

Curso de Pós-Graduação em Ciências da Fundação Antônio Prudente  
Área de Oncologia

**Desenvolvimento de Metodologia para Identificação de Fungos Através de *Metabarcoding* Aplicada em Amostras Humanas de Adenocarcinoma Gástrico e Carcinoma Epidermóide de Pênis**

Tese apresentada à Fundação Antônio Prudente para a obtenção do título de Doutor em Ciências  
Área de Concentração: Oncologia

Aluno: Bruno Spinetti Moda  
Nível: Doutorado  
Orientador: Israel Tojal da Silva  
Co-Orientador: Emmanuel Dias-Neto

São Paulo  
29 de Setembro de 2022

**Moda, Bruno .**

**Desenvolvimento de Metodologia para Identificação de Fungos Através de Metabarcoding Aplicada em Amostras Humanas de Adenocarcinoma Gástrico e Carcinoma Epidermóide de Pênis. / Bruno Moda. São Paulo, 2022.**

**145f.**

**Tese de Doutorado - Fundação Antônio Prudente. Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.**

**Orientador: Israel Tojal da Silva.**

**1. Microbioma fúngico, 2. Metabarcoding, 3. ITS2**

**CDU 616**

# Folha de Aprovação

**Nome:** Bruno Spinetti Moda

**Título:** Doutor em Ciências

**Aprovado em:** 29/08/022

## Banca Examinadora

**Orientador:** Israel Tojal da Silva

**Instituição:** Fundação Antônio Prudente

**Membro da banca:** Fabiano Cordeiro Moreira

**Instituição:** Universidade Federal do Pará

**Membro da banca:** Kelly Ishida

**Instituição:** Universidade São Paulo - São Paulo

**Membro da banca:** Nilce Maria Martinez Rossi

**Instituição:** Universidade de São Paulo - Ribeirão Preto

**Membro da banca:** Helano Carioca Freitas

**Instituição:** Fundação Antônio Prudente

## **SUPORTE À PESQUISA POR AGÊNCIA DE FOMENTO**

Este trabalho recebeu apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (período de 03/2018 a 04/2018), da Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), através de auxílio à Pesquisa - 2018/02972-3, do Instituto Nacional para Ciência e Tecnologia em Oncogenômica e Inovação Terapêutica (INCiTO-INOTE) (Processo FAPESP - 2014/50943-1, Processo CNPq - 465682/2014-6) e do Programa Nacional de Atenção ao Paciente Oncológico, Departamento de Ciência e Tecnologia, Ministério da Saúde: PRONON – SIPAR 2500.055-167/2015-23.

*A Biologia permite, a Cultura proíbe.*

*Yuval Noah Harari*

## AGRADECIMENTOS

Agradeço a todo apoio da minha família, aos meus irmãos (Karina e Duda), Bianca e Thainara, e principalmente, aos meus pais Fátima e Moda que me deram todo o suporte ao longo de toda a minha vida e ao longo de toda minha Pós-Graduação.

Ao meu orientador Dr. Israel Tojal da Silva, pela oportunidade de realizar o Doutorado em uma instituição com a infraestrutura do A.C.Camargo, por ter me recebido em seu laboratório e por todos os ensinamentos em relação a programação e bioinformática.

Ao meu coorientador Dr. Emmanuel Dias-Neto, por oferecer todo o suporte para realizar todos os experimentos e pelas discussões para solucionar os diversos desafios ao longo do projeto.

A Dra. Diana Noronha Nunes por todo auxílio e discussões para solucionar diversas questões técnicas relacionadas a biologia molecular.

A todos do Laboratório de Bioinformática Computacional e Bioinformática, por todos o aprendizado, auxílio e amizade ao longo desses anos, Felipe, Jaqueline, Monize, Renan, Rodrigo Borges, Rodrigo Drummond e em particular ao Alexandre Defelicibus, aprendi muito com a sua organização e didática para repassar os conceitos relacionados a programação, agradeço a todo suporte que você sempre forneceu ao grupo, facilitando o trabalho de todos.

A equipe do Laboratório de Genômica Médica, por me dar todo aporte para realizar e me ajudar em meus experimentos. Em particular agradeço muito a Jordana, Gabriela Albuquerque, Marianna, Lucas, Maria, Teresa, Luana e Gabriela Branco e principalmente à Dr. Thais Fernanda Bartelli por todo a ajuda e suporte, em especial com os sequenciamentos das amostras de tumores utilizadas. Sua ajuda foi fundamental para a conclusão dos experimentos finais do projeto.

A equipe do Biobanco de Macromoléculas, em especial a Ana Paula por

sempre estar disposta a auxiliar nas buscas sem fim atrás das amostras dos tumores do projeto.

A Jennice por todo apoio, principalmente na recepção e instrução assim que entrei na pós-graduação, no auxílio para preparar toda a documentação do cadastro do projeto no programa.

A todos meu amigos de faculdade por todo apoio de sempre, principalmente nos momentos difíceis passados durante todos esses anos.

## RESUMO

Moda B.S.. **Desenvolvimento de Metodologia para Identificação de Fungos Através de *Metabarcoding* Aplicada em Amostras Humanas de Adenocarcinoma Gástrico e Carcinoma Epidermoide de Pênis.** [Tese]. São Paulo; Fundação Antônio Prudente; 2022.

O microbioma humano compreende material genético da microbiota de um local do corpo e tem influência direta ou indireta na manutenção da homeostase. O distúrbio da microbiota pode estar relacionado ao desenvolvimento de doenças. A população fúngica ainda é muito pouco estudada no contexto do microbioma. No presente estudo, foi desenvolvida uma metodologia para identificação de fungos por *metabarcoding*. A metodologia desenvolvida foi aplicada em amostras de pacientes portadores de adenocarcinoma gástrico ou carcinoma epidermoide de pênis. De modo geral, em ambos os tumores foi verificada a redução de diversidade fúngica conforme a evolução do estadiamento patológico. Também foram verificados resultados não concordantes ao analisar espécies diferencialmente abundantes em dados de sequenciamento da região ITS2 e de WGS nas amostras de lavado gástrico. Este trabalho reforça a importância em se estudar os fungos e sua associação com doenças como o câncer e incentiva próximos estudos através do desenvolvimento de uma metodologia específica para o microbioma.

**Palavras-chave:** Microbioma fúngico, Micobioma, ITS2, *Metabarcoding*, Adenocarcinoma Gástrico, Carcinoma Epidermoide de Pênis

## ABSTRACT

Moda B.S.. [Development of Methodology for Fungi Identification Through Metabarcoding Applied in Gastric Adenocarcinoma and Penile Epidermoid Carcinoma of Human Samples]. [Tese]. São Paulo; Fundação Antonio Prudente; 2022.

The human microbiome comprises genetic material from the microbiota of a body site and has a direct or indirect influence on the maintenance of homeostasis. The disturbance of the microbiota may be related to the development of diseases. The fungal population is still very little studied in the context of the microbiome. In this study, a methodology was developed to identify fungi by *metabarcoding*. The methodology developed was applied to samples from patients with gastric adenocarcinoma or squamous cell carcinoma of the penis. In general, in both tumors, a reduction in fungal diversity was observed according to the evolution of the pathological staging. Discordant results were also found when analyzing differentially abundant species in sequencing data from the ITS2 region and WGS in gastric lavage samples. This work reinforces the importance of studying fungi and their association with diseases such as cancer and encourages further studies through the development of a specific methodology for the mycobiome.

**Keywords:** Fungi Microbiome, Mycobiome, ITS2, Metabarcoding, Gastric Adenocarcinoma, Penile Epidermoid Carcinoma

## LISTA DE FIGURAS

Figura 1 – Fluxograma do método de seleção dos oligonucleotídeos a serem utilizados para amplificação da região de interesse . . . . .	21
Figura 2 – Fluxograma do <i>pipeline</i> de processamento das leituras obtidas do sequenciamento da região ITS2 . . . . .	27
Figura 3 – Fluxograma do <i>pipeline</i> de processamento das <i>reads</i> obtidas no sequenciamento WGS . . . . .	31
Figura 4 – Variabilidade do tamanho dos amplicons (bp) recuperados pelo par de primers selecionados . . . . .	37
Figura 5 – Eletroforese das reações de PCR utilizando o par de <i>primers</i> 5.8Sf_khot/2 . . . . .	38
Figura 6 – Eficiência de diferentes protocolos de extração de DNA de material fecal provenientes de dois indivíduos do projeto “MOSAIC Challenge”. . . . .	40
Figura 7 – Amplificação da região ITS2 de amostras de DNA extraídas de material fecal provenientes de dois indivíduos do projeto “MOSAIC Challenge” . . . . .	42
Figura 8 – Curvas de rarefação de OTUs observadas dos <i>amplicons</i> gerados a partir de amostras fecais extraídas por diferentes protocolos . . . . .	43
Figura 9 – Comparação da diversidade fúngica entre os protocolos, na presença ou ausência de MAC4L. . . . .	44
Figura 10 – Comparação da diversidade fúngica entre presença e ausência de MAC4L para cada protocolo de extração. . . . .	45
Figura 11 – Perfil da população fúngica presente entre as amostras fecais extraídas por diferentes protocolos . . . . .	46

Figura 12 – Amplificação das regiões ITS2 presentes nas comunidades <i>mocks</i> . . . . .	48
Figura 13 – Eletroferograma das bibliotecas produzidas da região ITS2 amplificada por PCR das amostras Mock fúngicas	49
Figura 14 – Comparação entre as abundâncias observadas e esperadas das espécies presentes nas comunidades <i>mock</i> . .	51
Figura 15 – Análise de diferentes sequenciamentos de comunidades <i>mock</i> fúngicas . . . . .	54
Figura 16 – Testes para otimização do protocolo de PCR para região ITS2 em amostras de lavado gástrico . . . . .	58
Figura 17 – Amplificação da região ITS2 para amostras de lavado gástrico utilizando protocolo otimizado . . . . .	59
Figura 18 – Variabilidade do perfil de amplificação da região ITS2 de acordo com o tipo de amostra de origem . . . . .	61
Figura 19 – Curvas de Rarefação - Amostras de Pênis - ITS2 . . . .	68
Figura 20 – Perfil taxonômico dos microbioma fúngico das amostras de pênis a nível de Gênero . . . . .	70
Figura 21 – Diversidade alfa dos fungos identificados nas amostras de pênis . . . . .	71
Figura 22 – Diversidade beta dos fungos identificados nas amostras de pênis . . . . .	73
Figura 23 – Distribuição de sequências das amostras de lavado gástrico sequenciadas - ITS2 . . . . .	80
Figura 24 – Curvas de rarefação - amostras de lavado gástrico - ITS2	81
Figura 25 – Diversidade alfa dos fungos identificados nas amostras de lavado gástrico . . . . .	82
Figura 26 – Sequências Classificadas e Não Classificadas pelo <i>software</i> Kraken2 dos dados obtidos do TCGA . . . . .	90

Figura 27 – Distribuição das sequências fúngicas e humanas das amostras de STAD - TCGA . . . . .	91
Figura 28 – Comparação da presença de gêneros fúngicos em relação às informações clínicas e moleculares das amostras de STAD - TCGA . . . . .	94
Figura 29 – Frequência relativa de gêneros fúngicos identificados nos dados de WGS em relação a informações das amostras STAD - TCGA . . . . .	95
Figura 30 – Comparação da presença de gêneros fúngicos identificados no sequenciamento de DNA total em relação às informações clínicas e moleculares das amostras de lavado gástrico - ACCCC . . . . .	97
Figura 31 – Frequência relativa de gêneros fúngicos identificados no sequenciamento de DNA total em relação a informações das amostras de lavado gástrico - ACCCC . . . . .	98

## LISTA DE TABELAS

Tabela 1 – Bancos de sequências utilizados para classificação taxonômica de fungos . . . . .	33
Tabela 2 – Similaridade entre o banco UNITE e outros bancos. . . . .	34
Tabela 3 – Lista dos dez pares de oligonucleotídeos mais abrangentes em relação ao banco UNITE . . . . .	35
Tabela 4 – Porcentagem de sequências recuperadas do banco UNITE pelo par de primers 5.8Sf_khot/28S-1r . . . . .	36
Tabela 5 – Abundância relativa das <i>reads</i> geradas a partir do sequenciamento das amostras de fezes. . . . .	46
Tabela 6 – Sequenciamento de comunidades <i>mock</i> fúngicas. . . . .	53
Tabela 7 – Dados clínicos - Amostras de pênis . . . . .	62
Tabela 8 – Número de reads sequenciadas de 25 amostras de pênis . . . . .	66
Tabela 9 – Abundância diferencial de fungos nas amostra de pênis em relação ao tipo da amostra. . . . .	74
Tabela 10 – Abundância diferencial de fungos nas amostra de pênis em relação à presença ou ausência de metástase. . . . .	76
Tabela 11 – Dados clínicos - Conjunto amostral de LG . . . . .	77
Tabela 12 – Comparação da diversidade beta das amostras de lavado gástrico. . . . .	83
Tabela 13 – Abundância diferencial dos fungos presentes nas amostras de lavado gástrico, a nível de Gênero, entre amostras de gastrite superficial e amostras de câncer gástrico . . . . .	87
Tabela 14 – Análise de abundância diferencial dos fungos das amostras de lavado gástrico relativo às amostras com doença localizada e metastática. . . . .	88
Tabela 15 – Oligonucleotídeos levantados na literatura . . . . .	136

Tabela 16 – Número de <i>reads</i> do terceiro sequenciamento das amostras de pênis. . . . .	144
--	-----

## LISTA DE SIGLAS E ABREVIATURAS

<b>ACCCC</b>	A.C.Camargo Cancer Center
<b>AD</b>	Abundância Diferencial
<b>AM</b>	Amazonas
<b>AMPtk</b>	Amplicon Toolkit
<b>ANCOM-BC</b>	Análise de Composição de Microbiomas com Correção de Viés (do inglês, Analysis of Compositions of Microbiomes with Bias Correction)
<b>ANOSIM</b>	Análise de Similaridades (do inglês, Analysis of Similarities)
<b>ASV</b>	Amplicon Sequence Variants
<b>BAM</b>	Binary Alignment Map
<b>BLAST</b>	Basic Local Alignment Search Tool
<b>bx</b>	Biópsia
<b>CaPe</b>	Câncer de Pênis
<b>CE</b>	Carcinoma Epidermoide
<b>CG</b>	Câncer Gástrico
<b>DMSO</b>	Dimetilsulfóxido
<b>DNA</b>	Ácido Desoxirribonucleico (do inglês, Deoxyribonucleic Acid)
<b>EBV</b>	Vírus Epstein-Barr (do inglês, Epstein-Barr Virus)
<b>Eda</b>	Endoscopia alta

<b>FCECON</b>	Fundação Centro de Controle de Oncologia
<b>GRCh</b>	Genome Reference Consortium Human
<b>HBV</b>	Vírus da Hepatite B (do inglês, Hepatitis B Virus)
<b>HCV</b>	Vírus da Hepatite C (do inglês, Hepatitis C Virus)
<b>hg</b>	human genome
<b>HPV</b>	Papilomavírus humano (do inglês, Human Papillomavirus)
<b>IARC</b>	Agência Internacional de Pesquisa em Câncer (do inglês, International Agency for Research on Cancer)
<b>IMT</b>	Instituto de Medicina Tropical
<b>INCA</b>	Instituto Nacional do Câncer
<b>ISP</b>	Ion Spheres Particles
<b>ITS</b>	Espaçador Interno Transcrito (do inglês, Internal Transcript Spacer)
<b>Lasso</b>	Least Absolute Shrinkage and Selection Operator
<b>LFC</b>	Log Fold Change
<b>LG</b>	lavado gástrico
<b>MAC4L</b>	MetaPolzyme
<b>NGS</b>	Sequenciamento de Nova Geração (do inglês, New Generation Sequencing)
<b>NIH</b>	Instituto Nacional de Saúde (do inglês, National Institute Health)
<b>nt</b>	nucleotídeo

<b>NTC</b>	No Template Control
<b>OTU</b>	Unidade Operacional Taxonômica (do inglês, Operational Taxonomic Unit)
<b>pb</b>	pares de base
<b>PBS</b>	Tampão fosfato-salino (do inglês, Phosphate-buffered saline)
<b>PCoA</b>	Análise de Coordenada Principal (do inglês Principal Coordinate Analysis)
<b>PCR</b>	Reação em Cadeia da Polimerase (do inglês, Polymerase Chain Reaction)
<b>PMA</b>	Monoazida de Propídio (do inglês, Propidium Monoazide)
<b>RNA</b>	Ácido Ribonucleico (do inglês, Ribonucleic Acid)
<b>SECOM</b>	Estimação Esparça de Correlação entre Microbiomas (do inglês, Sparse Estimation of Correlations among Microbiomes)
<b>SP</b>	São Paulo
<b>TBE</b>	Tris-Ácido Bórico-EDTA
<b>TCGA</b>	The Cancer Genome Atlas
<b>USP</b>	Universidade de São Paulo
<b>WGS</b>	Sequenciamento Total do Genoma (do inglês, Whole Genome Sequencing)

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	O Estudo do Microbioma Humano	1
1.1.1	Microbioma e Saúde	8
1.1.2	Microbioma e Câncer	10
1.1.3	O Estudo dos Fungos e o Microbioma	11
<b>2</b>	<b>OBJETIVOS</b>	<b>17</b>
2.1	Objetivo Principal	17
2.2	Objetivos específicos	17
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>19</b>
3.1	Casuística	19
3.1.1	Seleção <i>in silico</i> de oligonucleotídeos de interesse para amplificação da região genômica dos fungos	20
3.1.2	Validação dos oligonucleotídeos selecionados	21
3.1.3	Seleção dos protocolos de extração de DNA de acordo com o tipo da amostra	22
3.1.4	Otimização da Reação de Amplificação da região ITS2	23
3.1.5	Eletroforese dos Produtos de PCR da Região ITS2	24
3.1.6	Sequenciamento da Região ITS2 Amplificada por PCR	24
3.1.7	Análise Bioinformática do Sequenciamento da Região ITS2	25
3.1.8	Desenvolvimento de um Modelo Matemático para Correção da Distribuição das Regiões ITS2 Sequenciadas via <i>Metabarcoding</i>	27
3.1.9	Sequenciamento e Análise do DNA total	29
3.1.10	Análises Estatísticas	31
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>33</b>

4.1	Seleção dos oligonucleotídeos para Amplificação da Região ITS2 em Fungos . . . . .	33
4.1.1	Seleção do protocolo de extração de DNA . . . . .	38
4.1.2	Rendimento dos protocolos testados . . . . .	39
4.1.3	Sequenciamento ITS2: Diversidade fúngica recuperada (inter e intra-amostral) . . . . .	40
4.1.4	Sequenciamento WGS: Abundância geral observada . . . . .	41
4.1.5	Influência da variabilidade da região ITS2 no sequenciamento . . . . .	47
4.1.6	Sequenciamento da Região ITS2 . . . . .	55
4.1.7	Resultados - Amostras de Pênis . . . . .	62
4.1.8	Resultados - Amostras de Lavado Gástrico . . . . .	77
<b>5</b>	<b>CONCLUSÕES . . . . .</b>	<b>99</b>
<b>6</b>	<b>REFERÊNCIAS . . . . .</b>	<b>103</b>
	<b>APÊNDICES . . . . .</b>	<b>123</b>
	<b>Apêndice A – PUBLICAÇÕES . . . . .</b>	<b>124</b>
	<b>Apêndice B – OLIGONUCLEOTÍDEOS LEVANTADOS NA LITERATURA . . . . .</b>	<b>136</b>
	<b>Apêndice C – TERCEIRO SEQUENCIAMENTO ITS2 COM 51 AMOSTRAS DE PÊNIS . . . . .</b>	<b>144</b>

# 1 INTRODUÇÃO

## 1.1 O ESTUDO DO MICROBIOMA HUMANO

O termo microbioma humano foi definido por Turnbaugh e colaboradores como o genoma da microbiota que habita o corpo humano (Turnbaugh et al. 2007), que se trata da comunidade de microrganismos comensais, simbióticos e patogênicos que habitam o corpo humano (Lederberg e McCray 2001) (bactérias, arqueobactérias, fungos, protozoários/parasitas e vírus). O corpo humano se assemelha a um ecossistema que apresenta, além de suas próprias células que formam os órgãos e tecidos, bilhões de células de bactérias e outros microrganismos, em menor proporção. Microrganismos habitam praticamente todas as regiões do corpo humano, incluindo o aparelho gastrointestinal, respiratório, urogenital e a pele. Esse ecossistema humano é resultado da coexistência entre a comunidade microbiana e o corpo humano ao longo da evolução. Acreditou-se por muitos anos que as células bacterianas excedem as células humanas de um indivíduo na proporção de 10:1 (Savage 1977), mas estudos recentes propuseram que, na realidade, esta proporção está mais próxima de 1:1 (Sender et al. 2016).

A microbiota humana vem sendo alvo de estudo desde o século XVII, quando Antonie van Leeuwenhoek comparou amostras de suas próprias microbiotas oral e fecal (Leeuwenhoek ; Dobell 1920). Os métodos de cultivo e isolamento foram, por muitos anos, o padrão ouro para a identificação e caracterização dos microrganismos (Grice e Segre 2012). Esta abordagem, ainda em uso, se baseia em uma análise fenotípica das colônias formadas em placas de cultivo, dependendo das condições necessárias para o crescimento do microrganismo. No entanto, várias evidências foram surgindo, demonstrando que apenas uma pequena fração da diversidade total microbiana que existe na natureza pode ser cultivada em laboratório, isso devido a dificuldade em replicar características essenciais do

ambiente que permitem seu crescimento (Stewart 2012).

A partir do avanço das tecnologias moleculares, novas evidências da existência de microrganismos não cultiváveis foram observadas. Ao obter sequências de DNA de amostras ambientais a partir da amplificação por PCR (Reação em Cadeia da Polimerase), seguido do sequenciamento destas sequências, foi possível caracterizar marcadores moleculares filogenéticos independentemente da viabilidade do organismo portador do DNA (Amann et al. 1995). Essas análises revelaram uma enorme diversidade que nunca havia sido verificada através das técnicas de cultivo. Antes do surgimento dessas ferramentas moleculares só haviam sido descritos 11 filos de bactérias (Woese 1987), e após essa inovação o número de filos aumentou para cerca de 90, sendo que a grande maioria dos novos filos não possuía representantes cultivados em laboratório (Woese 1987; Rappé e Giovannoni 2003; Keller e Zengler 2004; Achtman e Wagner 2008).

Os marcadores moleculares filogenéticos são sequências de DNA compartilhadas entre organismos que possuem tanto regiões altamente conservadas, que são utilizadas como alvos para amplificação por PCR, como regiões com alta taxa de mutação, que são utilizadas para diferenciação taxonômica. A ubiquidade de determinados genes os tornam alvos ideais para estudos filogenéticos e classificação taxonômica (Woese 1987). Os marcadores mais utilizados para identificar procariotos são os genes ribossomais do RNA 16S (16S rRNA), que estão presentes com pelo menos uma cópia no genoma dos procariotos (Wang e Qian 2009), e até mesmo em todos os genomas mitocondriais, que perderam a maior parte de seus genes ancestrais ao longo da evolução simbiótica (Gray et al. 1999). O estudo de outros microrganismos utilizam de outras regiões de interesse, como a região de espaçador interno transcrito (ITS) presente nos fungos, localizada entre os genes 18S e o 28S rRNAs, também utilizados no estudo do microbioma de microrganismos eucariontes (Schoch et al. 2012).

Outra técnica desenvolvida para estudar microbiomas é o sequenciamento total do DNA, também chamado de *shotgun*, metagenômica, ou WGS (do inglês

*Whole Genome Sequencing*. Ao invés de amplificar via PCR focar em um locus genômico específico para amplificação, todo o DNA é fragmentado e sequenciado independentemente, sem etapas prévias de amplificação (Sharpton 2014). Diferente da metodologia de *metabarcoding* que informam apenas em relação a classificação taxonômica e na maioria dos casos só alcança o nível de Gênero (Huttenhower et al. 2012; Claesson et al. 2017), as sequências obtidas no WGS, além de serem capazes de fornecer a classificação taxonômica de espécie (Seol et al. 2019; Simon et al. 2019), e em alguns casos até o nível de cepa (Sharpton 2014; Ventolero et al. 2022), também fornecem informações a respeito das funções biológicas, ao obter sequências codificantes dos microrganismos (Song et al. 2018). Todos os grupos de microrganismos podem ser estudados ao utilizar essa metodologia, mas devido ao custo ser mais elevado que a dos marcadores filogenéticos, ainda não utilizado de forma ampla. Apesar disso, alguns grupos de microrganismos são preferencialmente estudados utilizando a metagenômica, como o viroma e o arqueoma (microbioma viral e das arqueias, respectivamente) (Lecuit e Eloit 2013; Borrel et al. 2020).

A metagenômica trata-se do estudo realizado em larga escala de genomas provenientes de comunidades complexas, incluindo microrganismos não cultiváveis, e tem a capacidade de caracterizar a microbiota e determinar seu possível papel no ambiente em que está inserido (Riesenfeld et al. 2004), já a análise baseada em marcadores filogenéticos, como o gene 16S rRNA também é chamada de *metabarcoding* (Taberlet et al. 2012) ou metataxonomia (Marchesi e Ravel 2015). O desenvolvimento da metagenômica foi um passo essencial para a descoberta das relações, sejam elas causais ou não, entre o microbioma humano e diversas doenças, pois estimativas apontam que de 20 a 60% do microbioma humano não é cultivável (Pei et al. 2004; Zhou et al. 2004). Na última década as pesquisas relacionadas com microbioma e metagenômica cresceram exponencialmente (Young 2017), isso devido a evolução das técnicas moleculares independentes de cultivo, e o interesse em descobrir como comunidades complexas de microrganismos

mos podem influenciar a patogênese de diversas doenças e suas implicações na prevenção, diagnóstico e tratamento.

Também é importante considerar a metodologia adotada para a investigação do microbioma e o impacto que a mesma pode gerar no resultado final da análise por sequenciamento de nova geração (NGS). Uma revisão de estudos voltados para a microbiota bacteriana intestinal identificou a alteração da abundância de diferentes grupos a depender das alterações no protocolo de coleta e armazenamento das amostras, mas não foi verificada alteração significativa da diversidade microbiana em relação aos métodos de extração testados (Song et al. 2018). A seleção do tipo de sequenciamento também acarreta em diferentes impactos, como o *metabarcoding*, que a depender da região de interesse selecionada, possui informações diferenciadas quanto a especificidade taxonômica (Rintala et al. 2017). Já para o caso da metagenômica, apesar de não haver seleção de uma única região de interesse, também existem desafios em sua aplicação. O elevado custo ainda é uma das principais barreiras para a adoção dessa tecnologia, além disso, a quantidade de dados geradas é muito superior, o que faz necessário maior investimento para o armazenamento e análise destes dados (Song et al. 2018).

Outro fator que pode implicar em um viés de interpretação dos resultados de estudos do microbioma humano é o fato da maioria dos estudos ser baseada no DNA, que pode não refletir a população residente do determinado local de análise, já que os protocolos de extração do DNA comumente utilizados não diferenciam a viabilidade dos microrganismos (Ercolini 2013). Para solucionar essa questão, outros estudos utilizam de outras técnicas, como o uso da monoazida de propídio (PMA), composto que não penetra a membrana celular e se liga ao DNA livre, evitando que DNA de microrganismos mortos sejam mantidos durante a extração (Nocker et al. 2006; Nguyen et al. 2016; Tantikachornkiat et al. 2016; Li et al. 2017). Também existe a possibilidade de realizar o sequenciamento baseado no RNA, também chamado de meta-transcritômica (Chandler et al. 2015), que apesar de mostrar maior eficiência na manutenção dos microrganismos viáveis

(Li et al. 2017), também apresenta limitações em relação a custo e rendimento da extração, por ser uma molécula de maior instabilidade (Sheridan et al. 1998; Bikel et al. 2015; Bashiardes et al. 2016). Ainda assim, o método baseado no DNA voltado para o *metabarcoding* é o mais comumente utilizado até o momento, principalmente devido ao custo e a ampla base de dados já construída, quando nos referimos ao estudo do microbioma bacteriano (Song et al. 2018).

Além da seleção da metodologia de coleta, extração e tipo do sequenciamento, também existem diferentes abordagens em relação ao processamento das sequências obtidas. Para o *metabarcoding*, a depender da região e microrganismo de interesse a ser analisada, existem ferramentas e bancos de dados apropriados para seu processamento. Existem *pipelines open source* que contemplam todas as etapas de processamento e análise das regiões de interesse, como o QIIME2 (Bolyen et al. 2018), Mothur (Schloss et al. 2009) voltados mais para análise da região 16S, e também o AMPtk (Palmer et al. 2018) desenvolvido para análise das regiões ITS. A princípio, é realizada uma etapa de controle de qualidade, que identifica a região de interesse através dos oligonucleotídeos utilizados durante a etapa de PCR e remoção de sequências com baixa qualidade. Em seguida, é realizada a etapa de *clustering*, que se trata da de-replicação das sequências e o uso de uma metodologia de agrupamento de modo a gerar sequências representativas considerando uma porcentagem mínima de identidade (comumente delimitado em 97%), conhecidas como operational taxon units (OTUs) (Schloss 2010; Knight et al. 2018; Chiarello et al. 2022), realizado pelo algoritmo *Mothur* (Schloss et al. 2009) ou UPARSE (Edgar 2013), por exemplo. Também pode ser utilizada a metodologia de *denoising*, que promove a remoção de sequências erroneamente geradas durante o sequenciamento e distinguem as sequências com variações de até uma base (Overview... ; Jeske e Gallert 2022; Chiarello et al. 2022), nomeadas como *amplicon sequence variants* (ASVs). Para essa abordagem, temos como exemplos os algoritmos DADA2 (Callahan et al. 2016), Deblur (Amir et al. 2017) e UNOISE (Edgar e Flyvbjerg 2015; Edgar 2016b). As duas abordagens

podem ser realizadas por diferentes algoritmos, mas ambas registram a frequência de cada sequência representativa ao longo das amostras em uma tabela *feature table*. Atualmente o método de *denoising* tem ganhado destaque devido a sua maior resolução em diferenciar as sequências, e dessa forma, espera-se maior resolução taxonômica (Boers et al. 2019; Chiarello et al. 2022), e também maior acurácia quanto a medição da diversidade intra-específica (também chamada de diversidade alfa) em relação a abordagem baseada em OTU, apesar de que a medição da diversidade inter-específica (diversidade beta) é mais similar entre as duas metodologias (Nearing et al. 2018; Prodan et al. 2020; Joos et al. 2020; Moossavi et al. 2020).

Para a classificação taxonômica, a metodologia de *metabarcoding* utiliza um algoritmo que realiza a classificação das sequências representativas obtidas (OTU ou ASV). Existem diversos classificadores, como o RDP (Wang et al. 2007), UTX (Edgar ) e SINTAX (Edgar 2016a), CONSTAX (Gdanetz et al. 2017), entre outros. Além dos classificador é necessário um banco com sequências de referência contendo a classificação taxonômica. É necessário selecionar o banco de referência de acordo com a região e microrganismo de interesse. Para a região 16S de bactérias e arqueias temos como exemplo os bancos SILVA (Quast et al. 2012) e Greengenes (McDonald et al. 2012). Já para região para a região 18S de eucariotos, temos também o banco SILVA e para região ITS de fungos e plantas temos o banco UNITE (Nilsson et al. 2018). Com a *feature table* e a classificação taxonômica é possível realizar análises em relação a diversidade alfa e beta. A diversidade alfa avalia a diversidade presente dentro de uma amostra, sendo capaz de quantificar a riqueza de espécies (*richness*), ou a equitabilidade (*evenness*), ou ambas dependendo do índice de diversidade utilizado (Whittaker 1972). A diversidade Beta, por sua vez, avalia a similaridade entre comunidades de habitats diferentes em relação à variação das espécies identificadas nelas (Whittaker 1972). No contexto do microbioma humano, essa medida é utilizada para comparar a população de microrganismos de diferentes indivíduos. Além disso, sempre é

de interesse ao pesquisador investigar se existe algum táxon significativamente diferente em relação a abundância em um grupo de amostras. Para isso é realizada uma análise de abundância diferencial (AD), e existem diferentes algoritmos para esse propósito. Um ponto que deve ser levado em consideração para essa análise é a realização ou não da rarefação das amostras (*subsampling* ou subamostragem). Essa etapa de rarefação é utilizada para corrigir as diferentes coberturas de sequências entre as amostras (Weiss et al. 2017). Ainda não há um consenso em relação a melhor prática a ser adotada, mas comparações estão sendo realizadas entre as abordagens mencionadas (Pollock et al. 2018; Allaband et al. 2019; Nearing et al. 2022). A rarefação é criticada já que a subamostragem remove dados do conjunto amostral, reduz poder estatístico e pode introduzir um viés na análise (Nearing et al. 2022). Algoritmos que utilizam da rarefação são LEfSe, por exemplo (Segata et al. 2011), e aqueles que não fazem uso dessa abordagem são ALDEx2 (Fernandes et al. 2014), ANCOM (Mandal et al. 2015), DESeq2 (Love et al. 2014), edgeR (Robinson e Oshlack 2010), metagenomeSeq (Paulson et al. 2013), entre outros. Em um estudo recente, foram avaliados 14 algoritmos diferentes para análise de AD em 38 *datasets* de sequenciamentos da região 16S e os algoritmos ALDEx2 e ANCOM-II tiveram os resultados mais consistentes entre os diferentes *datasets* e com a maior concordância em relação as outras ferramentas (Nearing et al. 2022).

Apesar da limitação de informação obtida com o sequenciamento do tipo *metabarcoding*, foram criadas ferramentas para buscar contornar essa limitação, como as ferramentas desenvolvidas para prever as funções metabólicas dos microrganismos identificados a partir do perfil dos marcadores filogenéticos identificados. Para as sequências 16S de bactérias, foram desenvolvidas ferramentas como o PICRUST2 (Douglas et al. 2020) e Tax4Fun (Abuhauer et al. 2015). Para fungos, usualmente a caracterização metabólica só era obtida a partir de WGS, já que as ferramentas como o PICRUST2 não apresentam um banco relevante de sequências fúngicas devidamente anotadas (Douglas et al. 2020), mas recente-

mente foi desenvolvida uma nova ferramenta específica para predição metabólica em sequências ITS de fungos chamada FunFun, publicada em um artigo recente ainda não revisto pelos pares (Krivonos et al. 2022).

Todos os fatores mencionados, desde a coleta da amostra até a análise das sequências geradas, são importantes a serem considerados quando surge o questionamento em relação a existência de um microbioma comum entre os seres humanos para uma determinada região do corpo, também chamado de microbioma *core* (Turnbaugh et al. 2007; Turnbaugh et al. 2009; Sharon et al. 2022). O estudo de indivíduos saudáveis é importante para a definição deste microbioma *core*, de modo a ser possível comparar com o microbioma de indivíduos acometidos por doenças.

### 1.1.1 Microbioma e Saúde

A microbiota humana é reconhecida por realizar funções essenciais na manutenção da saúde do hospedeiro, incluindo o metabolismo, desenvolvimento de tecidos, processo inflamatório e a imunidade (Lee e Mazmanian 2010). Alguns dos benefícios proporcionados pela nossa microbiota incluem a síntese de certas vitaminas (como o complexo B e K), a prevenção contra a infecção por patógenos através da competição direta por superfícies corporais, ou pela imunidade induzida por espécies comensais e seu impacto no desenvolvimento normal do sistema imune (Gill et al. 2006). Bactérias da pele têm um papel crucial na prevenção de infecções por espécies patogênicas (Kong e Segre 2012), enquanto bactérias da boca e do trato gastrointestinal tem um importante papel na fisiologia digestiva, estando envolvidas na metabolização de carboidratos e proteínas (Macfarlane e Macfarlane 2012).

Devido a importância que a microbiota apresenta para o hospedeiro é possível entender que alterações em sua composição podem interferir na homeostase existente entre os microrganismos e as células do hospedeiro, influenciando no desenvolvimento de doenças (Turnbaugh et al. 2009; Muegge et al. 2011; Bulga-

relli et al. 2012; Smith et al. 2013). Entretanto, a distinção entre “microbiota saudável” e “microbiota doente” não é simples de se definir. Para responder a essa questão, uma grande iniciativa do Instituto Nacional de Saúde (NIH) foi realizada, o Projeto Microbioma Humano (*Human Microbiome Project*)<sup>1</sup>, com o principal objetivo de verificar como mudanças no microbioma humano estariam associadas com a saúde ou à doenças. Um dos estudos desse projeto demonstrou a presença, em indivíduos saudáveis, de comunidades bacterianas nicho específicas que apresentam variações entre outros indivíduos e que persistem ao longo do tempo em cada indivíduo (Huttenhower et al. 2012).

Outras iniciativas buscaram estabelecer relações causais entre doenças e o microbioma, como foi o caso da relação entre perfis da microbiota oral e doenças específicas, como a periodontite (Faveri et al. 2008), ou a doença infecciosa promovida por *Clostridium difficile* que está associada com a perda da diversidade e função da microbiota gastrointestinal, e que pode ser revertida pela reposição da microbiota através do transplante de microbiota fecal, (Rao e Young 2015). Também foi verificado que doenças inflamatórias intestinais têm maior risco de surgir quando variações genéticas na maquinaria do hospedeiro interagem com a microbiota intestinal, alterando-a (sem que ocorra a colonização por um patógeno) (Hugot et al. 2001; Ogura et al. 2001; Sartor 2008).

Trabalhos recentes começaram a identificar alguns dos mecanismos utilizados pela microbiota que podem influenciar o metabolismo energético do hospedeiro (Greiner e Bäckhed 2016; Parséus et al. 2017), e até mesmo alterar a resposta imune local quando a composição normal da microbiota é alterada, em casos de rinosinusite provocada pela infecção de *Pseudomonas aeruginosa* (Cope et al. 2016). Além disso, verificou-se que a infecção do trato respiratório por rinovírus pode alterar a microbiota, e foi sugerido que isso pode levar a um aumento da suscetibilidade a infecções no resto do trato respiratório, como otite média e pneumonia (Hofstra et al. 2015).

---

<sup>1</sup> <https://hmpdacc.org/>

Como é possível notar, o microbioma apresenta relações com diversas funções do hospedeiro, podendo influenciar direta ou indiretamente o surgimento de diversas doenças, sejam elas causadas por agentes infecciosos ou não. Muitos fatores, como a genética do hospedeiro e características ambientais, influenciam a composição e mudanças dinâmicas na microbiota humana (Sekirov et al. 2010). Dessa maneira, torna-se interessante estudar doenças complexas que são influenciadas por fatores genéticos e ambientais, como o câncer, a fim de revelar a relação entre os microrganismos e o desenvolvimento e a evolução dos tumores.

### 1.1.2 Microbioma e Câncer

O câncer é uma doença genética e influenciada por fatores ambientais. Cerca de 20% das neoplasias estão associadas a microrganismos (Parkin 2006; Martel et al. 2012). A Agência Internacional de Pesquisa em Câncer (IARC) classificou 10 agentes microbianos (7 vírus, 2 parasitas e 1 bactéria) como carcinogênicos ao ser humano. A bactéria *Helicobacter pylori*, os vírus da hepatite B e C (HBV, HCV), e o papilomavírus humano (HPV) juntos são responsáveis por 90% de todos os tipos de câncer gerados por infecções (Martel et al. 2012). A infecção gástrica por *H. pylori* (única bactéria classificada como carcinogênica pela IARC), produz diversos fatores de virulência que podem desregular a proliferação, polaridade e migração celular do hospedeiro.

Apesar de apenas uma bactéria ser classificada como carcinogênica, diversos estudos estão obtendo evidências da contribuição de outros microrganismos no processo de carcinogênese a partir de outros mecanismos, como o estímulo de processos inflamatórios. Esse processo geralmente ocorre quando há o rompimento das barreiras existentes entre as células do hospedeiro e os microrganismos, o que proporciona o estímulo de processos inflamatórios ou imunossupressores, afetando assim, a manutenção da homeostase (Hooper et al. 2012).

Outra hipótese estudada para estabelecer uma associação entre o microbioma e câncer são as perturbações geradas por mudanças na dieta, inflamações,

respostas imunes inatas ou infecções, que podem afetar a composição e a riqueza microbiana, podendo ter alguma relação com o desenvolvimento do tumor (Arthur et al. 2012; Costello et al. 2012; Holmes et al. 2012). Em um recente estudo, desenvolvido por pesquisadores da instituição A.C. Camargo Cancer Center (ACCCC) revelou que o consumo de álcool e tabaco são grandes fatores de risco para o desenvolvimento de carcinoma de células escamosas bucais, e também têm um efeito devastador sobre as bactérias encontradas na mucosa oral de indivíduos saudáveis (Thomas et al. 2014). Outros estudos sugerem que a periodontite, a má higiene bucal e microrganismos orais estão associados com o surgimento de câncer das regiões de cabeça e pescoço (Printz 2013; Schmidt et al. 2014; Chi et al. 2015; Galvão-Moreira e Cruz 2016).

Sendo assim, hoje se reconhece que a microbiota contribui para a carcinogênese, seja aumentando ou diminuindo o risco para o hospedeiro, por três mecanismos: (i) alterando o equilíbrio entre proliferação e morte da célula do hospedeiro, (ii) atuando no equilíbrio do sistema imunológico, e (iii) influenciando no metabolismo de diversas substâncias, podendo gerar compostos pró ou antitumorais (Garrett 2015).

### **1.1.3 O Estudo dos Fungos e o Microbioma**

Os fungos estão presentes em todo o nosso ambiente e participam de processos naturais e industriais como a produção de antibióticos, pães, queijos e bebidas alcoólicas; decomposição de detritos naturais, produção de nutrientes no solo para plantas, entre outros. Do estimado número de 5,1 milhões de diferentes espécies de fungos no mundo, cerca de 300 causam doenças de modo regular nos humanos (Taylor et al. 2001; Blackwell 2011). As doenças causadas por essas espécies vão de infecções superficiais da pele e unhas a infecções invasivas dos pulmões, sangue e cérebro (Brown et al. 2012).

Apesar da importância de seus representantes, o microbioma fúngico, também conhecido por micobioma, é um componente pouco estudado dentro do mi-

crobioma humano, tendo em vista que os fungos compõe uma pequena porção do microbioma humano total ( $\leq 0,1\%$ ) (Qin et al. 2010; Huffnagle e Noverr 2013). Apesar disso, a partir do métodos independentes de cultura utilizando sequenciamento de larga escala foi possível iniciar o descobrimento de fungos comensais presentes no corpo humano e determinar sua relação com a saúde e doenças.

Em uma recente publicação de Nash e colaboradores (2017), foi estudado o micobioma intestinal de indivíduos saudáveis de um dos trabalhos do Projeto Microbioma Humano e, com o sequenciamento da região ITS2, foi verificado que, diferentemente das comunidades bacterianas, as comunidades fúngicas exibiram alta diversidade tanto inter como intra-individual (Nash et al. 2017). O gênero mais abundante encontrado nesses indivíduos saudáveis foi *Saccharomyces*, seguido por *Malassezia* e *Candida*, evidência corroborada pelo sequenciamento do gene 18S rRNA e pelas sequências metagenômicas das mesmas amostras (Nash et al. 2017). Outros estudos buscaram identificar a composição da microbiota fúngica de outros locais do corpo, como pele, cavidade oral, pulmões e também a cavidade vaginal (Wu et al. 2015; Bradford e Ravel 2017; Diaz et al. 2017; Jo et al. 2017; Tipton et al. 2017).

A disbiose da microbiota fúngica, ou seja, qualquer mudança na composição das comunidades comensais residentes em um determinado ambiente em relação às composição encontrada em indivíduos saudáveis (Petersen e Round 2014), pode estar associada com doenças, como o caso do aumento da população de *Candida tropicalis* observado em pacientes com a doença de Crohn, uma doença inflamatória que afeta todo o trato gastrointestinal (Iliev et al. 2012; Sokol et al. 2016). Também foi observada a presença de fungos patogênicos oportunistas do gênero *Phoma* em biópsias de adenomas colorretais, que não estavam presentes em biópsias de tecido adjacente aos adenomas (Luan et al. 2015). Mais de 80% dos casos esporádicos de câncer colorretal são induzidos por esse tipo de adenoma (Jass 2007; Ullman e Itzkowitz 2011). Recentemente a disbiose fúngica foi associada mais diretamente com o tumores, como foi o caso do estudo que

identificou um papel ativo de espécies do gênero *Malassezia* com a progressão do adenocarcinoma pancreático ductal, que migram do intestino para o pâncreas e promovem a disbiose do microbioma do pâncreas (Aykut et al. 2019).

Esses estudos reforçam a importância do estudo dos fungos presentes no microbioma, devido a sua capacidade de estar relacionado com diversos tipos de doença, incluindo o câncer. Apesar de existirem diversos trabalhos sendo realizados atualmente, o conhecimento existente sobre o microbioma e sua relação com doenças ainda é muito limitado, o que incentiva a busca do papel dos fungos no microbioma humano e sua relação com a saúde humana, permitindo assim, descobrir possíveis biomarcadores, preditores de prognóstico ou até mesmo novos tratamentos utilizando a microbiota de indivíduos saudáveis.

Diante da grande diversidade do microbioma humano e de sua possível associação com várias formas de tumores, a pesquisa do microbioma do câncer humano apresenta-se como área de pesquisa extremamente relevante e atual. Embora a associação entre o câncer e bactérias sejam alvo de intensa investigação (Huttenhower et al. 2012; Garrett 2015; Wang e Jia 2016; Coker et al. 2017), a relação entre fungos e o câncer ainda é pouco conhecida, sobretudo, em tumores raros.

#### **1.1.3.1 Tumores de interesse: Câncer Gástrico e Câncer de Pênis**

De particular interesse, alguns tumores como o gástrico (adenocarcinoma gástrico), e o de pênis (carcinoma epidermoide), representam uma oportunidade de investigação, seja pela maior exposição ao ambiente externo, como também por se tratarem de serem raros (no caso do câncer de pênis).

Os carcinomas gástricos (CG) representam a sexta forma mais frequente de câncer e também a quarta maior causa de morte devido a câncer ao redor do mundo (Sung et al. 2021). No Brasil, o câncer gástrico apresentou uma estimativa de 21.230 novos casos em 2020, sendo 13.360 homens e 7.870 mulheres, segundo estimativas do Instituto Nacional do Câncer (INCA). Esse tipo de tumor maligno

é o quarto lugar na incidência entre homens e o sexto, entre as mulheres. Cerca de 65% pacientes diagnosticados com câncer gástrico têm mais de 50 anos, mas em sua maioria atinge homens na faixa etária de 60-70 (INCA, 2020).

Dentre os tipos tumorais, o adenocarcinoma representa 95% dos casos de câncer gástrico (Ajani et al. 2017). Estes são divididos entre cardia e não-cardia de acordo com sua localização no estômago, sendo que a região da cardia é definida pela região da junção gastroesofágica e a região denominada fundo do estômago e a região não-cardia trata-se da região inferior a cardia contemplando as regiões corpo, antro e piloro (Ajani et al. 2017). O adenocarcinoma não-cardia tem relação com gastrite crônica e inflamação do revestimento do estômago, que pode ser causada por uma variedade de fatores ambientais (Rawla e Barsouk 2019). Já a patogênese dos adenocarcinomas do tipo cardia ainda não possui uma definição clara, apesar de que existem duas etiologias consideradas: i) associação com doença do refluxo gastroesofágico e ii) associação com a gastrite provocada pela infecção de *H. pylori* (Mukaisho et al. 2015).

O adenocarcinoma gástrico também é classificado em relação a sua histologia de acordo com a classificação de Lauren, podendo ser intestinal (bem diferenciado) ou difuso (pouco diferenciado) (Rawla e Barsouk 2019). O tipo intestinal é o tipo mais frequente, com maior incidência pessoas do sexo masculino mais velhas e apresenta melhor prognóstico que o tipo difuso (Liu et al. 2013), que tem maior incidência em pessoas mais jovens do sexo feminino (Chon et al. 2017). Apesar disso, no Brasil, estudos relataram uma redução da ocorrência de tumores do tipo intestinal e um aumento do tipo difuso, predominantemente em mulheres com menos de 45 anos de idade, mas sem resultar no aumento da incidência deste tipo tumoral nas populações estudadas (Rampazzo et al. 2012; Amorim et al. 2014).

O diagnóstico de adenocarcinoma gástrico, é frequentemente tardio, já que 80% dos casos são assintomáticos nas etapas iniciais do câncer (Layke e Lopez 2004). Já nos estágios mais avançados os pacientes podem apresentar os sinto-

mas de perda de peso, dor abdominal, náusea e vômitos e úlcera péptica (Layke e Lopez 2004). O diagnóstico é feito pelo exame de endoscopia alta (EDA) ou esofagogastroduodenoscopia (Low et al. 1994), e a maioria dos pacientes se encontram em estágios avançados da doença ao primeiro diagnóstico (Jemal et al. 2010; Murray et al. 2012) e se não tratado, 63% dos casos evoluem para doença incurável em cinco anos (Wild et al. 2014). O prognóstico e tratamento para o quadro de CG é definido pela localização, estadiamento do tumor e número de linfonodos ressecados e acometidos (Medeiros et al. 2021). Por isso, se faz essencial aprimorar os métodos de diagnóstico para essa doença. Em um estudo recente, foi identificada a disbiose do microbioma fúngico da mucosa gástrica tumoral quando comparada com a mucosa gástrica normal dos pacientes avaliados, e identificaram uma maior abundância do gênero *Solicoccozyma* nas amostras tumorais (Zhang et al. 2022).

Já o câncer de pênis (CaPe) é uma doença rara em nações desenvolvidas, correspondendo de 0,4 a 0,6% de todas neoplasias malignas (Ries et al. 2003). No entanto, nos países em desenvolvimento o CaPe continua sendo um problema importante de saúde pública (Parkin 2006). No Brasil, dentre as neoplasias malignas, corresponde a aproximadamente 2% de todos os tumores no homem (Aita et al. 2015; Barreira<sup>1</sup> et al. 2014), podendo alcançar até 10% na região norte e nordeste (Maia et al. 2022). Os casos variam de 2.9 a 6.8 casos por 100.000 habitantes, sendo um dos países com as maiores incidências para esta doença (Monteiro et al. 2021), e no ano de 2020 foram registradas 463 mortes no Brasil (INCA, 2020). Dentre os tipos de CaPe, o carcinoma epidermoide (CE), que também pode ser denominado espinocelular ou escamoso, representa 95% dos casos de câncer peniano (Reis et al. 2010)

A doença do CaPe, assim como CG, é multifatorial e pode apresentar relação com higiene inadequada, fimose em adultos, tabagismo, muitos parceiros sexuais, HPV, principalmente do tipo 16, e também estuda-se a relação do vírus de Epstein-Barr (EBV) com a doença (Afonso et al. 2012; Costa et al.

2013). O HPV já pôde ser detectado em 15% a 71% dos casos de CaPe. Cerca de 74% dos pacientes com CP tem histórico de fimose e também já foi verificado que quando a circuncisão ocorre ainda quando recém nascido, a probabilidade de desenvolvimento da doença é muito baixa (<1%) (Maia et al. 2022). A circuncisão permite a higienização adequada diminuindo a infecção por *Mycobacterium smegatis* (Costa et al. 2013). A alta mortalidade da doença ocorre devido ao atraso na procura ao atendimento de saúde, que costuma ocorrer em média, um ano após o aparecimento dos primeiros sintomas (Pontillo et al. 2011). Indivíduos com baixa escolaridade e condição socioeconômica são mais afetados pela doença, o que também pode justificar o atraso na procura por serviços de saúde e conseqüentemente um diagnóstico precoce (Costa et al. 2013).

Em relação ao estudo do microbioma no contexto do CaPe, já foi verificado que o procedimento de circuncisão promove a redução de bactéria anaeróbicas pró-inflamatórias no tecido (Price et al. 2010). Um trabalho mais recente, ainda no formato de *preprint* comparou o microbioma bacteriano e fúngico de amostras periuretrais pediátricas e pré e pós-circuncisão e verificaram a redução de ordens bacterianas (*Clostridiales*, *Bacteroidales*, *Campylobacterales*) e fúngicas (*Saccharomycetales* e *Pleosporales*) após a circuncisão, que podem estar associadas com ativação de cascatas pro-inflamatórias (Mishra et al. 2022).

Dessa forma, se faz relevante os estudos voltados para o microbioma fúngico, com o intuito de caracterizar essa porção do microbioma humano para diferentes tecidos em diferentes contexto, no caso do atual projeto no contexto do câncer, considerando tumores com ocorrência relevante no Brasil, principalmente tumores raros como o câncer de pênis, que apesar de raros apresentam uma elevada incidência no país quando consideramos o contexto mundial.

## 2 OBJETIVOS

### 2.1 OBJETIVO PRINCIPAL

Desenvolver uma metodologia para identificação de fungos desde a seleção do oligonucleotídeos até o sequenciamento da região de interesse e aplicar o método em amostras humanas de pacientes portadores de adenocarcinoma gástrico e carcinoma epidermóide de pênis e comparar com amostras de indivíduos sem tumor ou amostras de tecido normal adjacente ao tumor.

### 2.2 OBJETIVOS ESPECÍFICOS

- Seleção e validação de oligonucleotídeos para amplificação da região ITS2 baseada em investigação *in silico* de bancos de dados públicos;
- Seleção do protocolo de extração de DNA voltado para microrganismos, com foco nos fungos;
- Estabelecimento e otimização da reação de PCR da região ITS2 utilizando os oligonucleotídeos selecionados;
- Desenvolver uma abordagem analítica-computacional para detecção de fungos em dados de sequenciamento baseado em *amplicons* (região ITS2);
- Desenvolver uma abordagem analítica-computacional para detecção de fungos em dados de sequenciamento baseado em DNA total utilizando dados públicos e aplicar em dados gerados *in house*;
- Padronizar a estratégia de sequenciamento do DNA fúngico baseada em *amplicons*;

- Identificar os fungos presentes em dados de sequenciamento genômico total proveniente de tumores de adenocarcinoma gástrico obtidos do repositório *The Cancer Genome Atlas* (TCGA)
- Sequenciamento DNA total de um subconjunto das amostras de lavado gástrico selecionadas para o projeto, com foco na análise da população fúngica;
- Comparar as populações fúngicas provenientes de amostras do tipo tumoral e normal por meio de análises de diversidade alfa e beta;
- Buscar grupos fúngicos diferencialmente abundantes entre amostras tumorais e normais e também entre outros subgrupos presentes no conjunto amostral.

## 3 MATERIAL E MÉTODOS

### 3.1 CASUÍSTICA

- **Amostras de Pênis:** Ao total foram obtidas 64 amostras de pênis, sendo 27 provenientes do A.C.Camargo Cancer Center (ACCCC), localizado em São Paulo (SP), e 37 da Fundação Centro de Controle de Oncologia (FCECON), localizada em Manaus (AM). Dessas amostras, 54 apresentavam informações clínicas que poderiam ser utilizadas nas análises posteriores. As amostras foram coletadas de peças cirúrgicas de pacientes com carcinoma epidermoide, sendo 46 de tecido tumoral e 8 de tecido normal adjacente ao tumor, que serão utilizadas como "controle" para as análises do microbioma fúngico, além de outros tipos histológicos.
- **Amostras de Lavado Gástrico:** As amostras de lavado gástrico (LG) foram obtidas do projeto temático FAPESP 14/26897-0, que permitiu a criação de uma grande biblioteca de amostras de lavado gástrico. Foram selecionadas 150 amostras para o projeto, contendo amostras originárias de pacientes diagnosticados com adenocarcinoma gástrico, metaplasia intestinal, ou gastrite superficial. A seleção das amostras buscou parear as amostras de acordo com a idade e sexo dos pacientes, em relação ao diagnóstico inicial do exame de endoscopia. Como critério de exclusão foram desconsideradas amostras de pacientes que tivessem feito uso recente de antibióticos a pelo menos um mês da coleta do exame endoscópico com a coleta do lavado gástrico.
- **Amostras de Adenocarcinoma Gástrico do repositório TCGA:** Devido à pandemia do COVID-19, as atividades presenciais referentes a este projeto (extração de DNA, quantificação e amplificação da região ITS2, produção e sequenciamento das bibliotecas de todas as amostras) ficaram

paralisadas entre o final do mês de Março até o início de Setembro de 2020. Decidimos assim, investir no desenvolvimento da análise de dados disponíveis na literatura para complementar o projeto proposto. Foram obtidos do repositório *The Cancer Genome Atlas* (TCGA Archive) <sup>1</sup>, dados de WGS provenientes de pacientes diagnosticados com adenocarcinoma gástrico (utilizando os termos de busca *Disease Type IS Stomach Adenocarcinoma AND Program Name IS TCGA AND Project Id IS TCGA-STAD AND Data Category IS Raw sequencing data AND Experimental Strategy IS WGS*). Ao total, foi realizado o download de dados provenientes de 111 amostras, sendo 82 provenientes de amostras de tumor, 29 de amostras de tecido normal adjacente, sendo que dessas, 25 compartilhavam dados dos dois tipos de materiais.

### 3.1.1 Seleção *in silico* de oligonucleotídios de interesse para amplificação da região genômica dos fungos

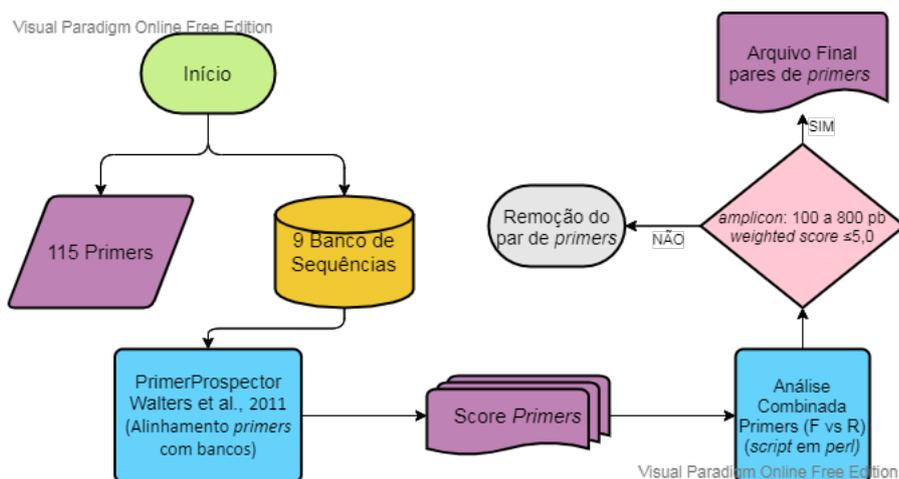
Os oligonucleotídeos analisados foram levantados a partir de uma extensa busca na literatura em bancos de dados voltado para classificação de fungos. O banco selecionado para servir como referência da classificação das sequências ITS foi checado em relação a similaridade com os demais bancos analisados utilizando a ferramenta BLASTn (Altschul et al. 1990; Ye et al. 2006).

Em seguida, os *primers* levantados na literatura foram alinhados individualmente, utilizando o *software* PrimerProspector (Walters et al. 2011), com o banco de referência selecionado. Os arquivos gerados pelo *software* apresentam informações do alinhamento do oligonucleotídeo em relação a cada sequência presente no banco de dados, como posição do oligonucleotídeo na sequência alvo, número de *mismatches*, *gaps*, e também um *weighted score* que leva em consideração todos os scores atribuídos ao alinhamento do oligonucleotídeo atribuindo diferentes pesos para cada fator (1).

<sup>1</sup> <https://portal.gdc.cancer.gov/legacy-archive/search/f>

$$\begin{aligned} \text{WeightedScore} = & (\text{non3'mismatches} \times 0.4) + (\text{3'mismatches} \times 1) + \\ & (\text{non3'gaps} \times 1) + (\text{3'gaps} \times 3) \end{aligned} \quad (1)$$

Os arquivos gerados foram analisados a partir de um *script* desenvolvido na linguagem Perl (Wall et al. 1994) com o objetivo de avaliar todas as combinações de *primers* possíveis de modo a obter pares de oligonucleotídeos capazes de alinhar com a maior quantidade possível de sequências de cada banco de dados, considerando os seguintes fatores: tamanho mínimo do amplicon = 100 pares de base (pb), tamanho máximo do amplicon = 800 pb, *weighted score*  $\leq$  5.0. A figura 1 contém o fluxograma de todo o processo de seleção dos oligonucleotídeos.



**Figura 1 – Fluxograma do método de seleção dos oligonucleotídeos a serem utilizados para amplificação da região de interesse** (Figura gerada com auxílio da ferramenta <https://online.visual-paradigm.com/pt/diagrams/templates/flowchart/>).

### 3.1.2 Validação dos oligonucleotídeos selecionados

Para validar a análise *in silico* o par de *primers* selecionado foi testado para amplificar a região ITS2 de 10 espécies de fungos (*Candida albicans*, *Aspergillus fumigatus*, *Fusarium solani*, *Trichosporon asahii*, *Candida neoformans*, *Histoplasma capsulatum*, *Rhizopus arrhizus*, *Paracoccidioides brasiliensis*, *Sporothrix brasiliensis*, e um fungo patogênico da ordem *Mucorales* pertencente ao grupo conhecido como "Fungo Negro" causador da doença mucormicose (Rahman

et al. 2021), gentilmente cedidas pelo Instituto de Medicina Tropical da Universidade de São Paulo (IMT-USP). Para definir os parâmetros da reação foram testadas diferentes condições para a concentração dos oligonucleotídeos, volume da reação, temperatura de anelamento e número de ciclos.

Os oligonucleotídeos foram validados para as amostras de DNA genômico isolado de fungos utilizando enzima KAPA HiFi HotStart ReadyMix (Roche) 1x, 0,2  $\mu$ M de cada oligonucleotídeo, 1 ng de DNA genômico, e o restante do volume completado para 15  $\mu$ L com H<sub>2</sub>O MiliQ. As amostras foram metidas às seguintes condições otimizadas: desnaturação inicial a 95°C por 5 min, seguidos de 30 ciclos de 98°C por 30 s (desnaturação) e 58°C por 30 s (anelamento), e uma extensão final a 72°C por 5 min.

### **3.1.3 Seleção dos protocolos de extração de DNA de acordo com o tipo da amostra**

A seleção do protocolo para a extração do DNA das amostras pênis foi feita a partir da comparação de 3 *kits* comerciais testados em amostras de fezes de 2 indivíduos cedidas pelo projeto “MOSAIC Challenge” (<https://platform.mosaicbiome.com/challenges/8>). Os *kits* avaliados foram: DNeasy PowerSoil Kit (Qiagen), ZymoBIOMICS DNA Miniprep Kit (Zymo Research) e E.Z.N.A.® Universal Pathogen Kit (Omega). Além destes, avaliamos também os efeitos do produto MetaPolzyme (Sigma) (MAC4L) para maior obtenção de DNA e mais fidedigna representação dos microrganismos da amostra.

O desenho experimental foi organizado da seguinte forma: **i)** para os kits Qiagen e Zymo, 250 *mu*L de cada amostra foram reservados para extração (alíquota 1), em seguida 250 *mu*L foram centrifugados (12.000 g por 5 min). O sobrenadante foi reservado para extração (alíquota 2) e o precipitado foi ressuspendido em 250 *mu*L de PBS 1X, incubado a 80°C por 5 min (para inativar possíveis inibidores enzimáticos), seguida da adição de 5 *mu*L de MAC4L (preparado conforme recomendações do fabricante) e azida de sódio a 0,02% (inibição

de crescimento bacteriano durante a incubação) e novamente incubado a 35°C *overnight* (alíquota 3); **ii**) Para o kit Omega houve uma modificação do desenho experimental devido a falta de reações disponíveis para realizar todas as variações de extrações mencionadas. Para esse protocolo foi realizada a extração com e sem adição da enzima MAC4L apenas da alíquota 1, mas seguindo o mesmo tratamento utilizado para a alíquota 3.

Para as amostras de lavado gástrico, uma diferente metodologia foi adotada: inicialmente é adicionado 500  $\mu\text{L}$  de tampão de digestão *Cell Lysis Solution* (Gentra Puregene Blood kit – Qiagen) e 15  $\mu\text{L}$  de Proteinase K a 20  $\mu\text{g}/\mu\text{L}$  (Invitrogen) em 800  $\mu\text{L}$  de lavado gástrico, seguido de incubação a 300 rpm/55°C por 18 horas. Após essa etapa, os passos restantes são baseados no que utiliza fenol-clorofórmio-álcool isoamílico para extração de ácidos nucleicos (Sambrook et al. 1989).

#### **3.1.4 Otimização da Reação de Amplificação da região ITS2**

Para testar a eficiência do protocolo de PCR para região ITS2, foram utilizadas amostras de DNA já extraídas anteriormente, de diferentes materiais e tumores, sendo estes: lavado gástrico (LG) e biópsia (bx) de pacientes com adenocarcinoma gástrico e *swab* oral (SW) de pacientes com carcinoma epidermoide bucal. Diferentes protocolos foram utilizados, conforme abaixo, sendo o último o que apresentou resultados os melhores dentre os demais avaliados.

- KAPA HiFi HotStart ReadyMix (Roche) 1x, 0,2  $\mu\text{M}$  de cada oligonucleotídeo, 20 ng de DNA extraído, em um volume final de 25  $\mu\text{L}$ ; desnaturação inicial a 95°C por 5 min, seguidos de 30 ciclos de 98°C por 30 s, 58°C por 30 s, seguido de extensão final a 72°C por 5 min;
- KAPA2G Robust HotStart ReadyMix (Sigma-Aldrich) 1x, 0,2  $\mu\text{M}$  de cada oligonucleotídeo, 20 ng de DNA extraído, em um volume final de 25  $\mu\text{L}$ ;

desnaturação inicial a 95°C por 5 min, seguidos de 30 ciclos de 98°C por 30 s, 58°C por 30 s, seguido de extensão final a 72°C por 5 min;

- KAPA HiFi HotStart ReadyMix 1x, 0,2  $\mu$ M de cada oligonucleotídeo, 20 ng de DNA extraído, DMSO a 5%, em um volume final de 25  $\mu$ L; desnaturação inicial a 95°C por 5 min, seguidos de 30 ciclos de 98°C por 30 s e 58°C por 30 s, e uma extensão final a 72°C por 5 min;
- KAPA HiFi HotStart ReadyMix 1x, 0,5  $\mu$ M de cada oligonucleotídeo, 50 ng de DNA extraído, em um volume final de 25  $\mu$ L; desnaturação inicial a 95°C por 5 min, seguidos de 34 ciclos de 98°C por 30 s e 58°C por 30 s, e uma extensão final a 72°C por 5 min.

### **3.1.5 Eletroforese dos Produtos de PCR da Região ITS2**

Inicialmente a checagem do resultado da amplificação da região ITS2 foi feita a partir de 5  $\mu$ L da reação de PCR submetido a eletroforese em gel de agarose 2% diluído em tampão TBE 1x corado com fluoróforo (syberSafe, Invitrogen®, EUA). Para algumas amostras o volume de amostra utilizado não era suficiente para visualização utilizando este protocolo, já que houve pouca ou nenhuma amplificação da região de interesse. Por esse motivo, decidimos realizar a eletroforese em gel de poliacrilamida 4% não denaturante diluído em tampão TBE 1x utilizando um protocolo de coloração por prata adaptado de Bassam e Gresshoff (2007), que se mostrou mais sensível que o método anterior.

### **3.1.6 Sequenciamento da Região ITS2 Amplificada por PCR**

O sequenciamento da região ITS2 amplificada por PCR foi realizada em dois equipamentos diferentes (Ion PGM™ e Ion S5™ - ThermoFisher), sendo o sequenciamento feito na plataforma Ion PGM para o teste para seleção do protocolo de extração do DNA e no Ion S5, o sequenciamento das amostras selecionadas para a validação do método desenvolvido. Inicialmente, o produto de

PCR foi purificado utilizando *beads* magnéticas (AMPure XP, Beckman Coulter®). A quantificação da concentração do DNA purificado foi feita por meio de reação fluorimétrica (Qubit 2.0 - ThermoFisher®). A massa inicial utilizada para o preparo das bibliotecas de sequenciamento variou de 10-100 ng eluída em 79  $\mu$ L de *H<sub>2</sub>O Nuclease-Free*. O preparo seguiu o protocolo de *Ion Plus Fragment Library Kit* (Cat. No.4471252) sem a etapa de fragmentação, com a diferenciação apenas do tamanho do fragmento selecionado. A verificação da integridade das bibliotecas foi realizada no equipamento Bioanalyzer 2100 (Agilent). Para o sequenciamento na plataforma Ion PGM foi selecionado o tamanho máximo do fragmento para 400 pb, já para a plataforma Ion S5, foi selecionado o tamanho máximo do fragmento para 600 pb. Em seguida, a preparação e enriquecimento das ISPs (*Ion Sphere Particles*) seguiu o protocolo Ion PGM™ Hi-Q™ View OT2 Kit (utilizando o equipamento Ion Touch™), ou o protocolo Ion 510™ & Ion 520™ & Ion 530™ Kit (utilizando o equipamento Ion Chef™), para o sequenciamento no Ion PGM e Ion S5, respectivamente. O carregamento do chip de sequenciamento com as ISPs enriquecidas foi feito de forma manual para o sequenciamento na plataforma Ion PGM (Ion Chip 316™), e de forma automatizada (Ion Chef), para o sequenciamento no Ion S5 (Ion Chip 520™).

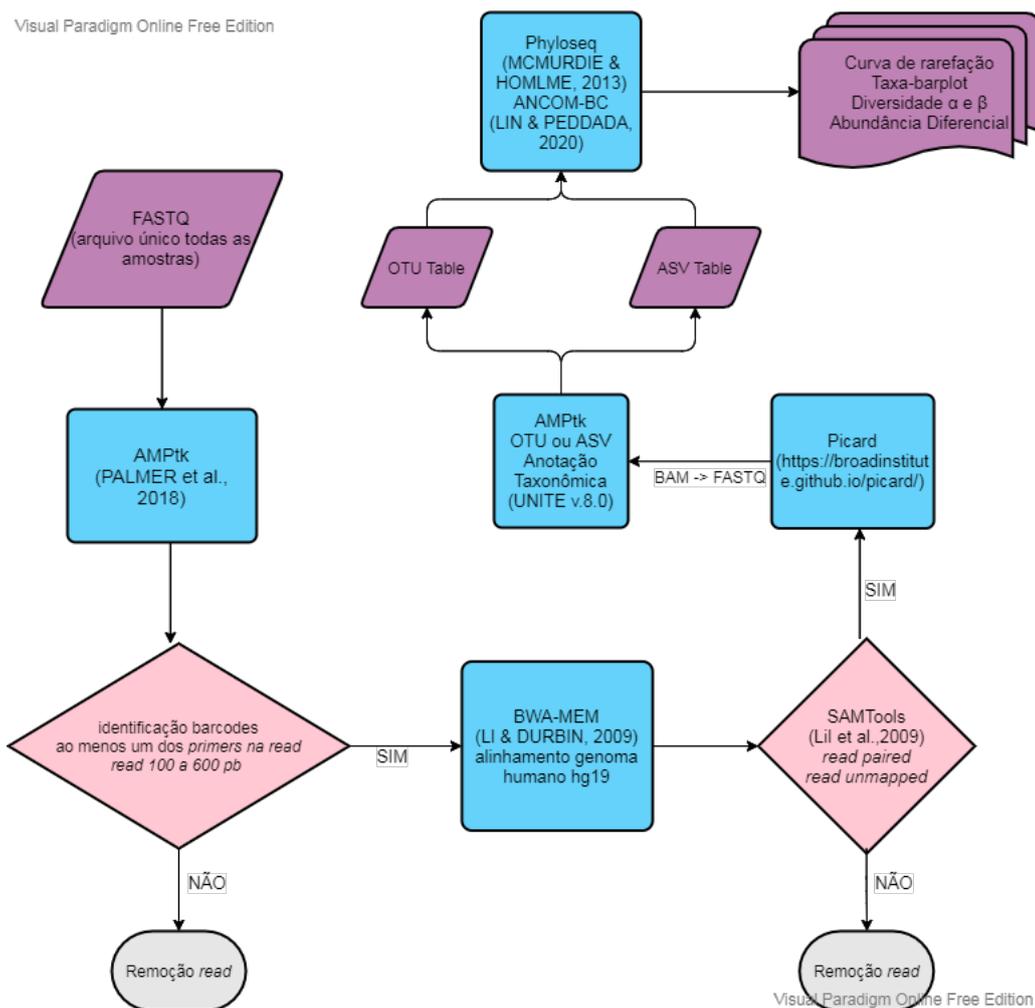
### **3.1.7 Análise Bioinformática do Sequenciamento da Região ITS2**

A análise de bioinformática foi realizada através de um *pipeline* desenvolvido durante este projeto 2. Os dados brutos do sequenciamento foram obtidos no formato FASTQ, sem a remoção dos *barcodes* que identificam cada amostra. O *software* AMPtk (Palmer et al. 2018) foi utilizado para as etapas de identificação das sequências (reconhecimento dos *barcodes*), seleção de sequências de 100 a 600 pb com pelo menos um dos oligonucleotídeos utilizados na reação de PCR. Em seguida, as sequências resultantes deste pré-processamento são alinhadas com o genoma humano utilizando a ferramenta BWA-MEM (Li e Durbin 2009), seguida da extração de sequências que não foram alinhadas ao

genoma do hospedeiro pelo *software* SAMTools (Li et al. 2009). Antes de realizar a etapa de clusterização das sequências, é necessária a re-conversão dos dados para o formato FASTQ, o mesmo é realizado através da ferramenta Picard <https://broadinstitute.github.io/picard/>. A etapa de clusterização foi realizada a nível de OTUs com 97% de identidade, ou ASVs, utilizando o *software* USEARCH (Edgar 2010) (versão 9.2.64) implementado no *software* AMPtk. Em seguida, foi realizado um filtro com valor 0.005 (valor padrão do AMPtk) para reduzir o *index-bleed* (sequências erroneamente identificadas em outras amostras), e classificação taxonômica utilizando os bancos de sequências ITS2 customizados a partir dos *primers* utilizados na reação de PCR utilizando como referência o banco de sequências ITS UNITE (versão 8.0) (Nilsson et al. 2018). Para avaliar a manutenção das amostras na análise, foi utilizada a curva de rarefação (função *ggrare*<sup>2</sup>), para verificar a saturação do número de ASVs obtidas. Em seguida, realizamos um filtro de abundância mínima de 0,001%, com o objetivo de reduzir o número de contaminantes do sequenciamento (Jeske e Gallert 2022), utilizando o pacote *phyloseq* (McMurdie e Holmes 2013). Análises posteriores utilizando esses resultados foram realizadas utilizando o pacote *phyloseq*, *tidyverse* (Wickham et al. 2019), *dplyr* (Wickham et al. 2015), *EnvStats* (Millard 2013) e *ggpubr* (Kassambara 2020), implementados na linguagem R (versão 4.0.2).

---

<sup>2</sup> <https://rdr.io/github/gauravsk/ranacapa/man/ggrare.html>



**Figura 2 – Fluxograma do *pipeline* de processamento das leituras obtidas do sequenciamento da região ITS2** (Figura gerada com auxílio da ferramenta <https://online.visual-paradigm.com/pt/diagrams/templates/flowchart/>).

### 3.1.8 Desenvolvimento de um Modelo Matemático para Correção da Distribuição das Regiões ITS2 Sequenciadas via *Metabarcoding*

Para o desenvolvimento do modelo, foram utilizados dados de sequenciamentos da região ITS2 de amostras controle, ou comunidades *mock*, compostas por diferentes espécies fúngicas, obtidas na literatura e também realizado durante o projeto. As comunidades *mock* criadas neste projeto utilizaram o DNA genômico isolado de 4 espécies fúngicas cedidas pelo IMT com tamanhos da região ITS2 variando de 260 a 350 pb aproximadamente, incluindo os oligonucleotídeos. As espécies selecionadas foram: *Fusarium solani*, apresentando a região

ITS2 com cerca de 240 pb (FS), *Trichosporon asahii*, com cerca de 260 pb (TA), *Cryptococcus neoformans* com aproximadamente 280 pb (CN) e *Paracoccidioides brasiliensis* com aproximadamente 300 pb (PB). Cada amostra foi composta por 10 ng de DNA genômico isolado, utilizando diferentes massas de cada DNA fúngico selecionado.

As três comunidades *mock* criadas receberam diferentes denominações conforme a sua composição: **Mock Iguais** - 2,5 ng de DNA genômico de cada espécie (25% para cada espécie); **Mock Crescente** - 1 ng FS, 2 ng TA, 3 ng CN e 4 ng PB (10, 20, 30 e 40%, respectivamente); **Mock Decrescente** - 4 ng FS, 3 ng TA, 2 ng CN e 1 ng PB (40, 30, 20 e 10%, respectivamente). Além destas, também foi utilizado 10 ng de uma comunidade *mock* comercial (ZymoBIOMICS *Microbial DNA Community Standard*) que tem em sua composição o DNA genômico de 2 espécies fúngicas, *Saccharomyces cerevisiae* e *Cryptococcus neoformans* (cada espécie representando 2% do total da comunidade *mock*). A reação de PCR da região ITS2, preparação das bibliotecas para sequenciamento, sequenciamento na plataforma Ion PGM e análise dos dados foram realizados conforme o descrito nos itens 3.1.4 e 3.1.6.

Para a criação do modelo matemático, utilizamos um *script* desenvolvido em Python para obter das sequências ITS2 classificadas pelo *pipeline* as informações dos *amplicons* que podem afetar a reação de amplificação por PCR, como o tamanho da sequência, porcentagem de bases GC e o número de sequências homopoliméricas. A metodologia utilizada foi a de análise multivariada conhecida como Lasso (*least absolute shrinkage and selection operator*), um método de regressão que emprega regularização e seleção de variáveis para melhorar a precisão e a interpretabilidade do modelo obtido (Kukreja et al. 2006). O algoritmo foi alimentado com dados de contagem total de sequências nas amostras, contagem de sequências correspondentes a cada OTU, tamanho do *amplicon*, conteúdo GC e contagens de homopolímeros de 5, 6 ou 8 pb (todos em  $\log_2$ ), e buscou estimar através de uma regressão linear, as frequências relativas dos OTUs nas amostras

sequenciadas.

### 3.1.9 Sequenciamento e Análise do DNA total

Foi realizado o sequenciamento do DNA total de todas as amostras de pênis (64) e de 90 amostras de lavado gástrico, das 150 selecionadas para o projeto. A partir de 100 ng de DNA de cada amostra realizamos a preparação das bibliotecas de sequenciamento, utilizando o protocolo *Nextera DNA FlexLibrary Prep* (Illumina) e o sequenciamento foi realizado na plataforma Illumina NextSeq 550, 2x 150 pb.

Inicialmente, para a análise dos dados obtidos do repositório TCGA, foi desenvolvido um *pipeline* utilizando o *software* Kraken2 (Wood et al. 2019) para anotação taxonômica das sequências obtidas. Inicialmente, as sequências humanas foram removidas dos arquivos BAMs (*Binary Alignment Map* utilizando o *software* SAMTOOLS (*flag=5*) (Li et al. 2009). As sequências restantes foram submetidas a classificação taxonômica pelo *software* Kraken2 utilizando os parâmetros *-minimum-base-quality=10*, *-confidence=0.1*, e o banco baseado no repositório *Nucleotide* (nt) (download realizado dia 10/11/2020). A análise da tabela de frequências obtida do *software* Kraken2 contendo o número de frequências por táxon classificado (considerando diferentes níveis taxonômicos) foi realizada utilizando os pacotes implementados em R *tidyverse*, *dplyr*, *ggpubr*, *EnvStats*, *jbebgler/tidyheatmap*<sup>3</sup>. Para análise diferencial dos táxons anotados foi utilizado o pacote *DESeq2* (Love et al. 2014), também implementado em R.

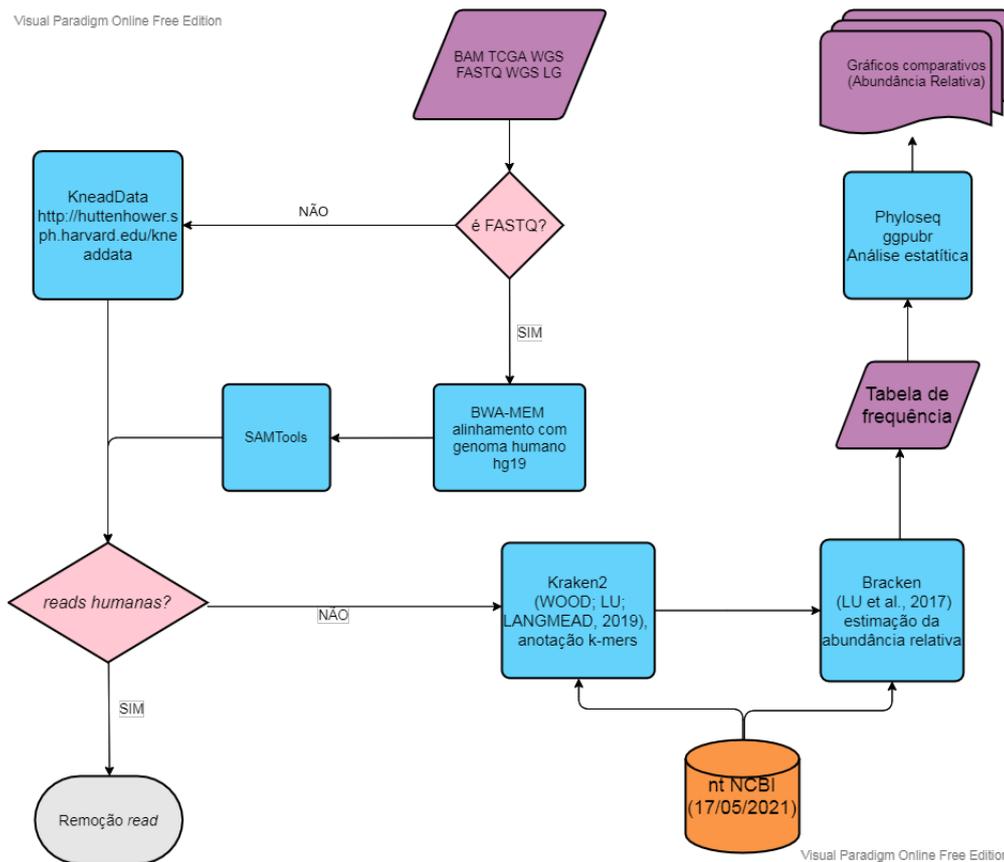
Também foi desenvolvido um segundo *pipeline*, utilizado para o processamento dos dados obtidos do sequenciamento de DNA total das amostras de lavado gástrico e de pênis, que compreende a mesma lógica, mas utiliza diferentes ferramentas para realizar o processamento das sequências<sup>3</sup>. Este *pipeline* foi utilizado em outro projeto em vigência do laboratório, que tem como alvo principal a análise da população bacteriana das amostras, mas é possível

<sup>3</sup> <https://github.com/jbebgler/tidyheatmap>

utilizar os dados processados para a análise da população fúngica. Os arquivos FASTQ gerados no sequenciamento tem as sequências do hospedeiro removidas através do *software* KneadData <sup>4</sup> tendo o genoma humano como referência (GRCh37 - hg19), seguida da classificação taxonômica das sequências remanescentes pelo *software* Kraken2 (Wood et al. 2019), com seus parâmetros padrões utilizando o banco PlusPF (*download* realizado em 17/05/2021), contendo o genoma humano, de arqueas, bactérias, vírus, protozoários e fungos (<https://benlangmead.github.io/aws-indexes/k2>). Em seguida, a tabela de frequência gerada é normalizada utilizando o *software* Bracken, método estatístico que computa a abundância dos indivíduos de um determinado nível taxonômico anotados em uma amostra (Lu et al. 2017).

---

<sup>4</sup> <http://huttenhower.sph.harvard.edu/kneaddata>



(Figura gerada com auxílio da ferramenta <https://online.visual-paradigm.com/pt/diagrams/templates/flowchart/>).

**Figura 3 – Fluxograma do *pipeline* de processamento das *reads* obtidas no sequenciamento WGS.** As leituras obtidas são submetidas a uma série de etapas de processamento para gerar os dados que contêm as informações taxonômicas do microbioma. Este *pipeline* foi aplicado nas sequências obtidas do sequenciamento do DNA total das amostras do TCGA e das amostras sequenciadas no projeto

### 3.1.10 Análises Estatísticas

Para análise dos dados metagenômicos utilizamos o pacote *phyloseq* no programa RStudio (versão 1.1.453; RStudio: *Integrated Development for R*. RStudio, Inc., Boston, MA, EUA), utilizando a linguagem R (versão 4.2.0). Assumindo a não-normalidade da amostra foi utilizado o teste U de Mann-Whitney para avaliar a existência de diferenças estatisticamente significativas entre os grupos testados, e a diversidade alfa (Observed, Shannon e Simpson). E em relação à diversidade beta (índice de Bray-Curtis e distâncias *unweighted* e *weighted* UniFrac), foi usado o teste multivariado *Analysis of Similarities* (ANOSIM) (Clarke

1993). Para análise de abundância diferencial utilizamos a metodologia *Analysis of Compositions of Microbiomes with Bias Correction* (ANCOM-BC) (Lin e Peddada 2020) e para a análise de correlação entre os microrganismos avaliados, utilizamos a metodologia *Sparse Estimation of Correlations among Microbiomes* (SECOM), capaz de detectar relações lineares e não lineares entre dois táxons dentro de uma amostra ou entre diferentes amostras.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 SELEÇÃO DOS OLIGONUCLEOTÍDEOS PARA AMPLIFICAÇÃO DA REGIÃO ITS2 EM FUNGOS

Ao todo foram levantados 115 oligonucleotídeos (Apêndice B) e 9 bancos de sequências ITS contendo diferentes composições quanto a região (Tabela 1). A tabela 2 apresenta o resultado do alinhamento feito pelo *software* BLASTn das sequências presentes no banco UNITE contra as sequências presentes nos demais bancos levantados. Com exceção dos bancos Findley-ITS (Findley et al. 2013), SIS (Usyk et al. 2017), SILVA-18S (Quast et al. 2012) e RDP-28S (Cole et al. 2013) onde o banco UNITE apresenta similaridade de 58.1%, 29.7%, 12% e 2,9% respectivamente, foi possível observar que o banco UNITE contém a grande maioria das sequências presentes nos outros bancos e por essa razão foi selecionado como banco de referência para as demais análises (Tabela 2). É importante ressaltar que a baixa similaridade entre os banco SILVA-18S e RDP-28S está no fato dos bancos mencionados serem compostos pelas regiões conservadas adjacentes à região ITS.

**Tabela 1** – Bancos de sequências utilizados para classificação taxonômica de fungos

Banco de dados	Versão	Data de Acesso	Sequências	Referência
Findley	-	08/2018	23456	(Findley et al. 2013)
NCBI-ITS	-	08/2018	9166	(Schoch et al. 2014)
ntF-ITS1	-	08/2018	13949	(Motooka et al. 2017)
RDP-28S	Release 11	08/2018	14725	(Cole et al. 2013)
SILVA-18S	v.132	08/2018	55145	(Quast et al. 2012)
SIS	-	08/2018	5789	(Usyk et al. 2017)
THF	v1.6.1	08/2018	3982	(Tang et al. 2015)
UNITE	v.8.0	02/2019	71042	(Nilsson et al. 2018)
Warcup	v.2	08/2018	17878	(Deshpande et al. 2016)

**Tabela 2** – Similaridade entre o banco UNITE (v.8.0) e os demais bancos de sequências para classificação taxonômica de fungos. Foi utilizado o *software* BLASTn para realizar o alinhamento das sequências do banco UNITE contra as sequências dos demais bancos avaliados.

Banco de dados	Sequências (totais)	Alinhadas - UNITE	% de Similaridade
Findley	23456	13631	58.1
NCBI-ITS	9166	8136	88.8
ntF-ITS1	13949	11547	82.8
RDP-28S	14725	1763	12.0
SILVA-18S	55145	1586	2.9
SIS	5789	1717	29.7
THF	3982	3433	86.2
Warcup	17878	17516	98.0

Após a seleção do banco UNITE os 115 *primers* levantados da literatura foram alinhados individualmente com todas as sequências fúngicas contidas no banco (versão 8.0) utilizando o *software* PrimerProspector. A partir dos arquivos de alinhamento gerados pelo *software* foi realizada uma comparação generalizada com o objetivo obter pares de oligonucleotídeos capazes de alinhar com a maior quantidade possível de sequências presentes no banco de referência, respeitando os seguintes fatores: tamanho do *amplicon* variando de 100 a 800 pb e score de alinhamento (*weighted score*)  $\leq 5.0$  (detalhes descritos no item 3.1.1).

Dentre os 10 melhores pares de *primers*, se encontrava o par 5.8Sf\_khot e 28S-1r (Tabela 3), desenvolvido com mínima homologia ao gene rRNA humano e capaz de gerar *amplicons* com grande informação filogenética (Khot et al. 2009). Por não haver uma grande diferença de sequências recuperadas entre esse par de *primers* (46151 de 71042 sequências) e aquele que recuperou a o maior número de sequências (46531 de 71042 sequências), e também pelo alinhamento dos *primers* junto ao banco UNITE não apresentarem grandes quantidades de *mismatches* e *gaps*, principalmente na região 3' do *primer* (Figura 2), optamos por selecionar o par de oligonucleotídeos 5.8Sf\_khot (5'-GTGAATCATCGARTCTTTGAAC-3')/28S-1r (5'-TATGCTTAAGTTCAGCGGGTA-3') que tem como alvo a região ITS2.

Também foi investigada a diversidade taxonômica dos possíveis fungos a

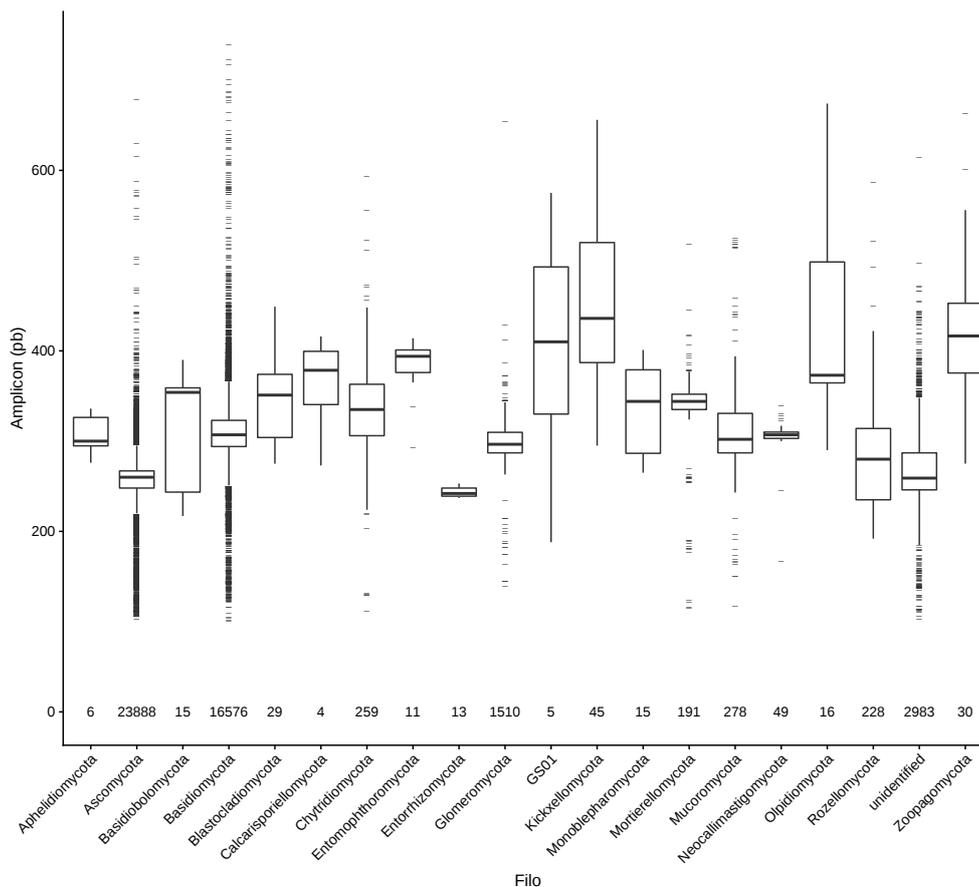
serem amplificados pelos *primers* selecionados. Ao menos 50% de todos os filamentos do reino Fungi presentes no banco UNITE (versão 8.0) foram recuperados (Tabela 4). Além disso, verificamos a variação dos tamanhos dos possíveis *amplicons*, tendo em vista que a região ITS, diferentemente do gene 16S rRNA dos procariotos, apresenta uma alta variabilidade no tamanho em pb (Palmer et al. 2018). A Figura 4 apresenta a variação da região ITS2 recuperada pelos *primers* selecionados dos filamentos com maior número de sequências (*Ascomycota*, *Basidiomycota* e *Glomeromycota*). As sequências recuperadas do filo *Ascomycota* apresentam menor tamanho em relação aos demais filamentos. Essa característica de variabilidade do tamanho da região ITS2 afeta diretamente a reação de PCR e também o sequenciamento, de modo que as menores regiões são mais amplificadas do que as regiões maiores, o que se torna um viés na análise da abundância das espécies encontradas nas amostras após o sequenciamento (Filippis et al. 2017).

**Tabela 3** – Lista dos dez pares de oligonucleotídeos mais abrangentes em relação ao banco UNITE

<b>Primer F</b>	<b>Primer R</b>	<b>Sequências Recuperadas</b>	<b>% banco UNITE</b>
ITS3KYO2f	28S-1r	46531	65.5
ITS3f	28S-1r	46500	65.5
ITS3KYO1f	28S-1r	46491	65.4
58A2f	28S-1r	46477	65.4
ITS3mix15f	28S-1r	46475	65.4
ITS3Rf	28S-1r	46370	65.3
fITS7f	28S-1r	46360	65.3
58A1f	28S-1r	46344	65.2
5.8SRf	28S-1r	46303	65.2
5.8Sf_khot	28S-1r	46151	65.0

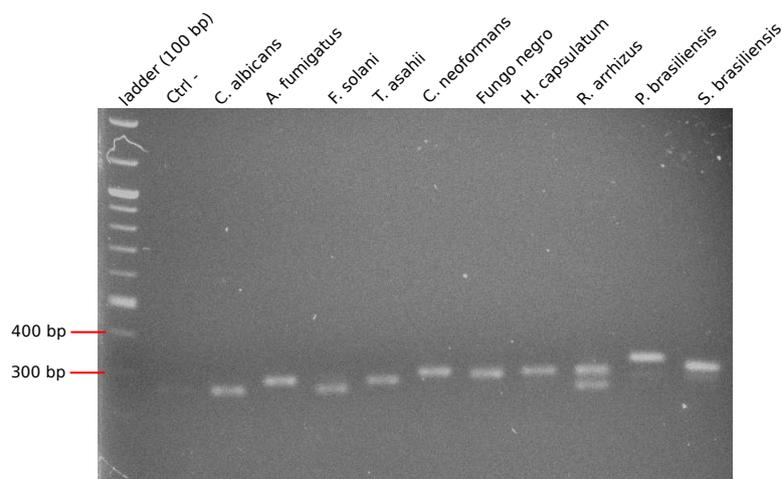
**Tabela 4** – Porcentagem de sequências recuperadas do banco UNITE (v.8.0) pelo par de primers 5.8Sf\_khot/28S-1r em relação ao nível taxonômico (Reino e Filo, representados por "k" e "p", respectivamente).

Nível Taxonômico	Sequências Amplificadas	Sequências Totais	%
k__Fungi	46151	71042	65.0
p__Ascomycota	23888	37077	64.4
p__Basidiomycota	16576	26048	63.6
p__unidentified	2983	4395	67.9
p__Glomeromycota	1510	1750	86.3
p__Mucoromycota	278	518	53.7
p__Chytridiomycota	259	391	66.2
p__Rozellomycota	228	251	90.8
p__Mortierellomycota	191	297	64.3
p__Neocallimastigomycota	49	75	65.3
p__Kickxellomycota	45	50	90.0
p__Zoopagomycota	30	32	93.8
p__Blastocladiomycota	29	39	74.4
p__Olpidiomycota	16	20	80.0
p__Monoblepharomycota	15	23	65.2
p__Basidiobolomycota	15	24	62.5
p__Entorrhizomycota	13	13	100.0
p__Entomophthoromycota	11	19	57.9
p__Aphelidiomycota	6	6	100.0
p__GS01	5	7	71.4
p__Calcarisporiellomycota	4	6	66.7



**Figura 4 – Variabilidade do tamanho dos amplicons (bp) recuperados pelo par de primers selecionados.** Resultado *in silico* da distribuição em pares de base dos amplicons gerados pelo par de *primers* 5.8Sf\_khot/28S-1r em relação aos filios do Reino Fungi presentes no banco UNITE (v.8.0). Os valores acima de cada nome de filo representam o número de seqüências recuperadas pelo par de primers escolhido.

Em seguida, o par de *primers* foi encomendado para síntese (IDT) e validado a partir da amplificação da região ITS2 dos DNAs genômicos isolados de fungos de interesse clínico cedidos pelo IMT-USP, utilizando as condições iniciais de PCR descritas no item 3.1.2. É possível observar que todos os DNAs genômicos foram amplificados e também a variabilidade da região ITS2 (Figura 5), verificada pela análise *in silico* e corroborada pelos achados na literatura citado anteriormente.



**Figura 5 – Eletroforese das reações de PCR utilizando o par de *primers* 5.8Sf\_khot/28S-1R. Amplificação da região ITS2 de 10 espécies fúngicas provenientes de DNAs genômicos isolados cedidos pelo IMT-USP.**

#### 4.1.1 Seleção do protocolo de extração de DNA

Tendo selecionado e validado os *primers*, a próxima etapa foi a seleção do protocolo de extração de DNA das amostras. Este passo é de particular importância principalmente para o estudo de fungos onde a parede celular de certas espécies pode apresentar resistência à lise, o que ocasionaria uma representação alterada das mesmas. Devido aos diferentes tipos de material de origem das amostras (lavado gástrico e tecido peniano) foram utilizados protocolos diferentes para cada tipo de amostra.

Outros estudos realizados na instituição relacionados ao microbioma gástrico (mais especificamente o microbioma bacteriano) já haviam estabelecido um protocolo de extração de DNA total eficiente para esse tipo de material. Como descrito no item 3.1.3, este protocolo não apresenta uma etapa de lise mecânica, mas se mostrou eficaz para extração do DNA fúngico presente nesse tipo de material, resultado verificado em um artigo publicado pelo grupo durante o período do projeto (Albuquerque et al. 2022), em que o material fúngico e bacteriano foram quantificados por PCR quantitativo (qPCR). Acreditamos que este protocolo relativamente simples foi bem sucedido devido ao material ser naturalmente

exposto ao pH ácido do estômago, que pode ter contribuído com o rompimento das estruturas celulares fúngicas.

No caso das amostras de pênis o protocolo de extração utilizado foi diferente, tendo em vista a diferença do material. Pela dificuldade em se obter amostras (optamos por selecionar o protocolo de extração a partir da extração de DNA de amostras de fezes, um tipo de amostra com grande riqueza de microrganismos e que em geral apresenta maiores dificuldades durante a etapa de amplificação do DNA, devido à presença de inibidores da DNA polimerase usada durante a PCR (Bourgeois et al. 2019). As amostras de fezes foram cedidas pelo projeto “MOSAIC Challenge”. Este experimento teve como intuito comparar os protocolos em diferentes quesitos:

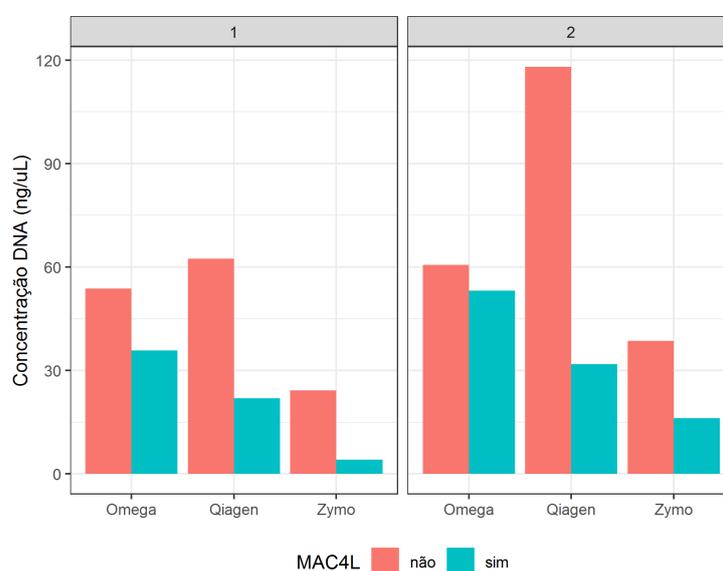
- Rendimento do protocolo
- Sequenciamento ITS2: Diversidade fúngica recuperada (inter e intra-amostral)
- Sequenciamento WGS: Abundância geral observada

#### **4.1.2 Rendimento dos protocolos testados**

Foram testados três kits comerciais usados para extração de DNA de diferentes microrganismos (bactérias, protozoários, fungos, vírus, entre outros), sendo eles: Qiagen, Zymo e Omega, na presença ou ausência do complexo enzimático MAC4L, complexo de seis enzimas com a função de potencializar a etapa de lise celular dos microrganismos, para maior obtenção de DNA e mais fidedigna representação dos microrganismos da amostra (protocolo detalhado no item 3.1.3). Verificamos que, ao contrário do esperado, a adição do mistura de enzimas MAC4L resultou em redução da eficiência da extração de DNA para todos os protocolos comerciais utilizados, principalmente para o protocolo Qiagen, considerando a concentração de DNA extraído e que protocolo da marca Zymo apresentou a menor eficiência dentre os protocolos testados, enquanto os Omega

e Qiagen tiveram rendimento acima de  $50 \text{ ng}/\mu\text{L}$  para as duas amostras (Figura 6).

Em seguida, a região ITS2 foi amplificada, utilizando os *primers* selecionados (5.8Sf\_khot/28S-1r) e as condições de PCR estabelecidas durante a validação (3.1.2). Foi possível observar que apenas as amostras submetidas ao protocolo Omega com adição da mistura enzimática MAC4L apresentaram um padrão diferente de amplificação em relação aos outros protocolos (Figura ??), mais um indício de algum fator neste protocolo que prejudicou a extração.



**Figura 6 – Eficiência de diferentes protocolos de extração de DNA de material fecal provenientes de dois indivíduos do projeto “MOSAIC Challenge”.** A eficiência dos diferentes protocolos de extração utilizados foi avaliada através da quantificação da concentração de DNA medido em  $\text{ng}/\mu\text{L}$  pelo fluorômetro Qubit 2.0 (Invitrogen). As amostras estão divididas por amostra de fezes utilizada (1 e 2), por marca de protocolo utilizado (Omega, Qiagen ou Zymo), e também pela ausência ou presença do complexo enzimático MAC4L (ausência: vermelho, presença: azul).

#### 4.1.3 Sequenciamento ITS2: Diversidade fúngica recuperada (inter e intra-amostral)

O preparo das bibliotecas, sequenciamento e análise foram realizados de acordo com o item 3.1.6, para a plataforma Ion-PGM. Inicialmente, foi possível verificar que as amostras sequenciadas alcançaram um nível de saturação de OTUs com uma quantidade baixa de *reads* (aproximadamente 2500 *reads*) (Fi-

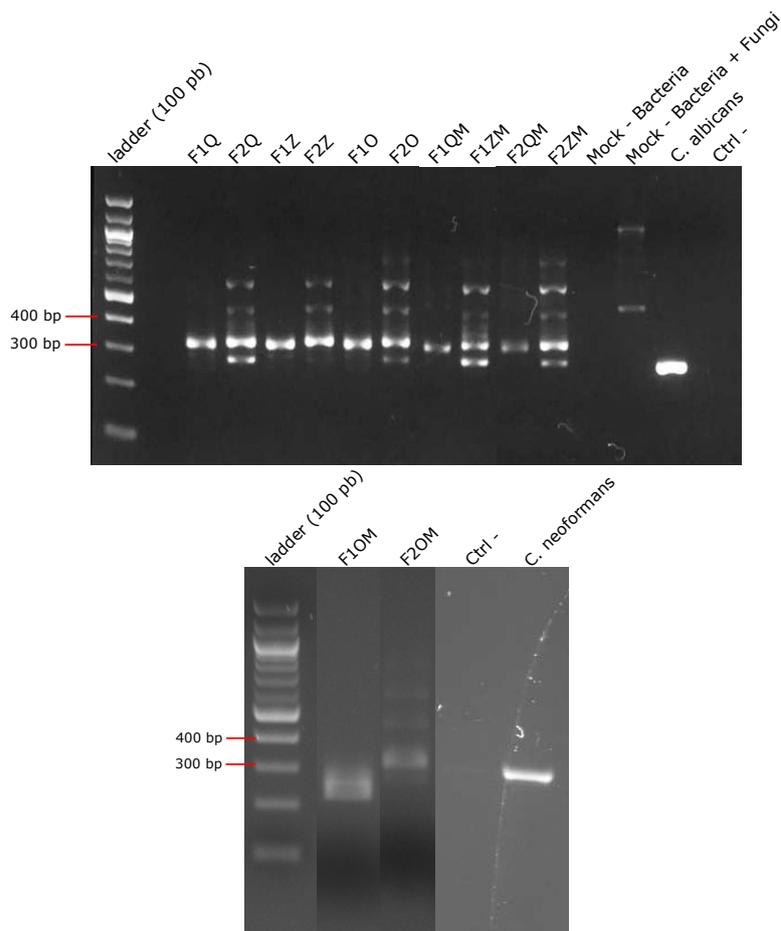
gura 8). É visível que, mesmo em amostras cujo processamento pelo *software* AMPtk mantiveram um número menor que 5000 *reads*, o nível de saturação foi alcançado e por isso as mesmas não foram excluídas da análise. Esse resultado está de acordo com valores reportados por estudos anteriores e com o fato de existir uma menor riqueza de fungos no microbioma humano, não sendo necessária uma grande profundidade de sequenciamento para se ter uma visão abrangente do microbioma fúngico (Nash et al. 2017; Aykut et al. 2019).

Ao comparar a diversidade intra-individual (diversidade alfa) aos níveis de OTU e gênero, não foi verificada uma diferença significativa tanto para os diferentes protocolos testados (Figura 9), como para presença ou ausência de MAC4L (Figura 10). Já a comparação da diversidade interindividual (diversidade beta), mostrou que a o perfil da população fúngica, a nível de gênero, é semelhante entre as amostras extraídas com diferentes protocolos, com exceção das amostras extraídas com o kit Omega na presença de MAC4L (Figura 11). Estas se mostraram mais semelhantes ao controle negativo do que as demais amostra, ao verificar o resultado da Análise de Coordenada Principal (PCoA) das métricas de Bray-Curtis, que se baseia na frequência de cada OTU, e *Weighted Unifrac*, que utiliza a distância filogenética entre as sequências considerando frequência de cada OTU .

Estes resultados indicam que qualquer um dos três protocolos testados, juntamente com o uso dos oligonucleotídeos selecionados neste trabalho, foi capaz de produzir uma diversidade semelhante de fungos, com exceção da combinação do protocolo Omega com o complexo enzimático MAC4L. Também houveram indícios de que o uso da mistura enzimática MAC4L não proporcionou um aumento da diversidade intra-individual (diversidade alfa).

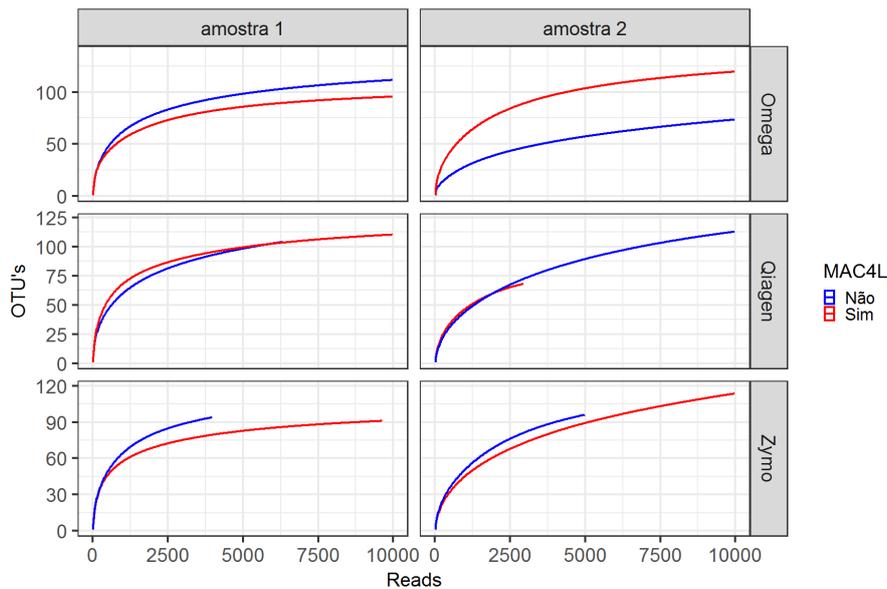
#### 4.1.4 Sequenciamento WGS: Abundância geral observada

Para concluir a comparação entre os protocolos testados, foi realizado o sequenciamento genômico total a partir do DNA total extraído das amostras



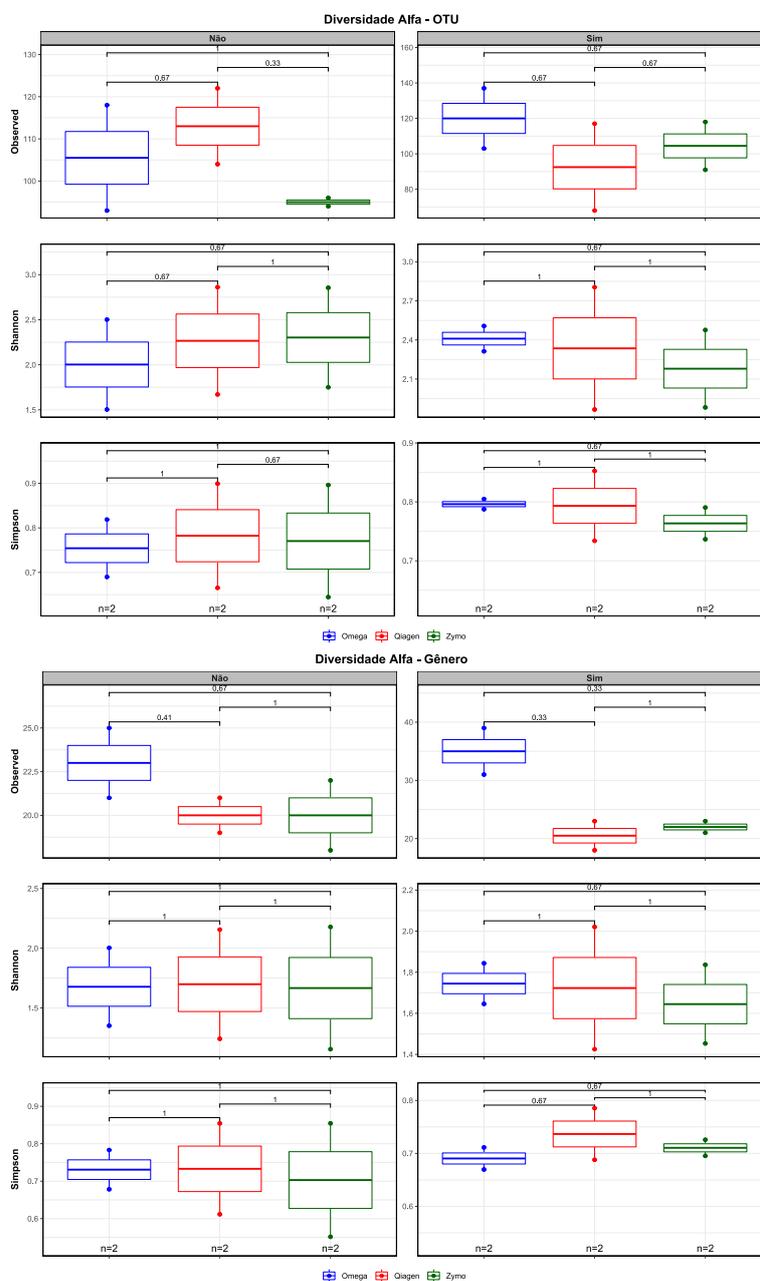
**Figura 7** – Amplificação da região ITS2 de amostras de DNA extraídas de material fecal provenientes de dois indivíduos do projeto “MOSAIC Challenge”. A nomenclatura das amostras está de acordo com a amostra de origem (F1 ou F2) e com o protocolo de extração utilizado (Qiagen - Q, Zymo - Z, ou Omega - O) na presença ou ausência de MAC4L (M). Os controles positivos utilizados foram a comunidade microbiana ZymoBIOMICS Microbial DNA Community Standard (Mock-Bacteria + Fungi) e o DNA genômico isolado de *Candida neoformans* e *Candida neoformans* cedido pelo Instituto de Medicina Tropical (USP). Como controle negativo, além das reações utilizando H<sub>2</sub>O MiliQ como template, também foi utilizada uma comunidade bacteriana, também cedida pelo “MOSAIC Challenge”. A eletroforese foi realizada em gel de agarose 2% preparado em TBE 1x.

extraídas pelos protocolos Omega e Qiagen, com e sem a adição de MAC4L, seguindo as etapas de preparo e análise descritas no item 3.1.9. As amostras de DNA extraídas com o protocolo Zymo foram excluídas dessa análise, já que não apresentaram diferenças quanto a diversidade fúngica e por apresentar baixo rendimento de DNA, quando comparado com os demais protocolos (Figura 6). Com o sequenciamento total do material genético, é possível ter uma visão mais abrangente do microbioma presente na amostra, tendo informações não só de bactérias, como fungos, vírus e arqueobactérias.

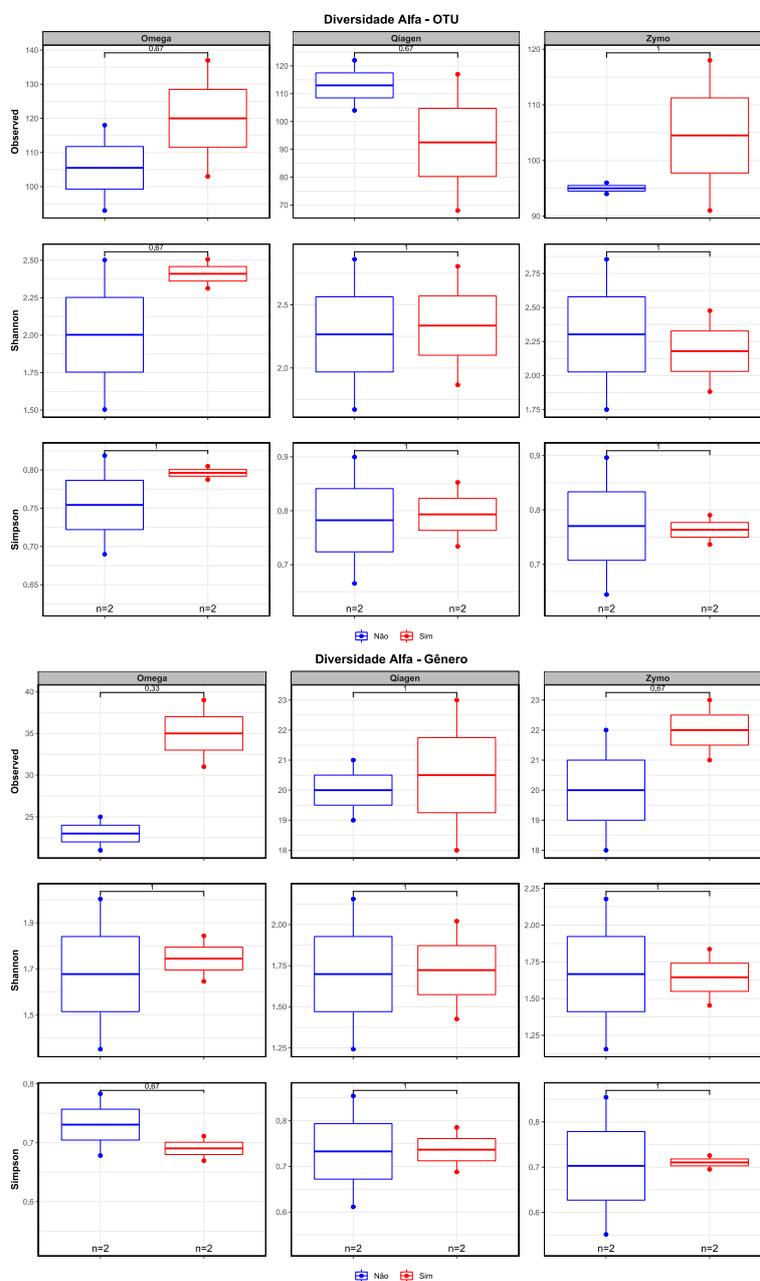


**Figura 8 –** Curvas de rarefação de OTUs observadas dos *amplicons* gerados a partir de amostras fecais extraídas por diferentes protocolos. As curvas de rarefação indicam o número de OTUs totais observadas de acordo com o número de *reads* presentes nas amostras, separadas de acordo com a amostra de origem, protocolo de extração de DNA utilizado e ausência ou presença da mistura enzimática MAC4L. É possível verificar que todas as amostras alcançam um nível de saturação de OTUs a partir de 2500 *reads*.

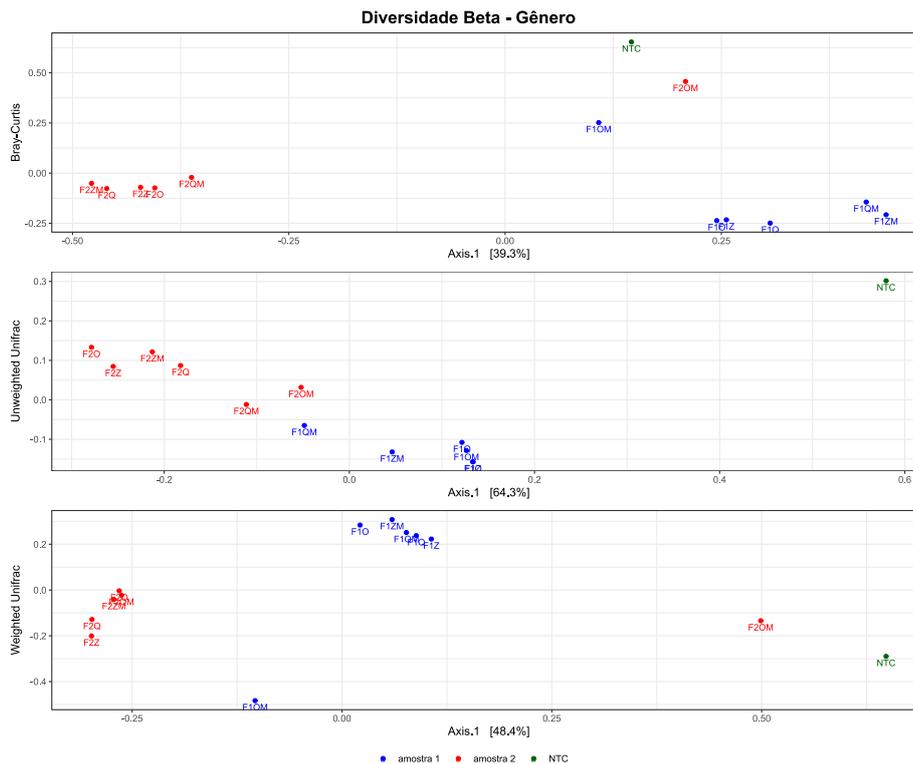
Na tabela 5 é possível verificar a predominância do reino Bacteria em todas as amostras, resultado verificado em outros estudos para amostras de fezes (Breitbart et al. 2003; Sender et al. 2016; Nash et al. 2017). Os demais grupos identificados se encontram todos com abundância inferiores a 10%, sendo que para o grupo dos fungos todas as amostras apresentaram valores inferiores a 1%, valores também observados em outros estudos investigando fezes humanas (Huffnagle e Noverr 2013; Nash et al. 2017). Apesar dos valores de frequência relativa entre as amostras serem semelhantes, isso não significa que o perfil microbiano detectado nas amostras são semelhantes, tendo em vista a situação observada com as amostras extraídas com o protocolo Omega com adição de MAC4L para o marcador ITS2 (Figura 11. Este experimento reforça o resultado anterior que os protocolos Omega e Qiagen produzem resultados semelhantes e satisfatórios. Logo, como não há evidências de uma diferença significativa entre os dois protocolos mencionados, o protocolo selecionado para ser utilizado na extração de



**Figura 9 – Comparação da diversidade fúngica entre os protocolos, na presença ou ausência de MAC4L.** acima: Comparação da diversidade alfa para o nível de OTU; abaixo: Comparação da diversidade alfa para o nível de Gênero. Para essa análise as amostras foram agrupadas de acordo com os protocolos de extração utilizados, diferenciando-as de acordo com o uso ou não da enzima MAC4L. Foi realizado o teste estatístico não paramétrico de Wilcoxon-Mann-Whitney para testar se havia diferença significativa entre os protocolos tanto na ausência, como na presença de MAC4L. Os gráficos foram gerados utilizando o pacote *phyloseq* implementado na linguagem R.



**Figura 10 – Comparação da diversidade fúngica entre presença e ausência de MAC4L para cada protocolo de extração. acima.** Comparação da diversidade alfa para o nível de OTU; **abaixo.** Comparação da diversidade alfa para o nível de Gênero. Para essa análise as amostras foram agrupadas de acordo com os protocolos de extração utilizados, diferenciando-as de acordo com o uso de MAC4L (presença em vermelho, ausência em azul). Foi realizado o teste estatístico não paramétrico de Wilcoxon-Mann-Whitney para testar a diferença entre os protocolos. Os gráficos foram gerados utilizando o pacote *phyloseq* implementado na linguagem R.



**Figura 11 – Perfil da população fúngica presente entre as amostras fecais extraídas por diferentes protocolos.** A diversidade interindividual (diversidade beta) foi analisada utilizando o pacote *phyloseq*, implementado na linguagem R. Para análise foram utilizados três índices de medição da diversidade: *Bray-Curtis*, *Weighted* e *Unweighted Unifrac*. As diferentes amostras estão nomeadas de acordo com a origem da amostra e do protocolo de extração utilizado, na ausência ou presença da enzima MAC4L. Em azul temos a amostra 1, em vermelho, a amostra 2 e em verde, o controle negativo (NTC).

DNA das amostras de pênis foi o Omega sem adição de MAC4L, considerando que o mesmo apresentou um bom rendimento de DNA (em  $\text{ng}/\mu\text{L}$ ) e tinha o melhor custo-benefício na época em que foram realizados os testes.

**Tabela 5 – Abundância relativa das *reads* geradas a partir do sequenciamento genômico total das amostras de fezes.**

Amostra	% Archaea	% Bacteria	% Eukaryota	% Fungi	% Virus
F1-Qiagen	0.056	96.107	3.397	0.328	0.44
F1-Omega	0.86	94.226	4.415	0.437	0.499
F1-Qiagen-MAC4L	0.101	94.153	5.353	0.535	0.393
F1-Omega-MAC4L	0.059	95.921	3.562	0.347	0.458
F2-Qiagen	0.965	93.381	5.414	0.455	0.239
F2-Omega	1.684	93.371	4.683	0.454	0.262
F2-Qiagen-MAC4L	3.407	90.919	5.462	0.520	0.213
F2-Omega-MAC4L	1.791	92.582	5.39	0.467	0.237

#### 4.1.5 Influência da variabilidade da região ITS2 no sequenciamento

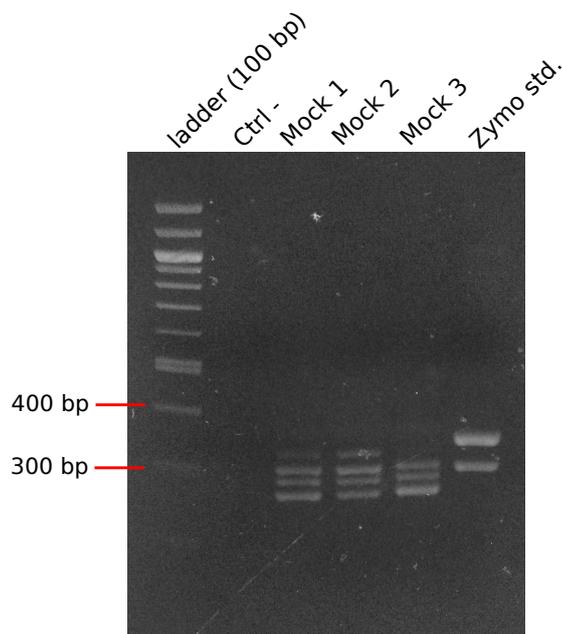
##### 4.1.5.1 Preparo das comunidades *Mock* e PCR da região ITS2

Durante a etapa de seleção *in silico* dos oligonucleotídeos verificamos uma alta variabilidade da região ITS em fungos em relação ao comprimento em pares de bases (Figura 4, que foi confirmada durante a validação dos *primers* e também durante a seleção do protocolo de extração de DNA (Figuras 5 e 7, respectivamente). Para melhor compreender essa questão, buscamos mais informações na literatura sobre a região ITS e suas particularidades, de modo compreender como essa variabilidade poderia impactar os resultados do sequenciamento.

Trabalhos prévios discutem essa característica da região ITS como um dos maiores desafios para estimar a abundância dos fungos identificados por meio de sequenciamento de *amplicons* (Palmer et al. 2018). Regiões menores tendem a ser favorecidas em reações de PCR, o que acaba distorcendo a quantidade real de indivíduos presentes em uma amostra (Filippis et al. 2017; Palmer et al. 2018). Para investigar essa questão realizamos o sequenciamento da região ITS2, amplificada por PCR, de 3 comunidades fúngicas preparadas com diferentes concentrações de 4 DNAs genômicos fúngicos cedidos pelo IMT-USP possuindo diferentes tamanhos da região ITS2 (descrição completa das comunidades *mocks* no item 3.1.8).

Pela intensidade das bandas obtidas na reação de PCR, é possível notar a que a amplificação não ocorre de maneira homogênea, mesmo quando utilizamos a mesma massa de DNA genômico (Mock 1 - Figura 12). Também é possível observar que, em todas as amostras *Mock* preparadas (Mock 1, 2 e 3), a região ITS2 de maior tamanho sempre teve menor intensidade de banda, mesmo quando possuía maior concentração que os demais isolados fúngicos (Mock 2). A exceção para essa amplificação favorecida às menores regiões ocorreu apenas na amostra comercial, que possuía a mesma concentração para os dois genomas fúngicos e a banda de maior comprimento (cerca de 350 pb) apresentou maior intensidade

do que a outra. Este último resultado revela a importância da análise de outras características das sequências que afetam a reação de PCR, como a porcentagem de bases GC na região de interesse. Uma região rica em GC apresenta formação de estruturas secundárias mais resistentes a desnaturação e prejudicam o anelamento do *primer* (Frey et al. 2008). Neste caso, a banda de maior comprimento corresponde a espécie *S. cerevisiae*, contendo cerca de 40% de GC na região de ITS2 e a espécie *C. neoformans* contendo cerca 60% de GC.

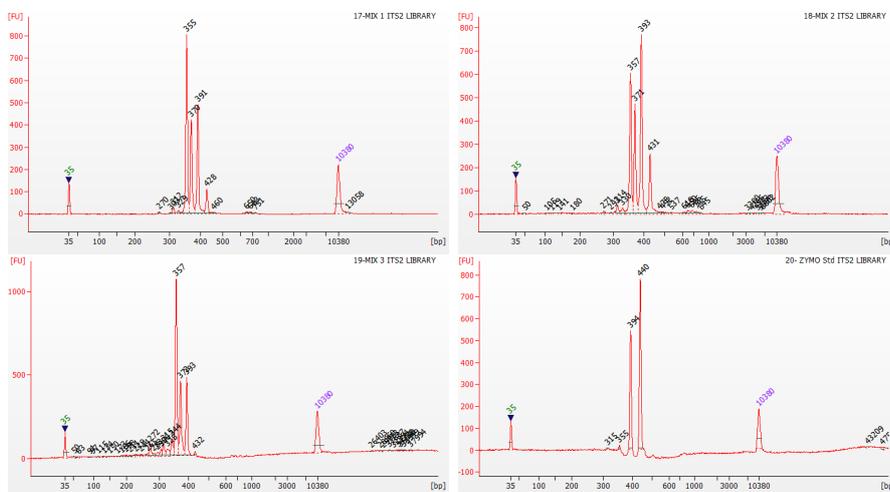


**Figura 12 – Amplificação das regiões ITS2 presentes nas comunidades *mocks*.** Mock 1, 2 e 3 correspondem as amostras *Mock* Iguais, Crescente e Decrescente, respectivamente (descrição detalhada no item 3.1.8); Zymo std - comunidade microbiana *ZymoBIOMICS Microbial DNA Community Standard*, contendo duas espécies fúngicas: *Saccharomyces cerevisiae* e *Cryptococcus neoformans*. Ctrl - H<sub>2</sub>O MiliQ.

#### 4.1.5.2 Sequenciamento e análise das amostras *mock*

O preparo das bibliotecas para o sequenciamento foi feito de acordo com o protocolo descrito no item 3.1.6, para a plataforma ION PGM. O mesmo padrão de amplificação dos diferentes fragmentos se manteve para as comunidades *mock* criadas para esse experimento (Figura 13). Os quatro fragmentos de interesse apresentaram um tamanho maior do que o esperado devido ao acréscimo dos adaptadores utilizados na construção das bibliotecas. O processamento e classifi-

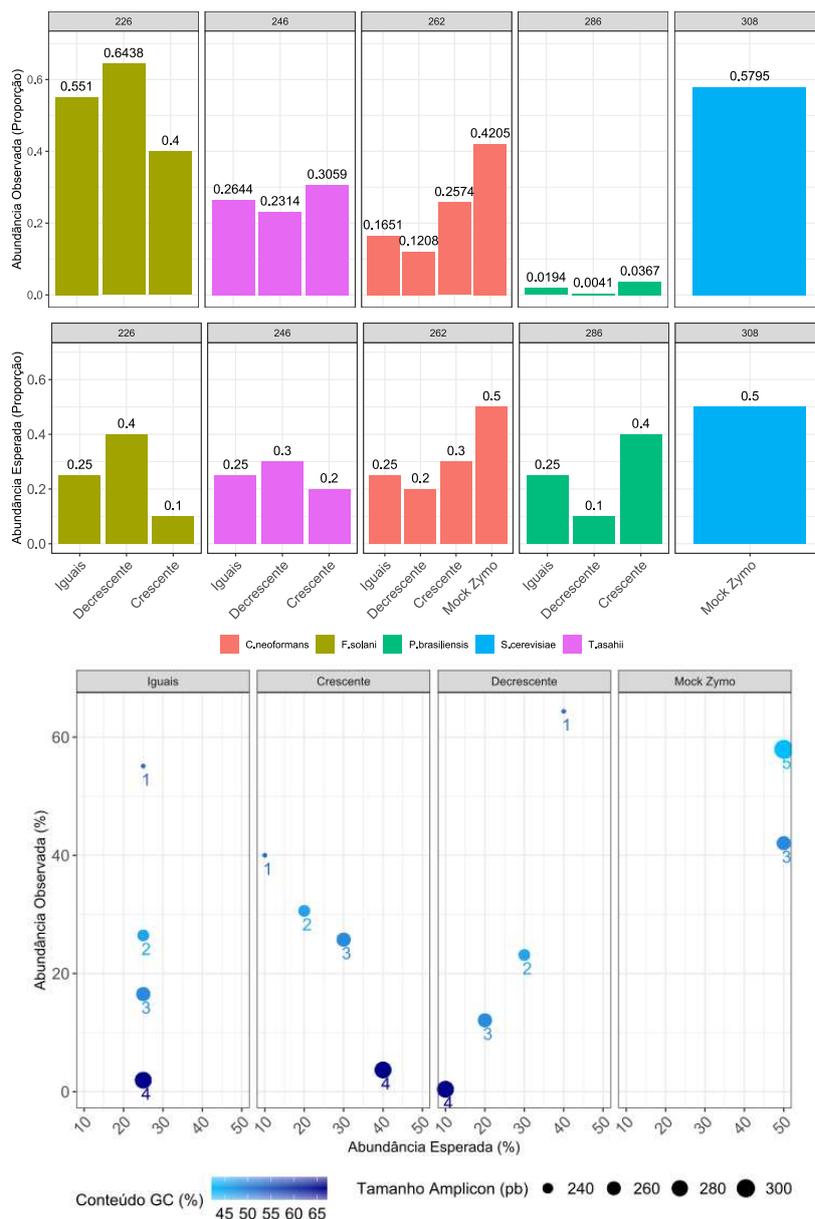
cação taxonômica das sequências obtidas após o sequenciamento foram realizados utilizando o *software* AMPtk, seguindo o *pipeline* descrito no item 3.1.6.



**Figura 13 – Eletroferograma das bibliotecas produzidas da região ITS2 amplificada por PCR das amostras Mock fúngicas. acima à esquerda** Mock Iguais - 2,5 ng de DNA genômico de cada espécie; **acima à direita** Mock Crescente - 1 ng *F.solani*, 2 ng *T.asahii*, 3 ng *C.neoformans* e 4 ng *P.brasiliensis*; **abaixo à esquerda** Mock Decrescente - 4 ng *F.solani*, 3 ng *T.asahii*, 2 ng *C.neoformans* e 1 ng *P.brasiliensis*; **abaixo à direita** Mock comercial (ZymoBIOMICS *Microbial DNA Community Standard*) que tem em sua composição o DNA genômico de 2 espécies fúngicas em iguais concentrações, *Saccharomyces cerevisiae* e *Cryptococcus neoformans* (cada espécie representando 2% do total da comunidade Mock). Para todas as corridas foi utilizado 1 ng da biblioteca para checar o perfil de fragmentos.

Em seguida, recuperamos informações das sequências de ITS2 das espécies que compuseram as amostras *mock*, como tamanho em pb, conteúdo GC e sequências homopoliméricas, características essas que são conhecidas por afetar o preparo das bibliotecas de sequenciamento e resultar na alteração da abundância observada entre as espécies (Ross et al. 2013; Laursen et al. 2017; Boers et al. 2019). Ao comparar a abundância observada com a esperada, incluindo as informações de tamanho em pb nota-se o viés gerado pela amplificação da região ITS2 (Figura 14, acima). A menor região (226 pb) da espécie *Fusarium solani* apresentou abundâncias observadas superiores às esperadas, mas em particular, a sequência ITS2 de *Paracoccidioides brasiliensis* (286 pb) obteve um resultado esperado muito inferior ao esperado em todas as situações. Mas, quando se observa a região de *Saccharomyces cerevisiae* com 308 pb e abundância observada

muito próxima da esperada na amostra Mock Zymo fica claro que outros fatores podem estar afetando a amplificação das regiões de interesse, como o conteúdo GC de cada sequência, ou até mesmo a complexidade da composição de cada amostra, já que a amostra Mock Zymo contém apenas o DNA genômico de duas espécies fúngicas. Ao adicionar a informação de conteúdo GC na análise anterior é possível verificar que a região ITS2 de *F.solani* é composta de 53.54% de bases GC, enquanto a região de *P.brasiliensis* possui a maior porcentagem (67.13%), e a região de *S.cerevisiae*, 42.20% (Figura 14, abaixo). Por apresentar tanto o maior tamanho como a maior porcentagem de bases GC, a região ITS2 de *P.brasiliensis* obteve uma quantidade *reads* muito inferior ao resultado esperado. Da mesma maneira, é possível que a baixa porcentagem de bases GC da região de *S.cerevisiae* compensou o maior tamanho em pb da região e gerou um resultado próximo do esperado.



**Figura 14 – Comparação entre as abundâncias observadas e esperadas das espécies presentes nas comunidades *mock*. acima.** As diferentes espécies presentes nas comunidades *mocks* estão separadas pelo tamanho de *amplicon* de ITS2. Cada amostra está indicada pelos nomes definidos de acordo com as concentrações utilizadas de cada DNA genômico detalhados no item 3.1.8. **abaixo.** Complementação da análise com o conteúdo GC. As espécies presentes nas amostras estão indicadas por números, sendo 1- *Fusarium solani*, 2- *Trichosporon asahii*, 3- *Cryptococcus neoformans*, 4- *Paracoccidioides brasiliensis* e 5- *Saccharomyces cerevisiae*.

#### 4.1.5.3 Desenvolvimento de um modelo estatístico para estimar com maior precisão a abundância dos representantes do microbioma fúngico de uma determinada amostra

Com o intuito de identificar a existência de correlação entre o resultado esperado com o resultado observado e as diferentes propriedades das regiões ITS2, buscamos outros trabalhos na literatura com dados de sequenciamento baseado em *amplicons* de comunidades *mock* fúngicas (Tabela 6). Os dados obtidos da literatura foram submetidos ao mesmo processamento da região ITS2, descrito no item 3.1.6. Foram obtidas as características das sequências das espécies fúngicas utilizadas, coletando também outras informações relevantes, como a abundância esperada de cada espécie, número de ciclos usado durante a etapa de amplificação por PCR, região de interesse sequenciada e plataforma de sequenciamento utilizada. Todos os estudos analisados apresentam uma grande diferença entre a abundância observada após a análise e a abundância esperada para cada espécie presente nas diferentes comunidades *mock* (Figura 15, acima).

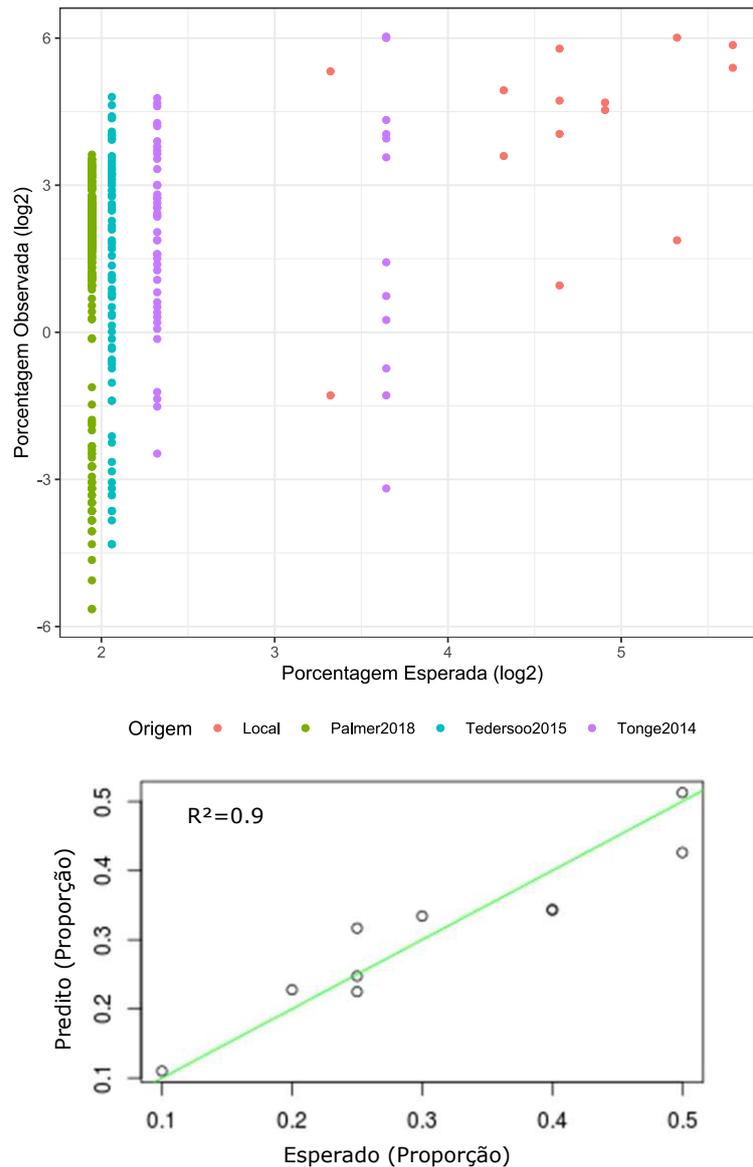
Numa etapa inicial de análise, foram utilizados apenas os dados gerados pelo sequenciamento das amostras *mock* feito *inhouse* para treinar um modelo para predição da abundância (frequência relativa) real de cada espécie presente nas amostras. A versão final do modelo (descrito em detalhes no item 3.1.8) utilizou apenas 5 das 7 variáveis de entrada (contagens de homopolímeros de 5 e 6 pb não foram consideradas) e foi capaz de realizar predições razoavelmente acuradas ( $R^2 = 0.9$ ; Figura 15, abaixo).

Em seguida o modelo foi aplicado para o conjunto de dados obtidos na literatura, mas o mesmo não performou de maneira satisfatória como o observado com os dados *inhouse* (análise ainda não finalizada). Acreditamos que outros fatores estão impactando no resultado ao agregar dados de diferentes laboratórios. Estamos trabalhando na normalização dos dados de sequenciamento, para remoção dos efeitos em lote, ou *batch effects*, presentes em análises de dados ge-

rados por diferentes plataformas biológicas *high-throughput* (Goh et al. 2017), que são gerados por diferentes fatores, como manipulação da amostra, reagentes utilizados, e também plataformas de sequenciamento (Leek et al. 2010). Após a normalização, será realizada a incorporação dos dados da literatura junto ao modelo de modo a obter um preditor que utilize a abundância observada de *reads* correspondentes a cada OTU fúngica, em conjunto com as demais propriedades da sequência, para estimar de forma mais eficiente a abundância das OTUs encontradas em um determinado microbioma.

**Tabela 6** – Sequenciamento de comunidades *mock* fúngicas encontrados na literatura.

Nome	Região Alvo	Plataforma de Sequenciamento	Referência
Tonge2014	ITS1, ITS2 e LSU	Ion PGM	(Tonge et al. 2014)
Tedersoo2015	ITS1, ITS2 e LSU	MiSeq Illumina	(Tedersoo et al. 2015)
Palmer2018	ITS2	MiSeq Illumina e Ion PGM	(Palmer et al. 2018)



**Figura 15 – Análise de diferentes sequenciamentos de comunidades *mock* fúngicas.** **acima** Comparação entre os resultados individuais observados e esperados de espécies fúngicas em comunidades *mock*. A porcentagem esperada foi definida de acordo com a composição de cada comunidade *mock* descrita em cada estudo. Os diferentes estudos analisados estão indicado pelas diferentes cores. **abaixo.** Frequências relativas estimadas pelo modelo obtido (Lasso) vs frequências esperadas nas amostras *mock*. O gráfico mostra a comparação entre os valores de frequências relativas esperadas de cada espécie fúngica que compõe as três comunidades *mock* sequenciadas no atual projeto em relação aos valores preditos pelo modelo descrito na seção 3.1.8. A regressão linear entre os dados esperados e os preditos possui o coeficiente de determinação igual 0.9 ( $R^2 = 0.9$ ).

#### 4.1.6 Sequenciamento da Região ITS2

##### 4.1.6.1 Otimização do protocolo de PCR da região ITS2

Apesar da validação dos oligonucleotídeos já ter sido realizada com sucesso, tanto utilizando amostras de DNA genômico de isolados fúngicos (cedidos pelo IMT-USP), como em amostras fecais ("MOSAIC Challenge"), foi necessário otimizar a reação de PCR de acordo com o material utilizado. Como já visto em trabalhos anteriores, a microbiota humana varia de acordo com a região do corpo (Huttenhower et al. 2012). Também há a presença de compostos que podem atuar como inibidores da amplificação durante a etapa de PCR e apresentam variação de acordo com o tipo da amostra (Schrader et al. 2012). Diferentes amostras de DNA já extraídas, provenientes de materiais diferentes, foram utilizados para verificar a eficiência do protocolo de PCR estabelecido anteriormente. Foram utilizadas amostras de lavado gástrico (LG) e biópsia (bx) de pacientes com adenocarcinoma gástrico e amostras de *swab* oral (SW) de pacientes com carcinoma epidermoide bucal.

Inicialmente foi testado o protocolo utilizado para a validação dos oligonucleotídeos, em que foi alterado a massa de DNA e o volume da reação, conforme o descrito no item 3.1.4 . Este protocolo resultou em amplificação apenas do controle positivo e uma amostra de lavado gástrico (LG131), que possivelmente deve apresentar uma quantidade muito elevada de DNA fúngico, pois a intensidade da banda obtida é similar a dos controles positivos (Figura 16, acima à esquerda).

Com o intuito de verificar se a ausência de amplificação estaria ocorrendo devido a presença de inibidores nas amostras, foram testadas variações do protocolo de PCR buscando aumentar a eficiência da reação de amplificação. Para isso, foram selecionadas a amostra LG131 como mais um controle interno de amplificação e a amostra LG789, que não apresentou amplificação pelo protocolo inicial. A Figura 16 apresenta o resultado do PCR utilizando a da enzima KAPA2G Robust HotStart ReadyMix (Sigma-Aldrich), recomendada por ter maior tole-

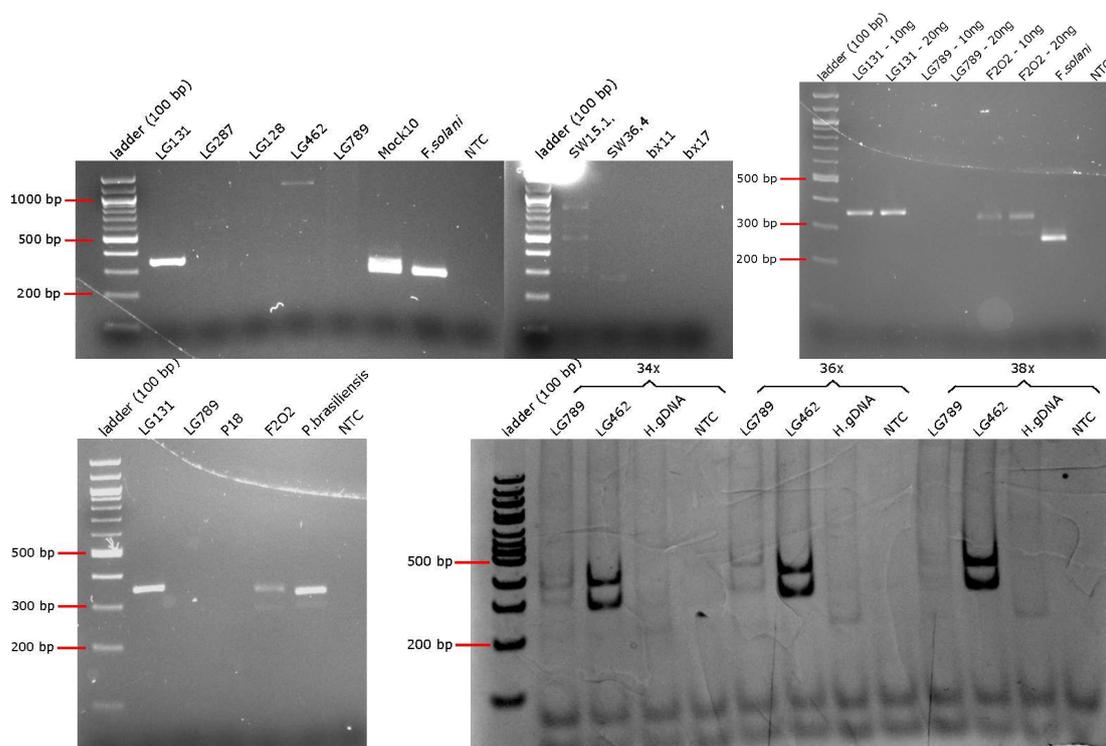
rância a inibidores e também por ser capaz de amplificar regiões ricas em GC e AT (protocolo descrito no item 3.1.4). É possível verificar que não houveram diferenças em relação ao resultado da Figura 16 (acima à direita) em relação às amostras em questão. Para verificar a presença de inibidores PCR nas amostras, foram repetidas as reações reduzindo a massa de DNA (10 ng) o que também não resultou em diferenças no resultado final.

Além do teste com a KAPA2G Robust, foi realizado outro teste para verificar se a amplificação não estava ocorrendo na amostra LG789 devido a presença de regiões ricas em GC. Para tal, foi adicionado Dimetilsulfóxido (DMSO), um solvente orgânico capaz de facilitar o rompimento de ligações GC durante a etapa de desnaturação da PCR, e também evita a formação de estruturas secundárias do DNA simples fita em sequências ricas em GC (protocolo descrito no item 3.1.4). A Figura 16 (abaixo à esquerda) apresenta resultado semelhante ao anterior, com amplificação apenas da amostra LG131 e do controle positivo, com maior intensidade das bandas.

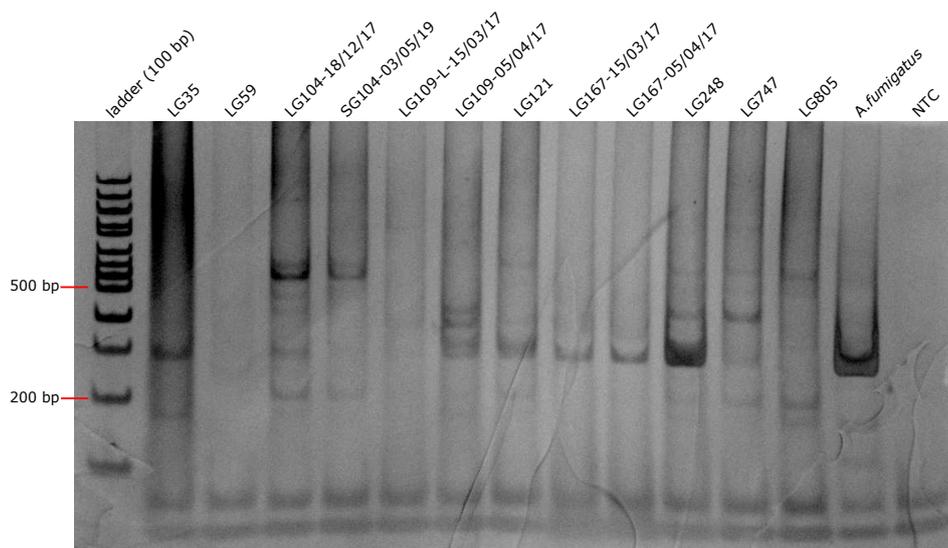
Pelos resultado obtidos até esse momento, foi decidido retornar ao protocolo inicial e trabalhar com os parâmetros internos do protocolo, tendo em vista que a ausência de produtos de amplificação das amostras não parece ser devido a uma quantidade exacerbada de inibidores ou a predominância de regiões ITS2 ricas em bases GC. Em seguida, vários testes foram realizados alterando a concentração dos oligonucleotídeos utilizados, massa de DNA, temperatura de anelamento e número de ciclos da reação de PCR. A Figura 16 (abaixo à direita) apresenta o resultado mais relevante para essa etapa de otimização do protocolo de PCR para amostras de LG. Após verificarmos a necessidade de aumentar a concentração dos oligonucleotídeos ( $0.5 \mu\text{M}$ ) e da massa de DNA (50 ng), foi avaliada a influência do número de ciclos da PCR (Figura 16, abaixo à direita). Neste experimento, é possível verificar a presença de bandas para a amostra LG789, que ainda não tinha sido observado em nenhum experimento anterior. A amostra LG462, assim como a amostra LG131, serviu de controle interno positivo para

a reação. Foi decidido alterar o número de ciclos para 34, tendo em vista que um valor maior começa a produzir amplificação inespecífica no DNA humano (H.gDNA). As bandas que aparecem em todas as amostras abaixo de 100 pb são dímeros de *primers* resultantes do aumento da concentração dos oligonucleotídeos. Apesar de não serem desejados em uma reação de PCR, estes dímeros são eliminados durante as etapas seguintes do protocolo para o sequenciamento.

O protocolo otimizado (item 3.1.4) foi testado em diferentes amostras de LG para verificar a eficiência da reação em um maior número de amostras (Figura 17). Observa-se que a amplificação se deu em praticamente todas as amostras e ainda que existe uma grande diversidade entre os tamanhos da região ITS2 no microbioma gástrico. É possível que amostras como a LG59, que não apresentou amplificação, apresentem uma quantidade tão baixa de fungos que inviabiliza o estudo do micobioma. Esse experimento também testou a conservação do DNA de algumas amostras que tiveram mais uma extração realizada (LG104, LG109 e LG167). Verificou-se que, de um modo geral, ocorre uma redução das bandas amplificadas em extrações mais recentes. Dessa forma, buscou-se priorizar a primeira extração realizada das amostras selecionadas, sempre que houvesse material disponível.



**Figura 16 – Testes para otimização do protocolo de PCR para região ITS2 em amostras de lavado gástrico. acima à esquerda.** PCR estabelecido para amostras de DNA genômico de isolados fúngicos e amostras de fezes. **acima à direita.** PCR utilizando enzima KAPA2G Robust HotStart ReadyMix (Sigma-Aldrich) **abaixo à esquerda.** PCR utilizando a enzima KAPA HiFi HotStart ReadyMix (Roche) e DMSO a 5%. **abaixo à direita.** PCR utilizando a enzima KAPA HiFi HotStart ReadyMix (Roche), oligonucleotídeos a  $0.5 \mu\text{M}$ , 50 ng de DNA e variações no número de ciclos (34, 36 ou 38). Tipos de amostras utilizadas: lavado gástricos (LG), biópsia de adenocarcinoma gástrico (bx), *swab* oral (SW), fezes (F) e peça cirúrgica peniana (P). Os números que seguem o tipo da amostra indicam diferentes pacientes. Controles positivos: DNA genômico isolado das espécies *Fusarium solani* (F.solani), *Paracoccidioides brasiliensis* (P.brasiliensis); controle negativo: reações sem DNA (NTC) e DNA genômico humano (H.gDNA) (Promega - número de catálogo: G1471). Detalhes do protocolo da PCR estão descritos na seção 3.1.4. A eletroforese foi realizada em gel 4% poliacrilamida (TBE 1X) na figura abaixo à direita e em gel 2% agarose (TBE 1X) nas demais figuras.



**Figura 17 – Amplificação da região ITS2 para amostras de lavado gástrico utilizando protocolo otimizado.** Diferentes amostras de LG foram utilizadas para testar o protocolo otimizado. O DNA genômico isolado da espécie *Aspergillus fumigatus* foi utilizado como controle positivo (IMT-USP) e como controle negativo, reações sem DNA (NTC). Detalhes da PCR estão descritos nas seção 4.1.6.1

#### 4.1.6.2 Amplificação da região ITS2 - Amostras de Lavado Gástrico e Pênis

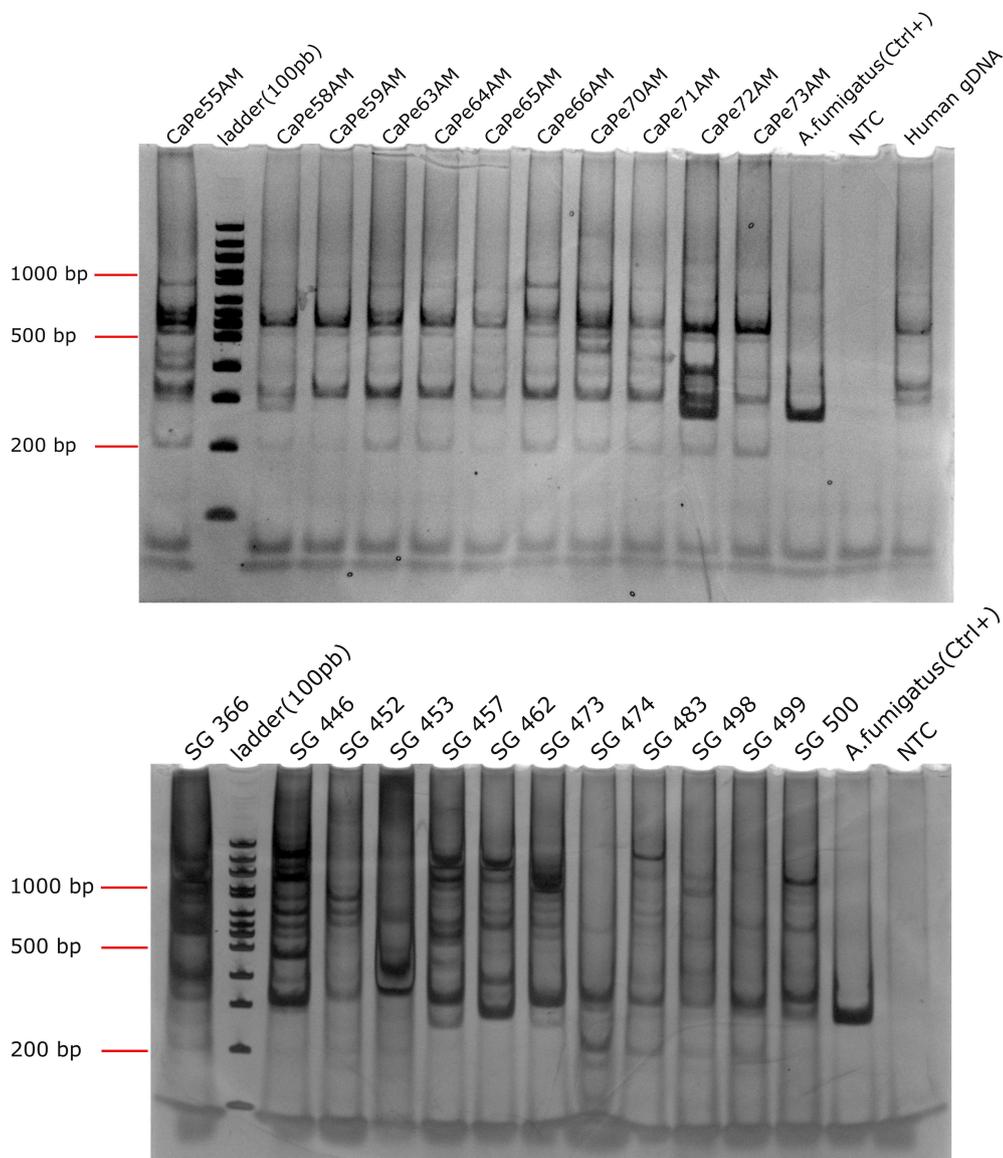
Conforme o descrito no item 4.1.1, foram utilizados dois protocolos de extração das amostras dependendo do tipo de material utilizado (tecido congelado para amostras de pênis ou alíquota de lavado gástrico). Após a extração e quantificação das amostras, foi realizado o protocolo de PCR otimizado para amplificação da região ITS2. Mesmo tendo utilizado a massa de 50 ng nos testes, verificamos a necessidade de aumentar a massa para 80 ng para as amostras de lavado gástrico, tendo em vista a ausência de amplificação de várias amostras durante as PCRs iniciais. Dessa forma, foram utilizados para amostras de pênis a massa de 50 ng de DNA total extraído e para as amostras LG foi utilizada a massa de 80 ng.

Como mencionado no item 4.1, o padrão de amplificação da região ITS2 é muito heterogêneo. Na Figura 18 podemos observar padrões diferentes dependendo da amostra de origem, sendo um padrão mais similar para amostras

de pênis (Figura 18, acima), que é justificável tendo em vista a amplificação de bandas inespecíficas quando é observado o resultado da PCR da amostra de DNA genômico humano (número de catálogo G1471 Promega ®). É importante ressaltar que essa contaminação não foi observada durante a validação dos oligonucleotídeos, que foi realizada com menores massas de DNA, já que os controles positivos utilizados se tratavam de DNA genômicos isolados de colônias fúngicas (cedidas pelo IMT-USP). Para contornar essa amplificação inespecífica, o *pipeline* de análise contém uma etapa de processamento prévio em que as sequências são alinhadas com o genoma humano e somente sequências não alinhadas seguem para a análise posterior 3.1.6. Para as amostras de LG, observa-se um padrão de amplificação mais heterogêneo (Figura 18, abaixo), o que também acreditamos ser esperado, já que o número de células do hospedeiro é menor quando comparado com uma amostra de tecido. Após a PCR, o preparo das bibliotecas e o sequenciamento das mesmas foi realizado de acordo com o item 3.1.6, e o resultados serão apresentados nos próximos itens.

#### **4.1.6.3 Problemas com o Sequenciamento utilizando a plataforma Ion S5**

Durante a etapa de sequenciamento enfrentamos dificuldades que inviabilizaram 2 experimentos realizados, resultando em uma porcentagem muito alta de sequências de baixa qualidade, e também de cobertura muito desigual entre as amostras sequenciadas, muito abaixo do esperado. Após a investigação com a equipe da Thermo, verificamos que o lote dos reagentes de sequenciamento utilizados estava com problema e também a concentração final do *pool* das bibliotecas a 60 pM, recomendado pela equipe da Thermo, não era adequada para o sequenciamento que estávamos realizando (sequenciamento de regiões de até 600 pb). Segundo a nova recomendação, deveríamos utilizar a concentração final a 40 pM. Infelizmente só foi possível repetir os experimentos utilizando essa nova concentração para as amostras de lavado gástrico.



**Figura 18 – Variabilidade do perfil de amplificação da região ITS2 de acordo com o tipo de amostra de origem. acima.** Exemplo de resultado da eletroforese da PCR de DNA total extraído de amostras de pênis; **abaixo** Exemplo de resultado da eletroforese da PCR de DNA total extraído de lavado gástrico. Amostras de pênis indicadas pelo sufixo "CaPe" e amostras de LG indicadas pelo sufixo "SG". Como controle temos o DNA genômico extraído de *Aspergillus fumigatus*, NTC para amostra ausente em DNA e *Human gDNA* para a amostra de DNA humano comercial (número de catálogo G1471 Promega®). A eletroforese foi realizada em gel 4% poliacrilamida (TBE 1X).

#### 4.1.7 Resultados - Amostras de Pênis

Das 54 amostras de pênis selecionadas, foram preparadas bibliotecas de 52 amostras (as outras 22 amostras não apresentaram amplificação da região ITS2 ou não apresentaram material suficiente para a reação de PCR). Como será explicado no próximo item, ao todo foram utilizadas 25 amostras para o sequenciamento da região ITS2. A média de idade dos pacientes foi de 52 anos e a tabela 7 apresenta as informações clínicas das amostras utilizadas. Por se tratar de um estudo retrospectivo e também por se tratarem de conjuntos diferentes de amostras, não foi possível obter o preenchimento das informações para todas as variáveis. Não foi possível utilizar a informação do estadiamento clínico para essas amostras, devido a falta de informação da maioria das amostras e também por utilizarem diferentes edições da classificação.

**Tabela 7** – Dados clínicos - Amostras de pênis

<b>Variáveis</b>	<b>n</b>	<b>%</b>
<b>Origem Amostra</b>		
ACCCC (São Paulo)	17	68
FCECON (Amazonas)	8	32
<b>Estado Civil</b>		
casado	13	52
solteiro	7	28
união estável	2	8
viúvo	2	8
não responderam	1	4
<b>Cor da pele</b>		
não informado	17	68
parda	7	28
branca	1	4
<b>alfabetizado</b>		

não informado	15	60
sim	9	36
não	1	4
<hr/>		
<b>Atendimento</b>		
SUS	18	72
Convênio	6	24
não informado	1	4
<hr/>		
<b>Tabagismo</b>		
não	13	52
sim	8	32
não informado	3	12
ex-tabagista	1	4
<hr/>		
<b>Fimose</b>		
sim	13	52
não informado	10	40
não	2	8
<hr/>		
<b>HPV</b>		
negativo	10	40
positivo	9	36
não informado	6	24
<hr/>		
<b>EBV</b>		
negativo	16	64
não informado	6	24
positivo	3	12
<hr/>		
<b>Histórico IST</b>		
não informado	18	72
não	6	24
sim	1	4
<hr/>		

<b>Tipo da amostra</b>		
neoplásico (tumor)	20	80
não neoplásico (tecido adjacente ao tumor)	5	20
<b>Local do tumor</b>		
glânde	12	48
outras especificações	5	20
prepúcio	3	12
não informado	3	12
corpo cavernoso e/ou esponjoso	1	4
glânde e prepúcio	1	4
<b>Tipo histológico</b>		
carcinoma epidermoide	21	84
outros tipos histológicos	3	12
carcinoma verrucoso	1	4
<b>Grau histopatológico</b>		
moderadamente diferenciado	14	56
bem diferenciado	5	20
não informado	3	12
indiferenciado	2	8
pouco diferenciado	1	4
<b>Metástase linfonodal</b>		
ausente	13	52
presente	10	40
não informado	2	8

#### 4.1.7.1 Sequenciamento - Região ITS2

Das 52 bibliotecas preparadas para as amostras de pênis, realizamos um sequenciamento inicial com 25 amostras. Os resultados obtidos deste sequenciamento e dos demais foram analisados pelo *pipeline* descrito no item 3.1.6, considerando aqui ASV ao invés de OTU. Diferentemente do método de OTU, que utiliza da clusterização das *reads* em um determinado valor de identidade (geralmente 97%, como foi utilizado nos experimentos anteriores) (Westcott e Schloss 2015; Kopylova et al. 2016), o método de ASV utiliza da tecnologia *denoising* para inferir as sequências biológicas nas amostras antes da introdução dos erros de amplificação e sequenciamento, o que permite a diferenciação das sequências a partir de um único nucleotídeo (Callahan et al. 2017). Tendo em vista as vantagens esperadas do uso de ASV e por estar sendo mais utilizado em trabalho recentes (Fricker et al. 2019), optamos por seguir com o uso de ASV para as amostras do projeto.

Neste sequenciamento, foram obtidas 2.129.785 sequências totais, sendo que 87.904 viáveis (sequências com pelo menos 100 pb e apresentando pelo menos um dos oligonucleotídeos utilizados na PCR). Os resultados obtidos após este processamento inicial do *pipeline* revelaram que não foi alcançado a cobertura média esperada mínima de 31.250 mil sequências por amostra (Tabela 8), quando consideramos que o sequenciamento é capaz de produzir de 3 a 4 milhões de *reads* ao utilizamos o Ion 520 Chip para sequências de 600 pb. Este processamento inicial teve como média 3143.5 *reads* por amostra e mediana de 2.418. Dessas, foram removidas na sequência 48.946 sequências por se alinharem ao genoma humano (GRCh37), restando 38.958 sequências para composição da tabela de ASVs e posteriores análise taxonômicas.

Devido ao baixo número de sequências do primeiro sequenciamento, decidimos repetir as amostras e adicionar as amostras restantes, dessa vez diminuindo o número total de amostras do sequenciamento (no primeiro sequenciamento, ao

todo existiam 96 amostras, incluindo as 25 de pênis), com o intuito de aumentar o número de sequências viáveis de cada amostra para confirmar os resultados observados nas análises realizadas. Infelizmente, esse sequenciamento foi feito antes do ajuste na molaridade final do *pool* de bibliotecas feito pela Thermo e o resultado obtido foi pior àquele obtido anteriormente, conforme o apêndice C. Para este sequenciamento, obtivemos uma média de 1685.86 *reads* por amostra com mediana de 1.234. Por esses motivos decidimos seguir apenas com o dados das 25 amostras iniciais sequenciadas.

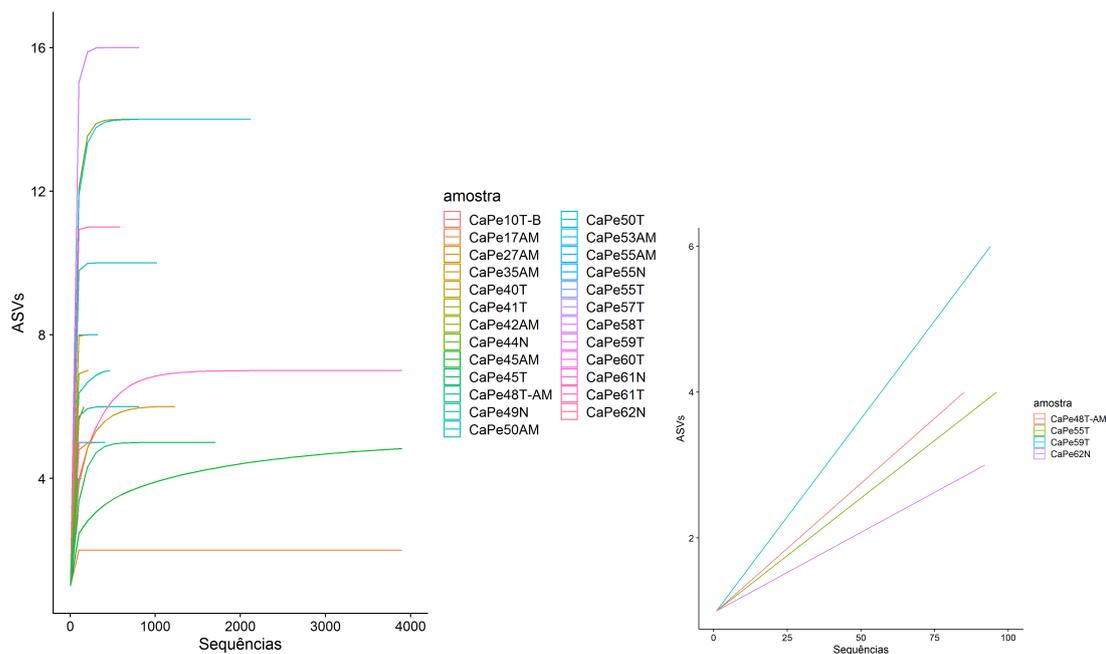
**Tabela 8** – Sequenciamento ITS2 com 25 amostras de pênis - Número de *reads* de ITS2 por amostra de pênis obtidas após a etapa de demultiplexação, identificação dos *primers*, considerando o tamanho mínimo de 100 pb, realizado pelo *software* AMPtk

Amostra	Barcode	Número de Sequências
CaPe10T-B	BC.1	2909
CaPe17AM	BC.2	13521
CaPe27AM	BC.3	2391
CaPe40T	BC.4	3026
CaPe41T	BC.5	2585
CaPe44N	BC.6	1367
CaPe45AM	BC.7	6962
CaPe45T	BC.8	3045
CaPe48T-AM	BC.9	1802
CaPe49N	BC.10	3162
CaPe50AM	BC.11	2393
CaPe50T	BC.12	2443
CaPe55N	BC.13	1860
CaPe55T	BC.14	2358
CaPe57T	BC.15	1273
CaPe59T	BC.16	2022
CaPe35AM	BC.17	4817
CaPe42AM	BC.18	4434
CaPe53AM	BC.19	7012
CaPe55AM	BC.20	3134
CaPe58T	BC.21	2953
CaPe60T	BC.22	2379
CaPe61N	BC.23	5770
CaPe61T	BC.24	2049
CaPe62N	BC.25	2237

Através da curva de rarefação (Figura 19, à esquerda), é possível verifi-

car que, com exceção de quatro amostras (CaPe48T-AM, CaPe55T, CaPe59T e CaPe62N) (Figura 19, à direita), todas as amostras atingiram o platô em relação ao número de ASVs. As amostras que não atingiram a saturação de ASVs foram removidas das análises posteriores. Essa métrica é uma forma de verificar se foi possível sequenciar de forma representativa o microbioma fúngico presente na amostra, através da checagem das ASVs presentes em um grupo de sequências selecionadas aleatoriamente das amostras (sem reposição) (McMurdie e Holmes 2014; Gotelli e Colwell 2001). Ao consultar trabalhos anteriores (Persoon et al. 2017; Nash et al. 2017; Raimondi et al. 2019; Diaz et al. 2019; Aykut et al. 2019), estimamos que seria necessário em média de 2 a 10 mil sequências válidas por amostra (sequências com qualidade, maior que 100 pares de base, com a presença de ao menos um dos oligonucleotídeos utilizados na reação de PCR), para se atingir esse platô de saturação (ponto que mesmo com o aumento do número de sequências, não é acompanhado por um aumento do número de ASVs). Também é importante observar que o número de ASVs identificadas nas amostras foi baixo, variando de 2 ASVs (amostra CaPe17AM) até 16 (amostra CaPe58T). Essa pequena diversidade observada está de acordo com outros estudo do microbioma humano (Persoon et al. 2017; Huseyin et al. 2017; Nash et al. 2017; Raimondi et al. 2019; Aykut et al. 2019).

Em relação a classificação taxonômica, dividimos as amostras obtidas de tecido adjacente ao tumor das amostras tumorais, mas devido ao pequeno número de amostras que restaram para as análises, não é possível identificar um padrão entre os fungos observados (Figura 20, acima). Verificamos também se havia alguma diferença no perfil fúngico das amostras obtidas do Amazonas comparadas com as amostras de São Paulo. A princípio, temos uma maior presença do gênero *Aspergillus* nas amostras coletadas no ACCCC do que em FCECON (Figura 20, abaixo). Além disso, as amostras recebidas de Amazonas aparentam ser menos diversas em termos gênero quando comparadas com as de São Paulo. Por se tratar de um pequeno grupo de amostras, não é possível concluir se realmente há



**Figura 19 – Curvas de Rarefação - Amostras de Pênis - ITS2.** à esquerda. ASVs observadas em relação ao número de sequências obtidas a partir do sequenciamento de 25 amostras de pênis; à direita. Curvas de rarefação das amostras que não alcançaram o ponto de saturação de diversidade de ASVs.

uma diferença entre a diversidade de fungos entre os dois grupos de amostras.

Além da classificação taxonômica, também é possível comparar a diversidade fúngica presente nas amostras através de métricas comumente utilizadas na área da Ecologia e também no estudo do microbioma. A análise da diversidade alfa realizada, foram avaliados os seguintes índices: **i.** Táxons Observados (*Observed*): avalia a riqueza, ou seja os diferentes táxons presentes em uma determinada amostra; **ii.** índice de Shannon: considera riqueza e equitabilidade, atribuindo o valor zero para a existência de apenas uma espécie o valor 1 quando todas as espécies possuem a mesma abundância (Sagar e Sharma 2012); **iii.** índice de Simpson: em oposição ao índice anterior, este mede a força da dominância, atribuindo maior peso para a espécie mais abundante e tem relação indiretamente proporcional com a diversidade de espécies (Whittaker 1972), sendo zero uma amostra sem dominância e um, o valor máximo, contendo apenas uma espécie na amostra (Greenberg 1956).

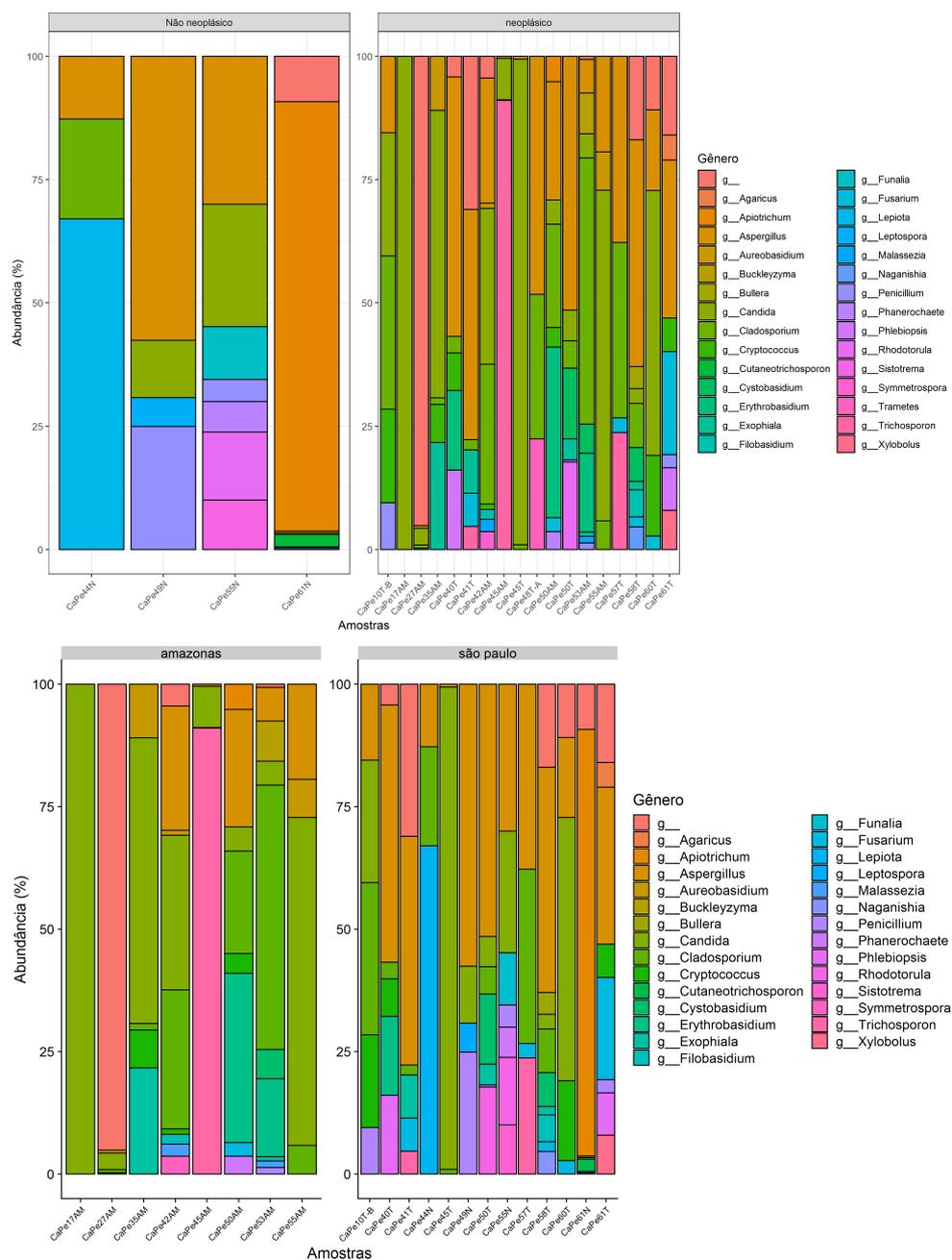
A Figura 21 apresenta alguns resultados obtidos para diferentes métricas

da diversidade alfa a partir das amostras de pênis sequenciadas. Verificamos que as amostras de São Paulo apresentaram uma maior diversidade intra-amostral em relação ao índice de Shannon quando comparada às amostras do Amazonas em relação nível Filo, ou seja, as amostras de SP são mais homogêneas em relação a abundância dos Filos do que as amostras de AM (Figura 21, à esquerda). É possível que o perfil fúngico varie de acordo com a região, como já notificado em um estudo recente, que identificou assinaturas microbianas de acordo com a cidade de origem das amostras ambientais sequenciadas (Danko et al. 2021). Além disso, independente da origem da amostra, pacientes que apresentaram metástase nos linfonodos possuem menor diversidade (a nível de Família) em relação a pacientes sem metástase, considerando o índice Shannon (Figura 21, à direita). Foram realizadas comparações entre outras características (tabagismo, fimose, HPV, EBV, grau histopatológico, raça cor ou etnia, etc), para cada nível taxonômico (ASV, Filo, Classe, Ordem, Família e Gênero), mas não foi verificada nenhuma outra diferença significativa. Também foi feita a comparação entre as amostras neoplásicas e não neoplásicas (tumor vs. tecido adjacente ao tumor) e não foi verificada diferença significativa, mas como foi dito anteriormente, esse resultado já era esperado, pelo pequeno número de amostras analisadas.

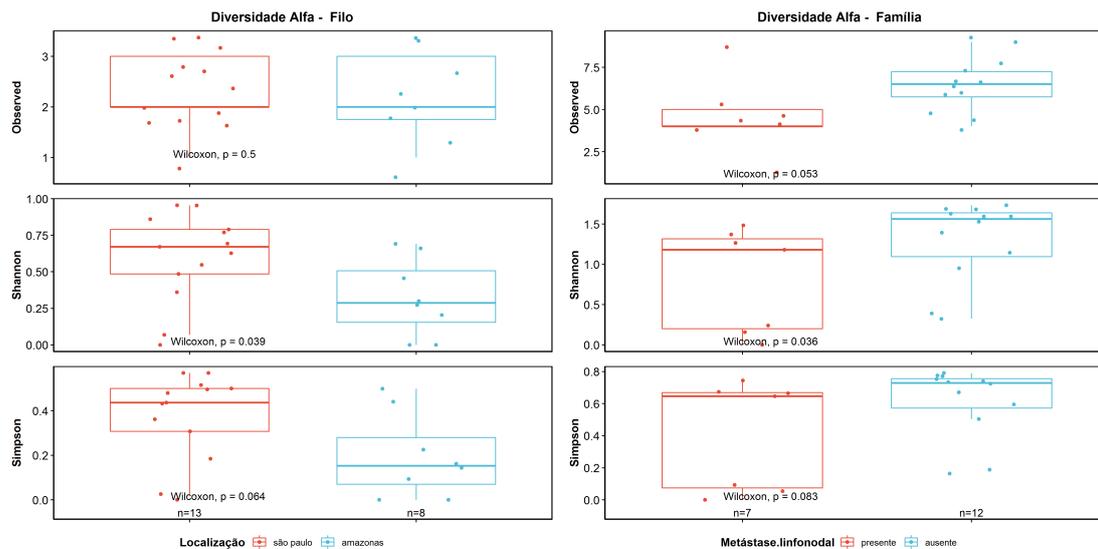
Para a análise da diversidade beta, foi utilizada a metodologia ANOSIM (*Analysis of Similarities*), que utiliza uma permutação da matriz de dissimilaridade para comparar a população fúngica entre dois ou mais grupos. O valor de R calculado, quando próximo de 1.0, sugere uma dissimilaridade entre os grupos, enquanto um valor de R próximo de zero sugere uma distribuição uniforme dentro e entre os grupos <sup>1</sup>.

Assim como a diversidade alfa, a diversidade beta pode ser avaliada por diferentes métricas, como: **i)** Dissimilaridade de Bray-Curtis: considera a abundância de microrganismos compartilhados entre duas amostras e o número de microrganismos identificados em cada amostra, medido na escala de zero a um,

<sup>1</sup> <https://sites.google.com/site/mb3gustame/hypothesis-tests/anosim>



**Figura 20 – Perfil taxonômico dos microbioma fúngico das amostras de pênis a nível de Gênero.** As ASVs classificadas taxonomicamente pelo *software* AMPTk foram convertidas em frequência relativa, representado no eixo Y (Abundância). As amostras foram separadas de acordo com o local de origem (Amazonas ou São Paulo). O termo “g\_\_” representam seqüências que não foram classificadas com o nível taxonômico em questão. A imagem foi gerada utilizando a função `plot_bar()` do pacote *phyloseq*.



**Figura 21 – Diversidade alfa dos fungos identificados nas amostras de pênis. à esquerda.** Análise de acordo com a origem das amostras coletadas em relação ao nível taxonômico de Filo. **à direita.** Análise de acordo a presença ou ausência de metástase linfonodal em relação ao nível taxonômico de Família, respectivamente. Foram avaliadas as métricas de táxons observados (*Observed*), índices de Shannon e Simpson.

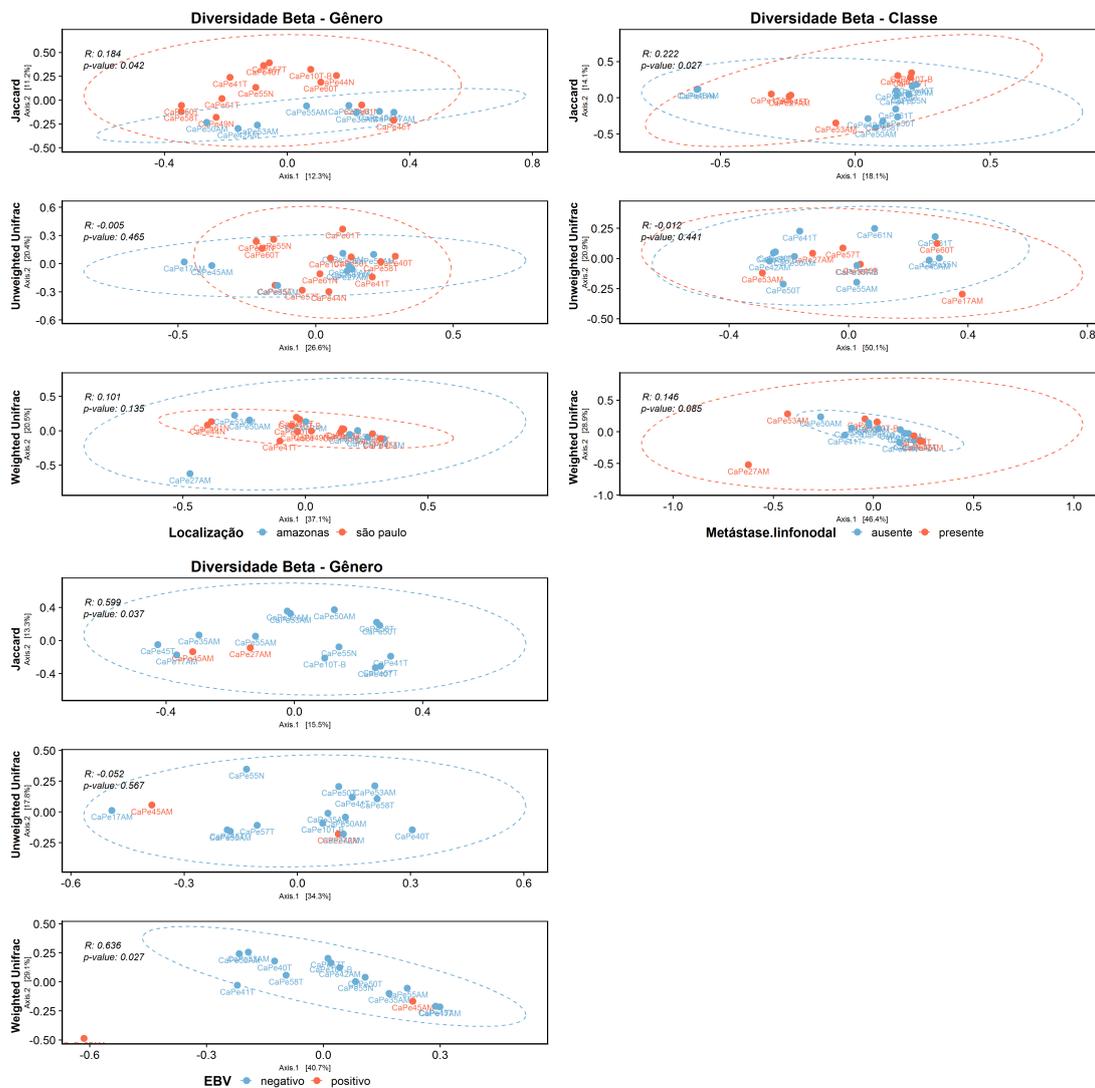
sendo zero atribuído para amostras com mesmo número de microrganismos e um quando não compartilham nenhum microrganismo (Bray e Curtis 1957); **ii**) Distância Jaccard: semelhante a métrica anterior, mas que não leva consideração a informação da abundância, apenas a presença de microrganismos em uma ou ambas as amostras (Jaccard 1900); **iii**) Unifrac Qualitativo (*Unweighted*): adiciona a informação de distâncias filogenéticas entre os representantes do microbioma da amostra para a realização do cálculo. Esta métrica considera que microrganismos taxonomicamente mais próximos podem desempenhar funções mais similares entre si do que espécies de grupos taxonômicos diferentes (Lozupone et al. 2006); **iv**) Unifrac Quantitativo (*Weighted*): segue a mesma regra que a medida anterior e adiciona abundância dos microrganismos observados (Lozupone et al. 2007).

Em relação à diversidade beta, verificamos para as variáveis EBV (presença/ausência) e metástase linfonodal, uma maior similaridade entre os grupos analisados (R próximos a zero com valor  $p < 0.05$ ) (Figura 22, acima à direita), a população de fungos dos dois grupos é muito similar (Figura 22, inferior à es-

querda). Esse resultado indica que apesar das amostras ausentes de metástase linfonodal apresentarem maior diversidade intra-amostral em relação às amostras positivas. Já a análise da diversidade beta em relação a origem de cada amostra indicou uma maior dissimilaridade entre as amostras de Amazonas e São Paulo em todas as métricas avaliadas (Figura 22, acima à esquerda). Em relação a presença/ausência de EBV (Epstein-Barr Virus), apenas as métricas que consideram a abundância de cada grupo fúngico (Jaccard e Unifrac Quantitativo) tiveram evidências estatísticas significativas para afirmar que a população de fungos nas amostras negativas para EBV tem maior grau de dissimilaridade em relação a população das amostras positivas para EBV (Figura 22, abaixo à esquerda). Também foram avaliadas as outras informações clínicas das amostras analisadas, inclusive tipo da amostra (tumor ou tecido adjacente ao tumor), mas não foram verificadas diferenças significativas.

Também foi realizada uma análise de abundância diferencial para verificar se um ou mais grupos de fungos estão mais ou menos presentes em diferentes subgrupos das amostras de carcinoma de pênis analisadas. Para esta análise foi utilizada a metodologia ANCOM-BC (*Analysis of Compositions of Microbiomes with Bias Correction*), desenvolvida para corrigir o problema de realizar amostragens desiguais, observados na metodologia de ANCOM (Lin e Peddada 2020).

A tabela 9 apresenta os dados observados para a análise comparando a abundância dos gêneros presentes entre as amostras neoplásicas e não neoplásicas. O gênero *Cladosporium* apresentou cerca de 4 vezes maior abundância nas amostras neoplásicas e os gêneros *Exophiala*, *Cryptococcus*, *Fusarium* e *Phanerochaete* apresentaram menor abundância, ou seja, são mais abundantes nas amostras não neoplásicas. Apesar da análise constar com um pequeno número de amostras, principalmente para as amostras não neoplásicas, foi possível verificar uma diferença significativa para os gêneros mencionados.



**Figura 22 – Diversidade beta dos fungos identificados nas amostras de pênis.** Comparação da população fúngica entre as amostras as amostras de carcinoma de pênis. Análise de Coordenadas Principais (PCoA) do microbioma fúngico classificado em diferentes níveis taxonômicos (Classe e Gênero). A variável clínica analisada, assim como o nível taxonômico estão indicados em cada Figura. As métricas avaliadas estão indicadas ao lado do eixo y (Jaccard, Unifrac Qualitativo ou Unifrac Quantitativo), e no canto superior esquerdo de cada gráfico estão indicados o valor de R estatístico ANOSIM.

**Tabela 9** – Análise de abundância diferencial dos fungos presentes nas amostras de pênis, a nível de Gênero, em relação às informações ao tipo da amostral (tumor ou tecido adjacente ao tumor). A metodologia de ANCOM-BC foi utilizada para análise de abundância diferencial, e a tabela consta com os valores de LFC (*Log Fold Change*), p, com nível de significância de 0.05. O asterisco (\*) indica valores p significativos corrigidos (considerando o nível de significância sendo 5%).

Gênero	LFC - Amostra: neoplásico	p
g__Cladosporium	2.0305	0.0000*
g__Exophiala	0.2433	0.0000*
g__Fusarium	-0.5600	0.0000*
g__Phanerochaete	-2.3593	0.0000*
g__Cryptococcus	0.6833	0.0000*
g__Penicillium	-2.7629	0.3739
g__Aspergillus	-1.4751	0.9232
g__Aureobasidium	-0.1225	1.0000
g__Candida	-0.6651	1.0000
g__Cystobasidium	0.0730	1.0000
g__Erythrobasidium	0.1751	1.0000
g__Malassezia	-0.1852	1.0000
g__Trichosporon	0.1636	1.0000

Também comparamos a abundância dos gêneros presentes nas amostras tumorais considerando a presença/ausência de metástase linfonodal (Tabela 10). Para essa análise incorporamos às informações de idade, presença/ausência de HPV e EBV, por serem fatores que tem relação com o câncer de pênis (Favorito et al. 2008; Afonso et al. 2012). O gênero *Penicillium* se mostrou cerca de 2x mais abundante em amostras com presença de metástase linfonodal. Já os gêneros *Aureobasidium*, *Exophiala*, *Fusarium* e *Cryptococcus* apresentaram menor abundância nas mesmas em relação às amostras ausentes para esta condição. É interessante observar que o gênero *Fusarium* também apresenta menor abundância nas amostras neoplásicas, pela análise anterior (Tabela 10). Pelo formato do estudo, não é possível estabelecer uma relação causal entre o microbioma e a evolução do tumor, mas é possível que a progressão do tumor desfavoreça a

manutenção de fungos pertencentes a esse gênero. Não é possível correlacionar o resultado observado com a presença de infecções por patógenos fúngicos devido a ausência dessa informação. Existem trabalhos na literatura que reportaram a infecção por representantes do gênero *Fusarium* em pacientes oncológicos (Jr et al. 1997; Arnoni et al. 2018).

**Tabela 10** – Análise de abundância diferencial dos fungos presentes nas amostras de pênis, a nível de Gênero, em relação à informação de presença ou ausência de metástase linfonodal. A metodologia de ANCOM-BC foi utilizada para análise de abundância diferencial, considerando as informações de idade, presença/ausência de HPV e EBV como fatores influenciadores para essa análise. As colunas contendo o termo LFC representam valores de (*Log Fold Change*), e as colunas adjacentes, representam os valores p, com nível de significância de 0.05. O asterisco (\*) indica valores p significativos corrigidos.

Gênero	LFC-Metast.linf: presente	p Metast.linf	LFC Idade	p Idade	LFC-HPV: positivo	p HPV	LFC-EBV: positivo	p EBV
g__Aureobasidium	-0.5647	0.0000*	-0.0511	0.4214	0.1526	1	-0.8658	1
g__Exophiala	-0.3288	0.0000*	0.0101	1	-0.5421	1	-0.5218	1
g__Fusarium	-0.1745	0.0000*	-0.0235	1	0.0544	1	-0.6829	1
g__Cryptococcus	-0.0111	0.0000*	-0.0035	1	-0.4307	1	0.0264	1
g__Penicillium	1.4975	0.0458	0.0015	1	-1.2329	0.4636	0.1964	1
g__Cladosporium	1.0716	0.9324	0.0555	0.1352	-1.379	0.7207	-0.9731	1
g__Trichosporon	-0.5153	1	-0.0996	0.1352	1.4024	1	2.7745	1
g__Candida	1.4127	1	-0.0976	0.4214	0.7015	1	0.0263	1
g__Cystobasidium	0.6552	1	0.0292	1	-2.1077	0.4636	0.4756	1
g__Aspergillus	-0.8568	1	0.017	1	-1.819	0.7052	0.3159	1
g__Erythrobasidium	0.6048	1	0.0413	0.9429	-0.4506	1	-0.3126	1
g__Malassezia	0.4564	1	-0.0105	1	0.1061	1	0.5932	1
g__Filobasidium	-0.1525	1	-0.0215	1	-0.8015	1	0.0115	1

#### 4.1.8 Resultados - Amostras de Lavado Gástrico

Diferentemente do ocorrido com as amostras de pênis, foi possível repetir e o sequenciamento das amostras de lavado gástrico após as novas recomendações da empresa Thermo a respeito da concentração otimizada do *pool* final das bibliotecas (40 pM). Dessa forma, foi possível realizar o sequenciamento das 150 amostras selecionadas para este projeto. A tabela 11 apresenta as informações das amostras selecionadas. Buscamos selecionar amostras que já haviam sido sequenciadas para a região 16S de bactérias, para em um trabalho futuro relacionarmos os fungos identificados neste trabalho com as bactérias identificadas anteriormente. A média de idade é de 56.6 anos para as amostras de câncer, e de 56.3 anos para as amostras de gastrite superficial. Para simplificar a apresentação dos dados, agrupamos as respostas "não se aplica" com "não informado", tendo em vista que existem informações exclusivas para as amostras provenientes de pacientes com câncer. Todas as amostras de câncer foram diagnosticadas como adenocarcinoma gástrico.

**Tabela 11** – Dados clínicos - Conjunto amostral de LG

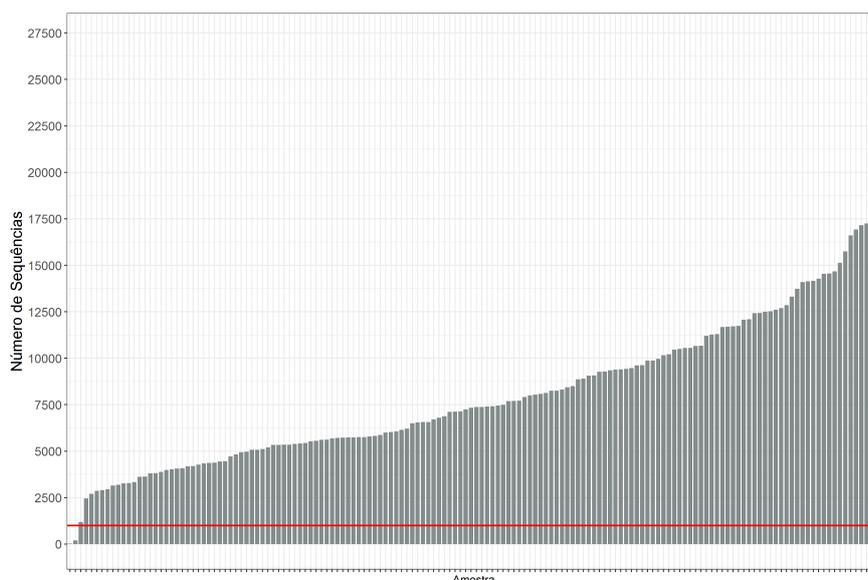
<b>Variáveis</b>	<b>n</b>	<b>%</b>
<b>Sexo</b>		
Masculino	89	58.94
Feminino	62	41.06
<b>Etilismo</b>		
sim	76	50.33
nao	74	49.01
não informado	1	0.66
<b>Tabagismo</b>		
nao	75	49.67
ex-fumante	58	38.41
sim	17	11.26

não informado	1	0.66
<b>Diabetes</b>		
Não	128	84.77
Sim	22	14.57
não informado	1	0.66
<b>H.pylori</b>		
Negativo	116	76.82
Positivo	26	17.22
não informado	9	5.96
<b>Uso antibióticos</b>		
mais de 1 ano	69	45.7
não se lembra	39	25.83
entre 1 e 6 meses	30	19.87
mais de 6 meses	7	4.64
pelo menos 1 ano	3	1.99
cerca de 1 mês	2	1.32
não informado	1	0.66
<b>Uso antiácidos</b>		
Não	127	84.11
Sim	23	15.23
não informado	1	0.66
<b>Uso inibidores proteases</b>		
Sim	90	59.6
Não	59	39.07
não informado	2	1.32
<b>Diagnóstico Endoscopia</b>		
Câncer	74	49.01
Gastrite Superficial	50	33.11

Metaplasia Intestinal	22	14.57
Gastrite Crônica	4	2.65
Exame normal	1	0.66
<hr/>		
<b>Local da lesão</b>		
não informado	77	50.99
Corpo	34	22.52
Cardia/Fundo	20	13.25
Esôfago Distal	9	5.96
Antro	8	5.3
Boca Anastomótica	3	1.99
<hr/>		
<b>Estado doença</b>		
não informado	78	51.66
Doença localizada	59	39.07
Doença metastática	14	9.27
<hr/>		
<b>Grau histológico</b>		
não informado	100	66.23
GH3 (pouco diferenciado)	33	21.85
GH2 (moderadamente diferenciado)	13	8.61
GH1 (bem diferenciado)	4	2.65
indiferenciado	1	0.66
<hr/>		
<b>Classificação Lauren</b>		
não informado	81	53.64
Difuso	44	29.14
Intestinal	21	13.91
Misto	5	3.31
<hr/>		

#### 4.1.8.1 Sequenciamento - Região ITS2

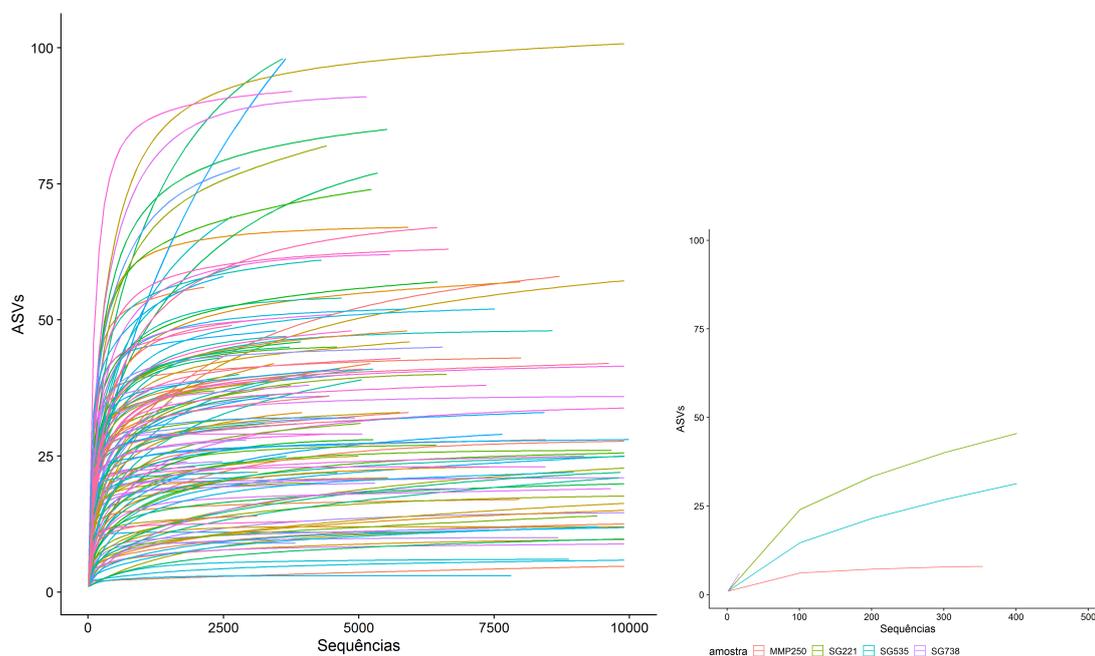
Após a alteração no protocolo de sequenciamento, observamos uma melhoria significativa em relação a cobertura de *reads* por amostra. Considerando os dois sequenciamentos realizados, obtivemos uma média e mediana de 8.022 e 7.355 *reads* por amostra. A Figura 23 apresenta a distribuição de seqüências após processamento inicial do *pipeline* de análise. Apenas duas amostras apresentaram um número inferior a 1.000 seqüências (SG738 e MMP250, com 17 e 354 *reads*, respectivamente), mas apenas a amostra SG738 foi removida das análises posteriores, já que a curva de saturação da amostra MMP250 atingiu o platô de saturação (Figura 24, à direita).



**Figura 23 – Distribuição de seqüências das amostras de lavado gástrico sequenciadas - ITS2.** Quantidade de *reads* obtidas após a etapa de demultiplexação, identificação dos *primers* e considerando o tamanho mínimo de 100 pb, realizado pelo *software* AMPtk.

Podemos observar também pela curva de saturação uma grande variedade em relação a saturação do número de ASVs comparando com o número de seqüências necessárias para atingir essa saturação (Figura 24, à esquerda). Existem amostras que atingiram o platô de saturação com um pequeno número de *reads*, como é o caso da amostras SG738, mas também temos casos de amostras

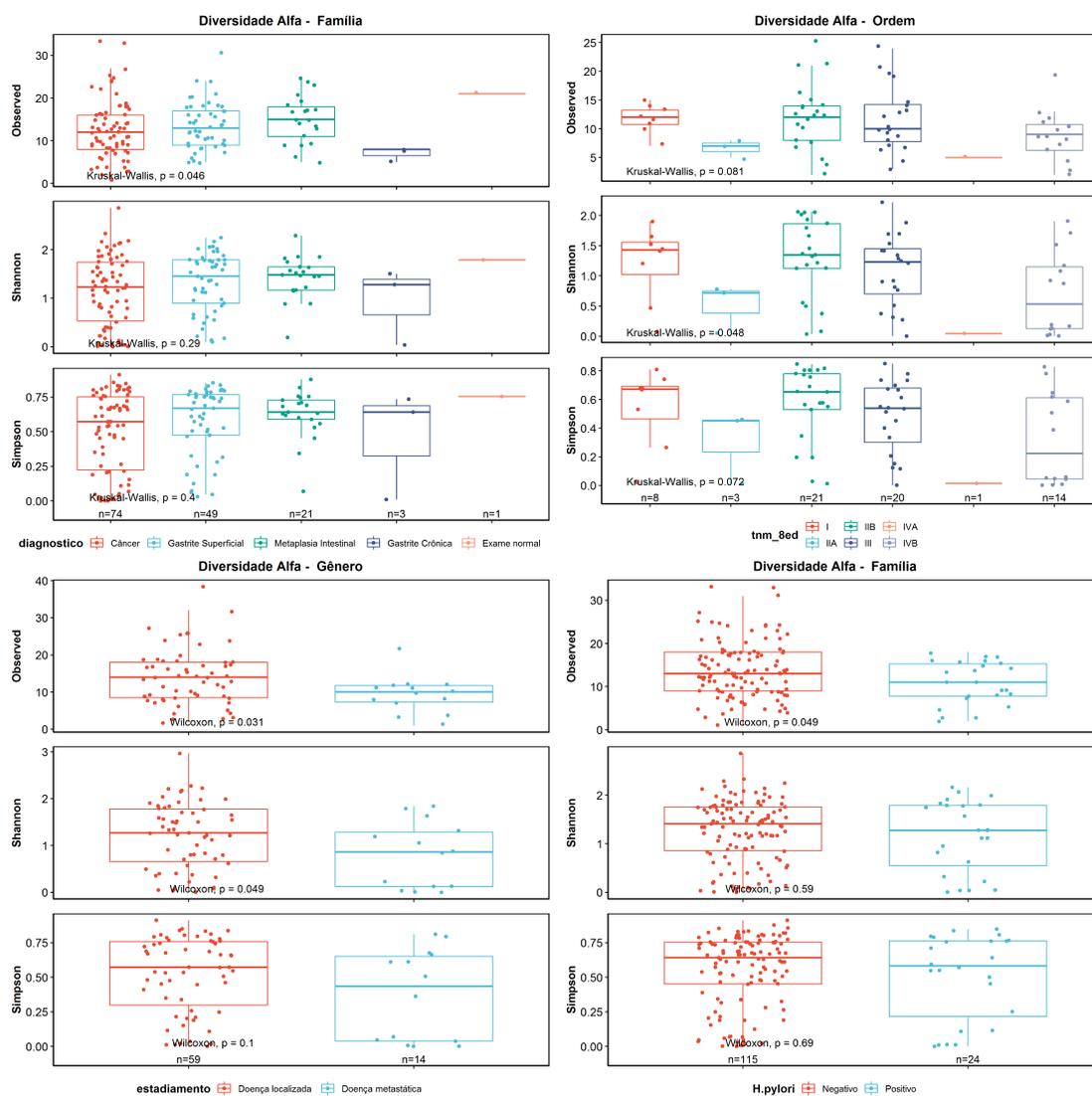
com um número elevado de *reads*, que não atingiram o platô (amostras SG221 e SG535) (Figura 24, à direita), que foram removidas das análises posteriores juntamente com a amostra SG738.



**Figura 24 – Curvas de rarefação - amostras de lavado gástrico - ITS2. acima.** ASVs observadas em relação ao número de sequências obtidas a partir do sequenciamento de 151 amostras de LG; **abaixo.** Visão detalhada das amostras que não alcançaram o ponto de saturação de ASVs e com número inferior a 1.000 *reads*.

A Figura 25 apresenta as principais diferenças obtidas para a análise da diversidade alfa da população fúngica observada nas amostras de lavado gástrico. Verificamos que em relação ao diagnóstico obtido da endoscopia dos pacientes, existe uma diferença intra-amostral a nível de Família dos fungos identificados nas amostras (Figura 25, acima à esquerda). Também foi possível verificar uma diferença na diversidade alfa na população fúngica quando aos estágios da doença, tanto em relação ao estadiamento patológico (TNM 8<sup>a</sup> edição), quanto ao agrupamento dos estágios iniciais e mais avançados (Figuras 25, acima à direita e abaixo à esquerda, respectivamente). Em termos absolutos, as amostras com estágios mais avançados do câncer apresentam menor diversidade intra-amostral em relação aos estágios iniciais e às doenças pré-neoplásicas. Em relação a outras informações clínicas das amostras, observamos uma maior diversidade alfa dos

fungos nas amostras negativas para *H.pylori* em relação às amostras positivas, para o nível de Família (Figura 25, abaixo à direita). Como já dito anteriormente, a *H.pylori* já tem uma relação bem estabelecida com o desenvolvimento do câncer gástrico (CG) e a alteração promovida por esta espécie no microbioma bacteriano pode ter relação com uma alteração do microbioma fúngica, tendo em vista a interação bacterioma-micobioma já investigada em outros trabalhos, inclusive no contexto do câncer (Oever e Netea 2014; Sam et al. 2017; Shay et al. 2020).



**Figura 25 – Diversidade alfa dos fungos identificados nas amostras de lavado gástrico. acima.** Análise de acordo com a origem das amostras coletadas em relação ao nível de classificação sendo ASV ou Gênero, respectivamente. **abaixo.** Análise de acordo a presença ou ausência de metástase linfonodal em relação ao nível de classificação sendo ASV ou Gênero, respectivamente. Foram avaliadas as métricas de táxons observados (*Observed*), índices de Shannon e Simpson.

Em relação comparação da diversidade inter-amostal (beta), também fizemos a análise utilizando as informações apresentadas na tabela 11, mas para facilitar a visualização, foi realizado o filtro das comparações com valor p significativo em pelo menos uma das métricas testadas (Tabela 12). Além das amostras positivas para *H.pylori* apresentarem menor diversidade intra-amostal, vemos que existe uma diferença entre os representantes do microbioma entre as amostras positivas ou negativas para essa bactéria, quando consideramos os níveis Filo e Ordem (valores de R - *Unweighted* Unifrac de 0.2238 e 0.1072 e valor p de 0.002 e 0.022, respectivamente). Quando analisamos a variável Estadiamento, que apresenta os subgrupos "doença localizada" e "doença metastática", verificamos uma maior diferença entre as populações fúngicas (R mais próximo de 1) ao nível de Classe (R Unw.Unifrac= 0.1708, p = 0.040), quando consideramos os demais resultados. Também verificamos diferenças no microbioma quando comparamos indivíduos com ou sem diabetes para o nível de Filo (R Unw.Unifrac = 0.1064, p = 0.039). Já para a variável tabagismo, as Ordens fúngicas detectadas nas amostras de indivíduos fumantes ou não, tem menor diferença entre seus indivíduos (R Jaccard = 0.0396, p = 0.048).

**Tabela 12** – Comparação da diversidade beta das amostras de lavado gástrico. Foram avaliadas diferentes informações clínicas considerando diferentes métricas para a análise (valor R estatísticos e p valor indicados nas colunas). Foram selecionadas as comparações com pelo menos uma métrica com valor significativo (p < 0.05). As métricas analisadas foram Jaccard, *Unweighted* Unifrac (Unw.Unifrac) e *Weighted* Unifrac (W.Unifrac). O asterisco (\*) indica valores p significativos (considerando o nível de significância sendo 5%).

Variável	Nível Taxonômico	R - Jaccard	p - Jaccard	R - Unw.Unifrac	p - Unw.Unifrac	R - W.Unifrac	p - W.Unifrac
<i>H.pylori</i>	Filo	0.0271	0.312	0.2238	0.002*	0.0106	0.394
<i>H.pylori</i>	Ordem	-0.0159	0.623	0.1072	0.022*	-0.0604	0.870
Estadiamento	Filo	0.1038	0.043*	0.0989	0.062	0.0054	0.430
Estadiamento	Classe	0.0946	0.073	0.1708	0.040*	0.0306	0.314
Estadiamento	Ordem	0.0901	0.086	0.1049	0.047*	0.0413	0.251
Estadiamento	Família	0.1425	0.045*	0.0956	0.076	-0.0031	0.471
Estadiamento	Gênero	0.1409	0.036*	0.0698	0.143	0.0060	0.448
TNM (Sed)	ASV	0.0801	0.025*	0.0550	0.098	0.0448	0.126
Tabagismo	Ordem	0.0396	0.048*	-0.0187	0.831	0.0318	0.092
Diabetes	Filo	0.0653	0.136	0.1064	0.039*	0.0473	0.203

Para a análise de abundância diferencial realizamos duas comparações, para a primeira, foram selecionadas as amostras de pacientes diagnosticados com gastrite superficial e as amostras de pacientes com câncer gástrico, des-

considerando assim as amostras das outras doenças pré-neoplásicas (gastrite crônica e metaplasia intestinal). Para esta análise, as variáveis sexo, IMC e presença/ausência para *H.pylori* foram consideradas como dependentes devido a achados em estudos anteriores, tanto para o microbioma bacteriano (Haro et al. 2016; Noto e Jr 2017; Pereira-Marques et al. 2019), como para o fúngico (Dash et al. 2019). Por isso, as mesmas foram incorporadas na análise de abundância diferencial em questão e não foram encontradas evidências estatísticas de gêneros fúngicos diferencialmente abundantes entre as amostras de gastrite superficial e câncer gástrico (Tabela 13). Diferentemente de outros estudos comparativos entre amostras tumorais e normais, o atual estudo utilizou de amostras provenientes de pacientes portadores de alguma doença gástrica para servir como controle. Ao verificar a análise da diversidade beta, não vemos a informação de diagnóstico na tabela 12, isso indica que para essa variável não foram encontradas diferenças significativas entre a variável em questão e algum nível taxonômico, o que reforça essa ausência de diferença para análise de abundância diferencial.

Para a segunda análise, filtramos apenas as amostras tumorais e realizamos a comparação entre as amostras de pacientes com a doença localizada e aquelas de pacientes com doença metastática, considerando as mesmas variáveis citadas anteriormente como fatores dependentes para a análise de abundância diferencial. Verificamos uma maior abundância dos gêneros *Symmetrospora* e *Cutaneotrichosporon*, que apresentaram uma abundância cerca de 120% e 145% maior nas amostras originadas de pacientes acometidos pela metástase do câncer gástrico. Em contrapartida, verificamos uma redução na abundância de vários gêneros (*Phlebia*, *Perenniporia*, *Erythrobasidium*, *Naganishia*, *Yarrowia*, *Rhodotorula*, *Debaryomyces*, *Filobasidium*, *Bjerkandera*, *Cystobasidium*, *Exophiala*, *Peniophora*, *Cryptococcus*, *Fusarium*, *Torulasporea*, *Humicola*, *Trametes* e *Trichosporon*), nas amostras destes mesmos pacientes (Tabela 14).

Dos gêneros que se mostraram mais abundantes em pacientes metastáticos, temos *Cutaneotrichosporon*, também relatado em um estudo recente, que estava

enriquecido em amostras de biópsia de 22 pacientes acometidos por CG quando comparados com amostras de 11 indivíduos saudáveis (Yang et al. 2022). Apesar disso, no mesmo estudo, também é reportada a maior abundância do gênero *Cystobasidium* na amostras de CG, mas aqui foi verificada a redução em amostras de LG de pacientes com metástase. Temos também o gênero *Exophiala*, que se mostrou reduzido em amostras metastáticas e foi relatado, no mesmo estudo, sua maior abundância nas biópsias dos indivíduos saudáveis (Yang et al. 2022). Como se trata de um estudo menor e também com amostras de outra população (China), é esperado que haja variações, tendo em vista a dificuldade de se estabelecer um microbioma *core* de um determinado ambiente, tendo em vista a grande variabilidade devido a distância geográfica entre as populações comparadas (Neu et al. 2021).

Considerando os gêneros enriquecidos nas amostras de lavado gástrico dos pacientes com doença localizada temos um destaque o gênero *Fusarium*, um patógeno já associado a etiologia do câncer de esôfago, devido a maior concentração de suas micotoxinas em milho e trigo em uma região da China com maior risco a desenvolver este tumor (Marasas et al. 1988; Sydenham et al. 1990), mas também esse gênero já foi associado com a ocorrência de infecções em pacientes com tumores pediátricos no Brasil (Arnoni et al. 2018). Também temos o gênero *Trichosporon*, mais abundante em amostras com o CG localizado. Este grupo de fungos apresentam espécies oportunistas, que foram associadas a disbiose fúngica em amostras de câncer colorretal (Każmierczak-Siedlecka et al. 2020) e estavam em maior abundância em amostras de CG no estudo citado (Neu et al. 2021). É possível que a redução da diversidade fúngica intra-amostral em pacientes metastáticos ocorra devido ao agravamento da disbiose já presente em amostras tumorais em estágios iniciais, favorecendo apenas algumas espécies com potencial patogênico que podem estar associadas ao desenvolvimento do câncer, ou o avanço do estadiamento tumoral pode comprometer a imunidade do paciente, permitindo que algumas espécies se desenvolvam mais, agravando ainda

mais a disbiose fúngica. Pelo desenho do estudo (retrospectivo), não é possível estabelecer uma relação direta de causa-efeito para o estudo em questão.

Os demais gêneros identificados como diferencialmente abundantes não foram mais explorados, por não haver estudos na literatura associando-os a doenças humanas. Apesar disso, é importante ter o registro desses fungos, para eventuais trabalhos futuros que possam associá-los novamente ao CG ou até mesmo outras doenças.

**Tabela 13** – Abundância diferencial dos fungos presentes nas amostras de lavado gástrico, a nível de Gênero, entre amostras de gastrite superficial e amostras de câncer gástrico. A metodologia de ANCOM-BC foi utilizada para análise, considerando as informações de sexo, IMC e presença/ausência de *H.pylori* como fatores influenciadores para essa análise. As colunas contendo o termo LFC representam valores de (*Log Fold Change*), e as colunas adjacentes, contém os valores p ajustados, com nível de significância de 0.05.

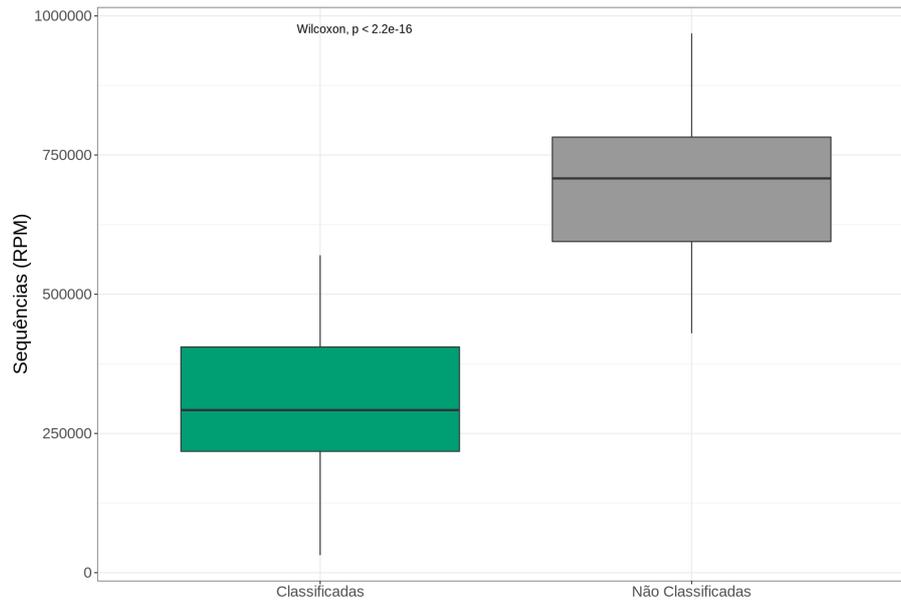
Gênero	LFC-Sexo:M	p-Sexo:M	LFC-IMC	p-IMC	LFC- <i>H.pylori</i> :positivo	p- <i>H.pylori</i> :positivo	LFC-Gastr.Superf.	p-Gastr.Superf.
g_Peniophora	-0.5104	1	0.0739	0.7801	0.0350	1.0000	-0.7911	0.2477
g_Sphaeropsis	0.2273	1	-0.0247	1.0000	-0.6152	1.0000	0.4149	1.0000
g_Cladosporium	-0.4374	1	0.0372	1.0000	0.5148	1.0000	-0.1630	1.0000
g_Aureobasidium	-0.3369	1	-0.0103	1.0000	-0.3944	1.0000	0.2959	1.0000
g_Exophiala	0.2302	1	-0.0001	1.0000	0.2230	1.0000	0.1239	1.0000
g_Aspergillus	0.7510	1	-0.1000	1.0000	0.4569	1.0000	0.5747	1.0000
g_Penicillium	0.1147	1	-0.0231	1.0000	0.4818	1.0000	-0.1521	1.0000
g_Candida	-0.5270	1	-0.0093	1.0000	-0.3261	1.0000	0.2840	1.0000
g_Debaryomyces	-0.0872	1	0.0202	1.0000	0.4532	1.0000	-0.1399	1.0000
g_Issatchenkia	0.1701	1	-0.0204	1.0000	0.3968	1.0000	-0.1515	1.0000
g_Torulaspota	-0.2123	1	-0.0176	1.0000	0.0134	1.0000	0.0836	1.0000
g_Cytospora	0.7183	1	-0.0423	1.0000	-0.5871	1.0000	0.4416	1.0000
g_Fusarium	0.0777	1	-0.0107	1.0000	-0.1175	1.0000	-0.6369	1.0000
g_Humicola	0.2538	1	0.0080	1.0000	0.6673	1.0000	0.1024	1.0000
g_Hyphodontia	-0.2261	1	0.0147	1.0000	0.1945	1.0000	-0.1249	1.0000
g_Perenniporia	-0.1209	1	0.0046	1.0000	-0.0253	1.0000	0.0304	1.0000
g_Phlebia	0.2300	1	-0.0116	1.0000	0.0724	1.0000	0.3130	1.0000
g_Bjerkandera	-0.0769	1	0.0281	1.0000	-0.2229	1.0000	-0.1727	1.0000
g_Trametes	0.3432	1	-0.0116	1.0000	0.0276	1.0000	0.3254	1.0000
g_Cystobasidium	-0.0084	1	0.0164	1.0000	-0.1490	1.0000	0.5094	1.0000
g_Erythrobasidium	0.0262	1	-0.0012	1.0000	-0.3424	1.0000	-0.1434	1.0000
g_Malassezia	-0.3429	1	0.0305	1.0000	1.2440	1.0000	-0.2458	1.0000
g_Rhodotorula	-0.3706	1	0.0477	1.0000	0.3363	1.0000	0.0604	1.0000
g_Filobasidium	-0.3020	1	0.0596	1.0000	1.2189	1.0000	-0.4611	1.0000
g_Naganishia	-0.0500	1	-0.0404	1.0000	1.1559	1.0000	0.1192	1.0000
g_Cryptococcus	0.3491	1	0.0136	1.0000	0.0886	1.0000	-0.5361	1.0000
g_Cutaneotrichosporon	-0.1067	1	-0.0329	1.0000	-0.3406	1.0000	-0.1078	1.0000
g_Trichosporon	-0.1377	1	0.0115	1.0000	-0.9280	0.2793	-0.9468	1.0000

**Tabela 14** – Análise de abundância diferencial dos fungos presentes nas amostras de lavado gástrico, a nível de Gênero, comparando as amostras tumorais com a doença localizada ou doença metastática. A metodologia de ANCOM-BC foi utilizada para análise de abundância diferencial, considerando as informações de sexo, IMC e presença/ausência de *H.pylori* como fatores influenciadores para essa análise. As colunas contendo o termo LFC representam valores de (*Log Fold Change*), e as colunas adjacentes, representam os valores p ajustados, com nível de significância de 0.05. O asterisco (\*) indica valores p significativos corrigidos.

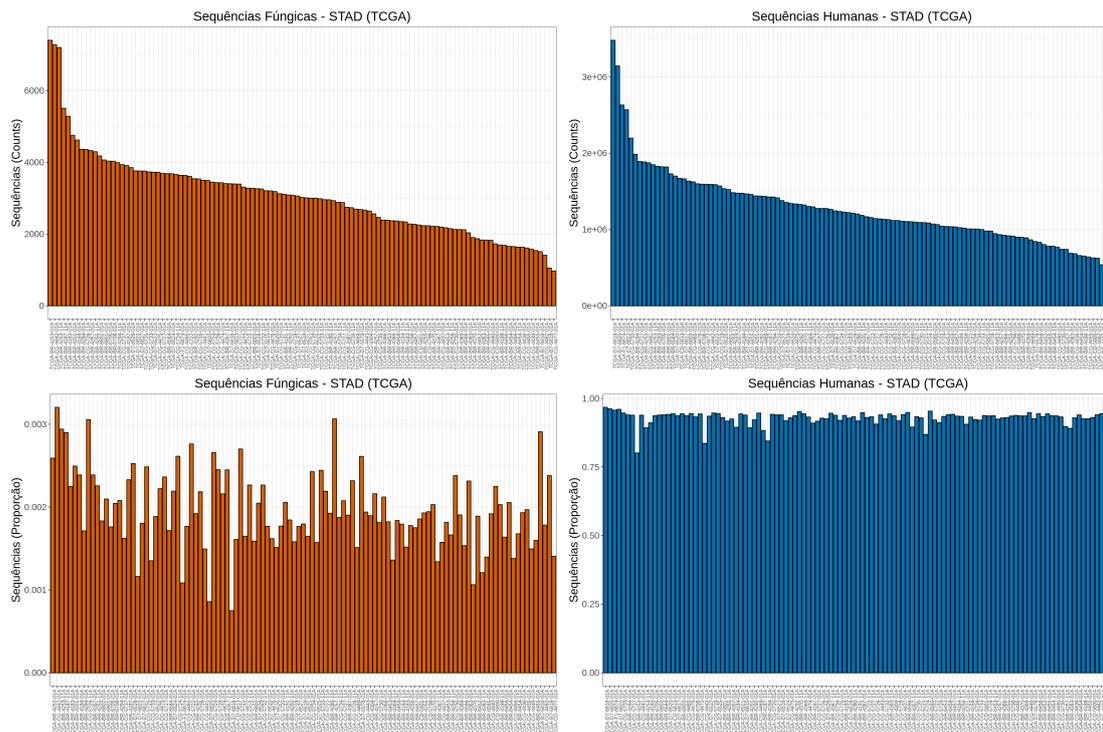
Gênero	LFC-Sexo:M	p-Sexo:M	LFC-IMC	p-IMC	LFC- <i>H.pylori</i> :positivo	p- <i>H.pylori</i> :positivo	LFC-Doença.Metast.	p-Doença.Metast.
g_ Exophiala	0.4933	1	-0.0292	1.0000	0.1378	1	0.7873	0.0000*
g_ Debaryomyces	-0.6098	1	0.1021	1.0000	-0.2397	1	0.3743	0.0000*
g_ Yarrowia	-0.3935	1	0.0014	1.0000	0.5745	1	0.1913	0.0000*
g_ Torulaspora	-0.4967	1	-0.0223	1.0000	0.4085	1	-0.4566	0.0000*
g_ Fusarium	-0.6256	1	0.0701	1.0000	-0.7956	1	-0.4597	0.0000*
g_ Humicola	0.3605	1	-0.0194	1.0000	0.3484	1	-0.2457	0.0000*
g_ Perenniporia	-0.0609	1	0.0133	1.0000	0.0050	1	0.0828	0.0000*
g_ Phlebia	-0.3779	1	0.0397	1.0000	-0.5366	1	0.0775	0.0000*
g_ Bjerkandera	-0.5647	1	0.0622	1.0000	-0.4899	1	0.4950	0.0000*
g_ Trametes	0.6005	1	-0.0468	1.0000	0.3502	1	-0.1865	0.0000*
g_ Peniophora	-0.8682	1	0.1311	0.1681	-0.1577	1	0.8362	0.0000*
g_ Symmetrospora	0.2268	1	0.0075	1.0000	0.1931	1	1.2370	0.0000*
g_ Cystobasidium	0.1178	1	-0.0078	1.0000	0.0393	1	0.7335	0.0000*
g_ Erythrobasidium	0.7120	1	-0.0492	1.0000	0.3498	1	0.1212	0.0000*
g_ Rhodotorula	-0.1307	1	0.0601	1.0000	1.0098	1	0.3483	0.0000*
g_ Filobasidium	-0.4665	1	0.0265	1.0000	0.9948	1	0.4694	0.0000*
g_ Naganishia	-0.1461	1	-0.0457	1.0000	0.6678	1	0.1466	0.0000*
g_ Cryptococcus	1.4538	1	-0.0628	1.0000	0.2682	1	-0.9640	0.0000*
g_ Cutaneotrichosporon	0.0433	1	0.0208	1.0000	-0.4408	1	1.4631	0.0000*
g_ Trichosporon	-0.0744	1	0.0125	1.0000	-1.4808	1	-0.0740	0.0000*
g_ Candida	-0.2943	1	-0.0138	1.0000	-1.1150	1	2.0445	0.3087
g_ Cladosporium	-0.0175	1	0.0279	1.0000	1.5092	1	-0.3589	1.0000
g_ Aureobasidium	0.5726	1	-0.0234	1.0000	-0.6243	1	0.3882	1.0000
g_ Aspergillus	-0.0364	1	-0.0065	1.0000	-0.0152	1	0.8940	1.0000
g_ Penicillium	0.2679	1	-0.0011	1.0000	0.6858	1	-0.5768	1.0000
g_ Issatchenkia	-0.3033	1	0.0126	1.0000	0.6841	1	1.9416	1.0000
g_ Malassezia	-0.5905	1	0.0981	1.0000	1.2880	1	-0.4159	1.0000

#### 4.1.8.2 Análise do microbioma fúngico a partir do sequenciamento total de amostras de adenocarcinoma gástrico do TCGA

Como foi descrito no item 3.1, foram obtidas do repositório TCGA um total de 111 amostras, sendo 82 provenientes de amostras de tumor, 29 de amostras de tecido normal adjacente, e 25 dessas amostras compartilhavam os dados dos dois tipos amostrais foram recuperados. A partir do *pipeline* descrito no item 3.1.9. A Figura 26 apresenta as diferenças das sequências classificadas e não classificadas pelo *software* Kraken2. Mesmo após a remoção das sequências do hospedeiro o *software* classifica a maior parte das sequências como sendo do hospedeiro (Figura 27, abaixo à direita). Ainda não está claro o motivo desse resultado, mas novos testes serão realizados para confirmar esse resultado, como teste com outros parâmetros de remoção pelo *software* SAMTOOLS, uso de outros bancos de referência e também aumentar a especificidade da classificação realizada pelo Kraken2. Apesar disso, a proporção de sequências classificadas como pertencentes ao Reino *Fungi* são condizentes com o observado em outros estudos ( $\leq 1\%$ ) (Figura 27, abaixo à esquerda) (Qin et al. 2010; Huffnagle e Noverr 2013; Nash et al. 2017; Chiu e Miller 2019) .



**Figura 26** – Sequências Classificadas e Não Classificadas pelo *software* **Kraken2** dos dados obtidos do **TCGA**. Boxplot das sequências das 111 amostras de **STAD** obtidas do repositório **TCGA** após filtro para remoção de sequências do hospedeiro.



**Figura 27 – Distribuição das sequências fúngicas e humanas das amostras de STAD - TCGA.** **acima à esquerda e acima à direita** apresentam o número total de sequências classificadas (*Counts*) como pertencentes ao Reino *Fungi* ou a espécie *Homo sapiens* pelo *software* Kraken2, respectivamente; **abaixo à esquerda e abaixo à direita** apresentam os valores da frequência relativa (*Proporção*) em relação as sequências classificadas como pertencentes ao Reino *Fungi* ou a espécie *Homo sapiens*, respectivamente. No eixo x estão identificadas as amostras ordenadas de acordo com os valores em RPM das sequências classificadas.

A partir da classificação taxonômica das sequências obtidas, foram selecionadas as sequências fúngicas (classificadas dentro do Reino *Fungi*) para análises posteriores. O número total de sequências classificadas para cada nível taxonômico foi convertido em frequência relativa ao total de sequências fúngicas. Em seguida, foi verificado se havia diferença significativamente estatística entre a frequência relativa de cada fungo identificado (de acordo com cada nível taxonômico) dentro de cada grupo de informação clínica obtida das amostras do TCGA. Inicialmente, realizamos um teste mais abrangente para ter uma visão ampla da população fúngica identificada. A Figura 28 apresenta um *heatmap* contendo todos os gêneros fúngicos identificados, sendo que os valores p (na escala  $-\log_{10}$ , nível de significância de 5%) estão representados em cores diferentes. As cores

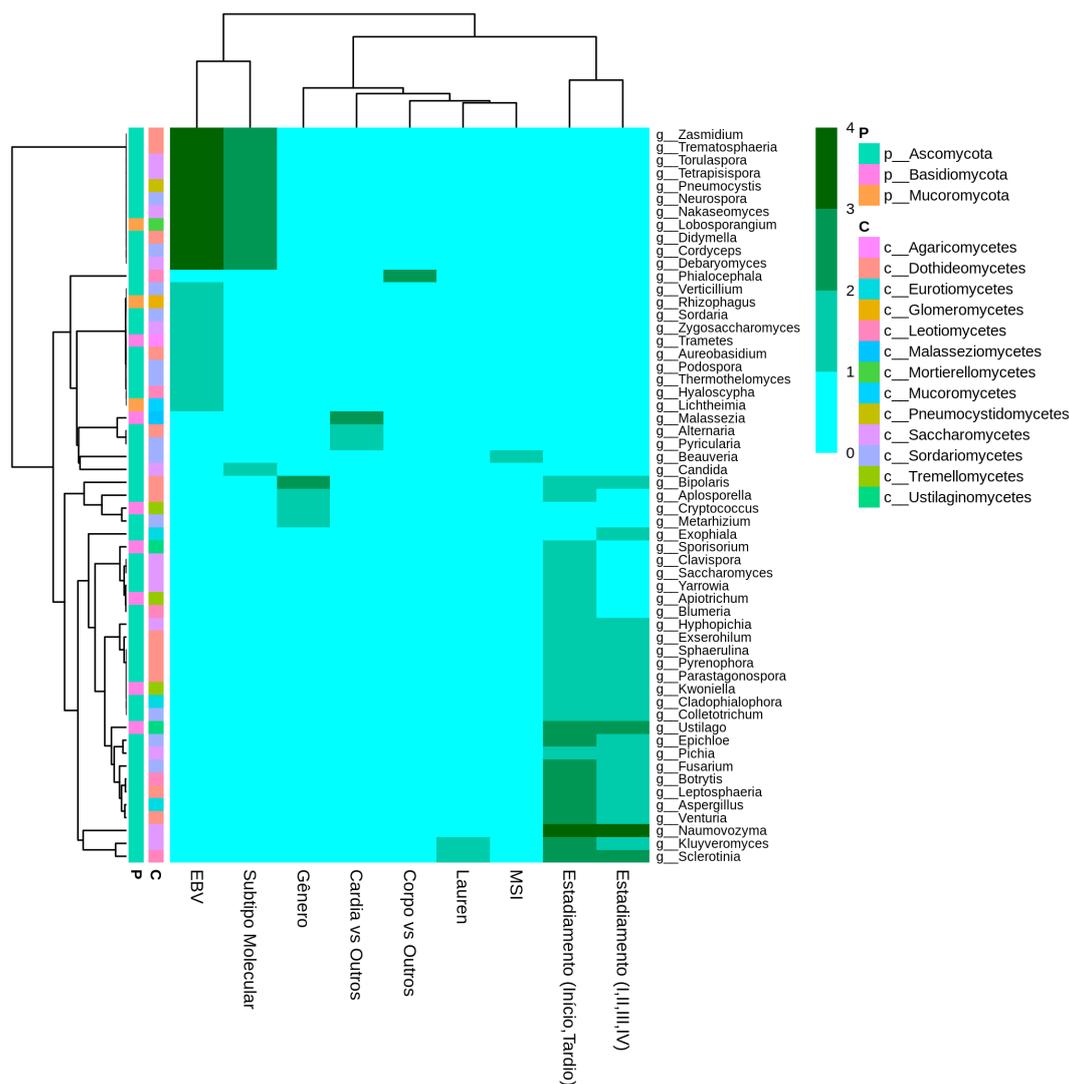
mais escuras representam menores valores de p, sendo os valores acima de 1,3 na escala apresentada representam valores menores que 0,05. Podemos observar que os gêneros *Kluyveromyces* e *Sclerotinia* apresentam frequências diferencialmente significativas entre os diferentes tipos de classificação histológica baseada nos critérios de Lauren (Lauren 1965). Essa classificação foi associada a diferentes prognósticos em relação a sobrevida geral de 5 anos, sendo o tipo intestinal apresentando melhor sobrevida em relação aos demais tipos (Chen et al. 2016).

Em seguida, aqueles gêneros fúngicos que apresentaram diferenças estatisticamente significativas foram analisados em maiores detalhes dentro das respectivas. A Figura 29 apresenta exemplos dessa análise detalhada. Os gêneros *Sclerotinia* e *Kluyveromyces* apresentaram diferenças em relação à classificação histológica de Lauren (Figura 29, acima à esquerda e acima à direita, respectivamente). Apesar da diferença na frequência relativa ser pequena, existe uma tendência de maior abundância dos gêneros em questão no tipo Misto do que nos tipos Difuso e Intestinal. Em particular, temos a espécie *Kluyveromyces fragilis*, já reportada anteriormente como um patógeno oportunista por causar um quadro de infecção pulmonar em um paciente cardíaco transplantado imunossuprimido (Lutwick et al. 1980).

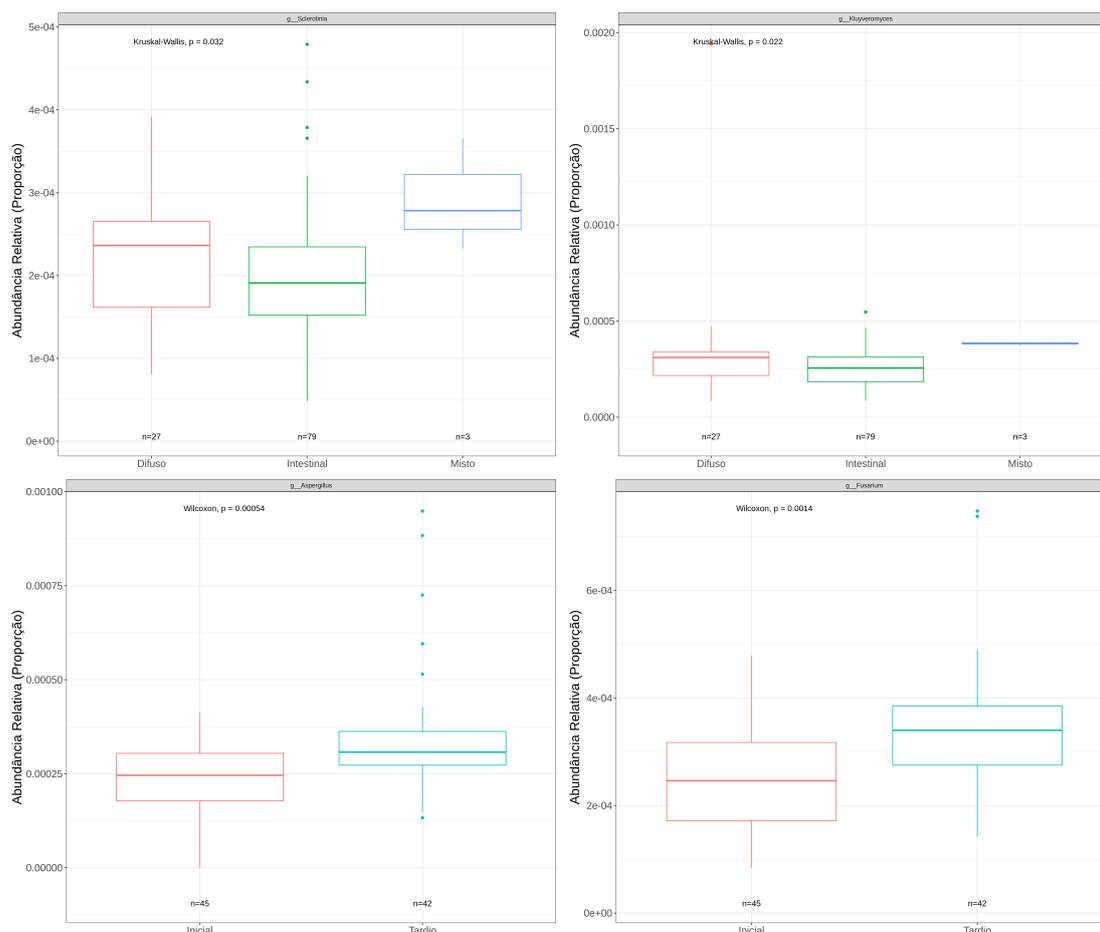
Também é possível verificar que existe uma diferença de abundância nos gêneros *Aspergillus* e *Fusarium* quando dividimos o estadiamento tumoral clínico em dois grupos: estágios iniciais (I e II) e tardios (III e IV) (Figura 28), sendo que ambos os gêneros são mais abundantes nos tumores de estadiamento mais avançado. O gênero *Aspergillus* possui espécies capazes de desenvolver uma micose oportunista, chamada de Aspergilose, em indivíduos imunossuprimidos, sendo a espécie *A.fumigatus* responsável por 90% das infecções no ser humano (McCormick et al. 2010; Jiang et al. 2013). Do mesmo modo, não é possível estabelecer relações diretas entre a maior abundância de um fungo em um determinado grupo de amostras, mas existem indícios que a microbiota bacteriana tumoral é menos diversa que amostras de indivíduos livres de tumor (Rajagopala

et al. 2017), é possível que o microbioma de tumores em estágios mais avançados apresentem menor diversidade fúngica, como observado nas amostras de lavado gástrico sequenciadas para a região ITS2 (Figura 25, abaixo à esquerda). Além disso, para o conjunto de amostras do TCGA o gênero *Fusarium* se mostrou mais abundante em amostras de estadiamento mais avançado, em concordância com o estudo recente de microbioma e CG (Neu et al. 2021), mas discordante do achado na análise das sequências de ITS2 (Figura 14).

Não é possível relacionar diretamente a maior abundância de um determinado fungo com uma redução da sobrevida geral ou com o agravamento do estadiamento do tumor, mas o objetivo deste estudo é identificar variações no microbioma fúngico que possam complementar o conhecimento em relação ao diagnóstico e também ao prognóstico dos tumores investigados, principalmente considerando a variação presente na população brasileira.



**Figura 28 – Comparação da presença de gêneros fúngicos em relação às informações clínicas e moleculares das amostras de STAD - TCGA.** *Heatmap* com os p valores na escala  $-\log_{10}$  em relação a comparação de cada gênero (linhas) em cada variável analisada (colunas). Foram realizados os testes estatísticos Wilcoxon-Mann-Whitney ou Kruskal-Wallis, de acordo com o número de grupos de cada variável, considerando o nível de significância sendo 5%. Uma diferença significativa é representada por valores maiores que 1,3, representados por cores mais escuras. Também foram adicionadas as classificações taxonômicas em relação aos níveis de Filo e Classe (P e C, respectivamente).



**Figura 29 –** Frequência relativa de gêneros fúngicos identificados nos dados de WGS em relação a informações das amostras STAD - TCGA. acima à esquerda, acima à direita. Abundância relativa dos gêneros *Sclerotinia*, *Kluyveromyces* em relação à classificação histológica de Lauren, respectivamente; abaixo à esquerda, abaixo à direita. Abundância relativa dos gêneros *Aspergillus* e *Fusarium* em relação ao estadiamento clínico tumoral, subdivido em estágios Inicial (compreende os estágios I e II) e Tardio (compreende os estágios III e IV), respectivamente. Os testes estatísticos realizados foram Wilcoxon-Mann-Whitney ou Kruskal-Wallis, de acordo com o número de grupos de cada variável, considerando o nível de significância sendo 5%.

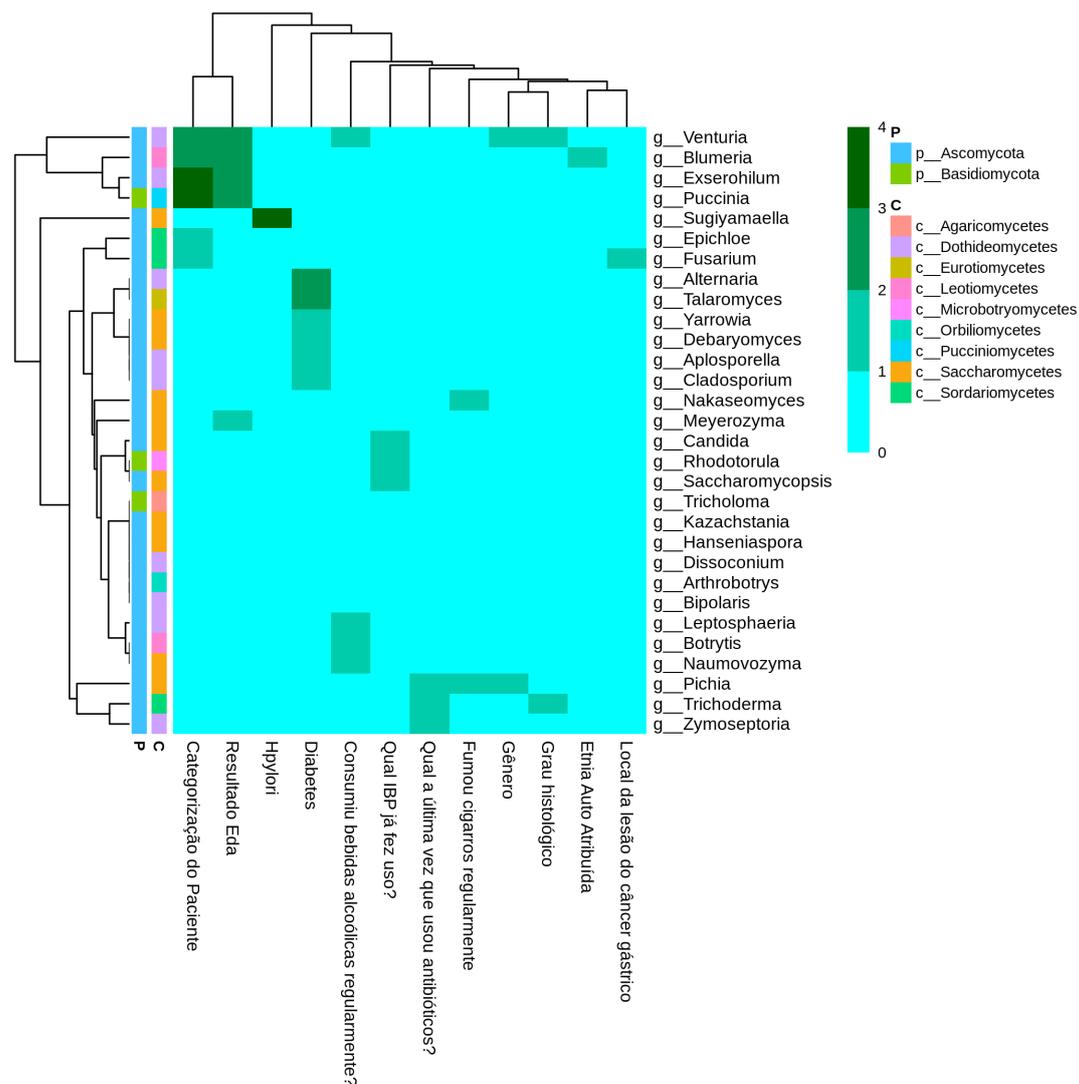
#### 4.1.8.3 Sequenciamento Genômico Total de Amostras de Lavado Gástrico (ACCC) para investigação do microbioma fúngico

Assim como realizado com os dados obtidos do TCGA, foi realizada uma análise mais abrangente a partir das sequências fúngicas obtidas do WGS das 90 amostras de lavado gástrico selecionadas para este experimento (32 casos de gastrite superficial, 1 gastrite crônica, 25 metaplasia Intestinal, 35 casos de CG). Foi realizada a normalização das sequências para frequência relativa e compara-

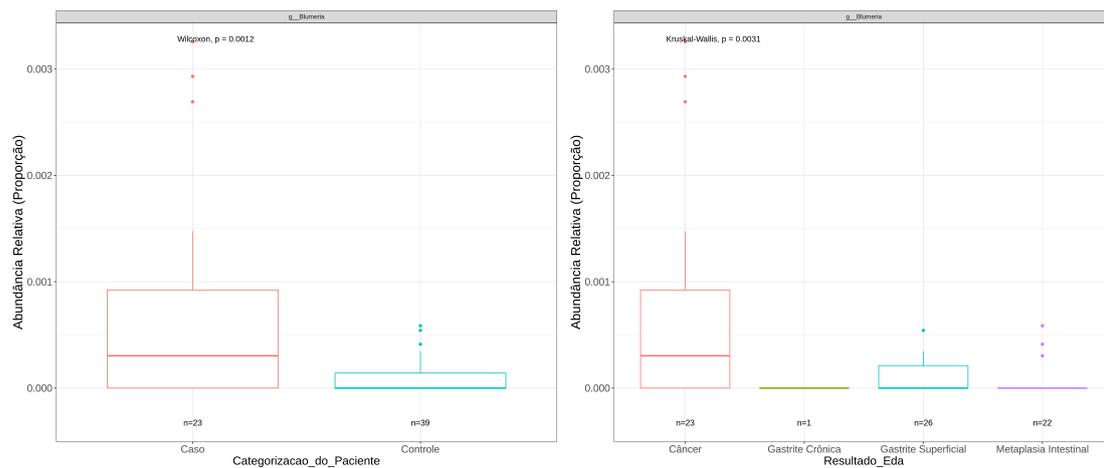
mos as abundâncias dos fungos anotados em relação às informações disponíveis (tabela 11). É possível observar uma concordância com o resultado obtido, já que a diferença verificada nos gêneros dentro do grupo de categorização do paciente também tem, em sua maioria, abundância diferenciada de acordo com o resultado obtido na endoscopia alta (Resultado Eda) (Figura 30). Os gêneros destacados nos dois grupos são comumente reportados como patógenos de plantas (*Venturia*, *Blumeria*, *Exserohilum* e *Puccinia*) (Gauthier e Keller 2013; Santra e Banerjee 2020), sendo que suas micotoxinas são agentes potenciais que podem levar a doenças em humanos, como câncer, dano ao fígado, falha dos rins e paralisia (Miller 1995). Verificamos ainda que, a maior abundância do gênero *Blumeria* ocorre em amostras provenientes de pacientes portadores de adenocarcinoma gástrico quando comparadas com amostras de indivíduos livres de tumor (Figura 31).

Apesar de não ser possível definir uma relação de causalidade, esses resultados contribuem para o entendimento que a microbiota fúngica também sofre alterações, assim como a microbiota bacteriana, amplamente estudada, que revelaram a diferença do microbioma de amostra tumorais em relação a amostras controle para câncer gástrico (Coker et al. 2018; Ferreira et al. 2018). Esses resultados também são promissores pela inovação do uso do material do lavado gástrico coletado do procedimento de endoscopia, sem a necessidade de utilizar métodos mais invasivos como a coleta de material para biópsia. Apesar de haver uma complementariedade entre as amostras processadas para o *metabarcoding* e as amostras WGS (73 amostras em comum), não foi verificado nenhum gênero em comum em relação as variáveis analisadas. Um resultado similar também foi observado na investigação do micobioma do Projeto Microbioma Humano, em que os achados em WGS não corresponderam ao observado para os resultados de ITS2, sendo que se tratava de um maior conjunto amostral provenientes de fezes de indivíduos saudáveis (317 amostras de 147 voluntários) (Nash et al. 2017). Assim como verificado neste trabalho, apenas uma pequena porcentagem das sequências geradas pelo WGS são alinhadas aos genomas fúngicos (0.01%), o que

pode indicar que o dados de WGS não atingiram a profundidade necessária para compreender toda a diversidade fúngica presente nas amostras e a dificuldade de se obter bons resultados para esses tipos amostrais, em que há um predomínio de DNA bacteriano (Nash et al. 2017).



**Figura 30 – Comparação da presença de gêneros fúngicos identificados no sequenciamento de DNA total em relação às informações clínicas e moleculares das amostras de lavado gástrico - ACCCC. Heatmap com os valores p na escala  $-\log_{10}$  em relação a comparação de cada gênero (linhas) em cada variável analisada (colunas). Foram realizados os testes estatísticos Wilcoxon-Mann-Whitney ou Kruskal-Wallis, de acordo com o número de grupos de cada variável, considerando o nível de significância sendo 5%. Uma diferença significativa é representada por valores maiores que 1.3, representados por cores mais escuras. Também foram adicionadas as classificações taxonômicas em relação aos níveis de Filo e Classe (P e C, respectivamente).**



**Figura 31 – Frequência relativa de gêneros fúngicos identificados no sequenciamento de DNA total em relação a informações das amostras de lavado gástrico - ACCCC. à esquerda e à direita.** Abundância relativa do gênero *Blumeria* em relação à categorização do paciente e resultado da endoscopia ao diagnóstico, respectivamente; Os testes estatísticos realizados foram Wilcoxon-Mann-Whitney ou Kruskal-Wallis, de acordo com o número de grupos de cada variável, considerando o nível de significância sendo 5%.

## 5 CONCLUSÕES

Durante o todo o projeto trabalhamos para a construção de um *pipeline* voltado para o estudo do microbioma humano, *pipeline* este desenvolvido desde a seleção dos oligonucleotídeos para amplificação da região de interesse, até a análise bioinformática mais apropriada para o estudo de sequências fúngicas. Foram selecionados o par de *primers* mais abrangente e específico dentre a diversidade que levantamos da literatura para a amplificação da região de interesse com maior informação taxonômica (ITS2). Também foram avaliados diversos bancos de sequências fúngicas para selecionar aquele mais abrangente para o nosso estudo, tendo em vista que o objetivo principal do projeto era caracterizar uma população fúngica pouco estudada, principalmente em relação ao câncer de pênis tumor de incidência rara, mas relevante para o cenário do Brasil.

Também padronizamos o protocolo de extração do DNA fúngico para os tipos de materiais a serem avaliados no projeto, realizando uma extensa avaliação de protocolos, tanto em questão de rendimento, quanto em relação aos resultados obtidos, de modo a garantir protocolos eficientes para o processamento das amostras do projeto. Após a seleção dos oligonucleotídeos e padronização dos protocolos de extração, foram realizados inúmeros testes para otimizar a reação de PCR da região ITS2, tendo em vista a grande diversidade de rendimento de cada amostra, principalmente das amostras de lavado gástrico. Por fim adaptamos um *pipeline* de processamento e análise bioinformática tanto para dados de *metabarcoding*, quanto de WGS voltados para o microbioma.

Paralelamente ao estabelecimento dos procedimentos, trabalhamos para a criação de um modelo estatístico para estimar com maior precisão a frequência dos representantes fúngicos de uma amostra, considerando as variações intrínsecas da região sequenciada (ITS2) que resultam em amplificação desigual durante a etapa de PCR e durante o sequenciamento. Apesar do modelo apresentar

resultados animadores para alguns conjuntos separados de dados obtidos da literatura, o modelo se mostrou ter eficácias diferentes a depender do conjunto de dados testado. Acreditamos que a composição das amostras e o efeito de *batch* possam influenciar no modelo e não conseguimos avançar a partir deste ponto. De todo modo, será interessante voltar nesse ponto pois a comparação de dados de sequenciamento de diferentes estudos é uma prática importante e verificamos que as sequências fúngicas tem maior propensão de não serem corretamente estimadas quando comparadas com sequências de marcadores bacterianos (16S). Além disso, foi possível publicar um artigo como segundo autor que trabalhou no desenvolvimento de uma metodologia baseada em qPCR capaz de estimar a razão entre o DNA bacteriano ou fúngico presente em diferentes tipos de amostra e o DNA do hospedeiro, de modo a avaliar qual a melhor metodologia para o estudo do microbioma de determinado tipo amostral (*metabarcoding* ou WGS) (Apêndice A).

Com os protocolos estabelecidos realizamos o processamento e sequenciamento, e análise das amostras. Infelizmente tivemos um problema com o sequenciador para as bibliotecas geradas para a região ITS2 das amostras de pênis, mas ainda assim conseguimos identificar algumas diferenças interessante entre as amostras, como a presença de diferença entre a diversidade fúngica entre as amostras de São Paulo e Amazonas, o que indica a grande variabilidade existente no país e reforça a dificuldade de se definir um microbioma *core* geral para cada local do corpo humano. Também verificamos uma menor diversidade em amostras com maior estadiamento patológico e que a população fúngica também é diferente para essas amostras. E ainda, identificamos gêneros fúngicos diferencialmente abundantes tanto nas amostras neoplásicas como nas não neoplásicas, o que reforça que sejam realizados mais estudos confirmar os achados deste projeto e investigar sua relação com o desenvolvimento e progressão do câncer de pênis.

Já para o estudo do microbioma em amostras de lavado gástrico foi possível realizar uma análise mais abrangente, apresentando dados da região ITS2, dados

de WGS e dados de WGS obtidos para esse tumor da literatura (TCGA). Apesar de não haver grande concordância entre os resultados, foi possível observar que a diversidade fúngica intra-amstral vai diminuindo conforme a progressão da doença, além disso, é possível que exista alguma interação bacteriana-fúngica, tendo em vista que as amostras diagnosticadas como *H.pylori* positivas apresentaram uma maior diversidade fúngica intra-amstral em relação às amostras negativas. Esse resultado reforça a importância realizar mais estudos voltados para a interação entre os dois grupos de microrganismos e suas possíveis implicações para manutenção da homeostase. Ao analisar a diversidade inter-amstral, verificamos que a população fúngica também varia de acordo a presença de *H.pylori*, estadiamento (amostra localizada ou metastática), tabagismo e diabetes, em diferentes níveis taxonômicos, mas isso mostra a importância de se considerar a porção fúngica nas análises de microbioma humano. Já a análise de abundância diferencial apresentou resultados diferentes entre os sequenciamentos da região ITS2 e WGS feito no projeto, e também quando comparados com dados do TCGA e de outro trabalho recente que estudou o microbioma associado ao câncer gástrico. Apesar das diferenças os resultados só reforçam que mais estudos sejam feitos, com amostras maiores e grupos mais bem definidos para buscar uma caracterização mais detalhada do microbioma fúngico associado a tumores.

Como limitações do projeto tivemos que superar desafios ao trabalhar com uma área não tão desenvolvida quanto a análise do microbioma bacteriano, mas o ponto principal foi o tempo e insumos perdidos que ocorreu durante os sequenciamentos das amostras, o que acarretou no comprometimento do sequenciamento das amostras de pênis, apesar de ter sido possível utilizar os dados, acabamos não tendo o melhor dado possível para estudar esse tumor raro. Essa questão com o sequenciamento acabou demandando um tempo muito maior que o planejado, pois foi necessário repetir o preparo de todas as bibliotecas já realizada, muitas vezes da PCR, mas em alguns casos desde a etapa de extração do DNA, o que acabou afetando o tempo para nos dedicarmos a análise dos da-

dos gerados. Para futuras publicações serão realizadas análises mais detalhadas, principalmente para os dados de WGS, como a análise das funções metabólicas do microbioma identificado, buscando também correlacionar os fungos anotados com os demais grupos presentes (bactérias, vírus, protozoários, etc).

É importante ressaltar que o principal objetivo deste experimento foi o caracterizar inicialmente a população fúngica dos tumores de interesse, com amostras da população brasileira, além de comparar com amostras livres de tumor, buscando a identificação de marcadores biológicos, tanto de desenvolvimento, quanto para progressão da doença, mas que devem ser comprovados com estudos desenhados especificamente para esse propósito.

## 6 REFERÊNCIAS

Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*. 2008; 6:431–440.

Afonso LA, Moyses N, Alves G, Ornellas AA, Passos MRL, Oliveira LdHdS, et al. Prevalence of human papillomavirus and epstein-barr virus dna in penile cancer cases from brazil. *Memórias do Instituto Oswaldo Cruz*. 2012; 107:18–23.

Aita G, Costa WH, Cassio Zequi S, Cunha IW, Soares F, Guimaraes GC, et al. Pattern of invasion is the most important prognostic factor in patients with penile cancer submitted to lymph node dissection and pathological absence of lymph node metastasis. *BJU international*. 2015; 116:584–589.

Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric adenocarcinoma. *Nature reviews Disease primers*. 2017; 3:1–19.

de Albuquerque GE, Moda BS, Serpa MS, Branco GP, Defelicibus A, Takenaka IK, et al. Evaluation of bacteria and fungi dna abundance in human tissues. *Genes*. 2022; 13:237.

Allaband C, McDonald D, Vázquez-Baeza Y, Minich JJ, Tripathi A, Brenner DA, et al. Microbiome 101: studying, analyzing, and interpreting gut microbiome data for clinicians. *Clinical gastroenterology and hepatology*. 2019; 17:218–230.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990; 215:403–410.

Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*. 1995; 59:143–169.

Amend AS, Seifert KA, Samson R, Bruns TD. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences*. 2010; 107:13748–13753.

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*. 2017; 2:e00191–16.

Amorim CA, Moreira JP, Rial L, Carneiro AJ, Fogaça HS, Elia C, et al. Ecological study of gastric cancer in brazil: geographic and time trend analysis. *World journal of gastroenterology: WJG*. 2014; 20:5036.

Arnoni MV, Paula CR, Auler ME, Simões CCN, Nakano S, Szeszs MW, et al. Infections caused by fusarium species in pediatric cancer patients and review of published literature. *Mycopathologia*. 2018; 183:941–949.

Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *science*. 2012; 338:120–123.

Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4fun: predicting functional profiles from metagenomic 16s rRNA data. *Bioinformatics*. 2015; 31:2882–2884.

Aykut B, Pushalkar S, Chen R, Li Q, Abengozar R, Kim JI, et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of mbl. *Nature*. 2019; 574:264–267.

Barreira<sup>1</sup> MA, de Oliveira Lima L, Júnior JJA, Silva LFG, Lima MVA. Experiência do hospital Haroldo Juaçaba com reconstrução utilizando retalhos miocutâneos em cirurgia para tratamento do câncer de pênis locorregionalmente avançado. *Revista Brasileira de Cancerologia*. 2014; 60:43–50.

Bashiardes S, Zilberman-Schapira G, Elinav E. Use of metatranscriptomics in microbiome research. *Bioinformatics and biology insights*. 2016; 10:BBI-S34610.

Bassam BJ, Gresshoff PM. Silver staining DNA in polyacrylamide gels. *Nature protocols*. 2007; 2:2649–2654.

Bergeron MJ, Feau N, Stewart D, Tanguay P, Hamelin RC. Genome-enhanced detection and identification of fungal pathogens responsible for pine and poplar rust diseases. *PloS one*. 2019; 14:e0210952.

Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, et al. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and structural biotechnology journal*. 2015; 13:390–401.

Blackwell M. The fungi: 1, 2, 3... 5.1 million species? *American journal of botany*. 2011; 98:426–438.

Boers SA, Jansen R, Hays JP. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *European Journal of Clinical Microbiology & Infectious Diseases*. 2019; 38:1059–1070.

Bokulich NA, Mills DA. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl Environ Microbiol*. 2013; 79:2519–2526.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*. [S.l.], 2018.

Borrel G, Brugère JF, Gribaldo S, Schmitz RA, Moissl-Eichinger C. The host-associated archaeome. *Nature Reviews Microbiology*. 2020; 18:622–636.

- Bourgeois S, Kaden J, Senn H, Bunnefeld N, Jeffery KJ, Akomo-Okoue EF, et al. Improving cost-efficiency of faecal genotyping: New tools for elephant species. *PloS one*. 2019; 14:e0210811.
- Bradford LL, Ravel J. The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence*. 2017; 8:342–351.
- Bray JR, Curtis JT. An ordination of the upland forest communities of southern wisconsin. *Ecological monographs*. 1957; 27:326–349.
- Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *Journal of bacteriology*. 2003; 185:6220–6223.
- Brown GD, Denning DW, Gow NA, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Science translational medicine*. 2012; 4:165rv13–165rv13.
- Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for arabidopsis root-inhabiting bacterial microbiota. *Nature*. 2012; 488:91–95.
- Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal*. 2017; 11:2639–2643.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. Dada2: High-resolution sample inference from illumina amplicon data. *Nature methods*. 2016; 13:581–583.
- Chandler JA, Liu RM, Bennett SN. Rna shotgun metagenomic sequencing of northern california (usa) mosquitoes uncovers viruses, bacteria, and fungi. *Frontiers in microbiology*. 2015; 6:185.
- Chen YC, Fang WL, Wang RF, Liu CA, Yang MH, Lo SS, et al. Clinicopathological variation of lauren classification in gastric cancer. *Pathology & Oncology Research*. 2016; 22:197–202.
- Chi AC, Day TA, Neville BW. Oral cavity and oropharyngeal squamous cell carcinoma—an update. *CA: a cancer journal for clinicians*. 2015; 65:401–421.
- Chiarello M, McCauley M, Villéger S, Jackson CR. Ranking the biases: The choice of otus vs. asvs in 16s rrna amplicon data analysis has stronger effects on diversity measures than rarefaction and otu identity threshold. *PloS one*. 2022; 17:e0264443.
- Chiu CY, Miller SA. Clinical metagenomics. *Nature Reviews Genetics*. 2019; 20:341–355.
- Chon HJ, Hyung WJ, Kim C, Park S, Kim JH, Park CH, et al. Differential prognostic implications of gastric signet ring cell carcinoma: stage adjusted analysis from a single high-volume center in asia. *Annals of surgery*. 2017; 265:946.

- Claesson MJ, Clooney AG, O'toole PW. A clinician's guide to microbiome analysis. *Nature Reviews Gastroenterology & Hepatology*. 2017; 14:585–595.
- Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Australian journal of ecology*. 1993; 18:117–143.
- Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut*. 2017;;gutjnl–2017.
- Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut*. 2018; 67:1024–1032.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rrna analysis. *Nucleic acids research*. 2013; 42:D633–D642.
- Cope EK, Goldberg AN, Pletcher SD, Lynch SV. A chronic rhinosinusitis–derived isolate of *Pseudomonas aeruginosa* induces acute and pervasive effects on the murine upper airway microbiome and host immune response. In: WILEY ONLINE LIBRARY. *International forum of allergy & rhinology*. [S.l.], 2016. 6, n. 12, 1229–1237.
- Costa S, da Silva Rodrigues RS, Barbosa L, dos Santos Silva J, de Caldas Brandão JO, de Medeiros C. Câncer de pênis: epidemiologia e estratégias de prevenção. In: . [S.l.: s.n.], 2013.
- Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science*. 2012; 336:1255–1262.
- Danko D, Bezdan D, Afshin EE, Ahsanuddin S, Bhattacharya C, Butler DJ, et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell*. 2021;:.
- Dash NR, Khoder G, Nada AM, Al Bataineh MT. Exploring the impact of *Helicobacter pylori* on gut microbiome composition. *PloS one*. 2019; 14:e0218274.
- Delhaes L, Monchy S, Fréalle E, Hubans C, Salleron J, Leroy S, et al. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community—implications for therapeutic management. *PloS one*. 2012; 7:e36313.
- Deshpande V, Wang Q, Greenfield P, Charleston M, Porrás-Alfaro A, Kuske CR, et al. Fungal identification using a bayesian classifier and the warcup training set of internal transcribed spacer sequences. *Mycologia*. 2016; 108:1–5.
- Diaz PI, Hong BY, Dupuy AK, Strausbaugh LD. Mining the oral mycobioime: Methods, components, and meaning. *Virulence*. 2017; 8:313–323.
- Diaz PI, Hong BY, Dupuy AK, Choquette L, Thompson A, Salner AL, et al. Integrated analysis of clinical and microbiome risk factors associated with the development of oral candidiasis during cancer chemotherapy. *Journal of Fungi*. 2019; 5:49.

- Dobell C. The discovery of the intestinal protozoa of man. 1. Proceedings of the Royal Society of Medicine. 1920; 13:1–15.
- Dollive S, Peterfreund GL, Sherrill-Mix S, Bittinger K, Sinha R, Hoffmann C, et al. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome biology*. 2012; 13:R60.
- Douglas GM, Maffei VJ, Zaneveld J, Yurgel SN, Brown JR, Taylor CM, et al. Picrust2: An improved and customizable approach for metagenome inference. *BioRxiv*. 2020;:672295.
- Edgar RC. Utax accurately predicts taxonomy of marker gene sequences;:.
- Edgar RC. Search and clustering orders of magnitude faster than blast. *Bioinformatics*. 2010; 26:2460–2461.
- Edgar RC. Uparse: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*. 2013; 10:996–998.
- Edgar RC. Syntax: a simple non-bayesian taxonomy classifier for 16S and its sequences. *bioRxiv*. 2016;:074161.
- Edgar RC. UNOISE2: improved error-correction for Illumina 16S and its amplicon sequencing. *BioRxiv*. 2016;:081257.
- Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*. 2015; 31:3476–3482.
- Egger KN. Molecular analysis of ectomycorrhizal fungal communities. *Canadian Journal of Botany*. 1995; 73:1415–1422.
- Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, Riaz T, et al. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular ecology*. 2012; 21:1821–1833.
- Ercolini D. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied and environmental microbiology*. 2013; 79:3148–3155.
- Faveri M, Mayer M, Feres M, De Figueiredo L, Dewhirst F, Paster B. Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Molecular Oral Microbiology*. 2008; 23:112–118.
- Favorito LA, Nardi AC, Ronalsa M, Zequi SC, Sampaio FJ, Glina S. Epidemiologic study on penile cancer in Brazil. *International braz j urol*. 2008; 34:587–593.
- Feau N, Vialle A, Allaire M, Maier W, Hamelin RC. DNA barcoding in the rust genus *Chrysomyxa* and its implications for the phylogeny of the genus. *Mycologia*. 2011; 103:1250–1266.

- Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing rna-seq, 16s rna gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014; 2:1–13.
- Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, et al. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut*. 2018; 67:226–236.
- De Filippis F, Laiola M, Blaiotta G, Ercolini D. Different amplicon targets for sequencing-based studies of fungal diversity. *Appl Environ Microbiol*. 2017; 83:e00905–17.
- Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, et al. Topographic diversity of fungal and bacterial communities in human skin. *Nature*. 2013; 498:367.
- Frey UH, Bachmann HS, Peters J, Siffert W. Pcr-amplification of gc-rich regions: 'slowdown pcr'. *Nature protocols*. 2008; 3:1312–1317.
- Fricker AM, Podlesny D, Fricke WF. What is new and relevant for sequencing-based microbiome research? a mini-review. *Journal of advanced research*. 2019; 19:105–112.
- Galvão-Moreira LV, da Cruz MCFN. Oral microbiome, periodontitis and risk of head and neck cancer. *Oral oncology*. 2016; 53:17–19.
- Garbelotto MM, Lee HK, Slaughter G, Popenuck T, Cobb FW, Bruns TD. Heterokaryosis is not required for virulence of heterobasidion annosum. *Mycologia*. 1997; 89:92–102.
- Gardes M, Bruns TD. Its primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*. 1993; 2:113–118.
- Garrett WS. Cancer and the microbiota. *Science*. 2015; 348:80–86.
- Gauthier GM, Keller NP. Crossover fungal pathogens: the biology and pathogenesis of fungi capable of crossing kingdoms to infect plants and humans. *Fungal Genetics and Biology*. 2013; 61:146–157.
- Gdanetz K, Benucci GMN, Vande Pol N, Bonito G. CONSTAX: A tool for improved taxonomic resolution of environmental fungal ITS sequences. *BMC Bioinformatics*. 2017; 18:1–9.
- Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. *science*. 2006; 312:1355–1359.
- Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends in biotechnology*. 2017; 35:498–507.

- Gotelli NJ, Colwell RK. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters*. 2001; 4:379–391.
- Gray MW, Burger G, Lang BF. Mitochondrial evolution. *Science*. 1999; 283:1476–1481.
- Greenberg JH. The measurement of linguistic diversity. *Language*. 1956; 32:109–115.
- Greiner TU, Bäckhed F. Microbial regulation of glp-1 and l-cell biology. *Molecular metabolism*. 2016; 5:753–758.
- Grice EA, Segre JA. The human microbiome: our second genome. *Annual review of genomics and human genetics*. 2012; 13:151.
- Hanson B, Zhou Y, Bautista EJ, Urch B, Speck M, Silverman F, et al. Characterization of the bacterial and fungal microbiome in indoor dust and outdoor air samples: a pilot study. *Environmental Science: Processes & Impacts*. 2016; 18:713–724.
- Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. *PloS one*. 2016; 11:e0154090.
- Hofstra J, Matamoros S, Van De Pol M, De Wever B, Tanck M, Wendt-Knol H, et al. Changes in microbiota during experimental human rhinovirus infection. *BMC infectious diseases*. 2015; 15:336.
- Holmes E, Li JV, Marchesi JR, Nicholson JK. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell metabolism*. 2012; 16:559–564.
- Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012; 336:1268–1273.
- Huffnagle GB, Noverr MC. The emerging world of the fungal microbiome. *Trends in microbiology*. 2013; 21:334–341.
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, et al. Association of nod2 leucine-rich repeat variants with susceptibility to crohn's disease. *Nature*. 2001; 411:599–603.
- Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD. Forgotten fungi—the gut mycobiome in human health and disease. *FEMS microbiology reviews*. 2017; 41:479–511.
- Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. *nature*. 2012; 486:207.

- Ihrmark K, Bödeker I, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, et al. New primers to amplify the fungal *its2* region—evaluation by 454-sequencing of artificial and natural communities. *FEMS microbiology ecology*. 2012; 82:666–677.
- Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, et al. Interactions between commensal fungi and the c-type lectin receptor dectin-1 influence colitis. *Science*. 2012; 336:1314–1317.
- Jaccard P. Contribution au problème de l’immigration post-glaciare de la flore alpine. *Bull Soc Vaudoise Sci Nat*. 1900; 36:87–130.
- Jass J. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*. 2007; 50:113–130.
- Jemal A, Ward E, Thun M. Declining death rates reflect progress against cancer. *PloS one*. 2010; 5:e9584.
- Jeske JT, Gallert C. Microbiome analysis via otu and asv-based pipelines—a comparative interpretation of ecological data in wwtp systems. *Bioengineering*. 2022; 9:146.
- Jiang Z, Wang Y, Jiang Y, Xu Y, Meng B. Vertebral osteomyelitis and epidural abscess due to *aspergillus nidulans* resulting in spinal cord compression: Case report and literature review. *Journal of international medical research*. 2013; 41: 502–510.
- Jo JH, Kennedy EA, Kong HH. Topographical and physiological differences of the skin mycobiome in health and disease. *Virulence*. 2017; 8:324–333.
- Joos L, Beirinckx S, Haegeman A, Debode J, Vandecasteele B, Baeyen S, et al. Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC genomics*. 2020; 21:1–17.
- Krcmery Jr V, Jesenska Z, Spanik S, Gyarfaz J, Nogova J, Botek R, et al. Fungaemia due to *fusarium* spp. in cancer patients. *Journal of hospital infection*. 1997; 36:223–228.
- Kassambara A. *ggpubr: 'ggplot2' Based Publication Ready Plots*. [S.l.], 2020. R package version 0.4.0. Disponível em: <https://CRAN.R-project.org/package=ggpubr>.
- Kaźmierczak-Siedlecka K, Dvořák A, Folwarski M, Daca A, Przewłócka K, Makarewicz W. Fungal gut microbiota dysbiosis and its role in colorectal, oral, and pancreatic carcinogenesis. *Cancers*. 2020; 12:1326.
- Keller M, Zengler K. Tapping into microbial diversity. *Nature Reviews Microbiology*. 2004; 2:141–150.

- Khot PD, Ko DL, Fredricks DN. Sequencing and analysis of fungal rna operons for development of broad-range fungal pcr assays. *Appl Environ Microbiol.* 2009; 75:1559–1565.
- Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol.* 2018 jul.; 16:410–422.
- Kong HH, Segre JA. Skin microbiome: looking back to move forward. *Journal of Investigative Dermatology.* 2012; 132:933–939.
- Kopylova E, Navas-Molina JA, Mercier C, Xu ZZ, Mahé F, He Y, et al. Open-source sequence clustering methods improve the state of the art. *MSystems.* 2016; 1:e00003–15.
- Krivosos DV, Konanov DN, Iliina EN. Funfun: Its-based functional annotator of fungal communities. *bioRxiv.* 2022;.
- Kukreja SL, Löfberg J, Brenner MJ. A least absolute shrinkage and selection operator (lasso) for nonlinear system identification. *IFAC proceedings volumes.* 2006; 39:814–819.
- Kurtzman C, Robnett C. Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5'end of the large-subunit (26s) ribosomal dna gene. *Journal of clinical microbiology.* 1997; 35:1216–1223.
- Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma: an attempt at a histo-clinical classification. *Acta Pathologica Microbiologica Scandinavica.* 1965; 64:31–49.
- Laursen MF, Dalgaard MD, Bahl MI. Genomic gc-content affects the accuracy of 16s rna gene sequencing based microbial profiling due to pcr bias. *Frontiers in microbiology.* 2017; 8:1934.
- Layke JC, Lopez PP. Gastric cancer: diagnosis and treatment options. *American family physician.* 2004; 69:1133–1140.
- Lecuit M, Eloit M. The human virome: new tools and concepts. *Trends in microbiology.* 2013; 21:510–515.
- Lederberg J, McCray AT. Ome sweetomics—a genealogical treasury of words. *The scientist.* 2001; 15:8–8.
- Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science.* 2010; 330:1768–1773.
- Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics.* 2010; 11:733–739.
- van Leeuwenhoek A. An abstract of a letter from antonie van leeuwenhoek, sep. 12, 1683. about animals in the scruff of the teeth. *Philos Trans R Soc Lond;* 14: 568–574.

- Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome biology*. 2010; 11:R15.
- Levesque CA, De Cock AW. Molecular phylogeny and taxonomy of the genus *pythium*. *Mycological research*. 2004; 108:1363–1383.
- Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*. 2009; 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. *Bioinformatics*. 2009; 25:2078–2079.
- Li R, Tun HM, Jahan M, Zhang Z, Kumar A, Dilantha Fernando W, et al. Comparison of dna-, pma-, and rna-based 16s rna illumina sequencing for detection of live bacteria in water. *Scientific reports*. 2017; 7:1–11.
- Liguori G, Lamas B, Richard ML, Brandi G, Da Costa G, Hoffmann TW, et al. Fungal dysbiosis in mucosa-associated microbiota of crohn’s disease patients. *Journal of Crohn’s and Colitis*. 2015; 10:296–305.
- Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nature communications*. 2020; 11:1–11.
- Liu L, Wang Z, Ji J, Zhang J, Yan M, Zhang J, et al. A cohort study and meta-analysis between histopathological classification and prognosis of gastric carcinoma. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*. 2013; 13:227–234.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with *deseq2*. *Genome biology*. 2014; 15:1–21.
- Low V, Levine M, Rubesin S, Laufer I, Herlinger H. Diagnosis of gastric carcinoma: sensitivity of double-contrast barium studies. *AJR American journal of roentgenology*. 1994; 162:329–334.
- Lozupone C, Hamady M, Knight R. Unifrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC bioinformatics*. 2006; 7:1–14.
- Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative  $\beta$  diversity measures lead to different insights into factors that structure microbial communities. *Applied and environmental microbiology*. 2007; 73:1576–1585.
- Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*. 2017; 3:e104.
- Luan C, Xie L, Yang X, Miao H, Lv N, Zhang R, et al. Dysbiosis of fungal microbiota in the intestinal mucosa of patients with colorectal adenomas. *Scientific reports*. 2015; 5:7980.

- Lutwick LI, Phaff HJ, Stevens DA. *Kluyveromyces fragilis* as an opportunistic fungal pathogen in man. *Sabouraudia: Journal of Medical and Veterinary Mycology*. 1980; 18:69–73.
- Macfarlane GT, Macfarlane S. Bacteria, colonic fermentation, and gastrointestinal health. *Journal of AOAC International*. 2012; 95:50–60.
- Maia APV, de Sousa Silva AB, de Queiroz ICA, de Souza Filho JRA, Cavalcante LV, Rebouças MES, et al. Incidência do câncer de pênis no brasil. *Brazilian Journal of Science*. 2022; 1:1–8.
- Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*. 2015; 26:27663.
- Marasas W, Jaskeiwicz K, Venter F, van Schalkwyk D. *Fusarium moniliforme* contamination of maize in oesophageal cancer areas in transkei. *South African Medical Journal*. 1988; 74:110–114.
- Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015; 3:1–3.
- De Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The lancet oncology*. 2012; 13:607–615.
- Martin KJ, Rygiewicz PT. Fungal-specific pcr primers developed for analysis of the its region of environmental dna extracts. *BMC microbiology*. 2005; 5:28.
- McCormick A, Loeffler J, Ebel F. *Aspergillus fumigatus*: contours of an opportunistic human pathogen. *Cellular microbiology*. 2010; 12:1535–1543.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*. 2012; 6:610–618.
- McMurdie PJ, Holmes S. phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*. 2013; 8:e61217.
- McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS computational biology*. 2014; 10:e1003531.
- de Oliveira Medeiros M, de Lima LJ, Ferreira Filho FJ, et al. Câncer gástrico: revisão narrativa. *Revista Interdisciplinar Encontro das Ciências-RIEC| ISSN: 2595-0959|*. 2021; 4:.
- Millard SP. *EnvStats: An R Package for Environmental Statistics*. New York: Springer, 2013. ISBN 978-1-4614-8455-4. Disponível em: <https://www.springer.com>.

- Miller JD. Fungi and mycotoxins in grain: implications for stored product research. *Journal of stored products research*. 1995; 31:1–16.
- Mishra K, Isali I, Sindhani M, Prunty M, Bell S, Mahran A, et al. Characterization of Changes in Penile Microbiome Following Pediatric Circumcision. *SSRN Electronic Journal*. 2022; 7:285–292.
- Monteiro LL, Skowronski R, Brimo F, Carvalho Paulo da C N, Vasconcelos RAL, Pacheco CRCV, et al. Erectile function after partial penectomy for penile cancer. *International braz j urol : official journal of the Brazilian Society of Urology*. 2021; 47:515–522.
- Moossavi S, Atakora F, Fehr K, Khafipour E. Biological observations in microbiota analysis are robust to the choice of 16s rRNA gene sequencing processing algorithm: case study on human milk microbiota. *BMC microbiology*. 2020; 20: 1–9.
- Motooka D, Fujimoto K, Tanaka R, Yaguchi T, Gotoh K, Maeda Y, et al. Fungal ITS1 deep-sequencing strategies to reconstruct the composition of a 26-species community and evaluation of the gut mycobiota of healthy Japanese individuals. *Frontiers in microbiology*. 2017; 8:238.
- Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*. 2011; 332:970–974.
- Mukaisho Ki, Nakayama T, Hagiwara T, Hattori T, Sugihara H. Two distinct etiologies of gastric cardia adenocarcinoma: interactions among pH, *Helicobacter pylori*, and bile acids. *Frontiers in microbiology*. 2015; 6:412.
- Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012; 380:2197–2223.
- Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, et al. The gut mycobiome of the Human Microbiome Project Healthy Cohort. *Microbiome*. 2017; 5:153.
- Nearing JT, Douglas GM, Comeau AM, Langille MG. Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ*. 2018; 6:e5364.
- Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*. 2022; 13:1–16.
- Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy of Sciences*. 2021; 118:e2104429118.

- Nguyen LDN, Deschaght P, Merlin S, Loywick A, Audebert C, Van Daele S, et al. Effects of propidium monoazide (pma) treatment on mycobiome and bacteriome analysis of cystic fibrosis airways during exacerbation. *PloS one*. 2016; 11: e0168860.
- Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The unite database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic acids research*. 2018; 47:D259–D264.
- Nocker A, Cheung CY, Camper AK. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of dna from dead cells. *Journal of microbiological methods*. 2006; 67:310–320.
- Noto JM, Peek Jr RM. The gastric microbiome, its interaction with helicobacter pylori, and its potential role in the progression to stomach cancer. *PLoS pathogens*. 2017; 13:e1006573.
- Oever Jt, Netea MG. The bacteriome–mycobiome interaction and antifungal host defense. *European journal of immunology*. 2014; 44:3182–3191.
- Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in nod2 associated with susceptibility to crohn’s disease. *Nature*. 2001; 411:603–606.
- OVERVIEW of QIIME 2 plugin workflows. Disponível em: <https://docs.qiime2.org/2022.2/tutorials/overview/> .
- Palmer JM, Jusino MA, Banik MT, Lindner DL. Non-biological synthetic spike-in controls and the amptk software pipeline improve mycobiome data. *PeerJ*. 2018; 6:e4925.
- Parkin DM. The global health burden of infection-associated cancers in the year 2002. *International journal of cancer*. 2006; 118:3030–3044.
- Parséus A, Sommer N, Sommer F, Caesar R, Molinaro A, Ståhlman M, et al. Microbiota-induced obesity requires farnesoid x receptor. *Gut*. 2017; 66:429–437.
- Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nature methods*. 2013; 10:1200–1202.
- Pei Z, Bini EJ, Yang L, Zhou M, Francois F, Blaser MJ. Bacterial biota in the human distal esophagus. *Proceedings of the National Academy of Sciences*. 2004; 101:4250–4255.
- Pereira-Marques J, Ferreira RM, Pinto-Ribeiro I, Figueiredo C. Helicobacter pylori infection, the gastric microbiome and gastric cancer. *Helicobacter pylori in Human Diseases*. 2019;:195–210.
- Persoon IF, Buijs MJ, Özok AR, Crielaard W, Krom BP, Zaura E, et al. The mycobiome of root canal infections is correlated to the bacteriome. *Clinical oral investigations*. 2017; 21:1871–1881.

- Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cellular microbiology*. 2014; 16:1024–1033.
- Pollock J, Glendinning L, Wisedchanwet T, Watson M. The madness of microbiome: attempting to find consensus “best practice” for 16s microbiome studies. *Applied and environmental microbiology*. 2018; 84:e02627–17.
- Pontillo H, Fernández A, León H. Experiencia de 15 años en el manejo del cáncer de pene. *Revista Venezolana de Oncología*. 2011; 23:26–33.
- Price LB, Liu CM, Johnson KE, Aziz M, Lau MK, Bowers J, et al. The effects of circumcision on the penis microbiome. *PloS one*. 2010; 5:e8422.
- Printz C. Poor oral health tied to oral hpv infection. *Cancer*. 2013; 119:4215–4215.
- Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16s rRNA amplicon sequencing. *PLoS One*. 2020; 15:e0227434.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *nature*. 2010; 464:59.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*. 2012; 41:D590–D596.
- Rahman FI, Islam MR, Bhuiyan MA. Mucormycosis or black fungus infection is a new scare in south asian countries during the covid-19 pandemic: Associated risk factors and preventive measures. *Journal of Medical Virology*. 2021;:.
- Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, et al. Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping, and colonization. *Frontiers in Microbiology*. 2019; 10:1–12.
- Rajagopala SV, Vashee S, Oldfield LM, Suzuki Y, Venter JC, Telenti A, et al. The human microbiome and cancer. *Cancer Prevention Research*. 2017; 10:226–234.
- Rampazzo A, Mott GL, Fontana K, Fagundes RB. Gastric adenocarcinoma trends in the central region of rio grande do sul (southern brazil): what has changed in 25 years? *Arquivos de Gastroenterologia*. 2012; 49:178–183.
- Rao K, Young VB. Fecal microbiota transplantation for the management of *Clostridium difficile* infection. *Infectious Disease Clinics*. 2015; 29:109–122.
- Rappé MS, Giovannoni SJ. The uncultured microbial majority. *Annual Reviews in Microbiology*. 2003; 57:369–394.
- Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Gastroenterology Review/Przegląd Gastroenterologiczny*. 2019; 14:26–38.

- Reis AAdS, Paula LBd, Paula AAPd, Saddi VA, Cruz ADD. [clinico-epidemiological aspects associated with penile cancer]. *Ciencia amp; saude coletiva*. 2010 June; 15 Suppl 1:1105—1111.
- Ries LAG, Reichman ME, Lewis DR, Hankey BF, Edwards BK. Cancer survival and incidence from the surveillance, epidemiology, and end results (seer) program. *The oncologist*. 2003; 8:541–552.
- Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2004; 38:525–552.
- Rintala A, Pietilä S, Munukka E, Eerola E, Pursiheimo JP, Laiho A, et al. Gut microbiota analysis results are highly dependent on the 16s rRNA gene target region, whereas the impact of DNA extraction is minor. *Journal of biomolecular techniques: JBT*. 2017; 28:19.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*. 2010; 11:1–9.
- Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome biology*. 2013; 14:1–20.
- Sagar R, Sharma GP. Measurement of alpha diversity using Simpson index (1/ $\lambda$ ): the jeopardy. *Environmental Skeptics and Critics*. 2012; 1:23.
- Sam QH, Chang MW, Chai LYA. The fungal mycobiome and its interaction with gut bacteria in the host. *International journal of molecular sciences*. 2017; 18:330.
- Sambrook J, Fritsch EF, Maniatis T, et al. *Molecular cloning: a laboratory manual*. [S.l.]: Cold Spring Harbor Laboratory Press, 1989.
- Santra HK, Banerjee D. Natural products as fungicide and their role in crop protection. *Natural Bioactive Products in Sustainable Agriculture*. 2020;:131.
- Sartor RB. Microbial influences in inflammatory bowel diseases. *Gastroenterology*. 2008; 134:577–594.
- Savage DC. Microbial ecology of the gastrointestinal tract. *Annual Reviews in Microbiology*. 1977; 31:107–133.
- Schadt CW, Martin AP, Lipson DA, Schmidt SK. Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science*. 2003; 301:1359–1361.
- Schloss PD. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLOS Computational Biology*. 2010 07; 6:1–16.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*. 2009; 75:7537–7541.

- Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, Queiroz EL, et al. Changes in abundance of oral microbiota associated with oral cancer. *PloS one*. 2014; 9:e98741.
- Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, et al. Finding needles in haystacks: linking scientific names, reference specimens and molecular data for fungi. *Database*. 2014; 2014:.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (its) region as a universal dna barcode marker for fungi. *Proceedings of the National Academy of Sciences*. 2012; 109:6241–6246.
- Schrader C, Schielke A, Ellerbroek L, Johne R. Pcr inhibitors—occurrence, properties and removal. *Journal of applied microbiology*. 2012; 113:1014–1026.
- Schurko AM, Mendoza L, Lévesque CA, Désaulniers NL, De Cock AW, Klassen GR. A molecular phylogeny of *pythium insidiosum*. *Mycological research*. 2003; 107:537–544.
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome biology*. 2011; 12:1–18.
- Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. *Physiological reviews*. 2010; 90:859–904.
- Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? revisiting the ratio of bacterial to host cells in humans. *Cell*. 2016; 164:337–340.
- Seol D, Jhang SY, Kim H, Kim SY, Kwak HS, Kim SH, et al. Accurate and strict identification of probiotic species based on coverage of whole-metagenome shotgun sequencing data. *Frontiers in microbiology*. 2019; 10:1683.
- Sharon I, Quijada NM, Pasolli E, Fabbri M, Vitali F, Agamennone V, et al. The core human microbiome: Does it exist and how can we find it? a critical review of the concept. *Nutrients*. 2022; 14:2872.
- Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science*. 2014; 5:.
- Shay E, Sangwan N, Padmanabhan R, Lundy S, Burkey B, Eng C. Bacteriome and mycobiome and bacteriome-mycobiome interactions in head and neck squamous cell carcinoma. *Oncotarget*. 2020; 11:2375.
- Sheridan G, Masters C, Shallcross J, Mackey B. Detection of mrna by reverse transcription-pcr as an indicator of viability in *escherichia coli* cells. *Applied and Environmental Microbiology*. 1998; 64:1313–1318.
- Simon HY, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. *Cell*. 2019; 178:779–794.

- Smith MI, Yatsunenکو T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut microbiomes of malawian twin pairs discordant for kwashiorkor. *Science*. 2013; 339:548–554.
- Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, et al. Fungal microbiota dysbiosis in ibd. *Gut*. 2016;:gutjnl–2015.
- Song EJ, Lee ES, Nam YD. Progress of analytical tools and techniques for human gut microbiome research. *Journal of Microbiology*. 2018; 56:693–705.
- Stewart EJ. Growing unculturable bacteria. *Journal of bacteriology*. 2012; 194: 4151–4160.
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2021; 71:209–249.
- Sydenham EW, Thiel PG, Marasas WF, Shephard GS, Van Schalkwyk DJ, Koch KR. Natural occurrence of some fusarium mycotoxins in corn from low and high esophageal cancer prevalence areas of the transkei, southern africa. *Journal of Agricultural and Food Chemistry*. 1990; 38:1900–1903.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using dna metabarcoding. *Molecular ecology*. 2012; 21:2045–2050.
- Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. Mycobiome: approaches to analysis of intestinal fungi. *Journal of immunological methods*. 2015; 421:112–121.
- Tantikachornkiat M, Sakakibara S, Neuner M, Durall DM. The use of propidium monoazide in conjunction with qpcr and illumina sequencing to identify and quantify live yeasts and bacteria. *International Journal of Food Microbiology*. 2016; 234:53–59.
- Taylor LH, Latham SM, Mark E. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2001; 356:983–989.
- Tedersoo L, Anslan S, Bahram M, Pölme S, Riit T, Liiv I, et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycKeys*. 2015;:1–43.
- Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, et al. Strong host preference of ectomycorrhizal fungi in a tasmanian wet sclerophyll forest as revealed by dna barcoding and taxon-specific primers. *New Phytologist*. 2008; 180:479–490.
- Thomas AM, Gleber-Netto FO, Fernandes GR, Amorim M, Barbosa LF, Francisco ALN, et al. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *BMC microbiology*. 2014; 14:250.

- Tipton L, Ghedin E, Morris A. The lung mycobiome in the next-generation sequencing era. *Virulence*. 2017; 8:334–341.
- Toju H, Tanabe AS, Yamamoto S, Sato H. High-coverage its primers for the dna-based identification of ascomycetes and basidiomycetes in environmental samples. *PloS one*. 2012; 7:e40863.
- Tonge DP, Pashley CH, Gant TW. Amplicon-based metagenomic analysis of mixed fungal samples using proton release amplicon sequencing. *PloS one*. 2014; 9:e93849.
- Turenne CY, Sanche SE, Hoban DJ, Karlowsky JA, Kabani AM. Rapid identification of fungi by using the its2 genetic region and an automated fluorescent capillary electrophoresis system. *Journal of clinical microbiology*. 1999; 37:1846–1851.
- Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *nature*. 2009; 457:480–484.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007; 449:804–810.
- Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. *Gastroenterology*. 2011; 140:1807–1816.
- Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel its1 fungal primers for characterization of the mycobiome. *mSphere*. 2017; 2:e00488–17.
- Vancov T, Keen B. Amplification of soil fungal community dna using the its86f and its4 primers. *FEMS microbiology letters*. 2009; 296:91–96.
- Ventolero MF, Wang S, Hu H, Li X. Computational analyses of bacterial strains from shotgun reads. *Briefings in bioinformatics*. 2022; 23:.
- Vialle A, Feau N, Allaire M, Didukh M, Martin F, MONCALVO JM, et al. Evaluation of mitochondrial genes as dna barcode for basidiomycota. *Molecular Ecology Resources*. 2009; 9:99–113.
- Wall L, et al. *The Perl programming language*. [S.l.]: Prentice Hall Software Series, 1994 .
- Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. Primerprospector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*. 2011; 27:1159–1161.
- Wang J, Jia H. Metagenome-wide association studies: fine-mining the microbiome. *Nature Reviews Microbiology*. 2016; 14:508–522.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*. 2007; 73:5261–5267.

- Wang Y, Qian PY. Conservative fragments in bacterial 16s rRNA genes and primer design for 16s ribosomal DNA amplicons in metagenomic studies. *PloS one*. 2009; 4:e7401.
- Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. 2017; 5:1–18.
- Westcott SL, Schloss PD. De novo clustering methods outperform reference-based methods for assigning 16s rRNA gene sequences to operational taxonomic units. *PeerJ*. 2015; 3:e1487.
- White TJ, Bruns T, Lee S, Taylor J, et al. Amplification and direct sequencing of fungal ribosomal rRNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*. 1990; 18:315–322.
- Whittaker RH. Evolution and measurement of species diversity. *Taxon*. 1972; 21: 213–251.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *Journal of Open Source Software*. 2019; 4:1686.
- Wickham H, François R, Henry L, Müller K, et al. dplyr: A grammar of data manipulation. R package version 04. 2015; 3:p156.
- Wild CP, Stewart BW, Wild C. *World cancer report 2014*. [S.l.]: World Health Organization Geneva, Switzerland, 2014.
- Willger SD, Grim SL, Dolben EL, Shipunova A, Hampton TH, Morrison HG, et al. Characterization and quantification of the fungal microbiome in serial samples from individuals with cystic fibrosis. *Microbiome*. 2014; 2:40.
- Woese CR. Bacterial evolution. *Microbiological reviews*. 1987; 51:221.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome biology*. 2019; 20:257.
- Wu G, Zhao H, Li C, Rajapakse MP, Wong WC, Xu J, et al. Genus-wide comparative genomics of *Malassezia* delineates its phylogeny, physiology, and niche adaptation on human skin. *PLoS genetics*. 2015; 11:e1005614.
- Yang P, Zhang X, Xu R, Adeel K, Lu X, Chen M, et al. Fungal microbiota dysbiosis and ecological alterations in gastric cancer. *Frontiers in microbiology*. 2022; 13:.
- Ye J, McGinnis S, Madden TL. Blast: improvements for better sequence analysis. *Nucleic acids research*. 2006; 34:W6–W9.
- Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ*. 2017; 356:j831.

Zhang Z, Feng H, Qiu Y, Xu Z, Xie Q, Ding W, et al. Dysbiosis of gastric mucosal fungal microbiota in the gastric cancer microenvironment. *Journal of Immunology Research*. 2022; 2022:.

Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology*. 2004; 150:2565–2573.

## APÊNDICES

## Apêndice A – Publicações

## Article

# Evaluation of Bacteria and Fungi DNA Abundance in Human Tissues

Gabriela E. de Albuquerque<sup>1</sup>, Bruno S. Moda<sup>2</sup>, Marianna S. Serpa<sup>1</sup>, Gabriela P. Branco<sup>1</sup>, Alexandre Defelicibus<sup>2</sup> , Isabella K. T. M. Takenaka<sup>1</sup>, Maria G. de Amorim<sup>1</sup>, Elizabeth C. Miola<sup>1</sup>, Valquiria C. A. Martins<sup>3</sup>, Katia L. Torres<sup>3</sup>, Stephania M. Bezerra<sup>4</sup>, Laura C. L. Claro<sup>4,5,6</sup>, Adriane G. Pelosof<sup>7</sup>, Claudia Z. Sztokfisz<sup>7</sup>, Lais L. S. Abrantes<sup>8</sup>, Felipe J. F. Coimbra<sup>9</sup>, Luiz P. Kowalski<sup>10,11</sup>, Fábio A. Alves<sup>12</sup> , Stênio C. Zequi<sup>13,14</sup>, Klas I. Udekwi<sup>15,16</sup>, Israel T. Silva<sup>2</sup> , Diana N. Nunes<sup>1</sup>, Thais F. Bartelli<sup>1,\*</sup> , and Emmanuel Dias-Neto<sup>1,17,\*</sup> 

- <sup>1</sup> Laboratory of Medical Genomics, A.C. Camargo Cancer Center, Sao Paulo 01508-010, SP, Brazil; gabi.albuquerque0602@gmail.com (G.E.d.A.); mariannaserpa@hotmail.com (M.S.S.); gbranco@gmail.com (G.P.B.); isabellakuniko@gmail.com (I.K.T.M.T.); maria.amorim@thermofisher.com (M.G.d.A.); liza.cris06@gmail.com (E.C.M.); dnoronha@accamargo.org.br (D.N.N.)
- <sup>2</sup> Laboratory of Computational Biology and Bioinformatics, A.C. Camargo Cancer Center, Sao Paulo 01508-010, SP, Brazil; brunomoda@gmail.com (B.S.M.); alexandre.defelicibus@accamargo.org.br (A.D.); itoal@accamargo.org.br (I.T.S.)
- <sup>3</sup> Department of Education and Research, Fundação Centro de Controle de Oncologia do Estado do Amazonas, Manaus 69040-010, AM, Brazil; alvesvalquiria@yahoo.com.br (V.C.A.M.); katialuztorres@hotmail.com (K.L.T.)
- <sup>4</sup> Department of Pathology, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; stephania.bezerra@accamargo.org.br (S.M.B.); lauralopez502@gmail.com (L.C.L.C.)
- <sup>5</sup> Rede D'Or São Luiz S/A, Sao Paulo 04321-130, SP, Brazil
- <sup>6</sup> Santa Casa de Misericórdia de São Paulo, Sao Paulo 01221-010, SP, Brazil
- <sup>7</sup> Endoscopy, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; adriane.pelosof@accamargo.org.br (A.G.P.); claudia.sztokfisz@accamargo.org.br (C.Z.S.)
- <sup>8</sup> International Research Center, A.C. Camargo Cancer Center, Sao Paulo 01508-010, SP, Brazil; lais.senda@accamargo.org.br
- <sup>9</sup> Director Department of Abdominal Surgery, Head Upper GI Oncology Reference Center, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; felipe.coimbra@accamargo.org.br
- <sup>10</sup> Department of Head and Neck Surgery and Otorhinolaryngology, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; lpkowalski@accamargo.org.br
- <sup>11</sup> Department of Head and Neck Surgery, University of Sao Paulo Medical School, Sao Paulo 01246-903, SP, Brazil
- <sup>12</sup> Department of Stomatology, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; falves@accamargo.org.br
- <sup>13</sup> Department of Urology, A.C. Camargo Cancer Center, Sao Paulo 01509-001, SP, Brazil; stenio.zequi@accamargo.org.br
- <sup>14</sup> National Institute for Science and Technology in Oncogenomics and Therapeutic Innovation, A.C. Camargo Cancer Center, São Paulo 01509-001, SP, Brazil
- <sup>15</sup> Department of Aquatic Sciences and Assessment, Swedish University of Agriculture, P.O. Box 7050, 75007 Uppsala, Sweden; klas.udekwi@slu.se
- <sup>16</sup> Department of Medical Sciences, Gastroenterology/Hepatology, Uppsala University Akademiska Sjukhuset, Ingång 40, 75185 Uppsala, Sweden
- <sup>17</sup> Laboratório de Neurociências Alzira Denise Hertzog Silva, Instituto de Psiquiatria, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, Sao Paulo 05403-010, SP, Brazil
- \* Correspondence: thais.bartelli@accamargo.org.br (T.F.B.); emmanuel@accamargo.org.br (E.D.-N.)



**Citation:** de Albuquerque, G.E.; Moda, B.S.; Serpa, M.S.; Branco, G.P.; Defelicibus, A.; Takenaka, I.K.T.M.; de Amorim, M.G.; Miola, E.C.; Martins, V.C.A.; Torres, K.L.; et al. Evaluation of Bacteria and Fungi DNA Abundance in Human Tissues. *Genes* **2022**, *13*, 237. <https://doi.org/10.3390/genes13020237>

Academic Editor: Christopher E. Mason

Received: 6 January 2022

Accepted: 24 January 2022

Published: 27 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Whereas targeted and shotgun sequencing approaches are both powerful in allowing the study of tissue-associated microbiota, the human: microorganism abundance ratios in tissues of interest will ultimately determine the most suitable sequencing approach. In addition, it is possible that the knowledge of the relative abundance of bacteria and fungi during a treatment course or in pathological conditions can be relevant in many medical conditions. Here, we present a qPCR-targeted approach to determine the absolute and relative amounts of bacteria and fungi and demonstrate their relative DNA abundance in nine different human tissue types for a total of 87 samples. In these tissues, fungi genomes are more abundant in stool and skin samples but have much lower levels in

other tissues. Bacteria genomes prevail in stool, skin, oral swabs, saliva, and gastric fluids. These findings were confirmed by shotgun sequencing for stool and gastric fluids. This approach may contribute to a more comprehensive view of the human microbiota in targeted studies for assessing the abundance levels of microorganisms during disease treatment/progression and to indicate the most informative methods for studying microbial composition (shotgun versus targeted sequencing) for various samples types.

**Keywords:** microbiome; mycobiome; metagenomics; 16S; qPCR; shotgun

## 1. Introduction

The human microbiota consists of trillions of microorganisms, including viruses, archaea, bacteria, fungi, protists, and other eukaryotes that may impact human health. The transition to an unbalanced, dysbiotic state needs to be explored, as it may be either the cause or the result of numerous pathological states. The reduction in sequencing costs and the continuous development of high throughput bioinformatics pipelines have contributed immensely not only to study human-associated microorganisms, but also to a deeper view of the microbiota associated with different niches, including plants [1,2], soils [3,4], urban environment [5,6], hospitals [7,8], and even the international space-station [9]. Notably, advances in the study of human-associated microbiota allowed the investigation of numerous health and disease aspects, ranging from the impact of C-section versus vaginal delivery on the microbiota of newborns [10,11], to neurodegenerative diseases [12,13] and the microbiota impact in various cancers [14,15], including treatment response [16].

In general, the most desirable and informative protocols for microbiota studies involve total DNA extraction, followed by DNA fragmentation, library construction, and sequencing, which is denominated shotgun metagenomics. This approach, depending on the efficacy of the DNA/RNA extraction, would allow the capture of all nucleic acids (total nucleic acids—TNA) available and provide a less skewed view of the microbiota [17]. From a comparative standpoint, the human genome is much larger (10–10,000×) than those of viruses, bacteria, fungi, or parasites, thus confounding comparison between studies. Furthermore, depending on the sample source, the human cell fraction will significantly outnumber the microbiota, when the host DNA can easily dominate the resultant TNA pool. This confounding issue may affect the proper inference of the microbiota and impede its effective characterization. Particularly, when the ratio of human to non-human DNA is less favorable, shotgun sequencing would be severely impacted, as most sequencing reads (>99.9%) would originate from the host [18]. In this sense, several strategies have recently emerged as alternatives to provide enrichment of microbial versus host DNA. Most of these methods rely on eliminating host cells prior to DNA extraction or require an intact, high-quality molecular weight DNA, which may depend on the sample source. Furthermore, the potential bias when determining microbial community profiles has not been thoroughly explored and their efficacy is variable and still debatable [19].

Analyses of the tissue-associated microbiota remain limited to cultivation protocols or the study of targeted regions of the genome, such as 16S rRNA for bacteria or ITS for fungi [20–23]. While the former is impacted by the lack of protocols that would allow the unbiased growth of all microorganisms in the sample, as to reflect their sample-specific abundance, the latter imposes limitations such as representation bias and lower phylogenetic resolution [24,25]. These biases have affected both sample-source and experimental approach selection (shotgun or targeted) for microbiota investigations. Consequently, most publications have focused on the study of the gut microbiota, where the bacteria: human ratios are very favorable for microbiome studies. Of particular importance, the tissue microbiota, which has been shown to impact immunomodulation and chemotherapy response in oncology [26], would benefit from characterization via more comprehensive (shotgun) approaches. Body tissues contain important and unique microbiota, whose function and

interest are being unraveled [27,28], and would likewise benefit. In accordance with this, were non-human: human DNA ratios quantified, appropriate samples for targeted or shotgun metagenomic studies could readily be determined, allowing a cost-effective, and more comprehensive or less biased microbiota evaluation. In addition to complementing and strengthening the potential of target and shotgun sequencing for microbiome composition analysis, determining the ratio of non-human DNA in a sample can be informative of the fluctuations in the abundance of the microbiota in several biological conditions, including therapeutic interventions such as chemotherapy and radiotherapy in cancer patients [16,29].

Here, we propose a unifying approach to determine the relative and absolute abundance of fungi and bacteria in different human sample types. This combines the comparison of targeted gene amplification and shotgun sequencing, both assessing the relative proportions of each DNA source for nine tissue types.

## 2. Materials and Methods

### 2.1. Samples and DNA Extraction Methods

A total of 87 samples were included, representing nine different sample types: oral mucosa swabs ( $n = 10$ ), non-stimulated saliva ( $n = 10$ ), gastric fluids collected during endoscopy ( $n = 16$ ), stool ( $n = 5$ ), biopsies of salivary glands ( $n = 6$ ), gastric biopsies ( $n = 10$ ), biopsies of rectal tumors ( $n = 10$ ), tissue fragments from penile tumors ( $n = 10$ ), and skin swabs ( $n = 10$ ). Some tissue specimens were selected as they were simple to obtain and would represent different tissue types (such as oral and skin swabs, and saliva) or when they would allow the characterization of body niches that are often explored in oncology, such as biopsies, tissue samples, and gastric fluids. Subjects enrolled were volunteers from ongoing studies in our group, to evaluate the microbiota of cancer patients and controls, or members of our research group.

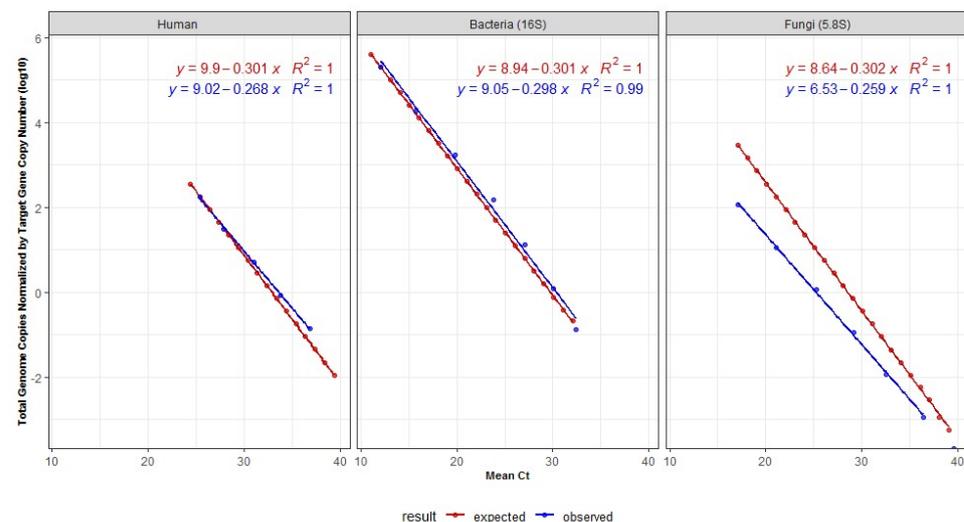
Gastric fluids and gastric biopsies were collected during endoscopy for diagnostic purposes at the A.C. Camargo Cancer Center and stored at  $-20\text{ }^{\circ}\text{C}$ . DNA from oral and skin swabs were extracted with DNeasy PowerSoil Kit (Qiagen, Hilden, NW, Germany); saliva and gastric fluids with phenol:chloroform:isoamyl alcohol 25:24:1 (Merck, Darmstadt, HE, Germany); salivary gland, rectum, penis, and stool DNA was obtained with E.Z.N.A.<sup>®</sup> Bacterial DNA Kit (Omega Bio-Tek, Norcross, GA, USA), and gastric biopsies DNA was extracted with AllPrep DNA/RNA Mini Kit (Qiagen). DNA samples were quantified with Qubit<sup>™</sup> dsDNA HS Assay Kit (ThermoFisher, Waltham, MA, USA) and qPCR amplified.

### 2.2. Primer Pairs and qPCR DNA Amplification

Bacteria- and fungi-derived DNA, were respectively evaluated by DNA amplification of a short fragment of the V1 region of the 16S rRNA (V1\_F: 5'-AGAGTTTGATCMTGGCTCAG-3' and V1\_R: 5'-TTACTCACCCGTICGCCRCT-3') or a portion of the 5.8S rRNA region (5.8S\_F: 5'-CARCAAYGGATCTCTTGG-3' and 5.8S\_R: 5'-TGTGCGTTCAAAGATTTCGAT-3'). To evaluate the human DNA levels, used to normalize human:fungi:bacteria ratios, we used *ACTB* primers (*ACTB\_F*: 5'-CCATCTACGAGGGTATGC-3' and *ACTB\_R*: 5'-GGTGAGGATCTTCATGAGGTA-3') that amplify two targets in the human genome (chrms 5 and chrms 7), generating amplicons of similar lengths (88 and 92 nt). The qPCR mix for *ACTB* and 16S-V1 genes consisted of 0.2  $\mu\text{M}$  of each primer, 5  $\mu\text{L}$  of Fast SYBR Green 2 $\times$  (ThermoFisher) and 3  $\mu\text{L}$  of the eluted DNA in various concentrations, in a final volume of 10  $\mu\text{L}$ . Cycling conditions consisted of an initial denaturation ( $95\text{ }^{\circ}\text{C}$  for 20 s) followed by 40 cycles at  $95\text{ }^{\circ}\text{C}$  for 3 s and  $60\text{ }^{\circ}\text{C}$  for 30 s. For 5.8S gene fragment, the qPCR mix included 0.2  $\mu\text{M}$  of each primer, 5  $\mu\text{L}$  of the GoTaq Master Mix 2 $\times$  (Promega, Madison, WIS, USA) and 4.2  $\mu\text{L}$  of eluted DNA in a final volume of 10  $\mu\text{L}$ . Cycling conditions for qPCR were initial denaturation at  $95\text{ }^{\circ}\text{C}$  for 3 min, followed by 40 cycles at  $95\text{ }^{\circ}\text{C}$  for 15 s and  $58\text{ }^{\circ}\text{C}$  for 30 s. All experiments were conducted in duplicates and qPCR was performed in a 7500 Real-Time PCR System (ThermoFisher).

### 2.3. Standard Curves

For absolute quantitative analysis by qPCR, we utilized commercially available human genomic DNA (Promega—G1471) to generate a standard curve derived from human amplicons. The ZymoBIOMICS™ Microbial Community DNA Standard (Zymo Research—D6305/D6306,) composed of eight bacteria and two fungi species (*Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Cryptococcus neoformans*), was used to prepare standard curves for bacteria and fungi, starting with a DNA mass of 5 ng. The number of bacteria, fungi, or human genomes were calculated per ng of DNA according to the formula: number of copies = mass of DNA used (ng)  $\times$   $6.022 \times 10^{23}$  / average genome size  $\times$  ( $1 \times 10^9$ )  $\times$  650 (average weight of a base pair)—divided by the median number of copies of the target genes. Mean Cts corresponding to each dilution point were used for plotting the standard curves. Mean genome size and target copy number were considered for the bacteria and fungi present in the mock community. A median of six copies of 16S rRNA for bacteria, 102.5 copies of rRNA for the haploid fungi (<https://rrndb.umms.med.umich.edu>, accessed on 13 April 2020), and four for the human calculations (two copies of the two loci amplified by the ACTB primers in a diploid human genome). The genome sizes considered were 6.4 Gb for the human diploid genome, 3.58 Mb for bacteria, and 15.6 Mb for fungi haploid genomes (Figure 1, Table 1). These values represent the median of the genome sizes for the components of the mock according to the National Center for Biotechnology (NCBI) database at the time of analysis (<https://www.ncbi.nlm.nih.gov/genome/>, accessed on 13 April 2020).



**Figure 1.** Standard DNA dilution curves for human, bacteria (V1-16S) and fungi (5.8S) qPCR, showing expected (red) and observed (blue) Ct-values considering the genome size and target copy number of the bacteria and fungi present in the mock community. Calculations do not consider the variable ploidy observed in many fungi species. Values used for plotting are shown in Table 1. In the equation, y denotes the genome copy number while x denotes the mean Ct value.

For calculating the absolute genome copy number for the specimens used here, additional standard curves were created considering the genome sizes and target gene copy number of an increased number of bacteria and fungi, not limited to the ones present at the mock community. For the construction of these standard curves and calculations, the average genome sizes were considered as 6.4 Gb for humans (human diploid genome), 12.3 Mb (fungi haploid genome), and 4.15 Mb (bacteria). The values for bacteria and fungi were respectively calculated based on the median length of 25,923 bacteria and 177 fungi genomes available at <https://www.ncbi.nlm.nih.gov/genome/>, accessed on 2 December 2021. Only complete genome sequences in the database were considered. The number of copies of the target genes was five for the bacterial 16S rRNA (based on a total of 20,255 full bacteria

genomes from the rrnDB database (as of Oct. 2021-<https://rrnDB.umms.med.umich.edu>, accessed on 2 December 2021); and 82 copies of the rRNA for fungi (based on a total of 91 fungal taxa, as described by Lofgren et al., 2019). Calculations of the genome copies (Log10) were obtained as follow: (i) human genome copies =  $9.02 - (0.268 \times \text{mean Ct value})$ ; (ii) bacteria genome copies =  $9.06 - (0.298 \times \text{mean Ct value})$ , and (iii) fungi genome copies  $6.73 - (0.259 \times \text{mean Ct value})$ . Ploidy was not considered for the overall calculations in fungi and may have impacted calculations in real-life samples. Absolute genome copies were calculated for bacteria and fungi according to the Cts obtained for each sample after comparing them to the standard curves. Standard curves correlating the mean Cts with the total genome copies/ng of DNA were constructed. Each standard curve is given as: (i) expected Cts—considered according to the lowest Ct-value, under the assumption that a difference of one Ct would correspond to a  $2 \times$  difference in the dilution of the template DNA mass and (ii) observed Cts—real values observed after qPCR amplification for the same theoretical dilutions.

**Table 1.** Calculated correlations between DNA mass, mean Cts and number of copies of the genomes and the target genes for the mock DNA and human DNA.

Organism Group	DNA Mass in qPCR (ng)	Mean Ct	Genome Length (Mb)	Number of Genome Copies	Number of Genomes Normalized by Copies of the Target	Log10 of the Number Genomes
Mock-bacteria	4.800000	12.1	3.6	1,242,233.8	207,038.9	5.3
	0.432432	15.8	3.6	111,913.0	18,652.16	4.3
	0.038958	19.8	3.6	10,082.2	1680.37	3.2
	0.003510	23.8	3.6	908.3	151.39	2.2
	0.000316	27.1	3.6	81.8	13.64	1.1
	0.000028	30.0	3.6	7.4	1.23	0.1
	0.000003	32.4	3.6	0.7	0.11	−1.0
Human	5.000000	25.4	64,000	723.9	180.96	2.3
	0.833333	27.8	64,000	120.6	30.16	1.5
	0.138889	31.1	64,000	20.1	5.03	0.7
	0.023148	33.8	64,000	3.4	0.84	−0.1
	0.003858	36.9	64,000	0.6	0.14	−0.9
Mock-fungi	0.200000	17.1	15.6	11,875.1	115.9	2.1
	0.020000	21.1	15.6	1187.5	11.6	1.1
	0.002000	25.3	15.6	118.8	1.2	0.06
	0.000200	29.2	15.6	11.9	0.1	−0.9
	0.000020	32.6	15.6	1.2	0.01	−1.9
	0.000002	36.4	15.6	0.1	0.001	−2.9

Notes: 1—The total mass of the mock contained 96% of bacteria DNA (12% of each of the eight species) and 4% from fungi DNA (2% of each one of the two species). 2—Calculations have considered the average number of copies of the amplification targets in the respective genomes used in this mock and the diploid human genome (four for human DNA, six for bacteria, and 102.5 for fungi).

#### 2.4. Shotgun Sequencing

As another metric to evaluate bacteria: human and fungi: human ratios, and to determine the accuracy of our calculations, we compared qPCR ratios to shotgun metagenomic sequencing data of 10 samples of gastric fluids and two stool samples. For this, we used 100 ng of DNA template for library preparation for each sample (Nextera DNA Flex kit, Illumina, San Diego, CA, USA), followed by sequencing on the NextSeq 500 platform (Illumina). The results were analyzed by Kraken2 [30], and the number of reads for bacteria, fungi and human were used to calculate the relative frequency of each group, and to compare with qPCR data.

#### 2.5. Statistical Analysis

Possible significant differences between the genome copy number of the different samples were assessed by the Mann–Whitney test. For the correlation analyses, we used Spearman’s correlation. For all statistical tests, a level of 5% significance was considered ( $p < 0.05$ ). Statistical analysis and graphs were generated with the appropriate packages using R Studio.

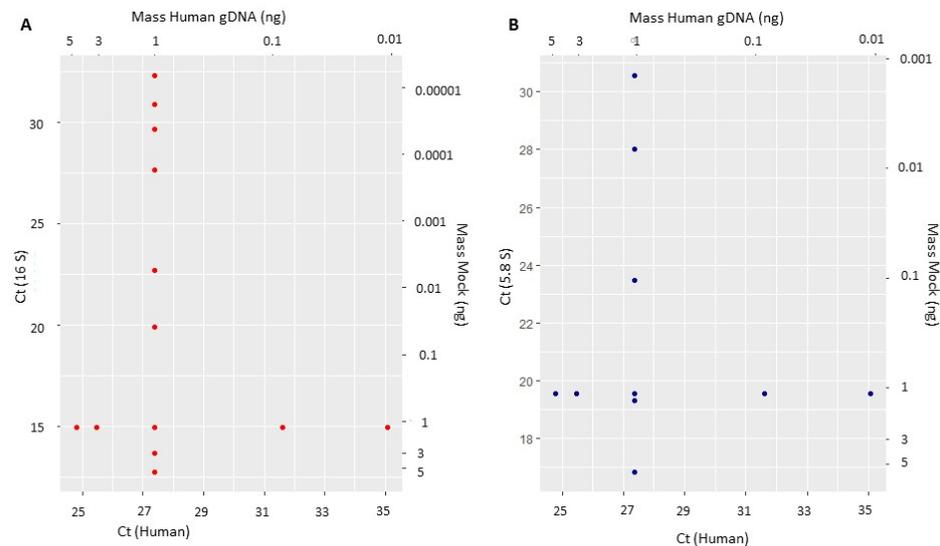
### 3. Results

#### 3.1. Standard Curves Correlating Genome Copies and Cts

Standard curves were plotted, allowing to correlate the approximate absolute number of genome copies for bacteria and fungi according to the respective Cts achieved in qPCR amplification (Table 1). As shown in Figure 1, all standard curves showed very high correlations ( $R^2$  equal to 0.99 or 1), despite some discrepancies observed for fungi.

#### 3.2. Impacts of Varying Bacteria:Human and Fungi:Human DNA-Ratios over Amplification/Quantification Capabilities

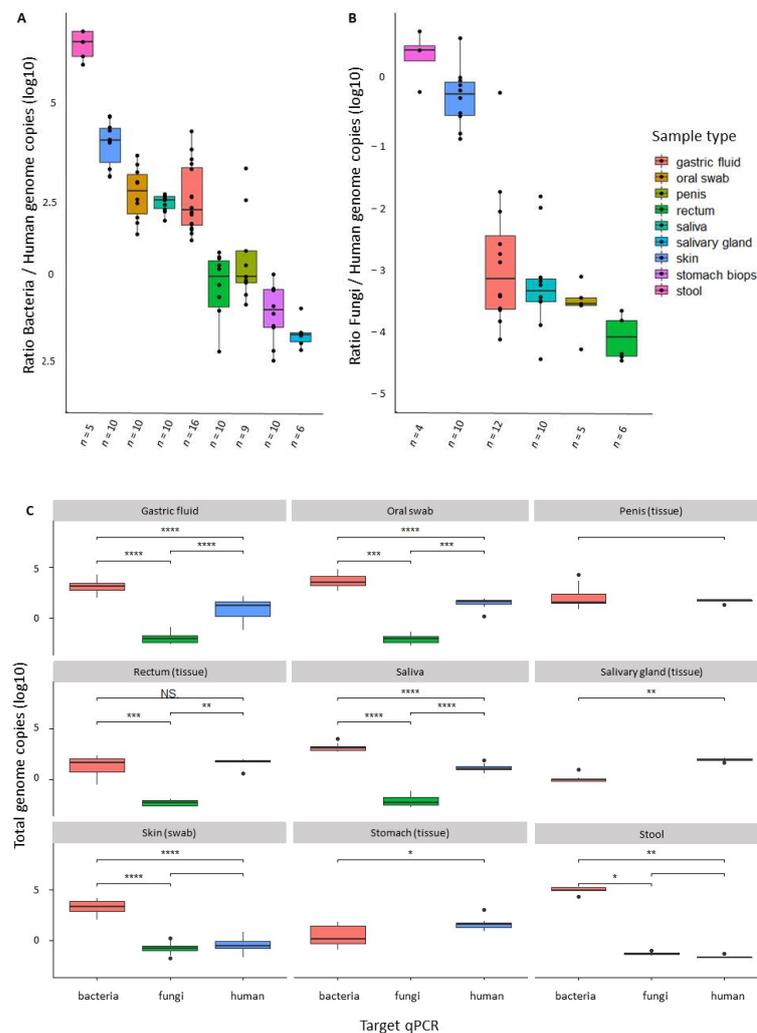
In order to simulate real-life analysis of samples that present significant variations in the ratios of target microorganisms and host-derived DNA, we evaluated the Ct fluctuations in face of different human: bacteria and fungi: human DNA ratios and their possible impact on the quantification of these microorganisms. When using a fixed amount of human DNA (1 ng) and up to 500 $\times$  variation (5 ng to 0.01 ng) of the mock DNAs from bacteria or fungi, the Ct values for the human ACTB gene remained within the expected range of variation, suggesting no inhibition or off-target amplification. The same was observed when the amount of mock DNA was fixed at 1 ng, and the human DNA varied 500 $\times$  (0.01 ng to 5 ng), with no significant impact on either the bacteria 16S-, or the fungi 5.8S-detection (Figure 2).



**Figure 2.** Effects of varying DNA mass over amplification capabilities of bacteria (A) and fungi (B), as compared to human DNA amounts. Horizontal axes indicate the Ct value for human DNA and vertical axes indicate Cts for bacteria (16S) or fungi (5.8S). Each dot represents the mean of a triplicate experiment.

#### 3.3. Genome Copies in Different Sample-Types

After running the standard curves for different amounts of DNA and showing no significant variations of background human DNA to affect the expected Cts, we were able to determine the approximate number of bacteria or fungi genome copies for nine different types of human-derived samples. Figure 3A shows, as expected, that stool samples have the highest bacteria: human ratios—making this a very good source for metagenome shotgun sequencing, as published by many different groups. Our data also show that skin and oral swabs, saliva, and gastric fluids also contain more bacteria genomes than human genomes and, thus, should provide informative microbiota data using this same sequencing approach, despite the much larger length of the human genome. By contrast, some tissues contain similarly high levels of human and bacterial genomes—as with rectal and penile biopsies—suggesting that metagenomics shotgun sequencing could still be informative, but human DNA will largely dominate due to its length.



**Figure 3.** Genome-ratios for bacteria:human (A) and fungi:human (B) and total genome copies for bacteria (red), fungi (green) and human (blue) (C) in different human sample types. Mann–Whitney test \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$  significant.

After comparing the approximate number of human and bacteria genomes in all samples, we observed that biopsies from salivary gland and stomach contained much less bacteria genomes as compared to human genomes (Figure 3C). Rectal and penile biopsies appear to have similar amounts of human and bacterial genome copies, whereas the other sample types have significantly more bacteria, stool being the sample with most bacterial DNA as expected. From our limited set of samples, only stool and skin samples seem to have more significant fungi: human genome-ratios, and good variation was seen for gastric fluids and saliva. Oral swabs, saliva, and gastric fluids have similar proportions of genomes from each type of organism, which makes sense as the gastric microbiota is heavily influenced by the large amounts of microorganisms present in the saliva (Figure 3C).

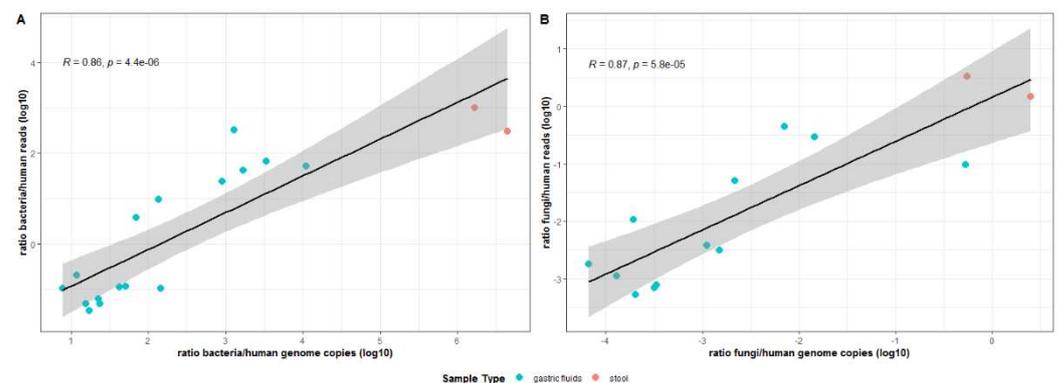
The quantitative analysis of fungi by qPCR showed stool and skin to be the samples with more fungi genome copies, with a reduction in saliva, oral swabs, rectal biopsies, all presenting similar amounts of fungi genomes. Interestingly, very small amounts of human genome were found in the skin and stool, as compared with the other tissues and samples evaluated here as DNA sources (Figure 3B,C).

### 3.4. Shotgun Analysis

As shotgun sequencing is less biased, it was used to compare the resulting data with our qPCR data, to analyze the relative abundance of bacteria, fungi, and human-derived

DNA in 16 samples of gastric fluids and two stool samples. Our data has shown human-derived reads varying from 0.5 and 67% in the gastric fluids, and to be  $\leq 0.1\%$  in the stool. As expected, we observed a smaller contribution of fungi-DNA, as compared to bacteria for both sample types. This low frequency of fungi was true for all samples, ranging from 0.03 to 0.6% of the total reads in gastric fluids, and 0.1% in the two stool samples.

Our shotgun sequencing results were also compared in terms of bacteria: human and fungi: human genome ratios (obtained from qPCR analysis of the same samples). As can be seen in Figure 4, very good correlations were obtained between qPCR data and the ratios of bacteria: human reads (Figure 4A) and fungi: human reads (Figure 4B), from shotgun sequencing analysis. This reinforces the validity of the herein presented qPCR approach, as well as the supporting calculations that allow the amounts of the genomes of these organisms to be determined.



**Figure 4.** Correlation between the ratios of bacteria:human reads observed by shotgun sequencing data and the relative bacteria:human genome copies calculated by qPCR (A). The same is shown for fungi (B). Each circle represents a distinct sample. For statistical analysis, Spearman’s correlation was used, and the closer it is to 1, the greater the correlation between the variables. The shaded area represents the confidence interval. Correlations were considered as statistically significant when  $p < 0.05$ .

#### 4. Discussion

Most human microbiota studies are focused on the composition and diversity of the microbiota and do not consider that the amounts of bacteria or the human: bacteria ratios could be informative. Additionally, very few authors have focused on the mycobiome composition, not to mention the different fungi colonization enrichment in sample types and disease settings, and possible impacts of this biomass in regulating or having any role in biological processes. Moreover, the quantitative variation of the microbiota may not only result from disease processes but from the use of medicines, pre- and pro-biotics, and drugs such as alcohol and tobacco, as well as levels of mucosa immunity and other situations. Here we observed that only stool and skin samples had more fungi genomes, but still much lower than bacteria, as compared to the other sample types. This observation should, however, be viewed with caution, as very few specimens were evaluated for some tissue types. On the other hand, fungi such as *Candida albicans* are known as important members of the human gastrointestinal tract and may also be opportunistic pathogens [31]. In this sense, it seems feasible that in some pathologies and for different tissue types, variations in the number of microorganisms may have a physiological role and may suggest the impact of certain drugs or therapies on the microbiota, their role in drug metabolism, or even serve as a proxy of immune surveillance status where microorganisms appear to influence immune regulation [32–34].

Other applications of our approach are more technical and include the use of this tool to effectively indicate the best sequencing approach for determining microbiota composition, and how many reads one can expect from a shotgun metagenomics study. In this sense, our results demonstrated that gastric tissues had a significantly higher number of human genomes per bacterial or fungi genomes as compared to the gastric fluids from the same

patients, indicating that a shotgun approach for gastric microbiome studies should be more efficient if based on gastric fluids. From our data (Figure 3) it is also clear that samples from skin, oral swabs, and gastric fluids have amounts of non-human DNA that would allow a direct shotgun sequencing approach, with no need to first amplify the target genes. Indeed, this was confirmed from our shotgun sequencing of gastric fluids. As seen here and for other studies, the quantification of fungi may be less precise due to ploidy variation, which may occur according to species composition and developmental stage [35,36], as well as the higher complexity of the fungal cell wall, which is usually more resistant to lysis [37], and would also affect fungi quantification and analysis when shotgun approaches are used. Overall, it seems that our approach is reliable, as our amplifications showed no detectable fluctuations when facing 500× excess of human-host DNA or 500× excess of mock microbiota template DNA, and it is also capable to provide a precise microbiota quantification, as seen from the expected/observed number of reads for bacteria or fungi as well as the amount of human-derived DNA in shotgun metagenomics. A good demonstration of this is the evaluation of the microbiota of gastric fluids by shotgun metagenomics, as predicted by our quantitative analysis, which shows that >50% of samples have more bacteria-derived DNA than host-DNA. Our data also indicates samples for which the shotgun approach would be inadequate, when lesser amounts of bacteria- or fungi-derived DNA are present, precluding the microbiota evaluation with shotgun sequencing.

The application of this method for medical/biological situations of interest is envisioned. An example may include the monitoring of bacteria biomass after antibiotic treatment and the resolution of the septic shock [38], the observation of variations in the abundance of protective or opportunistic microorganisms that may impact surgery, radiotherapy, and chemotherapy, including side-effects such as infections of surgical sites, mucositis, and even treatment efficacy [39–41]. In this regard, the use of specific microorganisms to protect the intestinal epithelium has been suggested to manage collateral damage related to radiotherapy [42,43]. Our approach could thus be applied to investigate a number of treatment-related scenarios including the impact of ‘protective’ bacteria inoculation in pre-radiotherapy versus post-radiotherapy stages, treatment efficacy, and to assess later development of mucositis in different grades. We envisage the widespread application of this simple tool to rapidly assess bacterial levels in diverse pathophysiological and other clinically relevant scenarios of metagenomics interest, not limited to human samples.

## 5. Conclusions

In this study, we proposed a method for estimating the relative and absolute abundance of fungi and bacteria in human samples. We observed that the ratio of human per bacteria or fungi DNA is variable depending on the sample collection method (tissue, swab, body fluids) and source (tissues such as skin, rectum, stomach, and penis). Additionally, although bacteria outnumber fungi in most samples, we show that these microorganisms can be more abundant in a few sample types (stool and skin), indicating their importance in the microbiome of different body sites. This approach provides a valuable resource for the improvement of metagenomic studies, enabling a more comprehensive view of the human microbiota in targeted studies and a quantitative assessment of the microbiota in several clinical scenarios.

**Author Contributions:** Experimental design: G.E.d.A., B.S.M., M.S.S., D.N.N., T.F.B., E.D.-N. Results discussion: G.E.d.A., B.S.M., M.S.S., D.N.N., T.F.B., E.D.-N. Patient recruitment and sample collection: L.L.S.A., I.K.T.M.T., V.C.A.M., K.L.T., E.C.M. Classification and microdissection of tissue samples: L.C.L.C., S.M.B. Sample preparation and experimental procedures: G.E.d.A., B.S.M., M.S.S., G.P.B., M.G.d.A., T.F.B. Bulk data processing: A.D., I.T.S. Data analysis: G.E.d.A., B.S.M., M.S.S., T.F.B. Clinical discussion: L.P.K., F.A.A., S.C.Z., C.Z.S., A.G.P., F.J.F.C., E.D.-N. Manuscript writing: G.E.d.A., T.F.B., E.D.-N. Manuscript editing: G.E.d.A., K.I.U., D.N.N., T.F.B., E.D.-N. All authors have read and agreed to the published version of the manuscript.

**Funding:** E.D.-N. is a researcher from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq—Brasil). This project was funded in part by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 14/26897-0; 18/14267-2; 18/02972-3) and Programa Nacional de Atenção ao Paciente Oncológico (PRONON—SIPAR 2500.055-167/2015-23), Departamento de Ciência e Tecnologia, Ministério da Saúde (DECIT, MS), Brazil.

**Institutional Review Board Statement:** The A.C.Camargo Cancer Center ethics committee previously approved the collection and analysis of the samples used in this study (protocols 2134/15, 2188/16, 2689/19, 2484/18, 2446/17, 2514/18).

**Informed Consent Statement:** All subjects were invited to join the study as volunteers and participants have signed an informed consent form also approved by the institutional ethics committee.

**Data Availability Statement:** The data that support the findings of this study are available on request from the corresponding authors (T.F.B.; E.D.-N.).

**Acknowledgments:** We thank the A.C.Camargo Cancer Center Biobank for sample storage and processing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Khan, Z.; Doty, S.L. Characterization of bacterial endophytes of sweet potato plants. *Plant Soil* **2009**, *322*, 197–207. [CrossRef]
2. Hardoim, P.R.; van Overbeek, L.S.; van Elsas, J.D. Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* **2008**, *6*, 463–471. [CrossRef] [PubMed]
3. Baldrian, P.; Kolarik, M.; Štursová, M.; Kopecký, J.; Valášková, V.; Větrovský, T.; Žifčáková, L.; Šnajdr, J.; Rídl, J.; Vlček, Č.; et al. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* **2012**, *6*, 248–258. [CrossRef] [PubMed]
4. Auffret, M.D.; Karhu, K.; Khachane, A.; Dungait, J.; Fraser, F.; Hopkins, D.; Wookey, P.; Singh, B.; Freitag, T.E.; Hartley, I.P.; et al. The Role of Microbial Community Composition in Controlling Soil Respiration Responses to Temperature. *PLoS ONE* **2016**, *11*, e0165448. [CrossRef] [PubMed]
5. Mason, C.; Afshinnikoo, E.; Ahsannudin, S.; Ghedin, E.; Read, T.; Fraser, C.; Dudley, J.; Hernandez, M.; Bowler, C.; Stolovitzky, G.; et al. The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. *Microbiome* **2016**, *4*, 24.
6. Danko, D.; Bezdan, D.; Afshin, E.E.; Ahsanuddin, S.; Bhattacharya, C.; Butler, D.J.; Chng, K.R.; Donnellan, D.; Hecht, J.; Jackson, K.; et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell* **2021**, *184*, 3376–3393.e17. [CrossRef]
7. Lax, S.; Sangwan, N.; Smith, D.; Larsen, P.; Handley, K.M.; Richardson, M.; Guyton, K.; Krezalek, M.; Shogan, B.D.; Defazio, J.; et al. Bacterial colonization and succession in a newly opened hospital. *Sci. Transl. Med.* **2017**, *9*, eaah6500. [CrossRef]
8. Chng, K.R.; Li, C.; Bertrand, D.; Ng, A.H.Q.; Kwah, J.S.; Low, H.M.; Tong, C.; Natrajan, M.; Zhang, M.H.; Xu, L.; et al. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. *Nat. Med.* **2020**, *26*, 941–951. [CrossRef]
9. Checinska Sielaff, A.; Urbaniak, C.; Mohan, G.B.M.; Stepanov, V.G.; Tran, Q.; Wood, J.M.; Minich, J.; McDonald, D.; Mayer, T.; Knight, R.; et al. Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome* **2019**, *7*, 50. [CrossRef]
10. Biasucci, G.; Rubini, M.; Riboni, S.; Morelli, L.; Bessi, E.; Retetangos, C. Mode of delivery affects the bacterial community in the newborn gut. *Early Hum. Dev.* **2010**, *86* (Suppl. 1), 13–15. [CrossRef]
11. Mueller, N.T.; Whyatt, R.; Hoepner, L.; Oberfield, S.; Dominguez-Bello, M.G.; Widen, E.M.; Hassoun, A.; Perera, F.; Rundle, A. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *Int. J. Obes.* **2015**, *39*, 665–670. [CrossRef] [PubMed]
12. Poole, S.; Singhrao, S.K.; Kesavalu, L.; Curtis, M.A.; Crean, S. Determining the Presence of Periodontopathic Virulence Factors in Short-Term Postmortem Alzheimer’s Disease Brain Tissue. *J. Alzheimer’s Dis.* **2013**, *36*, 665–677. [CrossRef] [PubMed]
13. Scheperjans, F.; Aho, V.; Pereira, P.A.B.; Koskinen, K.; Paulin, L.; Pekkonen, E.; Haapaniemi, E.; Kaakkola, S.; Eerola-Rautio, J.; Pohja, M.; et al. Gut microbiota are related to Parkinson’s disease and clinical phenotype. *Mov. Disord.* **2015**, *30*, 350–358. [CrossRef] [PubMed]
14. Thomas, A.M.; Jesus, E.C.; Lopes, A.; Aguiar, S.J.; Begnami, M.D.; Rocha, R.M.; Carpinetti, P.A.; Camargo, A.A.; Hoffmann, C.; Freitas, H.C.; et al. Tissue-Associated Bacterial Alterations in Rectal Carcinoma Patients Revealed by 16S rRNA Community Profiling. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 179. [CrossRef]
15. Metagenomic Analysis of Colorectal Cancer Datasets Identifies Cross-Cohort Microbial Diagnostic Signatures and a Link with Choline Degradation | Nature Medicine. Available online: <https://www.nature.com/articles/s41591-019-0405-7> (accessed on 20 January 2022).

16. Lehouritis, P.; Cummins, J.; Stanton, M.; Murphy, C.T.; McCarthy, F.; Reid, G.; Urbaniak, C.; Byrne, W.L.; Tangney, M. Local bacteria affect the efficacy of chemotherapeutic drugs. *Sci. Rep.* **2015**, *5*, 14554. [[CrossRef](#)]
17. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **2017**, *35*, 833–844. [[CrossRef](#)]
18. Marotz, C.A.; Sanders, J.G.; Zuniga, C.; Zaramela, L.S.; Knight, R.; Zengler, K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **2018**, *6*, 1–9. [[CrossRef](#)]
19. Heravi, F.S.; Zakrzewski, M.; Vickery, K.; Hu, H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* **2020**, *170*, 105856. [[CrossRef](#)]
20. Gophna, U.; Sommerfeld, K.; Gophna, S.; Doolittle, W.F.; van Zanten, S.J.O.V. Differences between Tissue-Associated Intestinal Microfloras of Patients with Crohn’s Disease and Ulcerative Colitis. *J. Clin. Microbiol.* **2006**, *44*, 4136–4141. [[CrossRef](#)]
21. Thomas, A.M.; Gleber-Netto, F.O.; Fernandes, G.R.; Amorim, M.; Barbosa, L.F.; Francisco, A.L.N.; de Andrade, A.G.; Setubal, J.C.; Kowalski, L.P.; Nunes, D.N.; et al. Alcohol and tobacco consumption affects bacterial richness in oral cavity mucosa biofilms. *BMC Microbiol.* **2014**, *14*, 250. [[CrossRef](#)]
22. Esposito, S.; De Simone, G.; Gioia, R.; Noviello, S.; Pagliara, D.; Campitiello, N.; Rubino, C.; Pardo, D.L.; Boccia, G.; DE Caro, F.; et al. Deep tissue biopsy vs. superficial swab culture, including microbial loading determination, in the microbiological assessment of Skin and Soft Tissue Infections (SSTIs). *J. Chemother.* **2017**, *29*, 154–158. [[CrossRef](#)] [[PubMed](#)]
23. Aykut, B.; Pushalkar, S.; Chen, R.; Li, Q.; Abengozar, R.; Kim, J.I.; Shadaloey, S.A.; Wu, D.; Preiss, P.; Verma, N.; et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* **2019**, *574*, 264–267. [[CrossRef](#)] [[PubMed](#)]
24. Donachie, S.P.; Foster, J.S.; Brown, M.V. Culture clash: Challenging the dogma of microbial diversity. *ISME J.* **2007**, *1*, 97–99. [[CrossRef](#)] [[PubMed](#)]
25. Campanaro, S.; Treu, L.; Kougiaris, P.G.; Zhu, X.; Angelidaki, I. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Sci. Rep.* **2018**, *8*, 1926. [[CrossRef](#)] [[PubMed](#)]
26. Anfossi, S.; Calin, G.A. Gut microbiota: A new player in regulating immune- and chemo-therapy efficacy. *Cancer Drug Resist.* **2020**, *3*, 356–370. [[CrossRef](#)]
27. Chng, K.R.; Chan, S.H.; Ng, A.H.Q.; Li, C.; Jusakul, A.; Bertrand, D.; Wilm, A.; Choo, S.P.; Tan, D.M.Y.; Lim, K.H.; et al. Tissue Microbiome Profiling Identifies an Enrichment of Specific Enteric Bacteria in *Opisthorchis viverrini* Associated Cholangiocarcinoma. *eBioMedicine* **2016**, *8*, 195–202. [[CrossRef](#)]
28. Poore, G.D.; Kopylova, E.; Zhu, Q.; Carpenter, C.; Fraraccio, S.; Wandro, S.; Kosciulek, T.; Janssen, S.; Metcalf, J.; Song, S.J.; et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **2020**, *579*, 567–574. [[CrossRef](#)]
29. Kim, Y.S.; Kim, J.; Park, S.-J. High-throughput 16S rRNA gene sequencing reveals alterations of mouse intestinal microbiota after radiotherapy. *Anaerobe* **2015**, *33*, 1–7. [[CrossRef](#)]
30. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [[CrossRef](#)]
31. Pérez, J.C. Fungi of the human gut microbiota: Roles and significance. *Int. J. Med. Microbiol.* **2021**, *311*, 151490. [[CrossRef](#)]
32. Geller, L.T.; Barzily-Rokni, M.; Danino, T.; Jonas, O.H.; Shental, N.; Nejman, D.; Gavert, N.; Zwang, Y.; Cooper, Z.A.; Shee, K.; et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **2017**, *357*, 1156–1160. [[CrossRef](#)] [[PubMed](#)]
33. Yuan, L.; Zhang, S.; Li, H.; Yang, F.; Mushtaq, N.; Ullah, S.; Shi, Y.; An, C.; Xu, J. The influence of gut microbiota dysbiosis to the efficacy of 5-Fluorouracil treatment on colorectal cancer. *Biomed. Pharmacother.* **2018**, *108*, 184–193. [[CrossRef](#)] [[PubMed](#)]
34. Guthrie, L.; Gupta, S.; Daily, J.; Kelly, L. Human microbiome signatures of differential colorectal cancer drug metabolism. *NPJ Biofilms Microbiomes* **2017**, *3*, 27. [[CrossRef](#)] [[PubMed](#)]
35. Todd, R.T.; Forche, A.; Selmecki, A. Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. *Microbiol. Spectr.* **2017**, *5*, 599–618. [[CrossRef](#)]
36. Wertheimer, N.B.; Stone, N.; Berman, J. Ploidy dynamics and evolvability in fungi. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20150461. [[CrossRef](#)]
37. Fredricks, D.N.; Smith, C.; Meier, A. Comparison of Six DNA Extraction Methods for Recovery of Fungal DNA as Assessed by Quantitative PCR. *J. Clin. Microbiol.* **2005**, *43*, 5122–5128. [[CrossRef](#)]
38. Jimenez, M.F.; Marshall, J.C. Source control in the management of sepsis. *Intensiv. Care Med.* **2001**, *27*, S49–S62. [[CrossRef](#)]
39. Wenzel, R.P. Surgical site infections and the microbiome: An updated perspective. *Infect. Control Hosp. Epidemiol.* **2019**, *40*, 590–596. [[CrossRef](#)]
40. Subramaniam, N.; Muthukrishnan, A. Oral mucositis and microbial colonization in oral cancer patients undergoing radiotherapy and chemotherapy: A prospective analysis in a tertiary care dental hospital. *J. Investig. Clin. Dent.* **2019**, *10*, e12454. [[CrossRef](#)]
41. Irfan, M.; Delgado, R.Z.R.; Frias-Lopez, J. The Oral Microbiome and Cancer. *Front. Immunol.* **2020**, *11*, 591088. [[CrossRef](#)]
42. He, J.-Y.; Wang, W.-Z.; Qi, H.-Z.; Ma, Y.; He, S.-Y. Use of recombinant *Lactobacillus sakei* for the prevention and treatment of radiation-induced enteritis. *Med. Hypotheses* **2018**, *119*, 37–40. [[CrossRef](#)]
43. Riehl, T.E.; Alvarado, D.; Ee, X.; Zuckerman, A.; Foster, L.; Kapoor, V.; Thotala, D.; Ciorba, M.A.; Stenson, W.F. *Lactobacillus rhamnosus* GG protects the intestinal epithelium from radiation injury through release of lipoteichoic acid, macrophage activation and the migration of mesenchymal stem cells. *Gut* **2018**, *68*, 1003–1013. [[CrossRef](#)]

## Apêndice B – Oligonucleotídeos levantados na literatura

**Tabela 15** – Oligonucleotídeos levantados na literatura

Primer	Sequência (3' - 5')	Direção	Referência
3271-ITS2r	GATATGCTTAAGTTCAGCGGGT	reverse	(Delhaes et al. 2012)
3271ITS2f	CARCAAYGGATCTCTTGG	forward	(Delhaes et al. 2012)
5.8SRf	TCGATGAAGAACGCAGCG	forward	Vilgalys Lab*
5.8Sf_khot	GTGAATCATCGARTCTTTGAAC	forward	(Khot et al. 2009)
5.8Sf	CGCTGCGTTCTTCATCG	forward	Vilgalys Lab*
5.8Sbr	TCGATGAAGAACGCAGCG	reverse	Vilgalys Lab*
5.8Sar_opt	CAAGAGATCCGTTGYTGAAAGTT	reverse	esse trabalho
5.8Sar	CAAGAGATCCGTTGTTGAAAGTT	reverse	(Epp et al. 2012)
58A2r	CTGCGTTCTTCATCGAT	reverse	(Martin e Ryגיעiewicz 2005)
58A2f	ATCGATGAAGAACGCAG	forward	(Martin e Ryגיעiewicz 2005)
58A1f	GCATCGATGAAGAACGC	forward	(Martin e Ryגיעiewicz 2005)
28S-4r	CCAGSTTTCAAATTTGAGCT	reverse	(Khot et al. 2009)

28S-3r	CCGCTTCACTCGCCGRTACT	reverse	(Khot et al. 2009)
28S-2r	TTTCCTCCGCTTATTGATATGC	reverse	(Khot et al. 2009)
28S-1r	TATGCTTAAGTTCAGCGGGTA	reverse	(Khot et al. 2009)
18Sr_rp1	ATTCCTCGTTGAAGAGCA	reverse	(Hanson et al. 2016)
18S0067adegf	AAGCCATGCATGYCTAAGTATMA	forward	(Dollive et al. 2012)
UN.UP18S42f	CGTAACAAGGTTTCCGTAGGTGAAC	forward	(Levesque e Cock 2004)
UN.LO28S22r	GTTTCTTTTCCTCCGCTTATTGATATG	reverse	(Schurko et al. 2003)
TW14r	GCTATCCTGAGGGAAACTTC	reverse	T.J. White, unpublished
TW13r	GGTCCGTGTTTCAAGACG	reverse	T.J. White, unpublished
SR6Rf_2	AAGTANAAGTCGTAACAAGG	forward	Vilgalys Lab*
SR6Rf_1	AAGWAAAAGTCGTAACAAGG	forward	Vilgalys Lab*
SR1Rf	TACCTGGTTGATNCTGCCAGT	forward	Vilgalys Lab*
SLG1f	TTGCGCAACCTGCGGAAGGAT	forward	AFTOL*
PNS1f	CCAAGCTTGAATTCGTAGTCATATGCTTGTC	forward	AFTOL*
NSR399r	TCTCAGGCTCCYTCTCCGG	reverse	(Dollive et al. 2012)
NSI1f	GATTGAATGGCTTAGTGAGG	forward	(Martin e Rygiewicz 2005)
NSA3f	AAACTCTGTCGTGCTGGGGATA	forward	(Martin e Rygiewicz 2005)

NS7af	CAATAACAGGTCTGTGATGC	forward	UNITE-primers*
NS4r	CTTCCGTCAATTCCTTTAAG	reverse	(White et al. 1990)
NS41r	CCCGTGTTGAGTCAAATTA	reverse	AFTOL*
NS24Rf	TAAAAGTCGTAACAAGGTTT	forward	AFTOL*
NS1f	GTAGTCATATGCTTGTCTC	forward	(White et al. 1990)
nLSU1221r	CTAGATGAACYAACACCTT	reverse	(Schadt et al. 2003)
NLC2r	GAGCTGCATTCCCAAACAACCTC	reverse	(Martin e Rygiewicz 2005)
NLB4r	GGATTCTCACCCCTCTATGAC	reverse	(Martin e Rygiewicz 2005)
NLB3r	GGATTCTCACCCCTCTATGA	reverse	(Martin e Rygiewicz 2005)
NL6Bmunr	CAAGCGTTTCCCTTTCAACA	reverse	(Egger 1995)
NL4r	GGTCCGTGTTTCAAGACGG	reverse	(Kurtzman e Robnett 1997)
NL1f	GCATATCAATAAGCGGAGGAAAAG	forward	(Kurtzman e Robnett 1997)
LRORr	CCGCTGAACTTAAGCATATCAATA	reverse	(Amend et al. 2010)
LR7r	TACTACCACCAAGATCT	reverse	Vilgalys Lab*
LR6r	CGCCAGTTCTGCTTACC	reverse	Vilgalys Lab*
LR5r	TCCTGAGGGAAACTTCG	reverse	Vilgalys Lab*
LR5f	CGATCGATTTGCACGTCAGA	reverse	(Tedersoo et al. 2015)

LR3Rf	GTCTTGAAACACGGACC	forward	Vilgalys Lab*
LR3r	CCGTGTTTCAAGACGGG	reverse	Vilgalys Lab*
LR21r	ACTTCAAGCGTTTCCCTTT	reverse	Vilgalys Lab*
LR1r	GGTTGGTTTCTTTTCCT	reverse	Vilgalys Lab*
LR16r	TTCCACCCAAACTCG	reverse	Vilgalys Lab*
LR15r	TAAATTACAACCTCGGAC	reverse	Vilgalys Lab*
LR0Rf	ACCCGCTGAACTTAAGC	forward	Vilgalys Lab*
LR0Br	GGTAGTCCTACCTGATTTG	reverse	UNITE-primers*
LF402r	TTCCCTTTYARCAATTTAC	reverse	(Tedersoo et al. 2015)
LF402f	GTGAAATTGYTRAAAGGGAA	forward	(Tedersoo et al. 2015)
LF340r	TACTTGTKCGCTATCGG	reverse	(Tedersoo et al. 2008)
ITSOf	ACTTGGTCATTTAGAGGAAGT	forward	(Tedersoo et al. 2008)
ITS9munf	TGTACACACCGCCCGTCG	forward	(Egger 1995)
ITS8munr	CTTCACTCGCCGTTACTA	reverse	(Egger 1995)
ITS86r	TTCAAAGATTCGATGATTCAG	reverse	(Vancov e Keen 2009)
ITS86f	GTGAATCATCGAATCTTTGAA	forward	(Turenne et al. 1999)
ITS5Rr	CCTTGTTACGACTTTTACTTCC	reverse	AFTOL*

ITS5f	GGAAGTAAAAGTCGTAACAAGG	forward	(White et al. 1990)
ITS4Sr	CCTCCGCTTATTGATATGCTTAAG	reverse	AFTOL*
ITS4r	TCCTCCGCTTATTGATATGC	reverse	(White et al. 1990)
ITS4ngsr	TCCTSCGCTTATTGATATGC	reverse	(White et al. 1990)
ITS4KYO3r	CTBTTVCCKCTTCACTCG	reverse	(Toju et al. 2012)
ITS4KYO2r	RBTTTCTTTTCCTCCGCT	reverse	(Toju et al. 2012)
ITS4KYO1r	TCCTCCGCTTWTTGWTWTGC	reverse	(Toju et al. 2012)
ITS4BRr	TCAACAGACTTGACATGGTCC	reverse	(Vialle et al. 2009)
ITS4Br	CAGGAGACTTGACACGGTCCAG	reverse	(Gardes e Bruns 1993)
ITS4BR2r	GGATTATCACCCCTCAATGAT	reverse	(Bergeron et al. 2019)
ITS3Rf	ATCGATGAAGAACACAG	forward	Bruns Lab*
ITS3R3f	CACATCGATGAAGAACACAGT	forward	(Feau et al. 2011)
ITS3R2f	AAGGTACACCTGTTTGAGTGTC	forward	(Feau et al. 2011)
ITS3mix15f	CANCGATGAAGAACGYRG	forward	(Tedersoo et al. 2015)
ITS3KYO2f	GATGAAGAACGYAGYRAA	forward	(Toju et al. 2012)
ITS3KYO1f	AHCGATGAAGAACRYAG	forward	(Toju et al. 2012)
ITS3f	GCATCGATGAAGAACGCAGC	forward	(White et al. 1990)

ITS2r	CGCTGCGTTCTTCATCG	reverse	Vilgalys Lab*
ITS2R2r	GACTCTCAAACAGGTGTACCTT	reverse	(Vialle et al. 2009)
ITS2Melampsorar	CACTGTGTTCTTCATCGACG	reverse	Bruns Lab*
ITS2KYO2r	TTYRCTRCGTTCTTCATC	reverse	(Toju et al. 2012)
ITS2KYO1r	CTRYGTTCTTCATCGDT	reverse	(Toju et al. 2012)
ITS2f	GTGARTCATCGAATCTTT	forward	(Liguori et al. 2015)
ITS2br	GCTGCGTTCTTCATCGATGC	reverse	(White et al. 1990)
ITS2ar	TGTGTTCTTCATCGATG	reverse	Bruns Lab*
ITS1Pf	TTATCATTTAGAGGAAGGAG	forward	UNITE-primers*
ITS1P2f	CTTTATCATTTAGAGGAAGGAG	forward	Bruns Lab*
ITS1ngsf	TCCGTAGGTGAACCTGC	forward	(Tedersoo et al. 2015)
ITS1KYO2f	TAGAGGAAGTAAAAGTCGTAA	forward	(Toju et al. 2012)
ITS1KYO1f	CTHGGTCATTTAGAGGAASTAA	forward	(Toju et al. 2012)
ITS1f	TCCGTAGGTGAACCTGCGG	forward	(White et al. 1990)
ITS1Fngsf	GGTCATTTAGAGGAAGTAA	forward	(Tedersoo et al. 2015)
ITS1FKYO2f	TAGAGGAAGTAAAAGTCGTAA	forward	(Toju et al. 2012)
ITS1Ff	GTAAAAGTCGTAACAAGGTTTC	forward	(Willger et al. 2014; Lennon et al. 2010)

ITS1ar	GTTCAAAGAYTCGATGATTCAC	reverse	(Willger et al. 2014)
ITS1af	CTTGGTCATTTAGAGGAAGTAA	forward	(Gardes e Bruns 1993)
ITS148f	ACACACCGCCCGTCGCTACT	forward	(Usyk et al. 2017)
ITS134f	CTGCCCTTTGTACACACCGC	forward	(Usyk et al. 2017)
ITS130f	GTCCCTGCCCTTTGTACACA	forward	(Usyk et al. 2017)
ITS127f	TACGTCCCTGCCCTTTGTAC	forward	(Usyk et al. 2017)
ITS1217r	TTTCGCTGCGTTCTTCATCG	reverse	(Usyk et al. 2017)
ITS10munr	GCTGCGTTCTTCATCGAT	reverse	(Egger 1995)
Illumina18S.775F.1f	TGARRAAATTAGAGTGTTCAAAGCAG	forward	(Nash et al. 2017)
Illumina18S.1121R.1r	CCTTYAAGTTTCAGCCTTGCG	reverse	(Nash et al. 2017)
gITS7f	GTGAATCATCGARTCTTTG	forward	(Ihrmark et al. 2012)
fITS7r	CAAAGATTCGATGAYTCAC	reverse	(Tedersoo et al. 2015)
fITS7f	GAACACAGCGAAATGTGA	forward	(Ihrmark et al. 2012)
fITS7af	GTGARTCATCGAATCTTTG	forward	(Ihrmark et al. 2012)
End18Sf	GTAAAAGTCGTAACAAGGTTTC	forward	(Findley et al. 2013)
CTB6f	GCATATCAATAAGCGGAGG	forward	(Garbelotto et al. 1997)
BITSf	ACCTGCGGARGGATCA	forward	(Bokulich e Mills 2013)

\*

AFTOL - <https://lutzonilab.org/aftol/primers/>

Bruns Lab - <https://nature.berkeley.edu/brunslab/tour/primers.html>

UNITE-Primers - <https://unite.ut.ee/primers.php>

Vilgalys Lab - <https://sites.duke.edu/vilgalyslab>

## Apêndice C – Terceiro sequenciamento ITS2 com 51 amostras de pênis

**Tabela 16** – Número de *reads* obtidas por amostra a partir do terceiro sequenciamento, após a etapa de demultiplexação, identificação dos primers, considerando o tamanho mínimo de 100 pb, realizado pelo software AMPtk.

Amostra	Barcode	Número de Sequências
CaPe10T-B	BC.1	540
CaPe17AM	BC.2	4740
CaPe25AM	BC.93	1356
CaPe27AM	BC.3	956
CaPe35AM	BC.17	1763
CaPe37AM	BC.33	2052
CaPe40T	BC.4	721
CaPe41T	BC.5	692
CaPe42AM	BC.18	1111
CaPe44N	BC.6	654
CaPe45AM	BC.7	2181
CaPe45T	BC.8	719
CaPe46AM	BC.34	1547
CaPe46T-B	BC.95	5586
CaPe48AM	BC.35	1550
CaPe48T-AM	BC.9	325
CaPe49AM	BC.36	2879
CaPe49N	BC.10	305
CaPe50AM	BC.11	693

**Tabela 16 – continuação da página anterior**

<b>Amostra</b>	<b>Barcode</b>	<b>Número de Sequências</b>
CaPe50T	BC.12	770
CaPe51AM	BC.37	1503
CaPe52T	BC.38	1471
CaPe53AM	BC.19	2515
CaPe53T	BC.39	1802
CaPe54T	BC.40	548
CaPe55AM	BC.20	745
CaPe55N	BC.13	560
CaPe55T	BC.14	1070
CaPe57AM	BC.41	1091
CaPe57T	BC.15	604
CaPe58T	BC.21	1650
CaPe59AM	BC.42	1083
CaPe59N	BC.43	1631
CaPe59T	BC.16	668
CaPe60T	BC.22	1045
CaPe61N	BC.23	1322
CaPe62N	BC.25	498
CaPe62T	BC.26	1759
CaPe63AM	BC.44	767
CaPe64AM	BC.45	980
CaPe65N	BC.27	2737
CaPe66AM	BC.28	3045
CaPe67AM	BC.29	4364
CaPe68AM	BC.30	994
CaPe70AM	BC.31	957

**Tabela 16 – continuação da página anterior**

<b>Amostra</b>	<b>Barcode</b>	<b>Número de Sequências</b>
CaPe71AM	BC.46	1249
CaPe72AM	BC.96	8203
CaPe73AM	BC.47	1864
CaPe74AM	BC.48	3564
CaPe75AM	BC.32	826
CaPe76AM	BC.94	2705