

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROBIOLOGIA
AGRÍCOLA E DO AMBIENTE

Raíssa Nunes dos Santos

**IDENTIFICAÇÃO DE VÍRUS EM AMOSTRAS DE MACACOS-PREGO
COLETADAS NO SUL DO BRASIL UTILIZANDO METAGENÔMICA**

Porto Alegre

2020

Raíssa Nunes dos Santos

**IDENTIFICAÇÃO DE VÍRUS EM AMOSTRAS DE MACACOS-PREGO
COLETADAS NO SUL DO BRASIL UTILIZANDO METAGENÔMICA**

Tese apresentada ao Programa de Pós-Graduação em Microbiologia Agrícola e do Ambiente do Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Doutora em Microbiologia Agrícola e do Ambiente.

Orientador(a): Profa. Dr. Ana Cláudia Franco

Coorientador(a), se houver: Prof. Dr. Fabrício Souza Campos

Porto Alegre

202

iv

CIP - Catalogação na Publicação

Santos, Raíssa
IDENTIFICAÇÃO DE VÍRUS EM AMOSTRAS DE MACACOS-PREGO
COLETADAS NO SUL DO BRASIL UTILIZANDO METAGENÔMICA /
Raíssa Santos. -- 2020.

56 f.

Orientadora: Ana Claudia Franco.

Coorientadora: Fabrício Campos.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Programa de Pós-Graduação em Microbiologia Agrícola e do Ambiente, Porto Alegre, BR-RS, 2020.

1. Viroma. 2. Sequenciamento de alta performance. 3. Macacos-prego. 4. Vírus em primatas. 5. Análise metagenômica. I. Franco, Ana Claudia, orient. II. Campos, Fabrício, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

“Em algum lugar algo incrível está esperando para ser descoberto”

[Carl Sagan](#)

AGRADECIMENTOS

Ser cientista é algo que carrego desde a infância então, não poderia começar os agradecimentos sem dedicar estes anos de estudo aos meus pais. Eles foram a base, o refúgio e o incentivo para sempre seguir no caminho que decidi. Junto a eles, meu irmão, que também esteve presente em todos estes momentos. Mas que sorte a minha de ainda ter mais dois casais de pais, Sílvio e Leonor e meus pais uruguaios, Carlos Julio e Selva, onde parte deste trabalho foi escrito em suas casas.

Agradeço aos meus colegas de laboratório por todo suporte nestes anos, principalmente ao Fernando Finoketti e à Anne Caroline dos Santos. Vocês foram muito importantes nesta jornada e com certeza estarão presentes nas próximas. E no quesito coletas e aventuras, gostaria de agradecer a grande “chefa” Aline Campos, com todos seus ensinamentos profissionais e pessoais. Muito obrigada!

Agradeço a todos os meus amigos que estiveram comigo nestes momentos aos quais gostaria de citar Janie Silva, Paolo Cerri, Karolina Hernandez, Heloísa Giacomelli e Mariane Facco por alegrarem meus dias. Dentro do grupo dos amigos gostaria de agradecer também à Thabata Barbosa por me presentear com uma linda afilhada neste período. Vejam, são 4 anos, muitas coisas acontecem na nossa vida e nem só de doutorado se vive!

Agradeço ao meu co-orientador, Prof. Dr. Fabrício Souza Campos, pelos ensinamentos e suporte durante esta tese. Agradeço também a minha orientadora professora Dra. Ana Cláudia Franco, por acreditar no meu trabalho desde o mestrado e me ajudar a concluir esta etapa em meio à pandemia de Covid-19 onde o Laboratório de Virologia (ICBS/UFRGS) encontra-se operando para realização dos testes de diagnóstico.

Por fim, gostaria de agradecer à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ou Conselho Nacional de Desenvolvimento Científico (CNPq) pela concessão da bolsa.

IDENTIFICAÇÃO DE VÍRUS EM AMOSTRAS DE MACACOS-PREGO COLETADAS NO SUL DO BRASIL UTILIZANDO METAGENÔMICA

RESUMO

A espécie *Sapajus nigritus*, compreende primatas não-humanos bem distribuídos na América do Sul. Eles são onívoros e vivem em grupos de até 40 indivíduos. As superfícies mucosas são um portal perfeito de entrada para vírus e bactérias do ambiente para acessar novos hospedeiros. A cavidade oral é um portal de entrada para vários microrganismos, incluindo vírus, e compreende a gengiva, língua, paladar, a superfície dos dentes e saliva. A descoberta periódica de novos vírus sugere que subestimamos o número total e a diversidade destes na natureza. Por outro lado, a crescente melhoria da detecção de vírus contribui para a descoberta de um mundo microscópico ainda inexplorado. Bacteriófagos (ou fagos) são o grupo mais abundante de vírus. Eles desempenham um papel dinâmico no ecossistema à medida que se replicam em hospedeiros bacterianos que habitam todos os tipos de ambientes, incluindo superfícies corporais de diferentes espécies de animais. Este estudo tem como objetivo apresentar a composição dos vírus em suabes orais, sangue e urina de macacos-capuchinhos selvagens (*Sapajus nigritus*) coletados de 12 animais em dois pontos de coleta distintos. As amostras foram coletadas seguindo as orientações do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (Ibama), armazenadas no gelo e transportadas para o laboratório. Posteriormente, foram processadas em três *pools* distintas (cotonete, sangue e urina), e a extração de ácidos nucleicos foi realizada utilizando-se Trizol®, de acordo com as instruções do fabricante. Após a extração, os ácidos nucleicos foram aleatoriamente amplificados e submetidos à plataforma Illumina para sequenciamento da próxima geração. Os resultados revelaram a presença de bacteriófagos pertencentes às famílias *Siphoviridae*, *Myoviridae* e *Podovirida* e 11 famílias virais eucarióticas: *Herpesviridae*, *Parvoviridae*, *Papillomaviridae*, *Polyomaviridae*, *Caulimoviridae*, *Iridoviridae*, *Astroviridae*, *Poxviridae* e *Baculoviridae*, além de dois genomas virais completos de *Anelloviridae* e *Genomoviridae*. Este estudo contribui para o acesso às sequências virais de macacos capuchinhos, melhorando nosso conhecimento restrito sobre comunidades microbianas em primatas não humanos.

¹Tese de Doutorado em Microbiologia Agrícola e do Ambiente – Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (56 p.) Dezembro, 2020.

VIRUS IDENTIFICATION IN SAMPLES OF CAPUCHIN MONKEYS COLLECTED IN SOUTH OF BRAZIL USING METAGENOMIC APPROACH

ABSTRACT

Sapajus nigritus, is a specie of non-human primates which are widespread in South America. They are omnivores and live in troops of up to 40 individuals. Mucosal surfaces are a perfect portal of entry for viruses and bacteria from the environment to access new hosts. The oral cavity is a portal of entry for several microorganisms including viruses, and it contains the gingiva, tongue, palate, the surface of teeth and saliva. The periodic discovery of new viruses suggests that we underestimate the total number and diversity of viruses existing in nature. On the other hand, the growing improvement of virus detection contribute to the discovery of a microbiology world yet unexplored. Bacteriophages (or phages) are the most abundant group of viruses. They play a dynamic role in the ecosystem as they replicate in bacterial hosts which inhabit all types of environments, including body surfaces of different species of animals. This study aims unveiling the viruses composition in oral swabs, blood and urine of wild capuchin monkeys (*Sapajus nigritus*) collected from 12 animals in two distinct collection points. Samples were collected following the Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) guidelines, stored on ice and transported to the lab. Samples were processed in three distinct pools (swab, blood and urine), and the extraction of nucleic acids was performed using Trizol®, according to the manufacturer's instructions. Following extraction, nucleic acids were randomly amplified and submitted to Illumina platform for next generation sequencing. The results revealed the presence of bacteriophages belonging to the families *Siphoviridae*, *Myoviridae*, and *Podovirida* and 11 eukaryotic viral families: *Herpesviridae*, *Parvoviridae*, *Papillomaviridae*, *Polyomaviridae*, *Caulimoviridae*, *Iridoviridae*, *Astroviridae*, *Poxviridae* and *Baculoviridae*, in addition to two complete viral genomes of *Anelloviridae* and *Genomoviridae*. This study contributes to access the viral sequences from capuchin monkeys, lead to improve our restrict knowledge regarding microbial communities in non-human primates.

¹Doctoral Dissertation in Agricultural and Environmental Microbiology – Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, 56p, december, 2020.

SUMÁRIO

1.	INTRODUÇÃO	15
2.	OBJETIVOS	17
2.1	Objetivo Geral	17
2.2	Objetivos Específicos.....	17
3.	REVISÃO DA LITERATURA	18
3.1	Primatas não-humanos objetos deste estudo: <i>Sapajus nigritus</i>	18
3.2	Metagenômica: contextualização do termo e viroma.....	19
3.3	Amostras biológicas.....	19
3.3.1	<i>Suabe oral</i>	20
3.3.2	<i>Sangue total</i>	21
3.3.3	<i>Suabe retal</i>	22
3.3.4	<i>Urina</i>	22
3.4	Bacteriófagos.....	22
3.5	Sequenciamento de alta performance.....	23
3.6	A era dos giga dados: como analisar?.....	25
4	MATERIAL E MÉTODOS.....	28
4.1	Coletas.....	28
4.2	Processamento de amostras.....	28
4.3	Extração de DNA e RNA para processamento de vírus.....	29
4.4	Enriquecimento do RNA extraído.....	29
4.5	Enriquecimento do DNA extraído.....	29
4.6	Purificação das amostras: AMPure ®.....	30
4.7	Preparo da biblioteca genômica.....	30
4.8	Análise dos dados gerados no sequenciamento.....	32
4.9	Construção de árvores filogenéticas.....	32
5.	ARTIGOS.....	33
5.1.	Artigo 1:	33
5.2.	Artigo 2:.....	34
5.3.	Resultados não apresentados na forma de artigo.....	47
6.	DISCUSSÃO GERAL.....	50
7.	CONCLUSÃO.....	53
8.	REFERÊNCIAS.....	54

LISTA DE TABELAS

	Página
Tabela 1. Relação das amostras coletas, processadas ou armazenadas para futuros estudos.	68
Tabela 2. Compilação do volume de dados de sequenciamento gerados neste estudo e seu status de análise.	71

LISTA DE FIGURAS

	Página
Figura 1. <i>Sapajus nigritus</i> no momento da captura em gaiola Tomahack. Os animais foram observados por 10 minutos para as condições de bem-estar animal	19
Figura 2. Cavidade oral de um chimpanzé e seus órgãos	21
Figura 3. Fluxograma esquemático dos procedimentos para obtenção das amostras para sequenciamento de alta performance.	69
Figura 4. Fluxograma gerado para análise das amostras conforme descrito no material e métodos deste manuscrito	70

LISTA DE ABREVIATURAS E SIGLAS

BLAST	Basic Local alignment Search tool
DNA	Ácido desoxirribonucleico
GC	Guanina-Citosina
Kg	Quilogramas
rRNA	RNA ribossomal
RNA	Ácido ribonucleico

1. INTRODUÇÃO

Os primeiros estudos de comunidades microbianas priorizavam a detecção dos micro-organismos utilizando amplificação e sequenciamento do gene do 16SrRNA. Recentemente, com as atualizações e acessibilidade ao sequenciamento de alta performance, mais estudos têm sido realizados em diferentes amostras e incluindo não só procariontes, mas também fungos e vírus. Estas técnicas permitiram a detecção de microrganismos ainda não identificados em um menor tempo, gerando uma quantidade de dados muito maior do que o que permite o sequenciamento guiado por *primers* com sequência conhecida, sendo completamente independente de cultivo.

Os primatas não humanos da espécie *Sapajus nigritus* são pequenos mamíferos (0,5 kg – 5,0 kg) distribuídos na América do Sul. A dieta destes animais é composta de pequenos roedores, insetos, folhas e frutos. Normalmente vivem em grupos de aproximadamente 40 animais (Dunn, 2016) habitando matas e fragmentos de matas. Entretanto, interações próximas com seres humanos tem sido relatadas ligadas à pressão antropogênica estabelecida principalmente nos setores de agricultura, caça e turismo. A proximidade entre esses animais e humanos aumenta o risco de acidentes, levando a possibilidade de transmissão de patógenos (Wolfe, 1998).

Estudos demonstram que a quantidade de vírus no planeta fica na ordem de 10^{31} partículas, o que representa mais que o número de estrelas no universo. Apesar desse número, devido ao seu pequeno tamanho, os vírus contribuem com uma biomassa muito pequena, comparado com outros organismos (Weitz & Wilhelm, 2013). Embora os vírus que infectam eucariontes sejam mais conhecidos por estarem associados a patologias, não só em humanos mas também em outras espécies, a maioria dos vírus desempenha um papel fundamental nos ciclos biogeoquímicos, controlando populações microbianas e tornando nutrientes disponíveis (Suttle, 2007).

A possibilidade de utilizar recursos de sequenciamento de alto desempenho, juntamente com o trabalho em conjunto com os órgãos de fiscalização e de saúde do estado consolidaram a ideia desta pesquisa.

Pesquisas tem sido conduzidas para a caracterização da composição bacteriana de animais silvestres, porém pouco tem se feito em relação à comunidade viral. Isso ocorre em parte porque a quantidade de material genético de vírus é proporcionalmente menor do que o genoma bacteriano ou fúngico, o que representa um grande desafio à comunidade científica para recuperação e identificação das sequencias. Atualmente temos na região Sul do Brasil alguns pontos onde a interação com primatas não-humanos é intensa e estimulada pelo turismo informal. Estes locais são focos que precisam de atenção e representam um alerta ao risco de acidentes e a identificação dos vírus que circulam nesses animais se torna importante por causa disso. Visando compreender melhor sobre a circulação de vírus em primatas não humanos brasileiros, este trabalho aborda a caracterização de vírus em amostras de *Sapajus nigritus* de ocorrência no estado do Rio Grande do Sul.

2. OBJETIVOS

2.1 Objetivo Geral

Este estudo tem como objetivo caracterizar a composição viral de amostras de macacos-prego de vida livre da espécie *Sapajus nigritus*, utilizando o sequenciamento de alta performance

2.2 Objetivos Específicos

2.2.1 Identificação de vírus presentes em amostras de suabe oral de um grupo de macacos-prego de vida livre coletados no Estado do Rio Grande do Sul (artigo 1)

2.2.2 Identificação da composição de bacteriófagos em amostras de suabe oral, urina e sangue de dois grupos de macacos-prego de vida livre coletados no Estado do Rio Grande do Sul (artigo 2)

2.2.3 Estabelecimento de um protocolo para recuperação de ácidos nucleicos virais para análise metagenômica.

2.2.4 Estabelecimento de um pipeline de análise de sequências virais obtidas através do sequenciamento de alta performance na plataforma *Illumina MiSeq*.

3. REVISÃO DA LITERATURA

Nesta seção estarão descritos temas relevantes à execução deste projeto de pesquisa.

3.1 Primatas não-humanos objetos deste estudo: *Sapajus nigritus*

Os macacos-prego pertencem à família *Cebidae*, que inclui o gênero *Sapajus*, a qual compreende 29 espécies. Estes animais apresentam um porte pequeno a médio (0,5 kg – 5,0 kg), e uma longevidade medida em cativeiro de cerca de 40 anos. Atualmente, sua densidade populacional encontra-se na faixa entre comum e abundante, mas vem decaindo em decorrência da fragmentação das florestas, seu habitat natural. A espécie *Sapajus nigritus*, popularmente conhecida por macaco-prego, é ubiquamente encontrada pela América do Sul, abrangendo os países Colômbia, Venezuela, Guianas, Paraguai, Argentina e Brasil (IUCN, 2018). O rápido crescimento e pressão das atividades antrópicas como agricultura, caça e turismo de exploração animal tem estreitado as interações entre os primatas não humanos e a população humana, aumentando o risco de exposição à microbiota desses primatas.

A sua dieta é basicamente composta de insetos, frutos e pequenos mamíferos. Estes animais são habilidosos na manipulação de ferramentas, galhos e pedras (IUCN, 2018). Em busca de alimento, é comum invadirem áreas de lazer como parques, uma vez que muitas pessoas estimulam este comportamento ao oferecer comida aos animais de vida livre. A captura desses animais é simples e feita por intermédio de gaiolas, uma vez que eles são facilmente atraídos pelo barulho e cheiro de alimentos oferecidos. Na imagem representada pela Figura 1 podemos ver um dos animais capturados individualmente nas gaiolas antes da sedação.



Figura 1. *Sapajus nigritus* no momento da captura em gaiola Tomahack. Os animais foram observados por 10 minutos para as condições de bem-estar animal

3.2 Metagenômica: contextualização do termo e viroma

A utilização do termo “metagenoma” refere-se à presença de genomas em uma amostra e foi cunhado primeira vez em 1998, quando um grupo de pesquisadores descreveu a importância dos microrganismos do solo como geradores de novos compostos essenciais (Handelsman et. al, 1998). Este conceito abriu uma nova geração de pesquisas que visa a prospecção de microrganismos totais em uma amostra através da detecção do seu material genético, mesmo que estes não sejam cultiváveis em condições laboratoriais (Sleator et al., 2008). Os estudos de metagenoma podem ter enfoques distintos, que podem incluir a composição da população demicroorganismos e sua dinâmica, definida como estudo estrutural (Haldelsman, 2005) ou a identificação de genes de interesse, que envolve a produção de bibliotecas oriundas da expressão destes genes, denominada de análise funcional (Vieites et. al., 2009).

Os vírus são as entidades biológicas mais numerosas do Planeta Terra, com uma população estimada de 10^{31} partículas. Eles habitam diversos ambientes incluindo oceanos, rios, animais aquáticos e os próprios sistemas do corpo humano. Historicamente, as pesquisas são focadas em vírus que causam impacto à saúde humana, como os vírus influenza ou vírus que

causam infecções rápidas no organismo como o vírus Ebola. No entanto, as evidências científicas sugerem que os vírus de procariontes também podem impactar a saúde, afetando diretamente a estrutura e função das comunidades bacterianas que simbioticamente interagem entre si e com organismos mais complexos como os mamíferos (Garmaeva, 2019). Os vírus que infectam bactérias são chamados de bacteriófagos e desempenham um papel importante em ecossistemas com alta abundância de procariontes (Rohwer, 2009). Recentemente, com a evolução de técnicas de sequenciamento e a expansão de estudos caracterizando as comunidades virais, os vírus ganharam terminologias “ome” e “omics”, cujos termos se referem a estudos que englobam todos os vírus habitando um mesmo ecossistema a partir de estudos de ácidos nucleicos.

3.3 Amostras biológicas

Nesta seção serão descritas as características dos locais escolhidos para amostragem nos animais. Trabalhou-se com amostras de sangue, suabes coletados na cavidade oral e retal e de alguns animais foi possível coletar urina.

3.3.1 Suabe oral

A cavidade oral é composta por diferentes tipos de tecido que tornam disponíveis uma variedade de superfícies para colonização microbiana. Esta cavidade é a entrada para o canal alimentar do trato digestivo. Alguns estudos demonstram que a cavidade oral abriga sua própria e única comunidade microbiana. As áreas supra gengival são continuamente expostas a mudanças ambientais de acordo com a alimentação, nos animais, e também pela higiene, em seres humanos. Já a área sub gengival consiste em superfícies de epitélio que são constantemente descamadas e embebidas por saliva. A composição microbiana da área sub gengival é selecionada pela falta de oxigênio e disponibilidade de aminoácidos e peptídeos necessários para o crescimento dos microrganismos (Gaetti-Jardim et al.,2012). A imagem a seguir (Figura 2) demonstra a

representação de um chimpanzé e suas estruturas.

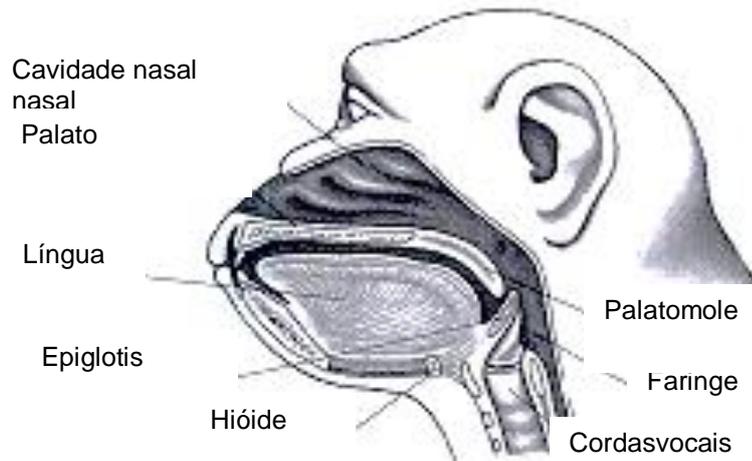


Figura 2. Cavidade oral de um chimpanzé e seus órgãos. Figura adaptada de <http://home.clear.net.nz/pages/robert.fromont/ling/408_tract.htm>

3.3.2 Sangue total

O sangue total é composto basicamente por células e plasma. O plasma é formado por água, proteínas (albumina, globulinas, e proteínas de coagulação) e sais inorgânicos. Já a parte celular, que tem origem na medula óssea, é constituída de eritrócitos, plaquetas e leucócitos. Alguns estudos mostram a presença de sequencias virais em amostras de sangue periférico, incluindo membros das famílias *Herpesviridae* e *Anelloviridae*, mesmo na ausência de quadros relativos a potenciais doenças (Breitbard et al., 2005). Os primatas não humanos são historicamente fontes potenciais de vírus zoonóticos e modelos experimentais para pesquisas de muitos patógenos que infectam humanos. No entanto, com exceção do vírus da imunodeficiência de primatas (SIV) pertencente à família *Retroviridae*, os vírus que podem ser transmitidos e circulam no sangue nestes animais

selvagens continuam pouco caracterizados (Bailey et al., 2016).

3.3.3 *Suabe retal*

O Sistema gastrointestinal é complexo, iniciando na cavidade oral e continuando através do estômago e intestino, tendo seu final no ânus. As condições químicas e físicas, bem como as influências ambientais desempenham funções determinantes na composição e densidade microbiana presente (Zarate et al., 2018). Bactérias ocorrem desde a cavidade oral até o fim do intestino. Em relação ao volume dos órgãos, os humanos apresentam cerca de 10-24% do volume do intestino no tamanho do seu estômago, 56-67% em relação ao intestino delgado e 17-23% relação ao colon. Já os chimpanzés e orangotangos apresentam 17-20% do volume em estômago, 23-28% para o intestino delgado e 52-54% para o cólon aproximadamente (Milton, 1987).

3.3.4 *Urina*

O trato urinário é constituído por rins, uereteres, bexiga e uretra. Os rins são considerados parte do trato urinário superior enquanto as outras estruturas parte do trato inferior. Normalmente é constituído de dois rins, localizados retro-peritoniais na parte posterior da parede abdominal ambos ao lado da coluna vertebral (Elbadawi, 1996). Ao contrário do trato gastrintestinal e respiratório, o trato urinário, em circunstâncias normais é tipicamente considerado estéril. Em seres humanos, as infecções do trato urinário ao longo da vida de uma pessoa saudável são infrequentes (Hickling, 2015).

3.4 Bacteriófagos

Os bacteriófagos (ou fagos) desempenham um papel dinâmico no ecossistema, sendo o grupo mais abundante de vírus, replicando-se em hospedeiros bacterianos. Durante a infecção os fagos interagem com a superfície bacteriana e injetam seu material genético para dentro da célula. Em seguida, o bacteriófago segue o ciclo de multiplicação por uma das duas principais rotas que compreendem seus ciclos de multiplicação: via lítica ou lisogênica.

Os ciclos líticos são letais para a célula do hospedeiro e culminam na produção de novos fagos, a exemplo dos bem descritos fagos Mu e T7 (que infectam *Escherichia coli*). Eles desempenham um importante papel na regulação da abundância do seu hospedeiro. Em contraste, o ciclo lisogênico se refere à replicação que não resulta na produção direta de virions, característica dos fagos temperados. Sob certas condições, como por exemplo danos no DNA ou baixa concentração de nutrientes, estes fagos espontaneamente removem seu DNA do genoma do hospedeiro e dão início ao ciclo lítico (Casjens, 2015). Essa remoção é chamada de indução, podendo ocorrer a remoção também de regiões específicas do genoma bacteriano. Já a habilidade dos fagos de transferir genes de uma bactéria para outra a partir da conversão lisogênica ou transdução pode aumentar a diversidade de espécies. Este fenômeno pode causar a disseminação de toxinas, genes relacionados à virulência e mesmo a possibilidade de genes de resistência à antimicrobianos entre a população bacteriana (Touchon, 2017).

3.5 Sequenciamento de alta performance

O sequenciamento de alto desempenho necessita das etapas de extração de ácidos nucleicos, amplificação e purificação. A biblioteca para o sequenciamento de alto desempenho é definida como um conjunto de fragmentos que podem representar tanto o total de ácidos nucleicos como uma região alvo. Cada plataforma de sequenciamento tem suas especificações, mas de maneira geral a preparação da biblioteca começa com a fragmentação do material inicial, seguida da adição de sequências adaptadoras que irão se conectar a estes fragmentos. Uma biblioteca de boa qualidade é formada por todos os fragmentos equalizados, ou seja, igualmente representados em quantidade, bem como não deve conter erros gerados no processo de amplificação das sequências, o que poderia gerar produtos não-específicos (Aird, 2011).

A primeira etapa para o preparo das bibliotecas na maioria dos fluxogramas de trabalho é a fragmentação, mencionada anteriormente. Este procedimento pode ser realizado por métodos físicos ou enzimáticos. Os

métodos físicos incluem cisalhamento acústico, sonicação e cisalhamento hidrodinâmico. Já os métodos enzimáticos incluem a digestão utilizando DNase I, Fragmentase e/ou Knierim. O resultado entre a fragmentação física ou química é similar e fica a critério do pesquisador e das facilidades do laboratório na escolha de qual método utilizar (Head, 2014; Van Dijk, 2014).

Uma vez que o DNA tenha sido fragmentado, a etapa seguinte consiste na junção dos fragmentos adaptadores. Os adaptadores são introduzidos para a criação de sequências com início e fim conhecidos dentro do processo de sequenciamento. Uma estratégia alternativa foi desenvolvida e combina o processo de fragmentação e ligação dos adaptadores em um único passo, tornando o processo mais simples, rápido e necessitando de menos quantidade de amostra inicial. Este processo é conhecido como tagmentação, e é baseado na tecnologia de utilização de transposons.

A próxima etapa consiste no enriquecimento, onde o total de material alvo é amplificado para a criação da biblioteca a ser sequenciada. Basicamente, dois métodos são comumente utilizados para este objetivo: o método de captura de sequenciamento híbrido e o método baseado em *amplicon* (Samorodnitsky, 2015). No método híbrido de captura, as moléculas fragmentadas são hibridizadas especificamente com fragmentos de DNA completos as regiões alvo de interesse. Isso pode ser realizado por diferentes métodos, como *microarray* ou o uso de oligonucleotídeos biotinizados que visam a captura física e isolamento das sequências de interesse (Horn, 2012). Já os métodos baseados em *amplicon* consistem no uso de oligonucleotídeos complementares às regiões flancas do DNA a ser sequenciado. Este método tem algumas limitações intrínsecas, como a formação de duplicatas, competição de oligonucleotídeos e amplificação não uniforme das regiões alvo devido à variação no conteúdo de GC (Hung, 2018).

A plataforma Illumina, escolhida para o desenvolvimento deste trabalho, utiliza como princípio a detecção de sinais dependentes de fluorescência. O mecanismo de detecção é a partir de medições fluorescentes de intensidades de sinal, resultantes dos nucleotídeos incorporados durante cada ciclo de amplificação. A empresa desenvolvedora afirma que sua tecnologia oferece uma maior porcentagem de leituras sem erros. As

versões mais recentes foram otimizadas para permitir a chamada de base precisa em regiões genômicas difíceis, como porções ricas em GC, homopolímeros e palíndromos. Além disto, os dideoxynucleotídeos (DNTPs), foram quimicamente modificados para conter um grupo de bloqueio reversível que age como um exterminador temporário para a polimerização do DNA. Após a incorporação de cada dNTP, a imagem é processada para identificar a base correspondente e, em seguida, clivar enzimaticamente para permitir a incorporação da próxima (Illumina, 2020).

3.6 A era dos giga dados: como analisar?

A evolução das plataformas de sequenciamento, possibilitou a geração de dados de forma mais eficiente e menos custosa, tornando possível o sequenciamento do genoma humano em uma semana a um preço considerável de U\$1000,00. Consequentemente, a quantidade de dados gerados demanda habilidades computacionais para analisar e interpretar as sequências produzidas. Devido a isto, um desenvolvimento considerável em bioinformática vem sendo aplicado, contudo, dependente da capacidade computacional (hardware) e de algoritmos e aplicativos para auxiliar todas as etapas necessárias que incluem o processamento de dados brutos bem como análises mais detalhadas.

A etapa inicial da análise dos dados, e primordial para dar seguimento a análises mais robustas, é a avaliação da qualidade dos dados gerados pelo sequenciamento, que consiste em filtrar as reads e remover as sequências adaptadoras (*trimming*). O parâmetro que mede a qualidade das bases sequenciadas nas plataformas Illumina e IonTorrent é o Phred score. Um Phred score de 10 (Q10) se refere à probabilidade de uma base estar incorreta a cada 10 sequenciadas, ou seja, uma acurácia de 90%, já um Q30 significa uma base errada a cada 1000 ou 99,9% de acurácia (Ewing, 1998). Os arquivos utilizados para este controle inicial são do formato Fastq, que contém as sequências brutas, os nomes e os valores atribuídos às qualidades das bases sequenciadas. A qualidade das sequências brutas é um ponto crucial para o sucesso das análises do sequenciamento, e devido a isto, existem várias ferramentas de bioinformática que foram desenvolvidas para avaliar a qualidade dos dados. A ferramenta FastQC, desenvolvida por

Andrews e colaboradores, é comumente empregada e sua análise gera informações gráficas e bem estruturadas sobre diferentes aspectos da qualidade das *reads*. Se as sequencias obtiverem qualidade suficiente, então podem ser utilizadas para as próximas etapas, que incluem alinhamento e mapeamento com bancos de dados genômicos.

Após a checagem da qualidade e remoção dos adaptadores, a próxima etapa que envolve a análise dos dados é o alinhamento das sequencias. Preferencialmente, uma vez que as sequencias geradas tem tamanhos curtos, o melhor método de montagem é utilizando genomas de referência. No entanto, o crescente número de estudos de metagenomas e a identificação inédita de muitas sequencias não torna possível este tipo de abordagem. Esse método faz um mapeamento das sequencias nos locais de alinhamento com as sequencias referência, tolerando um número (que pode ser definido pelo usuário) de “*mismatches*”, ou seja, variações nos nucleotídeos alinhados. Existe uma gama de ferramentas que fazem este procedimento, dentre estas o Burrows-Wheeler Aligners (BWA) (Li, 2010) e o Bowtie (Langmead, B, 2010). O BWA é utilizado tanto para sequencias longas quanto curtas e é um eficiente e rápido algoritmo para alinhamento e conversão de sequencias em outros formatos. Já o Bowtie, traz como vantagens ser mais rápido em algumas formas de alinhamento, no entanto, este atributo pode comprometer a qualidade, diminuindo a sensibilidade e a acurácia.

A montagem “*De Novo*”, é utilizada sem genomas referência e se baseia sobreposição de regiões iguais, seguida de sobreposição pelos finais fragmentados das sequencias. Esta estratégia traz limitações, principalmente por erros devido a sequencias pequenas e dificuldade de identificar as regiões que serão sobrepostas. Esta técnica de montagem é mais lenta e requer mais recursos computacionais (principalmente memória) se comparada às técnicas de mapeamento com genomas referência (Flicek, 2009). Apesar de apresentar vieses inerentes à técnica, esta é a montagem utilizada para análises de metagenomas, demandando horas de verificação manual para verificação das sequencias geradas.

Finalmente, uma análise funcional deve ser aplicada de modo a contextualizar os achados. A análise funcional consiste na anotação de

genes e verificação das variações genéticas. Estas análises podem ser feitas a nível de sequencias curtas, dados brutos pós checagem de qualidade, ou após passarem por softwares de montagem que irão agrupar as *reads* em sequencias maiores, denominadas de *contigs*. Um dos maiores desafios nesta etapa de análise é a caracterização de genes/genomas ainda não identificados e depositados nos bancos de dados como o GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

4 MATERIAL E MÉTODOS

As etapas aqui descritas foram subdivididas em seções, a fim de facilitar a visualização de cada um dos procedimentos realizados. A manipulação de todas as amostras aqui descritas foi realizada em cabine de fluxo laminar nível 2, e todos os resíduos gerados foram descontaminados antes do descarte.

4.1 Coletas

Animais da espécie *Sapajus nigritus* (n total= 33) foram capturados utilizando gaiolas Tomahawk em colaboração com o IBAMA, de acordo com seus protocolos de trabalho. As coletas deste trabalho foram realizadas de setembro de 2016 a junho de 2018, descritos na Tabela 1, apresentada nos resultados. Após 20 minutos de captura, foram observadas as condições de bem-estar animal por uma avaliação de um médico veterinário. Este procedimento é aplicado com o objetivo de minimizar o stress e dor dos animais durante a manipulação. Após observadas as condições, os animais foram submetidos a anestesia utilizando-se Quetamina (100 mg/mL) e Xilazina (20mg/mL) como descrito por Miranda, 2011 e colaboradores. Após sedação, iniciou-se as coletas das amostras biológicas. Neste estudo foram coletados suabe oral, retal, sangue e urina (quando possível). Os suabes foram coletados individualmente e estocados em tubos estéreis contendo meio mínimo essencial (MEM) suplementado com 0,5 mg/mL de Penicilina/Estreptomicina. Após, os tubos foram mantidos em caixas térmicas e transportados ao laboratório. Todos os animais se recuperaram do procedimento anestésico e foram liberados no local de captura. Estas amostras foram coletas em acordo com o Instituto Brasileiro do Meio Ambiente e Recursos Naturais (IBAMA), Centro Epidemiológico de Vigilância em Saúde (CEVS), sob a licença emitida pelo SISBIO (56540-1).

4.2 Processamento das amostras

As amostras coletadas foram processadas em *pools*. Inicialmente foram clarificadas por centrifugação a 10000 xg por 10 minutos. Após, o sobrenadante foi coletado e filtrado a 0,45 µM. Após esta etapa inicial, o filtrado

foi coletado e transferido para um tubo de ultra centrifuga contendo 5mL de solução de sacarose a 25%. Os tubos foram ultracentrifugados em centrífuga SORVAL (Combi Rotor AH-650) por 2h, a 14000 x g a 4° C. Após, os *pellets* foram ressuspensos em 1 mL de tampão composto de Tris, EDTA e NaCl (TEN). Após ressuspendido, o material foi congelado para posterior extração de ácidos nucleicos.

4.3 Extração de DNA e RNA para processamento de vírus

Uma alíquota de 500 µL resultante da etapa anterior foi tratada com RNase (100 U) e DNase (100 U) por 1 h a 37° C para eliminação de DNA e RNA exógeno que não esteja protegido por capsídeo. A extração de DNA foi realizada com Fenol e clorofórmio, conforme protocolo estabelecido por Sambrook e colaboradores. O RNA foi extraído com Trizol®, de acordo com as instruções do fabricante. Após extração de RNA, foi feita uma etapa de síntese de cDNA e amplificação do material extraído utilizando o kit WTA Qiagen conforme as instruções do fabricante e descritas a seguir. A concentração de RNA extraído foi estimada em espectrofotômetro (LQuant ®).

4.4 Enriquecimento do RNA extraído

Aproximadamente 10 ng (1,5 µL de cada amostra) foram adicionados ao tampão de lise contido no kit REPLI-g Qiagen® que foi incubado a 24°C por 5 min, seguido de incubação à 95°C por 3 minutos e armazenamento em caixa contendo gelo. Após este procedimento, adicionou-se 0,5 µL da solução nomeada *gDNAwipo out*, procedido de homogeneização em equipamento vortex, centrifugação e incubação a 42°C por 10 minutos. A etapa seguinte constitui na adição da enzima de transcriptase reversa, sua solução de reação, bem como primers randômicos. A reação foi montada conforme instruções do fabricante e incubada por 60 min a 42°C. A reação foi parada após este período, incubando-se a 95°C por 3 minutos. Após, procedeu-se a reação de ligação, com a adição da solução de reação acrescida da enzima Ligase. A última etapa deste kit de enriquecimento consiste na amplificação com a enzima sensi-phi DNA polimerase por 2 horas e 30 minutos a 30°C. Ao final desta etapa, as amostras foram visualizadas em corrida eletroforética utilizando gel de agarose 0,8%.

4.5 Enriquecimento do DNA extraído

Para amplificação dos DNA extraído foi realizada uma reação randômica utilizando a enzima phi29® DNA polimerase (Berhet et al., 2016). As amostras foram checadas e diluídas na proporção 1:4 em água ultrapura. Desta diluição foram utilizados 10µL que foram incubados a 95°C por 5 minutos e em seguida armazenados em caixa contendo gelo. Após, foram adicionados os seguintes reagentes a fim de compor o mix de reação conforme as instruções do fabricante: 0,5 µL de BSA 10X; 2,5 µL de tampão 10X; 0,5 µL de DDT (0,1 M); 1,5µL de primers randômicos exo-resistentes (500 µM); 2,5 µL de dNTPs (10 mM); 20 U da enzima phi29 DNA polimerase e H₂O ultrapura para completar um volume final de 25 µL. Os tubos foram incubados por 18 horas a 30°C, seguidos de uma etapa de inativação da enzima feita a 65°C por 10 minutos. Todas as reações foram checadas em corrida eletroforética, utilizando-se 2µL da amostra e gel de agarose 0,8%, a 60V por 45 minutos.

4.6 Purificação das amostras: AMPure ®

Após as reações de enriquecimento das amostras de DNA com phi29 e de RNA com WTA, os produtos foram submetidos à etapa de purificação com *beads* magnéticas utilizando Ampure XP. Esta etapa visou a remoção dos reagentes e enzimas (mesmo que inativadas) do material a ser sequenciado. A etapa constitui na adição de 1,8µL de *beads* vezes o volume total de amostra obtido. As amostras foram homogeneizadas e incubadas por 10 minutos à temperatura ambiente. Em seguida as amostras foram alocadas em estante magnética para que houvesse o deslocamento das *beads* em direção à placa magnética. O sobrenadante foi descartado e a purificação procedeu com uma lavagem com solução de etanol 70%. Ao final do processo os ácidos nucleicos purificados foram eluídos das *beads* em 20 µL de água ultrapura para utilização nos procedimentos descritos nas próximas etapas.

4.7 Preparo da biblioteca genômica

Para o sequenciamento, 70 ng das amostras a uma concentração de 2,5ng/uL (previamente quantificado em Qubit®) e com qualidade aferida dentro das razões de 1,8 e 2,0 medidas foram utilizados. As bibliotecas genômicas

foram preparadas com o kit Nextera XT (2x150pb) conforme descrições do fabricante. O sequenciamento foi feito em plataforma IlluminaMiSeq®. Esta etapa foi realizada por equipe treinada no Instituto de Pesquisas Veterinárias DesidérioFinamor.

4.8 Análise dos dados gerados no sequenciamento

A primeira etapa de verificação das *reads* obtidas no sequenciamento consistiu na verificação da qualidade destas através do software Fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Após esta verificação, as sequencias obtidas passaram por um processo de trimmagem utilizando *Trimmomatic*, etapa que realiza a remoção das pequenas sequencias repetidas em todas as reads (que constituem os oligonucleotídeos randômicos utilizados nas reações de amplificação). A partir destas primeiras etapas, as sequencias pré-processadas foram submetidas aos servidores Kaiju e MetaVir, que fornecem um panorama geral dos resultados.

A montagem destas pequenas sequências em sequencias consenso (contigs) foi feita utilizando o software St. Petersburg genome assembler (SPAdes) 3.10.1. Os contigs foram filtrados por tamanho (≥ 200 bp) e submetidos à ferramenta BLASTx, que fez uma busca das proteínas com identidade viral através do software Blas2GO (<https://www.blast2go.com/>). Após esta verificação inicial, todas as sequencias identificadas pelo software com identidade viral foram submetidos a uma conferência manual, incluindo-se a ferramenta BLASTn e selecionando-se a opção de banco de dados não redundante dentro do acervo disponível no NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Sequencias consenso identificadas pertencentes à mesma família foram agrupadas e submetidas ao software PRICE (<http://derisilab.ucsf.edu/software/price>), com o objetivo de aumentar as interações entre as reads e produzir sequencias consenso maiores. Finalmente, contigs representativos de genes ou genomas completos foram anotados utilizando-se o software ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>).

4.9 Construção de árvores filogenéticas

Após as etapas de verificação dos contigs finais gerados, sequências de espécies representativas das famílias virais foram obtidas em formato fasta dentro do banco de dados do GenBank (<https://www.ncbi.nlm.nih.gov/genbank>). Estas sequências foram utilizadas para os alinhamentos realizados entre as sequências geradas neste estudo e as previamente depositadas. Os alinhamentos foram realizados pela ferramenta MAFFT2 disponível dentro do Software Geneious 9.1.3 (<https://www.geneious.com/>). Os alinhamentos foram revisados manualmente e submetidos à predição dos modelos evolucionários e de substituição utilizando a ferramenta MEGA 7 (<http://www.megasoftware.net/>), conferindo robustez às análises.

5 ARTIGOS

Seção destinada a apresentação dos artigos produzidos dentro das pesquisas conduzidas neste manuscrito.

5.1 Virome profile of capuchin monkeys oral cavity by metagenomic analyses –

Publicado: <https://doi.org/10.1007/s42770-020-00350-w>

VIROME PROFILE OF CAPUCHIN MONKEYS ORAL CAVITY BY METAGENOMIC ANALYSES

Raissa Nunes dos Santos¹, Fabricio Souza Campos², Fernando Finoketti¹, Anne Caroline dos Santos¹, Aline Alves Scarpellini Campos^{1,2}, Paulo Guilherme CarnielWagner⁴, Paulo Michel Roehle¹, Helena Beatriz de Carvalho Ruthner Batista⁵, Ana Claudia Franco¹.

¹ Laboratório de Virologia – Departamento de Microbiologia, Imunologia e Parasitologia. Universidade Federal do Rio Grande do Sul. Rua Sarmento Leite, 500, Bairro Farroupilha. CEP 90050-170. Porto Alegre. Rio Grande do Sul. Brasil.

² Universidade Federal de Tocantins. Rua Badejós, lote 7, Zona Rural. CEP 77402-970. Gurupi. Tocantins. Brasil.

³ Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul. Rua Domingos Crescêncio, 132, Bairro Santana. CEP 90650-090. Porto Alegre. Rio Grande do Sul. Brasil.

⁴IBAMA – Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis. Rua Miguel Teixeira, 216, Bairro Cidade Baixa. CEP 90050-250. Porto Alegre. Rio Grande do Sul. Brasil.

⁵Instituto Pasteur – Secretaria da Saúde do Governo do estado de São Paulo. Av. Paulista, 393, BairroCerqueira César. CEP 01311-000. São Paulo. Brasil.

5.2 Phageome of oral cavity, blood and urine from healthy new world monkey *Sapajus nigritus* in South America *

Abstract

Bacteriophages (or phages) are the most abundant group of viruses. They play a dynamic role in the ecosystem as they replicate in bacterial hosts which inhabit all types of environments, including body surfaces of different species of animals. Scientific evidences suggest that prokaryote viruses affect animal health and directly modulate the structure and function of bacterial communities. This study aims at the unveiling the bacteriophage composition (phageome) in oral swabs, blood and urine of wild capuchin monkeys (*Sapajus nigritus*) collected from 12 animals in two distinct collection points. Samples were collected following the Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) guidelines, stored on ice and transported to the lab. Samples were processed in three distinct pools (swab, blood and urine), and the extraction of nucleic acids was performed using Trizol®, according to the manufacturer's instructions. Following extraction, nucleic acids were randomly amplified and submitted to Illumina platform for next generation sequencing. In total, 1.757.810 reads were *DE NOVO* assembled into >200 bp contigs using SPAdes assembler generating 3487 contigs. The consensus sequences were analyzed in MetaPhinder, a webserver from the Center for Genomic Epidemiology, based on the comparison of whole genome DNA sequences of phages. The results revealed the presence of bacteriophages belonging to the families *Siphoviridae*, *Myoviridae*, and *Podoviridae*. These phages belong to the *Caudovirales*, and about one quarter of them typically display structures to improve the adherence in mucosal surfaces, leading to the enhance of antimicrobial defense. Our findings contribute to characterize the phageome of biological samples from capuchin monkeys and corroborate with previous studies that show the phages as a major constituent of many microbiomes.

Key words: Capuchin monkeys. Phageome. Caudovirales. Virus. Bacteria.

INTRODUCTION

The exploration of microbiomes in recent years was revolutionized by high-throughput technologies, which detect a vast array of microorganisms. Most of the microbiome studies published have focused on bacterial composition (Harmsen, 2002), however it also includes viruses (Breitbart, 2003). Although viruses in biomass represent only a tiny fraction, in numbers they are comparable to cellular microbial symbionts (Shkoporov *et al.*, 2018). To be more exact, there are billions of viruses, about 10^{31} , which is more than the number of stars in the universe (Weitz and Wilhelm 2013). The majority of viruses play a role in biogeochemical cycles, infecting microbes (bacteria, archaea and microeukaryotes) which interfere with the organic matter cycling (Suttle, 2007). Among viruses, bacteriophages – viruses that infect bacteria are the most abundant biological entities at the earth.

In humans, the ratio phage/bacteria in gums is estimated to be >35:1 compared to a <10:1 for the adjacent planktonic environment, indicating that oral mucosa provides an extensive surface for phage adhesion. The highly persistent phage virions in mucus layers lead to consequences for their hosts, including a protective function against invasive bacteria (Edlund, 2015). In addition, phage genomes were detected in blood of immunosuppressed individuals, which suggests that phages are able to translocate mucosal barriers through cavities. Furthermore, mucosal sites such as oral cavity, which is composed by mouth (lips, tongue, teeth and the its roof and floor), include an extensive microbiome (Yolken, 2015). The development and acquisition of oral microbiome is directly influenced by habits, ecosystem and external perturbations.

Phages replicate through either the lytic or lysogenic cycle. In the lytic cycle the phage infects the bacteria cell and redirect its metabolism to replicate and produce new phage particles. In the lysogenic cycle the phage integrates its DNA into the bacterial genome, without causing cell lysis (lysogenic cycle) (Los, M. and Wegrzyn, G., 2012). In accordance to the host specificity, bacteriophages and their hosts live in the same ecosystems in a co-evolutionary balance. The phages force bacteria to develop defense mechanisms; on the

other side, they challenge the bacterial fitness to develop resistance mechanisms to complete their replication cycle (Golais, F., 2013)

It's already known that the composition of the oral microbiome is directly influenced by feeding habits, ecosystem and external perturbations (near contact with humans leading to unnatural feeding with chemical compounds, for example). The bacteriophage community, currently denominated phageome, was determined in a few studies including samples from human fecal samples (Manrique, 2016;). In human gut, most of the phages present are temperate, introducing new genes and alter phenotypes by the lysogenic cycle (Minot, S. et al., 2011). The analysis of the phageome is a powerful tool to study the relationship between the viruses and their bacteria host. Eventually, the population of bacteriophages has as a main role the control of infections risks, displaying a relationship with environmental stress conditions. The diversity of virus had been described in wild mammals using a next generation sequencing techniques, focused only in databases for eukaryotic viruses (Duarte, 2019). The aim of this study is to characterize the phageome of oral swabs, blood and urine collected from *Sapajus nigritus* to understand the composition of local microorganism's population.

MATERIAL AND METHODS

Study area – The present study was conducted in 2017 in Santa Cruz do Sul (29° 43' 04" S 52° 25' 33" O), and São Sebastião do Caí (29° 35' 12" S 51° 22' 32" W) municipalities of Rio Grande do Sul state, Brazil, showed in figure 1. The Figure 1 also show the fragment of Atlantic forest, characteristic of theses collection points. The Santa Cruz do Sul locality is a park, situated 2 km in distance from the center of city and São Sebastião is a mata fragment nearby to the hospital.

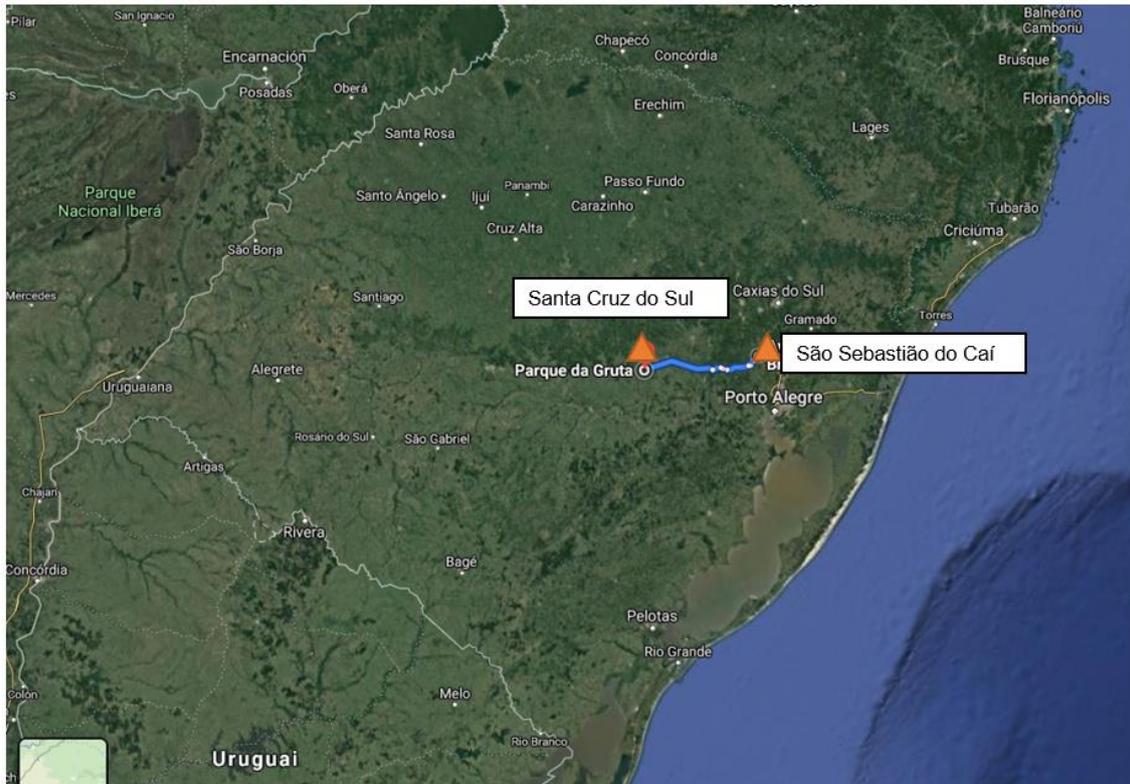


Figure 1. Map of samples' collection. The two municipalities were labeled.

Samples collection

Oral swab samples from five (5) specimens of *S. nigritus* from locality 1 and blood and urine from a total of seven (7) specimens from locality 2 were collected. These animals were captured using Tomahawk-type cages and then observed to assure animal welfare conditions. The veterinarians proceeded to anesthesia with Ketamine (100mg/mL) and Xylazine (20mg/mL) as previously described by Miranda et al. 2011. The swabs were collected individually in oral cavities and stored in sterile tubes with 2mL of Minimal Essential Medium (MEM) supplemented with 0.5 mg/mL penicillin-streptomycin (SIGMA). The tubes were kept on ice and stored at -80°C. All samples collections from this study were performed in strict accordance with the Instituto Brasileiro do Meio Ambiente e Recursos Naturais (IBAMA) and Centro Estadual de Vigilância em Saúde (CEVS) under a license emitted by SISBIO/Ministry of the Environment (License number: 6623382).

Nucleic acid extraction

Samples were gently vortexed for 2 min. Following, 1 mL of each sample was pooled into a single 15 mL conic plastic tube (one tube for urine, one for blood and one for oral swab), which was clarified by centrifugation for 5 min at 10 000 x g in a microcentrifuge. Next, the pools were filtered through a 0.45 µM Millipore syringe filter® and the filtered material was submitted to ultracentrifugation at 33 000 x g for 3 h at 4° C under a 25% sucrose cushion. The viral pellet was resuspended in ultrapure water, treated with 100 U of recombinant DNase I (Roche ®) and purelinkRNAse A (Thermo Fisher Scientific ®) and submitted to nucleic acids extraction with Trizol® and Phenol. DNA was randomly amplified with phi29 DNA polymerase 10U/µL (Thermo Fisher Scientific®) and products were used for pair-ended sequencing using the Illumina sequencing platform.

Randomly library preparation for Illumina MiSeq sequencing

Samples were gently vortexed for 2 min in their collection tubes. Following, 1 mL of each sample was clarified by centrifugation for 5 min at 12 000 x g in a microcentrifuge tube and then pooled into a single 15 mL conic tube. Next, the pool was filtered through a 0.45 µM Millipore syringe filter® and the filtered suspension was submitted to ultracentrifugation at 33 000 x g for 3 h at 4° C on a 25% sucrose cushion. The viral pellet was suspended in 500µL of ultrapure water, treated with nucleases (100 U of DNase I recombinant Roche® and PureLink 100 U of RNAse A Thermo Fisher Scientific®) and submitted to nucleic acids extraction (UltraPure Buffer-Saturated Phenol, Life Technologies®). The quality of DNA extraction was checked on a 0.7% agarose gel, and the amounts of both nucleic acids obtained were measured using Qubit®. The extracted DNA was randomly amplified with 10U/µL phi29 DNA polymerase (Thermo Fisher Scientific®). After purification, the extractions were quantified and normalized to construct the library using the Nextera XT DNA Library Prep Kit®. The sequences were generated by Illumina® MiSeq sequencer.

Bacteriophage assembly and check

Reads were *De Novo* assembled using MIRA and SPAdes assembler. The data generated were blasted using MetaPhinder and manually checked according to the default paramets. All the data are available at the National Center of Biotechnology Information (NCBI): Bioprojetc SUB6191473 entitled “Sapajus nigritus raw sequence reads” coding a Sequence Read Archive (SRA) SUB6191751.

RESULTS

We performed metagenomic sequencing of oral swabs samples to detect the phages composition of healthy adult and young capuchin monkeys (4 male and 1 female collected in site 1). From site 2, seven (7) animals adult and young (2 male and 5 female) had the phage composition analyzed by urine and blood samples. The body temperature was checked and no one presented fever. No animal had lesions or detectable signs of disease.

Sequencing data

A total of 1.757.810 reads were generated from all samples. These reads pass to a quality filter in 99% with FastQC and removing the 19 starting nucleotides of each sequence before assembly. In total, 396 of potential phages contigs were generated.

Phageome assembly and family classification from urine samples

In total, the assembly generated 920 contigs > 200 bp. After the analyses using the MetaPhinder searcher, we obtained 49 contigs with putative sequences of phages ranging in length between 515 bp to 31897 bp. Among these, 16% are still unclassified and represented unculturable or unknown phages. The majority of the sequences fall in the family *Siphovidae*(16%), followed by 8% of *Myoviridae* and 2% of *Podoviridae*. These results are represented in figure 2A.

Phageome assembly and family classification from whole blood samples

In total, we generated 835 contigs > 200 bp. After MetaPhinder searcher, we obtained 58 contigs with assumed sequences of phages, ranging in length

between 517 bp to 17230 bp. Among these 9% are still unclassified, 50% belong to *Siphoviridae*, 34% to *Myoviridae* and 7% to *Podoviridae* showed in figure 2B.

Phageome assembly and family classification from oral swab samples:

After trimming and assemblage of reads, 289 contigs from oral swab displayed amino acid identity (10-85%) to putative bacteriophages sequences. Among the total of bacteriophages identified, 17% are still unclassified according to the ICVT, whereas most of this fraction belongs to *Caudovirales* order. Figure 2C shows the percentual of each family found for each sample according to the genus described previously.

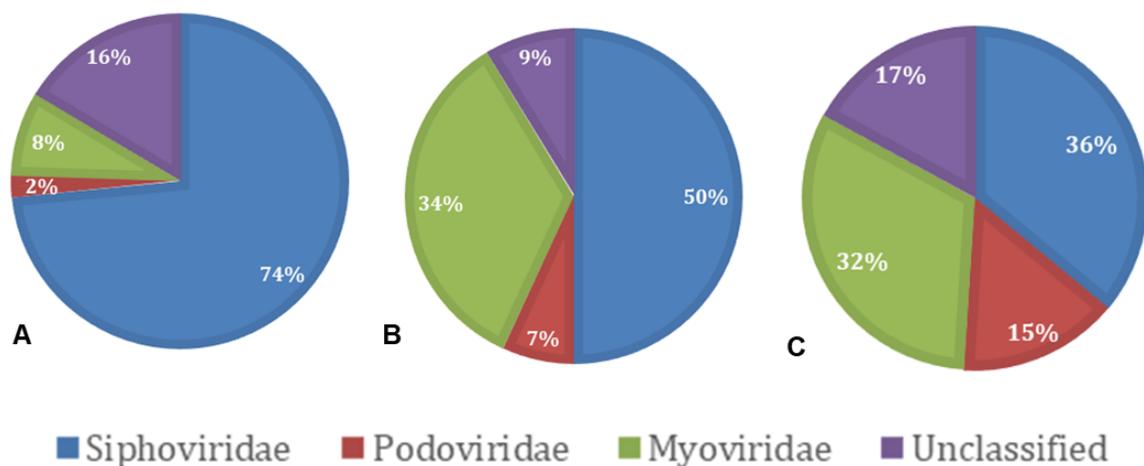


Figure 2. Distribution of bacteriophage families found in samples collected from specimens of *Sapajus nigritus*. A) bacteriophages from urine samples. B) bacteriophages from whole blood samples. C) bacteriophages from oral swabs samples.

Phageome host association:

We decided to analyze the oral swab sample, which present the highest percentages of putative phages according the host- specificity. Figures 3A, 3B and 3C show relative quantities of bacteriophage-related contigs in each host species, according to the best blast hits. The blast top hits performed by *MetaPhinder* showed that *Pseudomonas aeruginosa*, *Streptococcus*

thermophiles, *Streptococcus pneumoniae*, *Salmonella enterica*, *Klebsiella pneumoniae* and *Escherichia coli* are the main bacterial hosts for these bacteriophages. Among these species, most of them are gram negatives.

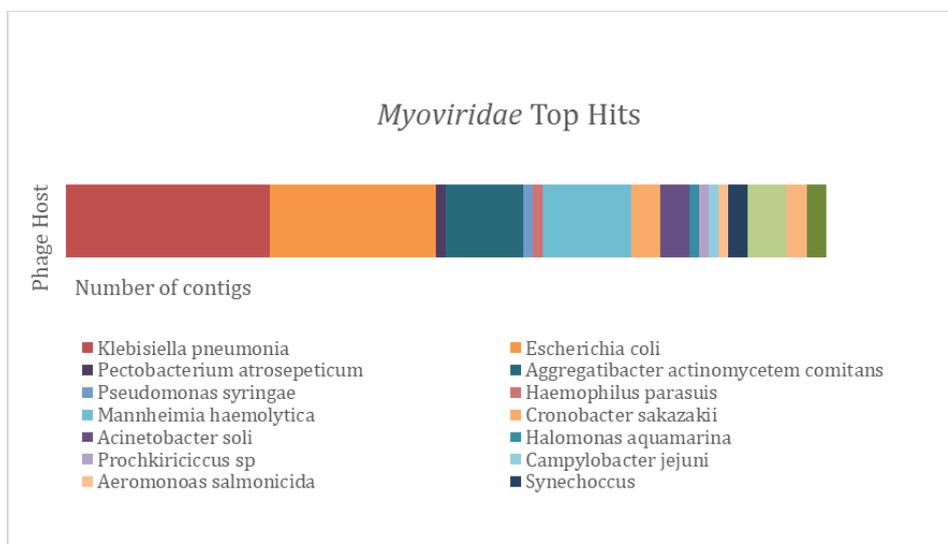
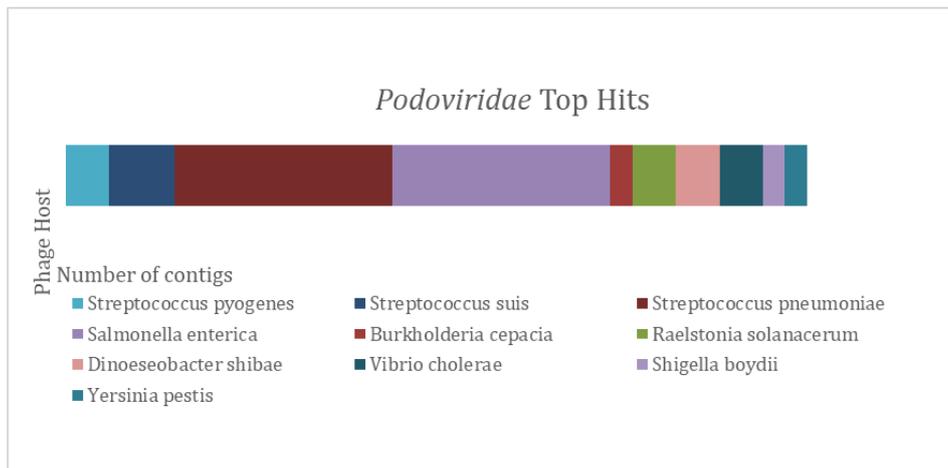
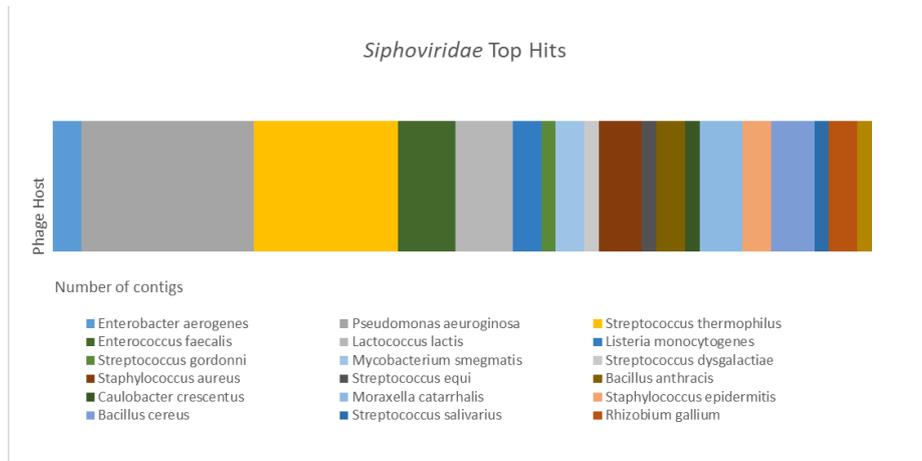


Figure 3. Host-specificity distribution among bacteriophage families found in the oral cavity of *S. nigritus*. A) *Siphoviridae* top hits; B) *Podoviridae* top hits; C) *Myoviridae* top hits.

DISCUSSION

Mucosal surfaces are a perfect portal of entry for viruses and bacteria from the environment to access new hosts. A large part of the viral community found on these surfaces are bacteriophages which replicate in the resident bacteria. The role of these viral communities remains unclear; however, they may have active roles in regulating microbial ecosystems, and thus, interfere with the health state of the host (Edlund et al., 2015; Wang et al., 2016; Baker et al., 2017; Garreto et al., 2019).

Among the sequences obtained here, three families of the *Caudovirales* were found, *Myoviridae* (generally larger dsDNA genomes with contractile tails), *Siphoviridae* (mid-sized dsDNA genomes with non-contractile tails), and *Podoviridae* (smaller dsDNA genomes with short tail stubs). The majority of the siphoviruses have primarily lysogenic lifestyles, while myoviruses and podoviruses often have lytic lifestyles (Wichels et al., 1998; Sullivan et al., 2003). Independent from their life cycle, siphoviruses were the most frequent bacteriophages identified, independent from the clinical sample (74, 50 and 36% in urine, whole blood and oral cavity samples).

Urine

Most of the bacteriophages found in the urine samples analyzed here belong to the family *Siphoviridae*, followed by members of the *Myoviridae* (8%) and *Podoviridae* (2%). To date, a few published studies have employed a metagenomic approach to sequence the viral fraction of the urinary microbiota of humans. Different results have been achieved, depending on the from a large number of unidentified sequences to the identification of a number of chlamydiamicroviruses, which infect *Chlamydia* spp. (Santiago-Rodriguez et al., 2015; Thannesberger et al., 2017; Miller-Esminger et al., 2018). Such studies confirm the existence of a still unexplored phage community within the urinary tract of humans and, probably the same is true for other species.

Blood

From the blood samples obtained, 9% of the bacteriophage-related sequences did not show significant hits with the database, 50% are siphoviruses, 34 % are myoviruses and 7%, podoviruses. These data agree with previous studies which indicates that human blood, once thought to be sterile in healthy individuals, has been found to contain both eukaryotic and prokaryotic viruses (Ghose et al., 2019).

Oral cavity

In this study, 17% of the bacteriophage-related sequences, found in the oral cavity of capuchin monkeys, did not display significant hits with the sequence database. Among the identifiable sequences, 32 % are part of the *Myoviridae*, 36 are part of the *Siphoviridae*, and 15%, to the *Podoviridae*. These data agree with previous studies which described bacteriophages as the main contributors to the oral virome, while *Siphoviridae*, *Myoviridae*, and *Podoviridae* as the most common phage families identified (Baker et al., 2017; Ghose et al., 2019). Here, we didn't identify the bacterial community of the oral cavity of these animals. However, using an *in silico* tool, we have found that *Pseudomonas aeruginosa*, *Streptococcus thermophiles*, *Streptococcus pneumoniae*, *Salmonella enterica*, *Klebsiella pneumoniae* and *Escherichia coli* are the main bacterial hosts for these bacteriophages. The *Myoviridae* and *Podoviridae* are predominantly lytic, and rapidly eliminate their hosts, thus they are important in shaping the local microbiome. Interestingly, analyses of the relative abundances of salivary phages and their respective putative hosts showed both direct and inverse relationships, indicating that mutualistic and antagonistic co-evolutionary relationships between oral phage and their bacterial hosts exist (Pride et al., 2012).

The metagenomics approach provides data based in viral analysis, while the major part of sequences, depending on the sample analyzed, and are bacteriophages-related. A recent study explored the whole virome of fecal samples obtained from wild animals in Brazil, however, bacteriophages were not targeted in these samples (Duarte, et al., 2019).

Although there are not enough data about the virome of non-human primate of the New World to compare to our data, this study corroborate with previous data described mainly in humans. We reinforce the lack of data on the microorganisms that are part of these animal's microbiota, as a way to understand the circulation of viruses and bacteria among these animals. Such data would bring light to ecological interactions among the components of their microbiota and interactions between microorganisms and their hosts, which may influence their health status.

REFERENCES

- A Duarte M, F Silva JM, R Brito C, S Teixeira D, L Melo F, M Ribeiro B, et al. Faecal Virome Analysis of Wild Animals from Brazil. *Viruses*. 2019;11(9).
- Bailey AL, Lauck M, Ghai RR, Nelson CW, Heimbruch K, Hughes AL, et al. Arteriviruses, Pegiviruses, and Lentiviruses Are Common among Wild African Monkeys. 2016 [cited 2020 May 3]; Available from: <https://github.com/hugheslab>
- Bodier-Montagutelli E, Morello E, L'Hostis G, Guillon A, Dalloneau E, Respaud R, et al. Inhaled phage therapy: a promising and challenging approach to treat bacterial respiratory infections. *Expert Opin Drug Deliv* [Internet]. 2017;14(8):959–72. Available from: <http://dx.doi.org/10.1080/17425247.2017.1252329>
- Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J Bacteriol*. 2003;185(20):6220–3.
- d'Humières C, Touchon M, Dion S, Cury J, Ghozlane A, Garcia-Garcera M, et al. A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci Rep*. 2019;9(1):25–8.
- Dalmasso M, Hill C, Ross RP. Exploiting gut bacteriophages for human health. *Trends Microbiol*. 2014;22(7):399–405.
- Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol*. 2020;12.
- Edlund A, Santiago-Rodriguez TM, Boehm TK, Pride DT. Bacteriophage and their potential roles in the human oral cavity. *J Oral Microbiol*. 2015;7:27423. Published 2015 Apr 9. doi:10.3402/jom.v7.27423

Gaetti-Jardim E, Monti LM, Nicolas Ciesielski FI, Gaetti-Jardim EC, Okamoto AC, Schweitzer CM, et al. Subgingival microbiota from Cebusapella (capuchin monkey) with different periodontal conditions. *Anaerobe* [Internet]. 2012;18(3):263–9. Available from: <http://dx.doi.org/10.1016/j.anaerobe.2012.02.002>

Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biol.* 2019;17(1):1–14.

Galan W, Bąk M, Jakubowska M. Host Taxon Predictor - A Tool for Predicting Taxon of the Host of a Newly Discovered Virus. *Sci Rep.* 2019;9(1):1–13.

Garretto A, Miller-Ensminger T, Wolfe AJ, Putonti C. Bacteriophages of the lower urinary tract. *Nat Rev Urol.* 2019;16(7):422–432. doi:10.1038/s41585-019-0192-4

Hickling DR, Sun T-T, Wu X-R. Anatomy and Physiology of the Urinary Tract: Relation to Host Defense and Microbial Infection. *Microbiol Spectr.* 2015 Aug 6;3(4).

Jonge PA De, Meijenfeldt FAB Von, Rooijen LE Van, Brouns SJJ. Evolution of BACON Domain Tandem Repeats in crAssphage and Novel Gut Bacteriophage Lineages. 2019;1–17.

Li H, Li H, Wang J, Guo L, Fan H, Zheng H, et al. The altered gut virome community in rhesus monkeys is correlated with the gut bacterial microbiome and associated metabolites. *Virol J.* 2019;16(1):1–14.

Li XF, Dong HL, Huang XY, Qiu YF, Wang HJ, Deng YQ, et al. Characterization of a 2016 Clinical Isolate of Zika Virus in Non-human Primates. *EBioMedicine* [Internet]. 2016;12:170–7. Available from: <http://dx.doi.org/10.1016/j.ebiom.2016.09.022>

Manrique P, Bolduc B, Walk ST, Van Oost J Der, De Vos WM, Young MJ. Healthy human gut phageome. *Proc Natl Acad Sci U S A.* 2016;113(37):10400–5.

Moustafa A, Xie C, Kirkness E, Biggs W, Wong E, Turpaz Y, et al. The blood DNA virome in 8,000 humans. Belshaw R, editor. *PLOS Pathog* [Internet]. 2017 Mar 22 [cited 2020 May 3];13(3):e1006292. Available from: <https://dx.plos.org/10.1371/journal.ppat.1006292>

Niedrig M, Patel P, El Wahed AA, Schädler R, Yactayo S. Find the right sample: A study on the versatility of saliva and urine samples for the diagnosis of emerging viruses. *BMC Infect Dis.* 2018;18(1):1–14.

Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med*

[Internet]. 2020;9(1). Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/31947757>

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol*. 2009;7(11):828–36.

Santiago-Rodriguez TM. Identification and quantification of DNA viral populations in human urine using next-generation sequencing approaches. *Methods Mol Biol*. 2018;1838:191–200.

Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome*. 2018;6(1):68.

Sharma S, Chatterjee S, Datta S, Prasad R, Dubey D, Prasad RK, et al. Bacteriophages and its applications: an overview. *Folia Microbiol (Praha)* [Internet]. 2017;62(1):17–55. Available from: <http://dx.doi.org/10.1007/s12223-016-0471-x>

Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *CurrProtocMol Biol*. 2018;122(1):1–15.

Virgin HW. NIH Public Access. *Cell*. 2014;23(1):142–50.

Virgin HW. The virome in mammalian physiology and disease. *Cell* [Internet]. 2014;157(1):142–50. Available from: <http://dx.doi.org/10.1016/j.cell.2014.02.032>

Wang J, Gao Y, Zhao F. Phage-bacteria interaction network in human oral microbiome. *Environ Microbiol*. 2016;18(7):2143-2158. doi:10.1111/1462-2920.12923

27. Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, et al. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome*. 2018;6(1):1–14.

28. Yolken RH, Severance EG, Sabunciyan S, Gressitt KL, Chen O, Stallings C, et al. Metagenomic sequencing indicates that the oropharyngeal phageome of individuals with schizophrenia differs from that of controls. *Schizophr Bull*. 2015;41(5):1153–61.

5.3 Resultados não apresentados em forma de artigo

Nesta seção serão apresentados resultados que estão sob análise e/ou foram utilizados como base para alguns dos artigos previamente descritos. A seguir estão descritas a relação dos grupos de animais capturados nesta pesquisa. A Tabela 1 descreve o número de animais por localidade e as amostras coletas/já processadas. Nota-se que não foi possível a coleta de urina em 3 localidades.

Tabela 1. Relação das amostras coletas, processadas ou armazenadas para futuros estudos.

Local	Nº animais	Período de coleta	Swab oral	Swab retal	Urina	Sangue
São Sebastião do Caí - RS	7	Setembro 2016	Sequenciado	Armazenado	Sequenciado	Sequenciado
Santa cruz do Sul - RS	5	Maio 2017	Sequenciado	Armazenado	Não coletado	Armazenado
Passo de Torres - RS	13	Junho 2018	Sequenciado	Armazenado	Não coletado	Armazenado
Maracajá - SC	8	Junho 2018	Sequenciado	Armazenado	Não coletado	Armazenado

5.3.1 Estabelecimento de um protocolo para recuperação de ácidos nucleicos virais

Com base nas ferramentas apresentadas na metodologia e nos conceitos abordados na revisão da literatura deste trabalho, um protocolo foi elaborado a fim de se recuperar ácidos nucleicos virais. Este protocolo consistiu na concentração de amostras e quebra de DNA oriundos do hospedeiro/ou de outros microrganismos que não de interesse deste estudo, através de uma etapa de tratamento com nucleases após o procedimento de ultracentrifugação. O protocolo completo foi descrito no material e métodos deste manuscrito. Nesta seção apresentamos uma forma esquemática de como realizamos os procedimentos descritos.

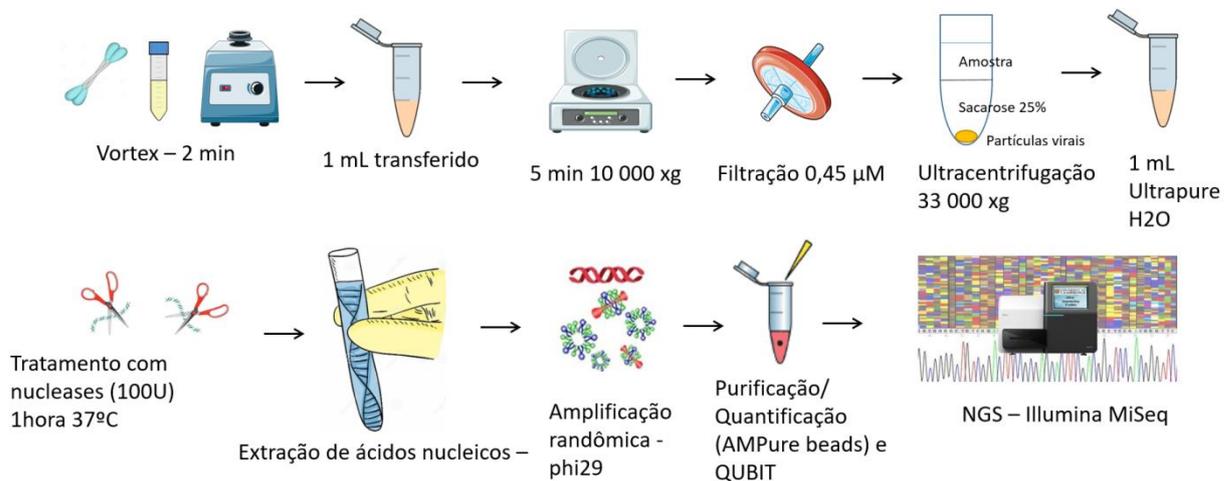


Figura 3. Fluxograma esquemático dos procedimentos para obtenção das amostras para sequenciamento de alta performance.

5.3.2 Estabelecimento de um pipeline de análise de seqüências virais obtidas através do sequenciamento de alta performance

Dando seguimento à padronização e busca de métodos adequados para a realização desta pesquisa, o presente projeto também elaborou um pipeline de análises, o qual está descrito abaixo. O pipeline desenvolvido foi denominado FRriDA, cuja sigla tem como significado “*ForcedRawreads in De Novo Assembly*”.

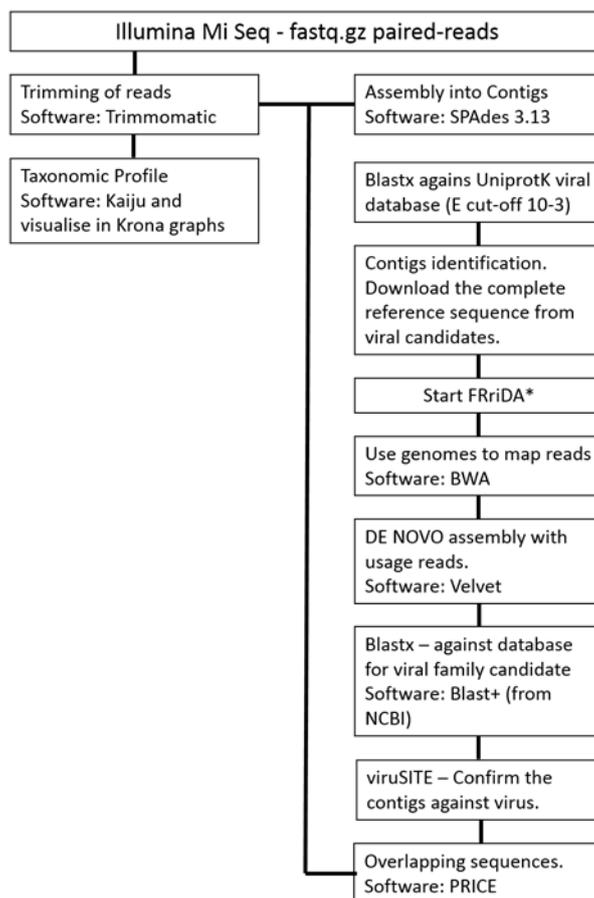


Figura 4. Fluxograma gerado para análise das amostras conforme descrito no material e métodos deste manuscrito.

5.3.3 Identificação de vírus em amostras de urina e sangue de um grupo de macacos de vida livre

Além da coleta de suabes orais, também foram coletados suabes retais, urina e sangue dos animais. A tabela 2 apresenta a quantidade de dados obtidos nesta pesquisa. As amostras de urina, sangue total da localidade de São Sebastião do Caí e a amostra de suabe oral de Santa Cruz do Sul foram utilizadas na preparação dos dois artigos aqui apresentados. As demais amostras, encontram-se em fase de análise e servirão de base para novos estudos a serem realizados pelo grupo de pesquisa do Laboratório de Virologia. Na tabela 2, podemos ver na coluna status, a situação da utilização destes dados gerados.

Tabela 2. Compilação do volume de dados de sequenciamento gerados neste estudo e seu status de análise.

Identificação	Local de coleta	Tipo de amostra	Nº de reads geradas	Status
RN1	29° 35' 12" S 51° 22' 32" W	Sangue total	635.562	Artigo 2
RN2	29° 20' 06" S 49° 43' 22" W	Suabe oral	397.434	Em análise
RN3	29° 43' 03" S 52° 25' 33" W	Suabe oral	903.430	Artigo aceito + artigo 2
RN4	29° 35' 12" S 51° 22' 32" W	Suabe oral	428.656	Artigo 2
RN5	29° 35' 12" S 51° 22' 32" W	Urina	218.818	Em análise
RN6	29° 20' 07" S 49° 43' 37" W	Suabe oral	369.904	Em análise

6 DISCUSSÃO GERAL

As análises da presença de sequências virais nas amostras em estudo permitiram a detecção tanto de genomas de RNA quanto de DNA classificados em 14 famílias, incluindo 3 famílias de bacteriófagos. As superfícies das mucosas são um portal perfeito de entrada para vírus e bactérias do ambiente para acessar novos hospedeiros. Grande parte da comunidade viral encontrada nessas superfícies são bacteriófagos que se replicam nas bactérias residentes. O papel dessas comunidades virais ainda não está claro; no entanto, podem ter papéis ativos na regulação dos ecossistemas microbianos e, assim, interferir no estado de saúde do hospedeiro (Edlund et al., 2015; Wang et al., 2016; Baker et al., 2017; Garreto et al., 2019).

Entre as seqüências obtidas aqui, três famílias de bacteriófagos foram encontradas, *Myoviridae* (genomas dsDNA geralmente maiores com caudas contraídas), *Siphoviridae* (genomas dsDNA de médio porte com caudas não contraídas) e *Podoviridae* (genomas dsDNA menores com caudas curtas). A maioria dos sifovírus tem ciclos de multiplicação principalmente lisogênicos, enquanto os miovírus e podovírus geralmente têm ciclos de multiplicação líticos (Wichels et al., 1998; Sullivan et al., 2003). Independente do ciclo de vida, os bacteriófagos foram os vírus mais frequentemente identificados neste estudo, independente da amostra clínica (74, 50 e 36% em amostras de urina, sangue

total e cavidade oral).

Os materiais e métodos utilizados nesta pesquisa foram todos avaliados de forma que se chegou a metodologia mais adequada e possível de ser realizada dentro das condições e insumos que foram disponibilizados. Entretanto, apesar de todos os esforços para obtenção sequências virais, observamos uma predominância de sequências do hospedeiro e bacterianas. Apenas 0,2% das sequências obtidas foram utilizadas para a montagem de contigs virais apresentados no artigo 1, onde apresentamos os dados de análise de um dos grupos. Após o trimming e a montagem, foram identificadas dezessete (17) sequências de consenso, relacionadas e que mostraram identidades variáveis para onze famílias virais, incluindo *Anelloviridae*, *Astroviridae*, *Parvoviridae*, *Herpesviridae*, *Genomoviridae*, *Papillomaviridae*, *Poxviridae* e *Polyomaviridae*.

Em relação as amostras analisadas para a composição de bacteriófagos, tivemos as seguintes proporções em relação as diferentes amostras, conforme a descrição a seguir. Nas amostras de urina a maioria pertenceu à família *Siphoviridae* seguida por membros do *Myoviridae* (8%) e *Podoviridae* (2%). Até o momento, alguns estudos publicados utilizaram uma abordagem metagenômica para sequenciar a fração viral da microbiota urinária dos seres humanos. Diferentes resultados foram alcançados, dependendo do grande número de sequências não identificadas até a identificação de um número de clamídiomicrovírus, que infectam a *Clamídia* spp. (Santiago-Rodriguez et al., 2015; Thannesberger et al., 2017; Miller-Esminger et al., 2018). Tais estudos confirmam a existência de uma comunidade de fagos ainda inexplorada dentro do trato urinário dos seres humanos e, provavelmente, o mesmo se reflete para outras espécies.

Já nas amostras de sangue analisadas 9 % das sequências relacionadas com bacteriófagos não apresentaram impactos significativos com o banco de dados, 50% são sifovírus, 34 % são miovírus e 7%, podovírus. Esses dados concordam com estudos anteriores que indicam que o sangue humano, antes considerado estéril em indivíduos saudáveis pode conter tanto vírus eucarióticos quanto procarióticos (Moustafa et al., 2017; Ghose et al., 2019).

Para as amostras da cavidade oral, onde a recuperação de ácidos nucleicos gerou mais dados de sequenciamento, encontramos 17% de sequências relacionadas a bacteriófagos sem identidade com sequências conhecidas no banco de dados. Entre as sequências identificáveis, 32% fazem parte da família *Myoviridae*, 36% fazem parte do *Siphoviridae*, e 15%, para o *Podoviridae*. Esses dados concordam com estudos anteriores que descreveram os bacteriófagos como os principais contribuintes para o viroma oral, tendo as famílias descritas em ambas as amostras como as mais comumente (Baker et al., 2017; Ghose et al., 2019). Curiosamente, análises das abundâncias relativas de fagos detectados na saliva e seus respectivos hospedeiros putativos mostraram relações diretas e inversas, indicando que existem relações mutualísticas e antagônicas entre a composição de fagos na cavidade oral e seus hospedeiros bacterianos (Pride et al., 2012).

Por fim, a abordagem de metagenomas mostrou-se ser uma ferramenta útil análise viral. Um estudo recente explorou todo o viroma de amostras fecais obtidas de animais silvestres no Brasil, porém, os bacteriófagos não foram alvo dessas amostras (Duarte, et al., 2019). Embora não haja dados suficientes sobre viromas de primatas não humanos do Novo Mundo para comparar com nossos dados, este estudo corrobora com dados anteriores descritos principalmente em humanos. Reforçamos a falta de dados sobre os microrganismos que fazem parte da microbiota desses animais, como forma de entender a circulação de vírus e bactérias entre esses animais. Os dados aqui apresentados para estes grupos de macacos-prego trazem indicações às interações ecológicas entre os componentes de sua microbiota e interações entre microrganismos e seus hospedeiros, o que pode influenciar seu estado de saúde, conforme descritos nos dois *papers* propostos.

7 CONCLUSÃO

O estudo aqui apresentado, caracterizou a composição viral das amostras de macacos-prego identificando vírus presentes em amostras de suabe oral, composição de bacteriófagos em amostras de suabe oral, sangue e urina; estabeleceu um protocolo para recuperação de ácidos nucleicos bem como um *pipeline* de análise.

Pode-se concluir que a hipótese inicial de encontrar vírus em animais aparentemente saudáveis foi atingida, e amostras diferentes dos grupos de macaco-prego foram analisadas.

8 REFERÊNCIAS BIBLIOGRÁFICAS

A Duarte M, F Silva JM, R Brito C, S Teixeira D, L Melo F, M Ribeiro B, et al. Faecal Virome Analysis of Wild Animals from Brazil. *Viruses*. 2019;11(9).

Bailey AL, Lauck M, Ghai RR, Nelson CW, Heimbruch K, Hughes AL, et al. Arteriviruses, Pegiviruses, and Lentiviruses Are Common among Wild African Monkeys. 2016 [cited 2020 May 3]; Available from: <https://github.com/hugheslab>

Bodier-Montagutelli E, Morello E, L'Hostis G, Guillon A, Dalloneau E, Respaud R, et al. Inhaled phage therapy: a promising and challenging approach to treat bacterial respiratory infections. *Expert Opin Drug Deliv* [Internet]. 2017;14(8):959–72. Available from: <http://dx.doi.org/10.1080/17425247.2017.1252329>

d'Humières C, Touchon M, Dion S, Cury J, Ghoulane A, Garcia-Garcera M, et al. A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci Rep*. 2019;9(1):25–8.

Dalmaso M, Hill C, Ross RP. Exploiting gut bacteriophages for human health. *Trends Microbiol*. 2014;22(7):399–405.

Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol*. 2020;12.

Gaetti-Jardim E, Monti LM, Nicolas Ciesielski FI, Gaetti-Jardim EC, Okamoto AC, Schweitzer CM, et al. Subgingival microbiota from *Cebus apella* (capuchin monkey) with different periodontal conditions. *Anaerobe* [Internet]. 2012;18(3):263–9. Available from: <http://dx.doi.org/10.1016/j.anaerobe.2012.02.002>

Gałań W, Bąk M, Jakubowska M. Host Taxon Predictor - A Tool for Predicting Taxon of the Host of a Newly Discovered Virus. *Sci Rep*. 2019;9(1):1–13.

Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biol*. 2019;17(1):1–14.

Jonge PA De, Meijenfeldt FAB Von, Rooijen LE Van, Brouns SJJ. Evolution of BACON Domain Tandem Repeats in crAssphage and Novel Gut Bacteriophage Lineages. 2019;1–17.

Li H, Li H, Wang J, Guo L, Fan H, Zheng H, et al. The altered gut virome community in rhesus monkeys is correlated with the gut bacterial microbiome and associated metabolites. *Virology*. 2019;16(1):1–14.

Li XF, Dong HL, Huang XY, Qiu YF, Wang HJ, Deng YQ, et al. Characterization of a 2016 Clinical Isolate of Zika Virus in Non-human Primates. *EBioMedicine* [Internet]. 2016;12:170–7. Available from:

<http://dx.doi.org/10.1016/j.ebiom.2016.09.022>

Manrique P, Bolduc B, Walk ST, Van Oost J Der, De Vos WM, Young MJ. Healthy human gut phageome. *Proc Natl AcadSci U S A*. 2016;113(37):10400–5.

Milton, K Primate diets and gut morphology: implications for human evolution. In *Food and Evolution: Toward a Theory of Human Food Habits*, pp. 93–116 [Harris, M and Ross, EB, editors]. 1987 Philadelphia, PA: Temple University Press.

Moustafa A, Xie C, Kirkness E, Biggs W, Wong E, Turpaz Y, et al. The blood DNA virome in 8,000 humans. Belshaw R, editor. *PLOS Pathog* [Internet]. 2017 Mar 22 [cited 2020 May 3];13(3):e1006292. Available from: <https://dx.plos.org/10.1371/journal.ppat.1006292>

Niedrig M, Patel P, El Wahed AA, Schädler R, Yactayo S. Find the right sample: A study on the versatility of saliva and urine samples for the diagnosis of emerging viruses. *BMC Infect Dis*. 2018;18(1):1–14.

Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med* [Internet]. 2020;9(1). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31947757>

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol*. 2009;7(11):828–36.

Santiago-Rodriguez TM. Identification and quantification of DNA viral populations in human urine using next-generation sequencing approaches. *Methods Mol Biol*. 2018;1838:191–200.

Sharma S, Chatterjee S, Datta S, Prasad R, Dubey D, Prasad RK, et al. Bacteriophages and its applications: an overview. *Folia Microbiol (Praha)* [Internet]. 2017;62(1):17–55. Available from: <http://dx.doi.org/10.1007/s12223-016-0471-x>

Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome*. 2018;6(1):68.

Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *CurrProtocMol Biol*. 2018;122(1):1–15.

Virgin HW. NIH Public Access. *Cell*. 2014;23(1):142–50.

Virgin HW. The virome in mammalian physiology and disease. *Cell* [Internet]. 2014;157(1):142–50. Available from: <http://dx.doi.org/10.1016/j.cell.2014.02.032>

Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, et al. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome*. 2018;6(1):1–14.

Yolken RH, Severance EG, Sabunciyar S, Gressitt KL, Chen O, Stallings C, et al. Metagenomic sequencing indicates that the oropharyngeal phageome of individuals with schizophrenia differs from that of controls. *Schizophr Bull*. 2015;41(5):1153–61.