UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

Francine Cezar Bandeira Timm

ANÁLISE METAGENÔMICA DO VIROMA DE TONSILAS DE BÚFALOS DA ILHA DE MARAJÓ, PARÁ

Porto Alegre
2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

Francine Cezar Bandeira Timm

ANÁLISE METAGENÔMICA DO VIROMA DE TONSILAS DE BÚFALOS DA ILHA DE MARAJÓ, PARÁ

Este trabalho foi escrito conforme as normas da revista Archives of Virology (ISSN: 1432-8798)

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas do Instituto de Biociências da Universidade Federal do Rio Grande do Sul.
Orientador: Prof. Dr. Paulo Michel Roehe
Coorientadores: Prof. Dr. Fabrício Souza Campos; Drª. Raíssa Nunes do Santos.

Porto Alegre
2022

CIP - Catalogação na Publicação

Timm, Francine Cezar Bandeira
Análise Metagenômica do Viroma de Tonsilas de Búfalos da Ilha de Marajó, Pará / Francine Cezar Bandeira Timm. -- 2022.
28 f.
Orientador: Paulo Michel Roehe.

Coorientadores: Fabrício Souza Campos, Raíssa Nunes dos Santos.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Bacharelado em Ciências Biológicas, Porto Alegre, BR-RS, 2022.

1. Metagenômica. 2. Viroma. 3. Búfalos. 4. Sequenciamento de Nova Geração. 5. Amplificação por Círculo Rolante. I. Roehe, Paulo Michel, orient. II. Campos, Fabrício Souza, coorient. III. dos Santos, Raíssa Nunes, coorient. IV. Título.

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

Agradecimentos

Aos meus orientadores.

A todos colegas e professores do Laboratório de Virologia do Instituto de Ciências Básicas da Saúde.

Aos pesquisadores Maria Audiléia da Silva Teixeira (ADEPARÁ), Gabriela Riet Correa Rivero (UFPA/Castanhal), Valíria Duarte Cerqueira (UFPA/Castanhal) e Pedro Soares Bezerra Júnior (UFPA/Castanhal) por terem enviado gentilmente o material a nós.

Aos pesquisadores do Centro de Desenvolvimento Científico e Tecnológico, Centro Estadual de Vigilância em Saúde da Secretaria Estadual de Saúde do Rio Grande do Sul.

E aos meus pais, irmã e marido pelo incondicional apoio.

FOLHA DE APROVAÇÃO

Francine Cezar Bandeira Timm

ANÁLISE METAGENÔMICA DO VIROMA DE TONSILAS DE BÚFALOS DA ILHA DE MARAJÓ, PARÁ

Trabalho de Conclusão de Curso apresentado com requisito parcial à obtenção do título de bacharel em Ciências Biológicas do Instituto de Biociências da Universidade Federal do Rio Grande do Sul.
Orientador: Prof. Dr. Paulo Michel Roehe
Coorientadores: Prof. Dr. Fabrício Souza Campos; Drª. Raíssa Nunes do Santos

Aprovado em: ___/___/___

_____
Prof. Dr. Paulo Michel Roehe

_____
Prof. Dr. Fabrício Souza Campos

_____
Drª. Raíssa Nunes dos Santos

_____
Profª. Drª. Ana Claúdia Franco

_____
Profª. Drª. Ana Paula Guedes Frazzon

Porto Alegre
2022

Resumo

Dado o potencial da criação de búfalos domésticos (*Bubalus bubalis*) e seu crescimento recente no Brasil, nosso grupo tem focado suas atenções no estudo do viroma de bubalinos da região norte do Brasil, região que possui o maior rebanho dessa espécie no País. Como parte desse projeto, no presente trabalho foi realizada uma análise metagenômica para avaliar a diversidade viral em tonsilas palatinas de búfalos domésticos de diferentes fazendas da ilha de Marajó, Pará. Fragmentos de tonsilas de 60 bubalinos foram selecionados aleatoriamente a partir de uma amostragem de 302 animais. Genomas virais foram purificados, extraídos e amplificados randomicamente com DNA polimerase do fago phi29. Após a amplificação, os produtos foram purificados e sequenciados através da plataforma Illumina. Após a montagem de *reads*, os *contigs* virais foram submetidos ao alinhamento usando a ferramenta BLAST+. Vírus de DNA circular foram predominantes no viroma, incluindo um genoma de poliomavírus bubalino não previamente reportado. Além disso, sequências de segmentos de genomas virais representativos de membros dos gêneros *Alphapolyomavirus*, *Gammaretrovirus*, *Gemykibivirus*, *Gemykrogvirus* e *Porprismacovirus* foram identificados, além de outros genomas de vírus circulares ainda não classificados. Estudos futuros serão realizados para aprimorar e aprofundar o conhecimento sobre o viroma e microbioma bubalino, permitindo um melhor entendimento sobre interações vírus-hospedeiro, bem como examinar o papel de vírus bubalinos como potenciais agentes de patologias em outras espécies, incluindo o homem.

Palavras-chave: Metagenômica. Viroma. *Bubalus bubalis bubalis*. Sequenciamento. Poliomavírus.

Abstract

Given the potential of raising domestic buffalo (*Bubalus bubalis bubalis*) and its recent growth in Brazil, our group has focused its attention on studying the bubaline virome in the northern region of Brazil, the area that has the largest herd of this species in the country. As part of this project, in the present work, we carried out a metagenomic analysis to evaluate the viral diversity in palatine tonsils of domestic buffaloes from different farms on Marajó Island, Pará. Tonsil fragments from 60 bubalines were randomly selected from a sample of 302 animals. Viral genomes were purified, extracted, and randomly amplified with phage phi29 DNA polymerase. After amplification, the products were purified and sequenced using the Illumina platform. After assembling the reads, viral contigs were subjected to alignment using the BLAST+ tool. Circular DNA viruses were predominant in the virome, including a previously unreported bubaline polyomavirus genome. In addition, sequences of representative viral genome segments from members of the genera *Alphapolyomavirus*, *Gammaretrovirus*, *Gemykibivirus*, *Gemykrogvirus*, and *Porprismacovirus* were identified, in addition to other yet unclassified circular virus genomes. Future studies will be conducted to enhance and deepen knowledge of the bubaline virome and microbiome, allowing for a better understanding of virus-host interactions, as well as examining the role of bubaline viruses as potential agents of pathologies in other species, including humans.

Keywords: Metagenomics. Virome. *Bubalus bubalis bubalis*. Sequencing. Polyomavirus.

SUMÁRIO

1. INTRODUÇÃO

A metagenômica é uma técnica que permite a análise da diversidade microbiana de determinado ecossistema [1]. Em virologia, tem sido usada para ampliar o conhecimento sobre viromas, isto é, o conjunto de vírus que se encontram em determinado nicho, órgão ou hospedeiro. Essa abordagem, é uma alternativa aos métodos clássicos, que envolvem o isolamento de vírus em animais de experimentação ou cultivos celulares [2], pois permite identificar vírus não cultiváveis, que de fato constituem a grande maioria dos vírus que compõem a virosfera [1], vírus estes sequer detectados, permanecendo desconhecidos. A complexidade desta tarefa se deve ao fato de que os vírus não apresentam um padrão genômico único comum [4]. Através da metagenômica, portanto, torna-se possível identificar genomas virais por meio de sequenciamento de nucleotídeos, sem a necessidade de conhecimento prévio de quaisquer regiões genômicas dos agentes investigados.

Em muitos casos, o número de cópias de genomas de determinados vírus presentes em uma amostra é muito pequeno para permitir análises subsequentes. Em vista disso, muitas vezes, antes da análise genômica, faz-se necessária uma etapa de enriquecimento do material genético contido na amostra. Nesse sentido, no laboratório de Virologia do ICBS/UFRGS (LABIVIR/UFRGS) tem sido utilizado um sistema de amplificação “a frio” denominado “amplificação por círculo rolante” (RCA). A técnica consiste na amplificação isotérmica de DNA ou RNA de fita simples utilizando a DNA polimerase do bacteriófago Φ29 (phi29) – que atua preferencialmente em DNA e RNA de fita simples – e primers randômicos promovendo um robusto enriquecimento do material genético presente na amostra [5]. No presente trabalho foram examinadas 60 amostras de tonsilas de bubalinos, como parte de uma amostragem maior (n=302), para contribuir com o conhecimento da diversidade viral presente nas amostras estudadas.

1.1 Rebanho Bubalino no Brasil

Os búfalos domésticos, também chamados de “búfalos de rio”, pertencem à família Bovidae, subfamília Bovinae, espécie *Bubalus bubalis* [6]. Estes animais são frequentemente criados em habitats pantanosos, mas possuem adaptabilidade a ambientes diversos. A sua elevada taxa de fertilidade e menor taxa de reposição, por exemplo, tornam a bubalinocultura uma atividade de menor custo de produção, quando comparada à bovinocultura [7]. Os animais são criados majoritariamente em pequenas e médias propriedades, em que a produção varia entre carne e leite [8]. Nutricionalmente, o leite das fêmeas bubalinas apresenta elevado teor nutricional quando comparado ao produzido pela fêmea bovina. A carne de búfalo contém baixo teor de gorduras, maior teor protético e elevado conteúdo de ômega 3/ omega 6 do que a de outras espécies utilizadas na pecuária [8].

O Brasil tem o maior rebanho bubalino entre as américas, onde a maior concentração de cabeças é estabelecida na Ilha de Marajó, no estado do Pará. A região marcou a entrada da primeira raça de búfalos no Brasil no final do século XIX. Entre os rebanhos bubalinos no país, quatro raças distintas são identificadas: Mediterrânea, Murrah, Jafarabadi e Carabao [9]. Búfalos do tipo baio (*Bubalus bubalis Var. bubalis*), cujas tonsilas foram coletadas para o estudo, não são reconhecidos como uma raça pela Associação Brasileira de Criadores de Bubalinos (ABCB) e apresentam, assim como a raça Carabao, risco de extinção e descaracterização [10]. Os bubalinos, em geral, são de fácil manejo e reconhecidamente resistentes a doenças [11]. No entanto, muito pouco se sabe sobre a sua sensibilidade a patógenos virais, que oferecem potencial risco a rebanhos inteiros e, consequentemente, prejuízos financeiros e produtivos aos produtores. Além disso podem ocasionalmente servir de fonte de infecção para outras espécies animais, incluindo bovinos [12]. Tendo em vista o potencial da bubalinocultura na indústria agropecuária e a expansão destas criações em todo o território brasileiro, um estudo sobre o viroma de bubalinos é de grande relevância econômica e sanitária.

No presente estudo, a técnica de amplificação por círculo rolante (RCA) foi empregada para amplificar os genomas de DNA e RNA nas amostras de tecido tonsilar de búfalos [13]. O ácido nucleico foi submetido ao enriquecimento e posteriormente

purificado para o sequenciamento genômico, sendo os resultados analisados e comparados com sequências disponíveis em bancos de dados públicos de genomas virais. Todos os genomas obtidos serão depositados no banco de dados GenBank.

2. OBJETIVOS

2.1 Objetivo geral

Contribuir para o conhecimento da diversidade do viroma bubalino.

2.2 Objetivos específicos

a. Identificar genomas virais presentes nas amostras de tonsilas examinadas.
b. Contribuir para a prevenção e profilaxia de doenças virais na bubalinocultura.

3. METODOLOGIA

3.1 Coleta de Amostras

Foram coletadas tonsilas palatinas de 302 bubalinos enviados para o abate. As coletas de amostras ocorreram no ano de 2020, em visitas realizadas a um frigorífico localizado na cidade de Belém, estado do Pará (PA), Brasil. Os búfalos eram procedentes de fazendas da Ilha de Marajó, PA. Após coleta, as tonsilas foram acondicionadas sob refrigeração e enviadas ao laboratório de Virologia para análise do viroma. Destas, 60 amostras foram aleatoriamente selecionadas para a execução do presente trabalho.

3.2 Processamento de Amostras

As amostras de tonsilas armazenadas a -80ºC foram descongeladas e fragmentos de 4 mg de tecido foram separados para confecção de pools. Seis pools contendo 10 fragmentos de tecido de cada amostra foram preparados. As amostras de tonsilas foram fragmentadas com o auxílio de lâminas de bisturi estéreis e acondicionados em tubos de ensaio de 15 mL. Os fragmentos foram homogeneizados em 8 mL de PBS com 10 esferas de vidro (2 mm de diâmetro) esterilizadas. Após

agitação por 1 minuto, os tecidos foram centrifugados por 5 minutos a 5000 x g, sendo 5 mL de sobrenadante coletados para posterior filtração (0,22 uM). O produto filtrado foi ultracentrifugado por 3 horas a 29 000 x g a 4ºC. O pellet assim obtido foi diluído em 500 µL de água ultrapura. Após processamento, as amostras foram tratadas como descrito a seguir.

3.3 Extração de ácidos nucleicos

Uma alíquota de 350 µL de cada pool foi transferida para um microtubo contendo 40 µL de tampão de DNAse, 2,5 µL de DNAse e 5 µL de RNAse [13]. A amostra foi incubada a 37ºC por uma hora. Após, 12,6 µL de EDTA 0,5 mM e 7,4 µL de água ultrapura foram adicionados às amostras e incubadas a 75ºC por 10 minutos. Para obtenção de DNA e RNA viral foram utilizadas beads magnéticas MagMAXTM CORE (Thermo Fischer Scientific®) conforme protocolo do fabricante. Para a obtenção do cDNA, foi utilizado o kit High Capacity cDNA (Thermo Fischer Scientific®).

3.4 Amplificação por Círculo Rolante

O enriquecimento do genoma viral foi realizado através da amplificação por círculo rolante, ou "rolling circle amplification" (RCA) [5]. Essa técnica utiliza primers randômicos e DNAse do fago phi29 a uma temperatura constante (30ºC por 12 horas) para amplificação de genomas. Para a reação, inicialmente foi realizada a desnaturação do DNA utilizando 0,5 µL de Reaction Buffer 10X, 100 µM n-hexamer primer, água livre de nuclease e o DNA extraído das amostras, submetidas a 95ºC por 3 minutos e posterior resfriamento em gelo de 3 a 5 minutos. Para a amplificação, 5 µL de DNA desnaturado foram adicionados a 1,5 µL de Reaction Buffer 10X, 2 µL de 10mM dNTP Mix, 1 µL de 10U phi29 polimerase e 10,5 µL de água livre de nucleases sob as seguintes condições: 36ºC por 18 horas e 65ºC por 10 minutos para inativação de DNAses. Os fragmentos amplificados foram visualizados por eletroforese, purificados, quantificados, enviados para montagem da biblioteca e sequenciamento no Centro de Desenvolvimento Científico e Tecnológico (CDCT) no Centro Estadual de Vigilância em Saúde (CEVS) da Secretaria Estadual de Saúde do Rio Grande do Sul (SESRS).

3.5 Biblioteca e Sequenciamento

Para a preparação da biblioteca genômica, foi utilizado o protocolo *Illumina Library Prep* (Illumina®), de acordo com o fluxo de trabalho da plataforma de sequenciamento de DNA *MiSeq*. Para a construção da biblioteca são necessários fragmentos de DNA em tamanho adequado (200 pb). Para tanto, durante o processo de tagmentação ocorre fragmentação, e normalização do tamanho do fragmento, bem como a ligação dos adaptadores e marcadores de amostra, permitindo que várias amostras sejam sequenciadas em um só *pool*. Durante o sequenciamento, o equipamento é capaz de identificá-las de acordo com seu marcador específico, facilitando a leitura e geração de *reads* para posterior montagem e análise dos dados.

3.6 Montagem dos genomas

Para montagem *de novo*, foram utilizadas as ferramentas *SPAdes 3.10 Genome Assembler* (http://bioinf.spbau.ru/spades) e *Genome Detective Virus Tool* (https://www.genomedetective.com/) para obter de forma rápida e acurada informações acerca dos dados de sequência. A metodologia de montagem por referência (“map to reference”), foi feita através do *Geneious v9.1.3* (https://www.geneious.com/). O *software BLAST+* (https://blast.ncbi.nlm.nih.gov/Blast.cgi) alinhou os *contigs* gerados na montagem com sequências de referência disponíveis em bancos de dados como *UniprotKB* e o banco de dados viral do *NCBI*. O perfil taxonômico em gráfico interativo foi plotado através dos *softwares Kaiju* (https://kaiju.binf.ku.dk) e *KRONA* (https://hpc.nih.gov/apps/kronatools.html).

3.7 Análise filogenética

O alinhamento de sequências conservadas - aminoácidos ou nucleotídeos - de genes/proteínas virais foi realizado utilizando *MAFFT* (https://mafft.cbrc.jp/alignment/software/). Sequências adicionais foram recuperadas do *GenBank* (https://www.ncbi.nlm.nih.gov/genbank/) após pesquisa inicial através do *BLAST+*. Análises de máxima verosimilhança em árvore, possibilitando uma inferência

filogenética foi realizada com o auxílio do software *MEGA v7* (https://www.megasoftware.net/). Foram atribuídos os alinhamentos a modelos evolutivos independentes de acordo com os critérios do ICTV (https://icetv.global/).

4. TRABALHO EXPERIMENTAL EM FORMATO DE ARTIGO CIENTÍFICO

Este manuscrito foi escrito conforme as normas da revista Archives of Virology (ISSN: 1432-8798)

Viral metagenomics of bubaline tonsils

Francine C. B. Timm[1], Raíssa Nunes dos Santos[3], Fabrício Souza Campos[1] Bruna Paredes Galarza[1], Nicole Vieira Stone[1], Gisele Barcelos Seberino[1], Lina Marcela Violet Lozano[1], Fernanda Marques de Souza Godinho[2], Richard Steiner Salvato[2], Ana Cláudia Franco[1], Maria Audiléia da Silva Teixeira[4] , Gabriela Riet-Correa Rivero[5], Valíria Duarte Cerqueira[5] Pedro Soares Bezerra Júnior[5], Paulo Michel Roehe[1].
f-timm@outlook.com

[1] Laboratório de Virologia, Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, 90050-170, Brasil.
[2] Centro de Desenvolvimento Científico e Tecnológico (CDCT), Centro Estadual de Vigilância em Saúde (CEVS) da Secretaria Estadual da Saúde do Rio Grande do Sul (SESRS), Porto Alegre, Rio Grande do Sul, 90450-190, Brasil .
[3] Laboratório de Bioinformática e Biotecnologia, Campus Gurupi, Universidade Federal do Tocantins, Gurupi, Tocantins, 77410-570, Brasil.
[4] Agência Estadual de Defesa Agropecuária do Estado do Pará (ADEPARÁ), Belém, Pará, 66085-023, Brasil.
[5] Laboratório de Patologia Animal, Instituto de Medicina Veterinária, Universidade Federal do Pará (UFPA), Castanhal, Pará, 68741-071, Brasil.

Abstract Water buffalo (Bubalus bubalis) farming is increasing in many regions of the world due to the species' ability to multiply in environments where bovine cattle would struggle to thrive. Despite water buffaloes being known for their resistance to diseases, data about the diversity of the microbiome of water buffaloes are still scarce. In this study, we examined the virome diversity in palatine tonsils collected from animals from the island of Marajó, in northern Pará state, Brazil, where the largest bubaline flock in the country is farmed. Tonsil fragments from 60 bubalines were randomly selected from a sample of 302 animals. Viral genomes were purified, extracted, and randomly amplified with phage phi29 DNA polymerase. After amplification, the products were purified and sequenced using the Illumina platform. After assembling the reads, viral contigs were subjected to alignment using the BLAST+ tool. Circular DNA viruses were predominant in the virome, including a previously unreported bubaline polyomavirus genome. In addition, sequences of representative viral genome segments from members of

the genera *Alphapolyomavirus*, *Gammaretrovirus*, *Gemykibivirus*, *Gemykrogvirus*, and *Porprismacovirus* were identified, in addition to other yet unclassified circular virus genomes.

Introduction The domestic water buffalo (*Bubalus bubalis*) belongs to the Bovidae family and Bovinae subfamily, which includes cattle, bison, African buffalo, and water buffalo [1]. Native to Asia, the domestic buffalo uses are predominantly for milk and meat production. For animal husbandry, its high fertility and lower replacement rate reduce production costs compared to cattle [2]. Nutritionally, the milk from buffalo females has a high nutritional content when compared to that produced by bovine females. Buffalo meat has low-fat content, higher protein content, and high omega 3/omega 6 content than other species used in livestock farming [3]. Tonsils are a pair of lymphoid tissue mass located in the lateral wall of the oropharynx and play an important role in immunological functions. As a rote of entry and replications site for many pathogens, tonsils appear as a potential assay for pathogen diagnosis [4]. Anatomically, tonsils have an elongated kidney shape with a central tonsilar fossa, leading to a central opening where mucous secretors are based [5]. Histologically, this organized mucosa-associated lymphoid mass (MALT) presents dendritic cells within germinal centers. Dendritic cells are antigen-present cells that produce type-I interferons (IFN-I) during viral infections and actively stimulate natural killer cells (NK) [6], cytotoxic lymphocytes that exhibit a diversity of anti-viral functions [7]. The characterization of the virome tonsils has been studied in previous works. A study [9] suggests that a combination of stress and changes in nutrition between newborn piglets can influence the composition of tonsilar microbiota. Also, a viral community might elucidate the etiology behind diseases and subclinical infections, as demonstrated in tonsilar assay research [9]. Despite current research focused on identifying the diversity of microbiome in such farm animals has been presented, the absence of studies on viral composition in buffalo's tonsils is a gap in our knowledge. The present research aims to identify and characterize viral diversity in bubaline tonsils. The sampling included animals from two different farms located in northern Brazil, which are pooled for high-throughput sequencing (HTS). To investigate the biological significance of viral diversity in the palatine tonsils of buffaloes, this study will provide results for further research and epidemiological issues.

Material and Methods

Animal sampling and ethics Sampling of tonsils was performed in 2020 in Marajó (0º 58’ S 49º 34’ W), an island in the state of Pará, Brazil **Fig 1** (see appendix A). Marajó island lies in the estuary where the Amazon river meets the Atlantic Ocean. It is the largest river island in northern Brazil and the world’s largest fluvial island. Tissue samples

of palatine tonsils of buffaloes (n=300) were collected in a slaughterhouse. The animals were from eight different farms on the island and were identified as *Bubalus bubalis* var. *bubalis*. However, the present study was concentrated on two different farms located in Ponte de Pedras (1º 23' 26'' S 48º 52' 13'' W) and Cachoeira do Arari (1º 0' 16'' S 48º 57' 27'' W), both in Marajó. The tonsils were collected individually and stored in sterile tubes kept on ice and transferred to -80ºC at the arrival at the laboratory.

Viral concentration, nucleic acid extraction, and library preparation Samples were thawed and gently vortexed for 2 minutes while still in the collection tubes. Subsequently, 4 mg of tissues from each of the 10 samples were pooled and mixed with 8 mL of PBS 1X in a plastic conic flask with 10 glass microspheres (diameter 2 mm). The tubes were then vortexed for 2 minutes and clarified by centrifugation for 5 min at 5000 × g. Next, each pool (n=6 pools of 10 samples) was filtered through a 0.22 μm *Millipore syringe filter*®. The filtered suspensions were submitted to ultracentrifugation at 33000 × g for 3 hours at 4 °C on a 25% sucrose cushion. The viral pellet was resuspended in 500 μL of ultrapure water, and treated with 100 U of *recombinant DNAse I* (Roche®), 100 U of *PureLink RNAse A* (Thermo Fisher Scientific®), and submitted to nucleic acid extraction *MagMaxx CORE* (Thermo Fisher Scientific®). The quality of DNA and RNA extraction was checked in a 0.7% agarose gel, and the amounts of both nucleic acids obtained were measured using *Qubit*®. The extracted DNA was randomly amplified with 10 U/ μL of *phi29 DNA polymerase* (New BioEnglands). RNA was used for cDNA synthesis whit the *High Capacity cDNA kit* (Thermo Fisher Scientific®). According to the manufacturer's instructions, the cDNA was submitted to phi 29 DNA polymerase amplification. The quality of the final amplification products was checked in a 0.8% agarose gel and submitted to *AMPure XP Magnetic Bead Purification* (Thermo Fisher Scientific®). After purification, the DNA was quantified and a library was constructed using the *Illumina Library Prep* protocol. The sequences were generated in an *Illumina*® *MiSeq* sequencer (MiSeq Reagent kit V3).

Assembly of viral genomes Reads were de novo assembled using *SPAdes 3.10 assembler* and *Genome Detective*. The contigs generated were blasted with *BLAST+* against the UniprotKB database, as well as the Viral database from NCBI. Sequences were manually reviewed, using a cutoff e-value of 0.001.

Taxonomy profile The interactive metagenomics visualization was applied using *Kaiju* and *KRONA* tools to generate taxonomy profile graphs. We used the *fastq.gz* reads as inputs to produce a read-based estimate of taxonomical groups and compare it with our viral analysis.

Phylogenetic analysis Sequence alignment of amino acid or nucleotide sequences of the conserved viral genes/proteins was carried out using *MAFFT*. Additional sequences were retrieved from *GenBank* after the initial Blast search. Maximum likelihood tree analyses were carried out using *MEGA.7*®. Alignments were assigned to independent evolutionary models as shown in each subsequent phylogenetic tree according to ICTV criteria (https://ictv.global/).

Results A total of 4,013,794 raw reads were generated. After the *Kaiju* check, 1,133,776 reads (~28%) were classified. About 4% of the reads represented viral sequences **Fig 2** (see appendix A).

Assembly of viral genomes NGS of the 60 buffaloes tonsils samples resulted in a total of 4,013,794 reads, of which 123,928 (3%) did not pass in Quality Control (QC) reports and were removed. A total of 1,133,776 were further classified. The sequence lengths were between 35 to 151 bp using *SPAdes* of total reads. About 40% of reads were assembled into viral contigs. Several 11 shown in **Table 1** (see appendix B) viral consensuses were identified with high NT identity and/or high genome coverage. All of the recovered genomes could be assigned to five genera: *Alphapolyomavirus*, *Gammaretrovirus*, *Gemykibivirus*, *Gemykrogvirus*, and *Porprismacovirus*, in addition to other yet unclassified circular virus genomes.

Identification of a novel virus of the genus Polyomavirus *Polyomavirus* belongs to the *Polyomaviridae* family which includes six genera: *Alpha-*, *Beta-*, *Delta-*, *Gamma-*, *Epsilon-*, and *Zetapolyomavirus*. Members of this family are non-enveloped viruses with a circular, double-stranded DNA (dsDNA) genome. The natural hosts of members of these genera are mammals, birds, and, as recently reported, fish and arthropods [10]. Each family member has a restricted host range and some members are known to cause symptomatic infection or cancers in their host [11]. We recovered two contigs encompassing coding regions of the viral Large Tumor antigen (Large T-ag), required by a correct phylogenetic analysis **Fig 3** (see appendix A). The contigs obtained comprised 67,112 reads, covering > 98% of the genome. In a preliminary review, such polyomavirus seems genetically more closely related to, but distinct from, *Sus scrofa polyomavirus*. Based on this finding, we propose that this novel poliomavírus be called *Bubaline Alphapolyomavirus 1* **Fig 4** (see appendix A).

Retroviridae The *Retroviridae* family comprises two subfamilies and eleven genera [12]. Members of this family have enveloped, and spherical, with 80 to 100 nm in virion diameter. Retroviral genomes are composed of diploid, single-strand, positive-sense RNA (ssRNA+), whose replication requires reverse transcription, which is performed by the viral reverse transcriptase. In this study, one contig was recovered that spans 2,128 NT, with 11,071 reads.

Such contig spans nearly 24,5% of the vírus genome. The recovered sequence starts at genome position 2323 and finishes at position 4360 relative to the NC_001501.1 reference *Moloney murine leukemia virus* genome. Although the coverage was somewhat lower than expected, the coding frame for *gag-pol-env* [13], the main region used for phylogenetic inferences, was included in the sequenced region here recovered.

Genomoviridae Members of the *Genomoviridae* family are non-enveloped viruses of about 1,8 -2,4 Kb long of circular ssDNA. The genome encodes two proteins the capsid protein (CP) and the rolling-circle replication initiation protein (Rep) [14]. The CP does not appear to be similar to those of other viruses, whereas *Rep* shows homology to other ssDNA viruses and is similar to that of plant viruses of the *Geminiviridae* Family [14]. Has been concerned in previous studies [15] that 78% of genome-wide pairwise identity is a species demarcation threshold. Here, were found two different genera, and contigs were obtained with NT identity between 64%-98% approximately **Table 1** (see appendix B). The species *Gemykibivirus humas2* (human-associated gemykibivirus 2), whose coverage is defined as 100% in our findings, was predicted in a previous study [16] to have a low zoonotic potential.

Smacoviridae Viruses belonging to the *Smacoviridae* family have a small (2.3-3 Kb), circular ssDNA genome, which encodes proteins *Rep* and *CP* [17]. Twelve smacovirus genera have been recognized so far. The members of the *Smacoviridae* **Fig 5** (see appendix A) have been found predominantly in the fecal matter of an extended host range of vertebrates, including humans. However, previous studies revealed that such genomes could be identified in sérum, tracheal swab samples [18], and insects [19]. In this study, sequencing analyses found two contigs that covered conserved genomic regions with similarity to members of the *Porprismacovirus* genus (name derived from Porcine and Primate Smacovirus, hosts for the virus) were identified.

Unclassified viruses According to ICTV [20], a considerable number of well-characterized viruses remain unclassified. Apart from the findings reported above, two other sequences were identified that present a high degree of NT identity. Bovine serum-associated circular virus and Sewage-associated circular DNA virus-18. Here, such genomes were recovered with between 79-96% coverage, respectively. Sewage-associated circular DNA virus-18 is an ssDNA genome that has been identified in both terrestrial and aquatic environments. Sewage consists of wastewater generated by human excrement, mainly, and substantial studies that evolved sewage viruses of clinical importance to humans, such as poliovirus have been conducted [21]. Metagenomics has been widely employed to identify the viral diversity associated with sewage in recent years. Bovine serum-associated circular virus, a circular

Rep-encoding single-stranded DNA (CRESS DNA), was previously recovered in the serum of calves persistently infected with Pestivirus A, which suggests support viremia of CRESS DNA genome [22]. As reported in recent research, several CRESS DNA viral genomes were present in the sérum of commercial pigs from southern Brazil, including nine unclassified genomes [23]. A number of other genome sequences, probably of viral origin, were identified in this study. However, the reads obtained were too short to warrant subsequent assemblies with reasonable statistical support and are not presented here.

Discussion A growing number of previous works based on metagenomics have been used to unveil the wide diversity of viruses in farm animals. These have allowed the discovery of novel viruses. Meanwhile, the methods have been optimized and applied for further deepening of knowledge in the field. Given the growing stimuli to water buffalo farming in many countries, including Brasil, we were requested by colleagues from the state of Pará to investigate viral pathogens that might be associated with disease in water buffaloes, since these may become causes of important economic losses. In this study, we focused on the analyses of the viral diversity in palatine tonsils of water buffaloes of clinically healthy animals. Samples were collected in abbatoirs. We have been able to identify viral genomes corresponding to members of, at least, four distinct viral families, including a novel polyomavirus for which we propose the name *Bubaline Alphapolyomavirus* 1. This novel virus has a close genetic relationship with *Sus scrofa polyomavirus* 1, which has the Suidae Family as hosts. It is also related to, but phylogenetically distinct from bovine polyomavirus **Fig 3** (see appendix A). Our findings differ from common viremic findings in farm animals, which often include genomes of members of *Picornaviridae*, *Parvoviridae*, and *Astroviridae*. Such differences might be related to details of the methodology applied here and to the sampling of tonsils only, the first organs to be examined in a wider range of studies involving the water buffalo virome, to be continued. Thus, further studies shall be conducted to widen the observations of the virome of water buffaloes. Metagenomics has an important role in broadening our understanding of the virosphere and vírus host interactions. However, the sizes of reads and other limitations associated with this methodology must be overcome to allow even more sensitive and precise results. Future work shall focus on the retrieval of complete viral genome sequences and searches in other host tissues. The results presented here are expected to encourage research on the virome of bubaline.

Conclusion Metagenomics allowed identifications of at least part of the virome in palatine tonsils of water buffaloes (*Bubalus bubalis* var. *bubalis*). We identified four viral families Poliomaviridae, Retroviridae, Genomoviridae, and Smacoviridae unveiling important information about the previously unexplored virome of bubaline tonsils. About

28% of the reads were classified by Kaiju including bacteria and cellular fragments. These data reveal the sensibility of HTS and the difficulties of removing non-viral particles during sample processing. Further studies should be performed to improve and deepen our knowledge of the *Bubalus bubalis* virome and microbiome, allowing a better understanding of vírus-host interactions, as well as to examine the role of viruses of bubaline as potential agents of disease in other species, including man.

References

1. ITIS - Report: *Bubalus bubalis*. https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=552476#null.

2. Soroprevalência de estomatite vesicular, diarreia viral bovina e infecção por herpesvírus bovino tipo 1 em búfalos (*Bubalus bubalis*) no Distrito Federal | VETTESES. https://pesquisa.bvsalud.org/portal/resource/pt/vtt-220192. Accessed 25 Jun 2022

3. Bernardes O (2007) Bubalinocultura no Brasil: situação e importância econômica. Rev Bras Reprod Anim, Belo Horizonte, 31:.293-298.http://www.cbra.org.br/pages/publicacoes/rbra/download/293.pdf

4. Liebler-Tenorio EM, Pabst R (2006) MALT structure and function in farm animals. Vet Res 37:257–280. https://doi.org/10.1051/vetres:2006001

5. Zidan M, Pabst R (2011) The microanatomy of the palatine tonsils of the buffalo (*Bos bubalus*). Vet Immunol Immunopathol 139:83–89. https://doi.org/10.1016/J.VETIMM.2010.08.006

6. Polak ME, Borthwick NJ, Gabriel FG, et al (2008) Activation of tonsil dendritic cells with immuno-adjuvants. BMC Immunol 9:1–12. https://doi.org/10.1186/1471-2172-9-10/FIGURES/5

7. Frank K, Paust S (2020) Dynamic Natural Killer Cell and T Cell Responses to Influenza Infection. Front Cell Infect Microbiol 10:425. https://doi.org/10.3389/FCIMB.2020.00425

8. Pena Cortes LC, Leveque RM, Funk J, et al (2018) Development of the tonsillar microbiome in pigs from newborn through weaning. BMC Microbiol 18: https://doi.org/10.1186/S12866-018-1176-X

9. Blomström AL, Ye X, Fossum C, et al (2018) Characterisation of the Virome of Tonsils from Conventional Pigs and from Specific Pathogen-Free Pigs. Viruses 10. https://doi.org/10.3390/V10070382

10. Buck CB, van Doorslaer K, Peretti A, et al (2016) The Ancient Evolutionary History of Polyomaviruses. PLoS Pathog 12:1005574. https://doi.org/10.1371/JOURNAL.PPAT.1005574

11. Moens U, Calvignac-Spencer S, Lauber C, et al (2017) ICTV virus taxonomy profile: Polyomaviridae. Journal of General Virology 98:1159–1160. https://doi.org/10.1099/JGV.0.000839

12. Family: Retroviridae | ICTV. https://ictv.global/report/chapter/unclassified/unclassified-viruses. Accessed 22 Sep 2022

13. Genus: Gammaretrovirus | ICTV. https://ictv.global/report/chapter/retroviridae/retroviridae/gammaretrovirus. Accessed 22 Sep 2022

14. Varsani A, Krupovic M (2021) Family Genomoviridae: 2021 taxonomy update. Arch Virol 166:2911–2926. https://doi.org/10.1007/S00705-021-05183-Y/FIGURES/4

15. Varsani A, Krupovic M (2017) Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. Virus Evol 3: https://doi.org/10.1093/VE/VEW037

16. Mollentze N, Babayan SA, Streicker DG (2021) Identifying and prioritizing potential human-infecting viruses from their genome sequences. PLoS Biol 19: https://doi.org/10.1371/JOURNAL.PBIO.3001390

17. Krupovic M, Varsani A (2021) A 2021 taxonomy update for the family Smacoviridae. Arch Virol 166:3245–3253. https://doi.org/10.1007/S00705-021-05224-6/FIGURES/3

18. Wang H, Li S, Mahmood A, et al (2018) Plasma virome of cattle from forest region revealed diverse small circular ssDNA viral genomes. Virol J 15:1–8. https://doi.org/10.1186/S12985-018-0923-9/FIGURES/3

19. Rosario K, Mettel KA, Benner BE, et al (2018) Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates. PeerJ 2018:e5761. https://doi.org/10.7717/PEERJ.5761/SUPP-3

20. Unclassified Viruses | ICTV. https://ictv.global/report/chapter/unclassified/unclassified-viruses. Accessed 22 Sep 2022

21. Kraberger S, Argüello-Astorga GR, Greenfield LG, et al (2015) Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. Infection, Genetics, and Evolution 31:73–86. https://doi.org/10.1016/J.MEEGID.2015.01.001

22. Weber MN, Cibulski SP, Silveira S, et al (2018) Evaluation of the serum virome in calves persistently infected with Pestivirus A, presenting or not presenting mucosal disease. Virus Genes 54:768–778. https://doi.org/10.1007/S11262-018-1599-3

23. Tochetto C, Cibulski SP, Varela APM, et al (2021) A variety of highly divergent eukaryotic ssDNA viruses in sera of pigs. Journal of General Virology 102:001706. https://doi.org/10.1099/JGV.0.001706/CITE/REFWORKS

24. dos Santos RN, Campos FS, Finoketti F, et al (2020) Viral diversity in oral cavity from Sapajus nigritus by metagenomic analyses. Brazilian Journal of Microbiology 51:1941. https://doi.org/10.1007/S42770-020-00350-W

APPENDIX A

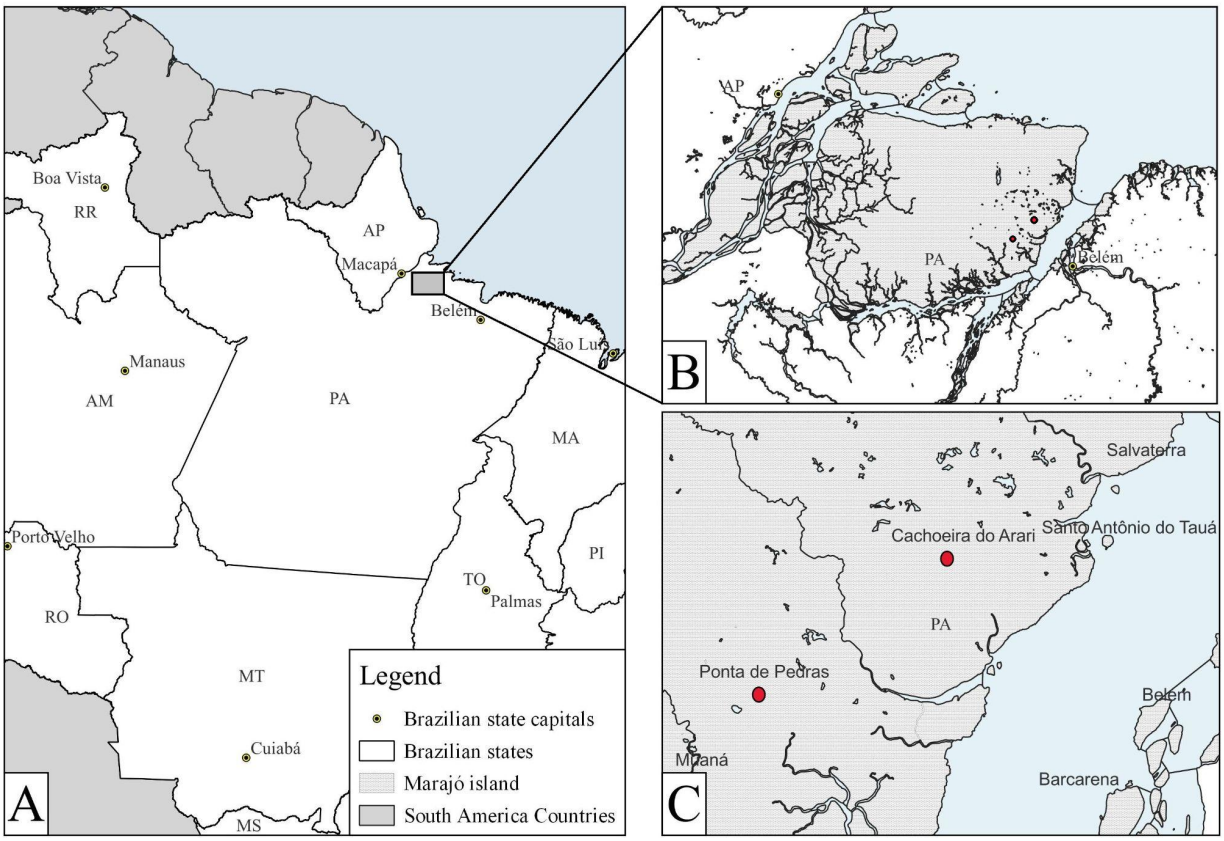

Fig 1 Map of sample collection. (**A**) Northern Brazil. (**B**) Marajó island, state of Pará. (**C**) Municipalities of Cachoeira do Arari and Ponta de Pedras [red dots], the region where the buffaloes were raised.

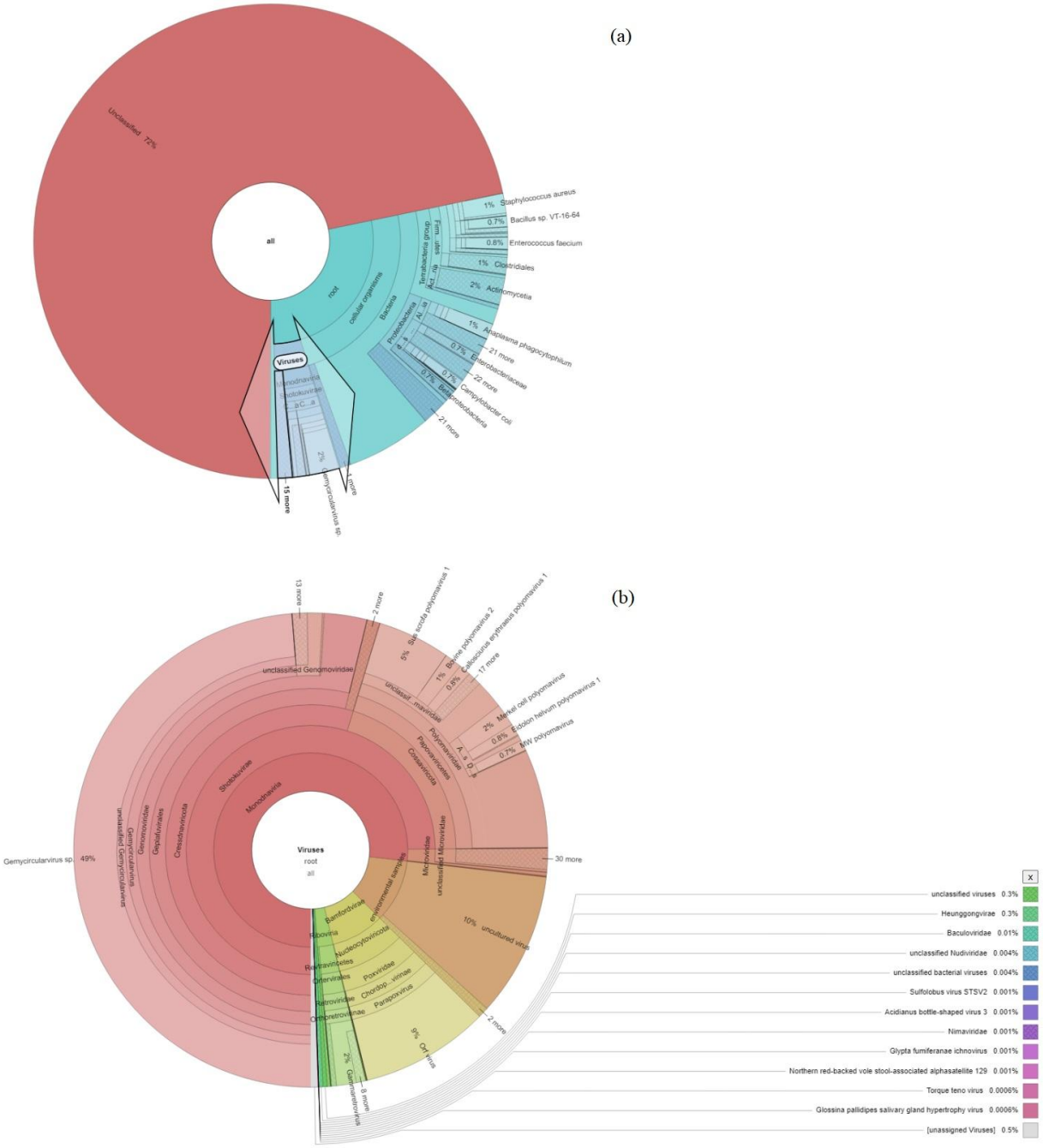

Fig 2 The KRONA display. (**a**) KRONA visualization of total reads; (**b**) Viruses classification.

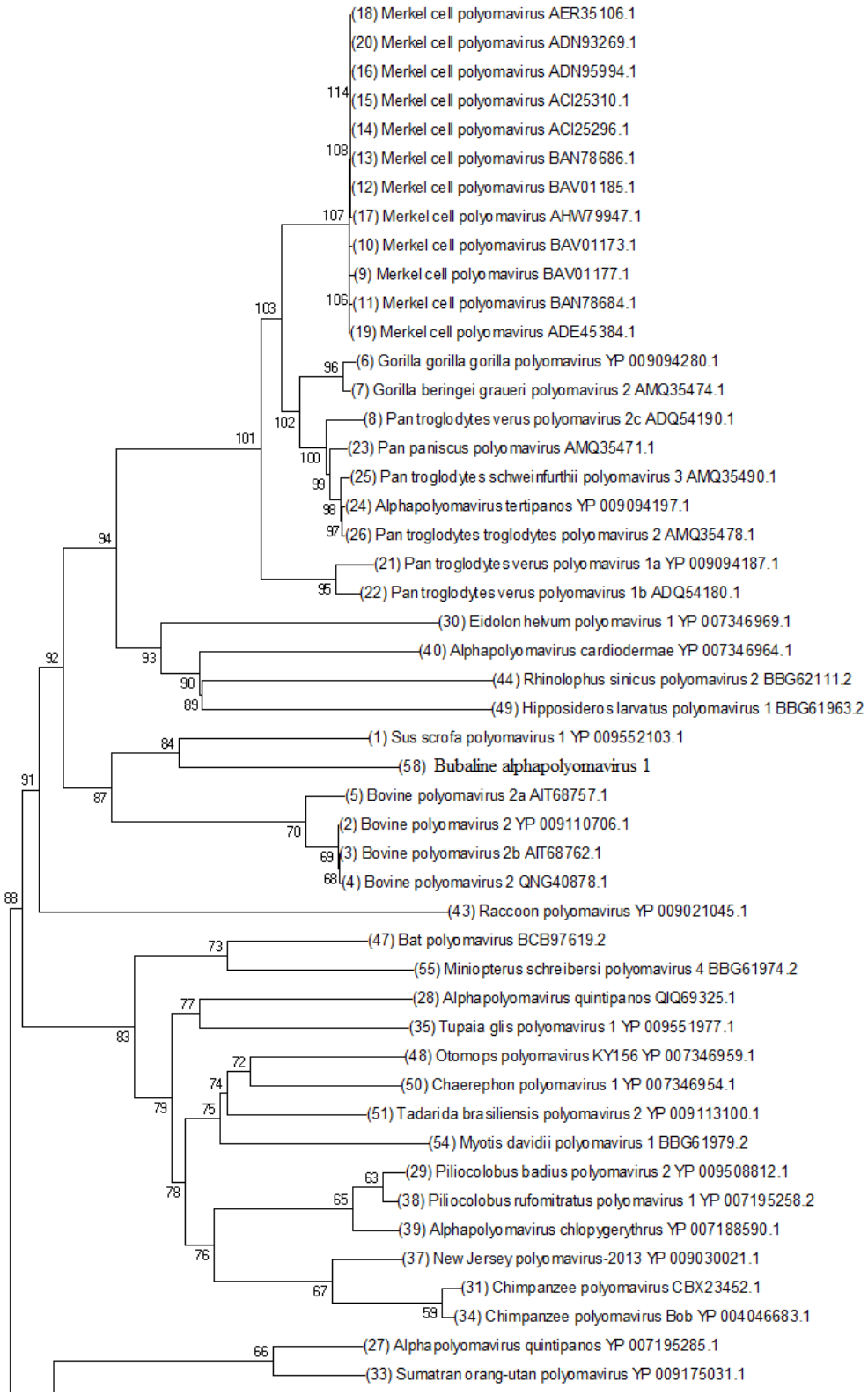

Fig 3 Phylogenetic three of some members of the Polyomaviridae family. The sequence recovered in the present study is numbered as (**58**). The analysis involved 58 amino acid sequences. All ambiguous positions were removed for each sequence pair. A total of 1413 positions were in the final dataset. Evolutionary analyses were conducted in MEGA.7.

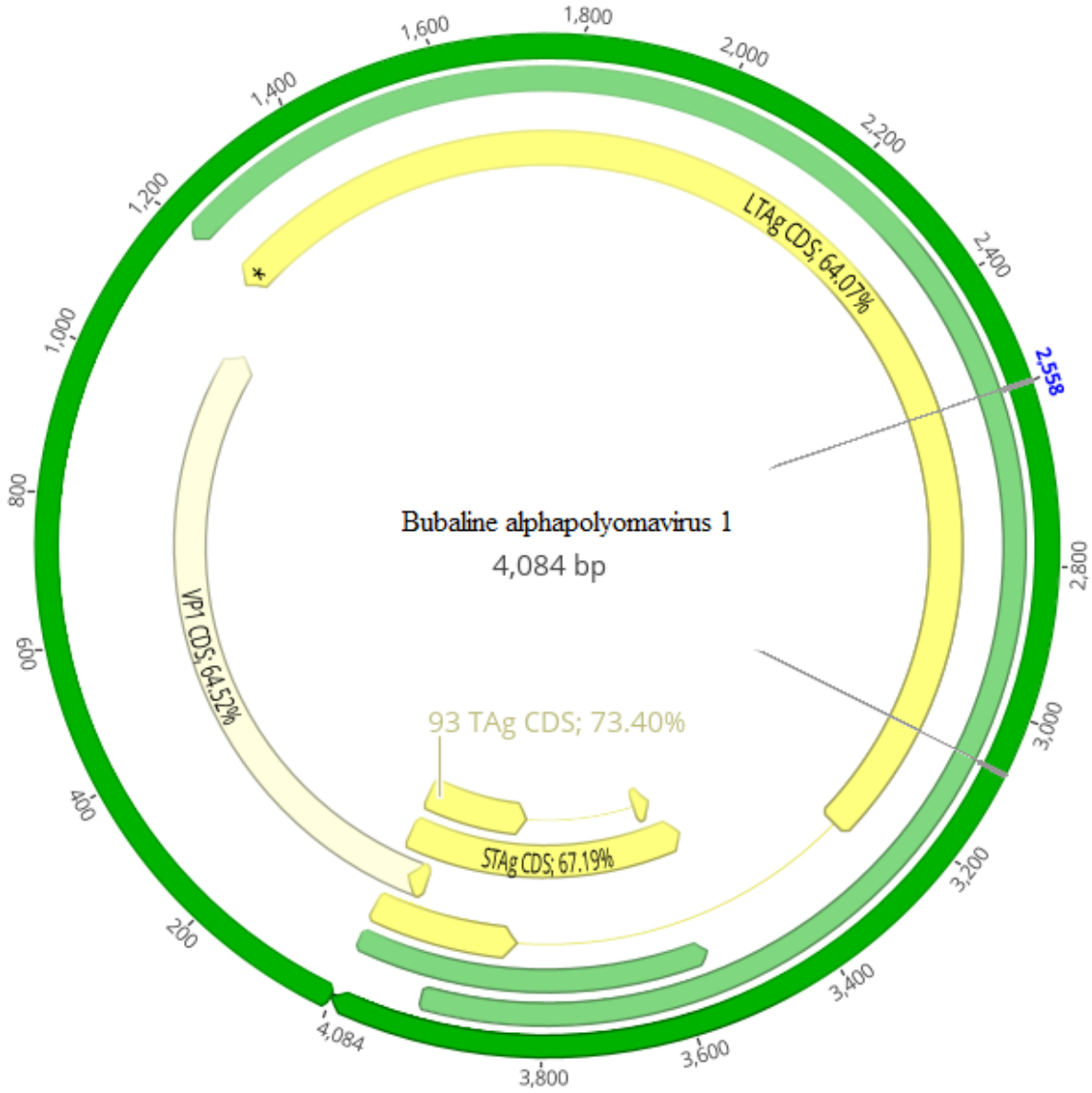

Fig 4 The complete genome of Bubaline alphapolyomavirus. At least, 4,084 base pairs were recovered in this study.

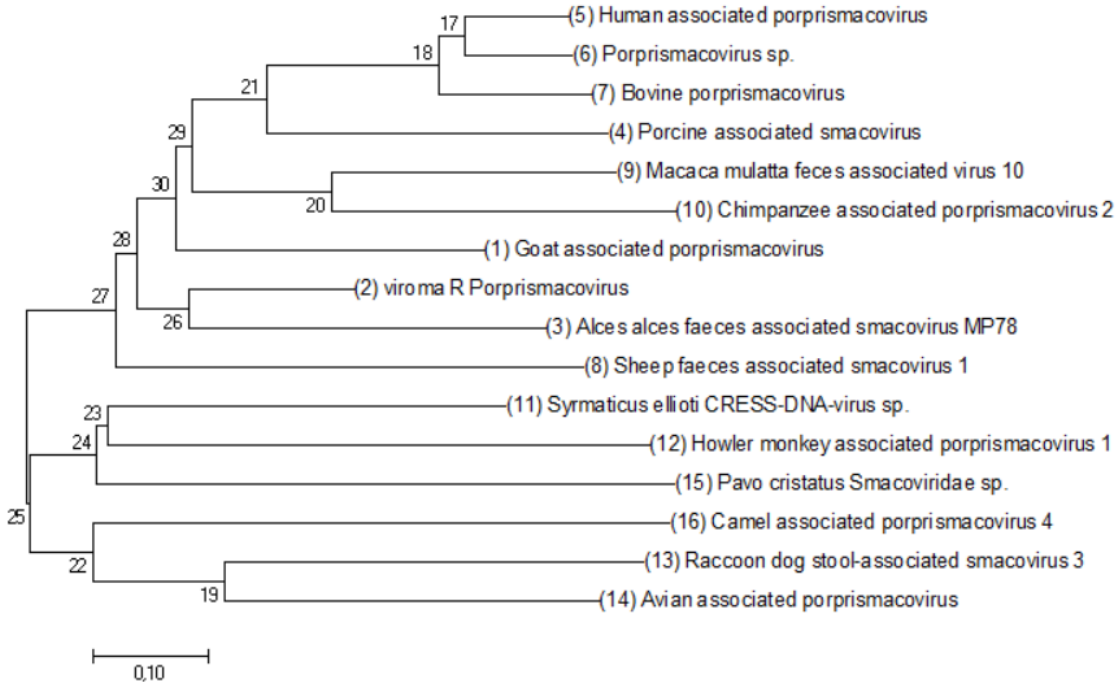

Fig 5 Phylogenetic tree of some members of the Smacoviridae family. The analysis involved 16 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 721 positions in the final dataset. Evolutionary analyses were conducted in MEGA.7.

APPENDIX B

Table 1 Most likely assignment of viral genome sequences recovered from bubaline tonsils.

Assignment	# Contigs	# Reads	Coverage (%)	Depth of Coverage	NT Identity (%)	AA Identity (%)
Sus scrofa polyomavirus 1	**2**	**67112**	**98,4215**	**1650.4**	**63,78**	**65,70**
Moloney murine leukemia virus	**1**	**11071**	**24,4599**	**695.3**	**76,10**	**99,70**
Gemykrogvirus	**2**	**4220**	**56,3229**	**445.3**	**84,74**	**82,90**
Gemykibivirus humas2	**1**	**1560**	**100**	**91.4**	**81,83**	**89,85**
Sewage-associated circular DNA virus-18	**2**	**909**	**96,0699**	**53.3**	**64,46**	**62,66**
Gemykibivirus blaro1	**1**	**131**	**82,1987**	**9.5**	**63,89**	**68,31**
Porprismacovirus	**2**	**112**	**40,0264**	**11.6**	**58,63**	**58,20**
Gemykibivirus humas4	**1**	**69**	**21,2988**	**18.8**	**81,02**	**70,44**
Gemykibivirus humas2	**1**	**60**	**65,075**	**5.4**	**97,69**	**98,89**
Gemykibivirus humas2	**2**	**47**	**12,2783**	**22.3**	**89,25**	**90,47**
Gemykrogvirus	**1**	**42**	**29,1843**	**8.7**	**88,69**	**83,98**
Bovine serum-associated circular virus	**1**	**35**	**79,0927**	**12.4**	**58,44**	**48,33**

REFERÊNCIA BIBLIOGRÁFICA REFERENTE À INTRODUÇÃO

1. Taş, N et al.(2021) Metagenomic tools in microbial ecology research. Current Opinion in Biotechnology 67, 184–191.

2. Hai chang Y, Hong yan C, Li li Z et al. (2020) Metagenomic Analysis of the Virome in Lung Tissues and Guts of Wild Mice. https://doi.org/10.21203/rs.2.22024/v1]

3. Burrell CJ, Howard CR, Murphy FA. (2017) History and Impact of Virology. Fenner and White's Medical Virology 3–14. doi: 10.1016/B978-0-12-375156-0.00001-1

4. Tisza MJ et al (2020) Discovery of several thousand highly diverse circular DNA viruses. Elife 9.

5. Gu L et al (2018) Research Progress on Rolling Circle Amplification (RCA)-Based Biomedical Sensing. Pharmaceuticals 11.

6. ITIS-Report: Bubalus bubalis. https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=552476#null

7. Soroprevalência de estomatite vesicular, diarreia viral bovina e infecção por herpesvírus bovino tipo 1 em búfalos (*Bubalus bubalis*) no Distrito Federal | VETTESES. https://pesquisa.bvsalud.org/portal/resource/pt/vtt-220192.

8. Bernardes O (2007) Bubalinocultura no Brasil: situação e importância econômica. Rev Bras Reprod Anim, Belo Horizonte, 31:.293-298.http://www.cbra.org.br/pages/publicacoes/rbra/download/293.pdf

9. dos Santos JCG. et al (2020) Analysis of the population structure of buffaloes in Brazil. Trop Anim Health Prod 52, 3889–3892.

10. Cassiano, LAP et al (2003) Caracterização fenotípica de raças bubalinas nacionais e do tipo Baio. Pesquisa Agropecuária Brasileira. 38:11 https://doi.org/10.1590/S0100-204X2003001100013

11. Minervino AHH, Zava M, Vecchio D, Borghese A (2020) Bubalus bubalis: A Short Story. Front Vet Sci 7.

12. Cerqueira LR, Bastianetto E (2009) Doenças Infecciosas em Búfalos. Ciência Animal Brasileira, supl.1.

13. Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11, 1095–1099.

14. Sklaviadis T, Akowitz, A, Manuelidis, EE, Manuelidis L (1990) Nuclease treatment results in high specific purification of Creutzfeldt-Jakob disease infectivity with a density characteristic of nucleic acid-protein complexes. Arch Virol 112, 215–228.