

**CARACTERIZAÇÃO DO VIROMA FECAL DE *Didelphis aurita* EM
PAISAGENS FLORESTAIS E URBANAS DO ESTADO DO RIO DE
JANEIRO**

MARIA EDUARDA SOARES LEITE RANGEL

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO (UENF)**

CAMPOS DOS GOYTACAZES - RJ

JULHO/2024

**CARACTERIZAÇÃO DO VIROMA FECAL DE *Didelphis aurita* EM
PAISAGENS FLORESTAIS E URBANAS DO ESTADO DO RIO DE
JANEIRO**

MARIA EDUARDA SOARES LEITE RANGEL

Dissertação apresentada ao Centro de Biociências e Biotecnologia, no Programa de Pós-Graduação em Ecologia e Recurso Naturais da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Ecologia e Recursos Naturais.

Orientadora: Dra. Caryne A. de Carvalho Braga
Coorientadora: Dra. Mirela D'arc Ferreira da Costa

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO (UENF)

CAMPOS DOS GOYTACAZES - RJ

JULHO/2024

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pela autora.

R196

Rangel, Maria Eduarda Soares Leite.

CARACTERIZAÇÃO DO VIROMA FECAL DE *Didelphis aurita* EM PAISAGENS FLORESTAIS E URBANAS DO ESTADO DO RIO DE JANEIRO / Maria Eduarda Soares Leite Rangel. - Campos dos Goytacazes, RJ, 2024.

95 f. : il.

Bibliografia: 64 - 73.

Dissertação (Mestrado em Ecologia e Recursos Naturais) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2024.

Orientadora: Caryne Aparecida de Carvalho Braga.

Coorientadora: Mirela D'Arc Ferreira da Costa.

1. Didelphidae. 2. Diversidade Viral. 3. Ecologia de Comunidades. 4. Marsupiais Didelfídeos. 5. Vírus. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 577

CARACTERIZAÇÃO DO VIROMA FECAL DE *Didelphis aurita* EM PAISAGENS FLORESTAIS E URBANAS DO ESTADO DO RIO DE JANEIRO

MARIA EDUARDA SOARES LEITE RANGEL

Dissertação apresentada ao Centro de Biociências e Biotecnologia, no Programa de Pós-Graduação em Ecologia e Recurso Naturais da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Ecologia e Recursos Naturais.

Aprovada em 10/07/2024

Comissão examinadora:

Documento assinado digitalmente
 GISELE REGINA WINCK
Data: 15/08/2024 12:41:59-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Gisele Regina Winck – FIOCRUZ

Documento assinado digitalmente
 MARINA GALVAO BUENO
Data: 13/08/2024 13:10:33-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Marina Galvão Bueno – FIOCRUZ

Documento assinado digitalmente
 PAULA RIBEIRO PRIST
Data: 13/08/2024 14:52:01-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Paula Ribeiro Prist - EcoHealth Alliance

Documento assinado digitalmente
 CARYNE APARECIDA DE CARVALHO BRAGA
Data: 11/07/2024 11:26:22-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Caryne Aparecida de Carvalho Braga – UENF (Orientadora)

Documento assinado digitalmente
 MIRELA D ARC FERREIRA DA COSTA
Data: 11/07/2024 10:46:08-0300
Verifique em <https://validar.iti.gov.br>

Dr^a Mirela D'arc Ferreira da Costa – UFRJ (Coorientadora)

“Só há duas opções nesta vida: se resignar ou se indignar. E eu não vou me resignar nunca.”

(Darcy Ribeiro)

AGRADECIMENTOS

Aos meus avós maternos e à minha mãe, que até hoje não conseguem entender com clareza o que é fazer um mestrado, mas seguem orgulhosos das minhas conquistas.

Ao Henrique Simfrone, também conhecido como meu namorado, que me encontrou quando eu já estava na metade dessa jornada, mas não deixou de me apoiar e escutar todos os meus lamentos.

À minha orientadora, Dr^a Caryne Braga que, para minha surpresa, aceitou me orientar e esteve sempre disposta a me ajudar, apoiar e incentivar a buscar me aperfeiçoar cada vez mais.

À minha coorientadora, Dr^a Mirela D'arc, que mesmo a mais de 200Km de distância sempre se fez presente, nunca hesitou em me ajudar e, assim como a Caryne, sempre tolerou o meu "jeitinho" particular de perguntar absolutamente tudo.

Aos meus colegas do laboratório *Vertebrate Ecology Lab* (VertEco) que sempre estiveram comigo nos campos e foram muitas vezes corajosos o suficiente para triar as amostras de fezes quando eu não estava presente. Especialmente deixo meu agradecimento para Amanda Braga, Victoria Bartolome e Larissa Terra tanto pela presença na minha vida acadêmica quanto pessoal.

Aos meus colegas do Laboratório de Diversidade e Doenças Virais (LDDV), da UFRJ, que me acolheram como parte do laboratório. Especialmente a Amanda Coimbra e o Matheus Cosentino que me ajudaram bastante nessa jornada pela virologia.

À banca que, com certeza, irá fazer sugestões que irão aprimorar esse trabalho e muito gentilmente se dispuseram a participar desse processo.

À CAPES e a FAPERJ, agências de fomento que forneceram a bolsa e os financiamentos necessários para que esse projeto ocorresse.

Ao Programa de Pós-graduação em Ecologia e Recursos Naturais (PPG-ERN/UENF) e seu corpo docente pelo acolhimento e conhecimento compartilhados.

A todos que participaram, direta e indiretamente, desse processo deixo aqui o meu muito obrigada!

SUMÁRIO

RESUMO.....	xi
ABSTRACT.....	xii
1. INTRODUÇÃO	1
2. OBJETIVOS	6
2.1 OBJETIVO GERAL.....	6
2.2 OBJETIVOS ESPECÍFICOS	6
3. HIPÓTESES	7
4. MATERIAL E MÉTODOS	7
4.1 ÁREAS DE ESTUDO	7
4.2 DELINEAMENTO AMOSTRAL.....	11
4.2.1 Agrupamento das Amostras.....	17
4.3 VIROMA	18
4.3.1 Enriquecimento Viral.....	19
4.3.2 Extração dos Ácidos Nucleicos Virais	20
4.3.3 Síntese da Dupla Fita de DNAc e Purificação	21
4.3.4 Confeção e Purificação das Bibliotecas.....	23
4.3.5 Avaliação da Biblioteca	24
4.3.6 Plataforma Illumina - MiSeq	27
4.4 ANÁLISES GENÉTICAS	29
4.5 ANÁLISES DOS DADOS.....	31
5. RESULTADOS	32
5.1 Riqueza e composição de famílias virais das áreas	34
5.2 Riqueza e composição de famílias virais entre os sexos.....	41
5.3 Detalhamento das bibliotecas separadas por sexo.....	43
5.4 Composição de famílias e tipos virais entre áreas.....	44
5.5 Composição de famílias e tipos virais entre sexos.....	46
5.6 Detalhamento dos tipos virais das bibliotecas	48
5.7 Jaccard e Similaridade das Bibliotecas.....	51
6. DISCUSSÃO	56
7. CONCLUSÃO.....	64
REFERÊNCIAS BIBLIOGRÁFICAS.....	64
GLOSSÁRIO	74
APÊNDICES.....	75
APÊNDICE A – Famílias virais identificadas nesse trabalho, com seus hospedeiros de acordo com o determinado pelo <i>site ViralZone</i> (Hospedeiro) e o hospedeiro padrão (Hosp. padrão) usado para construção dos mapas de calor.....	76
APÊNDICE B – Número de leituras (em milhões - M) que foram classificadas e não classificadas (“N classificadas”) pelas ferramentas <i>Diamond</i> e <i>Kraken2</i> , em cada uma das bibliotecas analisadas: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes.....	78

APÊNDICE C – Número de leituras (em milhões) atribuídas a cada grande domínio taxonômico pelas ferramentas *Diamond* e *Kraken2*, em cada uma das bibliotecas analisadas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes)..... 79

APÊNDICE D – Tabela do número de *contigs* identificados pelo *Kraken2*, para todas as bibliotecas analisadas: UA_F – Fêmeas da União A; U_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. O hospedeiro que cada família infecta pode ser encontrado no Apêndice A. 80

LISTA DE FIGURAS

Figura 1 - Representação das áreas de coleta de amostras de fezes da espécie <i>Didelphis aurita</i> na Reserva Biológica União, entre 2019 e 2023.	8
Figura 2 - Representação das áreas de coleta de amostras de fezes da espécie <i>Didelphis aurita</i> na APA da Bacia do Rio São João/ Mico-Leão-Dourado, entre 2019 e 2023.	9
Figura 3 - Representação das áreas de coleta de amostras de fezes da espécie <i>Didelphis aurita</i> em Campos dos Goytacazes/RJ, entre 2022 e 2023.	10
Figura 4 - Localização de todas as áreas de captura de <i>D. aurita</i> (ReBIO União; APA da Bacia do Rio São João e Campos dos Goytacazes) dentro do estado do Rio de Janeiro. ...	10
Figura 5 - Desenho amostral utilizado nas coletas na Reserva Biológica União e na APA Bacia do Rio São João, representando os transectos e cada círculo laranja representa uma estação amostral. As armadilhas utilizadas foram Tomahawk® (A), Sherman® (B) e de queda (C). Os círculos azuis representam as armadilhas de queda.	12
Figura 6 - Trecho da área de coleta ReBIO União A, Silva Jardim (RJ).	13
Figura 7 - Trecho da área ReBIO União B, Silva Jardim (RJ).	14
Figura 8 - Ponto de coleta no Horto Municipal no município de Campos dos Goytacazes (RJ).	16
Figura 9 - Resumo dos protocolos de bancada e bioinformática, estabelecidos pelo Laboratório de Diversidade e Doenças Virais (UFRJ), utilizados para análise das amostras de fezes de <i>D. aurita</i>	19
Figura 10 - Curva de rarefação baseada em indivíduos para bibliotecas analisadas: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. A linha contínua indica os valores obtidos enquanto a linha tracejada valores extrapolados, demonstrando o número de leituras virais necessário para atingir a assíntota naquelas bibliotecas.	34
Figura 11 - Riqueza de famílias virais compartilhadas e exclusivas entre as áreas analisadas. Na diagonal, em verde, temos o número de famílias virais exclusivas de cada área. Na parte superior direita da matriz temos o número de famílias virais compartilhadas entre cada par de área analisado.	39
Figura 12 - Riqueza de famílias virais compartilhadas e exclusivas entre as áreas analisadas, com as famílias encontradas em ambas as ferramentas (Kraken2 e Diamond). Na diagonal, em verde, temos o número de famílias virais exclusivas de cada área. Na parte superior direita da matriz temos o número de famílias virais compartilhadas entre cada par de área analisado.	41
Figura 13 – Quantidade de famílias virais exclusivas de cada sexo, assim como as compartilhadas. Na esquerda temos a quantidade de famílias exclusivas das fêmeas e na direita as exclusivas dos machos. A intercessão demonstra a quantidade de famílias compartilhadas entre os dois sexos.	42
Figura 14 - Quantidade de famílias virais exclusivas de cada sexo, assim como as compartilhadas. Na esquerda temos a quantidade de famílias exclusivas das fêmeas e na direita as exclusivas dos machos. A intercessão demonstra a quantidade de famílias compartilhadas entre os dois sexos. Aqui usamos os dados referentes as famílias virais confirmadas pelo Diamond.	43
Figura 15 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada área estudada. $\log_{10}(\text{Leituras})$ representa o número de leituras logaritimizado na base 10.	45
Figura 16 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada área, com os dados das famílias encontradas em ambas as ferramentas (Kraken2 e Diamond). $\log_{10}(\text{Leituras})$ representa o número de leituras logaritimizado na base 10.	46
Figura 17 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada sexo. $\log_{10}(\text{Leituras})$ representa o número de leituras logaritimizado na base 10.	47

Figura 18 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada sexo, com os dados das famílias encontradas em ambas as ferramentas (Kraken2 e Diamond). Log10(Leituras) representa o número de leituras logaritimizado na base 10.	48
Figura 19 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada biblioteca: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. Log10(Leituras) representa o número de leituras logaritimizado na base 10.	49
Figura 20 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada biblioteca: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes, considerando as famílias confirmadas no <i>Diamond</i> . Log10(Leituras) representa o número de leituras logaritimizado na base 10.	50
Figura 21 - Dendrograma feito a partir do índice de dissimilaridade de Jaccard entre as bibliotecas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes), baseado na presença e ausência de famílias virais em cada uma das bibliotecas.	51
Figura 22 - Dendrograma feito a partir do índice de dissimilaridade de Jaccard entre as bibliotecas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes), baseado na presença e ausência de famílias virais em cada uma das bibliotecas, usando os dados das famílias encontradas em ambas as ferramentas (Kraken2 e Diamond).	52
Figura 23 - Matriz de presença e ausência de famílias virais identificadas pela ferramenta Kraken2 ordenada por média recíproca na análise de elementos de estrutura de metacomunidades. Legenda: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes.	55
Figura 24 - Matriz de presença e ausência de famílias virais identificadas pelas ferramentas <i>Kraken2</i> e <i>Diamond</i> ordenada por média recíproca na análise de elementos de estrutura de metacomunidades. Legenda: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes).	56

LISTA DE TABELAS

Tabela 1 - Esforço amostral entre todas as áreas amostradas. Armadilhas/noite se refere a quantidade de armadilhas multiplicado pelo número de campanhas e noites de armadilha aberta.	12
Tabela 2 - Número de amostras utilizadas em cada conjunto e as siglas correspondentes.	17
Tabela 3 - Resumo dos resultados obtidos no <i>Kraken2</i> , para todas as bibliotecas analisadas. Entre parênteses está a porcentagem que as leituras virais representam dentro do total de leituras obtidos para cada biblioteca. Os resultados “pós confirmação” se referem àquelas famílias virais que foram identificadas pelo <i>Kraken2</i> que também estavam presentes no <i>Diamond</i>	33
Tabela 4 - Famílias virais identificadas pelo <i>Kraken2</i> e número de leituras atribuídas a cada uma nas bibliotecas analisadas (UA_F - Fêmeas da ReBIO União A; UA_M - Machos da ReBIO União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM – APA; CG_F – Fêmeas de Campos dos Goytacazes; CG_M – Machos de Campos dos Goytacazes). A coluna “ <i>Diamond</i> ” indica se aquela família viral também foi identificada (S) ou não (N) na ferramenta <i>Diamond</i> . Os hospedeiros de cada família viral podem ser encontrados no Apêndice A. Marcadas com * estão as famílias virais conhecidas por terem vírus que infectam de vertebrados.	36
Tabela 5 - Análise da estrutura de metacomunidades baseada na distribuição das famílias virais entre as bibliotecas de indivíduos de <i>Didelphis aurita</i> . “Aus.” se refere as ausências embutidas, “Sub.” a quantidade de substituição de espécies e “ <i>I</i> ” é o índice de Morisita. A média se refere ao valor médio obtido nas matrizes simuladas com os dados.	53

RESUMO

Viroma é o termo utilizado para representar praticamente toda a diversidade genética viral sequenciada e presente em um determinado organismo ou táxon de interesse. Por meio dessa técnica, são identificadas famílias virais patogênicas e não-patogênicas, e pode ser utilizada como ferramenta de monitoramento de risco zoonótico. O viroma também pode ser avaliado quanto aos fatores responsáveis por sua modulação, como a antropização do ambiente e sexo dos indivíduos, a fim de entendermos quais as possíveis fontes de variação na diversidade viral. Neste caso, é importante utilizar como modelo de estudo do viroma animais que são capazes de habitar áreas urbanas, ao mesmo passo que habitam áreas florestadas. Sendo assim, o gambá (*Didelphis aurita*) é um bom candidato para este tipo de estudo, um marsupial neotropical da família Didelphidae, com ocorrência na Mata Atlântica e que pode ser encontrado em fragmentos florestais preservados e com diferentes graus de perturbação, assim como em ambientes urbanos. O objetivo principal desse trabalho foi caracterizar a diversidade viral presente em *D. aurita* em três locais com diferentes graus de perturbação. Foram coletadas 62 amostras de fezes de animais capturados em armadilhas do tipo *Tomahawk*[®], *Sherman*[®] e de queda. As amostras foram agrupadas em sete grupos: quatro da ReBIO União, um da APA e dois de Campos dos Goytacazes. Os grupos passaram pelo sequenciamento de alta performance e a identificação taxonômica foi feita até o nível de família viral por duas ferramentas: *Kraken2* e *Diamond*. Com os resultados, identificamos a riqueza de famílias virais em cada conjunto e comparamos a composição entre esses conjuntos por meio do índice de dissimilaridade de Jaccard. Nós usamos os dados obtidos apenas pelo *Kraken2* e informações sobre as famílias que foram identificadas em ambas as ferramentas, para contornar possíveis falsos positivos do *Kraken2*. No geral, obtivemos uma riqueza total de 50 famílias virais, 23 dessas sendo confirmadas pelo *Diamond*. Em sua maioria essas eram famílias de bacteriófagos, algo usual em estudos do trato gastrointestinal, por ser uma área do corpo repleta de bactérias e vírus que as infectam. Por fim, ao analisar a composição identificada pela ferramenta *Kraken2*, o sexo do hospedeiro foi um fator mais importante para a composição do viroma do que a antropização da área. Essa relação já havia sido observada em seres humanos e animais de laboratório, mas é a primeira vez que esse resultado é observado em animais silvestres.

ABSTRACT

Virome is the word we use to represent virtually all the genetic viral diversity within a certain organism or taxon. Through this technique all viral families, pathogenic or not, are identified and we can use it as a tool to monitor zoonotic risk. The virome can also be studied in the sense to find which factors are responsible for its variation, such as the amount of anthropic intervention in the area or the sex of the individuals sampled, as a mean to understand which factors contribute to changes in viral diversity. It is important to choose, as a study model, animals that can inhabit both urban and forested areas. The opossum (*Didelphis aurita*), a neotropical marsupial from the family Didelphidae found in the Atlantic Forest, can be considered a good candidate for such studies because is found in intact and with different degrees of disturbance forest fragments and in urban environments. Our main goal was to characterize the gut viral diversity of *D. aurita* in three places with different degrees of disturbance. We obtained 62 fecal samples of animals captured in Tomahawk®, Sherman® and pitfall traps. Samples were pooled into seven different groups: four from *ReBIO União*, one from *APA*, and two from *Campos dos Goytacazes*. The pools went through High Throughput Sequencing and taxonomic identification, up until the viral family level, using both Kraken2 and Diamond. With the results we identified viral family richness in each pool and compared community composition using Jaccard's dissimilarity index. We used the data obtained only by Kraken2 and we also used data referring to the viral families that were identified in both tools to minimize false positives by Kraken2. We got a total richness of 50 viral families with 23 of those being confirmed by Diamond. They were, mostly, families of bacteriophages something usual in studies of the gut, an area of the body with plenty of bacteria and viruses that can infect them. Lastly, with community composition found by Kraken2, we saw that sex was more important for viral community structure than anthropic intervention in the area. Such relationship has been seen before in humans and lab animals, but this is the first time this finding is observed in wild animals.

1. INTRODUÇÃO

Vírus são partículas microscópicas formadas por um capsídeo proteico que protege um segmento de material genético (RNA e/ou DNA) em seu interior. Sua característica principal é que são parasitas celulares obrigatórios, por isso necessariamente fazem uso de um hospedeiro para se reproduzir (Manrubia e Lázaro, 2006; Segre, 2023). São considerados onipresentes no ambiente e são capazes de infectar diversas formas de vida, desde bactérias até animais. Também são considerados a forma de vida mais abundante do planeta (McCaughey, 2010). Porém, uma infecção viral não significa, de maneira estrita, uma doença, havendo vírus que compõem a microbiota natural de um organismo e vírus considerados patogênicos, que causam problemas de saúde (perturbam a homeostase) do indivíduo que infectam (Howe *et al.*, 2016; Lynch, 2014; Lysholm *et al.*, 2012).

Conhecer os vírus e suas particularidades é importante em questões evolutivas e de saúde. Ao estudar o padrão de coevolução dos vírus e de seus hospedeiros podemos traçar o seu local e período de surgimento, bem como trazer informações relevantes para a história evolutiva desses organismos junto de seus hospedeiros (D'arc *et al.*, 2020; Forterre e Prangishvili, 2009; King *et al.*, 2022; Lauber *et al.*, 2017). Um exemplo é o caso do Vírus da Imunodeficiência Humana tipo 1 (HIV-1, do inglês *Human Immunodeficiency Virus type 1*). Sabemos que a origem de dois dos seus quatro grupos veio do Vírus da Imunodeficiência Símia de Gorila (SIVgor, do inglês *Gorilla Simian Immunodeficiency Virus*) (D'arc *et al.*, 2015). O vírus HIV-1 já era conhecido por ter quebrado a barreira entre espécies e a identificação dessas origens símias permitiram que sua origem em humanos fosse traçada, demonstrando que o vírus foi capaz de saltar entre um hospedeiro e outro (D'arc *et al.*, 2017).

É importante ressaltar que o mundo vem sofrendo com diversas endemias e pandemias causadas por vírus, como no caso do SARS-CoV-2 (Platto *et al.*, 2021), Ebola (Saéz *et al.*, 2015) e M-Pox (Kaler *et al.*, 2022). Todas essas infecções têm em comum o fato de que foram originadas a partir de um transbordamento zoonótico. O transbordamento zoonótico ocorre quando um patógeno que afeta um determinado hospedeiro animal entra em contato, se estabelece e passa a se replicar e a infectar ativamente um novo hospedeiro animal (Honce e Schultz-Cherry,

2020). O registro de determinado vírus patogênico em uma espécie auxilia na avaliação do risco de transbordamento entre esse animal e os seres humanos, permitindo desenvolver medidas de prevenção e monitoramento. Essas estratégias são importantes para prevenir a ocorrência de surtos de doenças virais e para mitigar os efeitos caso um surto ocorra (Honce e Schultz-Cherry, 2020; Morse *et al.*, 2012).

He e colaboradores (2022) identificaram, por exemplo, a presença de um subtipo de Influenza A em animais exóticos comercializados como alimento na China. O subtipo em questão é conhecido por ter origem em aves domésticas e é capaz de infectar seres humanos (Bi *et al.*, 2020). Bi e colaboradores (2020) monitoraram o vírus e suas variantes durante três anos. Nesse período, conseguiram monitorar a prevalência de subtipos e compreender quais têm maior capacidade de infectar o ser humano (Bi *et al.*, 2020). Sendo assim, fica evidente a importância de monitorar os vírus a fim de prever possíveis fontes de endemias e pandemias.

Um exemplo de transbordamento viral epizootico é o caso do Papillomavirus que causa tumores em leões (*Panthera leo*) cativos. O transbordamento ocorreu a partir do consumo de carne bovina e a espécie de bovino consumida é um reservatório do vírus em questão (Munday e Knight, 2010; Orbell *et al.*, 2011). Algo similar ocorreu com leões asiáticos (*Panthera leo persica*) em um santuário na Índia. Neste caso, indivíduos da população começaram a morrer subitamente, infectados pelo Vírus da Cinomose Canina, transmitido por cães de rua (Mourya *et al.*, 2019). Esses dois exemplos destacam a influência da ação antrópica como facilitadora da transmissão dos vírus entre animais domésticos e silvestres.

O transbordamento de patógenos entre humanos, animais domésticos e animais silvestres se torna ainda mais crítico considerando a situação atual do planeta, onde a ação do homem tem aumentado o contato entre esses reservatórios, ameaçando a sobrevivência de todas as partes e especialmente de populações de animais silvestres (Carlson *et al.* 2022). Pode-se pensar no contexto das mudanças climáticas, onde a modificação das condições ambientais forçará as espécies a modificarem sua distribuição, de modo que o contato entre animais, antes isolados, será facilitado (Carlson *et al.* 2022). De forma mais direta, o contexto de fragmentação de ambientes naturais aumenta a quantidade de bordas nas paisagens e, portanto, aumenta a zona de contato entre os animais silvestres e os

domésticos e população humana residente no entorno. Albery e colaboradores (2020) determinam que o compartilhamento viral entre mamíferos depende tanto da proximidade filogenética quanto geográfica. Ou seja, além de variações espécie-específicas, quanto mais permitimos o contato entre diversas fontes, maiores serão as chances de determinados vírus ampliarem o número de hospedeiros que normalmente infectam (Albery *et al.*, 2020; Brooks, Boeger e Hoberg, 2022).

A descoberta desses transbordamentos zoonóticos, assim como o desenvolvimento de estudos relacionados à vírus e diversidade viral só foi possível a partir dos avanços das tecnologias de análise genética, como a metagenômica, que busca sequenciar teoricamente todo o material genético de interesse em uma amostra. As técnicas de metagenômica passaram por diversas fases, avançando até o que chamamos hoje de sequenciamento de alta performance (do inglês *High Throughput Sequencing*) (Breitwieser *et al.*, 2019; Reuter *et al.*, 2015; Roumpeka *et al.*, 2017). Esse tipo de sequenciamento permite que uma grande quantidade de leituras (do inglês *reads*), mesmo que de fragmentos pequenos, sejam obtidas simultaneamente (Metzker, 2010; Slatko *et al.*, 2018).

O avanço dessas tecnologias permitiu um aumento no conhecimento de espécies virais, e da diversidade viral como um todo, principalmente por meio dos estudos dos viomas de diferentes espécies. O viroma é o conjunto de material genético de todos os vírus presentes na amostra obtida de um indivíduo ou população. Seu estudo é útil tanto para conhecer a diversidade viral, quanto para a descoberta de novos vírus (patogênicos ou não) que circulam em uma população, bem como para prever o risco zoonótico de um determinado vírus emergente (Harvey e Holmes, 2022; Lu *et al.*, 2022). Logo, o viroma também pode sinalizar se uma espécie deve ser monitorada de maneira contínua, devido ao risco zoonótico (Harvey e Holmes, 2022).

O viroma não é algo único (Zuo *et al.*, 2020), podendo sofrer influência de diversos fatores, como o local em que o hospedeiro está inserido. O grau de intervenção antrópica, no que se trata do local amostrado, pode ser um dos causadores de variação. Para populações de diversas espécies de morcego na Croácia ambientes afetados pela atividade humana apresentaram riqueza mais baixa, quando comparados com ambientes tidos como naturais pelos autores (Šimić *et al.*, 2020). Ou seja, o quanto uma área é urbanizada pode se refletir no viroma dos animais nela encontrados. Os autores também utilizaram diferentes tipos de amostra

(fezes, guano e *swab* oral) para estudar o viroma e obtiveram valores diferentes de riqueza e número de sequências indicando que outros fatores, como o tipo de amostra, também podem ser responsáveis por variação (Šimić *et al.*, 2020). Neste caso, a maior diversidade de famílias virais foi encontrada nas amostras de guano (um tipo de material fecal) dos morcegos (Šimić *et al.*, 2020). Outro fator reconhecido de variação na diversidade viral é o sexo do hospedeiro. Feng e colaboradores (2022) demonstraram que, em mosquitos da espécie *Culex pipens*, machos e fêmeas compartilham a maioria dos componentes do viroma, mas também preservam componentes exclusivos de cada sexo. O alto compartilhamento de vírus foi associado à época de coleta, coincidente com altas temperaturas e com a época reprodutiva, de modo que o contato entre machos e fêmeas era maior, aumentando compartilhamento viral (Feng *et al.*, 2022).

Em relação a grupos com potencial relevante para monitoramento e descrição da diversidade viral podemos destacar os marsupiais (Metatheria), devido seu potencial de contribuir no entendimento da evolução dos vírus em mamíferos. Os metatérios divergiram dos eutérios há cerca de 125 milhões de anos (Luo *et al.*, 2003). Assim, o estudo do viroma desses animais pode ajudar a entender melhor a evolução dos vírus em mamíferos. No entanto, os estudos sobre viroma de marsupiais ainda são raros, e restritos a marsupiais australianos (Chong *et al.*, 2019; Dunowska *et al.*, 2022), excluindo uma parcela importante da diversidade, composta por marsupiais neotropicais. Os marsupiais neotropicais ocorrem desde o sul da América do Sul até a região central da América do Norte, e no Brasil são representados pela família Didelphidae, que compõe mais de 90% das espécies neotropicais (Voss e Jansa, 2021).

Embora não existam ainda estudos sobre viroma de marsupiais didelfídeos, existem estudos sobre vírus específicos, como os registros da presença de *Anelovirus* (de Souza *et al.*, 2018) e *Vaccinia* (Peres *et al.*, 2016) em *Didelphis albiventris*. Outros estudos incluem relatos de arbovírus em *D. marsupialis* e *Philander opossum* (Bernal *et al.*, 2021) e do Vírus do Nilo Ocidental em *D. virginiana* (Lamglait e Lair, 2019). Alguns trabalhos identificaram anticorpos contra o Vírus da Raiva em *D. aurita* (Araujo *et al.*, 2014), *D. albiventris*, *D. marsupialis* e *Lutreolina crassicaudata* (Almeida *et al.*, 2001), mas no geral não houve detecção de infecção ativa pelo vírus. Apenas recentemente tivemos um relato de *D. albiventris* infectado pelo vírus da raiva (Ferreira-Machado *et al.*, 2023). O gênero *Didelphis* é

particularmente interessante para esse tipo de estudo por ser composto de espécies que têm sucesso em ocupar ambientes antropizados e naturais, mantendo contato próximo com seres humanos e animais domésticos, também podendo carregar vírus de uma área para a outra (de Castro *et al.*, 2017). Além disso, muitas vezes são também utilizados como fonte de alimento pelos humanos, aumentando o risco de servirem como fonte/reservatório zoonótico (Winck *et al.*, 2022).

Como marsupial didelfídeo que necessita de atenção nos estudos de viroma e monitoramento zoonótico podemos destacar a espécie *Didelphis aurita*, conhecido na literatura científica como gambá-de-orelhas-pretas. Esse animal é ampla e predominantemente distribuído pela Mata Atlântica, podendo ser encontrado em todos os estados da região sul do Brasil; parte do Mato Grosso do Sul; todos os estados do sudeste; e, parte dos estados do nordeste, como Bahia, Sergipe, Alagoas, Pernambuco e Paraíba (Astúa *et al.*, 2023; Faria, Lanes e Bonvicino, 2019; Paglia *et al.*, 2012). A espécie habita tanto locais florestados e preservados quanto áreas urbanas com alto grau de degradação ambiental, assim como seus congêneres (Bezerra-Santos *et al.*, 2020; Melo e Sponchiado, 2012). A espécie é majoritariamente noturna, onívora e com hábitos escansoriais, usando tanto substrato terrestre quanto arbóreo para locomoção e nidificação (Ferreira, 2023; Paglia *et al.*, 2012). Apresenta dimorfismo sexual, com as fêmeas sendo, em geral, menores que os machos (Astúa, 2010; Bubadué *et al.*, 2021). Machos e fêmeas também apresentam distinção em suas áreas de vida, com as fêmeas sendo mais territorialistas e tendo áreas de vida mais restritas do que machos (Brum, Melo e Cáceres, 2022; Cáceres e Monteiro-Filho, 2001). As fêmeas já foram descritas como tendo uma área de vida de cerca de 0,2 hectares (ha), enquanto os machos, na mesma área de 5 ha, tinham uma área de vida de cerca de 3 ha (Cáceres e Monteiro-Filho, 2001). Por isso, assumimos que machos e fêmeas apresentam hábitos ligeiramente distintos, além de fisiologias distintas, que podem refletir em variações no viroma de cada sexo.

Por muito tempo marsupiais didelfídeos foram considerados completamente solitários (Cáceres *et al.*, 2023), com machos e fêmeas interagindo apenas durante a época reprodutiva. Além disso, as fêmeas estarão em contato com seus filhotes até que esses amadureçam. Entretanto, as interações intraespecíficas não parecem ser restritas a essas duas situações. Para *Didelphis aurita*, já foi demonstrado o compartilhamento de tocas por irmãos recém desmamados (Astúa *et al.*, 2015).

Apesar do registro anedótico dessas interações, ainda não há outras evidências de um maior nível de sociabilidade para a espécie e do quanto esses indivíduos interagem fora das situações apresentadas, portanto é difícil inferir a que ponto isso poderia se refletir no viroma.

A adaptabilidade de *D. aurita* para habitar locais florestais preservados e antropizados propicia um importante modelo de estudo do viroma, permitindo avaliar de que forma o contato com humanos e animais domésticos pode afetar o viroma de um marsupial neotropical. Não só isso como podemos sugerir riscos de exposição à vírus patogênicos para humanos e animais domésticos das áreas considerando que os gambás agiriam como sentinelas da diversidade viral dos locais, já que espécies desse mesmo gênero já são consideradas sentinelas de outros patógenos (Silva, Estevão e Villela, 2024; Zitelli *et al.*, 2021). Consideramos como espécie sentinela aquela espécie que, por algum parâmetro fisiológico, indica a saúde, qualidade ou, neste caso, os vírus que circulam no ambiente em que ela está inserida (Tabor e Aguirre, 2004).

2. OBJETIVOS

2.1 OBJETIVO GERAL

Caracterizar o viroma gastrointestinal de populações de *D. aurita* e seu potencial como reservatório de zoonoses em áreas com diferentes níveis de perturbação antrópica no estado do Rio de Janeiro.

2.2 OBJETIVOS ESPECÍFICOS

1. Caracterizar a diversidade viral presente em *D. aurita*;
2. Avaliar o impacto de diferentes níveis de antropização e do sexo na composição do viroma de *D. aurita*;
3. Avaliar se *D. aurita* é um reservatório de vírus potencialmente zoonótico nas áreas de estudo.

3. HIPÓTESES

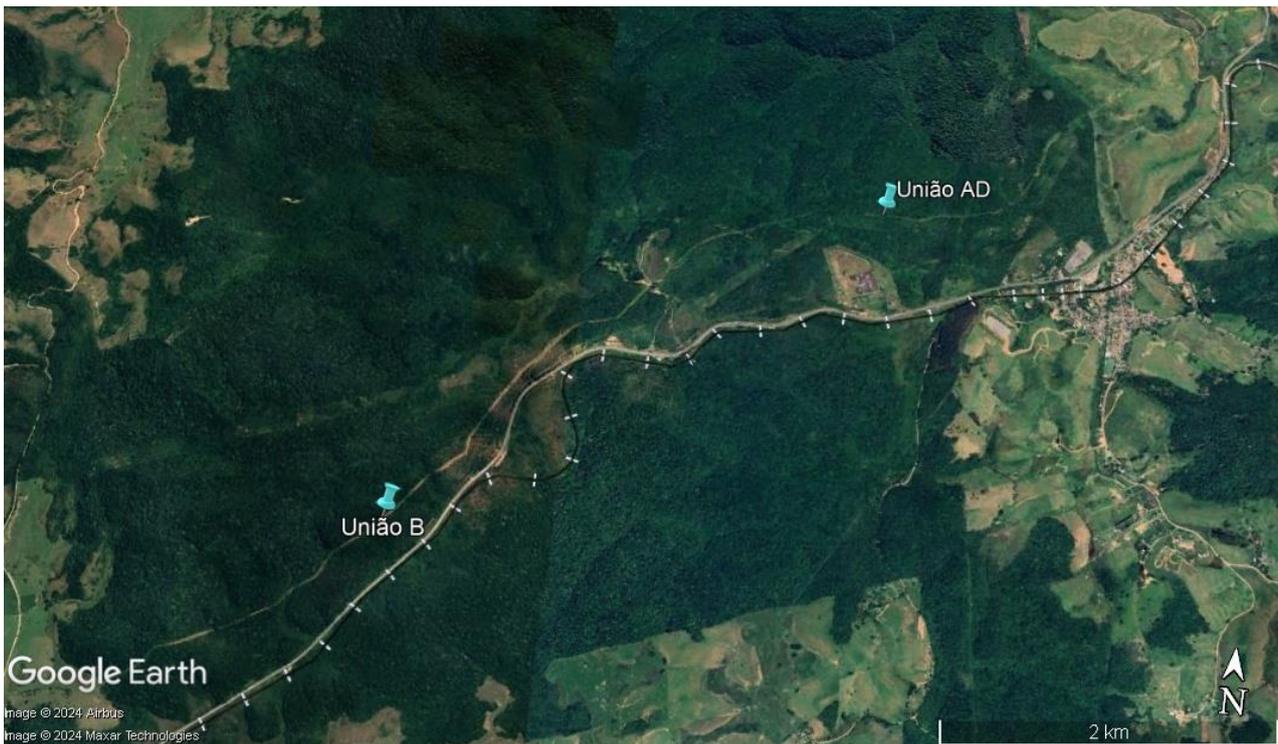
1. A composição de famílias virais nos conjuntos de amostra refletirá o nível de antropização das áreas de coleta, uma vez que a antropização do local facilita o contato dos hospedeiros com fontes virais diferentes dos ambientes naturais, por exemplo animais domésticos, esgoto, lixo etc.;
2. Conjuntos de amostras do mesmo sexo serão mais similares entre si, devido às diferenças inerentes aos sexos tanto ecológicas quanto fisiológicas;
3. A riqueza de famílias virais nos conjuntos de amostras provenientes de área urbana será menor, pois uma maior perturbação antrópica pode perturbar o equilíbrio da microbiota do animal e o tornar suscetível à disbiose, onde uma quantidade reduzida de famílias virais domina o viroma;

4. MATERIAL E MÉTODOS

4.1 ÁREAS DE ESTUDO

A coleta de amostras para o estudo foi realizada na Reserva Biológica União (ReBIO União; **Figura 1**), na Área de Proteção Ambiental da Bacia do Rio São João (APA da Bacia do Rio São João/Mico-Leão-Dourado; **Figura 2**) e no município de Campos dos Goytacazes (**Figura 3**). Todas as áreas pertencem ao estado do Rio de Janeiro (**Figura 4**), a ReBIO União estando nos municípios de Rio das Ostras, Casimiro de Abreu e Macaé (ICMBio, 2023). A APA da Bacia do Rio São João/Mico-Leão-Dourado está alocada nos municípios de Silva Jardim, Casimiro de Abreu, Rio das Ostras, Cachoeiras de Macacu, Rio Bonito, Araruama e Cabo Frio (ICMBio, 2008).

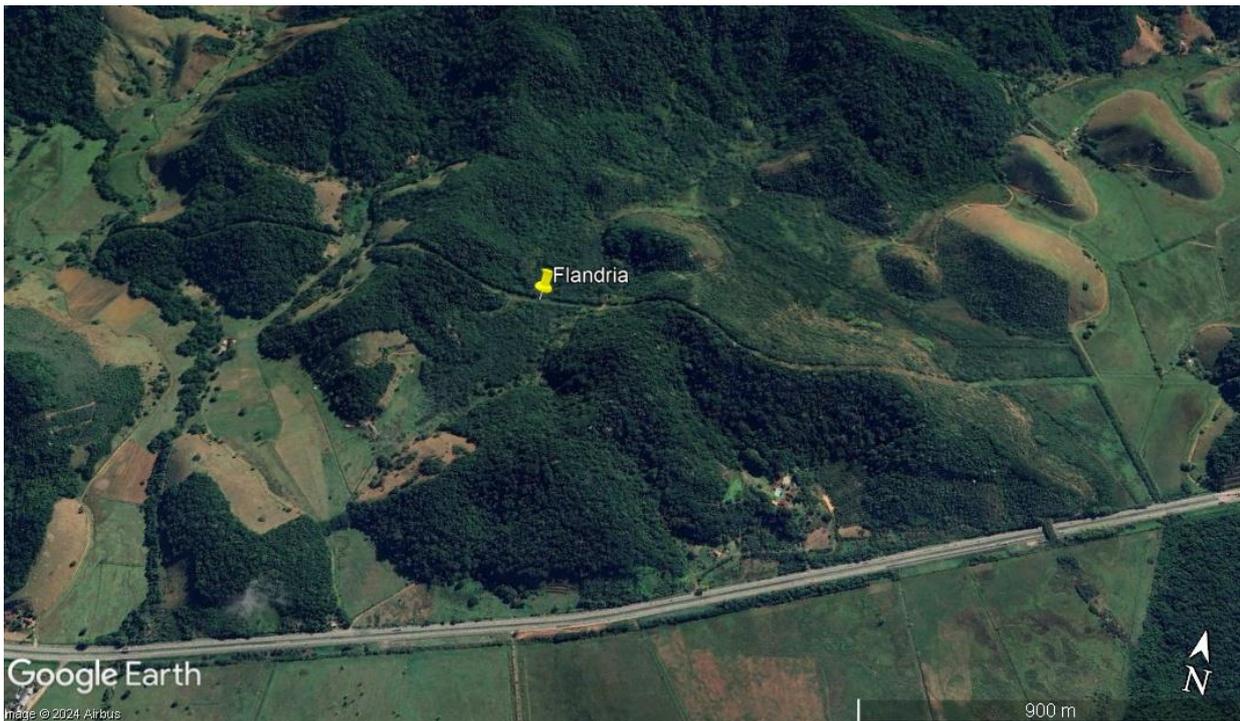
Figura 1 - Representação das áreas de coleta de amostras de fezes da espécie *Didelphis aurita* na Reserva Biológica União, entre 2019 e 2023.



Fonte: Google Earth (2024).

A ReBIO União é uma área de proteção integral que abrange 7.756,76 ha (ICMBio, 2023). Consideramos essa área como a que apresenta o menor grau de antropização. Isso se deve a proibição legal de uso dessa área para atividades extrativistas, mesmo que de uso sustentável. Logo, os animais dessa área têm menor contato com a ação humana e menores efeitos da antropização do seu hábitat. Por outro lado, a APA da Bacia do Rio São João/Mico-Leão-Dourado, com 150.700 ha (ICMBio, 2008), é categorizada como de uso sustentável. Por consequência, as áreas de coleta da APA estão inseridas em matrizes rurais, dentro de fazendas de uso privado. Sendo assim, nessas áreas os animais têm um contato intermediário tanto com o ser humano quanto com outros animais (de companhia e/ou de produção), com o animal de produção mais abundante sendo o gado.

Figura 2 - Representação das áreas de coleta de amostras de fezes da espécie *Didelphis aurita* na APA da Bacia do Rio São João/ Mico-Leão-Dourado, entre 2019 e 2023.



Fonte: *Google Earth* (2023).

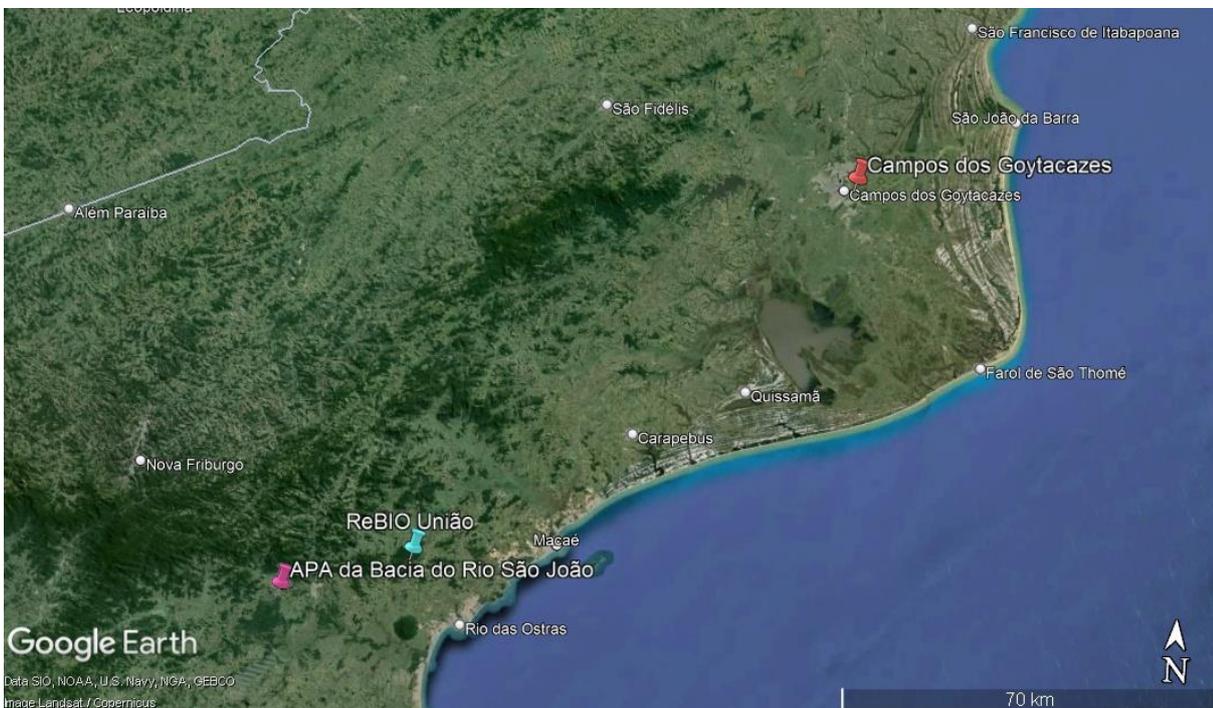
Por fim, temos o município de Campos dos Goytacazes, que está localizado na região Norte Fluminense do estado do Rio de Janeiro e é o maior município do estado, em extensão territorial, com 4.032,487 quilômetros quadrados (km²) (IBGE, 2022). O município foi nossa área de maior grau de antropização, justamente por ser uma área urbana que apresenta alguns pequenos fragmentos de floresta. As coletas no município ocorreram em locais que apresentavam pequenos fragmentos de mata, como no Horto Municipal e na própria Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF).

Figura 3 - Representação das áreas de coleta de amostras de fezes da espécie *Didelphis aurita* em Campos dos Goytacazes/RJ, entre 2022 e 2023.



Fonte: Google Earth (2024).

Figura 4 - Localização de todas as áreas de captura de *D. aurita* (ReBIO União; APA da Bacia do Rio São João e Campos dos Goytacazes) dentro do estado do Rio de Janeiro.

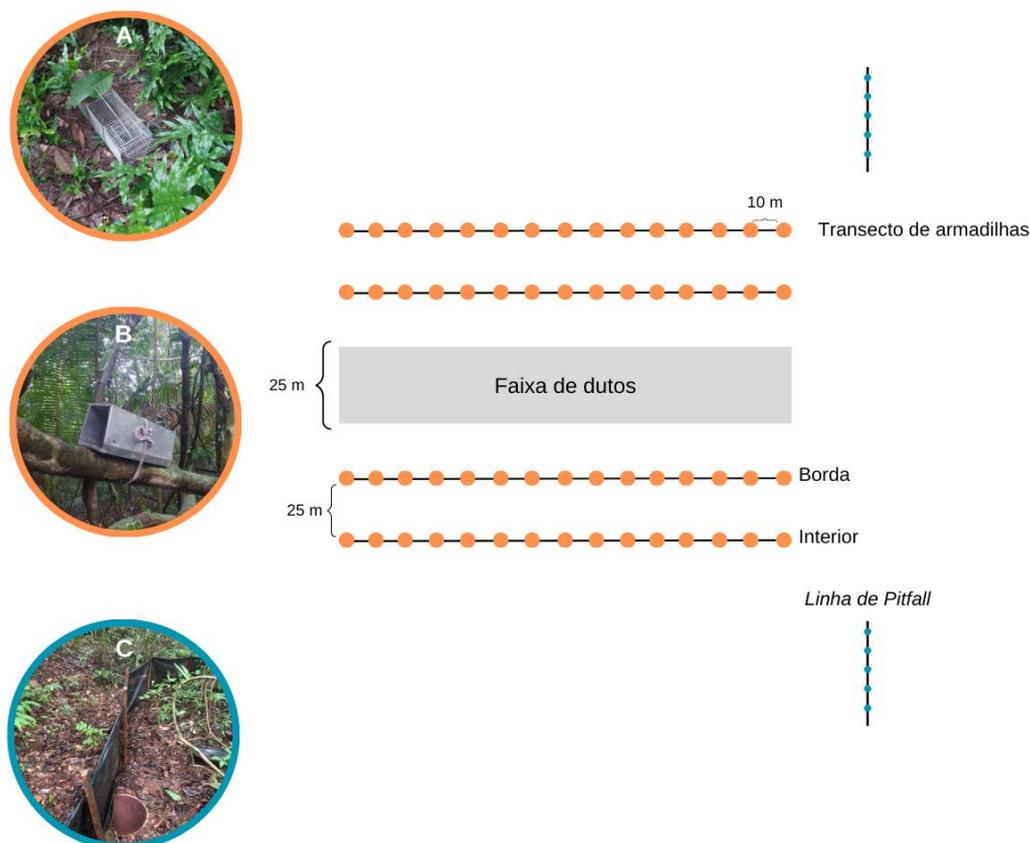


Fonte: Google Earth (2024).

4.2 DELINEAMENTO AMOSTRAL

O delineamento amostral na ReBIO União e na APA se baseou em quatro transectos por área amostrada, tendo cada transecto 15 estações de captura com duas armadilhas cada (*Sherman*[®] e *Tomahawk*[®]), uma no chão e outra no sub-bosque. Armadilhas de queda (do inglês *pitfall*), formadas por baldes de 60 litros (L) instalados na altura do solo (**Figura 5c**) e cercados por lona para guiar a trajetória do animal em direção ao balde, também foram instaladas. Cada área possuía duas linhas dessas armadilhas, com cinco baldes instalados no interior da floresta. Como essas duas áreas eram cortadas por uma faixa de dutos (gasodutos de petróleo) montamos dois transectos do lado esquerdo e dois transectos do lado direito do duto. Um transecto representava o que denominamos de borda, adjacente ao gasoduto, e outro representava o que chamamos de interior, distante da borda cerca de 25m (largura do gasoduto) (**Figura 5**).

Figura 5 - Desenho amostral utilizado nas coletas na Reserva Biológica União e na APA Bacia do Rio São João, representando os transectos e cada círculo laranja representa uma estação amostral. As armadilhas utilizadas foram *Tomahawk*® (A), *Sherman*® (B) e de queda (C). Os círculos azuis representam as armadilhas de queda.



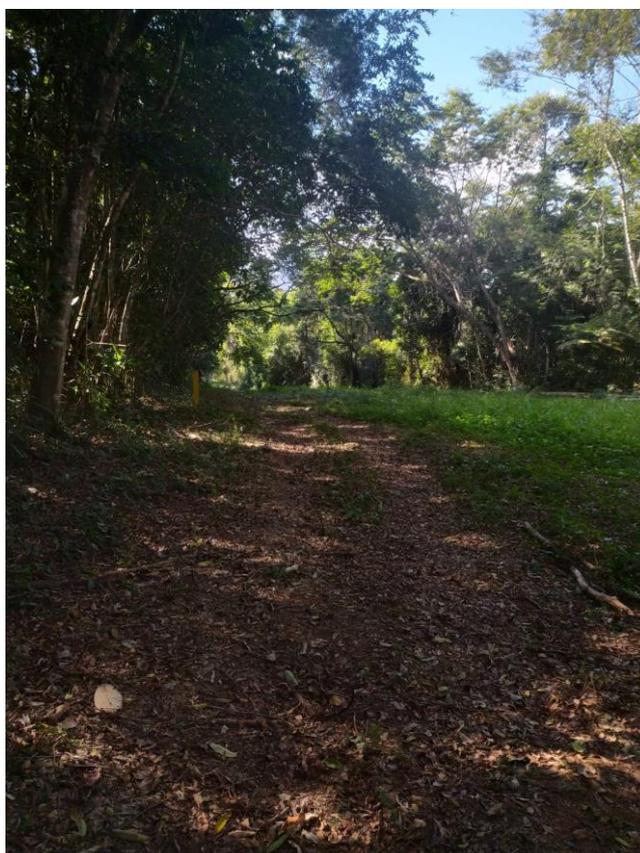
Em todas as áreas, as armadilhas ficavam abertas durante sete noites, com isca atrativa para pequenos mamíferos (uma mistura de farinha de milho para cuscuz, amendoim moído, banana e sardinha). Esse projeto foi realizado entre 2019 e 2023, com as amostragens nessas áreas ocorrendo nos anos de 2018, 2019, 2022 e janeiro de 2023. O salto temporal, entre 2020 e 2021, aconteceu devido a pandemia de coronavírus. No total oito campanhas foram realizadas na ReBIO União (quatro em 2019; três em 2022; e uma em 2023) e nove na APA da Bacia do Rio São João (uma em 2019; quatro em 2019; e quatro em 2022) (**Tabela 1**).

Tabela 1 - Esforço amostral entre todas as áreas amostradas. Armadilhas/noite se refere a quantidade de armadilhas multiplicado pelo número de campanhas e noites de armadilha aberta.

Área	Nº de armadilhas	Nº de campanhas	Armadilhas/noite
ReBIO União	360	8	20.160
APA da Bacia do Rio São João	120	9	7.560

Na ReBIO União, foram amostradas duas áreas (União A e União B). A União A (latitude -22,422339; longitude -42,021964) é uma das áreas onde os animais têm menor contato com o ser humano. Por sua localização ser na sede da reserva as normas referentes à necessidade de autorizações e permissões para que se a adentre são reforçadas, dificultando o contato dos animais com o público geral. A área como um todo é úmida, majoritariamente plana, e com um sub-bosque mais fechado (**Figura 6**).

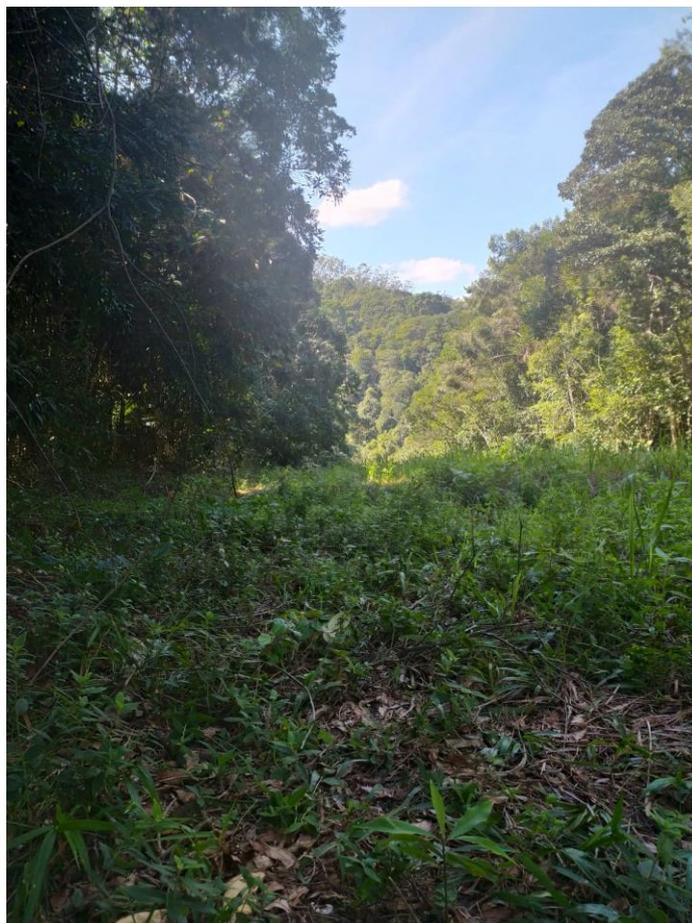
Figura 6 - Trecho da área de coleta ReBIO União A, Silva Jardim (RJ).



A União B (latitude -22,443004; longitude -42,050378) dista cerca de 3,7 quilômetros (km) da área anterior e está localizada próxima à rodovia. Essa proximidade com a rodovia faz com que a área seja mais acessível para a população e mais vulnerável a caçadores de animais silvestres, atividades essas que são consideradas ilegais. Sua fisionomia é similar à da área anterior, assim como a vegetação (**Figura 7**). Nela os animais possuíam um pouco mais de contato

com a antropização, mas ainda em menor escala do que as áreas da APA e Campos dos Goytacazes, por isso ainda a tratamos como uma das áreas que representa menor intervenção antrópica. Assim, a ReBIO União foi representada por duas áreas, amostradas com 360 armadilhas no total.

Figura 7 - Trecho da área ReBIO União B, Silva Jardim (RJ).

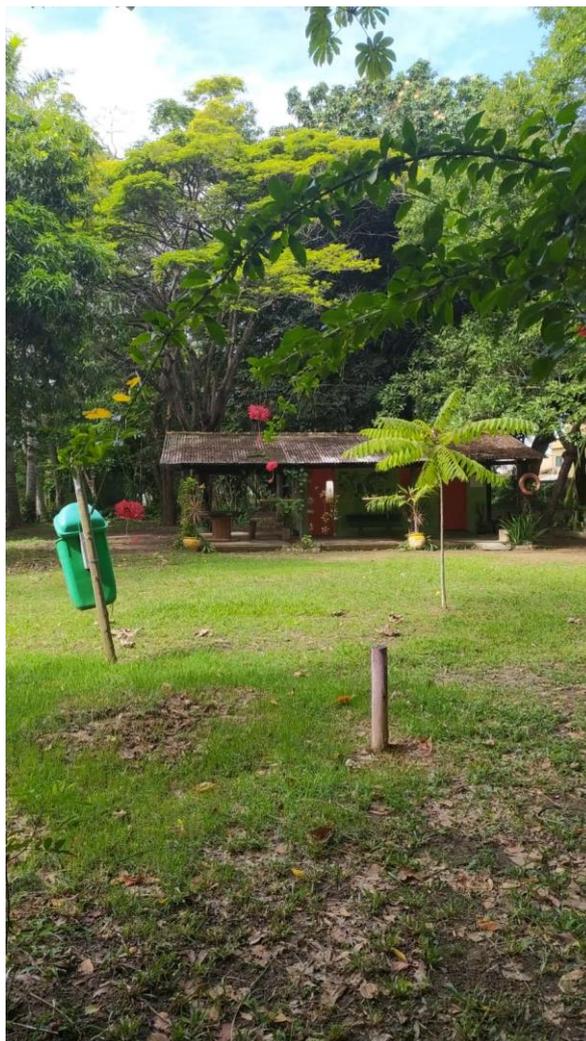


Na APA uma área (Fazenda Flandria) foi amostrada e 120 armadilhas foram instaladas no total. A Fazenda Flandria (latitude -22,50525; longitude -42,31626) é plana e extremamente úmida, com um de seus lados formando um brejo. Sua vegetação é a mais distinta das demais. Seu dossel tendia a ser mais aberto, porém com sub-bosque bem fechado. Nosso local de coleta estava inserido em matriz agrícola, por isso o nome da área é o nome de uma propriedade privada. Esta área dista cerca de 31,58 km da União A e 28,44 km da União B.

As coletas em Campos dos Goytacazes (**Figura 8**) foram realizadas em conjunto com um projeto de levantamento da fauna Norte Fluminense, também do VertEco/UENF (E-26/201.313/2022), e não seguiram o delineamento amostral

apresentado anteriormente. Os animais também foram capturados com o uso de armadilhas do tipo *Sherman*[®] e *Tomahawk*[®], porém não houve a instalação de armadilhas de queda e a distribuição das armadilhas variou conforme a área amostrada. Na UENF tivemos a definição de transectos de armadilhas em áreas da universidade com, em geral, dois tipos de armadilha em cada ponto. No horto amostrávamos sempre os mesmos pontos, mas sem definição de transectos. No condomínio Bosque das Acácias a disposição das armadilhas era definida conforme a área do condomínio que tínhamos permissão de amostrar. No total 6 campanhas foram realizadas no município de Campos, cada uma delas com cerca de 40 armadilhas. Campos dista cerca de 104,6 km da União A, 109,27 km da União B e 134,36 km da APA. A distância entre as áreas é o suficiente para não termos a sobreposição de área de vida e captura de indivíduos de uma área na outra. Como os indivíduos são marcados com anilhas numeradas temos a certeza de que essa sobreposição não ocorre.

Figura 8 - Ponto de coleta no Horto Municipal no município de Campos dos Goytacazes (RJ).



Todos os animais capturados tiveram suas medidas (tamanho do corpo, cauda e peso) anotadas. No momento da triagem as fezes eram coletadas em potes estéreis, próprios para coleta desse tipo de amostra, sempre que notávamos que o animal estava evacuando. Desse modo, as fezes caíam diretamente no pote, sem contato com o ambiente. Após a triagem os animais eram liberados novamente. As fezes frescas, pós atividade de campo, foram armazenadas em tubos do tipo *Falcon*[®] de 15 mililitros (mL), diluídas em *RNA/later*[®] (Thermo Fisher Scientific, Massachusetts, EUA) na proporção 1:1 em relação à quantidade de fezes, a fim de preservar o material genético ali presente. As amostras foram identificadas com nome da espécie, número do animal, data e o ponto de coleta, podendo permanecer até 3 meses em temperatura ambiente (TA) antes de serem transportadas para o laboratório de análise. No laboratório, as amostras foram armazenadas no freezer - 80°C.

Este trabalho foi realizado em conjunto com as atividades de três projetos distintos, sendo dois responsáveis pelas coletas das amostras e um responsável pelas análises de metagenômica. As coletas e organização geral do projeto ficaram sob responsabilidade do Laboratório de Ecologia de Vertebrados (VertEco) da UENF. A captura dos animais, coleta de amostras, transporte e processamento foram aprovadas pelos comitês responsáveis, SISBIO n° 64807-9 / CEUA-UENF n° 508.

4.2.1 Agrupamento das Amostras

Um total de 7 conjuntos de amostras (do inglês *pools*) foram formados de forma a permitir a análise da influência do sexo e da antropização sobre a composição e diversidade de famílias virais. Assim, os conjuntos foram compostos de amostras de indivíduos do mesmo sexo e da mesma área, exceto pelo conjunto da Flandria, em que foram incluídas amostras de machos e fêmeas devido ao baixo número de amostras de cada sexo, separadamente, disponíveis. A composição dos conjuntos de amostras é apresentada na tabela 2. Utilizamos apenas uma amostra por indivíduo para formar os conjuntos de amostras, de modo que os indivíduos não foram representados mais de uma vez.

Tabela 2 - Número de amostras utilizadas em cada conjunto e as siglas correspondentes.

Área	Sexo	N° de amostras	Sigla
ReBIO União A	F	11	UA_F
ReBIO União A	M	10	UA_M
ReBIO União B	F	8	UB_F
ReBIO União B	M	7	UB_M
APA	F e M	8	AP_FM
Campos dos Goytacazes	F	10	CG_F
Campos dos Goytacazes	M	8	CG_M

Os conjuntos foram formados dessa maneira para que comparações entre os viromas das áreas amostradas fossem feitas, assim como comparações por sexo. As comparações por sexo foram feitas entre os animais das áreas União A, União B, e aqueles capturados no município de Campos. Isso foi possível devido ao número de amostras dessas áreas permitir a separação por sexo. A quantidade de amostras compondo cada conjunto foi similar, permitindo que as comparações fossem feitas

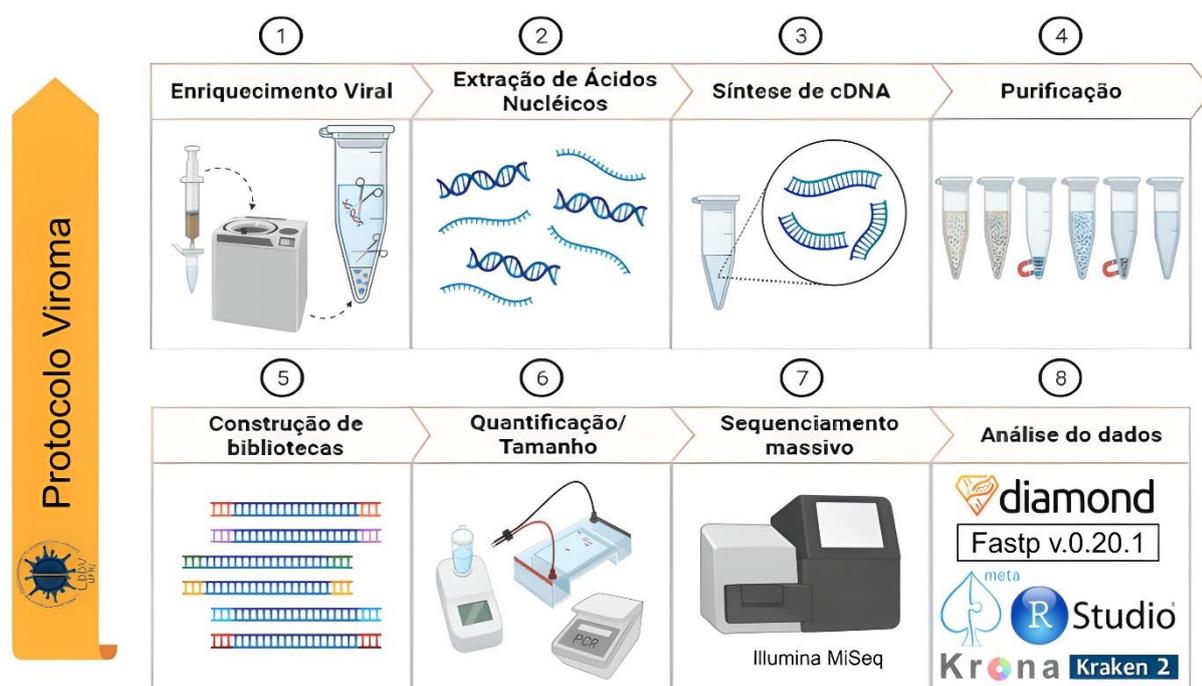
de modo equilibrado, sem que uma biblioteca fosse muito mais representada do que a outra. Desse modo também controlamos as diferenças de esforço amostral entre as áreas, considerando que as áreas tiveram, no geral, esforços amostrais de captura parecidos, mas selecionamos um recorte das áreas da ReBIO e da APA para amostrarmos o viroma.

4.3 VIROMA

As análises laboratoriais foram cobertas pelo projeto Rede Virofera Brasilis (VIBRA; E-26/211.355/2021), que tem como objetivo estudar o viroma de animais da fauna brasileira, sob responsabilidade do Laboratório de Diversidade e Doenças Virais (LDDV) da Universidade Federal do Rio de Janeiro (UFRJ). Neste trabalho realizamos duas levas de sequenciamento. A primeira, em fevereiro de 2023, com os conjuntos de amostras referentes as fêmeas das áreas ReBIO União A e B. A segunda leva foi feita em novembro de 2023, com os conjuntos restantes.

O protocolo laboratorial utilizado já é bem padronizado e demonstrado em trabalhos prévios para análises de viroma (*e.g.* D'arc *et al.*, 2020; Miranda *et al.*, 2022). As diversas etapas do protocolo de bancada e de bioinformática tiveram por objetivo elucidar a diversidade viral contida nos conjuntos de amostras (**Figura 9**). O protocolo fez uso de diversos *kits*, que foram utilizados conforme as recomendações do fabricante, com algumas pequenas modificações ao longo do processo. Vale destacar que ao longo de todo o processo foi feito um controle negativo, com água pura ou RNA^{later}[®], que passou por todas as etapas que as outras bibliotecas que sequenciamos. Desse modo, foi possível identificar se houve contaminação entre os conjuntos de amostras ou designação errônea das sequências.

Figura 9 - Resumo dos protocolos de bancada e bioinformática, estabelecidos pelo Laboratório de Diversidade e Doenças Virais (UFRJ), utilizados para análise das amostras de fezes de *D. aurita*.



Fonte: Coimbra, 2023 (adaptado de Schiffler, 2022).

4.3.1 Enriquecimento Viral

Começamos o nosso protocolo com o enriquecimento viral pois, em animais saudáveis, o material genético viral é representativamente menor do que os outros componentes genéticos. Para isso foi feito o completo descongelamento das amostras e homogeneização via vórtex. No caso de amostras fecais, utiliza-se também a homogeneização por esferas (do inglês *beads*), que consistiu na transferência de cerca de um mL de cada amostra para um tubo contendo esferas de extração (*Lysing Matrix E*; MP Biomedicals, Irvine, EUA). Esses tubos foram colocados no disruptor (*Super FastPrep-2*; MP Biomedicals) e homogeneizados por cerca de um minuto (min) e 30 segundos (s), com paradas a cada 30 s para incubação rápida em banho de gelo, para evitar superaquecimento das amostras. Por fim, as amostras foram centrifugadas a 6.000 forças gravitacionais (g), por 10 min, a quatro graus celsius (°C), para se recuperar o sobrenadante.

A partir deste momento os sobrenadantes foram unidos para a formação de um conjunto de amostras, visando o melhor custo-benefício da análise. O passo seguinte foi alocar 200 microlitros (µL) do sobrenadante de cada amostra no corpo

de uma mesma seringa com um filtro Millipore (*Millex-HV* – 0,45µm – SLHV0135L; Merck, Darmstadt, Alemanha) acoplado na ponta.

Para concentrar as partículas virais, foi realizada uma ultracentrifugação com gradiente de sacarose, o que nos permitiu selecionar a camada específica contendo a porção viral. Para isso, 500 µL do conjunto foram alocados em um novo tubo de 1,5 mL, contendo 200 µL de Sacarose 20%. O tubo contendo o material filtrado e a sacarose foi ultracentrifugado a 35.000 rotações por minuto (rpm), a quatro °C, por uma hora (h) e 30 min.

Por fim, foi feita a degradação do material genético livre no sobrenadante, através da utilização de diversas enzimas de digestão, uma vez que o material genético viral se encontra protegido dentro das partículas virais. Essas enzimas foram preparadas de modo a formar uma combinação de reagentes contendo: duas DNases, quatro unidades (U) de *RNase-Free DNase* (Promega, Madison, EUA) e seis U de *Baseline Zero DNase* (Epicentre, San Diego, EUA); uma RNase, 200 U de *RNase I* (Ambion, Waltham, EUA); e, uma nuclease, ~200 U de *Benzonase Nuclease* (Sigma, San Luis, EUA). Na combinação também foi adicionado 25 µL do *10x Baseline-Zero Buffer*, que garante as condições ideais para que as enzimas atuem. Após uma centrifugação rápida, 40 µL dessa combinação de reagentes foi transferido para o tubo contendo o conjunto de amostras. A reação foi então incubada por uma h e 30 min, a 37 °C, com agitação de 300 rpm.

Após a incubação foi adicionado dois µL do reagente de inativação (*10x Baseline-Zero Stop Solution*), que ajuda a parar a ação das enzimas, e novamente foi feita incubação, por 10 min a 65 °C. Essa etapa foi finalizada com uma centrifugação a 10.000 g por dois min, em TA, para que o reagente de inativação fosse sedimentado e pudesse ser retirado da amostra. Logo em seguida o sobrenadante (~200 µL) foi transferido para um novo tubo de 1,5 mL e passou-se para a etapa seguinte.

4.3.2 Extração dos Ácidos Nucleicos Virais

Realizamos a extração do material genético viral através da lise das partículas virais. Os capsídeos e envelopes foram submetidos a ação de enzimas e reagentes que degradam proteínas e ácidos graxos, fazendo com o que o material genético viral se tornasse livre no meio. Quando a integridade estrutural dos vírus é afetada, esses também se tornam inativos, não infectantes. Nessa etapa foi utilizado o

QIAamp MinElute Virus Spin Kit (QIAGEN, Hilden, Alemanha), com algumas modificações do protocolo sugerido pelo fabricante.

Resumidamente, 25 µL de protease e 200 µL de *Buffer AL* foram adicionados no tubo de amostra e a solução incubada a 56 °C, durante 15 min. Em seguida, adicionou-se 250 µL de etanol absoluto no tubo e foi feita novamente a incubação, por cinco min, TA, finalizando com uma centrifugação rápida do tubo. Como o objetivo principal foi fazer com que o material genético viral livre no meio fosse separado e isolado de outras partículas, o conjunto foi transferido para uma coluna de purificação. Desse modo, enquanto o líquido e outros componentes considerados contaminantes conseguem passar diretamente pela membrana, os ácidos nucleicos ficam nela retidos. Para que isso ocorresse realizamos algumas lavagens. Primeiro, centrifugamos o tubo a 8.000 rpm, por um min, TA, fizemos a lavagem da membrana com o tampão *AW2* e foi repetida a centrifugação, nas mesmas configurações. Para terminar a lavagem foi adicionado 500 µL etanol à coluna de purificação e repetida a centrifugação. Por fim, a membrana foi seca com uma centrifugação a 14.000 rpm, durante três min, ainda em TA. Para a liberação do material genético viral limpo na solução, foi adicionado 20 µL de água ultrapura na coluna, seguida de incubação (um min, TA) e centrifugação (14.000 rpm por um min).

4.3.3 Síntese da Dupla Fita de DNAC e Purificação

Com o material genético viral purificado, foi feita essa etapa para que o RNA presente fosse convertido em DNA complementar (DNAC). Deste modo, os vírus de DNA fita dupla permaneceram no conjunto sem serem alterados e continuamos com toda a diversidade viral que estava ali presente na forma de RNA, porém em formato de DNAC. Essa padronização em DNAC é necessária para que a máquina responsável pelo sequenciamento seja capaz de efetivamente sequenciar o que está presente no conjunto de amostras. Utilizamos a *Superscript IV First Strand Synthesys* (LifeTech, Waltham, EUA) e a *Klenow Fragment, 3'-5' exo-* (NEB, Ipswich, EUA).

Essa etapa depende do uso de três combinações de reagentes (*mixes*) para a Transcrição Reversa seguida da Reação em Cadeia da Polimerase (RT-PCR, do inglês *Reverse Transcriptase Polymerase Chain Reaction*), cada um com uma função específica. A primeira combinação (M1) foi responsável pelo anelamento e

continha sequências de seis dinucleotídeos degenerados, formados por todas as combinações possíveis de ATGC, capazes de se ligar a qualquer sequência. A segunda combinação (M2) foi responsável pela síntese da primeira fita do DNAC. Nessa combinação de reagentes está contida a transcriptase reversa (RT, do inglês *Reverse Transcriptase*), enzima capaz de converter a informação genética de uma molécula de RNA em DNAC fita simples, complementar ao RNA molde. A terceira combinação (M3) foi responsável pela síntese da segunda fita de DNAC.

Sendo assim, o M1 foi formado por 100 ng *Random Hexamers* e 20 mM *dNTP mix* e então misturado no conjunto de amostras (~20 µL). O material foi levado ao termociclador por cinco min a 65 °C e incubado no gelo por dois min. O M2 foi então preparado com oito µL de *5x SSIV Buffer*, 200 mM *DTT*, dois µL de *Ribonuclease Inhibitor* e 400 U de *SuperScript IV RT*. Em seguida este foi adicionado ao tubo e levado para um novo ciclo no termociclador, sendo esse de 10 min a 23 °C, 10 min a 50 °C e 10 min a 80 °C; e incubado no gelo. O M3 era composto por 5,5 µL de *Klenow Buffer 10x*, 25mM de *DNTPs* e 12,5 U de *Klenow Enzyme*. Essa etapa foi finalizada no termociclador, com um ciclo de uma hora a 37 °C, 20 min a 75 °C e manutenção a quatro °C.

Após a síntese do DNAC dupla fita partiu-se para a limpeza do DNAC e concentração dele, pois no sequenciamento utiliza-se um volume pequeno de material inicial. O *Agencourt AMPure XP Kit* (Beckman Coulter Genomics, Indianapolis, EUA) foi utilizado para essa etapa e se baseia na purificação do material genético da solução a partir do uso de esferas magnéticas, em diferentes quantidades. Quando em solução ideal, as esferas conseguem se unir ao DNA, por diferença de carga, e permitir que a solução seja lavada/descartada sem perda do material genético. Além disso, quanto maior o volume/quantidade de esferas, maior será a capacidade de apreender fragmentos de diversos tamanhos.

A etapa foi iniciada com a alocação do conjunto em um poço de uma placa de 96 poços e adição de 85 µL de *beads* (1,8x). A placa foi então incubada por 15 min em TA e levada para a estante magnética por três a cinco min. Sem remover a placa da estante o poço foi lavado com 200 µL de etanol 80% fresco, duas vezes, e incubado por cinco a 10 min, em TA, até evaporação completa do etanol. Após secas, as esferas magnéticas foram ressuspensas com 17 µL de *Resuspension-Buffer* (RSB) do *Tagmentation Kit* (Illumina, San Diego, EUA), um reagente responsável tanto por hidratar quanto por fazer com que as esferas perdessem

afinidade ao DNA. A placa foi então removida da estante, o conteúdo do poço homogeneizado e a placa incubada por dois min em TA. Essa etapa foi finalizada com a alocação da placa na estante magnética, incubação e remoção do material genético limpo sem as esferas.

4.3.4 Confeção e Purificação das Bibliotecas

A formação da biblioteca é importante para o sequenciamento em si e a identificação das sequências durante o processo de análise, pois esse é feito com diversas bibliotecas ao mesmo tempo, por isso cada uma deve ter um código que a identifique. O que chamamos de biblioteca é a construção molecular que une sequências conhecidas, como o adaptador e o código de barras (do inglês *barcode/index*) molecular, e a sequência de nucleotídeos desconhecida de interesse. Cada biblioteca recebe dois códigos de barras, um no sentido senso e outro no antisenso. Essa etapa foi realizada com o *Nextera XT DNA Library Preparation Kit* (Illumina), que possui uma etapa de fragmentação enzimática, onde é anelado o código de barras, e uma etapa de PCR, responsável por finalizar a construção da biblioteca e sua amplificação.

O *kit* continha reagentes como *Tagment DNA Buffer* (TD) e a *Amplicon Tagment Mix* (ATM), que fragmenta o DNA ao mesmo tempo que anela sequências conhecidas de DNA nas extremidades do fragmento. Essas sequências são conhecidas como adaptadores e auxiliam a introdução do código de barras na biblioteca. Por isso, primeiro foi colocado cinco μL de TD, depois 2,5 μL do *pool* e 2,5 μL de ATM no poço de uma nova placa e realizada uma centrifugação por um min a 300 g e 20 °C. Depois a placa foi levada ao termociclador e incubada por cinco min a 55 °C e removida rapidamente ao atingir 10 °C para dar continuidade ao procedimento.

Após retirar a placa do termociclador a reação de fragmentação foi interrompida com a adição de 2,5 μL de *Neutralize Tagment Buffer* (NT), centrifugação por um min a 300 g e 20 °C e incubação por cinco min em TA. Após decorrido este tempo foi adicionado 2,5 μL do código um e 2,5 μL do código dois em cada poço e 7,5 μL de *Nextera PCR Master Mix* (NPM). A placa foi levada para o termociclador e passou por uma ciclagem consistindo em três min a 72 °C, seguidos de 30 s a 55 °C e o ciclo repetido 18x de 10 s a 95 °C, 30 s a 55 °C e 30 s a 72 °C. A

ciclagem foi finalizada com cinco min a 72 °C e em seguida a temperatura baixou para oito °C até que a placa foi retirada do termociclador. Por fim, esta etapa foi finalizada com uma centrifugação a 300 g, durante um min, a 20 °C e avançou-se para a etapa seguinte.

Este passo foi similar a purificação e concentração do material genético, seguindo os mesmos passos, porém com algumas distinções. A principal diferença está na quantidade de esferas magnéticas utilizada, pois aqui utilizamos uma quantidade menor, o que permitiu que fragmentos grandes (>800 pares de bases - pb) fossem removidos da nossa solução, considerando que nossos fragmentos de interesse devem ter tamanho entre 200 e 800 pb, como estabelecido pela Illumina. Utilizamos novamente o *Agencourt AMPure XP Kit (Beckman Coulter Genomics)*.

Primeiro foi adicionado cinco µL de esferas magnéticas (0,2x) ao poço com a biblioteca e este foi incubado da mesma maneira que anteriormente. Depois o sobrenadante (~25 µL) foi transferido para outro poço da placa e a placa foi retirada da estante magnética. Neste novo poço foram alocados 15 µL de esferas (0,6x). Após decorrido o tempo de incubação, todo o sobrenadante deste poço foi descartado. Em seguida foi feita a lavagem e ressuspensão das esferas. Essa etapa foi finalizada com a alocação do conteúdo dos poços em dois tubos distintos. Um dos tubos contendo fragmentos menores que 800 pb e o outro com fragmentos acima de 800 pb. Passou-se então para a etapa seguinte apenas com o tubo contendo os fragmentos menores que 800 pb.

4.3.5 Avaliação da Biblioteca

Após a montagem das bibliotecas, foi necessário avaliar sua qualidade, uma característica que depende tanto da concentração quanto do tamanho dos fragmentos nela presentes. Esse passo é importante pois a qualidade é crucial para decidir se a biblioteca está apta para prosseguir para o sequenciamento. Para avaliar a concentração, foi feita uma quantificação utilizando o QuBit 2.0 com o *QuBit dsDNA HS Assay Kit (LifeTech)*, que se baseia em marcar os fragmentos de DNA com uma molécula fluorescente. Essa molécula, quando excitada pela luz emitida no QuBit, fluoresce e a máquina é capaz de estimar a concentração de DNA na biblioteca a partir da intensidade dessa fluorescência.

Primeiro foi feita uma combinação de reagentes com *Qubit® dsDNA HS Buffer* (199 μ L) e *Qubit® dsDNA HS Reagent* (um μ L), desta combinação foi retirado volume suficiente para o misturar com a alíquota da biblioteca obter um volume final de 200 μ L de reação. Foram realizadas duas aferições por biblioteca, uma contendo três μ L e outra contendo cinco μ L da biblioteca, com 197 μ L e 195 μ L da combinação de reagentes, respectivamente. Desse modo o resultado foi uma média desses dois resultados, o que permitiu maior confiabilidade. Antes de passar as amostras pelo equipamento, esse foi calibrado com dois padrões, um trazendo a concentração mínima e outro a máxima a ser detectada pelo QuBit.

Para avaliar o tamanho da biblioteca foi feita uma PCR com visualização em gel de agarose. Fazemos a PCR para amplificar a biblioteca novamente, facilitando sua visualização, já que utilizamos uma quantidade diluída da biblioteca. Foi necessário diluir para otimizar a reação de PCR, pois assim permitimos que as enzimas conseguissem encontrar os fragmentos durante o tempo da reação. Uma quantidade muito grande de DNA faria com que as enzimas fossem saturadas muito cedo. Para isso foi utilizado o *NEBNext® Library Quant Kit* (NEB, Ipswich, EUA) e preparado um tampão de diluição 1x com 45 μ L de água e cinco μ L de *NEBNext Dilution Buffer 10x* (por biblioteca). O tampão foi então homogeneizado via vórtex, centrifugado rapidamente e adicionado 49 μ L deste em um μ L de biblioteca. No novo tubo foi feita homogeneização via vórtex e centrifugação rápida.

Depois, dois μ L de biblioteca diluída foram alocados em um tubo com oito μ L de *NEBNext Master Mix + Primer*. Fizemos o mesmo com o controle positivo, composto de um fragmento de DNA de tamanho conhecido, e com o negativo, o tampão de diluição preparado anteriormente. A ciclagem da PCR foi então iniciada no termociclador com um ciclo de um min a 95 °C, seguido de uma repetição de 35x de 15 s a 95 °C e 45 s a 63 °C, e manutenção a oito °C até remoção dos tubos do equipamento. Os tubos foram retirados do termociclador, centrifugados rapidamente e preparados para visualização no gel.

O gel de agarose é um polímero poroso que permite a passagem do DNA pelo seu interior devido à natureza negativa da molécula que é atraída para o polo positivo da cuba de eletroforese, quando aplicada uma corrente elétrica. Quanto mais concentrado for este gel, mais difícil será a migração do DNA. Aqui foi utilizado um gel 1,5% pois essa concentração é eficaz para visualização de fragmentos entre

200 e 3.000 pb. As moléculas foram marcadas com um corante fluorescente (*dye*) que permite visualização após ser excitado com luz ultravioleta (UV).

Foram alocados oito μL do produto de PCR da biblioteca, dos controles e do que chamamos de *ladder* (escada, em português), um marcador de peso molecular conhecido que mostra bandas entre 100 e 1.000 pb, em poços do gel. Todos foram previamente misturados com o corante. O gel foi então colocado na cuba e a corrida ocorreu por ~uma h e 30 min. Após a corrida, o gel foi levado para o transiluminador *L-Pix EX* (*Loccus*, Cotia, Brasil) e excitado com luz UV, gerando uma imagem com as bandas identificadas no gel. A identificação do tamanho médio da biblioteca consistiu em observar o *ladder* e comparar com o arrasto deixado pela biblioteca pois essa não deixa bandas definidas, já que possui fragmentos de diversos tamanhos. Entretanto, existe um tamanho médio onde a maioria dos fragmentos se localizam e este que foi identificado com a comparação ao *ladder*.

Como a avaliação do tamanho por visualização em gel depende do conhecimento do usuário e pode ser subjetiva, foi realizada uma segunda avaliação. Dessa vez utilizando o *BioAnalyzer* e o *Agilent High Sensitivity DNA Kit* (*Agilent*, Santa Clara, EUA). O *BioAnalyzer* funciona de maneira similar ao gel de agarose, também sendo pautado no uso de um gel que permite que o DNA o percorra durante a passagem de uma corrente e seja visualizado por fluorescência. Entretanto, o gel agora se localiza no interior de um *chip* específico para ser colocado no equipamento. A diferença crucial está no fato de que a máquina nos retorna o tamanho médio da biblioteca de forma automática e por ser uma tecnologia muito mais sensível e precisa.

Primeiro foi preparado o gel, onde 15 μL de *High Sensitivity DNA dye - HSD*, previamente homogeneizado via vórtex e centrifugado rapidamente, foram misturados ao tubo com nosso gel (*HSD gel matrix*). Essa mistura foi homogeneizada via vórtex rapidamente e centrifugada a 6.000 rpm, durante 10 min, TA. O gel foi adicionado ao *chip* em três poços distintos (nove μL em cada poço) e em seguida foi adicionado cinco μL de um marcador do tamanho máximo e mínimo detectado pelo *kit* (*HSD marker*) em todos os poços. Por fim, foi adicionado um μL de *ladder* no seu respectivo poço e um μL de biblioteca em um dos poços livres. A preparação do *chip* terminou com uma homogeneização via vórtex por um min, a 2.400 rpm.

Por fim, o chip foi inserido no *BioAnalyzer* e os resultados obtidos por meio dos gráficos e relatórios fornecidos pelo equipamento. Essa informação foi unida ao tamanho estimado pela etapa anterior e o tamanho médio final das bibliotecas foi um consenso entre os dois resultados.

4.3.6 Plataforma Illumina - MiSeq

Como a plataforma Illumina - MiSeq possui apenas uma porta de entrada para sequenciar todas as bibliotecas ao mesmo tempo, elas foram previamente unidas em um único conjunto. O volume total desse conjunto não é formado por um volume equivalente de cada biblioteca, pois isso depende tanto do tamanho médio daquela biblioteca quanto da sua concentração. De maneira geral, bibliotecas que possuem um tamanho médio entre 300 e 500 pb, e uma molaridade acima de quatro nM podem ser consideradas ótimas. As bibliotecas que não cumprem esses requisitos ainda podem ser sequenciadas, porém essas irão compor uma parte menor do conjunto.

Isso é necessário para o bom funcionamento do equipamento que iremos utilizar. Uma quantidade muito grande de uma biblioteca que possui molaridade baixa, faria a máquina travar pois essa rapidamente se esgota de sequências para ler. O mesmo ocorre para sequências de tamanho muito pequeno. Para o preparo do conjunto de bibliotecas foi decidido previamente o volume de cada biblioteca que seria pipetado em um único tubo, levando em consideração as informações de qualidade. Por fim, foi feita a quantificação do conjunto com o QuBit, seguindo o mesmo protocolo descrito anteriormente, para confirmar sua concentração e diluí-lo, com água destilada, para quatro nM, o que é exigido pela empresa como massa ideal para o início do protocolo de sequenciamento.

Por fim, os sequenciamentos foram realizados com o *MiSeq v2 2x150 cycles Reagent kit* (Illumina). O preparo do conjunto de bibliotecas para ser colocado no sequenciador começou com a desnaturação da dupla fita de DNA, necessária para o anelamento das bibliotecas nos adaptadores fixados na placa de vidro, chamada *flow cell*, fornecida pelo *kit*. Foi adicionado cinco μL de NaOH a 0,2M em um tubo com cinco μL do conjunto de bibliotecas, que foi homogeneizado via vórtex e incubado por cinco min em TA para que a reação ocorresse. Após esse tempo, a reação foi parada com a adição de 990 μL de HT1, fornecido pelo *kit*. Em seguida, 240 μL do conjunto foram transferidos para outro tubo e mais 360 μL de HT1

adicionados ao tubo. Neste momento a molaridade de oito pM foi atingida, uma molaridade que já se mostrou eficaz para o sequenciamento de alta performance em outros estudos de viroma realizados pelo laboratório.

Para ser colocado na máquina, 600 µL do conjunto foram pipetados no que chamamos de cartucho do MiSeq, uma espécie de caixa com diversos poços e todos os reagentes necessários para que as etapas do sequenciamento ocorram. Antes de iniciar o sequenciamento a máquina foi ajustada com todos os componentes sendo colocados nos compartimentos corretos e a quantidade de ciclos desejada foi configurada (2x151 ciclos). Cada ciclo corresponde ao número de pares de base da fita no sentido senso e antisenso da biblioteca de DNA que serão lidos. Ou seja, sabendo o tamanho médio das nossas bibliotecas, devemos ajustar o número de ciclos para que esse esteja de acordo com as bibliotecas de menor tamanho. Por fim, o cartucho foi colocado na máquina e a corrida iniciada.

Essa plataforma usa o método de sequenciamento por síntese, pautado em uma PCR. Esse processo consiste na síntese de novas fitas de DNA a partir das sequências presentes na biblioteca, e leitura de fluorescência. Os adaptadores ligados a placa de vidro (*flow cell*) são complementares aos adaptadores que foram adicionados à biblioteca. Quando temos esse anelamento, em um primeiro momento, ocorre o que chamamos de clusterização, ou agregação. Essa etapa é baseada em uma amplificação por ponte, formando agregados (do inglês *clusters*) de material genético clone de um único fragmento.

Após formação dos agregados em ambos os sentidos das fitas, temos a clivagem das fitas de sentido antisenso da placa e ficamos apenas com as fitas senso (R1). Então, partimos para o sequenciamento propriamente dito. Este se inicia com a identificação do iniciador presente no nosso fragmento e a cada ciclo temos a ligação de um nucleotídeo complementar ao que temos no nosso fragmento. Esses nucleotídeos são fluorescentes e cada um possui uma cor distinta do outro, de modo que quando esse se liga ao seu complementar uma luz é emitida e o sequenciador é capaz de distinguir qual nucleotídeo emitiu aquela luz. Como estamos lidando com agregados, a máquina consegue perceber de maneira clara o nucleotídeo sinalizado pois o agregado inteiro fluoresce de forma sincronizada.

Após o sequenciamento da fita senso temos o mesmo processo ocorrendo para a fita antisenso (R2), recuperada após mais um ciclo de amplificação por ponte. Entre a leitura de uma fita e outra também temos a leitura de cada código de barras,

o que permite à máquina atribuir as sequências de maneira correta para as bibliotecas. O próprio sequenciador, após finalizar a corrida, nos retorna arquivos com todas as sequências correspondentes a cada biblioteca. As corridas duraram entre 28h e 36h. Após finalizada passou-se para a etapa de análise de bioinformática.

4.4 ANÁLISES GENÉTICAS

Depois que os trabalhos de bancada “molhada” foram finalizados, as sequências passaram por um fluxo de análises bioinformáticas (*pipeline*) para a identificação taxonômica das sequências presentes na biblioteca. As leituras brutas, antes de serem identificadas, passaram por um controle de qualidade tanto do tamanho quanto da confiabilidade da leitura de cada nucleotídeo (*QScore*). As leituras que apresentaram tamanho maior que 50 pb e *QScore* maior que 30 passaram para as próximas etapas e as reprovadas foram retiradas das análises posteriores. Esse controle foi feito com três ferramentas distintas. Primeiro as leituras passaram pelo *FastQC* 0.11.3 (Andrews, 2010), uma ferramenta elaborada justamente para fazer uma avaliação da qualidade dos dados oriundos de HTS. Depois, foi usado o *fastp* v.0.20.1 (Kearse *et al.*, 2012), para filtrar os dados baseado nos parâmetros definidos pelo usuário. Por fim, as leituras pós-filtro passaram pelo *FastQC* novamente e todos os relatórios de qualidade foram agrupados utilizando o *Multiqc* v1.14 (Ewels *et al.*, 2016), responsável por compilar esses dados em um único arquivo.

As análises bioinformáticas também incluíram um genoma de referência. Este pode ser tanto do hospedeiro alvo do estudo, ou do organismo mais próximo filogeneticamente, desse hospedeiro, que tenha seu genoma completo descrito. Essa etapa é utilizada como forma de “limpar” as sequências e excluir aquelas que são oriundas do hospedeiro. Desse modo, garantimos que as sequências identificadas representam genomas que são de outras fontes que não o próprio hospedeiro. No nosso caso, por não haver o genoma completo de *Didelphis aurita* disponível, utilizamos como genoma de referência o da espécie *Monodelphis domestica* (RefSeq: GCF_027887165.1).

As leituras passaram para a identificação taxonômica, onde as sequências foram atribuídas a algum táxon. Essa análise foi feita com as ferramentas *Kraken2*

v.2.1.2 (Wood *et al.*, 2019) e *Diamond* v.2.0.14 (Edgar, 2004). A análise realizada pelo *Kraken2* é baseada na assinatura de cada sequência, ou seja, ele busca por padrões que indicam que aquela sequência seria pertencente a um determinado táxon. Chamamos essas assinaturas de *k-mers*, então o *Kraken2* trabalha de maneira que ele busca correspondências exatas para os *k-mers* que ele encontra (Cárdenas, Neuenschwander e Malaspinas, 2022). Entretanto, por trabalhar dessa maneira, o *Kraken2* é menos conservador, ao atribuir as sequências a um determinado táxon, do que ferramentas que fazem busca por similaridade.

Por outro lado, o *Diamond* compara as sequências que nós temos com aquelas presentes em bancos de dados públicos. O *Diamond* também funciona com um algoritmo do tipo Ancestral Comum Mais Próximo (do inglês *Lowest Common Ancestor*), onde uma leitura que for identificada como sendo muito similar a mais de uma espécie será classificada no nível taxonômico comum a todas essas espécies (Cárdenas, Neuenschwander e Malaspinas, 2022). Logo, o *Kraken2* é uma ferramenta mais suscetível a sinalizar falsos positivos, enquanto o *Diamond* é mais suscetível a retornar falsos negativos. Ambas as ferramentas retornam resultados ao nível de espécie, entretanto, a confiabilidade tende a ser menor devido ao tamanho reduzido das sequências disponíveis para análise, por isso foi utilizado o nível de família viral para este estudo.

As mesmas análises também foram realizadas com fragmentos de maior tamanho (*contigs*). Esses fragmentos são formados a partir do programa *SPAdes* v3.15.3 (Rice *et al.*, 2000), que utiliza as leituras presentes na biblioteca e alinha aquelas que possuem um determinado grau de sobreposição, formando um fragmento de maior tamanho que as leituras originais (*contig*). Os dados obtidos foram resumidos graficamente e tabelados com o uso do *Krona* v.2.7.1 (Capella-Gutiérrez *et al.*, 2009), que é capaz de criar gráficos com camadas múltiplas do nível taxonômico mais abrangente até o menor táxon identificado. A ferramenta retorna os mesmos dados em um arquivo “.txt”.

Após finalizada a etapa de bioinformática, os números de leituras listados no arquivo .txt passaram por um ajuste metodológico, onde foi aplicado um corte de 1% para a plataforma *Kraken2*. Ou seja, o número de leituras de uma família viral em cada biblioteca só foi mantido caso este número fosse maior do que 1% do valor apresentado para a biblioteca com o maior número de leituras para aquela família. Isso é necessário para compensar qualquer erro de atribuição durante o

sequenciamento, por isso o corte foi feito com todas as bibliotecas sequenciadas ao mesmo tempo. Ou seja, o corte não foi feito apenas com as bibliotecas aqui apresentadas, mas sim com todas que forem sequenciadas ao mesmo tempo, independente do hospedeiro de interesse, já que estamos buscando minimizar erros do sequenciador.

Os dados obtidos pela plataforma *Kraken2* foram o alvo principal das análises no presente projeto. De forma a lidar com o menor conservadorismo dessa plataforma, além de realizar as análises para as famílias virais identificadas pelo *Kraken2* pós corte de 1%, fizemos também as análises considerando apenas as famílias virais identificadas simultaneamente por ambas as ferramentas *Kraken2* e *Diamond*. Essa escolha foi feita devido à possível existência de uma taxa de falsos positivos do *Kraken2*, de cerca de 1% (Cárdenas, Neuenschwander e Malaspinas, 2022; Johnson, Sun e Fodor, 2022). Dessa maneira, para fins de descrever a presença de famílias virais, teríamos maior confiabilidade, já que confirmamos a presença das famílias no *Diamond*, uma ferramenta conhecidamente mais conservadora. Os dados do *Kraken2*, apesar de ainda não serem amplamente usados para identificação de novos vírus, cumprem o propósito de descrever amplamente a diversidade viral e assim são úteis para serem utilizados em análises ecológicas.

4.5 ANÁLISES DOS DADOS

A riqueza foi calculada pelo número de famílias virais identificadas em cada biblioteca. Para avaliar a eficiência dos métodos para amostrar as famílias virais presentes na amostra, foi calculada uma curva de rarefação baseada em indivíduos para cada biblioteca pós corte de 1% (da Silva *et al.*, 2022). Aqui tratamos cada leitura como um indivíduo. Decidimos tratar cada leitura como um indivíduo distinto porque essa é a menor unidade amostral que nos permite identificar a presença de uma família em uma biblioteca.

Para algumas representações dos dados nós utilizamos uma classificação das famílias por hospedeiros foi feita baseada nos hospedeiros já identificados para cada família viral, de acordo com a base de dados do *Viral Zone* (Hulo *et al.*, 2011). Famílias virais conhecidas por infectar apenas um grupo, ou dois grupos similares, de organismos foram separadas como tal (*e.g.* Bacteria; Archea; Algae; Plantae;

Protozoa - **Apêndice A**). Famílias virais que infectam diferentes grupos de animais foram classificadas como Metazoa, as que infectam diferentes grupos de eucariotos como Eukaryota, e as indicadas com infectantes de vertebrados foram classificadas como Craniata. Famílias virais com hospedeiro tipo como desconhecido foram classificadas como "N/A".

A diferença na composição de famílias virais entre conjuntos de amostras de acordo com a localidade e o sexo dos hospedeiros foi calculada a partir do índice de dissimilaridade de Jaccard (Magurran, 2021), baseado na presença ou ausência das famílias virais nas bibliotecas. A partir da matriz de dissimilaridade de Jaccard, foi construído um dendrograma, baseado em médias aritméticas não-ponderadas (UPGMA) (Magurran, 2021), de forma a observar se as bibliotecas tenderam a se agrupar de acordo com o índice de antropização das áreas e/ou com o sexo dos indivíduos.

Para explorar mais a forma que a composição de famílias virais se distribui entre os hospedeiros de ambos os sexos e em localidades sob diferentes níveis de antropização, foi empregada análise de elementos de estrutura de metacomunidades (Leibold e Mikkelsen, 2002; Presley, Higgins e Willig, 2010). Essa análise avalia se existe algum gradiente, neste caso o sexo ou antropização da área, estruturando a distribuição das famílias virais e como elas se estruturam ao longo desse suposto gradiente. Seguimos a nomenclatura proposta por Braga, Oliveira e Cerqueira (2017).

5. RESULTADOS

Foi reconhecido um total de 36.223.147 leituras pelo *Kraken2*. Dessas, 7.643.459 foram identificadas com sucesso. As outras 28.579.688 leituras reconhecidas pela ferramenta não puderam ser identificadas a nenhum nível taxonômico. Essas são chamadas de leituras "não classificadas". O *Diamond* reconheceu 36.220.147 leituras, identificando com sucesso 7.899.596. As outras 28.320.551 leituras que restaram não foram classificadas (**Apêndice B**). Tanto no *Kraken2* (0,71%) quanto no *Diamond* (1,65%) a maior parte dessas leituras identificadas não eram de vírus, assim, estes representaram uma das menores porções classificadas (**Apêndice C**).

Após o corte de 1% ficamos com 193.288 leituras virais. O menor número de leituras virais foi atribuído a biblioteca de machos da União B com 466 leituras e o maior para a biblioteca de fêmeas de Campos, com 150.800 leituras (**Tabela 3**). Em relação aos *contigs* virais, foram registrados 2.884 no total. A biblioteca com a menor quantidade de *contigs* foi a da APA, com 12 *contigs* e a maior foi atribuída a biblioteca de fêmeas de Campos, com 1.473 *contigs* (**Apêndice D**). Algo esperado quando se considera que quanto maior o número de leituras maiores chances temos de que esses formem *contigs*.

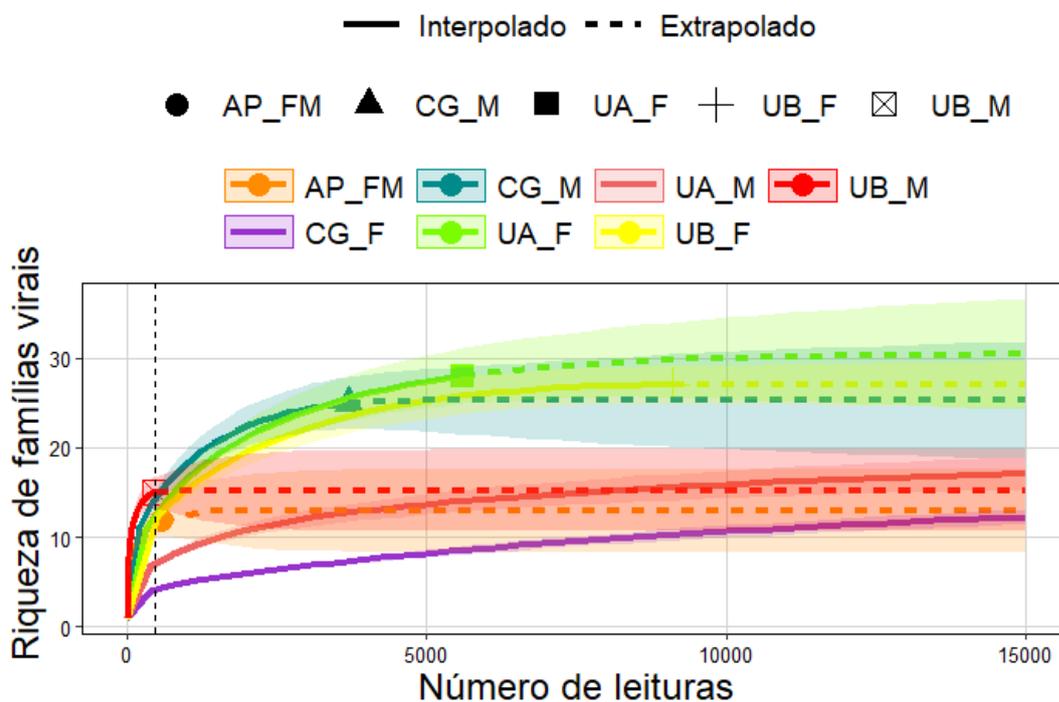
Ao longo de todo o estudo foram identificadas 50 famílias virais distintas usando a plataforma *Kraken2*, sendo 23 identificadas também pela plataforma *Diamond* (**Tabela 3**). A maior riqueza foi identificada para a biblioteca de fêmeas da União A (28 famílias virais) e a menor para biblioteca da APA (12 famílias virais).

Tabela 3 - Resumo dos resultados obtidos no *Kraken2*, para todas as bibliotecas analisadas. Entre parênteses está a porcentagem que as leituras virais representam dentro do total de leituras obtidos para cada biblioteca. Os resultados “pós confirmação” se referem àquelas famílias virais que foram identificadas pelo *Kraken2* que também estavam presentes no *Diamond*.

Sexo	União A		União B		APA	Campos	
	F	M	F	M	F/M	F	M
Leituras brutas	9.705.510	10.142.844	6.982.452	10.503.542	2.898.848	3.469.428	3.671.970
Leituras pós filtro	7.582.603	7.724.705	4.785.645	7.770.762	2.415.281	2.688.760	3.255.391
Aproveitamento	78,1%	76,1%	68,5%	74,0%	83,3%	75,5%	88,6%
Leituras virais (%)	5.601 (0,07)	23.056 (0,3)	9.102 (0,19)	466 (0,006)	566 (0,02)	150.800 (5,6)	3.697 (0,11)
Leituras virais (%) - pós confirmação	5.503 (0,07)	22.818 (0,3)	8.635 (0,11)	299 (0,003)	161 (0,006)	150.702 (5,6)	3.559 (0,10)
<i>Contigs</i> virais	273	497	434	13	12	1.473	182
Riq. famílias virais	28	18	27	15	12	21	25
Riq. famílias virais - pós confirmação	12	7	14	5	4	8	14
Total de famílias identificadas	50	-	-	-	-	-	-
Total de famílias - pós confirmação	23	-	-	-	-	-	-

A partir da curva de rarefação (**Figura 10**) consideramos que as bibliotecas foram bem amostradas, com a maioria das bibliotecas apresentando curvas que estavam próximas de atingir a assíntota.

Figura 10 - Curva de rarefação baseada em indivíduos para bibliotecas analisadas: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. A linha contínua indica os valores obtidos enquanto a linha tracejada valores extrapolados, demonstrando o número de leituras virais necessário para atingir a assíntota naquelas bibliotecas.



5.1 Riqueza e composição de famílias virais das áreas

As bibliotecas da ReBIO União apresentaram 43 das 50 famílias virais identificadas pelo *Kraken2*. A maior riqueza foi encontrada na biblioteca de fêmeas da União A (28) e o menor para a biblioteca de machos da União B (15). Nas fêmeas da União A, das famílias identificadas pelo *Kraken2*, 12 também foram identificadas pelo *Diamond* (Tabela 4). O maior número de leituras nessa biblioteca foi da família Drexlerviridae (75,5%).

A biblioteca de machos da União A apresentou riqueza de 18 famílias identificadas pelo *Kraken2*, sendo sete famílias virais em comum com o *Diamond*. A família Straboviridae concentrou o maior número de leituras (91,6%). Na biblioteca de fêmeas da União B, a riqueza identificada pelo *Kraken2* foi de 27 famílias virais e por ambas as ferramentas de 14, sendo novamente o maior número de leituras para a família Drexlerviridae (29,8%) para essa biblioteca. Destaca-se também a identificação, em ambas as ferramentas, de 203 leituras da família Circoviridae, que abriga diversos vírus causadores de doenças em vertebrados. A biblioteca de machos da União B apresentou maior incongruência na riqueza entre as

plataformas, variando de 15 famílias virais identificadas pelo *Kraken2* para cinco identificadas por ambas as plataformas. Nesta biblioteca, a família com maior número de leituras (62,5%) também foi a família Drexlerviridae.

Na APA tivemos uma biblioteca composta por amostras de hospedeiros de ambos os sexos, mas esta apresentou poucas leituras totais e baixa riqueza, quando comparada com as outras. Aqui 12 famílias virais foram identificadas pelo *Kraken2* e apenas quatro foram em comum com o Diamond. A família Suoliviridae acumulou o maior número de leituras (48,4%).

Juntas as bibliotecas de Campos somaram 32 famílias virais diferentes, das 50 identificadas pelo *Kraken2*. A biblioteca de fêmeas obteve o maior número de leituras entre todas as bibliotecas analisadas nesse estudo, tanto pelo *Kraken2* individualmente (150.800) quanto por ambas as ferramentas (150.702). Sendo que 134.967 leituras (89,5%) estão concentradas em uma única família de bacteriófagos (Straboviridae). Os machos de Campos tiveram sua riqueza variando entre 25 famílias identificadas pelo *Kraken2* e 14 famílias virais identificadas por ambas as ferramentas. A família viral com maior número de leituras foi Straboviridae (57,6%). Destaca-se a presença da família Parvoviridae em ambas as ferramentas, com 420 leituras, que é notória por abrigar vírus que causam doenças em vertebrados.

Tabela 4 - Famílias virais identificadas pelo *Kraken2* e número de leituras atribuídas a cada uma nas bibliotecas analisadas (UA_F - Fêmeas da ReBIO União A; UA_M - Machos da ReBIO União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM – APA; CG_F – Fêmeas de Campos dos Goytacazes; CG_M – Machos de Campos dos Goytacazes). A coluna “*Diamond*” indica se aquela família viral também foi identificada (S) ou não (N) na ferramenta *Diamond*. Os hospedeiros de cada família viral podem ser encontrados no Apêndice A. Marcadas com * estão as famílias virais conhecidas por terem vírus que infectam de vertebrados.

Família viral	UA_F		UA_M		UB_F		UB_M		AP_FM		CG_F		CG_M	
	<i>Kraken2</i>	<i>Diamond</i>												
Ackermannviridae	772	S	0	N	8	S	0	N	0	N	362	S	0	N
Adenoviridae*	0	N	2	N	0	N	0	N	0	N	2	N	0	N
Ascoviridae	1	N	0	N	0	N	0	N	0	N	0	N	0	N
Asfarviridae*	0	N	0	N	3	N	0	N	0	N	0	N	0	N
Autographiviridae	94	S	503	S	1764	S	56	S	0	N	19	S	80	S
Baculoviridae	2	N	8	N	2	N	0	N	10	N	3	N	11	N
Casjensviridae	9	S	7	N	3	N	0	N	1	N	3	N	8	S
Chaseviridae	0	N	0	N	2	N	2	N	0	N	14	N	0	N
Circoviridae*	6	S	0	N	203	S	2	S	0	N	0	N	0	N
Coronaviridae*	11	N	22	N	0	N	14	N	36	S	23	N	4	N
Demereciviridae	0	N	0	N	1157	S	0	N	0	N	12712	S	0	N
Dicistroviridae	0	N	0	N	5	S	0	N	0	N	10	S	0	N
Drexleriviridae	4156	S	934	S	2573	S	187	S	0	N	2586	S	504	S
Fiersviridae	0	N	2	N	0	N	0	N	0	N	0	N	0	N
Fimoviridae*	3	N	0	N	0	N	0	N	0	N	0	N	0	N
Flaviviridae*	0	N	0	N	0	N	0	N	0	N	0	N	2	N
Genomoviridae*	0	N	0	N	0	N	0	N	0	N	3	N	2	S
Guelinviridae	0	N	0	N	6	S	15	S	0	N	0	N	0	N
Herelleviridae	9	S	22	S	40	S	39	S	45	S	2	S	6	S

Família viral	UA_F		UA_M		UB_F		UB_M		AP_FM		CG_F		CG_M	
	Kraken2	Diamond												
Herpesviridae*	4	N	2	N	0	N	4	N	9	N	4	N	2	N
Hytrosaviridae	0	N	0	N	2	N	0	N	0	N	0	N	0	N
Iflaviridae	2	N	0	N	0	N	0	N	0	N	0	N	0	N
Inoviridae	2	S	0	N	0	N	0	N	0	N	0	N	0	N
Intestiviridae	2	N	2	N	4	N	6	N	23	N	14	N	39	S
Iridoviridae*	0	N	0	N	0	N	0	N	0	N	0	N	2	N
Kyanoviridae	3	N	0	N	7	N	0	N	0	N	0	N	5	N
Mesyanzhinovviridae	0	N	3	N	0	N	0	N	0	N	0	N	2	S
Microviridae	125	S	1	S	3	S	0	N	2	S	0	N	55	S
Mimiviridae	3	N	0	N	5	N	0	N	0	N	4	N	1	N
Nudiviridae	2	N	0	N	0	N	0	N	0	N	0	N	0	N
Pachyviridae	0	N	0	N	2	N	0	N	0	N	0	N	0	N
Paramyxoviridae*	1	N	0	N	0	N	0	N	0	N	0	N	0	N
Parvoviridae*	0	N	0	N	0	N	0	N	0	N	0	N	420	S
Peduoviridae	0	N	17	N	0	N	8	N	344	N	0	N	0	N
Phenuiviridae*	1	N	0	N	0	N	0	N	0	N	0	N	0	N
Phycodnaviridae	3	N	0	N	3	N	0	N	0	N	1	N	0	N
Picornaviridae*	0	N	0	N	0	N	0	N	0	N	0	N	4	S
Potyviridae	5	N	0	N	0	N	0	N	0	N	0	N	0	N
Poxviridae*	1	N	0	N	0	N	1	N	1	N	0	N	2	N
Retroviridae*	0	N	6	S	0	N	0	N	2	N	0	N	2	N
Roniviridae	0	N	0	N	0	N	0	N	0	N	2	N	0	N
Rountreeviridae	0	N	0	N	2	N	37	N	0	N	0	N	0	N

Família viral	UA_F		UA_M		UB_F		UB_M		AP_FM		CG_F		CG_M	
	<i>Kraken2</i>	<i>Diamond</i>												
Salasmaviridae	0	N	0	N	4	S	0	N	0	N	0	N	0	N
Schitoviridae	10	S	35	N	14	S	0	N	0	N	3	N	12	S
Smacoviridae	1	S	0	N	48	S	0	N	0	N	0	N	3	S
Solinviridae	0	N	0	N	0	N	0	N	0	N	0	N	1	N
Steigviridae	54	N	138	N	430	N	72	N	15	N	22	N	106	N
Straboviridae	206	S	20913	S	356	S	0	N	0	N	134967	S	2049	S
Suoliviridae	113	S	439	S	2454	S	20	N	78	S	44	S	375	S
Xinmoviridae	0	N	0	N	2	N	3	N	0	N	0	N	0	N

Quando observamos a riqueza compartilhada e exclusiva entre as áreas vemos que a área com maior número de famílias exclusivas foi a União A (nove famílias exclusivas) (**Figura 11**). Essa é seguida pela União B e por Campos, ambas com sete famílias exclusivas. A área da APA não apresentou nenhuma família viral que ocorria apenas naquela área. As famílias exclusivas da UA foram: Ascoviridae, Fiersviridae, Fimoviridae, Iflaviviridae, Inoviridae, Nudiviridae, Paramyxoviridae, Phenuiviridae e Potyviridae.

A área UB teve como famílias exclusivas: Asfarviridae, Guelinviridae, Hytrosaviridae, Pachyviridae, Rountreeviridae, Salasmaviridae e Xinmoviridae. Em Campos as famílias exclusivas foram: Flaviviridae, Genomoviridae, Iridoviridae, Parvoviridae, Picornaviridae, Roniviridae e Solinviridae.

Figura 11 - Riqueza de famílias virais compartilhadas e exclusivas entre as áreas analisadas. Na diagonal, em verde, temos o número de famílias virais exclusivas de cada área. Na parte superior direita da matriz temos o número de famílias virais compartilhadas entre cada par de área analisado.

	UA	UB	APA	Campos
UA	9	21	12	22
UB		7	11	22
APA			0	11
Campos				7

Nas comparações par a par vemos que as áreas com mais famílias virais compartilhadas são Campos e União A (22 famílias). O mesmo acontece entre Campos e a União B. O terceiro par com maior número de famílias compartilhadas é entre as áreas da União, que compartilham 21 famílias virais. A área da APA

compartilha 12 famílias virais com a União A, e 11 com as outras áreas. Isso significa que todas as famílias presentes na APA também estão presentes na União A e que esta área difere das outras apenas por uma família viral.

Usando os dados das famílias virais que foram encontradas em ambas as ferramentas vimos que a área com maior número de famílias exclusivas, com cinco famílias ocorrendo apenas naquela área, foi Campos, em vez de a União A, como vimos anteriormente (**Figura 12**). Em seguida tivemos as áreas da União com duas famílias exclusivas cada. Também diferindo do que vimos anteriormente, a APA, que não tinha nenhuma família exclusiva, agora possui uma família que ocorre apenas nessa área. Antes a APA era um subconjunto de quase todas as bibliotecas. Entre todas as 23 famílias identificadas em ambas as ferramentas tivemos apenas três que ocorreram em todas as áreas.

Aqui as famílias exclusivas da área UA foram: Inoviridae e Retroviridae. Na União B tivemos Guelinviridae e Salasmaviridae. Para a APA a família exclusiva foi Coronaviridae. Em Campos as famílias exclusivas foram: Genomoviridae, Intestiviridae, Mesyanhinoviridae, Parvoviridae e Picornaviridae.

Figura 12 - Riqueza de famílias virais compartilhadas e exclusivas entre as áreas analisadas, com as famílias encontradas em ambas as ferramentas (*Kraken2* e *Diamond*). Na diagonal, em verde, temos o número de famílias virais exclusivas de cada área. Na parte superior direita da matriz temos o número de famílias virais compartilhadas entre cada par de área analisado.

	UA	UB	APA	Campos
UA	2	10	3	10
UB		2	3	11
APA			1	3
Campos				5

Quando comparamos par a par vemos que agora as áreas que compartilham um maior número de famílias virais são Campos e a União B, seguidas por Campos e União A. Isso difere do que vimos anteriormente em que Campos compartilhava o mesmo número de famílias virais com as outras duas áreas. Mantivemos que o terceiro par com maior número de famílias compartilhadas foi aquele entre as áreas da União. A APA compartilhou três famílias virais com todas as áreas o que significa que, neste caso, ela difere de todas as áreas em apenas uma família. Já havíamos visto algo parecido anteriormente onde a APA foi um subconjunto de quase todas as outras áreas.

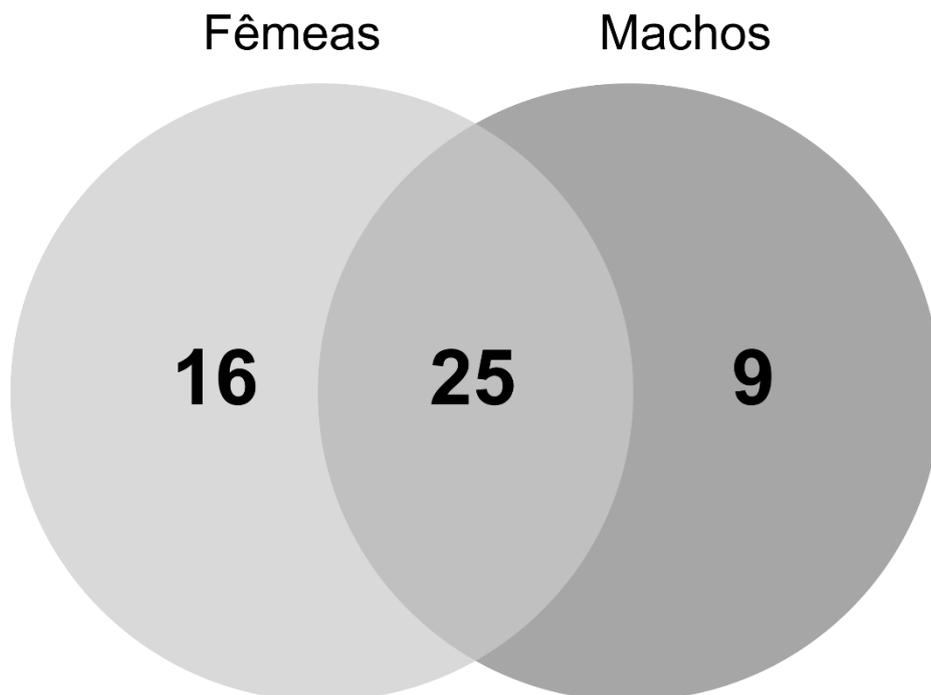
5.2 Riqueza e composição de famílias virais entre os sexos

Das sete bibliotecas que foram feitas seis foram divididas entre os sexos. No geral as fêmeas tiveram uma riqueza de 41 famílias virais, com 16 sendo encontradas em ambas as ferramentas. Os machos tiveram uma riqueza de 34

famílias virais e 17 dessas foram encontradas em ambas as ferramentas. Usando os dados completos do *Kraken2*, observamos que metade das 50 famílias identificadas são compartilhadas entre os sexos (**Figura 13**). Fêmeas de *D. aurita* possuíram mais famílias virais exclusivas (32%), sendo esse quase o dobro do número de famílias identificadas apenas nos machos (18%). Logo, fêmeas de *D. aurita* apresentaram uma riqueza de famílias virais maior do que machos.

Entre as fêmeas a família viral com maior número de leituras foi Straboviridae (135.529/82,2%), a mesma que teve mais de 130 mil leituras nas fêmeas de Campos, e, por isso, também teve muitas leituras nesta localidade. Para os machos a família Straboviridae também foi a que obteve o maior número de leituras com 22.962 (86,0%).

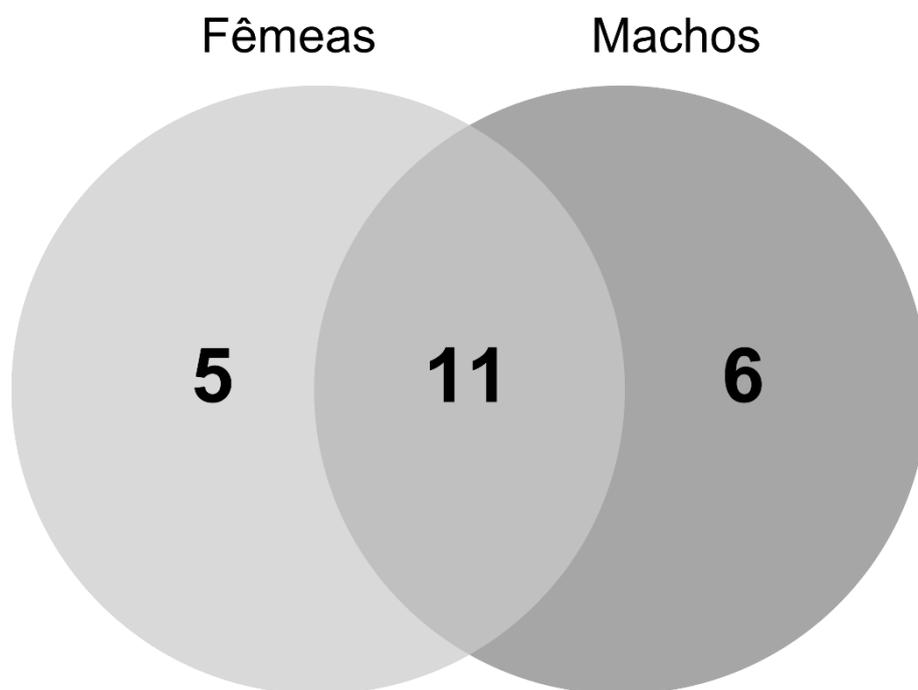
Figura 13 – Quantidade de famílias virais exclusivas de cada sexo, assim como as compartilhadas. Na esquerda temos a quantidade de famílias exclusivas das fêmeas e na direita as exclusivas dos machos. A intercessão demonstra a quantidade de famílias compartilhadas entre os dois sexos.



O padrão de variação na composição de famílias virais, considerando apenas aquelas identificadas por ambas as ferramentas, mantém parte do padrão geral observado para aquelas identificadas pelo *Kraken2*. A maior parte (11 famílias virais) permaneceu sendo compartilhada entre machos e fêmeas (**Figura 14**). Seis foram exclusivas de machos e cinco de fêmeas. A confirmação das famílias no *Diamond*

fez com que machos tivessem mais famílias exclusivas do que fêmeas. Anteriormente, havíamos visto o contrário, com fêmeas tendo quase o dobro de famílias exclusivas, quando comparadas com os machos. Nesta comparação o total é 22 em vez de 23 famílias, pois essa comparação descartou a APA (por não ser separada por sexo) que, com os dados confirmados do *Diamond*, tinha uma família viral exclusiva.

Figura 14 - Quantidade de famílias virais exclusivas de cada sexo, assim como as compartilhadas. Na esquerda temos a quantidade de famílias exclusivas das fêmeas e na direita as exclusivas dos machos. A intercessão demonstra a quantidade de famílias compartilhadas entre os dois sexos. Aqui usamos os dados referentes as famílias virais confirmadas pelo *Diamond*.



5.3 Detalhamento das bibliotecas separadas por sexo

No padrão geral para o sexo, observamos que as fêmeas têm uma riqueza maior que os machos, exceto quando consideramos os dados de ambas as ferramentas (*Kraken2* e *Diamond*). Quando avaliamos cada um dos locais separados por sexo com mais detalhes, observamos um padrão similar. Nas áreas da União, as fêmeas tiveram uma riqueza maior que os machos, algo que se manteve para ambas as ferramentas de identificação taxonômica. Na União A os machos e as fêmeas compartilharam 13 famílias virais, fugindo do padrão de que a maior parte das famílias identificadas seria compartilhada entre os sexos. Com os

dados das famílias encontradas em ambas as ferramentas esse compartilhamento foi de seis famílias virais, representando a maior parte das famílias identificadas entre essas bibliotecas.

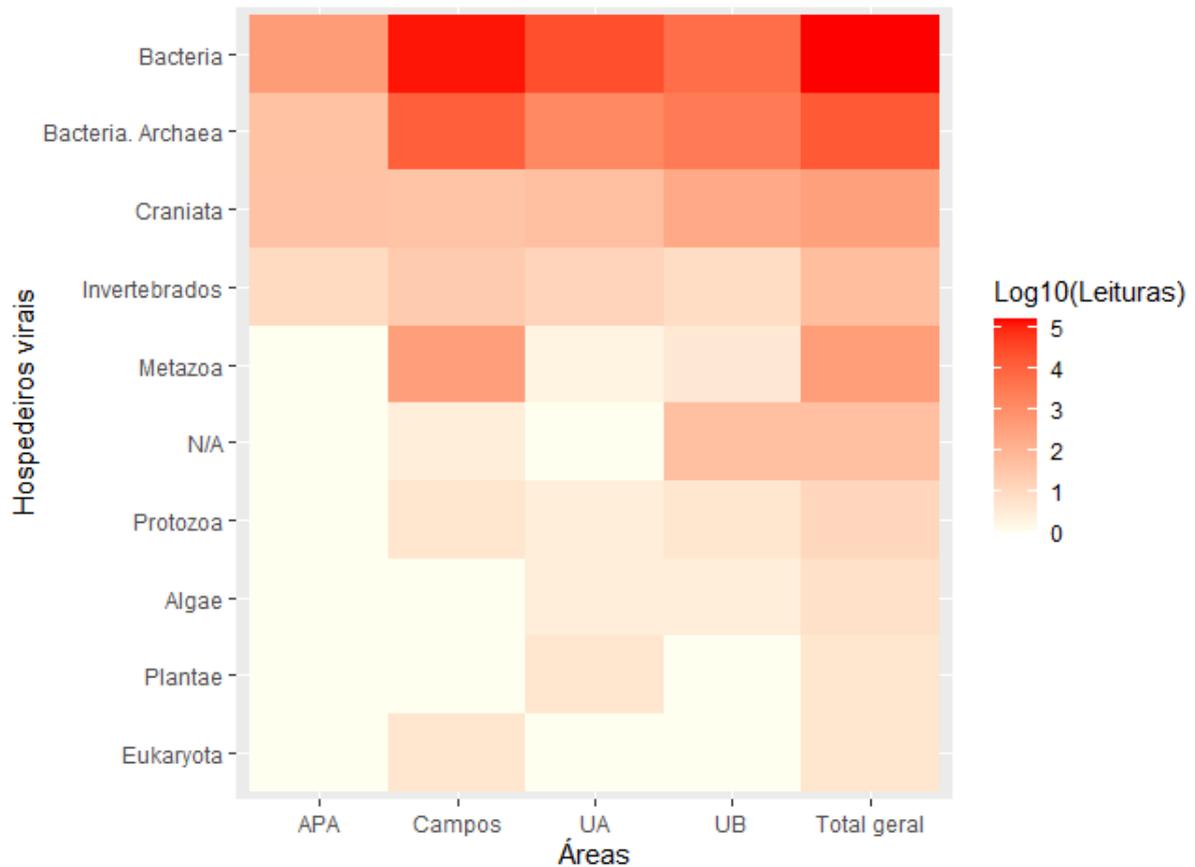
Na União B, machos e fêmeas compartilharam 11 famílias. Nesta área apenas os machos tiveram grande parte das suas famílias sendo compartilhadas. Considerando as famílias encontradas em ambas as ferramentas esse compartilhamento foi de cinco famílias. Nesse caso, todas as famílias encontradas nos machos da União B também foi encontrada nas fêmeas.

Os resultados das localidades de Campos fugiram do padrão visto para os dados do *Kraken2* e se assemelharam um pouco mais ao padrão dos dados das famílias encontradas em ambas as ferramentas. As fêmeas tiveram uma riqueza menor do que os machos tanto nos dados apenas do *Kraken2* quanto nas famílias identificadas em ambas as ferramentas. Essas bibliotecas compartilharam 14 famílias. Neste caso as bibliotecas seguiram o padrão visto anteriormente de a maior parte das famílias serem compartilhadas em vez de exclusivas. Para as famílias encontradas em ambas as ferramentas o compartilhamento foi de cinco famílias e apenas as fêmeas tiveram grande parte da sua riqueza compartilhada.

5.4 Composição de famílias e tipos virais entre áreas

As famílias virais foram classificadas com base nos hospedeiros conhecidos por abrigá-las, o que foi denominado tipo viral (**Apêndice A**). As 50 famílias virais identificadas pelo *Kraken2* no presente estudo se dividiram entre 10 tipos virais. A frequência dos tipos virais entre as áreas pode ser visualizada na figura 15. A tendência observada foi de a maioria das famílias identificadas serem infectantes de bactérias. Em todas as áreas tivemos que vírus infectantes de bactérias foram os mais prevalentes, estes se mostraram dominantes no viroma gastrointestinal dos gambás. Além dos vírus infectantes de bactérias, também observamos vírus que infectam animais, porém, é notável que esses apresentaram menor prevalência, pois a cor é claramente menos quente do que a cor que representa os bacteriófagos.

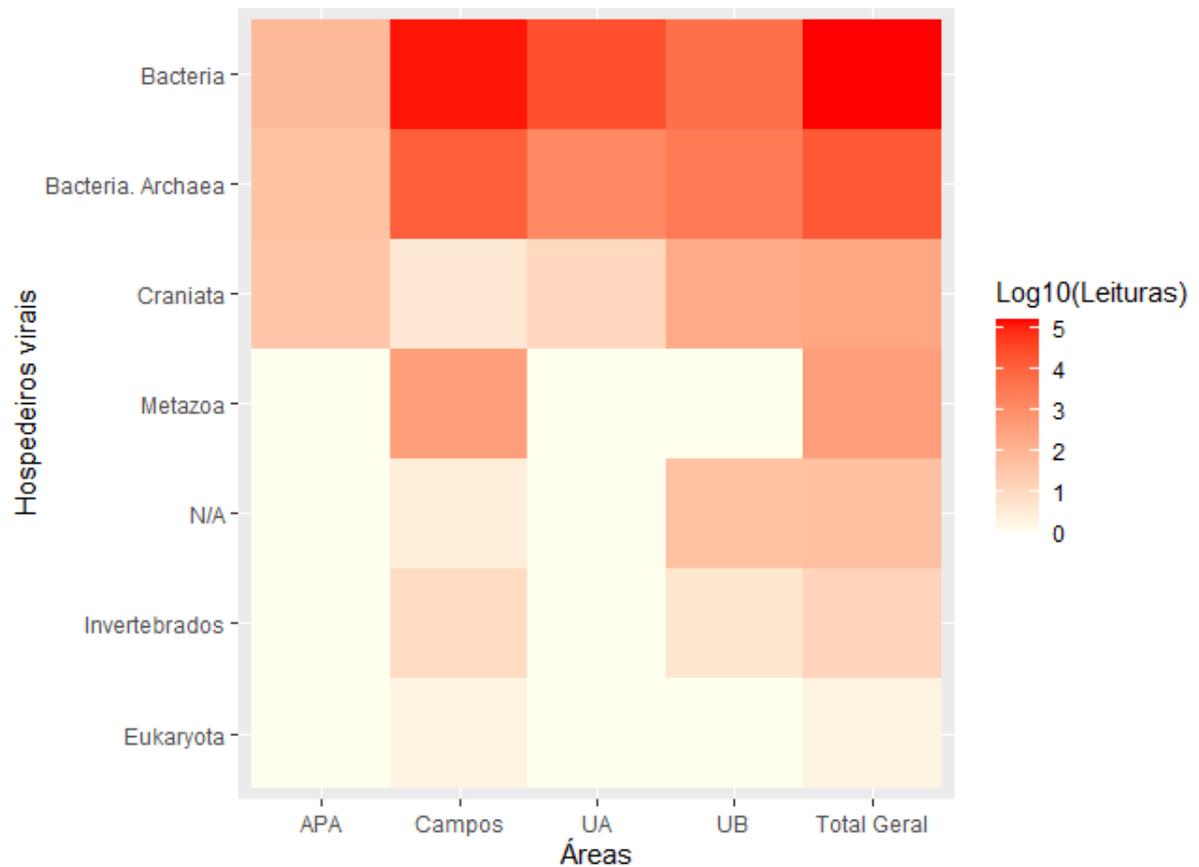
Figura 15 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada área estudada. Log10(Leituras) representa o número de leituras logaritmizado na base 10.



Nota-se que a APA é a área que apresentou a menor diversidade de tipos virais, com quatro de 10 tipos virais, sendo esses os vírus que infectam bactérias, aqueles que infectam bactérias e arqueas, os que infectam Craniata e os que infectam invertebrados. Todas as outras áreas apresentaram oito de 10 tipos virais identificados, mas não os mesmos. Em Campos, dos 10 tipos virais não encontramos aqueles que infectam algas e plantas. Para a União A não tivemos vírus que infectam eucariotos, nem aqueles que não tem seu hospedeiro definido. E na União B não tivemos vírus que infectam eucariotos nem plantas.

Para os dados referentes as famílias que foram identificadas em ambas as ferramentas continuamos observando uma maior prevalência de vírus bacteriófagos em todas as áreas. Em vez de 10 tipos virais, nesta comparação, tivemos sete tipo virais no total, com três tipos virais ausentes (Algae, Plantae e Protozoa) (**Figura 16**).

Figura 16 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada área, com os dados das famílias encontradas em ambas as ferramentas (*Kraken2* e *Diamond*). $\text{Log}_{10}(\text{Leituras})$ representa o número de leituras logaritmizado na base 10.

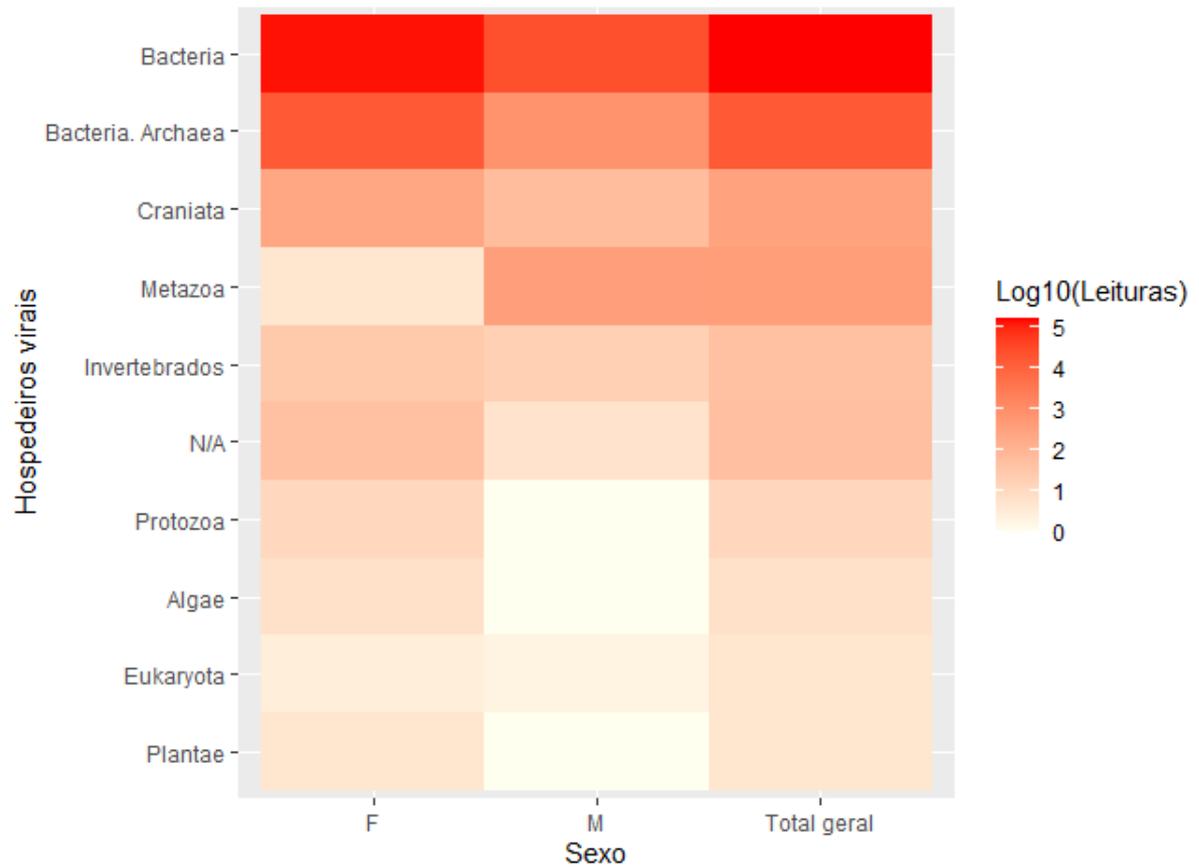


As áreas com menor diversidade de tipos virais foram a APA e a União A, com três tipos virais cada (Bacteria; Bacteria, Archea; e Craniata), diferente do que vimos anteriormente em que a APA tinha consideravelmente menos tipos virais do que o resto das áreas. Campos foi a área com a maior quantidade de tipos virais, tendo todos os sete tipos diferentes aqui identificados. A União B teve cinco dos sete tipos virais, não apresentando os vírus infectantes de metazoários e eucariotos.

5.5 Composição de famílias e tipos virais entre sexos

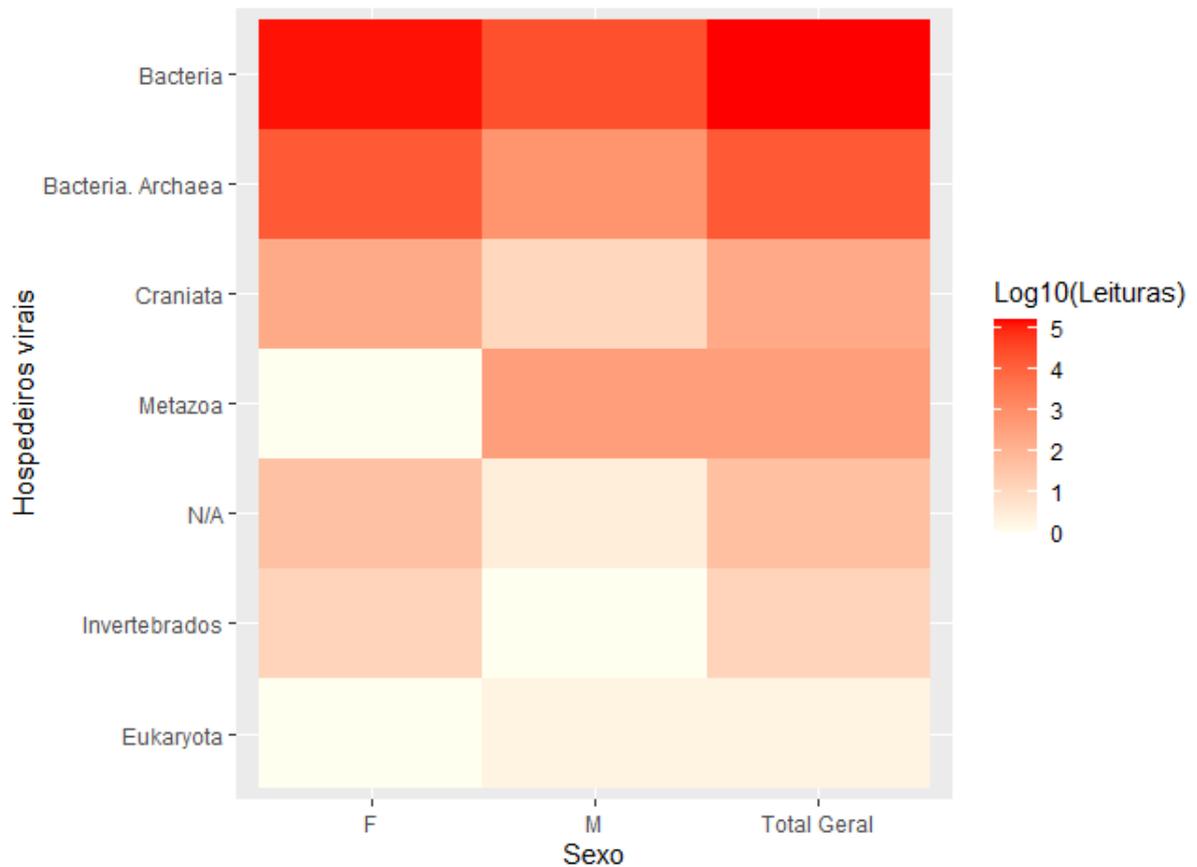
Ao comparar a prevalência de tipos virais entre sexos continuamos observando a maior prevalência vírus bacteriófagos (**Figura 17**), os outros tipos virais apresentaram cores menos quentes e, conseqüentemente, menor prevalência. O número de tipos virais dos machos (sete de 10) tendeu a ser menor do que as fêmeas (10 de 10). Os machos não tiveram vírus que infectam protozoários, plantas e algas, diferente das fêmeas.

Figura 17 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada sexo. Log10 (Leituras) representa o número de leituras logaritimizado na base 10.



O padrão de uma prevalência de vírus bacteriófagos continuou sendo visto para os dados considerando as famílias encontradas em ambas as ferramentas (**Figura 18**). As fêmeas (cinco de sete tipos virais), assim como na riqueza de famílias exclusivas e compartilhadas, apresentaram menor diversidade de tipos virais do que os machos (seis de sete tipos virais). Os mesmos três tipos virais que deixaram de ocorrer na comparação entre as áreas (Algae, Plantae e Protozoa) também não ocorreram aqui então esse desaparecimento, provavelmente, é um resultado do tratamento dos dados e não de um padrão relacionado ao sexo.

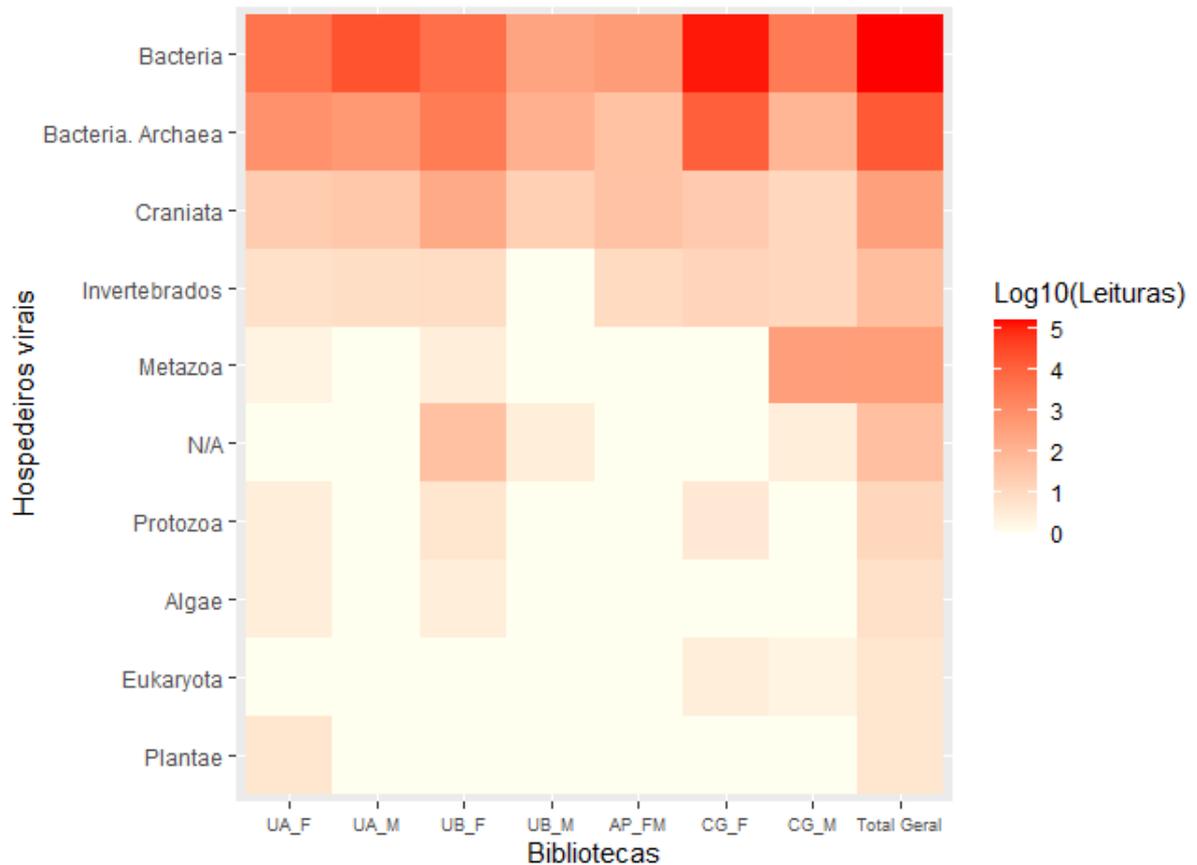
Figura 18 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada sexo, com os dados das famílias encontradas em ambas as ferramentas (*Kraken2* e *Diamond*). $\text{Log}_{10}(\text{Leituras})$ representa o número de leituras logaritmizado na base 10.



5.6 Detalhamento dos tipos virais das bibliotecas

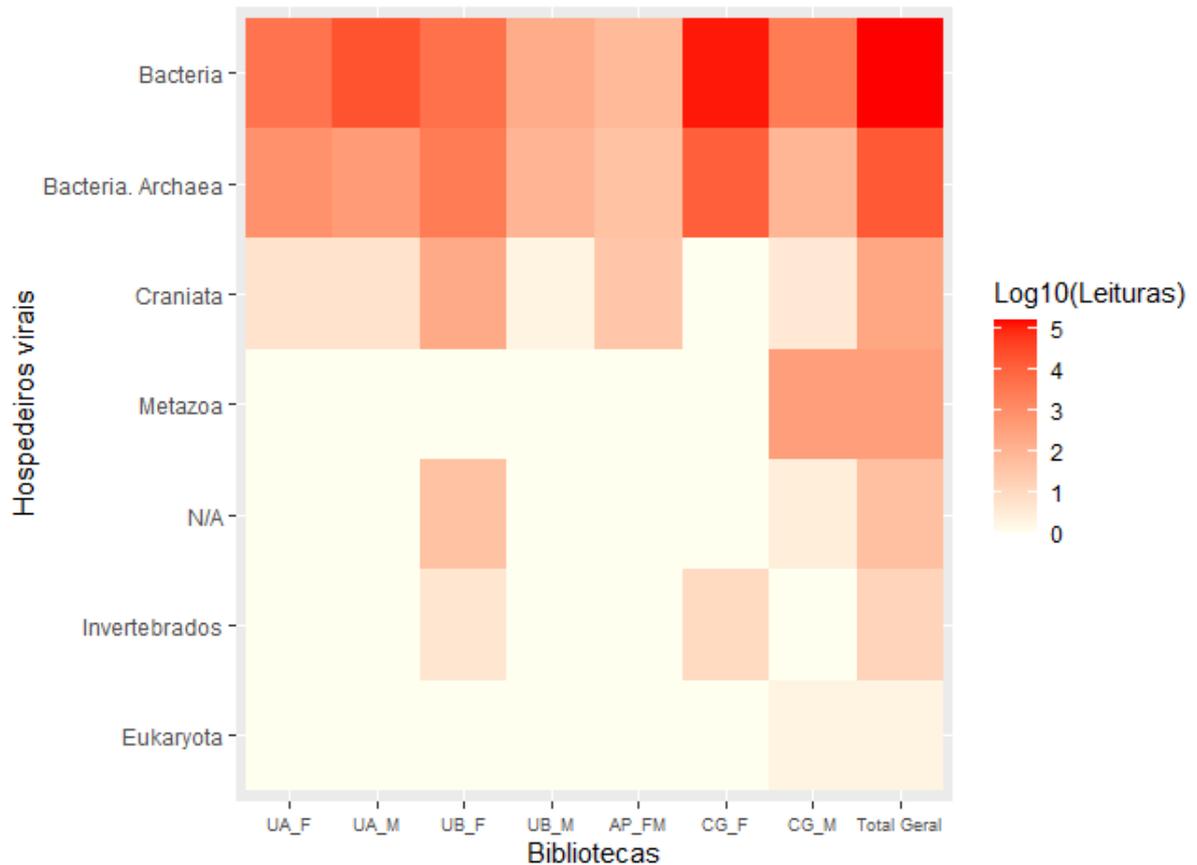
O número de tipos virais variou de três a oito entre as bibliotecas (**Figura 19**). Os menores valores de diversidade de tipos virais foram identificados para a biblioteca da APA com três tipos virais e as bibliotecas de machos da União, ambas com quatro tipos virais cada. Por outro lado, as bibliotecas de fêmeas da União A e de fêmeas da União B apresentaram a maior diversidade de tipos virais, ambas com oito tipos virais diferentes, sendo seguidas pela biblioteca de fêmeas de Campos, com seis tipos virais distintos. No geral, conseguimos observar que, independentemente da biblioteca, o tipo viral prevalente novamente foram os bacteriófagos. Famílias que infectam outros hospedeiros, como vertebrados, também estavam presentes, porém em menores quantidades.

Figura 19 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada biblioteca: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. Log10(Leituras) representa o número de leituras logaritmizado na base 10.



Com os dados das famílias que também foram identificadas pelo *Diamond* observamos que a prevalência se manteve nos bacteriófagos, assim como em todas as outras comparações e explorações que fizemos dos dados (**Figura 20**). Aqui, a maioria das bibliotecas apresentou a mesma quantidade de tipos virais, três dos sete tipos identificados. Apenas as bibliotecas de fêmeas da União B (cinco de sete) e de machos de Campos (seis de sete) apresentaram maiores quantidades de tipos virais, com os machos de Campos tendo quase todos os tipos presentes.

Figura 20 - Prevalência de famílias virais, de acordo com o tipo de organismo que essas infectam, para cada biblioteca: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes, considerando as famílias confirmadas no *Diamond*.
Log10(Leituras) representa o número de leituras logaritmizado na base 10.

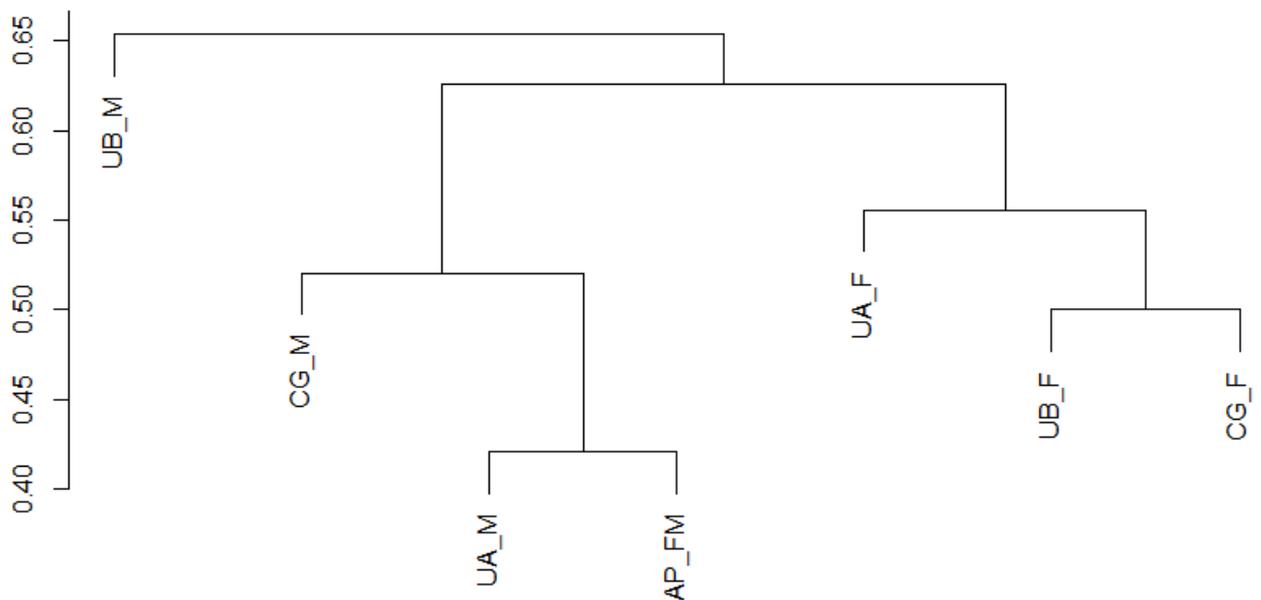


As fêmeas da União A, os machos da mesma área, os machos da União B e a APA todos tiveram os mesmos tipos virais ocorrendo (Bacteria; Bacteria, Archea e Craniata), reforçando a grande prevalência de vírus bacteriófagos. As fêmeas de Campos apesar de também terem apresentado três tipos virais esses foram: Bacteria; Bacteria, Archea e vírus que infectam vertebrados. Nas fêmeas da União B os tipos virais ausentes foram os que infectam metazoários e os que infectam eucariotos. Para os machos de Campos apenas não identificamos vírus que infectam invertebrados.

5.7 Jaccard e Similaridade das Bibliotecas

O índice de similaridade de Jaccard aplicado sobre a composição de famílias identificadas pelo *Kraken2* indicou um agrupamento das bibliotecas por sexo e não pela área, sugerindo o grau de intervenção antrópica nas áreas é menos importante do que o sexo do hospedeiro para determinar a composição viral (**Figura 21**). É possível identificar um grupo formado pelas bibliotecas de fêmeas de Campos e das duas áreas da União. O segundo grupo é formado pelas bibliotecas de machos de Campos, da União A e pela biblioteca da APA que incluiu machos e fêmeas. A biblioteca da APA foi formada majoritariamente por amostras de machos (seis das oito amostras), o que pode explicar seu agrupamento. Apenas a biblioteca de machos da União B ficou isolada de todas as outras bibliotecas.

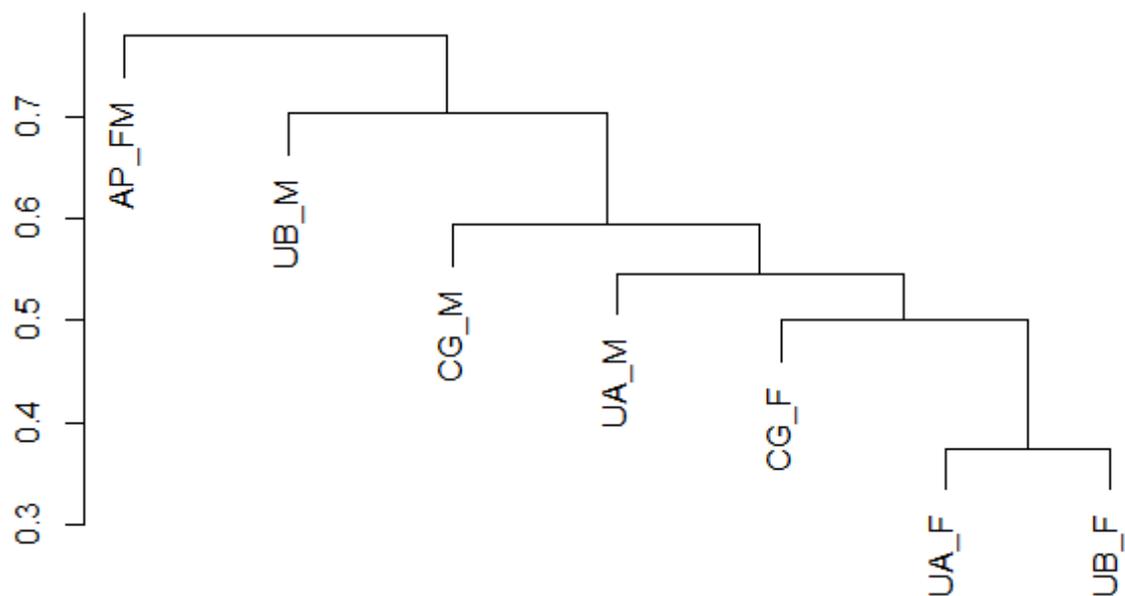
Figura 21 - Dendrograma feito a partir do índice de dissimilaridade de Jaccard entre as bibliotecas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes), baseado na presença e ausência de famílias virais em cada uma das bibliotecas.



Para os dados referentes as famílias identificadas em ambas as ferramentas, observamos que os resultados e os padrões vistos anteriormente se modificaram. O agrupamento por sexo visto nos dados “brutos” do *Kraken2* ficou menos claro (**Figura 22**). Ainda houve a formação de um grupo das bibliotecas de fêmeas, mas

as bibliotecas de machos não formaram um grupo único e ficaram como grupos externos ao grupo de fêmeas. Assim, as fêmeas parecem ter uma composição do viroma mais parecida entre si do que entre os machos. Esse resultado pode estar ligado a menor diversidade viral identificada nos machos, que dificulta a formação de um grupo coeso entre eles, principalmente após a exclusão de parte das famílias quando trabalhamos apenas com aquelas identificadas por ambas as ferramentas.

Figura 22 - Dendrograma feito a partir do índice de dissimilaridade de Jaccard entre as bibliotecas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes), baseado na presença e ausência de famílias virais em cada uma das bibliotecas, usando os dados das famílias encontradas em ambas as ferramentas (*Kraken2* e *Diamond*).



Elementos de estrutura de metacomunidade

A análise dos elementos de estrutura de metacomunidade identificou uma coerência positiva, com substituição de espécies positivas não significativa e índice de Morisita superior a um, significativo para a análise baseada na matriz de famílias do *Kraken2* e não significativo para aquela baseada na matriz confirmada pelo *Diamond* (**Tabela 5**). Isso indica uma estrutura de metacomunidade quasi-clementisiana para a matriz do *Kraken2* (**Figura 23**) e uma estrutura quase-geasoniana para a matriz confirmada pelo *Diamond*. Uma estrutura de metacomunidades quasi-Clementisiana sugere que as famílias virais estão ordenadas de forma a gerar grupos distintos ao longo de um determinado gradiente,

embora a substituição de famílias ao longo deste gradiente não seja tão forte quanto em uma estrutura Clementsiana. Nessa estrutura nossos dados aparentam se organizar em um gradiente determinado pelo sexo dos indivíduos. Na estrutura quasi-gleasoniana, existe o mesmo problema da menor distinção da substituição de famílias ao longo do gradiente, mas nesse caso, não há formação de grupos ao longo do gradiente, sendo a substituição mais gradual.

Tabela 5 - Análise da estrutura de metacomunidades baseada na distribuição das famílias virais entre as bibliotecas de indivíduos de *Didelphis aurita*. “Aus.” se refere as ausências embutidas, “Sub.” a quantidade de substituição de espécies e “I” é o índice de Morisita. A média se refere ao valor médio obtido nas matrizes simuladas com os dados.

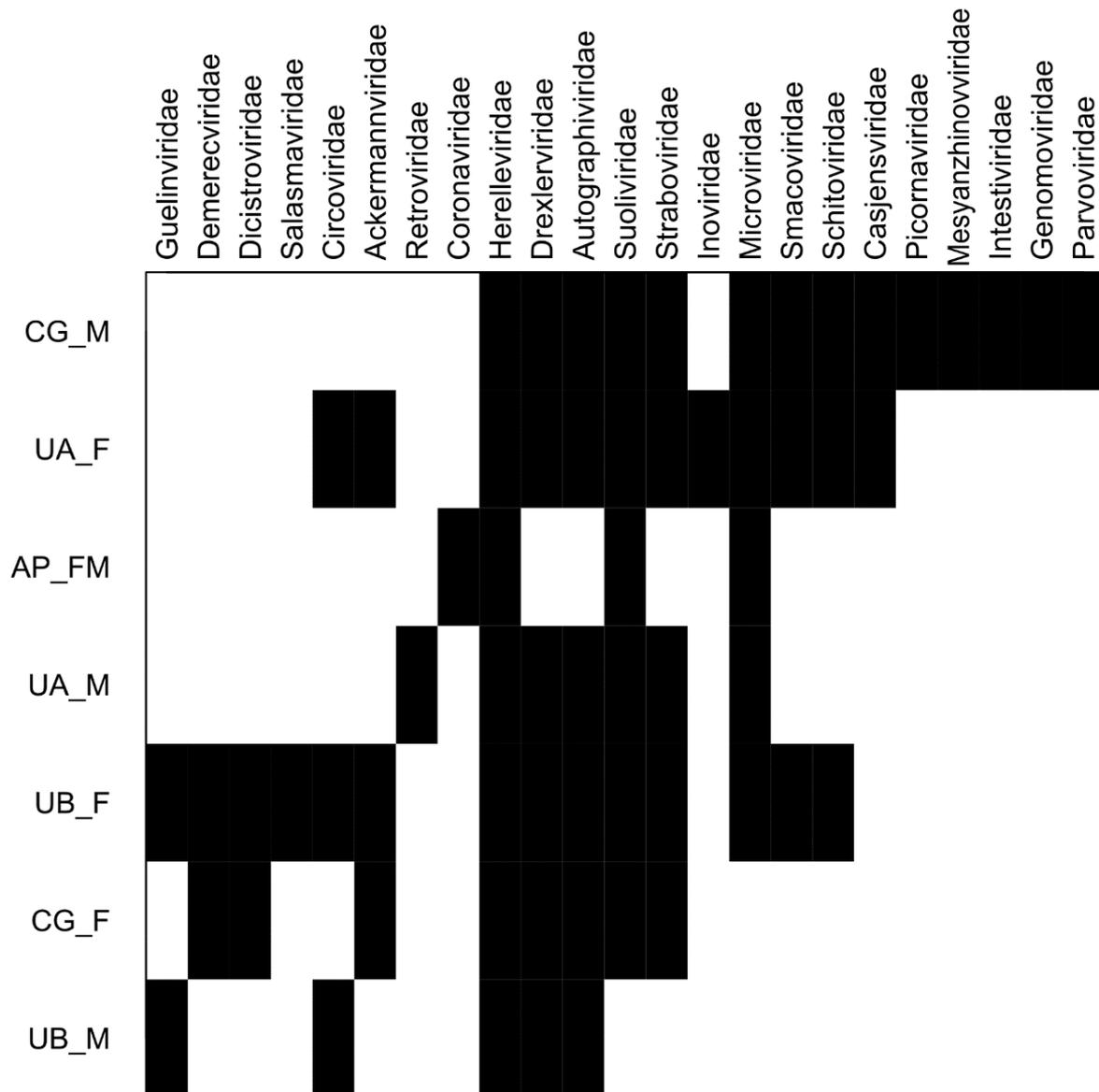
Dados	Coerência			Substituição de espécies			Coincidência de limites		Estrutura de Metacomunidades
	Aus.	P	Média	Sub.	P	Média	I	P	
<i>Kraken2</i>	6,80E+07	<0,001	1,73E+08	1,06E+09	0,07	9,00E+08	1,37	<0,001	Quasi-clementisiana
<i>Kraken2</i> e <i>Diamond</i>	3,10E+07	<0,001	7,00E+07	236	0,97	234,566	1,08	0,22	Quasi-gleasoniana

O gradiente identificado na análise de elementos de estrutura de metacomunidades executada sobre a matriz de famílias identificadas pelo *Kraken2* evidenciou uma variação na composição em função do sexo similar àquela encontrada no dendrograma da seção anterior (**Figura 21**). A ordenação das bibliotecas agrupou os machos na extremidade inferior do gráfico, incluindo a biblioteca da APA e as bibliotecas de fêmeas tenderam a ficar na extremidade superior. Apenas a biblioteca de machos da União B (UB_M) ficou entre as bibliotecas de fêmeas. Pelo gráfico, é possível observar que as famílias virais que a biblioteca UB_M compartilha com as demais bibliotecas de machos também estão presentes nas bibliotecas de fêmeas. No entanto, ela inclui cinco famílias presentes apenas em bibliotecas de fêmeas (Chaseviridae, Circoviridae, Guelinviridae, Rountreeviridae e Xinmoviridae), o que explica sua ordenação entre as fêmeas. Nessa figura também conseguimos observar quais são as famílias que estão presentes em todas as bibliotecas, formando o que podemos chamar de viroma central (do inglês *core virome*). O viroma central seriam as famílias que formam o viroma de todos os gambás analisados, independentemente de qualquer outra variável, sendo elas: Steigviridae, Herelleviridae, Suoliviridae e Intestiviridae.

A ordenação na análise de estrutura de metacomunidades para as famílias identificadas por ambas as plataformas diferiu daquela com todas as famílias

identificadas pelo *Kraken2*. As bibliotecas não se ordenaram por sexo e nem por área (**Figura 24**). Também é possível identificar um viroma central, mas esse não é composto pelas mesmas famílias do anteriormente, sendo que nesse caso a única família presente em todas as bibliotecas foi Herelleviridae, mas três famílias estiveram presentes em seis das sete bibliotecas (Drexelviriidae, Autographiviridae e Suoliviridae).

Figura 24 - Matriz de presença e ausência de famílias virais identificadas pelas ferramentas *Kraken2* e *Diamond* ordenada por média recíproca na análise de elementos de estrutura de metacomunidades. Legenda: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes).



6. DISCUSSÃO

Neste trabalho, descrevemos a diversidade viral do trato gastrointestinal de *Didelphis aurita*. Estudos sobre vírus nessa espécie são escassos, um dos poucos exemplos foi o estudo de Caserta e colaboradores (2023) que também buscou descrever a diversidade viral de *D. aurita*, no estado de São Paulo. Esse estudo identificou famílias virais que também estavam presentes nas nossas bibliotecas,

como Retro, Parvo e Smacoviridae. A família Picornaviridae, que abriga o vírus causador da hepatite, também já foi encontrada em *Didelphis aurita* (Carneiro *et al.*, 2018) e a encontramos nos machos de Campos. A presença dessas famílias em outro estudo fortalece os nossos achados, entretanto, os autores não trazem dados sobre os vírus bacteriófagos. Isso nos faz acreditar que este estudo é o primeiro em que descrevemos de maneira completa o viroma de *D. aurita*. Outros registros de vírus infectando a espécie são de famílias virais não identificadas em nossas bibliotecas. Araújo e colaboradores (2012) encontraram hantavírus (família Hantaviridae) em *D. aurita*, mas também já houve relatos do vírus vaccinia (família Poxviridae) (Peres *et al.*, 2018) infectando a espécie. Apesar de termos encontrado vírus patogênicos nas nossas bibliotecas, até onde se sabe, a maioria dos vírus que identificamos não eram patogênicos para os gambás.

Das famílias que identificamos algumas já foram identificadas em animais domésticos, como Picornaviridae, Herpesviridae, Circoviridae e Genomoviridae, encontradas em viromas gastrointestinais de gatos domésticos (Van Brussel *et al.*, 2022; Zhang *et al.*, 2014). Picornaviridae, Genomoviridae, Circoviridae e Smacoviridae também já foram identificadas no viroma gastrointestinal de cães domésticos (Jiang *et al.*, 2023; Moreno *et al.*, 2017). Dessas cinco famílias já descritas em animais domésticos, quatro ocorreram nos gambás da área urbana de Campos, sendo duas exclusivas das bibliotecas dessa área (Genomoviridae e Picornaviridae). Isso sugere que o contato com animais domésticos pode estar influenciando a composição do viroma da espécie. Esses artigos não trazem informações sobre vírus bacteriófagos então não sabemos se há famílias de bacteriófagos em comum entre os gambás e esses animais domésticos.

Quando avaliamos as famílias encontradas no presente estudo que já foram registradas em animais de produção, especialmente gado bovino, destacamos Flaviridae, Picornaviridae, Parvoviridae, Dicistroviridae, Mimiviridae, Smacoviridae, Genomoviridae, Baculoviridae, Herpesviridae, Circoviridae, Retroviridae, e Coronaviridae já foram identificadas (Ambrose *et al.*, 2023; Zhu *et al.*, 2024). Dessas famílias, quatro (Baculoviridae, Coronaviridae, Herpesviridae, Retroviridae) foram encontradas na biblioteca da APA, a área que era adjacente a fazendas. Nesse caso, o sinal é mais fraco do que aquele visualizado para as bibliotecas de Campos, uma vez que apenas um terço das famílias foi em comum e não foram exclusivas de bibliotecas da área onde esperamos maior contato.

Todas as bibliotecas avaliadas, bem como os subgrupos por área ou sexo tiveram vírus bacteriófagos como dominantes na composição e frequência. Esse tipo viral não é considerado patogênico para os gambás ou qualquer outro vertebrado, por terem como hospedeiro classicamente as bactérias (Griffiths *et al.*, 2016). Esse resultado era esperado porque as amostras representam principalmente os vírus que ocupam o trato gastrointestinal, um local repleto de bactérias e, por consequência, de vírus que as infectam (Chaffringeon *et al.*, 2022). No geral, estudos de viroma gastrointestinal em mamíferos trazem resultados parecidos, como o de demônios da tasmânia (Chong, 2019), de seres humanos (Cao *et al.*, 2022), de cães domésticos e lobos cinza (Conceição-Neto *et al.*, 2017).

Nossos resultados indicam que, provavelmente, a maioria dos animais amostrados estavam saudáveis no momento da captura já que a microbiota refletiu o padrão da grande presença de bacteriófagos visto para o viroma gastrointestinal de outras espécies. Entretanto, tivemos duas bibliotecas (Machos da União A e fêmeas de Campos) que demonstraram indícios de disbiose, ou seja, possível sinal de patologia (Walker, 2017). A disbiose é caracterizada pela predominância de uma ou duas famílias virais no viroma, indicando um desequilíbrio (Carding *et al.*, 2015). Em cada uma dessas bibliotecas tivemos uma família concentrando quase todo o número de leituras. Na biblioteca de machos da União A 91,6% das leituras foram atribuídas a família Straboviridae e nas fêmeas de Campos essa mesma família acumulou 89,5% das leituras.

Não podemos confirmar nenhum quadro de doença pois, visualmente, os indivíduos estavam saudáveis. Os indícios de disbiose que observamos também podem ser momentâneos, considerando que a amostragem que fizemos é um recorte de como o indivíduo estava naquele dado momento. Além disso, não temos nenhum outro estudo de referência que indique como o viroma gastrointestinal dessa espécie deveria se comportar para assinalar que este esteja fugindo de um padrão esperado. No entanto, temos as outras bibliotecas descritas para o presente estudo que apresentam maior diversidade e menor dominância de uma família específica, o que é mais condizente com o padrão visualizado para viromas gastrointestinais de outras espécies de mamíferos (Amanbayeva *et al.*, 2021; Lojkić *et al.*, 2016). Esperávamos encontrar esses indícios de disbiose nas bibliotecas de Campos, pois o ambiente urbano supostamente favoreceria a disbiose e a riqueza de famílias virais como visto por Tirera e colaboradores (2021). Tirera e

colaboradores (2021) avaliaram o viroma de quatro espécies de roedores em florestas intactas, florestas perturbadas, savana e ambiente periurbano e observaram que a riqueza de famílias virais para essas espécies, em geral, era menor nos roedores capturados em florestas perturbadas e a menor riqueza de todas foi vista no ambiente periurbano. No nosso caso, vimos que a disbiose poderia estar ocorrendo tanto na União A, uma área mais preservada, quanto em Campos, a área mais antropizada. Assim não é possível atribuir esta potencial disbiose aos efeitos da antropização. Além disso, uma das bibliotecas foi de machos, enquanto a outra foi de fêmeas, então também não parece estar relacionada ao sexo dos hospedeiros.

Das famílias patogênicas identificadas é importante destacar as famílias Circoviridae, nas fêmeas da União B e Parvoviridae, nos machos de Campos. Daremos destaque a essas famílias porque foram as famílias com maior número de leituras em suas bibliotecas, excluindo bacteriófagos. Além disso, ambas apareceram nos resultados confirmados pelo *Diamond*. Entre as famílias patogênicas encontradas nas bibliotecas essas famílias foram as únicas que apresentaram um número de leituras >40. Consideramos essa quantidade de leituras abaixo de 40 como sendo baixa. A família *Coronaviridae*, que foi encontrada em quase todas as bibliotecas, não foi destacada aqui tanto por ter <40 leituras quanto devido ao fato de que o Laboratório de Diversidade e Doenças Virais trabalha com amplificação desse vírus, logo, a fonte da família viral pode ter sido contaminação.

A família Circoviridae esteve presente em três bibliotecas da União (fêmeas da União A e ambas as bibliotecas da União B), sendo que o número de leituras mais alto foi para a biblioteca de fêmeas da União B (n=203). As fêmeas da União A e os machos da União B apresentaram, respectivamente seis e duas leituras. Apesar de essa família ser comum em porcos domésticos ela não foi identificada na APA, uma área adjacente a matrizes agrícolas. Contudo, sabemos que na ReBIO União temos a presença de catetos (*Dicotyles tajacu*) (Mayer *et al.*, 2023) e talvez esse contato possa ter sido a fonte do vírus ou essa família possa ter vindo de algum item alimentar dos gambás. Essa família é composta por vírus de DNA circular, de fita simples e foi descrita pela primeira vez nos anos 90, com espécies infectantes de porcos e aves (Studdert, 1993).

Os circovírus infectam vertebrados, sendo conhecidos por infectar mamíferos, aves e peixes (Breitbart *et al.*, 2017; Delwart e Li, 2012). Infecções por circovírus em porcos causam sintomas relacionados aos tecidos do sistema linfático e cardiorrespiratório, podendo causar a morte do animal (Segalés e Sibila, 2022). Também há relatos de problemas no trato gastrointestinal (Yang *et al.*, 2021), com sua detecção ocorrendo desde o sangue até nas fezes (Segalés, Allan e Domingo, 2005). Essa família também já foi previamente detectada no viroma gastrointestinal do Vison-Americano (*Neogale vison*), da gineta (*Genetta genetta*), do texugo-europeu (*Meles meles*) e da lontra-europeia (*Lutra lutra*) (Bodewes *et al.*, 2014).

Não existem relatos de circovírus infectando *D. aurita*, mas temos um relato desta família viral presente no marsupial didelfídeo *Monodelphis domestica* (Katzourakis e Gifford, 2010). Em *M. domestica* o circovírus identificado é considerado endógeno, ou seja, o genoma viral já está integrado ao genoma da espécie (Katzourakis e Gifford, 2010). Este poderia ser o caso deste estudo, porém, fizemos o filtro de genoma do hospedeiro justamente com o genoma completo de *M. domestica*. Logo, as sequências da família Circoviridae aqui vistas podem ser de origem da dieta ou de uma infecção real na espécie. Seria necessária a realização de mais análises para se chegar a uma conclusão.

A família Parvoviridae foi exclusiva da biblioteca de machos de Campos e apresentou 420 leituras. Os parvovírus são vírus de DNA fita simples, com genomas pequenos (Cotmore *et al.*, 2019). Essa família possui duas subfamílias, Parvovirinae, que abriga os vírus que infectam vertebrados; e Densovirinae, com vírus que infectam artrópodes (Cotmore *et al.*, 2014). Os parvovírus infectantes de vertebrados causam doenças como a parvovirose canina, comum em cães domésticos (Sykes, 2014). Essa infecção é caracterizada pela presença de diarreia com sangue e outras alterações, como vômito e desidratação, podendo culminar na morte do animal (Ahmad e Waheed, 2020; Reddy *et al.*, 2015). Os vírus circulam pelo sistema linfático, mas aparecem nas fezes (Ahmad e Waheed, 2020), por isso essa família já apareceu no viroma fecal das espécies de mamíferos *Genetta genetta* e *Mustela lutreola* (Bodewes *et al.*, 2014). Logo após a finalização das coletas das amostras no município de Campos, foi noticiado que a cidade estaria passando por um surto de parvovirose canina. Considerando que os gambás que habitam centros urbanos estão em contato direto com cachorros e outros animais domésticos, a presença de parvovírus nos gambás pode indicar que estes estejam agindo como reservatório da

doença ou estejam sendo infectados. Essa família também já foi identificada no viroma de gatos e cães domésticos (Jiang *et al.*, 2023; Moreno *et al.*, 2017; Van Brussel *et al.*, 2022; Zhang *et al.*, 2014). Porém, como existem parvovírus que infectam artrópodes (Cotmore *et al.*, 2014), um item alimentar consumido por *D. aurita*, não é possível afirmar que a presença dessa família está atrelada a um parvovírus infectante de vertebrados. Outro ponto é que apenas com a detecção desses ácidos nucleicos não é possível definir a presença de infecção ou doença e por isso ressalta-se a necessidade de aprofundamento da análise nessas famílias virais patogênicas.

Apresentamos a hipótese de que a urbanização afetaria a composição viral dos gambás. As únicas evidências encontradas foram a potencial disbiose e a identificação da família Parvoviridae, que pode indicar a presença de uma espécie viral característica de animais doméstico, exclusivamente em uma biblioteca de área urbana. As demais análises não indicaram uma ordenação na composição de famílias virais em função do espaço ou do nível de antropização das áreas. Foi observada uma maior importância do sexo do hospedeiro como determinante da composição viral. Esse resultado contradiz o que já foi visualizado para análises da influência das atividades antrópicas sobre o viroma de mamíferos. Para a espécie de roedor *Zygodontomys brevicauda* a riqueza de famílias virais foi menor nos animais capturados em ambiente periurbano (Tirera *et al.*, 2021). Heni e colaboradores (2023) demonstraram que a microbiota das espécies *D. marsupialis*, *Philander opossum* e *Proechimys semispinosus*, varia conforme o grau de perturbação antrópica no local, com a diversidade alpha sendo impactada negativamente pela perturbação.

Sabemos que diversas variáveis são capazes de modular o viroma, mas, devido a evidências na literatura, acreditamos que a antropização da área fosse a mais importante fonte de variação nas bibliotecas (Šimić *et al.*, 2020; Tirera *et al.*, 2021). Um outro motivo é que ambientes perturbados são provedores de fontes virais distintas daquelas dos ambientes intactos (e.g. esgoto, animais domésticos) e que estes seriam mais suscetíveis a favorecer o surgimento de doenças (Keesing e Ostfeld, 2024). No entanto, esse padrão não se manifestou nas amostras avaliadas no presente estudo e pode ser reflexo da alta resistência de *D. aurita* a antropização.

Em relação ao sexo, sabemos que machos de *D. aurita* têm áreas de vida maiores (Brum, Melo e Cáceres, 2022; Cáceres e Monteiro-Filho, 2001) e, portanto,

poderiam estar mais suscetíveis a entrar em contato com novas fontes virais do que as fêmeas. Mas, mesmo com essas diferenças, esperávamos que machos e fêmeas de uma mesma área tivessem a composição do viroma mais parecida entre si do que com indivíduos de outras áreas, justamente pela proximidade e exposição a fontes virais similares. No entanto, essa hipótese não se confirmou e o sexo se mostrou como um preditor da composição mais forte do que a área ou o nível de antropização. Essa influência do sexo na composição viral já foi descrita para seres humanos e animais de laboratório (Jašarević, Morrison e Bale, 2016). Dominianni e colaboradores (2015) demonstraram que, em seres humanos, o sexo está fortemente relacionado com a composição da microbiota intestinal. Esse achado também foi visto em estudos explorando a relação entre sexo, doenças autoimunes e a microbiota intestinal, no qual mulheres são mais suscetíveis a doenças autoimunes e isso pode ser uma consequência de como hormônios sexuais interagem com as bactérias do intestino (Gomez, Luckey e Taneja, 2015; Markle *et al.*, 2013; Rizzetto *et al.*, 2018). A riqueza da microbiota também seria afetada, com fêmeas tendo uma riqueza maior do que os machos (Sisk-Hackworth, Kelley e Thackray, 2023). Isso foi o que observamos na maioria das nossas bibliotecas em que tivemos essa divisão por sexo. Nas áreas da União, as fêmeas sempre tiveram uma riqueza de famílias maior do que os machos, apesar do mesmo não ter sido visto em Campos.

Os gambás são conhecidamente onívoros e generalistas, sem diferenças na dieta determinadas pelo sexo (Ceotto *et al.*, 2009; Faria; Lanes e Bonvicino, 2019). Isso quer dizer que eles não possuem um item alimentar preferido e sim variam sua dieta conforme a disponibilidade de alimento (Cáceres e Monteiro-Filho, 2001). Essa possibilidade de adaptação dos gambás pode ter sido o suficiente para que a antropização das áreas não afetasse tanto a composição das comunidades virais e o efeito causado pelo sexo se tornasse mais visível.

Quando usamos apenas os dados das famílias confirmadas no *Diamond*, os padrões identificados no presente estudo se tornaram menos claros. Nós decidimos fazer essa etapa de “confirmação” e comparar os resultados pois as ferramentas funcionam de maneira distinta e, conseqüentemente, erram de maneiras distintas. O *Kraken2*, que pode retornar falso positivos, e o *Diamond*, tende a retornar resultados com uma maior taxa de falsos negativos (Cárdenas, Neuenschwander e Malaspina, 2022). A confirmação nas duas plataformas seria um esforço para contornar os erros

de identificação do *Kraken2* que indicaria a presença de famílias que na verdade não estavam na biblioteca.

Com os dados confirmados continuamos observando a prevalência de bacteriófagos que vimos entre todas as bibliotecas. O maior impacto desse tratamento foi na riqueza de famílias, que mudou consideravelmente, com algumas bibliotecas sendo mais afetadas do que outras. Por consequência, também tivemos um impacto na estrutura de comunidades e no padrão de separação por sexo que nós havíamos observado. Assim, por mais que a plataforma *Kraken2* seja suscetível a falsos positivos, talvez seus dados completos sejam os mais adequados para serem utilizados em análises ecológicas. Como não temos certeza da taxa de erro do *Kraken2*, que parece chegar a até 1% do total (Cárdenas, Neuenschwander e Malaspina, 2022; Johnson, Sun e Fodor, 2022), a confirmação que fizemos com o *Diamond* pode ter sido reducionista demais, uma vez que não podemos controlar para os falsos negativos gerados por ele. O impacto na riqueza, por exemplo, foi consequência da exclusão de famílias raras e que tendiam a ser exclusivas de determinadas bibliotecas. Logo, essa exclusão reduziu as diferenças e particularidades entre as bibliotecas, impedindo detectar os padrões observados com o uso do *Kraken2* apenas.

Outra questão que podemos destacar é a importância de estudos a longo prazo, como o Programa de Pesquisas Ecológicas de Longa Duração (PELDs). Esse programa fomenta, justamente, projetos que duram anos (CNPq, 2024). Seria de extrema importância que esse tipo de programa fosse implementado no que tange o estudo dos vírus e o monitoramento da diversidade viral em animais silvestres. Desse modo, não estaríamos apenas fazendo um recorte do viroma no momento da coleta das amostras, mas sim o avaliando de maneira contínua. Isso nos permitiria tirar melhores conclusões sobre as famílias patogênicas que foram encontradas e quaisquer desequilíbrios que encontramos.

7. CONCLUSÃO

Este estudo traz a primeira descrição do viroma gastrointestinal completo da espécie *D. aurita*, listando as famílias virais que ocorrem na espécie e definindo uma base para estudos futuros. Demonstramos que o viroma gastrointestinal é composto majoritariamente de vírus bacteriófagos e que a composição de famílias virais é fortemente influenciada pelo sexo dos indivíduos, com o nível de antropização da área de coleta tendo aparentemente pouca influência. Até onde sabemos, este é o primeiro estudo que demonstra a importância do sexo na composição de famílias virais em hospedeiros vertebrados de vida livre, já que todos os outros relatos são para humanos e animais de laboratório.

Destacamos aqui o encontro de famílias virais conhecidamente patogênicas e a importância de estudos futuros avaliarem especialmente estas famílias em mais detalhes. Dessa maneira, será possível avaliar se as espécies dessas famílias virais são realmente as espécies patogênicas e avaliar mais diretamente o papel de *D. aurita* no ciclo dessas doenças.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmad, M., Waheed, A. (2020). Case of Canine Parvovirus Infection. *Research Journal of Veterinary Practitioners*. 8(2): 23-25. <http://dx.doi.org/10.17582/journal.rjvp/2020/8.2.23.25>
- Albery, G. F., Eskew, E. A., Ross, N., Olival, K. J. (2020). Predicting the global mammalian viral sharing network using phylogeography. *Nature Communications*. 11: 2260. <https://doi.org/10.1038/s41467-020-16153-4>
- Almeida, M. F., Massad, E., Aguiar, E. A. C., Martorelli, L. F. A., Joppert, A. M. S. (2001). Neutralizing Antirabies Antibodies in Urban Terrestrial Wildlife in Brazil. *Journal of Wildlife Diseases*. 37(2): 394-398. <https://doi.org/10.7589/0090-3558-37.2.394>
- Amanbayeva, M., Anarkulova, E., Bogoyavlenskiy, A., Alexyuk, M., Imangazy, A., Berezin, V. (2021). Metagenomic Exploration of *Atelerix albiventris* Gut Microbiome. *Microbiol Resour Announc*. <https://doi.org/10.1128/mra.01342-20>
- Ambrose, R. K., Blakebrough-Hall, C., Gravel, J. L., Gonzalez, L. A., Mahony, T. J. (2023). Characterisation of the Upper Respiratory Tract Virome of Feedlot Cattle and Its Association with Bovine Respiratory Disease. *Viruses*. 15(2): 455. <https://doi.org/10.3390/v15020455>

- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Araujo, D. B. *et al.* (2014). Antibodies to rabies virus in terrestrial wild mammals in native rainforest on the north coast of São Paulo state, Brazil. *Journal of Wildlife Diseases*. 50(3): 469-477. <https://doi.org/10.7589/2013-04-099>
- Araújo, J. *et al.* (2012). Detection of hantavirus in bats from remaining rain forest in São Paulo, Brazil. *BMC Research Notes*. 5: 690. <https://doi.org/10.1186/1756-0500-5-690>
- Astúa, D. (2010). Cranial sexual dimorphism in New World marsupials and a test of Rensch's rule in Didelphidae. *Journal of Mammalogy*. 91(4): 1011-1024. <https://doi.org/10.1644/09-MAMM-A-018.1>
- Astúa, D., Carvalho, R. A., Maia, P. F., Magalhães, A. R., Loretto, D. (2015). First evidence of gregarious denning in opossums (Didelphimorphia, Didelphidae), with notes on their social behaviour. *Biology Letters*. 11: 20150307. <http://dx.doi.org/10.1098/rsbl.2015.0307>
- Astúa, D., Cherem, J. J., Teta, P. (2023). Taxonomic Checklist of Living American Marsupials. In: Cáceres, N. C., Dickman, C. R. (eds.), *American and Australasian Marsupials An Evolutionary, Biogeographical, and Ecological Approach*. Switzerland: Springer. p. 115-162.
- Bernal, M. K. M., Chiang, J. O., Mendes, F. F., Andrade, S. L. S., da Silva, S. K. S. M., Pereira, W. L. A. (2021). Study of Arboviruses in *Philander opossum*, *Didelphis marsupialis* and *Nectomys rattus* captured from forest fragments in the municipality of Belém, Pará, Brazil. *Ciência Rural*. 51(4): e20200515. <http://doi.org/10.1590/0103-8478cr20200515>
- Bezerra-Santos, M. A. *et al.* (2020). Gastrointestinal parasites in the opossum *Didelphis aurita*: Are they a potential threat to human health? *J. Parasit. Dis.* 44(2): 355-363. <https://doi.org/10.1007/s12639-020-01205-9>
- Bi, Y. *et al.* (2020). Dominant subtype switch in avian influenza viruses during 2016-2019 in China. *Nature Communications*. 11: 5909. <https://doi.org/10.1038/s41467-020-19671-3>
- Bodewes, R., Ruiz-Gonzalez, A., Schapendonk, C. ME., van den Brand, J. MA., Osterhaus, A. DME., Smits, S. L. (2014). Viral metagenomic analysis of feces of wild small carnivores. *Virology Journal*. 11:89. <https://doi.org/10.1186/1743-422X-11-89>
- Braga, C., Oliveira, J. A., Cerqueira, R. (2017). Metacomunidades: uma introdução aos termos e conceitos. *Oecologia Australis*. 21(2): 108-118. <https://doi.org/10.4257/oeco.2017.2102.02>
- Breitbart, M., Delwart, E., Rosario, K., Segalés, J., Varsani, A., ICTV Report Consortium. (2017). ICTV Virus Taxonomic Profile: *Circoviridae*. *Journal of General Virology*. 98: 1997-1998. DOI: 10.1099/jgv.0.000871

- Breitwieser, F. P., Lu, J., Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*. 20(4): 1125-1139. doi: 10.1093/bib/bbx120
- Brooks, D. R., Boeger, W. A., Hoberg, E. P. (2022). The Stockholm Paradigm: Lessons for the Emerging Infectious Disease Crisis. *MANTER: Journal of Parasite Biodiversity*. 22. doi: 10.32873/unl.dc.manter22
- Brum, M. N., Melo, G. L., Cáceres, N. C. (2022). Ecological evidence of hierarchy and competition in *Didelphis aurita* females. *Mammalian Biology*. 102: 1709-1721. <https://doi.org/10.1007/s42991-022-00269-9>
- Bubadué, J., Meloro, C., Hendges, C., Battistela, T., Carvalho, R., Cáceres, N. (2021). Clinal and Allometric Variation in the Skull of Sexually Dimorphic Opossums. *Journal of Mammalian Evolution*. 28: 185-198. <https://doi.org/10.1007/s10914-020-09513-w>
- Cáceres, N. C., Delciellos, A. C., Prevedello, J. A., Brum, M. N., Albanese, M. S. (2023). Movement, Habitat Selection, and Home Range of American Marsupials. In: Cáceres, N. C., Dickman, C. R. (eds.), *American and Australasian Marsupials An Evolutionary, Biogeographical, and Ecological Approach*. Switzerland: Springer. p. 1005-1031
- Cáceres, N. C., Monteiro-Filho, E. L. A. (2001). Food Habits, Home Range and Activity of *Didelphis aurita* (Mammalia, Marsupialia) in a Forest Fragment of Southern Brazil. *Studies on Neotropical Fauna and Environment*. 36(2): 85-92. <https://doi.org/10.1076/snfe.36.2.85.2138>
- Cao, Z., Sugimura, N., Burgermeister, E., Ebert, M. P., Zuo, T., Lan, P. (2022). The gut virome: A new microbiome component in health and disease. *EBioMedicine*. 81: 104113. <https://doi.org/10.1016/j.ebiom.2022.104113>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15): 1972-1973. doi:10.1093/bioinformatics/btp348
- Cárdenas, Y. O. A., Neuenschwander, S., Malaspinas, A. S. (2022). Benchmarking metagenomics classifiers on ancient viral DNA: a simulation study. *PeerJ*. 10: e12784. <https://doi.org/10.7717/peerj.12784>
- Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. *Microbial Ecology in Health and Disease*. 26:1. <https://doi.org/10.3402/mehd.v26.26191>
- Carlson, C. J. *et al.* (2022). Climate change increases cross-species viral transmission risk. *Nature*. 607: 555-562. <https://doi.org/10.1038/s41586-022-04788-w>
- Carneiro, I. O. *et al.* (2018). A Novel Marsupial Hepatitis A Virus Corroborates Complex Evolutionary Patterns Shaping the Genus Hepatovirus. *Journal of Virology*. 92(13): e00082-18. <https://doi.org/10.1128/jvi.00082-18>

- Caserta, L. C. *et al.* (2023). Bacterial and Viral Diversity of Didelphid Opossums from Brazil. *EcoHealth*. 20: 362-369. <https://doi.org/10.1007/s10393-023-01667-2>
- Ceotto, P., Finotti, R., Santori, R., Cerqueira, R. (2009). Diet variation of the marsupials *Didelphys aurita* and *Philander frenatus* (Didelphimorphia, Didelphidae) in a rural area of Rio de Janeiro state, Brazil. *Mastozoología neotropical*. 16(1): 49:58.
- Chaffringeon, L., Lamy-Besnier, Q., Debardieux, L., Sordi, L. D. (2021). The intestinal virome: lessons from animal models. *Current Opinion in Virology*. 51: 141-158. <https://doi.org/10.1016/j.coviro.2021.09.016>
- Chong, R., Shi, M., Grueber, C. E., Holmes, E. C., Hogg, C. J., Belov, K., Barrs, V. R. (2019). Fecal Viral Diversity of Captive and Wild Tasmanian Devils Characterized Using Virion-Enriched Metagenomics and Metatranscriptomics. *Journal of Virology*. 93(11): 00205-19. <https://doi.org/10.1128/JVI.00205-19>
- Conceição-Neto, N. *et al.* (2017). Viral gut metagenomics of sympatric wild and domestic canids, and monitoring of viruses: Insights from an endangered wolf population. *Ecology and Evolution*. 7(12): 4135-4146. <https://doi.org/10.1002/ece3.2991>
- Conselho Nacional de Desenvolvimento Científico e Tecnológico. (2024). *PELD*. Disponível em: <https://www.gov.br/cnpq/pt-br/acesso-a-informacao/acoes-e-programas/programas/peld>
- Cotmore, S. F. *et al.* (2014). The family *Parvoviridae*. *Arch Virol*. 159: 1239-1247. DOI 10.1007/s00705-013-1914-1
- Cotmore, S. F. *et al.* (2019). ICTV Virus Taxonomy Profile: *Parvoviridae*. *Journal of General Virology*. 100: 367-368. DOI 10.1099/jgv.0.001212
- Da Silva, F. R., Gonçalves-Souza, T., Paterno, G. B., Provete, D. B., Vancine, M. H. (2022). *Análises Ecológicas no R*. Bauru: Canal 6 Editora. 640p.
- D'arc, M. *et al.* (2015). Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proceedings of the National Academy of Sciences of the United States of America*. 112(11): E1343-E1352. <https://doi.org/10.1073/pnas.1502022112>
- D'arc, M., Etienne, L., Delaporte, E., Peeters, M. (2017). Origin and Emergence of HIV/AIDS. In: Tibayrenc, M. (org.), *Genetics and Evolution of Infectious Diseases*. Elsevier. p. 573-599.
- D'arc, M. *et al.* (2020). The characterization of two novel neotropical primate papillomaviruses supports the ancient within-species diversity model. *Virus Evolution*. 6(1): veaa036. doi: 10.1093/ve/veaa036
- De Castro, R. G. B. M., Neto, S. F. C., Maldonado Júnior, A., Gentile, R. (2017). Ecological aspects of nematode parasites of *Didelphis aurita* (Didelphimorphia, Didelphidae) in urban-sylvatic habitats in Rio de Janeiro, Brazil. *Oecologia Australis*. 21(1): 54-61. <https://doi.org/10.4257/oeco.2017.2101.06>

- De Souza, W. M. *et al.* (2018). Discovery of novel anelloviruses in small mammals expands the host range and diversity of the Anelloviridae. *Virology*. 514: 9-17. <https://doi.org/10.1016/j.virol.2017.11.001>
- Delwart, E., Li, L. (2012). Rapidly expanding genetic diversity and host range of the *Circoviridae* viral family and other Rep encoding small circular ssDNA genomes. *Virus Research*. 164: 114-121. doi:10.1016/j.virusres.2011.11.021
- Dominianni, C., Sinha, R., Goedert, J. J., Pei, Z., Yang, L., Hayes, R. B., Ahn, J. (2015). Sex, Body Mass Index, and Dietary Fiber Intake Influence the Human Gut Microbiome. *PLoS ONE*. 10(4): e0124599. doi:10.1371/journal.pone.0124599
- Dunowska, M., Perrott, M., Biggs, P. (2022). Identification of a novel polyomavirus from a marsupial host. *Virus Evolution*. 8(2): 1-9. <https://doi.org/10.1093/ve/veac096>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5): 1792-1797. DOI: 10.1093/nar/gkh340
- Ewels, P., Magnusson, M., Lundin, S., Källér, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 32(19): 3047-3048. doi: 10.1093/bioinformatics/btw354
- Faria, M. B., Lanes, R. O., Bonvicino, C. R. (2019). *Guia dos marsupiais do Brasil: guia de identificação com base em caracteres morfológicos externos e cranianos*. 1.ed. São Caetano do Sul : Amélie Editorial. 84p.
- Feng, Y. *et al.* (2022). A time-series meta-transcriptomic analysis reveals the seasonal, host, and gender structure of mosquito viromes. *Virus Evolution*. 8(1): 1-14. DOI: <https://doi.org/10.1093/ve/veac006>
- Ferreira, M. S. (2023). Activity Patterns of American Marsupials. In: Cáceres, N. C., Dickman, C. R. (eds.), *American and Australasian Marsupials An Evolutionary, Biogeographical, and Ecological Approach*. Switzerland: Springer. p. 1189-1219.
- Ferreira-Machado, E. *et al.* (2023). Naturally Acquired Rabies in White-Eared Opossum, Brazil. *Emerging Infectious Diseases*. 29(12): 2541-2545. DOI: <https://doi.org/10.3201/eid2912.230373>
- Forterre, P., Prangishvili, D. (2009). The origin of viruses. *Research in Microbiology*. 160(7): 466-472. doi:10.1016/j.resmic.2009.07.008
- Gomez, A., Luckey, D., Taneja, V. (2015). The gut microbiome in autoimmunity: Sex matters. *Clinical Immunology*. 159: 154-162. <http://dx.doi.org/10.1016/j.clim.2015.04.016>
- Griffiths, A. J. F., Wessler, S. R., Carroll, S. B., Doebley, J. (2015). *Introdução à genética*. Rio de Janeiro: Guanabara Koogan. 2036pp.
- Harvey, E., Holmes, E. C. (2022). Diversity and evolution of the animal virome. *Nat. Rev. Microbiol.* 20(6): 321-334. <https://doi.org/10.1038/s41579-021-00665-x>

- He, W. *et al.* (2022). Virome characterization of game animals in China reveals a spectrum of emerging pathogens. *Cell*. 185 (7): 1117-1129.
<https://doi.org/10.1016/j.cell.2022.02.014>
- Heni, A. C. *et al.* (2023). Wildlife gut microbiomes of sympatric generalist species respond differently to anthropogenic landscape disturbances. *Animal Microbiome*. 5: 26. <https://doi.org/10.1186/s42523-023-00237-9>
- Honce, R., Schultz-Cherry, S. (2020). Recipe for Zoonosis: How Influenza Virus Leaps into Human Circulation. *Cell Host & Microbiome*. 28: 506-508.
<https://doi.org/10.1016/j.chom.2020.09.008>
- Howe, A. *et al.* (2016). Divergent responses of viral and bacterial communities in the gut microbiome to dietary disturbances in mice. *The ISME Journal*. 10(5): 1217-1227.
<https://doi.org/10.1038/ismej.2015.183>
- Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. 39: D576-82. doi: 10.1093/nar/gkq901
- ICMBio (2008). Plano de Manejo da Área de Proteção Ambiental da Bacia do Rio São João/Mico-Leão-Dourado. Rio de Janeiro. 349p.
- ICMBio (2023). Plano De Manejo Reserva Biológica União. 2023. 75p.
- Instituto Brasileiro de Geografia e Estatística (IBGE) (2022). Campos dos Goytacazes: <https://cidades.ibge.gov.br/brasil/rj/campos-dos-goytacazes/panorama-em-13/05/23>
- Jašarević, E., Morrison, K. E., Bale, T. L. (2016). Sex differences in the gut microbiome – brain axis across the lifespan. *Phil. Trans. R. Soc. B*. 371: 20150122.
<http://dx.doi.org/10.1098/rstb.2015.0122>
- Jiang, X. *et al.* (2023). Virome of high-altitude canine digestive tract and genetic characterization of novel viruses potentially threatening human health. *mSphere*. 8:e00345-23. <https://doi.org/10.1128/msphere.00345-23>
- Kaler, J., Hussain, A., Flores, G., Kheiri, S., Desrosiers, D. (2022). Monkeypox: A Comprehensive Review of Transmission, Pathogenesis, and Manifestation. *Cureus*. 14(7): e26531. DOI 10.7759/cureus.26531
- Katzourakis, A., Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genetics*. 6(11): e1001191. doi:10.1371/journal.pgen.1001191
- Kearse, M. *et al.* (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28(12): 1647-1649. doi:10.1093/bioinformatics/bts199
- Keesing, F., Ostfeld, R. S. (2024). The more, the healthier: Tree diversity reduces forest pests and pathogens. *PLoS Biology*. 22(2): e3002525.
<https://doi.org/10.1371/journal.pbio.3002525>

- King, K. *et al.* (2022). Coevolutionary Analysis Implicates Toll-Like Receptor 9 in Papillomavirus Restriction. *mBio*. 13(2): e0005422. doi: 10.1128/mbio.00054-22
- Lamglait, B., Lair, S. (2019). Fatal west nile virus infection in a Virginia opossum (*Didelphis Virginiana*) with pulmonary lepidic-predominant adenocarcinoma. *Journal of Wildlife Diseases*. 55(4): 990-994. <https://doi.org/10.7589/2018-12-284>
- Lauber, C. *et al.* (2017). Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. *Cell Host and Microbe*. 22(3): 387-399. doi: 10.1016/j.chom.2017.07.019
- Leibold, M. A., Mikkelsen, G. M. (2002). Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos*. 97(2): 237-250. <https://doi.org/10.1034/j.1600-0706.2002.970210.x>
- Lojkić, I., Biđin, M., Prpić, J., Šimić, I., Krešić, N. (2016). Faecal virome of red foxes from peri-urban areas. *Comparative Immunology, Microbiology and Infectious Diseases*. 45: 10-15. <http://dx.doi.org/10.1016/j.cimid.2016.01.005>
- Lu, J. *et al.* (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*. 17: 2815-2839. <https://doi.org/10.1038/s41596-022-00738-y>
- Luo, Z., Ji, Q., Wible, J. R., Yuan, C. (2003). An Early Cretaceous Tribosphenic Mammal and Metatherian Evolution. *Science*. 302: 1934-1940. <https://doi.org/10.1126/science.1090718>
- Lynch, S. V. (2014). Viruses and microbiome alterations. *Annals of the American Thoracic Society*. 11(1): S57-S60. <https://doi.org/10.1513/AnnalsATS.201306-158MG>
- Lysholm, F. *et al.* (2012). Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS ONE*. 7(2): e30875. doi:10.1371/journal.pone.0030875
- Magurran, A. E. (2021). Measuring biological diversity. *Current Biology*. 31(19): R1174-R1177. <https://doi.org/10.1016/j.cub.2021.07.049>.
- Manrubia, C. S., Lázaro, E. (2006). Viral evolution. *Physics of Life Reviews*. 3: 65-92. <https://doi.org/10.1016/j.plrev.2005.11.002>
- Markle, J. G. M. *et al.* (2013). Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity. *Science*. 339: 1084-1088. <https://doi.org/10.1126/science.1233521>
- Mayer, L. B., Mellado, B., Ruiz-Miranda, C. R., Nogueira, M. R., Monteiro, L. R. (2023). Diversity profiles of medium and large-size mammals in an Atlantic Forest remnant: seasonal and spatial patterns. *Studies on Neotropical Fauna and Environment*. <https://doi.org/10.1080/01650521.2023.2180198>
- McCaughey, C. (2010). Influenza: a virus of our times. *Ulster Med. J.* 79(2): 46-51.

- Melo, G. L., Sponchiado, J. (2012). Distribuição geográfica de marsupiais no Brasil. *In: Cáceres, N. C., Os marsupiais do Brasil*. 2.ed. Campo Grande: Editora UFMS. p. 95-112.
- Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*. 11: 31-46. <https://doi.org/10.1038/nrg2626>
- Miranda, T. S. *et al.* (2022). Metagenomic analysis reveals novel dietary-related viruses in the gut virome of marmosets hybrids (*Callithrix jacchus x Callithrix penicillata*), Brazil. *Virus Research*. 325: 199017. <https://doi.org/10.1016%2Fj.virusres.2022.199017>
- Moreno, P. S., Wagner, J., Mansfield, C. S., Stevens, M., Gilkerson, J. R., Kirkwood, C. D. (2017). Characterisation of the canine faecal virome in healthy dogs and dogs with acute diarrhoea using shotgun metagenomics. *PLoS One*. 12(6): e0178433. <https://doi.org/10.1371/journal.pone.0178433>
- Morse, S. S. *et al.* (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet*. 385: 1956-1965. [https://doi.org/10.1016/S0140-6736\(12\)61684-5](https://doi.org/10.1016/S0140-6736(12)61684-5)
- Mourya, D. T. *et al.* (2019). Canine Distemper Virus in Asiatic Lions of Gujarat State, India. *Emerging Infectious Diseases*. 25(11): 2128-2130. <https://doi.org/10.3201/eid2511.190120>
- Munday, J. S., Knight, C. G. (2010). Amplification of feline sarcoid-associated papillomavirus DNA sequences from bovine skin. *Veterinary Dermatology*. 21: 341-344. <https://doi.org/10.1111/j.1365-3164.2010.00872.x>
- Orbell, G. M. B., Young, S., Munday, J. S. (2011). Cutaneous Sarcoids in Captive African Lions Associated With Feline Sarcoid-Associated Papillomavirus infection. *Veterinary Pathology*. 48(6): 1176-1179. <https://doi.org/10.1177/0300985810391111>
- Paglia, P. A. *et al.* (2012). *Lista anotada dos mamíferos do Brasil*. 2.ed. Arlington: Conservation International. 76p.
- Peres, M. G. *et al.* (2016). Dogs and opossums positive for vaccinia virus during outbreak affecting cattle and humans, São Paulo State, Brazil. *Emerging Infectious Diseases*. 22(2): 271-273. <http://dx.doi.org/10.3201/eid2202.140747>
- Peres, M. G. *et al.* (2018). Vaccinia Virus in Blood Samples of Humans, Domestic and Wild Mammals in Brazil. *Viruses*. 10(1): 42. <https://doi.org/10.3390/v10010042>
- Platto, S., Wang, Y., Zhou, J., Carafoli, E. (2021). History of the COVID-19 pandemic: Origin, explosion, worldwide spreading. *Biochemical and Biophysical Research Communications*. 538: 14-23. <https://doi.org/10.1016/j.bbrc.2020.10.087>
- Presley, S. J., Higgins, C. L., Willig, M. R. (2010). A comprehensive framework for the evaluation of metacommunity structure. *Oikos*. 119(6): 908-971. doi: 10.1111/j.1600-0706.2010.18544.x

- Reddy, K. B., Shobhamani, B., Sreedevi, B., Prameela, D. R., Reddy, B. S. (2015). Canine Parvo Viral Infection in Dogs and Their Treatment. *International Journal of Veterinary Science*. 4(3): 142-144.
- Reuter, J. A., Spacek, D. V., Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*. 58(4): 586-597.
<http://dx.doi.org/10.1016/j.molcel.2015.05.004>
- Rice, P., Longden, I., Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*. 16(6): 276-277.
[https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2)
- Rizzetto, L., Fava, F., Tuohy, K. M., Selmi, C. (2018). Connecting the immune system, systematic chronic inflammation and the gut microbiome: the role of sex. *Journal of Autoimmunity*. 92: 12-34. DOI: 10.1016/j.jaut.2018.05.008
- Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., Watson, M. (2017). A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Frontiers in Genetics*. 8:23. doi: 10.3389/fgene.2017.00023
- Saéz, A. M. *et al.* (2015). Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine*. 7: 17-23.
<https://doi.org/10.15252/emmm.201404792>
- Segalés, J., Allan, G. M., Domingo, M. (2005). Porcine circovirus diseases. *Animal Health Research Reviews*. 6(2): 119-142. DOI: 10.1079/AHR2005106
- Segalés, J., Sibila, M. (2022). Revisiting porcine circovirus disease diagnostic criteria in the current *Porcine Circovirus 2* Epidemiological Context. *Veterinary Sciences*. 9(3): 110. <https://doi.org/10.3390/vetsci9030110>
- Segre, J. (2023). Virus: <https://www.genome.gov/genetics-glossary/Virus>
- Silva, R. A., Estevão, V. A. O., Villela, E. F. M. (2024). Circulation of *Trypanosoma cruzi* in triatomines and *Didelphis* sp. in urban areas: Transmission risk assessment in the Metropolitan Region. *Veterinary Parasitology: Regional Studies and Reports*. 52: 101059. <https://doi.org/10.1016/j.vprsr.2024.101059>
- Šimić, I. *et al.* (2020). Viral Metagenomic Profiling of Croatian Bat Population Reveals Sample and Habitat Dependent Diversity. *Viruses*. 12(8): 891.
<https://doi.org/10.3390/v12080891>
- Sisk-Hackworth, L., Kelley, S. T., Thackray, V. G. (2023). Sex, puberty, and the gut microbiome. *Reproduction*. 165: R61-R74. doi: 10.1530/REP-22-0303
- Slatko, B. E., Gardner, A. F., Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*. 122: e59. doi: 10.1002/cpmb.59

- Studdert, M. J. (1993). *Circoviridae*: new viruses of pigs, parrots and chickens. *Australian Veterinary Journal*. 70(4): 121-122. <https://doi.org/10.1111/j.1751-0813.1993.tb06100.x>
- Sykes, J. E. (2014). Canine Parvovirus Infection and Other Viral Enteritides. *In*: Sykes, J. E. (A.), *Canine and Feline Infectious Diseases*. Elsevier. p. 141-151.
- Tabor, G. M., Aguirre, A. A. (2004). Ecosystem Health and Sentinel Species: Adding an Ecological Element to the Proverbial “Canary in the Mineshaft”. *EcoHealth*. 1: 226-228. DOI: 10.1007/s10393-004-0092-8
- Tirera, S., de Thoisy, B., Donato, D., Bouchier, C., Lacoste, V., Franc, A., Lavergne, A. (2021). The influence of habitat on viral diversity in neotropical rodent hosts. *Viruses*. 13(9): 1690. <https://doi.org/10.3390/v13091690>
- Van Brussel, K. *et al.* (2022). The enteric virome of cats with feline panleukopenia differs in abundance and diversity from healthy cats. *Transboundary and Emerging Diseases*. 69: e2952–e2966. DOI: 10.1111/tbed.14646
- Voss, R. S., Jansa, S. A. (2021). *Opossums: an adaptative radiation of new world marsupials*. Baltimore: Johns Hopkins University Press. 338p.
- Walker, W. A. (2017). Dysbiosis. *In*: Floch, M. H., Ringel, Y., Walker, A. W. (Eds.), *The Microbiota in Gastrointestinal Pathophysiology*. Elsevier. p. 227-232.
- Winck, G. R. *et al.* (2022). Socioecological vulnerability and the risk of zoonotic disease emergence in Brazil. *Science Advances*. 8: eabo5774. <https://doi.org/10.1126/sciadv.abo5774>
- Wood, D. E., Lu, J., Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*. 20(1): 257. <https://doi.org/10.1186/s13059-019-1891-0>
- Yang, K. *et al.* (2021). Epidemiology and evolution of emerging porcine circovirus-like viruses in pigs with hemorrhagic dysentery and diarrhea symptoms in central China from 2018 to 2021. *Viruses*. 13(11): 2282. <https://doi.org/10.3390/v13112282>
- Zhang, W., Li, L., Deng, X., Kapusinszky, B., Pesavento, P. A., Delwart, E. (2014). Faecal virome of cats in an animal shelter. *Journal of General Virology*. 95: 2553-2564. DOI 10.1099/vir.0.069674-0
- Zhu, Q. *et al.* (2024). A survey of fecal virome and bacterial community of the diarrhea-affected cattle in northeast China reveals novel disease-associated ecological risk factors. *mSystems*. 9:e00842-23. <https://doi.org/10.1128/msystems.00842-23>
- Zitelli, L. C. *et al.* (2021). Serological investigation of protozoan pathogens (*Trypanosoma cruzi*, *Toxoplasma gondii* and *Neospora caninum*) in opossums from southern Brazil. *Veterinary Parasitology: Regional Studies and Reports*. 24: 100546. <https://doi.org/10.1016/j.vprsr.2021.100546>
- Zuo, T. *et al.* (2020). Human-Gut-DNA Virome Variations across Geography, Ethnicity, and Urbanization. *Cell Host & Microbe*. 28: 741-751. <https://doi.org/10.1016/j.chom.2020.08.005>

GLOSSÁRIO

Biblioteca: Fragmentos de genoma dentro de um conjunto de amostras que estão marcados com um código de barras e um adaptador molecular que tanto permitem o sequenciador reconhecer a presença do fragmento quanto reconhecer o conjunto de amostras a qual esse pertence;

Conjunto(s) de amostras: Junção de amostras de indivíduos diferentes, vindos de uma mesma área, para análise do viroma. Análogo a *pool(s)*;

Disbiose: Desequilíbrio desfavorável da microbiota/viroma tanto da composição quanto diversidade de vírus;

Doença: Condição que afeta negativamente o organismo e sua saúde;

Epizootico: Se refere a doenças que ocorrem em animais não humanos;

Espécie sentinela: Espécie que sua saúde pode ser usada como indicador da saúde do ambiente em que esta está inserida;

Fragmentos de maior tamanho: Fragmento de genoma formado a partir da junção de leituras que apresentam uma sequência de pares de bases complementar. Análogo à *contig(s)*;

Infecção viral: Penetração e desenvolvimento ou multiplicação de um vírus nas células dos hospedeiros (não necessariamente acarreta doença);

Leitura(s): Fragmento curto de genoma. Análogo à *read(s)*;

Transbordamento zoonótico: Quando um patógeno, que afeta um determinado hospedeiro animal, entrar em contato, se estabelece e passa a se replicar e a infectar ativamente um novo hospedeiro animal. Usado aqui de forma análoga à “transbordamento viral”;

Viroma: Toda a diversidade genética viral presente em uma amostra ou conjunto de amostras.

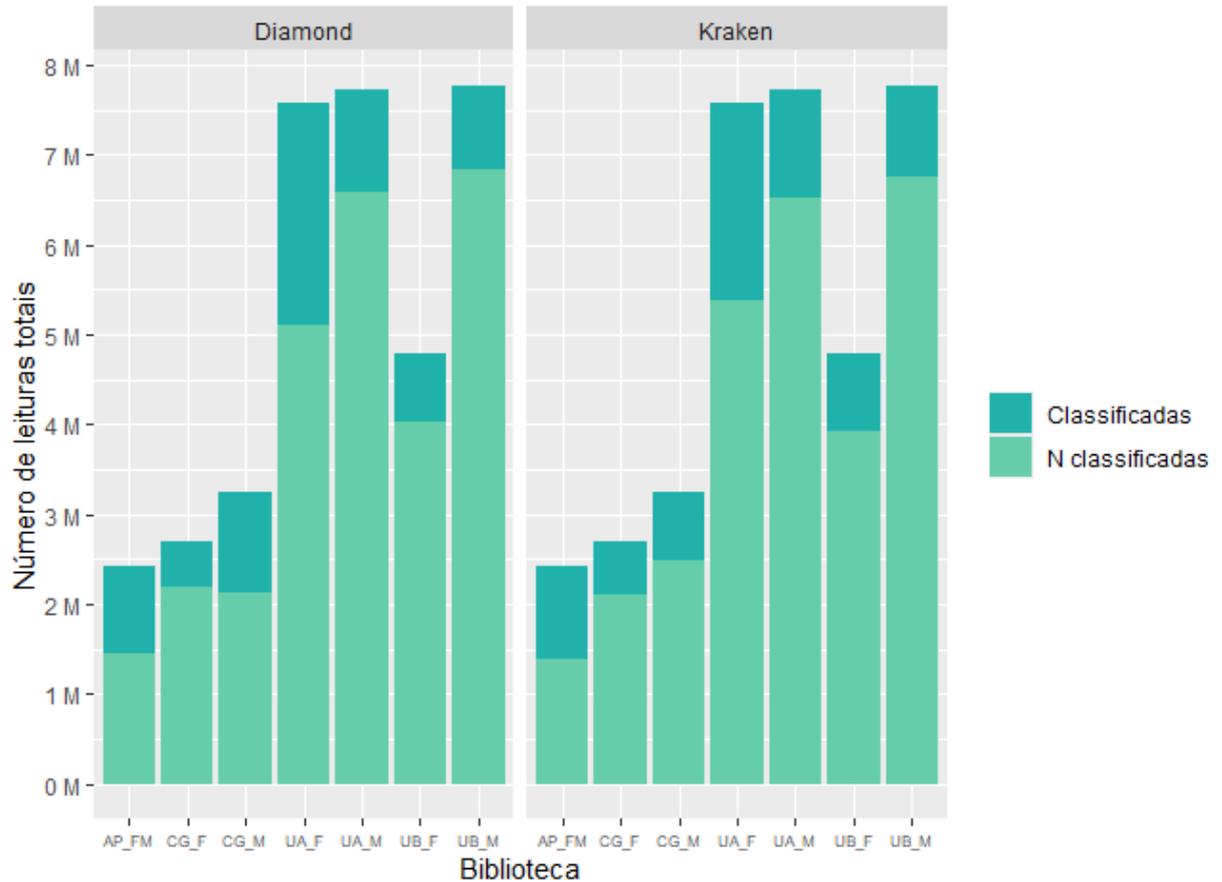
APÊNDICES

APÊNDICE A – Famílias virais identificadas nesse trabalho, com seus hospedeiros de acordo com o determinado pelo *site ViralZone* (Hospedeiro) e o hospedeiro padrão (Hosp. padrão) usado para construção dos mapas de calor

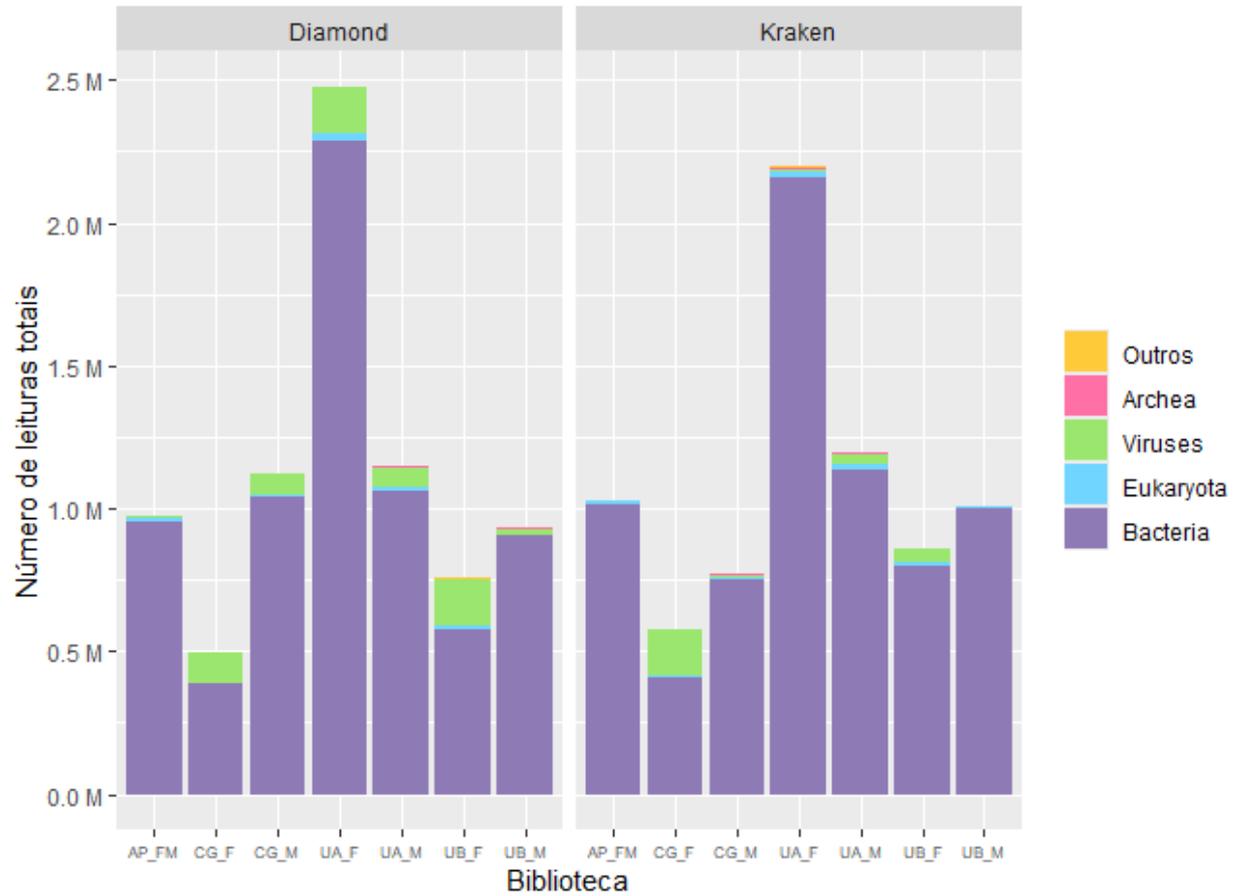
Família viral	Hospedeiro	Hosp. padrão
Ackermannviridae	Bacteria, Archaea	Bacteria, Archaea
Adenoviridae	Vertebrate	Craniata
Ascoviridae	Invertebrate	Invertebrados
Asfarviridae	Vertebrate, Invertebrate	Metazoa
Autographiviridae	Bacteria, Archaea	Bacteria, Archaea
Baculoviridae	Invertebrate	Invertebrados
Casjensviridae	Bacteria	Bacteria
Chaseviridae	Bacteria	Bacteria
Circoviridae	Vertebrate	Craniata
Coronaviridae	Vertebrate	Craniata
Demereciviridae	Bacteria, Archaea	Bacteria, Archaea
Dicistroviridae	Invertebrate	Invertebrados
Drexleriviridae	Bacteria	Bacteria
Fiersviridae	Bacteria	Bacteria
Fimoviridae	Vertebrate	Craniata
Flaviviridae	Vertebrate, Invertebrate	Metazoa
Genomoviridae	Vertebrate, Fungi	Eukaryota
Guelinviridae	Bacteria, Archaea	Bacteria, Archaea
Herelleviridae	Bacteria, Archaea	Bacteria, Archaea
Herpesviridae	Vertebrate	Craniata
Hytrosaviridae	Invertebrate	Invertebrados
Iflaviridae	Insect	Invertebrados
Inoviridae	Bacteria	Bacteria
Intestiviridae	Bacteria	Bacteria
Iridoviridae	Vertebrate, Invertebrate	Metazoa
Kyanoviridae	Bacteria	Bacteria
Mesyanzhinoviridae	Bacteria	Bacteria
Microviridae	Bacteria	Bacteria
Mimiviridae	Protozoa	Protozoa
Nudiviridae	Invertebrate	Invertebrados
Pachyviridae	Bacteria	Bacteria
Paramyxoviridae	Human, Vertebrate	Craniata
Parvoviridae	Vertebrate, Invertebrate	Metazoa
Peduviridae	Bacteria	Bacteria
Phenuiviridae	Ruminants, camels and humans	Metazoa
Phycodnaviridae	Algae	Algae
Picornaviridae	Vertebrate	Craniata
Potyviridae	Plant	Plantae

Poxviridae	Human, vertebrate, arthropods	Metazoa
Retroviridae	Vertebrate	Craniata
Roniviridae	Crustaceans, mostly prawns	Invertebrados
Rountreeviridae	Bacteria, Archaea	Bacteria, Archaea
Salasmaviridae	Bacteria	Bacteria
Schitoviridae	Bacteria, Archaea	Bacteria, Archaea
Smacoviridae	Unknown (2018), discovered using viral metagenomics from vertebrate faeces and insects.	N/A
Solinviridae	Insects	Invertebrados
Steigviridae	Bacteria	Bacteria
Straboviridae	Bacteria	Bacteria
Suoliviridae	Bacteria	Bacteria
Xinmoviridae	N/A	N/A

APÊNDICE B – Número de leituras (em milhões - M) que foram classificadas e não classificadas (“N classificadas”) pelas ferramentas *Diamond* e *Kraken2*, em cada uma das bibliotecas analisadas: UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes.



APÊNDICE C – Número de leituras (em milhões) atribuídas a cada grande domínio taxonômico pelas ferramentas *Diamond* e *Kraken2*, em cada uma das bibliotecas analisadas (UA_F – Fêmeas da União A; UA_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes).



APÊNDICE D – Tabela do número de *contigs* identificados pelo *Kraken2*, para todas as bibliotecas analisadas: UA_F – Fêmeas da União A; U_M - Machos da União A; UB_F - Fêmeas da União B; UB_M - Machos da União B; AP_FM - APA; CG_F - Fêmeas de Campos dos Goytacazes; e CG_M - Machos de Campos dos Goytacazes. O hospedeiro que cada família infecta pode ser encontrado no Apêndice A.

Família viral	UA_F	UA_M	UB_F	UB_M	AP_FM	CG_F	CG_M
Ackermannviridae	46	0	0	0	0	17	0
Adenoviridae	0	0	0	0	0	0	0
Ascoviridae	0	0	0	0	0	0	0
Asfarviridae	0	0	1	0	0	0	0
Autographiviridae	12	33	58	0	0	0	2
Baculoviridae	0	0	0	0	0	0	0
Casjensviridae	0	0	0	0	0	0	1
Chaseviridae	0	0	0	0	0	0	0
Circoviridae	0	0	1	0	0	0	0
Coronaviridae	1	0	0	0	1	1	0
Demereciviridae	0	0	68	0	0	304	0
Dicistroviridae	0	0	0	0	0	1	0
Drexleriviridae	169	94	128	7	0	7	86
Fiersviridae	0	0	0	0	0	0	0
Fimoviridae	0	0	0	0	0	0	0
Flaviviridae	0	0	0	0	0	0	0
Genomoviridae	0	0	0	0	0	0	0
Guelinviridae	0	0	0	3	0	0	0
Herelleviridae	1	1	2	1	0	0	0
Herpesviridae	0	1	0	0	0	0	0

Hytrosaviridae	0	0	0	0	0	0	0
Iflaviridae	0	0	0	0	0	0	0
Inoviridae	0	1	0	0	0	0	0
Intestiviridae	0	0	0	0	1	0	2
Iridoviridae	0	0	0	0	0	0	0
Kyanoviridae	0	0	0	0	0	0	0
Mesyanzhinovviridae	0	0	0	0	0	0	0
Microviridae	3	0	1	0	0	0	0
Mimiviridae	0	0	0	0	0	0	0
Nudiviridae	0	0	0	0	0	0	0
Pachyviridae	0	0	0	0	0	0	0
Paramyxoviridae	0	0	0	0	0	0	0
Parvoviridae	0	0	0	0	0	0	1
Peduoviridae	1	0	0	0	5	0	0
Phenuiviridae	0	0	0	0	0	0	0
Phycodnaviridae	1	0	0	0	0	0	0
Picornaviridae	0	0	0	0	0	0	0
Potyviridae	0	0	0	0	0	0	0
Poxviridae	0	0	0	0	0	0	0
Retroviridae	0	0	0	0	0	0	0
Roniviridae	0	0	0	0	0	0	0
Rountreeviridae	0	0	0	0	0	0	0
Salasmaviridae	1	0	0	0	0	0	0
Schitoviridae	1	0	0	0	0	1	2

Smacoviridae	1	0	1	0	0	0	1
Solinviridae	0	0	0	0	0	0	0
Steigviridae	3	3	24	1	2	1	9
Straboviridae	23	338	3	0	0	1138	49
Suoliviridae	11	26	147	0	3	3	29
Xinmoviridae	0	0	0	1	0	0	0