



**UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PÓS-GRADUAÇÃO EM BIOLOGIA MICROBIANA**

**ESTUDO DE VIROMA EM ÁGUA DE ESGOTO INFLUENTE E
EFLUENTE NA ESTAÇÃO DE TRATAMENTO DE ESGOTO
NORTE EM BRASÍLIA**

Nataly Mitie Natsume Moriya

**Brasília
2018**

NATALY MITIE NATSUME MORIYA

**ESTUDO DE VIROMA EM ÁGUA DE ESGOTO INFLUENTE E EFLUENTE NA
ESTAÇÃO DE TRATAMENTO DE ESGOTO NORTE EM BRASÍLIA**

Dissertação apresentada ao
Programa de Pós-Graduação em
Biologia Microbiana, do Instituto de
Ciências Biológicas da Universidade
de Brasília como requisito parcial
para obtenção do Título de Mestre
em Biologia Microbiana.

Orientador: Prof. Dr. Tatsuya Nagata

Brasília
2018

Dissertação de autoria de Nataly Mitie Natsume Moriya, intitulada
“ESTUDO DE VIROMA EM ÁGUA DE ESGOTO INFLUENTE E EFLUENTE NA
ESTAÇÃO DE TRATAMENTO DE ESGOTO NORTE EM BRASÍLIA”,
apresentada como requisito parcial para obtenção do grau de Mestre em Biologia
Microbiana da Universidade de Brasília, em 01 de março de 2018, defendida e
aprovada pela banca examinadora abaixo assinada:

Prof. Dr. Tatsuya Nagata (orientador)

Prof. Dr. Ricardo Henrique Krüger (examinador)

Prof. Dr. Bergmann Morais Ribeiro (examinador)

Profa. Dra. Eliane Ferreira Noronha (suplente)

Brasília
2018

*“Natura nusquam magis est tota
quam in minimis”*

(Em parte alguma encontramos a
natureza na sua totalidade como nas
suas menores criaturas)

Plínio

RESUMO

A contaminação por patógenos humanos na água de esgoto é um assunto importante, principalmente em locais onde as condições de saneamento não são ideais. A tecnologia de *Next Generation Sequencing* (NGS) tem sido utilizada em diversos estudos metagenômicos virais (viromas). Além de descrever a diversidade genética, é uma ferramenta útil para analisar o viroma em amostras de esgoto. Apesar do interesse no tratamento de água, a cidade de Brasília não possui um estudo de censo e informações focado na contaminação viral do Lago Paranoá. Com este propósito, os patógenos humanos astrovírus (HAstV), norovírus (NoV) e sapovírus (SaV) foram selecionados e investigados em amostras de água não-tratada (influyente) e tratada (efluente) utilizando o NGS. Amostras de influente e efluente (de março, maio de 2016 e fevereiro de 2017) foram coletadas na estação de tratamento de água de esgoto em Brasília, Brasil. As amostras foram submetidas a processos de extração de RNA e as bibliotecas de cDNA foram sequenciadas através do Illumina HiSeq 2000, com a condição de 100 bases *paired-end*. Os *reads* foram trimados através do Trimmomatic e os *contigs* foram montados com os programas Megahit e Velvet. As sequências foram analisadas por tBLASTX e Megablast, respectivamente, contra o "RefSeq Virus" utilizando o software Geneious v.8.1. Os *reads* resultantes também foram enviados e analisados na plataforma online Kaiju. Os SaV foram detectados nas seis amostras sequenciadas, HAstV e NoV foram encontrados em todas as amostras, exceto no efluente de maio. Os resultados encontrados foram discutidos e comparados com estudos brasileiros e internacionais.

Palavras-chave: *Next-generation sequencing*. Esgoto. Viroma.

ABSTRACT

Contamination of human pathogen in wastewater is an important matter, especially where sanitation is not ideal condition. Next Generation Sequencing (NGS) has been applied in several viral metagenomes (viromes) studies. Besides describing gene diversity, it is a helpful tool to analyze virome in sewage water. Despite the interest in wastewater treatment, Brasilia city lacks a census study and information aiming to assess viral contamination in Paranoa lake. For this purpose, the human pathogens astrovirus (HAstV), norovirus (NoV) and sapovirus (SaV) were selected in untreated (influent) and treated (effluent) wastewater and investigated using NGS technology. Samples of influent and effluent stream (from March, May of 2016 and February of 2017) were collected at the treatment station of wastewater in Brasilia, Brazil. Samples were submitted to RNA extraction process and the cDNA libraries were sequenced using Illumina HiSeq 2000 with the condition of 100 base paired-end. The NGS reads were trimmed by Trimmomatic and the contigs were assembled with Megahit and Velvet. The assembled contigs were analyzed by tBLASTX and Megablast, respectively, against RefSeq Virus using Geneious Software v.8.1. Resulted reads were also uploaded and analyzed on Kaiju online platform. SaV were detected in the six sequenced samples, HAstV and NoV were found in all samples except the May untreated water sample. The results found were discussed and compared with Brazilian and international studies.

Keywords: Next Generation Sequencing. Sewage. Virome

LISTA DE ABREVIATURAS

CAESB – Companhia de Saneamento Ambiental do Distrito Federal

DF – Distrito Federal

ETE – Estação de Tratamento de Esgoto

HAstV – Astrovírus humano

NGS – Next Generation Sequencing

NoV – Norovírus

ORF – Open reading frame

PCR – Polimerase Chain Reaction

RT – Reverse Transcription

RT-qPCR – Real Time Quantitative reverse transcription *PCR*

SaV – Sapovírus

THM - Trihalometanos

WHO – World Health Organization

SUMÁRIO

Resumo.....	04
Abstract.....	05
Lista de Abreviatura.....	06
1 INTRODUÇÃO.....	09
1.1 ESTAÇÃO DE TRATAMENTO DE ESGOTO.....	12
1.2 ASTROVÍRUS.....	16
1.3 NOROVÍRUS.....	18
1.4 SAPOVÍRUS.....	22
2 OBJETIVO GERAL.....	26
2.1 Objetivos específicos.....	26
3 MATERIAIS E MÉTODOS.....	27
3.1 AMOSTRAGEM.....	27
3.2 ENRIQUECIMENTO DA FRAÇÃO VIRAL.....	27
3.3 EXTRAÇÃO DO RNA TOTAL.....	28
3.4 NEXT GENERATION SEQUENCING (NGS).....	28
3.5 ANÁLISE DAS SEQUÊNCIAS.....	28
3.5.1 Construção dos gráficos.....	30
3.6 MONTAGEM DAS ÁRVORES FILOGENÉTICAS.....	30
4 RESULTADOS.....	31
4.1 AVALIAÇÃO DE QUALIDADE DE LEITURA COM FASTQC, TRIMAGEM E LIMPEZA DAS SEQUÊNCIAS DE <i>READS</i>	31
4.2 CONSTRUÇÃO DOS GRÁFICOS METAGENÔMICOS ATRAVÉS DA PLATAFORMA KAIJU	34
4.3 <i>DE NOVO ASSEMBLY</i> DOS <i>READS</i> TRIMADOS UTILIZANDO O PROGRAMA MEGAHIT E ANÁLISE DOS <i>CONTIGS</i> GERADOS COM TBLASTX.....	39

4.4	<i>DE NOVO ASSEMBLY</i> DOS <i>READS</i> TRIMADOS UTILIZANDO O PROGRAMA VELVET E ANÁLISE DOS <i>CONTIGS</i> GERADOS COM MEGABLAST	41
4.5	MONTAGEM DOS GENOMAS E ÁRVORE FILOGENÉTICA ATRAVÉS DO PROGRAMA GENEIOUS	44
5	DISCUSSÃO	47
6	CONCLUSÃO	51
7	PERSPECTIVAS	52
	REFERÊNCIAS	53
	APÊNDICES	64

1. INTRODUÇÃO

Os vírus causadores de gastroenterite são destaque dentre os patógenos importantes circulantes na população humana e fundamentais no sistema de monitoramento da saúde pública. A gastroenterite é caracterizada pela inflamação do trato gastrointestinal, gerando sintomas como diarreia, vômito e cólicas abdominais (Schlossberg, 2015). Seu diagnóstico pode ser complexo, pois diversos vírus podem ser potenciais patógenos, apresentando sintomas clínicos semelhantes (Nakamura *et al.*, 2009).

De acordo com a Organização Mundial de Saúde, a diarreia é a segunda maior causa de morte em crianças menores de cinco anos, gerando cerca de 525 mil óbitos anualmente no mundo, na maioria das vezes resultantes da ingestão de água e/ou comida contaminadas. Também pode ser transmitida de pessoa-a-pessoa, situação agravada pela higiene pessoal precária (WHO, 2017).

A água contaminada por fezes humanas através do esgoto é de preocupação em particular. Em caso de tratamento ineficiente ou desvio de água não tratada, o esgoto pode contaminar fontes de água potável, gerando riscos epidemiológicos e surtos diarreicos (Le Guyader *et al.*, 2008). Prevost e colaboradores (2015) monitoraram a água do rio Sena em Paris ao longo de um ano à procura de vírus gastroentéricos. Foi obtido, por meio desse estudo, uma enorme diversidade viral, fornecendo uma visão geral da dinâmica do rio ao redor da cidade. Concluiu-se também que a maior parte da fonte contaminante provia das estações de tratamento de águas residuais. É possível que os vírus entéricos possuam maior tempo de vida quando livres no meio ambiente, quando comparados com os demais vírus, devido a sua resistência ao calor, alta pressão e pH (Rzezutka e Cook, 2004).

Os vírus entéricos são excretados em grande quantidade nas fezes de um indivíduo infectado (cerca de 10^{11} partículas virais por grama de fezes) (Fong e Lipp, 2005), fazendo com que o monitoramento ambiental utilizando apenas bactérias não seja um bom indicador de qualidade da água.

Para Gensberger e Kostić (2013), existem quatro grandes desafios ao analisar uma amostra ambiental: a baixa concentração dos vírus alvos, a presença de inibidores de reação enzimática, a emergência de novas cepas/variantes virais, além da correlação entre a presença dos ácidos nucleicos virais com a infectividade. A análise de metagenomas virais (viromas) através do *Next-Generation Sequencing* (NGS) é uma boa alternativa para superar tais desafios e tem sido utilizada em amostras ambientais, como oceanos, rios e água reutilizável. A tabela 1 exemplifica alguns estudos, realizados em diversas partes do mundo, que utilizam plataformas de sequenciamento para detecções de vírus patógenos humanos em diferentes tipos de resíduos.

Tabela 1 – Estudos envolvendo metagenômica e vírus patógenos humanos.

Tipo de amostra e país de coleta	Exemplos de diversidade viral detectada	Plataforma de sequenciamento	Referência
Águas residuais (EUA)	<i>Adenoviridae</i> , <i>Polyomaviridae</i> , <i>Papillomaviridae</i> , <i>Picornaviridae</i>	Illumina HiSeq	(Aw <i>et al.</i> , 2014)
Esgoto bruto (Dinamarca)	<i>Adenoviridae</i> , <i>Caliciviridae</i> , <i>Astroviridae</i> , <i>Picobirnaviridae</i>	Illumina MiSeq	(Hjelmsø <i>et al.</i> , 2017)
Amostra fecal (Japão)	<i>Reoviridae</i> (Rotavírus)	Illumina MiSeq	(Mizukoshi <i>et al.</i> , 2014)
Soro, plasma, fezes (China)	<i>Flaviviridae</i> (Dengue), <i>Hepadnaviridae</i> (Hepatite A, C) <i>Retroviridae</i> (HIV), <i>Caliciviridae</i> (Norovirus)	Illumina HiSeq Illumina MiSeq 454 GS Junior	(Li <i>et al.</i> , 2016)
Lodo de esgoto (EUA)	<i>Astroviridae</i> , <i>Caliciviridae</i> , <i>Hepadnaviridae</i> , <i>Reoviridae</i>	Illumina HiSeq 454 Life Science	(Bibby <i>et al.</i> , 2011)
Água “reciclada” (EUA)	<i>Picornaviridae</i> (Aichi vírus), Enterovirus, Parechovirus, Rhinovirus	454 Life Science	(Rosario <i>et al.</i> , 2009)

A metagenoma tem seu custo ter reduzido a cada ano desde 2001, o que a torna uma ferramenta mais acessível (Mardis , 2011). O custo do sequenciamento por genoma (humano) que chegava a custar cerca de cem mil dólares em 2010, atualmente pode custar menos de mil dólares, conforme dados gerados pelo *National Human Genome Research Institute* (Wetterstrand, 2017). Além disso há a vantagem da metagenômica utilizar um pool genômico de

microrganismos de um determinado ambiente por técnicas independentes de cultivo, não necessitando de uma sequência genômica já publicada.

Os principais vírus de RNA de fita simples causadores de gastroenterite encontrados na amostra de água de esgoto a serem estudados são: astrovírus (*Astroviridae*), norovírus e sapovírus (ambos *Caliciviridae*), indicados na Tabela 2. Os rotavírus não foram incluídos neste estudo por serem vírus de RNA fita dupla.

Tabela 2 – Vírus de RNA patógenos humanos incluídos no estudo.

Nome comum	Nome científico	Gênero	Família	Ácido Nucleico	Tamanho de genoma (kb)
Astrovírus	<i>Human astrovirus</i>	<i>Mamastrovirus</i>	<i>Astroviridae</i>	ssRNA	6.8~7
Norovírus	<i>Norwalk virus</i>	<i>Norovirus</i>	<i>Caliciviridae</i>	ssRNA	7.3~8.3
Sapovírus	<i>Sapporo virus</i>	<i>Sapovirus</i>	<i>Caliciviridae</i>	ssRNA	7.3~8.3

1.1. ESTAÇÃO DE TRATAMENTO DE ESGOTO

O esgoto é definido como todos os resíduos líquidos industriais e domésticos que necessitam de tratamento adequado, buscando remover as impurezas para que assim as águas resultantes desse processo possam ser devolvidas ao meio ambiente sem causar danos a sua fauna e flora.

O esgoto doméstico é composto principalmente dos dejetos domiciliares provenientes das edificações. Além disso, o esgoto hospitalar é incluído nessa categoria, pois não há separação específica para esses rejeitos no Brasil. A água pluvial também se encaixa na categoria de esgoto doméstico. Já os esgotos industriais devem ser previamente tratados antes de despejados junto ao esgoto doméstico pois, podem ter grandes teores de compostos tóxicos e metais pesados, devido aos processos produtivos das fábricas.

Aproximadamente 99,8% ~ 99,9% do esgoto é composto por água, os sólidos representam de 0,1% a 0,2%. Dos sólidos, 70% são orgânicos e o restante inorgânico. Na porção orgânica são encontrados diversos

microrganismos, que auxiliam na decomposição e estabilização da matéria, porém podem ser potenciais patógenos (Batista, 2015).

No presente estudo as amostras foram coletadas na Estação de Tratamento de Esgoto (ETE) Norte da Companhia de Saneamento Ambiental do Distrito Federal – CAESB (Figura 1). O complexo tem capacidade de vazão de 920 L/s, com média anual de 450 L/s, atendendo cerca de 260 mil habitantes, incluindo as regiões administrativas da Asa Norte, Vila Planalto, parte da área central de Brasília, Lago Norte, Taquari, Vila Estrutural, Vila Varjão e Torto (CAESB, 2017).

O tratamento de esgoto é dividido em fase preliminar, tratamento primário, secundário e terciário. No tratamento preliminar é empregado o uso de caixas de areia ou peneiras para reter resíduos maiores, como plásticos e lixo, impedindo que estes causem danos aos equipamentos do sistema e facilitando o transporte e fluidez do esgoto. Aqui, o esgoto é separado em fase sólida e líquida. Os sólidos são os denominados lodos e recebem um tratamento à parte, através de decantadores e digestores, para depois serem desidratados e despejados em locais apropriados ou reutilizados, por exemplo, em agricultura e reflorestamento.

A fase líquida ainda possui bastante matéria orgânica, logo é encaminhada para a etapa de tratamento secundário/terciário, onde são empregados organismos aeróbios, anaeróbios e facultativos para assimilar os nutrientes e a matéria. Tais microrganismos podem ser reutilizados em outro processo ou descartados juntamente com o lodo na fase sólida.

O líquido que é clarificado nos decantadores secundários passa então por um tratamento químico (polimento final), onde o fósforo e sólidos remanescentes são retirados, através de processos físico-químicos. Há a separação por flotação e finalmente o efluente é lançado no lago Paranoá (Batista, 2015).



Figura 1 – ETE Norte. Fonte: Google Maps.

Existem diversas opções técnicas para a desinfecção de esgotos sanitários. A escolha das melhores técnicas deve levar em consideração o uso posterior da água a ser tratada, assim como os riscos de saúde associados a ela. Deve haver uma avaliação das alternativas disponíveis para o controle dos esgotos contaminados por patógenos, além da avaliação dos impactos ambientais que as medidas de desinfecção podem ocasionar (Chernicharo *et al.*, 2001). Os patógenos a serem considerados abrangem vírus, bactérias, protozoários e helmintos.

Na desinfecção dos esgotos com compostos clorados, a inativação viral ocorre pela modificação nos ácidos nucleicos e na capa proteica. Porém, a presença de sólidos no efluente pode proteger os microrganismos da ação do cloro. Globalmente é feita a desinfecção de efluentes através do cloro, porém a adição deste pode causar efeitos indesejáveis: ao reagir com ácidos fúlvicos e húmicos, provenientes da decomposição de matéria orgânica, o cloro faz com que tais ácidos liberem radicais de cetona, formando halofórmios como o triclorometano, bromodiclorometano, dibromoclorometano e tribromometano, conhecidos como trihalometanos (THM), sendo o clorofórmio o mais facilmente detectado. Os THM são importantes pois foram associados à certos tipos de

câncer como o de estômago, intestino e cólon (Meyer, 1994). Ademais, o cloro não é um oxidante poderoso o bastante para eliminar completamente os vírus.

O ozônio possui maior poder oxidativo do que o cloro, sendo bastante eficiente na destruição de vírus, bactérias e protozoários. Nos vírus, o ozônio ataca as proteínas do capsídeo, estas responsáveis pela ligação do vírus na célula hospedeira (no caso vírus não-envelopado), diminuindo assim a capacidade infectante. Em altas concentrações, o ozônio pode destruir completamente o capsídeo e agir sobre os ácidos nucleicos (Meyer, 1994).

Quanto à desinfecção por luz ultravioleta (UV), de modo geral, bactérias e vírus são muito sensíveis à radiação UV, sendo doses na ordem de 20mWs/cm² bastante efetivas. Davies-Colley e colaboradores (1999) identificaram, em lagoas de estabilização, que os componentes do espectro solar UVA (320-400nm), UVB (290-320nm) e a o espectro visível azul/verde (400-550nm) colaboram com a desinfecção do efluente, pois foram capazes de inativar *Escherichia coli* e fagos. A tabela 3 resume a efetividade de alguns processos de sanitização sobre os organismos patógenos encontrados no esgoto.

Tabela 3 – Efetividade dos processos sobre os organismos patógenos.

Efeito sobre organismo	Cloração	Cloração/descloração	Ozônio	UV	Lagoas de Estabilização
Bactérias	Bom	Bom	Bom	Bom	Bom
Vírus	Ruim	Ruim	Bom	Bom	Bom
Protozoários e helmintos	Regular	Regular	Regular	Pouco	Bom

Adaptado de: Programa de Pesquisas em Saneamento Básico (PROSAB).

Os vírus que causam gastroenterite como astrovírus, norovírus e sapovírus são capazes de sobreviver em torno de duas semanas na água e um pouco mais nos esgotos, devido à adsorção a partículas em suspensão e partículas virais resistentes que não possuem envelope (Vieira *et al.*, 2012). A ausência de envelope lipídico, assim como o fato de que o capsídeo proteico apresenta carga negativa em pH natural das águas (5 a 9), facilita a adsorção viral a partículas do meio e confere maior proteção contra degradação/inativação química, tornando-os capazes de resistir aos processos de tratamento de água

e esgoto aplicados no controle bacteriano, inclusive cloração (Bosch 1998, Griffin et al. 1999).

1.2 ASTROVÍRUS

Os astrovírus humanos (HAstVs) são agentes de gastroenterites com distribuição mundial, sendo predominantemente considerados causadores de casos pediátricos e geriátricos, embora também ocorra em adultos (Bosch *et al.*, 2014, Vu *et al.*, 2017). Alguns estudos determinaram os HAstVs como segunda maior causa de gastroenterites em crianças, atrás do rotavírus. A incidência varia de 4,2% a 7,3% (Dalton *et al.*, 2002.; Walter e Mitchell., 2003). No Brasil houve um relato de surto, no qual a incidência chegou a 56% (Gabbay *et al.*, 2006). Em geral as infecções em humanos ocorrem no inverno, em regiões temperadas, e na estação chuvosa, no caso de regiões tropicais (Cruz *et al.*, 1992; Palombo *et al.*, 1996, Walter e Mitchell., 2003).

Foi primeiramente reconhecido em amostras de fezes diarreicas em crianças hospitalizadas no Reino Unido, através de microscopia eletrônica. O termo *astrovírus* foi dado para descrever vírus pequenos, arredondados com uma estrutura distinta de cinco ou seis pontas, com aparência de estrela (*astron*, estrela em grego), de cerca de 28 a 30nm em diâmetro (Madeley e Cosgrove, 1975).

A família *Astroviridae* inclui astrovírus humanos, com cerca de 6.17kb e animais de até 7.72kb. Com a morfologia icosaédrica, estes vírus não-envelopados apresentam genoma de ssRNA (RNA fita simples) senso positivo, composto por três *open-reading frames* (ORFs) (Figura 2). As duas ORFs localizadas na ponta 5', chamadas ORF1a e ORF1b, codificam proteínas não-estruturais, envolvidas no processo de transcrição e replicação viral. A ORF2, localizada na extremidade 3' do genoma codifica a proteína do capsídeo.

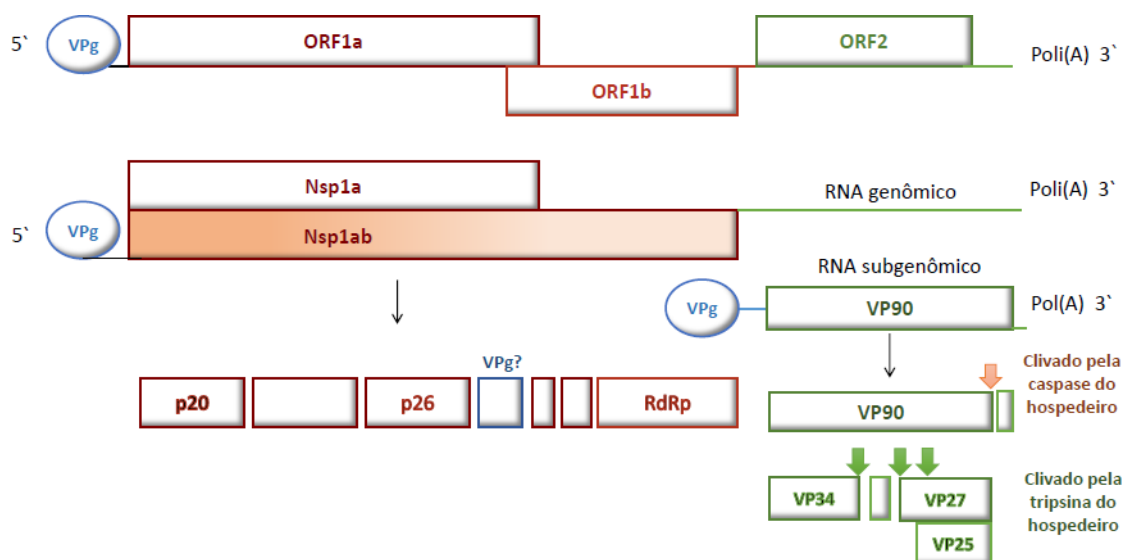


Figura 2 – Esquema representativo do genoma do Astrovírus. VPg= Viral Protein genome-linked, Nsp=Non-structural protein, p20=membrane associated protein, p26= serine protease, RdRp=RNA-dependent RNA polymerase, VP90=capsid polyprotein, VP34= capsid protein p34, VP27=capsid protein p27 e VP25=capsid protein p25.

Os astrovírus de mamíferos infectam geralmente as células epiteliais próximas às microvilosidades do intestino delgado, não sendo observada no estômago (Sebire *et al.*, 2004). Porém, em 2010 o HAstV foi detectado como patógeno único no sistema nervoso central (SNC) de um paciente imunodeficiente, que veio a falecer de encefalite. As estruturas virais foram encontradas nos astrócitos (Quan *et al.*, 2010). O vírus que afeta o SNC foi geneticamente ligado à uma nova cepa de HAstV (cepa VA1) associada à um surto de gastroenterite (Finkbeiner *et al.*, 2009). Astrovírus que afetam o SNC foram descritos em mamíferos mustelídeos (*mink* ou *vison*) através de análise metagenômica (Blostrom *et al.*, 2010).

Os HAstVs clássicos são classificados em oito sorotipos (HAstV-1 a HAstV-8), porém desde 2008, dois tipos de astrovírus diferentes, nomeados MLB (Melbourne) e VA/HMO (Virginia/Human-Mink-Ovine like) foram identificados em amostras fecais humanas, através de NGS (Kapoor *et al.*, 2009; Finkbeiner *et al.*, 2008). O HAstV tipo 1 é o mais comum em crianças (Monroe *et al.*, 2003).

O HAstV foi detectado em várias fontes de água, como água potável e recreacional, rios, esgotos e água tratada (Le Cann *et al.*, 2004; Maunula *et al.*, 2004; Nadan *et al.*, 2003; Pusch *et al.*, 2005). A infectividade do HAstV pode ser mantida durante meses, além disso, é resistente à cloração, que é uma das principais formas de tratamento de esgoto (Espinosa *et al.*, 2008). O HAstV também é resistente à desinfetantes normais (Schultz-Cherry *et al.*, 2001).

Porém, a maior fonte de infecção por HAstV é a comida contaminada. O vírus foi encontrado, juntamente com outros causadores de gastroenterite aguda (GA), em moluscos (Vilariño *et al.*, 2009), o que mostra que frutos do mar podem contribuir na incidência da doença.

1.3 NOROVÍRUS

Norovírus (NoV) é nome comum do gênero *Norovirus* que possui única espécie tipo, *Norwalk virus* é o causador de gastroenterites virais mais importante em vários países. De acordo com o "Centro de Controle e Prevenção dos Estados Unidos da América" (*Centers for Disease Control and Prevention - CDC, EUA*), 93% dos casos estudados entre 1997 e 2000 foram atribuídos aos NoV (Frankhauser *et al.*, 2002). NoV são responsáveis por 19 a 21 milhões de casos de gastroenterite aguda (GA) anualmente nos EUA (<https://www.cdc.gov/norovirus/trends-outbreaks.html>). Os surtos ocorrem frequentemente em locais de confinamento, como por exemplo asilos, campus, áreas recreacionais, escolas primárias, restaurantes, navios de cruzeiro, exército, prisões e creches. (Bitler *et al.*, 2013; Blanton *et al.*, 2006; Frankhauser *et al.*, 2002; Lopman *et al.*, 2003).

É o segundo maior agente causador de GA no mundo, atrás apenas do rotavírus, tal posição faz com que aproximadamente duzentas mil mortes de crianças menores de cinco anos ocorram anualmente em países em desenvolvimento. É responsável pela internação por GA de 12% dos pacientes pediátricos, também são associados a infecções e doenças em todas as faixas etárias (Patel *et al.*, 2008), inclusive causando surtos em idosos, em especialos que vivem em instituições como asilos. (Frankhauser *et al.*, 2002).

Estima-se que o NoV cause cerca de 60 bilhões de dólares de prejuízo anualmente. Apesar de o NoV não ser letal na maioria dos casos, em alguns

países entre 85% e 90% dos prejuízos causados pelo vírus são devido à perda de produtividade, ou seja, cidadãos que não podem contribuir com seu trabalho para com a sociedade por estarem muito doentes (Bartsch *et al.*, 2016). Tais valores podem ser subestimados, já que a maioria dos casos não é reportada e muitos países, como o Brasil, não possuem uma rede de monitoramento de calicivírus humanos.

É sabido que ocorrem surtos mais frequentes na época do inverno (Patel *et al.*, 2008), com exceção de um surto de GII.4 em 2002/3, que ocorreu na primavera/verão da Europa (Lopman *et al.*, 2004). Ahmed e colaboradores (2013) em sua revisão sobre a sazonalidade dos NoV, estabeleceram que dentre os 68 estudos analisados, 52% dos casos e 41,2% dos surtos ocorreram em meses de inverno (dezembro a fevereiro no hemisfério norte e junho a agosto no hemisfério sul), porém não houve estudos o suficiente para mostrar dados consistentes sobre surtos no inverno ou verão em países do hemisfério sul como Austrália e Brasil. Estações chuvosas parecem ser um importante fator para a sazonalidade dos NoVs, devido à sua fonte de transmissão através de água contaminada.

Devido à dificuldade de propagação viral em cultura celular pouco se sabe acerca das respostas imunológicas aos NoV. Em alguns surtos virais, mais de 80% dos adultos de uma comunidade ficam doentes (Kaplan *et al.*, 1982). O vírus é associado a infecções prolongadas em pacientes imunocomprometidos (Kaufman *et al.*, 2005; Nilsson *et al.*, 2003; Vilariño *et al.*, 2009), podendo aumentar o risco de desenvolvimento de pneumatose intestinal (cistos gasosos entre as paredes intestinais) (Kim *et al.*, 2011).

A transmissão pode ocorrer entre pessoas, pela rota fecal-oral. A transmissão através da via respiratória não foi estabelecida. Em trabalhos com voluntários infectados, foi determinado que o vômito pode ser considerado uma fonte de transmissão (Kirby *et al.*; 2016).

Os NoV também são encontrados em alimentos contaminados, como ostras e frutos do mar. Porém, na maioria dos casos, não é possível determinar a fonte contaminante (Widdowson *et al.*, 2005). Sabe-se que os NoV são resistentes ao tratamento de esgoto, sendo detectados em amostras de água poluída (Laverick *et al.*, 2004; Le Guyader *et al.*, 2000; Myrmel *et al.*, 2006; Ueki *et al.*, 2005).

Estudos mostram que os vírions mantêm-se estáveis na presença dos ácidos estomacais, fazendo com que consigam sobreviver à passagem através do estômago (Calderon-Margalit *et al.*, 2005). Dolin e colaboradores (1972), em estudos em voluntários expostos a filtrado de fezes com pH menor que 3 por 3 horas a temperatura ambiente; tratamento com éter 20% por 18 horas à 4°C ou incubação a 60°C por 30 minutos, determinaram haver capacidade viral infectante sobre os pacientes. Os vírions sobrevivem em água subterrânea em temperatura ambiente por cerca de 61 dias (Seitz *et al.*, 2011). Apresentam resistência a detergentes e alvejantes a base de etanol, usados na limpeza de superfícies e fômites, portanto é necessário a utilização de outros químicos na desinfecção. Alguns químicos como hipoclorito (5000 ppm), alvejantes a base de peróxido de hidrogênio e limpadores a base de fenol se mostraram eficientes na desinfecção (Hall, 2011).

Os NoV pertencem à família *Caliciviridae*, composto por vírus pequenos, icosaédricos e não envelopados, que possuem genoma de ssRNA positivo, de aproximadamente 7,3 a 7,5kb. A ORF1 codifica uma poliproteína, que é processada em proteínas não-estruturais maduras (Komolafe *et al.*, 1986.; Liu *et al.*, 1996) e essenciais para replicação de genoma (Figura 3). A ORF2 codifica a proteína do capsídeo maior, VP1 e a ORF3 codifica a estrutura do capsídeo VP2, esta é idêntica em todos os calicivírus.

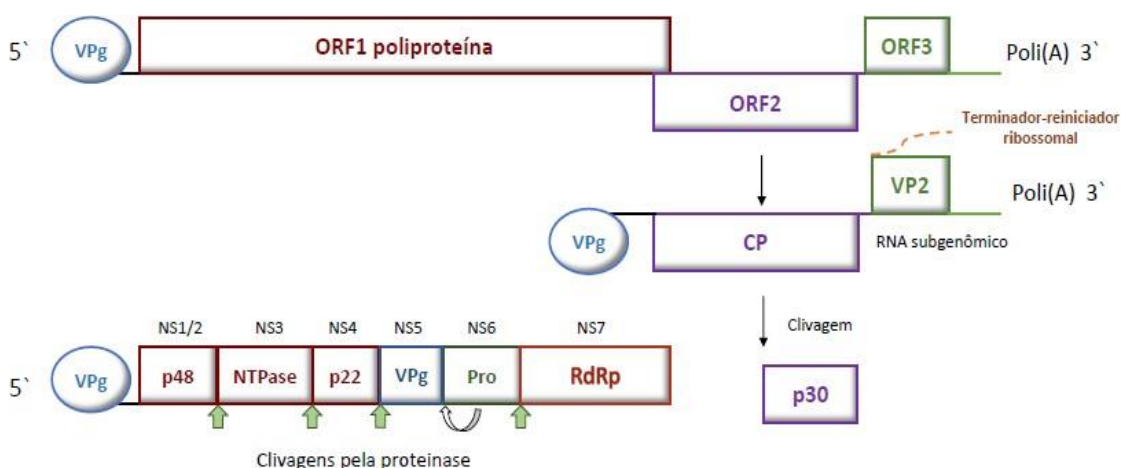


Figura 3 - Esquema representativo do genoma do Norovírus. VPg= Viral Protein genome-linked, CP= coat protein, VP2= minor structure protein, NS1/2 à NS7= mature non-structural proteins, RdRp=RNA-dependent RNA polymerase, NTPase = nucleoside triphosphatase, Pro= chymotrypsin-like cysteine proteinase, p48= protein 48, p22= protein 22, p30= protein 30.

Os norovírus são divididos em cinco genogrupos, do GI ao GV, sendo que GI, GII e GIV infectam humanos (Zheng *et al.*, 2006). Cada genogrupo pode em ser dividido em genótipos, pelo menos 36 deles são reconhecidos até agora. A tabela 4 resume os genogrupos e genótipos de norovírus humanos. O GII é responsável por aproximadamente 92% dos casos (Lopman *et al.*, 2004). Os vírus do grupo GII.4 são os mais predominantes mundialmente. De acordo com o CDC, 81,5% dos casos de GA de 1994 a 2006 foram causados pelo GII, e destes, 44,2% pelo GII.4 (Zheng *et al.*, 2010). É frequente a ocorrência de mudanças entre os genótipos circulantes, podendo haver infecções por múltiplos genótipos em um único indivíduo ou em um mesmo surto (Bull *et al.*, 2005).

Atualmente a técnica mais utilizada na detecção de NoV é a Transcrição Reversa – Reação em cadeia da Polimerase (RT-PCR), em amostras clínicas, água e alimentos contaminados. O RT-qPCR (RT-PCR quantitativa) também é bastante usado por permitir detecção rápida e comparação dos níveis de RNA viral. O uso destas técnicas juntamente com o sequenciamento tem sido aplicado na detecção e caracterização de NoV em surtos (Butot *et al.*; 2010; Kageyama *et al.*, 2003; Laverick *et al.*, 2004).

Tabela 4 – Genogrupos e genótipos de Norovírus humano determinados pela relação da VP1.

Vírus de referência	Genogrupo	Genótipo	GenBank
Hu/NoV/GI.1/Norwalk/1968/US	I	1	M87661
Hu/NoV/GI.2/Southampton/1991/UK	I	2	L07418
Hu/NoV/GI.3/Desert Shield 395/1990/SA	I	3	U04469
Hu/NoV/GI.4/Chiba 407/1987/JP	I	4	AB022679
Hu/NoV/GI.5/Musgrove/1989/UK	I	5	AJ277614
Hu/NoV/GI.6/BS5(Hesse3)/1997/DE	I	6	AF093797
Hu/NoV/GI.7/Winchester/1994/UK	I	7	AJ277609
Hu/NoV/GI.8/Boxer/2001/US	I	8	AF538679
Hu/NoV/GI.9/Vancouver730/2004/CA	I	9	HQ637267
Hu/NoV/GII.1/Hawaii/1971/US	II	1	U07611
Hu/NoV/GII.2/Melksham/1994/UK	II	2	X81879
Hu/NoV/GII.3/Toronto 24/1991/CA	II	3	U02030
Hu/NoV/GII.4/Bristol/1993/UK	II	4	X76716
Hu/NoV/GII.5/Hillingdon/1990/UK	II	5	AJ277607
Hu/NoV/GII.6/Seacroft/1990/UK	II	6	AJ277620
Hu/NoV/GII.7/Leeds/1990/UK	II	7	AJ277608
Hu/NoV/GII.8/Amsterdam/1998/NL	II	8	AF195848
Hu/NoV/GII.9/VA97207/1996/US	II	9	AY038599
Hu/NoV/GII.10/Erfurt546/2000/DE	II	10	AF427118
Hu/NoV/GII.12/Wortley/1990/UK	II	12	AJ277618
Hu/NoV/GII.13/Fayetteville/1998/US	II	13	AY113106
Hu/NoV/GII.14/M7/1999/US	II	14	AY130761
Hu/NoV/GII.15/J23/1999/US	II	15	AY130762
Hu/NoV/GII.16/Tiffin/1999/US	II	16	AY502010
Hu/NoV/GII.17/CS-E1/2002/US	II	17	AY502009
Hu/NoV/GII.20/Luckenwalde591/2002/DE	II	20	EU373815
Hu/NoV/GII.21/IF1998/2003/IR	II	21	AY675554
Hu/NoV/GII.22/Yuri/2002/JP	II	22	AB083780
Hu/NoV/GIV.1/Alphatron 98–2/1998/NL	IV	1	AF195847

1.4 SAPOVÍRUS

Os sapovírus (SaV) são nomes comuns para os vírus do gênero *Sapovirus* que possui única espécie, *Sapporo virus* e causam GA em animais e humanos em qualquer faixa etária. Os sintomas clínicos são indistinguíveis daqueles causados por norovírus e incluem náusea, diarreia, vômito e febre. Podem durar de 1 a 6 dias (Noel *et al.*, 1997; Rock *et al.*, 2002; Johansson *et al.*, 2005). Em termos morfológicos, o SaV se distingue (junto com norovírus) de outros

patógenos de gastroenterites devido à sua morfologia típica em forma de “Estrela de Davi” (Madeley, 1979, Cubitt *et al.*, 1987).

Eles pertencem ao gênero *Sapovirus*, da família *Caliciviridae*. Foram descobertos em 1976, por Madeley e Cosgrove, através de microscopia eletrônica de amostras de fezes humanas. O vírus utilizado como modelo foi o de um surto na cidade de Sapporo, no Japão em 1982 (Chiba *et al.*, 1979). Surtos como este frequentemente ocorrem em ambientes fechados, como navios de cruzeiro (Blanton *et al.*, 2006), creches (Hedlund *et al.*, 2000), asilos, abrigos de emergência (Pang *et al.*, 2009), dentre outros. Também há relatos de transmissões através de água e alimentos contaminados, como por exemplo ostras (Ueki *et al.*, 2010), além da transmissão entre pessoas através da rota fecal-oral.

O sapovírus é estável em pH de 3 a 8 em temperatura ambiente. Possui sensibilidade à etanol 70% em temperatura ambiente por 30 segundos e é inativado por 200mg/l de hipoclorito por 30 minutos e por calor a 56°C por 2 horas (Wang *et al.*, 2012).

Seu genoma é de fita simples RNA senso positivo, com aproximadamente 7.1 a 7.7kb de tamanho, com uma cauda poli(A) na sua extremidade 3'. O genoma contém duas ORFs, sendo que a ORF1 codifica uma poliproteína, com as proteínas não-estruturais NS3, NTPase, NS5/VPg, NS6 protease e NS7 RdRp seguidos pelas proteínas do capsídeo maior, VP1 (Figura 4). A ORF2 codifica a proteína VP2, que possivelmente é o capsídeo menos abundante (*minor capsid*) traduzido pela maquinaria de "*ribosomal termination-reinitiation*". A ORF3 não tem papel bem elucidado.

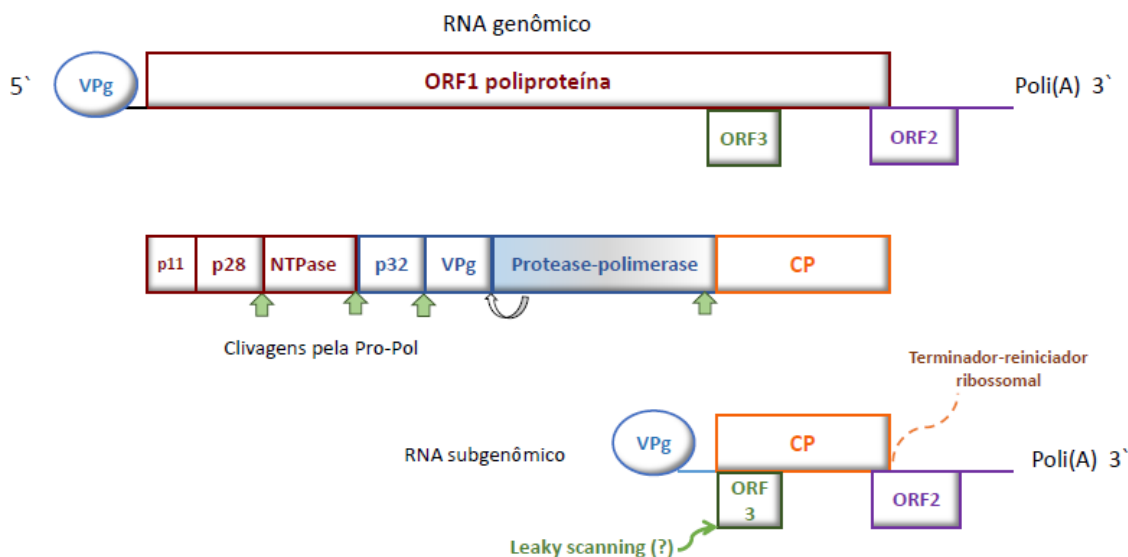


Figura 4: esquema representativo do genoma do sapovírus. VPg= Viral Protein genome-linked, CP= coat protein, NTPase = nucleoside triphosphatase p11 = protein 11, p28 = protein 28, p32= protein 32.

A classificação dos SaV é feita a partir da análise da sequência completa do capsídeo viral (Oka *et al.*, 2012). Cinco genogrupos (GI a GV) são reconhecidos, todos infectam humanos, com exceção do GIII que infecta suínos (Farkas *et al.*, 2004; Oka *et al.*, 2012). Recentemente pelo menos mais três genogrupos (GVI, GVII e GVIII) foram propostos, baseados na sequência completa de seus capsídeos, além de vários outros vírus suínos ainda não classificadas (Scheuer *et al.*, 2013). Um dos fatores preocupantes que requer mais atenção aos SaVs é a possibilidade de haver recombinação intra- e inter-genogrupos e transmissão intra-espécies também (Hansman *et al.*, 2005).

Para a detecção de SaV em amostras clínicas, uma técnica de ELISA foi desenvolvida, porém não é muito utilizada como método diagnóstico pela dificuldade de detecção gerada pela reatividade cruzada (Hansman *et al.*, 2006). O RT-qPCR é mais utilizado devido a sua sensibilidade, especificidade e ampla reatividade. Também pode ser utilizado em conjunto com técnicas como PCR de fluorescência para detecção do SaV juntamente com outros vírus gastroentéricos, mas não se sabe ao certo quantos vírus diferentes essas técnicas podem detectar (Liu *et al.*, 2011). O vírus também pode ser detectado através de técnicas de metagenômica, conforme já descrito em estudos

envolvendo esgoto não tratado, águas residuais, fezes de humanos, cachorros e até mesmo leões marinhos (Bibby *et al.*, 2013; Ng *et al.*, 2012; Li *et al.*, 2011).

No presente estudo, a ferramenta poderosa de NGS foi utilizada para estudar viroma em água de esgoto não tratado e tratado. Foi focado o estudo dos vírus de RNA de fita simples senso positivo, causadores de gastroenterite. Além desta categoria, nos resultados foram detectados também vírus animais, de plantas, protozoários e bacteriófagos.

2. OBJETIVO GERAL

Detecção e identificação de vírus patogênicos circulantes na população é de suma importância no sistema de monitoramento da saúde pública e as águas de esgoto oferecem uma fonte indicadora da saúde populacional. Ainda não existem muitos estudos envolvendo NGS e virologia ambiental relacionados a amostras do Brasil. Por meio desse trabalho pode-se obter informações sobre os vírus que causam gastroenterites, analisando por NGS dos vírus de RNA em água de esgoto não-tratada e tratada, comparando sua diversidade e o efeito do tratamento de água.

2.1 OBJETIVOS ESPECÍFICOS

- Coletar amostras de água de esgoto não-tratada e tratada
- Sequenciar por NGS os RNAs totais presentes nas amostras.
- Analisar a sequência de vírus relevantes à saúde humana, incluindo os causadores de gastroenterite: astrovírus, sapovírus e norovírus.
- Analisar e discutir as diferenças observadas entre as amostras pré e pós tratamento de esgoto.

3. MATERIAIS E MÉTODOS

3.1. AMOSTRAGEM

Amostras de água de esgoto não-tratadas (influyente) e tratadas (efluentes) foram coletadas na Estação de Tratamento de Esgoto da CAESB (ETE-Norte), em Brasília. As amostras foram coletadas em dois pontos diferentes da ETE: o influente primário (esgoto bruto) foi coletado na parte de tratamento preliminar, no início da manhã, enquanto o efluente foi retirado da zona de polimento final, no mesmo dia, porém ao final da tarde, antes da água ser despejada no lago Paranoá, tempo aproximado que a água leva para ser processada. As amostras foram armazenadas em recipientes de vidro com tampa (1L) estéreis e mantidos em gelo para transporte e processamento no laboratório de Virologia da Universidade de Brasília.

Tabela 5 - nomenclatura das amostras.

Mês de coleta	Influente	Efluente
Março de 2016	1NT	1T
Maio de 2016	2NT	2T
Fevereiro de 2017	3NT	3T

3.2. ENRIQUECIMENTO DA FRAÇÃO VIRAL

Primeiramente, as amostras foram colocadas em tubos de polipropileno (8 x 50mL, 400mL total) e precipitadas por centrifugação à baixa velocidade (5000 x g) por 20 minutos, para remoção de bactérias e detritos. O sobrenadante foi submetido à ultracentrifugação (142.000 x g) (Rotor type 45 Ti) (Beckman Coulter, Brea, EUA) com colchão de sacarose 20% por 1 hora.

Para as amostras de efluente, foi necessário realizar as etapas de centrifugação duas vezes (800mL total), pois o precipitado resultante era escasso em comparação com a amostra de influente.

3.3. EXTRAÇÃO DO RNA TOTAL

O precipitado remanescente da ultracentrifugação foi ressuspenso e o RNA foi extraído utilizando o ZR Soil/Fecal RNA MicroPrep kit (Zymo Research, Irvine, EUA), conforme instruções do fabricante. O kit contém filtros que podem eliminar inibidores de RT-PCR como polifenóis e ácido húmico, abundantes em amostras de água de esgoto e fezes.

3.4. NEXT GENERATION SEQUENCING (NGS)

Os RNAs totais purificados foram mantidos congelados a -80°C até serem preparados com *RNAstable kit* (Biomatrix, San Diego, EUA) e enviados à MacroGen Inc. (Seul, Coreia do Sul) para sequenciamento.

O RNA total foi tratado com Ribo-Zero rRNA *removal kit* para bactérias (Illumina, San Diego, EUA) e a construção do cDNA foi realizada utilizando *TruSeq Stranded Total RNA Library Prep Kit* (Illumina) no MacroGen Inc.

Nas amostras de água tratada a quantidade de RNA era muito baixa, portanto foi necessário amplificar o cDNA por *SMARTer Universal low RNA library kit* (Clontech, Mountain View, EUA).

As bibliotecas de cDNA de ambas as amostras foram sequenciadas utilizando Illumina HiSeq 2000, com a condição de *paired-end* de 100 bases.

3.5. ANÁLISE DAS SEQUÊNCIAS

A análise das sequências geradas para as amostras foi realizada no Laboratório de Microscopia Eletrônica e Virologia da Universidade de Brasília, utilizando computadores com processadores CPU Intel(R) Xeon(R) E5-2620 com 2.10Gz, 36GB RAM.

Para montar os genomas virais completos utilizando dados de NGS, as seguintes etapas foram realizadas: (1) avaliação de qualidade de leitura com FastQC, (2) trimagem e limpeza das sequências de *reads*, (3) *de novo assembly* dos *reads* trimados utilizando o programa MEGAHIT (Li et al., 2015) e Velvet (Zerbino e Birney 2008) (4) análise dos *Contigs* gerados com tBLASTX e

Megablast (5) montagem dos genomas usando o software Geneious R8.1 (Biomatters, Auckland, Nova Zelândia) através da extensão dos *contigs*, ou realizar o mapeamento usando genomas de referência para obtenção do genoma viral completo.

A qualidade das sequências foi analisada através do programa FastQC (Bioinformatics Babraham, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e os adaptadores e sequências de baixa qualidade foram removidos com o Trimmomatic (Usadel lab <http://www.usadellab.org/cms/?page=trimmomatic>). O comando, “ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 MINLEN:36” foi utilizado na trimagem das bibliotecas para a remoção dos adaptadores (TruSeq3-PE) e das sequências de baixa qualidade (menor que 20 Phred) dos dados brutos.

Os seguintes valores de k (k-mer) foram utilizados para o *de novo assembly* com MEGAHIT: 21 e 51 (<https://github.com/voutcn/megahit>). O K (k-mer) 21 foi utilizado para o *assembly com o Velvet*.

Os *contigs* (nodes) obtidos com o Megahit foram importados para o programa Geneious e submetidos a análise de tBLASTX (Matrix Blosum 62, E-value e-10) usando o banco de dados de RefSeq Virus publicado em setembro de 2016 pelo GenBank (<http://www.ncbi.nlm.nih.gov/genome/viruses/>). Os hits correspondentes à fagos foram removidos das listas. Em seguida, os nodes remanescentes foram classificados em pastas de acordo com a similaridade viral.

Os *contigs* (nodes) obtidos com o Velvet foram importados para o programa Geneious e submetidos a análise de Megablast, usando o banco de dados de RefSeq Virus publicado em setembro de 2016 pelo GenBank. Os hits correspondentes à fagos foram removidos das listas. Em seguida, os nodes remanescentes foram classificados em pastas de acordo com a similaridade viral.

Foi utilizado o recurso *Map to reference* (utilizando configurações de médium-low/fast sensitivity) do Geneious R8.1 para fazer a extensão de cada *contig*, através do mapeamento dos *reads* trimados ao *contig* de referência. O recurso foi repetido diversas vezes, manualmente, até que o comprimento não pudesse mais ser estendido.

3.5.1 Construção dos gráficos

A construção dos gráficos metagenômicos foi realizada através do programa Kaiju (Menzel *et al.*, 2016), que faz classificações taxonômicas dos *reads*, traduzindo-os em sequências de aminoácidos e comparando-os com um banco de referência microbiano, que encontra a quantidade exata de combinações (*Maximum exact matches -MEMs*). As abundâncias taxonômicas foram ilustradas através dos Krona Charts (Ondov *et al.*, 2011), que permite que os dados hierárquicos sejam dispostos em gráficos de pizza com facetas interativas.

3.6. MONTAGEM DAS ÁRVORES FILOGENÉTICAS

A sequência de aminoácidos das poliproteínas previstas foram montadas (MUSCLE) e as árvores filogenéticas (Maximum Likelihood, JTT matrix-based model, G15, 1000 bootstrap) construídas no MEGA 7.0 (Tamura *et al.*, 2013) utilizando genomas completos das sequências virais de referência.

4. RESULTADOS

4.1 AVALIAÇÃO DE QUALIDADE DE LEITURA COM FASTQC, TRIMAGEM E LIMPEZA DAS SEQUÊNCIAS DE *READS*

Através do programa FastQC foram obtidos os gráficos indicando os valores de qualidade de cada base no formato FastQ. O eixo Y representa o índice de qualidade, sendo que quanto maior melhor a leitura de base. As cores de fundo indicam a qualidade em muito boa (verde), mediana (laranja) e pobre (vermelho). O eixo X indica a posição em *read* (bp). Cada amostra possui dois gráficos, pois o sequenciamento foi realizado em *Paired-end*.

Os gráficos *Per base quality* das amostras Não-tratadas e tratadas de março (1NT e 1T), maio de 2016 (2NT e 2T) e não tratada de fevereiro de 2017 (3NT) apresentaram boa qualidade, conforme as figuras 5, 6, 7, 8 e 9.

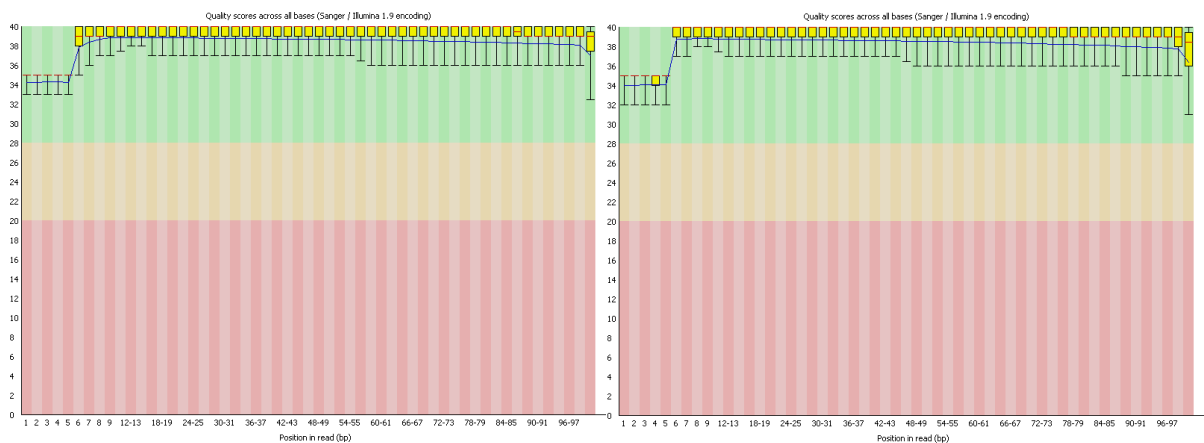


Figura 5 – Gráficos *Per base quality* obtido através do FastQC para a amostra 1NT. As duas figuras representam cada "*pair*" de leitura "*paired-end*".

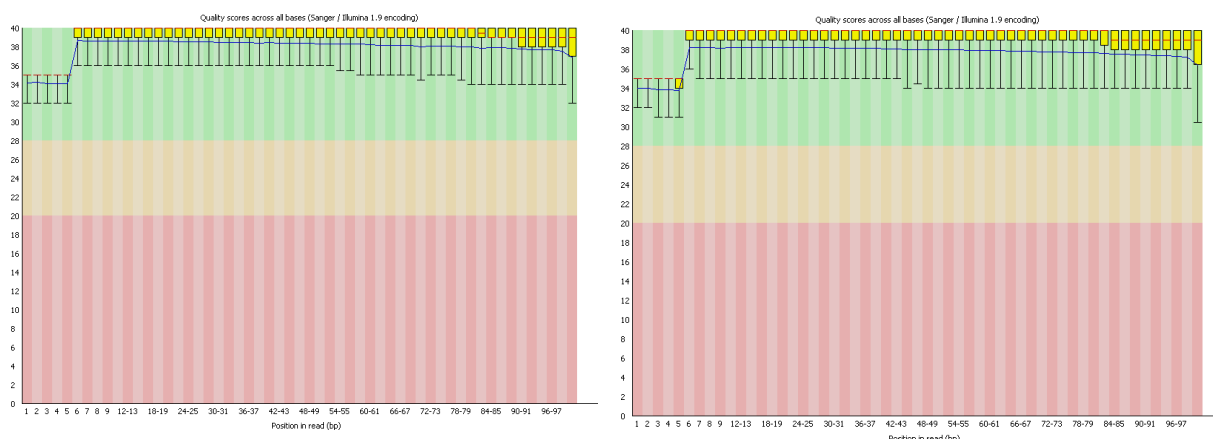


Figura 6 – Gráficos *Per base quality* obtido através do FastQC para a amostra 1T. As duas figuras representam cada "pair" de leitura "paired-end".

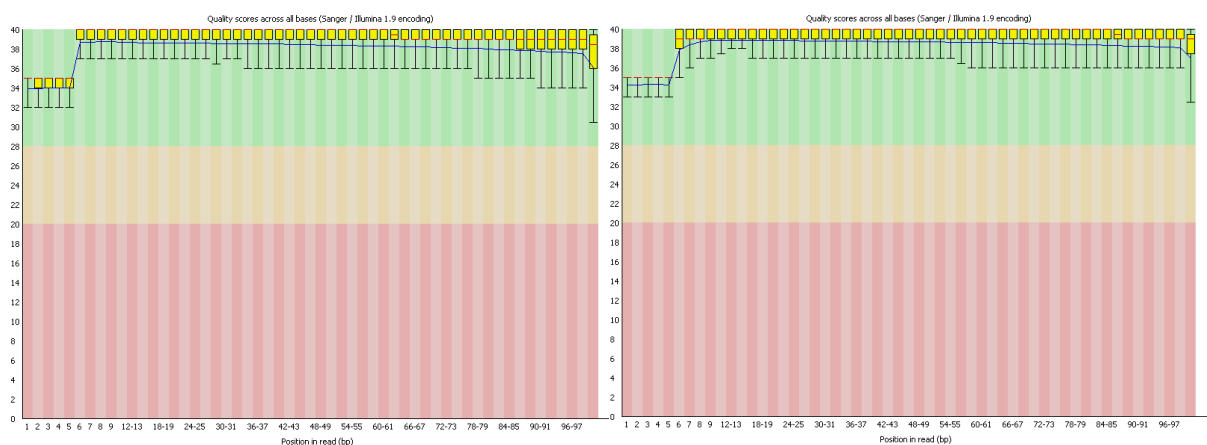


Figura 7 – Gráficos *Per base quality* obtido através do FastQC para a amostra 2NT. As duas figuras representam cada "pair" de leitura "paired-end".

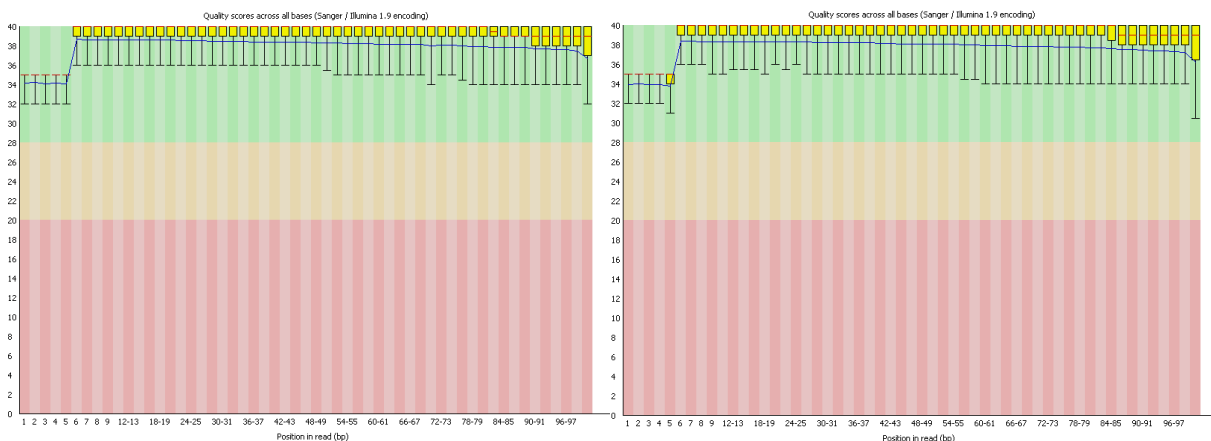


Figura 8 – Gráficos *Per base quality* obtido através do FastQC para a amostra 2T. As duas figuras representam cada "pair" de leitura "paired-end".

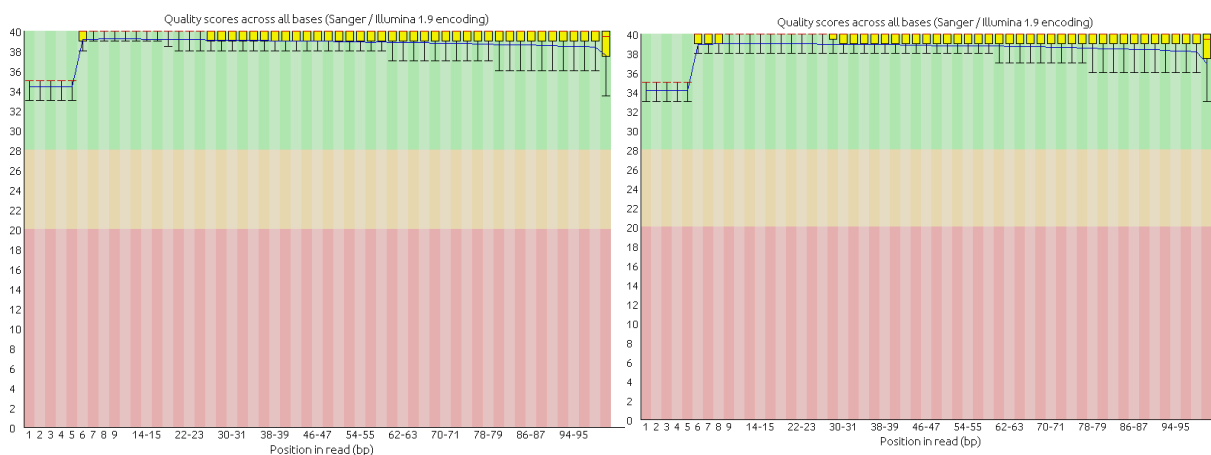


Figura 9 – Gráficos *Per base quality* obtido através do FastQC para a amostra 3NT. As duas figuras representam cada "pair" de leitura "paired-end".

Os gráficos correspondentes à amostra tratada de fevereiro (3T) mostraram estatísticas básicas ruins (com baixa qualidade). O conteúdo de adaptadores falhou com as amostras, mostrando uma contagem percentual cumulativa com uma proporção de biblioteca com sequências de adaptadores em várias posições, conforme ilustrado na figura 10.

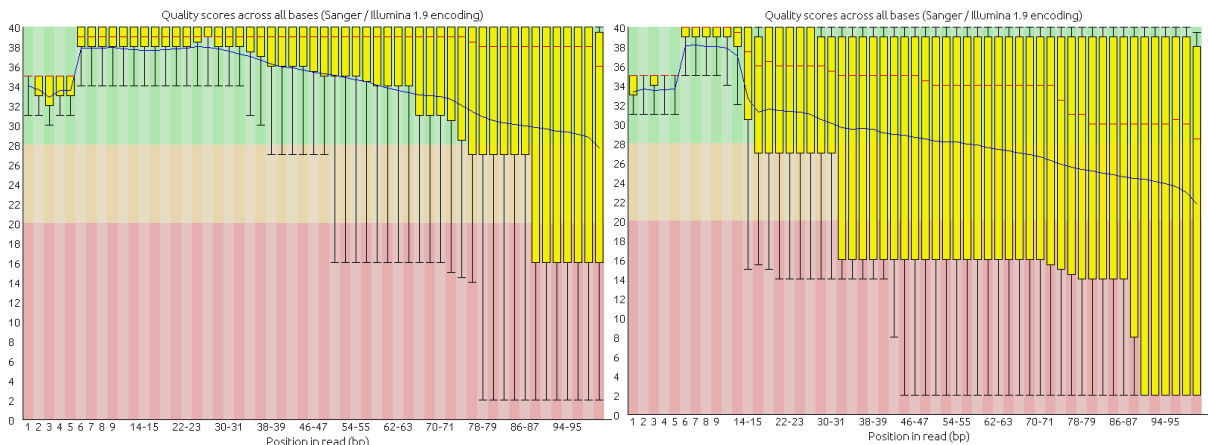


Figura 10 – Gráficos *Per base quality* obtido através do FastQC para a amostra 3T. As duas figuras representam cada "pair" de leitura "paired-end".

4.2 CONSTRUÇÃO DOS GRÁFICOS METAGENÔMICOS ATRAVÉS DA PLATAFORMA KAIJU

Foi feito o *upload* das sequências dos *reads*, em formato FASTA no servidor online do programa Kaiju. Por ser um sequenciamento *Paired-end*, cada amostra teve um par de arquivos a serem enviadas ao servidor. O Kaiju realiza as análises em *Greedy mode* e dentre os parâmetros estão o *Minimum match length:11*, *Minimum match score: 75* e com permissão de *5 mismatches*. As figuras abaixo (figuras 11, 12, 13, 14 e 15) mostram os resultados obtidos para a amostra 1NT (influyente, março de 2016) em formato de *Krona Chart*. Os demais gráficos obtidos através da plataforma Kaiju para as amostras 1T, 2NT, 2T, 3NT e 3T encontram-se nas figuras apresentadas nos **APÊNDICES**.

Nos gráficos da amostra 1NT, de um total de 27747530 de *reads* de organismos encontrados, 19718199 foram classificados, os demais 29% não são classificados (fração azul na figura 11). Os vírus correspondem a 204277 organismos detectados, o que significa aproximadamente 1% do total (figura 12). Deste total viral, 49% correspondem a vírus ssRNA senso positivo, sendo 2% *Caliciviridae* e apenas 1% *Astroviridae* (figura 13). Dos calicivírus, destacam-se o Sapovírus Mc10, o Sapovírus Hu de Kashiwa 2010 e, também, o NoV GI.2 de Jingzhou (figura 14), enquanto os HAstV humanos correspondem a apenas 6% do total de astrovírus (figura 15).

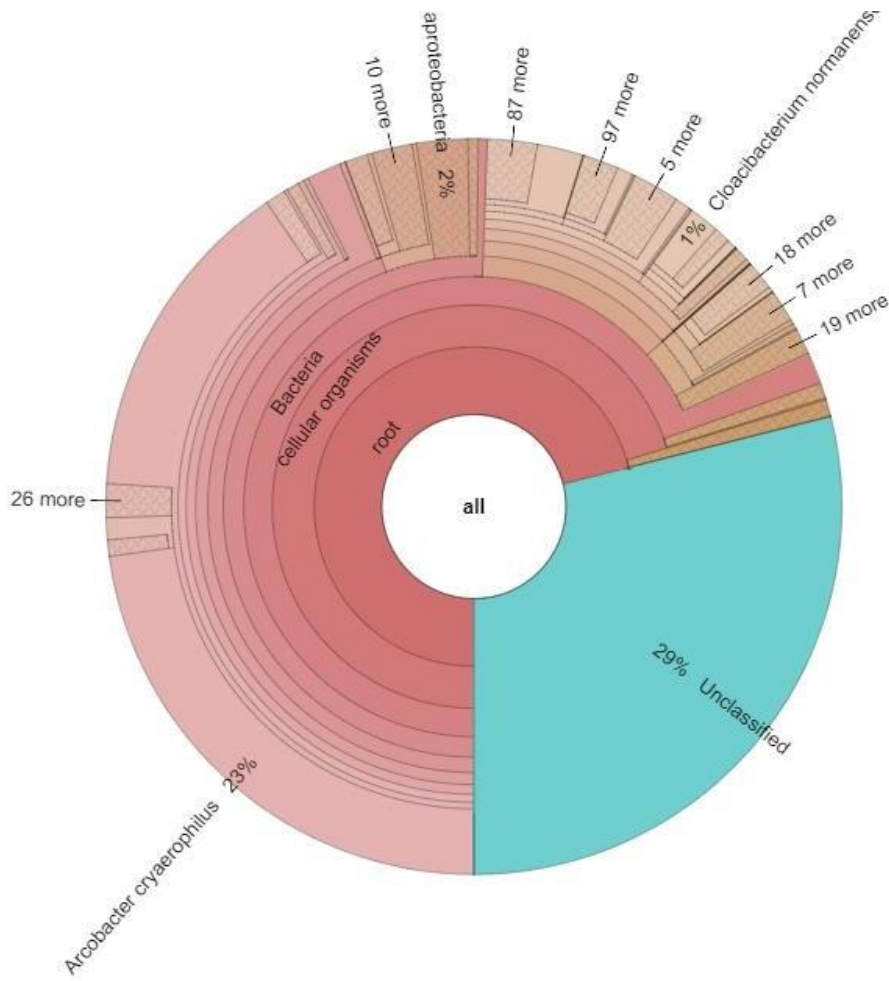


Figura 11 – Krona Chart obtido através da plataforma Kaiju para a amostra 1NT. De um total de 27747530 organismos encontrados, 19718199 foram classificados, os demais 29% não são classificados.

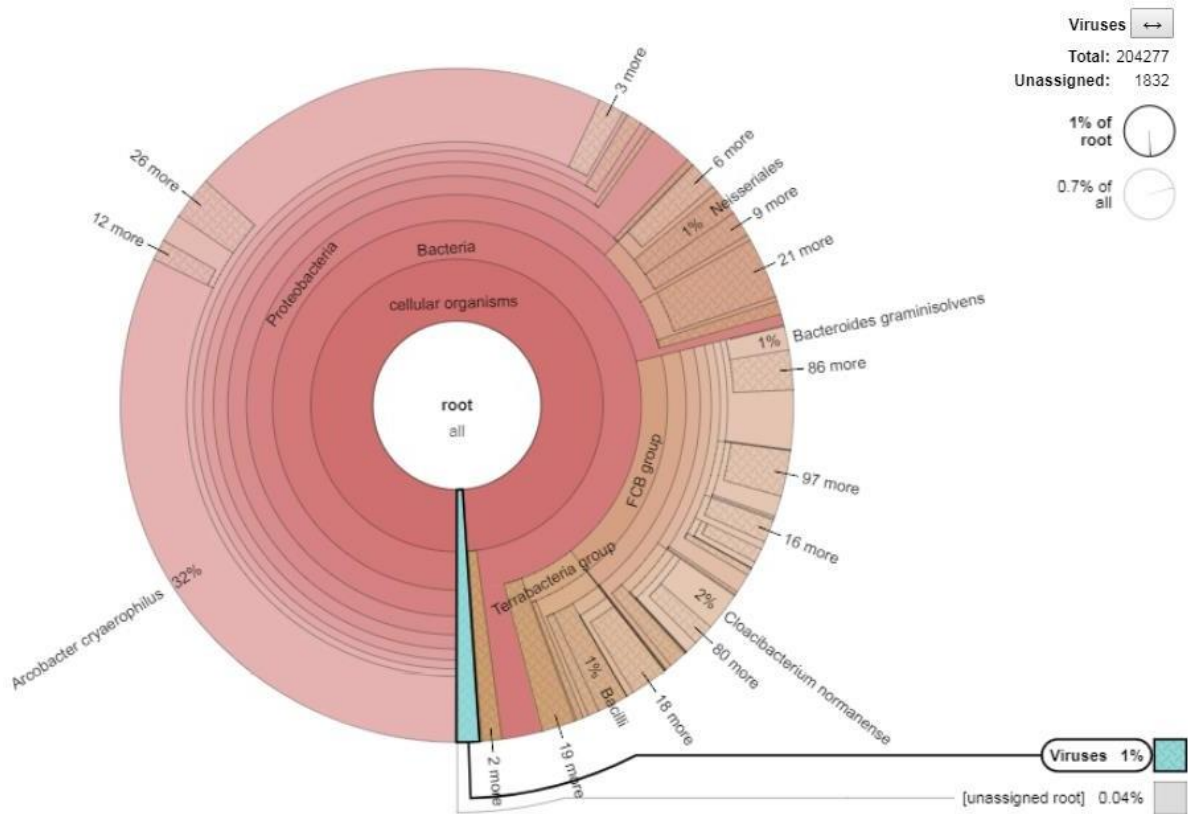


Figura 12 – Krona Chart obtido através da plataforma Kaiju para a amostra 1NT. Os vírus correspondem a 204277 organismos detectados, o que significa aproximadamente 1% do total.

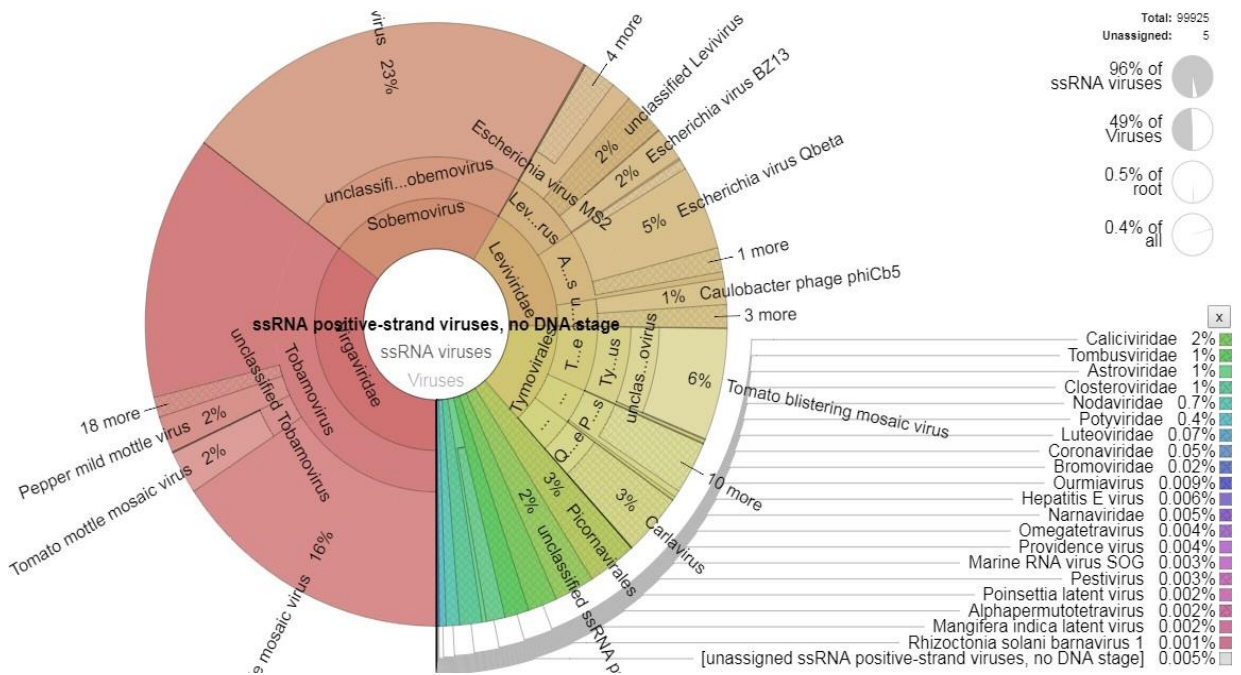


Figura 13 – Krona Chart obtido através da plataforma Kaiju para a amostra 1NT. Do total de vírus, 49% correspondem a vírus ssRNA senso positivo, sendo 2% *Caliciviridae* e apenas 1% *Astroviridae*.

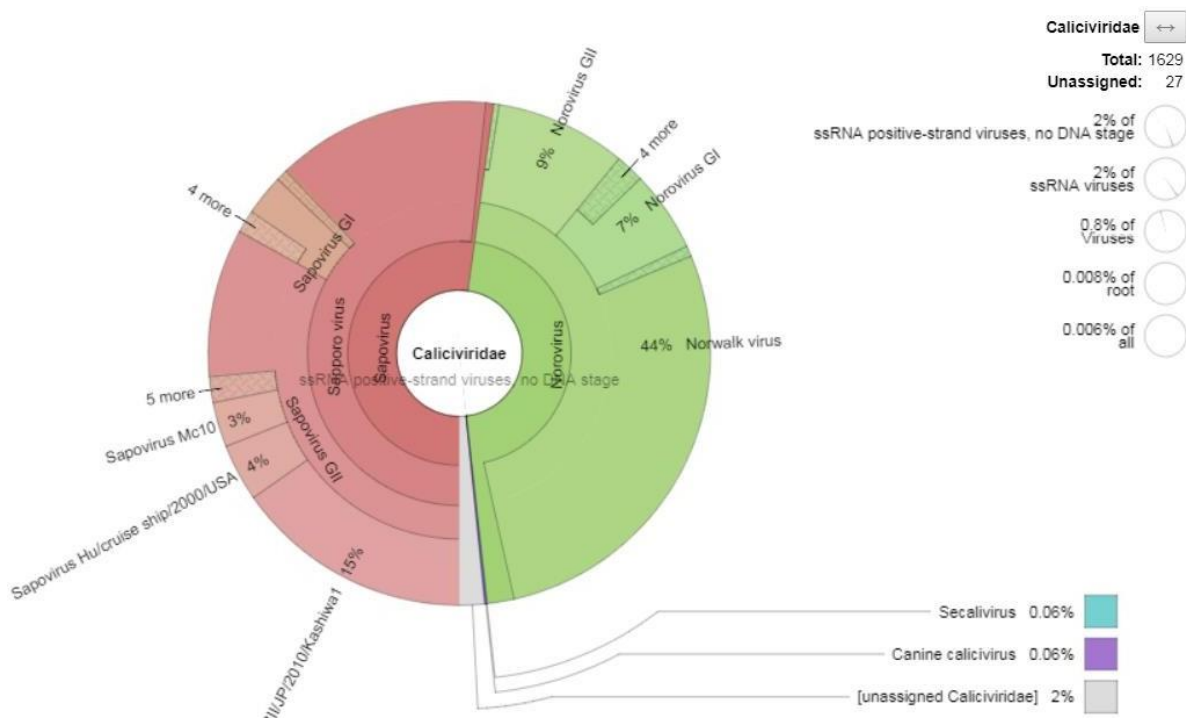


Figura 14 – Krona Chart obtido através da plataforma Kaiju para a amostra 1NT. Dos calcivírus, destacam-se o Sapovirus Mc10, o Sapovirus Hu de Kashiwa. 2010 e, também, o NoV GI.2 de Jingzhou.

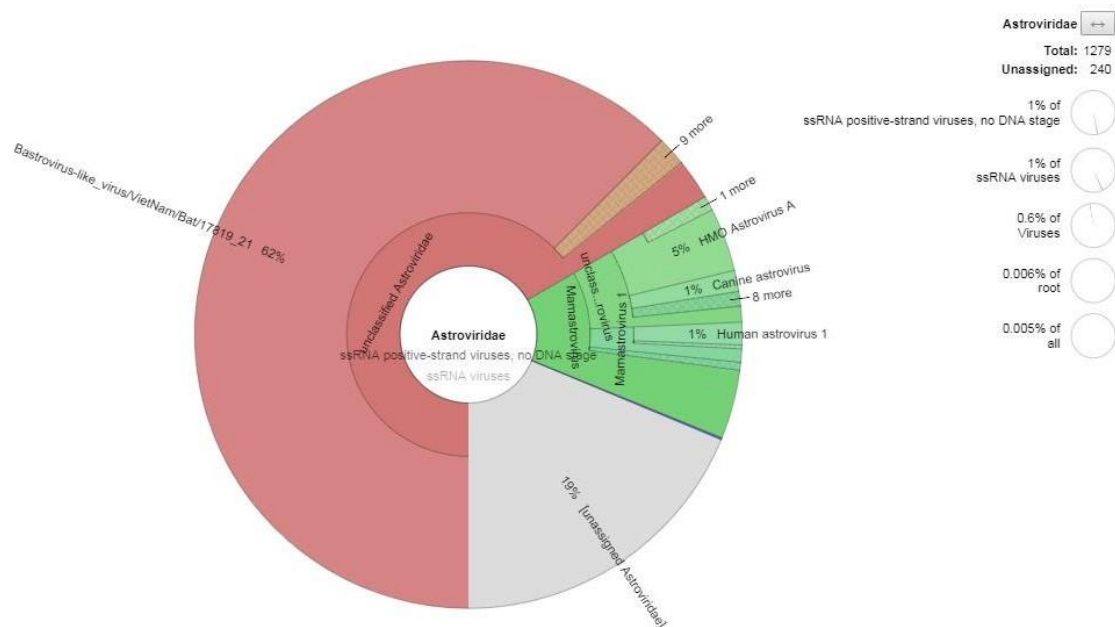


Figura 15 – Krona Chart obtido através da plataforma Kaiju para a amostra 1NT. Os HAstV humanos correspondem a apenas 6% do total de astrovírus.

As tabelas 6, 7, 8, 9, 10 e 11 resumem as estatísticas obtidas na plataforma Kaiju. As tabelas comparam os *reads* totais virais em relação à percentagem total obtida com todos os *reads* classificados. Também compara a quantidade de *reads* específico dos *Caliciviridae* e *Astroviridae*.

Observando a porcentagem de *reads* virais em relação à raiz, é possível notar que, em geral, a proporção viral aumenta após o tratamento de água, demonstrando que o processo é mais eficiente em eliminar outros microrganismos do que vírus. A proporção de *reads* de cada vírus é muito variável, independentemente de ser uma amostra de influente ou efluente e do mês de coleta.

Tabela 6 – Resumo de dados do Kaiju para a amostra 1NT.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	204277	1	0,7	-
ssRNA (+)	104060	0,5	0,4	-
<i>Caliciviridae</i>	1629	0,008	0,006	2
NoV	750	0,004	0,003	0,8
SaV	850	0,004	0,003	0,9
<i>Astroviridae</i>	1279	0,006	0,005	1

Tabela 7 – Resumo de dados do Kaiju para a amostra 1T.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	106835	8	0,9	-
ssRNA (+)	36410	3	0,5	-
<i>Caliciviridae</i>	16	0,001	0,0001	0,04
NoV	6	0,0005	0,00005	0,02
SaV	10	0,0008	0,00008	0,03
<i>Astroviridae</i>	22	0,002	0,00002	0,06

Tabela 8 – Resumo de dados do Kaiju para a amostra 2NT.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	136956	0,7	0,5	-
ssRNA (+)	72259	0,4	0,3	-
<i>Caliciviridae</i>	856	0,005	0,003	1
NoV	399	0,002	0,002	0,6
SaV	415	0,002	0,002	0,6
<i>Astroviridae</i>	17926	0,1	0,07	25

Tabela 9 – Resumo de dados do Kaiju para a amostra 2T.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	102131	6	1	-
ssRNA (+)	18022	1	0,2	-
Caliciviridae	13	0,0007	0,0002	0,07
NoV	-	-	-	-
SaV	9	0,005	0,0001	0,05
Astroviridae	-	-	-	-

Tabela 10 – Resumo de dados do Kaiju par a amostra 3NT.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	59539	0,5	0,4	-
ssRNA (+)	19922	0,2	0,1	-
Caliciviridae	260	0,02	0,02	1
NoV	182	0,02	0,001	0,9
SaV	73	0,0007	0,0005	0,4
Astroviridae	1541	0,01	0,01	8

Tabela 11 – Resumo de dados do Kaiju para a amostra 3T.

Descrição	Reads totais	% da raiz	% do total	% dos ssRNA (+)
Vírus totais	72052	2	1	-
ssRNA (+)	5104	0,1	0,07	-
Caliciviridae	373	0,01	0,005	7
NoV	365	0,01	0,005	7
SaV	8	0,0002	0,0001	0,2
Astroviridae	2	0,00006	0,00003	0,04

4.3 DE NOVO ASSEMBLY DOS READS TRIMADOS UTILIZANDO O PROGRAMA MEGAHIT E ANÁLISE DOS CONTIGS GERADOS COM TBLASTX

As tabelas 12, 13, 14 e 15 abaixo, mostram os resultados do *assembler* Megahit e depois tBLASTX, com os nomes das sequências, além da média de identidade dos *hits* virais e a média de comprimento dos mesmos, para os vírus de GA detectados em amostras de água não-tratada (1NT, 2NT e 3NT) e, também, na amostra 3T. Não foram detectados HAstV, NoV ou SaV nas

amostras tratadas 1T e 2T utilizando este montador. Isso demonstra indiretamente a eficácia de eliminação de vírus por tratamento de água.

O NoV GI foi encontrado somente na amostra 1NT, enquanto o NoV GII em todas as amostras onde algum NoV foi detectado.

Tabela 12 – Vírus de gastroenterite encontrados na amostra 1NT – Megahit.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
NC_011400	Astrovirus MLB1	5	94,7	157,2
NC_013443	HMO Astrovirus A	3	93,4	114
NC_001943	Human astrovirus	4	94,1	106,2
NC_001959	Norovirus GI	4	82,6	359
NC_029646	Norovirus GII	10	79,8	130
NC_006554	Sapovirus C12	4	94	386
NC_006269	Sapovirus Hu/Dresden/pJG-Sap01/DE	4	77,5	173
NC_010624	Sapovirus Mc10	6	94	245,3

*Comprimento médio da sequência em aminoácidos

Tabela 13 – Vírus de gastroenterite encontrados na amostra 2NT – Megahit.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio
NC_011400	Astrovirus MLB1	3	97	107
NC_001943	Human astrovirus	17	86	339
NC_029646	Norovirus GII	1	92	533
NC_029647	Norovirus GII	1	84	510
NC_029648	Norovirus GII	1	69	251
NC_029649	Norovirus GII	1	80	71
NC_029650	Norovirus GII	1	87	118
NC_029651	Norovirus GII	1	62	137
NC_029652	Norovirus GII	1	80	30
NC_006554	Sapovirus C12 strain C12	1	99	303
NC_006269	Sapovirus Hu/Dresden/pJG-Sap01/DE	10	77	170
NC_010624	Sapovirus Mc10	8	95	176

*Comprimento médio da sequência em aminoácidos

Tabela 14 – Vírus de gastroenterite encontrados na amostra 3NT – Megahit.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
NC_011400	Astrovirus MLB1	3	100	89
NC_016155	Astrovirus MLB2	7	94,4	226
NC_001943	Human astrovirus	1	75,3	882
NC_029646	Norovirus GII	5	73,7	68

*Comprimento médio da sequência em aminoácidos

Tabela 15 – Vírus de gastroenterite encontrados na amostra 3T – Megahit.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
NC_029646	Norovirus GII	3	94,1	211

*Comprimento médio da sequência em aminoácidos

4.4 DE NOVO ASSEMBLY DOS READS TRIMADOS UTILIZANDO O PROGRAMA VELVET E ANÁLISE DOS CONTIGS GERADOS COM MEGABLAST

Os dados recebidos pelo sequenciamento também foram montados através da *de novo contig assembler* Velvet, com valor de k igual a 21. Em seguida, os *contigs* foram transferidos para o Geneious e foi feito um Megablast das sequências obtidas. As tabelas 16, 17, 18 e 19 abaixo, mostram os resultados do Megablast, com os nomes das sequências de acordo com o GenBank, além da média de identidade dos *hits* virais e a média de comprimento dos mesmos, para as amostras de influente e efluente de março e maio de 2016.

Tabela 16 – Vírus de gastroenterite encontrados na amostra 1NT utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Média do comprimento*
JX857868	Astrovirus VA3 isolate VA3/human/Vellore/28054/2005	1	99,3	152
GQ415660	HMO Astrovirus A isolate NI-295	10	95,1	190,6
KF211475	Human astrovirus 1 isolate JZ	1	87,7	57
KF859964	Human astrovirus BF34	1	92,4	132
LC064152	Human astrovirus MLB1 genomic RNA: NAGANO1545	4	97,4	116,7
L23513	Human astrovirus type 1 strain Oxford	2	97,3	91,5
KM358468	Human astrovirus UK1	2	99,2	121
FJ222451	Astrovirus MLB1	1	97,8	131
AB823731	Astrovirus MLB1, isolate: BtnMLB1-40	1	98,3	357
AB823732	Astrovirus MLB1, isolate: BtnMLB1-86	15	98,2	174
HM450380	Astrovirus MLB1 HK05	2	98,4	284,5
JQ086552	Astrovirus MLB1 isolate SY071	2	95,2	125
KT224358	Astrovirus MLB2 isolate MLB2/human/Geneva/2014	4	98,8	148
AB829252	Astrovirus MLB2, isolate: GUP187	1	98,5	134
GQ502193	Astrovirus VA2 isolate VA2/human/Stl/WD0680/2009	10	93,1	201
KJ196292	Norovirus GI/Hu/JP/2007/GI.P3_GI.3/Shimizu/KK2866	4	92,7	137
LN854563	Norovirus GI/Hu/NL/2011/GI.4/Groningen	1	98,6	146
LN854570	Norovirus GII/Hu/NL/2014/GII.2/Groningen	5	96,4	166,2
KF944111	Norovirus GII.3/Novosibirsk/Nsk-N1648/2011/RUS	1	99,4	156
KT202794	Norovirus GII Hu/Guangzhou/GZ2014-L106/CHN/2014	3	98,7	186
KF306212	Norovirus Hu/GI.2/Jingzhou/2013401/CHN	19	99,1	386,2
KU757050	Norovirus Hu/GII/CN/2014/GII.17/1401184	1	99,4	522
KU757051	Norovirus Hu/GII/CN/2014/GII.17/FX141210-5	1	99,6	273
KU757052	Norovirus Hu/GII/CN/2014/GII.17/FX141226-31	3	99,7	115
LC175468	Norovirus JP/2016/_GII.4_Sydney2012/Kawasaki194	15	98,8	178
AB972504	Norovirus Hu/GII.4/GII4-NG1343/2011/JP	2	99	256,5
AB972505	Norovirus Hu/GII.4/GII4-NG1365/2011/JP	4	99,4	188,5
KC577175	Norovirus Hu/GII.4/Jiangsu2/2012/CHN	2	99,7	126,5
AB972484	Norovirus Hu/GII.4/OsakaCity1/2012/JP	1	98,6	139
KC464499	Norovirus Hu/GII.12/CGMH41/2010/TW	2	95,7	148,5
AY603425	Sapovirus C12	1	95,2	126
KP067444	Sapovirus GII strain PHL-TGO12-028	3	91,5	220
JX993277	Sapovirus Hu/G1/BE-HPI01/DE/2012	10	97,2	223,3
KP298674	Sapovirus Hu/GI.1/Seoul/ROK62/2013/KOR	7	96,4	135
AY237419	Sapovirus Mc2	3	93,6	167
AY646856	Sapovirus NongKhai-24/Thailand	2	95,6	112
AY646855	Sapovirus SaKaeo-15/Thailand	1	88,4	181

*Comprimento médio da sequência em aminoácidos

Tabela 17 – Vírus de gastroenterite encontrados na amostra 1T utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
LC064152	Human astrovirus MLB1: NAGANO1545	1	99,1	115
KF306212	Norovirus Hu/GI.2/Jingzhou/2013401/CHN	2	99	144
AB614356	Sapovirus Hu/GI.2/BR-DF01/BRA/2009	3	99,6	153,6

*Comprimento médio da sequência em aminoácidos

Tabela 18 – Vírus de gastroenterite encontrados na amostra 2NT utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio *
AB823731	Astrovirus MLB1, isolate: BtnMLB1-40	1	98,5	130
AB823732	Astrovirus MLB1, BtnMLB1-86	5	98	192,6
HM450380	Astrovirus MLB1 HK05	6	98,6	209,3
JQ086552	Astrovirus MLB1 isolate SY071	2	96,2	169,5
KT224358	Astrovirus MLB2/human/Geneva/2014	4	99,1	120,6
JF742759	Astrovirus MLB2/human/Stl/WD0559/2008	2	98,5	140,5
AB829252	Astrovirus MLB2, isolate: GUP187	3	98,3	117,3
FJ973620	Astrovirus VA1	1	98,4	128
GQ502193	Astrovirus VA2 isolate VA2/human/Stl/WD0680/2009	9	97,2	177,4
JX857868	Astrovirus VA3 isolate VA3/human/Vellore/28054/2005	1	95,9	122
GQ415660	HMO Astrovirus A isolate NI-295	3	96,2	157,6
HQ398856	Astrovirus 1 strain Hu/Nyergesujfalu/HUN4520/2010/HUN	9	98,1	113,3
KF039910	Astrovirus 2 isolate Rus-Nsc05-430,	1	93,6	202
KF039911	Human astrovirus 2 isolate Rus-Nsc06-1029	18	96,6	259,6
KF039912	Human astrovirus 4 isolate Rus-Nsc05-623	1	99,1	113
JQ403108	Human astrovirus 5 isolate DL030	2	95,4	166,5
LC064152	Human astrovirus MLB1, strain: NAGANO1545	2	98	101,5
JF327666	Human astrovirus strain Pune/063681/India	1	96,6	146
KM358468	Human astrovirus UK1	4	98,4	135,2
KJ196292	Norovirus GI/Hu/JP/2007/GI.P3_GI.3/Shimizu/KK2866	3	93,1	166,3
KJ196295	Norovirus GII/Hu/JP/2010/GII.P7_GII.7/Musashimurayama/TAKAsanKimchi	6	94,9	193,6
LN854570	Norovirus GII/Hu/NL/2014/GII.2/Groningen	1	98,1	108
LN854567	Norovirus GII/Hu/NL/2014/GII.4/Groningen02	2	99,6	138,5
KT970376	Norovirus GII isolate Hu/Guangzhou/GZ2015-L343/CHN/2015	1	100	117
KF429783	Norovirus Hu/GI.1/8K/1979/USA	1	86,6	97
KF306212	Norovirus Hu/GI.2/Jingzhou/2013401/CHN	3	99,5	161,6
LC175468	Norovirus JP/2016/GII.P16_GII.4_Sydney2012/Kawasaki194	16	98,4	324,1
AB972499	Norovirus GII4-HK01/2011/JP RNA	1	99,1	112
AB972504	Norovirus Hu/GII.4/GII4-NG1343/2011/JP	2	99,5	154
AB972505	Norovirus Hu/GII.4/GII4-NG1365/2011/JP	4	97,1	166,5
AB972473	Norovirus Hu/GII.4/Iwate5/2012/JP	2	97,8	210,5
AB972484	Norovirus Hu/GII.4/OsakaCity1/2012/JP	1	98,3	121
HQ449728	Norovirus Hu/GII.12/HS210/2010/USA	1	94,4	179
GU594162	Norovirus Hu/GII.14/8610/Saga/2008/JPN	1	96,2	105
KP067444	Sapovirus GII strain PHL-TGO12-028	1	82,2	163
JX993277	Sapovirus Hu/G1/BE-HPI01/DE/2012	7	97,2	227
KP298674	Sapovirus Hu/GI.1/Seoul/ROK62/2013/KOR	16	96,5	217,6
AY237423	Sapovirus N21	1	96,5	115

*Comprimento médio da sequência em aminoácidos

Tabela 19 – Vírus de gastroenterite encontrados na amostra 2T utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
AB614356	Sapovirus Hu/GI.2/BR-DF01/BRA/2009	3	99	166,3
-	Norovirus	-	-	-
-	Astrovirus	-	-	-

*Comprimento médio da sequência em aminoácidos

Tabela 20 – Vírus de gastroenterite encontrados na amostra 3NT utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
NC_001943	Human astrovirus, complete genome	9	90,9	161
NC_011400	Astrovirus MLB1, complete genome	16	95,8	164
NC_013060	Astrovirus VA1, complete genome	2	96,3	96
NC_013443	HMO Astrovirus A, complete genome	2	95,3	161
NC_016155	Astrovirus MLB2, complete genome	33	98	192
NC_019026	Astrovirus VA3/human/Vellore/28054/2005	1	97,6	126
NC_001959	Norovirus GI, complete genome	2	91,5	123
NC_029646	Norovirus GII, complete genome	2	83,9	186
NC_006269	Sapovirus Hu/Dresden/pJG-Sap01/DE	1	91,4	174

Tabela 21 – Vírus de gastroenterite encontrados na amostra 3T utilizando o Velvet.

Nome	Descrição	Hits	Média da identidade (%)	Comprimento médio*
NC_029646	Norovirus GII, complete genome	7	92	251
NC_006269	Sapovirus Hu/Dresden/pJG-Sap01/DE	1	87,3	63
-	Astrovirus	-	-	-

Nas tabelas obtidas pelo Velvet (*de novo assembly*) e Megablast (tipo de Blast), é possível notar que a média de identidade de sequência viral em geral é mais alta do que da análise realizada com o Megahit e tBLASTX. Além disso, a quantidade de *hits* obtidas em cada amostra pelo Velvet é maior, deixando a lista com maior diversidade viral.

NoV, HAstV e SaV não foram detectados com o Megahit para as amostras de efluente 1T e 2T, porém com o Velvet foi encontrado o SaV HU/G1.2/BR-DF01/BRA/2009, NoV Jingzhou e o Astrovirus MLB1, para a amostra 1T (tabela 17). O mesmo SaV anteriormente citado também foi detectado na amostra 2T (tabela 19),

4.5 MONTAGEM DOS GENOMAS E ÁRVORE FILOGENÉTICA ATRAVÉS DO PROGRAMA GENEIOUS

Foi montado uma sequência de Norovirus GI.2, resultante na cepa denominada Norovirus_Distrito Federal (Figura 16), descrita e sublinhada como CP DISTRITO FEDERAL na árvore da figura 17.

Através da comparação do NCBI, o vírus mais próximo, com identidade de 98%, foi a sequência de Norovírus GI.2 de Jingzhou, China (número de acesso KF306212) (Figura 17).

Os vírus do mesmo *cluster* do NoV GI.2 possuem isolados de diversos países, como Bélgica, Inglaterra, Brasil e China, um indicativo da rápida disseminação do genótipo GI em um curto período.

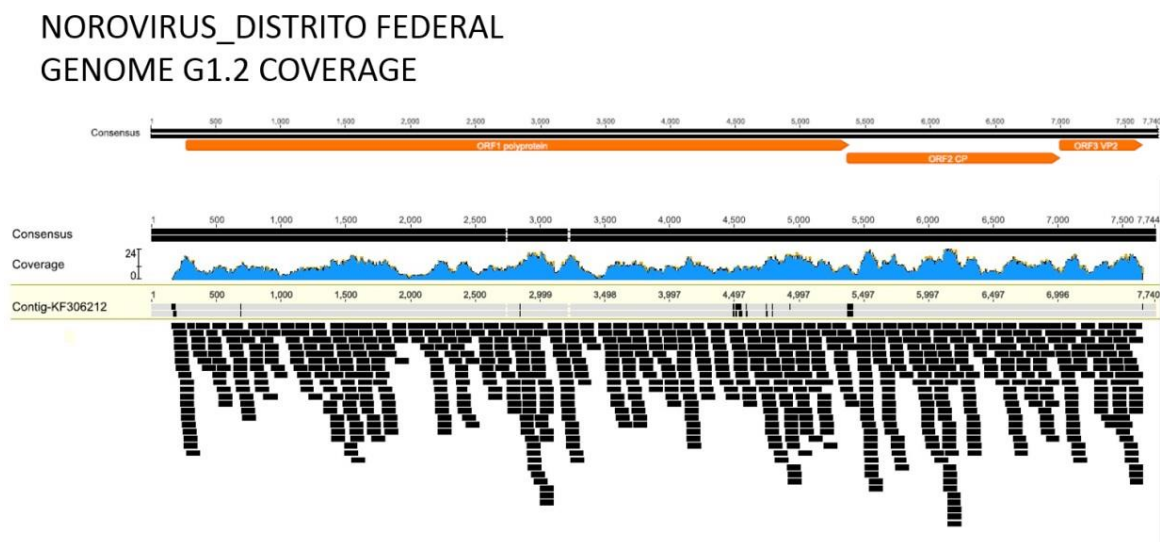


Figura 16 – Cobertura do genoma Norovírus_Distrito Federal, obtido através do software Geneious.

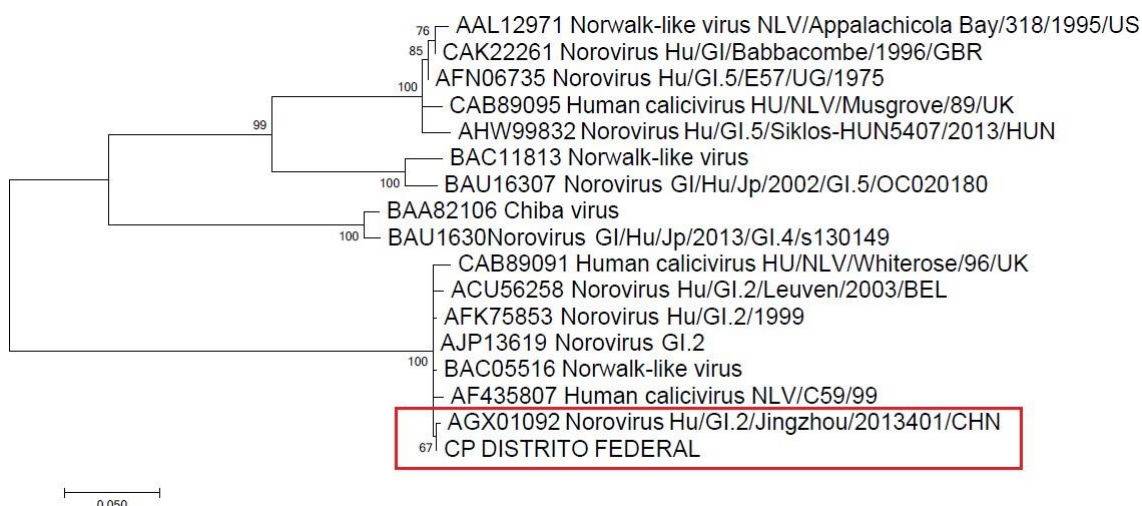


Figura 17 – Árvore filogenética do Norovírus detectado na amostra 1NT. A árvore foi construída ao comparar as diferentes cepas no MEGA6. O vírus detectado nesse estudo está realçado no quadro vermelho. *Bootstrap values* (baseados em 1000 replicatas) acima de 70% estão mostradas no *branch*, realçado no quadro vermelho.

Um genoma de SaV foi montado, com base no genoma do SaV Mc10, encontrado na amostra 1NT, conhecido por seus isolados detectados em surtos na Tailândia e no Japão (acesso: AY237420).

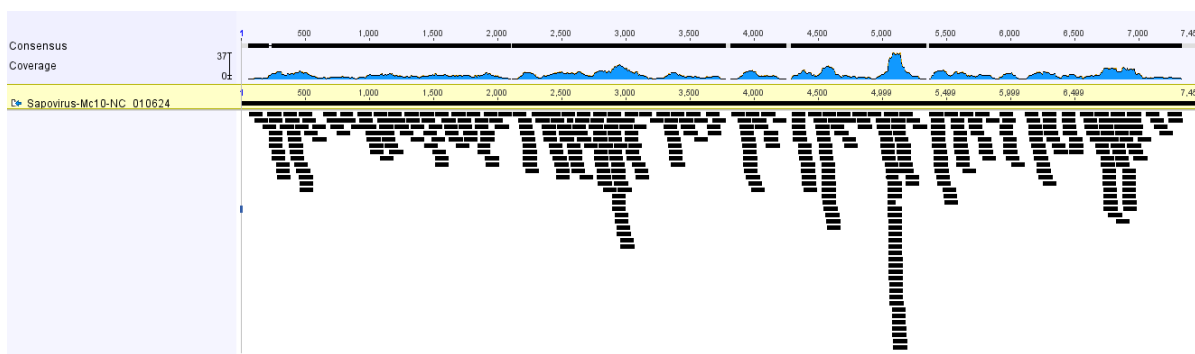


Figura 18 – Cobertura do genoma Sapovírus MC10, obtido através do software Geneious.

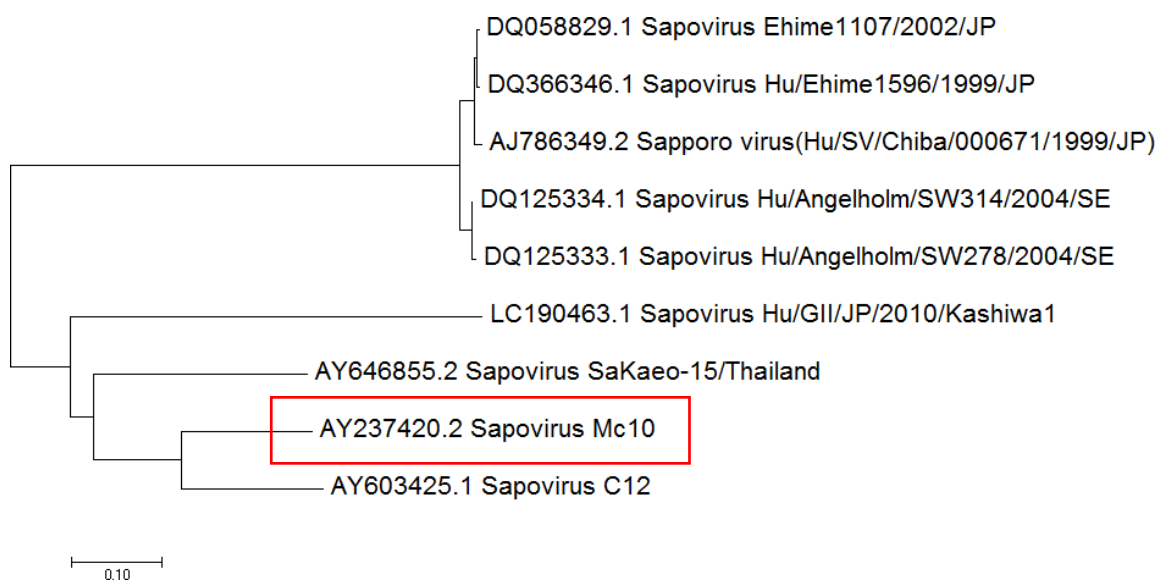


Figura 19 – Árvore filogenética do Sapovírus detectado na amostra 1NT. A árvore foi construída ao comparar as diferentes cepas no MEGA6.

5. DISCUSSÃO

A trimagem é um processo importante para reconstrução das sequências genômicas. Neste momento se retiram as sequências com baixa qualidade, para que não interfiram no *de novo assembly*, mas tomando cuidados para não perder informações, sempre avaliando a qualidade da sequência antes cada comando de trimagem, através do FastQC. O Trimmomatic foi escolhido pois ele pode ser utilizado nos principais sistemas operacionais como Linux, Windows e Mac. Além disso, possui a capacidade de remover adaptadores e bases N de baixa qualidade, oferecendo ferramentas interessantes de filtragem para realizar trimagens de qualidade, como as filtragens utilizadas nesse estudo, por exemplo o *Sliding window*, que corta sequências que ficam abaixo de um *threshold* determinado e o *Minlen*, que remove o *read* se ele é abaixo do desejado. Os gráficos *per base quality* do FastQC das amostras do presente estudo mostraram o mesmo perfil, com exceção da amostra 3NT, que mostrou estatísticas baixas. Uma das alternativas seria repetir o sequenciamento, ou tentar remover o máximo possível de adaptadores e sequências de baixa qualidade, como realizado, porém pode ser que algum material tenha sido perdido no momento da trimagem.

Para o *de novo assembly* dos *contigs*, foram utilizados algoritmos baseados nos gráficos *de Bruijn*, estes são considerados mais eficientes na montagem dos genomas. Alguns dos principais montadores que utilizam os gráficos *de Bruijn* são o Velvet, MEGAHIT, SPAdes (Bankevich *et al.*, 2012), IDBA (Peng *et al.*, 2012). Um dos montadores escolhido neste estudo foi o MEGAHIT, pois é rápido e requer pouca memória para a montagem. Ele é adequado para montagens de dados utilizando múltiplos valores de *k*. Em estudos feitos pelo Laboratório de Virologia da Universidade de Brasília, onde o estudo descrito nessa dissertação foi realizado, demonstrou-se que, dentre os montadores citados, o MEGAHIT gera *contigs* individuais mais longos para cada sequência pertencente à determinada família, além de trabalhar de forma semelhante como o SPAdes. Outro montador escolhido foi o Velvet, pois é desenvolvido para *reads* curtos e gera grande quantidade de *contigs* (Blawid *et al.*, 2017).

Os estudos de metagenoma para amostras ambientais tem se mostrado eficientes para detectar não somente vírus humanos, mas também vírus ambientais e bacteriófagos. Nos sequenciamentos aqui apresentados, a grande maioria de *hits* virais foi de fagos, condizendo com trabalhos como o de Cantalupo e colaboradores (2011), que demonstraram que dos *hits* virais, mais de 80% era relacionado a bacteriófagos. Já a grande maioria de *hits* virais encontrados nas amostras provém de vírus de plantas, conforme já demonstrado por outros autores (Zhang *et al.*, 2006; Rosario *et al.*, 2009). Patógenos de planta de importância comercial foram encontrados nas amostras deste estudo, como por exemplo *Melon yellowing-associated virus*, *Garlic common latent virus*, *Pepper mild mottle virus*, dentre outros (dados não mostrados). Tais vírus também foram detectados a partir de amostra fecais e descritos em trabalhos anteriores (amostras de esgoto não-tratado) realizados pelo Laboratório de Virologia Geral (UnB) (Anjos, 2017).

A quantidade de publicações a respeito do HAstV no Brasil, até o presente momento, é escassa, quando comparada com os calicivírus, apesar do HAstV ser associado a surtos de GA envolvendo crianças pequenas e pessoas de outras faixas etárias em diferentes partes do mundo (Matsui e Greenberg, 2001). No sequenciamento realizado no presente estudo, foram obtidos diversos *hits* virais para HAstV como o MLB1 [isolado BtnMLB1-86 (GenBank: AB823732)] e o HMO HAstV A, conforme relatado em análises de amostras fecais (Kapoor *et al.*, 2009; Finkbeiner *et al.*, 2008). Santos e colaboradores (2007) coletaram 1244 amostras fecais de crianças hospitalizadas no Centro Oeste do Brasil, incluindo as cidades de Brasília e Goiânia. Do total, 46 (3,7%) dos espécimes foram positivas para HAstV, sendo 25 espécimes de Brasília e 21 de Goiânia.

Os NoV foram encontrados em todas as amostras de esgoto bruto, dado condizente com o trabalho de Haramoto e colaboradores (2006), que também compararam a quantidade de NoV de influente e efluente numa ETE de Tóquio, no Japão. O número de *hits* de NoV GI é abaixo do genogrupo GII, corroborando com outros estudos brasileiros, que também acharam baixa prevalência (Ferreira *et al.*, 2010; Fioretti *et al.*, 2011; Andrade *et al.*, 2014). Andrade e colaboradores (2014) descreveram em estudos de amostra de surtos de GA no Rio Grande do Sul que o genótipo GII.4 correspondia a 72% dos NoV detectados enquanto apenas duas amostras correspondiam à NoV GI.1 e GI.3. No presente estudo,

um dos genótipos de NoV GI detectados também corresponde ao NoV GI.3, variante de Shimizu, Japão 2007 (GenBank KJ196292), enquanto o outro genótipo correspondeu ao NoV GI.4, caracterizado em 2011, Groeningen na Holanda (GenBank LN854563), além de grande quantidade do NoV GI.2 de Jingzhou, China (KF306212), quando analisados por identidade de aminoácidos do referente ao gene do capsídeo viral.

Em todas as amostras foram obtidos *hits* virais para o NoV mais epidêmico, o GII.4, também condizendo com artigos publicados no Brasil (Barreira *et al.*, 2010; Ferreira *et al.*, 2010; Ferreira *et al.*, 2012; Andrade *et al.*, 2014). Assim como no estudo de Haramoto e colaboradores (2006), os NoV GII sempre foram mais abundantes no esgoto bruto do que o NoV GI. Os NoV GII.4 foram relatados na maioria dos surtos mundiais desde a década de 1990 (Siebenga *et al.*, 2009; Zheng *et al.*, 2010; Marshal e Bruggunk, 2011). Entre 2004 e 2012 dentre as pandemias de GA relatadas, algumas variantes de GII.4 foram Hunter_2004, Den Haag_2006b, New Orleans_2009 e Sidney_2012 (Kroneman *et al.*, 2013). A variante Sidney_2012 de GII.4 (GenBank LC175468) foi o NoV que mais obteve *hits* nas amostras 1NT e 2NT.

Eftim e colaboradores (2017) buscaram entender as variações nas distribuições de NoV de acordo com a região geográfica e estação do ano. Após extensa revisão bibliográfica, foi determinado que as maiores ocorrências de NoV eram associadas com a primavera e inverno, porém, o estudo não incluiu dados da América do Sul. Em nosso estudo, as amostras sequenciadas foram coletadas nos meses de fevereiro e março, durante o período de chuva e maio, período de seca. No Brasil, foram relatados aumento na incidência durante a estação de seca, porém o padrão de sazonalidade ainda é controverso (Soares *et al.*, 2007; Andreasi *et al.*, 2008; Ferreira *et al.*, 2010).

No Brasil há poucos estudos clínicos que avaliam a circulação de SaV, dentre eles, há um relato de casos de SaV GI.1 e GI.3 em 4,6% de crianças assintomáticas em uma creche no Centro Oeste (Oliveira *et al.*, 2014). Aragão e colaboradores (2010) coletaram 305 espécimes fecais de pacientes com GA em Belém/PA: do total 15 (4,9%) foram positivas para SaV, do tipo GI.1, GI.2 e GII.1. Nas amostras coletadas, também foi detectado o SaV GI. A variante SaV GI/Be-HPI01 foi registrada em um surto de GA na Alemanha em 2012 (GenBank JX993277). O SaV GI.1 Seoul/ROK62 foi caracterizado por Choi e colaboradores

(2015), na Coréia do Sul (GenBank: KP298674). Os *hits* para SaV GII correspondem ao SaV GII variante PHL-TGO12-028 (Liu *et al.*, 2015).

No presente estudo, foram detectados sapovírus nas amostras de efluente final (2T e 3T, Tabelas 19 e 21), com *hits* virais para o sapovírus HU/G1.2/BR-DF01/BRA/2009 (GenBank: AB614356), variante descrita pelo Laboratório de Virologia Geral (UnB) (Anjos *et al.*, 2011). Porém, tais achados contradizem com resultados da pesquisa de Fioretti e colaboradores (2016), que analisando amostras provenientes da ETE do Rio de Janeiro, não encontraram SaV no efluente final.

Nos resultados obtidos, é notável a diferença de *hits* virais entre as amostras de influente e efluente: as amostras tratadas possuem menos *hits* virais detectados, condizendo com outros artigos, onde foi observado que a concentração de NoV humano do esgoto bruto (influyente) diminui quando comparado ao efluente (Lodder e Husman, 2005; van den Berg *et al.*, 2005; Haramoto *et al.*, 2006). Van der Berg e colaboradores (2005) relataram que em treze de quatorze amostras de efluente testadas foram detectados a presença de NoV, os autores ressaltaram ainda que a ingestão e uso da água contaminadas pelo efluente gera potenciais riscos à saúde. Com a crise hídrica no Distrito Federal, desde outubro de 2017 iniciou-se a captação de água do Lago Paranoá para o abastecimento de regiões administrativas como a Asa Norte, Lago Norte, Itapoã, dentre outras. Há relatos de que a água captada tem gerado diversos casos de GA nos habitantes do DF (Ferraz, 2017). Apesar de ainda não haver estudos comprovando a relação entre esses fatores, em Brasília, sabe-se que o efluente despejado no Lago Paranoá possui sequencias virais de causadores de GA, como mostrado e discutido nesta dissertação.

6. CONCLUSÃO

A tecnologia de NGS confirmou ser uma ferramenta interessante quando deseja-se avaliar um *pool* de amostras tão ricas em informações quanto as coletadas na estação de tratamento de esgoto de Brasília. Através da coleta, extração, sequenciamento e processamento das análises com ferramentas da bioinformática, foi possível adquirir uma infinidade de informações, não só sobre os vírus de gastroenterite selecionados, como também uma diversidade de vírus animais, de plantas e bacteriófagos.

Neste trabalho focado nos principais vírus de GA, foi observado a presença dos vírus tanto em amostras de esgoto bruto, quanto em águas tratadas e prontas para o despejo no Lago Paranoá. O processo de desinfecção do esgoto parece ser bastante efetivo, pois a quantidade de *hits* virais encontrados em amostras de efluente foi bem menor quando comparado às de influente.

Com a crise hídrica no Distrito Federal, a água proveniente do Lago Paranoá tem sido utilizada no abastecimento de várias regiões administrativas, devido a isso é importante manter um monitoramento constante do viroma presente na água processada através de uma ETE, para assim avaliar a saúde populacional, através de amostras não-tratadas, e evitar possíveis contaminações que venham a gerar surtos de GA, com a observação das amostras tratadas.

7. PERSPECTIVAS

O NGS gera muitos dados sobre o pool genômico estudado, porém não se pode ter certeza da presença viral apenas com o número de *hits*, por isso, é importante analisar as amostras de maneira quantitativa através do RT-qPCR.

Espera-se que seja realizada a análise das amostras por RT-qPCR, sendo que o sistema de detecção e quantificação a ser utilizado deve ser o método por sonda, onde se utiliza uma sonda que será hibridizada à uma região específica do *amplicon*. O sistema de sonda por hidrólise é mais utilizado em análises clínicas devido à maior precisão e especificidade, essenciais para a área da saúde. Os vírus escolhidos para a realização da RT-qPCR são os principais vírus de GA descritos neste estudo, detectados nas amostras de água tratada e não tratada: astrovírus humano, norovírus GI, norovírus GII.4, norovírus GII.17 e sapovírus, devido à compatibilidade de metodologia de RT-qPCR e sua relevância como vírus prejudiciais à saúde humana.

Os *primers* a serem utilizados para a clonagem e para o qPCR foram desenhados seguindo as regras de tamanho, temperatura e conteúdo de GC requeridas. As sondas *single-quencer probes* também foram feitas de acordo com os requisitos necessários. A tabela de *Primers* e sondas encontra-se disponível no Anexo 25.

Espera-se que com os *primers* e sondas já desenhados, seja possível realizar o RT-qPCR, à princípio dos cinco vírus escolhidos, espera-se também confirmar a quantidade destes vírus circulando no DF e futuramente corroborar esses dados com possíveis dados clínicos com intuito não somente de desenvolver uma ferramenta de detecção, mas também como alerta para a circulação desse patógeno em águas tratadas.

REFERÊNCIAS

AHMED, S. M., LOPMAN, B. A., & LEVY, K. (2013). A Systematic Review and Meta-Analysis of the Global Seasonality of Norovirus, 8(10), 1–7. <https://doi.org/10.1371/journal.pone.0075922>

ANDRADE, J. DA S. R., ROCHA, M. S., CARVALHO-COSTA, F. A., FIORETTI, J. M., XAVIER, M. DA P. T. P., NUNES, Z. M. A., ... MIAGOSTOVICH, M. P. (2014). Noroviruses associated with outbreaks of acute gastroenteritis in the State of Rio Grande do Sul, Brazil, 2004-2011. *Journal of Clinical Virology*, 61(3), 345–352. <https://doi.org/10.1016/j.jcv.2014.08.024>

ANDREASI MS, CARDOSO DD, FERNANDES SM, TOZZETI IA, BORGES AM, FIACCADORI FS, ET AL. Adenovirus, calicivirus and astrovirus detection in fecal samples of hospitalized children with acute gastroenteritis from Campo Grande MS, Brazil. *Mem Inst Oswaldo Cruz* 2008;103:741–4

ANJOS, K., LIMA, L. M. P., SILVA, P. A., INOUE-NAGATA, A. K., & NAGATA, T. (2011). The possible molecular evolution of sapoviruses by inter- and intra-genogroup recombination. *Archives of virology*, 156(11), 1953–1959. <https://doi.org/10.1007/s00705-011-1079-8>

ANJOS, KAROLINE DOS. Calicivírus humanos: pesquisa em amostra de esgoto, construção de clone e modelagem estrutural. 2017. vii, 116 f., il. Tese (Doutorado em Biologia Molecular) —Universidade de Brasília, Brasília, 2017.

ARAGÃO, G.C., OLIVEIRA, D.S., SANTOS, M.C., ET AL., 2010. Molecular characterization of norovirus, sapovirus and astrovirus in children with acute gastroenteritis from belém, pará, brazil. *Rev. Pan-amaz. Saúde* 1 (1), 149–157.

BANKEVICH A, NURK S, ANTIPOV D, ET AL. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021.

BARREIRA DM, FERREIRA MS, FUMIAN TM, CHECON R, DE SADOVSKY AD, LEITE JP, ET AL. Viral load and genotypes of noroviruses in symptomatic and asymptomatic children in Southeastern Brazil. *J Clin Virol* 2010; 47: 60–4.

BARTSCH SM, LOPMAN BA, OZAWA S, HALL AJ, LEE BY (2016) Global Economic Burden of Norovirus Gastroenteritis. *PLoS ONE* 11(4): e0151219. <https://doi.org/10.1371/journal.pone.0151219>

BATISTA, L. F. (2015). Lodos gerados nas Estações de Tratamento de Esgotos no Distrito Federal: Um estudo de sua aptidão para o condicionamento, utilização e disposição final. *Universidade de Brasília. Faculdade de Tecnologia. Departamento de Engenharia Civil E Ambiental*, 197

BIBBY K, PECCIA J. 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* 47:1945– 1951. <http://dx.doi.org/10.1021/es305181x>.

BITLER, E. J., MATTHEWS, J. E., DICKEY, B. W., & EISENBERG, J. N. S. (2013). Norovirus outbreaks: a systematic review of commonly implicated transmission routes and vehicles, 141 (8): 1563–1571. <https://doi.org/10.1017/S095026881300006X>

BLANTON, L. H., ADAMS, S. M., BEARD, R. S., WEI, G., BULENS, S. N., WIDDOWSON, M., MONROE, S. S. (2006). Molecular and Epidemiologic Trends of Caliciviruses Associated with Outbreaks of Acute Gastroenteritis in the United States, 2000 – 2004, 193(3), 413-21.

BLAWID, R., SILVA, J. M. F., & NAGATA, T. (2017). Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline, 170, 301–314. <https://doi.org/10.1111/aab.12345>

BLOSTROM AL, WIDEN F, HAMMER AS, ET AL. Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome by use of viral metagenomics. *J Clin Microbiol* 2010;48:4392–4396.

BOSCH, A. 1998. Human enteric viruses in the water environment: a minireview. *International microbiology*, 1: 191-196

BOSCH, A.; PINTO, R.M.; GUIX, S. Human astroviruses. *Clin. Microbiol. Rev.* 2014, 27, 1048–1074.

BULL RA, HANSMAN GS, CLANCY LE, et al. Norovirus recombination in ORF1/ORF2 overlap. *Emerg Infect Dis* 2005;11:1079–1085

BUTOT, S., LE, F. S., KROL, J., PUTALLAZ, T., AMOROSO, R., & SANCHEZ, G. (2010). Evaluation of various real-time RT-PCR assays for the detection and quantitation of human norovirus, 167, 90–94. <https://doi.org/10.1016/j.jviromet.2010.03.018>

CALDERON-MARGALIT R, SHEFFER R, HALPERIN T, ET AL. A large-scale gastroenteritis outbreak associated with norovirus in nursing homes. *Epidemiol Infect* 2005; 133: 35–40.

CANTALUPO P.G, et al. 2011. Raw sewage harbors diverse viral populations. *mBio* 2(5):e00180-11. doi:10.1128/mBio.00180-11

CDC, <https://www.cdc.gov/norovirus/trends-outbreaks.html>

CDC. Updated norovirus outbreak management and disease prevention guidelines. *MMWR Recomm Rep* 2011; 60: 1–20.

CHERNICHARO, C. A. de L.; DANIEL, L. A.; SENS, M.; CORAUCCI FILHO, B. Pós- tratamento de efluentes de reatores anaeróbios por sistemas de desinfecção. In: CHENICHARO, C. A. (Coord.). Pós-tratamento de efluentes de reatores anaeróbios. Belo Horizonte, 2001. p. 377-454.

CHIBA, S., SAKUMA Y., KOGASAKA, R., AKIHARA, M., HORINO, K., NAKAO, T., FUKUI, S. 1979. An outbreak of gastroenteritis associated with calicivirus in an infant home. *J med virol* 4:249–254.

CHOI HL, SUH C-I, PARK S-W, JIN J-Y, CHO H-G, PAIK S-Y (2015) Whole-Genome Sequencing Analysis of Sapovirus Detected in South Korea. *PLoS ONE* 10(7): e0132328. doi:10.1371/journal.pone.0132328

CRUZ JR, BARTLETT AV, HERRMANN JE, Et al. Astrovirus associated Diarrhea among Guatemalan ambulatory rural children. *J Clin Microbiol* 1992; 30:1140–1144

CUBITT WD, BLACKLOW NR, HERRMANN JE, NOWAK NA, NAKATA S, CHIBA S. 1987. Antigenic relationships between human caliciviruses and Norwalk virus. *J Infect Dis* 156:806–814. <http://dx.doi.org/10.1093/infdis/156.5.806>

DALTON, R.M., ROMAN, E.R., NEGREDO, A.A., WILHELMI, D. I., GLASS. R.I., SANCHEZ-FAUQUIER A. Astrovirus acute gastroenteritis among children in Madrid, Spain. *Pediatr Infect Dis J*, 2002;21:1038-4

DAVIES-COLLEY, R. J.; DONNISON, A. M.; SPEED, D. J.; ROSS, C. M.; NAGELS, J. W. Inactivation of fecal indicator microorganisms in waste stabilization ponds: interactions of environmental factors with sunlight. *Water research*, v. 33, n. 5, p.1220-1230, 1999.

DOLIN R, BLACKLOW NR, DUPONT H, ET al. Biological properties of Norwalk agent of acute infectious nonbacterial gastroenteritis. *Proc Soc Exp Biol Med* 1972;140:578–583.

EFTIM, SORINA E, HONG, TAO, SOLLER, JEFFREY, BOEHM, ALEXANDRIA, WARREN, ISAAC, ICHIDA, AUDREY, NAPPIER, SHARON P. Occurrence of norovirus in raw sewage - A systematic literature review and meta-analysis. *Water research*, ISSN: 1879-2448, Vol: 111, Page: 366-374, 2017

ESPINOSA AC, MAZARI-HIRIART M, ESPINOSA R, et al. Infectivity and genome persistence of rotavirus and astrovirus in groundwater and surface water. *Water Res* 2008;42:2618–2628.

FARKAS T, ZHONG WM, JING Y, HUANG PW, ESPINOSA SM, MARTINEZ N, MORROW AL, RUIZ-PALACIOS GM, PICKERING LK, JIANG X. 2004. Genetic diversity among sapoviruses. *Arch Virol* 149:1309–1323. <http://dx.doi.org/10.1007/s00705-004-0296-9>.

FERRAZ, I. Multiplicam-se relatos de brasilienses que passam mal com água do lago. *METRÓPOLES*. Brasília, 22 dez. 2017. Disponível em: <https://www.metropoles.com/distrito-federal/apos-passar-mal-moradores-se-previnem-contra-agua-suspeita>. Acesso em: 30 jan. 2018

FERREIRA MS, VICTORIA M, CARVALHO-COSTA FA, VIEIRA CB, XAVIER MP, FIORETTI JM, et al. Surveillance of norovirus infections in the state of Rio de Janeiro Brazil 2005–2008. *J Med Virol* 2010; 82: 1442–8.

FERREIRA. M.S.R., GARCIA RCC, XAVIER MPTP, RIBEIRO RL, ASSIS RM, MOTA MCMS, LEITE JPG, MIAGOSTOVICH MP, DE OLIVEIRA SA 2012c. Genotyping of gastroenteric viruses in hospitalised children: first report of norovirus gii.21 in Brazil. *Mem Inst Oswaldo Cruz* 107: 1064-1067

FINKBEINER SR, LI Y, RUONE S, et al. Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. *J Virol* 2009; 83: 10836–10839

FINKBEINER, S.R.; KIRKWOOD, C.D.; WANG, D. Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. *Viol. J.* 2008, 5, 117

FIORETTI JM, FERREIRA MS, VICTORIA M, VIEIRA CB, XAVIER MDAP, LEITE JP, et al. Genetic diversity of noroviruses in Brazil. *Mem Inst Oswaldo Cruz* 2011;106:942–7.

FIORETTI, J. M., ROCHA, M. S., FUMIAN, T. M., GINUINO, A., DA SILVA, T. P., DE ASSIS, M. R., ... MIAGOSTOVICH, M. P. (2016). Occurrence of human sapoviruses in wastewater and stool samples in Rio de Janeiro, Brazil. *Journal of Applied Microbiology*, 121(3), 855–862. <https://doi.org/10.1111/jam.13205>

FONG, T., & LIPP, E. K. (2005). Enteric Viruses of Humans and Animals in Aquatic Environments: Health Risks, Detection, and Potential Water Quality Assessment Tools, 69(2), 357–371. <https://doi.org/10.1128/MMBR.69.2.357>

FRANKHAUSER RL, MONROE SS, NOEL JS, ET AL. Epidemiologic and molecular trends of “Norwalk-like viruses” associated with outbreaks of gastroenteritis in the United States. *J Infect Dis* 2002; 186: 1–7.

GABBAY YB, CHAMONE CB, NAKAMURA LS, et al. Characterization of an astrovirus genotype 2 strain causing an extensive outbreak of gastroenteritis among Maxakali Indians, Southeast Brazil. *J Clin Virol* 2006; 37:287–292.

GENSBERGER, E. T., & KOSTIC, T. Novel tools for environmental virology. *Current Opinion in Virology* 2013, 3: 61-68.

GRIFFIN, D. W.; GIBSON, C. J. LII; LIPP, E. K.; RILEY, K.; PAUL, J. H. LII, ROSE, J. B. 1999. Detection of viral pathogens by reverse transcriptase PCR and of microbial indicators by standard methods in the canals of the Florida Keys. *Applied and Environmental Microbiology*, 65: 4118-4125. Erratum in: *Applied and Environmental Microbiology*, 2000. 66: 876.

HANSMAN GS, GUNTAPONG R, PONGSUWANNA Y, NATORI K, KATAYAMA K, TAKEDA N. 2006. Development of an antigen ELISA to detect sapovirus in

clinical stool specimens. *Arch Virol* 151:551–561.
<http://dx.doi.org/10.1007/s00705-005-0630-x>.

HANSMAN GS, TAKEDA N, OKA T, OSETO M, HEDLUND KO, KATAYAMA K. 2005. Intergenogroup recombination in sapoviruses. *Emerg. Infect. Dis.* 11:1916–1920.

HARAMOTO, E., KATAYAMA, H., OGUMA, K., YAMASHITA, H., TAJIMA, A., NAKAJIMA, H., ... Management, I. (2006). Seasonal profiles of human noroviruses and indicator bacteria in a wastewater treatment plant in Tokyo, Japan, 301–308. <https://doi.org/10.2166/wst.2006.888>

HEDLUND KO, RUBILAR-ABREU E, SVENSSON L. 2000. Epidemiology of calicivirus infections in Sweden, 1994–1998. *J Infect Dis* 181: S275–S280

JOHANSSON PJ, BERGENTOFT K, LARSSON PA, MAGNUSSON G, WIDELL A, THORHAGEN M, HEDLUND KO. 2005. A nosocomial sapovirus- associated outbreak of gastroenteritis in adults. *Scand J Infect Dis* 37:200–204.

KAGEYAMA, T., KOJIMA, S., SHINOHARA, M., UCHIDA, K., FUKUSHI, S., HOSHINO, F. B., KATAYAMA, K. (2003). Broadly reactive and highly sensitive assay for norwalk-like viruses based on real-time quantitative reverse transcription-pcr, 41(4), 1548–1557. <https://doi.org/10.1128/jcm.41.4.1548>

KAPLAN JE, GARY GW, BARON RC, et al. Epidemiology of Norwalk gastroenteritis and the role of Norwalk virus in outbreaks of acute nonbacterial gastroenteritis. *Ann Intern Med* 1982; 96: 756–761.

KAPOOR, A.; LI, L.; VICTORIA, J.; ODERINDE, B.; MASON, C.; PANDEY, P.; ZAIDI, S.Z.; DELWART, E. Multiple novel astrovirus species in human stool. *J. Gen. Virol.* 2009, 90, 2965–2972

KAUFMAN SS, CHATTERJEE NK, FUSCHINO ME, et al. Characteristics of human calicivirus enteritis in intestinal transplant recipients. *J Pediatr Gastroenterol Nutr* 2005; 40: 328–333.

KIM MJ, KIM YJ, LEE JH, ET AL. Norovirus: a possible cause of pneumatosis intestinalis. *J Pediatr Gastroenterol Nutr* 2011;52:314–318.

KIRBY, A. E., STREBY, A., & MOE, C. L. (2016). Vomiting as a Symptom and Transmission Risk in Norovirus Illness: Evidence from Human Challenge Studies, 1–10. <https://doi.org/10.5061/dryad.sk800>

KOMOLAFE OO, JARRETT O. A possible maturation pathway of calicivirus particles. *Microbios* 1986;46:103–111.

KRONEMAN A, VEGA E, VENNEMA H, VINJE J, WHITE PA, HANSMAN G, GREEN K, MARTELLA V, KATAYAMA K, KOOPMANS M. Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* 2013;158:2059–2068.

LAVERICK, M. A., A. P. WYN-JONES, AND M. J. CARTER. 2004. Quantitative RT-PCR for the enumeration of noroviruses (Norwalk-like viruses) in water and sewage. *Lett. Appl. Microbiol.* 39:127–136.

LE CANN P, RANARIJAONA S, MONPOEHO S, et al. Quantification of human astroviruses in sewage using real-time RT-PCR. *Res Microbiol* 2004; 155:11–15.

LE GUYADER, F. S., L. HAUGARREAU, L. MIOSSEC, E. DUBOIS, AND M. POM- MEPUY. 2000. Three-year study to assess human enteric viruses in shellfish. *Appl. Environ. Microbiol.* 66:3241–3248

LE GUYADER, F. S., LE SAUX, J. C., AMBERT-BALAY, K., KROL, J., SERAIS, O., PARNAUDEAU, S., ... ATMAR, R. L. (2008). Aichi virus, norovirus, astrovirus, enterovirus, and rotavirus involved in clinical cases from a French oyster-related gastroenteritis outbreak. *Journal of Clinical Microbiology*, 46(12), 4011–4017. <https://doi.org/10.1128/JCM.01044-08>

LI L, PESAVENTO PA, SHAN T, LEUTENEGGER CM, WANG C, DELWART E. 2011. Viruses in diarrhoeic dogs include novel kobuviruses and sapo viruses. *J Gen Virol* 92:2534–2541. <http://dx.doi.org/10.1099/vir.0.034611-0>.

LI L, SHAN T, WANG C, COTE C, KOLMAN J, ONIONS D, GULLAND FM, DELWART E. 2011. The fecal viral flora of California sea lions. *J Virol* 85:9909–9917. <http://dx.doi.org/10.1128/JVI.05026-11>.

LI, D., LIU, C-M., LUO, R., SADAKANE, K., AND LAM, T-W., (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, doi: 10.1093/bioinformatics/btv033 [PMID: 25609793]

LIU B, CLARKE IN, LAMBDEN PR. Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by in vitro mutagenesis. *J Virol* 1996;70:2605–2610.

LIU J, KIBIKI G, MARO V, MARO A, KUMBURU H, SWAI N, TANIUCHI M, GRATZ J, TONEY D, KANG G, HOUPPT E. 2011. Multiplex reverse transcription PCR Luminex assay for detection and quantitation of viral agents of gastroenteritis. *J Clin Virol* 50:308–313. <http://dx.doi.org/10.1016/j.jcv.2010.12.009>.

LODDER, W.J. AND DE RODA HUSMAN, A.M. (2005). Presence of noroviruses and other enteric viruses in sewage and surface waters in The Netherlands. *Applied and Environmental Microbiology*, 71, 1453–1461.

LOPMAN B, VENNEMA H, KOHLI E, POTHIER P, SANCHEZ A, et al. (2004) Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* 363: 682–688.

LOPMAN BA, ADAK GK, REACHER MH, BROWN DW (2003) Two epidemiologic patterns of norovirus outbreaks: surveillance in England and Wales, 1992–2000. *Emerg Infect Dis* 9: 71–77.

MADELEY CR. 1979. Comparison of the features of astroviruses and caliciviruses seen in samples of feces by electron microscopy. *J Infect Dis* 139:519–523. <http://dx.doi.org/10.1093/infdis/139.5.519>

MADELEY, C.R.; COSGROVE B.P. Viruses in infantile gastroenteritis. *The Lancet*, 306(7925), p124 (1975) [http://dx.doi.org/10.1016/S0140-6736\(75\)90020-3](http://dx.doi.org/10.1016/S0140-6736(75)90020-3)

MARDIS, E. R. (2011). PERSPECTIVE A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198–203. <https://doi.org/10.1038/nature09796>

MARSHALL JA, BRUGGUNK LD. The dynamics of norovirus outbreak epidemics: recent insights. *Int J Environ Res Public Health* 2011;8:1141–9.

MATSUI, S. M., & GREENBERG, H. B. (2001). Immunity to caliciviruses. *Journal of infectious diseases*, 181(suppl. 2), s331–s335.

MAUNULA L, KALSO S, VON BONSDORFF CH, et al. Wading pool water contaminated with both noroviruses and astroviruses as the source of a gastroenteritis outbreak. *Epidemiol Infect* 2004;132:737–743.

MENZEL, P. ET AL. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.*7:11257 doi: 10.1038/ncomms11257 (2016)

MEYER, S. T. O uso de cloro na desinfecção de águas, a formação de trihalometanos e os riscos potenciais à saúde pública. *Cad. Saúde Públ.*, Rio de Janeiro, 10 (1): 99-110, jan/mar, 1994.

MONROE SS. Molecular epidemiology of human astroviruses. In: Desselberger U, Gray JJ, eds. *Perspectives in Medical Virology. Viral Gastroenteritis, vol. 9*. Amsterdam: Elsevier; 2003:607–616.

MYRMEL, M., E. M. M. BERG, B. GRINDE, AND E. RIMSTAD. 2006. Enteric viruses in inlet and outlet samples from sewage treatment plants. *J. Water Health* 4:197–209.

NADAN S, WALTER JE, GRABOW WO, ET AL. Molecular characterization of astroviruses by reverse transcriptase PCR and sequence analysis: comparison of clinical and environmental isolates from South Africa. *Appl Environ Microbiol* 2003;69:747–753.

NAKAMURA, S., YANG, C. S., SAKON, N., UEDA, M., TOUGAN, T., YAMASHITA, A., ... NAKAYA, T. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput

sequencing approach. PLoS ONE, 4(1), 1–8.
<https://doi.org/10.1371/journal.pone.0004219>

NG TF, MARINE R, WANG C, SIMMONDS P, KAPUSINSZKY B, BODHIDATTA L, ODERINDE BS, WOMMACK KE, DELWART E. 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* 86:12161–12175. <http://dx.doi.org/10.1128/JVI.00869-12>.

NILSSON M, HEDLUND KO, THORHAGEN M, et al. Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J Virol* 2003; 77: 13117–13124.

NOEL JS, LIU BL, HUMPHREY CD, RODRIGUEZ EM, LAMBDEN PR, CLARKE IN, DWYERDM, ANDO T, GLASSRI, MONROE SS. 1997. Parkville virus: A novel genetic variant of human calicivirus in the Sapporo virus clade, associated with an outbreak of gastroenteritis in adults. *J Med Virol* 52:173–178.

OKA, T., MORI, K., IRITANI, N. et al. Human sapovirus classification based on complete capsid nucleotide sequences *Arch Virol* (2012) 157: 349. <https://doi.org/10.1007/s00705-011-1161-2>

OLIVEIRA, D. M. M., SOUZA, M., FIACCADORI, F. S., SANTOS, H.C.P., CARDOSO, D.D.P. (2014) Monitoring of Calicivirus among day-care children: evidence of asymptomatic viral excretion and first report of GI.7 Norovirus and GI.3 Sapovirus in Brazil. *J Med Virol* 86, 1569–1575

ONDOV BD, BERGMAN NH, AND PHILLIPPY AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011 Sep 30; 12(1):385.

PALOMBO EA, BISHOP RF. Annual incidence, serotype distribution, and genetic diversity of human astrovirus isolates from hospitalized children in Melbourne, Australia. *J Clin Microbiol* 1996;34:1750–1753.

PANG XL, LEE BE, TYRRELL GJ, PREIKSAITIS JK. 2009. Epidemiology and genotype analysis of sapovirus associated with gastroenteritis outbreaks in Alberta, Canada: 2004–2007. *J Infect Dis* 199:547– 551.

PATEL MM, WIDDOWSON MA, GLASS RI, et al. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg Infect Dis* 2008; 14: 1224–1231.

PENG, Y., LEUNG, H. C. M., YIU, S. M., CHIN, F. Y. I. (2012) IDBA-ud: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28, 1420–1428

PREVOST, B., LUCAS, F. S., GONCALVES, A., RICHARD, F., MOULIN, L., & WURTZER, S. (2015). Large scale survey of enteric viruses in river and waste water underlines the health status of the local population. *Environment International*, 79, 42–50. <https://doi.org/10.1016/j.envint.2015.03.004>

- PUSCH D, OH DY, WOLF S, et al. Detection of enteric viruses and bacterial indicators in German environmental waters. *Arch Virol* 2005;150(5):929–947.
- QUAN PL, WAGNER TA, BRIESE T, et al. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerg Infect Dis* 2010;16:918–925.
- ROCKX B, DEWIT M, VENNEMA H, VINJE J, DE BRUIN E, VAN DUYNHOVEN Y, KOOPMANS M. 2002. Natural history of human calicivirus infection: A prospective cohort study. *Clin Infect Dis* 35:246–253.
- ROSARIO, K., NILSSON, C., LIM, Y. W., RUAN, Y., & BREITBART, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*, 11(11), 2806–2820. <https://doi.org/10.1111/j.1462-2920.2009.01964.x>
- RZEZUTKA A, COOK N. 2004. Survival of human enteric viruses in the environment and food. *FEMS Microbiol Rev* 28:441-453.
- SANTOS, R. A. T., BORGES, A. M. T., COSTA, P. S. S., TEIXEIRA, J. M. S., GIUGLIANO, L. G., LEITE, J. P. G., CARDOSO, D. D. P.. Astrovirus infection in children living in the central west region of Brazil, 102(march), 209–213.
- SCHEUER KA, OKA T, HOET AE, GEBREYES WA, MOLLA BZ, SAIF LJ, WANG Q. 2013. Prevalence of porcine noroviruses, molecular characterization of emerging porcine sapoviruses from finisher swine in the United States, and unified classification scheme for sapoviruses. *J Clin Microbiol* 51: 2344–2353. <http://dx.doi.org/10.1128/JCM.00865-13>
- SCHLOSSBERG, D. (2015). *Clinical infectious disease* (Second ed.). p. 334. ISBN 9781107038912.
- SCHULTZ-CHERRY S, KING DJ, KOCI MD. Inactivation of an astrovirus associated with poult enteritis mortality syndrome. *Avian Dis* 2001;45:76–82.
- SEBIRE NJ, MALONE M, SHAH N, et al. Pathology of astrovirus associated diarrhoea in a paediatric bone marrow transplant recipient. *J Clin Pathol* 2004;57:1001–1003.
- SEITZ SR, LEON JS, SCHWAB KJ, et al. Norovirus infectivity in humans and persistence in water. *Appl Environ Microbiol* 2011;77:6884–6888.
- SIEBENGA JJ, VENNEMA H, ZHENG DP, VINJÉ J, LEE BE, PANG XL, et al. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J Infect Dis* 2009;200:802–12.
- SOARES CC, SANTOS N, BEARD RS, ALBUQUERQUE MC, MARANHÃO AG, ROCHA LN, et al. Norovirus detection and genotyping for children with gastroenteritis, Brazil. *Emerg Infect Dis* 2007;13:1244–6

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A., KUMAR, S. Mega6: molecular evolutionary genetics analysis version 6.0, molecular biology and evolution, volume 30, issue 12, 1 december 2013, pages 2725–2729, <https://doi.org/10.1093/molbev/mst197>

UEKI, Y., D. SANO, T. WATANABE, K. AKIYAMA, AND T. OMURA. 2005. Norovirus pathway in water environment estimated by genetic analysis of strains from patients of gastroenteritis, sewage, treated wastewater, river water and oysters. *Water Res.* 39:4271–4280

UEKI, Y., SHOJI, M., OKIMURA, Y., MIYOTA, Y., MASAGO, Y., OKA, T., OMURA, T. (2010). Detection of Sapovirus in oysters, (March), 483–486. <https://doi.org/10.1111/j.1348-0421.2010.00239.x>

Van den Berg, H., Lodder, W., van der Poel, W., Vennema, H. And de Roda Husman, A.M. (2005). Genetic diversity of noroviruses in raw and treated sewage water. *Research in Microbiology*, 156, 532–54

VIEIRA, C. B., CAROLINA, A., MENDES, D. O., OLIVEIRA, J. M. DE, GASPARG, M. C., PAULO, J., MIAGOSTOVICH, M. P. (2012). Vírus entéricos na lagoa rodrigo de Freitas, 16(August 2007), 540–565.

VILARIÑO ML, LE GUYADER FS, POLO D, et al. Assessment of human enteric viruses in cultured and wild bivalve molluscs. *Int Microbiol* 2009; 12:145–151

VU, D.-L., BOSCH, A., PINTO, R., & GUIX, S. (2017). Epidemiology of Classic and Novel Human Astrovirus: Gastroenteritis and Beyond. *Viruses*, 9(2), 33. <https://doi.org/10.3390/v9020033>

WALTER JE, MITCHELL DK. Astrovirus infection in children. *Curr Opin Infect Dis* 2003;16:247–253.

WANG Q, ZHANG Z, SAIF LJ. 2012. Stability of and attachment to lettuce by a culturable porcine sapovirus surrogate for human calici- viruses. *Appl Environ Microbiol* 78:3932–3940. <http://dx.doi.org/10.1128/AEM.06600-11>

WETTERSTRAND KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata. Accessed [29 Jan. 2018].

WIDDOWSON, M. A., A. SULKA, S. N. BULENS, R. S. BEARD, S. S. CHAVES, R. HAMMOND, E. D. P. SALEHI, E. SWANSON, J. TOTARO, R. WORON, P. S. MEAD, J. S. BRESEE, S. S. MONROE, AND R. I. Glass. 2005. Norovirus and foodborne disease, United States, 1991-2000. *Emerg. Infect. Dis.* 11:95–102

WORD HEALTH ORGANIZATION. Diarrhoeal disease. Fact Update. 2017. <http://www.who.int/mediacentre/factsheets/fs330/en/>

XIAOFANG LIU, DAI YAMAMOTO, MARIKO SAITO, TOSHIFUMI IMAGAWA, ADRIANNE ABLOLA, AMADO O. TANDOC, 3RD, EDELWISA SEGUBRE-

MERCADO, SOCORRO P. LUPISAN, MICHIKO OKAMOTO, YUKI FURUSE, ET AL. Molecular detection and characterization of sapovirus in hospitalized children with acute gastroenteritis in the Philippines. *J Clin virol.* 2015 jul; 68: 83–88. Published online 2015 may 6. Doi: 10.1016/j.jcv.2015.05.001

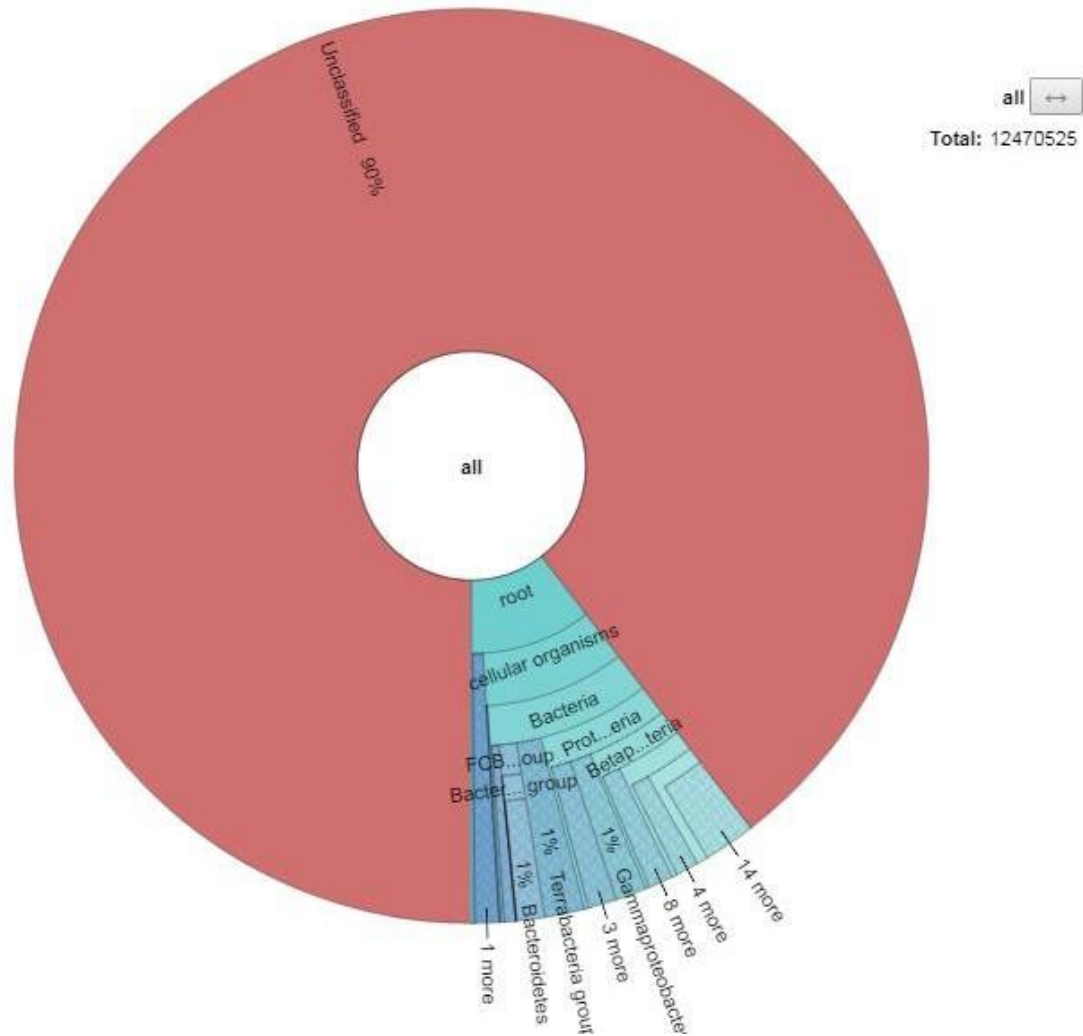
ZHANG T, et al. 2006. RNAviral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4:e3.

ZHENG DP, WIDDOWSON MA, GLASS RI, et al. Molecular epidemiology of genogroup II-genotype 4 noroviruses in the United States between 1994 and 2006. *J Clin Microbiol* 2010;48:168–177.

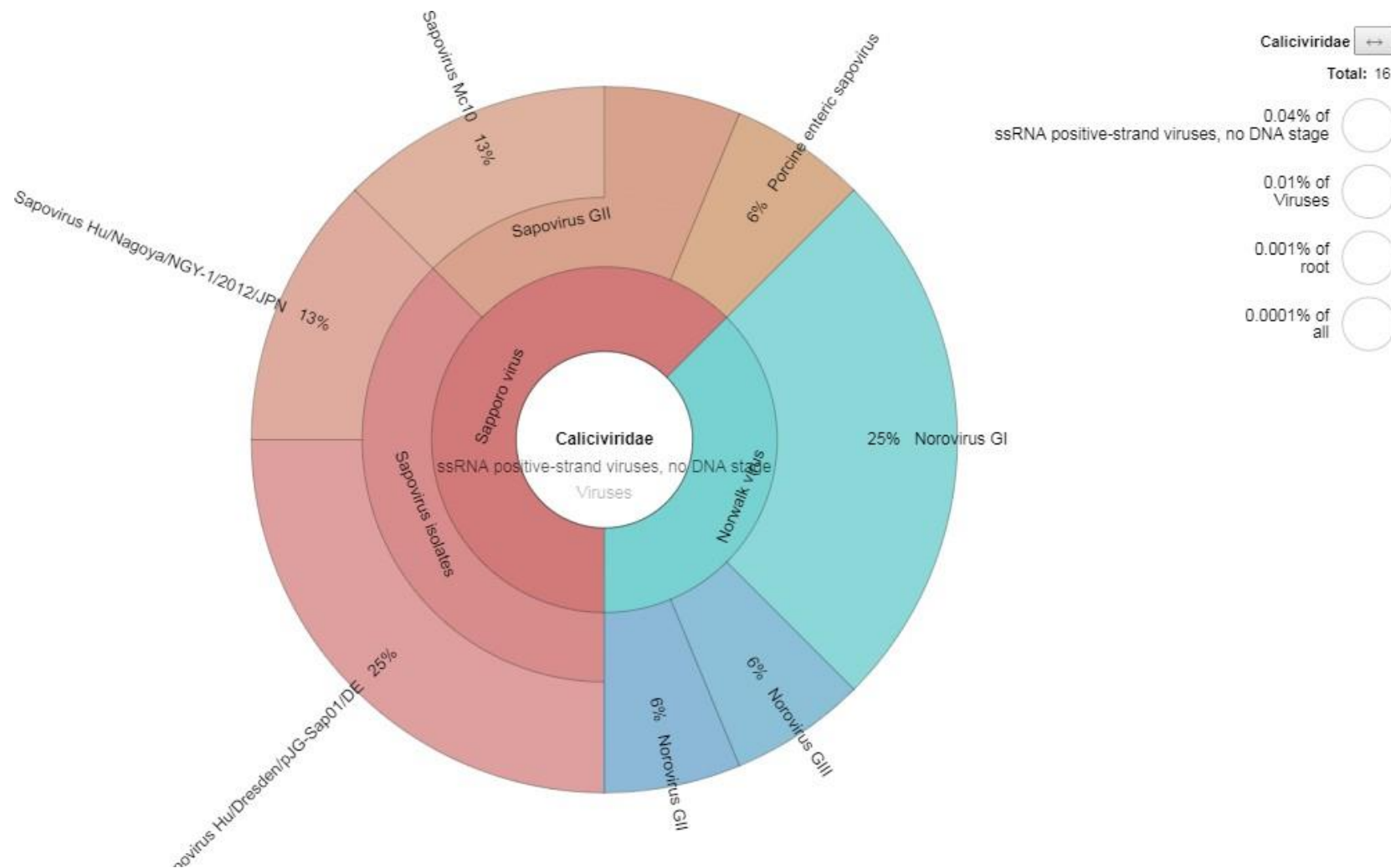
ZHENG, D. P., T. ANDO, R. L. FANKHAUSER, R. S. BEARD, R. I. GLASS, AND S. S. MONROE. 2006. Norovirus classification and proposed strain nomenclature. *Virology* 346:312–323

APÊNDICES

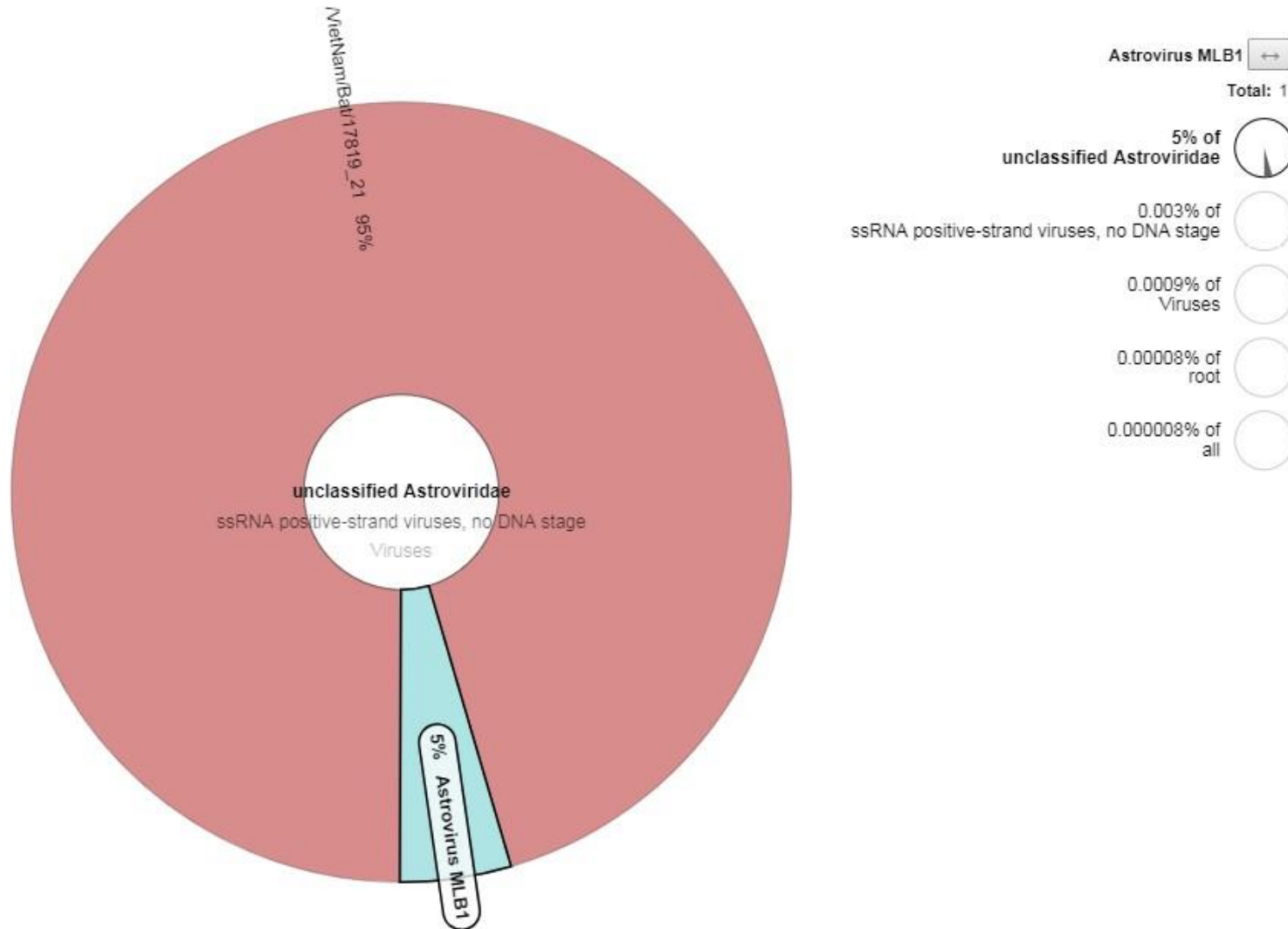
Apêndice 1 – Krona Chart obtido através da plataforma Kaiju para a amostra 1T. Do total de 12470525 *reads*, 1305828 foram classificados.



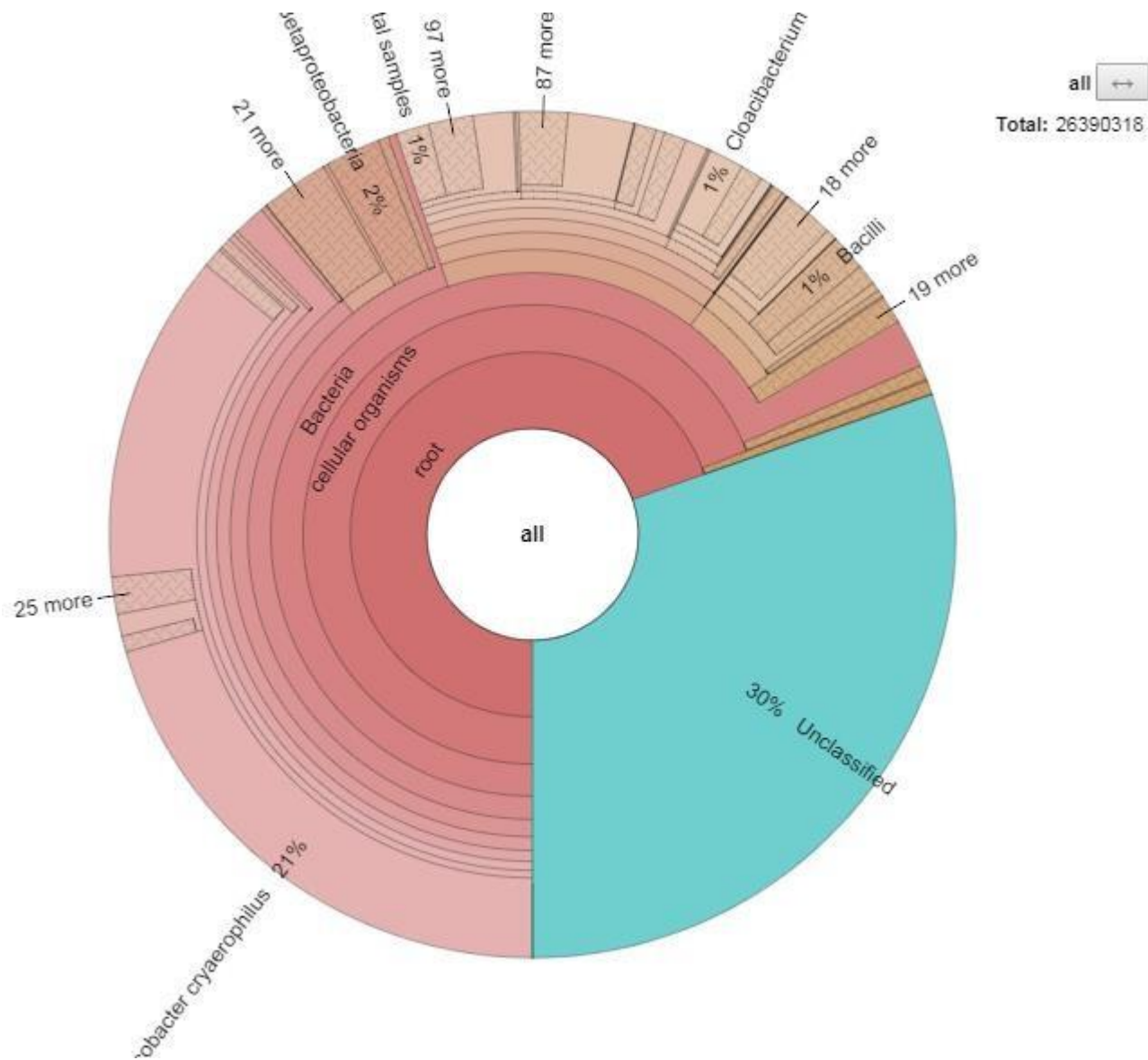
Apêndice 4 – Krona Chart obtido através da plataforma Kaiju para a amostra 1T. Os calcivírus correspondem a 0,04% do total de vírus de ssRNA senso positivo. Do total de *Caliciviridae*, 25% são Norovirus GI e 25% Sapovírus variante Dresden. Há destaque também para os SaV variantes de Nagoya e Mc10.



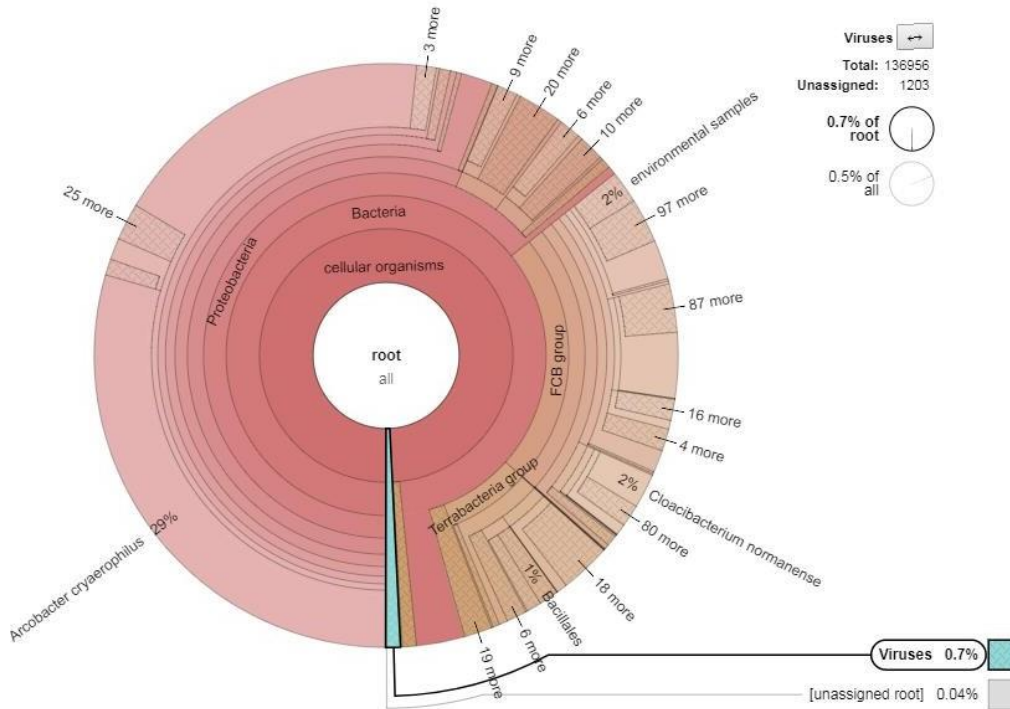
Apêndice 5 – Krona Chart obtido através da plataforma Kaiju para a amostra 1T. 0,06% do total de *reads* de vírus ssRNA senso positivo foi classificado como *Astroviridae*, a maioria correspondente a astrovírus que afetam morcegos. Foi classificado também o HAsTV MLB1.



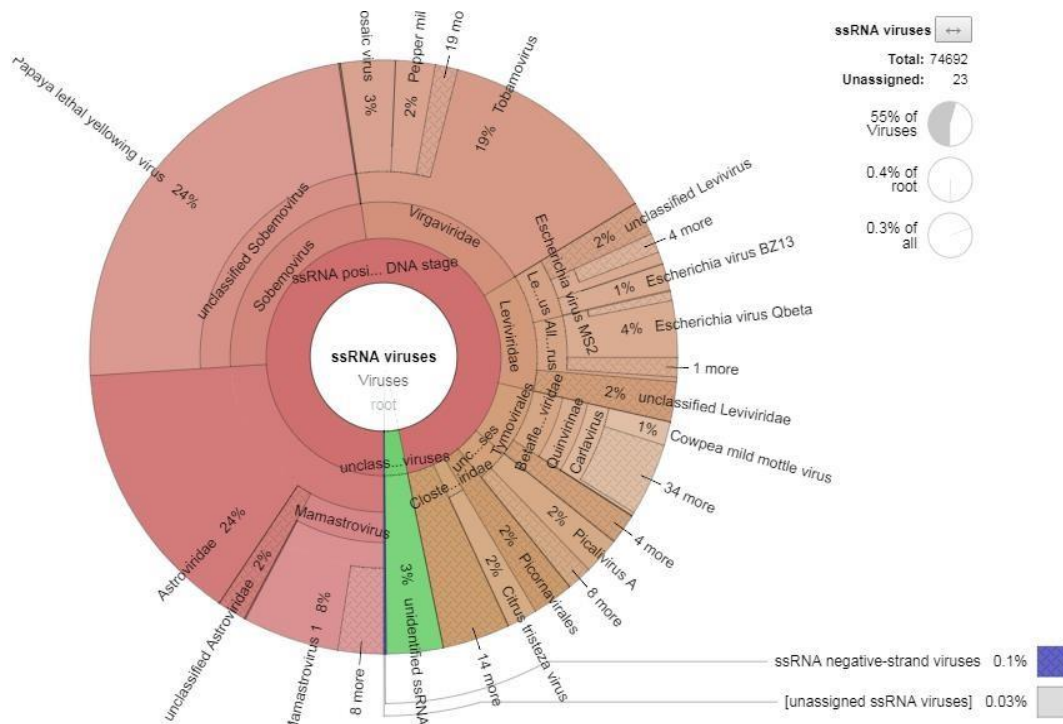
Apêndice 6 – Krona Chart obtido através da plataforma Kaiju para a amostra 2NT. 18376080 pares de *reads* de um total de 26390318 foram classificados.



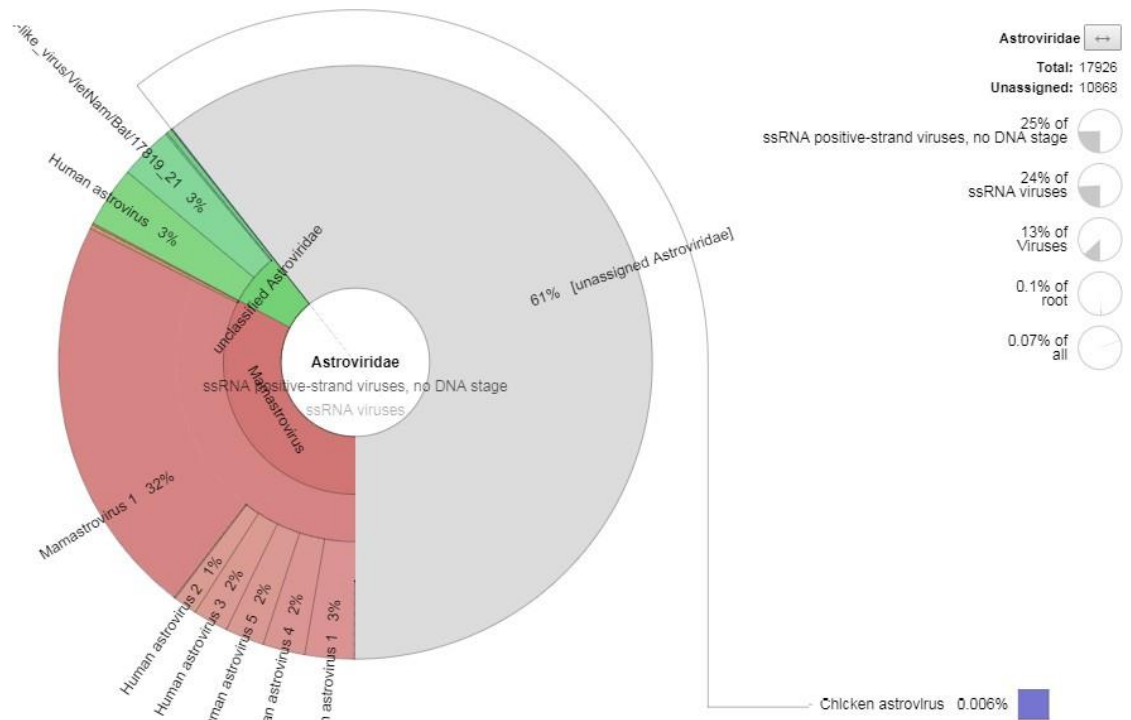
Apêndice 7 – Krona Chart obtido através da plataforma Kaiju para a amostra 2NT. Os vírus, com destaque em azul, correspondem a 0,7% do total de organismos.



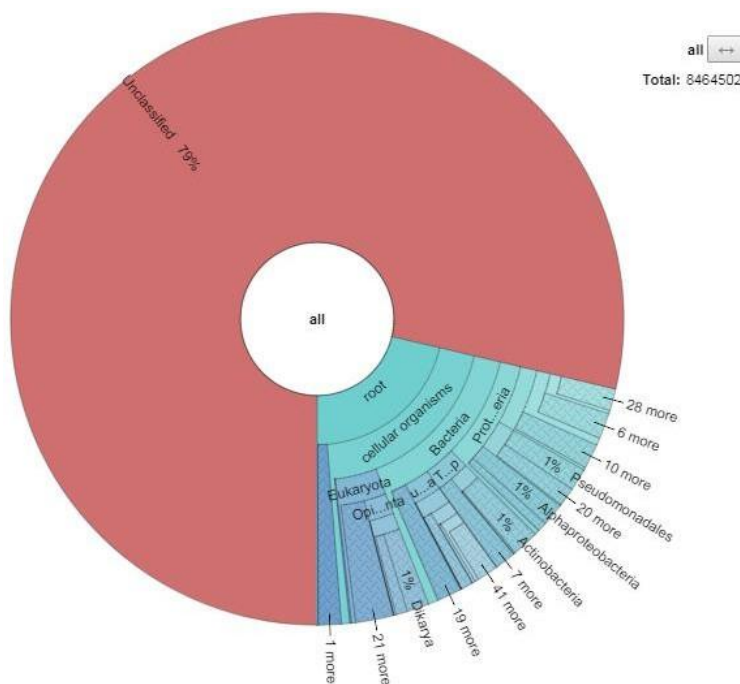
Apêndice 8 – Krona Chart obtido através da plataforma Kaiju para a amostra 2NT. 55% dos vírus são ssRNA. Os *Astroviridae* correspondem a 24%, equivalentes ao Papaya lethal yellowing vírus.



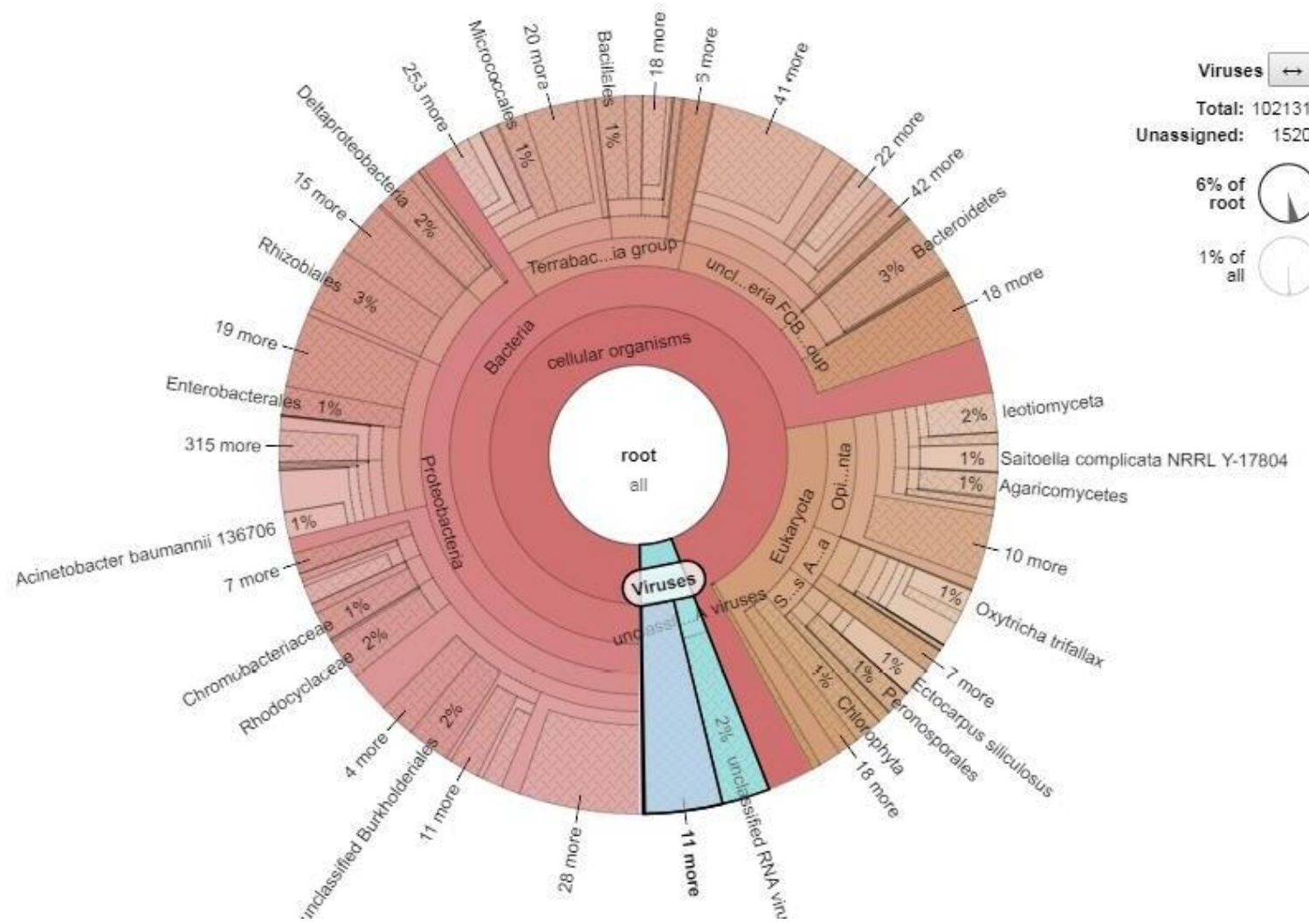
Apêndice 10 – Krona Chart obtido através da plataforma Kaiju para a amostra 2NT. Dos *Astroviridae* classificados, 32% correspondem a Mamastrovirus.



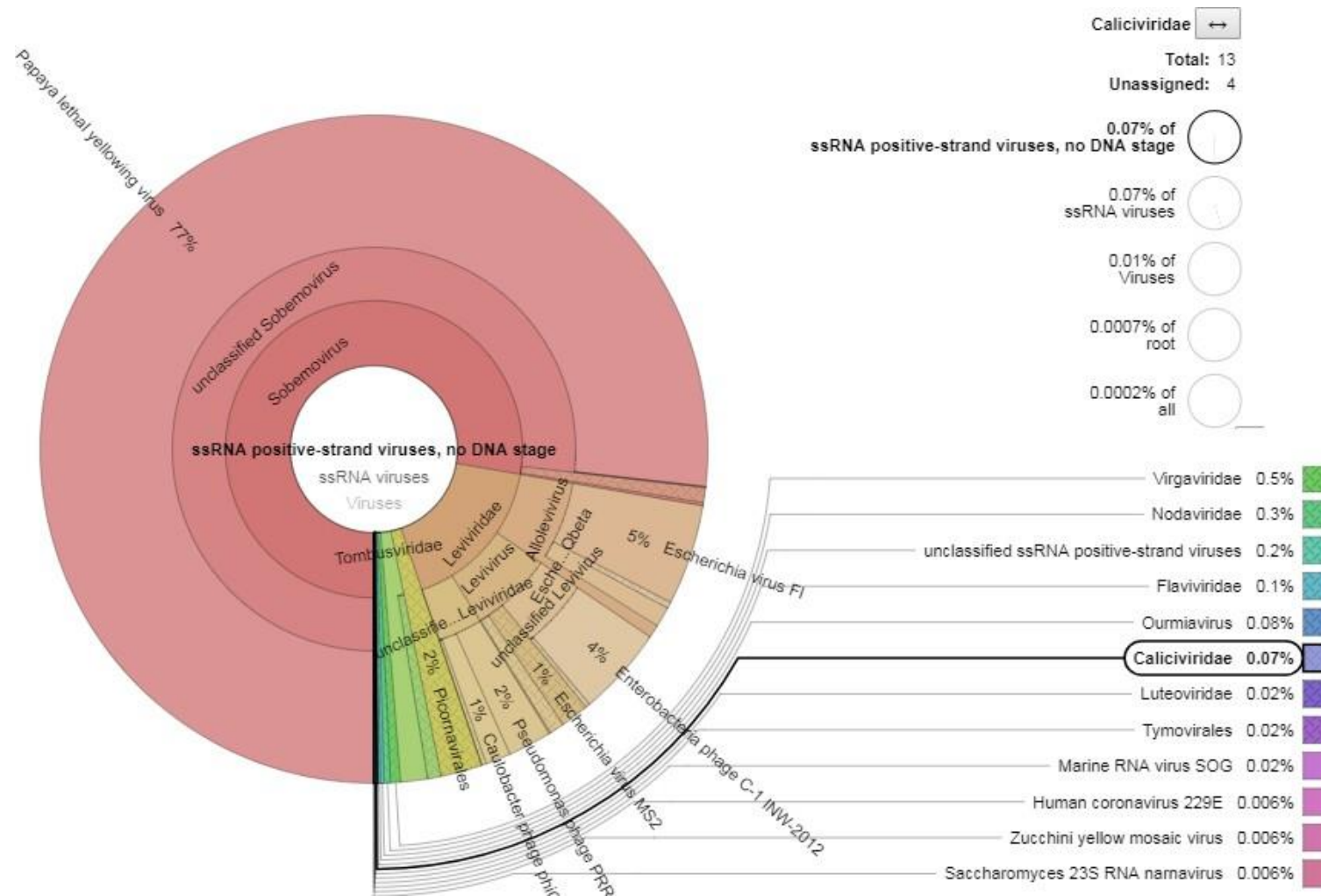
Apêndice 11 – Krona Chart obtido através da plataforma Kaiju para a amostra 2T. 1804758 de um total de 8464502 pares de *reads* foram classificados.



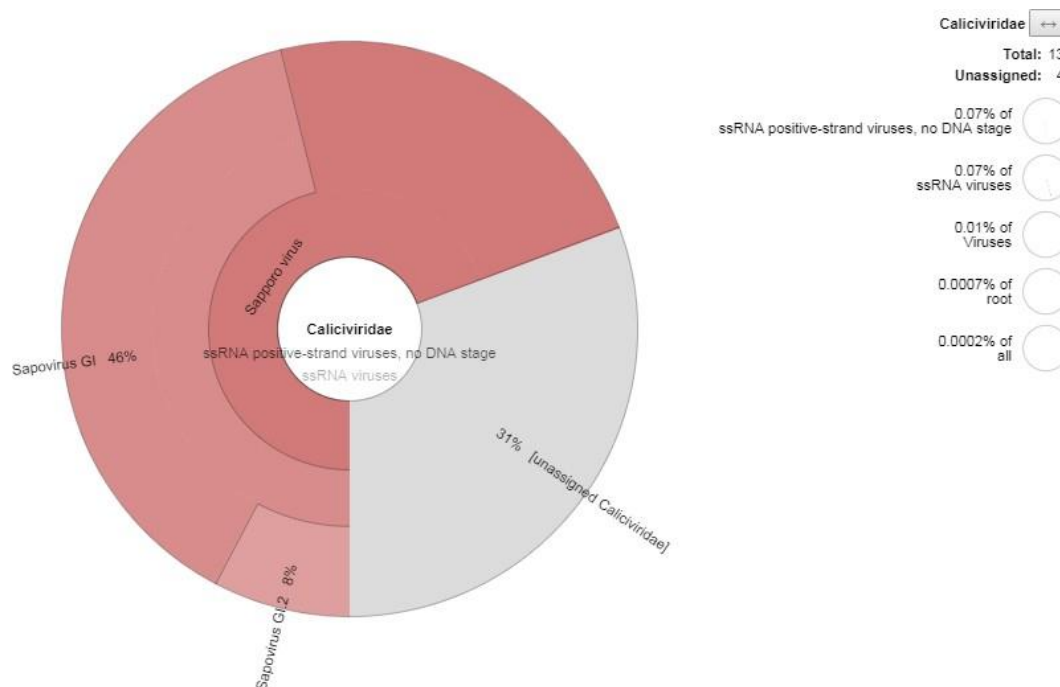
Apêndice 12 – Krona Chart obtido através da plataforma Kaiju para a amostra 2T. Os vírus, com destaque em azul, correspondem a 1% do total de organismos.



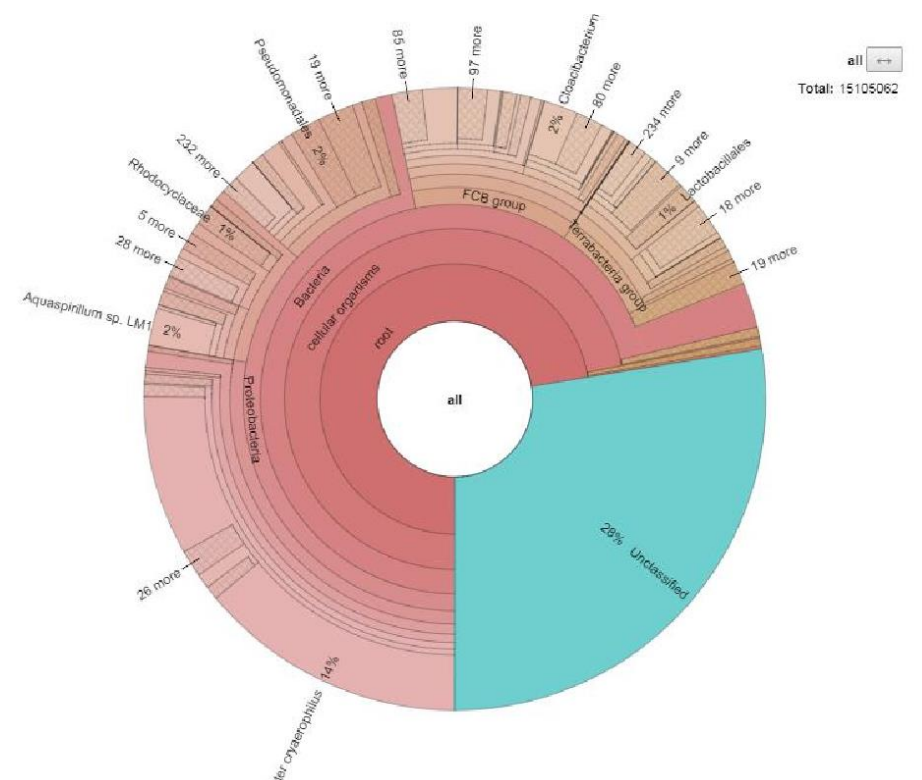
Apêndice 13 – Krona Chart obtido através da plataforma Kaiju para a amostra 2T. 77% dos vírus ssRNA senso positivo classificados pertencem ao Papaya lethal yellowing vírus, enquanto os *Caliciviridae* representam apenas 0,07%. Não foram encontrados *Astroviridae*.



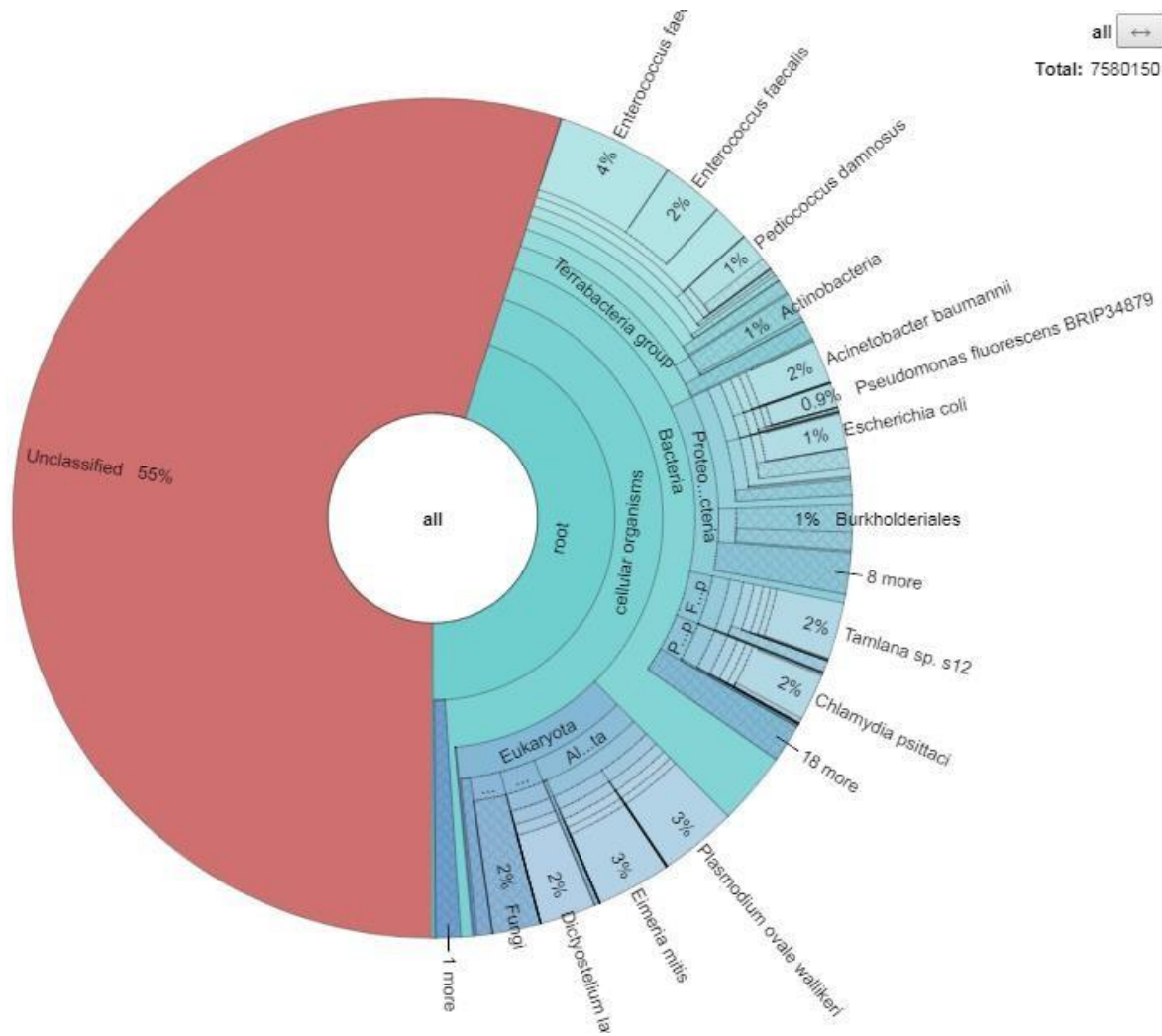
Apêndice 14 – Krona Chart obtido através da plataforma Kaiju para a amostra 2T. Dentre os *Caliciviridae*, encontra-se o Sapovírus GI. Não foram registrados norovírus.



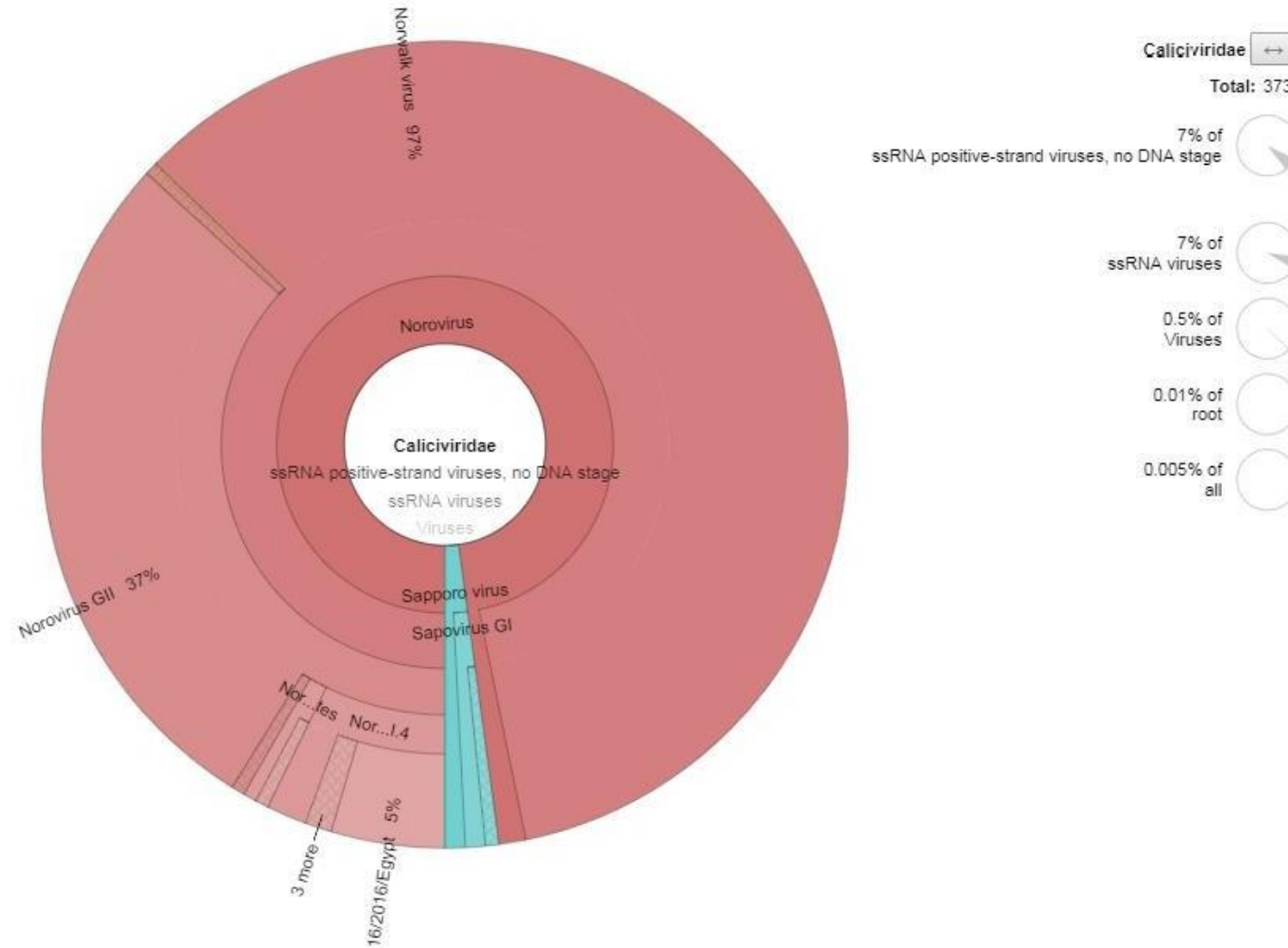
Apêndice 15 – Krona Chart obtido através da plataforma Kaiju para a amostra 3NT. 10938859 de 15105062 pares de *reads* foram classificados.



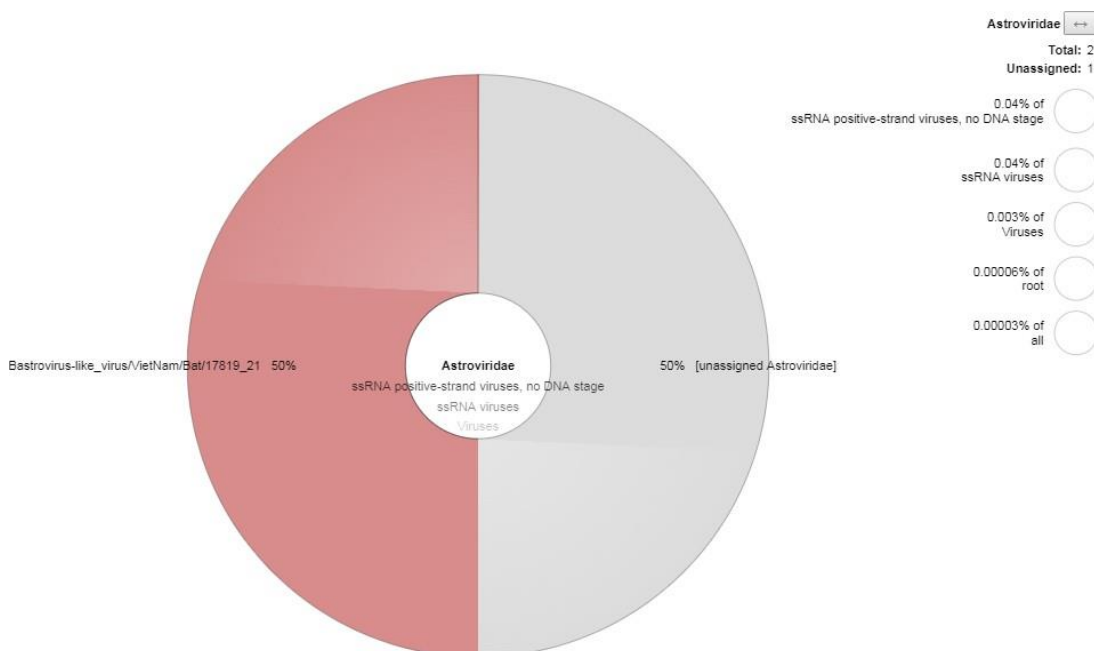
Apêndice 20 – Krona Chart obtido através da plataforma Kaiju para a amostra 3T. 3413760 de 7580150 pares de *reads* foram classificados. Mais da metade permaneceu sem classificação.



Apêndice 23 – Krona Chart obtido através da plataforma Kaiju para a amostra 3T. Apenas 2% dos *Caliciviridae* detectados pertence ao SaV, dentre eles está o SaV GI.2/BR-DF01/BRA/2009, caracterizado pelo nosso laboratório. Os destaques de NoV GI.4 são as variantes Cairo16/2016/Egypt e Taoyuan/CGMH51/TW.



Apêndice 24 – Krona Chart obtido através da plataforma Kaiju para a amostra 3T. Não foram classificados astrovírus humanos, apenas o Bastrovirus like, representado pela porção vermelha do gráfico.



Apêndice 25 – Tabela de *primers* e sondas desenhadas para o experimento de qPCR.

Primers e sondas desenhados para o experimento de qPCR

Nome do primer/sonda	Sequência
HAstro Cloning 3969 For* ¹	TGGAAGAMCTATGAYACTGT
HAstro qPCR 4155 For	TCCTRTGCTAYCAGTTGCT
HAstro Probe 4280 Rev* ²	5' FAM-CCA AAT GCG ATG GAG TTG CTC TTC-BHQ1 3'
HAstro qPCR 4327 Rev.	ATC RCA CTT CTT TGG TCC TC
HAstro Cloning 4543 Rev	ACA CCY TGT TTC CTG AGT TG
NoroGI Cloning 5192 For	GGTGAAAAGTTYTACAGGAAGAT
NoroGI qPCR 5251 For	<u>Y</u> CTTGAAATGTATGTGCCAG
NoroGI Probe 5338 Rev.	5' TAMRA-GAG ATT GCG ATC TCC TGT CCA CAA - BHQ2 3'
NoroGI qPCR 5372 Rev.	CTT AGA CGC CAT CAT CAT TTC

NoroGI Cloning 5668 Rev	CCG ACC CAR CCA TTA TAC AT
NoroGII.4 Cloning 4969 For	GAAGGTGGCATGGATTTTTACGT
NoroGII.4 qPCR 5144 For	GTCAACAATGAGGTTATGGC
NoroGII.4 Probe 5259 Rev	5' FAM- CCA CCA GGG GCT TGT ACA AAA TT -BHQ 3'
NoroGII.4 qPCR 5294 Rev	GTT TCT AGG GGA TAC TGT AAA CT
NoroGII.4 Cloning 5457 Rev	GGT GGG ACT GCT GCA AAT AT
Noro GII.17 Cloning 5599 For	TGA GAT TAG TGG CTA TGC TCT A
NoroGII.17 qPCR 5775 For	TTTTAACCCCTTTCTGAGCTCA
NoroGII.17 Probe 5868 Rev	5'JOE - ACC TGC CAT TTT GAC ACT GCA C - BHQ' 3'
NoroGII.17 qPCR 5905 Rev	CTG GGT TGT GCC TTG TAA CT
Noro GII.17 Cloning 6080 Rev	GCC CTT GAA GTC AGG TGT AC
Sapo Cloning 4659 For	TYC ACA CST ACG GKG ATG A
Sapo qPCR 4925 For	CCA CCY GCY TYC GAT CG
Sapo Probe 5098Rev	5' FAM - GTA GGT gGC GAG AGC CTG GTC BHQ1 3'
Sapo qPCR 5122 Rev	TGT ACC aCC TAT GAA CCA mG
Sapo Cloning 5181Rev	CCC TCC ATY TCA AAC ACT AWT TTG

Números em nomes dos oligonucleotídeos representam posição nucleotídica de genomas baseadas em referências. Primers com nome de "Cloning" são para clonar fragmentos de cDNA em plasmídeo. "qPCR" são para qPCR e o termo "Probe" é para sondas. *1For = primers de senso e *2Rev, antisenso.