UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MICROBIANA

Expressão heteróloga de antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificação de coronavírus em água de esgoto

IKARO ALVES DE ANDRADE

BRASÍLIA – DF
FEVEREIRO / 2024

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MICROBIANA

Expressão heteróloga de antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificação de coronavírus em água de esgoto

IKARO ALVES DE ANDRADE

Tese apresentada ao Programa de Pós-Graduação em Biologia Microbiana, Instituto de Ciências Biológicas, Universidade de Brasília, como requisito para obtenção do título de Doutor em Biologia Microbiana.

Orientador: Prof. Dr. Tatsuya Nagata.

BRASÍLIA – DF
FEVEREIRO / 2024

Ikaro Alves de Andrade

Expressão heteróloga de antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificação de coronavírus em água de esgoto

Banca Examinadora:

Dr. Nicolau Brito da Cunha
Departamento de Agronomia – Universidade de Brasília – UnB
Membro Interno

Dr.ª Valéria Christina de Rezende Féres
Departamento de Farmácia – Universidade Federal de Goiás – UFG
Membro Externo

Drª. Ana Cláudia de Souza
Centro Universitário de Brasília - UniCEUB
Membro Externo

Dr. Eliane Ferreira Noronha
Departamento de Biologia Celular – Universidade de Brasília – UnB
Suplente

Dr. Tatsuya Nagata
Departamento de Biologia Celular – Universidade de Brasília – UnB
Orientador e Presidente da Banca

BRASÍLIA – DF
FEVEREIRO / 2024

Dedico o presente trabalho a aqueles que sempre foram o meu porto seguro: minha família.

AGRADECIMENTOS

Ao longo de minha trajetória acadêmica, eu passei por inúmeras experiências, que possibilitaram o meu desenvolvimento pessoal e profissional, no entanto, cursar o doutorado durante um período extremamente complexo como o de uma pandemia me possibilitou ressignificar muitos sentimentos, e evidenciar o papel da ciência e educação. Através deste presente espaço venho agradecer às instituições e as pessoas que contribuíram para o desenvolvimento deste trabalho.

Inicialmente, agradeço à CAPES (Comissão de Aperfeiçoamento de Pessoal do Nível Superior), e a Universidade de Brasília (UnB), por intermédio do Programa de Pós-Graduação em Biologia Microbiana (PPGBM), e pela Diretoria de Atenção à Saúde da Comunidade Universitária (DAC | DASU), pelo apoio financeiro e humano, que foram cruciais na execução deste projeto que apresenta um impacto significativo no auxílio à sociedade em períodos de crise, como evidenciado pela pandemia pelo COVID-19.

De forma extremamente especial, eu gostaria de agradecer ao meu orientador, o professor Dr. Tatsuya Nagata por todos os ensinamentos, as ‘‘broncas’’, os conselhos, e em especial, pelo acolhimento fraterno e humano quando mais precisei de ajuda. Agradeço também ao professor Dr. Fernando Lucas Melo por todos os ensinamentos, parceira e acolhimento no decorrer do doutorado. Tenham certeza de que eu desenvolvi uma forte admiração pelos profissionais de excelência que ambos são!

Ao professor Dr. Bergmann Ribeiro, o meu mais sincero agradecimento por possibilitar a participação em um dos projetos voltados ao COVID-19, e uso das estruturas pertencentes ao laboratório de Baculovírus. Muito obrigado também a todos o grupo de pesquisa em Baculovírus, em especial ao Bruno, Leonardo e Brenda.

A professora Dr.ª Eliane Noronha, por toda paciência e parceira enquanto estava na coordenação do PPGBM, o que estimulou a minha persistência na pós-graduação. Ademais, incluindo a referida professora juntamente aos professores Dr. Ricardo Krüger e Dr. Edivaldo, muito obrigado por possibilitarem o uso dos equipamentos pertencentes aos laboratórios de Enzimologia.

A professora Dr.ª Sônia Nair Baó, por toda paciência e prestatividade em permitir o uso de equipamentos destinados ao laboratório de Microscopia.

Muito obrigado à professora Dr.ª Samira Bührer, juntamente ao Leonardo e Mateus por todo o suporte e parceira na execução da parte imunológica do trabalho nas dependências da Universidade Federal de Goiás (UFG).

Aos meus queridos amigos e amigas da UnB: Macária, Caterine, Pedro, Jéssica, Giovana, Bruno, Karen, Marina, Vilson, Sandrinha, Luc, Alfonso, Jefferson, Ezequiel, Manu, Philipe, Mateus Ramos, Jonathan, Vittória; muito obrigado por todo o suporte nesta caminhada.

Aos meus amigos da docência: Joselita, Juliana, Melissa, Pablo, Karen, Bento, Marcela, Rosana, Marina, Anna Maly, Andrea, Daniela e aos demais, muito obrigado por todo o suporte!

A minha família, por todo o suporte e carinho, e por me ensinarem a ser forte mesmo nos momentos mais desafiadores. Obrigado por todo amor!

O meu mais profundo e sincero agradecimento, a minha querida amiga e professora M.ª Karine Watanabe, que me ensinou as bases da docência e o vasto mundo da Biomedicina, o qual sou muito grato. Mesmo estando longe, você tem grande parte nesta conquista!

Aos meus queridos amigos e professores M.ª Rosângela Reis e Dr. Rodrigo Lima, por todo o carinho e suporte desde a época da graduação.

Aos queridos professores Dr. Maurício Rossato e Dr.ª Thaís pelo suporte e sábios conselhos! Obrigado por todo o direcionamento!

Por fim, obrigado a Deus e a todos os guias de luz por todo o livramento, todas as bençãos, e possibilitar mais conquistas!

GRUPO DE PESQUISA

O Laboratório de Virologia Vegetal situado no departamento de Biologia Celular, e liderado pelo professor Dr. Tatsuya Nagata abarca o desenvolvimento de pesquisas nos segmentos de agronomia e biotecnologia. A partir do emprego de estratégias moleculares, o referido professor orienta alunos de iniciação científica, mestrado e doutorado em projetos que envolvem a compreensão do viroma de água de esgoto por intermédio de ferramentas de bioinformática, atividade de vírus vegetais e suas implicações em grandes culturas e uso de vetor viral para a produção de proteínas heterólogas em plantas. O presente trabalho foi fruto do uso de técnicas moleculares direcionadas ao campo biotecnológico mediante o cenário da pandemia por COVID-19. A partir do exposto, as atividades derivadas desta tese resultaram na consequente produção de proteínas heterólogas de SARS-CoV-2 em planta (*Nicotiana benthamiana*) e na quantificação do material genético de coronavírus oriundo de água de esgoto.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

RESUMO

ALVES-ANDRADE, IKARO. Universidade de Brasília, fevereiro de 2024. **Expressão heteróloga de antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificação de coronavírus em água de esgoto.** Orientador: Prof. Dr. Tatsuya Nagata.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pertence à família *Coronaviridae* e abrange um vírus que afeta mamíferos e aves. O vírus é o agente etiológico da doença COVID-19, que causou uma pandemia de extrema gravidade e até o momento resultou em aproximadamente 774.075.242 casos no mundo, deste total, 7,012,986 mortes. Neste cenário, diversos grupos de pesquisa em todo o mundo estão aprimorando ferramentas científicas para o desenvolvimento de estratégias diagnósticas, como a produção de proteínas heterólogas aplicáveis à testes rápidos. Neste sentido, para contribuir para o controle da pandemia, o presente trabalho buscou expressar de forma heteróloga o antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificar coronavírus em água de esgoto no intuito de contribuir no combate das epidemias. Em uma etapa posterior, foi realizado um estudo metagenômico em uma unidade de tratamento de esgoto de Brasília (DF) – Unidade Asa Norte (CAESB), mediante a coleta de amostras de água de esgoto nos meses de março e maio de 2016, e maio e agosto de 2020, com o intuito de descrever as principais sequências virais relacionadas à *Coronaviridae* presentes neste ambiente. Neste sentido, as amostras foram representadas como E163 e 165 para àquelas coletadas em março e maio de 2016, e E205 e E208 para os elementos coletados em maio e agosto de 2020. Para o teste de detecção, buscou-se a expressão de genes que codificam parte das proteínas estruturais S (S1-N) e NC do SARS-CoV-2 em *Nicotiana benthamiana*, empregando o vetor viral vegetal baseado no agente *pepper ringspot virus*. Seguiu-se com a purificação das proteínas recombinantes e sua avaliação em testes sorológicos para uma detecção em escala laboratorial de SARS-CoV-2. Clones recombinantes foram produzidos e agroinoculados em plantas de fumo, obtendo-se 39.5 µg de S1-N e 46.0 µg de NC por grama de folha seca. Essas proteínas reagiram positivamente (n=5) com soros de pacientes positivos ao vírus, demonstrando o potencial de aplicação em kits de detecção. Posteriormente, pela análise metagenômica de amostras do esgoto, foram identificados fragmentos de genoma relacionados a espécies como *Alphacoronavirus* 1 e SARS-CoV-2, de forma que foi possível a quantificação de RNA viral da última espécie nas amostras coletadas em 2020, sendo que para E205 correspondeu à 4931 cópias por µL de amostra, e para E208 foi de 3367 copies por µL de amostra. Os dados presentes neste trabalho apontam inicialmente a viabilidade da produção destas proteínas recombinantes e sua possível aplicação no segmento diagnóstico, uma vez que a estratégia de produção pode ser facilmente replicada. Além disso, os dados oriundos das amostras de esgoto podem contribuir para a identificação e consequente monitoramento dos vírus presentes neste ambiente.

PALAVRAS–CHAVE: Coronavírus, Expressão Heteróloga, Água de esgoto.

ABSTRACT

ALVES-ANDRADE, IKARO. Universidade de Brasília, março de 2020. **Heterologous expression of sars-cov-2 antigen in *Nicotiana benthamiana* and identification of coronavirus in sewage water.** Orientador: Prof. Dr. Tatsuya Nagata.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the Coronaviridae family and encompasses a virus that affects mammals and birds. The virus is the etiological agent of the COVID-19 disease, which caused an extremely serious pandemic and to date has resulted in approximately 774,075,242 cases worldwide, of which 7,012,986 deaths. In this scenario, several research groups around the world are improving scientific tools for the development of diagnostic strategies, such as the production of heterologous proteins applicable to rapid tests. In this sense, to contribute to the control of the pandemic, the present work sought to heterologously express the SARS-CoV-2 antigen in Nicotiana benthamiana and identify coronaviruses in sewage water in order to contribute to combating epidemics. In a subsequent stage, a metagenomic study was carried out in a sewage treatment unit in Brasília (DF) – Asa Norte Unit (CAESB), by collecting sewage water samples in the months of March and May 2016, and May and August 2020, with the aim of describing the main viral sequences related to Coronaviridae present in this environment. In this sense, the samples were represented as E163 and 165 for those collected in March and May 2016, and E205 and E208 for the elements collected in May and August 2020. For the detection test, the expression of genes that encode part of the S (S1-N) and NC structural proteins of SARS-CoV-2 in Nicotiana benthamiana, using the plant viral vector based on the agent pepper ringspot virus. This was followed by the purification of the recombinant proteins and their evaluation in serological tests for laboratory-scale detection of SARS-CoV-2. Recombinant clones were produced and agroinoculated in tobacco plants, obtaining 39.5 µg of S1-N and 46.0 µg of NC per gram of dry leaf. These proteins reacted positively (n=5) with sera from virus-positive patients, demonstrating the potential for application in detection kits. Subsequently, through metagenomic analysis of sewage samples, genome fragments related to species such as Alphacoronavirus 1 and SARS-CoV-2 were identified, so that it was possible to quantify viral RNA from the latter species in samples collected in 2020, and for E205 corresponded to 4931 copies per µL of sample, and for E208 it was 3367 copies per µL of sample. The data present in this work initially point to the viability of producing these recombinant proteins and their possible application in the diagnostic segment, since the production strategy can be easily replicated. Furthermore, data from sewage samples can contribute to the identification and subsequent monitoring of viruses present in this environment.

***KEY-WORDS*:** *Coronavirus, Heterologous Expression, Sewage.*

CAPÍTULO I. REVISÃO DE LITERATURA

1.1 INTRODUÇÃO

1.1 SARS-CoV-2: Caracterização e abordagem diagnóstica

Coronavírus (CoV) é o nome comum de um grupo de vírus da família *Coronaviridae* e gêneros *Alphacoronavirus*, *Betacoronavirus, Deltacoronavirus* e *Gammacoronavirus*, sendo que os membros do CoV são agentes patogênicos conhecidos por afetarem mamíferos e aves (CORMAN et al., 2018; BANERJEE et al., 2019; RAHMAN et al., 2021), causando doenças respiratórias e gastrointestinais (RENU et al., 2020; SALATA et al., 2020). O genoma de um CoV corresponde a um RNA de fita simples e senso positivo com aproximadamente 30 quilobases (kb). A partícula viral, em forma de coroa (*corona*, em latim), apresenta proteínas de espícula (S), membrana (M), envelope (E) e nucleocapsídeo (NC) (SALATA et al., 2020).

Normalmente, as espécies que causam doenças em humanos correspondem a human coronavirus 229E (HCoV-229E) e human coronavirus NL63 (HCoV-NL63); severe acute respiratory syndrome coronavirus (SARS-CoV), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Middle east respiratory syndrome coronavirus (MERS-CoV), human coronavirus OC43 (HCoV-OC43) e human coronavirus HKU1 (HCoV-HKU-1) (RENU et al., 2020; SALATA et al., 2020). Patologias promovidas por HCoV-229E, HCoV-NL63, HCoV-HKU-1e HCoV-OC43 geralmente ocasionam febre, tosse, coriza e dor de garganta (SEAH; AGRAWAL, 2020); por outro lado; SARS-CoV-2, SARS-CoV e MERS-CoV estão associados ao desconforto respiratório, pneumonia e presença de infiltrado pulmonar (GRALINSKI; BARIC, 2015; ASTUTI; YSRAFIL, 2020).

SARS-CoV-2 pertence ao gênero *Betacoronavirus* e subgênero *Sarbecovirus* (SALATA et al., 2020; MAHMOOD et al., 2021) e apresenta identidade nucleotídica com espécies virais que afetam morcegos (ZHOU et al., 2020) e pangolins (LAM et al., 2020). Dentre suas proteínas, sabe-se que quatro delas são estruturais: glicoproteína de espícula (S) (141 kDa), proteína de envelope (E) (8 kDa), proteína de membrana (M) (25 kDa), e proteína de nucleocapsídeo (NC) (45 kDa) (HU et al., 2021). A proteína S é um homotrímero que apresenta o domínio de ligação ao receptor e interage com o receptor humano da enzima conversora de angiotensina 2 (ECA2) (MEYER et al., 2014; GRIFONI et al., 2020; ROSALES-MENDOZA et al., 2020). A proteína NC é responsável pela formação do nucleocapsídeo viral

e auxilia nos eventos de replicação de RNA e transcrição de mRNA (DEMURTAS et al., 2016; RAVI; SAXENA; PANDA, 2022). Sabe-se que é produzido em altas quantidades durante todo o período de infecção e é altamente imunogênico (ZENG et al., 2020), devido a isto, a referida proteína é um dos principais candidatos a um possível marcador diagnóstico (WANG, Y. et al., 2021).

O referido patógeno é o agente etiológico da doença COVID-19, que resulta em sinais clínicos gripais leves à moderados que pode ocasionar em uma síndrome respiratória aguda grave (SARS) que confere (CHAN et al., 2020; HUANG et al., 2020). Os primeiros relatos ocorreram em 2019 na cidade de Wuhan (China) (HUANG et al., 2020), sendo que novos casos começaram a ocorrer exponencialmente e o vírus se dispersou rapidamente para outras localidades, antes que as medidas de restrição de locomoção tornassem-se mais rígidas (CRODA et al., 2020). Posteriormente, mediante melhor caracterização dos efeitos da doença, o vírus foi declarado como emergência de saúde pública internacional pela organização mundial de saúde (OMS) em 30 de janeiro de 2020 (CANDIDO et al., 2020).

No Brasil, a doença foi considerada como emergência de saúde nacional em três (03) de fevereiro de 2020 (SOUZA et al., 2020), e a primeira normativa voltada a medidas de quarentena e investigação epidemiológica correspondeu à Lei 13.979 de seis (06) de fevereiro de 2020, que dispõe sobre as medidas para enfrentamento da emergência de saúde pública de importância internacional decorrente do coronavírus responsável pelo surto de 2019 (CRODA et al., 2020). Mediante a integração da atividade dos laboratórios centrais de saúde pública, houve a confirmação do primeiro caso em 26 de fevereiro de 2020, sendo este uma pessoa que retornava a São Paulo após uma viagem à região da Lombardia (Itália) (DE JESUS et al., 2020).

Durante o período de pandemia, os laboratórios centrais de saúde pública atuaram fortemente na compreensão dos padrões de dispersão do vírus em território nacional. No entanto, fatores como atraso na notificação dos casos, comunicação oficial ineficiente, disseminação de informações sem base científica sólida, a dificuldade no acesso aos insumos e, consequentemente, aos testes por parte da população comprometeram a avaliação em tempo real dos casos de COVID-19 (SOUZA et al., 2020; CARDOSO; FERNANDES; SANTOS, 2021).

Atrelado aos demais fatores que comprometem a avaliação dos casos, pode-se destacar também a influência de eventos de mutações no genoma viral. Vírus apresentam a capacidade inata de sofrer mutações ao longo de sua replicação, o que resulta no aparecimento de variantes. A depender da região gênica afetada, pode-se observar novos fenótipos com maior

transmissibilidade e virulência, além de menor proteção vacinal e cobertura terapêutica. Outra característica perceptível é a competição entre as linhagens, de maneira que à medida que novas variantes aparecem, estas se sobressaem em relação à ancestral e perpetuam-se no ambiente (CANTÓN et al., 2021).

No tocante a SARS-CoV-2, como orientado pela OMS, as variantes são comumente agrupadas mediante os possíveis riscos e atreladas à filogenia, de forma que se deve evitar fazer menção a particularidades geográficas (ABBASI, 2021). O intuito de não referenciar tais características tem como objetivo evitar qualquer problema geopolítico, no entanto, é praticamente inevitável não as referir por origem geográfica. Neste sentido, têm-se aplicado o sistema de nomenclatura PANGOLIN (do inglês, *Phylogenetic Assignment of Named Global Outbreak*) (RAMBAUT et al., 2020). As classificações foram estipuladas para proporcionar um sistema simples e objetivo de notificação, de forma que correspondem a: variantes sendo monitoradas (*variants being monitored*, VBM) variantes de interesse (*variants of interest*, VOI), variantes de preocupação (*variants of concern*, VOC) e variante de alta consequência (*variant of high consequence*, VOHC) (DHAMA et al., 2023). O quadro abaixo (**Quadro 01**) apresenta as principais variantes de SARS-CoV-2.

Quadro 1. Principais variantes de SARS-CoV-2

OMS	PANGO	Classificação de variantes	Detecção Inicial	Amostras Iniciais	Principais Mutações Típicas	Referência
Alpha	B.1.1.7	VOC	Reino Unido	Setembro de 2020	H69/V70del, Y144del, N501Y, A570D, P681H, S106/G107/F108del	HOFFMANN et al., 2021
Beta	B.1.351	VOC	África do Sul	Outubro de 2020	L242/A243/L244del, K417N, E484K, N501Y, S106/G107/F108del	ZHOU et al., 2021
Gamma	B.1.1.28.1 (P1)	VOC	Brasil e Japão	Dezembro de 2020	K417T, E484K, N501Y, S106/G107/F108del	SABINO et al., 2021
Delta	B.1.617.2	VOC	Índia	Outubro de 2020	T19R, (G142D), 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N	CHERIAN et al., 2021
Epsilon	B.1.427/ B.1.429	VOI	Estados Unidos da América	Junho de 2020	L452R, W152C, S13I, D614G	DENG et al., 2021
Zeta	B.1.1.28.2 (P2)	VOI	Brasil	Abril de 2020	L18F, T20N, P.26S, F157L, E484K, D614G, S929I, V1176F	VOLOCH et al., 2021
Eta	B.1.525	VOI	Múltiplos países	Dezembro de 2020	H69-V70del, Y144del, Q52R, E484K, Q677H, D616G, S929I, V1176F	PEREIRA et al., 2021
Tetha	B.1.1.28.3 (P3)	VOI	Filipinas e Japão	Fevereiro de 2021	141-143del, E484K, N501Y, P681H	ROSE et al., 2022
Iota	B.1.526	VOI	Estados Unidos da América	Novembro de 2020	LSF, T95I, D253G, D614G, V483A, H655Y, G669S, Q949R, N1187D	PETRONE et al., 2022
Kappa	B.1.617.1	VOI	Índia	Outubro de 2020	(T95I), G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H	SAVILLE et al., 2022
Lambda	C.37	VOI	Peru	Dezembro de 2020	246-252del, G75V, T76I, L452Q, F490S, D614G e T859N	ROMERO et al., 2021
Omicron	B.1.1.529	VOC	África do Sul	Novembro de 2021	Q498R, S477N, N501Y, K417N,	LAMBROU et al., 2022

Fonte: Adaptado de CHERIAN et al., 2021; DENG et al., 2021; HOFFMANN et al., 2021; PEREIRA et al., 2021; ROMERO et al., 2021; SABINO et al., 2021; VOLOCH et al., 2021; ZHOU et al., 2021; PETRONE et al., 2022; LAMBROU et al., 2022; ROSE et al., 2022; SAVILLE et al., 2022.

Em períodos de pandemia, os testes diagnósticos auxiliam na confirmação dos casos e estabelecimento de políticas públicas que visem o controle da doença. Caso tais ferramentas apresentem elevado custo de produção, maior será a dificuldade em utilizá-la de forma generalizada (YUSUF et al., 2021). No contexto da pandemia por COVID-19, a técnica de reação da transcriptase reversa seguida pela reação em cadeia da polimerase quantitativa (RT-qPCR) é a mais utilizada no diagnóstico molecular da doença, todavia, em decorrência do alto valor agregado (reagente, equipamento e mão de obra) não se pode executá-la de forma ampliada (CHERKAOUI et al., 2021; FOURATI et al., 2021).

Neste sentido, uma metodologia alternativa ao RT-qPCR, todavia, mais rápida, simples de usar, e aplicável especialmente em regiões com recursos limitados corresponde ao emprego de antígenos (SHERIDAN, 2020; YUSUF et al., 2021). Dentre os sistemas de produção aplicáveis, o uso de plantas para expressão transiente de proteínas é visto como uma ótima alternativa em comparação com algumas plataformas bem estabelecidas, devido à possibilidade de ampliação da escala de produção em um curto período, fator crucial em períodos de pandemia (CAPELL et al., 2020). As principais plataformas utilizadas foram o sistema MagnICON (DIEGO-MARTIN et al., 2020) vetor binário baseado em pEAQ (LINDSAY et al., 2020; MARGOLIN; VERBEEK; et al., 2020) e vetores pCBP (MAKATSA et al., 2021).

Além disso, esta estratégia pode oferecer maior segurança quanto à não propagação de patógenos que afetam mamíferos (PARK; WI, 2016). A maquinaria pós-traducional em plantas também corrobora com seu uso como uma plataforma biológica viável para a expressão genética heteróloga (CAPELL et al., 2020). A N-glicosilação é uma modificação pós-traducional que influencia características estruturais e funcionais de proteínas (O'CONNOR; IMPERIALI, 1996; PRABHU et al., 2021), o que impacta em sua atividade biológica (SOLÁ; GRIEBENOW, 2010). Plantas, enquanto eucariotos apresentam mais robustez deste sistema e permite melhor processamento de proteínas, o que torna mais atrativo para a síntese de proteínas recombinantes (WALSH; JEFFERIS, 2006; BOSCH; SCHOTS, 2010; WOLFERT; BOONS, 2013; DONINI; MARUSIC, 2019). A **Figura 1**, adaptado de Fischer e Buyel (2020) corresponde uma síntese dos benefícios e desafios no emprego de plantas como sistema de expressão (FISCHER; BUYEL, 2020).

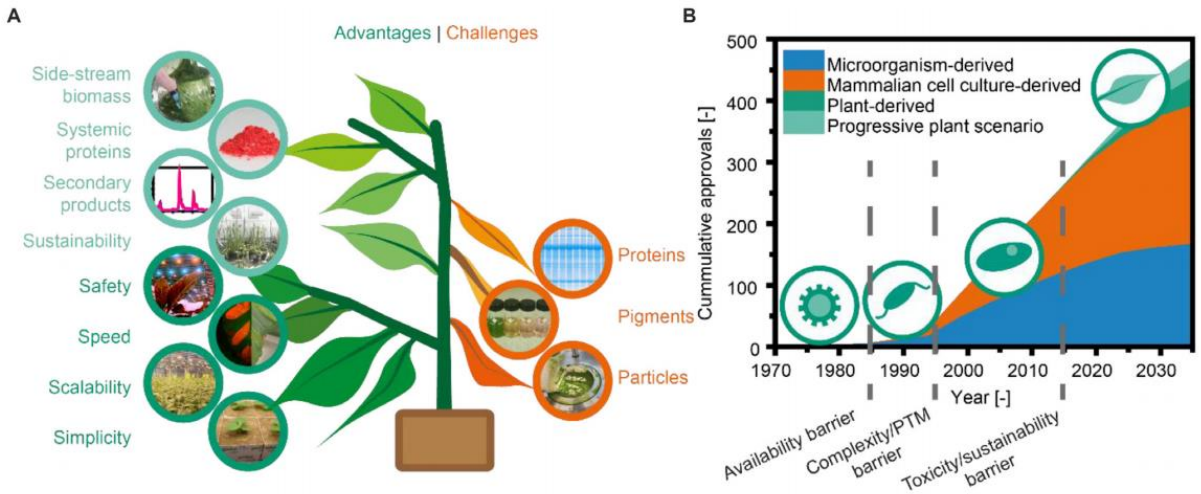


Figura 1. Principais vantagens e desafios atrelados ao emprego de plantas em sistema de produção de proteína heteróloga. Em **A** pode-se observar as vantagens, como a segurança no manuseio de material foliar, rápida produção de biomassa, produção de proteínas e produtos secundários, velocidade e escalabilidade e simplicidade. Enquanto os desafios observados normalmente compreendem a produção de proteínas que podem falsear os resultados, presença de pigmentos e partículas que podem inviabilizar a execução de algumas técnicas, como a cromatografia. Em B observa-se uma linha do tempo com os principais sistemas de expressão utilizados, e a tendencia positiva do uso de plantas a partir do ano de 2010 e que se intensificou no período posterior ao ano de 2020. Tal característica pode ser correlata com o período de pandemia que impulsionou o desenvolvimento de plantas como sistema de expressão (Adaptado de FISCHER e BUYEL, 2020).

Os kits de diagnóstico sorológico comercialmente disponíveis para COVID-19 são baseados em proteínas S ou NC como antígenos (GHAFFARI; MEURANT; ARDAKANI, 2020; LEE et al., 2020). Achados recentes indicam que pacientes com SARS-CoV-2 tendem a produzir mais anticorpos contra proteínas S e NC durante a infecção (HOULIHAN; BEALE, 2020). Estudos anteriores mostraram que a proteína SARS-CoV-2 NC pode ser usada para ensaios imunológicos e vacinação devido à sua antigenicidade (AHLÉN et al., 2020).

1.1.2 Plantas como sistema de produção heteróloga de proteínas

Atualmente, a produção de proteínas recombinantes é muito importante para a indústria farmacêutica e biotecnológica devido às suas diversas aplicações. A produção de proteínas para diagnósticos sorológicos de doenças infecciosas representa uma parcela significativa do mercado (BURNETT; BURNETT, 2020; SHANMUGARAJ; BULAON; PHOOLCHAROEN, 2020). Atualmente, existem variados sistemas para produção, que empregam bactérias (HAYAT et al., 2018; ZHANG, K.; SU; WU, 2020), leveduras (KARBALAEI; REZAEE; FARSIANI, 2020), plantas (FISCHER; BUYEL, 2020), culturas de células de insetos (GEISLER; JARVIS, 2018) e células de mamíferos (O'FLAHERTY et al., 2020).

A agricultura molecular (*molecular farming*) é atrelada ao uso de plantas para produção de proteínas recombinantes farmacêuticas que apresentam interesse industrial (RYBICKI, 2020a), iniciada no final da década de 1980 com a primeira produção de anticorpos em plantas, *Nicotiana tabacum* (LARRICK et al., 1989). As melhorias subsequentes nas metodologias de biologia molecular permitiram a produção de proteínas farmacêuticas em tecidos vegetais (SANTOS et al., 2016), musgos (DECKER; RESKI, 2012), algas (ROSALES-MENDOZA; PAZ-MALDONADO; SORIA-GUERRA, 2012a) e plantas aquáticas (EVERETT et al., 2012a).

Além disso, este movimento permitiu que empresas como *Medicago* (Québec, Canadá) e *Kentucky Bioprocessing* (Owensboro, EUA) se especializassem neste segmento de síntese de proteínas derivadas de plantas (HOLTZ et al., 2015; RYBICKI, 2020). Por exemplo, *Nicotiana benthamiana* tem sido explorada como uma plataforma biológica para a produção de proteínas recombinantes, incluindo a produção de antígenos de SARS-CoV (DEMURTAS et al., 2016), anticorpos anti-Ebola (HUANG et al., 2020), anti-Zika (DIAMOS et al., 2020), e anti-HIV (ZISCHEWSKI; SACK; FISCHER, 2016; LICO et al., 2020).

Anticorpos (Ac) são proteínas que reconhecem um antígeno (Ag), sendo o último, qualquer molécula capaz de estimular a atividade do sistema imunológico do hospedeiro. Corriqueiramente, empregam-se os Ac em atividades clínicas e biotecnológicas, de forma que os principais tipos consistem em monoclonais à base de hibridoma (mAc), monoclonais recombinantes (rAc), e policlonais (pAc) (CAMPBELL, 1991; ASCOLI; AGGELER, 2018). A síntese destas proteínas no organismo humano decorre de uma resposta imunológica natural ou por intermédio de processos de imunização. Estas moléculas são conhecidas por apresentarem características variadas, dentre elas, afinidade e especificidade. Tal variação, em certa parte, é decorrente da produção de induzida por proteínas antigênicas que interagem com diferentes epítopos do antígeno imunizante (PELTOMAA et al., 2022).

Anticorpos policlonais, por sua vez, são empregados desde o século XIX, com o intuito inicial de proteção contra doenças, toxinas e agentes patológicos. Estas moléculas caracterizam-se por apresentar uma abrangência maior a diversos epítopos, tal característica é útil para aplicações biotecnológicas (PELLETIER; MUKHTAR, 2020). No entanto, o mesmo caráter heterogêneo pode conferir certas desvantagens, uma vez que cada conjunto (lote) de Ac é único, cada remessa de anticorpos precisa ser produzida por respostas imunológicas semelhantes que, mesmo com adequado processo de validação, os anticorpos subsequentes podem apresentar variações de desempenho (ASCOLI; AGGELER, 2018). Anticorpos policlonais são

inconstantes, no sentido de que as características dos anticorpos purificados refletem o exato momento no qual o sangue foi retirado do animal, apesar de serem muito mais baratos e fáceis de obter (CAMPBELL, 1991; NING et al., 2021).

Dada a urgência na obtenção de proteínas heterólogas em baixo custo para o desenvolvimento de testes diagnósticos sorológicos precisos para o COVID-19, o objetivo deste trabalho foi primeiro: a expressão de genes que codificam a parte das proteínas S e NC do SARS-CoV-2, empregando novo vetor viral vegetal baseado agente pepper ringspot virus (PepRSV, membro do *gênero Tobravirus*), relatado anteriormente pelo nosso grupo de pesquisa (TAVARES-ESASHIKA et al., 2020) em plantas *de N. benthamiana*. O PepRSV tem genoma bissegmentado, RNA1 e RNA2, o que torna a clonagem mais fácil do que a do genoma mono segmentado devido ao menor tamanho de plasmídeos (TAVARES-ESASHIKA et al., 2022).

1.1.3 Identificação de coronavírus em viroma de esgoto

O esgoto é caracterizado como o conjunto de águas residuárias em uma determinada comunidade (CORSI et al., 2014; CHOI et al., 2018). Este tipo de água apresenta mudança em suas características naturais após a utilização humana, e a sua composição normalmente abrange despejos resultantes de atividade doméstica, industrial ou comercial que são veiculados pela água (KILIÇ, 2021). Além disso, também pode apresentar detritos presentes na superfície que em certo momento adentra a rede de esgoto (NUNES et al., 2021; TOPARE; ATTAR; MANFE, 2011), como excretas de humanos e animais (GERBA; BETANCOURT; KITAJIMA, 2017).

A compreensão de elementos dispostos na água de esgoto pode proporcionar dados relevantes sobre uma população. Neste sentido a metodologia de epidemiologia baseada no esgoto - EBE (em inglês *Wastewater-Based Epidemiology* – WBE) foi elaborada em 2001 (DAUGHTON, 2001) e adotada inicialmente em 2005 para o rastreio de drogas ilícitas (ZUCCATO et al., 2005) e monitoramento de medicamentos durante a pandemia de *Influenza* (H1N1) em 2009 (LEKNES; STURTZEL; DYE, 2012).

A premissa da EBE discorre na hipótese de que qualquer substância excretada por humanos é estável em água de esgoto e pode ser empregada para o cálculo a concentração original excretada pela população contribuinte (CHOI et al., 2018). O referido conceito é fortemente aplicável para a vigilância epidemiológica de vírus (BERCHENKO et al., 2017; DE FREITAS BUENO et al., 2022).

A vigilância epidemiológica pode ser definida como o monitoramento da prevalência de patógenos em uma população, com a finalidade de conhecer e monitorar o padrão de disseminação de doenças, assim como a identificar a emergência de novos patógenos (MARTIN et al., 2020; DE ARAUJO et al., 2021; YANIV et al., 2021). No contexto de doenças causadas por vírus, EBE pode proporcionar dados relativos à uma população e apontar regiões onde a concentração viral pode ser aumentada (REF, PECCIA et al 2020). Tais resultados resultam informações menos tendenciosas ao verificar indivíduos sintomáticos e assintomáticos (REF MURAKAMI et al., 2020). Dessa forma, se aplicável em consonância com dados clínicos, a EBE corresponde a uma valiosa estratégia metodológica para a elaboração de ações de saúde pública(FUMIAN et al., 2019; GUAJARDO-LEIVA et al., 2020; NIEUWENHUIJSE et al., 2020).

O viroma relacionado ao esgoto é complexo e compreende agentes patogênicos e comensais (FORTERRE; PRANGISHVILI, 2009; CORPUZ et al., 2020). A descrição da diversidade e abundância viral na água de esgoto pode contribuir para uma melhor compreensão de padrões de infecção em humanos (MONTAZERI et al., 2015; WHITNEY et al., 2021). A análise de RNA viral em água de esgoto pode ser aplicada como uma ferramenta diagnóstica indireta do nível da população infectada e de possíveis tendências de aumento ou diminuição do quantitativo de infectados (REF).

Os coronavírus podem ser detectados com recorrência em amostras de esgoto (BIVINS et al., 2020; MICHAEL-KORDATOU; KARAOLIA; FATTA-KASSINOS, 2020; ORIVE; LERTXUNDI; BARCELO, 2020), inclusive SARS-CoV-2 (AHMED et al., 2020; AMOAH; KUMARI; BUX, 2020; CORPUZ et al., 2020; DE ARAUJO et al., 2021; ROTHMAN et al., 2021). Neste sentido, a verificação de concentrações de partículas virais ou de fragmentos gênicos neste ambiente pode ser aplicável como possível ferramenta epidemiológica para estimar a prevalência da COVID-19 em uma população (POLIO et al 2020).

Para entender os padrões de disseminação de SARS-CoV-2, em certas populações, foram adotadas abordagens de EBE combinada com sequenciamento de alto rendimento (em inglês, *High-Throughput Sequencing* - HTS) (MARTIN et al., 2020; PECCIA et al., 2020; WHITNEY et al., 2021). Amostras de esgoto bruto são usadas como um sistema de observação e alerta para possíveis surtos de SARS-CoV-2, pois o RNA pode ser identificado nas fezes e posteriormente na água de esgoto semanas antes dos primeiros relatos oficiais da doença (HATA; HONDA; HONDA, 2020; HOLSHUE et al., 2020). Através dos avanços da bioinformática, a aplicação do HTS permite uma melhor compreensão da dinâmica viral em

ambiente residuário, pois pode possibilitar o estudo de uma maior quantidade de amostras proporcionando uma visão global da ocorrência e/ou circulação do vírus (NG et al., 2012; CORPUZ et al., 2020).

1.2 JUSTIFICATIVA

A pandemia de COVID-19 promoveu diversos momentos de reflexão na comunidade científica. Tais reflexões resultaram no desenvolvimento de variadas estratégias biotecnológicas e moleculares com o intuito de aprimorar ferramentas diagnósticas e de identificação de patógenos.

Uma das estratégias para desenvolvimento de novos testes diagnósticos corresponde à obtenção de proteínas heterólogas. A elaboração de um meio eficaz de produção em escala garante um suprimento consistente e padronizado de antígenos para a formulação de testes rápidos, por exemplo, que são passíveis de aplicação no contexto de saúde pública. A precisão e sensibilidade desses testes são diretamente influenciadas pela qualidade dos reagentes empregados, sendo as proteínas antigênicas componentes fundamentais nesse processo.

A expressão de proteína em plantas permite a produção em grande escala e minimiza condições atreladas à contaminações. Além disso, a depender das estratégias de purificação, corrobora-se os níveis de pureza dos produtos, o que minimiza a presença de contaminantes que possam interferir nos resultados de diferentes estudos.

Em relação à identificação de patógenos, a análise de material genético em águas residuárias pode corroborar atividades de vigilância epidemiológica, o que promove melhor compreensão sobre a dispersão de determinado agente etiológico em uma comunidade. Com estas informações, pode-se rastrear e monitorar a disseminação geográfica de vírus em uma determinada região, o que facilita a implementação de medidas preventivas e a alocação eficiente de recursos de saúde pública. O procedimento proporciona uma visão ampla e abrangente da diversidade viral presente em uma determinada população. Ao contrário de amostras clínicas individuais, por exemplo, que podem ser limitadas a casos sintomáticos e assintomáticos identificados.

O presente estudo enfoca o desenvolvimento de um sistema de produção de proteínas virais de SARS-CoV-2 de forma eficiente, rápida, de baixo custo e alta qualidade para uso em testes de detecção viral e de avaliar um método de identificação viral a partir de amostras de

água de esgoto para contribuir para a compreensão da epidemiologia do vírus no local e subsidiar ações de prevenção e combate.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

- Expressar de forma heteróloga o antígeno de SARS-CoV-2 em *Nicotiana benthamiana* e identificar coronavírus em água de esgoto no intuito de contribuir no combate das epidemias.

1.3.2 Objetivos Específicos

- Produzir proteínas estruturais (S e NC) de SARS-CoV-2 em plantas utilizando vetor viral.
- Produzir anticorpos policlonais contra S e NC de SARS-CoV-2 produzidos em plantas *Nicotiana benthamiana*.
- Avaliar a aplicabilidade das proteínas recombinantes e anticorpos policlonais em ensaios imunocromatográficos.
- Analisar as sequências e identificar os coronavírus circulantes na água de esgoto na região de Brasília, DF

CAPÍTULO II. *Practical use of tobravirus-based vector to produce SARS-CoV-2 antigens in plants*

O presente capítulo é referente à publicação do artigo ‘‘***Practical use of tobravirus-based vector to produce SARS-CoV-2 antigens in plants***’’ publicado no periódico ***Journal of Virological Methods*** – DOI: doi.org/10.1016/j.jviromet.2023.114710

1 INTRODUCTION

The use of plant systems presents an attractive alternative for pharmaceutical protein production due to their ability of a large-scale production within a short period. Additionally, the production of heterologous proteins in plants offers biosafety advantages as there is typically no need to use human-infecting pathogens (PARK; WI, 2016). Furthermore, this system provides a eukaryotic cell-based platform, which is important for certain types of protein production (CAPELL et al., 2020).

Historically, the molecular farming, defined as the use of plants to produce pharmaceutical recombinant proteins of industrial interest (FISCHER; BUYEL, 2020; RYBICKI, 2020b), started in the late 1980s with the first production of antibodies in plants of *Nicotiana tabacum* (HIATT; CAFFFERKEY; BOWDISH, 1989). Then, other systems, such as plant tissues (SANTOS et al., 2016), mosses (DECKER; RESKI, 2012), algae (ROSALES-MENDOZA; PAZ-MALDONADO; SORIA-GUERRA, 2012b), and aquatic plants (EVERETT et al., 2012b), were also reported.

Our expression system relies on the use of the gene expression vector based on pepper ringspot virus (PepRSV), a tobravirus found in Brazil. PepRSV has a bi-segmented genome, RNA1 and RNA2. Tavares-Esashika et al. (2020) constructed an expression vector of PepRSV, by modifying the RNA2 genome for insertion of the gene of interest, which makes the cloning step easier than when compared to a larger sized plasmid (TAVARES-ESASHIKA et al., 2020b). They demonstrated the efficiency of this vector using the modified green fluorescent protein (mGFP) gene as a reporter gene (TAVARES-ESASHIKA et al., 2022). However, this vector system had not been applied for other genes of interest, such as those useful for pharmaceutical purposes.

In the context of the pandemic caused by SARS-CoV-2, the use of plant-based platforms for SARS-CoV-2 antigen production has increased (CAPELL et al., 2020). The major platforms used for transient expression were of the binary vector systems as pEAQ-based (LINDSAY et al., 2020; MARGOLIN et al., 2020; MAMEDOV et al., 2021; JUNG et al., 2022), pCAMBIA-based (pUPD) (DIEGO-MARTIN et al., 2020), pCBP (MAKATSA et al., 2021) and pBYR2fp (MOON et al., 2022) vectors. Furthermore, plant virus-based vectors, including those tobamovirus-based (MAHARJAN et al., 2021; WILLIAMS et al., 2021), potexvirus-based (MAHARJAN et al., 2021; MARDANOVA; KOTLYAROV; RAVIN, 2021) and potyvirus-based vectors (ACHS; GLASA; ŠUBR, 2022), were also used to produce SARS-CoV-2 antigens.

For a diagnostic use, mostly the structural proteins of SARS-CoV-2 were chosen, i.e., the spike (S) and the nucleocapsid (NC) proteins. The S protein forms a homotrimer that presents the receptor-binding domain for the human angiotensin-converting enzyme 2 receptor (MEYER; DROSTEN; MÜLLER, 2014; GRIFONI et al., 2020; ROSALES-MENDOZA et al., 2020). This viral protein is the most important antigen to produce subunit vaccines to induce neutralizing antibodies (CHEN et al., 2020; ZHANG, J. et al., 2020). The NC protein is responsible for the formation of the viral nucleocapsid complex, assists in the RNA replication and the mRNA transcription events (MCBRIDE; VAN ZYL; FIELDING, 2014). It is known that it is produced in high amounts throughout the infection period and is highly immunogenic (ZENG et al., 2020). Hence, it is one of the preferred viral proteins for the diagnostic marker (WANG, H. et al., 2021). Recent findings indicate that patients with SARS-CoV-2 tend to produce more antibodies against S and N proteins during the onset of the disease (HOULIHAN; BEALE, 2020).

Our objective was to characterize the performance of the PepRSV vector for SARS-CoV-2 genes expression. We intend to help developing a system to produce reliable, large-scale, and low-cost SARS-CoV-2 detection tests.

2 MATERIALS AND METHODS

2.1 Vector construction and cloning

cDNA constructs of RNA1 and modified RNA2 of PepRSV (**Figure S1**; Tavares-Esashika et al., 2022) containing the gene of interest were agro-infiltrated in the plant for

transient expression. Antigenic regions of S and NC proteins were selected based on the analyses of epitope prediction, which was carried out by the online tool ElliPro (https://www.iedb.org/) after prediction of the three-dimensional structure by Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). By the epitope analyses of S (accession number YP_009724390) and N (YP_009724397) proteins of Wuhan isolate, the partial gene fragments based on the N-terminus of the S protein, named S1-N (12 – 316 aa) (which is a conserved region of the S protein) and the NC (37 – 402 aa) were synthesized with codon-optimization for *N. benthamiana* codon usage (Epoch Life Science, Sugar Land, USA) (GenBank accession numbers ON890138 - ON890143).

Three constructs for each fragment were prepared to compare the productivity of the target proteins, named Construct 1 (C1), C2, and C3 (**Figure 1**) in the gene of interest (GI) region of RNA 2 construct (**Figure S1**). C1 (ON890138) contains the signal peptide sequence from the basic chitinase gene of Arabidopsis thaliana (AAK96819) at the 5′-end, replacing the S protein signal peptide (the first 11 amino acids of the S protein) sequence, followed by the S1-N sequence and the endoplasmic reticulum retention signal (KDEL) peptide in the C-terminus (**Figure 1**). C2 (ON890139) is designed as the fused protein of the N-terminus of the coat protein (CP) of PepRSV before S1-N to be translated in the plant cell cytoplasm (without signal peptide and no KDEL peptide in C-terminus). In general, N-terminus of viral coat protein provides stable features in the infected cells, thus, we expected to have higher protein accumulation with C2 (**Figure 1**). C3 (ON890140) construct was also designed for translation in the cytoplasm, but as a fusion protein with the N and C-terminal regions of the CP of PepRSV flanking a modified green fluorescent protein (mGFP). The 2A autocleavage peptide sequence of foot-and-mouse disease virus (GSGEGRGSLLTCGDVEENPGP) was added between mGFP and S1-N fragments (**Figure 1**).

The constructs of the NC protein coding sequence were also designed and prepared using the same strategies as S1-N (**Figure 1**) (ON890141, ON890142 and ON890143). The hexahistidine tag (His-tag) sequence was added in the 3′-end of all constructs (in the C1 constructs of both sequences, the His-tag sequences were added prior to the KDEL sequences). Oligonucleotides used in this study are described in **Table S1**. Gene cassettes and backbones were amplified by PCR using Q5 High Fidelity DNA polymerase (New England Biolabs, Ipswich, USA). Amplicons obtained by PCR were purified using the QIAEX II Gel extraction kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The resulting fragments were ligated by the Gibson assembly reaction (Blawid and Nagata, 2015, Tavares-

Esashika et al., 2020) using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs), following the manufacturer's instruction. *E. coli* cells of the DH10B strain (Thermo Fisher Scientific, Waltham, USA) were transformed with these constructs by electroporation. Plasmids were purified from the resulting transformants using the Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, USA) and sequenced using the Sanger method.

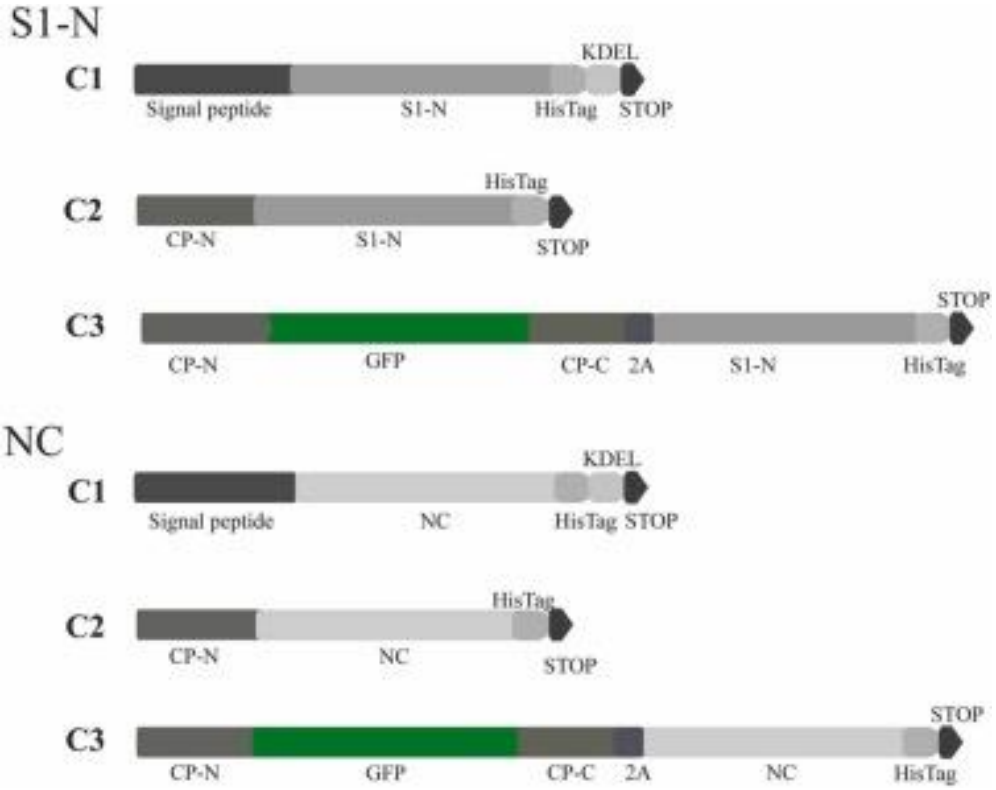


Figure 1. Scheme of the S1-N and NC protein sequence constructs in RNA2 segment of PepRSV. *S1-N_C* and *NC_C1* contain the signal peptide sequence of the chitinase gene of *Arabidopsis thaliana*, a hexahistidine (His)-tag, and the KDEL peptide sequence in addition to the gene of interest. S1-N_C2 and NC_C2 contains the sequence of the N-terminus of PepRSV CP sequence for protein stability, and a His-tag in addition to the gene of interest. S1-N_C3 and NC_C3 contains the mGFP gene sequence flanked by the N- and C-terminal peptides of the PepRSV CP protein, an autocatalytic 2 A peptide from of foot-and-mouse disease virus, and a His-tag. These constructs were inserted in the open reading frame of *CP* gene of the RNA2 segment of PepRSV.

2.2 Agroinfiltration in *N. benthamiana* plants and transient expression.

Agrobacterium tumefaciens GV3101:pMP90 was transformed by electroporation with the constructs described in the previous section; the transformed cells were maintained in the modified Luria-Bertani medium (LB3 medium: 2 g/L yeast extract, 10 g/L peptone, 4 g/L NaCl, 1 g/L KCl, 3 g/L $MgSO_4 \cdot 7H_2O$), with 50 µg/mL kanamycin, 25 µg/mL gentamicin, and 50 µg/mL rifampicin and preserved at −70 ºC in 25 % (v/v) glycerol solution. Briefly, *A. tumefaciens* was cultured for 12 h at 28 °C in an orbital incubator set at 180 rpm in LB3 medium with antibiotics. The cultures were centrifuged for 10 min at 3000 × g at 8 ºC and resuspended

to an OD_{600}= 2.0 with Murashige-Skoog medium containing 150 μM acetosyringone for each RNA1 and RNA2 constructs (TAVARES-ESASHIKA et al., 2020b). This relatively high OD_{600} value was determined by previous work (TAVARES-ESASHIKA et al., 2022). Agrobacteria containing RNA1 and RNA2 constructs were mixed with equal volume and agro-infiltrated using a 1 mL syringe without a needle (BLAWID; NAGATA, 2015) to the abaxial surface of *N. benthamiana* leaves (5–6 weeks old). The plants were incubated at 25 ºC with an air moisture content of 70 % and photoperiod of 16 h per day. An initial evaluation of recombinant protein production (S1-N and NC) was performed by SDS-PAGE and Western blot as described below.

Vacuum agroinfiltration of the whole plant with the selected constructs was conducted. In brief, agrobacterial cultures were prepared, as described above, but in larger volume of 500 mL each (RNA1 and RNA2 constructs). Almost all aerial parts of the *N. benthamiana* plant were submerged in an agrobacterial suspension and subjected to a vacuum of 66 kPa for 1 min, followed by a quick vacuum release for efficient infiltration. Finally, the plants were incubated for four days in the growth chamber. The mGFP production as a gene expression indicator harboring the C3 was monitored using a portable ultraviolet lamp (365 nm; Blak-Ray® - UVP, Upland, USA).

2.3. Recombinant protein purification and Western blot analysis

For the initial analysis of protein production, leaf tissues of the agro-infiltrated plants were collected and ground in the proportion of 1:1 (weight/volume) with phosphate buffered saline (PBS, pH 7.4) containing 2 mM dithiothreitol, and 2 mM of the protease inhibitor phenylmethylsulfonyl fluoride (PMSF). The homogenized tissue was clarified by centrifugation for 10 min at 10,000 × *g* at 4 °C, and 10 μL of the clarified supernatant were applied to a 12 % SDS-PAGE using a Mini-Protean Tetra Cell (Bio-Rad, Hercules, USA). The visualization of proteins in SDS-PAGE was performed by silver staining; later, proteins were analyzed by Western blot using an anti His-tag monoclonal antibody (1:10,000) (Sigma-Aldrich, San Luis, USA), and anti-mouse IgG (H+L) conjugated with alkaline phosphatase produced in goat (1:10,000) (Thermo Fisher Scientific). The colorimetric visualization was performed using an NBT/BCIP substrate system (Sigma-Aldrich).

In order to obtain a higher protein amount, vacuum-agroinfiltrated leaves (40–60 g of leaves, either fresh or stored in the ultrafreezer at −70 °C) were ground with liquid nitrogen,

and the resulted powder was mixed with the extraction buffer 1:5 (w/v) [50 mM Tris-HCl, 150 mM NaCl, 30 mM imidazole, 0.2 % (v/v) Triton X-100, pH 7.1] supplemented with 1 mM PMSF, and 2 mM β-mercaptoethanol. After mixing, the clarified supernatant containing the soluble recombinant protein was obtained by centrifuging twice for 20 min at 10,000 × *g* at 4 ºC; the clarified supernatant was stored at 4 ºC until further use in purification steps. S1-N and N proteins were purified using the HisPur Ni-NTA resin (Thermo Fisher Scientific). First, the resin was equilibrated with the extraction buffer [50 mM Tris-HCl, 150 mM NaCl, 30 mM imidazole, 0.2 % (v/v) Triton X-100, pH 7.1]. Then, 25 mL of the clarified supernatant were mixed with 500 µL of equilibrated resin for 1 h at 4 ºC; the resin was precipitated by centrifugation for 5 min at 700 × *g* at 10 ºC.

The most stringent resin-washing step was evaluated using washing buffer [50 mM Tris-HCl, 150 mM NaCl, 0.2 % (v/v) Triton X-100, pH 7.1] containing stepwise concentrations of imidazole (40, 60, 80, 120, 250, and 500 mM), confirmed by SDS-PAGE with silver-staining and Western blot. After determination of the most stringent washing condition (80 mM imidazole for S1-N and 120 mM for N), the purified protein was eluted by applying 1 mL of elution buffer (50 mM Tris-HCl, 500 mM imidazole, pH 7.1) into the column. The purified protein was analyzed using SDS-PAGE with silver-staining and Western blot. For protein concentration, the purified pooled fraction was applied to an Amicon® Ultra-15 Centrifugal Filter (30 kDa cut-off - Millipore, Burlington, USA) and centrifuged for 40 min at 4000 × *g* at 12 ºC, resulting in a solution with four-fold concentrated protein. Protein concentration was evaluated by the Bradford protocol, using bovine serum albumin (BSA) as a standard.

To estimate molecular mass of the S1-N protein without oligosaccharide chains, the semi-purified S1-N protein was treated with endoglycosidase H (Promega) following the manufacturer's instructions, with subsequent visualization on SDS-PAGE by silver staining and Western blot.

2.4. Reactivity of antigens to patients' sera

Dot-immunobinding assay (DIBA) and enzyme-linked immunosorbent assay (ELISA) were carried out to evaluate the reactivity of the produced antigens to patients' sera. Antigen solution containing 1 µg to 1 ng (dilution series of 10 times) was spotted onto a nitrocellulose membrane and dried on air for more than 4 h. Then, the membrane was incubated with previously diagnosed patient sera by RT-qPCR as positive and negative, with 1:1000 dilution

for 1 h. After washing with PBS containing 0.05 % (w/v) Tween-20 (PBS-T) three times, the membrane was incubated with goat anti-human IgG (0.3 µg/mL in PBS) conjugated with alkaline phosphatase (Sigma-Aldrich) for 1 h. After the washing step with PBS-T, the membrane was incubated with NBT/BCIP substrate in alkaline buffer (330 µg/mL NBT 165 µg/mL BCIP in 100 mM Tris-HCl pH 9.5 buffer containing, 100 mM NaCl and 5 mM $MgCl_2$) for color development. ELISA was carried out in a 96-well plate (Nunc MaxiSorp, Thermo Fisher Scientific).

Overnight antigen coating was performed at 4 °C by applying 100 µL/well of the recombinant S1-N and N in bicarbonate/carbonate buffer (0.1 M Na_2CO_3, 0.1 M $NaCO_3$, pH 9.6) for a final concentration of 50 ng/100 µL. Subsequently, the wells were washed thrice with PBS-T, and blocked with PBS-T containing 1 % (w/v) BSA at 37 °C for 1 h; 100 µL of patient sera (1:200 dilution) in PBS-T supplemented with 0.05 % (w/v) BSA was added and kept at 37 °C for 1 h. The plates were washed thrice with PBS and incubated with goat anti-human IgG (Fc-specific) peroxidase conjugate (Sigma-Aldrich) (1:50,000) in PBS-T plus 0.5 % (w/v) BSA, for 1 h at 37 °C. The wells were washed six times with PBS-T, then, TMB substrate (Sigma-Aldrich) was added; the reaction was stopped by the addition of 2 M sulfuric acid.

3. RESULTS

3.1. Production and purification of SARS-CoV-2 S1-N and NC proteins

The sequences coding for the S1-N and N proteins of SARS-CoV-2 containing multiple complementary domains: His-tag, signal peptide, KDEL peptide, PepRSV CP N-terminal (CP-N) and/or C-terminal (CP-C) peptides, and mGFP (**Figure 1**) were expressed *in N. benthamiana* plants by transient gene expression. In the first step, small-scale gene expression was done by syringe-based agroinfiltration.

All agroinfiltrated parts of the leaves showed yellowish color due to the reaction to agrobacteria at two days post-infiltration (dpi); however, leaves infiltrated with the *C2* and *C3* constructs of *S1-N* (*S1-N_C2* and *S1-N_C3*) and *C1* of *NC* (*NC_C1*) showed very severe cell death responses in the agroinfiltrated regions at 3 dpi. This host response was considered as a hypersensitive-like programmed cell death by plant innate immunity (pattern-triggered immunity, PTI) and the production of recombinant proteins was expected to be very low. The initial analysis in SDS-PAGE and Western blot using plant leaves of 4 dpi demonstrated that

the S1-N protein production was confirmed only in one of two clones of *C1* but not in *C2* nor *C3* (**Figure 2**).

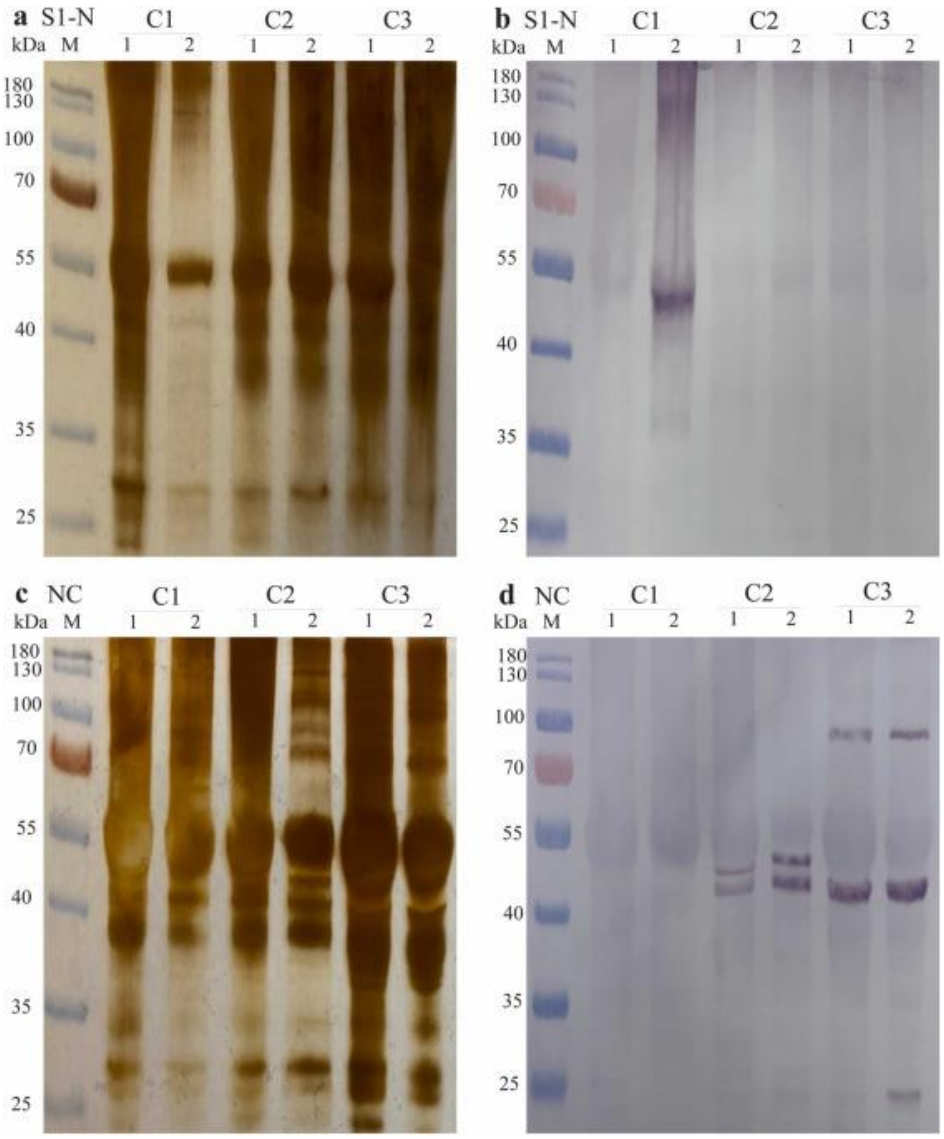


Figure 2. Protein profile of recombinant S1-N and N proteins produced in *N. benthamiana*. Silver stained 12 % SDS-PAGE of S1-N (a) and NC (c) proteins. Western blot (with anti His-tag) S1-N (b) and NC (d) proteins. The numbers (1 and 2) showed clone numbers. PageRuler™ Prestained Protein Ladder was used as the molecular mass standard.

Although the insert position and the sequence were confirmed by Sanger sequencing, no production was observed in clone 1 of *S1-N_C1* for an unknown reason, whereas N protein production was observed in both constructs of *C2* and *C3* (**Figure 2**) but not in *C1*. The molecular mass of S1-N_C1 was predicted as 38 kDa. However, the protein labeled with an anti His-tag antibody was about 53 kDa. This discrepancy was expected to be the glycosylation that occurred in the S1-N_C1 protein. Hence, the deglycosylation treatment was done using

semi-purified S1-N_C1 protein. After the treatment with endoglycosidase H, the protein reacted with anti His-tag antibody was confirmed at around 38 kDa (**Figure 3**). Therefore, it was concluded that the discrepancy in molecular mass was due to glycosylation. With these experiments, the clone 2 of S1-N_C1 was chosen for further study.

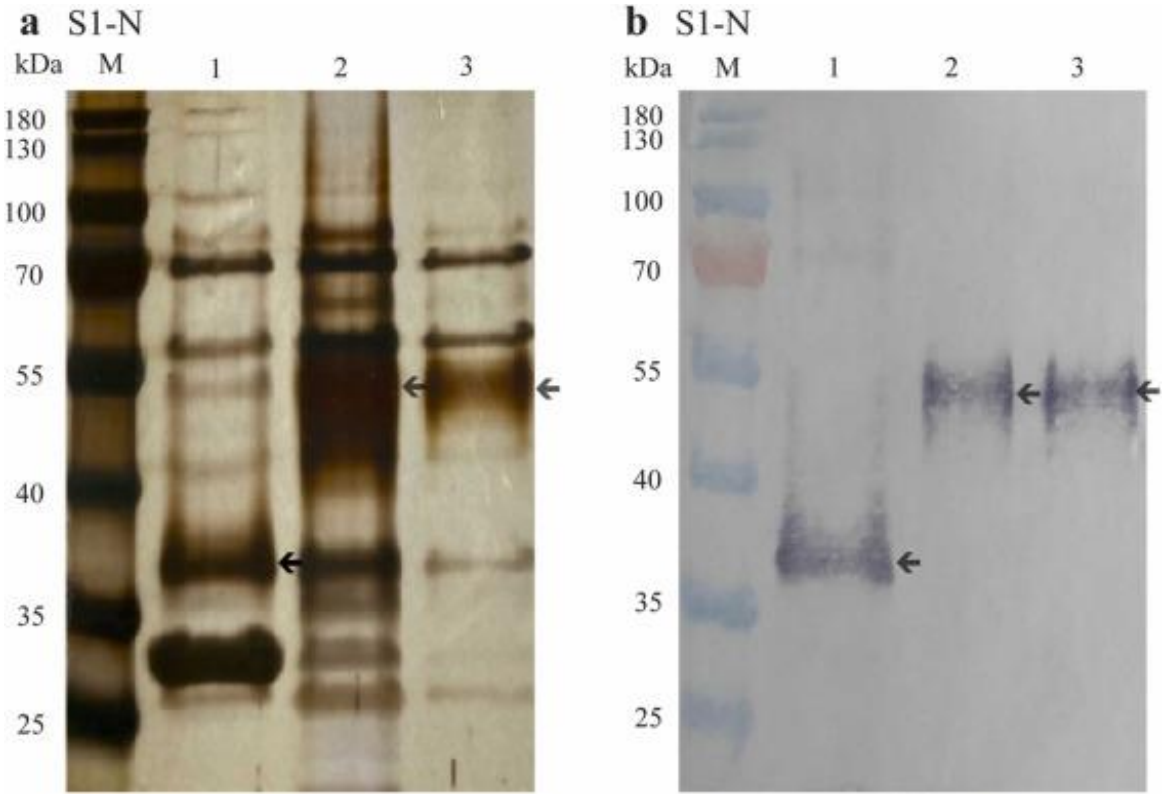


Figure 3. Profile of the S1-N protein submitted to endoglycosidase H treatment. Silver stained 12 % SDS-PAGE of purified S1-N (a) and Western blot (with anti His-tag) of purified S1-N (b). Black arrows indicate the reacted protein with anti His-tag antibody. Lane 1, enzyme-treated protein; lane 2, the protein treated only with enzyme buffer without endoglycosidase; and lane 3, S1-N protein without any treatment. PageRuler™ Prestained Protein Ladder was used as the molecular mass standard.

The molecular mass of NC was predicted as 45 kDa in NC_C2 and 74 kDa in NC_C3 as the fused protein with mGFP; and 39 kDa as the free NC_C3 protein, which was expected when autocatalytic peptide 2 A worked (**Figure 2d**). The NC_C2 protein band appeared near the expected size, and an additional larger protein was also detected by Western blot (**Figure 2d**). The presence of the protein with a larger size in NC_C2 is discussed later. In NC_C3, both fused protein and possible free NC proteins were found by Western blot with larger sizes than expected. In clone 2 of *NC-C3*, one more additional target protein at around 25 kDa was shown which was thought to be a cleaved protein. The protein conformational features can cause these slight differences of molecular mass. In this experiment, the *NC_C3* construct was selected for further study due to higher protein accumulation and having a mGFP indicator, which allows

easy temporal monitoring of the gene expression. The strong mGFP signaling of NC_C3 in *N. benthamiana* under UV light was shown in **Figure S2**.

After determining the construct (*C1* for *S1-N* and *C3* for *NC*), the optimal condition for Ni-NTA column washing was evaluated, i.e., the concentration of imidazole in the washing buffer. For this purpose, the stepwise imidazole concentrations (40–500 mM) were tested for column washing. To prepare purer protein, 80 mM imidazole for S1-N and 120 mM for N were chosen for the column washing procedure and 500 mM for elution (**Figure S3**). Then, a vacuum agroinfiltration procedure was applied for the preparation of larger amount of purified S1-N and NC proteins. After purification with the condition determined above, the purified proteins were evaluated by SDS-PAGE and Western blot. A more prominent band of purified S1-N protein with approximately 53 kDa was observed in SDS-PAGE and confirmed by Western blot, as shown in **Figure 4a**. One more prominent protein band was recognized near 110 kDa by Western blot (**Figure 4a**, lane 2). This protein band may represent the dimeric S1-N protein according to the molecular mass.

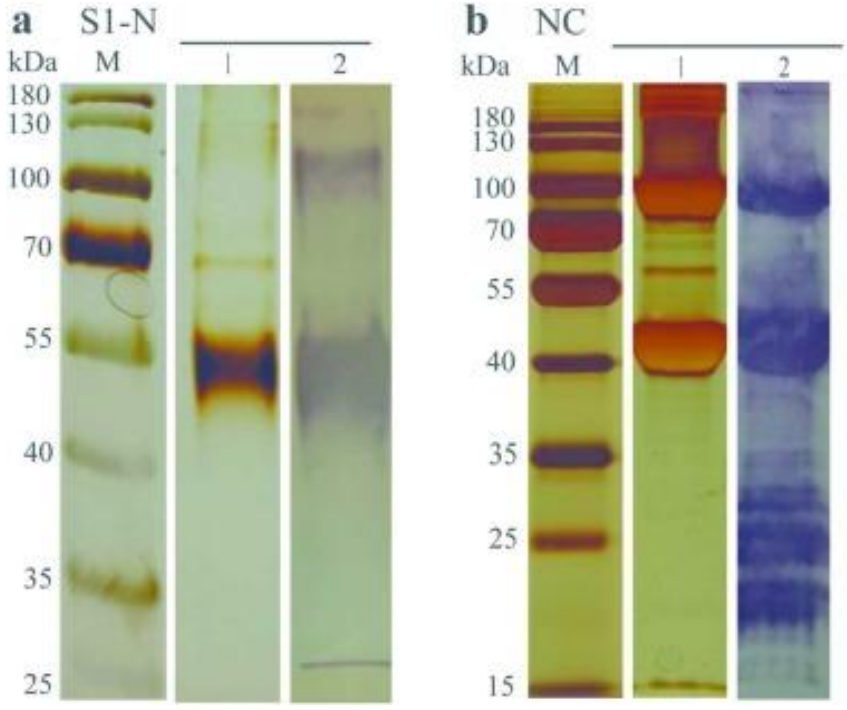


Figure 4. Profile of the purified recombinant proteins of S1-N and NC. Silver-stained 12 % SDS-PAGE (1) and Western blot (2) of the purified S1-N (a) and NC protein (b). PageRuler™ Prestained Protein Ladder was used as the molecular mass marker.

Regarding NC protein expression analysis shown in **Figure 4b**, the observed prominent band corresponds to nearly expected size of 39 kDa. The higher molecular mass band corresponds to the mGFP-fused N protein (74 kDa) was also observed. Smaller proteins below

35 kDa were also found in Western blot (**Figure 4b**, lane 2), which were considered as fragmented proteins generated during the process. In terms of protein yields obtained after purification, 40 or 60 µg of S1-N and NC (together with fused with mGFP) purified proteins were obtained from 1 g of agroinfiltrated fresh leaves, respectively.

3.2. Serological evaluation of recombinant proteins of S1-N and NC

DIBA was performed to assess the reactivity of S1-N and NC proteins as antigens to antisera of the patients with COVID-19. N protein showed significant reactivity to patient's serum from antigen loads of 1, 0.1, and 0.01 µg (**Figure S4**); however, the S1-N protein did not show clear signals (data not shown). Then, the ELISA protocol was performed using these two antigens with dilutions. The results showed that sera from patients diagnosed with SARS-CoV-2 by RT-PCR (n = 5) recognized the S1-N and NC antigens, while the negative control samples (sera obtained pre-pandemic, n = 3) showed extremely low ELISA values, which were considered negative reactions (**Figure 5**). Out of the five positive sera tested, three showed a strong reaction and two showed a weak reaction. Interestingly, the serum that reacted strongly to one antigen also showed a high level of reaction to another antigen. Despite the pilot test with a small number of samples, our study showed that the SARS-CoV-2 antigens made by a plant platform have high reactivity and almost no background level.

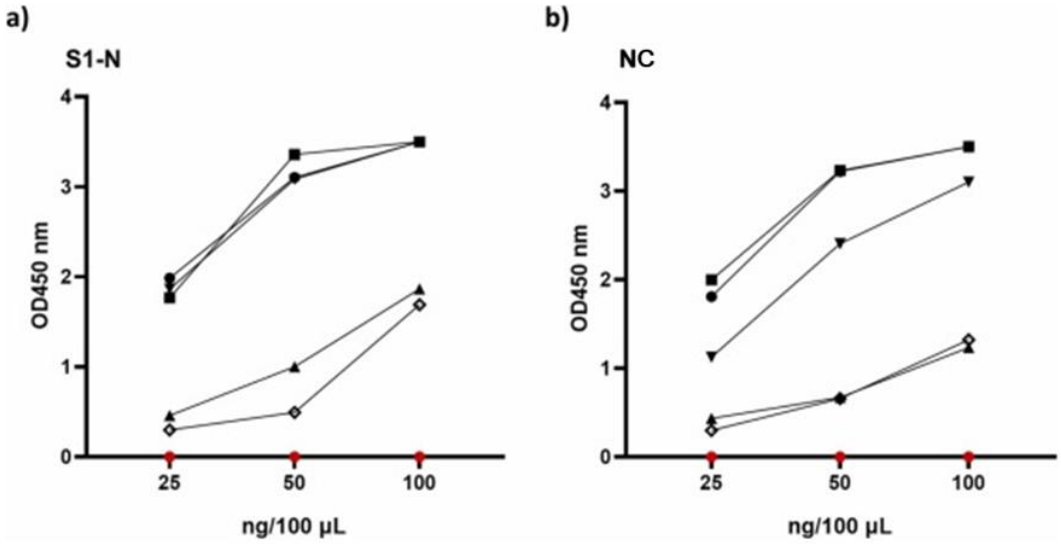


Figure 5. ELISA for detection of antibodies against SARS-CoV-2 in patients' sera using a range of dilution of S1-N (a) and NC (b) proteins as antigens. The same serum was used for both assays with S1-N and NC antigens, as indicated by the same symbol. Positive patients' sera by RT-qPCR (n = 5) were denoted by black symbols, while pre-pandemic sera without prior exposure to SARS-CoV-2 antigens (n = 3), served as negative controls, were denoted by red symbols. Results were reported as absorbance at 450 nm.

4. DISCUSSION

In this study, S1-N and NC proteins were successfully produced in plants of *N. benthamiana* using the PepRSV vector. They were produced as soluble proteins in high yields, and enabled purification by Ni-NTA column. The production of protein was dependent on the construct, as for some constructs, almost no protein was produced. The agro-infiltrated leaves with the constructs of *S1-N_C2*, *S1-N_C3* and *NC_C1* showed the hypersensitive-like programmed cell death response, resulting in almost no production of the target protein. Similar hypersensitive-like reactions on the agroinfiltrated leaves were reported in the interactions of TMV-based vector and *N. tabacum* plants, causing the reduction of recombinant protein productions (LI et al., 2006; LI et al., 2010).

It is speculated that recombinant proteins containing transmembrane or hydrophobic domains caused more hypersensitive-like reactions (LI et al., 2006). It is also reported that some defense response-related genes were upregulated with the hypersensitive-like reactions (LI et al., 2010). In general, the mechanisms responsible for the induction of this hypersensitive-like reaction remains largely unknown. We attempted to produce larger regions or other segments of the S protein as receptor binding domain (RBD) (position 317–544 aa of the S protein), S-1 N + RBD (12–544 aa), RBD+S1-C (311–652 aa) and S2 (888–1212 aa). However, all attempts failed because the PTI reactions resulted in local necrotic lesions, which prevented the production of recombinant protein. We are now investigating how these differences occur, characterizing PTI reactions.

Low or no recombinant protein production by PTI can be caused by the endoplasmic reticulum stress or "unfolded protein response" (UPR). This hypothesis was supported by the observation that co-expression of SARS-CoV-2 proteins with human chaperone proteins calnexin and calreticulin to reduce UPR led to increased production of the S protein, particularly when calreticulin was co-infiltrated, as demonstrated in a study by Margolin et al. (2020) (MARGOLIN; VERBEEK; et al., 2020). In our case, however, co-expression of calnexin and calreticulin did not result in improvement of the protein production (data not shown).

The construct harboring the mGFP gene had a higher NC protein production, and consequently, a high mount of purified protein was obtained. Initially, we expected that the production of a large fusion protein would result in reduction of the protein accumulation by competition in translation, but it was not the case in this study. One possible explanation for the higher protein yield is that the mGFP helps to stabilize the NC protein in the cell. Furthermore,

the co-translation of *mGFP* was particularly useful for monitoring the recombinant protein accumulation in plant, thus, this fusion protein approach proved to be a very useful strategy.

The synthesis of complex glycoproteins that maintain their natural conformations, especially viral envelope glycoproteins, is an obstacle in molecular agriculture (MARGOLIN et al., 2018). This difficulty involves glycosylation events, which in plants differ from mammalian cells (STRASSER, 2016). It is also noted that enveloped virus glycoproteins normally have extensive disulfide bonds, which leads to significant dependence on host chaperones (ALONZI et al., 2017).

Viral envelope glycoproteins tend to be at low levels when plants are used as heterologous systems (KANG et al., 2018; MARGOLIN et al., 2018), which typically results in misfolded recombinant proteins with aberrant glycosylation (STRASSER; ALTMANN; STEINKELLNER, 2014; MARGOLIN et al., 2021, 2022). These patterns are linked to the lack of consistent support observed in plant metabolic pathways for more accurate patterns of glycosylation (MARGOLIN et al., 2021), with consequent folding and processing mediated by chaperones (WILBERS et al., 2016; MARGOLIN et al., 2020), which are points necessary to produce various viral glycoproteins.

We concluded that the S1-N protein, which had both the signal peptide and the KDEL ER retention signal (S1-N_C1), was glycosylated by the endoglycosidase H digestion. This provides evidence that animal virus proteins can be glycosylated in a plant cell system, which opens up new possibilities for producing proteins that are more similar to their native ones. Based on the computational prediction, the protein would contain eight N-glycosylation sites. The produced glycosylated S1-N_C1 was reactive to the positive sera to SARS-CoV-2 by ELISA. This is indicative that the protein folding and glycosylation are enough to be recognized to some extent by immune sera in the context of an ELISA test. Due to the addition of the retention signal, we assume that glycans were added to the protein in ER, but not in Golgi apparatus.

Another evidence was provided by the endoglycosidase H treatment. If the S1-N_C1 protein was modified in the Golgi apparatus, the endoglycosidase H would not be able to remove the complex glycans due to its limited activity for complex glycans. The presence of two protein bands of similar sizes were observed in the protein NC_C2. This was also observed in a previous study using the same plant virus vector (TAVARES-ESASHIKA et al., 2022), when the mGFP containing the N-terminus of PepRSV CP showed double-sized proteins with slightly different molecular masses. We speculate that these double-sized proteins are produced

by two different translation initiations in the same open reading frame. To answer this question, we are currently focusing on this aspect.

In this study, we demonstrated that the tobravirus PepRSV-based vector was useful for production of recombinant protein of pharmaceutical use. This vector is small (only the RNA2), making it easy for molecular cloning steps. To improve this vector system, we are currently studying the plant innate immunity induced by the production of the recombinant proteins.

5. APPENDIX

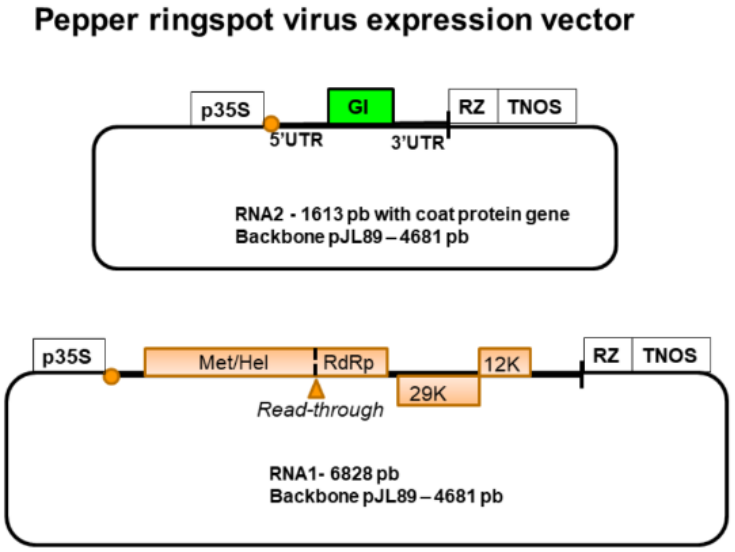


Figure S1. Genetic map of the pepper ringspot virus vector in the plasmid pJL89. p35S, Cauliflower mosaic virus 35S protein promoter; GI, Gene of interest (changed from coat protein gene); UTR, untranslated region; RZ, ribozyme of hepatitis D virus; TNOS, terminator of Nopaline synthase; Met/Hel, methyltransferase and Helicase; RdRp, RNA-dependent RNA polymerase; 29K, 29K protein; 12K, 12K protein.

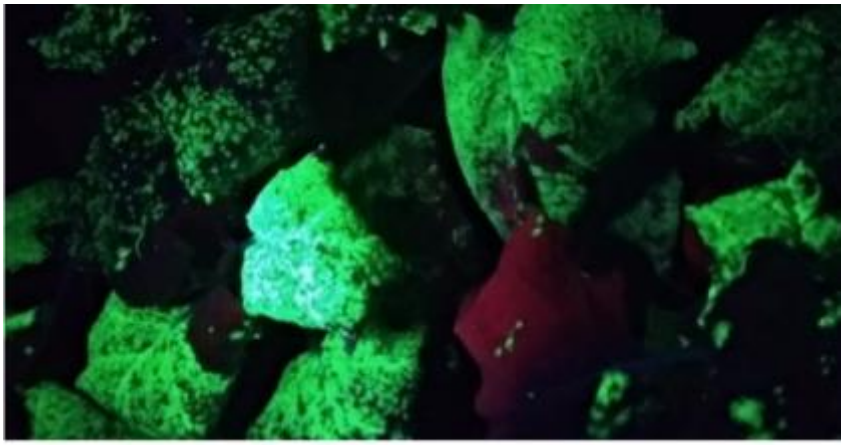


Figure S2. mGFP signals observed as the indicator of NC protein production in agro-infiltrated *N. benthamiana* plants. *N. benthamiana* plants showing GFP fluorescence under UV irradiation at four days post-agroinfiltration with the N-C3 construct.

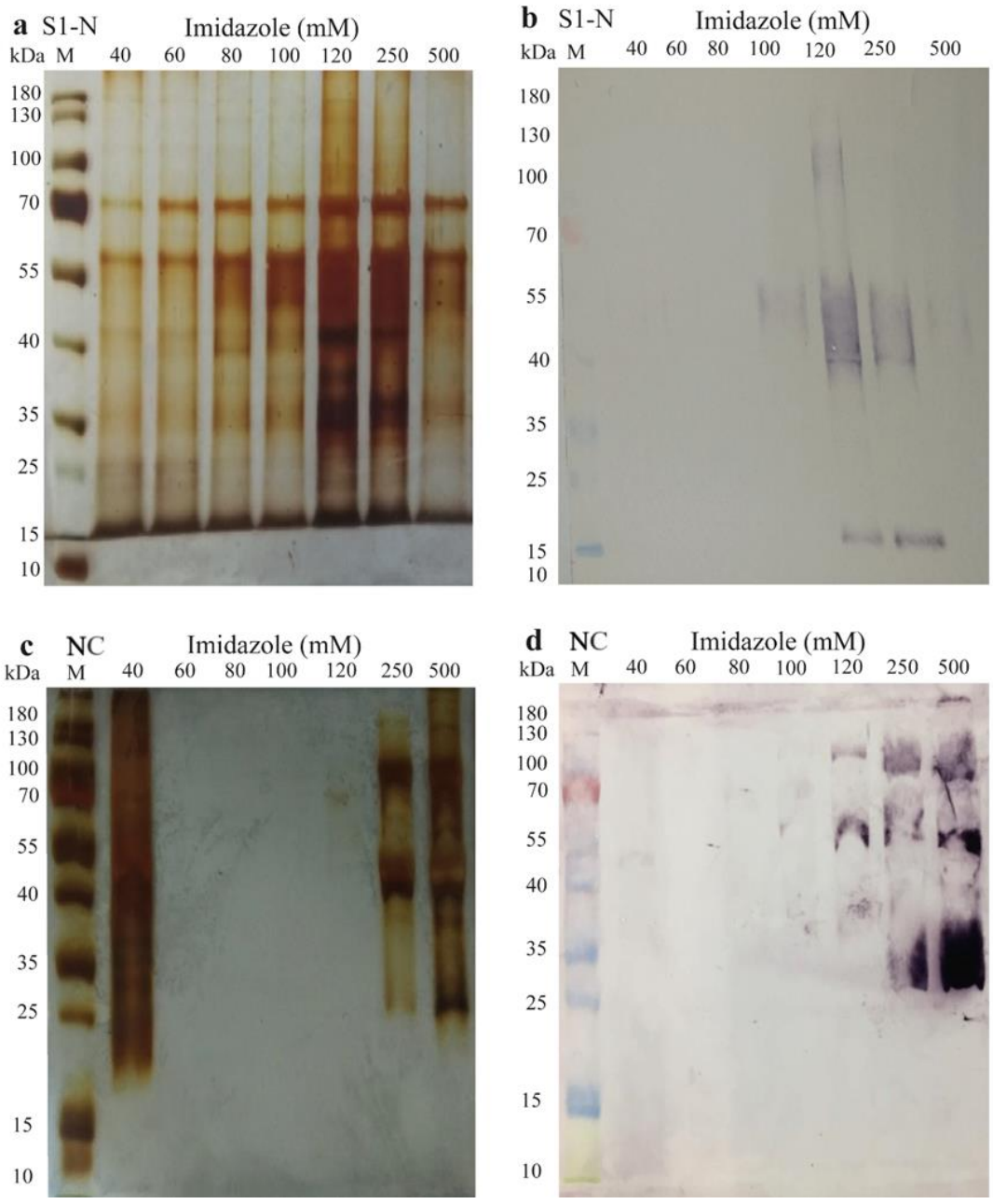


Figure S3. Elution profile by series of imidazole concentration washing. Silver stained 12% SDS-PAGE profile of S1-N (a) and NC (c) proteins using washing or elution buffer containing from 40 to 500 mM imidazole. Western blotting (anti His-tag) of S1-N (b), and NC proteins (d). PageRuler™ Prestained Protein Ladder was used as the protein mass standard.

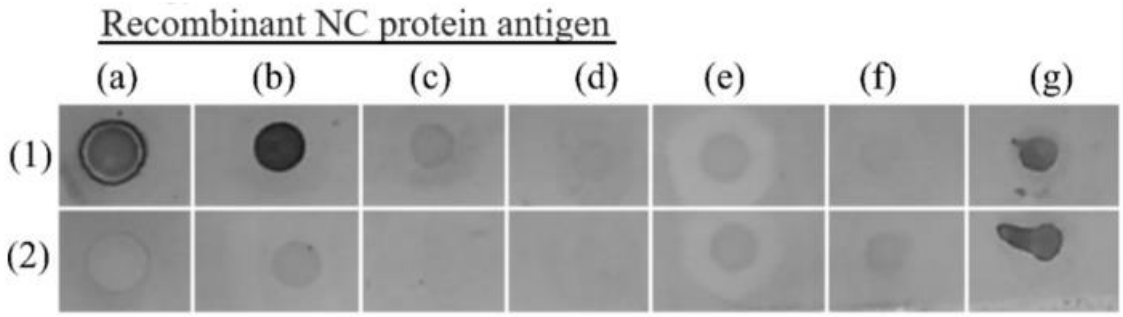


Figure S4. Dot blot immunobinding assay using the NC protein as an antigen against positive SARS-CoV-2 patient serum (1) and negative SARS-CoV-2 patient sera (2). Columns (a) to (d) represents a serial dilution of N protein: (a) 1 µg, (b) 0.1 µg, (c) 0.01 µg, and (d) 0.001 µg. Columns (e) and (f) are non-reactive controls, semi-purified proteins, and crude plant extract from non-agroinfiltred *N. benthamiana*, respectively. Column (g) represents *E. coli* whole cell lysate. Human sera normally have reactive antibodies to *E. coli* proteins, serving as positive control for SARS-CoV-2 negative serum.

Table S1. Primers used in this study

Primer name	Sequence
S1-N C1 For	CTGCAATCACCTGCGCAAAATGAAAACCAACTTGTTCTTGTTTTTGATCTTCTCATT GTTGTTGTCATTGTCATCTGCCGAGTTCGTGAATTTGACTACAAGAACCCA
S1-N C1 Rev	AGCTCCACTTGGGGCAGATTAAAGTTCATCCTTATGATGGTGGTGATGATGAGAA GTCTGATAGATCCCCT
Backbone S1-N C1 For	TCTGCCCCAAGTGGAGCT
Backbone S1-N C1 Rev	TTGCGCAGGTGATTGCAG
S1-N C2 For	GATGTTACTACTAGGTTGAGAGCGGTGAATTTGACTACAAGAACCC
S1-N C2 Rev	AGCTCCACTTGGGGCAGACTAATGATGGTGGTGATGATGAGAAGTCTGATAGATC CCCTT
Backbone S1-N C2 For	TCTGCCCCAAGTGGAGCT
Backbone S1-N C2 Rev	CGCTCTCAACCTAGTAGTAACATC
S1-N C3 For	AAGTTCTGCCGGTGCGATTGGTTCTGGTGAAGGTAGAGGTTCTCTTCTTACTTGTG GTGATGTTGAAGAAAATCCTGGTCCTGTGAATTTGACTACAAGAACCCA
S1-N C3 Rev	GTCCGTAAAGCAGACAACTCTTAATGATGGTGGTGATGATGAGAAGTCTGATAGA TCCCCTT
Backbone S1-N C3 For	GAGTTGTCTGCTTTACGGAC
Backbone S1-N C3 Rev	AATCGCACCGGCAGAACTT
NC C1 For	TTGTCATCTGCCGAGTTCTCTAAACAAAGAAGACCACAA
NC C1 Rev	ATGATGGTGGTGATGATGATCATCCAAATCGGCAGC
Backbone NC C1 For	CATCATCACCACCATCAT
Backbone NC C1 Rev	GAACTCGGCAGATGACAA
NC C2 For	TTACTACTAGGTTGAGAGCGTCTAAACAAAGAAGACCACAAG
NC C2 Rev	ATGATGGTGGTGATGATGATCATCCAAATCGGCAGC
Backbone NC C2 For	CATCATCACCACCATCAT
Backbone NC C 2 Rev	CGCTCTCAACCTAGTAGTAA
NC C3 For	TTGAAGAAAATCCTGGTCCTTCTAAACAAAGAAGACCACAAG
NC C3 Rev	ATGATGGTGGTGATGATGATCATCCAAATCGGCAGC
Backbone NC C3 For	CATCATCACCACCATCAT
Backbone NC C3 Rev	AGGACCAGGATTTTCTTCAA

CAPÍTULO III. *Metagenomic approach to genetic sequence coronaviruses in untreated sewage water*

O presente capítulo é referente ao texto direcionado à *short communication* em produção para posterior submissão à publicação no periódico ***Brazilian Journal of Microbiology***.

1 INTRODUCTION

SARS-CoV-2 is a positive sense RNA virus belonging to the family *Coronaviridae*, genus *Betacoronavirus* (HU et al., 2021; DHAMA et al., 2023). The agent is initially considered a respiratory pathogen; however, it can also cause gastrointestinal complications (ZERBATO et al., 2021). It is believed that gastrointestinal pathogenesis is due to a greater migration of the virus present in the circulative system, resulting from an increase in lung permeability caused by the cytokine storm (AHLAWAT; ASHA; SHARMA, 2020). In this situation, the virus is stimulated to bind to ACE-2 (Angiotensin Converting Enzyme 2) receptors located on various cell types, such as enterocytes (SAJDEL-SULKOWSKA, 2021).

An immune response in the intestinal region can cause dysbiosis, which alters the dynamics of the microbiota and promotes the dispersion of bacterial agents and toxins, reinforcing gastrointestinal symptoms (GUAN et al., 2020; WANG et al., 2020). This type of pathogenesis increases the presence of the viral RNA in stool samples from patients with COVID-19 (SYED et al., 2020), even in situations of negative detection through analysis of respiratory tract samples (XIAO; SUN; et al., 2020; WANG, D. et al., 2020; ZHANG, W. et al., 2020).

Viruses characterized by fecal-oral transmission or excretion through fecal matter can be more effectively identified using epidemiological techniques centered on the analysis of sewage wastewater (GUAJARDO-LEIVA et al., 2020; CHILD et al., 2023). These analyzes use treated and/or untreated wastewater as a procedure aimed at epidemiological monitoring of human pathogens in each community (AMOAH; KUMARI; BUX, 2020; KLAPSA et al., 2022). Such analyzes are used to check possible outbreaks (MANOR et al., 2014). Due to the COVID-19 pandemic, wastewater has been used to monitor the spread of viral agents such as SARS-CoV-2 and enteric pathogens (CHILD et al., 2023; WOLFE et al., 2022), in addition to verify the occurrence of antimicrobial resistance genes in bacterial populations (NAQUIN et

al., 2015; HENDRIKSEN et al., 2019; LARSSON; FLACH; LAXMINARAYAN, 2023). Due to the pandemic, research for SARS-CoV-2 in wastewater also works as a form of early notification, which helps detect the pathogen before COVID-19 cases appear in the locality, allowing monitoring of the spread of the disease (PEINADO et al., 2022).

This study aimed to verify the analysis of WBE for detection SARS-CoV-2 and related viruses in raw sewage collected at the North Sewage Treatment Station of the Environmental Sanitation Company of the Federal District (CAESB) in Brasília, at two moments: (1) before the pandemics in 2016; and (2) during the pandemics in 2020. Furthermore, we analyzed the presence of non-SARS-CoV-2 coronaviruses in 2020 through HTS. To date, this is the first report on the identification of coronavirus in sewage in Brasília.

2 MATERIAL AND METHODS

2.1 Sample collection, preparation and sequencing

A wastewater sample (one liter of untreated sewage water) was collected in March (sample E163) and May 2016 (sample E165), and in May (sample E205) and August (sample E208) 2020, at the North Sewage Treatment Station (ETE-Norte) of CAESB in the Asa Norte region, in Brasília, the Federal District. From the time of collection until processing, samples were stored in refrigerated sterile glass containers.

For each sample, preparations enriched with virus particles were produced based on the following procedure: collected water was aliquoted in eight polypropylene tubes (50 mL), total of 400 mL; then subsequently centrifuged at low speed (5000 × g) for 20 minutes, to remove debris and microorganisms. The supernatant (50 mL per tube) was transferred to six tubes with 16 mL of 20% sucrose cushion and subjected to ultracentrifugation (140,000 × g) (Rotor 45Ti) (Beckman Coulter, Brea, USA) for 75 minutes.

The precipitate was resuspended with lysis buffer of the ZR Soil/Fecal RNA MicroPrep kit (Zymo Research, Irvaine, USA) and total RNA was extracted according to the manufacturer's instructions. The RNA samples were dried in the RNAstable tubes (Biomatrica, USA) and sent for high-throughput sequencing (Macrogen Inc.) (Seoul, South Korea), using the Illumina Novaseq 6000, with 100 pb-piared end 5G scale.

2.2 Sequence Analyses

Data analysis was performed using computers with an Intel (R) Xeon (R) E5-2620 CPU with 98GB of RAM. The sequence reads were trimmed with BBDuk software (https://sourceforge.net/projects/bbmap/), and the contigs assembled with MEGAHIT (LI et al., 2016) with k (k-mer) values of 21 and 71, and analyzed using tBLASTX using the RefSeq viral databases (BRISTER et al., 2015) retrieved in January 2021 (https://www.ncbi.nlm.nih.gov/genome/viruses/). Contigs were imported into Geneious R.8.1 software (Biomatters, New Zealand) for genome assembly. Kaiju read mapper program was used for taxonomic classification of reads (MENZEL; NG; KROGH, 2016) performed with default parameters (https://github.com/bioinformatics-centre/kaiju.git) using a Viral Genome RefSeq database and visualized by Krona viewer (https://github.com/marbl/Krona/wiki/Krona Tools).

2.3 Detection of SARS-CoV-2 in sewage water

In order to corroborate the identification of SARS-CoV-2 in sewage water, pathogens that could be identified in this environment were also selected as control species. The first corresponded to human norovirus *G2* (NoV GII) (EFTIM et al., 2017), agent related to the development of gastrointestinal diseases (PLAZA-GARRIDO et al., 2023). In turn, the second element was pepper mild mottle virus (PMMoV), an important plant virus presents in human feces (ZHANG, T. et al., 2006) and a valid viral indicator targeting fecal contamination in urban sewage samples (KITAJIMA; SASSI; TORREY, 2018; BONANNO FERRARO et al., 2021).

For confirmation of the presence of SARS-CoV-2 in the sewage water sample, total RNA preparations from the 2020 collections were analyzed in duplicate by RT-qPCR using the GoTaq® Probe 1-Step RT-qPCR System kit (Promega, Madison, USA). Detection was directed to the N gene, according to the protocol recommended by the Centers for Disease Control and Prevention (CDC) (LU et al., 2020). The RT-qPCR procedure was performed on the QuantStudio TM 5 Real-Time PCR system (Thermo Fisher Scientific, Waltham, USA) under the conditions above: reverse transcription for 15 minutes at 45 °C, reverse transcriptase inactivation and DNA polymerase activation GoTaq® for 2 minutes at 95°C and 40 cycles of denaturation (15 seconds at 95°C) and trapping and extension (1 minute at 60°C). For viral quantification, the standard curve was performed using serial dilutions of the positive control,

with subsequent amplification of the N gene. The positive control corresponded to the construct Bac-N-CoV2, a bacmid with insertion of the SARS-CoV-2 N gene, kindly provided by Dr. Bergmmam Morais Ribeiro (Baculovirus Laboratory - UnB).

3. RESULTS

3.1 Virus species identified in wastewater samples

Sewage water samples were collected four times, two in 2016 and other two in 2020. Total RNA was subsequently extracted, and sequencing was done via HTS. **Table 1** shows the number of raw and processed reads using the BBDuk software, and the contigs formed using the MEGAHIT tool.

Table 1. Raw and processed dataset

Data Set	Raw Data (bp)	Data Processed (pb)	Contigs
E163	55.495.060	55.160,880	130.972
E165	52.780.636	52.316,276	131.962
E205	57.649.960	56.942,430	127.475
E208	73.676.392	72.876,508	148.051

Regarding the taxonomic results, initially, analyzes with Kaiju read mapper and BLAST showed that in sample E163 (March 2016) 557,487 reads (out of 55,160,880 reads) and 7,223 contigs (out of 130,972 contigs) were classified as viruses. For E165 (May 2016) 478,410 reads (out of 52,780,636 reads) and 6,785 contigs (out of 131,962 contigs) were associated with viruses. Sample E205 (May 2020) presented 520,421 reads (out of 57,649,960 reads) and 8,003 contigs (out of 127,475 contigs) related to viruses. Finally, in E208 (August 2020) 767,850 reads (out of 73,676,392 reads) and 9,784 contigs (out of 148,051 contigs) linked to the virus were observed.

Subsequently, the results obtained by Kaiju showed that in sample E163, 27 reads (0.005%) corresponded to the family *Coronaviridae*, and 24 out of 27 reads were of *Orthocoronavirinae* sub-family. Of the 24 reads, it was observed that 23 reads were linked to the genus *Alphacoronavirus*, with mention of the species *Alphacoronavirus 1* 17 reads), human coronavirus (HCoV) NL63 (3 reads) and bat coronavirus (BatCoV) CDPHE15 (1 read). The remaining sequences were not associated with specific taxa. In *Alphacoronavirus 1* species group, 10 out of 17 reads were identified as feline infectious peritonitis virus (FIPV).

Furthermore, the species swine enteric coronavirus (SeCoV) (1 read) and transmissible gastroenteritis virus (TGEV) (1 read) were also identified. In relation to *Gammacoronavirus*, this genus was represented only by the species avian coronavirus (AvCov) (1 read), specifically the species infectious bronchitis virus (IBV) that infects birds.

In sample E165, 71 reads (0.02%) corresponded to *Coronaviridae* and classified in the subfamily *Orthocoronavirinae*. Of these, 66 reads were classified to the genus *Alphacoronavirus*, with mention of the species *Alphacoronavirus 1* (0.01%, 61 reads). 16 out of 61 reads were identified as FIPV, 2 reads were of FCoV (0.02%) and 1 read was of BatCoV CDPHE15. Furthermore, the species SeCoV (3 reads) and TGEV (6 reads) were identified. Still belonging to this group, 36 reads couldn´t be classified to any specific taxa. Four reads in E165 sample were linked to the genus *Betacoronavirus*. Two reads were classified as *Betacoronavirus 1*, 1 read as BatCoV *Pipistrellus* HKU5 and 1 read as bat Hp- betacoronavirus Zhejiang2013. Within the group of species related to *Betacoronavirus 1*, the 2 reads identified as HCoV OC43, a species that causes respiratory problems in humans. *Deltacoronavirus* presented only one read, but this read couldn´t be classified as specific species.

In sample E163, 416 reads (0.08%) corresponded to NoV genogroup I (NoV GI); 205 reads (0.04%) were equivalent to NoV GII, more specifically to the genotypes GII.17 (82 reads, 0.02%) and GII.2 (1 read, 0.002%). Regarding contigs, the analyzes indicated 3 (0.04%) to NoV GI; and 12 (0.09%) linked to NoV GII, of which 6 corresponded to GII.17 and the remainder were not associated with a specific genotype. E_165 presented 28 reads (0.006%) corresponding to NoV GI; 297 reads (0.06%) were equivalent to NoV GII, more specifically to the genotypes GII.17 (8 reads, 0.002%) and GII.2 (1 read, 0.002%), the remainder were not associated with a specific genotype. Both NoV GIII (NoV GIII) and NoV GIV genotype I (NoV GIV.1) presented 1 read (0.002%). NoV GV was linked to 3 reads (0.006%). Among the contigs, the results indicated 6 (0.09%) to NoV GII without mentioning the specific genotype; and 1 (0.02%) linked to NoV GV (NoV GV).

In sample E205, 79 reads were identified as to the reads belonged to *Coronaviridae*, and among these, 3 reads were not classified to species. Thus, 76 reads were classified as *Orthocoronavirinae* subfamily. Out of 76 reads, 57 reads were classified in the *Alphacoronavirus* genus, and among these, 3 reads were also not taxonomically classified. Among *Alphacoronavirus*, the species *Alphacoronavirus 1* (51 reads), 25 reads were classified as FIPV, 2 reads to HCoV NL63, and one read to FCoV. Three 3 reads were not classified. Furthermore, two reads were classified as SeCoV and two reads to TGEV. Still belonging to

this group, 22 reads were not associated with any specific species.

In the E205 sample, 18 reads were classified in the genus *Betacoronavirus*. Out of 18 reads, 12 reads were classified of SARS-CoV-2, One of SARS-CoV Tor2 and five reads were not associated to any specific species. *Deltacoronavirus* presented only a single read, however, it did not present reference to a specific species.

E_205 presented 9 reads (0.002%) corresponding to NoV GI; 310 reads (0.06%) were equivalent to NoV GII without mention of a specific genotype. NoV GIV.1 presented 1 read (0.002%). Among the contigs, the results indicated 6 (0.09%) to NoV GII without mentioning the specific genotype. E_208 presented 221 reads (0.03%) corresponding to NoV GI; 15 reads (0.002%) were equivalent to NoV GII, more specifically to GII.17 genotypes (7 reads, 0.0009%), the remainder were not associated with a specific genotype. NoV GIV (NoV GIV) presented 3 reads (0.0004%), and NoV GV was linked to 2 reads (0.0003%). Among the contigs, the results indicated 5 (0.06%) to NoV GI without mentioning the specific genotype.

In sample E208, 89 reads corresponded to *Coronaviridae*, and among these, 7 reads were not taxonomically classified. Thus, 79 reads were directly classified to the *Orthocoronavirinae* subfamily. Out of these 79 reads, 40 reads were classified to the *Alphacoronavirus* genus. Among alphacoronavirus, 32 reads were classified to *Alphacoronavirus 1*, 5 reads to ferret coronavirus, 1 read to BatCoV CDPHE15 (0.0001%, 1 read) and one read to *Myotis ricketti* alphacoronavirus Sax-2011. The Wencheng Sm shrew coronavirus species was also identified (1 read), in turn, it does not yet have a classification in the National Center for Biotechnology Information (NCBI) database.

In the E208 sample, a total of 37 reads were classified to the *Betacoronavirus* genus. Out of 37 reads, 24 reads were of SARS-CoV-2 and 6 reads were of Betacoronavirus *1*. The species rabbit coronavirus HKU14 (1 read) was also identified. The remaining sequences were not associated with specific species. Within Betacoronavirus 1, 1 read was identified as HCoV OC43, 2 reads to *Bovine coronavirus*, and the rest were not classified. Out of 37 reads of *Betacoronavirus* genus, 3 reads were classified to BatCoV and 2 reads to Shrew coronavirus.

According to the International Committee on Taxonomy of Viruses (ICTV), a genomic sequence corresponds to a coronavirus when it shares 90% amino acid sequence identity in the conserved domains of the replicase. This limit corresponds to the only species demarcation criterion (DE SABATO et al., 2019; GORBALENYA et al., 2020; LAU et al., 2020). Through studying the four datasets, tBLASTx analysis identified 60 coronavirus species (**Figure 6**) (**Table S2, S3, S4 and S5**). Of these, the majority are classified among the described genera

(*Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*), and a small portion of unclassified sequences. Among the unclassified ones, due to some issue of subfamily nomenclature linked to the database, the species were separated into unclassified *Coronavinae* and *Orthocoronavirnae*. So far, according to ICTV regulations, *Orthocornavirnae* is used as a subfamily (JACOB MACHADO et al., 2021).

Species that affect animals such as Ferret coronavirus, SeCoV, TGEV, bovine coronavirus (BCoV), FIPV were identified, however, a constant pattern was not observed, so only TGEV was present in all samples. In addition, the species HCoV-229E (E163, E208), HCoV-NL63 (E205), HCoV-OC43 (E165) and HCoV-HKU-1 that affect humans have also been identified. Only the HCoV-HKU-1 species was present in all four datasets.

Given these allied factors and to the low identity of hits observed by tBLASTx and the presence of stop-codon (*) throughout the sequences (data not shown). A new analysis was performed, now with BlastN (**Figure 7**) **(Table S6, S8, S10 and S11)** and BlastX (**Figure 8**) **(Table S7, S9, S12 and S13)** with the contigs from MEGAHIT, using nr/nt (non-redundant protein sequence) database, however, was restricted to sequences that did not present stop-codons. Figures 7 and 8 indicate that the majority of sequences classified in tBLASTx as viruses are actually directed to bacteria and unclassified sequences. This characteristic observed in the four samples collected is consistent with false positive findings.

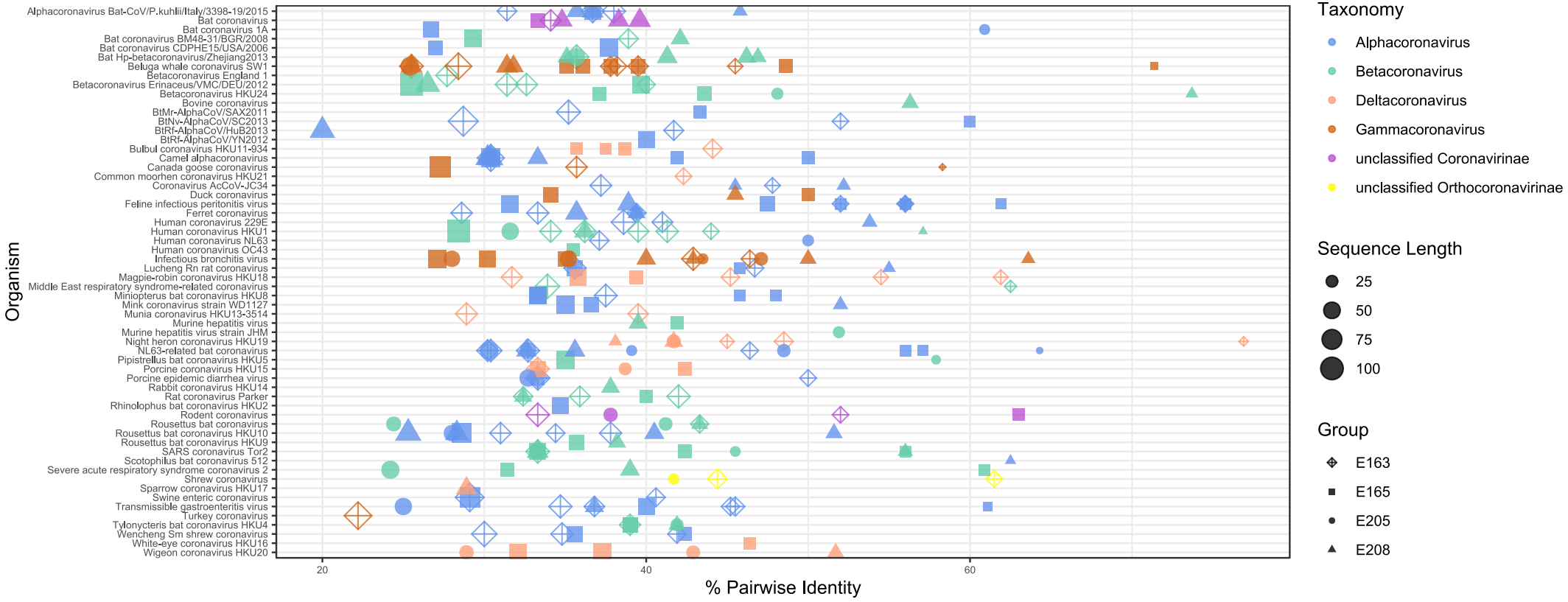


Figure 6. Comparison between nucleotide identity and sequence size of *Coronaviridae* identified via TBlastX. The figure corresponds to a bubble heatmap where the nucleotide identity (in percentage) of the sixty (60) coronavirus species is observed in line with their respective sequence fragments. The samples (groups) are represented as geometric shapes, being E163 (unfilled square with internal lines), E165 (filled square), E205 (circle) and E208 (triangle). The genera are represented by different colors, Alphacoronavirus (blue), Betacoronavirus (green), Deltacoronavirus (pink) and Gammacoronavirus (brown), unclassified Coronavirinae (purple) and unclassified Orthocoronavirinae (yellow). The size of the observed forms is proportional to the sequence size, ranging from 25 to 100 nucleotides (nt). Most of the nucleotide identities observed ranged between 30 and 50%, in line with reads with larger sequence sizes.

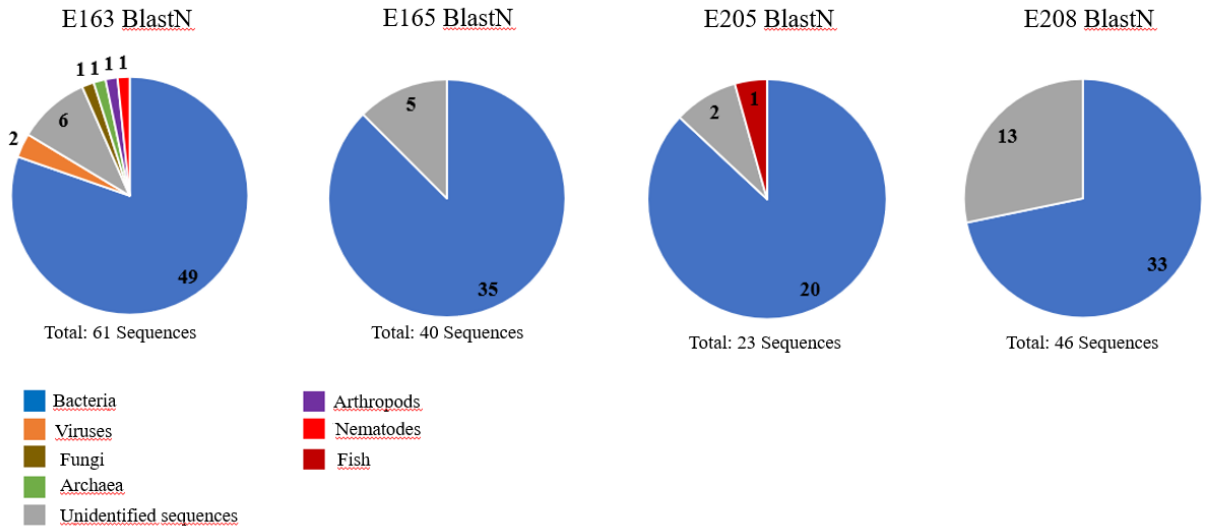


Figure 7. Sequence analysis with the BlastN tool from data obtained initial analysis with tBlastX. The sequences obtained via initial analysis with tBlastX were analyzed again via BlastN for data checking. It is observed that most of the sequences are strongly associated with bacteria, followed by sequences that did not show any similarity with others available in the database. The resulting sequences were selected due to the lowest E value, as listed in the supplementary tables.

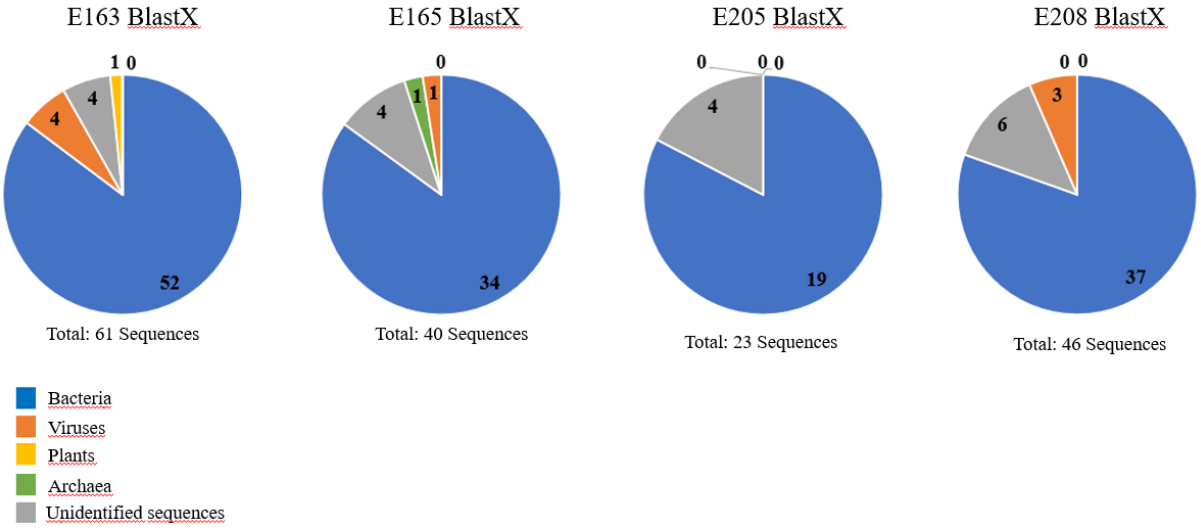


Figure 8. Sequence analysis with the BlastX tool of data obtained from initial analysis with tBlastX. The sequences obtained via initial analysis with tBlastX were analyzed again via BlastX for data checking. It is observed that most of the sequences are strongly associated with bacteria, followed by sequences that did not show any similarity with others available in the database. The resulting sequences were selected due to the lowest E value, as listed in the supplementary tables.

3.2 RT-qPCR analyses of viral material in sewage water

The detection of SARS-CoV-2 by RT-qPCR was carried out using samples E_163, E_165, E_205 and E_208. For greater robustness of the results, the samples were prepared in duplicate and thus the analyzed data corresponded to E163_1 and E163_2, E165_1 and E165_2, E205_1 and E205_2, and finally, E208_1 and E208_2. To carry out RT-qPCR and subsequent preparation of the standard curve, PMMoV, NoV GII and the bacmid (Bac-N-CoV) were considered as control samples. **Table 2** presents the equations and R^2 values that corresponded to 0.964 (Bac-N-CoV), 0.999 (PMMoV) and 0.997 (NoV GII). The CT (Cycle threshold) values found are shown in **table 3**. In this condition of RT-qPCR, Ct value below 35 is not considered as positive. About SARS-CoV-2, in the samples in 2016 (E133 and E165) were considered as negative and in the samples in 2020 (E205 and E208), SARS-CoV-2 were positive. The mean Ct value of E205 for SARS-CoV-2 was 30.476. Using the equation (Table 2), the x value for the Ct of 30.476 was -0.1397, showing 0.691 pg of the standard DNA (Bac-N-CoV). This value corresponds to the bacmid with insert (about 137k bps). SARS-CoV-2 has 30 kb, so Ct = 30.476 of SARS-CoV-2 corresponds to 0.1513 pg. Then, the copy number of SARS-CoV-2 with Ct = 30.476 was calculated as 4931 copies per µL of sample. The calculation of E208 (Ct = 31.364) was 0.1033 pg, then, 3367 copies per µL of sample.

Table 2. Equations for calculating the concentration of genetic material in wastewater

Control	**Equation**	**Value of R^2**
Bac-N-CoV	y = -4,7597x + 29,81	0,964
PMMoV	y = -2,6854x + 25,657	0,999
NoV GII	y = -2,7631 + 22,881	0,997

Table 3. CT Values of Viral Species Analyzed in RT-qPCR

		Control Species	
Sample	SARS-CoV-2 CT	PMMoV CT	NoV GII CT
E163_1	36,546	---	36,078
E163_2	---	---	35,602
E165_1	---	26,902	30,074
E165_2	36,543	27,329	29,951
E205_1	30,510	25,684	30,972
E205_2	30,441	30,709	30,709
E208_1	31,364	25,311	33,284
E208_2	31,364	25,040	32,434

4. DISCUSSION

Coronaviruses are important RNA viruses that infect humans and animals. The presence of these in wastewater may be the result of the release of viral material into aquatic ecosystems, especially in areas with low access to basic sanitation (WARTECKI; RZYMSKI, 2020). In water environments, the survival of these viruses is normally dependent on water temperature; availability of light; level of organic matter; and predation (PRATELLI, 2008; AUSAR et al., 2013; PRUSSIN et al., 2018). And in general, the viability of coronaviruses in wastewater is lower than in other water ecosystems because the envelop of the virion is not resistant for such condition (WARTECKI; RZYMSKI, 2020).

In this study, reads associated with SARS-CoV-2 and other coronaviruses were identified through HTS, in addition to quantifying the number of copies per mL of SARS-CoV-2 was shown. This result was also presented in other works (AHMED et al., 2020; LA ROSA et al., 2020; PRADO et al., 2020; FONGARO et al., 2021; MOTA et al., 2021; BAR-OR et al., 2022). The methods observed in the literature for quantifying SARS-CoV-2 were adapted from techniques aimed at detecting non-enveloped enteric viruses. Due to differences in the structure of the viral particle, the efficiency in identifying the SARS-CoV-2 viral RNA may differ from that of other species (AHMED et al., 2020). Furthermore, the way in which sewage water is processed has been identified as one of the main barriers to the quantification of SARS-CoV-2 (DE ARAUJO et al., 2021). For example, Bar-or and collaborators (2022) carried out tests with different processing techniques and RNA extraction kits with the aim of improving CT values, considering the N and E gene of SARS-CoV-2 as targets.

Metagenome techniques applied to wastewater samples provide new perspectives in the field of virology. One of reports describes about better understanding of the viral sequences present in sewage and how they can affect human health (SCHAEFFER et al., 2023). In virome studies in this environment, most abundant viruses are phages, followed by plant, animal and human viruses (FERNANDEZ-CASSI et al., 2018; MARTÍNEZ-PUCHOL et al., 2020). Through metagenome results, reads associated with *Alphacoronavirus 1* were identified, especially the FCoV species; *Betacoronavirus 1*, in particular HCoV-OC43; and SARS-CoV-2, which was also confirmed by Martínez-Punchol and collaborators (2021) (MARTÍNEZ-PUCHOL et al., 2021). Unlike the study, reads associated with *Lucheng Rn rat coronavirus* (LRNV) and *Canine coronavirus* (CCoV) were not found.

The presence of reads associated with *Alphacoronavirus 1* in sewage water environments is also reinforced by other studies (BLANCO et al., 2019; LA ROSA;

BONADONNA; et al., 2020). FCoV was found in all samples. Normally, the species is of veterinary importance as it affects domestic cats and stimulates enteric complications (GUNDY; GERBA; PEPPER, 2009; DELAPLACE et al., 2021). *Alphacoronavirus 1* and *Betacoronavirus 1* cover species that affect humans and cause simple colds to more serious complications. Reads associated with HCoV-OC43 were found in E165, which was also confirmed in other studies (CHILD et al., 2023; MARTÍNEZ-PUCHOL et al., 2021). The species is most common during winter (VAN DER HOEK, 2007) and it is noteworthy that rodents are considered natural hosts, while bovine are intermediaries (YE et al., 2020). The identification of the other human respiratory coronaviruses HCoV-229E, HCoV-HKU-1 and HCoV-NL63 in the samples is also consistent with the others studies (GUNDY; GERBA; PEPPER, 2009; BIBBY; VIAU; PECCIA, 2011; MCCALL et al., 2020; CHILD et al., 2023). HCoV NL63 were found in set E208. Literature focused on this species in Brazil is scarcer (CABEÇA; BELLEI, 2012; CABEÇA; GRANATO; BELLEI, 2013), in line with its characterization in residual samples.

In the country, most of the studies analyzing wastewater linked to SARS-CoV-2 are mainly concerned with quantifying the species in that environment, as for example in the work of Fongaro and collaborators (2021) (FONGARO et al., 2021). Therefore, few detail the presence of other coronavirus species, except SARS-CoV-2 (MARTIN et al., 2020; ELSAMADONY et al., 2021; MORA et al., 2021; SCHAEFFER et al., 2023). Most commo studies on wastewater virome (CANTALUPO et al., 2011; FERNANDEZ-CASSI et al., 2018; GUAJARDO-LEIVA et al., 2020; ROTHMAN et al., 2021) discuss the presence of taxa commonly present in wastewater, in addition to the dominance of bacterial sequences and plant viruses (DUARTE et al., 2023).

Regarding RT-qPCR analyses, the SARS-CoV-2 genome could be detected in studies in the samples of 2020. Such early detection of SARS-CoV-2 in sewage water were reported in countries such as Australia (AHMED et al., 2020), United States (WU, F. et al., 2020), and Holand (MEDEMA et al., 2020). In Brazil, the SARS-C-V-2 analyses in sewage water were carried out in the states of Santa Catarina (FONGARO et al., 2021), Rio de Janeiro (PRADO et al., 2020), São Paulo (CLARO et al., 2021).

Another relevant virus identified via Kaiju and BLAST analyses in all samples was Norwalk virus (NoV). This virus belongs to the genus *Norovirus* (family *Caliciviridae*), which includes positive-sense RNA viruses and promotes episodes of acute gastroenteritis (GUAJARDO-LEIVA et al., 2020; PLAZA-GARRIDO et al., 2023). By analyzing the amino

acids of the viral capsid protein (VP1), these viruses are classified into seven genogroups (GI to GVII) (KRONEMAN et al., 2013).

The identification of NoV GII is in line with other studies, as this genogroup is frequently identified in metagenomic studies has a global prevalence in sewage water (FARKAS et al., 2018; FUMIAN et al., 2019), in addition to being linked to sporadic outbreaks of gastroenteritis (RAJKO-NENOW et al., 2013; CORNEJO-SÁNCHEZ et al., 2023).

PMMoV copy number quantification via RT-qPCR occurred only in E_208, while the identification of reads and contigs occurred in all datasets. So, in E163, 2093 reads (0.4%) and 4 contigs (0.06%) were found; E_165 presented 1,725 reads (0.4%) and 8 contigs (0.1%), and among the samples, E_165 presented the largest number of contigs. E_205 presented 2,938 reads (0.6%) and 3 contigs (0.04%). E_208 presented 12,923 reads (2%), being the largest quantity among the analyzed samples, while 2 contigs were observed (0.02%). This finding is in line with other previously published work. (BAČNIK et al., 2020; DUARTE et al., 2023; GYAWALI et al., 2019; ROSARIO et al., 2009). In this study, we sought to demonstrate the possibility of next-generation sequencing tools related to the monitoring of viruses belonging to *Coronaviridae*. The identification of these viruses can help in understanding viral dynamics in each community, given the possibility of monitoring this virus family with possible seasonal behavior.

5. APPENDIX

Table S2. Analysis of E163 data using tBLASTX

Alphacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_002306	56.0%	1.16e-02	k119_1855	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGKSIKEARDRAYALCG
NC_002306	56.0%	1.61e-02	k119_59380	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGKSIKEARDRAYALCG
NC_002306	52.0%	3.89e-02	k119_27145	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGKSIKEARDRAYTLCG
NC_028833	52.0%	8.81e-02	k119_90324	*BtNv-AlphaCoV/SC2013*	*Alphacoronavirus*	*Nyctacovirus*	VVTIHCCHVLVFTNILITKHTITGT
NC_003436	50.0%	3.10e-02	k119_41384	*Porcine epidemic diarrhea virus*	*Alphacoronavirus*	*Pedacovirus*	PAQPVDQTVITCGKITQFTAAQHNPL
NC_034972	47.8%	9.76e-02	k119_63042	*Coronavirus AcCoV-JC34*	*Alphacoronavirus*	*Luchacovirus*	ICLAKFCNFACFFQCNSHCVFKV
NC_032730	46.7%	7.05e-02	k119_128631	*Lucheng Rn rat coronavirus*	*Alphacoronavirus*	*Luchacovirus*	SLRGGVGSLVDIYLYPIPCCSRLPATSFTR
NC_048216	46.4%	8.39e-02	k119_124653	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	VIEYKNKPWTIIKTNFVKPGKGGAFMQV
NC_038861	45.5%	3.36e-02	k119_106652	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	LSTSVVFPWSTWAMIAMFRIFCI*YLLFG CKGT
NC_038861	45.2%	3.13e-02	k119_47935	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	YCWRHQYQ*SIRKSKKIFWRASSRTSCN KIH
NC_038861	41.9%	3.36e-02	k119_99232	Wencheng Sm shrew coronavirus	*Alphacoronavirus*	unclassified	DTKEVEHEYNLKLIKELNIAKYESIVLAV AH
NC_028814	41.7%	2.14e-02	k119_50041	*BtRf-AlphaCoV/HuB2013*	*Alphacoronavirus*	*Decacovirus*	VMAIVFAVSFYKKDMEISLLLIGRIKLIYL ALFTLG
NC_002645	41.0%	6.78e-02	k119_22562	*Human coronavirus 229E*	*Alphacoronavirus*	*Duvinacovirus*	RPVKRMVSVSFSQSCPAKIIKLLISSVAVF IVRSCFLTL
NC_028806	40.6%	7.97e-02	k119_96071	*Swine enteric coronavirus*	*Alphacoronavirus*	*Tegacovirus*	VFLLPIVLITSISYFRLLILISTGCCASCGC C
NC_030292	39.4%	2.58e-02	k119_31290	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	FCCIQSGLCICCGLVQSIPVGSCKTETCRV GFF
NC_002645	38.6%	1.78e-02	k119_86558	*Human coronavirus 229E*	*Alphacoronavirus*	*Duvinacovirus*	LRDMKNPSSISLLFTTLRLHFLMVIKCQH TLSTSSTYIYTAFPCKTHITRKIDSTRF
NC_046964	38.0%	6.34e-02	k119_31388	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	FDFSKNLHLRIQDICRIQVTCNI*DTWFM FIDNGEITLIKNLFLSLNKNA

NC_018871	37.8%	1.08e-02	k119_60723	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	LS*FYSFYDIKELTKFTFCDSFIFFNKSCFK EFLSLI*DSTNSFL
NC_010438	37.5%	2.19e-02	k119_107283	*Miniopterus bat coronavirus HKU8*	*Alphacoronavirus*	*Minunacovirus*	INKKESATIGIWLKHFYGTRKNTNRSSIF VICSSEIVVKLYILCKSTQ
NC_034972	37.2%	3.80e-02	k119_67203	*Coronavirus AcCoV-JC34*	*Alphacoronavirus*	*Luchacovirus*	RQGKARGKPLSAYPHHQRFLGCGWQRL HEADHTGQLQPNQGRC
NC_005831	37.1%	1.22e-03	k119_75929	*Human coronavirus NL63*	*Alphacoronavirus*	*Setracovirus*	CTICWYIYCYCLITYFEAYSFCCFRWICK CYCIRN
NC_038861	36.8%	3.71e-02	k119_55453	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	TNTFLAGFKNPFSWQYNVLLVLLIIIFSYF YTAITIPV
NC_046964	36.7%	2.04e-03	k119_73552	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	FSILLWIKLIYYDVFTSIFVCFYADIQIRKL IP*LFQGYLYQNTVQNFC
NC_032730	35.6%	8.30e-02	k119_15360	*Lucheng Rn rat coronavirus*	*Alphacoronavirus*	*Luchacovirus*	IRKQAATAPIRYIEDYGKFLVDIDKKALE DGFYYIGRGKPLGKYE
NC_028811	35.2%	5.48e-02	k119_53416	*BtMr-AlphaCoV/SAX2011*	*Alphacoronavirus*	*Myotacovirus*	RQQY*MFQQQDYLLRNLHQHQLHYLPL KQ*L*PYRLMHKHSRRRHDKERPEYQE
NC_048211	34.8%	8.33e-03	k119_20009	Wencheng Sm shrew coronavirus	*Alphacoronavirus*	unclassified	VHHILNIPQH*KPLSCVQNHHFHELYKG NFVLLPQYHHKWNANLRF
NC_038861	34.7%	6.28e-02	k119_120235	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	TSEVRTDCHYNRVLITSYFNVCIVFAGRH IRCFFEYINRFTPKLYRSYF
NC_018871	34.4%	9.72e-02	k119_91260	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	SFTFSQCYLRCNWNFDCY*PYGFYCRWY CPWF
NC_003436	33.3%	4.32e-02	k119_58399	*Porcine epidemic diarrhea virus*	*Alphacoronavirus*	*Pedacovirus*	GKQHAEPIGACMKSSWVVPSHSTWVAR TAFRSSSSVVPRKRNSTSCKDTRSLAS
NC_003436	33.3%	9.64e-02	k119_38356	*Porcine epidemic diarrhea virus*	*Alphacoronavirus*	*Pedacovirus*	WTLLEQRLIRVQFLFLLKSNLSGIDNFHY SIIL
NC_030292	33.3%	9.76e-02	k119_104995	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	YNFNGTSLINNNWIIEHTVIKHCFCR*YS* HNLTNNFINVIDIAL

NC_032107	32.7%	5.97e-02	k119_6681	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	RLFSTAETPPGNCTDLFTLMPPRIISRTMR NLPLSSSHESDDSEPCTSV
NC_046964	31.4%	9.99e-02	k119_77092	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	REHLVNLMKRPEVKCSPKLLPTCVRILQ KRPSSIR
NC_018871	31.0%	8.39e-02	k119_12657	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	CLIWLSGISFSSALSRHCQFQTCCRNSISA TPSNGRTTFIGK
NC_028752	30.4%	9.08e-02	k119_82915	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	VSPWIGQVRFHPEDIYYYVKNPIYYTSGG QQVTPFISSPKAGVPSSFSLVDQDFKMPS VWRSSLGIDYK
NC_048216	30.4%	9.80e-02	k119_90706	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	CYINLYINWHPSLCFQEHRCYRCNGNTL VTECTIRCHVYCICNIIF
NC_048216	30.2%	5.04e-02	k119_54631	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	YMSFFFYCLYFWIKPIM*IPIFNILVIFSIRS TTWIENYMSWL
NC_048211	30.0%	6.81e-02	k119_76488	Wencheng Sm shrew coronavirus	*Alphacoronavirus*	unclassified	LLTSIMAWISCSDPSIIKNQLGVSPVTVRP LLHVTAVLGHLLIW*SIRCPGLYKALVAP G
NC_028806	29.1%	3.70e-02	k119_61001	*Swine enteric coronavirus*	*Alphacoronavirus*	*Tegacovirus*	IFLLSSLMLALPYIVLFTAFSLFTNPSVIPF VTFSLMVFNIAL*SNFTPVKNVLSVLIFVF QYFFIHLISLQLSVPVII
NC_028833	28.7%	9.70e-02	k119_9792	*BtNv-AlphaCoV/SC2013*	*Alphacoronavirus*	*Nyctacovirus*	TIFHILPTISKTHSRCTQRDIICPHCRLCSY VPFSFC*LLGVELSDVMNVSRFTIHVRTY LLNFVSMIHHHMSWMFNGTIFPVGTYV
NC_030292	28.6%	2.85e-02	k119_77192	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	ESRSYDIHNGMVTISGSCQNYSVLAPCLC RQRSRRITASDVL

Betacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_019843	62.5%	3.76e-03	k119_61753	*Middle East respiratory syndrome-related coronavirus*	*Betacoronavirus*	*Merbecovirus*	RICNAC*RCCSRLWCC
NC_006577	44.0%	9.06e-02	k119_128581	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	WLSDFIVLFIYSIYRDCYPVYSCFY
NC_030886	43.3%	3.00e-02	k119_99392	*Rousettus bat coronavirus*	*Betacoronavirus*	*Nobecovirus*	PDQCLFGHLRPNSRKRICSNCIPWSRTWRV
NC_012936	42.0%	1.39e-03	k119_7586	*Rat coronavirus Parker*	*Betacoronavirus*	*Embecovirus*	DVCGCCSTISITDYVR*RIYCSYESCCRYISIGSISIQLNCSCIGNRISC
NC_006577	41.3%	9.28e-02	k119_23785	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	FCEFLLVSLVAVILRTT*SDIFFSSLVETECQTVVFVCLQLCMLFM
NC_039207	40.0%	5.25e-02	k119_81210	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	IHNCTININLFIY*HFIF*NKRYFC*I*YR
NC_006577	39.5%	2.87e-02	k119_18459	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	NCTSLLITIIVLTVQRLSCSIVDESTILTDLAILPVLYENTEL
NC_009019	39.0%	5.59e-03	k119_21400	*Tylonycteris bat coronavirus HKU4*	*Betacoronavirus*	*Merbecovirus*	QNLYFVYLHQKHHFVKYIVYLNLIQQNHHYFPQLLKILS*I
NC_014470	38.9%	6.95e-02	k119_64073	*Bat coronavirus BM48-31/BGR/2008*	*Betacoronavirus*	*Sarbecovirus*	SFVNAELYLTSFCVFNSSCNVWSYCTYFWVRHQATW
NC_006577	36.2%	2.78e-02	k119_90627	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	KI*KLFCYKAILIDFSIKFIFNTLVLKYILKEFYEKKIN*F*IYSYY
NC_012936	35.9%	8.55e-02	k119_76433	*Rat coronavirus Parker*	*Betacoronavirus*	*Embecovirus*	NYMQGGVMHKLHNYLQKPLRNIHLYNFISIVILAIGCLV
NC_025217	35.7%	7.64e-02	k119_78113	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	NFRYQFSVCNLSILTDNNYCTRQQATQWSICHTHTVTFIKIRETEIRQRNYIFNTF
NC_006577	34.1%	2.31e-02	k119_610	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	NDSIYYLNKEQTRETFCFTSLFFILSFRLKITFCFYVCSFLVFF

NC_019843	33.9%	3.87e-03	k119_77541	*Middle East respiratory syndrome-related coronavirus*	*Betacoronavirus*	*Merbecovirus*	GHCLSALVANHLANQGALEGCFSFRHV GSLLGGLLFSKNGLGTSDIAASSLQSRSV
NC_004718	33.3%	2.70e-03	k119_50984	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	WLSANFWVFCCTSSDHLYKACKC*VNC SLCLLAH*DTSLVFSCGLVVFLKK
NC_039207	32.6%	1.57e-02	k119_44350	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	IKHCRFCKSWNLTVSTRHNSCRTIYTIHN LCCSTLKFTHITSCQCC
NC_012936	32.4%	6.36e-02	k119_33234	*Rat coronavirus Parker*	*Betacoronavirus*	*Embecovirus*	KSTCAATCDTLAIERYVLVNVDAFHNSV EQILLL
NC_039207	31.4%	3.28e-02	k119_37699	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	VLQYPERHLHSVPVYRHNLQLYKKHRFP DNEHVSTEIRFLSTQANHPLQLQ
NC_038294	27.7%	9.92e-02	k119_126399	*Betacoronavirus England 1*	*Betacoronavirus*	*Merbecovirus*	LMCIEVTLCCRTVLLPILNCETYISCHCLV RFLILVISCLYLNCLIL
				Deltacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_016994	76.9%	1.72e-02	k119_109101	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	YIYRCKIRCEWCW
NC_016993	61.9%	4.36e-02	k119_118139	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	LHYNRWSGCRESNPIYHLCAY
NC_016993	54.5%	1.35e-02	k119_3797	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	RCPDDWSG*RGCWYC*PSGRWH
NC_016994	48.5%	5.73e-02	k119_181	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	NTIPMILRITSRSPVGLSLASVRDARPRNI LTT
NC_016993	45.2%	1.99e-02	k119_74941	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	WQLGCHCYQL*WWCHQFLQRRYKWCR HHCWE
NC_016994	45.0%	9.87e-02	k119_72939	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	CLEHINHCSNNYDYHCYQLW
NC_011547	44.1%	2.99e-02	k119_62003	*Bulbul coronavirus HKU11-934*	*Deltacoronavirus*	*Buldecovirus*	SFPRFCSVRDAGVCRVTVSSFKHWNKVI ARFLYV
NC_016996	42.3%	4.07e-02	k119_104164	*Common moorhen coronavirus HKU21*	*Deltacoronavirus*	*Buldecovirus*	EQFHYHISNNNHLQLVRLLKQFDSHY

NC_011550	39.5%	3.20e-02	k119_130354	*Munia coronavirus HKU13-3514*	*Deltacoronavirus*	*Buldecovirus*	TF*ICHIRIEILFQSFVHIPNVSNSTAHSSSEIFTNST
NC_039208	33.3%	4.56e-02	k119_9247	*Porcine coronavirus HKU15*	*Deltacoronavirus*	*Buldecovirus*	LTLSCGGQITDYGLSDNYQSDTSFSCDSYSLGVGAKIKMNKHLNLNVG
NC_016993	31.7%	5.67e-02	k119_51859	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	GTGDRCISRGECRDGGENGRCTYRGAGERCGRNDLYEFVVF
NC_011550	28.9%	6.84e-02	k119_40089	*Munia coronavirus HKU13-3514*	*Deltacoronavirus*	*Buldecovirus*	SRCRYRRKLSRCTRRYRGFVKQFRMVAGQEVRPGYSVCRTQHR*R
				Gammacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_046965	58.3%	3.70e-03	k119_93732	Canada goose coronavirus	*Gammacoronavirus*	unclassified	CIVTACCFCYCC
NC_001451	46.4%	2.90e-02	k119_91351	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	HYLLHNLTWEGLLKETILWFAVWMLWQH
NC_010646	45.5%	2.20e-05	k119_2249	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	KIDDNMILILEGIHALNPELTP
NC_001451	42.9%	4.62e-02	k119_35599	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	VIRSIP*VLFSTRTKAASSPWMNFC*RLAICAISSFDCDALPSSRALYV
NC_010646	39.5%	1.13e-05	k119_29749	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	SGEYDYESIYALDLDLINEQFNALFRGEEVELPKYDFQ
NC_010646	38.2%	4.16e-03	k119_29547	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	LNADNVLVIEGIHALNPELTSAIPRQAMFKVYVS
NC_010646	37.8%	1.09e-04	k119_103558	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	GEYDYESIYALDLQLINDQFNALFRGEEVELPKYDFQ
NC_046965	35.7%	2.54e-02	k119_82666	Canada goose coronavirus	*Gammacoronavirus*	unclassified	WENIQEVWYQCTYTEHFTSCYNDDYDWILILCTHRDTFYSKY
NC_010646	28.4%	4.61e-02	k119_129574	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	YRELCN*RSYFCCSAESHNC*ICCSSVKIHFNSTSFTISKNVINISKLSTERNITCSIVKLSCYILK
NC_010646	25.5%	4.89e-03	k119_75869	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	VDREDTPRDENGNYDYESLYALDLELFNTQLQALLRGEEVELPRFNFNLGK
NC_010800	22.2%	1.71e-02	k119_122414	*Turkey coronavirus*	*Gammacoronavirus*	*Igacovirus*	CSNCFNAKVVSVSCKKRTMPWLIGCVELLTNAYNKEWSKFCRYRCSTTSGFPLVIVPVLSSTNVSHFAILSI

conclusão

Não classificadas							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_046955	61.5%	1.01e-02	k119_74653	Shrew coronavirus	unclassified	-	QGWIGFEVAAAMVLGENIGTTITANL
NC_046954	52.0%	3.98e-02	k119_16079	Rodent coronavirus	unclassified	-	FKNGSICYIVTSTFLVNCYVSAFIK
NC_046955	44.4%	1.66e-02	k119_2959	Shrew coronavirus	unclassified	-	CTCFCYYFALVKLNNFTCHICITNSTFSC L*VFSCN
NC_034440	34.1%	5.12e-02	k119_127254	Bat coronavirus	unclassified	--	TVITPTLGSIVQKGKLADCAFALDKQLK RVDFPTLGSPTIP
NC_046954	33.3%	3.25e-02	k119_102638	Rodent coronavirus	unclassified		IAFTYLNDSLKKRVVPIAEEFESGDLVYP CDVSKPEEIKALKESLEKDLGQ

Table S3. Analysis of E165 data using tBLASTX

Alphacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_002306	61.9%	7.33e-02	k119_86773	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGKTIKEARDNAY
NC_038861	61.1%	1.66e-02	k119_30422	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	WCSLHHRCFSWWQLGRRF
NC_028833	60.0%	2.34e-02	k119_102903	*BtNv-AlphaCoV/SC2013*	*Alphacoronavirus*	*Nyctacovirus*	LRHHWYTRSNRSLRCILRILLVIFS
NC_032107	57.1%	2.08e-02	k119_123078	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	AFNLFVIDRNCCLLGLSCWCC
NC_002306	56.0%	3.50e-03	k119_27984	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGSSIKEARDRAYALCG
NC_048216	56.0%	2.39e-02	k119_35534	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	HHH*R*SYATESCTESACNLYPLLH
NC_002306	52.0%	2.87e-02	k119_88464	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGFGKSIKEARDRAYTLCG
NC_028752	50.0%	9.74e-04	k119_56822	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	LLLH*ALSCGELNLSVPVNILLCVFTLIM QME
NC_010438	48.0%	2.26e-02	k119_36736	*Miniopterus bat coronavirus HKU8*	*Alphacoronavirus*	*Minunacovirus*	HKRSGSLKPYQISL*HLQEQHHTFR
NC_002306	47.5%	4.62e-03	k119_123393	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	LANLFSFVSQMAKE*CDAYQCITSVVTS RINYSTISLTAD

NC_032730	45.8%	4.75e-02	k119_67111	*Lucheng Rn rat coronavirus*	*Alphacoronavirus*	*Luchacovirus*	CSSYRKRSLRKLHFSSDEYHHGAD
NC_010438	45.8%	2.06e-02	k119_89557	*Miniopterus bat coronavirus HKU8*	*Alphacoronavirus*	*Minunacovirus*	YHNPSGYLQHLQLEHYWYNACNPN
NC_028811	43.3%	4.17e-02	k119_75653	*BtMr-AlphaCoV/SAX2011*	*Alphacoronavirus*	*Myotacovirus*	SAT*FYCRRCRRVSMPWW*CCRWTCVRNTS
NC_048211	42.4%	6.24e-02	k119_95390	*Wencheng Sm shrew coronavirus*	*Alphacoronavirus*	*unclassified*	PLSPYGKRKYFIFSKFSLPRVLITHFLYSLKSS
NC_028752	41.9%	1.83e-02	k119_95629	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	LHLLISALSCRNILHNIFRIDQCKFLCICCH
NC_028824	40.0%	5.16e-02	k119_57403	*BtRf-AlphaCoV/YN2012*	*Alphacoronavirus*	*Rhinacovirus*	SITKNLLISKIIDNGSSNTGQASVHALHVVQAHNSSFVI*SLRSFLPSST
NC_038861	40.0%	3.03e-02	k119_12045	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	SCAVWLSQSSLLARSLVWLIAVWRLTACAIVRWL*NCLMKLDRVMRFVTA
NC_022103	37.7%	4.15e-02	k119_84321	*Bat coronavirus CDPHE15/USA/2006*	*Alphacoronavirus*	*Colacovirus*	SHFTLSTSFQGLFGEKFFGTVNTITGITFILQIHTKAIRVASTNGSRDQQSKCS*KKYFLY
NC_046964	36.7%	4.08e-03	k119_78721	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	FSILLWIKLIYYDVFTSIFVCFYADIQIRKLIP*LFQGYLYQNTVQNFC
NC_023760	36.6%	4.48e-02	k119_108802	*Mink coronavirus strain WD1127*	*Alphacoronavirus*	*Minacovirus*	KRCTCRIGIDCSAICLKDTVCLYVCATRCSTTICRRICALV
NC_032730	35.6%	7.82e-02	k119_34055	*Lucheng Rn rat coronavirus*	*Alphacoronavirus*	*Luchacovirus*	IKKQAATAPIRYIEDYGKFLVDIDKKALEDGFYYIGRGKPLGKYE
NC_048211	35.6%	8.14e-03	k119_35828	Wencheng Sm shrew coronavirus	*Alphacoronavirus*	unclassified	HHILNIPQH*KPLSCVQNHHFHELYKGNFVLQPQYHHKWNANLRF
NC_023760	35.0%	3.72e-02	k119_59789	*Mink coronavirus strain WD1127*	*Alphacoronavirus*	*Minacovirus*	KLATFYLKVVTTNSNTSFYVVFFFGLLALVWIIFIICIVYWFHVVAQSFFIYHYAGSFFL
NC_009988	34.7%	1.35e-02	k119_37089	*Rhinolophus bat coronavirus HKU2*	*Alphacoronavirus*	*Rhinacovirus*	ILTAILVSCCSIFVFFPYVGSLISLFLFFVCTNNLHGCFVVCFIYEYSY

NC_010438	33.3%	2.90e-02	k119_109671	*Miniopterus bat coronavirus HKU8*	*Alphacoronavirus*	*Minunacovirus*	FYADCCDRASFNCTFRIYDQCDCLTCW*TCYSRLRNQVSIR*NCLRNFCSCIHT
NC_010438	33.3%	8.27e-02	k119_29632	*Miniopterus bat coronavirus HKU8*	*Alphacoronavirus*	*Minunacovirus*	LLKIQLTVCRMRWLSRLFSNKSLMVVSFWRLEKNITTNLKKS*LLERFIESLKL
NC_003436	33.3%	7.62e-02	k119_50484	*Porcine epidemic diarrhea virus*	*Alphacoronavirus*	*Pedacovirus*	WTLLEQRLIRVQFLFLLKSNLSGIDNFHYSIIL
NC_002306	31.6%	4.12e-03	k119_82588	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	THALPSHSKIEVIRDSAITSRKEGLVHSCDRIVIFCSRCSLSSDSRHNRLSINSKLR
NC_028752	30.4%	7.45e-02	k119_100517	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	VSPWIGQVRFHPEDIYYYVKNPIYYTSGGQQVTPFISSPKAGVPSSFSLVDQDFKMPSVWRSSLGIDYK
NC_028806	29.1%	6.99e-02	k119_3736	*Swine enteric coronavirus*	*Alphacoronavirus*	*Tegacovirus*	IFLLSSLMLALPYIVLFTAFSLFTNPSVIPFVTFSLMVFNIAL*SNFTPVKNVLSVLIFVFQYFFIHLISLQLSVPVII
NC_018871	28.6%	4.85e-02	k119_99841	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	ILISVEILMTVFILV*LLAFKMYIIIATLFIQRIWKIVSIVIPLNHCVLQAQVLMLSLVLFFVPLLIHLF
NC_022103	27.0%	5.56e-02	k119_6757	*Bat coronavirus CDPHE15/USA/2006*	*Alphacoronavirus*	*Colacovirus*	HWIFVPLTLGLAVVMGIMETLYYKTGNEFWKKTAQFW
NC_010437	26.7%	9.85e-03	k119_133145	*Bat coronavirus 1A*	*Alphacoronavirus*	*Minunacovirus*	LSTFYEKIMFVLSCLVVGCCATVLCVLTDVVSVYPILKSSTIVRL
				Betacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_045512	60.9%	8.70e-02	k119_95268	*Severe acute respiratory syndrome coronavirus 2*	*Betacoronavirus*	*Sarbecovirus*	LSFFKSCNWFSFSTF*FLYILFS
NC_004718	56.0%	7.65e-02	k119_74880	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	IQHYVCFHHRNCQNCACFRIHQNSC
NC_026011	43.6%	5.97e-02	k119_109221	*Betacoronavirus HKU24*	*Betacoronavirus*	*Embecovirus*	YGFCY*QSDIFCCCSHRQYHN*LRKLYQTRSYIRQRRKN

NC_004718	42.4%	4.96e-02	k119_126891	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	VFLLSENC*MLVLHHCFCSCKFFSVWYC QSRFC
NC_001846	41.9%	1.68e-02	k119_112272	*Murine hepatitis virus*	*Betacoronavirus*	*Embecovirus*	PCSCLTKKTYRTKIFLYFICHPLPTCHPVP D
NC_012936	40.0%	5.81e-02	k119_31524	*Rat coronavirus Parker*	*Betacoronavirus*	*Embecovirus*	HIHLSEKE*FRIQPCRRHEDSETYHTHLPH
NC_039207	39.7%	3.31e-02	k119_112743	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	KTFTEKPNLEMAKIFVSSGEFYWNSGMF LWNLDAILEAFNTYLPEVAEKFAAGAAV YN
NC_009019	39.0%	6.48e-03	k119_66797	*Tylonycteris bat coronavirus HKU4*	*Betacoronavirus*	*Merbecovirus*	QNLYFVYLHQKHHFVKYIVYLNLIQQNH HYFPQLLKILS*I
NC_009019	39.0%	1.03e-02	k119_66796	*Tylonycteris bat coronavirus HKU4*	*Betacoronavirus*	*Merbecovirus*	QNLYFVYLHQKHHFVKYIVYLNLIQQNH HYFPQLLKILS*I
NC_026011	37.1%	9.35e-02	k119_125274	*Betacoronavirus HKU24*	*Betacoronavirus*	*Embecovirus*	VALNIE*IALNSGICQPTHSRFFHWVNIM RICACN
NC_009021	35.7%	2.87e-02	k119_58236	*Rousettus bat coronavirus HKU9*	*Betacoronavirus*	*Nobecovirus*	SGTCWCIASYNNSSRFWLYFTLCLYNRS *ITTLSQTIVSPTI
NC_006213	35.5%	3.32e-02	k119_69287	*Human coronavirus OC43*	*Betacoronavirus*	*Embecovirus*	HRNLIKTLCNINHLVIGGSCHLCLMEFLL HS
NC_009020	35.0%	9.16e-02	k119_26763	*Pipistrellus bat coronavirus HKU5*	*Betacoronavirus*	*Merbecovirus*	TSPLRRKPLFLVASGTRVVPTSITTAPGFT ISAVTKSGLPIAAMMISASRHSSFKFLE*E
NC_004718	33.3%	2.73e-03	k119_6450	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	WLSANFWVFCCTSSDHLYKACKC*VNC SLCLLAH*DTSLVFSCGLVVFLKK
NC_045512	31.4%	7.61e-02	k119_12084	*Severe acute respiratory syndrome coronavirus 2*	*Betacoronavirus*	*Sarbecovirus*	LYRNDVCNLQFLRNHIFVTIVNKNHFIRF SFSKII
NC_014470	29.3%	9.39e-02	k119_99122	*Bat coronavirus BM48-31/BGR/2008*	*Betacoronavirus*	*Sarbecovirus*	FHYDIKKRYPISFLLSLHFIPHNPPLTAFIA *KNDTPYRFFYSILSVSNGNGGLFGVS

NC_006577	28.4%	3.47e-02	k119_25546	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	SPVATLTELLSVATILPSVLTVTAAPGTK LVTAPLSSSACADVPGAVCSGCESATSL ASYKTTSTLAKLV*PFSTATK*TVLATFA YPSKSTVTTCCPAVRP
NC_039207	25.5%	7.63e-03	k119_61985	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	CCKEALLVRKLLSSAVALCRFFFAPALP WFSSFCLANSCSAKVILLAVLKYSSCNCI NCGLAIVAITCPFFT*SPICTFSVFITPPYK EVALALLLLG
				Deltacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_016991	46.4%	9.56e-02	k119_94815	*White-eye coronavirus HKU16*	*Deltacoronavirus*	*Buldecovirus*	APLADNPSEKVKVKIVPLQFLKAVLPNP
NC_039208	42.4%	7.75e-02	k119_58860	*Porcine coronavirus HKU15*	*Deltacoronavirus*	*Buldecovirus*	YVSAIVILNYTRFNICSTTVWACIIVRYKT YCW
NC_016993	39.4%	3.49e-02	k119_36966	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	HQPEKPSPTGHCCISHWHNPDGYPSAH* YCGGC
NC_011547	38.7%	8.02e-02	k119_29497	*Bulbul coronavirus HKU11-934*	*Deltacoronavirus*	*Buldecovirus*	CSGSVCRRRSSFTVVVVHCSAVFRSCVV CNG
NC_011547	37.5%	4.30e-02	k119_124029	*Bulbul coronavirus HKU11-934*	*Deltacoronavirus*	*Buldecovirus*	CGREKIHQKRRCRNGSKRCI*CYY
NC_016995	37.3%	9.78e-03	k119_116189	*Wigeon coronavirus HKU20*	*Deltacoronavirus*	*Andecovirus*	AVTCNPKAFASRS*VNMLGSA**GLLIVL W*VLTSLVITKRFISIPFNRFRNDRRVTNV
NC_016993	35.8%	5.14e-02	k119_129923	*Magpie-robin coronavirus HKU18*	*Deltacoronavirus*	*Buldecovirus*	YINLSLLIWKSQLMRLPPLWMRTAISQH VLLSIRCVEEVIPPGYVTLIHYVFV
NC_011547	35.7%	4.77e-02	k119_112447	*Bulbul coronavirus HKU11-934*	*Deltacoronavirus*	*Buldecovirus*	CWLIFYKELVVGINRYSYILLCHCLS*C
NC_039208	33.3%	1.36e-02	k119_44523	*Porcine coronavirus HKU15*	*Deltacoronavirus*	*Buldecovirus*	LTLSCGGQITDYGLSDNYQSDTSFSCDSY SLGVGAKIKMNKHLNLNVG
NC_016995	32.1%	3.86e-02	k119_108732	*Wigeon coronavirus HKU20*	*Deltacoronavirus*	*Andecovirus*	CLAQNQTAPTCSQLTVKQVILLGMLPVIL ML*ITNSFENQYYLLSSASSKFFYKISQR

Gammacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_010646	71.4%	3.31e-04	k119_68888	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	VLVVEGIHALNPEL
NC_048214	50.0%	4.35e-02	k119_51409	*Duck coronavirus*	*Gammacoronavirus*	*Igacovirus*	TWFLNFTIKCRFIFKNWNTIIFTCKFFIFNCF
NC_010646	48.6%	4.38e-02	k119_9176	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	NK*QFLIPLSSKPSQKLLINMRFLMILQRNTISDN
NC_010646	39.5%	2.30e-05	k119_57225	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	SGEYDYESIYALDLDLINEQFNALFRGEEVELPKYDFQ
NC_010646	37.8%	1.71e-04	k119_87143	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	GEYDYESIYALDLQLINDQFNALFRGEEVELPKYDFQ
NC_010646	36.1%	3.12e-03	k119_120713	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	GDFDFESIYALNLDLLNEQFNALFRGEEVELPKYDF
NC_010646	35.1%	3.92e-03	k119_35150	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	GEYDFESLYALDLPYFNSDLQKAINGEEIALPTFSFE
NC_001451	35.0%	6.07e-02	k119_33667	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	TLFYRYLYHNHTILWQELHKQIPDCHHTIQSQKHKTSFLS
NC_048214	34.1%	9.51e-02	k119_35312	*Duck coronavirus*	*Gammacoronavirus*	*Igacovirus*	FLDMCHCIGYLVFSALLGRCNTGLYKWHL*SVLCNHRFSGS
NC_001451	30.2%	3.13e-02	k119_11559	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	CRILYAFQCNNLRQKETAPRLKLNFSIIDAGIMNGFEILCKLGGYIMIFSILL
NC_001451	27.1%	5.07e-02	k119_15889	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	LIQSIYASTDAAIISVFAPKP*YICPSYSTCMCTLPISSLPLLIAWIVNSFNTMWRSII
Não classificadas							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_046954	63.0%	7.77e-03	k119_89865	Rodent coronavirus	unclassified	-	YTKTVIKPGEKGTVTATYNAATPGAFH
NC_048212	33.3%	7.62e-02	k119_45049	Bat coronavirus	unclassified	-	LQTVEAELRCLDTYIVEIKEVLVWLLYWQTEHYAGSQLD
NC_046965	27.3%	6.92e-02	k119_9808	Canada goose coronavirus	*Gammacoronavirus*	-	CKDVNCPFHGTLPVRGQVITGKVVSDKMMGTVVVARDYLHYVRKYNRYEKRISKIHAHNPPCIQAKVGDLVKIAECRPLSKSTTYVVV

Table S4. Analysis of E205 data using tBLASTX

Alphacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_032107	64.3%	4.60e-02	k119_99774	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	SYRLWQRCHHRYQQ
NC_010437	60.9%	1.87e-02	k119_29496	*Bat coronavirus 1A*	*Alphacoronavirus*	*Minunacovirus*	C*WFPCW**CSRHKRAFPWARFC
NC_002306	56.0%	2.14e-02	k119_70683	*Feline infectious peritonitis virus*	*Alphacoronavirus*	*Tegacovirus*	GRVLVCVGSGKTIKEARDRAYALCG
NC_005831	50.0%	4.62e-02	k119_99131	*Human coronavirus NL63*	*Alphacoronavirus*	*Setracovirus*	CTILYCCK*LFRSY*GNFWYLIYYSS
NC_048216	48.5%	2.90e-02	k119_41045	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	VCIDTATFLVFTRSTCRVLWFDQFKL*L WHTIG
NC_030292	39.4%	4.81e-03	k119_55707	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	FCCIQSGLCICCGLVQSIPVGSCKTETCRV GFF
NC_048216	39.1%	1.72e-02	k119_2443	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	KVWHHQRLCHGTSSGYPQWRNHP
NC_048216	32.7%	8.12e-03	k119_46781	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	SAWRAFLFAHGFEYLAHGSFSLVHPRRG HCRNRRQNPIAGAFSGRPLQPKKR
NC_003436	32.7%	8.34e-02	k119_51041	*Porcine epidemic diarrhea virus*	*Alphacoronavirus*	*Pedacovirus*	LYSFCAP*SYI*LRFFSSSIDILIVSNQVSLL QHRKYVSCLEVLLLQGFC*SPQE
NC_028752	30.4%	4.09e-02	k119_74911	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	VSPWIGQVRFHPEDIYYYVKNPIYYTSGG QQVTPFISSPKAGVPSSFSLVDQDFKMPS VWRSSLGIDYK
NC_032107	30.3%	7.03e-02	k119_106283	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	FLISFCVFTSVLIFVLLVHHRSLIQVLLLQ QIFGFFQIIYLLVPLYVVFYKYLLVQLQS YLFNLIR
NC_028752	30.2%	1.01e-02	k119_74799	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	HWWCCTSTSIRCCIW*LIFGRTI*LRRHLT LSIHRQLLCIIES
NC_018871	28.0%	6.11e-02	k119_65003	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	WLVIWGSLKCRYCHSTYYSTCLICILNIE REVESFAKKSYTGRVRINTVN
NC_038861	25.0%	4.71e-02	k119_73595	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	IIYDIRH*IKTQIFGEIIASFEHNTNFIQIVIR DNTLCIRITCRSICIEVFRTSRS
Betacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_009020	57.9%	9.16e-02	k119_90845	*Pipistrellus bat coronavirus HKU5*	*Betacoronavirus*	*Merbecovirus*	LCCRNHYLCCFTNRCIRMP
AC_000192	51.9%	4.60e-02	k119_76818	*Murine hepatitis virus strain JHM*	*Betacoronavirus*	*Embecovirus*	VSDCRSMDLVTGQQFNGSTVQRFNGQK

NC_026011	48.1%	4.43e-02	k119_30949	*Betacoronavirus HKU24*	*Betacoronavirus*	*Embecovirus*	IPS*LQKITPAPLQQQTKNRLNAEATP
NC_004718	45.5%	8.57e-02	k119_42031	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	CRFCLIYSCGWWVICPVSDRCC
NC_009019	41.9%	3.73e-02	k119_43949	*Tylonycteris bat coronavirus HKU4*	*Betacoronavirus*	*Merbecovirus*	QNLYFVYLHQKHHFVKYIVYLNLIQQNH HYF
NC_030886	41.2%	1.56e-02	k119_79490	*Rousettus bat coronavirus*	*Betacoronavirus*	*Nobecovirus*	WSERLTRCWMHNSPSATLQVRSKLWLR VQITTSL
NC_025217	35.7%	5.93e-02	k119_29176	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	NFRYQFSVCNLSILTDNNYCTRQQATQW SICHTHTVTFIKIRETEIRQRNYIFNTF
NC_004718	33.3%	4.78e-03	k119_119336	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	WLSANFWVFCCTSSDHLYKACKC*VNC SLCLLAH*DTSLVFSCGLVVFLKK
NC_006577	31.6%	8.97e-02	k119_42175	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	KHDLGTSPYNNIYLYSSKL*NKWSYRVY KRTYGANTGYATYEYNRRTSKTNKHINE T
NC_030886	24.4%	8.66e-02	k119_96178	*Rousettus bat coronavirus*	*Betacoronavirus*	*Nobecovirus*	LFAQAVQSRTNQVVWIRRTQRLSHHVM NTHHIKYGTHRTAG
NC_045512	24.2%	5.80e-02	k119_91259	*Severe acute respiratory syndrome coronavirus 2*	*Betacoronavirus*	*Sarbecovirus*	IVFLFFISFMLLLHSDTSDRYKNETIWYK MGKWKIIIPSVFIVLLFLSNLVPSNKNHFL YIV
				Deltacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_016995	42.9%	9.68e-02	k119_47781	*Wigeon coronavirus HKU20*	*Deltacoronavirus*	*Andecovirus*	WTYKVLGIT*LKLSCSYCWSRECRLTKF DNIRCAI
NC_016994	41.7%	4.71e-03	k119_29555	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	NSGFTHNRNRFYLDHKDFTFLLEWSKKL IHHLESLF
NC_039208	38.7%	2.77e-02	k119_82377	*Porcine coronavirus HKU15*	*Deltacoronavirus*	*Buldecovirus*	VQNAFKATDCLFQSHVFTW*TGEYFRNV KWL
NC_016995	28.9%	9.45e-02	k119_61905	*Wigeon coronavirus HKU20*	*Deltacoronavirus*	*Andecovirus*	YLGRIRSPDCNHGWCYPVPFCSWRSFTS SSFTIRQYCR

conclusão

				Gammacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_001451	47.1%	9.08e-02	k119_88850	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	RHRLGRRTASVPSHGARLPKRSGARGKR

							AVYGDD
NC_048213	43.5%	4.32e-02	k119_20782	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	FGISYHWSGKWSKYYCGRRYIYN
NC_048213	35.2%	3.50e-02	k119_21350	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	STISAPWKPPTWKT*IFPVRAPCPSSTPSN KPTVSNTRRTVSVLVMASVFAPST
NC_001451	28.0%	3.70e-02	k119_122346	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	YKCCNHTYTALKFFYQITSTTCSSCFYICI FKSFMNLFIQINSICYYNDF
NC_010646	25.4%	5.38e-02	k119_10031	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	ETIGRSSIDITLVVRIVQQDFFIHQLTPSTG CILTCIVEAPRTYSRRSTTTVLSILIDGYK RIILEI
Não classificada							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_046955	41.7%	4.33e-02	k119_48116	Shrew coronavirus	unclassified	-	KYCYIFQDIYHEIQSSVRYSEPVC
NC_046954	37,80%	8.71e-02	k119_54624	Rodent coronavirus	-	-	NCCRAC*FYTNIWRNISIFNLKTTICFSWC YIHYRNY

Table S5. Analysis of E208 data using tBLASTX

				Alphacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_009657	62.5%	8.38e-03	k119_32933	*Scotophilus bat coronavirus 512*	*Alphacoronavirus*	*Pedacovirus*	GCCGCCGYLASTYQGC
NC_032730	55.0%	4.79e-02	k119_61303	*Lucheng Rn rat coronavirus*	*Alphacoronavirus*	*Luchacovirus*	EGSWKFYFLPSPYKRVKQLI
NC_002645	53.8%	3.17e-02	k119_133830	*Human coronavirus 229E*	*Alphacoronavirus*	*Duvinacovirus*	FTVTTIASCCCSDRCTIFFSCFIQSI
NC_034972	52.2%	7.21e-02	k119_96559	*Coronavirus AcCoV-JC34*	*Alphacoronavirus*	*Luchacovirus*	CTHSSNKQRNTCTNIRRHHCCST
NC_023760	52.0%	1.45e-02	k119_132029	*Mink coronavirus strain WD1127*	*Alphacoronavirus*	*Minacovirus*	CTRSCSYCRTYNTPYRSAHCFVLCI
NC_018871	51.6%	4.87e-02	k119_77851	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	IGVSLSTSGRCMNHIFLQGPRGLERYI*LCL
NC_046964	45.8%	7.61e-02	k119_43248	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	YRNWSHRHGQSSPCSLRFYSHQPL
NC_034972	45.5%	2.58e-02	k119_95989	*Coronavirus AcCoV-JC34*	*Alphacoronavirus*	*Luchacovirus*	LTLYMFTIVRSSL*VLTRPYMI
NC_018871	40.5%	3.55e-02	k119_38633	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	TKQDKLKPGAADRNVEVYKTKLLETWEDLPDIYITSAEKKMS
NC_030292	39.4%	5.74e-03	k119_106387	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	FCCIQSGLCICCGLVQSIPVGSCKTETCRVGFF
NC_002306	38.9%	5.28e-03	k119_53600	Feline infectious peritonitis virus	Alphacoronavirus	Tegacovirus	VSTYSICTSAYPNINLSQIYDSLNNGFLQTVAIDLKTTNKYVCLLIS*FPYIAV
NC_038861	36.8%	8.06e-02	k119_152091	*Transmissible gastroenteritis virus*	*Alphacoronavirus*	*Tegacovirus*	TNTFLAGFKNPFSWQYNVLLVLLIIIFSYFYTAITIPV
NC_046964	36.7%	2.77e-03	k119_119178	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	FSILLWIKLIYYDVFTSIFVCFYADIQIRKLIP*LFQGYLYQNTVQNFC
NC_030292	35.7%	7.06e-02	k119_64133	*Ferret coronavirus*	*Alphacoronavirus*	*Minacovirus*	PNRNECSNTTFFRGSCSLFPSTIFTTFECRNRKIVSSLSIDRTYNIFNEVGIIRMR
NC_046964	35.7%	6.24e-02	k119_107137	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Alphacoronavirus*	*Nyctacovirus*	THWISTHSTNREWKLTK*ICLHCLKNQALKNRPKRFVRSYKQ
NC_048216	35.6%	4.60e-02	k119_51346	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	WRHVCAIGYLKTSMLLAVTPATQRTFLTARLLIGGTTPRGVWTTL
NC_028752	33.3%	9.25e-03	k119_25681	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	EHNHFITNSKDRTQMNNQPHHPGKPSLHFYFTKLCHGTITSYSRH
NC_048216	32.6%	6.82e-02	k119_126531	*NL63-related bat coronavirus*	*Alphacoronavirus*	*Setracovirus*	LRNICCGLLNDALCHHRVGNFHEAGDISALHVVDIAVRLCTVL

NC_028752	30.4%	7.68e-02	k119_131562	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	VSPWIGQVRFHPEDIYYYVKNPIYYTSGG QQVTPFISSPKAGVPSSFSLVDQDFKMPS VWRSSLGIDYK
NC_028752	30.2%	1.03e-02	k119_100319	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	HWWCCTSTSIRCCIW*LIFGRTI*LRRHLT LSIHRQLLCIIES
NC_028752	30.2%	6.66e-02	k119_21520	*Camel alphacoronavirus*	*Alphacoronavirus*	*Duvinacovirus*	LSSLSHITTVLFSKSPDMAAEGPLYSVPAI GCAATSCSLFAPLIALDIFSFVLPTSIIVVF LF
NC_018871	28.3%	3.42e-02	k119_110646	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	TRRSKMPRQESMRLPDRLKNTSPCHGPR DEIYRLGVRRQSVRSEQRLPQPG*RRTRP HPS
NC_018871	25.3%	4.59e-02	k119_99954	*Rousettus bat coronavirus HKU10*	*Alphacoronavirus*	*Decacovirus*	LLQDSPVEAEKCHLGIAPVTLALTQIQAA LSEAVRPVYTVAQDLSRIAGLGAYTGEV SAELLKDSQIDYVLVGHS
NC_028814	20.0%	4.04e-02	k119_91265	*BtRf-AlphaCoV/HuB2013*	*Alphacoronavirus*	*Decacovirus*	KAIAGCTIESIVGNKILSNNHIRIISGN VLVGTKVSKTDYLGAYDNQITVIPEG DDNDEFFGWATPGLDK
				Betacoronavirus			
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_026011	73.7%	7.94e-02	k119_108162	*Betacoronavirus HKU24*	*Betacoronavirus*	*Embecovirus*	LSTYLNSDMEKVDAWNYGK
NC_006577	57.1%	3.81e-03	k119_153004	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	KHCRLRQHTHHHCH
NC_003045	56.3%	1.33e-02	k119_68771	*Bovine coronavirus*	*Betacoronavirus*	*Embecovirus*	ELCLLWKRLVFLLWQLTLCLIQRWLKIL LRNN
NC_004718	56.0%	7.65e-02	k119_140408	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	IQHYVCFHHRNCQNCACFRIHQNSC
NC_025217	46.9%	3.19e-03	k119_27445	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	NASYFWFGYRRYCYSSRSNWFSCYDFL NYCLY
NC_025217	46.2%	5.13e-02	k119_126537	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	PDLPNSEGRTFERARTNARKTLIFQRCLK THGCLNVLTL
NC_030886	43.3%	2.90e-02	k119_147278	*Rousettus bat coronavirus*	*Betacoronavirus*	*Nobecovirus*	PDQCLLGHLRPNSRKRICSNCIPWSRTWR V
NC_014470	42.1%	7.63e-02	k119_115257	*Bat coronavirus BM48-31/BGR/2008*	*Betacoronavirus*	*Sarbecovirus*	RPPERPYFVTRLPDVNWDLVAMTASLFT QKFSKLARLV

NC_009019	41.9%	4.11e-02	k119_61894	*Tylonycteris bat coronavirus HKU4*	*Betacoronavirus*	*Merbecovirus*	QNLYFVYLHQKHHFVKYIVYLNLIQQNH HYF
NC_025217	41.3%	4.23e-02	k119_43760	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	SVEVAYCATAATNSATTVPRLAPSFSAFT QKSIPAASCAYVLPSVL
NC_048217	39.5%	2.21e-02	k119_35537	*Murine hepatitis virus*	*Betacoronavirus*	*Embecovirus*	LRVIKMTPACWLV*ITILRTWMWMELPA MCATYEWPLS
NC_045512	39.0%	7.96e-03	k119_129387	*Severe acute respiratory syndrome coronavirus 2*	*Betacoronavirus*	*Sarbecovirus*	HTNKFCTSIPHRHRFVFSSCLHRTKPLFLP PRERITWFHFV
NC_009021	38.2%	5.08e-02	k119_118491	*Rousettus bat coronavirus HKU9*	*Betacoronavirus*	*Nobecovirus*	CTYSLALLNAILSCVLWLYTILVFSLINLV SMVL
NC_017083	37.8%	6.84e-02	k119_95592	*Rabbit coronavirus HKU14*	*Betacoronavirus*	*Embecovirus*	RVLVRHGAQKTVSQETNRNLCATKKAR MYTQPMLEIY
NC_006577	36.2%	4.18e-02	k119_88116	*Human coronavirus HKU1*	*Betacoronavirus*	*Embecovirus*	KI*KLFCYKAILIDFSIKFIFNTLVLKYILK EFYEKKIN*F*IYSYY
NC_025217	35.1%	3.14e-02	k119_65439	*Bat Hp-betacoronavirus/Zhejiang2013*	*Betacoronavirus*	*Hibecovirus*	SICTSASLTNGVCASKVTISIKLLTDSSCM YAISPCY
NC_004718	33.3%	5.27e-03	k119_51157	*SARS coronavirus Tor2*	*Betacoronavirus*	*Sarbecovirus*	WLSANFWVFCCTSSDHLYKACKC*VNC SLCLLAH*DTSLVFSCGLVVFLKK
NC_012936	32.4%	5.07e-02	k119_124248	*Rat coronavirus Parker*	*Betacoronavirus*	*Embecovirus*	KSTCAATCDTLAVERYVLVNIDAFHNSV EQILLL
NC_039207	26.5%	1.38e-03	k119_124689	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Betacoronavirus*	*Merbecovirus*	VKFAIFSKRLDVTYFFIPYFNVFKNFIME DKPIVKVIDNNFKAKRRSLAHQKFDLLV HFEVFKIFAQK

Deltacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_016995	51.7%	6.11e-02	k119_141842	*Wigeon coronavirus HKU20*	*Deltacoronavirus*	*Andecovirus*	TFSFSSIYASSFLTHKKRRTLSMIRTSCS
NC_016994	41.7%	6.94e-03	k119_32537	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	NSGFTHNRNRFYLHHKDFTFLLEWSKKL IHHLESLF
NC_016994	38.1%	4.25e-02	k119_12081	*Night heron coronavirus HKU19*	*Deltacoronavirus*	*Herdecovirus*	PTLIHHLHLFNNDPTLSVHTP
NC_016992	28.9%	5.16e-02	k119_59724	*Sparrow coronavirus HKU17*	*Deltacoronavirus*	*Buldecovirus*	GMTARQMNAWMYYGGGQKLLDEFYA NYNVVSFAGGNTGTQMGGWW
Gammacoronavirus							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_001451	63.6%	1.45e-02	k119_22294	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	KICRCINNSFFIFFTNRTFILF
NC_001451	50.0%	7.52e-02	k119_64572	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	YLFQCGDFRYQEPYSGYGAYSFFHCCNR IV
NC_048214	45.5%	1.90e-02	k119_35673	*Duck coronavirus*	*Gammacoronavirus*	*Igacovirus*	CKSCRYDF*RWQVESRLYQCAYHFELLP SSWSC
NC_001451	42.9%	3.36e-02	k119_116203	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	ITIC*NFGVFFDCCARWCVYF*LAASSS
NC_048213	40.0%	1.12e-02	k119_121099	*Infectious bronchitis virus*	*Gammacoronavirus*	*Igacovirus*	APVIVQEDPFSGHSSHAQLSANTLHGYR DASDNLYSQECQ
NC_010646	31.8%	4.41e-02	k119_22696	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	LGRLNAKGVELAESGIDALKLATLTKRIS DSTISGKAAKDVLDY
NC_010646	31.4%	4.23e-08	k119_142548	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	VNRVDTPLDENGEWDFEALHAVDLELF NTQLTQLLNGEEVEIPFYNFEDGK

conclusão

NC_010646	25.5%	3.59e-04	k119_146985	*Beluga whale coronavirus SW1*	*Gammacoronavirus*	*Cegacovirus*	VDREETPLDENGNYDYESLYALDLELFN QQLQALLRGEEVELPRFNFSLGK
Não classificada							
Name	***% Identity***	**E *value***	***Query***	***Organism***	***Genus***	***Subgenus***	***Sequence***
NC_048212	39.6%	8.64e-03	k119_147053	Bat coronavirus	unclassified	-	LVLYMAGTFLAALTAVVASFLFPTELILK AGEISQTPPSSLAEVIKVLLTNMV
NC_034440	38.3%	7.88e-02	k119_53162	Bat coronavirus	unclassified	-	ANRARSFGSPFSMSVSTSFLKQISCSATA ALMANMAPPQLADEPTAL
NC_048212	34.8%	6.70e-02	k119_4709	Bat coronavirus	unclassified	-	FENTLLGYSYKIFNTAGQLIDSQNIKANN KSAHIKFQQKLSPGIYY

Table S6. Comparison of E163 sequences from tBLASTX with the BlastN tool

tBLASTX					BlastN						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_018871	31.0%	8.39e-02	*Rousettus bat coronavirus HKU10*	*Bacteroides* sp. A1C1	210	210	38%	5,00E-50	100.00%	4502190	CP036491.1
NC_028814	41.7%	2.14e-02	*BtRf-AlphaCoV/HuB2013*	*Parabacteroides goldsteinii*	209	209	99%	2,00E-49	73.98%	6881351	CP064353.1
NC_002645	41.0%	6.78e-02	*Human coronavirus 229E*	*Phocaeicola vulgatus*	849	849	99%	0.0	99.37%	5165891	CP013020.1
NC_002645	38.6%	1.78e-02	*Human coronavirus 229E*	*bacterium*	59.0	59.0	12%	3,00E-04	87.23%	1451127	CP064957.1
NC_028752	30.4%	9.08e-02	*Camel alphacoronavirus*	*Epilithonimonas vandammei*	925	925	98%	0.0	72.63%	3276391	CP034161.1
NC_032730	35.6%	8.30e-02	*Lucheng Rn rat coronavirus*	*Methanocaldococcus* sp. SG7	57.2	57.2	6%	0.003	78.48%	1532285	LR792632.1
NC_032730	46.7%	7.05e-02	*Lucheng Rn rat coronavirus*	---	---	---	---	---	---	---	---
NC_034972	37.2%	3.80e-02	*Coronavirus AcCoV-JC34*	*Bacteroides zhangwenhongi*	211	282	80%	6,00E-50	77.90%	5244518	CP059856.1
NC_034972	47.8%	9.76e-02	*Coronavirus AcCoV-JC34*	---	---	---	---	---	---	---	---
NC_030292	39.4%	2.58e-02	*Ferret coronavirus*	*Cohaesibacter* sp. ES.047	86.9	86.9	18%	6,00E-12	71.20%	5110332	LT907844.1
NC_030292	28.6%	2.85e-02	*Ferret coronavirus*	*Phocaeicola vulgatus*	350	350	97%	9,00E-92	83.74%	5165891	CP013020.1
NC_010438	37.5%	2.19e-02	*Miniopterus bat coronavirus HKU8*	*Phocaeicola coprophilus*	97.8	97.8	26%	2,00E-15	73.03%	4113613	CP069440.1
NC_028833	52.0%	8.81e-02	*BtNv-AlphaCoV/SC2013*	*Prevotella* sp. WR041	501	501	100%	3,00E-137	89.60%	3524789	AP024484.1
NC_046964	31.4%	9.99e-02	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Bacteroides xylanisolvens*	932	932	100%	0.0	97.63%	6425332	CP072216.1

NC_003436	33.3%	9.64e-02	Porcine epidemic diarrhea virus	Phocaeicola salanitronis DSM 18170	71.6	71.6	24%	4,00E-08	77.53%	4242803	CP002530.1
NC_003436	33.3%	4.32e-02	*Porcine epidemic diarrhea virus*	ssRNA phage SRR5467090_10	932	932	100%	0.0	89.77%	3453	BK014119.1
NC_003436	50.0%	3.10e-02	*Porcine epidemic diarrhea virus*	---	---	---	---	---	---	---	---
NC_005831	37.1%	1.22e-03	*Human coronavirus NL63*	*Aliarcobacter lanthieri*	785	28468	99%	0.0	86.56%	2276457	CP053839.1
NC_032107	32.7%	5.97e-02	*NL63-related bat coronavirus*	---	---	---	---	---	---	---	---
NC_048216	30.4%	9.80e-02	*NL63-related bat coronavirus*	*Bacteroidales bacterium* CF	208	208	98%	2,00E-49	72.16%	2657245	CP006772.1
NC_048216	46.4%	8.39e-02	*NL63-related bat coronavirus*	uncultured *Alphaproteobacteria bacterium*	178	178	82%	3,00E-40	76.21%	130338	MN990730.1
NC_002306	56.0%	1.16e-02	*Feline infectious peritonitis virus*	*Arcobacter aquimarinus*	752	752	99%	0.0	91.79%	2520808	CP030944.1
NC_002306	56.0%	1.61e-02	*Feline infectious peritonitis virus*	*Arcobacter ellisii*	954	954	99%	0.0	91.27%	2799949	CP032097.1
NC_002306	52.0%	3.89e-02	*Feline infectious peritonitis virus*	*Arcobacter cloacae*	654	654	99%	0.0	88.84%	2621880	CP053833.1
NC_028806	40.6%	7.97e-02	*Swine enteric coronavirus*	*Bacteroidetes bacterium*	254	254	80%	2,00E-62	68.79%	3966474	CP065006.1
NC_038861	36.8%	3.71e-02	*Transmissible gastroenteritis virus*	*Cloacibacterium caeni*	487	487	96%	6,00E-133	88.22%	2979886	OU015319.1
NC_038861	34.7%	6.28e-02	*Transmissible gastroenteritis virus*	*Sulfurospirillum multivorans*	128	128	48%	8,00E-25	73.06%	3181530	CP042966.1
NC_048211	41.9%	3.36e-02	*Wencheng Sm shrew coronavirus*	*Aliarcobacter cryaerophilus*	480	480	99%	7,00E-131	94.48%	2115546	CP026655.1
NC_006577	44.0%	9.06e-02	*Human coronavirus HKU1*	*Campylobacter hyointestinalis*	379	379	99%	3,00E-100	78.13%	1753385	LR698977.1
NC_006577	39.5%	2.87e-02	*Human coronavirus HKU1*	*Prevotella copri*	631	1203	100%	2,00E-176	98.62%	3748729	CP042464.1

NC_006577	34.1%	2.31e-02	*Human coronavirus HKU1*	*Bacteroides* sp. A1C1	1361	1361	100%	0.0	93.27%	4502190	CP036491.1
NC_012936	35.9%	8.55e-02	*Rat coronavirus Parker*	*Candida orthopsilosis* Co 90-125	63.5	63.5	18%	2,00E-05	84.21%	984	XM_003867053.1
NC_012936	32.4%	6.36e-02	*Rat coronavirus Parker*	---	---	---	---	---	---	---	---
NC_025217	35.7%	7.64e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Bacteroides intestinalis*	962	962	100%	0.0	98.55%	5785761	CP041379.1
NC_019843	33.9%	3.87e-03	*Middle East respiratory syndrome-related coronavirus*	*Comamonas aquatica*	961	961	100%	0.0	95.83%	3969186	CP072916.1
NC_038294	27.7%	9.92e-02	*Betacoronavirus England 1*	*Prevotella copri*	1852	1852	100%	0.0	98.77%	3748729	CP042464.1
NC_039207	32.6%	1.57e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Sulfurospirillum multivorans*	279	1135	99%	2,00E-70	75.06%	3181530	CP042966.1
NC_039207	31.4%	3.28e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Parabacteroides merdae*	1290	1290	100%	0.0	99.72%	4697288	CP072229.1
NC_030886	43.3%	3.00e-02	*Rousettus bat coronavirus*	---	---	---	---	---	---	---	---
NC_014470	38.9%	6.95e-02	*Bat coronavirus BM48-31/BGR/2008*	Curvibacter putative symbiont of Hydra magnipapillata	888	888	92%	0.0	76.04%	564641	FN543107.1
NC_011547	44.1%	2.99e-02	*Bulbul coronavirus HKU11-934*	*Lachnospiraceae bacterium*	136	136	21%	1,00E-27	100.00%	3449685	LR699005.1
NC_016996	42.3%	4.07e-02	*Common moorhen coronavirus HKU21*	*Aliarcobacter cibarius*	632	632	100%	8,00E-177	90.77%	2336659	CP054051.1
NC_039208	33.3%	4.56e-02	*Porcine coronavirus HKU15*	*Phocaeicola vulgatus*	1686	1686	100%	0.0	99.68%	5356744	CP068489.1
NC_016993	31.7%	5.67e-02	*Magpie-robin coronavirus HKU18*	Coprobacter secundus subsp. similis	349	349	90%	1,00E-91	81.37%	4171466	AP023322.1

NC_016993	61.9%	4.36e-02	*Magpie-robin coronavirus HKU18*	*CrAss-like virus* sp.	471	653	100%	1,00E-128	99.25%	188037	BK035443.1
NC_016994	45.0%	9.87e-02	*Night heron coronavirus HKU19*	*Arcobacter venerupis*	266	266	47%	1,00E-66	87.96%	3209085	CP053840.1
NC_016994	76.9%	1.72e-02	*Night heron coronavirus HKU19*	*Arcobacter cloacae*	730	730	99%	0.0	89.05%	2621880	CP053833.1
NC_016994	48.5%	5.73e-02	*Night heron coronavirus HKU19*	*Fibrobacter succinogenes* subsp. succinogenes S85	224	224	95%	3,00E-54	75.60%	3843004	CP002158.1
NC_010646	39.5%	1.13e-05	*Beluga whale coronavirus SW1*	*Prevotella copri*	1278	1278	81%	0.0	96.88%	3748729	CP042464.1
NC_010646	38.2%	4.16e-03	*Beluga whale coronavirus SW1*	*Bacteroides ovatus*	86.0	86.0	51%	2,00E-12	70.93%	6571392	CP046397.1
NC_010646	37.8%	1.09e-04	*Beluga whale coronavirus SW1*	*Prevotella copri*	478	478	99%	4,00E-130	81.44%	3748729	CP042464.1
NC_010646	25.5%	4.89e-03	*Beluga whale coronavirus SW1*	*Bacteroides uniformis CL03T12C37*	970	970	98%	0.0	98.39%	4874733	CP072255.1
NC_010646	45.5%	2.20e-05	*Beluga whale coronavirus SW1*	*Phocaeicola vulgatus*	1373	1373	99%	0.0	100.00%	5165891	CP013020.1
NC_001451	46.4%	2.90e-02	*Infectious bronchitis virus*	*Bacteroides xylanisolvens*	588	588	100%	2,00E-163	98.52%	6499434	CP041230.1
NC_010800	22.2%	1.71e-02	*Turkey coronavirus*	*Sulfurospirillum* sp. UCH001	757	757	100%	0.0	97.96%	2606563	AP014723.1
NC_046965	35.7%	2.54e-02	*Canada goose coronavirus*	*Bacteroides intestinalis*	181	181	98%	3,00E-41	70.18%	5785761	CP041379.1
NC_046965	58.3%	3.70e-03	*Canada goose coronavirus*	*Roseburia hominis*	649	649	99%	0.0	98.66%	3592125	LR699011.1
NC_046955	61.5%	1.01e-02	*Shrew coronavirus*	*Cyprideis torosa*	129	129	26%	3,00E-25	77.64%	561	OB737022.1
NC_034440	34.1%	5.12e-02	*Bat coronavirus*	*Bacteroides heparinolyticus*	432	432	97%	2,00E-116	79.96%	3608975	CP027234.1
NC_046954	33.3%	3.25e-02	*Rodent coronavirus*	*Aliarcobacter butzleri*	864	864	99%	0.0	86.92%	2347607	LT906455.1
NC_046954	52.0%	3.98e-02	*Rodent coronavirus*	*Rhabditophanes* sp. KR3021	59.0	59.0	6%	5,00E-04	90.48%	40753	LK995684.1

Table S7. Comparison of E163 sequences from tBLASTX with the BlastX tool

tBLASTX					BlastX						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_018871	31.0%	8.39e-02	*Rousettus bat coronavirus HKU10*	*Bacteroides* sp. CAG:545	120	120	77%	2,00E-31	74.36%	252	CCZ43958.1
NC_028814	41.7%	2.14e-02	*BtRf-AlphaCoV/HuB2013*	*Parabacteroides* sp.	193	193	86%	2,00E-59	100.00%	252	MBP7939102.1
NC_002645	41.0%	6.78e-02	*Human coronavirus 229E*	*Phocaeicola vulgatus*	330	330	99%	5,00E-109	100.00%	511	WP_118935698.1
NC_002645	38.6%	1.78e-02	*Human coronavirus 229E*	*Prevotella* sp. AM23-5	143	143	69%	9,00E-42	100.00%	86	WP_118312741.1
NC_028752	30.4%	9.08e-02	*Camel alphacoronavirus*	*Chryseobacterium senegalense*	757	757	94%	0.0	74.54%	1111	WP_055042094.1
NC_032730	35.6%	8.30e-02	*Lucheng Rn rat coronavirus*	---	---	---	---	---	---	---	---
NC_032730	46.7%	7.05e-02	*Lucheng Rn rat coronavirus*	---	---	---	---	---	---	---	---
NC_034972	37.2%	3.80e-02	*Coronavirus AcCoV-JC34*	*Prevotella copri*	185	185	82%	3,00E-58	100.00%	111	MBM0145659.1
NC_034972	47.8%	9.76e-02	*Coronavirus AcCoV-JC34*	*Sulfurospirillum* sp.	310	310	70%	1,00E-104	98.82%	170	MBP9613378.1
NC_030292	39.4%	2.58e-02	*Ferret coronavirus*	*Proteobacteria bacterium*	296	296	85%	2,00E-95	67.48%	339	MBS4774248.1
NC_030292	28.6%	2.85e-02	*Ferret coronavirus*	*Bacteroides* sp. CAG:98	227	227	100%	3,00E-66	99.10%	1144	CDF15912.1
NC_010438	37.5%	2.19e-02	*Miniopterus bat coronavirus HKU8*	*Alistipes* sp. ZOR0009	331	331	100%	2,00E-112	70.04%	231	WP_047449358.1
NC_028833	52.0%	8.81e-02	*BtNv-AlphaCoV/SC2013*	*Prevotella* sp.	247	247	99%	1,00E-74	94.35%	838	MBP8935331.1
NC_046964	31.4%	9.99e-02	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Bacteroides xylanisolvens*	361	361	99%	3,00E-121	96.70%	479	KAB6275256.1
NC_003436	33.3%	9.64e-02	Porcine epidemic diarrhea virus	*Phocaeicola massiliensis*	246	246	99%	6,00E-78	100.00%	438	WP_195196069.1

NC_003436	33.3%	4.32e-02	*Porcine epidemic diarrhea virus*	ssRNA phage SRR5467090_10	456	456	99%	5,00E-157	93.48%	546	DAD52491.1
NC_003436	50.0%	3.10e-02	*Porcine epidemic diarrhea virus*	---	---	---	---	---	---	---	---
NC_005831	37.1%	1.22e-03	*Human coronavirus NL63*	*Arcobacter* sp. FW59	275	507	99%	3,00E-81	81.92%	1767	RBQ30770.1
NC_032107	32.7%	5.97e-02	*NL63-related bat coronavirus*	*Leviviridae* sp.	100	100	57%	2,00E-23	53.57%	166	QDH91199.1
NC_048216	30.4%	9.80e-02	*NL63-related bat coronavirus*	*Bacteroidales bacterium*	196	196	87%	6,00E-58	83.33%	446	MBP9585107.1
NC_048216	46.4%	8.39e-02	*NL63-related bat coronavirus*	*Alphaproteobacteria bacterium*	163	163	79%	8,00E-49	96.25%	187	MBQ8661341.1
NC_002306	56.0%	1.16e-02	*Feline infectious peritonitis virus*	*Arcobacter aquimarinus*	324	324	100%	8,00E-108	96.57%	421	WP_129094264.1
NC_002306	56.0%	1.61e-02	*Feline infectious peritonitis virus*	*Arcobacter* sp.	445	445	97%	2,00E-154	98.64%	421	MBD3829903.1
NC_002306	52.0%	3.89e-02	*Feline infectious peritonitis virus*	*Deltaproteobacteria bacterium* HGW-Deltaproteobacteria-24	332	332	98%	5,00E-111	97.59%	421	PKN14532.1
NC_028806	40.6%	7.97e-02	*Swine enteric coronavirus*	*Hevea brasiliensis*	256	256	95%	5,00E-81	48.77%	281	KAF2282056.1
NC_038861	36.8%	3.71e-02	*Transmissible gastroenteritis virus*	*Cloacibacterium* sp.	256	256	96%	3,00E-81	97.64%	459	MBP8084671.1
NC_038861	34.7%	6.28e-02	*Transmissible gastroenteritis virus*	*Campylobacterales bacterium*	194	194	87%	2,00E-60	75.76%	170	HGW31711.1
NC_048211	41.9%	3.36e-02	*Wencheng Sm shrew coronavirus*	*Aliarcobacter cryaerophilus*	187	187	89%	2,00E-55	98.91%	414	WP_141047624.1
NC_006577	44.0%	9.06e-02	*Human coronavirus HKU1*	*Sulfuricurvum* sp. PD_MW2	293	293	99%	1,00E-96	89.74%	359	PHM16577.1
NC_006577	39.5%	2.87e-02	*Human coronavirus HKU1*	*Prevotella copri*	243	296	99%	3,00E-73	99.17%	859	WP_118066204.1
NC_006577	34.1%	2.31e-02	*Human coronavirus HKU1*	*Bacteroides uniformis*	476	476	79%	1,00E-166	97.10%	356	WP_151880367.1

NC_012936	35.9%	8.55e-02	*Rat coronavirus Parker*	---	---	---	---	---	---	---	---
NC_012936	32.4%	6.36e-02	*Rat coronavirus Parker*	*Prevotella* sp. Marseille-P4119	78.6	78.6	78%	6,00E-15	47.19%	242	WP_106810860.1
NC_025217	35.7%	7.64e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Bacteroides*	328	328	88%	6,00E-113	100.00%	166	WP_021967221.1
NC_019843	33.9%	3.87e-03	*Middle East respiratory syndrome-related coronavirus*	*Comamonas*	178	178	44%	2,00E-54	98.88%	89	WP_042418916.1
NC_038294	27.7%	9.92e-02	*Betacoronavirus England 1*	*Prevotella* sp.	453	453	64%	4,00E-158	99.56%	251	MBN2917049.1
NC_039207	32.6%	1.57e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Sulfurospirillum multivorans*	209	209	99%	7,00E-63	77.54%	442	WP_025343811.1
NC_039207	31.4%	3.28e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Parabacteroides merdae*	454	454	93%	2,00E-157	99.56%	454	WP_055284858.1
NC_030886	43.3%	3.00e-02	*Rousettus bat coronavirus*	*Leviviridae* sp.	76.3	76.3	77%	2,00E-12	36.03%	542	QDH91152.1
NC_014470	38.9%	6.95e-02	*Bat coronavirus BM48-31/BGR/2008*	*Burkholderiales bacterium*	305	305	39%	9,00E-99	94.35%	231	MBN8556849.1
NC_011547	44.1%	2.99e-02	*Bulbul coronavirus HKU11-934*	[Eubacterium] rectale	51.6	51.6	27%	5,00E-05	87.10%	189	MBS6837328.1
NC_016996	42.3%	4.07e-02	*Common moorhen coronavirus HKU21*	*Arcobacter* sp. R-73987	317	317	99%	1,00E-100	95.36%	821	WP_198305419.1
NC_039208	33.3%	4.56e-02	*Porcine coronavirus HKU15*	*Salmonella enterica* subsp. enterica	324	324	50%	3,00E-108	100.00%	222	ECJ3915988.1
NC_016993	31.7%	5.67e-02	*Magpie-robin coronavirus HKU18*	*Eggerthella* sp. CAG:1427	218	218	87%	6,00E-63	100.00%	1052	CCY05671.1
NC_016993	61.9%	4.36e-02	*Magpie-robin coronavirus HKU18*	*CrAss-like virus* sp.	65.9	65.9	28%	7,00E-12	97.06%	34	DAH01655.1
NC_016994	45.0%	9.87e-02	*Night heron coronavirus HKU19*	*Arcobacter cloacae*	147	147	83%	7,00E-41	73.64%	229	WP_128985952.1
NC_016994	76.9%	1.72e-02	*Night heron coronavirus HKU19*	*Arcobacter aquimarinus*	360	360	100%	1,00E-122	91.94%	334	WP_129095768.1

conclusão

NC_016994	48.5%	5.73e-02	*Night heron coronavirus HKU19*	*Fibrobacteres bacterium* CG2_30_45_31	221	221	100%	8,00E-68	90.52%	453	OIP45784.1
NC_010646	39.5%	1.13e-05	*Beluga whale coronavirus SW1*	*Streptococcus* sp.	509	509	80%	2,00E-180	99.61%	273	MBN2940686.1
NC_010646	38.2%	4.16e-03	*Beluga whale coronavirus SW1*	*Paludibacteraceae bacterium*	211	211	98%	3,00E-63	89.09%	555	MBP7965346.1
NC_010646	37.8%	1.09e-04	*Beluga whale coronavirus SW1*	*Prevotella* sp.	319	319	99%	4,00E-104	97.59%	553	MBD9074338.1
NC_010646	25.5%	4.89e-03	*Beluga whale coronavirus SW1*	*Bifidobacterium pseudocatenulatum*	391	391	100%	4,00E-135	100.00%	339	MZS62412.1
NC_010646	45.5%	2.20e-05	*Beluga whale coronavirus SW1*	*Phocaeicola vulgatus*	520	520	100%	0.0	100.00%	557	KAB3854752.1
NC_001451	46.4%	2.90e-02	*Infectious bronchitis virus*	*Bacteroides* sp. D2	228	228	99%	9,00E-73	100.00%	290	EFS32121.1
NC_010800	22.2%	1.71e-02	*Turkey coronavirus*	*Sulfurospirillum* sp. UCH001	288	288	99%	5,00E-90	97.26%	806	WP_067177532.1
NC_046965	35.7%	2.54e-02	*Canada goose coronavirus*	*Phocaeicola paurosaccharolyticus*	186	186	99%	8,00E-52	100.00%	1116	WP_024992896.1
NC_046965	58.3%	3.70e-03	*Canada goose coronavirus*	*Roseburia* sp.	212	212	82%	2,00E-65	99.03%	372	HBD76780.1
NC_046955	61.5%	1.01e-02	*Shrew coronavirus*	*bacterium*	245	245	99%	8,00E-75	68.81%	562	MBO8439575.1
NC_034440	34.1%	5.12e-02	*Bat coronavirus*	*Bacteroides* sp.	280	280	95%	5,00E-94	95.51%	167	MBP9508060.1
NC_046954	33.3%	3.25e-02	*Rodent coronavirus*	*Aliarcobacter butzleri*	271	271	66%	7,00E-88	92.36%	273	WP_046996817.1
NC_046954	52.0%	3.98e-02	*Rodent coronavirus*	*Cetobacterium*	51.2	51.2	99%	0.002	25.41%	483	WP_023049678.1

Table S8. Comparison of E165 sequences from tBLASTX with the BlastN tool

tBLASTX					BlastN						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_010437	26.7%	9.85e-03	*Bat coronavirus 1A*	---	---	---	---	---	---	---	---
NC_022103	27.0%	5.56e-02	*Bat coronavirus CDPHE15/USA/2006*	*Bacteroides xylanisolvens*	613	613	100%	5,00E-171	99.71%	6499434	CP041230.1
NC_028833	60.0%	2.34e-02	*BtNv-AlphaCoV/SC2013*	*Prevotella copri*	505	505	93%	2,00E-138	92.69%	3748729	CP042464.1
NC_028752	41.9%	1.83e-02	*Camel alphacoronavirus*	*Bacteroides ovatus*	907	907	100%	0.0	99.03%	6571392	CP046397.1
NC_028752	30.4%	7.45e-02	*Camel alphacoronavirus*	*Epilithonimonas vandammei*	847	847	99%	0.0	74.90%	3276391	CP034161.1
NC_002306	61.9%	7.33e-02	*Feline infectious peritonitis virus*	*Arcobacter butzleri JV22*	1095	1095	100%	0.0	97.97%	2297844	CP040507.1
NC_002306	56.0%	3.50e-03	*Feline infectious peritonitis virus*	---	---	---	---	---	---	---	---
NC_002306	52.0%	2.87e-02	*Feline infectious peritonitis virus*	*Arcobacter cloacae*	531	531	96%	2,00E-146	89.47%	2621880	CP053833.1
NC_002306	31.6%	4.12e-03	*Feline infectious peritonitis virus*	*Tenacibaculum maritimum NCIMB 2154*	51.8	51.8	10%	0.043	86.05%	3435971	LT634361.1
NC_032730	45.8%	4.75e-02	*Lucheng Rn rat coronavirus*	*Prevotella copri*	316	316	98%	7,00E-82	77.28%	3748729	CP042464.1
NC_032730	35.6%	7.82e-02	*Lucheng Rn rat coronavirus*	*Methanocaldococcus* sp. SG7	57.2	57.2	6%	0.003	78.48%	1532285	LR792632.1
NC_010438	45.8%	2.06e-02	*Miniopterus bat coronavirus HKU8*	*Bacteroides fragilis*	132	132	97%	2,00E-26	67.34%	5330584	CP043610.1
NC_023760	36.6%	4.48e-02	*Mink coronavirus strain WD1127*	---	---	---	---	---	---	---	---
NC_023760	35.0%	3.72e-02	*Mink coronavirus strain WD1127*	*Bacteroides ovatus*	1764	1764	84%	0.0	93.79%	6571392	CP046397.1
NC_032107	57.1%	2.08e-02	*NL63-related bat coronavirus*	*Odoribacter splanchnicus*	599	599	100%	1,00E-166	99.70%	4392177	LT906459.1
NC_003436	33.3%	7.62e-02	*Porcine epidemic diarrhea virus*	*Phocaeicola salanitronis* DSM 18170	59.0	59.0	16%	2,00E-04	84.62%	4242803	CP002530.1

NC_009988	34.7%	1.35e-02	*Rhinolophus bat coronavirus HKU2*	*Bacteroides eggerthii*	799	799	100%	0.0	97.66%	4228380	CP069794.1
NC_038861	61.1%	1.66e-02	*Transmissible gastroenteritis virus*	*Prevotella copri*	89.7	154	78%	2,00E-13	65.78%	3748729	CP042464.1
NC_048211	42.4%	6.24e-02	*Wencheng Sm shrew coronavirus*	---	---	---	---	---	---	---	---
NC_039207	39.7%	3.31e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Adhaeribacter radiodurans*	167	167	97%	7,00E-37	70.57%	6799362	CP055153.1
NC_006213	35.5%	3.32e-02	*Human coronavirus OC43*	*Pedobacter cryoconitis*	59.9	59.9	11%	5,00E-04	77.92%	5950539	CP014504.1
NC_001846	41.9%	1.68e-02	*Murine hepatitis virus*	*Phocaeicola vulgatus*	629	629	100%	7,00E-176	99.16%	5298614	CP072234.1
NC_004718	56.0%	7.65e-02	*SARS coronavirus Tor2*	---	---	---	---	---	---	---	---
NC_045512	31.4%	7.61e-02	*Severe acute respiratory syndrome coronavirus 2*	*Paludibacter propionicigenes* WB4	497	497	95%	6,00E-136	77.13%	3685504	CP002345.1
NC_011547	38.7%	8.02e-02	*Bulbul coronavirus HKU11-934*	*Bacteroides* sp. M10	220	220	81%	1,00E-52	79.13%	6746034	CP060488.1
NC_016993	35.8%	5.14e-02	*Magpie-robin coronavirus HKU18*	*Tannerella forsythia 92A2*	189	189	88%	4,00E-43	68.34%	3405521	CP003191.1
NC_039208	42.4%	7.75e-02	*Porcine coronavirus HKU15*	*Prevotella scopos JCM 17725*	173	173	93%	2,00E-38	71.83%	1859892	CP016204.1
NC_039208	33.3%	1.36e-02	*Porcine coronavirus HKU15*	*Phocaeicola vulgatus*	678	678	100%	0.0	99.22%	5010342	CP043529.1
NC_016991	46.4%	9.56e-02	*White-eye coronavirus HKU16*	*Prevotella multiformis*	380	380	100%	6,00E-101	83.70%	1390346	CP072358.1
NC_010646	71.4%	3.31e-04	*Beluga whale coronavirus SW1*	*Alistipes shahii* WAL 8301	499	499	72%	1,00E-136	90.16%	3763317	FP929032.1
NC_010646	39.5%	2.30e-05	*Beluga whale coronavirus SW1*	*Prevotella copri*	3012	3180	100%	0.0	97.12%	3748729	CP042464.1
NC_010646	37.8%	1.71e-04	*Beluga whale coronavirus SW1*	*Prevotella copri*	403	403	93%	1,00E-107	74.29%	3748729	CP042464.1
NC_010646	36.1%	3.12e-03	*Beluga whale coronavirus SW1*	*Prevotella copri*	904	904	99%	0.0	84.26%	3748729	CP042464.1

conclusão

NC_010646	35.1%	3.92e-03	*Beluga whale coronavirus SW1*	*Parabacteroides merdae*	201	201	96%	4,00E-47	72.34%	4697288	CP072229.1
NC_048214	50.0%	4.35e-02	*Duck coronavirus*	*Cloacibacterium normanense*	541	541	99%	4,00E-149	87.41%	2759562	CP034157.1
NC_001451	35.0%	6.07e-02	*Infectious bronchitis virus*	*Bacteroides uniformis* CL03T12C37	602	602	95%	1,00E-167	98.55%	4874733	CP072255.1
NC_001451	30.2%	3.13e-02	*Infectious bronchitis virus*	*Roseburia intestinalis* L1-82	60.8	60.8	20%	9,00E-05	75.00%	4493348	LR027880.1
NC_010646	71.4%	3.31e-04	*Beluga whale coronavirus SW1*	*Alistipes shahii* WAL 8301	499	499	72%	1,00E-136	90.16%	3763317	FP929032.1
NC_048212	33.3%	7.62e-02	*Bat coronavirus*	*Prevotella copri*	1658	2477	93%	0.0	92.86%	3748729	CP042464.1
NC_046965	27.3%	6.92e-02	Canada goose coronavirus	*Methanoregula boonei* 6A8	545	545	100%	3,00E-150	91.09%	2542943	CP000780.1
NC_046954	63.0%	7.77e-03	Rodent coronavirus	*Flavobacterium indicum* GPTSA100-9 = DSM 17447	718	718	100%	0.0	99.75%	2993089	HE774682.1

Table S9. Comparison of E165 sequences from tBLASTX with the BlastX tool

tBLASTX					BlastX						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_010437	26.7%	9.85e-03	*Bat coronavirus 1A*	---	---	---	---	---	---	---	---
NC_022103	27.0%	5.56e-02	*Bat coronavirus CDPHE15/USA/2006*	*Bacteroides xylanisolvens*	226	226	99%	4,00E-74	100.00%	129	WP_202037006.1
NC_028833	60.0%	2.34e-02	*BtNv-AlphaCoV/SC2013*	*Prevotella copri*	228	228	92%	4,00E-74	99.11%	191	MQN34665.1
NC_028752	41.9%	1.83e-02	*Camel alphacoronavirus*	*Bacteroides ovatus*	332	332	95%	2,00E-114	100.00%	179	KXT47549.1
NC_028752	30.4%	7.45e-02	*Camel alphacoronavirus*	*Epilithonimonas hominis*	649	649	99%	0.0	78.92%	1100	WP_089768019.1
NC_002306	61.9%	7.33e-02	*Feline infectious peritonitis virus*	*Aliarcobacter butzleri*	345	345	86%	2,00E-115	98.37%	420	WP_151942821.1
NC_002306	56.0%	3.50e-03	*Feline infectious peritonitis virus*	*Deltaproteobacteria bacterium HGW-Deltaproteobacteria-24*	199	199	99%	6,00E-60	96.00%	421	PKN14532.1
NC_002306	52.0%	2.87e-02	*Feline infectious peritonitis virus*	*Deltaproteobacteria bacterium HGW-Deltaproteobacteria-24*	263	263	95%	3,00E-84	97.71%	421	PKN14532.1
NC_002306	31.6%	4.12e-03	*Feline infectious peritonitis virus*	---	---	---	---	---	---	---	---
NC_032730	45.8%	4.75e-02	*Lucheng Rn rat coronavirus*	*Prevotella* sp. CAG:386	268	268	99%	2,00E-88	89.78%	272	CDC28926.1
NC_032730	35.6%	7.82e-02	*Lucheng Rn rat coronavirus*	---	---	---	---	---	---	---	---
NC_010438	45.8%	2.06e-02	*Miniopterus bat coronavirus HKU8*	*Prevotella* sp. AM34-19LB	281	281	99%	2,00E-85	99.26%	1115	WP_119248859.1
NC_023760	36.6%	4.48e-02	*Mink coronavirus strain WD1127*	*Campylobacterales bacterium*	60.1	60.1	62%	4,00E-08	35.42%	145	HGW30632.1
NC_023760	35.0%	3.72e-02	*Mink coronavirus strain WD1127*	*Bacteroides*	502	502	53%	2,00E-175	97.95%	277	WP_008020874.1
NC_032107	57.1%	2.08e-02	*NL63-related bat coronavirus*	*Odoribacter splanchnicus*	195	195	83%	4,00E-60	100.00%	260	WP_013612994.1

NC_003436	33.3%	7.62e-02	*Porcine epidemic diarrhea virus*	*Phocaeicola massiliensis*	214	214	99%	1,00E-65	100.00%	438	WP_018452033.1
NC_009988	34.7%	1.35e-02	*Rhinolophus bat coronavirus HKU2*	*Siphoviridae* sp.	197	197	68%	8,00E-62	93.52%	138	DAI35265.1
NC_038861	61.1%	1.66e-02	*Transmissible gastroenteritis virus*	*Escherichia coli*	263	263	99%	1,00E-84	95.14%	396	MBL1008794.1
NC_048211	42.4%	6.24e-02	*Wencheng Sm shrew coronavirus*	*Hydrogenobaculum* sp.	85.5	85.5	86%	3,00E-17	42.45%	251	HHD12339.1
NC_039207	39.7%	3.31e-02	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*bacterium*	203	203	98%	5,00E-62	76.23%	357	MBO8440687.1
NC_006213	35.5%	3.32e-02	*Human coronavirus OC43*	*Prevotella* sp. AM42-24	234	234	52%	6,00E-76	100.00%	115	WP_147344092.1
NC_001846	41.9%	1.68e-02	*Murine hepatitis virus*	*Phocaeicola* sp.	244	244	99%	7,00E-81	100.00%	163	MBP8076174.1
NC_004718	56.0%	7.65e-02	*SARS coronavirus Tor2*	---	---	---	---	---	---	---	---
NC_045512	31.4%	7.61e-02	*Severe acute respiratory syndrome coronavirus 2*	*Bacteroidetes bacterium*	306	306	67%	3,00E-103	96.82%	182	MBP1663050.1
NC_011547	38.7%	8.02e-02	*Bulbul coronavirus HKU11-934*	*Barnesiella intestinihominis*	192	192	92%	9,00E-60	100.00%	228	WP_008860840.1
NC_016993	35.8%	5.14e-02	*Magpie-robin coronavirus HKU18*	*Parabacteroides* sp.	426	426	99%	2,00E-145	100.00%	551	MBP8760825.1
NC_039208	42.4%	7.75e-02	*Porcine coronavirus HKU15*	*Prevotella* sp.	213	213	97%	3,00E-65	91.96%	407	MBP7358749.1
NC_039208	33.3%	1.36e-02	*Porcine coronavirus HKU15*	*Salmonella enterica* subsp. enterica	239	239	91%	6,00E-78	100.00%	222	ECJ3915988.1
NC_016991	46.4%	9.56e-02	*White-eye coronavirus HKU16*	*Prevotella stercorea*	254	254	99%	3,00E-81	100.00%	408	WP_007900564.1
NC_010646	71.4%	3.31e-04	*Beluga whale coronavirus SW1*	*Prevotella stercorea*	265	265	99%	4,00E-83	80.72%	553	MBL6517762.1
NC_010646	39.5%	2.30e-05	*Beluga whale coronavirus SW1*	*Prevotella copri*	803	803	71%	0.0	99.77%	553	WP_119228984.1
NC_010646	37.8%	1.71e-04	*Beluga whale coronavirus SW1*	*Prevotellamassilia timonensis*	356	356	91%	2,00E-117	79.82%	557	WP_022101645.1

conclusão

NC_010646	36.1%	3.12e-03	*Beluga whale coronavirus SW1*	*Prevotella*	536	536	99%	0.0	99.64%	553	WP_117662607.1
NC_010646	35.1%	3.92e-03	*Beluga whale coronavirus SW1*	*Macellibacteroides fermentans*	267	267	99%	1,00E-84	100.00%	556	WP_179399854.1
NC_048214	50.0%	4.35e-02	*Duck coronavirus*	*Cloacibacterium normanense*	296	296	99%	1,00E-94	98.62%	666	WP_069796443.1
NC_001451	35.0%	6.07e-02	*Infectious bronchitis virus*	*Bacteroides uniformis*	217	217	87%	2,00E-69	99.06%	210	KXT38435.1
NC_001451	30.2%	3.13e-02	*Infectious bronchitis virus*	*Roseburia* sp. *OM04-10BH*	249	249	99%	1,00E-81	98.61%	211	WP_158567193.1
NC_010646	71.4%	3.31e-04	*Beluga whale coronavirus SW1*	*Prevotella stercorea*	265	265	99%	4,00E-83	80.72%	553	MBL6517762.1
NC_048212	33.3%	7.62e-02	*Bat coronavirus*	*Prevotella copri*	761	761	93%	0.0	97.59%	415	MQO97323.1
NC_046965	27.3%	6.92e-02	Canada goose coronavirus	*Methanoregula boonei*	200	200	76%	5,00E-64	96.00%	108	WP_012106080.1
NC_046954	63.0%	7.77e-03	Rodent coronavirus	*Flavobacterium indicum*	219	219	91%	4,00E-71	100.00%	147	WP_014387213.1

Table S10. Comparison of E205 sequences from tBLASTX with the BlastN tool

tBLASTX					BlastN						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_032107	64.3%	4.60e-02	*NL63-related bat coronavirus*	*Porphyromonas* sp. oral taxon 275	305	305	99%	4,00E-78	76.92%	2180921	CP072333.1
NC_048216	39.1%	1.72e-02	*NL63-related bat coronavirus*	*Sulfurospirillum* sp. UCH001	317	317	99%	5,00E-82	83.17%	2606563	AP014723.1
NC_048216	32.7%	8.12e-03	*NL63-related bat coronavirus*	*Neisseria dentiae*	244	244	100%	3,00E-60	78.37%	2755930	CP059570.1
NC_032107	30.3%	7.03e-02	*NL63-related bat coronavirus*	*Arcobacter ellisii*	1926	1926	100%	0.0	91.17%	2799949	CP032097.1
NC_002306	56.0%	2.14e-02	*Feline infectious peritonitis virus*	*Arcobacter ellisii*	338	338	76%	2,00E-88	88.55%	2799949	CP032097.1
NC_030292	39.4%	4.81e-03	*Ferret coronavirus*	---	---	---	---	---	---	---	---
NC_028752	30.4%	4.09e-02	*Camel alphacoronavirus*	*Epilithonimonas vandammei*	590	590	99%	5,00E-164	76.06%	3276391	CP034161.1
NC_018871	28.0%	6.11e-02	*Rousettus bat coronavirus HKU10*	*Prevotella* sp. WR041	860	860	100%	0.0	90.30%	3524789	AP024484.1
NC_009020	57.9%	9.16e-02	*Pipistrellus bat coronavirus HKU5*	*Bacteroides uniformis*	746	746	100%	0.0	99.52%	4734883	AP019724.1
AC_000192	51.9%	4.60e-02	*Murine hepatitis virus strain JHM*	*Chromobacterium violaceum*	83.3	83.3	27%	2,00E-11	78.85%	4757798	CP050992.1
NC_004718	45.5%	8.57e-02	*SARS coronavirus Tor2*	*Tolumonas auensis DSM 9187*	304	304	83%	4,00E-78	82.20%	3471292	CP001616.1
NC_009019	41.9%	3.73e-02	*Tylonycteris bat coronavirus HKU4*	*Aliarcobacter cibarius*	256	256	99%	2,00E-63	77.85%	2105974	CP043857.1
NC_030886	41.2%	1.56e-02	*Rousettus bat coronavirus*	*Ephemeroptericola cinctiostellae*	69.8	69.8	26%	1,00E-07	76.34%	2723407	CP031124.1
NC_030886	24.4%	8.66e-02	*Rousettus bat coronavirus*	*Ephemeroptericola cinctiostellae*	627	627	98%	5,00E-175	80.76%	2723407	CP031124.1
NC_025217	35.7%	5.93e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Bacteroides intestinalis*	887	887	100%	0.0	98.25%	5785761	CP041379.1
NC_045512	24.2%	5.80e-02	*Severe acute respiratory syndrome coronavirus 2*	*Cyprinus carpio*	59.0	59.0	8%	5,00E-04	83.61%	86312	LN594266.1

conclusão

NC_016995	28.9%	9.45e-02	*Wigeon coronavirus HKU20*	*Thauera* sp. MZ1T	96.0	96.0	19%	6,00E-15	77.50%	4496212	CP001281.2
NC_016994	41.7%	4.71e-03	*Night heron coronavirus HKU19*	*Chryseobacterium balustinum*	340	340	100%	5,00E-89	81.01%	4865895	CP033934.1
NC_001451	47.1%	9.08e-02	*Infectious bronchitis virus*	*Sulfurimonas* sp. CVO	402	402	100%	3,00E-107	75.08%	1923697	CP033720.1
NC_048213	43.5%	4.32e-02	*Infectious bronchitis virus*	*Oryzias latipes*	53.6	53.6	21%	0.011	76.62%	29950042	CP020791.1
NC_001451	28.0%	3.70e-02	*Infectious bronchitis virus*	[Arcobacter] porcinus	1068	1068	100%	0.0	94.72%	2018219	CP036246.2
NC_010646	25.4%	5.38e-02	*Beluga whale coronavirus SW1*	uncultured organism	361	361	71%	5,00E-95	92.98%	973	GQ877655.1
NC_046955	41.7%	4.33e-02	Shrew coronavirus	*Burkholderia pseudomallei*	83.3	83.3	17%	5,00E-11	73.19%	3257170	CP046583.1

Table S11. Comparison of E208 sequences from tBLASTX with the BlastN tool

tBLASTX					BlastN						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_018871	40.5%	3.55e-02	*Rousettus bat coronavirus HKU10*	*Cloacibacterium caeni*	549	549	98%	5,00E-152	98.73%	2979886	OU015319.1
NC_018871	25.3%	4.59e-02	*Rousettus bat coronavirus HKU10*	*Acinetobacter towneri*	567	567	100%	2,00E-157	97.60%	2789147	CP046045.1
NC_028814	20.0%	4.04e-02	*BtRf-AlphaCoV/HuB2013*	*Parabacteroides merdae*	469	469	98%	4,00E-127	71.83%	4697288	CP072229.1
NC_002645	53.8%	3.17e-02	*Human coronavirus 229E*	*Streptococcus suis*	620	620	100%	4,00E-173	98.60%	80407	KX077886.1
NC_028752	33.3%	9.25e-03	*Camel alphacoronavirus*	*Bacteroides ovatus V975*	1038	1038	99%	0.0	96.51%	6475296	LT622246.1
NC_028752	30.4%	7.68e-02	*Camel alphacoronavirus*	*Epilithonimonas vandammei*	825	825	98%	0.0	72.91%	3276391	CP034161.1
NC_028752	30.2%	6.66e-02	*Camel alphacoronavirus*	*Sulfurospirillum halorespirans* DSM 13726	86.9	86.9	13%	4,00E-12	78.10%	3029840	CP017111.1
NC_032730	55.0%	4.79e-02	*Lucheng Rn rat coronavirus*	---	---	---	---	---	---	---	---
NC_034972	52.2%	7.21e-02	*Coronavirus AcCoV-JC34*	*Parabacteroides merdae*	455	455	99%	2,00E-123	78.51%	4697288	CP072229.1
NC_023760	52.0%	1.45e-02	*Mink coronavirus strain WD1127*	*Aliarcobacter cryaerophilus*	702	702	100%	0.0	99.00%	160910	MK715471.1
NC_030292	39.4%	5.74e-03	*Ferret coronavirus*	---	---	---	---	---	---	---	---
NC_030292	35.7%	7.06e-02	*Ferret coronavirus*	*Bacteroides caecimuris*	1026	1026	96%	0.0	73.32%	4839927	CP015401.2
NC_046964	45.8%	7.61e-02	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Streptococcus respiraculi*	285	285	98%	3,00E-72	80.91%	2041601	CP022680.1
NC_009657	62.5%	8.38e-03	*Scotophilus bat coronavirus 512*	*Delftia acidovorans*	405	538	49%	8,00E-108	77.09%	6528979	CP022656.1
NC_048216	35.6%	4.60e-02	*NL63-related bat coronavirus*	---	---	---	---	---	---	---	---
NC_048216	32.6%	6.82e-02	*NL63-related bat coronavirus*	*Prevotella denticola*	393	506	59%	2,00E-104	79.30%	3027740	CP073339.1

NC_038861	36.8%	8.06e-02	*Transmissible gastroenteritis virus*	*Cloacibacterium normanense*	866	866	99%	0.0	87.92%	2759562	CP034157.1
NC_003045	56.3%	1.33e-02	Bovine coronavirus	*Sulfurospirillum* sp. JPD-1	361	361	99%	2,00E-95	77.39%	2814086	CP023275.1
NC_006577	57.1%	3.81e-03	*Human coronavirus HKU1*	---	---	---	---	---	---	---	---
NC_012936	32.4%	5.07e-02	*Rat coronavirus Parker*	---	---	---	---	---	---	---	---
NC_026011	73.7%	7.94e-02	*Betacoronavirus HKU24*	*Microvirgula aerodenitrificans*	193	193	85%	2,00E-44	71.58%	4121243	CP028519.1
NC_017083	37.8%	6.84e-02	*Rabbit coronavirus HKU14*	*Prevotella* sp. WR041	529	529	99%	6,00E-146	90.21%	3524789	AP024484.1
NC_025217	41.3%	4.23e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	---	---	---	---	---	---	---	---
NC_025217	35.1%	3.14e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Sulfurimonas autotrophica* DSM 16294	103	103	16%	7,00E-17	73.78%	2153198	CP002205.1
NC_025217	46.9%	3.19e-03	*Bat Hp-betacoronavirus/Zhejiang2013*	*Tolumonas auensis* DSM 9187	124	124	47%	7,00E-24	77.02%	3471292	CP001616.1
NC_025217	46.2%	5.13e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Sulfurospirillum* sp. JPD-1	361	361	99%	2,00E-95	77.39%	2814086	CP023275.1
NC_009019	41.9%	4.11e-02	*Tylonycteris bat coronavirus HKU4*	*Aliarcobacter cibarius*	249	249	89%	2,00E-61	78.39%	2105974	CP043857.1
NC_039207	26.5%	1.38e-03	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Sulfurospirillum* sp. UCH001	490	490	100%	4,00E-134	95.72%	2606563	AP014723.1
NC_009021	38.2%	5.08e-02	*Rousettus bat coronavirus HKU9*	*Draconibacterium orientale*	243	243	68%	2,00E-59	74.46%	5132075	CP007451.1
NC_030886	43.3%	2.90e-02	*Rousettus bat coronavirus*	---	---	---	---	---	---	---	---
NC_004718	56.0%	7.65e-02	*SARS coronavirus Tor2*	---	---	---	---	---	---	---	---

conclusão

NC_045512	39.0%	7.96e-03	*Severe acute respiratory syndrome coronavirus 2*	---	---	---	---	---	---	---	---
NC_014470	42.1%	7.63e-02	*Bat coronavirus BM48-31/BGR/2008*	---	---	---	---	---	---	---	---
NC_016995	51.7%	6.11e-02	*Wigeon coronavirus HKU20*	[Eubacterium] rectale M104/1	571	571	93%	2,00E-158	95.82%	3698419	FP929043.1
NC_016992	28.9%	5.16e-02	*Sparrow coronavirus HKU17*	uncultured Rhodocyclaceae bacterium	337	562	100%	4,00E-88	91.21%	3463890	OU015322.1
NC_016994	41.7%	6.94e-03	*Night heron coronavirus HKU19*	*Chryseobacterium balustinum*	354	354	99%	3,00E-93	81.13%	4865895	CP033934.1
NC_016994	38.1%	4.25e-02	*Night heron coronavirus HKU19*	---	---	---	---	---	---	---	---
NC_010646	31.8%	4.41e-02	*Beluga whale coronavirus SW1*	*Sulfurospirillum* sp. ACSTCE	177	177	51%	2,00E-39	74.64%	2685870	CP045453.2
NC_010646	31.4%	4.23e-08	*Beluga whale coronavirus SW1*	*Paludibacter propionicigenes* WB4	161	161	95%	4,00E-35	68.16%	3685504	CP002345.1
NC_010646	25.5%	3.59e-04	*Beluga whale coronavirus SW1*	*Bacteroides thetaiotaomicron*	531	531	89%	1,00E-146	98.68%	6487685	CP012937.1
NC_001451	63.6%	1.45e-02	*Infectious bronchitis virus*	*Arcobacter cryaerophilus* D2610	416	416	100%	5,00E-112	98.74%	2055914	CP032825.1
NC_001451	50.0%	7.52e-02	*Infectious bronchitis virus*	*Bacteroides eggerthii*	531	531	97%	2,00E-146	89.11%	5008416	CP072227.1
NC_048213	40.0%	1.12e-02	*Infectious bronchitis virus*	---	---	---	---	---	---	---	---
NC_034440	38.3%	7.88e-02	Bat coronavirus	*Dysgonomonas* sp. HDW5B	255	255	83%	3,00E-63	70.42%	4402003	CP049858.1
NC_048212	39.6%	8.64e-03	Bat coronavirus	*Moraxella osloensis*	591	591	99%	2,00E-164	97.43%	2434688	CP014234.1
NC_048212	34.8%	6.70e-02	Bat coronavirus	*Cloacibacterium caeni*	911	911	99%	0.0	94.14%	2979886	OU015319.1

Table S12. Comparison of E205 sequences from tBLASTX with the BlastX tool

tBLASTX					BlastX						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_032107	64.3%	4.60e-02	*NL63-related bat coronavirus*	*Acetobacteroides hydrogenigenes*	265	265	99%	4,00E-82	100.00%	699	WP_131838092.1
NC_048216	39.1%	1.72e-02	*NL63-related bat coronavirus*	*Sulfurospirillum* sp.	204	204	99%	5,00E-62	99.01%	407	MBP9493279.1
NC_048216	32.7%	8.12e-03	*NL63-related bat coronavirus*	*Uruburuella suis*	88.6	88.6	43%	2,00E-19	100.00%	190	WP_132953697.1
NC_032107	30.3%	7.03e-02	*NL63-related bat coronavirus*	*bacterium*	622	622	66%	0.0	98.01%	322	MBU0924069.1
NC_002306	56.0%	2.14e-02	*Feline infectious peritonitis virus*	*Arcobacter cloacae*	167	167	76%	2,00E-47	93.10%	421	WP_129013918.1
NC_030292	39.4%	4.81e-03	*Ferret coronavirus*	---	---	---	---	---	---	---	---
NC_028752	30.4%	4.09e-02	*Camel alphacoronavirus*	*Epilithonimonas hominis*	444	444	99%	6,00E-145	80.73%	1100	WP_089768019.1
NC_018871	28.0%	6.11e-02	*Rousettus bat coronavirus HKU10*	*Prevotella* sp.	385	385	99%	8,00E-130	99.04%	509	MBP9984259.1
NC_009020	57.9%	9.16e-02	*Pipistrellus bat coronavirus HKU5*	*Bacteroides* sp.	234	234	91%	4,00E-72	100.00%	491	MBP9505997.1
AC_000192	51.9%	4.60e-02	*Murine hepatitis virus strain JHM*	---	---	---	---	---	---	---	---
NC_004718	45.5%	8.57e-02	*SARS coronavirus Tor2*	*Vibrio metoecus*	98.2	98.2	85%	3,00E-21	66.35%	402	WP_134988116.1
NC_009019	41.9%	3.73e-02	*Tylonycteris bat coronavirus HKU4*	*Aliarcobacter cryaerophilus*	160	160	98%	1,00E-44	77.57%	446	WP_066403549.1
NC_030886	41.2%	1.56e-02	*Rousettus bat coronavirus*	*Burkholderiaceae bacterium*	228	228	99%	7,00E-74	93.22%	187	MBS1175044.1
NC_030886	24.4%	8.66e-02	*Rousettus bat coronavirus*	*Formosimonas limnophila*	298	298	74%	6,00E-100	94.74%	171	WP_189493581.1
NC_025217	35.7%	5.93e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Bacteroides*	327	327	95%	2,00E-112	100.00%	166	WP_021967221.1
NC_045512	24.2%	5.80e-02	*Severe acute respiratory syndrome coronavirus 2*	---	---	---	---	---	---	---	---

conclusão

NC_016995	28.9%	9.45e-02	*Wigeon coronavirus HKU20*	*Gammaproteobacteria bacterium*	143	143	74%	2,00E-35	52.98%	1019	NCA88786.1
NC_016994	41.7%	4.71e-03	*Night heron coronavirus HKU19*	*Chryseobacterium* sp. OV259	222	222	98%	3,00E-67	89.83%	563	WP_034754649.1
NC_001451	47.1%	9.08e-02	*Infectious bronchitis virus*	*Sulfurovum* sp. UBA12169	366	366	99%	4,00E-124	90.31%	403	DAB40956.1
NC_048213	43.5%	4.32e-02	*Infectious bronchitis virus*	---	---	---	---	---	---	---	---
NC_001451	28.0%	3.70e-02	*Infectious bronchitis virus*	*Campylobacteraceae bacterium*	430	430	99%	3,00E-144	99.12%	747	NLC28610.1
NC_010646	25.4%	5.38e-02	*Beluga whale coronavirus SW1*	*Alistipes putredinis*	243	243	98%	1,00E-73	100.00%	775	WP_040293015.1
NC_046955	41.7%	4.33e-02	Shrew coronavirus	Brachymonas sp.	113	113	20%	9,00E-28	98.11%	104	MBP8596478.1

Table S13. Comparison of E208 sequences from tBLASTX with the BlastX tool

tBLASTX					BlastX						
Name	***% Identity***	**E *value***	***Organism***	***Scientific Name***	***Max Score***	***Total Score***	***Query Cover***	***E Value***	***% Identity***	***Acc Leng***	***Acession***
NC_018871	40.5%	3.55e-02	*Rousettus bat coronavirus HKU10*	*Cloacibacterium normanense*	216	216	99%	2,00E-69	99.05%	206	WP_069797379.1
NC_018871	25.3%	4.59e-02	*Rousettus bat coronavirus HKU10*	*Acinetobacter seohaensis*	223	223	97%	5,00E-71	100.00%	263	GIT82862.1
NC_028814	20.0%	4.04e-02	*BtRf-AlphaCoV/HuB2013*	*Parabacteroides* sp.	599	599	99%	0.0	99.33%	449	MBP7954536.1
NC_002645	53.8%	3.17e-02	*Human coronavirus 229E*	*Streptococcus suis*	200	200	91%	9,00E-59	99.08%	561	WP_212575947.1
NC_028752	33.3%	9.25e-03	*Camel alphacoronavirus*	*Bacteroides* sp.	440	440	99%	4,00E-152	100.00%	465	MBS5443997.1
NC_028752	30.4%	7.68e-02	*Camel alphacoronavirus*	*Epilithonimonas hominis*	671	671	97%	0.0	75.78%	1100	WP_089768019.1
NC_028752	30.2%	6.66e-02	*Camel alphacoronavirus*	*Campylobacterales bacterium*	256	256	84%	1,00E-81	55.14%	275	MBL6969885.1
NC_032730	55.0%	4.79e-02	*Lucheng Rn rat coronavirus*	*Beihai levi-like virus 13*	45.8	45.8	49%	0.045	34.41%	569	APG77058.1
NC_034972	52.2%	7.21e-02	*Coronavirus AcCoV-JC34*	*Porphyromonadacea e bacterium*	277	277	73%	6,00E-91	100.00%	287	HAD02976.1
NC_023760	52.0%	1.45e-02	*Mink coronavirus strain WD1127*	*Aliarcobacter cryaerophilus*	96.3	96.3	99%	4,00E-20	99.24%	3184	WP_148572025.1
NC_030292	39.4%	5.74e-03	*Ferret coronavirus*	*Proteobacteria bacterium*	106	106	83%	7,00E-25	70.21%	339	MBS4774248.1
NC_030292	35.7%	7.06e-02	*Ferret coronavirus*	*Parabacteroides chartae*	1228	1228	98%	0.0	99.33%	1181	WP_079683632.1
NC_046964	45.8%	7.61e-02	*Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015*	*Streptococcus minor*	212	212	99%	2,00E-66	97.09%	324	WP_124777702.1
NC_009657	62.5%	8.38e-03	*Scotophilus bat coronavirus 512*	*Delftia tsuruhatensis*	471	471	89%	5,00E-158	54.95%	502	WP_046239165.1
NC_048216	35.6%	4.60e-02	*NL63-related bat coronavirus*	---	---	---	---	---	---	---	---
NC_048216	32.6%	6.82e-02	*NL63-related bat coronavirus*	*Prevotella* sp.	315	315	58%	3,00E-102	99.34%	445	MBD9246311.1

NC_038861	36.8%	8.06e-02	*Transmissible gastroenteritis virus*	*Cloacibacterium rupense*	355	355	76%	3,00E-118	99.43%	459	WP_188616928.1
NC_003045	56.3%	1.33e-02	Bovine coronavirus	*Sulfurospirillum* sp.	259	259	99%	3,00E-79	94.81%	748	MBP9492197.1
NC_006577	57.1%	3.81e-03	*Human coronavirus HKU1*	Prolixibacteraceae bacterium	133	346	99%	6,00E-34	79.55%	743	KAF0236500.1
NC_012936	32.4%	5.07e-02	*Rat coronavirus Parker*	*Sulfurospirillum* sp. UCH001	199	199	99%	8,00E-62	99.01%	279	WP_067174393.1
NC_026011	73.7%	7.94e-02	*Betacoronavirus HKU24*	*Pseudogulbenkiania* sp. NH8B	195	195	89%	2,00E-58	72.18%	356	WP_014088352.1
NC_017083	37.8%	6.84e-02	*Rabbit coronavirus HKU14*	*Prevotella* sp. WR041	271	271	99%	5,00E-87	99.22%	465	BCS85441.1
NC_025217	41.3%	4.23e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Campylobacterales bacterium*	127	127	83%	1,00E-34	67.01%	174	HHB95291.1
NC_025217	35.1%	3.14e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	*Campylobacterales bacterium*	224	224	89%	2,00E-66	47.06%	419	NWF67418.1
NC_025217	46.9%	3.19e-03	*Bat Hp-betacoronavirus/Zhejiang2013*	*Tolumonas osonensis*	77.8	77.8	50%	6,00E-14	89.47%	948	WP_188026195.1
NC_025217	46.2%	5.13e-02	*Bat Hp-betacoronavirus/Zhejiang2013*	---	---	---	---	---	---	---	---
NC_009019	41.9%	4.11e-02	*Tylonycteris bat coronavirus HKU4*	*Aliarcobacter cryaerophilus*	167	167	97%	3,00E-47	78.38%	446	WP_066403549.1
NC_039207	26.5%	1.38e-03	*Betacoronavirus Erinaceus/VMC/DEU/2012*	*Sulfurospirillum* sp. UCH001	199	199	99%	8,00E-62	99.01%	279	WP_067174393.1
NC_009021	38.2%	5.08e-02	*Rousettus bat coronavirus HKU9*	*Paludibacter* sp.	317	317	100%	7,00E-102	81.22%	677	MBP7611604.1
NC_030886	43.3%	2.90e-02	*Rousettus bat coronavirus*	*Leviviridae* sp.	66.2	66.2	43%	2,00E-09	31.20%	133	QDH87428.1
NC_004718	56.0%	7.65e-02	*SARS coronavirus Tor2*	---	---	---	---	---	---	---	---

conclusão

NC_045512	39.0%	7.96e-03	*Severe acute respiratory syndrome coronavirus 2*	---	---	---	---	---	---	---	---
NC_014470	42.1%	7.63e-02	*Bat coronavirus BM48-31/BGR/2008*	ssRNA phage SRR5467137_1	46.2	46.2	51%	0.008	37.93%	373	DAD52517.1
NC_016995	51.7%	6.11e-02	*Wigeon coronavirus HKU20*	[Eubacterium] rectale ATCC 33656	196	196	85%	2,00E-62	90.83%	107	ACR76992.1
NC_016992	28.9%	5.16e-02	*Sparrow coronavirus HKU17*	*Rhodocyclaceae bacterium*	159	159	99%	2,00E-45	96.20%	362	MBP9655621.1
NC_016994	41.7%	6.94e-03	*Night heron coronavirus HKU19*	*Chryseobacterium* sp. OV259	235	235	99%	3,00E-72	91.13%	563	WP_034754649.1
NC_016994	38.1%	4.25e-02	*Night heron coronavirus HKU19*	---	---	---	---	---	---	---	---
NC_010646	31.8%	4.41e-02	*Beluga whale coronavirus SW1*	*Campylobacterales bacterium*	234	234	83%	8,00E-72	76.82%	478	HGW31130.1
NC_010646	31.4%	4.23e-08	*Beluga whale coronavirus SW1*	*Paludibacteraceae bacterium*	300	300	99%	5,00E-97	90.26%	555	MBP7965346.1
NC_010646	25.5%	3.59e-04	*Beluga whale coronavirus SW1*	*Bacteroides xylanisolvens*	216	216	99%	3,00E-68	95.54%	280	KAB6329534.1
NC_001451	63.6%	1.45e-02	*Infectious bronchitis virus*	*Aliarcobacter cryaerophilus*	152	152	99%	6,00E-44	98.73%	244	WP_151202722.1
NC_001451	50.0%	7.52e-02	*Infectious bronchitis virus*	*Bacteroides fluxus* YIT 12057	219	219	82%	6,00E-70	96.49%	203	EGF58629.1
NC_048213	40.0%	1.12e-02	*Infectious bronchitis virus*	---	---	---	---	---	---	---	---
NC_034440	38.3%	7.88e-02	*Bat coronavirus*	*Sphingobacteriia bacterium*	408	408	99%	8,00E-136	86.67%	729	NCA79705.1
NC_048212	39.6%	8.64e-03	*Bat coronavirus*	*Moraxella* sp.	221	221	99%	1,00E-71	97.41%	171	MBP7234351.1
NC_048212	34.8%	6.70e-02	*Bat coronavirus*	*Cloacibacterium* sp.	199	199	52%	7,00E-57	95.24%	594	MBP7356995.1

REFERÊNCIAS BIBLIOGRÁFICAS

ABBASI, K. CoviD-19: Social murder, they wrote-elected, unaccountable, and unrepentant. **BMJ**, v. 372, n. 314, p. 1–3, 3 fev. 2021.

ACHS, A.; GLASA, M.; ŠUBR, Z. Plum Pox Virus Genome-Based Vector Enables the Expression of Different Heterologous Polypeptides in *Nicotiana benthamiana* Plants. **Processes**, v. 10, n. 8, p. 1–11, 3 ago. 2022.

AHLAWAT, S.; ASHA; SHARMA, K. K. Immunological co-ordination between gut and lungs in SARS-CoV-2 infection. **Virus Research**, v. 286, p. 1–10, 1 set. 2020.

AHLÉN, G. et al. The SARS-CoV-2 N Protein Is a Good Component in a Vaccine. **Journal of Virology**, v. 94, n. 18, ago. 2020.

AHMED, W.; BERTSCH, P. M.; et al. Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. **Science of the Total Environment**, v. 739, p. 1–9, 15 out. 2020.

AHMED, W.; ANGEL, N.; et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. **Science of the Total Environment**, v. 728, p. 1–8, 1 ago. 2020.

ALONZI, D. S. et al. Iminosugar antivirals: The therapeutic sweet spot. **Biochemical Society Transactions**, v. 45, n. 2, p. 571–582, 15 abr. 2017.

AMOAH, I. D.; KUMARI, S.; BUX, F. Coronaviruses in wastewater processes: Source, fate and potential risks. **Environment International**, v. 143, p. 1–13, 1 out. 2020.

ASCOLI, C. A.; AGGELER, B. Overlooked benefits of using polyclonal antibodies. **BioTechniques**, v. 65, n. 3, p. 127–136, 1 set. 2018.

ASTUTI, I.; YSRAFIL. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. **Diabetes and Metabolic Syndrome: Clinical Research and Reviews**, v. 14, n. 4, p. 407–412, 1 jul. 2020.

AUSAR, S. et al. Forced degradation studies: an essential tool for the formulation development of vaccines. **Vaccine: Development and Therapy**, v. 3, p. 11–33, jun. 2013.

BAČNIK, K. et al. Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. **Water Research**, v. 177, p. 1–11, 15 jun. 2020.

BANERJEE, A. et al. Bats and coronaviruses. **Viruses**, v. 11, n. 41, p. 1–15, 1 jan. 2019.

BAR-OR, I. et al. Regressing SARS-CoV-2 Sewage Measurements Onto COVID-19 Burden in the Population: A Proof-of-Concept for Quantitative Environmental Surveillance. **Frontiers in Public Health**, v. 9, p. 1–7, 3 jan. 2022.

BERCHENKO, Y. et al. Estimation of polio infection prevalence from environmental surveillance data. **Science Translational Medicine**, v. 9, p. 1–8, 2017.

BIBBY, K.; VIAU, E.; PECCIA, J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. **Letters in Applied Microbiology**, v. 52, n. 4, p. 386–392, 2011.

BIVINS, A. et al. Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight against COVID-19. **Environmental Science and Technology**, v. 54, n. 13, p. 7754–7757, 7 jul. 2020.

BLANCO, A. et al. Glass Wool Concentration Optimization for the Detection of Enveloped and Non-enveloped Waterborne Viruses. **Food and Environmental Virology**, v. 11, n. 2, p. 184–192, 1 jun. 2019.

BLAWID, R.; NAGATA, T. Construction of an infectious clone of a plant RNA virus in a binary vector using one-step Gibson Assembly. **Journal of Virological Methods**, v. 222, p. 11–15, 2015.

BONANNO FERRARO, G. et al. Pepper Mild Mottle Virus as Indicator of Pollution: Assessment of Prevalence and Concentration in Different Water Environments in Italy. **Food and Environmental Virology**, v. 13, n. 1, p. 117–125, 1 mar. 2021.

BOSCH, D.; SCHOTS, A. Plant glycans: Friend or foe in vaccine development? **Expert Review of Vaccines**, v. 9, n. 8, p. 835–842, ago. 2010.

BRISTER, J. R. et al. NCBI viral Genomes resource. **Nucleic Acids Research**, v. 43, p. 571–577, 28 jan. 2015.

BURNETT, M. J. B.; BURNETT, A. C. Therapeutic recombinant protein production in plants: Challenges and opportunities. **Plants People Planet**, v. 2, n. 2, p. 121–132, 1 mar. 2020.

CABEÇA, T. K.; BELLEI, N. Human coronavirus NL-63 infection in a Brazilian patient suspected of H1N1 2009 influenza infection: Description of a fatal case. **Journal of Clinical Virology**, v. 53, n. 1, p. 82–84, jan. 2012.

CABEÇA, T. K.; GRANATO, C.; BELLEI, N. Epidemiological and clinical features of human coronavirus infections among different subsets of patients. **Influenza and other Respiratory Viruses**, v. 7, n. 6, p. 1040–1047, nov. 2013.

CAMPBELL, A. M. **Monoclonal antibody and immunosensor technology The production and application of rodent and human monoclonal antibodies**. [S.l.]: Elsevier, 1991. v. 23.

CANDIDO, D. S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. **Science**, v. 369, p. 1255–1260, 2020.

CANTALUPO, P. G. et al. Raw sewage harbors diverse viral populations. **mBio**, v. 2, n. 5, p. 1–11, 2011.

CANTÓN, R. et al. New variants of SARS-CoV-2. **Revista Espanola de Quimioterapia**, v. 34, n. 5, p. 419–428, 1 out. 2021.

CAPELL, T. et al. Potential Applications of Plant Biotechnology against SARS-CoV-2. **Trends in Plant Science**, v. 25, n. 7, p. 635–643, 1 jul. 2020.

CHAN, J. F. W. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. **The Lancet**, v. 395, n. 10223, p. 514–523, 15 fev. 2020.

CHEN, W. H. et al. **The SARS-CoV-2 Vaccine Pipeline: an Overview**. **Current Tropical Medicine Reports**. [S.l.]: Springer. , 1 jun. 2020

CHERIAN, S. et al. Sars-cov-2 spike mutations, l452r, t478k, e484q and p681r, in the second wave of covid-19 in Maharashtra, India. **Microorganisms**, v. 9, n. 1542, p. 1–11, 1 jul. 2021.

CHERKAOUI, D. et al. Harnessing recombinase polymerase amplification for rapid multi-gene detection of SARS-CoV-2 in resource-limited settings. **Biosensors and Bioelectronics**, v. 189, p. 1–11, 1 out. 2021.

CHILD, H. T. et al. Comparison of metagenomic and targeted methods for sequencing human pathogenic viruses from wastewater. **mBio**, v. 14, n. 6, p. 1–19, 19 dez. 2023.

CHOI, P. M. et al. Wastewater-based epidemiology biomarkers: Past, present and future. **TrAC - Trends in Analytical Chemistry**, v. 105, p. 453–469, 1 ago. 2018.

CLARO, I. C. M. et al. Long-term monitoring of SARS-COV-2 RNA in wastewater in Brazil: A more responsive and economical approach. **Water Research**, v. 203, p. 1–12, 15 set. 2021.

CORMAN, V. M. et al. Hosts and Sources of Endemic Human Coronaviruses. **Adv Virus Res**. [S.l.]: Academic Press Inc., 2018. v. 100. p. 163–188.

CORNEJO-SÁNCHEZ, T. et al. Epidemiology of GII.4 and GII.2 norovirus outbreaks in closed and semi-closed institutions in 2017 and 2018. **Scientific Reports**, v. 13, n. 1, p. 1–8, 1 dez. 2023.

CORPUZ, M. V. A. et al. Viruses in wastewater: occurrence, abundance and detection methods. **Science of the Total Environment**, v. 745, p. 1–26, 25 nov. 2020.

CORSI, S. R. et al. Human and bovine viruses in the Milwaukee River watershed: Hydrologically relevant representation and relations with environmental variables. **Science of the Total Environment**, v. 490, p. 849–860, 15 ago. 2014.

CRODA, J. et al. Covid-19 in Brazil: Advantages of a socialized unified health system and preparation to contain cases. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 53, p. 1–6, 2020.

DAUGHTON, C. G. Pharmaceuticals and Care Products in the Environment. **ACS Symposium Series**, v. 791, p. 1–38, 2001.

DE ARAUJO, J. C. et al. **SARS-CoV-2 sewage surveillance in low-income countries: Potential and challenges**. **Journal of Water and Health**. [S.l.]: IWA Publishing. , 1 fev. 2021a

DE FREITAS BUENO, R. et al. Wastewater-based epidemiology: A Brazilian SARS-COV-2 surveillance experience. **Journal of Environmental Chemical Engineering**, v. 10, n. 5, p. 1–9, 1 out. 2022.

DE JESUS, J. G. et al. Importation and early local transmission of covid-19 in brazil, 2020. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 62, 2020.

DE SABATO, L. et al. Full genome characterization of two novel Alpha-coronavirus species from Italian bats. **Virus Research**, v. 260, p. 60–66, 15 jan. 2019.

DECKER, E. L.; RESKI, R. Glycoprotein production in moss bioreactors. **Plant Cell Reports**, v. 31, n. 3, p. 453–460, mar. 2012.

DELAPLACE, M. et al. Feline coronavirus antivirals: A review. **Pathogens**, v. 10, n. 9, p. 1–16, 1 set. 2021.

DEMURTAS, O. C.; MASSA, S.; ILLIANO, E.; MARTINIS, D. de. Antigen Production in Plant to Tackle Infectious Diseases Flare Up : The Case of SARS. v. 7, n. February, p. 1–12, 2016.

DEMURTAS, O. C.; MASSA, S.; ILLIANO, E.; DE MARTINIS, D.; et al. Antigen production in plantto tackle infectious diseases flare up: The case of SARS. **Frontiers in Plant Science**, v. 7, p. 1–12, 5 fev. 2016.

DENG, X. et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. **Cell**, v. 184, n. 13, p. 3426–3437, 24 jun. 2021.

DHAMA, K. et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. **Journal of Infection and Public Health**, v. 16, n. 1, p. 4–14, 1 jan. 2023.

DIAMOS, A. G. et al. High Level Production of Monoclonal Antibodies Using an Optimized Plant Expression System. **Frontiers in Bioengineering and Biotechnology**, v. 7, n. 472, p. 1–15, 2020.

DIEGO-MARTIN, B. et al. Pilot Production of SARS-CoV-2 Related Proteins in Plants: A Proof of Concept for Rapid Repurposing of Indoor Farms Into Biomanufacturing Facilities. **Frontiers in Plant Science**, v. 11, n. 612781, p. 1–14, 23 dez. 2020.

DONINI, M.; MARUSIC, C. Current state-of-the-art in plant-based antibody production systems. **Biotechnology Letters**, v. 41, n. 3, p. 335–346, 15 mar. 2019.

DUARTE, M. F. et al. Metagenomic analyses of plant virus sequences in sewage water for plant viruses monitoring. **Tropical Plant Pathology**, v. 48, n. 4, p. 408–416, 1 ago. 2023.

EFTIM, S. E. et al. Occurrence of norovirus in raw sewage – A systematic literature review and meta-analysis. **Water Research**, v. 111, p. 366–374, 2017.

ELSAMADONY, M. et al. Possible transmission of viruses from contaminated human feces and sewage: Implications for SARS-CoV-2. **Science of the Total Environment**, v. 755, 10 fev. 2021.

EVERETT, K. M. et al. Development of a plant-made pharmaceutical production platform. **BioProcess International**, v. 10, n. 1, p. 1–10, 2012a.

EVERETT, K. M. et al. Development of a Plant-Made Pharmaceutical Production Platform . **BioProcess International**, v. 10, n. 1, p. 16–26, 2012b.

FARKAS, K. et al. Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. **Environmental Science and Pollution Research**, v. 25, n. 33, p. 33391–33401, 1 nov. 2018.

FERNANDEZ-CASSI, X. et al. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. **Science of the Total Environment**, v. 618, p. 870–880, 15 mar. 2018.

FISCHER, R.; BUYEL, J. F. Molecular farming – The slope of enlightenment. **Biotechnology Advances**, v. 40, p. 1–16, 1 maio 2020.

FONGARO, G. et al. The presence of SARS-CoV-2 RNA in human sewage in Santa Catarina, Brazil, November 2019. **Science of the Total Environment**, v. 778, p. 1–4, 15 jul. 2021.

FORTERRE, P.; PRANGISHVILI, D. The origin of viruses. **Research in Microbiology**, v. 160, n. 7, p. 466–472, 2009.

FOURATI, S. et al. Performance of six rapid diagnostic tests for SARS-CoV-2 antigen detection and implications for practical use. **Journal of Clinical Virology**, v. 142, p. 1–5, 1 set. 2021.
FUMIAN, T. M. et al. Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. **Environment International**, v. 123, p. 282–291, 1 fev. 2019.
GEISLER, C.; JARVIS, D. L. Adventitious viruses in insect cell lines used for recombinant protein expression. **Protein Expression and Purification**, v. 144, p. 25–32, 1 abr. 2018.
GERBA, C. P.; BETANCOURT, W. Q.; KITAJIMA, M. How much reduction of virus is needed for recycled water: A continuous changing need for assessment? **Water Research**, v. 108, p. 25–31, 1 jan. 2017.
GHAFFARI, A.; MEURANT, R.; ARDAKANI, A. COVID-19 serological tests: how well do they actually perform? **Diagnostics**, v. 10, n. 7, p. 1–14, 1 jul. 2020.
GORBALENYA, A. E. et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nature Microbiology**, v. 5, n. 4, p. 536–544, 1 abr. 2020.
GRALINSKI, L. E.; BARIC, R. S. Molecular pathology of emerging coronavirus infections. **Journal of Pathology**, v. 235, n. 2, p. 185–195, 1 jan. 2015.
GRIFONI, A. et al. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. **Cell Host and Microbe**, v. 27, n. 4, p. 671–680, 8 abr. 2020.
GUAJARDO-LEIVA, S. et al. Metagenomic insights into the sewage RNA virosphere of a large city. **Viruses**, v. 12, n. 9, p. 1–15, 1 set. 2020.
GUAN, W. et al. Clinical Characteristics of Coronavirus Disease 2019 in China. **New England Journal of Medicine**, v. 382, n. 18, p. 1708–1720, 30 abr. 2020.
GUNDY, P. M.; GERBA, C. P.; PEPPER, I. L. Survival of Coronaviruses in Water and Wastewater. **Food and Environmental Virology**, v. 1, n. 1, p. 10–14, 1 mar. 2009.
GYAWALI, P. et al. Evaluation of pepper mild mottle virus as an indicator of human faecal pollution in shellfish and growing waters. **Water Research**, v. 154, p. 370–376, 2019.
HATA, A.; HONDA, R.; HONDA, R. Potential Sensitivity of Wastewater Monitoring for SARS-CoV-2: Comparison with Norovirus Cases. **Environmental Science and Technology**, v. 54, n. 11, p. 6451–6452, 2 jun. 2020.
HAYAT, S. M. G. et al. Recombinant Protein Expression in Escherichia coli (E.coli): What We Need to Know. **Current Pharmaceutical Design**, v. 24, p. 718–725, 2018.

HENDRIKSEN, R. S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. **Nature Communications**, v. 10, n. 1, p. 1–12, 1 dez. 2019.
HIATT, A.; CAFFFERKEY, R.; BOWDISH, K. Production of antibodies in transgenic plants. **Nature**, v. 342, n. 6245, p. 76–78, nov. 1989.
HOFFMANN, M. et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. **Cell**, v. 184, n. 9, p. 2384–2393, abr. 2021.
HOLSHUE, M. L. et al. First Case of 2019 Novel Coronavirus in the United States. **New England Journal of Medicine**, v. 382, n. 10, p. 929–936, 5 mar. 2020.
HOULIHAN, C. F.; BEALE, R. The complexities of SARS-CoV-2 serology. **The Lancet Infectious Diseases**, v. 20, n. 12, p. 1350–1351, 1 dez. 2020.
HU, B. et al. Characteristics of SARS-CoV-2 and COVID-19. **Nature Reviews Microbiology**, v. 19, n. 3, p. 141–154, 1 mar. 2021.
HUANG, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. **The Lancet**, v. 395, n. 10223, p. 497–506, 15 fev. 2020.
HUANG, Y. et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. **Acta Pharmacologica Sinica**, v. 41, n. 9, p. 1141–1149, 1 set. 2020.
JACOB MACHADO, D. et al. Fundamental evolution of all Orthocoronavirinae including three deadly lineages descendent from Chiroptera-hosted coronaviruses: SARS-CoV, MERS-CoV and SARS-CoV-2. **Cladistics**, v. 37, n. 5, p. 461–488, 1 out. 2021.
JUNG, J. et al. Plant-based expression and characterization of SARS-CoV-2 virus-like particles presenting a native spike protein. **Plant Biotechnology Journal**, v. 20, n. 7, p. 1363–1372, 6 jul. 2022.
KANG, H. et al. Fusion of a highly N-glycosylated polypeptide increases the expression of ER-localized proteins in plants. **Scientific Reports**, v. 8, n. 1, p. 1–10, 1 dez. 2018.
KARBALAEI, M.; REZAEE, S. A.; FARSIANI, H. Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. **Journal of Cellular Physiology**, v. 235, p. 5867–5881, 2020.
KILIÇ, Z. Water Pollution: Causes, Negative Effects and Prevention Methods. **Journal of the Institute of Science and Technology**, v. 3, n. 2, p. 129–132, 31 ago. 2021.
KITAJIMA, M.; SASSI, H. P.; TORREY, J. R. **Pepper mild mottle virus as a water quality indicator**. **npj Clean Water**. [S.l.]: Nature Research. , dez. 2018

KLAPSA, D. et al. Sustained detection of type 2 poliovirus in London sewage between February and July, 2022, by enhanced environmental surveillance. **The Lancet**, v. 400, n. 10362, p. 1531–1538, 29 out. 2022.
KRONEMAN, A. et al. Proposal for a unified norovirus nomenclature and genotyping. **Archives of Virology**, v. 158, n. 10, p. 2059–2068, out. 2013.
LA ROSA, G.; BONADONNA, L.; et al. Coronavirus in water environments: Occurrence, persistence and concentration methods - A scoping review. **Water Research**, v. 179, p. 1–11, 15 jul. 2020.
LA ROSA, G.; IACONELLI, M.; et al. First detection of SARS-CoV-2 in untreated wastewaters in Italy. **Science of the Total Environment**, v. 736, p. 1–5, 20 set. 2020.
LAM, T. T. Y. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. **Nature**, v. 583, n. 7815, p. 282–285, 9 jul. 2020.
LAMBROU, A. S. et al. Genomic Surveillance for SARS-CoV-2 Variants: Predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) Variants — United States, June 2021–January 2022. **Morbidity and Mortality Weekly Report**, v. 71, n. 6, p. 206–211, 11 fev. 2022.
LARRICK, J. W. et al. Production of antibodies in transgenic plants. **Nature**, v. 342, p. 76–78, 1989.
LARSSON, D. G. J.; FLACH, C. F.; LAXMINARAYAN, R. Sewage surveillance of antibiotic resistance holds both opportunities and challenges. **Nature Reviews Microbiology**, v. 21, n. 4, p. 213–214, 1 abr. 2023.
LAU, S. K. P. et al. Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2. **Emerging Infectious Diseases**, v. 26, n. 7, p. 1542–1547, 1 jul. 2020.
LEE, C. Y. P. et al. Serological Approaches for COVID-19: Epidemiologic Perspective on Surveillance and Control. **Frontiers in Immunology**, v. 11, p. 1–7, 24 abr. 2020.
LEKNES, H.; STURTZEL, I. E.; DYE, C. Environmental release of oseltamivir from a Norwegian sewage treatment plant during the 2009 influenza A (H1N1) pandemic. **Science of the Total Environment**, v. 414, p. 632–638, 1 jan. 2012.
LI, D. et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. **Methods**, v. 102, p. 3–11, 1 jun. 2016.
LI, M. et al. An Induced Hypersensitive-Like Response Limits Expression of Foreign Peptides via a Recombinant TMV-Based Vector in a Susceptible Tobacco. **PLoS ONE**, v. 5, n. 11, p. 1–8, 29 nov. 2010.

LI, Q. et al. TMV recombinants encoding fused foreign transmembrane domains to the CP subunit caused local necrotic response on susceptible tobacco. **Virology**, v. 348, n. 2, p. 253–259, maio 2006.
LI, W. et al. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. **Science**, v. 310, n. 5748, p. 676–679, 28 out. 2005.
LICO, C. et al. Plant Molecular Farming as a Strategy Against COVID-19 – The Italian Perspective. **Frontiers in Plant Science**, v. 11, n. December, p. 1–10, 2020.
LINDSAY, P. et al. Rapid expression of COVID-19 proteins by transient expression in tobacco. Author contributions. **bioRxiv**, p. 1–17, 2020.
LU, X. et al. US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome Coronavirus 2. **Emerging Infectious Diseases**, v. 26, n. 8, p. 1654–1665, 1 ago. 2020.
MAHARJAN, P. M. et al. Plant-Expressed Receptor Binding Domain of the SARS-CoV-2 Spike Protein Elicits Humoral Immunity in Mice. **Vaccines**, v. 9, n. 978, p. 1–15, 1 set. 2021.
MAHMOOD, N.; NASIR, S. B.; HEFFERON, K. Plant-based drugs and vaccines for COVID-19. **Vaccines**, v. 9, n. 15, p. 1–16, 2021.
MAKATSA, M. S. et al. SARS-CoV-2 Antigens Expressed in Plants Detect Antibody Responses in COVID-19 Patients. **Frontiers in Plant Science**, v. 12, n. 589940, p. 1–13, 31 mar. 2021.
MAMEDOV, T. et al. Production and Characterization of Nucleocapsid and RBD Cocktail Antigens of SARS-CoV-2 in *Nicotiana benthamiana* Plant as a Vaccine Candidate against COVID-19. **Vaccines**, v. 9, n. 11, p. 1–15, 17 nov. 2021.
MANOR, Y. et al. Intensified environmental surveillance supporting the response to wild poliovirus type 1 silent circulation in Israel, 2013. **Eurosurveillance**, v. 19, n. 7, p. 1–10, 20 fev. 2014.
MARDANOVA, E. S.; KOTLYAROV, R. Y.; RAVIN, N. V. High-yield production of receptor binding domain of sars-cov-2 linked to bacterial flagellin in plants using self-replicating viral vector peff. **Plants**, v. 10, n. 12, p. 1–11, 1 dez. 2021.
MARGOLIN, E. et al. Augmenting glycosylation-directed folding pathways enhances the fidelity of HIV Env immunogen production in plants. **Biotechnology and Bioengineering**, v. 119, n. 10, p. 2919–2937, 1 out. 2022.

MARGOLIN, E.; VERBEEK, M.; et al. Calreticulin co-expression supports high level production of a recombinant SARS-CoV-2 spike mimetic in *Nicotiana benthamiana*. **bioRxiv**, p. 1–9, 2020.
MARGOLIN, E.; OH, Y. J.; et al. Co-expression of human calreticulin significantly improves the production of HIV gp140 and other viral glycoproteins in plants. **Plant Biotechnology Journal**, v. 18, n. 10, p. 2109–2117, 1 out. 2020.
MARGOLIN, E. et al. Production of complex viral glycoproteins in plants as vaccine immunogens. **Plant Biotechnology Journal**, v. 16, n. 9, p. 1531–1545, 1 set. 2018.
_____. Site-Specific Glycosylation of Recombinant Viral Glycoproteins Produced in *Nicotiana benthamiana*. **Frontiers in Plant Science**, v. 12, p. 1–12, 22 jul. 2021.
MARTIN, J. et al. Tracking SARS-CoV-2 in sewage: Evidence of changes in virus variant predominance during COVID-19 pandemic. **Viruses**, v. 12, n. 10, p. 1–17, 1 out. 2020.
MARTÍNEZ-PUCHOL, S. et al. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. **Science of the Total Environment**, v. 713, p. 1–9, 15 abr. 2020.
_____. Exploring the diversity of coronavirus in sewage during COVID-19 pandemic: Don't miss the forest for the trees. **Science of the Total Environment**, v. 800, p. 1–7, 15 dez. 2021.
MCBRIDE, R.; VAN ZYL, M.; FIELDING, B. C. The coronavirus nucleocapsid is a multifunctional protein. **Viruses**, v. 6, n. 8, p. 2991–3018, 7 ago. 2014.
MCCALL, C. et al. Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. **Water Research**, v. 184, p. 1–13, out. 2020.
MEDEMA, G. et al. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in the Netherlands. **Environmental Science and Technology Letters**, v. 7, n. 7, p. 511–516, 14 jul. 2020.
MENZEL, P.; NG, K. L.; KROGH, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. **Nature Communications**, v. 7, p. 1–9, 13 abr. 2016.
MEYER, B.; DROSTEN, C.; MÜLLER, M. A. Serological assays for emerging coronaviruses: Challenges and pitfalls. **Virus Research**, v. 194, p. 175–183, 19 dez. 2014.
MICHAEL-KORDATOU, I.; KARAOLIA, P.; FATTA-KASSINOS, D. Sewage analysis as a tool for the COVID-19 pandemic response and management: The urgent need for optimised protocols for SARS-CoV-2 detection and quantification. **Journal of Environmental Chemical Engineering**, v. 8, n. 5, p. 1–24, 1 out. 2020.

MONTAZERI, N. et al. Pathogenic enteric viruses and microbial indicators during secondary treatment of municipal wastewater. **Applied and Environmental Microbiology**, v. 81, n. 18, p. 6436–6445, 2015.

MOON, K.-B. et al. Construction of SARS-CoV-2 virus-like particles in plant. **Scientific Reports**, v. 12, n. 1, p. 1–7, 19 jan. 2022.

MORA, M. et al. Explorative assessment of coronavirus-like short sequences from host-associated and environmental metagenomes. **Science of the Total Environment**, v. 793, p. 1–5, 1 nov. 2021.

MOTA, C. R. et al. Assessing spatial distribution of COVID-19 prevalence in Brazil using decentralised sewage monitoring. **Water Research**, v. 202, p. 1–10, 1 set. 2021.

NAQUIN, A. et al. Presence of antibiotic resistance genes in a sewage treatment plant in Thibodaux, Louisiana, USA. **Bioresource Technology**, v. 188, p. 79–83, 1 jul. 2015.

NG, T. F. F. et al. High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage. **Journal of Virology**, v. 86, n. 22, p. 12161–12175, 15 nov. 2012.

NIEUWENHUIJSE, D. F. et al. Setting a baseline for global urban virome surveillance in sewage. **Scientific Reports**, v. 10, n. 1, 1 dez. 2020.

NING, L. et al. Development and application of therapeutic antibodies against covid-19. **International Journal of Biological Sciences**, v. 17, n. 6, p. 1486–1496, 2021.

NUNES, N. et al. Review of sewage sludge as a soil amendment in relation to current international guidelines: A heavy metal perspective. **Sustainability (Switzerland)**, v. 13, n. 4, p. 1–20, 2 fev. 2021.

O'CONNOR, S. E.; IMPERIALI, B. Modulation of protein structure and function by asparagine-linked glycosylation Chemistry & Biology. **Chemistry & Biology**, v. 3, n. 10, p. 803–812, 1996.

O'FLAHERTY, R. et al. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. **Biotechnology Advances**, v. 43, p. 1–17, 1 nov. 2020.

ORIVE, G.; LERTXUNDI, U.; BARCELO, D. Early SARS-CoV-2 outbreak detection by sewage-based epidemiology. **Science of the Total Environment**, v. 732, p. 1–2, 25 ago. 2020.

PARK, K. Y.; WI, S. J. Potential of plants to produce recombinant protein products. **Journal of Plant Biology**, v. 59, n. 6, p. 559–568, 1 dez. 2016.

PECCIA, J. et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. **Nature Biotechnology**, v. 38, n. 10, p. 1164–1167, 1 out. 2020.

PEINADO, B. et al. Improved methods for the detection and quantification of SARS-CoV-2 RNA in wastewater. **Scientific Reports**, v. 12, n. 1, p. 1–11, 1 dez. 2022.

PELLETIER, J. P. R.; MUKHTAR, F. Passive Monoclonal and Polyclonal Antibody Therapies. **Immunologic Concepts in Transfusion Medicine**. [S.l.]: Elsevier, 2020. p. 251–348.

PELTOMAA, R. et al. Recombinant antibodies and their use for food immunoanalysis. **Analytical and Bioanalytical Chemistry**, v. 414, n. 1, p. 193–217, 21 jan. 2022.

PEREIRA, F. et al. Genomic surveillance activities unveil the introduction of the SARS-CoV-2 B.1.525 variant of interest in Brazil: Case report. **Journal of Medical Virology**, v. 93, n. 9, p. 5523–5526, 1 set. 2021.

PETRONE, M. E. et al. Combining genomic and epidemiological data to compare the transmissibility of SARS-CoV-2 variants Alpha and Iota. **Communications Biology**, v. 5, n. 1, p. 1–10, 1 dez. 2022.

PLAZA-GARRIDO, A. et al. Norovirus, Hepatitis A and SARS-CoV-2 surveillance within Chilean rural wastewater treatment plants based on different biological treatment typologies. **Science of the Total Environment**, v. 863, p. 1–11, 10 mar. 2023.

PRABHU, S. K. et al. Exploring a combined Escherichia coli-based glycosylation and in vitro transglycosylation approach for expression of glycosylated interferon alpha. **Bioorganic and Medicinal Chemistry**, v. 33, p. 1–18, 1 mar. 2021.

PRADO, T. et al. Preliminary results of SARS-CoV-2 detection in sewerage system in niterói municipality, Rio de Janeiro, Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 115, n. 7, p. 1–3, 2020.

PRATELLI, A. Canine coronavirus inactivation with physical and chemical agents. **Veterinary Journal**, v. 177, n. 1, p. 71–79, jul. 2008.

PRUSSIN, A. J. et al. Survival of the Enveloped Virus Phi6 in Droplets as a Function of Relative Humidity, Absolute Humidity, and Temperature. **Applied and Environmental Microbiology**, v. 84, n. 12, p. 1–10, 15 jun. 2018.

RAHMAN, Md. M. et al. Coronaviruses in wild birds – A potential and suitable vector for global distribution. **Veterinary Medicine and Science**, v. 7, n. 1, p. 264–272, 24 jan. 2021.

RAJKO-NENOW, P. et al. Norovirus genotypes present in oysters and in effluent from a wastewater treatment plant during the seasonal peak of infections in ireland in 2010. **Applied and Environmental Microbiology**, v. 79, n. 8, p. 2578–2587, abr. 2013.

RAMBAUT, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. **Nature Microbiology**, v. 5, n. 11, p. 1403–1407, 1 nov. 2020.

RAVI, V.; SAXENA, S.; PANDA, P. S. Basic virology of SARS-CoV 2. **Indian Journal of Medical Microbiology**, v. 40, n. 2, p. 182–186, 1 abr. 2022.

RENU, K.; PRASANNA, P. L.; VALSALA GOPALAKRISHNAN, A. Coronaviruses pathogenesis, comorbidities and multi-organ damage – A review. **Life Sciences**, v. 255, p. 1–15, 15 ago. 2020.

ROMERO, P. E. et al. The Emergence of Sars-CoV-2 Variant Lambda (C.37) in South America. **Microbiology Spectrum**, v. 9, n. 2, p. 1–3, 31 out. 2021.

ROSALES-MENDOZA, S. et al. What does plant-based vaccine technology offer to the fight against COVID-19? **Vaccines**, v. 8, n. 183, p. 1–19, 1 abr. 2020.

ROSALES-MENDOZA, S.; PAZ-MALDONADO, L. M. T.; SORIA-GUERRA, R. E. Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: Current status and perspectives. **Plant Cell Reports**, v. 31, p. 479–494, 2012a.

_____. Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: current status and perspectives. **Plant Cell Reports**, v. 31, n. 3, p. 479–494, 12 mar. 2012b.

ROSARIO, K. et al. Pepper mild mottle virus as an indicator of fecal pollution. **Applied and Environmental Microbiology**, v. 75, n. 22, p. 7261–7267, nov. 2009.

ROSE, R. et al. Outbreak of P.3 (Theta) SARS-CoV-2 emerging variant of concern among service workers in Louisiana. **Journal of Infection and Public Health**, v. 15, n. 1, p. 7–9, 1 jan. 2022.

ROTHMAN, J. A. et al. RNA viromics of Southern California wastewater and detection of SARS-CoV-2 single-nucleotide variants. **Applied and Environmental Microbiology**, v. 87, n. 23, p. 1–15, 1 nov. 2021.

RYBICKI, E. P. Plant molecular farming of virus-like nanoparticles as vaccines and reagents. **Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology**, v. 12, p. 1–22, 2020a.

RYBICKI, E. P. Plant molecular farming of virus-like nanoparticles as vaccines and reagents. **Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology**, v. 12, n. 2, p. 1–22, 1 mar. 2020b.

SABINO, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. **The Lancet**, v. 397, n. 10273, p. 452–455, 6 fev. 2021.

SAJDEL-SULKOWSKA, E. M. A Dual-Route Perspective of SARS-CoV-2 Infection: Lung- vs. Gut-specific Effects of ACE-2 Deficiency. **Frontiers in Pharmacology**, v. 12, p. 1–12, 11 jun. 2021.

SALATA, C. et al. Coronaviruses: A paradigm of new emerging zoonotic diseases. **Pathogens and Disease**, v. 77, n. 9, p. 1–5, 18 fev. 2020.

SANTOS, R. B. et al. Putting the spotlight back on plant suspension cultures. **Frontiers in Plant Science**, v. 7, p. 1–12, 2016.

SAVILLE, J. W. et al. Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. **Nature Communications**, v. 13, n. 1, p. 1–10, 1 dez. 2022.

SCHAEFFER, J. et al. Looking into sewage: how far can metagenomics help to detect human enteric viruses? **Frontiers in Microbiology**, v. 14, p. 1–14, 2023.

SEAH, I.; AGRAWAL, R. Can the Coronavirus Disease 2019 (COVID-19) Affect the Eyes? A Review of Coronaviruses and Ocular Implications in Humans and Animals. **Ocular Immunology and Inflammation**, v. 28, n. 3, p. 391–395, 2 abr. 2020.

SHANMUGARAJ, B.; BULAON, C. J. I.; PHOOLCHAROEN, W. Plant molecular farming: A viable platform for recombinant biopharmaceutical production. **Plants**, v. 9, n. 842, p. 1–19, 1 jul. 2020.

SHERIDAN, C. COVID-19 spurs wave of innovative diagnostics. **Nature biotechnology**, v. 38, n. 7, p. 769–772, 1 jul. 2020.

SOLÁ, R. J.; GRIEBENOW, K. Glycosylation of therapeutic proteins: An effective strategy to optimize efficacy. **BioDrugs**, v. 24, n. 1, p. 9–21, 2010.

SOUZA, W. M. de et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. **Nature Human Behaviour**, v. 4, n. 8, p. 856–865, 1 ago. 2020.

STRASSER, R. Plant protein glycosylation. **Glycobiology**, v. 26, n. 9, p. 926–939, 1 set. 2016.

STRASSER, R.; ALTMANN, F.; STEINKELLNER, H. Controlled glycosylation of plant-produced recombinant proteins. **Current Opinion in Biotechnology**, v. 30, p. 95–100, dez. 2014.

SYED, A. et al. Gastrointestinal pathophysiology of SARS-CoV2 – a literature review. **Journal of Community Hospital Internal Medicine Perspectives**, v. 10, n. 6, p. 523–528, 1 nov. 2020.

TAVARES-ESASHIKA, M. L. et al. Characterization of an infectious clone of pepper ringspot virus and its use as a viral vector. **Archives of Virology**, v. 165, p. 367–375, 2020a.

_____. Characterization of an infectious clone of pepper ringspot virus and its use as a viral vector. **Archives of Virology**, v. 165, n. 2, p. 367–375, 16 fev. 2020b.

_____. Development of a heterologous gene expression vector in plants based on an infectious clone of a tobravirus, pepper ringspot virus. **Annals of Applied Biology**, v. 181, n. 1, p. 107–116, 1 jul. 2022.

TAVARES-ESASHIKA, M. L. et al. Development of a heterologous gene expression vector in plants based on an infectious clone of a tobravirus, pepper ringspot virus. **Annals of Applied Biology**, v. 181, n. 1, p. 107–116, 28 jul. 2022.

TOPARE, N. S.; ATTAR, S. J.; MANFE, M. M. SEWAGE/WASTEWATER TREATMENT TECHNOLOGIES: A REVIEW. **Sci. Revs. Chem. Commun**, v. 1, n. 1, p. 18–24, 2011.

VAN DER HOEK, L. SARS-CoV, human coronavirus NL63 (HCoV-NL63) and HCoV-HKU1 were first described Human coronaviruses: what do they cause? v. 12, p. 651–658, 2007.

VOLOCH, C. M. et al. Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil. **Journal of Virology**, v. 95, n. 10, p. 1–5, 26 abr. 2021.

WALSH, G.; JEFFERIS, R. Post-translational modifications in the context of therapeutic proteins. **Nature Biotechnology**, v. 24, n. 10, p. 1241–1252, out. 2006.

WANG, D. et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. **JAMA - Journal of the American Medical Association**, v. 323, n. 11, p. 1061–1069, 17 mar. 2020.

WANG, H. et al. SARS-CoV-2 Nucleocapsid Plasma Antigen for Diagnosis and Monitoring of COVID-19. **Clinical Chemistry**, v. 68, n. 1, p. 204–213, 30 dez. 2021.

WANG, Y. et al. SARS-CoV-2 S1 is superior to the RBD as a COVID-19 subunit vaccine antigen. **Journal of Medical Virology**, v. 93, n. 2, p. 892–898, 1 fev. 2021.

WARTECKI, A.; RZYMSKI, P. On the coronaviruses and their associations with the aquatic environment and wastewater. **Water (Switzerland)**, v. 12, n. 6, p. 1–27, 1 jun. 2020.

WHITNEY, O. N. et al. Sewage, Salt, Silica, and SARS-CoV-2 (4S): An Economical Kit-Free Method for Direct Capture of SARS-CoV-2 RNA from Wastewater. **Environmental Science and Technology**, v. 55, n. 8, p. 4880–4888, 20 abr. 2021.

WILBERS, R. H. P. et al. Co-expression of the protease furin in *Nicotiana benthamiana* leads to efficient processing of latent transforming growth factor-β1 into a biologically active protein. **Plant Biotechnology Journal**, v. 14, n. 8, p. 1695–1704, 1 ago. 2016.

WILLIAMS, L. et al. The C-Terminal Half of SARS-CoV-2 Nucleocapsid Protein, Industrially Produced in Plants, Is Valid as Antigen in COVID-19 Serological Tests. **Frontiers in Plant Science**, v. 12, p. 1–10, 27 jul. 2021.

WOLFE, M. K. et al. Wastewater-Based Detection of Two Influenza Outbreaks. **Environmental Science and Technology Letters**, v. 9, n. 8, p. 687–692, 9 ago. 2022.

WOLFERT, M. A.; BOONS, G. J. Adaptive immune activation: Glycosylation does matter. **Nature Chemical Biology**, v. 9, n. 12, p. 776–784, dez. 2013.

WU, F. et al. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. **mSystems**, v. 5, n. 4, p. 1–9, 25 ago. 2020.

XIAO, F. et al. Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. **Emerging Infectious Diseases**, v. 26, n. 8, p. 1920–1922, 20 ago. 2020.

YANIV, K. et al. City-level SARS-CoV-2 sewage surveillance. **Chemosphere**, v. 283, p. 1–6, 1 nov. 2021.

YE, Z. W. et al. Zoonotic origins of human coronaviruses. **International Journal of Biological Sciences**, v. 16, n. 10, p. 1686–1697, 2020.

YUSUF, L. et al. Rapid, Cheap, and Effective COVID-19 Diagnostics for Africa. **Diagnostics**, v. 11, p. 1–10, 13 nov. 2021.

ZENG, W. et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. **Biochemical and Biophysical Research Communications**, v. 527, n. 3, p. 618–623, 30 jun. 2020.

ZERBATO, V. et al. High fecal calprotectin levels are associated with SARS-CoV-2 intestinal shedding in COVID-19 patients: A proof-of-concept study. **World Journal of Gastroenterology**, v. 27, n. 22, p. 3130–3137, 14 jun. 2021.

ZHANG, J. et al. Progress and prospects on vaccine development against sars-cov-2. **Vaccines**, v. 8, n. 153, p. 1–12, 1 abr. 2020.

ZHANG, K.; SU, L.; WU, J. Recent Advances in Recombinant Protein Production by Bacillus subtilis. **Annual Review of Food Science and Technology**, v. 11, n. 6, p. 1–24, 2020.

ZHANG, T. et al. RNA viral community in human feces: Prevalence of plant pathogenic viruses. **PLoS Biology**, v. 4, n. 1, p. 0108–0118, jan. 2006.

ZHANG, W. et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. **Emerging Microbes and Infections**, v. 9, n. 1, p. 386–389, 1 jan. 2020.

ZHOU, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. **Cell**, v. 184, n. 9, p. 2348–2361, 29 abr. 2021.

ZHOU, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, v. 579, n. 7798, p. 270–273, 12 mar. 2020.

ZISCHEWSKI, J.; SACK, M.; FISCHER, R. Overcoming low yields of plant-made antibodies by a protein engineering approach. **Biotechnology Journal**, v. 11, p. 107–116, 2016.

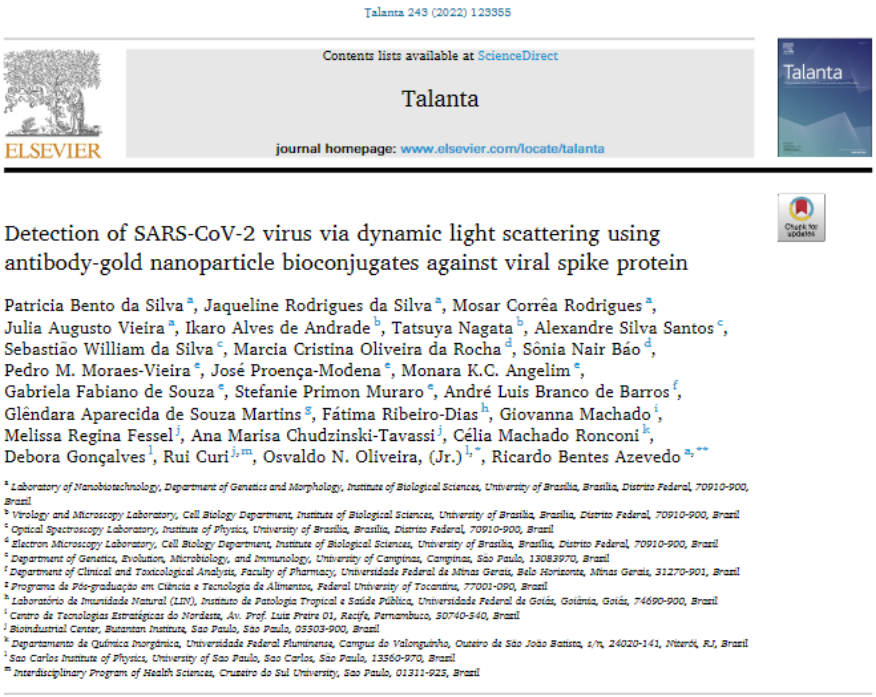
ZUCCATO, E. et al. Cocaine in surface waters: A new evidence-based tool to monitor community drug abuse. **Environmental Health: A Global Access Science Source**, v. 4, n. 14, p. 1–7, 5 ago. 2005.

ANEXOS

Participação em Publicações

Trabalhos publicados:

- SILVA, Patricia Bento et al. Detection of SARS-CoV-2 virus via dynamic light scattering using antibody-gold nanoparticle bioconjugates against viral spike protein. **Talanta**, v. 243, p. 1 - 7, 2022.



Talanta 243 (2022) 123355

Contents lists available at ScienceDirect

Talanta

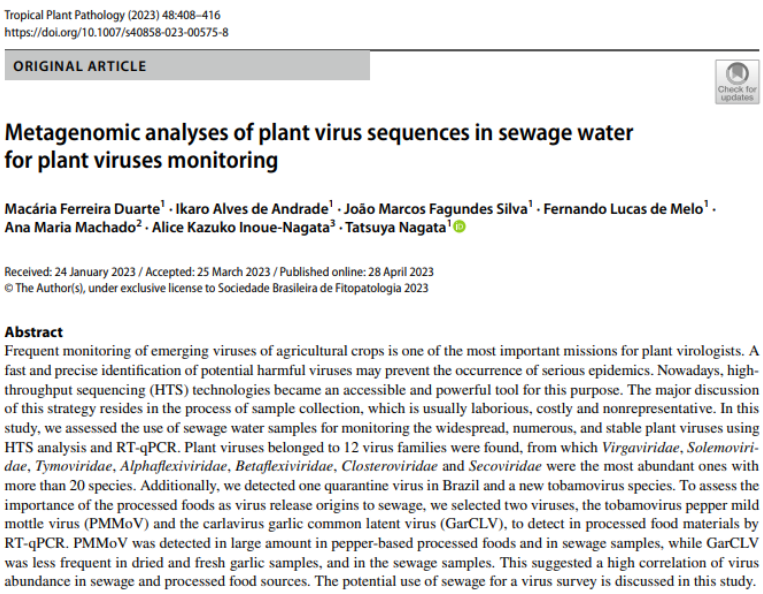
journal homepage: www.elsevier.com/locate/talanta

Detection of SARS-CoV-2 virus via dynamic light scattering using antibody-gold nanoparticle bioconjugates against viral spike protein

Patricia Bento da Silva [a], Jaqueline Rodrigues da Silva [a], Mosar Corrêa Rodrigues [a], Julia Augusto Vieira [a], Ikaro Alves de Andrade [b], Tatsuya Nagata [b], Alexandre Silva Santos [c], Sebastião William da Silva [c], Marcia Cristina Oliveira da Rocha [d], Sônia Nair Báo [d], Pedro M. Moraes-Vieira [e], José Proença-Modena [e], Monara K.C. Angelim [e], Gabriela Fabiano de Souza [e], Stefanie Primon Muraro [e], André Luis Branco de Barros [f], Glêndara Aparecida de Souza Martins [g], Fátima Ribeiro-Dias [h], Giovanna Machado [i], Melissa Regina Fessel [j], Ana Marisa Chudzinski-Tavassi [j], Célia Machado Ronconi [k], Debora Gonçalves [l], Rui Curi [j,m], Osvaldo N. Oliveira, (Jr.) [i,*], Ricardo Bentes Azevedo [a,**]

[a] Laboratory of Nanobiotechnology, Department of Genetics and Morphology, Institute of Biological Sciences, University of Brasília, Brasília, Distrito Federal, 70910-900, Brazil
[b] Virology and Microscopy Laboratory, Cell Biology Department, Institute of Biological Sciences, University of Brasília, Brasília, Distrito Federal, 70910-900, Brazil
[c] Optical Spectroscopy Laboratory, Institute of Physics, University of Brasília, Brasília, Distrito Federal, 70910-900, Brazil
[d] Electron Microscopy Laboratory, Cell Biology Department, Institute of Biological Sciences, University of Brasília, Brasília, Distrito Federal, 70910-900, Brazil
[e] Department of Genetics, Evolution, Microbiology, and Immunology, University of Campinas, Campinas, São Paulo, 13083970, Brazil
[f] Department of Clinical and Toxicological Analysis, Faculty of Pharmacy, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, 31270-901, Brazil
[g] Programa de Pós-graduação em Ciência e Tecnologia de Alimentos, Federal University of Tocantins, 77001-090, Brazil
[h] Laboratório de Imunidade Natural (LIN), Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, Goiás, 74690-900, Brazil
[i] Centro de Tecnologias Estratégicas do Nordeste, Av. Prof. Luis Freire 01, Recife, Pernambuco, 50740-540, Brazil
[j] Bioindustrial Center, Butantan Institute, Sao Paulo, São Paulo, 05503-900, Brazil
[k] Departamento de Química Inorgânica, Universidade Federal Fluminense, Campus do Valonguinho, Outeiro de São João Batista, s/n, 24020-141, Niterói, RJ, Brazil
[l] Sao Carlos Institute of Physics, University of Sao Paulo, Sao Carlos, São Paulo, 13560-970, Brazil
[m] Interdisciplinary Program of Health Sciences, Cruzeiro do Sul University, Sao Paulo, 01311-925, Brazil

- DUARTE, Macária Ferreira et al. Metagenomic analyses of plant virus sequences in sewage water for plant viruses monitoring. **Tropical Plant Pathology**, v. 48, n.4, p. 1-9, 2023.



Tropical Plant Pathology (2023) 48:408–416
https://doi.org/10.1007/s40858-023-00575-8

ORIGINAL ARTICLE

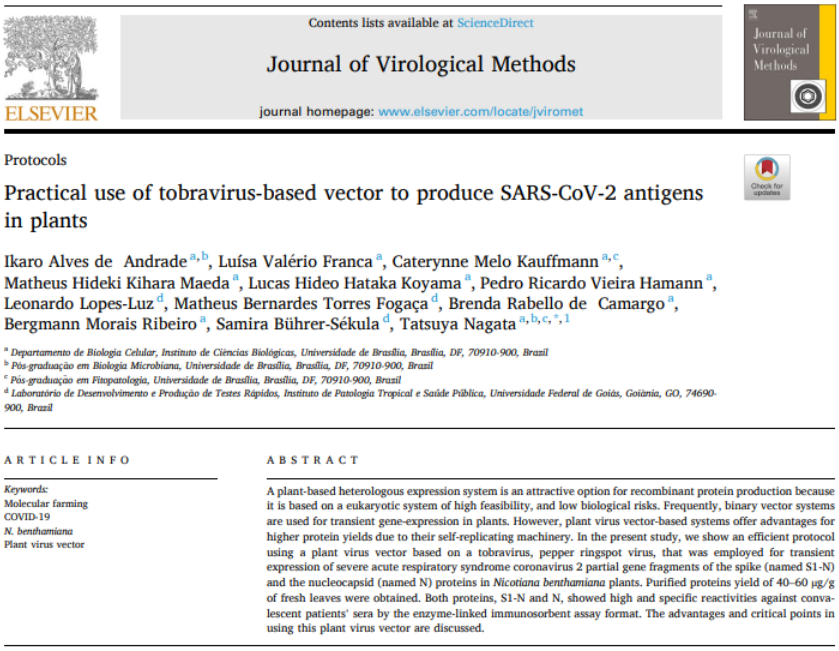
Metagenomic analyses of plant virus sequences in sewage water for plant viruses monitoring

Macária Ferreira Duarte[1] · Ikaro Alves de Andrade[1] · João Marcos Fagundes Silva[1] · Fernando Lucas de Melo[1] · Ana Maria Machado[2] · Alice Kazuko Inoue-Nagata[3] · Tatsuya Nagata[1]

Received: 24 January 2023 / Accepted: 25 March 2023 / Published online: 28 April 2023

Abstract
Frequent monitoring of emerging viruses of agricultural crops is one of the most important missions for plant virologists. A fast and precise identification of potential harmful viruses may prevent the occurrence of serious epidemics. Nowadays, high-throughput sequencing (HTS) technologies became an accessible and powerful tool for this purpose. The major discussion of this strategy resides in the process of sample collection, which is usually laborious, costly and nonrepresentative. In this study, we assessed the use of sewage water samples for monitoring the widespread, numerous, and stable plant viruses using HTS analysis and RT-qPCR. Plant viruses belonged to 12 virus families were found, from which *Virgaviridae*, *Solemoviridae*, *Tymoviridae*, *Alphaflexiviridae*, *Betaflexiviridae*, *Closteroviridae* and *Secoviridae* were the most abundant ones with more than 20 species. Additionally, we detected one quarantine virus in Brazil and a new tobamovirus species. To assess the importance of the processed foods as virus release origins to sewage, we selected two viruses, the tobamovirus pepper mild mottle virus (PMMoV) and the carlavirus garlic common latent virus (GarCLV), to detect in processed food materials by RT-qPCR. PMMoV was detected in large amount in pepper-based processed foods and in sewage samples, while GarCLV was less frequent in dried and fresh garlic samples, and in the sewage samples. This suggested a high correlation of virus abundance in sewage and processed food sources. The potential use of sewage for a virus survey is discussed in this study.

- ANDRADE, Ikaro Alves et al. Practical use of tobravirus-based vector to produce SARS-CoV-2 antigens in plants. **Journal of Virological Methods**, v. 315, p. 1 - 7, 2023.



Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Protocols

Practical use of tobravirus-based vector to produce SARS-CoV-2 antigens in plants

Ikaro Alves de Andrade[a,b], Luísa Valério Franca[a], Caterynne Melo Kauffmann[a,c], Matheus Hideki Kihara Maeda[a], Lucas Hideo Hataka Koyama[a], Pedro Ricardo Vieira Hamann[a], Leonardo Lopes-Luz[d], Matheus Bernardes Torres Fogaça[d], Brenda Rabello de Camargo[a], Bergmann Morais Ribeiro[a], Samira Bührer-Sékula[d], Tatsuya Nagata[a,b,c,*,1]

[a] Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, 70910-900, Brazil
[b] Pós-graduação em Biologia Microbiana, Universidade de Brasília, Brasília, DF, 70910-900, Brazil
[c] Pós-graduação em Fitopatologia, Universidade de Brasília, Brasília, DF, 70910-900, Brazil
[d] Laboratório de Desenvolvimento e Produção de Testes Rápidos, Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, GO, 74690-900, Brazil

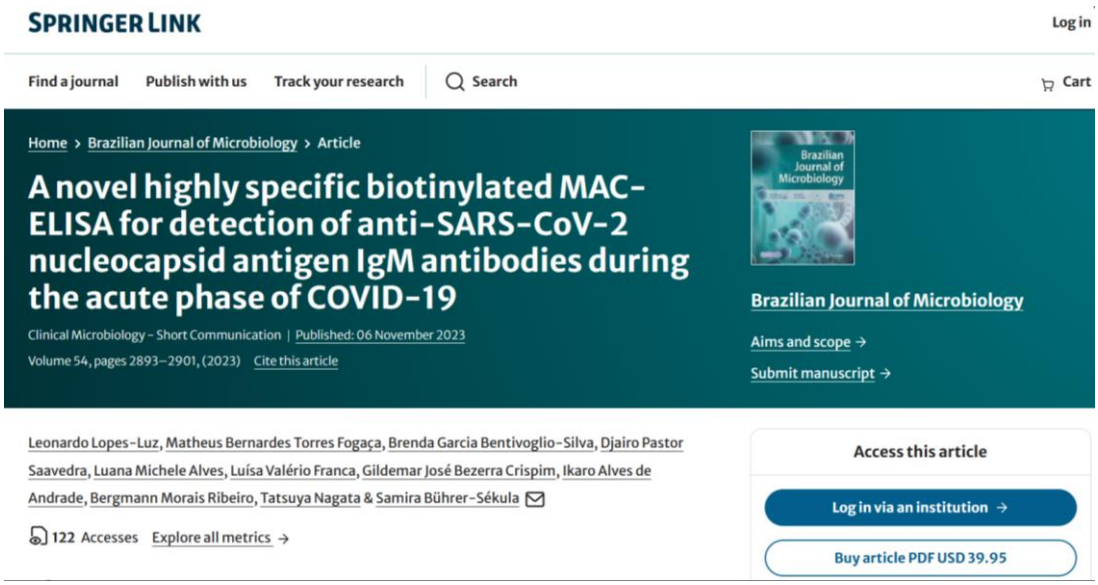
ARTICLE INFO

Keywords:
Molecular farming
COVID-19
N. benthamiana
Plant virus vector

ABSTRACT

A plant-based heterologous expression system is an attractive option for recombinant protein production because it is based on a eukaryotic system of high feasibility, and low biological risks. Frequently, binary vector systems are used for transient gene-expression in plants. However, plant virus vector-based systems offer advantages for higher protein yields due to their self-replicating machinery. In the present study, we show an efficient protocol using a plant virus vector based on a tobravirus, pepper ringspot virus, that was employed for transient expression of severe acute respiratory syndrome coronavirus 2 partial gene fragments of the spike (named S1-N) and the nucleocapsid (named N) proteins in *Nicotiana benthamiana* plants. Purified proteins yield of 40–60 μg/g of fresh leaves were obtained. Both proteins, S1-N and N, showed high and specific reactivities against convalescent patients' sera by the enzyme-linked immunosorbent assay format. The advantages and critical points in using this plant virus vector are discussed.

- LOPES-LUZ, Leonardo et al. A novel highly specific biotinylated MAC-ELISA for detection of anti-SARS-CoV-2 nucleocapsid antigen IgM antibodies during the acute phase of COVID-19. **Brazilian Journal of Microbiology**, v. 54, p. 1-9, 2023.



- KAUFFMANN, Caterynne M. et al. Specific antibody production using recombinant proteins to elucidate seed transmission and nuclear localization of Coguvirus citrulli and Coguvirus henanense in radicles of watermelon crop. **Journal of Virological Methods**, v. 325, p. 1 – 8, 2024.