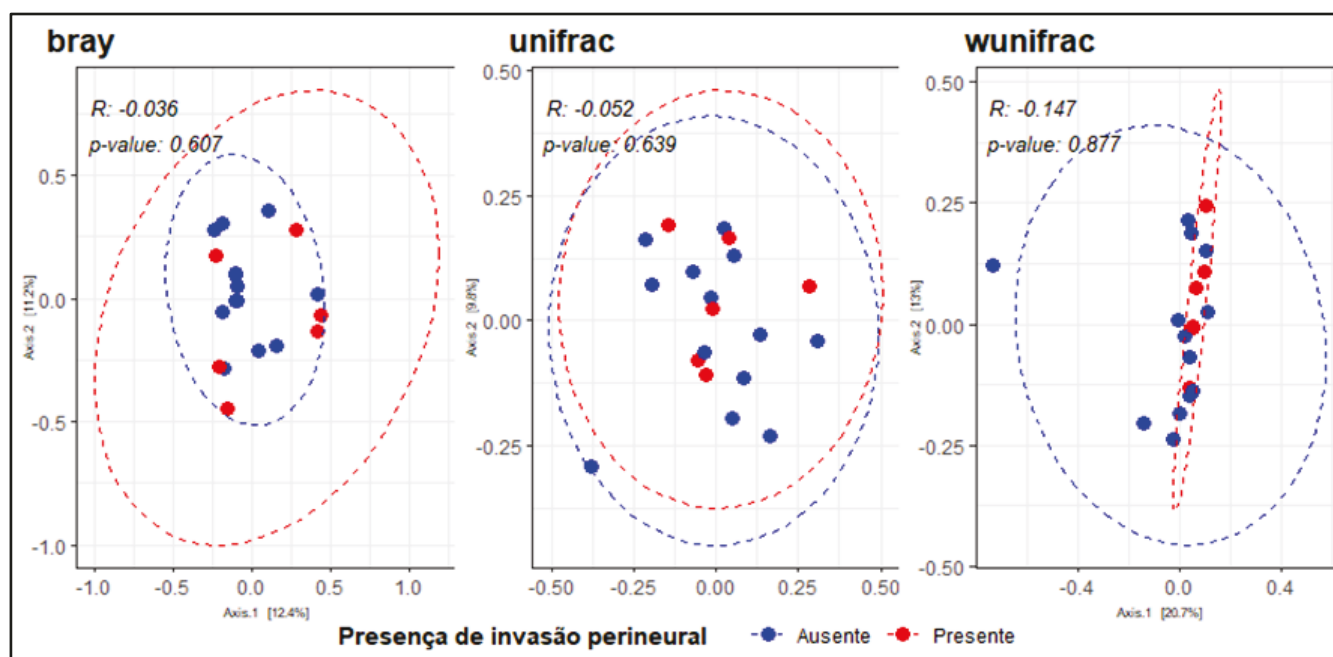


Figura 34 – Medidas de diversidade beta em relação a invasão perineural nos casos de câncer de pênis de São Paulo. **Legenda:** Teste estatístico utilizado ANOSIM, $p \leq 0.05$

A análise de LefSe para invasão perineural mostrou que ambos os gêneros *Alcaligenes* e *Sutterella* estão presentes de modo significativo apenas nas amostras de pacientes com invasão perineural (**Figura 35**). Curiosamente, o gênero *Alcaligenes* já havia sido previamente identificado como enriquecido nas amostras de São Paulo (**Figura 20**).

O Gênero *Sutterella*, apesar de comensal no microbioma intestinal, já foi descrito por ter propriedades pró-inflamatórias, tanto em doenças intestinais, quanto neurológicas, como autismo e síndrome de Down (105).

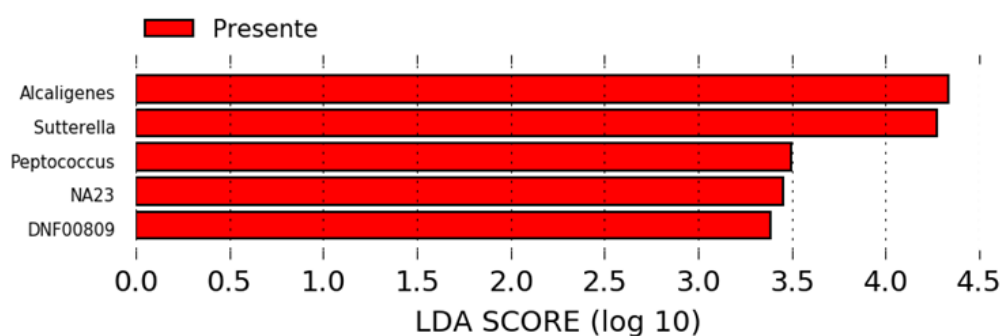


Figura 35 - Análise de LefSe entre os grupos de pacientes com câncer de pênis do estado de SP com e sem invasão perineural. **Legenda:** LDA score ≥ 4 ; $p \leq 0.05$

Em relação às metástases linfonodais, não foram encontradas diferenças estatísticas significativas para as diversidades alfa e beta (**Figura 36 e 37**). No entanto, o gênero *Alcaligenes* está enriquecido no grupo com metástase linfonodal, como mostra a análise LefSe (**Figura 38**). *Alcaligenes* já havia sido associado anteriormente aos pacientes com invasão perineural (106,107) (**Figura 35**).

Já o gênero *Escherichia Shiguella* foi mais frequente nos pacientes sem metástases linfonodais. Ambos os gêneros possuem espécies muito similares entre si e são difíceis de serem distinguidos, por isso sua anotação conjunta. Esses gêneros são em sua maioria inofensivos, mas algumas cepas podem causar infecções e diarreias, como seu enriquecimento observado em tumores colorretais (107).

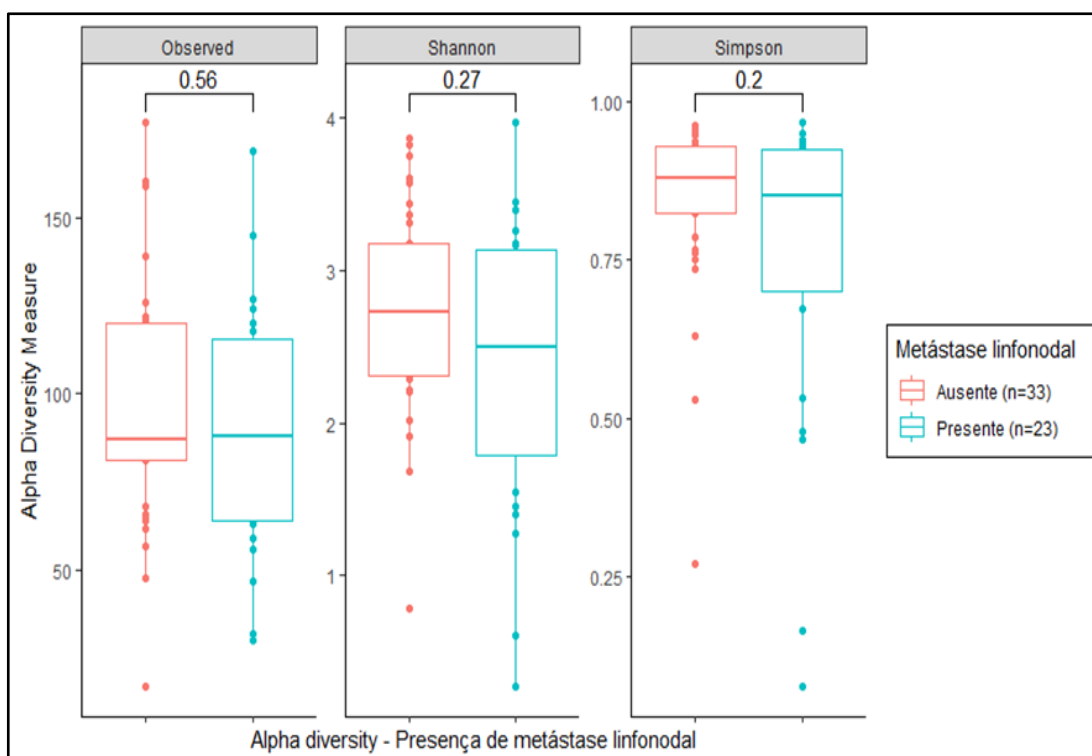


Figura 36 – Medidas de diversidade alfa em relação a metástase linfonodal nos casos de carcinoma de pênis de SP e AM. **Legenda:** Teste estatístico utilizado Mann Whitney, $p \leq 0.05$ significante.

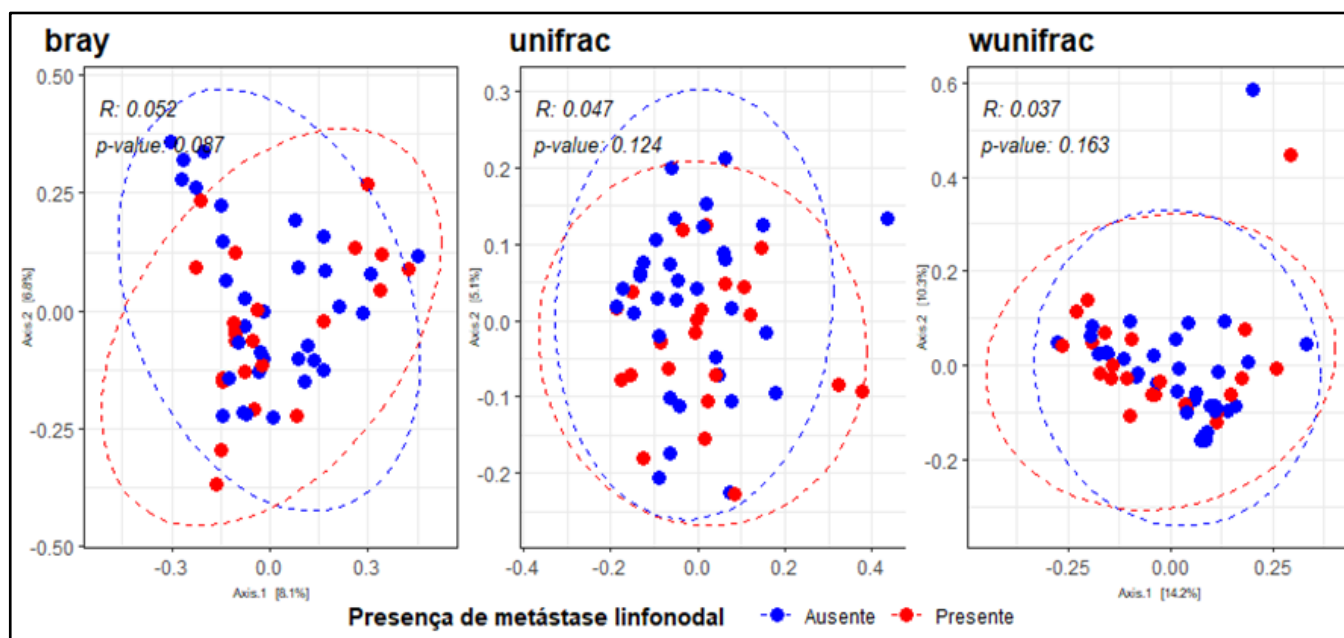


Figura 37 - Medidas de diversidade beta em relação a metástase linfonodal nos casos de câncer de pênis de SP e AM. **Legenda:** Teste estatístico utilizado ANOSIM, $p \leq 0.05$

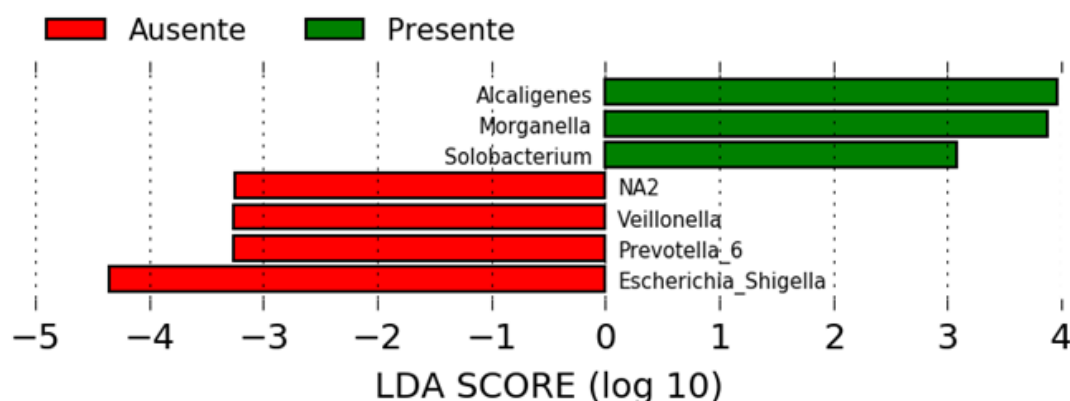


Figura 38 - Análise LefSe das amostras de câncer de pênis com e sem metástase linfonodal.
Legenda: LDA score ≥ 4 ; $p \leq 0.05$.

4.7.7 Análise do bacterioma dos carcinomas de pênis em relação ao tabagismo

O consumo de tabaco está relacionado ao desenvolvimento de muitos tipos de câncer, e também é um fator de risco importante no câncer de pênis, com aumento de até 4,5 vezes de desenvolvê-lo em fumantes (108). A avaliação das diversidades alfa (**Figura 39**) e beta (**Figura 40**) dos pacientes fumantes, ex-fumantes e não-fumantes não mostrou nenhuma diferença significativa entre os grupos. Na análise de LefSe um único gênero de bactéria se mostrou mais enriquecida no grupo não-fumante, com um LDA score próximo a 4. No entanto, este gênero ainda não foi descrito, ou anotado (NA) nos bancos de dados bacterianos. É interessante notar que em outras análises, outros gêneros de bactérias NA também foram encontradas, como na análise LefSe em pacientes com e sem fimose, onde dois gêneros NA12 e NA22 foram mais abundantes (LDA score ≥ 2).

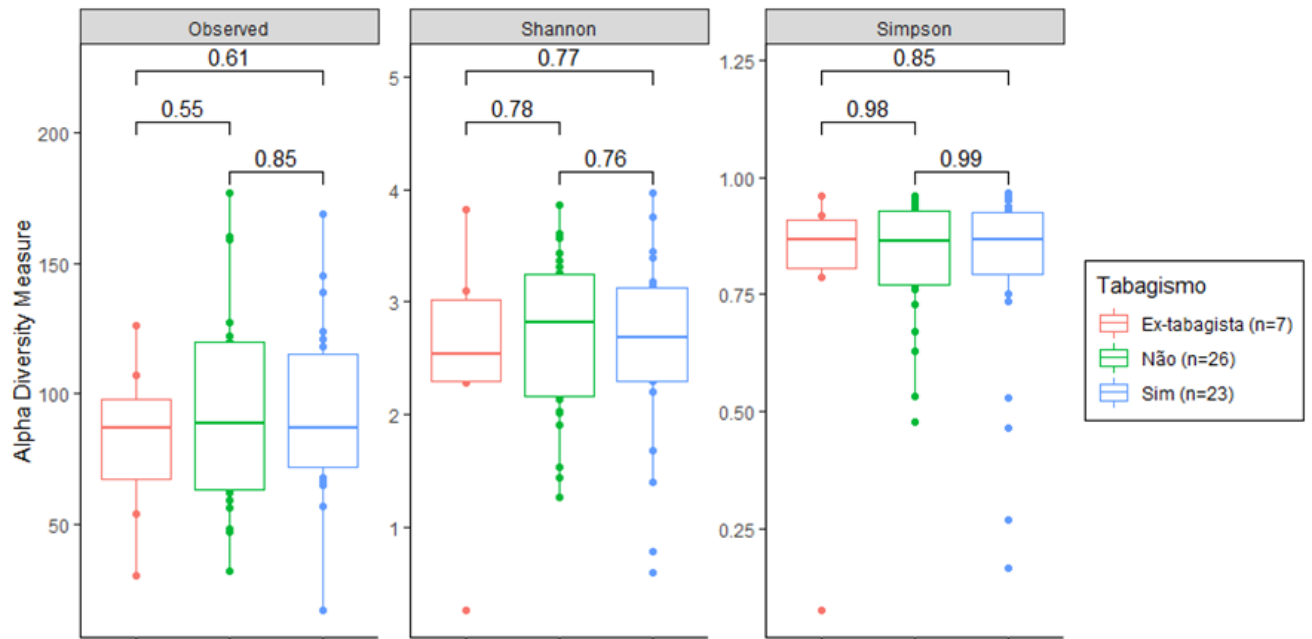


Figura 39 - Medidas de diversidade alfa em relação ao tabagismo nos casos de câncer de pênis. **Legenda:** Teste estatístico utilizado Mann Whitney, $p \leq 0.05$ significativo.

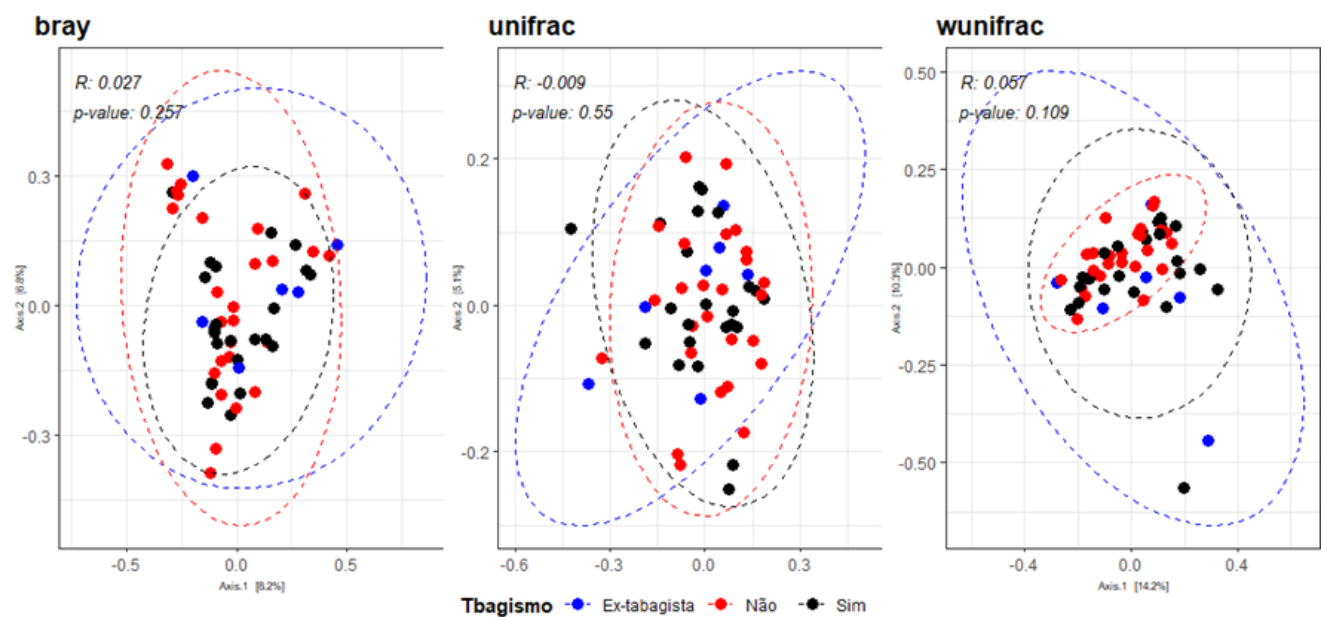


Figura 40 - Medidas de diversidade beta em relação ao tabagismo nos casos de câncer de pênis. **Legenda:** Teste estatístico utilizado ANOSIM, $p \leq 0.05$

4.7.8 Avaliação de associação dos dados de bacterioma com outros parâmetros clínicos

Outros dados clínicos dos pacientes, tanto da coorte de São Paulo quanto do Amazonas, ainda estão sendo revistos e re-adequados, para que futuras análises de associação com os resultados dos bacteriomas sejam gerados. Entre eles destacamos: avaliação do estadiamento clínico-patológico (de acordo com a 8ª edição da AJCC) e grau tumoral dos tumores, presença de invasão vasculolinfática, recidiva tumoral clínica, escolaridade, idade, zoofilia e grau de amputação peniana.

5. CONCLUSÕES

- Apenas 9/23 (39%) das amostras de carcinoma de pênis estudadas tiveram vírus identificados após sequenciamento, sendo que os gêneros *Circovirus* e *Parvovirus* os mais abundantes;
- A análise de ancestralidade molecular dos 24 pacientes que tiveram o viroma avaliado mostrou que o componente ameríndio foi mais enriquecido nas amostras do estado do Amazonas enquanto na coorte de São Paulo não observamos predominância de nenhuma ancestralidade específica, indicando maior miscigenação nessa região do país;
- O gênero bacteriano *Fusobacterium* foi encontrado com maior frequência em ambas as coortes (18,6% em São Paulo e em 12% da coorte do Amazonas);
- As amostras derivadas dos pacientes do Amazonas possuem maior diversidade bacteriana quando comparadas com as de São Paulo, com maior número de gêneros bacterianos identificados, independente da métrica utilizada;
- A diversidade Beta se mostrou distinta entre as coortes do Amazonas e São Paulo nas três métricas avaliadas, indicando diferenças quantitativas e qualitativas na composição bacteriana entre as duas coortes;
- Análises de bactérias diferencialmente abundantes entre os grupos mostraram que o gênero *Alcaligenes* é mais abundante nas amostras de São Paulo, enquanto *Pseudomonas* foi mais frequente nas amostras do Amazonas;
- Amostras de pacientes sem fimose apresentaram maior diversidade bacteriana em comparação ao grupo com fimose e apresentaram o gênero *Enterococcus* (anaeróbios facultativos) como mais abundante enquanto o gênero *Oligella* (aeróbio) foi mais enriquecido nos pacientes sem fimose.
- O gênero bacteriano *Alcaligenes* foi mais frequente nas amostras da coorte de São Paulo, e se apresentou como um possível marcador para presença tanto de invasão perineural,

quanto de metástase linfonodal, indicando que sua presença possa estar associada a um pior prognóstico na evolução do câncer de pênis.

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. A.C.Camargo Cancer Center. INCiTO-INOTE [Internet]. [cited 2019 Mar 11]. Available from: <https://www.accamargo.org.br/pesquisa/incito-inote>
2. Alemany L, Cubilla A, Halc G, Kasamatsu E, Quirós B, Masferrer E, et al. Role of Human Papillomavirus in Penile Carcinomas Worldwide. *Eur Urol* [Internet]. 2016;69(5):953–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26762611>
3. Coelho RWP, Pinho JD, Moreno JS, Garbis DVEO, do Nascimento AMT, Larges JS, et al. Penile cancer in Maranhão, Northeast Brazil: the highest incidence globally? *BMC Urol* [Internet]. 2018 May 29;18(1):50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29843769>
4. INCA. Atlas On-line de Mortalidade [Internet]. 2019. Available from: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-penis>
5. Calmon MF, Tasso Mota M, Vassallo J, Rahal P. Penile carcinoma: risk factors and molecular alterations. *ScientificWorldJournal* [Internet]. 2011 Feb 3;11:269–82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21298218>
6. Feber A, Arya M, de Winter P, Saqib M, Nigam R, Malone PR, et al. Epigenetics markers of metastasis and HPV-induced tumorigenesis in penile cancer. *Clin Cancer Res* [Internet]. 2015 Mar 1;21(5):1196–206. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25311166>
7. Zequi S de C, Guimarães GC, da Fonseca FP, Ferreira U, de Matheus WE, Reis LO, et al. Sex with animals (SWA): behavioral characteristics and possible association with penile cancer. A multicenter study. *J Sex Med* [Internet]. 2012 Jul;9(7):1860–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22023719>
8. Gloeckler Ries LA, Reichman ME, Lewis DR, Hankey BF, Edwards BK. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. *Oncologist* [Internet]. 2003;8(6):541–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14657533>
9. Noone AM, Howlader N, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ CK (eds). SEER Cancer Statistics Review [Internet]. Instituto Nacional do Câncer. 2018 [cited 2019 Feb 13]. Available from: https://seer.cancer.gov/csr/1975_2015/
10. A.C.Camargo Cancer Center. Pênis [Internet]. [cited 2019 Feb 13]. Available from: <https://accamargo.org.br/tipos-de-cancer/penis>
11. Thomas A, Necchi A, Muneer A, Tobias-Machado M, Tran ATH, Van Rompuy A-S, et al. Penile cancer. *Nat Rev Dis Prim* [Internet]. 2021 Dec 11;7(1):11. Available from: <http://www.nature.com/articles/s41572-021-00246-5>
12. INCA. Câncer de pênis [Internet]. Instituto nacional de câncer - ministério da saúde. 2018 [cited 2019 Feb 11]. Available from: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-penis>
13. Dias-Neto E. Estudo de perfis microbianos em tumores humanos: implicações para tumorigênese, prognóstico e resposta ao tratamento. São Paulo; 2018. p. 1–41.
14. American Cancer Society. What Is Penile Cancer? [Internet]. 2018 [cited 2019 Mar 11]. Available from: <https://www.cancer.org/cancer/penile-cancer/about/what-is-penile-cancer.html>
15. American Joint Committee on Cancer. Penis. In: Springer, editor. American Joint Committee on Cancer. New York, NY; 2017. p. 701–14.
16. G. Pizzocaro, F. Algaba, S. Horenblas, E. Solsona, S. Tana H, Van Der Poel NW. DIRETRIZES

- PARA O CÂNCER DE PÊNIS. Eur Urol [Internet]. 2010;6(57):1002–12. Available from: https://portaldaurologia.org.br/medicos/academia/assets/pdf/Diretrizes_para_o_cancer_de_penis.pdf
17. The American Cancer Society. Penile Cancer Stages [Internet]. [cited 2019 Feb 15]. p. 2018. Available from: https://www.cancer.org/cancer/penile-cancer/detection-diagnosis-staging/staging.html#written_by
 18. A.C.Camargo Cancer Center. Tipos de câncer [Internet]. [cited 2019 Mar 11]. Available from: <https://accamargo.org.br/tipos-de-cancer/penis>
 19. Stern J, Miller G, Li X, Saxena D. Virome and bacteriome: two sides of the same coin. Curr Opin Virol [Internet]. 2019 Aug;37:37–43. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1879625719300161>
 20. Janda JM AS. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol [Internet]. 2007;45(9):2761–4. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/17626177>
 21. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nat Rev Microbiol [Internet]. 2018 Mar 15;16(3):143–55. Available from: <http://www.nature.com/articles/nrmicro.2017.157>
 22. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, et al. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J [Internet]. 2010 May 21;4(5):642–7. Available from: <http://www.nature.com/articles/ismej2009153>
 23. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, et al. Evaluating the Information Content of Shallow Shotgun Metagenomics. Rawls JF, editor. mSystems [Internet]. 2018 Oct 30;3(6). Available from: <https://journals.asm.org/doi/10.1128/mSystems.00069-18>
 24. Regalado J, Lundberg DS, Deusch O, Kersten S, Karasov T, Poersch K, et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. ISME J [Internet]. 2020 Aug 13;14(8):2116–30. Available from: <http://www.nature.com/articles/s41396-020-0665-8>
 25. Liu CM, Prodger JL, Tobian AAR, Abraham AG, Kigozi G, Hungate BA, et al. Penile Anaerobic Dysbiosis as a Risk Factor for HIV Infection. Blaser MJ, editor. MBio [Internet]. 2017 Sep 6;8(4). Available from: <https://journals.asm.org/doi/10.1128/mBio.00996-17>
 26. Onywera H, Williamson A-L, Cozzuto L, Bonnin S, Mbulawa ZZA, Coetzee D, et al. The penile microbiota of Black South African men: relationship with human papillomavirus and HIV infection. BMC Microbiol [Internet]. 2020 Dec 6;20(1):78. Available from: <https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-020-01759-x>
 27. Mehta SD, Zhao D, Green SJ, Agingu W, Otieno F, Bhaumik R, et al. The Microbiome Composition of a Man’s Penis Predicts Incident Bacterial Vaginosis in His Female Sex Partner With High Accuracy. Front Cell Infect Microbiol [Internet]. 2020 Aug 4;10. Available from: <https://www.frontiersin.org/article/10.3389/fcimb.2020.00433/full>
 28. Onywera H, Williamson A-L, Ponomarenko J, Meiring TL. The Penile Microbiota in Uncircumcised and Circumcised Men: Relationships With HIV and Human Papillomavirus Infections and Cervicovaginal Microbiota. Front Med [Internet]. 2020 Jul 30;7. Available from: <https://www.frontiersin.org/article/10.3389/fmed.2020.00383/full>
 29. Price LB, Liu CM, Johnson KE, Aziz M, Lau MK, Bowers J, et al. The Effects of Circumcision on the Penis Microbiome. Bereswill S, editor. PLoS One [Internet]. 2010 Jan 6;5(1):e8422. Available from: <https://dx.plos.org/10.1371/journal.pone.0008422>
 30. Zárate S, Taboada B, Yocupicio-Monroy M, Arias CF. Human Virome. Arch Med Res [Internet]. 2017 Nov;48(8):701–16. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29398104>

31. Wylie KM, Weinstock GM, Storch GA. Virome genomics: a tool for defining the human virome. *Curr Opin Microbiol* [Internet]. 2013 Aug;16(4):479–84. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0966842X13001315>
32. Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* [Internet]. 2019 Dec 28;7(1):12. Available from: <https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0626-5>
33. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* [Internet]. 2001 Feb 15;409(6822):860–921. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11237011>
34. Cantalupo PG, Katz JP, Pipas JM. Viral sequences in human cancer. *Virology* [Internet]. 2018 Jan;513(1):208–16. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0042682217303677>
35. National Cancer Institute. The Cancer Genome Atlas Program [Internet]. [cited 2019 Mar 11]. Available from: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
36. Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* [Internet]. 2018;115(52):E12353–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30482864>
37. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu Rev Pathol* [Internet]. 2019 Jan 24;14:319–38. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30355154>
38. Damania B. DNA tumor viruses and human cancer. *Trends Microbiol* [Internet]. 2007 Jan;15(1):38–44. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0966842X06002666>
39. Prati B, Marangoni B, Boccardo E. Human papillomavirus and genome instability: from productive infection to cancer. *Clinics (Sao Paulo)* [Internet]. 2018;73(suppl 1):e539s. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6113919/?report=classic>
40. IARC Monographs. Human Cancer: Known Causes and Prevention by Organ Site [Internet]. 2021. p. 1–129. Available from: <https://monographs.iarc.who.int/wp-content/uploads/2019/12/OrganSitePoster.PlusHandbooks.pdf>
41. National Cancer Institute. HPV and Cancer [Internet]. National Cancer Institute. 2015 [cited 2019 Feb 11]. Available from: <https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-fact-sheet>
42. Scheiner MA, Campos MM, Ornellas AA, Chin EW, Ornellas MH, Andrada-Serpa MJ. Human papillomavirus and penile cancers in Rio de Janeiro, Brazil: HPV typing and clinical features. *Int braz j urol* [Internet]. 2008 Aug;34(4):467–76. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1677-55382008000400009&lng=en&tlng=en
43. Laura M Plata, Julián F Oviedo BR-O. Revisión sistemática: estrategias virales para la inducción de cáncer “virus de Epstein-Barr: latencia y mecanismos asociados a la oncogénesis viral.” *Revista de la Universidad Industrial de Santander*. 2018;50:257–8.
44. Ainsworth C. Building a better lymphoma vaccine. *Nature* [Internet]. 2018 Nov 14;563(7731):S52–4. Available from: <http://www.nature.com/articles/d41586-018-07366-1>
45. Martins V do CA, Cunha IW, Figliuolo G, Rondon HH de MF, de Souza PM, Torres Silva FL, et al. Presence of HPV with overexpression of p16INK4a protein and EBV infection in penile cancer—A series of cases from Brazil Amazon. *Siegel MO, editor. PLoS One* [Internet]. 2020 May 6;15(5):e0232474. Available from: <https://dx.plos.org/10.1371/journal.pone.0232474>

46. Rovedo M, Longnecker R. Epstein-Barr Virus Latent Membrane Protein 2B (LMP2B) Modulates LMP2A Activity. *J Virol* [Internet]. 2007 Jan;81(1):84–94. Available from: <https://journals.asm.org/doi/10.1128/JVI.01302-06>
47. Alves G, Macrini CMT, de Souza Nascimento P, Morais JC, Ornellas AA. Detection and expression of Epstein-Barr Virus (EBV) DNA in tissues from penile tumors in Brazil. *Cancer Lett* [Internet]. 2004 Nov 8;215(1):79–82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15374635>
48. Afonso LA, Moyses N, Alves G, Ornellas AA, Passos MRL, Oliveira L do H dos S, et al. Prevalence of human papillomavirus and Epstein-Barr virus DNA in penile cancer cases from Brazil. *Mem Inst Oswaldo Cruz* [Internet]. 2012 Feb;107(1):18–23. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762012000100003&lng=en&tlng=en
49. Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol* [Internet]. 2009;NA-NA. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ajhb.20976>
50. Liang Z, Bu L, Qin Y, Peng Y, Yang R, Zhao Y. Selection of Optimal Ancestry Informative Markers for Classification and Ancestry Proportion Estimation in Pigs. *Front Genet* [Internet]. 2019 Mar 11;10. Available from: <https://www.frontiersin.org/article/10.3389/fgene.2019.00183/full>
51. Rodrigues de Moura R, Coelho AVC, de Queiroz Balbino V, Crovella S, Brandão LAC. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol* [Internet]. 2015 Sep 10;27(5):674–80. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ajhb.22714>
52. Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy F de SG, et al. The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. Harpending H, editor. *PLoS One* [Internet]. 2011 Feb 16;6(2):e17063. Available from: <https://dx.plos.org/10.1371/journal.pone.0017063>
53. Slopnick EA, Kim SP, Kiechle JE, Gonzalez CM, Zhu H, Abouassaly R. Racial Disparities Differ for African Americans and Hispanics in the Diagnosis and Treatment of Penile Cancer. *Urology* [Internet]. 2016 Oct;96:22–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0090429516303806>
54. da Silva EM, Fernandes MR, de Carvalho DC, Leitao LPC, Cavalcante GC, Pereira EEB, et al. Effect of genetic ancestry to the risk of susceptibility to gastric cancer in a mixed population of the Brazilian Amazon. *BMC Res Notes* [Internet]. 2017 Nov 29;10(1):646. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29187240>
55. Leal LF, de Paula FE, De Marchi P, de Souza Viana L, Pinto GDJ, Carlos CD, et al. Mutational profile of Brazilian lung adenocarcinoma unveils association of EGFR mutations with high Asian ancestry and independent prognostic role of KRAS mutations. *Sci Rep* [Internet]. 2019 Mar 1;9(1):3209. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30824880>
56. Lewis DD, Cropp CD. The Impact of African Ancestry on Prostate Cancer Disparities in the Era of Precision Medicine. *Genes (Basel)* [Internet]. 2020 Dec 8;11(12):1471. Available from: <https://www.mdpi.com/2073-4425/11/12/1471>
57. Conti D V., Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet* [Internet]. 2021 Jan 4;53(1):65–75. Available from: <http://www.nature.com/articles/s41588-020-00748-0>
58. Gravitt PE, Peyton CL, Alessi TQ, Wheeler CM, Coutlée F, Hildesheim A, et al. Improved amplification of genital human papillomaviruses. *J Clin Microbiol* [Internet]. 2000 Jan;38(1):357–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10618116>

59. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* [Internet]. 2017;27(5):824–34. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28298430>
60. Bobrovnitchaia I, Valieris R, Drummond RD, Lima JP, Freitas HC, Bartelli TF, et al. APOBEC-mediated DNA alterations: A possible new mechanism of carcinogenesis in EBV-positive gastric cancer. *Int J Cancer* [Internet]. 2020 Jan 7;146(1):181–91. Available from: <http://www.nature.com/articles/s41598-018-26812-8>
61. Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep* [Internet]. 2015 Nov 12;5(August):16532. Available from: <http://dx.doi.org/10.1038/srep16532>
62. Hopkins C for CBJ. Kraken 2 [Internet]. Available from: <https://ccb.jhu.edu/software/kraken2/index.shtml>
63. NCBI. nucleotide [Internet]. [cited 2019 Sep 26]. Available from: <https://www.ncbi.nlm.nih.gov/nucleotide>
64. Andrews S. FASTQC. Uma ferramenta de controle de qualidade para dados de sequência de alto rendimento. 2010.
65. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
66. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* [Internet]. 2011 May 2;17(1):10. Available from: <http://journal.embnnet.org/index.php/embnnetjournal/article/view/200>
67. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*. 2017;18(Suppl 3).
68. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014;15(3).
69. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* [Internet]. 2012 May;19(5):455–77. Available from: <http://www.liebertpub.com/doi/10.1089/cmb.2012.0021>
70. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* [Internet]. 2015 Jan 17;12(1):59–60. Available from: <http://www.nature.com/articles/nmeth.3176>
71. Merkel D. Docker: lightweight linux containers for consistent development and deployment. *Linux J*. 2014;2014(239):2.
72. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* [Internet]. 2014 Apr 29;5:3513. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24781250>
73. Freitas H. Avaliação do perfil de mutações somáticas em uma coorte de pacientes com adenocarcinoma gástrico. Fundação Antônio Prudente; 2019.
74. broadinstitute. GATK [Internet]. Available from: <https://gatk.broadinstitute.org/hc/en-us>
75. broadinstitute. HaplotypeCaller [Internet]. Available from: <https://gatk.broadinstitute.org/hc/en-us/articles/360036897532-HaplotypeCaller>
76. Neoprosperta Microbiome Technol. Bacterial identification through accurate library preparation and highthroughput sequencing. Available from:

<https://marketing.neoprospecta.com/438d08988c3aa55de70e>

77. Hall M, Beiko RG. 16S rRNA Gene Analysis with QIIME2. *Methods Mol Biol* [Internet]. 2018;1849:113–29. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30298251>
78. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* [Internet]. 2017 Jul 3;45(W1):W180–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28449106>
79. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J GF. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. 2014; Available from: <https://www.arb-silva.de/>
80. Bioconductor. phyloseq [Internet]. Available from: <https://www.bioconductor.org/packages/release/bioc/html/phyloseq.html>
81. R Core Team. RStudio [Internet]. 2020. Available from: <https://www.rstudio.com/products/rstudio/>
82. Morgan XC, Huttenhower C. Chapter 12: Human Microbiome Analysis. Lewitter F, Kann M, editors. *PLoS Comput Biol* [Internet]. 2012 Dec 27;8(12):e1002808. Available from: <https://dx.plos.org/10.1371/journal.pcbi.1002808>
83. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol* [Internet]. 2011;12(6):R60. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-6-r60>
84. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, et al. Multiple Diverse Circoviruses Infect Farm Animals and Are Commonly Found in Human and Chimpanzee Feces. *J Virol*. 2010;84(4):1674–82.
85. Liu X, Ouyang T, Ouyang H, Liu X, Niu G, Huo W, et al. Human cells are permissive for the productive infection of porcine circovirus type 2 in vitro. *Sci Rep* [Internet]. 2019 Dec 4;9(1):5638. Available from: <http://www.nature.com/articles/s41598-019-42210-0>
86. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E DJ. Virus Identification in Unknown Tropical Febrile Illness Cases Using Deep Sequencing. *PLoS Negl Trop Dis* [Internet]. 2012;2(1485.):6. Available from: <https://doi.org/10.1371/journal.pntd.0001485>
87. Pattison JR PG. Parvoviruses. In: Baron S, editor. *Medical Microbiology* [Internet]. 1996. p. Chapter 64. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK7715/>
88. Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, et al. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. *Proc Natl Acad Sci U S A* [Internet]. 2013 Jun 18;110(25):10264–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23716702>
89. Smuts H, Kew M, Khan A, Korsman S. Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants in silica column-based nucleic acid extraction kits. *J Virol* [Internet]. 2014 Jan;88(2):1398. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24335290>
90. Falchook GS, Rady P, Hymes S, Nguyen HP, Tying SK, Prieto VG, et al. Merkel cell polyomavirus and HPV-17 associated with cutaneous squamous cell carcinoma arising in a patient with melanoma treated with the BRAF inhibitor dabrafenib. *JAMA dermatology* [Internet]. 2013 Mar;149(3):322–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23552670>
91. Tomic S, Warner TF, Messing E, Wilding G. Penile merkel cell carcinoma. *Urology* [Internet]. 1995 Jun;45(6):1062–5. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0090429599801344>
92. Aron M, Zhou M. Merkel Cell Carcinoma of the Genitourinary Tract. *Arch Pathol Lab Med* [Internet]. 2011 Aug 1;135(8):1067–71. Available from: <http://meridian.allenpress.com/aplm/article/135/8/1067/65078/Merkel-Cell-Carcinoma-of-the->

Genitourinary-Tract

93. Howard A, O'Donoghue M, Feeney A, Sleator RD. *Acinetobacter baumannii*. Virulence [Internet]. 2012 May 27;3(3):243–50. Available from: <http://www.tandfonline.com/doi/abs/10.4161/viru.19700>
94. Agarwal K, Robinson LS, Aggarwal S, Foster LR, Hernandez-Leyva A, Lin H, et al. Glycan cross-feeding supports mutualism between *Fusobacterium* and the vaginal microbiota. Cadwell K, editor. PLOS Biol [Internet]. 2020 Aug 25;18(8):e3000788. Available from: <https://dx.plos.org/10.1371/journal.pbio.3000788>
95. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, et al. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. Gastroenterology [Internet]. 2018 Aug;155(2):383–390.e8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S001650851830489X>
96. Komiya Y, Shimomura Y, Higurashi T, Sugi Y, Arimoto J, Umezawa S, et al. Patients with colorectal cancer have identical strains of *Fusobacterium nucleatum* in their colorectal cancer and oral cavity. Gut [Internet]. 2019 Jul;68(7):1335–7. Available from: <https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2018-316661>
97. Ruiz-Pino M, Foronda-García-Hidalgo C, Alarcón-Blanco P, Gutiérrez-Fernández J. Male genitourinary infections by *Corynebacterium glucuronolyticum*. A review and clinical experience. Rev Esp Quimioter [Internet]. 2019 Oct;32(5):479–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31515978>
98. Donald Peter , DVM, MS, DACT, Frontier Genetics, Hermiston O. Campylobacteriose genital bovina -Manual MDS [Internet]. 2020. Available from: <https://www.msddvetmanual.com/reproductive-system/bovine-genital-campylobacteriosis/bovine-genital-campylobacteriosis>
99. Kuhn KG, Hvass AK, Christiansen AH, Ethelberg S, Cowan SA. Sexual Contact as Risk Factor for *Campylobacter* Infection, Denmark. Emerg Infect Dis [Internet]. 2021 Apr;27(4):1133–40. Available from: https://wwwnc.cdc.gov/eid/article/27/4/20-2337_article.htm
100. Ramos JM, Domine M, Ponte MC, Soriano F. [Bacteremia caused by *Alcaligenes* (*Achromobacter*) *xylosoxidans*. Description of 3 cases and review of the literature]. Enferm Infecc Microbiol Clin [Internet]. 14(7):436–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8991439>
101. Larke NL, Thomas SL, dos Santos Silva I, Weiss HA. Male circumcision and penile cancer: a systematic review and meta-analysis. Cancer Causes Control [Internet]. 2011 Aug 22;22(8):1097–110. Available from: <http://link.springer.com/10.1007/s10552-011-9785-9>
102. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer [Internet]. 2013 Nov 17;13(11):800–12. Available from: <http://www.nature.com/articles/nrc3610>
103. Stein-Thoeringer CK, Nichols KB, Lazrak A, Docampo MD, Slingerland AE, Slingerland JB, et al. Lactose drives *Enterococcus* expansion to promote graft-versus-host disease. Science (80-) [Internet]. 2019 Nov 29;366(6469):1143–9. Available from: <https://www.sciencemag.org/lookup/doi/10.1126/science.aax3760>
104. Bush LMMTV-P. Infecções por *Oligella* [Internet]. 2020. Available from: <https://www.msddmanuals.com/pt-br/profissional/doencas-infecciosas/cocos-e-cocobacilos-gram-negativos/infecções-por-oligella>
105. Hiippala K, Kainulainen V, Kalliomäki M, Arkkila P, Satokari R. Mucosal Prevalence and Interactions with the Epithelium Indicate Commensalism of *Sutterella* spp. Front Microbiol [Internet]. 2016 Oct 26;7. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01706/full>
106. Chen Y-T, Peng H-L, Shia W-C, Hsu F-R, Ken C-F, Tsao Y-M, et al. Whole-genome sequencing and identification of *Morganella morganii* KT pathogenicity-related genes. BMC Genomics [Internet]. 2012;13(Suppl 7):S4. Available from: <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471->

2164-13-S7-S4

107. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and *Shigella* Species from Whole-Genome Sequences. Ledeboer NA, editor. *J Clin Microbiol* [Internet]. 2017 Feb;55(2):616–23. Available from: <https://journals.asm.org/doi/10.1128/JCM.01790-16>

108. Daling JR, Madeleine MM, Johnson LG, Schwartz SM, Shera KA, Wurscher MA, et al. Penile cancer: Importance of circumcision, human papillomavirus and smoking in situ and invasive disease. *Int J Cancer* [Internet]. 2005 Sep 10;116(4):606–16. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ijc.21009>

Anexo 1 – Carta de aprovação do Comitê de Ética em Pesquisa - (CEP)**A.C. Camargo Cancer Center**

Centro Integrado de Diagnóstico, Tratamento, Ensino e Pesquisa

**COMITÊ DE ÉTICA
EM PESQUISA - CEP****APROVAÇÃO**

Os membros do Comitê de Ética em Pesquisa da Fundação Antonio Prudente – A.C. Camargo Cancer Center, em sua última reunião de **21/05/2019**, após analisarem as respostas aos questionamentos realizados em reunião de **26/03/2019**, **aprovaram** a realização do projeto nº **2689/19** intitulado: “**Avaliação do viroma em carcinomas de pênis por sequenciamento de nova geração.**”

Pesquisador Responsável: Diana Noronha Nunes**Aluna:** Elizabeth Cristina Miola (Mestrado)

Informações a respeito do andamento do referido projeto deverão ser encaminhadas ao CEP dentro de 06 meses em relatório (modelo CEP).

São Paulo, 24 de maio de 2019.

Atenciosamente,

Dra. Sandra Caires Serrano**1ª. Vice-Coordenadora do Comitê de Ética em Pesquisa**

1/1

Anexo 2 - Carta de anuência da Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON)



CARTA DE ANUÊNCIA

Declaramos para os devidos fins que mantemos parceria com os pesquisadores Diana Noronha Nunes, Maria Galil de Amorim, Emmanuel dias Neto e Elizabeth Cristina Miola para a execução do estudo ***"Avaliação do viroma em carcinomas de pênis por sequenciamento de nova geração"***. Assim sendo, somos de acordo com a execução das análises necessárias ao referido estudo utilizando amostras de pacientes com câncer de pênis obtidas no estado do Amazonas no contexto do estudo já em andamento "Análise metagenômica do viroma das neoplasias penianas" aprovado pelo Cep da Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON) em 21 de Agosto de 2017, sob Número do Parecer: 2.230.007

Manaus, 26 de abril de 2019

Atenciosamente,

Kátia Luz Torres Silva
Diretora de Ensino e Pesquisa
Fundação Centro de Controle de Oncologia do Amazonas

Valquíria do Carmo Alves Martins
Pesquisadora
Fundação Centro de Controle de Oncologia do Amazonas

