

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization

International Bureau



(10) International Publication Number

WO 2024/015510 A1

(43) International Publication Date  
18 January 2024 (18.01.2024)

**Published:**

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(51) International Patent Classification:  
*A61K 39/12* (2006.01)      *A61P 31/14* (2006.01)

(21) International Application Number:  
PCT/US2023/027622

(22) International Filing Date:  
13 July 2023 (13.07.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
63/368,324      13 July 2022 (13.07.2022)      US

(71) Applicant: WISCONSIN ALUMNI RESEARCH FOUNDATION ("WARF") [US/US]; 614 Walnut Street, Middleton, Wisconsin 53562 (US).

(72) Inventors: KAWAOKA, Yoshihiro; 8722 Airport Road, Middleton, Wisconsin 53562 (US). HALFMANN, Peter J.; 13 Kiowa Court, Madison, Wisconsin 53713 (US). KURODA, Makoto; 202 N. Eau Claire Avenue #110, Madison, Wisconsin 53705 (US).

(74) Agent: PERDOK, Monique M. et al.; P.O. Box 2938, Minneapolis, Minnesota 55402 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: SARS-COV-2 LACKING THE ENVELOPE PROTEIN AS AN ATTENUATED VACCINE VIRUS AGAINST COVID-19

(57) Abstract: An isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope (E) protein and/or M protein, a vaccine comprising the recombinant genome and methods of using the vaccine are provided.

**SARS-COV-2 LACKING THE ENVELOPE PROTEIN AS AN  
ATTENUATED VACCINE VIRUS AGAINST COVID-19**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

5        This application claims the benefit of the filing date of U.S. application No. 63/368,324, filed on July 13, 2022, the disclosure of which is incorporated by reference herein.

**STATEMENT OF GOVERNMENT SUPPORT**

10      This invention was made with government support under AI165077 awarded by the National Institutes of Health. The government has certain rights in the invention.

**INCORPORATION BY REFERENCE OF SEQUENCE LISTING**

15      A Sequence Listing is provided herewith as an xml file, “2350480.xml” created on July 11, 2023 and having a size of 275,824 bytes. The content of the xml file is incorporated by reference herein in its entirety.

**BACKGROUND**

20      Most available vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) including mRNA vaccines, viral vector vaccines, and recombinant protein vaccines, induce serum antibodies to block the function of the spike (S) protein that is essential for viral entry. However, the induction of mucosal immunity in the upper respiratory tract is insufficient with current  
25      vaccines.

**SUMMARY**

30      To develop a vaccine that can elicit protective immune responses in mucosa, a coronavirus, e.g., SARS-CoV-2, vaccine based on an attenuated coronavirus was prepared. An attenuated virus demonstrates reduced virulence *in vivo*. In one embodiment, the attenuated coronavirus has a genome that does not encode all the viral proteins (it is a mutant viral genome) needed for viral replication but may still produce progeny, but does express spike (S) protein. An attenuated virus may be a “semi-virus” (or “semi-live virus”), which is a

virus that expresses viral proteins to invade cells and induce immunity for infection defense, but does not produce new infectious progeny particles, e.g., as a result of the lack of viral proteins for multiple rounds of replication and the generation of infectious progeny virus. Multiplication of a virus occurs when 5 the virus produces infectious progeny virus particles from cells that the virus enters, and this step can be repeated by the progeny viruses and their progeny for multiple generations. An attenuated virus that does not express one or more of the viral proteins necessary for viral replication may be employed to induce mucosal immunity. An attenuated vaccine virus based on a whole virus may 10 generate an immune response not only against the spike protein (the target of most SARS-CoV-2 vaccines), but also against other SARS-CoV-2 proteins, thereby eliciting a more robust and durable protection profile. The efficacy of a semi-live virus as a type of vaccine against SARS-CoV-2 in animal models and in clinical studies in humans may be enhanced relative to an attenuated virus that 15 produces some progeny virus.

Therefore, a coronavirus vaccine based on the attenuated virus has the following advantages over current vaccines: it can induce not only humoral but also cellular immunity as effectively as live-attenuated vaccines, e.g., FluMist (an influenza vaccine based on a cold-adapted live-attenuated influenza virus); 20 the risk of reversion to the wild-type virus with pathogenicity, which is a concern with live-attenuated vaccines, is low; local mucosal immunity can be induced through intranasal administration; because the attenuated virus is not a viral vector vaccine, multiple inoculations (vaccinations) are feasible and it would likely induce immune responses against structural proteins other than the 25 spike protein; and because innate immune responses can be activated after a single inoculation with the attenuated virus, there is no need for an adjuvant(s).

In one embodiment, the genome of the attenuated coronavirus is a mutant genome where expression of coronavirus S, E, M, N, ORF1, e.g., ORF 1a, ORF3, e.g., ORF3a, ORF6, ORF7, and/or ORF8, is knocked down or knocked 30 out, e.g., by a genetic modification including but not limited to one or more nucleotide deletion(s), substitution(s), insertion(s), or any combination thereof. In one embodiment, the coding region for E is deleted. In one embodiment, a portion of the coding region for E is deleted, e.g., a deletion of 5, 10, 20, 30, 40, 50, 60, 70 or more amino acids. In one embodiment, the coding region for M is

deleted. In one embodiment, a portion of the coding region for M is deleted, e.g., a deletion of 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 105, 110, 120, 130, 140, 150, 160, 170, 180, or more amino acids.

In one embodiment, the genome of the attenuated coronavirus is a mutant genome having one or more modifications that result in a cold-adapted coronavirus. In one embodiment, a cold-adapted coronavirus encodes one or more of nsp2 (non-structural protein 2) having amino acid residues from 82 to 84 (e.g., residues glycine (G), histidine (H), and valine (V)) deleted, and/or methionine (M) or valine (V) at position 85; nsp6 having 3609K (lysine), and/or 10 3671T (threonine)); nsp7 having 3926A (alanine); nsp13 having 5604F (phenylalanine); and/or S protein having 95I, 185K, and/or 968A, or any combination thereof. In one embodiment, a cold-adapted coronavirus encodes a 12-amino acid-deletion located in the junction of S1 and S2 region including the furin cleavage site (PRRAR); and/or a 371-nucleotide-deletion resulting in 15 partial orf7b (1–17 amino acid residues); the complete deletion of the orf8 protein; nsp3 having 494K, 579V, 763M, 793S, and/or 1456I; nsp16 having 69Y, and/or 813I; E having 32V; orf7a having 44L; and/or N having 198I, or any combination thereof.

Since the vaccine virus genome can be generated by reverse genetics, 20 the original S gene can easily be replaced with the S gene from other strains, which makes it possible to prepare a new seed virus quickly when a variant with different antigenic properties emerges. A semi-live SARS-CoV-2 that is effective in humans establishes a different vaccine modality and may be applied to infectious diseases other than COVID, e.g., immunogenic non-coronavirus 25 gene products may be expressed from a genome with a knock-out.

As described herein, a SARS-CoV-2 semi-live, attenuated vaccine virus based on the original Wuhan genome, e.g., the semi-live virus encodes S protein of the Wuhan strain, but lacking the envelope (E) open reading frame was prepared and this vaccine virus replicated efficiently in Vero cells that stably 30 express the E protein. To demonstrate the safety of this vaccine virus (CoV-2 ΔE), human (h)ACE2 transgenic mice were used, which are highly susceptible to infection and serve as a lethal animal model for SARS-CoV-2 infection. Infection with 10,000 plaque-forming units (pfu) of wild-type SARS-CoV-2 (Wuhan isolate generated by reverse genetics) of hACE2 mice resulted in

significant body weight loss, and all of the mice succumbed to infection by day 7. In contrast, hACE2 mice infected with the same dose of CoV-2 ΔE, had the same body weight and survival profiles as mock-infected animals.

To determine the protective efficacy of CoV-2 ΔE, Syrian hamsters were 5 vaccinated with 100,000 pfu of CoV-2 ΔE by intranasal inoculation. Two weeks after vaccination, the hamsters had antibody titers against the SARS-CoV-2 spike receptor-binding domain antigen ranging from 1:320 to 1:1280. At 4- weeks after vaccination, the hamsters were challenged with 1,000 pfu of an early 10 SARS-CoV-2 isolate. Three days after challenge, three of the four vaccinated hamsters had no detectable infectious virus in their lung tissue, and the fourth hamster had a viral load in its lung tissue of approximately  $10^4$  pfu/gram. In 15 contrast, the control hamsters had high virus titers, close to  $10^8$  pfu/gram in their lung tissue. Vaccine efficacy in the nasal turbinate (NT) tissues was less pronounced, but there was a significant reduction in viral load in the vaccinated 20 compared to control hamsters. The data demonstrate the near-complete protection of hamsters from infectious virus in the lungs after a single vaccination with CoV-2 ΔE.

The CoV-2 ΔE mutant virus is not completely replication-deficient. Other deletions in the CoV-2 genome, optionally in combination with one or 20 more other deletions in open reading frames including ΔE, may provide for enhanced attenuation of the virus. Those viruses with genomes having one or more knock outs of viral proteins, e.g., deletion of at least part of the open reading frame of one or more viral proteins (and the expression of those protein(s) in *trans*, for instance, in Vero cells during viral growth/amplification, 25 if needed) may provide for enhanced attenuation of the virus *in vivo*. For example, a CoV-2 ΔEM mutant virus is replication-deficient.

The disclosure thus provides for methods of making an attenuated virus.

In one embodiment, a recombinant CoV-2 is provided that completely lacks the E gene, e.g., from nucleotide 26,245 to 26,472, and/or the M gene, 30 e.g., from nucleotide 26,523 to 27,191 in the ancestral Wuhan reference sequence (NCBI Accession number MN908947.3). The intergenic regions flanking the 5' and 3' ends of the E gene (e.g., nucleotide 26,221 to 26,244 and 26,473 to 26,522, respectively) and/or M gene (e.g., 26,473 to 26,522 and 27,192 to 27,201, respectively) may also be deleted with the respective open-

reading frame. In one embodiment specific functional domains of the E and M gene may be deleted such as the transmembrane domain (e.g., amino acids 11 to 37 of E protein, and/or amino acids 20 to 38, 46 to 70, and/or 76 to 100 of M protein, or any combination thereof) or C-terminal intracellular region of M protein (e.g., amino acids 104 to 222) that interacts with N protein leading to efficient virion formation.

The disclosure also provides for isolated attenuated virus and compositions, for example, vaccines, having the isolated attenuated virus.

Also provided are isolated host cells that express one or more SARS-CoV-2 viral proteins, e.g., from an exogenously introduced vector, isolated host cells comprising an exogenous vector comprising a mutated SARS-CoV-2 viral genome, and isolated host cells that express one or more SARS-CoV-2 viral proteins in *trans* and comprise an exogenous vector comprising a mutated SARS-CoV-2 viral genome and virus obtained from those host cells. In one embodiment, the host cell comprises a vector comprising a nucleic acid sequence encoding an E protein, e.g., a nucleic acid sequence comprising SEQ ID NO:13 or a nucleic acid sequence having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99%v nucleotide sequence identity to SEQ ID NO:13, e.g., one that encodes an E protein with at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99%v amino acid sequence identity to a polypeptide encoded by SEQ ID NO:13. In one embodiment, the host cell comprises a vector comprising a nucleic acid sequence encoding a M protein, e.g., a nucleic acid sequence comprising SEQ ID NO:14 or a nucleic acid sequence having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99%v nucleotide sequence identity to SEQ ID NO:14, e.g., one that encodes a M protein with at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99%v amino acid sequence identity to a polypeptide encoded by SEQ ID NO:14. In one embodiment, the host cell comprises a vector comprising a nucleic acid sequence encoding a human ACE2 protein, e.g., a nucleic acid sequence comprising SEQ ID NO:17 or a nucleic acid sequence having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99%v nucleotide sequence identity to SEQ ID NO:13, e.g., one that encodes a hACE2 protein with at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or

99% v amino acid sequence identity to a polypeptide encoded by SEQ ID NO:17.

In one embodiment, the host cell has two or more vectors, e.g., to express E, M, and/or hACE2. In one embodiment, one or more of the vectors is/are integrated into the host cell genome.

- 5        Further provided is a method to induce an immune response in a mammal.

In one embodiment, an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope (E) protein is provided. In one embodiment, 10 an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein comprises SEQ ID NO:15 or a nucleic acid sequence having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% 15 nucleotide sequence identity to SEQ ID NO:15. In one embodiment, an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E and M proteins comprises SEQ ID NO:16 or a nucleic acid sequence having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% 20 v nucleotide sequence identity to SEQ ID NO:16.

- 20        In one embodiment, an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus integral membrane (M) protein is provided.

In one embodiment, the modification is a deletion of at least part of the open reading frame encoding the E protein. In one embodiment, the modification 25 is a deletion of the entire open reading frame encoding the E protein. In one embodiment, the modification is an insertion into the open reading frame encoding the E protein. In one embodiment, the modification is a substitution of one or more nucleotides in the open reading frame encoding E protein, e.g., that results in a termination codon. In one embodiment, the modification is a deletion 30 of the entire open reading frame encoding the E protein. In one embodiment, the isolated nucleic acid further comprises one or more genetic modifications that inhibit or prevent expression of coronavirus M protein. In one embodiment, the isolated nucleic acid comprises DNA. In one embodiment, the isolated nucleic acid comprises RNA. Also provided is a cell comprising the isolated nucleic

acid. In one embodiment, the cell is a mammalian cell, e.g., a Vero cell or other non-human primate cell. In one embodiment, the cell is a non-human primate cell. In one embodiment, the cell stably expresses coronavirus E protein. In one embodiment, the cell stably expresses hACE2.

5       In one embodiment, the modification is a deletion of at least part of the open reading frame encoding the M protein. In one embodiment, the modification is a deletion of the entire open reading frame encoding the M protein. In one embodiment, the modification is a deletion of at least part of the open reading frame encoding the E protein and a deletion of at least part of the  
10      open reading frame encoding the M protein. In one embodiment, the modification is a deletion of the entire open reading frame encoding the E protein and a deletion of at least part of the open reading frame encoding the M protein. In one embodiment, the modification is a deletion of at least part of the open reading frame encoding the E protein and a deletion of the entire open  
15      reading frame encoding the M protein. In one embodiment, the modification is a deletion of the entire open reading frame encoding the E protein and a deletion of the entire open reading frame encoding the M protein, optionally including the intergenic region therebetween. In one embodiment, the modification is an insertion into the open reading frame encoding the M protein. In one  
20      embodiment, the modification is a substitution of one or more nucleotides in the open reading frame encoding M protein, e.g., that results in a termination codon. In one embodiment, the modification is a deletion of the entire open reading frame encoding the M protein. In one embodiment, the isolated nucleic acid further comprises one or more genetic modifications that inhibit or prevent  
25      expression of coronavirus E protein. In one embodiment, the isolated nucleic acid comprises DNA. In one embodiment, the isolated nucleic acid comprises RNA. Also provided is a cell comprising the isolated nucleic acid. In one embodiment, the cell is a mammalian cell. In one embodiment, the cell is a non-human primate cell. In one embodiment, the cell stably expresses coronavirus M protein. In one embodiment, the cell stably expresses hACE2. In one embodiment, the entire open reading frame encoding the E protein, the entire open reading frame encoding the M protein and intergenic region between the E and M genes are deleted.

Further provided is a composition comprising an attenuated recombinant coronavirus comprising a coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope E protein, which virus comprises E protein embedded in the envelope. In one embodiment, the 5 coronavirus genome further comprises a genetic modification that inhibits or prevents expression of coronavirus M protein, which virus comprises M protein embedded in the envelope.

Also provided is a composition comprising an attenuated recombinant 10 coronavirus comprising a coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus M protein, which virus comprises M protein embedded in the envelope. In one embodiment, the coronavirus genome further comprises a genetic modification that inhibits or prevents expression 15 of coronavirus E protein, which virus comprises E protein embedded in the envelope.

The disclosure provides a system comprising: i) an isolated cell that stably expresses coronavirus E protein, or coronavirus E protein and coronavirus M protein; and ii) an isolated nucleic acid comprising a recombinant coronavirus 20 genome having a genetic modification that inhibits or prevents expression of coronavirus E protein, or an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein and M protein. In one embodiment, the isolated cell stably expresses coronavirus E protein and the isolated nucleic acid 25 comprises a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein. In one embodiment, the isolated cell stably expresses coronavirus E protein and M protein and the isolated nucleic acid comprises a recombinant coronavirus genome having a a 30 genetic modification that inhibits or prevents expression of coronavirus E protein and M protein.

A recombinant coronavirus is provided, wherein the genome of the recombinant coronavirus contains a deletion of one or more nucleotides in a polynucleotide sequence for a viral protein corresponding to SARS CoV-2 E protein which deletion is effective to prevent expression of a functional viral

protein corresponding to SARS CoV-2 E protein upon infection of a cell with the recombinant coronavirus, wherein the genome encodes one or more coronavirus glycoproteins, and wherein the coronavirus comprises E protein. In one embodiment, the cell that is infected does not express functional E protein.

- 5     In one embodiment, the recombinant coronavirus further comprises a deletion of one or more nucleotides in a polynucleotide sequence having an open reading frame for a viral protein corresponding to coronavirus M protein. In one embodiment, the recombinant coronavirus comprises M protein. In one embodiment, at least 90% of sequences corresponding to E or M protein coding
- 10    sequences, or any combination, in the viral genome of the virus, are deleted. In one embodiment, the recombinant genome further comprises a nucleotide sequence encoding a prophylactic or therapeutic heterologous gene product. A vaccine having an effective amount of the recombinant coronavirus is further provided. In one embodiment, the vaccine of is formulated for intranasal
- 15    delivery. In one embodiment, the vaccine is formulated for subcutaneous delivery.

- A recombinant coronavirus is provided, wherein the genome of the recombinant coronavirus contains a deletion of one or more nucleotides in a polynucleotide sequence for a viral protein corresponding to SARS CoV-2 M protein which deletion is effective to prevent expression of a functional viral protein corresponding to SARS CoV-2 M protein upon infection of a cell with the recombinant coronavirus, wherein the genome encodes one or more coronavirus glycoproteins, and wherein the coronavirus comprises M protein. In one embodiment, the cell that is infected does not express functional M protein.
- 20    In one embodiment, the recombinant coronavirus further comprises a deletion of one or more nucleotides in a polynucleotide sequence having an open reading frame for a viral protein corresponding to coronavirus E protein. In one embodiment, the recombinant coronavirus comprises E protein. In one embodiment, at least 90% of sequences corresponding to E or M protein coding
  - 25    sequences, or any combination, in the viral genome of the virus, are deleted. In one embodiment, the recombinant genome further comprises a nucleotide sequence encoding a prophylactic or therapeutic heterologous gene product. A vaccine having an effective amount of the recombinant coronavirus is further provided. In one embodiment, the vaccine of is formulated for intranasal
  - 30

delivery. In one embodiment, the vaccine is formulated for subcutaneous delivery.

A method to immunize a mammal is provided, comprising administering to the mammal an effective amount of the vaccine. In one embodiment, the 5 mammal is a human. In one embodiment, the method includes administering two or more doses.

In one embodiment, the method comprises administering one dose.

### BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1B. A) Genomes of the wild-type Wuhan genome (top) and 10 the CoV-2 ΔE open reading frame (ORF) vaccine virus. B) CoV-2 ΔE plaque formation on Vero cells stably expressing the E protein.

Figures 2A-2B. Body weight changes (A) and survival (B) of hACE2 mice infected with wild-type, CoV-2 ΔE, or control (mock-infected).

Figure 3. Replication of challenge virus in the lung and nasal turbinate 15 (NT) tissues of control hamsters and hamsters vaccinated once with CoV-2 ΔE.

Figure 4. Overview of semi-virus.

Figure 5. Constructs for ΔE and ΔEM genomes.

Figure 6. Pathogenicity and protective effect of ΔE virus vaccination.

Figure 7. Growth of ΔE virus in cell culture.

Figure 8. ΔEM with various spike proteins.

Figure 9A. Generation of cell clone stably expressing hACE2, E and M.

Figure 9B. Pathogenicity of ΔEM and potential for recombination.

Figure 10A-D. Immunity induction and infection protection in animals inoculated with ΔEM.

Figure 11. Testing of ΔEM vaccine in humans.  $10^6$  pfu = high dose; 25  $10^4$  pfu = low dose.

Figures 12A-12C. Exemplary SARS-CoV-2 sequences. A) Delta variant (SEQ ID NO:1 is amino acid sequence for E; SEQ ID NO:2 is amino acid sequence for M; SEQ ID NO:3 is amino acid sequence for N; SEQ ID NO:4 is 30 nucleotide sequence for viral genome) (SEQ ID NOs: 32-40). B) Omicron variant (SEQ ID NO:5 is amino acid sequence for E; SEQ ID NO:6 is amino acid sequence for M; SEQ ID NO:7 is amino acid sequence for N; SEQ ID NO:8 is nucleotide sequence for viral genome) (SEQ ID NOs: 41-49). C) Wuhan

variant (SEQ ID NO:9 is amino acid sequence for E; SEQ ID NO:10 is amino acid sequence for M; SEQ ID NO:11 is amino acid sequence for N; SEQ ID NO:12 is nucleotide sequence for viral genome) (SEQ ID NOs: 50-58).

Figure 13. Schematic of genome.

5 Figure 14. Assembly of infectious clone.

Figure 15. Sequences for an exemplary codon-optimized CoV-2 E gene (SEQ ID NO:13), codon-optimized CoV-2 M gene (SEQ ID NO:14), ΔE genome (SEQ ID NO:15), ΔEM genome (SEQ ID NO:16), and hACE2 open reading frame (SEQ ID NO:17).

10 Figures 16A-16D. Efficacy of one vaccination of CoV-2 ΔE+ΔM. Virus titers three days after challenge with the Delta variant or Omicron XBB variant in non-vaccinated, control hamsters or hamsters vaccinated once (prime) with CoV-2 ΔE+ΔM. Dotted line indicates limit of detection ( $1.3 \log_{10}$  pfu/g). Each dot in the bar graph indicates individual hamsters in each group.

15 Figures 17A-17D. Efficacy of two vaccinations of CoV-2 ΔE+ΔM. Virus titers three days after challenge with the Delta variant or Omicron XBB variant in non-vaccinated, control hamsters or hamsters vaccinated (prime + boost [P+B]) with CoV-2 ΔE+ΔM. Dotted line indicates limit of detection ( $1.3 \log_{10}$  pfu/g). Each dot in the bar graph indicates individual hamsters in each group.

20 Figure 18. NCBI Accession number MN908947.3 (SEQ ID NO: 59).

## DETAILED DESCRIPTION

### Definitions

A "vector" or "construct" (sometimes referred to as gene delivery or gene transfer "vehicle") refers to a macromolecule or complex of molecules comprising a polynucleotide or virus to be delivered to a host cell, either *in vitro* or *in vivo*. The polynucleotide or virus to be delivered may comprise a coding sequence of interest for gene therapy. Vectors include, for example, viral vectors (such as coronavirus, filovirus, adenovirus, adeno-associated virus (AAV), 25 lentivirus, herpesvirus and retrovirus vectors), liposomes and other lipid-containing complexes, and other macromolecular complexes capable of mediating delivery of a polynucleotide to a host cell. Vectors can also comprise other components or functionalities that further modulate gene delivery and/or gene expression, or that otherwise provide beneficial properties to the targeted 30

cells. Such other components include, for example, components that influence binding or targeting to cells (including components that mediate cell-type or tissue-specific binding); components that influence uptake of the vector nucleic acid by the cell; components that influence localization of the polynucleotide within the cell after uptake (such as agents mediating nuclear localization); and components that influence expression of the polynucleotide. Such components also might include markers, such as detectable and/or selectable markers that can be used to detect or select for cells that have taken up and are expressing the nucleic acid delivered by the vector. Such components can be provided as a natural feature of the vector (such as the use of certain viral vectors which have components or functionalities mediating binding and uptake), or vectors can be modified to provide such functionalities. A large variety of such vectors are known in the art and are generally available. When a vector is maintained in a host cell, the vector can either be stably replicated by the cells during mitosis as an autonomous structure, incorporated within the genome of the host cell, or maintained in the host cell's nucleus or cytoplasm.

A "recombinant viral vector" refers to a viral vector comprising one or more modifications, including deletions, insertions, substitutions, and/or heterologous genes or sequences. Since many viral vectors exhibit size constraints associated with packaging, the heterologous genes or sequences are typically introduced by replacing one or more portions of the viral genome. Such viruses may become replication-defective or replication-incompetent, e.g., requiring the deleted function(s) to be provided in *trans* during viral replication and encapsidation (by using, e.g., a helper virus or a packaging cell line carrying genes for replication and/or encapsidation). Modified viral vectors in which a polynucleotide to be delivered is carried on the outside of the viral particle have also been described.

"Gene delivery," "gene transfer," and the like as used herein, are terms referring to the introduction of an exogenous polynucleotide (sometimes referred to as a "transgene") into a host cell, irrespective of the method used for the introduction. Such methods include a variety of well-known techniques such as vector-mediated gene transfer (by, e.g., viral infection/transfection, or various other protein-based or lipid-based gene delivery complexes) as well as

techniques facilitating the delivery of "naked" polynucleotides (such as electroporation, "gene gun" delivery and various other techniques used for the introduction of polynucleotides). The introduced polynucleotide may be stably or transiently maintained in the host cell. Stable maintenance typically requires  
5 that the introduced polynucleotide either contains an origin of replication compatible with the host cell or integrates into a replicon of the host cell such as an extrachromosomal replicon (e.g., a plasmid) or a nuclear or mitochondrial chromosome. A number of vectors are known to be capable of mediating transfer of genes to mammalian cells, as is known in the art.

10 By "transgene" is meant any piece of a nucleic acid molecule (for example, DNA) which is inserted by artifice into a cell either transiently or permanently, and becomes part of the organism if integrated into the genome or maintained extrachromosomally. Such a transgene may include at least a portion of an open reading frame of a gene which is partly or entirely heterologous (i.e.,  
15 foreign) to the transgenic organism, or may represent at least a portion of an open reading frame of a gene homologous to an endogenous gene of the organism, which portion optionally encodes a polypeptide with substantially the same activity as the corresponding full-length polypeptide or at least one activity of the corresponding full-length polypeptide.

20 By "transgenic cell" is meant a cell containing a transgene. For example, a cell stably or transiently transformed with a vector containing an expression cassette is a transgenic cell that can be used to produce a population of cells having altered phenotypic characteristics. A "recombinant cell" is one which has been genetically modified, e.g., by insertion, deletion or replacement of  
25 sequences in a nonrecombinant cell by genetic engineering.

The term "wild-type" or "native" refers to a gene or gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form  
30 of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product that displays modifications in sequence and/or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by

the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "transduction" denotes the delivery of a polynucleotide to a recipient cell either *in vivo* or *in vitro*, via a viral vector and optionally via a 5 replication-defective viral vector.

The term "heterologous" as it relates to nucleic acid sequences such as gene sequences encoding a protein and control sequences, denotes sequences that are not normally joined together, and/or are not normally associated with a particular cell, e.g., are from different sources (for instance, sequences from a 10 virus are heterologous to sequences in the genome of an uninfected cell). Thus, a "heterologous" region of a nucleic acid construct or a vector is a segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a nucleic acid construct could include a coding sequence flanked by 15 sequences not found in association with the coding sequence in nature, i.e., a heterologous promoter. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, a cell transformed with a construct which is not normally present in the cell would be 20 considered heterologous for purposes of this disclosure.

By "DNA" is meant a polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in double-stranded or single-stranded form found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular DNA 25 molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence complementary to the mRNA). The term captures molecules that include the four bases adenine, guanine, thymine, or cytosine, as well as molecules that 30 include base analogues which are known in the art.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by

the base-pairing rules. For example, the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids.

- 5 The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

DNA molecules are said to have "5' ends" and "3' ends" because  
10 mononucleotides are reacted to make oligonucleotides or polynucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide or polynucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a  
15 mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide or polynucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or  
20 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements that direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding  
25 region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

A "gene," "polynucleotide," "coding region," "sequence," "segment," "fragment" or "transgene" which "encodes" a particular protein, is a nucleic acid molecule which is transcribed and optionally also translated into a gene product,  
30 e.g., a polypeptide, *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The coding region may be present in either a cDNA, genomic DNA, or RNA form. When present in a DNA form, the nucleic acid molecule may be single-stranded (i.e., the sense strand) or double-stranded. The boundaries of a coding region are determined by a start codon at the 5'

(amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A gene can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be 5 located 3' to the gene sequence.

The term "control elements" refers collectively to promoter regions, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites ("IRES"), enhancers, splice junctions, and the like, which collectively provide 10 for the replication, transcription, post-transcriptional processing and translation of a coding sequence in a recipient cell. Not all of these control elements need always be present so long as the selected coding sequence is capable of being replicated, transcribed and translated in an appropriate host cell.

The term "promoter" is used herein in its ordinary sense to refer to a 15 nucleotide region comprising a DNA regulatory sequence, wherein the regulatory sequence is derived from a gene which is capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding sequence.

By "enhancer" is meant a nucleic acid sequence that, when positioned 20 proximate to a promoter, confers increased transcription activity relative to the transcription activity resulting from the promoter in the absence of the enhancer domain.

By "operably linked" with reference to nucleic acid molecules is meant 25 that two or more nucleic acid molecules (e.g., a nucleic acid molecule to be transcribed, a promoter, and an enhancer element) are connected in such a way as to permit transcription of the nucleic acid molecule. "Operably linked" with reference to peptide and/or polypeptide molecules is meant that two or more peptide and/or polypeptide molecules are connected in such a way as to yield a single polypeptide chain, i.e., a fusion polypeptide, having at least one property 30 of each peptide and/or polypeptide component of the fusion. The fusion polypeptide may be chimeric, i.e., composed of heterologous molecules.

- "Homology" refers to the percent of identity between two polynucleotides or two polypeptides. The correspondence between one sequence and to another can be determined by techniques known in the art. For example, homology can be determined by a direct comparison of the sequence information
- 5      between two polypeptide molecules by aligning the sequence information and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single strand-specific nuclease(s), and size determination of the digested fragments.
- 10     Two DNA, or two polypeptide, sequences are "substantially homologous" to each other when at least about 80%, e.g., at least about 90%, such as at least about 95% of the nucleotides, or amino acids, respectively match over a defined length of the molecules, as determined using the methods above.

By "mammal" is meant any member of the class *Mammalia* including,

15     without limitation, humans and nonhuman primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats, rabbits and guinea pigs, and the like.

By "derived from" is meant that a nucleic acid molecule was either made

20     or designed from a parent nucleic acid molecule, the derivative retaining substantially the same functional features of the parent nucleic acid molecule, e.g., encoding a gene product with substantially the same activity as the gene product encoded by the parent nucleic acid molecule from which it was made or designed.

25     By "expression construct" or "expression cassette" is meant a nucleic acid molecule that is capable of directing transcription. An expression construct includes, at the least, a promoter. Additional elements, such as an enhancer, and/or a transcription termination signal, may also be included.

The term "exogenous," when used in relation to a protein, gene, nucleic

30     acid, or polynucleotide in a cell or organism refers to a protein, gene, nucleic acid, or polynucleotide which has been introduced into the cell or organism by artificial or natural means. An exogenous nucleic acid may be from a different

organism or cell, or it may be one or more additional copies of a nucleic acid which occurs naturally within the organism or cell. By way of a non-limiting example, an exogenous nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence  
5 than that found in nature.

The term "isolated" when used in relation to a nucleic acid, peptide, polypeptide or virus refers to a nucleic acid sequence, peptide, polypeptide or virus that is identified and separated from at least one contaminant nucleic acid, polypeptide or other biological component with which it is ordinarily associated  
10 in its natural source, e.g., so that it is not associated with *in vivo* substances, or is substantially purified from *in vitro* substances. Isolated nucleic acid, peptide, polypeptide or virus is present in a form or setting that is different from that in which it is found in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA  
15 sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. The isolated nucleic acid molecule may be present in single-stranded or double-stranded form. When an isolated nucleic acid molecule is to be utilized to express a protein, the molecule will contain at a minimum the  
20 sense or coding strand (i.e., the molecule may single-stranded), but may contain both the sense and anti-sense strands (i.e., the molecule may be double-stranded).

As used herein, the term "recombinant nucleic acid" or "recombinant DNA sequence, molecule or segment" refers to a nucleic acid, e.g., to DNA, that  
25 has been derived or isolated from a source, that may be subsequently chemically altered *in vitro*, and includes, but is not limited to, a sequence that is naturally occurring, is not naturally occurring, or corresponds to naturally occurring sequences that are not positioned as they would be positioned in the native genome. An example of DNA "derived" from a source, would be a DNA  
30 sequence that is identified as a useful fragment, and which is then chemically synthesized in essentially pure form. An example of such DNA "isolated" from a source would be a useful DNA sequence that is excised or removed from said source by chemical means, e.g., by the use of restriction endonucleases, so that it

can be further manipulated, e.g., amplified, for use in the disclosure, by the methodology of genetic engineering.

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule that is expressed from a recombinant nucleic acid molecule.

The term "peptide", "polypeptide" and "protein" are used interchangeably herein unless otherwise distinguished.

The term "sequence homology" means the proportion of base matches between two nucleic acid sequences or the proportion amino acid matches between two amino acid sequences. When sequence homology is expressed as a percentage, e.g., 50%, the percentage denotes the proportion of matches over the length of a selected sequence that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of 15 bases or less are usually used, 6 bases or less or 2 bases or less. When using oligonucleotides as probes or treatments, the sequence homology between the target nucleic acid and the oligonucleotide sequence is generally not less than 17 target base matches out of 20 possible oligonucleotide base pair matches (85%); e.g., not less than 9 matches out of 10 possible base pair matches (90%), or not less than 19 matches out of 20 possible base pair matches (95%).

The term "selectively hybridize" means to detectably and specifically bind. Polynucleotides, oligonucleotides and fragments of the disclosure selectively hybridize to nucleic acid strands under hybridization and wash conditions that minimize appreciable amounts of detectable binding to nonspecific nucleic acids. High stringency conditions can be used to achieve selective hybridization conditions as known in the art and discussed herein. Generally, the nucleic acid sequence homology between the polynucleotides, oligonucleotides, and fragments of the disclosure and a nucleic acid sequence of interest is at least 65%, and more typically with increasing homologies of at least about 70%, about 90%, about 95%, about 98%, and 100%.

Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for

maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less or 2 or less.

Alternatively, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used  
5 herein, if they have an alignment score of at more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. The two sequences or parts thereof may be homologous if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program.

10       The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (e.g., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence that encodes a polypeptide or its complement, or that a polypeptide sequence is identical in sequence or function to a reference polypeptide sequence. For illustration, the  
15 nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and  
20 "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length,  
25 frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more)  
30 polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence

5 (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by using local homology algorithms or by a search for similarity method, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA Genetics Software Package or by

10 inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denote a characteristic of a

15 polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, e.g., at least 90 to 95 percent sequence identity, or at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions,

20 frequently over a window of at least 20-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total

25 20 percent or less of the reference sequence over the window of comparison.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least about 80% sequence identity, at least about 90% sequence identity, at least about 95% percent sequence 5 identity, or at least about 99% sequence identity.

A "protective immune response" and "prophylactic immune response" are used interchangeably to refer to an immune response which targets an immunogen to which the individual has not yet been exposed or targets a protein associated with a disease in an individual who does not have the disease, such as 10 a tumor associated protein in a patient who does not have a tumor.

A "therapeutic immune response" refers to an immune response which targets an immunogen to which the individual has been exposed or a protein associated with a disease in an individual who has the disease.

The term "prophylactically effective amount" is meant to refer to the 15 amount, in the case of infectious agents, prevent an individual from developing an infection, and in the case of diseases, prevent an individual from developing a disease.

The term "therapeutically effective amount" is meant to refer to the amount, in the case of infectious agents, reduce the level of infection in an 20 infected individual in order to reduce symptoms or eliminate the infection, and in the case of diseases, to reduce symptoms or cure the individual.

"Inducing an immune response against an immunogen" is meant to refer to induction of an immune response in a naïve individual and induction of an immune response in an individual previously exposed to an immunogen wherein 25 the immune response against the immunogen is enhanced.

As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual species in the composition), and optionally a substantially purified fraction is a composition wherein the object species comprises at least 30 about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 percent of all macromolecular species present in the composition, more than

about 85%, about 90%, about 95%, and about 99%. For example, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species.

5 "Transfected," "transformed" or "transgenic" is used herein to include any host cell or cell line, which has been altered or augmented by the presence of at least one recombinant DNA sequence. The host cells of the present disclosure are typically produced by transfection with a DNA sequence in a plasmid expression vector, as an isolated linear DNA sequence, or infection with a  
10 recombinant viral vector.

#### Exemplary Vectors, Viruses and Methods

Most of the vaccines (mRNA vaccines, viral vector vaccines, recombinant protein vaccines, etc.) against SARS-CoV-2 currently implemented are intended to induce antibodies in the blood to inhibit the function of the spike  
15 protein on the virus particles by intramuscular administration. The purpose of these vaccines is to induce blood antibodies to inhibit the function of spike proteins on viral particles by intramuscular administration. However, the induction of immunity in the upper respiratory tract mucosa is not sufficient. A "semi-live virus" (attenuated) SARS-COV-2 vaccine that can induce immunity,  
20 e.g., in the nasal mucosa through intranasal inoculation, is described herein.

The "semi-viable viruses" are viruses that, by lacking the viral proteins essential for multiplication, invade cells and express viral proteins to induce immunity in the upper respiratory mucosa for infection defense, but do not produce new infectious progeny particles. As with other attenuated live viruses  
25 (e.g., FluMist vaccine using cold-acclimated influenza virus), it is possible to induce not only liquid immunity but also cellular immunity. In addition, since "semi-viable viruses" do not have proliferative capacity, the risk of reversion to virulence is low, and they are safer than attenuated live viruses.

Because certain attenuated viruses induce local mucosal immunity, they  
30 can be used through intranasal administration. And because it is not a viral vector vaccine, it can be administered multiple times. Moreover, unlike mRNA, viral vector, or recombinant protein vaccines that target only spike proteins, these vaccines are expected to induce immune responses against structural

proteins other than spike proteins. Further, since innate immunity can be activated by the establishment of a single infection, there is no need to use immunostimulants (adjuvants).

Since this vaccine is produced using reverse genetics, the S-protein gene 5 can be easily replaced, making it possible to respond to epidemics of mutant strains with different antigenic properties. Therefore, an attenuated virus such as a "semi-viable" vaccine can make a significant contribution to the development of vaccines against infectious diseases other than SARS-CoV-2.

The disclosure provides isolated vectors, e.g., plasmids, which encode 10 positive-sense, single stranded RNA viruses and/or express vRNA from recombinant nucleic acid corresponding to sequences for mutant positive-sense, single stranded RNA viruses. When introduced into a cell, a combination of these vectors is capable of yielding recombinant infectious but not necessarily replication competent virus after infection of a cell such as a non-helper cell. 15 Thus, the disclosure includes host cells that produce recombinant infectious, attenuated (semi-live) virus of the disclosure. In one embodiment, the disclosure provides isolated vectors, e.g., plasmids, which encode coronavirus proteins and/or express mutant coronavirus vRNA which, when introduced into a cell, are capable of yielding recombinant infectious, attenuated coronavirus. The 20 disclosure includes host cells that transiently or stably produce recombinant infectious, attenuated coronavirus, including helper cells, and isolated recombinant coronavirus prepared by the methods disclosed herein.

The vectors include those for mRNA production and vRNA production. 25 In one embodiment, the vectors include coronavirus DNA, for example, vectors for mRNA production with sequences corresponding to one or more open reading frames encoding coronavirus proteins, or vectors for vRNA production that include a genetic modification such as a deletion in the full-length genomic sequence, e.g., the modification may be a deletion including internal coronavirus sequences corresponding to at least a portion of one open reading frame. The 30 RNA produced from the vRNA vector is capable of being packaged into virions in the presence of coronavirus proteins but as part of the resulting virion, is not capable of being replicated and so does not result in virus production when that virion is introduced to a cell that otherwise supports coronavirus replication and

which cell does not express at least one coronavirus protein *in trans*, e.g., a cell that is not a coronavirus helper cell.

Candidate sequences for mutation including deletion, substitution or insertion, in any combination, and optional replacement with heterologous sequences include but are not limited to E, M or N encoding sequences or corresponding sequences in other positive-sense, single stranded RNA viruses, e.g., sequences for nonstructural, nonpolymerase and/or nonglycosylated viral proteins or non-coding regions. The vectors may include gene(s) or portions thereof other than those of a positive-sense, single stranded RNA virus such as a 5 coronavirus (heterologous sequences), which genes or portions thereof are intended to be expressed in a host cell, either as a protein or incorporated into vRNA. Thus, a vector may include in addition to viral sequences, for instance, 10 coronavirus sequences, a gene or open reading frame of interest, e.g., a heterologous gene for an immunogenic peptide or protein useful as a vaccine or 15 a therapeutic protein.

If more than one vector is employed, the vectors may be physically linked or each vector may be present on an individual plasmid or other, e.g., linear, nucleic acid delivery vehicle. The vectors or plasmids may be introduced to any host cell, e.g., a eukaryotic cell such as a mammalian cell, that supports 20 viral replication. Host cells useful to prepare virus of the disclosure include but are not limited to insect, avian or mammalian host cells such as canine, feline, equine, bovine, ovine, or primate cells including simian or human cells. In one embodiment, the host cell is one that is approved for vaccine production.

The viruses produced by methods described herein are useful in viral 25 mutagenesis studies, drug screening and in the production of vaccines and gene therapy vectors (e.g., for cancer, AIDS, adenosine deaminase, muscular dystrophy, ornithine transcarbamylase deficiency and central nervous system tumors). In particular, an attenuated coronavirus of the disclosure which induces strong humoral and cellular immunity may be employed as a vaccine vector.

30 Thus, a virus for use in medical therapy (e.g., for a vaccine or gene therapy) is provided. For example, the disclosure provides a method to immunize an animal against a pathogen, e.g., a virus, bacteria, or parasite, or a

malignant tumor. The method comprises administering to the animal an effective amount of at least one isolated virus of the disclosure which encodes and expresses, or comprises nucleic acid for an immunogenic peptide or protein of a pathogen or tumor, optionally in combination with an adjuvant, effective to 5 immunize the animal.

To prepare expression cassettes for transformation herein, the recombinant DNA sequence or segment may be circular or linear, double-stranded or single-stranded. A DNA sequence which encodes an RNA sequence that is substantially complementary to a mRNA sequence encoding a gene 10 product of interest is typically a "sense" DNA sequence cloned into a cassette in the opposite orientation (i.e., 3' to 5' rather than 5' to 3'). Generally, the DNA sequence or segment is in the form of chimeric DNA, such as plasmid DNA, that can also contain coding regions flanked by control sequences which promote the expression of the DNA in a cell. As used herein, "chimeric" means that a vector 15 comprises DNA from at least two different species, or comprises DNA from the same species, which is linked or associated in a manner which does not occur in the "native" or wild-type of the species.

Aside from DNA sequences that serve as transcription units, or portions thereof, a portion of the DNA may be untranscribed, serving a regulatory or a 20 structural function. For example, the DNA may itself comprise a promoter that is active in eukaryotic cells, e.g., mammalian cells, or in certain cell types, or may utilize a promoter already present in the genome that is the transformation target of the lymphotropic virus. Such promoters include the CMV promoter, as well as the SV40 late promoter and retroviral LTRs (long terminal repeat 25 elements), e.g., the MMTV, RSV, MLV or HIV LTR, although many other promoter elements well known to the art may be employed in the practice of the disclosure.

Other elements functional in the host cells, such as introns, enhancers, polyadenylation sequences and the like, may also be a part of the recombinant 30 DNA. Such elements may or may not be necessary for the function of the DNA, but may provide improved expression of the DNA by affecting transcription, stability of the mRNA, or the like. Such elements may be included in the DNA

as desired to obtain the optimal performance of the transforming DNA in the cell.

The recombinant DNA to be introduced into the cells may contain either a selectable marker gene or a reporter gene or both to facilitate identification and selection of transformed cells from the population of cells sought to be transformed. Alternatively, the selectable marker may be carried on a separate piece of DNA and used in a co-transformation procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers are well known in the art and include, for example, antibiotic and herbicide-resistance genes, such as *neo*, *hpt*, *dhfr*, *bar*, *aroA*, *puro*, *hyg*, *dapA* and the like. See also, the genes listed on Table 1 of Lundquist et al. (U.S. Patent No. 5,848,956).

Reporter genes are used for identifying potentially transformed cells and for evaluating the functionality of regulatory sequences. Reporter genes which encode for easily assayable proteins are well known in the art. In general, a reporter gene is a gene which is not present in or expressed by the recipient organism or tissue and which encodes a protein whose expression is manifested by some easily detectable property, e.g., enzymatic activity. Exemplary reporter genes include the chloramphenicol acetyl transferase gene (*cat*) from Tn9 of *E. coli*, the beta-glucuronidase gene (*gus*) of the *uidA* locus of *E. coli*, the green, red, or blue fluorescent protein gene, and the luciferase gene. Expression of the reporter gene is assayed at a suitable time after the DNA has been introduced into the recipient cells.

The general methods for constructing recombinant DNA which can transform target cells are well known to those skilled in the art, and the same compositions and methods of construction may be utilized to produce the DNA useful herein. For example, Sambrook et al., Molecular Cloning: A Laboratory Manual (2002) provides suitable methods of construction.

The recombinant DNA can be readily introduced into the host cells, e.g., mammalian, yeast or insect cells, by transfection with an expression vector comprising the recombinant DNA by any procedure useful for the introduction

into a particular cell, e.g., physical or biological methods, to yield a transformed (transgenic) cell having the recombinant DNA so that the DNA sequence of interest is expressed by the host cell. In one embodiment, at least one of the recombinant DNA which is introduced to a cell is maintained  
5 extrachromosomally. In one embodiment, at least one recombinant DNA is stably integrated into the host cell genome.

Physical methods to introduce a recombinant DNA into a host cell include calcium-mediated methods, lipofection, particle bombardment, microinjection, electroporation, and the like. Biological methods to introduce  
10 the DNA of interest into a host cell include the use of DNA and RNA viral vectors. Viral vectors, e.g., retroviral or lentiviral vectors, have become a widely used method for inserting genes into eukaryotic, such as mammalian, e.g., human, cells. Other viral vectors useful to introduce genes into cells can be derived from poxviruses, e.g., vaccinia viruses, herpes viruses, adenoviruses,  
15 adeno-associated viruses, baculoviruses, and the like.

To confirm the presence of the recombinant DNA sequence in the host cell, a variety of assays may be performed. Such assays include, for example, molecular biological assays well known to those of skill in the art, such as Southern and Northern blotting, RT-PCR and PCR; biochemical assays, such as  
20 detecting the presence or absence of a particular gene product, e.g., by immunological means (ELISAs and Western blots) or by other molecular assays.

To detect and quantitate RNA produced from introduced recombinant DNA segments, RT-PCR may be employed. In this application of PCR, RNA is reverse transcribed into DNA, using enzymes such as reverse transcriptase, and  
25 then the DNA is amplified through the use of conventional PCR techniques. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique demonstrates the presence of an RNA species and gives information about the integrity of that RNA. The  
30 presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the recombinant DNA segment in question, they do not provide information as to whether the recombinant DNA segment is being expressed. Expression may be evaluated by specifically identifying the peptide products of the introduced DNA sequences or 5 evaluating the phenotypic changes brought about by the expression of the introduced DNA segment in the host cell.

The recombinant viruses described herein have modifications in genomic sequences relative to a corresponding wild-type viral genome, i.e., the genome of the recombinant virus has a modification which includes a deletion, and 10 optionally an insertion, in a region corresponding to sequences for a viral protein that is associated with transcription, is nonstructural or is nonglycosylated. The mutation in the viral genome is effective to inhibit or prevent production of at least one functional viral protein from that genome, e.g., when those sequences are present in a nontransgenic cell which supports viral replication. In one 15 embodiment, the deletion includes from 1 up to thousands of nucleotides, e.g., 1%, 10%, 50%, 90% or more of sequences corresponding to the coding region for the viral protein. In one embodiment, the deleted sequences correspond to sequences with a substantial identity, e.g., at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, nucleic acid sequence 20 identity, to E sequences and/or M sequences. In one embodiment, the deletion includes from 1 up to hundreds of nucleotides, e.g., 1%, 10%, 50%, 90% or more of sequences corresponding to at N coding sequences.

In one embodiment, the viral genome provides for an attenuated, e.g., 25 replication-incompetent, positive-sense, single-stranded RNA virus, which genome includes a deletion in sequences corresponding to those in a wild-type viral genome for a protein that is associated with viral assembly and/or progeny production, and may include heterologous sequences that are nontoxic to host cells including cells in an organism to be immunized. In one embodiment, the heterologous sequence is a marker sequence, a selectable sequence or other 30 sequence which is detectable or capable of detection, e.g., GFP or luciferase, or a selectable gene such as an antibiotic resistance gene, e.g., a hygromycin B resistance gene or neomycin phosphotransferase gene, which marker gene or selectable gene is not present in the host cell prior to introduction of the vector.

Pharmaceutical Compositions

Pharmaceutical compositions, suitable for inoculation, e.g., nasal, parenteral or oral administration, such as by intravenous, intramuscular, intranasal, topical or subcutaneous routes, comprise one or more virus isolates, 5 e.g., one or more recombinant attenuated positive-sense, single stranded RNA virus isolates, optionally further comprising sterile aqueous or non-aqueous solutions, suspensions, and emulsions. The compositions can further comprise auxiliary agents or excipients, as known in the art. The composition is generally presented in the form of individual doses (unit doses). Preparations for 10 parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and/or emulsions, which may contain auxiliary agents or excipients known in the art. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Carriers or occlusive dressings can be used to 15 increase skin permeability and enhance antigen absorption. Liquid dosage forms for oral administration may generally comprise a liposome solution containing the liquid dosage form. Suitable forms for suspending liposomes include emulsions, suspensions, solutions, syrups, and elixirs containing inert diluents commonly used in the art, such as purified water. Besides the inert diluents, 20 such compositions can also include adjuvants, wetting agents, emulsifying and suspending agents, or sweetening, flavoring, or perfuming agents.

When a composition is used for administration to an individual, it can further comprise salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. For vaccines, 25 adjuvants, substances which can augment a specific immune response, can be used. Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same site of the organism being immunized.

The pharmaceutical compositions comprise a therapeutically effective 30 amount of the virus, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeiae for use in animals,

and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the pharmaceutical composition is administered. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable 5 pharmaceutical excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. These compositions can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained-release formulations and the like.

10 These compositions can be formulated as a suppository. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical Sciences" by E. W. Martin. Such compositions will contain a 15 therapeutically effective amount of the virus, e.g., in purified form, together with a suitable amount of carrier so as to provide the form for proper administration to the patient. The formulation should suit the mode of administration.

The compositions may be systemically administered, e.g., orally or intramuscularly, in combination with a pharmaceutically acceptable vehicle such 20 as an inert diluent. For oral administration, the virus may be combined with one or more excipients and used in the form of ingestible capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 25 to about 60% of the weight of a given unit dosage form. The amount of active compound in such useful compositions is such that an effective dosage level will be obtained.

The compositions may also contain the following: binders such as gum tragacanth, acacia, corn starch or gelatin; excipients such as dicalcium 30 phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, fructose, lactose or aspartame or a flavoring agent such as peppermint, oil of wintergreen, or cherry flavoring may be added. Various other materials may be present. For instance, a syrup or elixir may contain the virus,

sucrose or fructose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavoring such as cherry or orange flavor. Of course, any material used in preparing any unit dosage form, including sustained-release preparations or devices, should be pharmaceutically acceptable and substantially non-toxic in the amounts employed. The composition also can be administered intravenously or intraperitoneally by infusion or injection. Solutions of the virus can be prepared in water or a suitable buffer, optionally mixed with a nontoxic surfactant. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, triacetin, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of undesirable microorganisms.

The pharmaceutical dosage forms suitable for injection or infusion can include sterile aqueous solutions or dispersions or sterile powders comprising the active ingredient which are adapted for the extemporaneous preparation of sterile injectable or infusible solutions or dispersions, optionally encapsulated in liposomes. In all cases, the ultimate dosage form should be sterile, fluid and stable under the conditions of manufacture and storage. The liquid carrier or vehicle can be a solvent or liquid dispersion medium comprising, for example, water, ethanol, a polyol (for example, glycerol, propylene glycol, liquid polyethylene glycols, and the like), vegetable oils, nontoxic glyceryl esters, and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the formation of liposomes, by the maintenance of the particle size in the case of dispersions or by the use of surfactants. The prevention of the action of undesirable microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it may be preferable to include isotonic agents, for example, sugars, buffers or sodium chloride.

Sterile injectable solutions are prepared by incorporating the virus in the amount in the appropriate solvent with various of the other ingredients enumerated above, followed by filter sterilization.

Useful liquid carriers include water, alcohols or glycols or water-alcohol/glycol blends, in which the present viruses can be dissolved or dispersed at effective levels, optionally with the aid of non-toxic surfactants. Adjuvants such as fragrances and additional antimicrobial agents can be added to optimize

the properties for a given use. The resultant liquid compositions can be applied from absorbent pads, used to impregnate bandages and other dressings, or sprayed onto the affected area using pump-type or aerosol sprayers.

Useful dosages of the viruses of the disclosure can be determined by  
5 comparing their *in vitro* activity and *in vivo* activity in animal models.

#### Pharmaceutical Purposes

The administration of the composition may be for either a “prophylactic” or “therapeutic” purpose. When provided prophylactically, the compositions of the disclosure which are vaccines are provided before any symptom or clinical  
10 sign of a pathogen infection becomes manifest. The prophylactic administration of the composition serves to prevent or attenuate any subsequent infection. When provided prophylactically, the gene therapy compositions of the disclosure, are provided before any symptom or clinical sign of a disease becomes manifest. The prophylactic administration of the composition serves to  
15 prevent or attenuate one or more symptoms or clinical signs associated with the disease.

When provided therapeutically, a viral vaccine is provided upon the detection of a symptom or clinical sign of actual infection. The therapeutic administration of the compound(s) serves to attenuate any actual infection.  
20 When provided therapeutically, a gene therapy composition is provided upon the detection of a symptom or clinical sign of the disease. The therapeutic administration of the compound(s) serves to attenuate a symptom or clinical sign of that disease.

Thus, a vaccine composition of the present disclosure may be provided  
25 either before the onset of infection (so as to prevent or attenuate an anticipated infection) or after the initiation of an actual infection. Similarly, for gene therapy, the composition may be provided before any symptom or clinical sign of a disorder or disease is manifested or after one or more symptoms are detected.

30 A composition is said to be “pharmacologically acceptable” if its administration can be tolerated by a recipient mammal. Such an agent is said to be administered in a “therapeutically effective amount” if the amount

administered is physiologically significant. A composition of the present disclosure is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient, e.g., enhances at least one primary or secondary humoral or cellular immune response against at least one strain of a virus.

The “protection” provided need not be absolute, i.e., the infection need not be totally prevented or eradicated, if there is a statistically significant improvement compared with a control population or set of mammals. Protection may be limited to mitigating the severity or rapidity of onset of symptoms or clinical signs of the virus infection.

#### Pharmaceutical Administration

A composition may confer resistance to one or more pathogens, e.g., one or more virus, bacterium or parasite strains, by either passive immunization or active immunization. In active immunization, a live vaccine composition is administered prophylactically to a host (e.g., a mammal), and the host’s immune response to the administration protects against infection and/or disease. For passive immunization, the elicited antisera can be recovered and administered to a recipient suspected of having an infection caused by at least one virus strain.

The present disclosure thus includes methods for preventing or attenuating a disorder or disease, e.g., an infection by at least one strain of pathogen. As used herein, a vaccine is said to prevent or attenuate a disease if its administration results either in the total or partial attenuation (i.e., suppression) of a clinical sign or condition of the disease, or in the total or partial immunity of the individual to the disease.

At least one virus isolate of the present disclosure, may be administered by any means that achieve the intended purposes. For example, administration of such a composition may be by various parenteral routes such as subcutaneous, intravenous, intradermal, intramuscular, intraperitoneal, intranasal, oral or transdermal routes. Parenteral administration can be accomplished by bolus injection or by gradual perfusion over time.

A typical regimen for preventing, suppressing, or treating a viral related pathology, comprises administration of an effective amount of a vaccine

composition as described herein, administered as a single treatment, or repeated as enhancing or booster dosages, for instance, over a period up to and including between one week and about 24 months, or any range or value therein.

According to the present disclosure, an “effective amount” of a composition is one that is sufficient to achieve a desired effect. It is understood that the effective dosage may be dependent upon the species, age, sex, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect wanted. The ranges of effective doses provided below are not intended to limit the disclosure and represent dose ranges.

Exemplary doses include but are not limited to from about  $10^4$  to  $10^8$  virus particles (vp) or genomes (vg),  $10^6$  to  $10^8$  vp or vg,  $10^6$  to  $10^{10}$  vp or vg, or  $10^8$  to  $10^{12}$  vp or vg, or more, or from about  $10^6$  to  $10^8$  vp or vg,  $10^8$  to  $10^{10}$  vp or vg, or  $10^{10}$  to  $10^{12}$  vp or vg, or from about  $10^2$  to  $10^3$  plaque forming units (pfu) or TCID<sub>50</sub>,  $10^3$  to  $10^4$  pfu or TCID<sub>50</sub>,  $10^4$  to  $10^5$  pfu or TCID<sub>50</sub>,  $10^5$  to  $10^7$  pfu or TCID<sub>50</sub>,  $10^6$  to  $10^8$  pfu or TCID<sub>50</sub>,  $10^6$  to  $10^{10}$  pfu or TCID<sub>50</sub>, or  $10^8$  to  $10^{12}$  pfu or TCID<sub>50</sub>, or more, or from about  $10^6$  to  $10^8$  pfu or TCID<sub>50</sub>,  $10^8$  to  $10^{10}$  pfu or TCID<sub>50</sub>, or  $10^{10}$  to  $10^{12}$  pfu or TCID<sub>50</sub>.

#### Exemplary Coronavirus Proteins

In one embodiment, there is reduced or an absence of expression from the mutant viral genome of an E protein having SEQ ID NO:1, SEQ ID NO:5, or SEQ ID NO:9, or a protein having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 97%, 98% or 99%, amino acid sequence identity thereto.

In one embodiment, there is reduced or an absence of expression from the mutant viral genome of a M protein having SEQ ID NO:2, SEQ ID NO:6, or SEQ ID NO:10, or a protein having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 97%, 98% or 99%, amino acid sequence identity thereto.

In one embodiment, an isolated host cell expresses an E protein having SEQ ID NO:1, SEQ ID NO:5, or SEQ ID NO:9, or a protein having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 97%, 98% or 99%, amino acid sequence identity thereto.

In one embodiment, th an isolated host cell expresses a M protein having SEQ ID NO:2, SEQ ID NO:6, or SEQ ID NO:10, or a protein having at least 80%, 82%, 84%, 85%, 87%, 89%, 90%, 92%, 94%, 95%, 97%, 98% or 99%, amino acid sequence identity thereto.

5    Exemplary Embodiments

The disclosure provides a vaccine comprising an effective amount of a recombinant positive-sense, single stranded RNA virus, the genome of which contains, in one embodiment, a deletion of viral sequences corresponding to those for a structural, nonstructural and/or nonglycosylated viral protein that is 10 essential *in trans* for viral replication and/or progeny production and in one embodiment, one or more insertions of a nucleotide sequence encoding one or more heterologous gene products, wherein the insertions may be in coding or non-coding sequences. In one embodiment, the heterologous gene product is from a heterologous virus, or a bacteria or fungus. In one embodiment, the 15 heterologous gene product is a glycoprotein. In one embodiment, the insertions may replace coding sequences, or may replace non-coding sequences. In one embodiment, the deletion is effective to inhibit or prevent viral genome replication or progeny production upon infection of a cell with the recombinant positive-sense, single stranded RNA virus. For example, the deletion of viral 20 sequences corresponding to those for a structural, nonstructural and/or nonglycosylated viral protein that is essential *in trans* for viral replication or progeny production may be effective to prevent expression of a functional structural, nonstructural or nonglycosylated protein upon infection of a cell with the recombinant positive-sense, single stranded RNA virus. In one embodiment, 25 the deletion of viral sequences corresponds to those for a structural, nonstructural or nonglycosylated viral protein that is essential *in trans* for viral replication or progeny production, e.g., the deletion may be in coronavirus sequences for a viral protein corresponding to the E protein, the M protein, the N protein, or any combination thereof. In one embodiment, the genome of the 30 recombinant, attenuated coronavirus comprises heterologous sequences, for instance, positioned within the deletion in E protein, the M protein, the N protein, or any combination thereof, related sequences. Any of the deletions in viral sequences of a positive-sense, single stranded RNA virus may include a deletion of 1 or more nucleotides, e.g., a deletion of at least 0.1%, 1%, 5%, 10%,

50%, 60%, 70%, 80%, 90%, or any integer in between, and up to 100% of the viral coding sequences corresponding to those for a structural, nonstructural, glycosylated or nonglycosylated viral protein. The deletion of viral sequences corresponding to those for a structural, nonstructural or nonglycosylated viral

5 protein that is essential *in trans* for viral replication is one that is stable over multiple passages and is readily detectable, e.g., by RT-PCR. In one embodiment, the genome of the recombinant virus has a deletion in viral sequences for two or more structural, nonstructural or nonglycosylated proteins, for example, a deletion in coding sequences for viral proteins that are contiguous

10 with each other, such as sequences for a viral protein corresponding to E protein and for a viral protein corresponding to M protein. In one embodiment, the genome of the recombinant virus has a deletion in viral sequences for two or more structural, nonstructural or nonglycosylated proteins, for example, a deletion in coding sequences for viral proteins that are not contiguous with each

15 other, such as sequences for a viral protein corresponding to E protein and for a viral protein corresponding to N protein. In one embodiment, where the genome of the recombinant virus has a deletion in viral sequences for a structural, nonstructural, glycosylated or nonglycosylated protein, at least a portion of the deleted viral sequences may be replaced with a nucleotide sequence encoding an

20 antigen or other gene product that is expressed in the recombinant coronavirus which, when administered to a mammal, is prophylactic or therapeutic. In one embodiment, where the genome of the recombinant virus has a deletion in viral sequences for two or more proteins that are structural, nonstructural, glycosylated or nonglycosylated proteins, at least a portion of one of the deleted

25 viral sequences may be replaced with a nucleotide sequence encoding an antigen that is expressed in the recombinant coronavirus which, when administered to a mammal, is prophylactic or therapeutic. The vaccine of the disclosure may provide for subtype cross protection, for coronavirus cross protection and optionally as a bi- or multi-valent vaccine for pathogens other than coronavirus.

30 In one embodiment, a monovalent recombinant coronavirus vaccine comprises one or more adjuvants and a recombinant coronavirus, the expression of the genome results in a virus having a heterologous glycoprotein, e.g., inserted into sequences corresponding to coronavirus E, M or N.

In one embodiment, the mutant genome further comprises a nucleotide sequence encoding a prophylactic or therapeutic heterologous gene product. In one embodiment, the nucleotide sequence is inserted within 500 nucleotides of the deletion site or at the site of the deletion. In one embodiment, the nucleotide sequence is inserted into the coronavirus genome at a site other than the site of the deletion in the polynucleotide. In one embodiment, the nucleotide sequence replaces E or M sequences or a portion thereof. In one embodiment, the nucleotide sequence is inserted into E or M coding sequences. In one embodiment, the heterologous gene product comprises a heterologous glycoprotein. In one embodiment, the vaccine of further comprises a pharmaceutically acceptable carrier. In one embodiment, the recombinant coronavirus in the vaccine is inactivated.

A method to immunize a mammal using a composition having the recombinant coronavirus is also provided. In one embodiment, the mammal is a human. In one embodiment, two doses of the composition are administered. In one embodiment, a single dose is administered.

The disclosure provides for bi- or multi-valent vaccines to address combinations of diseases that impact particular areas. Monovalent vaccines may be particularly useful in response to any outbreaks that don't correspond well to other vaccines. Multivalent vaccines may be based on the addition of exogenous sequences into any of several positions in the coronavirus genome including but not limited to: 1) the E, M or N open reading frame, 2) the E open reading frame, or 3) the M open reading frame. In one embodiment, a bivalent vaccine virus may express a one or more nonglycosylated proteins, one or more glycosylated proteins, or at least one nonglycosylated protein and at least one glycosylated protein.

Thus, in one embodiment, a recombinant coronavirus, wherein the genome of the recombinant coronavirus contains a deletion of one or more nucleotides in a polynucleotide sequence for a viral protein corresponding to SARS-CoV-2 E protein which deletion is effective to prevent expression of a functional viral protein corresponding to SARS-CoV-2 E protein upon infection of a cell with the recombinant coronavirus, and the genome encodes one or more coronavirus glycoproteins.

Thus, in one embodiment, a recombinant coronavirus, wherein the genome of the recombinant coronavirus contains a deletion of one or more nucleotides in a polynucleotide sequence for a viral protein corresponding to SARS-CoV-2 M protein which deletion is effective to prevent expression of a functional viral protein corresponding to SARS-CoV-2 M protein upon infection of a cell with the recombinant coronavirus, and the genome encodes one or more coronavirus glycoproteins.

Further provided is a multivalent vaccine comprising an effective amount of a recombinant coronavirus, wherein the genome of the recombinant coronavirus contains a deletion in one or more nucleotides for a polynucleotide sequence for a viral protein corresponding to E M or N, or a combination thereof, which deletion is effective to prevent expression of a functional viral protein corresponding to E, M or N protein upon infection of a cell with the recombinant coronavirus, and wherein the genome encodes one or more coronavirus glycoproteins and at least one heterologous gene product.

In one embodiment, the prophylactic or therapeutic heterologous gene product is not a glycoprotein, e.g., a nonglycosylated protein. In one embodiment, the prophylactic or therapeutic heterologous gene does not encode a protein. In one embodiment, the gene product comprises a glycoprotein.

Further provided is a method to immunize a mammal, e.g., a human, by administering to the mammal an effective amount of the vaccine. For example, a human in contact with coronavirus infected individuals or inadvertently exposed to coronavirus, e.g., in a laboratory, may be administered the recombinant attenuated virus of the disclosure in an amount effective to inhibit or substantially eliminate coronavirus replication in the human.

Positive-sense, single stranded RNA viruses other than SARS-CoV-2 may likewise be manipulated, e.g., the genome of alphaletovirus, alphacoronavirus, betacoronavirus, gammacoronavirus, deltacoronavirus, nidovirales, and the like, may be manipulated to mutate or delete sequences corresponding to those for a nonstructural or nonglycosylated viral protein that may be required for viral genome replication or progeny production. Thus, genomes of viruses in the above-mentioned families may be manipulated to provide for an attenuated virus that resembles wild-type virus in its life cycle,

morphology, and growth properties, can be grown to reasonably high titers in helper cells, is genetically stable, and is safe.

The disclosure also provides a method to prepare an attenuated positive-sense, single stranded RNA virus, e.g., coronavirus. In one embodiment, the 5 method includes providing a host cell, e.g., a Vero cell, having one or a plurality of vectors which when expressed (stably or transiently) are effective to yield attenuated positive-sense, single stranded RNA virus. In one embodiment, the plurality of vectors includes a vector for vRNA production comprising a promoter operably linked to a virus DNA which contains a deletion of sequences 10 for a viral gene in the viral genome, which results in a mutant viral genome, which deletion is effective to prevent expression of a functional viral protein corresponding to, for example, E, M or N protein, linked to a transcription termination sequence, and optionally an insertion of heterologous sequences as discussed above. The host cell also includes a vector for mRNA production 15 comprising a promoter operably linked to a DNA segment encoding the viral protein that is not expressed from the mutant viral genome. Then attenuated virus is isolated from the cell. In one embodiment, the host cell is transiently transfected with the plurality of vectors and virus collected within 1, 2, 3, and up to 7 days post-transfection. In one embodiment, the host cell is one that is 20 approved for vaccine production. In one embodiment, additional heterologous sequences are included in the vRNA vector or in mRNA vectors subsequently introduced to the host cell, and/or are introduced to the host cell via a mRNA vector. In one embodiment, the additional heterologous sequences are for an immunogenic polypeptide or peptide of a pathogen, a tumor antigen, or a 25 therapeutic protein.

In one embodiment, a method to prepare a multivalent attenuated coronavirus is provided. The method includes providing a host cell comprising a plurality of coronavirus vectors which, when expressed in the host cell, are effective to yield attenuated coronavirus, wherein the plurality of vectors 30 includes a vector for vRNA production comprising a promoter operably linked to a coronavirus DNA which contains a viral genome having a deletion in sequences for a functional viral protein corresponding to, for example, E, M or N protein, which deletion is effective to prevent expression of the functional viral

protein linked to a transcription termination sequence, and a vector for mRNA production comprising a promoter operably linked to a DNA segment encoding the coronavirus protein corresponding to E, M or N, and a vector for mRNA production comprising a promoter operably linked to a DNA segment encoding a coronavirus protein corresponding to E, M or N; and isolating attenuated coronavirus from the host cell. In one embodiment, the cells are mammalian cells. In one embodiment, the cells are primate cells. In one embodiment, the cells are Vero cells. In one embodiment, the gene product sequences for an immunogenic polypeptide or peptide of a pathogen, a tumor antigen, or a therapeutic protein. In one embodiment, each vector encoding a coronavirus protein is on a separate plasmid.

#### Exemplary Mutations for Cold-Adaptation

Table 1

Mutation sites of SARS-CoV-2 TS11 compared with WA1 strain.

15

Nucleotide Position	WA1		TS-11		Protein
	Nucleotide	Amino Acid	Nucleotide	Amino Acid	
344	CTC	L	TTC	F	
548	ATT	I	CTT	L	nsp1
2393	GTC	V	ATC	I	
4200	ATG	M	AAG	K	
4455	GCC	A	GTC	V	
5007	ACG	T	ATG	M	nsp3
5097	ATT	N	AGT	S	
7086	ACT	T	ATT	I	
15,240	AAC	N	AAT	N	nsp12
16,411	GAT	D	AAT	N	nsp13
19,893	GAT	D	GAG	E	nsp15
20,863	CAT	H	TAT	Y	nsp16

Nucleotide Position	WA1		TS-11		Protein
	Nucleotide	Amino Acid	Nucleotide	Amino Acid	
22,120	TTC	F	TTT	F	
22,296	CAT	H	CGT	R	
23,594– 23,629		TNSPRRARSVAS	36-nt-Del	12-aa-Del	S
24,000	AGC	S	ATC	I	
24,554	ACA	T	GCA	A	
26,339	GCC	A	GTC	V	E
26,571	CTT	L	TTT	F	
26,907	CTG	L	TTG	L	M
27,524	TCA	S	TTA	L	orf7a
27,807– 28,177			371-nt-Del		deletion of aa 18–43 of orf7b; deletion of orf8
28,866	ACT	T	ATT	I	N

**Table 2.**

	ORF1ab										Structural protein genes			
	nsp3 D-37	nsp4 T6763C T7603A*	nsp5 C10973T	nsp6 C10973T	nsp9	nsp10 G12660A G12754T*	nsp12 G12754N C13126T	nsp13 C16228G*	nsp14	nsp16 T20743C C10937	S A21965C C23926G G24822C	E C21622T* A22943C C23926G A24467C T24959C	M C26213A C26671T	N G26671T G28830A
D-ca	T6763C	C6881T A8882C*	C10973T	-	G12660A G12754T*	C12861A* C13126T	T13285A	-	-	T20743C	C21622T* A22943C C23926G A24467C T24959C	C26671T	G28830A	
D-B4	T6763C	A8882C*	C10973T	-	G12754N	C13126T	T13285A	-	-	T20743C C10937	C21622T* A22943C C23926G C24467C* A24467C T24959C	C26213A C26671T	-	
D-D2	C3496G T6763C C7662T	C8881A	G10988T	T13285C	G12754N	-	T13285A	-	A16817Y A20918C*	A16817Y A20918C*	C21616T* C21621A* A22943C C23926G A24467C T24959C T25241G*	G26245A* C26262T	C26671T	-

**ORF** - open reading frame; **nsp** - nonstructural protein; structural proteins: **S** - spike, **E** - envelope, **M** - membrane, **N** - nucleocapsid; **N** - undefined nucleotide.

Green color highlights substitutions common for all virus variants; blue and orange color highlights substitutions common only for ca variants; orange color highlights substitutions characteristic only for a given variant. An "\*" indicates unique substitutions in SARS-CoV-2

5 variants in relation to viruses deposited in GenBank.'

**Table 3.**  
**ORF1a**

<b>ORF1ab</b>											<b>Structural proteins</b>			
<b>nsp3</b>	<b>nsp4</b>	<b>nsp5</b>	<b>nsp6</b>	<b>nsp9</b>	<b>nsp10</b>	<b>nsp12</b>	<b>nsp13</b>	<b>nsp14</b>	<b>nsp16</b>	<b>S</b>	<b>E</b>	<b>M</b>	<b>N</b>	
D-37	I2181T I2237W*	-	-	-	-	-	D6936G*	-	-	R349S P883R G39V* E475Q P883R Q893H F1147F	-	-	-	
D-ca	I2181T I2921H*	A2382V I3338F	I3338F	-	G4178A G4178D*	A6260P* I4382I	H4382K	-	-	-	-	T72	G3038	
D-B4	I2181T I2921H*	I2921H*	I3338F	-	G4178?	T4382I	H4382K	-	-	V683Q	G38V* E475D P883R Q883H* Q893H F1147F	V341	T72	-
D-D2	A3092G I2181T I2237W	G3724A I2181T I2237W	C8438F	I3338F	G4178?	-	H4382K	-	-	D6936A* T331T G393T E475D P883R Q893H F1147F C1241W*	V301	T72	-	

**ORF** - open reading frame; **nsp** - nonstructural protein; structural proteins: **S** - spike, **E** - envelope, **M** - membrane, **N** - nucleocapsid. ? - unspecified amino acid.

Green color highlights substitutions common for all virus variants; blue and orange color highlights substitutions common only for ca variants; orange color highlights substitutions characteristic only for a given variant. An "\*" indicates unique substitutions in SARS-CoV-2 variants in relation to viruses deposited in GenBank.

The invention will be described by the following non-limiting examples.

### Example 1

#### Method for the generation of delta E or E/M viruses

##### Stable cells

5        Stable E cells: HEK293T E cells (human embryonic kidney cell line stably expressing CoV-2 E) and Vero E/TMPRSS2 cells (African green monkey kidney cell line stably expressing CoV-2 E and human TMPRSS2) were generated as follows: a cDNA fragment encoding the codon-optimized CoV-2 E gene (Addgene) (SEQ ID NO:13; Figure 15) was cloned into the murine  
10      leukemia virus (MLV)-based retroviral vector pMXs-IRES-puromycin (pMXs-IP) (Cell Biolabs). To generate the retrovirus, Plat-GP cells (Cell Biolabs) were co-transfected with pMXs-IP vector encoding CoV-2 E along with an expression vector for VSV G by using Lipofectamine 2000 (Invitrogen). Two days later, the culture supernatants containing the retroviruses were collected and used to  
15      transduce HEK293T cells and Vero E6 TMPRSS2 cells (JCRB Cell Bank [1819]). Stable cells were selected with 2 µg/ml and 7 µg/ml puromycin (InvivoGen) for HEK293T E cells and Vero E/SS2 cells, respectively.

Stable E/M cells: HEK293T E/M cells (HEK293T cell line stably expressing CoV-2 E and M) and Vero E/M/TMPRSS2 cells (Vero cell line stably expressing CoV-2 E and M and human TMPRSS2) were generated in a similar manner as stable E cells: briefly, pMXs-IP vector encoding the codon-optimized CoV-2 M gene (Addgene) (SEQ ID NO:14; Figure 15) was used to generate the retrovirus. Then, a mixture of retroviruses encoding CoV-2 E and M was used to transduce HEK293T cells and Vero E6 TMPRSS2 cells. Stable cells  
25      were selected with 2 µg/ml and 7 µg/ml puromycin (InvivoGen) for HEK293T E/M cells and Vero E/M/TMPRSS2 cells, respectively.

HEK293T stable cells were maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS in the presence of 2 µg/ml of puromycin. Vero stable cells were maintained in DMEM containing  
30      10% FBS in the presence of 7 µg/ml puromycin and 1000 µg/ml G418 (InvivoGen). All cells were incubated at 37 °C and 5% CO<sub>2</sub>.

##### CPER fragment preparation

Six fragments (F1-6; Figure 13A) for the CPER reaction were amplified from the full-length cDNA of CoV-2 (Wuhan-Hu-1 isolate) cloned into the

pBeloBAC11 vector by using high-fidelity PrimeSTAR GXL DNA polymerase (TaKaRa Bio) and the corresponding primer pairs with overlapping sequences at the 5' end (see Table below), which enables sequence-specific assembly.

Fragment	Primers	Sequences (5' to 3')
1	F1_forward	TCCCAGGTAACAAACCAACCAACTTCG (SEQ ID NO:18)
	F1_reverse	CTTGC GTGTGGAGGTTAATGTTG TCACTG (SEQ ID NO:19)
2	F2_forward	CATTAACCTCCACACGCAAGTTGGACATG (SEQ ID NO:20)
	F2_reverse	GTCTGTCCTGGTTGAATGCGAACAAACTTATAC (SEQ ID NO:21)
3	F3_forward	CGCATTCAACCAGGACAGACTTTTCAGTG (SEQ ID NO:22)
	F3_reverse	GCCACACATGACCATTCACTCAATACTTGAG (SEQ ID NO:23)
4	F4_forward	GTGAAATGGTCATGTGTGGCGGTTCACTATATG (SEQ ID NO:24)
	F4_reverse	CCTGGTGCAACTCCTTATCAGAACAG (SEQ ID NO:25)
5	F5_forward	GATAAAGGAGTTGCACCAGGTACAGCTGTTAAG (SEQ ID NO:26)
	F5_reverse	GTCGTCGTCGGTTCATCAAATTGGTTCC (SEQ ID NO:27)
6	F6_forward	TATGATGAACCGACGACGACTACTAGCG (SEQ ID NO:28)
	F6_reverse	GTCATTCTCCTAAGAAGCTATTAAAATCACATGGGG (SEQ ID NO:29)
Linker	Linker_forward	CCATGTGATTTAATAGCTTCTAGGAGAATG (SEQ ID NO:30)
	Linker_reverse	CAAGAGATCGAAAGTTGGTTGGTTGTTACCTGGG (SEQ ID NO:31)

5

Fragment F6 for the ΔE virus, which lacks its entire ORF region, or for the ΔE/M virus, which lacks both the entire ORF regions including the intergenic region between the ORFs, was cloned into the pCAGGS vector.

The linker fragment (Figure 13B) used to connect fragments F1 and F6 contains a polyA tail (30 adenines) and the hepatitis delta virus ribozyme (HDVr) for generating the authentic 3' end of the viral RNA, a simian virus 40 (SV40) polyA signal for efficient termination of transcription, and a spacer sequence followed by a cytomegalovirus (CMV) promoter for viral RNA transcription.

Each PCR product was purified with a QIAquick Gel Extraction Kit (Qiagen) after separation by agarose gel electrophoresis, and then used for the CPER reaction.

CPER reaction

To generate an infectious cDNA clone, six CoV-2 fragments and a linker fragment were mixed at 0.1 pmol each in a 50- $\mu$ l reaction volume and used for the following PCR reaction with PrimeSTAR GXL DNA polymerase (TaKaRa Bio): initial denaturation at 98 °C for 1 min; 15 cycles of denaturation at 98 °C for 10 s, annealing at 55 °C for 20 s, and extension at 68 °C for 15 min; and a final extension at 68 °C for 15 min.

#### CPER transfection and virus rescue

The CPER product (30  $\mu$ l of a 50- $\mu$ l reaction volume) was directly transfected into HEK293T stable cells (E or E/M cells) seeded in a 6-well plate (8.0  $\times$  10<sup>5</sup> cells/well) by using TransIT-LT1 transfection reagent (Mirus Bio). The next day, the culture supernatant was replaced with fresh culture medium containing 5% FBS. On the fourth day after transfection, the supernatant was collected and 1 ml of supernatant was added to a T-25 flask of confluent Vero stable cells (E/TMPRSS2 or E/M/TMPRSS2 cells).

Supernatants containing viruses were harvested when cytopathic effect (CPE) appeared (4–7 days post-infection). To obtain high-titer virus stocks, the supernatant was passaged in fresh Vero stable cells if needed.

#### Results

Many vaccines against COVID-19 are either against the spike protein based on mRNA or virus vector platforms or inactivated whole-virus vaccines. A SARS-CoV-2 attenuated vaccine virus based on the original Wuhan genome but lacking the envelope (E) open reading frame was prepared (Figure 1A). This vaccine virus replicates efficiently and forms plaques on Vero cells that stably express the E protein (Figure 1B).

To demonstrate initial safety of this vaccine virus (CoV-2 ΔE), human (h)ACE2 transgenic mice were used, which are highly susceptible to infection and serve as a lethal animal model for SARS-CoV-2 infection. Infection with 10,000 plaque-forming units (pfu) of wild-type SARS-CoV-2 (Wuhan isolate generated by reverse genetics) of hACE2 mice resulted in significant body weight loss, and mice succumbed to infection by Day 7 (Figure 2A and 2B). In contrast, hACE2 mice infected with the same dose of CoV-2 ΔE, had the same body weight and survival profiles as mock-infected animals (Figure 2A and 2B).

To determine the protective efficacy of CoV-2 ΔE, Syrian hamsters were vaccinated with 100,000 pfu of CoV-2 ΔE by intranasal inoculation. Two weeks

after vaccination, the hamsters had antibody titers against the SARS-CoV-2 spike receptor-binding domain antigen ranging from 1:320 to 1:1280. At 4-weeks after vaccination, the hamsters were challenged with 1,000 pfu of an early SARS-CoV-2 isolate. Three days after challenge, three of the four vaccinated  
5 hamsters had no detectable infectious virus in their lung tissue, and the fourth hamster had a viral load in its lung tissue of approximately  $10^4$  pfu/gram (Figure 3). In contrast, the control hamsters had high virus titers, close to  $10^8$  pfu/gram in their lung tissue (Figure 3). Vaccine efficacy in the nasal turbinate (NT) tissues was less pronounced, but there was a significant reduction in viral load in  
10 the vaccinated compared to control hamsters (Figure 3). The data demonstrate the near-complete protection of hamsters from infectious virus in the lungs after a single vaccination with CoV-2 ΔE.

### Example 2

Most of the current socially implemented vaccines against SARS-CoV-2  
15 are aimed at inducing antibodies to inhibit the function of spike proteins on viral particles. Socially implemented vaccines include mRNA vaccines, viral vector vaccines, and recombinant protein vaccines. These vaccines induce spike protein-specific antibodies in the blood through intramuscular administration. However, the induction of immunity in the nasal mucosa is not sufficient.  
20 Therefore, a "semi-live virus" was developed as a new modality vaccine that can induce immunity in the nasal mucosa through intranasal inoculation. The "semi-viruses" are viruses that express viral proteins to invade cells and induce immunity for infection defense, but do not produce new infectious progeny particles by lacking viral proteins essential for multiplication (Figure 4).

25 Therefore, "half-living viruses" have the following features and advantages. As with attenuated live viruses (e.g., FluMist; a vaccine using cold-acclimated attenuated live virus of influenza), it is possible to induce not only liquid immunity but also cellular immunity, and since "semi-live viruses" do not have proliferative capacity, they have a low risk of virulence reversion and are  
30 safer compared to attenuated live viruses. Intranasal administration is expected to induce local mucosal immunity. Because it is not a viral vector vaccine, it can be administered multiple times. Unlike mRNA, viral vector, or recombinant protein vaccines that target only spike proteins, these vaccines are expected to induce immune responses against structural proteins other than spike proteins.

Since innate immunity can be activated by the establishment of a single infection, there is no need to use immunostimulants (adjuvants).

Does ΔE SARS-CoV-2 function as a "semi-viral" vaccine

ΔE SARS-CoV-2 (ΔE virus) was generated as a "half-live SARS-CoV2" candidate by deleting the region encoding the envelope (E) protein from SARS-CoV-2 (Figure 5). Vero cells expressing E protein were established to propagate the ΔE virus, and the ΔE virus was generated in E-Vero cells. When transgenic mice expressing human ACE2 (hACE2 mice) were inoculated with wild-type SARS-CoV-2, the mice showed severe weight loss and all individuals died, while mice inoculated intranasally with ΔE virus showed no weight loss and all individuals survived (Figure 6A). This clearly indicates that the ΔE virus is highly attenuated in virulence. Next, ΔE virus was administered intranasally to hamsters, and four weeks later, an attack test by wild-type SARS-CoV-2 was conducted. The results showed that the amount of virus in the respiratory tract of the group intranasally administered ΔE virus was significantly lower than that of the control group (Figure 6B). This indicates that the ΔE virus has a protective effect against infection. However, since the ΔE virus was found to be able to multiply even in cells that did not express the E protein (Figure 7), it was determined that the ΔE virus was not a "half-live virus".

ΔEM SARS-CoV-2, a "half-live virus"

ΔEM SARS-CoV-2 (hereafter referred to as ΔEM virus) was generated from ΔE virus by further deleting the region encoding the matrix (M) protein (Figure 5). ΔEM virus can grow in newly established Vero cells expressing E and M protein (EM-Vero cells), but not in wild-type cells. Thus, the ΔEM virus is a semi-living virus. Since the spike protein that contributes greatly to infection defense is the same as that of the ΔEM virus, it is expected to have the same level of infection defense ability as the ΔE virus. Therefore, a "half-live virus" based on this ΔEM SARS-CoV-2 is developed as a vaccine.

Materials and Methods

Using mice and hamsters, it is tested whether the ΔEM virus induces humoral and cellular immunity, and whether animals immunized with the ΔEM virus are protected against infection when infected with wild-type SARS-CoV-2.

The efficiency of ΔEM virus multiplication has a significant impact on facilities and production costs during vaccine production. Therefore, expression

cells are established in which ΔEM viruses multiply efficiently. hACE2 expression is predicted to improve virus multiplication, so cell clones expressing hACE2, E and M proteins, are established and screened based on ΔEM SARS-CoV-2 multiplication efficiency. Based on the screening results, cell clones with 5 high ΔEM SARS-CoV-2 proliferation efficiency are established.

Toxicity (safety) and pharmacology studies of the ΔEM virus are conducted, including whether cellular and/or humoral immunity (antibody production) is/are induced.

#### Experimental

10 Current mRNA, inactivated, and recombinant protein vaccines are insufficient to induce immunity in the upper respiratory tract mucosa. However, the ΔEM SARS-CoV-2 semi-live vaccine is expected to induce high mucosal immunity in the upper respiratory tract because it invades upper respiratory tract mucosal cells and expresses viral proteins. In addition, since this vaccine is 15 produced using reverse genetics, the S protein gene can be easily replaced, making it possible to respond to epidemics of mutant strains with different antigenic properties. Therefore, the efficacy of the semi-viral SARS-CoV-2 in humans supports that a "semi-viral" vaccine is a new modality, which will greatly contribute to the development of vaccines against infectious diseases 20 other than SARS-CoV-2.

A strain of SARS-CoV-2 is selected, e.g., the BA.2 strain of SARS-CoV-2 omicron mutant. The expression plasmid of the ΔEM virus is produced by utilizing the artificial chromosome (BAC) system of *E. coli* of the Wuhan strain. E and M protein expression plasmids are generated for the ΔEM virus. 293T 25 cells are transfected with the E and M protein expression plasmids and the ΔEM virus expression BAC to generate the ΔEM virus. In addition to producing ΔEM viruses in Wuhan strains, a platform is established to allow easy replacement of the S protein gene in order to respond quickly when a new epidemic strain with different antigenicity arises (Figure 8).

30 To ensure high vaccine production efficiency, cells in which the ΔEM virus can efficiently multiply are established. ΔEM virus multiplication occurs in cells expressing the E and M proteins of SARS-CoV-2. Expression of hACE2, the human receptor of SARS-CoV-2, in cells increases the efficiency of virus entry into cells and improves virus multiplication. On the other hand, the balance

of expression levels of hACE2, E protein and M protein is thought to affect the efficiency of ΔEM virus multiplication. Therefore, using gene transfer technology, we will establish a Vero cell line that constantly expresses hACE2, E protein, and M protein, and from this cell line, a cell clone with an increase in 5 ΔEM virus is selected (Figure 9).

The ΔEM viruses are inoculated into hamsters and hACE2 mice, which are highly susceptible to SARS-CoV-2, and the presence of infectious virus in respiratory tract of the mice and the weight changes are measured to determine if the ΔEM virus is pathogenic (Figure 9A). Wild-type cells are infected with 10 viruses obtained by repeated passages of ΔEM virus in hACE2, E and M protein-expressing cells to confirm that nonproliferative properties are maintained. Furthermore, the virus obtained by passaging is inoculated into hACE2 mice to confirm that it is non-pathogenic (Figure 9B).

To test whether ΔEM virus induces liquid and cellular immunity, hACE2 15 mice and hamsters are inoculated once or twice with ΔEM virus and it is tested whether SARS-CoV-2 specific antibodies and cellular immunity are induced. Furthermore, animals inoculated with the ΔEM virus are infected with the Wuhan strain and various mutant strains, and weight changes, survival rates, and 20 virus levels in the respiratory tract, are measured and compared to the control (PBS-inoculated) group to verify the protective effect of the ΔEM virus against infection (Figure 10A).

Many people have a certain level of immunity against SARS-CoV-2, either by vaccine or natural infection. To verify the efficacy of ΔEM virus as a booster vaccine, hACE2 mice or hamsters that had already been inoculated with 25 mRNA vaccine are inoculated with ΔEM virus and it is tested whether the booster effect was observed (e.g., whether humoral and cellular immunity to SARS-CoV-2 was induced more strongly than immediately before inoculation with ΔEM virus). The booster effect of the ΔEM virus is also tested by infecting the hamsters with the Wuhan strain and various mutant strains, then measuring 30 weight change, survival rate, and virus levels in the respiratory tract in those hamsters and comparing that data to the control (PBS inoculated) group (Figure 10B).

The ΔEM virus induces high mucosal immunity in the upper respiratory tract and suppresses viral replication. It is tested whether ΔEM virus has a protective effect against transmission.

ΔEM virus-inoculated animals are inoculated with the Wuhan strain or 5 various mutant strains, followed by cohabitation of uninfected animals. After several days of cohabitation, the amount of virus in the respiratory tract of uninfected animals is measured to verify whether transmission to uninfected animals is inhibited (Figure 10C). Naive animals are inoculated with the Wuhan strain or various mutants, and then cohabitation with ΔEM virus inoculated 10 animals is begun. After several days of cohabitation, the amount of virus in the respiratory tract of the ΔEM virus animals is measured to verify whether transmission and viral replication to the ΔEM virus-inoculated animals is suppressed (Figure 10D).

Creation of ΔEM virus and establishment of a cell bank to be used for 15 propagation

The hACE2/E/M-expressing Vero cell clones are used to prepare a cell bank in accordance with Good Manufacturing Practice (GMP) standards. The master and working cell bank is stored and managed in a vapor phase liquid nitrogen storage container.

20 Creation of ΔEM virus bank

Cells, e.g., from a portion of the working cell bank, are transfected with ΔEM virus expression plasmids to generate ΔEM viruses. Full-length sequencing of the ΔEM viruses may be conducted. At least 1 ml tubes of master virus banks of ΔEM virus with a titer of at least  $1 \times 10^6$  pfu/mL are prepared. The 25 master virus bank is stored and maintained in a freezer at -70°C or lower. The working virus bank is stored and maintained in a freezer at -70°C or below. Characteristic tests such as sterility test, mycoplasma negativity test, and stray virus negativity test may be conducted.

Non-clinical drug production

30 For nonclinical drugs, the working cell bank is inoculated with ΔEM virus from the virus bank and the virus is grown under established culture conditions. The resulting virus culture medium is concentrated by ultrafiltration after removing cellular residues by filtration. Non-clinical drugs with a titer of  $1 \times 10^6$  pfu/mL or higher are produced by cryopreservation after adding

appropriate additives thereto.

Pharmacodynamic studies (hamsters and monkeys: non-GLP)

Hamsters are inoculated intranasally with one or two doses of nonclinical drug, and blood is drawn 3-4 weeks later to determine neutralizing antibody titer.

- 5 After blood collection, intranasal inoculation with Wuhan strain ( $10^5$  pfu: calculated with EM-expressing cells) as a challenge infection is conducted and weight changes are observed for 2 weeks after infection. Three and six days after infection, hamsters are dissected to quantify virus levels in the lungs and nasal turbinates, and pathological analysis of the lungs, nasal turbinates, and major organs of the body are performed.

10 Monkeys are inoculated intranasally with one or two doses of a nonclinical drug, and blood is drawn 3-4 weeks later to see if neutralizing antibodies and cellular immunity have been induced. After blood sampling, intranasal and intratracheal inoculation with Wuhan strain ( $10^7$  pfu: calculated with EM-expressing cells) as a challenge infection are performed. Weight changes and general symptoms after infection are observed. Nasal, pharyngeal, and rectal swabs are collected at 1, 3, 5, and 7 days post-infection to quantify viral load and to obtain CT images to confirm the presence of pneumonia.

15 Monkeys are dissected at 3 and 7 days post-infection, and virus levels in the lungs, trachea, and pharynx are quantified. Pathological analysis is performed on the dissected monkeys' lungs, nasopharynx, and major organs throughout the body. Body temperature is measured as needed with an implantable telemetry transmitter implanted in each individual.

Biodistribution test (monkey: non-GLP)

20 Monkeys are inoculated intranasally with a nonclinical drug and dissected 6 days after inoculation to confirm the presence of ΔEM virus in the brain, olfactory bulb, nasal concha, pharynx, trachea, lungs, heart, liver, kidney, spleen, stomach, small and large intestine, genital organs, bladder, urine, blood, stool, oral and rectal swabs by RT-qPCR.

25 Repeated dose toxicity study (hamster: GLP)

Safety pharmacology and local irritation are evaluated in parallel. As for the safety pharmacology core battery (organ systems of vital importance), functions on the central nervous, cardiovascular and respiratory systems are evaluated. Local irritation is evaluated in the analysis of the nasal turbinates

during histopathological examination of repeated dose toxicity studies.

Specifically, hamsters are inoculated intranasally with the nonclinical drug two or three times and general symptoms are observed before and after inoculation, and hematology, blood biochemistry, and histopathology in the brain, olfactory

5 bulb, nasal concha, trachea, lung, heart, liver, kidney, spleen, stomach, small intestine, colon, and genital tract are determined. Heart rate and body temperature are measured as needed with an implanted telemetry transmitter.

Respiratory function is measured by prestimograph after each vaccination.

#### Subjects

10 Based on the doses studied in the non-clinical studies (safety, drug efficacy, etc.), the subjects are divided into the three groups: high dose, low dose, and placebo (Figure 11). Although it is desirable that eligible subjects should be those who have no history of novel coronavirus infection and vaccination against novel coronavirus, it is assumed that recruiting such

15 participants is difficult. Therefore, the safety and efficacy (immunogenicity) in boosted vaccinated healthy adult males, e.g., 20 to 64 years old, is studied.

As an example, the following exclusion criteria may be assigned

- (1) Persons with COVID-19 or in close contact with a person with COVID-19 at the time of vaccination with the clinical study drug
- 20 (2) Patients with a history of anti-SARS-CoV-2 monoclonal antibody administration within 3 months prior to clinical study drug inoculation
- (iii) Those with underlying diseases such as serious cardiovascular disease, kidney disease, liver disease, blood disease, developmental disorder, respiratory disease, and diabetes mellitus.
- 25 (4) Those who have been diagnosed with immunodeficiency in the past or those who have a close relative with congenital immunodeficiency.
- (5) Persons who are judged by the investigator to be unsuitable for participation in this clinical trial as a result of the screening test

Recruitment of clinical trial participants is done through contract  
30 research organizations (CROs).

#### Safety and tolerability

- Percentage of subjects reporting at least one adverse event of any kind
- Percentage of subjects reporting at least one relevant adverse event by degree (grade)

- Summary statistics of safety-related laboratory tests (subject background investigation, physical examination findings, clinical examination, vital signs, serious adverse events, specific adverse events, unspecified adverse events, and COVID-19 disease status)

5        Adverse events are defined as all unwanted or unintended illnesses or signs of illness (including abnormal laboratory values) that occur in subjects inoculated with an investigational drug, regardless of whether they are causally related to the investigational drug. Adverse events will be collected from the time of study drug immunization to the 4-week post-immunization examination, 10 but will continue to be collected for serious adverse events and COVID-19 until follow-up is completed. Adverse reactions are defined as reactions that have at least a reasonable possibility of being related to the clinical trial drug and for which an association cannot be ruled out.

Immunogenicity (neutralizing antibody titer)

15       Neutralizing antibody titer against SARS-CoV-2 strain and SARS-CoV-2 mutant strain after immunization with the study drug in each group and by subject is measured.

20       T-cell IFN- $\gamma$  production in response specifically to SARS-CoV-2 antigen after immunization with the study drug in each group and by subject is determined.

          S-protein, N-protein, and RBD protein antibody titers (ELISA method) of SARS-CoV-2 are determined.

Conclusion

25       A live-attenuated vaccine virus based on a whole virus generates an immune response not only against the spike protein (the target of most SARS-CoV-2 vaccines), but also against other SARS-CoV-2 proteins, thereby eliciting a more robust and durable protection profile. Moreover, a live-attenuated SARS-CoV-2 vaccine platform that can be readily updated with new SARS-CoV-2 sequences as needed, offers a robust and durable platform solution for 30 Covid immunizations.

**Example 3**

The M protein along with the E protein are essential for proper SARS-CoV-2 virus-like particle formation. To allow for CoV-2 ΔE+ΔM virus growth, Vero cells that stably express both the E and M proteins were generated.

To examine the vaccine efficacy of CoV-2 ΔE+ΔM, hamsters were first  
5 vaccinated once by intranasal inoculation of  $5 \times 10^4$  plaque-forming units (pfu) of CoV-2 ΔE+ΔM. Six weeks after the last vaccinations, hamsters were challenged intranasally with the SARS-CoV-2 Delta variant ( $10^3$  pfu) or a more recent and antigenically advanced variant, Omicron XBB ( $10^5$  pfu). On day three after infection, titers of the challenge viruses were determined by plaque assay in  
10 the lung and nasal turbinates tissues.

One vaccination resulted in a 10-fold reduction in virus titers in the lung tissue of hamsters challenged with the Delta variant with undetectable virus in the lung tissue of one of the vaccinated hamsters (Figure 16A). There was no reduction in the virus titers in the nasal turbinates tissue of the same animals  
15 compared to the control group (Figure 16B). In vaccinated hamsters challenged with Omicron XBB, there was a 10- to 100-fold reduction of the challenge virus in the lung and nasal turbinates tissues (Figure 16C and 16D).

Another group of hamsters received two doses of the vaccine virus, CoV-  
2 ΔE+ΔM, with four weeks between vaccinations. Six weeks after the last  
20 vaccination, hamsters were infected with either challenge virus. On day three after infection in vaccinated hamsters challenged with the Delta variant, the prime + boost (P+B) vaccine regimen provided better protection compared to the single vaccination in the lung tissue with no infectious Delta virus detected (Figure 17A). Virus replication of the Delta variant was also reduced by over  
25 1,000-fold in the nasal turbinates tissue of the vaccinated hamsters (Figure 17B).

Similar protective efficacy results were observed in CoV-2 ΔE+ΔM vaccinated hamsters after challenge with Omicron XBB. No infectious challenge virus was detected in half of vaccinated hamsters, while there was a 1,000 to 10,000-fold reduction in Omicron XBB virus titers in the remaining two animals  
30 (Figure 17C). In the nasal turbinates tissue, vaccination with CoV-2 ΔE+ΔM reduced challenge virus titers by 1,000-fold when compared to non-vaccinated control hamsters (Figure 17D).

#### References

Adachi et al., *J. Med. Virol.*, 94:1789 (2022).

- Bai et al., PloS One, \_\_\_\_:\_\_\_\_ (2008).  
(<https://doi.org/10.1371/journal.pone.0002685>)
- Efficacy of vaccination and previous infection with the Omicron BA1 variant in Syrian hamsters. Cell Rep. 39(3):110688, 2022
- 5 Eisfeld et al., Host & Microbe, 22:817 (2017).
- Halfmann et al., N. Engl. J. Med., 383:592 (2020).
- Halfmann et al., Nature, 603:687 (2022).
- Halfmann et al., Cell Rep., 38:110394 (2022).
- Halfmann et al., Cell Rep., 39:110688 (2022).
- 10 Hatakeyama et al., Virology, \_\_\_\_:\_\_\_\_ (2008).  
(<https://doi.org/10.1016/j.virol.2008.07.012>)
- He et al., PNAS, 118:e2025866118(2021)  
(<https://doi.org/10.1073/pnas.2025866118>)
- Ho et al., Biochem. Biophys. Res. Commun., \_\_\_\_:\_\_\_\_ (2004)
- 15 (https://doi.org/10.1016/j.bbrc.2004.04.111)
- Hou et al., Science, 370:1464 (2020).
- Hsieh et al., J. Virol., \_\_\_\_:\_\_\_\_ (2005).  
(<https://doi.org/10.1128/JVI.79.22.13848-13855.2005>)
- Huang et al., J. Virol., \_\_\_\_:\_\_\_\_ (2004).  
20 (https://doi.org/10.1128/JVI.78.22.12557-12565.2004)
- Ikeuchi et al., BMC Infect. Dis., 22:167 (2022).
- Imai et al., Proc. Natl. Acad. Sci. USA, 118:e2106535118 (2021).
- Imai et al., Proc. Natl. Acad. Sci. USA, 117:16587 (2020).
- Imai et al., Cell Host & Microbe, 22:615 (2017).
- 25 Imai et al., Nature Microbiol., 5:27 (2020).
- Ishizaka et al., Viruses, 13:2101 (2021).
- Ishizaka et al., Microbiol. Spectr., 9:e0070821 (2021).
- Jiang et al., Cell Biosci., \_\_\_\_:\_\_\_\_ (2021). (<https://doi.org/10.1186/s13578-021-00644-y>)
- 30 Ju et al., A novel cell culture system modeling the SARS-CoV-2 life cycle. PloS Pathog., \_\_\_\_:\_\_\_\_ (2021).  
(<https://doi.org/10.1371/journal.ppat.1009439>)
- Koga et al., Hepatol Res., 52:227 (\_\_\_\_).
- Kubota-Aizawa et al., S, Matsubara Y, Kanemoto H, Mimuro H, Uchida

- K, Chambers J, Tsuboi M, Ohno K, Fukushima K, Kato N, Yotsuyanagi H, Tsujimoto H. Transmission of Helicobacter pylori between a human and two dogs: A case report. Kanemoto Y, Kanemoto H, Mimuro H, Uchida K, Chambers J, Tsuboi M, Ohno K, Fukushima K, Kato N, Yotsuyanagi H, Tsujimoto H.
- 5 Transmission of Helicobacter pylori between a human and two dogs: A case report.
- Kuroda et al., PLoS Pathog., 16:e1008900 (2020).
- Kuroda et al., Nature Commun., 11:2953 (2020).
- Liu et al., bioRxiv, Feb. 15 (2022).
- 10 (<https://www.biorxiv.org/content/10.1101/2022.02.14.480460v1>)
- Lu et al., Nat. Comm., 12:502 (2021).
- Marzi et al., Science, 348:439 (2015).
- Yoshimura et al., \_\_\_\_\_, \_\_\_\_:\_\_\_\_ (2020).
- Minote et al., Pharma. Ther., 5:310 (2022).
- 15 Mizutani et al., Microbiol. Spectr., 7:e0168921 (2022).
- Nagamura-Inoue & Nagamura, Umbilical Cord Blood and Cord Tissue Bank as a Source for Allogeneic Use, 1-24(2020)
- Noda et al., Nature Commun., 9:54 (2018).
- Nojima et al., BMJ Open, 8:e021129 (2018).
- 20 Okushin et al., BMC Infect Dis., 21:399 (2021).
- Ricardo-Lax et al., Science, \_\_\_\_:\_\_\_\_ (2021)
- (<https://doi.org/10.1126/science.abj8430>)
- Saito et al., Nature, 602:300 (2022).
- Silvas et al., J. Virol., 95:e0040221 (2021).
- 25 (<https://doi.org/10.1128%2FJVI.00402-21>)
- Siu et al., J. Virol., \_\_\_\_:\_\_\_\_ (2008). (<https://doi.org/10.1128/JVI.01052-08>)
- Sonoda, Translat. Regulat. Sci., 3:120 (2021).
- Takada et al., Nature Microbiol., 4:1268 (2019).
- Takahashi et al., J. Infect. Chemother., 28:1 (2022).
- 30 Takashita et al., N. Engl. J. Med., 386:1475 (2022).
- Takashita et al., N. Engl. J. Med., 386:995 (2022).
- Taniguchi et al., J. Cancer, 149:646 (2021).
- Thi et al., Nature, 582:561 (2020) (<https://doi.org/10.1038/s41586-020-2294-9>)

- Ishigaki et al., J. Japanese Soc. Labor. Med., 69:1 (2021).
- Ueki et al., Proc. Natl. Acad. Sci. USA, 115:E6622 (2018).
- Ueki et al., Nature Protoc., 15:1041 (2020).
- Wang et al., Virologica Sinica, 36:890 (2021).
- 5      (<https://doi.org/10.1007/s12250-021-00369-9>)
- Xu et al., Front. Bioeng. Biotech., 8:862 (2020).
- Yasuhara et al., Nature Microbiol., 4:1024 (2019).
- Yoshimi et al., iScience, 25:103830 (2022).
- Yuki et al., Microbe, 2:e429 (2021).
- 10     Zhang et al., Antiviral Res., 185:104974 (2021).
- (<https://doi.org/10.1016/j.antiviral.2020.104974>)
- Zhang et al., J. Gen. Virol., 102:001583 (2021)
- (<https://doi.org/10.1099%2Fjgv.0.001583>).
- Zhang et al., Nat. Comm., 13:4399 (2022).

15

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification, this invention has been described in relation to certain preferred embodiments thereof, and many details have been 20 set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details herein may be varied considerably without departing from the basic principles of the invention.

**WHAT IS CLAIMED IS:**

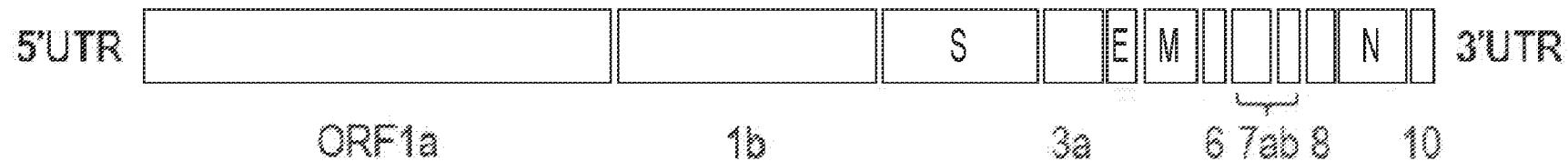
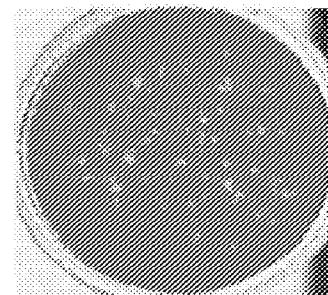
- 5      1. An isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope (E) protein.
- 10     2. The isolated nucleic acid of claim 1 wherein the modification is a deletion of at least part of the open reading frame encoding the E protein.
- 15     3. The isolated nucleic acid of claim 1 or 2 further comprising one or more genetic modifications that inhibit or prevent expression of coronavirus M protein.
- 20     4. The isolated nucleic acid of claim 1, 2 or 3 which comprises DNA.
- 25     5. The isolated nucleic acid of claim 1, 2 or 3 which comprises RNA.
- 30     6. An isolated cell comprising the isolated nucleic acid of any one of claims 1 to 5.
- 35     7. The isolated cell of claim 6 which is a mammalian cell.
- 40     8. The isolated cell of claim 7 which is a non-human primate cell.
- 45     9. The isolated cell of any one of claims 6 to 8 that stably expresses coronavirus E protein.
- 50     10. The isolated cell of any one of claims 6 to 8 that stably expresses hACE2 and optionally M protein.
- 55     11. An isolated cell that stably expresses coronavirus E protein.

12. The isolated cell of claim 11 which is a mammalian cell.
13. The isolated cell of claim 12 which is a non-human primate cell.
- 5 14. The isolated cell of any one of claims 11 to 13 that stably expresses hACE2.
- 10 15. The isolated cell of any one of claims 11 to 14 further comprising one or more genetic modifications that inhibit or prevent expression of coronavirus M protein.
16. The isolated cell of claim 15 that stably expresses coronavirus M protein.
17. A composition comprising an attenuated recombinant coronavirus comprising a coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope E protein, which virus comprises E protein embedded in the envelope.
- 20 18. The composition of claim 17 wherein the coronavirus genome further comprises a genetic modification that inhibits or prevents expression of coronavirus M protein, which virus comprises M protein embedded in the envelope.
19. A system comprising:
  - 25 i) an isolated cell that stably expresses coronavirus E protein, or coronavirus E protein and coronavirus M protein; and
  - ii) an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein, or an isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein and M protein.
- 30 20. The system of claim 19 wherein the isolated cell stably expresses coronavirus E protein and the isolated nucleic acid comprises a

recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein.

21. The system of claim 19 wherein the isolated cell stably expresses coronavirus E protein and M protein and the isolated nucleic acid comprises a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus E protein and M protein.
22. A recombinant coronavirus, wherein the genome of the recombinant coronavirus contains a deletion of one or more nucleotides in a polynucleotide sequence for a viral protein corresponding to coronavirus E protein which deletion is effective to prevent expression of a functional viral protein corresponding to coronavirus E protein upon infection of a cell with the recombinant coronavirus, wherein the genome encodes one or more coronavirus glycoproteins, and wherein the coronavirus comprises E protein.
23. The recombinant coronavirus of claim 22 wherein the cell that is infected does not express functional E protein.
24. The recombinant coronavirus of claim 22 or 23 further comprising a deletion of one or more nucleotides in a polynucleotide sequence having an open reading frame for a viral protein corresponding to coronavirus M protein.
25. The recombinant coronavirus of claim 22 which comprises M protein.
26. The recombinant coronavirus of claim 24 or 25 wherein at least 90% of sequences corresponding to E or M protein coding sequences, or any combination, in the viral genome of the virus, are deleted.
27. The recombinant coronavirus of any one of claims 22 to 26 wherein the recombinant genome further comprises a nucleotide sequence encoding a prophylactic or therapeutic heterologous gene product.
28. The recombinant coronavirus of any one of claims 22 to 27 wherein the genome encodes a heterologous S protein.

29. The recombinant coronavirus of any one of claims 22 to 28 which is cold adapted.
30. A vaccine having an effective amount of the recombinant coronavirus of any one of claims 22 to 29.
- 5 31. The vaccine of claim 30 which is formulated for intranasal delivery.
32. The vaccine of claim 30 which is formulated for subcutaneous delivery.
33. The vaccine of claim 30, 31 or 32 which includes a pharmaceutically acceptable carrier.
34. A method to immunize a mammal, comprising administering to the  
10 mammal an effective amount of the vaccine of any one of claims 30 to 33.
35. The method of claim 34 wherein the mammal is a human.
36. The method of claim 34 or 35 which comprises administering two doses.
37. The method of claim 34 or 35 which comprises administering one dose.
- 15 38. The method of claim 37 further comprising administering a different coronavirus vaccine.
39. The method of claim 38 wherein the different coronavirus vaccine is a mRNA vaccine.
40. The method of claim 38 wherein the different coronavirus vaccine is  
20 administered before the vaccine of any one of claims 30 to 33.
41. The method of any one of claims 34 to 40 wherein the mammal is immunocompromised.

**Wuhan wild-type****ΔE ORF virus***Fig. 1A**Fig. 1B*

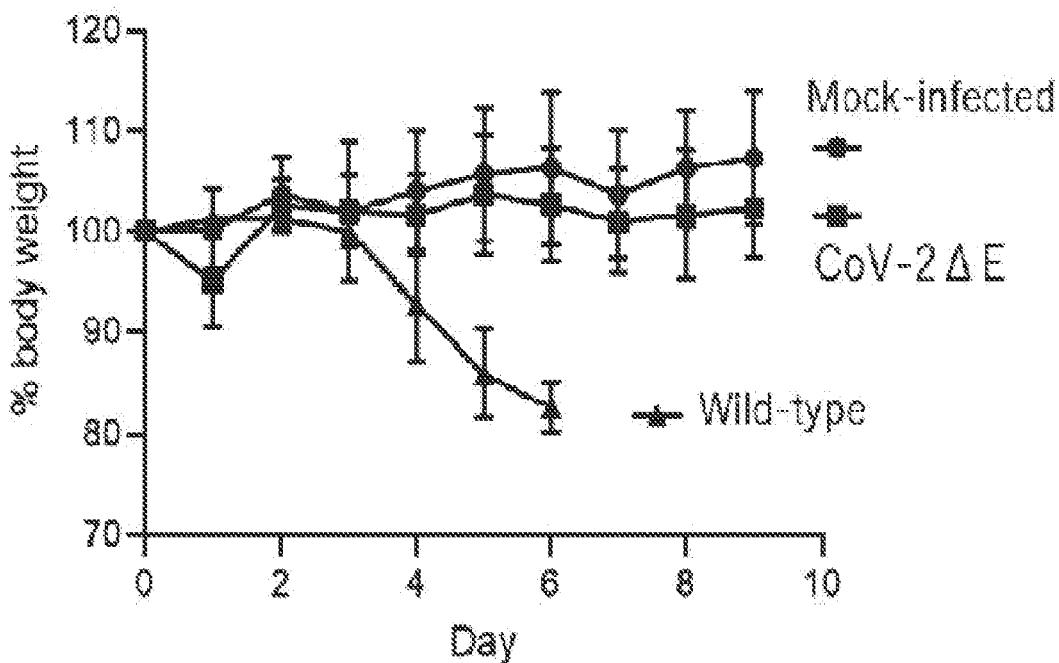


Fig. 2A

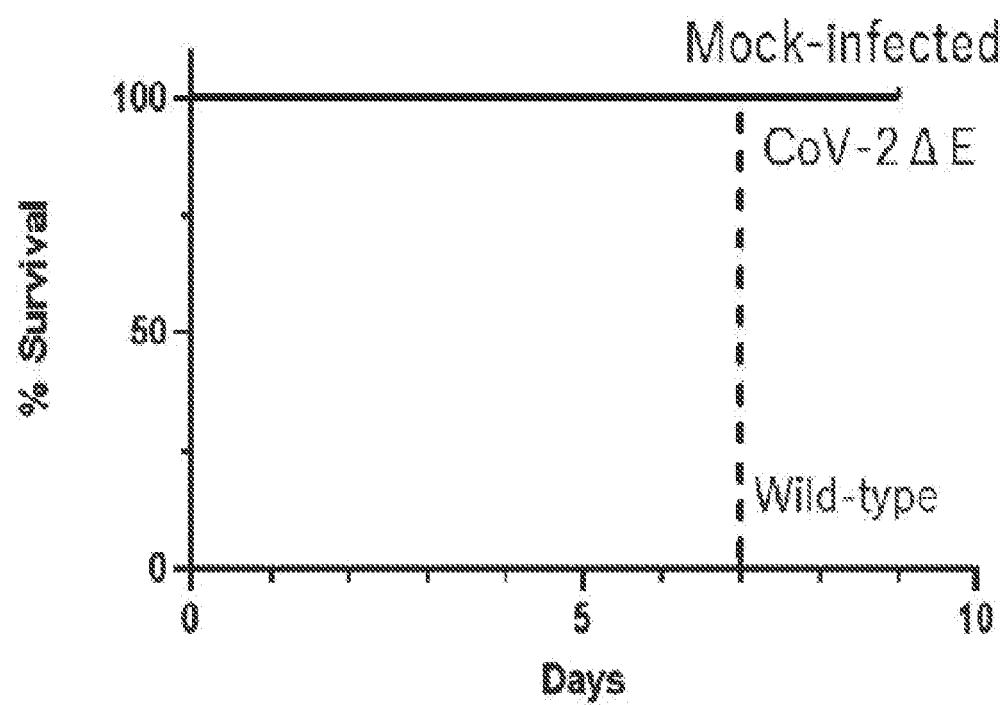


Fig. 2B

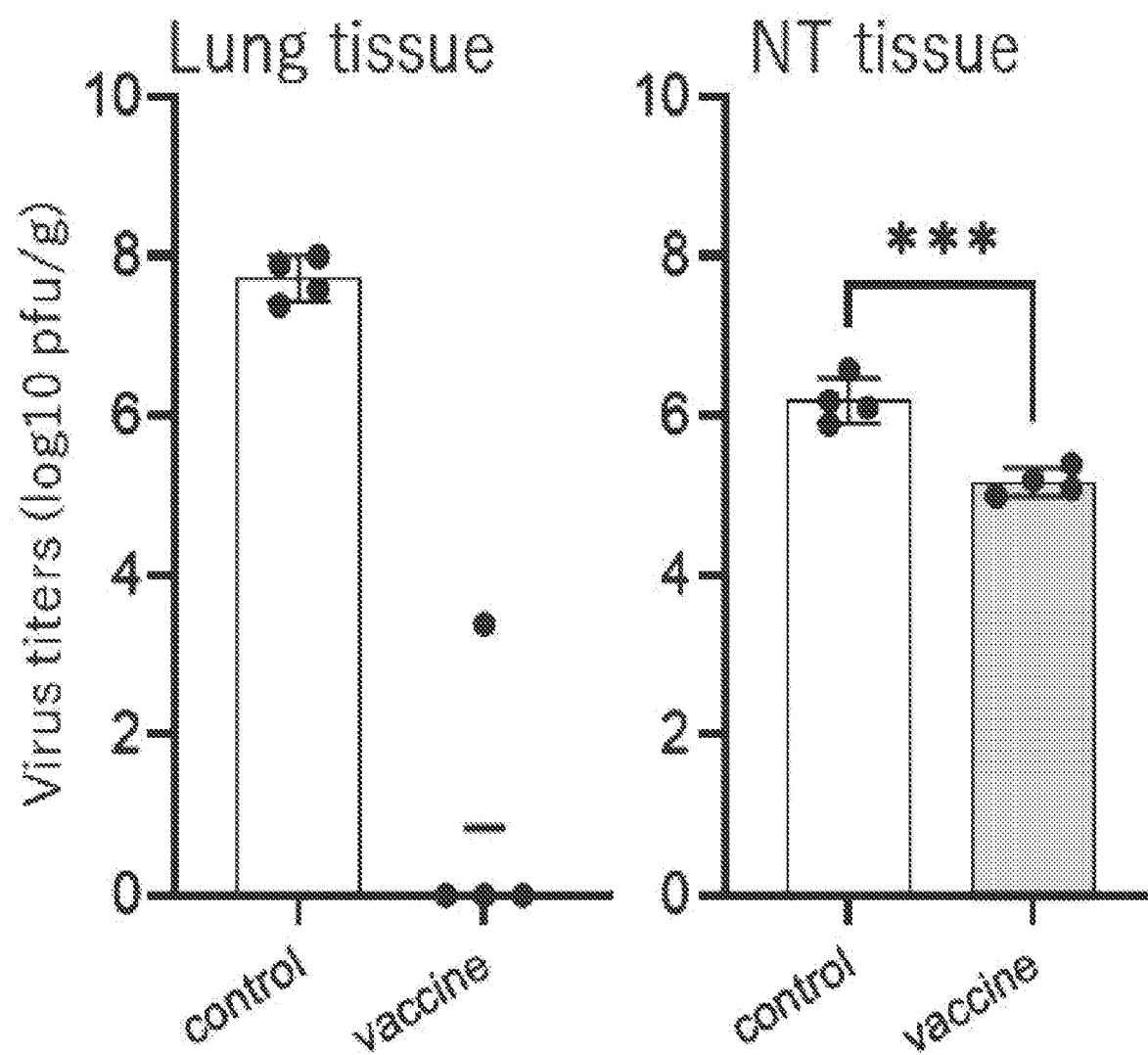


Fig. 3

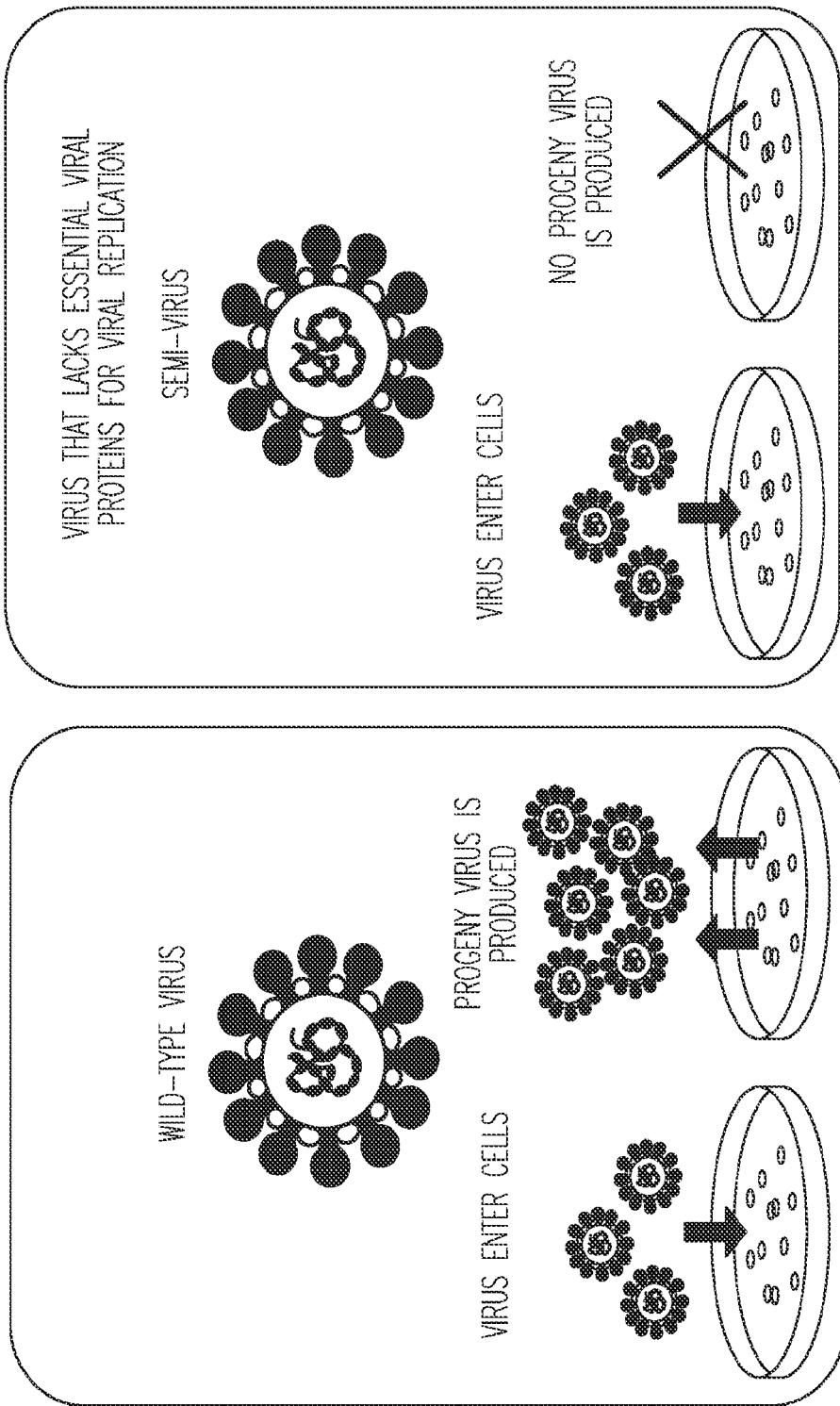


Fig. 4

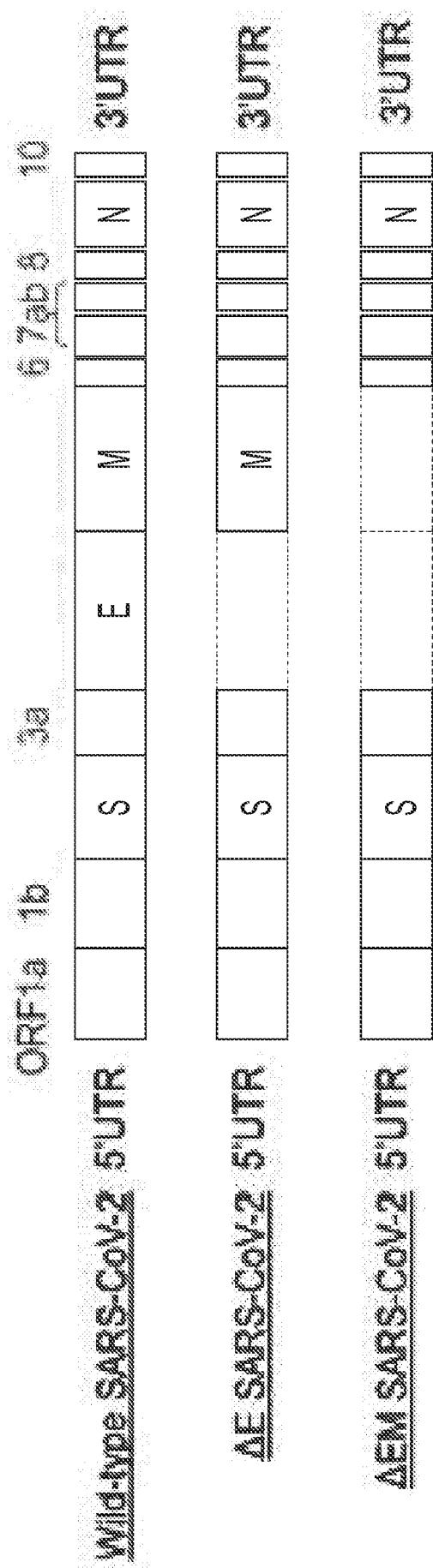


Fig. 5

## Evaluation of pathogenicity in ACE2 mouse model

### Intranasal inoculation with ΔE virus

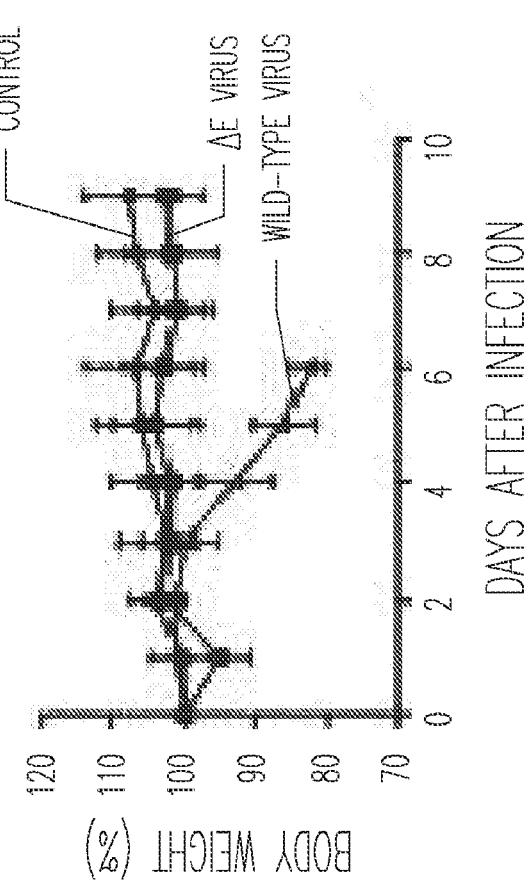
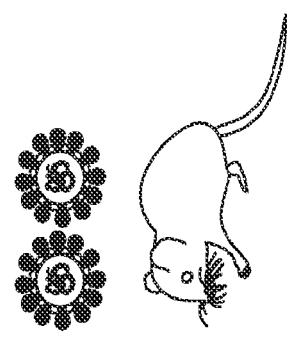


Fig. 6

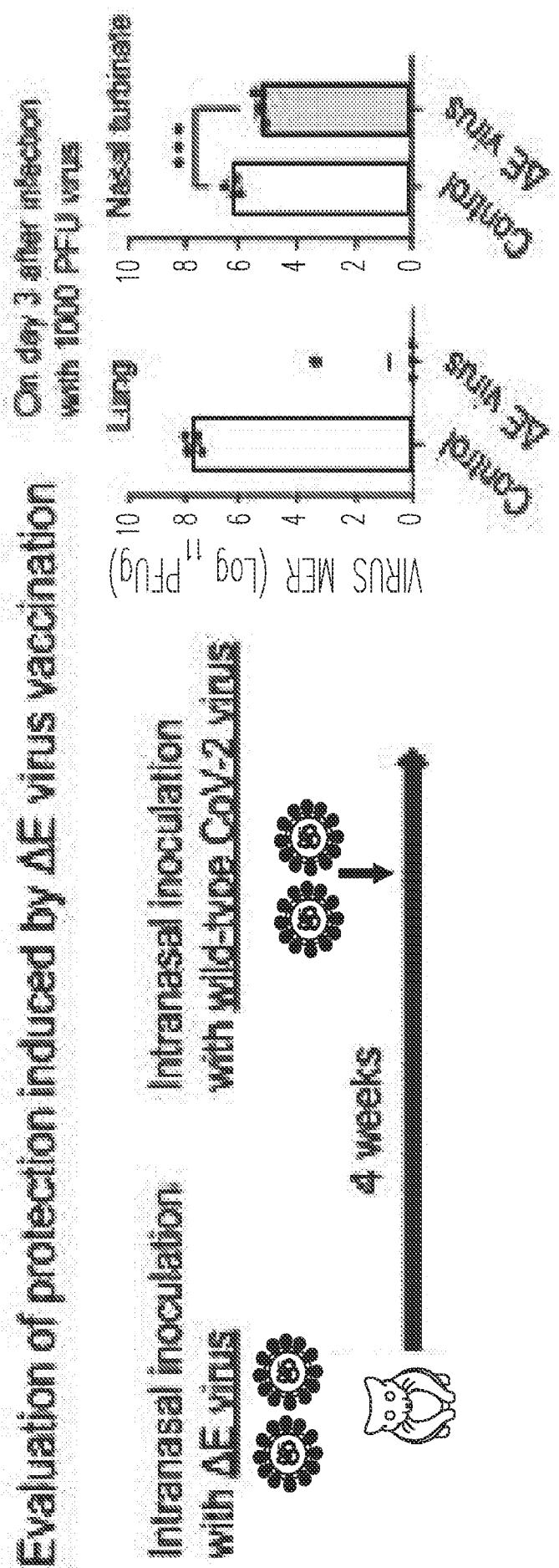


Fig. 6 (continued)

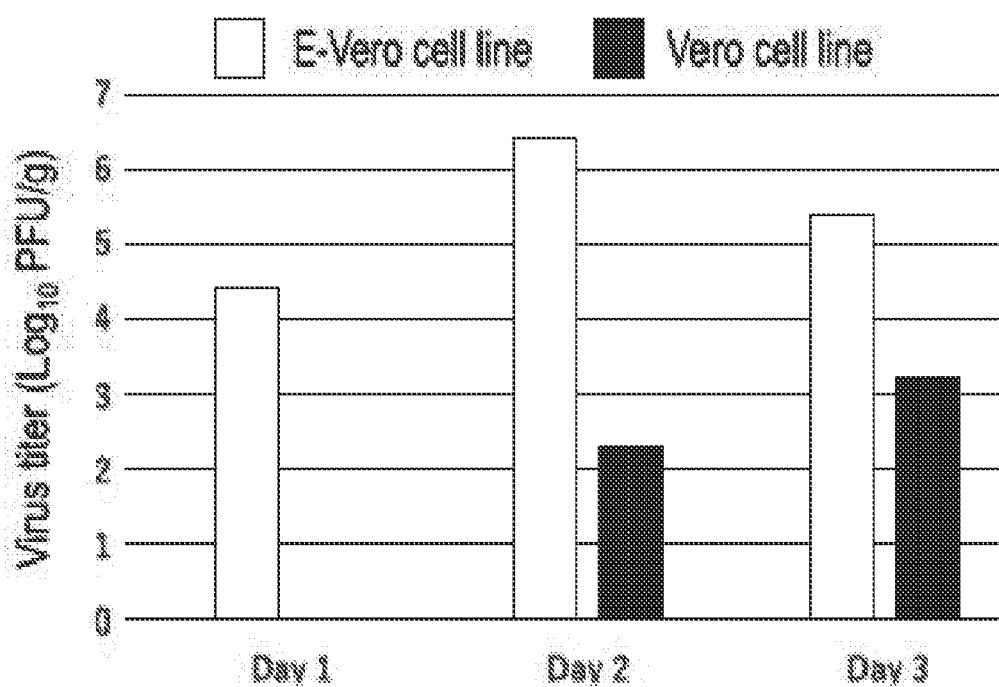
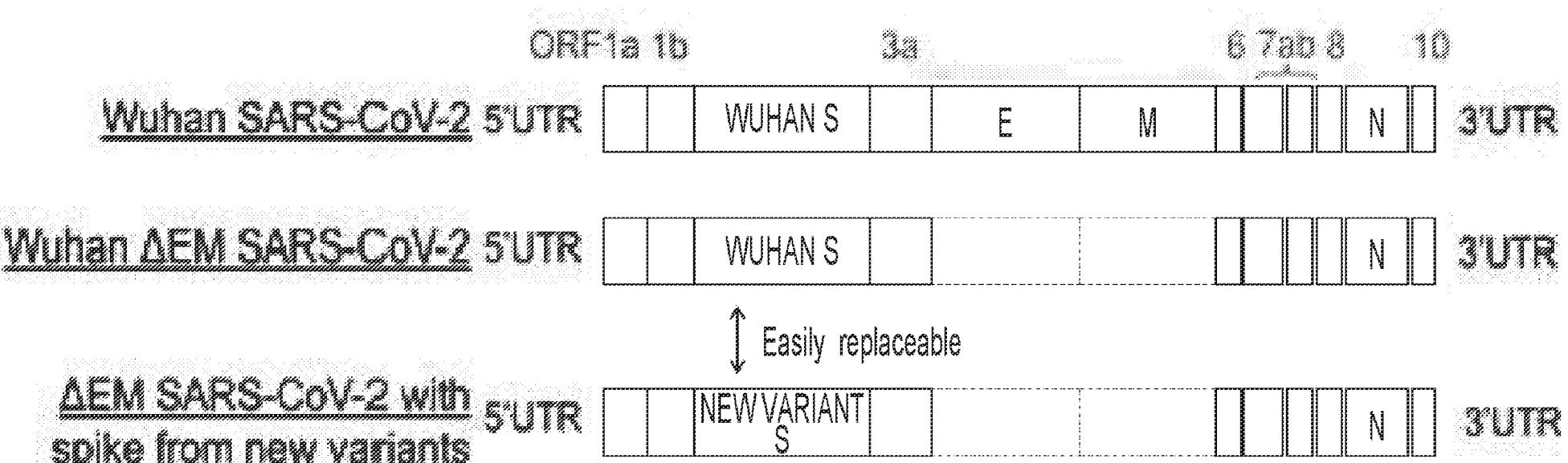


Fig. 7



*Fig. 8*

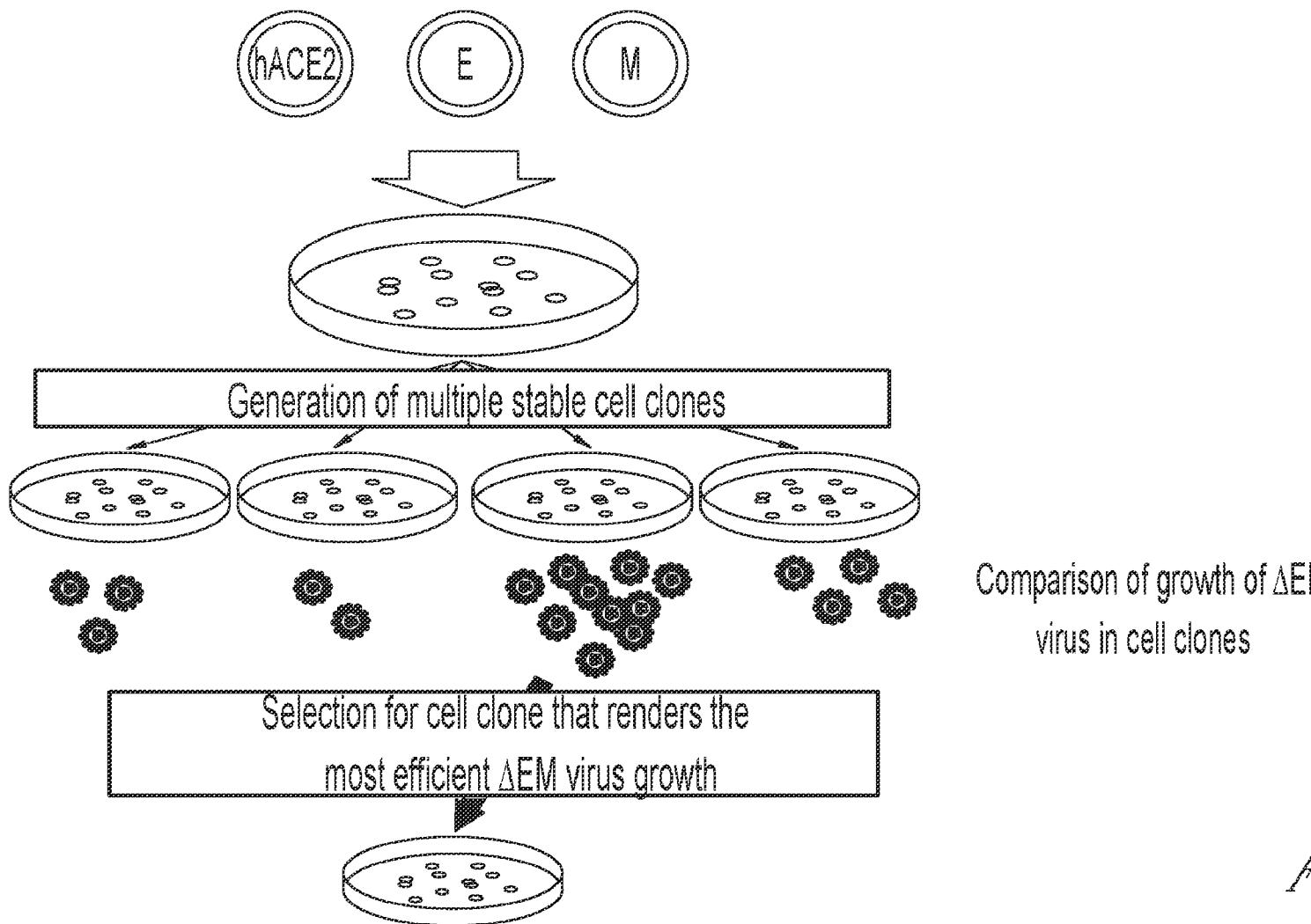


Fig. 94

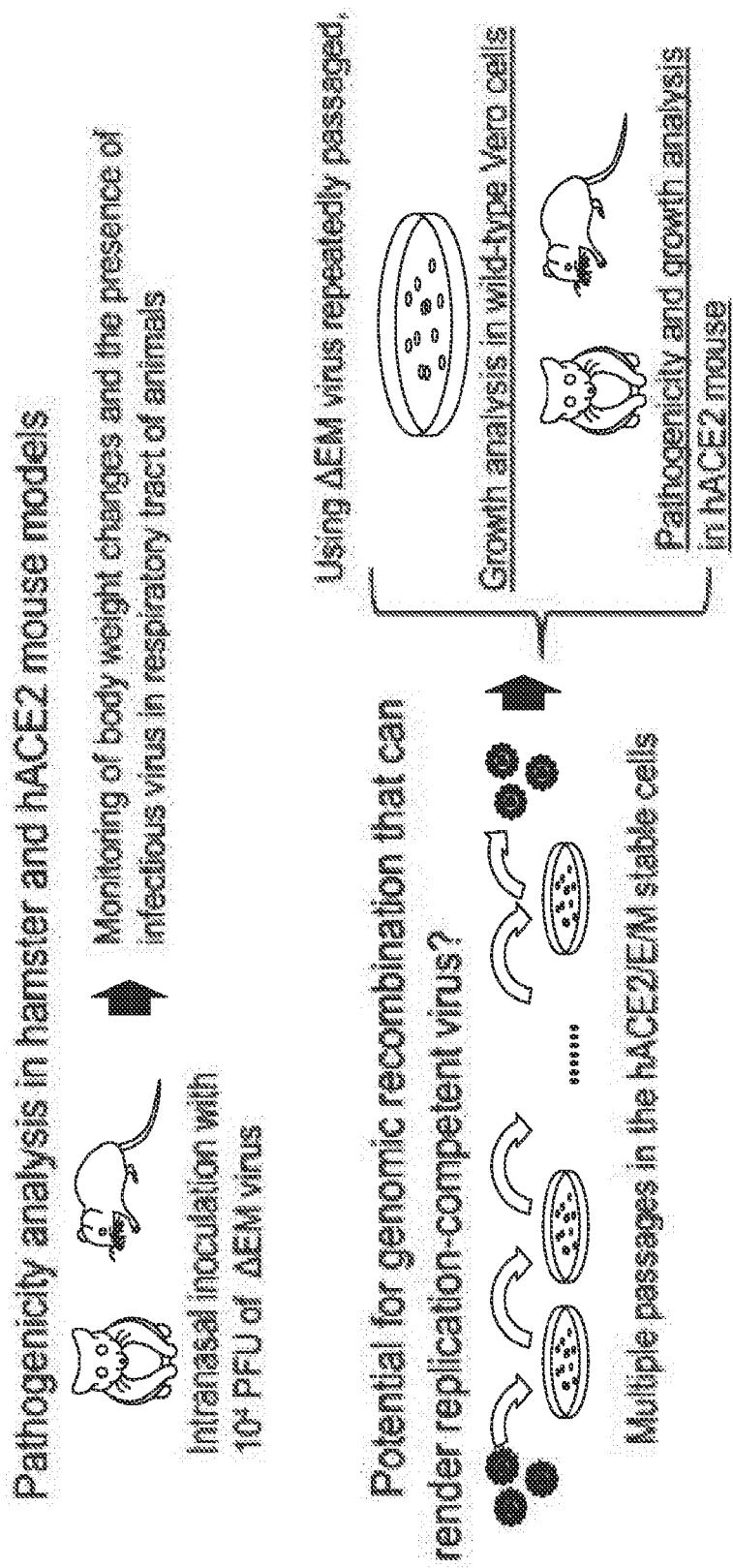


Fig. 9B

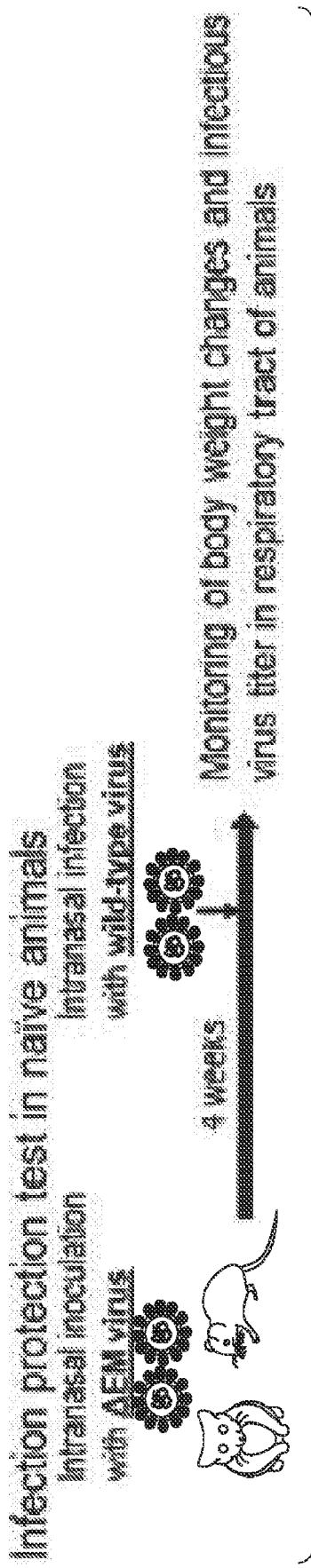


Fig. 10A

## Infection protection test in mRNA-vaccinated animals

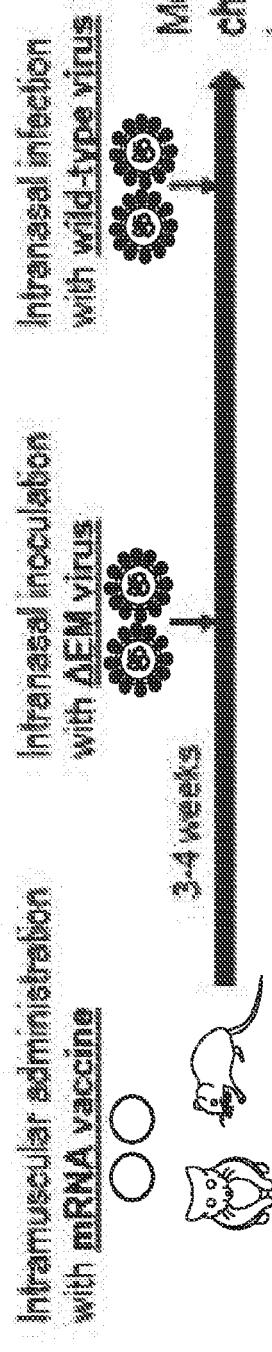
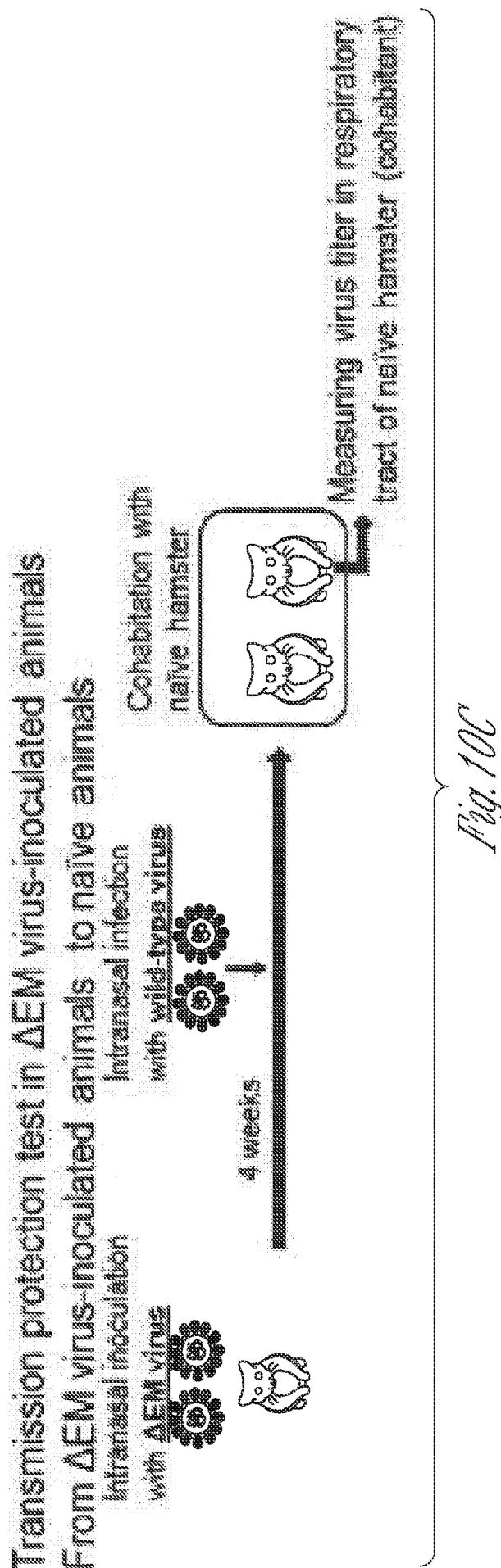


Fig. 10B



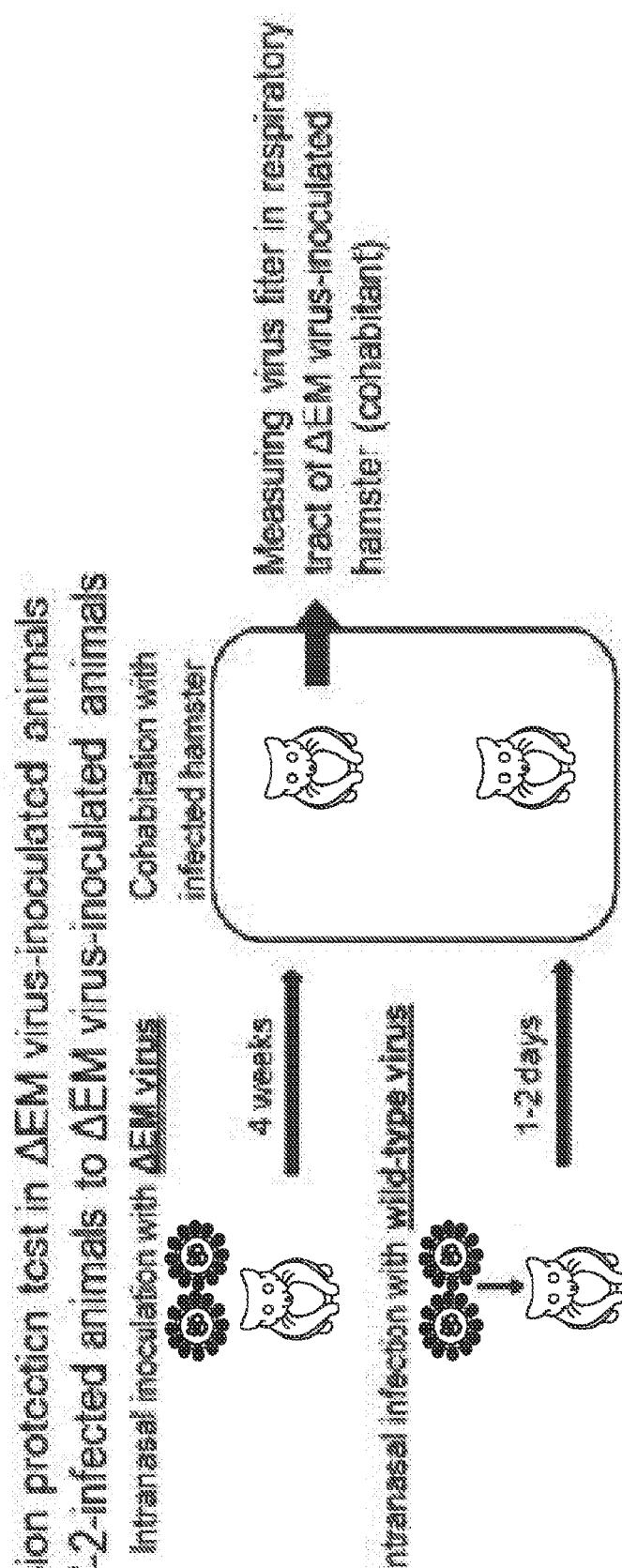


Fig. 100

	EXAMINEE HUMAN SUBJECT	SEX	AGE	VACCINATION RECORD/CoV-2 INFECTION RECORD
GROUP 1 ΔEM VIRUS HIGH DOSE	20	M	20-64	MORE THAN 2 TIMES WITH mRNA VACCINE (PFIZER OR TAKEDA/MODERNA)
GROUP 2 ΔEM VIRUS LOW DOSE	20	M	20-64	MORE THAN 2 TIMES WITH mRNA VACCINE (PFIZER OR TAKEDA/MODERNA)
GROUP 3 PLACEBO	20	M	20-64	MORE THAN 2 TIMES WITH mRNA VACCINE (PFIZER OR TAKEDA/MODERNA)

*Fig. 11*

# Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/SARS-CoV-2, B.1.617.2 lineage, Delta variant/2021, complete genome

GenBank: OK091006.1

## FASTA Graphics

Go to:

LOCUS OK091006 29836 bp RNA linear VRL 13-SEP-  
2021

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate  
SARS-CoV-2/human/JPN/SARS-CoV-2, B.1.617.2 lineage, Delta  
variant/2021, complete genome.

ACCESSION OK091006

VERSION OK091006.1

KEYWORDS .

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM Severe acute respiratory syndrome coronavirus 2

*Fig. 12A*

Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29836)

AUTHORS Rajib,M.S.A., Hossain,M.B., Satou,Y. and Ikeda,T.

TITLE Direct Submission

JOURNAL Submitted (13-SEP-2021) Joint Research Center for Human Retrovirus Infection, Kumamoto University, 2-2-1 Honjo, Chuo-ku, Kumamoto 8600811, Japan

COMMENT ##Assembly-Data-START##  
 Assembly Method :: Burrows-Wheeler Alignment (BWA-MEM) tool  
 v. 0.7.1  
 Sequencing Technology :: Illumina  
 ##Assembly-Data-END##

FEATURES Location/Qualifiers

source 1..29836  
 /organism="Severe acute respiratory syndrome coronavirus 2"  
 /mol\_type="genomic RNA"

*Fig. 12A continued*

/isolate="SARS-CoV-2/human/JPN/SARS-CoV-2, B.1.617.2  
lineage, Delta variant/2021"  
/isolation\_source="nasal swab"  
/host="Homo sapiens"  
/db\_xref="taxon:2697049"  
/country="Japan"  
/collection\_date="2021-08-24"  
gene 239..21528  
/gene="ORF1ab"  
CDS join(239..13441,13441..21528)  
/gene="ORF1ab"  
/ribosomal\_slippage  
/codon\_start=1  
/product="ORF1ab polyprotein"  
/protein\_id="UAL04645.1"  
  
/translation="MESLVPGFNEKTHVQLSLPVVLQVRDVLVRGFGDSVEEVLSearq  
HLKDGTGCLVEVEKGVLQPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

*Fig. 12A continued*

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKLDGFMRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG  
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG  
LKTIILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGLNDNL  
LEILQKEKVNIIVGDFKLNEEIAIIILASFSASTS AFVETVKGLDYKAFKQIVESCGN  
FKVTKGAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV  
NKFLALCADSIIIGGAALKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

*Fig. 124 continued*

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGMLLEIKDTEK  
YCALAPNMMVTNNNTFLKGGAFTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATTYLFDESGEF  
KLASHMYCSEFYPPDEDEEEGDCEEEEFEPSHQYEYGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQTVGQODGSEDNQTTIQTIVEVQPQLEMELTPVVQTIENVNSFSG  
YLKLTDNVYIKNADIVEEAKKVKP TVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGS CVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLA VFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTITLEETKFELTENLLYIDINGN  
LHPDSATLVSDIDITELKKDAPYIVGDVVQEGVLTAVVIPTKKS GGTTEMLAKALRKV

*Fig. 12A continued*

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM  
LAHAEETRKILMPVCETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATSVSSPDAVTAYNGYLTS  
KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPPTFHLDGEVIT  
FDNLKTLSSLREVRTIKVFTTVDNINLHTQVVDMMSMTYGQQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPNDTLRVEAFEYYHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK  
WADNNCYLATALLTLOQTTELKFNPALQDAYYRARAGEAANFCALILAYCNKTVGEIG  
DVRETMSYLFQHANLDSCKRVLNAVCKTCGQQOTTLLKGVEAVMYMGTLSYEQFKKGVQ  
IPCTCGKQATKYLVQQESPVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTIKPVTYKLDGVVCTEIDPKLDNYY

*Fig. 12A continued*

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT  
FFPDLNGDVVAIDYKHYTPSEKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWXX  
XXXXXXXXXVLKSEDAQGMNDNLACEDLKLVSEEVVENPTIQKDVLECNVKTTEVVGD  
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS  
VPWDITIANYAKPFLNKVVSTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI  
KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINTIIWFLLSVCLGSLIYSTA  
ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSISCSVCLSGLDSDLTYP SLET  
IQITISSEFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL  
MWLIINLVQMAPISAMVRMYIFFASEFYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE  
CTTIVNGVRRSEFYVYANGGKGFCKLHNWNVCNCDTECAGSTFISDEVARDLSIQEKFRP

*Fig. 12A continued*

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI  
NVIVFDGSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN  
TFSSTFNVPMEKLKTLVATAEAEELAKNVS LDNVLSTFISAARQGFVDSDVETKDVVEC  
LKLSHQSDIEVTGDSNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI  
WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKI VNNWL  
KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEITIGYKAIDGGVTRDIASTDTCF  
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHF  
RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPLPYCYDTNVLEG  
YESLRPDTRYVLMGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTGERSEAGVCVSTSG  
RWVLNNNDYYRSLPGVFCGVDAVNLLTNMFTPLIOPIGALDISASIIVAGGIVAI  
VVTCL

*Fig. 12A continued*

AYYFMRFRRAFGEYSHVVAFTNTLLFLMSFTVLCLTPVYSELPGVYSVIYLYLTFLTN  
DVSFLAHIQWMVMETPLVPFWITIAYIICISTKHFWFFSNYLKRRVVEANGVSESTFE  
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTSYREAACCHL  
AKALNDFNSNGSDVLYQPPQISITSAVLQSGFRKMAFP SGKVEGCMVQVTCGTTLNG  
LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNNFLVQAGNVQLRVI GHSMQNCVL  
KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC  
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNEYGPFVDRQTAQAAGTDTTITVN  
VLAWLYYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV  
LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFOQSAVKRTIKGTHHW  
LLLTIITSLLVLVQSTQWSLFFFLYENAFLEFAMCIIAMSASFAMMEVKHKHAFLCLFL

*Fig. 12A continued*

LP SLAAVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR  
TVYDDGARRVWTLMNVLTLYKVVYGNALDQAI SMWALIISVTSNYSGVVTTVMFLAR  
GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL  
VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTS VVL  
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEK MV SLLSVLLSMQGAVDINKL  
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKSLNVAK  
SEFDRDAAMQRKLEKMA DQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN  
NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTFTYASALWEIQQVVDA  
DSKIVQLSEISM DNSPNLAWPLIVTALRANS AVKLQNNELSPVALRQMSCAAGTTQTA  
CTDDNALAYYNTTKGGREVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

*Fig. 12A continued*

KGPKVKYLYFIKGLNNLNRCMVLGSLAATVRLQAGNATEVPANSSTVLSFCAFADVAAK  
AYKDYLASGGQPITNCVKMLCTHTGTGQAIVTPEANMDQESFGGASCCLYCRCHIDH  
PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA  
DAQSFLNRVCVGSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD  
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDEFKFRIDGDMVPHISRQR  
LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDYFNKKDWYDFVENPDILRVYA  
NLGERVRQALLKTVQFC DAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV  
DSYYSLMPILTILTRALTAESHVDIDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDO  
TYHPNCVNCLDDRCILHCANFNVLFSTVFPLTSFCPLVRKIFVVDGVPFVVSTGYHERE  
LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLLDKRTTCE SVAALTNNVAFO

*Fig. 12A continued*

TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFAQDGNAIASDYDYYRYNLP TMCDI  
RQLLFVVEVVDKYFDCYDGGCINANQVI VNNLDKSAGFPFNKGKARLYYDSMSYEDQ  
DALFAYTKRNVIPTITQMNLKYAI SAKNRARTVAGVSICSTMNRQFHQKLLKSIAAT  
RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR  
KHTTCCSLSHRFYRLANECAQVLSEMVMCGSSLYVKPGGTSSGDATTAYANSVFNICQ  
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHF SMM  
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS  
QHTMLVKQGDDYVYLPYPDPSSRILGAGCFVDDIVKTDTLMIERFVSLAIDAYPLTKH  
PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV  
LQAVGACVLCNSQTSLRCGACIRRFLCCKCCYDHVISTSHKLVLSVNRYVCNAPGCD

*Fig. 12A continued*

VTDVTQLYLGGMSYYCKSHKLPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW  
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP  
RPPLNRNYVFTGYRVTNSKVQIGEYTFEKGDYGDAVVYRGTTYKLNVDYFVLTSW  
TVMPPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK  
SHFAIGLALYYPSARIIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKE  
KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVNVNARLRAKHVVYIGDPAQ  
LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK  
AHKDKSAAQCFKMFYKGVIHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSONA  
VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD  
RDLYDKLQETSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT

*Fig. 12A continued*

EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNNGYPNMFITREEAIRHVRAWIGFDVEG  
CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP  
LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL  
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGETGNLQSNHDLYCQVHGNA  
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK  
FPVLHDIGNPKAIKCVPOADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL  
FWNCNVDRYPVNSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ  
LPFFYYSDSPCESHGKQVVSIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN  
MMIISAGEFSLWVYKQFDTYNLWNTFTRLQSLENVAENVVNKGHEDGQQGEVRVSIINNT  
VYTKVDGVDVELFENKTLPVNVAFELWAKRNTKPVPEVKILNNILGVDTAAANTVIWDY

*Fig. 12A continued*

KRDAPAHISTIGVCSMTDIAKKPTETICAPLTFFDGRVDGQVDLFRNARNGVLITEG

SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK

PRSQMEIDFLELAMDEFIERYKLEGYAFEHTIVYGDFSHSQLGGLHLLIGLAKRFKESP

FELEDFIPMDSTVKNYFITDAQTGSSKCVCSSVIDLLLDDFVEIIKSQDLSVVSKVVKV

TIDYTEISFMLWCKDGHVETFYPKLQSQAQPGVAMPNLYKMQRMLLEKCDLQNYGD

SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRCVIHFGAGSDKGVAPGTAVLRQWLP

TGTLLVDSLNDVFSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG

FFTYICGFIQQKLALGGSSVAIKITEHSWNADLYKLMGHFAWWTAEVTNVNASSSEAFL

IIGCNYLGKPREQIDGYVMHANYIFWRNTNPQLSSYSLFDMSKFPLKLRGTAVMSLKE

GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat\_peptide 239..778

*Fig. 12A continued*

	/gene="ORF1ab"
	/product="leader protein"
<u>mat_peptide</u>	779..2692
	/gene="ORF1ab"
	/product="nsp2"
<u>mat_peptide</u>	2693..8527
	/gene="ORF1ab"
	/product="nsp3"
<u>mat_peptide</u>	8528..10027
	/gene="ORF1ab"
	/product="nsp4"
<u>mat_peptide</u>	10028..10945
	/gene="ORF1ab"
	/product="3C-like proteinase"
<u>mat_peptide</u>	10946..11815
	/gene="ORF1ab"
	/product="nsp6"
<u>mat_peptide</u>	11816..12064
	/gene="ORF1ab"
	/product="nsp7"
<u>mat_peptide</u>	12065..12658

*Fig. 12A continued*

mat\_peptide /gene="ORF1ab"  
/product="nsp8"  
12659..12997  
mat\_peptide /gene="ORF1ab"  
/product="nsp9"  
12998..13414  
mat\_peptide /gene="ORF1ab"  
/product="nsp10"  
join(13415..13441,13441..16209)  
mat\_peptide /gene="ORF1ab"  
/product="RNA-dependent RNA polymerase"  
16210..18012  
mat\_peptide /gene="ORF1ab"  
/product="helicase"  
18013..19593  
mat\_peptide /gene="ORF1ab"  
/product="3'-to-5' exonuclease"  
19594..20631  
mat\_peptide /gene="ORF1ab"  
/product="endoRNase"

*Fig. 12A continued*

mat\_peptide 20632..21525  
/gene="ORF1ab"  
/product="2'-O-ribose methyltransferase"  
CDS 239..13456  
/gene="ORF1ab"  
/codon\_start=1  
/product="ORF1a polyprotein"  
/protein\_id="UAL04646.1"  
  
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLEARQ  
HLKDGTICGLVEVEKGVLPOLEQPYVFIKRSDARTAPHGCHVMVELVAELEGIQYGRSCE  
TLGVLVPHVGETPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKLDGMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

*Fig. 12A continued*

DFVKATCEF CGTENLTKEGATT CGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG  
LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL  
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTS AFVETVKGLDYKAFKQIVESCGN  
FKVTKGKA KKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIE GTVYEKL  
KPVLDWLEEKFKEGV EFL RDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV  
NKF LALCADSIIIGGA KLKALNLGETFVTHSKGLYRKC VKSREETGLLMP LKAPKEII  
FLEGETLPTEVLTEEVVLKTGDLQPL EQPTSEAVEAPLVGTPVC INGLMLLEIKDTEK  
YCALAPNMMVTNN TETLKGGA PTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLINEK  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLT P LGIDLDEWSMATYYLFDES GEF

*Fig. 12A continued*

KLASHMYCSFYPPDEDEEEGDCEEEEFPSTQYEGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQTVGQQDGSEDNQTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG  
YIKLTDNVYIKNADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAHENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFILEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTLEETKFLTENLLLIDINGN  
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPKKSGGTTEMLAKALRKV  
PTDNYITYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTWSWNLREM  
LAHAEETRKLMPCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLVTMPILGYVTHGLNIEEAARYMRSLKVPATSVSSPDAVTAYNGYLTSSS

*Fig. 12A continued*

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLRGDKSVYYTSNPTTFHLDGEVIT  
FDNLKTISSLREVRTIKVFTTVDNINLHTQVVDMMSMTYQQQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPNDDTLRVEAFYYHTDPSFLGRYMSALNHTKKWKPQVNGLTSIK  
WADNNCYLATALLTLQQIELKENPPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETMSYLFQHANLDSCKRVLNAVCKTCGQQQTTLKGVEAVMYSMGTLSYEQFKKGVQ  
IPCTCGKQATKYLVQQESPFVMMSAPP AQYELKHGTFTCASEYTGNYQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPIDVFYKENS YTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT  
FFPDILNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWXX  
XXXXXXXXXVLKSEDAQGMDNLACEDLKLVSEEVVENPTIQKDVLECNVKTTEVVGD

*Fig. 12A continued*

IILKPANNSLKITEEVGHTDILMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS  
VPWDTIANYAKPFLNKVVSTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTSNSRI  
KASMPPTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINI IIIWFLLLSVCLGSЛИYSTA  
ALGVILMSNLGMP SYCTGYREGYLNSTNVTIATYCTGSISCSVCLSGLDSLDTYPSET  
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL  
MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDG CNSSTCMMCYKRNRATRVE  
CTTIVNGVRRSFYVYANGGKGFCKLHNWNCSVNCDFCAGSTFISDEVARDSLQFKRP  
INPTDOSSYIVDSVTVKNGSIHLYFDKAGQTYERHSLSHFVNLDNLRANNTKGSLPI  
NVIVFDGKSKEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN  
TFSSTFNVPMEKLKTLVATAEAEELAKNVSLDNVLSTEISAARQGFVDSDVETKDVVEC

*Fig. 12A continued*

LKLHQSDIEVTGDSNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI  
WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVTTKIALKGGKIVNNWL.  
KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA  
NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFPLP  
RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPLPYCYDTNVLEGSVA  
YESLRPDTRYVLMGDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTGERSEAGVCVSTSG  
RWVLNNNDYYRSLPGVFCCGVDAVNLLTNMFTPLIQPIGALDISASIIVAGGIVAIWTCL  
AYYFMRFRRRAFGEYSHVVAFTNLFLMSFTVLCLTPVYSELPGVYSVIYLYLIFYLTN  
DVSEFLAHIQWMVMFTPLVPEWITIAYIICISTKHFWFFSNYLKRRVVNGVSESTEE  
EAALCTELINKEMYLKRSVDVLLPLTQYNRYLALYNKYKYFSGAMDTSYREAACCHL

*Fig. 12A continued*

AKALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFP SGKVEGCMVQVTCTTLLNG  
LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNNFLVQAGNVQLRVIGHSMQNCVL  
KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSP SGVYQCAMPNFTIKGSFLNGSC  
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFDVDRQTAQAAGTDTITVN  
VLAWLYYAAVINGDRWFLNRF TTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV  
LDMCASLKELLQNGMNGRTIILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW  
LLLTIITSLLVLVQSTQWSLFFFLYENAFLPFAMGT IAMS AFAMMFVKHKHAFLCLFL  
LPSLAAVAYFNMVYMPASWVMRIMTWILDMDTSLSGFKLKDCVMYASAVVLLILMTAR  
TVYDDGARRVWTLMNVLT L VYKVYYGNALDQAI SMWALI I SVTSNYSGVVTTVMFLAR  
GIVEMCVEYCPTEFFITGNTLQCIMLVYCELGYFCTCYFGLFCLLNRYFRRLTLGVYDYL

*Fig. 12A continued*

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL  
LSVLQQQLRVESSSKLWAQCVQLHNDILLAKDTIEAFEKMVSSLVLLSMQGAVDINKL  
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKSLNVAK  
SEFDRDAAMQRKLEKMDQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN  
NIINNARDGCVPNTIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA  
DSKIVQLSEIISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA  
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP  
KGPKVKYLYFIKGLNNLNRCMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK  
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH  
PNPKGFCDLKGYVQIPTCANDPGFTLKNTVCTVCGMWKGCGSCDQLREPMLQSA

*Fig. 12A continued*

DAQSFLNGFAV"

mat\_peptide 239..778  
/gene="ORF1ab"  
/product="leader protein"

mat\_peptide 779..2692  
/gene="ORF1ab"  
/product="nsp2"

mat\_peptide 2693..8527  
/gene="ORF1ab"  
/product="nsp3"

mat\_peptide 8528..10027  
/gene="ORF1ab"  
/product="nsp4"

mat\_peptide 10028..10945  
/gene="ORF1ab"  
/product="3C-like proteinase"

mat\_peptide 10946..11815  
/gene="ORF1ab"  
/product="nsp6"

mat\_peptide 11816..12064  
/gene="ORF1ab"

*Fig. 12A continued*

## SUBSTITUTE SHEET (RULE 26)

	/product="nsp7"
<u>mat_peptide</u>	12065..12658
	/gene="ORF1ab"
	/product="nsp8"
<u>mat_peptide</u>	12659..12997
	/gene="ORF1ab"
	/product="nsp9"
<u>mat_peptide</u>	12998..13414
	/gene="ORF1ab"
	/product="nsp10"
<u>mat_peptide</u>	13415..13453
	/gene="ORF1ab"
	/product="nsp11"
<u>stem_loop</u>	13449..13476
	/gene="ORF1ab"
	/note="Coronavirus frameshifting stimulation element stem-loop 1"
<u>stem_loop</u>	13461..13515
	/gene="ORF1ab"
	/note="Coronavirus frameshifting stimulation elemer

*Fig. 12A continued*

stem-loop 2"  
gene 21536..25357  
/gene="S"  
CDS 21536..25357  
/gene="S"  
/codon\_start=1  
/product="surface glycoprotein"  
/protein\_id="UAL04647.1"  
  
/translation="MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVYYPDKVER  
SSVLHSTQDLELPFFSNVTWFHATHVSGTNGTKRFDNPVLPFNDGVYFASXEKSNTIR  
GWIEGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPELDVYYHKNNKSWMESXXXVY  
SSANNCTFEYVSQPFILMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ  
GFSALEPLVDLPIGINITRFQTLLALHRSYLTGDSSSGWTAGAAAYVGYLQPRTFI  
LKYNENGTTDAVDCALDPLSETKCTLKSFTVEKGIVQTSNFRVQPTESIVRFPNITN

*Fig. 12A continued*

LCPFGEVFNATRFASVYAWNKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCE  
TNVYADSFVIRGDEVRQIAPGQTGKIAODYKLPDDFTGCVIAWNSNNLDSKVGGYN  
YRYRLFRKSNLKPFERDISTEIQAGSKPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY  
RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG  
RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQAVLYQGVNCTEVPAI  
HADQLTPTWRYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNSRR  
RARSVASQSIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMKTSVDCTM  
YICGDSTECSNLLQYGSFCTQLNRALTGIAVEQDKNTQEVAQVKQIYKTPIKDFG  
GFNFSQLPDPSKP SKRSFIEDLLFNKVTLADAGFIKQYGDCLGDI AARDLICAQKFN  
GLTVLPPILLDEMIAQYTSALLACTITSGWTFCAGAAIQLRFAMQMAYRFNGCIGVTQN

*Fig. 12A continued*

VLYENQKLIANQFNSAIGKIQDSLSSTIASALGKLQNVVNQNAQALNTLVQLSSNFGA  
ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS  
ECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAH  
FPREGVFVSNGTHWFVTQRNFYEPQIITTNTFVSGNCDVVIGIVNNTVYDPLQPELD  
SFKEELDKYFKNHTSPDVLDGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELG  
KYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCSCLKGCCSCGSCCKFDEDDSE  
PVLKGVKLHYT"  
gene 25366..26193  
/gene="ORF3a"  
CDS 25366..26193  
/gene="ORF3a"  
/codon\_start=1  
/product="ORF3a protein"  
/protein\_id="UAL04648.1"

*Fig. 12A continued*

/translation="MDLFMRIIFTIGTVTLKQGEIKDATPLDFVRATATIPIQASLPFG  
WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE  
APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY  
NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ

LSTDITGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVNPVMEPIYDEPTTTSVPL"

gene 26218..26445

/gene="E"

CDS 26218..26445

/gene="E"

/codon\_start=1

/product="envelope protein"

/protein\_id="UAL04649.1"

/translation="MYSFVSEETGTLIVNSVLLFLAFVVFLLVTILAILTALRLCAYCC  
NIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV"

*Fig. 12A continued*

gene 26496..27164  
 /gene="M"  
CDS 26496..27164  
 /gene="M"  
 /codon\_start=1  
 /product="membrane glycoprotein"  
 /protein\_id="UAL04650.1"  
 /translation="MADSNGTITVEELKKLLEQWNLVIGFLFTIWICLLQFAYANRNR  
 FLYIIKLIIFLWLLWPVTLACFVLAAYRINWITGGIAXAMACLVGIMWLSYFIASFRL  
 FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD  
 IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSDNIA  
 LLVQ"  
gene 27175..27360  
 /gene="ORF6"  
CDS 27175..27360  
 /gene="ORF6"  
 /codon\_start=1

*Fig. 12A continued*

/product="ORF6 protein"  
/protein\_id="UAL04651.1"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIKNLSKSL  
TENKYSQLDEEQPMEID"

gene 27367..27732  
/gene="ORF7a"

CDS 27367..27732  
/gene="ORF7a"  
/codon\_start=1  
/product="ORF7a protein"  
/protein\_id="UAL04652.1"

/translation="MKIIILFLALITLATCELYHYQECVRGTTVLLKEPCSSCTYEGNS

PFHPLADNKALTCFSTQFAFACPDGVKHVYQLRARSASPKLFIRQEEVQELYSPIFL  
IVAAIIVFITLCFTLKRKTE"

gene 27729..27860  
/gene="ORF7b"

CDS 27729..27860

*Fig. 12A continued*

/gene="ORF7b"  
/codon\_start=1  
/product="ORF7b"  
/protein\_id="UAL04653.1"

/translation="MIELSLIDEYLCFLAFLFLVLLIMLIIIFWFSSLELQDHNEICHA"

gene 27867..28232  
/gene="ORF8"  
CDS 27867..28232  
/gene="ORF8"  
/codon\_start=1  
/product="ORF8 protein"  
/protein\_id="UAL04654.1"

/translation="MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK  
WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRC  
SFYEDFLEYHDVRVVLVDLI"

gene 28247..29506  
/gene="N"  
CDS 28247..29506

*Fig. 12A continued*

/gene="N"  
/codon\_start=1  
/product="nucleocapsid phosphoprotein"  
/protein\_id="UAL04655.1"  
/translation="MSDNGPQNQRNAPRITFGPSDSTGSNQNGERSGARSKQRRPQC  
LPNNNTASWFTALTQHGKEGLKFPRGQGVPIINTNSSPDDQIGYYRRATRRIRGGDGKMK  
DLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAIVLQ  
LPQGTTLPKGFYAEGSRGGSQASSRSSRSRNSSRNSTPGSSMGTSPARMAGNGCDAA  
LALLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKPRQKRTATKAYNVTQAFGR  
RGPEQTOGNFGDQELIROGTDYKHWPOIAQFAPSASAFFGMSRIGMEVTPSGTWLTYT  
GAIKLDDKDPNFKDQVILLNKHIDAYKTEPPTEPKDKKKKAYETQALPQRQKKQQTV  
TLLPAADLDDFSKQLQQSMSSADSTQA"  
gene 29531..29647  
/gene="ORF10"

Fig. 12A continued

<u>CDS</u>	29531..29647
	/gene="ORF10"
	/codon_start=1
	/product="ORF10 protein"
	/protein_id=" <u>UAL04656.1</u> "
	/translation="MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT"
<u>stem_loop</u>	29582..29617
	/gene="ORF10"
	/note="Coronavirus 3' UTR pseudoknot stem-loop 1"
<u>stem_loop</u>	29602..29630
	/gene="ORF10"
	/note="Coronavirus 3' UTR pseudoknot stem-loop 2"
<u>stem_loop</u>	29701..29741
ORIGIN	/note="Coronavirus 3' stem-loop II-like motif (s2m)"
	1 caaaccaacc aactttcgat ctctttaga tctgttctct aaacgaactt taaaatctgt
	61 gtggctgtca ctcggctgca tgcttagtgc actcacgcag tataattaaat aactaattac
	121 tgtcgttgac aggacacgag taactcgtct atcttctgca ggctgcttac ggtttcgtcc
	181 gttttgcaga cgatcatcag cacatctagg ttttgtccgg gtgtgaccga aaggttaagat
	241 ggagagcctt gtccctggtt tcaacgagaa aacacacgtc caactcagtt tgccctgtttt

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

301 acaggttcgc gacgtgctcg tacgtggctt tggagactcc gtggaggagg tcttatcaga  
361 ggcacgtcaa catcttaaag atggcacttg tggcttagta gaagttgaaa aaggcgaaaa  
421 gcctcaactt gaacagccct atgtgttcat caaacgttcg gatgctcgaa ctgcaccta  
481 tggtcatgtt atggttgagc tggtagcaga actcgaaggc attcagtacg gtcgtagtgg  
541 tgagacactt ggtgtccttg tccctcatgt gggcgaaaata ccagtggctt accgcaaggt  
601 tcttcttcgt aagaacggta ataaaggagc tggtgccat agttacggcg ccgatctaaa  
661 gtcatttgac ttaggcgacg agcttggcac tgatccttat gaagatttc aagaaaactg  
721 gaacactaaa catagcagtg gtgttacccg tgaactcatg cgtgagctta acggaggggc  
781 atacactcgc tatgtcgata acaacttctg tggccctgat ggctaccctc ttgagtgcat  
841 taaagacctt ctagcacgtg ctggtaaagc ttcatgcact ttgtccgaac aactggactt  
901 tattgacact aagaggggtg tatactgctg ccgtgaacat gagcatgaaa ttgcttggta  
961 cacggaacgt tctgaaaaga gctatgaatt gcagacaccc tttgaaatta aattggcaaa  
1021 gaaatttgac accttcaatg gggaatgtcc aaattttgta tttcccttaa attccataat  
1081 caagactatt caaccaaggg ttgaaaagaa aaagcttcat ggctttatgg gttagaattcg  
1141 atctgtctat ccagttgcgt caccaaatga atgcaaccaa atgtgcctt caactctcat  
1201 gaagtgtgat cattgtggtg aaacttcatg gcagacgggc gattttgtta aagccacttg  
1261 cgaattttgt ggcactgaga atttgactaa agaaggtgcc actacttgtg gttacttacc  
1321 ccaaaaatgct gttgtaaaaa tttattgtcc agcatgtcac aattcagaag taggacctga  
1381 gcatagtctt gccgaatacc ataatgaatc tggcttgaaa accattcttc gtaagggtgg  
1441 tcgcactatt gcctttggag gctgtgtgtt ctcttatgtt ggttgccata acaagtgtac

*Fig. 12A continued*

SUBSTITUTE SHEET

(RULE 26)

1501 ctattgggtt ccacgtgcta gcgcttaacat aggttgtaac catacaggtg ttgttggaga  
1561 aggttccgaa ggtcttaatg acaaccttct taaaaactc caaaaagaga aagtcaacat  
1621 caatattgtt ggtgacttta aacttaatga agagatcgcc attatttgg cattttttc  
1681 tgcttccaca agtgctttg tggaaactgt gaaaggttt gattataaag cattcaaaca  
1741 aattgttgaa tcctgtggta atttaaaagt tacaaaagga aaagctaaaa aaggtgcctg  
1801 gaatatttgtt gaacagaaat caatactgag tcctctttat gcatttgcatt cagaggctgc  
1861 tcgtgttgta cgatcaattt tctccgcac tcttgaaact gtc当地attt ctgtgcgtgt  
1921 tttacagaag gccgctataa caatactaga tggaaatttca cagtattcac tgagactcat  
1981 ttagtgcataat atgttcacat ctgatttggc tactaacaat ctagttgtaa tggcttacat  
2041 tacaggtgggt gttgttcagt tgacttcgca gtggctaact aacatctttg gcactgttta  
2101 tggaaaaactc aaaccgtcc ttgattggct tgaagagaag tttaaggaag gtgttagagtt  
2161 tcttagagac ggttggaaa ttgttaaatt tatctcaacc tgtgcttgc aaattgtcgg  
2221 tggacaaaatt gtcacctgtt caaaggaaat taaggagagt gttcagacat tctttaaagct  
2281 tggatataaaa ttttggctt tggatgtgttgc ctctatcatt attgggtggag ctaaactttaa  
2341 agccttgaat ttaggtgaaa catttgcac gcactcaaag ggattgtaca gaaagtgtgt  
2401 taaatccaga gaagaaaactg gcctactcat gcctctaaaa gccccaaaag aaattatctt  
2461 cttagaggga gaaacacttc ccacagaagt gttAACAGAG gaagttgtct tgaaaactgg  
2521 tgatttacaa ccattagaac aacctactag tggatgtgtt gaaatccat tggttgtac  
2581 accagttgtt attaacgggc ttatgttgct cgaaatcaa gacacagaaa agtactgtgc  
2641 cttgcaccc aatatgatgg taacaaacaa tacttcaca ctcaaaggcg gtgcaccaac  
2701 aaaggttact ttgggtgatg acactgtgat agaagtgcac ggttacaaga gtgtgaat; Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

2761 cactttgaa cttgatgaaa ggattgataa agtacttaat gagaagtgt ctgcctatac  
2821 agttgaactc ggtacagaag taaatgagtt cgccctgtgtt gtggcagatg ctgtcataaa  
2881 aactttgcaa ccagtatctg aattacttac accactggc attgatttag atgagtggag  
2941 tatggctaca tactacttat ttgatgagtc tggtagttt aaattggctt cacatatgt  
3001 ttgttcttt tacccctccag atgaggatga agaagaaggt gattgtgaag aagaagagtt  
3061 tgagccatca actcaaatatg agtatggtac tgaagatgtat taccaaggta aacctttgga  
3121 atttggtgcc acttctgtcg ctcttcaacc tgaagaagag caagaagaag attggttaga  
3181 ttagatgatgt caacaaactg ttggtcaaca agacggcagt gaggacaatc agacaactac  
3241 tattcaaaca attgttggagg ttcaacctca attagagatg gaacttacac cagttgttca  
3301 gactattgaa gtgaatagtt ttagtggta tttaaaactt actgacaatg tatacattaa  
3361 aaatgcagac attgtggaaag aagctaaaaa ggtaaaaacca acagtggtg ttaatgcagc  
3421 caatgtttac cttaaacatg gaggaggtgt tgcaggagcc ttaaataagg ctactaaca  
3481 tgccatgcaa gttgaatctg atgattacat agtactaat ggaccactta aagtgggtgg  
3541 tagttgtgtt ttaagcggac acaatcttgc taaacactgt cttcatgttg tcggcccaaa  
3601 tgttaacaaa ggtgaagaca ttcaacttct taagagtgt tatgaaaatt ttaatcagca  
3661 cgaqttcta cttgcaccat tattatcagc tggtagttt ggtgctgacc ctatacattc  
3721 tttaagagtt tgttagata ctgttcgcac aaatgtctac ttagctgtct ttgataaaaaa  
3781 tctctatgac aaacttggtt caagctttt ggaaatgaag agtaaaaagc aagttgaaca  
3841 aaagatcgct gagattccta aagaggaagt taagccattt ataactgaaa gtaaaccttc  
3901 agttgaacag agaaaacaag atgataagaa aatcaaagct tgtgttgaag aagttacaac

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

3961 aactctggaa gaaaactaagt tcctcacaga aaacttgtta ctttatattg acattaatgg  
4021 caatcttcat ccagattctg ccactcttgt tagtgacatt gacatcactt tottaaagaa  
4081 agatgctcca tatatagtgg gtgatgttgt tcaagagggt gtttaactg ctgtggttat  
4141 acctactaaa aagtctggtg gcactactga aatgctagcg aaagcttga gaaaagtgcc  
4201 aacagacaat tatataacca cttaccggg tcagggtta aatggttaca ctgttagagga  
4261 ggcaaagaca gtgcttaaaa agtgtaaaag tgcctttac attctaccat ctattatctc  
4321 taatgagaag caagaaattc ttggaactgt ttcttggaat ttgcgagaaa tgcttgacaca  
4381 tgcagaagaa acacgcaa at taatgcctgt ctgtgtggaa actaaagcca tagttcaac  
4441 tatacagcgt aaatataagg gtattaaaat acaagagggt gtggttgatt atggtgctag  
4501 attttacttt tacaccagta aaacaactgt aqcgtaactt atcaacacac ttaacgatct  
4561 aaatgaaact cttgttacaa tgccacttgg ctatgtaca catggcttaa atttggaaaga  
4621 agctgctcgg tatatgagat ctctcaaagt gccagctaca gtttctgtt cttcacctga  
4681 tgctgttaca gcgtataatg gttatcttac ttcttcttct aaaacacctg aagaacattt  
4741 tattgaaacc atctcacttg ctggttccta taaagattgg tcctattctg gacaatctac  
4801 acaacttaggt atagaatttc ttaagagagg tgataaaaagt gtatattaca ctagtaatcc  
4861 taccacattc caccttagatg gtgaagttat caccttgac aatcttaaga cacttcttc  
4921 tttgagagaa gtgaggacta ttaaggtgtt tacaacagta gacaacattha acctccacac  
4981 gcaagttgtg gacatgtcaa tgacatatgg acaacagttt ggtccaaactt atttggatgg  
5041 agctgatgtt actaaaataa aacctcataa ttcacatgaa ggtaaaacat tttatgtttt  
5101 acctaataatgat gacactctac gtgttgaggc ttttgagttac taccacacaa ctgatcctag  
5161 ttttctgggt aggtacatgt cagcattaaa tcacactaaa aagtggaaat atccacaag Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

5221 taatggttta acttctatta aatgggcaga taacaactgt tatcttgcca ctgcattgtt  
5281 aacactccaa caaatagagt tgaagttaa tccacctgct ctacaagatg cttattacag  
5341 agcaaggct ggtgaagctg ctaacttttgc tgcacttatac ttgcctact gtaataagac  
5401 agtaggtgag tttaggtgatg tttagagaaaac aatgagttac ttgtttcaac atgccaattt  
5461 agattcttgc aaaagagtct tgaacgcggt gtgtaaaact tgtggacaac agcagacaac  
5521 ccttaagggt gtagaagctg ttatgtacat gggcacactt tcttatgaac aatttaagaa  
5581 aggtgttcag ataccttgc cgtgtggtaa acaagctaca aaatatctag tacaacagga  
5641 gtcacccccc gttatgtatgt cagcaccacc tgctcatgtat gaacttaagc atggcacatt  
5701 tacttgtgct agtgagtaca ctggtaatta ccagtgtggc cactataaac atataacttc  
5761 taaaqaaaact ttgtatttgc tagacqgtgc tttacttaca aagtccctcag aatacaaagg  
5821 tcctattacg gatgtttct acaaagaaaa cagttacaca acaaccataa aaccagttac  
5881 ttataaaattt gatgggtgtg tttgtacaga aattgaccct aagttggaca attattataa  
5941 gaaagacaat ttttatttca cagagcaacc aattgatctt gtacccaaacc aaccatatcc  
6001 aaacgcaagc ttcgataatt ttaagttgt atgtgataat atcaaatttgc ctgatgattt  
6061 aaaccagtta actggttata agaaacctgc ttcaagagag cttaaaagtta cattttcccc  
6121 tgacttaaat ggtgatgtgg tggctattga ttataaacac tacacaccct cttttaagaa  
6181 aggagctaaa ttgttacata aacctattgt ttggcatgtt aacaatgcaa ctaataaagc  
6241 cacgtataaaa ccaaataacct ggtgtatacg ttgttttgg annnnnnnnnn nnnnnnnnnnn  
6301 nnnnnnnnnnnnnnnnnn nnnnatgtac tgaagtcaga ggacgcgcag ggaatggata atcttgccctg  
6361 cgaagatcta aaacttagtct ctgaagaagt agtggaaaat cctaccatac agaaagacgt

*Fig. 12A continued*

SUBSTITUTE SHEET

(RULE 26)

6421 tcttgagtgt aatgtgaaaa ctaccgaagt tgttaggagac attatactta aaccagcaaa  
6481 taatagtttta aaaattacag aagagggtgg ccacacagat ctaatggctg cttatgtaga  
6541 caattctagt cttaactatta agaaaacctaa tgaattatct agagtattag gtttgaaaac  
6601 ccttgctact catggtttag ctgctgttaa tagtgtccct tgggatacta tagctaatta  
6661 tgctaaggcct tttcttaaca aagttgttag tacaactact aacatagtta cacgggtttt  
6721 aaaccgtgtt tgtactaatt atatgcctta tttctttact ttattgctac aattgtgtac  
6781 ttttactaga agtacaaaatt ctagaattaa agcatctatg ccgactacta tagcaaagaa  
6841 tactgttaag agtgtcggtt aattttgtct agaggcttca tttaattatt tgaagtcc  
6901 taattttctt aaactgataa atattataat ttgggtttta ctattaagtg tttgccttagg  
6961 ttcttaatc tactcaaccg ctgctttagg tgtttaatg tctaatttag gcatgccttc  
7021 ttactgtact ggttacagag aaggctattt gaactctact aatgtcacta ttgcaaccta  
7081 ctgtactggt tctatatott gtagtgtttg tcttagtggt ttagattctt tagacacacca  
7141 tccttcttta gaaactatac aaattaccat ttcatctttt aaatggatt taactgcttt  
7201 tggcttagtt gcagagtggt tttggcata tatttttc actaggttt tctatgtact  
7261 tggattggct gcaatcatgc aattgtttt cagctattt gcagtgacatt ttattagtaa  
7321 ttcttggctt atgtggtaa taattaatct tgtacaaatg gccccgattt cagctatggt  
7381 tagaatgtac atcttctttg catcattta ttatgtatgg aaaagttatg tgcatgttgt  
7441 agacggttgt aattcatcaa cttgtatgat gtgttacaaa cgtaatagag caacaagagt  
7501 cgaatgtaca actattgtta atgggttttag aaggccctt tatgtctatg ctaatggagg  
7561 taaaggcttt tgcaaactac acaattggaa ttgtgttaat tgtgatacat tctgtgctgg  
7621 tagtacattt attagtgtatg aagttgcgag agacttgtca ctacagttt aaagaccaa Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

7681 aaatcctact gaccagtctt cttacatcg ttagatgtttt acagtgaaga atggttccat  
7741 ccatctttac tttgataaaag ctggtaaaaa gacttatgaa agacatttc tctctcattt  
7801 tgttaactta gacaacctga gagctaataa cactaaagg tcaattgcctt ttaatgttat  
7861 agttttgtat ggttaaatcaa aatgtgaaga atcatctgca aaatcagcgt ctgtttacta  
7921 cagtcagctt atgtgtcaac ctatactgtt actagatcag gcattagtgt ctgatgttgg  
7981 tgatagtgcg gaagttgcag taaaatgtt tgatgcttac gttaatacgt tttcatcaac  
8041 ttttaacgta ccaatggaaa aactcaaaac actagttgca actgcagaag ctgaacttgc  
8101 aaagaatgtg tccttagaca atgtcttatac tacttttatt tcagcagctc ggcaagggtt  
8161 tgttgattca gatgtagaaa ctaaagatgt tgttgaatgt cttaaattgt cacatcaatc  
8221 tgacatagaa gttactggcg atagttgtaa taactatatg ctcacccata acaaagttga  
8281 aaacatgaca ccccgtgacc ttggtgcttg tattgactgt agtgcgcgtc atattaatgc  
8341 gcaggttagca aaaagtcaca acattgctt gatatggaac gttaaagatt tcatgtcatt  
8401 gtctgaacaa ctacgaaaac aaatacgtag tgctgctaaa aagaataact tacctttaa  
8461 gttgacatgt gcaactacta gacaagttgt taatgttgta acaacaaaaga tagcacttaa  
8521 gggtgttaaa attgttaata attgggtgaa gcagtttaatt aaagttacac ttgtgttact  
8581 ttttgttgc gctatttct attaataac acctgttcat gtcatgtcta aacatactga  
8641 ctttcaagt gaaatcatag gatacaaggc tattgatggt ggtgtcactc gtgacatagc  
8701 atctacagat acttgggtttg ctaacaaaca tgctgatttt gacacatggt ttagccagcg  
8761 tggtggtagt tatactaattg acaaagttg cccattgatt gctgcagtca taacaagaga  
8821 agtgggtttt gtcgtgcctg gtttgctgg cacgatatta cgcacaaacta atggtgactt

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

8881 tttgcatttc ttacctagag ttttagtgc agttggtaac atctgttaca caccatcaaa  
8941 acttatacagg tacactgatt ttgcaacatc agcttgtgtt ttggctgctg aatgtacaat  
9001 tttaaagat gttctggta agccattacc atattgttat gataccaatg tactagaagg  
9061 ttctgttgc tatgaaagtt tacgccctga cacacgttat gtgctcatgg atggctctat  
9121 tattcaattt cctaacacct accttgaagg ttctgttaga gtggtaacaa cttttgattc  
9181 tgagtactgt aggcacggca cttgtgaaag atcagaagct ggtgtttgtg tatctactag  
9241 tggtagatgg gtacttaaca atgattatta cagatctta ccaggagttt tctgtggtgt  
9301 agatgctgta aatttactta ctaatatgtt tacaccacta attcaaccta ttggtgcttt  
9361 ggacatatca gcatctatag tagctggtgg tatttagct atcgttagtaa catgccttgc  
9421 ctactattt atgaggtta gaagagctt tggtaatac agtcatgtag ttgcctttaa  
9481 tactttacta ttcccttatgt cattcactgt actctgttta acaccagttt actcatttt  
9541 acctgggttt tattctgttta tttacttgta cttgacattt tatcttacta atgatgttc  
9601 ttttttagca catattcagt ggatggttat gttcacacctt ttagtacctt tctggataac  
9661 aattgcttat atcatttgta tttccacaaa gcatattctat tggttcttta gtaattacct  
9721 aaagagacgt gtagtcttta atgggtttc ctttagtact tttgaagaag ctgcgcgtgtg  
9781 caccttttg ttaaataaaag aaatgtatct aaagttgcgt agtcatgtgc tattacctct  
9841 tacgcaatat aatagatact tagctttta taataagtac aagtatttttta gtggagcaat  
9901 ggatacaact agtacagag aagctgcttg ttgtcatctc gcaaaggctc tcaatgactt  
9961 cagtaactca ggttctgatg ttctttacca accaccacaa atctctatca cctcagctgt  
10021 tttgcagagt ggttttagaa aaatggcatt cccatctggt aaagttgagg gttgtatggt  
10081 acaagtaact tgtggtacaa ctacacttaa cggtcttgg cttgatgacg tagtttag Fig. 12A *continued*

10141 tccaaagacat gtgatctgca cctctgaaga catgcttaac cctaattatg aagattttact  
10201 cattcgtaag tctaattata atttcttggt acaggcttgt aatgttcaac tcagggttat  
10261 tggacattct atgcaaaaatt gtgtacttaa gcttaagggtt gatacagcca atcctaagac  
10321 acctaagtat aagtttgttc gcattcaacc aggacagact ttttcagtgt tagcttgta  
10381 caatggttca ccatctggtg tttaccaatg tgctatgagg cccaaattca ctattaaggg  
10441 ttcattcctt aatggttcat gtggtagtgt tggtttaac atagattatg actgtgtctc  
10501 ttttggttac atgcaccata tggaaattacc aactggagtt catgctggca cagacttaga  
10561 aggttaacttt tatggacctt ttgttgacag gcaaacagca caagcagctg gtacggacac  
10621 aactattaca gttaatgttt tagcttggtt gtacgctgct gttataaatg gagacaggtg  
10681 gtttctcaat cgatttacca caactcttaa tgactttaac cttgtggcta tgaagtacaa  
10741 ttatgaacct ctaacacacaag accatgttga catacttagga cctctttctg ctcaaactgg  
10801 aattgccgtt ttagatatgt gtgcttcatt aaaagaatta ctgaaaaatg gtatgaatgg  
10861 acgtaccata ttgggttagtg ctttattaga agatgaattt acaccctttg atgttggtag  
10921 acaatgctca ggtgttactt tccaaagtgc agtaaaaaga acaatcaagg gtacacacca  
26) 10981 ctgggttgtt ctcacaatti tgacttcact tttagttta gtccagagta ctcaatggc  
11041 tttgttcttt tttttgtatg aaaatgcctt tttacctttt gctatggta ttattgctat  
11101 gtctgctttt gcaatgatgt ttgtcaaaca taagcatgca tttctctgtt tggttttgtt  
11161 accttctctt gccgctgttag cttatTTAA tatggtctat atgcctgcta gttgggtgt  
11221 gcgtattatg acatggttgg atatggttga tactagtttgc tctggttta agctaaaaga  
11281 ctgtgttatg tatgcatcag ctgtgggttt actaattcctt atgacagcaa gaactgtgt

*Fig. 12A* continued

SUBSTITUTE SHEET

(RULE 26)

11341 tgatgatggt gctaggagag tgtggacact tatgaatgtc ttgacactcg tttataaaagt  
 11401 ttattatggt aatgctttag atcaagccat ttccatgtgg gctcttataa tctctgttac  
 11461 ttctaactac tcaggtgtag ttacaactgt catgttttg gccagaggtt ttgttttat  
 11521 gtgtgttag tattgcccta ttttcttcat aactggtaat acacttcagt gtataatgt  
 11581 agtttattgt ttcttaggct atttttgtac ttgttacttt ggctctttt gtttactcaa  
 11641 ccgcctacttt agactgactc ttgggtttt tgattactta gtttctcac aggagtttag  
 11701 atatatgaat tcacagggac tactcccacc caagaatagc atagatgcct tcaaactcaa  
 11761 cattaaatttggtgggtgttggtggcaaaacc ttgttatcaaa gtagccactg tacagtctaa  
 11821 aatgtcagat gtaaaagtgcacatcagtagt cttaactctca gttttgcaac aactcagagt  
 11881 agaatcatca tctaaatttgt gggctcaatgtgtccagtttacaatgaca ttctcttagc  
 11941 taaaagataact actgaagcct ttgaaaaat ggtttcaacta ctttctgttt tgctttccat  
 12001 gcaggggtgct gtagacataa acaagctttg tgaagaaaatg ctggacaaca gggcaacctt  
 12061 acaagctata gcctcagagt ttagttccct tccatcatat gcagctttg ctactgctca  
 12121 agaagcttat gagcaggctg ttgctaatgg tgattctgaa gttgttctta aaaagttgaa  
 12181 gaagtctttg aatgtggcta aatctgaatt tgaccgtat gcagccatgc aacgttaagtt  
 12241 ggaaaaagatg gctgatcaag ctatgaccca aatgtataaa caggctagat ctgaggacaa  
 12301 gagggcaaaa gttacttagtgc tcatgcagac aatgcctttc actatgctta gaaagttgga  
 12361 taatgatgca ctcaacaaca ttatcaacaa tgcaagagat ggtgtgttc ctttgaacat  
 12421 aataacctttt acaacagcag ccaaactaat ggttgtcata ccagactata acacatataa  
 12481 aaatacgtgt gatggtacaa catttactta tgcatcagca ttgtggaaa tccaacaggt  
 12541 tgttagatgca gatagtaaaa ttgttcaact tagtggaaatt agtatggaca attcacctaa

*Fig. 12A continued*

SUBSTITUTE SHEET

(RULE 26)

12601 ttttagcatgg cctcttattg taacagctt aaggccaat tctgctgtca aattacagaa  
12661 taatgagctt agtcctgttg cactacgaca gatgtcttgc gctgccggta ctacacaaaac  
12721 tgcttgcact gatgacaatg cgtagctta ctacaacaca acaaagggag gtaggtttgt  
12781 acttgcactg ttatccgatt tacaggattt gaaatggcgt agattcccta agagtgtatgg  
12841 aactggtaact atctatacag aactggaacc acctttagg tttgttacag acacacctaa  
12901 aggtcctaaa gtgaagtatt tatactttat taaaggatta aacaacctaa atagaggtat  
12961 ggtacttgggt agtttagctg ccacagtacg tctacaagct ggtaatgcaa cagaagtgcc  
13021 tgccaattca actgttattat ctttctgtgc ttttgcgtta gatgctgcta aagcttacaa  
13081 agattatcta gctagtgggg gacaaccaat cactaattgt gttaagatgt tgtgtacaca  
13141 cactggtaact ggtcaggcaa taacagttac accggaagcc aatatggatc aagaatcctt  
13201 tggtggtgca tcgtgttgc tgtactgccg ttgccacata gatcatccaa atcctaaagg  
13261 attttgtgac taaaaaggta agtatgtaca aataacctaca acttgtgcta atgaccctgt  
13321 gggttttaca cttaaaaaca cagtctgtac cgtctgcgggt atgtggaaag gttatggctg  
13381 tagttgtgat caactcccgcg aacccatgct tcagtcagct gatgcacaat cgtttttaaa  
13441 cgggtttgcg gtgttaagtgc agcccgtttt acaccgtgcg gcacaggcac tagtactgat  
13501 gtcgtataca gggctttga catctacaat gataaagtag ctggtttgc taaattccta  
13561 aaaactaatt gttgtcgctt ccaagaaaaag gacgaagatg acaatttaat tgattcttac  
13621 ttttagtttta agagacacac tttctctaac taccaacatg aagaaaacaat ttataattta  
13681 cttaaggatt gtccagctgt tgctaaacat gacttcttta agtttagaat agacggtgac  
13741 atggtaccac atatatcacg tcaacgtttt actaaataca caatggcaga cctcgtctat

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

13801 gctttaaggc attttgcatga aggttaattgt gacacattaa aagaaaatact tgtcacatac  
13861 aatttgttgtg atgatgatta tttcaataaaa aaggactggt atgattttgt agaaaaaccca  
13921 gatatattac gcttatacgc caacttaggt gaacgtgtac gccaaagctt gttaaaaaca  
13981 gtacaattct gtgatgccat gcgaaatgct ggtattgttgc gtgtactgac attagataat  
14041 caagatctca atggtaactg gtatgatttc ggtgatttca tacaaaccac gccaggtagt  
14101 ggagttcctg ttgttagattc ttattattca ttgttaatgc ctatattaac cttgaccagg  
14161 gctttaactg cagagtcaca ttttgacact gacttaacaa agccttacat taagtggat  
14221 ttgttaaaat atgacttcac ggaagagagg ttaaaaactct ttgaccgttta ttttaaatat  
14281 tgggatcaga cataccaccc aaattgtgtt aactgtttgg atgacagatg cattctgcat  
14341 tgtgcaaact ttaatgtttt attctctaca gtgttcccac ttacaagttt tggaccacta  
14401 gtgagaaaaa tatttgttga tggtgttcca tttgttagttt caactggata ccacttcaga  
14461 gagcttaggtg ttgtacataa tcaggatgt aacttacata gctctagact tagtttaag  
14521 gaattacttg tgtatgctgc tgaccctgtt atgcacgctg cttctggtaa tctattacta  
14581 gataaacgca ctacgtgott ttcagtagct gcacttacta acaatgttgc ttttcaaact  
14641 gtc当地acccg gtaattttaa caaagacttc tatgactttg ctgtgtctaa gggtttcttt  
14701 aaggaaggaa gttctgttga attaaaacac ttcttctttg ctcaggatgg taatgctgt  
14761 atcagcgatt atgactacta tcgttataat ctaccaacaa tgtgtgatat cagacaacta  
14821 ctatttgttgc ttgaagttgt tgataagtac tttgattgtt acgatggtgg ctgttataat  
14881 gctaaccagg tcatcgtaa caaccttagac aaatcagctg gttttccatt taataaatgg  
14941 ggtaaggcta gacttttata tgattcaatg agttatgagg atcaagatgc acttttgc  
15001 tataaaaaac gtaatgtcat ccctactata actcaaatga atcttaagta tgccattac Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

15061 gcaaagaata gagctcgac cgtagctggt gtctcttatct gtagtactat gaccaataga  
15121 cagtttcattc aaaaattatt gaaatcaata gccgccacta gaggagctac ttagtaatt  
15181 ggaacaagca aattctatgg tggttggcac aacatgttaa aaactgttt tagtgatgta  
15241 gaaaaccctc accttatggg ttgggattat cctaaatgtg atagagccat gcctaacatg  
15301 cttagaatta tggcctcaact tgttcttgct cgcaaacata caacgtgtg tagttgtca  
15361 caccgtttct atagattagc taatgagtgt gctcaagtat tgagtgaaat ggtcatgtgt  
15421 ggcagttcac tatatgttaa accaggtgga acctcatcag gagatgccac aactgcttat  
15481 gctaatagtg ttttaacat ttgtcaagct gtcacggcca atgttaatgc acttttatct  
15541 actgatggta acaaaattgc cgataaagtat gtccgcaati tacaacacag actttatgag  
15601 tgtctctata gaaatagaga tggtgacaca gactttgtga atgagttta cgcatatttg  
15661 cgtaaacatt tctcaatgat gatactctct gacgatgctg ttgtgtgtt caatagcact  
15721 tatgcatttc aaggtagtggt ggcttagcata aagaactttt agtcagttct ttattatcaa  
15781 aacaatgttt ttatgtctga agcaaaatgt tggactgaga ctgaccttac taaaggacct  
15841 catgaatttt gctctcaaca tacaatgcta gttaaacagg gtgatgatta tgtgtacctt  
15901 ctttacccag atccatcaag aatccttaggg gccggctgtt ttgttagatga tatcgtaaaa  
15961 acagatggta cacttatgat tgaacggttc gtgtcttttag ctataatgc ttacccactt  
16021 actaaacatc ctaatcagga gtatgctgat gtcttcatt tgcattaca atacataaga  
16081 aagctacatg atgagttAAC aggacacatg ttagacatgt attctgttt gcttactaat  
16141 gataaacatt caaggtattg ggaacctgag ttttatgagg ctatgtacac acogcataca  
16201 gtcttacagg ctgttgggc ttgtgttctt tgcaattcac agacttcatt aagatgtggt  
16261 gcttgcatac gtagaccatt cttatgttgt aaatgctgtt acgaccatgt catatcaac Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

16321 tcacataaaat tagtcttgtc tgttaatccg tatgttgca atgctccagg ttgtgatgtc  
 16381 acagatgtga ctcaacttta cttaggaggt atgagctatt attgtaaaatc acataaaacta  
 16441 cccattagtt ttccattgtg tgctaattgga caagttttg gtttatataa aaatacatgt  
 16501 gttggtagcg ataatgttac tgactttaat gcaattgcaa catgtgactg gacaaatgt  
 16561 ggtgattaca ttttagctaa cacctgtact gaaagactca agcttttgc agcagaaacg  
 16621 ctcaaagcta ctgaggagac atttaaactg tcttatggta ttgctactgt acgtgaagtg  
 16681 ctgtctgaca gagaattaca tcttcatgg gaagttggta aacctagacc accacttaac  
 16741 cgaaattatg tctttactgg ttatcgtgta actaaaaaca gtaaaagtaca aataggagag  
 16801 tacaccttg aaaaaggtaa ctatggtcat gctgttgtt accgaggtac aacaacttac  
 16861 aaattaaatg ttggtgatta ttttgtgctg acatcacata cagtaatgcc attaagtgca  
 16921 cctacactag tgccacaaga gcactatgtt agaattactg gcttataccc aacactcaat  
 16981 atctcagatg agttttctag caatgttgca aattatcaaa aggttggtat gcaaaagtat  
 17041 tctacactcc agggaccacc tggtaactggt aagagtcati ttgctattgg cctagctctc  
 17101 tactaccctt ctgctcgcat agtgtataca gcttgctctc atgccgctgt tgatgcacta  
 17161 tgtgagaagg cattaaaata tttgcctata gataaatgta gtagaattat acctgcacgt  
 17221 gctcgtgttag agtgtttga taaattcaaa gtgaattcaa cattagaaca gtatgtctt  
 17281 tgtactgtaa atgcattgcc tgagacgaca gcagatatacg ttgtctttga tgaaatttca  
 17341 atggccacaa attatgattt gagtggtgtc aatgccagat tacgtgctaa gcactatgtg  
 17401 tacattggcg accctgctca attacctgca ccacgcacat tgctaactaa gggcacacta  
 17461 gaaccagaat atttcaattc agtgtgtaga cttatgaaaa ctataggtcc agacatgttc

*Fig. 12A continued*

SUBSTITUTE SHEET

(RULE 26)

17521 ctcggaaactt gtcggcggttg tcctgctgaa attgttgaca ctgtgagtgc tttggtttat  
17581 gataataagc ttaaaggcaca taaagacaaa tcagctcaat gctttaaaat gttttataag  
17641 ggtgttatca cgcatgatgt ttcatctgca attaacaggc cacaaatagg cgtggtaaga  
17701 gaattcccta cacgtaaccc tgcttggaga aaagctgtct ttatttcacc ttataattca  
17761 cagaatgctg tagcctcaaa gatTTTggga ctaccaactc aaactgttga ttcacacag  
17821 ggctcagaat atgactatgt catattcaact caaaccactg aaacagctca ctcttgtaat  
17881 gtaaacagat ttaatgttgc tattaccaga gcaaaagtag gcatactttg cataatgtct  
17941 gataagagacc ttatgaccaa gttgcaattt acaagtcttg aaattccacg taggaatgtg  
18001 gcaactttac aagctgaaaaa tgtaacagga ctctttaaag attgttagtaa ggtaatca  
18061 gggttacatc ctacacaggc acctacacac ctcagtgttgc acactaaatt caaaactgaa  
18121 ggTTTATGTG ttgacatacc tggcataacct aaggacatga cctatagaag actcatct  
18181 atgatgggtt taaaatgaa ttatcaagtt aatggttacc ctaacatgtt tatcacccgc  
18241 gaagaagcta taagacatgt acgtgcattgg attggcttcg atgtcgaggg gtgtcatgct  
18301 actagagaag ctgttggtag caatttacct ttacagctag gttttctac aggtgttaac  
18361 ctagttgctg tacctacagg ttatgttcat acacctaata atacagattt ttccagagtt  
18421 agtgctaaac caccgcctgg agatcaattt aaacacctca taccacttat gtacaaagga  
18481 cttccttggaa atgttagtgcg tataaagatt gtacaaatgt taagtgacac actttaaaat  
18541 ctctctgaca gagtcgtatt tgtcttatgg gcacatggct ttgagttgac atctatgaag  
18601 tattttgtga aaataggacc tgagcgcacc tgTTTGTCTAT gtgatagacg tgccacatgc  
18661 ttttccactg cttcagacac ttatgcctgt tggcatcatt ctattggatt tgattacgta  
18721 tataatccgt ttatgattga tgttcaacaa tggggTTTA caggtAACCT acaaagca: Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

18781 catgatctgt attgtcaagt ccatggtaat gcacatgtag ctagttgtga tgcaatcatg  
18841 actagggtgc tagctgtcca cgagtgcctt gttaagcgtg ttgactggac tattgaatat  
18901 cctataattg gtgatgaact gaagattaat gcggcttgt aaaaagggtca acacatggtt  
18961 gttaaagctg cattattagc agacaaattc ccagttcttc acgacattgg taaccctaaa  
19021 gctattaagt gtgtacctca agctgatgt aaatggaaatgt tctatgtgc acagccttgt  
19081 agtgacaaag ctatataaaat agaagaatta ttcttattctt atgccacaca ttctgacaaa  
19141 ttcacagatg gtgtatgcctt attttggaaat tgcaatgtcg atagatatcc tgttaattcc  
19201 attgtttgtt gatttgacac tagagtgcta tctaaccctta acttgcctgg ttgtgatgg  
19261 ggcagtttgt atgtaaataa acatgcattc cacacaccag cttttgataa aagtgccttt  
19321 gttaatttaa aacaattacc atttttctat tactctgaca gtccatgtga gtctcatgga  
19381 aaacaagtag tgtcagatat agattatgt acaactaaatgt ctgctacgtg tataaacacgt  
19441 tgcaatttag gtgggtgtgt ctgttagacat catgctaattgt agtacagatt gtatctcgat  
19501 gcttataaca tggatgtatctc agctggctt agcttgggg tttacaaaca atttgataact  
19561 tataacctct ggaacacattt tacaagactt cagagtttag aaaatgtggc ttttaatgtt  
19621 gtaaataagg gacactttga tggacaacag ggtgaagtac cagtttctat cattaataac  
19681 actgtttaca caaaagttga tgggtttgtat gtagaattgt ttgaaaataa aacaacattt  
19741 cctgttaatg tagcatttga gctttggct aagcgcaaca tttaaaccagt accagaggtg  
19801 aaaatactca ataatttggg tgtggacatt gctgctaata ctgtgatctg ggactacaaa  
19861 agagatgtc cagcacatat atctactatt ggtgtttgtt ctatgactga catagccaag  
19921 aaaccaactg aaacgatttg tgcaccactc actgtctttt ttgatggtag agttgatgg

*Fig. 12A continued*

SUBSTITUTE SHEET

19981 caagtagact tatTTAGAAA tgcccgtaat ggtgttctta ttacagaagg tagtgttaaa  
20041 ggTTACAAC catCTGTagg tcccaAAACAA gctagtctta atggagtcac attaattgga  
20101 gaAGCCGtaa aaACACAGTT caATTATTAT aAGAAAGTTG atGGTGTGT ccaacaatta  
20161 CCTGAAACTT actttactca gagtagaaat ttacaagaat ttaaACCCAG gagtcaaATG  
20221 gaaATTGATT TCTTAGAATT agCTATGGAT gaATTcATTG AACGGTATAA ATTAGAAGGC  
20281 tatgccttcg aacatatacgt ttatggagat tttAGTCATA gtcAGTTAGG tggTTTACAT  
20341 ctactgattg gactagctaa acgttttaAG gaATCACCTT ttGAATTAGA agATTTATT  
20401 CCTATGGACA gtacAGTTAA AAACTATTTC ataACAGATG CGCAAACAGG ttcATCTAAG  
20461 tgtgtgtgtt ctgttattGA tttattACTT gatgATTTG ttGAAATAAT AAAATCCCAA  
20521 gatttatctg tagTTTCTAA ggTTGTCAAA gtGACTATTG ACTATAcAGA aATTcATTt  
20581 atgCTTTGGT gtaAAAGATGG ccatgtAGAA acATTTACC CAAAATTACA atCTAGTCaa  
20641 gcgtggcaac cgggtgtgc tatgcctaAT ctTTACAAAAA tgcaaAGAA gctattAGAA  
(20701 aagtgtgacc ttcaAAATTa tggtgatAGT gcaACATTAC ctAAAGGcAt aAtgAtGAAT  
20761 gtcgcaAAAT atactcaACT gtgtcaATAT ttaAAACACAT taACATTAGC tgtaccCTAT  
20821 aatATGAGAG ttATACATTt tggtgctGGT tctgataAAAG gagTTGcAcc aggtacAGCT  
20881 gttttaAGAC agtggttGCC tacGGGTACG ctgtttgtcg attcAGATCT taATGACTTT  
20941 gtctctgatG cAGATTCAAC tttgattGGT gattgtGCAA ctgtACATAC agctaATAAA  
21001 tggatctca ttattAGTGA tatgtacGAC CCTAAAGACTA AAAATGTTAC AAAAGAAAAT  
21061 gactctAAAG aggTTTTT cacttACATT tggGGTTA tacaACAAAAA gctAGCTCTT  
21121 ggaggttCCG tggCTATAAA gataACAGAA cattCTTGGA atgctgatCT ttataAGCTC  
21181 atgggacACT tcgcatggTG gacAGCCTT gttACTAATG tgaATGCGTC atcatCTG

*Fig. 12A continued*

SUBSTITUTE SHEET

(RULE

26)

21241 gcattttaa ttggatgtaa ttatcttggc aaaccacgcg aacaaataga tggttatgtc  
 21301 atgcatacgaa attacatatt ttggaggaat acaaataccaa ttcaagggttc ttccatttct  
 21361 ttatggaca tgtagtaaatt tccccttaaaa ttaagggttc ctgctgttat gtctttaaaaa  
 21421 gaaggtcaaa tcaatgatat gatTTtatct cttcttagta aaggttagact tataattaga  
 21481 gaaaacaaca gagttgttat ttcttagtgat gttcttgtta acaactaaac gaacaatgtt  
 21541 tgTTTTCTT gTTTATTGC cactagtctc tagtcagtgt gttaatctta gaaccagaac  
 21601 tcaattaccc cctgcataca ctaattttt cacacgtggt gtttattacc ctgacaaagt  
 21661 ttccagatcc tcagtttac attcaactca ggacttgttc ttacctttct tttccaatgt  
 21721 tacttggttc catgctatac atgtctctgg gaccaatggt actaagaggt ttgataaccc  
 21781 tgtoctacca tttaatgatg gtgtttattt tgcttccayt gagaagtcta acataataag  
 21841 aggctggatt ttgggtacta cttagattc gaagacccag tccctactta ttgttaataaa  
 21901 cgctactaat gttgttatta aagtctgtga atttcaattt tgtaatgatc catttttggaa  
 (21961 tgTTTATTAC cacaaaaaca acaaaggatg gatggaaagt gnnnnnnngag tttattctag  
 22021 tgctactaat tgcaCTTTG aatatgtctc tcagcctttt cttatggacc ttgaaggaaa  
 22081 acagggtaat ttcaaaaatc ttagggatt tgtgtttaag aatattgatg gttatTTAA  
 22141 aatatattct aagcacacgc ctattaattt agtgcgtgat ctccctcagg gttttcggc  
 22201 tttagaacca ttggtagatt tgccaatagg tattaacatc actaggttc aaactttact  
 22261 tgctttacat agaagttatt tgactctgg tgattttct tcaggttggc cagctggc  
 22321 tgcaGCTTAT tatgtgggtt atcttcaacc taggactttt ctattaaat ataatgaaaa  
 22381 tggaaccatt acagatgtg tagactgtgc acttgaccct ctctcagaaa caaagtgtac

*Fig. 12A continued*

22441 gttgaaatcc ttcactgtag aaaaaggaat ctatcaaact tctaacttta gagtccaacc  
22501 aacagaatct attgttagat ttccataat tacaacttg tgccctttg gtgaagttt  
22561 taacgccacc agatttgcatttgc ttggaacagg aagagaatca gcaactgtgt  
22621 tgctgattat tctgtcctat ataattccgc atcattttcc acttttaagt gttatggagt  
22681 gtctcctact aaattaaatg atctctgtt tactaatgtc tatgcagatt catttgttaat  
22741 tagaggtgat gaagtcagac aaatcgctcc agggcaaact ggaaagattg ctgattataa  
22801 ttataaaatta ccagatgatt ttacaggctg cggtatagct tggaaattcta acaatcttga  
22861 ttctaagggtt ggtggtaatt ataattaccg gtatagattt ttttaggaagt ctaatctcaa  
22921 accttttgag agagatattt caactgaaat ctatcaggcc ggttagcaaac cttgtatgg  
22981 tggtgaaggt tttaattgtt actttcctt acaatcatat ggtttccaac ccactaatgg  
23041 tggtggttac caaccataca gagtagtagt actttcttt gaacttctac atgcaccagg  
23101 aactgtttgt ggacctaaaa agtctactaa ttggtaaaa aacaaatgtg tcaatttcaa  
23161 cttcaatggc ttaacaggca caggtgttct tactgagtct aacaaaaagt ttctgccttt  
23221 ccaacaattt ggcagagaca ttgctgacac tactgatgct gtccgtgatc cacagacact  
23281 tgagattttt gacattacac catgttctt tggtgggtgc agtgttataa caccaggaac  
23341 aaataacttct aaccagggtt ctgttcttta tcagggtgtt aactgcacag aagtccctgt  
23401 tgctattcat gcagatcaac ttactcctac ttggcgtgtt tattctacag gttctaattgt  
23461 ttttcaaaca cgtgcaggct gtttaatagg ggctgaacat gtcaacaact catatgagt  
23521 tgacatacccc attgggtgcag gtatatgcgc tagttatcag actcagacta attctcgat  
23581 gcgggcacgt agtgttagcta gtcaatccat cattgcctac actatgtcac ttgggtgcaga  
23641 aaattcagtt gcttactcta ataactctat tgccatacccc acaaatttttta ctatttagtgt

SUBSTITUTE SHEET (RULE 26)

23701 taccacagaa attctaccag tgtctatgac caagacatca gtagattgt aatgtacat  
23761 ttgtggtgat tcaactgaat gcagcaatct tttgttgc aa tatggcagtt tttgtacaca  
23821 attaaaaccgt gctttaactg gaatacgctgt tgaacaagac aaaaacaccc aagaagttt  
23881 tgcacaagtc aaacaaattt aaaaaacacc accaattaaa gatTTTggtg gtttaattt  
23941 ttcacaaata ttaccagatc catcaaaacc aagcaagagg tcatttattt aagatctact  
24001 tttcaacaaa gtgacacttg cagatgctgg cttcatcaaa caatatggtg attgccttgg  
24061 tgatattgt gctagagacc tcatttgc acaaaagttt aacggccta ctgtttgcc  
24121 acctttgctc acagatgaaa tgattgctca atacacttct gcactgttag cgggtacaat  
24181 cacttctggc tgacacccgc tgccaggcgc tgcattacaa ataccatttgc tgcacaaat  
24241 ggcttatagg ttaatggta ttggagttac acagaatgtt ctctatgaga accaaaaatt  
24301 gattgccaac caatttaata gtgttattgg caaaattcaa gactcacttt cttccacagc  
24361 aagtgcactt ggaaaacttc aaaatgttgtt caaccaaaat gcacaagctt taaacacgct  
24421 tgtaaacaat ctttgttgc aatttcaatgtt gttttaaatgtt atatccttc  
24481 acgtttgac aaagttgagg ctgaagtgc aattgtatgg ttgtatcacag gcagacttca  
24541 aagttgcag acatatgttgc ctcaacaatt aatttagatgtt gcagaaatca gagttctgc  
24601 taatcttgc gctactaaaa tgtcagatgtt tgtacttggc caatcaaaaa gagttgattt  
24661 ttgtggaaag ggcttatcatc ttatgtcctt ccctcagtca gcacctcatg gtgttagtctt  
24721 cttgcattgtt acttatgtcc ctgcacaaga aaagaacttc acaactgctc ctgccatttgc  
24781 tcatgtatggc aaagcacact ttcctcgatg aggtgtcttt gttcaaatgc gcacacactg  
24841 gtttgcataaca caaaggaatt tttatgcacc acaaatttactt actacagaca acacatttgc

Fig. 12A continued

SUBSTITUTE SHEET  
(RULE 26)

24901 gtctggtaac tgtgatgtt taataggaat tgtcaacaac acagtttatg atccttgca  
24961 acctgaatta gactcattca aggaggagtt agataaaatat tttagaatc atacatcacc  
25021 agatgttcat tttagtgaca tctctggcat taatgcttca gttgtaaaca ttcaaaaaga  
25081 aattgaccgc ctcaatgagg ttgccaaagaa tttaaatgaa tctctcatcg atctccaaga  
25141 acttggaaag tatgagcagt atataaaatg gccatggtac atttggctag gttttatagc  
25201 tggcttgatt gccatagtaa tggtgacaat tatgctttgc tgtatgacca gttgctgttag  
25261 ttgtctcaag ggctgttgtt cttgtggatc ctgctgcaaa tttgatgaag acgactctga  
25321 gccagtgc tc aaaggagtca aattacatta cacataaaacg aacttatgga tttgtttatg  
25381 agaatcttca caattggaac tgtaactttg aagcaaggtg aaatcaagga tgctactcct  
25441 ttagattttg ttgcgcgtac tgcaacgata ccgataacaag cctcaactccc tttcgatgg  
25501 cttattgttgcgact tttgctgtt tttcagagcg cttccaaaat cataaccctc  
25561 aaaaagagat ggcaacttagc actctccaag ggtgttcact ttgtttgcaa cttgctgttg  
25621 ttgtttgtaa cagtttactc acacctttg ctcgttgctg ctggccgtga agcccccttt  
25681 ctctatctt atgcttttgt ctacttcttgc cagagtataa actttgtaaag aataataatg  
25741 aggctttggc ttgcgtggaa atgcccgttcc aaaaacccat tactttatga tgccaaactat  
25801 tttctttgtt ggcataactaa ttgttacgac tattgtatac cttacaatag tgtaacttct  
25861 tcaattgtca ttacttcagg tgatggcaca acaagtccta tttctgaaca tgactaccag  
25921 attgggtggtt atactgaaaa atggaaatct ggagtaaaag actgtgttgtt attacacagt  
25981 tacttcactt cagactatta ccagctgtac tcaactcaat tgagtacaga cactgggttt  
26041 gaacatgtta ccttcttcat ctacaataaa attgttcatg agcctgaaga acatgtccaa  
26101 attcacaccaa tcgacgggttc atccggagtt gttaatccag taatggaacc aatttatgtat

*Fig. 12A continued*

SUBSTITUTE SHEET  
(RULE 26)

26161 gaaccgacga cgactactag cgtgccttg taagcacaag ctgatgagta cgaacttatg  
26221 tactcatcg tttcgaaaga gacaggtacg ttaatagtta atagcgtact tcttttctt  
26281 gcttcgtgg tattcttgc agttacacta gccatccta ctgcgcttcg attgtgtgcg  
26341 tactgctgca atattgttaa cgtgagtctt gtaaaacctt cttttacgt ttactctgt  
26401 gttaaaaatc tgaattcttc tagagttcct gatcttctgg tctaaacgaa ctaaatatta  
26461 tattagttt tctgtttgga actttaattt tagccatggc agattccaac ggtactatta  
26521 ccgttgaaga gcttaaaaag ctcccttgaac aatggaaccc agtaataggt ttccatttcc  
26581 ttacatggat ttgtcttcta caatttgcct atgccaacag gaataggtt ttgttatataa  
26641 ttaagttaat ttccctctgg ctgttatggc cagtaacttt agcttgttt gtgcttgctg  
26701 ctgtttacag aataaattgg atcaccggtg gaattgctay cgcaatggct tgtctttag  
26761 gcttgatgtg gctcagctac ttcattgctt ctttcagact gttgcgcgt acgcgttcca  
26821 tgtggtcatt caatccagaa actaacattc ttctcaacgt gccactccat ggcactattc  
26881 tgaccagacc gcttctagaa agtgaactcg taatcgagc tgtgatcctt cgtggacatc  
26941 ttctgtattgc tggacaccat ctaggacgct gtgacatcaa ggacctgcct aaagaaatca  
27001 ctgttgctac atcacgaacg ctttcttatt acaaattgg agcttcgcag cgtgttagcag  
27061 gtgactcagg ttttgctgca tacagtcgct acaggattgg caactataaa ttaaacacag  
27121 accattccag tagcagtgac aatattgctt tgcttgtaca gtaagtgaca acagatgttt  
27181 catctcggtt actttcaggt tactatagca gagatattac taattattat gaggacttt  
27241 aaagtttcca tttggaatct tgattacatc ataaacctca taattaaaaa tttatctaag  
27301 tcactaactg agaataaata ttctcaatta gatgaagagc aaccaatgga gattgattaa

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

27361 acgaaacatga aaatttattct tttcttggca ctgataaacac tcgctacttg tgagctttat  
 27421 cactaccaag agtgtgttag aggtacaaca gtactttaa aagaaccttg ctcttcgtga  
 27481 acatacgagg gcaattcacc atttcatcct ctagctgata acaaatttgc actgacttgc  
 27541 tttagcactc aatttgctt tgcttgcct gacggcgtaa aacacgtcta tcagttacgt  
 27601 gccagatcag cttcacctaa actgttcatc agacaagagg aagttcaaga actttactct  
 27661 ccaatttttc ttattgtgc ggcaataagtg tttataacac tttgcttcac actcaaaaaga  
 27721 aagacagaat gattgaactt tcattaattt acttctattt gtgctttta gcctttctgc  
 27781 tattccttgc tttaattatg cttattatct tttggttctc acttgaactg caagatcata  
 27841 atgaaatttgc tcacgcctaa acgaaacatga aatttcttgc tttcttagga atcatcacaa  
 27901 ctgtagctgc atttcaccaa gaatgttagtt tacagtcatg tactcaacat caaccatatg  
 27961 tagttgatga cccgtgtcct attcacttct attctaaatg gtatattaga gtaggagcta  
 28021 gaaaatcagc accttaattt gaattgtgcg tggatgagggc tggttctaaa tcacccattc  
 28081 agtacatcga tatcgtaat tatacagttt cctgtttacc tttacaattt aattgccagg  
 28141 aacctaaattt gggtagtctt gtatgtgcgtt gttcggtctt tgaagacttt ttagagtata  
 28201 atgacgttcg tgggtttta gatttaatct aaacgaacaa actnaaatgt ctgataatgg  
 28261 accccaaaaat cagcgaaatg cacccccat tacgtttggg ggaccctcag attcaactgg  
 28321 cagtaaccag aatggagaac gcagtggggc gcgatcaaaa caacgtcgcc cccaaaggttt  
 28381 acccaataat actgcgtctt gttcaccgc tctcactcaa catggcaagg aaggccttaa  
 28441 attccctcga ggacaaggcg ttccaattaa caccaatagc agtccagatg accaaattgg  
 28501 ctactaccga agagctacca gacgaattcg tggtggtgac ggtaaaatga aagatctcag  
 28561 tccaaagatgg tatttctactt accttaggaac tggggccagaa gctggacttc cctatggtgc

*Fig. 12A continued*

SUBSTITUTE SHEET (RULE 26)

28621 taacaaaagac ggcatcatat gggttgcaac tgagggagcc ttgaatacac caaaagatca  
28681 cattggcacc cgcaatcctg ctaacaatgc tgcaatcgta ctacaacttc ctcaaggaac  
28741 aacattgcca aaaggcttct acgcagaagg gagcagaggc ggcagtcaag cctttctcg  
28801 ttcctcatca ctagtcgca acagttcaag aaattcaact ccaggcagca gtatggaaac  
28861 ttctcctgct agaatggctg gcaatggctg tgatgctgct cttgcttgc tgctgcttga  
28921 cagattgaac cagcttgaga gcaaaatgtc tggtaaaggc caacaacaac aaggccaaac  
28981 tgtcactaag aaatctgctg ctgaggcttc taagaaggct cggaaaaaac gtactgccac  
29041 taaagcatac aatgtAACAC aagcttcgg cagacgtgg ccagaacaaa cccaaaggaaa  
29101 ttttggggac caggaactaa tcagacaagg aactgattac aaacattggc cgcaaattgc  
29161 acaatttgcc cccagcgctt cagcgtttt cgaaatgtcg cgcatggca tggaaagtac  
29221 accttcggga acgtgggtga cctacacagg tgccatcaaa ttggatgaca aagatccaaa  
29281 tttcaaagat caagtcattt tgctgaataa gcatattgac gcatacaaaa cattcccacc  
29341 aacagagcct aaaaaggaca aaaagaagaa ggcttatgaa actcaagcct taccgcagag  
29401 acagaagaaa cagcaaactg tgactcttct tccctgctgca gatttggatg atttctccaa  
29461 acaattgcaa caatccatga gcagtgtga ctcaactcag gcctaaactc atgcagacca  
29521 cacaaggcag atgggctata taaacgttt cgctttccg tttacgatat atagtctact  
29581 cttgtgcaga atgaattctc gtaactacat agcacaagta gatgttagtta actttaatct  
29641 cacatagcaa tctttaatca gtgtgtaca ttagggagga cttgaaagag ccaccacatt  
29701 ttccaccgagg ccacteggag tacgatcgag tgtacagtga acaatgctag ggagagctgc  
29761 ctatatggaa gagccctaatt gtgtaaaatt aatttttagta gtgctatccc catgtgattt

*Fig. 12A continued*

*Fig. 12A continued*

29821 taatagttt ttagggaa

# Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/NLD/EMC-Omicron-1/2021, complete genome

GenBank: OM287553.1

[FASTA Graphics](#)

[Go to:](#)

LOCUS OM287553 29743 bp RNA linear VRL 17-FEB-  
2022 77/254  
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate  
(RULE SHEET) SARS-CoV-2/human/NLD/EMC-Omicron-1/2021, complete genome.  
ACCESSION OM287553  
VERSION OM287553.1  
26)  
KEYWORDS .  
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)  
ORGANISM Severe acute respiratory syndrome coronavirus 2  
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;

SUBSTITUTE SHEET (RULE 26)

Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;  
Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29743)

AUTHORS GeurtsvanKessel, C.H., Geers, D., Schmitz, K.S., Mykytyn, A.Z.,  
Lamers, M.M., Bogers, S., Scherbeijn, S., Gommers, L.,  
Sablerolles, R.S.G., Nieuwkoop, N.N., Rijsbergen, L.C., van  
Dijk, L.L.A., de Wilde, J., Alblas, K., Breugem, T.I.,  
Rijnders, B.J.A.,  
de Jager, H., Weiskopf, D., van der Kuy, P.H.M., Sette, A.,  
Koopmans, M.P.G., Grifoni, A., Haagmans, B.L. and de Vries, R.D.

TITLE Divergent SARS CoV-2 Omicron-reactive T- and B cell responses in  
COVID-19 vaccine recipients

JOURNAL Sci Immunol, eab02202 (2022) In press

PUBMED 35113647

REMARK Publication Status: Available-Online prior to print

REFERENCE 2 (bases 1 to 29743)

AUTHORS Lamers, M., Mykytyn, A., Bestebroer, T. and Haagmans, B.

TITLE Direct Submission

JOURNAL Submitted (18-JAN-2022) Viroscience, Erasmus MC, Wytemaweg 80,  
Rotterdam 3015CN, Netherlands

*Fig. 12B* *continued*

## SUBSTITUTE SHEET (RULE 26)

COMMENT           ##Assembly-Data-START##  
                  Assembly Method        :: CLC genomics workbench v. 21.0.3  
                  Sequencing Technology :: Sanger dideoxy sequencing; Illumina  
                  ##Assembly-Data-END##

FEATURES           Location/Qualifiers

source            1..29743  
                  /organism="Severe acute respiratory syndrome coronavirus  
                  2"  
                  /mol\_type="genomic RNA"  
                  /isolate="SARS-CoV-2/human/NLD/EMC-Omicron-1/2021"  
                  /isolation\_source="oronasopharynx"  
                  /host="Homo sapiens"  
                  /db\_xref="taxon:2697049"  
                  /country="Netherlands"  
                  /collection\_date="2021-11-29"

gene            212..21489  
                  /gene="ORF1ab"

CDS            join(212..13402,13402..21489)  
                  /gene="ORF1ab"  
                  /ribosomal\_slippage

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

/codon\_start=1  
/product="ORF1ab polyprotein"  
/protein\_id="UJD17627.1"  
/translation="MESLVPGFNEKTHVQLSLPVVLQVRDVVLVRGFGDSVEEVLSEARQ  
HLKDGTCGLVEVEKGVLPOLEQPYVFIRSDARTAPHGHVMVELVAELEGIQYGRSGE  
TLGVILVPHVGETPVAYRKVLLRKNGNKAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNCGAYTRYVDNNFCGRPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKLDGFNGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG  
DFVKATCEF CGTENLTKEGATT CGYL P QNAVV K TYCPA CHN SEV GP E H S L A E Y H N E SG  
LKTILRKGGRTIAFGGCVFSYVGCHNCAYWVPRASANIGCNHTGVV GEGSEG LNDNL  
LEILQKEKVNIIVGDFKLNEEIAIILASF SASTSAF VETVKGLDYKA F KQ I V E S C GN

*Fig. 12B continued*

FKVTKGAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFKEGVEFLRGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV  
NKEFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPPLKAPKEII  
FLEGETLPTEVLTEEVVLKTGDLOPЛЕOPTSEAVEAPLVGTPVCINGLMLLEIKDTEK  
YCALAPNMMVTNNNTFLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNER  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF  
KLASHMYCSFYPPDEDEEEGDCEEEEFEPSHQYEYGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQTVGQQDGSEDNQTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG  
YLKLTDNVYIKNADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

*Fig. 12B* *continued*

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTLEETKFLTENLLLVIDINGN  
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVVIPTKKAGGTTEMLAKALRKV  
PTDNYITYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM  
LAHAEETRKILMPVCETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSILKVPATSVSSPDAVTAYNGYLTSSS  
KTPEEHFIETISLAGSYKDWSYSQSTQLGIEFLRGDKSVYYTSNPPTFHLDGEVIT  
FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPNDDTLRVEAFEYYHTDPNFLGRYMSALNHTKKWKPQVNGLTSIK

*Fig. 12B* *continued*

WADNNCYLATALLTLQQIELKFNPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ  
IPCTCGKQATKYLVQQESPFVMMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVPNQYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT  
FFPDNLNGDVVAIDYKHYTPSFKKGAKLHKPIVWHVNNATNKATYKPNTWCIRCLWST  
KPVETSNSFDVLKSEDAQGMNDNLACEDLKPVSEEVVENPTIQKDVLNVKTTEVVGD  
IILKPANNIKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNSV  
PWDTIANYAKPFLNKVVSTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRIK  
ASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSЛИYSTAA

*Fig. 12B* *continued*

LGVILMSNLGMP SYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSDLTYPSE  
QITISSEFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLM  
WLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVEC  
TTIVNGVRRSFYVYANGGKGFCKLHNWNVCNCDFTCAGSTFISDEVARDLSLQFKRP  
NPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPIN  
VIVFDGKSKEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNT  
FSSTFNVPMEKLKTLVATAEAEELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVECL  
KLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNITLIW  
NVKDFMSLSEQLRKQIRSAAKKNLNFKLTCAATTRQVVNVVTTKIALKGGKIVNNWLK  
QLIKVTLVFLFVAATIFYLITPVHVM SKHTDFSSEIIIGYKAIDGGVTRDIASTDTCFAN

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

KHAFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTINGDFLHFILPR  
VFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTINVLEGSVAY  
ESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSGR  
WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCLA  
YYFMRFRRAFGEYSHVVAFTNLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFLTND  
VSFLAHIQWMVMFTPLVPFWITIAIICISTKHEYWFESNYLKRRVVFNGVSFSTFEE  
AALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLA  
KALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFP SGKVEGCMVQVTCGTTLNGL  
WLDDVVYCPRHVIC TSEDMLNP NYEDLLIRKSNNFLVQAGNVQLRVIGHSMQNCVLK  
LKVDTANPKTPKYKFVRIQPGQTF SVLACYNGSPSGVYQCAMRHNFТИKGSLNGSCG

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

SVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNV  
LAWLYAAVINGDRWFLNRFTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVL  
DMCASLKELLQNGMNGRTIILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWL  
LLTILTSLLVLVQS TQWSLFFFYENAFLPFAMGI IAMSAFAMMFVKHKHAFLCLFLL  
PSLATVAYFNMVYMPASWVMRIMTWLDMVDTSFKLKDCVMYASAVVLLILMTARTVYD  
DGARRVWTLMNVLT LVYKVYYGNALDQAISMWALII SVTSNYSGVVT TVMFLARGVVF  
MCVEYCPPIFFITGNTLQCIMLVYCFGLGYFCTCYFGLCLLNRYFRLTLGVYDYLVSTQ  
EFRYMNSQGLLPPKNSIDAFLNIKLLGVGGKPCIKVATVQSKMSDVKCTS VVLLSVL  
QQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSVLLSVLLSMQGAVDINKLCEEM  
LDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKSLNVAKSEFD

*Fig. 12B continued*

SUBSTITUTE SHEET (RULE 26)

RDAAMQRKLEKMAHQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNI IN  
NARDGCVPLNIIPPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQQVVDADSKI  
VQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDD  
NALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPK  
VKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFADVDAAKAYKD  
YLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPK  
GFCDLKKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSADAQS  
FLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKEDEDDN  
LIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQRLTKY  
TMADILVYALRFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGE

*Fig. 12B continued*

RVRQALLKTVQFC DAMRNAGIVGVL TLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSYY  
SLLMPILTLTRALTAESHVD TD LTKPYIKWDLKYDFTEERLKLFDRYFKYWDQTYHP  
NCVNCLDDRCILHCANEVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRELGVV  
HNQDVNLHSSRLSF KELLVYAADPAMHAASGNLLLDKRTTCSVAALTNNVAFQTVKP  
GNFNKDFYDFAVSKGFFKEGSSVELKHFFF AQDGNA AISDYDYYRYNLPTMCDIROLL  
FVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKGKARLYYDSMSYEDQDALE  
AYTKRNVIPITITQMLKYAISAKNRARTVAGVSICSTMNRQFHQKLLKSIAATRGAT  
VVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTT  
CCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTA  
NVNALLSTDGNKIADKYVRNLQHRLYECLYRNDVDIDFVNEEFYAYLRKHE SMMIILSD

*Fig. 12B* *continued*

DAVVCFNSTYASQGLVASICNFKSVLYYQNNVMSEAKCWTETDLTKGPHEFC5QHTM  
LVKQGDDYVYLPYPDPSSRILGAGCFVDDIVKTDTLMIERFVSLAIDAYPLTKHPNQE  
YADVPHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAV  
GACVLCLNSQTSLRCGACIRRPFLLCCKCCYDHVISTSHKLVL SVNPP YVCNAPGCDVTDV  
TQLYLGGMSYYCKSHKPPISFPLCANGQVFGGLYKNTCVGSDNVTDNFNAIATCDWTNAG  
DYILANTCTERLKLEAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPL  
NRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGITTYKLNVDYFVLTSHTVMP  
LSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFA  
IGLALYYPSARIIVYTACSHAADVDALEKALKYLPIDKCSRIIPARARVECFDKFKVNS  
TLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHVVYIGDPAQLPAP

*Fig. 12B continued*

RTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKD  
KSAQCFKMFYKGVIHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNQNAVASK  
ILGLPTQTVDSSQGSEYDYVIFTQTTEAHSCNVNRFNVAITRAKVGILCIMSDRDLY  
DKLQFTSLEIPRRNVATLQAENVITGLFKDCSKVITGLHPHQAPTHLSVDTKFKEGLC  
VDVPGIPKDMTYRRLISMGMFKMNYQVNNGYPNMFITREEAIRHVRAWIGFDVEGCHAT  
REAVGTNLPLQLGFSTGVNLVAVPTGYVDTNNNTDFSRVSAKPPPGDQFKHLIPLMYK  
GLPWNVVRRIKIVQMLSDLTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDRR  
ATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFITGNLQSNHDLYCQVHGNAHVAS  
CDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADKEPVL  
HDIGNPKAIKCVFQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCLFWNC

*Fig. 12B continued*

SUBSTITUTE SHEET (RULE 26)

NVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNKLQLPFF  
YYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHANEYRLYLDAYNMMS  
AGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNTVYTK  
VDGVDVLFENKTTLPVNVAFELWAKRNKPVPEVKILNNLGVDIAANTVIWDYKRDA  
PAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEGSVKG  
LQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFKPRSQ  
MEIDFLELAMDEFIERYKLEGYAFEHIIVYGDFSHSQLGGLHLLIGLAKRFKESPFLE  
DFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKVTIDY  
TEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATL  
PKGIMMNVAKYTQLCQYLNTLTLAVPYNMRCVIHEGAGSDKGVAPGTAVLRQWLPTGTL

*Fig. 12B* *continued*

LVDSDLNDFVSDADSTLIGDCATVHTANKWDEFIISDMYDPKTKNVTKENDSKEGFFTY

ICGFIQQKLALGGSVAIKITEHAWNADLYKLMGHFAWWTAFVTNVNASSSEAFLIGCN

YLGKPREQIDGYVMHANYIFWRNTNPQLSSYSLFDMSKEPLKLRGTAVMSLKEQQIN

DMILSLLSKGRLLIRENNRVISSDVLVNN"

mat\_peptide 212..751

/gene="ORF1ab"

/product="leader protein"

mat\_peptide 752..2665

/gene="ORF1ab"

/product="nap2"

mat\_peptide 2666..8497

/gene="ORF1ab"

/product="nsp3"

mat\_peptide 8498..9997

/gene="ORF1ab"

/product="nsp4"

mat\_peptide 9998..10915

/gene="ORF1ab"

*Fig. 12B* *continued*

mat\_peptide /product="3C-like proteinase"  
10916..11776  
/gene="ORF1ab"  
/product="nsp6"  
mat\_peptide 11777..12025  
/gene="ORF1ab"  
/product="nsp7"  
mat\_peptide 12026..12619  
/gene="ORF1ab"  
/product="nsp8"  
mat\_peptide 12620..12958  
/gene="ORF1ab"  
/product="nsp9"  
mat\_peptide 12959..13375  
/gene="ORF1ab"  
/product="nsp10"  
mat\_peptide join(13376..13402,13402..16170)  
/gene="ORF1ab"  
/product="RNA-dependent RNA polymerase"

*Fig. 12B continued*

## SUBSTITUTE SHEET (RULE 26)

mat\_peptide 16171..17973  
/gene="ORF1ab"  
/product="helicase"  
  
mat\_peptide 17974..19554  
/gene="ORF1ab"  
/product="3'-to-5' exonuclease"  
  
mat\_peptide 19555..20592  
/gene="ORF1ab"  
/product="endoRNase"  
  
mat\_peptide 20593..21486  
/gene="ORF1ab"  
/product="2'-O-ribose methyltransferase"  
  
CDS 212..13417  
/gene="ORF1ab"  
/codon\_start=1  
/product="ORF1a polyprotein"  
/protein\_id="UJD17628.1"  
  
/translation="MESLVPGFNEKTHVQLSLPVLDQVRDVLVRGFGDSVEEVLSearQ  
HLKDGTCAVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

TLGVILVPHVGEIPVAYRKVLLRKNGNKAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKLDGMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG  
DFVKATCEF CGTENLTKEGATT CGYL PQNAVVKIYCPACHN SEVGPEHSLAEYHNESG  
LKTILRKGGRTIAFGGCVF SYVGCHNK CAYWV PRASANIGCNHTGVV GEGSEG LNDNL  
LEILQKEKVNI NIVGDFKLNEEIAIILASF SASTSAF VETVKGLDYKAFKQ IVE SCGN  
FKVTKGAKKGAWNIGEQKSILSPLYAFASEAARVVR SIFS RTLETAQNSVRVLQKAA  
ITILDGISOYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPPLKAPKEII  
FLEGETLPTEVLTEEVVLKTGDLOPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK  
YCALAPNMMVTNNIFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNER  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF  
KLASHMYCSFYPPDEEEECDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSSQQTVGQQDGSEDNQTTIQTIVEVQPQLEMELTPVVQTIEVNNSFSG  
YLKLTDNVYIKNADIVEEAKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVIDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTLEETKFLTENLLLVIDINGN

*Fig. 12B continued*

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV  
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM  
LAHAEETRKILMPVCVETKAIVSTIQRKYKGIKIQEGVVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATSVSSPDAVTAYNGYLTS  
KTPEEHFIEITISLAGSYKDWSYSQOSTQLGIEFLRGDKSVYYTSNPTTFHLDGEVIT  
FDNLKTLSSLREVRTIKVFTTVDNINLHTQVVDMMSMTYQQQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPNDDTLRVEAFYYHTDPSFLGRYMSALNHTKKWKPQVNGLTSIK  
WADNNCYLATALLTLQQIELKFNPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETMMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ  
IPCTCGKQATKYLVQQESPFVMMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

*Fig. 12B* *continued*

EILYCIDGALLTKSSEYKGPITDVFYKENSYTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT  
FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNAATNKATYKPNTWCIRCLWST  
KPVEITSNSFDVLKSEDAQGMNDNLACEDLKPVSEEVVENPTIQKDVLECNVKTIEVVGD  
IILKPANNIKITEEVGHTDLMAAYVDNSSLTICKPNELSRVLGLKTLATHGLAAVNSV  
PWDTIANYAKPFLNKVVSTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRIK  
ASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLSVCLGSLIYSTAA  
LGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSDLTYPSELTII  
QITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLM  
WLIIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVDGCNSSTCMMCYKRNRATRVEC

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

TTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRPI  
NPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPIN  
VIVFDGKSKEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNT  
FSSTFNVPMEKLKTLVATAEAEELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVECL  
KLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNITLIW  
NVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWLK  
QLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFAN  
KHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLPR  
VFSAVGNICYTPSKLIEYTDFATSACVLAECTIFKDASGKPVPYCYDTNVLEGSVAY  
ESLRPDTRYVLMGSIIQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSGR

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

WVLNNNDYYRSLPGVFCGDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIIVVTCLA  
YYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFLTND  
VSFLAHIQWMVMFTPLVPFWITIAIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEE  
AALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLA  
KALNDFNSGSDVLYQPPQISITSAVLQSGFRKMAFP SGKVEGCMVQVTCGTTLNGL  
WLDDVVYCPRHVIC TSEDMLNP NYEDLLIRKSNNFLVQAGNVQLRVIGHSMQNCVLK  
LKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRHNFТИKGSFLNGSCG  
SVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFDVRQTAQAAGTDTTITVNV  
LAWLYAAVINGDRWF LNRF TTLLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVL  
DMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFOSAVKRTIKGTHWL

*Fig. 12B* *continued*

LLTILTSLLVLVQSTQWSLFFFYENAFLPFAMGI IAMSAFAMMFVKHKHAFLCLFLL  
PSLATVAYFNMVYMPASWVMRIMTWLDMVDTSFKLKDCVMYASAVVLLILMTARTVYD  
DGARRVWTLMNVLTLYKVYYGNALDQAI SMWALIISVTSNYSGVVTTVMFLARGVVF  
MCVEYCPPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQ  
EFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVLLSVL  
QQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSLLSVLLSMQGAVDINKLCEEM  
LDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGSEVVLKKLKSLNVAKSEFD  
RDAAMQRKLEKMAHQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIIN  
NARDGCVPLNIIPLTTAAKLMVIPDYNTYKNCDGIFTYASALWEIQQVVDADSKI  
VQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQACTDD

*Fig. 12B* *continued*

NALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPK

VKYLYFIKGLNNLNRMVILGSLAATVRLQAGNATEVPANSTVLSFCAFADVDAKAYKD

YLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPK

GFCDLKCKYVQIPTTCANDPVGFTLKNTVCTVCCMWKGYCCSCDQLREPMIQSADAQS

FLNGFAV"

mat\_peptide 212..751

/gene="ORF1ab"

/product="leader protein"

mat\_peptide 752..2665

/gene="ORF1ab"

/product="nsp2"

mat\_peptide 2666..8497

/gene="ORF1ab"

/product="nsp3"

mat\_peptide 8498..9997

/gene="ORF1ab"

/product="nsp4"

*Fig. 12B* *continued*

<u>mat_peptide</u>	9998..10915
	/gene="ORF1ab"
	/product="3C-like proteinase"
<u>mat_peptide</u>	10916..11776
	/gene="ORF1ab"
	/product="nsp6"
<u>mat_peptide</u>	11777..12025
	/gene="ORF1ab"
	/product="nsp7"
<u>mat_peptide</u>	12026..12619
	/gene="ORF1ab"
	/product="nsp8"
<u>mat_peptide</u>	12620..12958
	/gene="ORF1ab"
	/product="nsp9"
<u>mat_peptide</u>	12959..13375
	/gene="ORF1ab"
	/product="nsp10"
<u>mat_peptide</u>	13376..13414

*Fig. 12B* *continued*

/gene="ORF1ab"  
/product="nspl1"  
stem\_loop 13410..13437  
/gene="ORF1ab"  
/note="Coronavirus frameshifting stimulation element  
stem-loop 1"  
stem\_loop 13422..13476  
/gene="ORF1ab"  
/note="Coronavirus frameshifting stimulation element  
stem-loop 2"  
gene 21497..25309  
/gene="S"  
CDS 21497..25309  
/gene="S"  
/codon\_start=1  
/product="surface glycoprotein"  
/protein\_id="UJD17629.1"  
/translation="MFVFLVLLPLVSSQCVNLTRTQLPPAYNSFTRGVYPDKVFR  
SSVLHSTQDLFLPFFSNVIWFHVISGTNGIKRFDNPVLPFNDGVYFASIEKSNIIRGW

*Fig. 12B continued*

IFGTTLDSKTQSILLVNNAATNVVIKVCEFQFCNDPFLDHKNNKSWMESEFRVYSSANN  
CTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVREPEDLPQGFS  
ALEPLVDLPIGINITRFQTLLALHRSYLTGDSSSGWTAGAAAYVGYLQPRTFLLKY  
NENGTTDAVDCALDPLSETKCTLKSFTEKGIVQTSNFRVQPTESIVRFPNITNLCP  
FDEVFNATRFASVYAWNKRISNCVADYSVLYNLAPFFTFCYGVSPTKLNDLCFTNV  
YADSFVIRGDEVRQIAPGQTGNIADYNKLPDDFTGCVIAWNSNKLDISKVSGNYNYLY  
RLFRKSNLKPFERDISTEIQAGNKPCNGVAGFNCYFPLRSYSFRPTYGVGHQPYRVV  
VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLKGTGVLTESNKKFLPFQQFGRDI  
ADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQAVLYQGVNCTEVPAIHAD  
QLTPTWRYVYSTGSNVFQTRAGCLIGAEVNNSYECDIPIGAGICASYQTQTKSHRRAR

*Fig. 12B continued*

SUBSTITUTE SHEET (RULE 26)

SVASQSIIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMKTSVDCTMYIC  
GDSTECNSNLLQYGSFCTQLKRALTGIAVEQDKNTQEVAQVKQIYKTPPIKYFGGFN  
FSQILPDPSKPSSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIARADLICAQKFKGLT  
VLPLLTDDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
ENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNHNNAQALNTLVKQLSSKFGAISS  
VLNDIFSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV  
LGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFTIAPAICHDGKAHFPR  
EGVFVSNGTHWFVTQRNFYEPQIITTDNTVSGNCDVVIGIVNNTVYDPLQPELDSFK  
EELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE  
QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLGCCSCGSCCKFDEDDSEPVL  
KGVKLHYT"

*Fig. 12B continued*

SUBSTITUTE SHEET (RULE 26)

gene 25318..26145  
/gene="ORF3a"  
CDS 25318..26145  
/gene="ORF3a"  
/codon\_start=1  
/product="ORF3a protein"  
/protein\_id="UJD17630.1"  
  
/translation="MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFG  
  
WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE  
  
APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY  
  
NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWEKGVKDCVVVLHSYFTSDYYQLYSTQ  
  
LSTDGTVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPITYDEPITTTSVPL"  
  
gene 26170..26397  
/gene="E"  
CDS 26170..26397

*Fig. 12B continued*

/gene="E"  
/codon\_start=1  
/product="envelope protein"  
/protein\_id="UJD17631.1"

/translation="MYSFVSEEIGTLIVNSVLFLA**F**VVFLLVTLAILTALRLCAYCC  
NIVNVSLVKPSFYVYSRVKNLNSSRVPD**L**LV"

gene 26448..27116  
/gene="M"

CDS 26448..27116  
/gene="M"  
/codon\_start=1  
/product="membrane glycoprotein"  
/protein\_id="UJD17632.1"

/translation="MAGSNGTITVEELKKLLEEWNLVIGFLFLTWICLLQ**E**AYANRNR  
FLYI**I**KLIFLWLLWPVTLTCFVLA**A**YRINWITGGIAIAMACLVGLMWLSYFIASFRL  
FARTRS**M**WSFN**P**ETNILLNVPL**H**G**T**ILTRPL**E**SELVIGAVILRGHLRIAGHHLGRCD

*Fig. 12B continued*

IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSDNIA  
LLVQ"

gene 27127..27312  
/gene="ORF 6"  
CDS 27127..27312  
/gene="ORF 6"  
/codon\_start=1  
/product="ORF6 protein"  
/protein\_id="UJD17633.1"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIINKNLSKSL  
TENKYSQQLDEEQPMEID"

gene 27319..27684  
/gene="ORF7a"  
CDS 27319..27684  
/gene="ORF7a"  
/codon\_start=1  
/product="ORF7a protein"  
/protein\_id="UJD17634.1"

*Fig. 12B continued*

/translation="MKIILFLALITLATCELYHYQECVRGTVLLKEPCSSGTYEGNS  
PFHPLADNKALTCEFSTQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIEL  
IVAAIVFITLCFTLKRKTE"

gene 27681..27812  
/gene="ORF7b"  
CDS 27681..27812  
/gene="ORF7b"  
/codon\_start=1  
/product="ORF7b"  
/protein\_id="UJD17635.1"

/translation="MIELSLIDFYLCFLAFLFLVLIMLIIFWFSLELQDHNETCHA"  
gene 27819..28184  
/gene="ORF8"  
CDS 27819..28184  
/gene="ORF8"  
/codon\_start=1  
/product="ORF8 protein"  
/protein\_id="UJD17636.1"

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

/translation="MKFLVFLGIITVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK  
WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYIVSCLPFTINCQEPKLGSLVVRC  
SFYEDFLEYHDVRVVLDFI"  
gene 28199..29449  
/gene="N"  
CDS 28199..29449  
/gene="N"  
/codon\_start=1  
/product="nucleocapsid phosphoprotein"  
/protein\_id="UJD17637.1"  
  
/translation="MSDNGPQNQRNALRITFGGPSDSTGSNQNGGARSKQRRPQGLPN  
NTASWFTALTQHGKEDLKPRGQGVPIINTNSSPDDQIGYYRATRRIRGGDGKMKDLS  
PRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIIGTRNPANNAAIVLQLPQ  
GTTLPKGFYAEGSRGGSQASSRSSRSRNSSRNSTPGSSKRTSPARMAGNGGDAALAL

*Fig. 12B continued*

LLLLDRLNQLESKMSGKGQQQQGQTVKKSAAEASKPRQKRTATKAYNVTQAFGRGP  
EQTQGNFGDQEELIRQGTDYKHWPOIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAI  
KLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKDKKKKADETQALPQRQKKQQTVTLL  
PAADLDDFSKQLQQSMSSADSTQA"

gene 29474..29590  
/gene="ORF10"  
CDS 29474..29590  
/gene="ORF10"  
/codon\_start=1  
/product="ORF10 protein"  
/protein\_id="UJD17638.1"  
/translation="MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT"  
stem\_loop 29525..29560  
/gene="ORF10"  
/note="Coronavirus 3' UTR pseudoknot stem-loop 1"  
stem\_loop 29545..29573  
/gene="ORF10"  
/note="Coronavirus 3' UTR pseudoknot stem-loop 2"

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

stem\_loop 29644..29684  
 ORIGIN /note="Coronavirus 3' stem-loop II-like motif (s2m)"  
 1 agatctgttc tctaaacgaa cttaaaaatc tgtgtggctg tcactcggt gcatgcttag  
 61 tgcactcacg cagtataatt aataactaat tactgtcggt gacaggacac gagtaactcg  
 121 tctatcttct gcaggctgct tacggtttcg tccgtgttgc agccgatcat cagcacatct  
 181 aggaaaaatc cgggtgtgac cgaaaggtaa gatggagagc cttgtccctg gtttcaacga  
 241 gaaaacacac gtccaaactca gtttgccctgt tttacagggtt cgcgacgtgc tcgtacgtgg  
 301 ctttggagac tccgtggagg aggtcttatac agaggcacgt caacatctta aagatggcac  
 361 ttgtggctta gtagaagttt aaaaaggcgt tttgcctcaa cttgaacagc cctatgtgtt  
 421 catcaaacgt tcggatgctc gaactgcacc tcatggtcat gttatggttt agctggtagc  
 481 agaactcgaa ggcattcagt acggtcgtag tggtagagaca cttgggtgtcc ttgtccctca  
 541 tgtggcgaa ataccagtgg cttaccgcaa ggttcttctt cgtaagaacg gtaataaaagg  
 601 agctgggtggc catagttacg ggcgcgatct aaagtcattt gacttaggcg acgagcttgg  
 661 cactgatcct tatgaagatt ttcaagaaaa ctggaacact aaacatagca gtgggtttac  
 721 ccgtgaactc atgcgtgagc ttaacggagg ggcatacact cgctatgtcg ataacaactt  
 781 ctgtggccct gatggctacc ctcttggatgt cattaaagac cttctagcac gtgctggtaa  
 841 agcttcatgc actttgtccg aacaactgga ctttattgac actaagaggg gtgtataactg  
 901 ctgcgcgtgaa catgagcatg aaattgcttgc gtacacggaa cgttctgaaa agagctatga  
 961 attgcagaca ccttttgaaa ttaaattggc aaagaaaattt gacaccttca atggggatg

*Fig. 12B* *continued*

1021 tccaaatttt gtatccct taaattccat aatcaagact attcaaccaa gggttggaaa  
1081 gaaaaagctt gatggcttta tgggtagaat tcgatctgtc tatccagttg cgtcacccaa  
1141 tgaatgcaac caaatgtgcc tttcaactct catgaagtgt gatcattgtg gtgaaacttc  
1201 atggcagacg ggcgattttg ttaaagccac ttgcgaattt tgtggcactg agaatttgac  
1261 taaagaaggt gccactactt gtggtaactt accccaaaat gctgttgtt aaatttattt  
1321 tccagcatgt cacaattcag aagtaggacc tgagcatagt cttgccgaat accataatga  
1381 atctggcttgg aaaaccattc ttcgtaaggg tggtcgcact attgcctttg gaggctgtgt  
1441 gttcttttat gttgggttgcc ataacaagtgt tgccattttgg gttccacgtg ctagcgctaa  
1501 cataggttgt aaccatacag gtgttgggg agaaggttcc gaaggcttta atgacaaccc  
1561 tcttgaaaata ctccaaaaag agaaagtcaa catcaatatt gttggtgact ttaaacttaa  
1621 tgaagagatc gccatttattt tggcatctt ttctgcttcc acaagtgtt ttgtggaaac  
1681 tgtgaaaggt ttggattata aagcattcaa acaaaattgtt gaatcctgtg gtaattttaa  
(1741 agttacaaaa gaaaaagcta aaaaagggtgc ctggatatatt ggtgaacaga aatcaataact  
1801 gagtcctttt tatgcatttg catcagaggc tgctcgtgtt gtacgatcaa ttttctcccg  
1861 cactcttggaa actgctcaaa attctgtgcg tgtttacag aaggccgcta taacaataact  
1921 agatggaaatt tcacagtatt cactgagact cattgatgt atgatgttca catctgattt  
1981 ggctactaac aatctagttg taatggccta cattacaggt ggtgttgttc agttgacttc  
2041 gcagtggcta actaacatct ttggcactgt ttatggaaaaa ctc当地cccg tccttgattt  
2101 gcttgaagag aagtttaagg aaggtgtaga gtttctttaga gacgggtggg aaattgttaa  
2161 atttatctca acctgtgtt gtgaaattgt cggtggacaa attgtcacct gtgcaaagga  
2221 aattaaggqag agtgttcaga cattctttaa qcttqtaaat aaatttttqg ctttqgtqd

SUBSTITUTE SHEET  
(RULE 26)

2281 tgacttotatc attattggtg gagctaaact taaagccttg aatttaggtg aaacatttgc  
2341 cacgcactca aagggattgt acagaaaagtg tgttaaatcc agagaagaaaa ctggcctact  
2401 catgcctcta aaagccccaa aagaaattat cttcttagag ggagaaacac ttcccacaga  
2461 agtgttaaca gaggaagttg tcttgaaaac tggtgattta caaccattag aacaacctac  
2521 tagtgaagct gttgaagctc cattggttgg tacaccagtt tgtattaacg ggcttatgtt  
2581 gctcgaaatc aaagacacag aaaagtactg tgcccttgca cctaataatga tggtaacaaa  
2641 caataccttc acactcaaag gcgggtgcacc aacaaaagggtt acttttggtg atgacactgt  
2701 gatagaagtg caagggttaca agagtgtgaa tatcactttt gaacttgatg aaaggattga  
2761 taaagtactt aatgagaggt gctctgccta tacagttgaa ctcggcacag aagtaaatga  
2821 gttcgctgt gttgtggcag atgctgtcat aaaaactttt caaccagttt ctgaattact  
2881 tacaccactg ggcattgatt tagatgagtg gagtatggct acatactact tatttgatga  
2941 gtctggtgag tttaaattgg cttcacatat gtattgttct ttttaccctc cagatgagga  
3001 tgaagaagaa ggtgattgtg aagaagaaga gtttgagcca tcaactcaat atgagtatgg  
3061 tactgaagat gattaccaag gttaaaccttt ggaatttggt gccacttctg ctgctctca  
3121 acctgaagaa gagcaagaag aagattggtt agatgatgat agtcaacaaa ctggtggca  
3181 acaagacggc agtgaggaca atcagacaac tactattcaa acaatttttg aggttcaacc  
3241 tcaatttagag atggaactta caccagttgt tcagactatt gaagtgaata gtttttagtgg  
3301 ttatTTaaaaa cttaactgaca atgtatacat taaaaatgca gacattgtgg aagaagctaa  
3361 aaaggtaaaaa ccaacagtgg ttgttaatgc agccaatgtt taccttaaac atggaggagg  
3421 tggcagga gccttaata aggctactaa caatgccatg caagttgaat ctgatgatta

Fig. 12B *continued*

SUBSTITUTE SHEET  
(RULE 26)

3481 catagctact aatggaccac ttaaagtgg tgtagttgt gtttaagcg gacacaatct  
 3541 tgctaaacac tgtcttcatg ttgtcgccc aaatgttaac aaaggtaag acattcaact  
 3601 tcttaagagt gtttatgaaa attttaatca gcacgaagtt ctacttgcac cattattatc  
 3661 agctggatt tttggtgctg accctataca ttcttaaga gtttgttag atactgtcg  
 3721 cacaaatgtc tacttagctg tcttgataa aaatctctat gacaaacttg tttcaagctt  
 3781 tttggaaatg aagagtgaaa agcaagttga acaaaagatc gctgagattc ctaaagagga  
 3841 agttaagcca tttataactg aaagtaaacc ttcagttgaa cagagaaaaac aagatgataa  
 3901 gaaaatcaaa gcttgtgtt aagaagttac aacaactctg gaagaaaacta agttccctcac  
 3961 agaaaaacttg ttactttata ttgacattaa tggcaatctt catccagatt ctgccactct  
 4021 tgtagtgac attgacatca ctttcttaaa gaaagatgtc ccatatatag tgggtgtatgt  
 4081 tgttcaagag ggtgtttaa ctgctgttgt tatacctact aaaaaggctg gtggcactac  
 4141 tgaaatgcta gcgaaagctt tgagaaaaagt gccaacagac aattatataa ccacttaccc  
 4201 gggtcaggg ttaaatggtt acactgtaga ggaggcaaag acagtgccta aaaagtgtaa  
 4261 aagtgcctt tacattctac catctattat ctctaattgag aagcaagaaa ttcttggAAC  
 4321 tgttcttgg aatttgcgag aaatgcttgc acatgcagaa gaaacacgca aattaatgcc  
 4381 tgcgtgtgtg gaaactaaag ccatagttc aactatacag cgtaaatata agggtattaa  
 4441 aatacaagag ggtgtggttg attatggtgc tagattttac ttttacacca gtaaaacaac  
 4501 tgtagcgtca cttatcaaca cacttaacga tctaaatgaa actcttgccta caatgccact  
 4561 tggctatgta acacatggct taaatttgg aagaagctgtc cggtatata gatctctcaa  
 4621 agtgcgcagot acagttctg tttcttacc tgatgctgtt acagcgtata atggttatct  
 4681 tacttcttac tctaaaacac ctgaagaaca ttttattgaa accatctcac ttgctggttc

*Fig. 12B*  
*continued*

SUBSTITUTE SHEET (RULE 26)

4741 ctataaaagat tggtcctatt ctggacaatc tacacaacta ggtatagaat ttcttaagag  
4801 aggtgataaaa agtgttatatt acactagtaa tcctaccaca ttcccacctag atggtaagt  
4861 tatcacctt gacaatctt agacacttct ttctttgaga gaagtgagga ctattaagg  
4921 gtttacaaca gtagacaaca ttaacacctt cacgcaagtt gtggacatgt caatgacata  
4981 tggacaacag tttggtccaa cttatttgga tggagctgat gttactaaaa taaaacctca  
5041 taattcacat gaaggtaaaa cattttatgt tttaccta at gatgacactc tacgtgttga  
5101 ggottttgag tactaccaca caactgatcc tagtttctg ggttaggtaca tgtagcatt  
5161 aaatcacact aaaaagtgga aataccaca agttaatgg ttaacttcta ttaaatggc  
5221 agataacaac tgttatcttgc cactgcatt gttaacactc caacaaatag agttgaagtt  
5281 taatccaccc gctctacaag atgcttatta cagagcaagg gctggtgaag cggctaactt  
5341 ttgtgcactt atcttagcct actgtataa gacagtaggt gagttaggtg atgttagaga  
5401 aacaatgagt tacttgttc aacatgccaa tttagattct tgcaaaagag tcttgaacgt  
(5461 ggtgtgtaaa acttgtggac aacagcagac aacccttaag ggtgtagaag ctgttatgt  
5521 catggcaca ctttcttataa aacaatttaa gaaaggtgtt cagatacctt gtacgtgtgg  
5581 taaacaagct acaaaaatatc tagtacaaca ggagtccaccc tttgttatga tgtagcacc  
5641 acctgtcag tatgaactt agcgtggatc atttacttgc gctagtgtt acactgttac  
5701 ttaccagtgt ggtcactata aacatataac ttctaaagaa actttgtatt gcatacgg  
5761 tgcttactt acaaagtct cagaatacaa aggtcctatt acggatgtt tctacaaaga  
5821 aaacagttac acaacaacca taaaaccagt tacttataaa ttggatggtg ttgtttgtac  
5881 agaaattgac cctaagttgg acaattatta taagaaagac aattcttatt tcacagagca

*Fig. 12B* *continued*

Fig. 12B

*continued*SUBSTITUTE SHEET  
(RULE 26)

5941 accaattgat cttgtaccaa accaaccata tccaaacgca agcttcgata attttaagtt  
6001 tgtatgttat aatatcaaat ttgctgatga tttaaaccag ttaactggtt ataagaaaacc  
6061 tgcttcaaga gagcttaaag ttacatttt ccctgactta aatggtgatg tggggctat  
6121 tgattataaaa cactacacac cctctttaa gaaaggagct aaattgttac ataaacctat  
6181 tgtttggcat gttaacaatg caactaataa agccacgtat aaaccaaata cctggtgtat  
6241 acgttgtctt tggagcacaa aaccagttga aacatcaaatt tcgtttgatg tactgaagtc  
6301 agaggacgcg caggaaatgg ataatcttc ctgogaagat ctaaaaccag tctctgaaga  
6361 agtagtggaa aatcctacca tacagaaaaga cgttcttgag tgtaatgtga aaactaccga  
6421 agttgttagga gacattatac ttaaaccagc aaataatata aaaattacag aagaggttgg  
6481 ccacacagat ctaatggctg cttatgtaga caattctagt cttactatta agaaacctaa  
6541 tgaattatct agagtattag gtttggaaac ctttgctact catggtttag ctgctgttaa  
6601 tagtgtccct tgggatacta tagctaatta tgctaaggct tttcttaaca aagttgttag  
6661 tacaactact aacatagttt cacgggtttt aaaccgtgtt tgtactaatt atatgcotta  
6721 tttctttact ttattgctac aattgtgtac ttttactaga agtacaaatt ctagaattaa  
6781 agcatctatg ccgactacta tagcaaagaa tactgttaag agtgcggta aattttgtct  
6841 agaggcttca tttattttt tgaagtcacc taattttctt aaactgataa atattataat  
6901 ttgggtttta ctattaagtg tttgcctagg ttctttaatc tactcaaccg ctgctttagg  
6961 tgtttaatg tctaatttag gcatgccttc ttactgtact ggttacagag aaggctattt  
7021 gaactctact aatgtcacta ttgcaaccta ctgtactggt tctataccctt gtagtgtttg  
7081 tcttagtgggt ttagatttt tagacaccta tccttcttta gaaactatac aaattaccat  
7141 ttcatctttt aaatgggatt taactgctt tggcttagtt gcagagtggc tttggcata

SUBSTITUTE SHEET (RULE 26)

7201 tattctttc actaggaaaa totatgtact tggattggat gcaatcatgc aattgttttt  
 7261 cagctatTTT gcagtacatt ttattAGTAA ttcttggctt atgtggTTA taattaatct  
 7321 tgtacAAATg gccccgattt cagctatggT tagaatgtac atcttcttG catcatTTA  
 7381 ttatgtatgg aaaagttatg tgcatgttgt agacggTTgt aattcatcaa ctgttatgtat  
 7441 gtgttacAAA cgtaataAGAG caacaAGAGT cgaatgtaca actattgtta atggTgttag  
 7501 aaggTCCTT tatgtctatg ctaatggagg taaaggctt tgcaaactac acaattggaa  
 7561 ttgtgttaat tgtgatacat tctgtgctgg tagtacattt attagtgtatg aagttgcgag  
 7621 agacttGTCA ctacAGTTA aaAGACCAAT aaATCCTACT gaccAGTCTT cttacatcgt  
 7681 tgatagtgtt acagtGAAGA atggTTccat ccatCTTAc tttgataAAAG ctggTcaaaa  
 7741 gacttatgaa agacattCTC tctctcattt tgttaactta gacaACCTGA gagctaataa  
 7801 cactaaAGGT tcattgccta ttaatgttat agttttgtat ggtaaatcaa aatgtgaaga  
 7861 atcatctgca aaatcAGCGT ctgtttacta cagtcagctt atgtgtcaac ctataactgtt  
 7921 actagatcag gcattAGTGT ctgatgttgG tgatagtgcg gaagttgcag ttAAAATgtt  
 7981 tgatgCTTAC gttaatacgt tttcatcaac ttttaacgta ccaatggAAA aactcaAAAC  
 8041 actagttgca actgcagaAG ctgaacttgc aaagaatgtg tccttagaca atgtcttAtc  
 8101 tacttttatt tcagcagCTC ggcaagggtt tgTTgattca gatgtagAAA ctAAAGATgt  
 8161 tgTTgaatgt cttaaATTgt cacatcaATC tgacatAGAA gttactggcg atagttgtAA  
 8221 taactatATG ctCACCTATA acaaAGTTGA aaACATGACA ccccgtgacc ttggTgCTTG  
 8281 tattgactgt agtgcgcgtc atattaatgc gcaggtAGCA AAAAGTCACA acattacttt  
 8341 gatatggaac gttaaAGATT tcatgtcatt gtctgaacaa ctacgAAAAC aaatacgtAG

Fig. 12B *continued*

SUBSTITUTE SHEET  
(RULE 26)

8401 tgctgctaaa aagaataact tacctttaa gttgacatgt gcaactacta gacaagttgt  
8461 taatgttgta acaacaaaaga tagcacttaa gggtggtaaa attgttaata attggttgaa  
8521 gcagtttaatt aaagttcacac ttgtgttcct ttttggct gctattttct atttaataaac  
8581 acctgttcat gtcatgtcta aacatactga ctttcaagt gaaatcatag gatacaaggc  
8641 tattgatgg ggtgtcactc gtgacatagc atctacagat acttggtttg ctaacaaaca  
8701 tgctgattt gacacatggt ttagccagcg tggtagt tataactaatg acaaagctg  
8761 cccattgatt gctgcagtca taacaagaga agtgggtttt gtcgtgcctg gttgcctgg  
8821 cacgatatta cgoacacaacta atgggtgactt tttgcatttc ttacctagag ttttagtgc  
8881 agttggtaac atctgttaca caccatcaaa acttata>tag tacactgact ttgcaacatc  
8941 agcttgtgtt ttggctgctg aatgtacaat ttttaaagat gcttctggta agccagtagcc  
9001 atattgttat gataccaatg tactagaagg ttctgttgct tatgaaagtt tacgccctga  
9061 cacacgttat gtgctcatgg atggctctat tattcaattt cctaacacacct accttgaagg  
9121 ttctgttaga gtggtaacaa cttttgattc tgagtactgt aggcacggca cttgtgaaag  
9181 atcagaagct ggtgtttgtg tatctactag tggtagatgg gtacttaaca atgattatta  
9241 cagatcttta ccaggagttt tctgtgggtgt agatgctgta aatttactta ctaatatgtt  
9301 tacaccacta attcaaccta ttggtgctt ggacatataca gcatctatacg tagctgggtgg  
9361 tatttagtct atcgttagtaa catgccttgc ctactattt atgaggtttta gaagagctt  
9421 tggtaataac agtcatgttag ttgcctttaa tactttacta ttccttatgt cattcactgt  
9481 actctgttta acaccagttt actcattttt acctgggtgtt tattctgtta tttacttgc  
9541 cttgacattt tatcttacta atgatgttcc ttttttagca catattcagt ggatggttat  
9601 gttcacacacctt ttagtaccc ttggataac aattgcttac atcatttgc tttccacaaa

*Fig. 12B continued*

SUBSTITUTE SHEET (RULE 26)

9661 gcatttctat tggttcttta gtaattacct aaagagacgt gtagtcttta atggtgttcc  
9721 ctttagtact tttgaagaag ctgcgctgtg caccttttg ttcaaataaaag aaatgttatct  
9781 aaagttgcgt agtcatgtgc tattacctct tacgcaatat aatacgatact tagctcttta  
9841 taataagtac aagtatttta gtggagcaat ggatacaact agctacagag aagctgcttg  
9901 ttgtcatctc gcaaaggctc tcaatgactt cagtaactca gggtctgatg ttctttacca  
9961 accaccacaa atctctatca cctcagctgt tttgcagagt gggtttqaau aaatggcatt  
10021 cccatctggt aaagttgagg gttgtatggt acaagtaact tgtggtacaa ctacacttaa  
10081 cggtctttgg cttgatgacg tagttactg tccaagacat gtgatctgca cctctqaaaga  
10141 catgcttaac cctaattatg aagattact cattcgtaag tctaattcata atttcttggt  
10201 acaggctggt aatgttcaac tcagggttat tggacattct atgcaaaatt gtgtacttaa  
10261 gcttaagggt gatacagcca atcctaagac acctaagtat aagtttgttc gcattcaacc  
10321 aggacagact ttttcagtgt tagcttggta caatggttca ccatctggtg tttaccaatg  
10381 tgctatgagg cacaatttca ctattaaggg ttcattcctt aatggttcat gtggtagtgt  
10441 tggtttaac atagattatg actgtgtctc tttttgttac atgcaccata tggaaattacc  
10501 aactggagtt catgctggca cagacttaga aggttaactt tatggacctt ttgttgacag  
10561 gcaaacagca caagcagctg gtacggacac aactattaca gttaatgtt tagcttggtt  
10621 gtacgctgt gttataaaatg gagacaggtg gtttctcaat cgatttacca caactcttaa  
10681 tgactttaac cttgtggcta tgaagtacaa ttatgaacct ctaacacaag accatgttga  
10741 catacttagga cttttctg ctc当地actgg aattgccgtt ttagatatgt gtgcttcatt  
10801 aaaagaatata ctgcaaaaatg gtatgaatgg acgtaccata ttgggttagtg ctttattaga

Fig. 12B *continued*

*Fig. 12B*  
*continued*

SUBSTITUTE SHEET  
(RULE 26)

10861 agatgaattt acacctttg atgttgttag acaatgctca ggtgttactt tccaaagtgc  
 10921 agtggaaaaga acaatcaagg gtacacacca ctgggtgtta ctcacaattt tgacttcact  
 10981 tttagttta gtccagagta ctcaatggc tttgttctt ttttgtatg aaaatgcctt  
 11041 tttagcttt gctatggta ttattgctat gtctgcttt gcaatgatgt ttgtcaaaca  
 11101 taagcatgca tttctctgtt tgttttgtt acottctctt gccactgtag ottattttaa  
 11161 tatggtctat atgcctgcta gttgggtgat gcgtattatg acatggttgg atatgggtga  
 11221 tactagttt aagctaaaag actgtgttat gtatgcatca gctgttagtgt tactaatcct  
 11281 tatgacagca agaactgtgt atgatgatgg tgctaggaga gtgtggacac ttatgaatgt  
 11341 cttgacactc gttataaaag tttattatgg taatgctta gatcaagcca tttccatgtg  
 11401 ggctcttata atctctgtta cttctaacta ctcaggtgta gttacaactg tcatttttt  
 11461 ggccagaggt gttgtttta tgtgtgttgta gtattgccct attttcttca taactggtaa  
 11521 tacacttcag tgtataatgc tagtttattt tttcttaggc tattttgtt cttgttactt  
 11581 tggcctctt tgtttactca accgctactt tagactgact cttgggtttt atgattactt  
 11641 agtttctaca caggagttta gatatatgaa ttcacaggga ctactcccac ccaagaatag  
 11701 catagatgcc ttcaaactca acattaaatt gttgggtttt ggtggcaaacc cttgtatcaa  
 11761 agtagccact gtacagtcta aaatgtcaga tgtaaagtgc acatcagtag tcttactctc  
 11821 agtttgcaa caactcagag tagaatcatc atctaaattt ggggtcaat gtgtccagtt  
 11881 acacaatgac attctcttag ctaaagatac tactgaagcc tttgaaaaaaa tggtttact  
 11941 actttctgtt ttgctttcca tgcagggtgc tgttagacata aacaagctt gtgaagaaat  
 12001 gctggacaac agggcaacct tacaagctat agcctcagag tttagttccc ttccatcata  
 12061 tgcagcttt gctactgctc aagaagctt tgagcaggct gttgctaattt gtgattctga

SUBSTITUTE SHEET (RULE 26)

12121 agttgttctt aaaaagttga agaagtcttt gaatgtggct aaatctgaat ttgaccgtga  
 12181 tgcagccatg caacgtaagt tggaaaagat ggctgatcaa gctatgaccc aaatgtataa  
 12241 acaggctaga tctgaggaca agagggcaaa agttactagt gctatgcaga caatgcttt  
 12301 cactatgctt agaaaagttgg ataatgatgc actcaacaac attatcaaca atgcaagaga  
 12361 tggttgtgtt cccttgaaca taatacctct tacaacagca gccaaactaa tggttgtcat  
 12421 accagactat aacacatata aaaatacgtg tgatggtaca acatttactt atgcatcagc  
 12481 attgtggaa atccaacagg ttgttagatgc agatagtaaa attgttcaac ttagtgaat  
 12541 tagtatggac aattcaccta atttagcatg gcctcttatt gtaacagtt taagggccaa  
 12601 ttctgctgtc aaattacaga ataatgagct tagtcctgtt gcactacgac agatgtcttg  
 12661 tgctgccggt actacacaaa ctgcttgcac tgatgacaat gcgttagtt actacaacac  
 12721 aacaaaggga ggtaggttt tacttgcact gttatccgat ttacaggatt tgaaatggc  
 12781 tagattccct aagagtgtg gaactggtagt tatctataca gaactggAAC caccttgttag  
 12841 gtttgttaca gacacaccta aaggtcctaa agtgaagtat ttatacttta ttaaaggatt  
 12901 aaacaaccta aatagaggta tggtaacttgg tagtttagct gccacagtac gtctacaagg  
 12961 tggtaatgca acagaagtgc ctgccaattc aactgttatta tctttctgtg cttttgtgt  
 13021 agatgtgtct aaagcttaca aagattatct agcttagtggg ggacaaccaa tcactaattt  
 13081 tgttaagatg ttgtgtacac acactggtagtac tggtcaggca ataacagtca caccggaaagc  
 13141 caatatggat caagaatctt ttgggtggtagt acgtgttgtt ctgtactgcc gttgccacat  
 13201 agatcatcca aatcctaaaag gatTTTGTGA cttaaaaggta aagtatgtac aaatacctac  
 13261 aacttgtgtt aatgaccctg tgggttttac actaaaaaac acagtctgtt ccgtctgcgg

*Fig. 12B continued*

*Fig. 12B*  
*continued*

SUBSTITUTE SHEET (RULE 26)

13321 tatgtggaaa ggtttatggct gtagtgtga tcaactccgc gaacccatgc ttcagtcagc  
 13381 tcatgcacaa tcgttttaa acggggttgc ggtgtaagtg cagcccgtct tacaccgtgc  
 13441 ggcacaggca ctagtactga tgtcgatac agggctttg acatctacaa tgataaaagta  
 13501 gctggtttg ctaaattcct aaaaactaat tgggtcgct tccaagaaaa ggacgaagat  
 13561 gacaatttaa ttgattctta cttttagtt aagagacaca ctttctctaa ctaccaacat  
 13621 gaagaaaacaa ttataattt acttaaggat tgtccagctg ttgctaaaca tgacttctt  
 13681 aagtttagaa tagacggtga catggcacca catatatcac gtcaacgtct tactaaatac  
 13741 acaatggcag acctcgctca tgcttaagg cattttgatg aaggttaattg tgacacatta  
 13801 aaagaaaatac ttgtcacata caattgtgt gatgatgatt atttcaataa aaaggactgg  
 13861 tatgattttg tagaaaaccc agatatatta cgctgtatacg ccaacttagg tgaacgtgta  
 13921 cgccaagctt tgtaaaaaac agtacaattc tgtgatgcca tgcgaaatgc tggattgtt  
 13981 ggtgtactga cattagataa tcaagatctc aatggtaact ggtatgattt cggtgatttc  
 14041 atacaaaacca cgccaggtag tggagttcct gtttagattt ctttattttc attgttaatg  
 14101 cctatattaa ctttgaccag ggcttaact gcagagtcac atgttgacac tgacttaaca  
 14161 aagccttaca ttaagtggta ttgttaaaa tatgacttca cggaagagag gttaaaactc  
 14221 tttgaccgtt attttaataa ttgggatcag acataccacc caaattgtgt taactgtttg  
 14281 gatgacagat gcattctgca ttgtgcaaacc tttaatgttt tattctctac agtgttccca  
 14341 cttacaagtt ttggaccact agtgagaaaa atatttgttg atggtgttcc attttagtt  
 14401 tcaactggat accacttcag agagcttagt gttgtacata atcaggatgt aaacttacat  
 14461 agctctagac ttagtttaa ggaattactt gtgtatgctg ctgaccctgc tatgcacgct  
 14521 gcttctggta atctattact agataaacgc actacgtgct tttcagtagc tgcacttact

SUBSTITUTE SHEET  
(RULE 26)

14581 aacaatgttg cttttcaaacc tgtcaaaaccc ggtaattttt aaaaagactt ctatgacttt  
14641 gctgtgtcta agggtttctt taaggaagga agttctgttg aattaaaaca cttcttcttt  
14701 gctcaggatg gtaatgctgc tatcagcgat tatgactact atcgttataa tctaccaaca  
14761 atgtgtgata tcagacaact actatttgta gttgaagttg ttgataagta ctttgattgt  
14821 tacgatggtg gctgttattaa tgctaaccaa gtcatcgta acaacctaga caaatcagct  
14881 ggaaaaatcat ttaataaaatg gggtaaggct agactttatt atgattcaat gagttatgag  
14941 gatcaagatg cactttcgc atatacaaaa cgtaatgtca tccctactat aactcaaattg  
15001 aatcttaagt atgccatttag tgcaaagaat agagctcgca ccgttagctgg tgtctctatc  
15061 tgttagtacta tgaccaatacg acagtttcat caaaaattat tgaaatcaat agccgccact  
15121 agaggagcta ctgttagtaat tggaaacaagg aaattctatg gtgggtggca caatatgtta  
15181 aaaactgttt atagtgtatgt agaaaaaccct caccttatgg gttgggatta tcctaaatgt  
15241 gatagagcca tgcctaacat gcttagaaatt atggcctcac ttgttcttgc tcgcaaacat  
15301 acaacgtgtt gtagcttgcc acaccgttcc tatagattt ctaatgagt tgctcaagta  
15361 ttgagtgaaa tggcatgtg tggcggttca ctatatgtt aaccagggtgg aacctcatca  
15421 ggagatgcca caactgctt tgctaatagt gtttttaaca ttgtcaagc tgtcaacggcc  
15481 aatgttaatg cacttttatac tactgatggt aacaaaattt ccgataagta tgtccgcaat  
15541 ttacaacaca gactttatga gtgtctctat agaaaatagag atgttgacac agactttgtg  
15601 aatgagttt acgcatatcc gcgtaaacat ttctcaatga tgatacttc tgacgatgt  
15661 gttgtgtgtt tcaatagcac ttatgcacat caaggtctag tggctagcat aaagaacttt  
15721 aagtcaatgttcc ttattatca aaacaatgtt tttatgtctg aagcaaaatg ttggactgag

*Fig. 12B continued*

SUBSTITUTE SHEET

(RULE

26)

15781 actgacctta ctaaaggacc tcatgaattt tgctctcaac atacaatgct agttaaacag  
15841 ggtgatgatt atgtgtacct tccttaccca gatccatcaa gaatcctagg ggccggctgt  
15901 tttgttagatg atatcgtaaa aacagatggt acacttatga ttgaacggtt cgtgtcttta  
15961 gctatagatg cttacccact tactaaacat cctaattcagg agtatgctga tgtctttcat  
16021 ttgtacttac aatacataag aaagctacat gatgagttaa caggacacat gttagacatg  
16081 tattctgtta tgcttactaa tgataaacact tcaaggtatt gggAACCTGA gttttatgag  
16141 gctatgtaca caccgcatac agtcttacag gctgttgggg ctgtgttct ttgcaattca  
16201 cagacttcat taagatgtgg tgottgcata cgtagaccat tcttatgttg taaatgctgt  
16261 tacgaccatg tcatatcaac atcacataaa ttagtcttgt ctgttaatcc gtatgtttgc  
16321 aatgctccag gttgtgatgt cacagatgtg actcaacttt acttaggagg tatgagctat  
16381 tattgtaaat cacataaacc acccattagt tttccattgt gtgctaattgg acaagtttt  
16441 ggTTTATATA AAAATACATG TGTTGGTAGC GATAATGTtA CTGACTTTAA TGCAATTGCA  
16501 acatgtgact ggacaaatgc tggtgattac atttttagcta acacctgtac tgaaagactc  
16561 aagcttttg cagcagaaac gctcaaagct actgaggaga catTTAAACT gtcttatgg  
16621 attgctactg tacgtgaagt gctgtctgac agagaattac atctttcatg ggaagttgg  
16681 aaacctagac caccacttaa ccgaaattat gtctttactg gtatcgtgt aactaaaaac  
16741 agtaaaagtac aaataggaga gtacaccttt gaaaaagggtg actatggtga tgctgttgg  
16801 taccgaggta caacaactta caaattaaat gttggtgatt attttgtgct gacatcacat  
16861 acagtaatgc cattaagtgc acctacacta gtgccacaag agcactatgt tagaattact  
16921 ggcttataacc caacactcaa tatctcagat gagttttcta gcaatgttgc aaattatcaa  
16981 aaggttggta tgcaaaaagta ttctacactc cagggaccac ctggtaactgg taagagtcat

Fig. 12B  
continued

SUBSTITUTE SHEET  
(RULE 26)

17041 tttgcttattg gccttagctct ctactaccct tctgctcgca tagtgtatac agottgotct  
17101 catgccgctg ttgatgcact atgtgagaag gcattaaaat atttgcctat agataaatgt  
17161 agtagaatta tacctgcacg tgctcgtgta gagtgtttg ataaattcaa agtgaattca  
17221 acattagaac agtatgtctt ttgtactgta aatgcattgc ctgagacgac agcagatata  
17281 gttgtcttg atgaaatttc aatggccaca aattatgatt tgagtgtgt caatgccaga  
17341 ttacgtgcta agcactatgt gtacattggc gaccctgctc aattacctgc accacgcaca  
17401 ttgctaacta egggcacact agaaccagaa tatttcaatt cagtgtag acttatgaaa  
17461 actataggc cagacatgtt cctcggaact tgtcggttt gtccctgctga aattgttgac  
17521 actgtgagtg ctgggttta tgataataag cttaaagcac ataaagacaa atcagctcaa  
17581 tgctttaaaa tggtttataa ggggtttatc acgcatgatg tttcatctgc aattaacagg  
17641 ccacaaaatag gcgtggtaag agaattcctt acacgtAAC ctgcttggag aaaagctgtc  
17701 tttatttcac ctataattc acagaatgct gtagcctcaa agatTTGGG actaccaact  
17761 caaactgttg attcatcaca gggctcagaa tatgactatg tcataattcac tcaaaccact  
17821 gaaacagctc actcttgtaa tggtaaacaga tttaatgtt ctattaccag agcaaaaagta  
17881 ggcataacttt gcataatgtc tgatagagac ctttatgaca agttgcaatt tacaagtctt  
17941 gaaattccac gtaggaatgt ggcaacttta caagctgaaa atgtAACAGG actctttaaa  
18001 gattgttagta aggttaatcac tgggttacat cctacacagg cacctacaca cctcagtgtt  
18061 gacactaaat tcaaaaactga aggTTTATGT gttgacgtac ctggcatacc taaggacatg  
18121 acctatagaa gactcatctc tatgatgggt ttAAAATGA attatcaagt taatggttac  
18181 cctaacatgt ttatcaccgg cgaagaagct ataagacatg tacgtgcattt gattggctt

*Fig. 12B continued*

Fig. 12B  
continued

18241 gatgtcgagg ggtgtcatgc tactagagaa gctgttggta ccaatttacc tttacagcta  
18301 ggaaaaatc caggtgttaa cctagttgct gtacctacag gttatgttga tacacctaatt  
18361 aatacagatt tttccagagt tagtgctaaa ccaccgcctg gagatcaatt taaacacaccc  
18421 ataccactta tgtacaaagg acttccttgg aatgttagtgc gtataaaagat tgtacaaatg  
18481 ttaagtgaca cactaaaaaa tctctctgac agagtcgtat ttgtcttatg ggcacatggc  
18541 tttgagttga catctatgaa gtattttgtg aaaataggac ctgagcgcac ctgttgtcta  
18601 tgtgatagac gtgccacatg cttttccact gcttcagaca cttatgcctg ttggcatcat  
18661 tctattggat ttgattacgt ctataatccg tttatgattg atgttcaaca atggggtttt  
18721 acaggttaacc tacaaagcaa ccatgatctg tattgtcaag tccatggtaa tgcacatgta  
18781 gctagttgtg atgcaatcat gactaggtgt ctagctgtcc acgagtgtt tgttaagcgt  
18841 gttgactgga ctattgaata tcctataatt ggtgatgaac tgaagattaa tgccggcttgt  
18901 agaaagggttc aacacatggt tgttaaagct gcattattag cagacaaatt cccagttctt  
(RULE 26) 18961 cacgacattg gtaaccctaa agctattaag tgtgtacctc aagctgatgt agaatggaag  
19021 ttctatgtg cacagccttg tagtgacaaa gcttataaaa tagaagaatt attctattct  
19081 tatgccacac attctgacaa attcacagat ggtgtatgcc tattttggaa ttgcaatgtc  
19141 gataagatatc ctgctaattc cattgtttgt agatttgaca ctagagtgtc atctaaccctt  
19201 aacttgcctg gttgtgatgg tggcagtttgc tatgtaaata aacatgcatt ccacacacca  
19261 gctttgata aaagtgcctt tgttaattta aaacaattac cattttcttgc ttactctgac  
19321 agtccatgtg agtctcatgg aaaacaagta gtgtcagata tagattatgt accactaaag  
19381 tctgctacgt gtataacacg ttgcaattta ggtggtgctg tctgttagaca tcattgttaat  
19441 gagtacagat tgtatctcga tgcttataac atgatgatct cagctggctt tagttgtgg

SUBSTITUTE SHEET

(RULE 26)

19501 gtttacaaac aatttgatac ttataaacctc tggaacactt ttacaagact tcagagttta  
 19561 gaaaatgtgg ctttaatgt tggaaataag ggacactttg atggacaaca gggtaagta  
 19621 ccagtttcta tcattaataa cactgttac acaaaagttg atgggtttga tgtagaattg  
 19681 tttgaaaata aaacaacatt acctgttaat gtagcattt agctttggc taagcgcaac  
 19741 attaaaccag taccagaggt gaaaatactc aataatttg gtgtggacat tgctgtaat  
 19801 actgtgatct gggactacaa aagagatgct ccagcacata tatctactat tggtgtttgt  
 19861 tctatgactg acatagccaa gaaaccaact gaaacgattt gtgcaccact cactgtctt  
 19921 tttgatggta gagttgatgg tcaagtagac ttattnagaa atgcccgtaa tggtgtttctt  
 19981 attacagaag gtagtgttaa aggttacaa ccatctgttag gtcccaaaca agctagtctt  
 20041 aatggagtca cattaattgg agaagccgt aaaaacacagt tcaattatta taagaaagtt  
 20101 gatgggttg tccaacaatt acctgaaact tactttactc agagtagaaaa tttacaagaa  
 20161 tttaaaccca ggagtcaaattt ggaaattgat ttcttagaat tagctatggta tgaatttcatt  
 20221 gaacggtata aattagaagg ctatgccttc gaacatatcg tttatggaga ttttagtcat  
 20281 agtcagttag gtggtttaca tctactgatt ggactagcta aacgtttaa ggaatcacct  
 20341 tttgaatttag aagattttat tcctatggac agtacagttt aaaaactatcc cataacagat  
 20401 gcgcaaacag qtcatctaa gtgtgtgtgt tctgttattt atttattact tgatgatttt  
 20461 gttgaaataa taaaatccca agatttatct gtagtttcta aggttgtcaa agtactatt  
 20521 gactatacag aaatttcatt tatgctttgg tgtaaagatg gccatgtaga aacatttac  
 20581 ccaaaattac aatctagtca agcgtggcaa cgggtgttg ctatgcotaa tctttacaaa  
 20641 atgcaaagaa tgctattaga aaagtgtgac cttcaaaattt atgggtgatag tgcaacattt  
 20701 cctaaaggca taatgatgaa tgtcgaaaa tataactcaac tgtgtcaata tttaaacaca

*Fig. 12B*  
*continued*

SUBSTITUTE SHEET  
(RULE 26)

20761 ttaacattag ctgtacccta taatatgaga gttatacatt ttgggtgtgg ttctgataaa  
20821 ggagttgcac caggtacagc tgtttaaga cagtggttgc ctacgggtac gctgcttgcc  
20881 gattcagatc ttaatgactt tgtctctgat gcagattcaa ctttgattgg tgattgtgca  
20941 actgtacata cagctaataa atgggatttc attatttagtg atatgtacga ccctaagact  
21001 aaaaatgtta caaaagaaaa tgactctaaa gagggtttt tcacttacat ttgtgggtt  
21061 atacaacaaa agcttagctct tggagggtcc gtggctataa agataaoaga acattttgg  
21121 aatgctgatc ttataaagct catggacac ttgcgtatggt ggacagoctt tgtaactaat  
21181 gtgaatgcgt catcatctga agcattttta attggatgta attatcttgg caaaccacgc  
21241 gaacaaaatag atggttatgt catgcatgca aattacatat tttggaggaa tacaaatcca  
21301 attcagttgt cttccttattc tttatggac atgagtaaat ttcccccttaa attaagggggt  
21361 actgctgtta tgtctttaaa agaaggtaaa atcaatgata tgattttatc ttttttttgt  
21421 aaaggttagac ttataattag agaaaacaac agagttgtta ttcttagtga tgttcttgg  
21481 aacaactaaa cgaacaatgt ttgtttttct tgttttattg ccactagtct ctagtcagtg  
21541 tgtaatctt acaaccagaa ctcaattacc ccctgcatac actaattctt tcacacgtgg  
21601 tgtttattac cctgacaaaag tttcagatc ctcatggatc cattcaactc aggacttgg  
21661 cttaccttc tttccaaatg ttacttggtt ccattttatc tctgggacca atggtaactaa  
21721 gaggtttgat aaccctgtcc taccatttaa tgatgggttt tattttgttt ccattgagaa  
21781 gtctaacata ataagaggct ggattttgg tactacttta gattcgaaga cccagtcct  
21841 acttattgtt aataacgcta ctaatgttgc tattaaagtc tgtgaatttc aattttgtaa  
21901 tgatccattt ttggaccaca aaaacaacaa aagttggatg gaaagtgagt tcagagtttta

*Fig. 12B* *continued*

SUBSTITUTE SHEET (RULE 26)

21961 ttcttagtgcg aataattgca cttttgaata tgtctctcag cttttctta tggaccttga  
22021 aggaaaaacag ggtaatttca aaaatcttag ggaatttgtg tttaagaata ttgatggta  
22081 tttaaaata tattctaagg acacgcctat tatagtgcgt gagccagaag atctccctca  
22141 gggttttcg gctttagaac cattggtaga tttgccaaata ggtattaaca tcacttaggtt  
22201 tcaaacttta ctgccttac atagaagtta tttgactcct ggtgatttt cttcagggtt  
22261 gacagcttgt gctgcagctt attatgtggg ttatcttcaa cctaggactt ttctattaa  
22321 atataatgaa aatggAACCA ttacagatgc ttagactgt gcacttgacc ctctctcaga  
22381 aacaaaagtgt acgttggaaat cttcaactgt agaaaaagga atctatcaaa cttctaaactt  
22441 tagagtccaa ccaacagaat ctattgttag atttcctaattt attacaaact tggcccttt  
22501 tgatgaagtt tttaacgcca ccagatttgc atctgtttat gcttggaaaca ggaagagaat  
22561 cagcaactgt gtgtctgatt attctgtcct atataatctc gcaccatttt tcactttaa  
22621 gtgttatgga gtgtctccta ctaaattaaa tgatctctgc ttactaatg tctatgcaga  
22681 ttcatTTGTA attagaggtg atgaagtcaG acaaATCGt ccagggcaaa ctggaaatat  
22741 tgctgattat aattataaaat taccagatga ttttacaggc tgcttataAG cttggaaattc  
22801 taacaagctt gattctaagg ttagtggtaa ttataattac ctgtatagat tgTTTtaggaa  
22861 gtctaatttc aaaccttttG agagagatAT ttcaactgaa atctatcagg ccggtaacaa  
22921 accttgtaat ggtgttgcag gttttatttG ttactttcct ttacgatcat atagtttccg  
22981 acccaattat ggtgttggtc accaaccata cagagtagta gtacttttt ttgaacttct  
23041 acatgcacca gcaactgttt gtggacctaa aaagtctact aatttggta aaaacaaaatg  
23101 tgtaatttc aacttcaatg gttaaaagg cacaggttt ctactgagt ctaacaaaaaa

*Fig. 12B continued*

*Fig. 12B*  
*continued*

SUBSTITUTE SHEET  
 (RULE 26)

```

23161 gtttctgcct ttccaaacaat ttggcagaga cattgctgac actactgatg ctgtccgtga
23221 tccacagaca cttgagattc ttgacattac accatgttct tttgggtggtg tcagtgttat
23281 aacaccagga acaaatactt ctaaccagggt tgctgttctt tatcagggtg ttaactgcac
23341 agaagtccct gttgctattc atgcagatca acttactcct acttggcgtg tttattctac
23401 aggttctaattt gtttttcaaa cacgtgcagg ctgtttaata ggggctgaat atgtcaacaa
23461 ctcatatgag tgtgacatac ccattggtgc aggtatatgc gctagttatc agactcagac
23521 taagtctcat cggcgggcac gtagtgttagc tagtcaatcc atcattgcct acactatgtc
23581 acttggtgca gaaaattcag ttgcttactc taataactct attgccatac ccacaaattt
23641 tactatttagt gttaccacag aaattctacc agtgtctatg accaagacat cagtagattg
23701 tacaatgtac atttgtggtg attcaactga atgcagcaat cttttgttgc aatatggcag
23761 ttttgtaca caattaaaac gtgcttaac tggaatagct gttgaacaag aaaaaaacac
23821 ccaagaagtt ttgcacaaag tcaaacaaat ttacaaaaca ccaccaatta aatattttgg
23881 tggtttaat tttcacaaa tattaccaga tccatcaaaa ccaagcaaga ggtcattttat
23941 tgaagatcta ctttcaaca aagtgacact tgcagatgct ggcttcatca aacaatatgg
24001 tgattgcctt ggtgatattg ctgctagaga cctcatttgc gcacaaaagt ttaaaaggcct
24061 tactgtttg ccaccttgc tcacagatga aatgattgct caatacactt ctgcactgtt
24121 agcgggtaca atcaattctg gttggacctt tgggtgcaggt gctgcattac aaataccatt
24181 tgctatgcaa atggcttata ggttaatgg tattggagtt acacagaatg ttctctatga
24241 gaaccaaaaa ttgattgcc accaattaa tagtgcatt ggcaaaaattc aagactcact
24301 ttcttccaca gcaagtgcac ttggaaaact tcaagatgtg gtcaaccata atgcacaagg
24361 tttaaacacg cttgttaaac aacttagctc caaatttggt gcaatttcaa gtgtttaaa

```

SUBSTITUTE SHEET (RULE 26)

24421 tgatatctt tcacgtcttgc acaaaagttga ggctgaagtgc caaattgata ggttgatcac  
 24481 aggcagactt caaagttgc agacatatgt gactcaacaa ttaatttagag ctgcagaaat  
 24541 cagagcttct gctaattttgc ctgctactaa aatgtcagag tgtgtacttg gacaatcaaa  
 24601 aagagtttat ttttgtggaa agggotatca tcttatgtcc ttccctcagt cagcacctca  
 24661 tggtgttagtc ttcttgcatt tgacttatgt ccctgcacaa gaaaagaact tcacaactgc  
 24721 tcctgccatt tgcattgtatg gaaaagcaca ctttcctcggt gaagggtgtct ttgtttcaaa  
 24781 tggcacacac tggtttgtaa cacaaaggaa ttttatgaa ccacaaatca ttactacaga  
 24841 caacacattt gtgtctggta actgtatgt tgcataatagga attgtcaaca acacagtttca  
 24901 tgatccttgc caacctgaat tagattcatt caaggaggag tttagataaat atttttaagaa  
 24961 tcatacatca ccagatgttgc atttaggtga catctctggc attaatgtt cagttgtaaa  
 25021 cattcaaaaaa gaaattgacc gcctcaatga gggtgcacaa aattttaaatg aatctctcat  
 25081 cgatctccaa gaacttggaa agtatgagca gtatataaaa tggccatggt acatttggct  
 25141 aggtttata gctggcttgc ttgcataatgt aatggtgaca attatgttt gctgtatgac  
 25201 cagttgtgtt agttgtctca agggctgttgc ttcttgcgttgc aattttatgt  
 25261 agacgactct gagccagtgc tcaaaggagt caaattacat tacacataaaa cgaactttatg  
 25321 gatttgcattt tgagaatctt cacaattggaa actgtatgtt tgaagcaagg tgaaatcaag  
 25381 gatgctactc cttcagattt tggtcgcttgc actgcaacgc taccgataca agcctcactc  
 25441 ctttcggat ggcttattgt tggcgttgc cttcttgcgttgc tttttcagag cgcttccaaa  
 25501 atcataactc tcaaaaaagag atggcaacta gcaatctcca agggtgttca ctttgc  
 25561 aacttgcgttgc tggtgtttgtt aacagtttac tcacacaccc ttgcgttgc tgctggcctt

Fig. 12B *continued*

*Fig. 12B*  
continued

SUBSTITUTE SHEET  
(RULE 26)

```

25621 gaagcccctt ttctctatct ttatgcttta gtctacttct tgcaaggtat aaactttgtat
25681 agaataataa tgaggctttg gctttgctgg aaatgccgtt ccaaaaaaccc attactttat
25741 gatgccaact attttctttg ctggcatact aattgttacg actattgtat accttacaat
25801 agtgtaactt cttcaattgt cattacttca ggtgatggca caacaagtcc tatttctgaa
25861 catgactacc agattggtgg ttatactgaa aaatggaaat ctggagtaaa agactgtgtt
25921 gtattacaca gttacttcac ttcagactat taccagctgt actcaactca attgagtaaca
25981 gacactggtg ttgaacatgt taccttcttc atctacaata aaattgttga tgagcctgaa
26041 gaacatgtcc aaattcacac aatcgacggt tcatccggag ttgttaatcc agtaatggaa
26101 ccaatttatg atgaaccgac gacgactact agcgtgcctt tgtaagcaca agctgatgag
26161 tacgaactta tgtactcatt cgtttcggaa gagataggtt cgtaatagt taatacgta
26221 cttcttttc ttgctttcgt ggtattcttgc ttagttacac tagccatct tactgcgcctt
26281 cgattgtgtg cgtactgctg caatattgtt aacgtgagtc ttgtaaaacc ttctttttac
26341 gtttactctc gtgttaaaaaa tctgaattct tctagagttc ctgatcttct ggtctaaacg
26401 aactaaatat tatatttagtt tttctgtttg gaactttaat tttagccatg gcaggttcca
26461 acggtactat taccgttcaa gagctaaaaa agctccttga agaatggAAC ctagtaatag
26521 gtttcctatt ccttacatgg atttgtcttc tacaatttgc ctatgccaaac aggaataggt
26581 ttttgtatataat aattaagtta attttcctct ggctgttatg gccagtaact ttaacttgtt
26641 ttgtgcttgc tgctgtttac agaataaaatt ggatcaccgg tggaattgtt atcgcaatgg
26701 cttgtcttgtt aggcttgatg tggctcagct acttcattgc ttctttcaga ctgtttgcgc
26761 gtacgcgttc catgtggtca ttcaatccag aaactaacat tcttctcaac gtgccactcc
26821 atggcactat tctgaccaga ccgcttctag aaagtgaact cgtaatcgga gctgtgatcc

```

SUBSTITUTE SHEET (RULE 26)

26881 ttctgtggaca tttttgttatt gctggacacc atctaggacg ctgtgacatc aaggacactgc  
 26941 ctaaaagaaaat cactgttgct acatcacgaa cgctttctta ttacaaaattg ggagcttcgc  
 27001 agcgtgttagc aggtgactca ggtttgctg catacagtcg ctacaggatt ggcaactata  
 27061 aattaaacac agaccattcc agtagcagtg acaatattgc tttgcttgta cagtaagtga  
 27121 caacagatgt ttcatctcg tgactttcag gttactatacg cagagatatt actaattatt  
 27181 atgcggactt ttaaagtttc catttggaaat cttgattaca tcataaacct cataattaaa  
 27241 aatttatcta agtcactaac tgagaataaa tattctcaat tagatgaaga gcaaccaatg  
 27301 gagattgatt aaacgaacat gaaaattatt cttttcttgg cactgataac actcgctact  
 27361 tgtgagctt atcactacca agagtgtgtt agaggtacaa cagtactttt aaaagaacct  
 27421 tgctttctg gaacatacga gggcaattca ccatttcattc ctctagctga taacaaattt  
 27481 gcactgactt gcttagcac tcaatttgct tttgcttgc ctgacggcgt aaaacacgtc  
 27541 tatcagttac gtgccagatc agttcacct aaactgttca tcagacaaga ggaagttcaa  
 27601 gaactttact ctccaatttt tcttattgtt gcggcaatag tgtttataac actttgttc  
 27661 acactcaaaa gaaagacaga atgattgaac tttcattaat tgacttctat ttgtgtttt  
 27721 tagcctttct gttattcctt gtttaatta tgcttattat cttttggttc tcacttgaac  
 27781 tgcaagatca taatgaaact tgtaacgcct aaacgaacat gaaattttt gttttcttag  
 27841 gaatcatcac aactgttagct gcatttcacc aagaatgtag tttacagtca tgtactcaac  
 27901 atcaaccata tgttagttgat gacccgtgtc ctattcactt ctattctaaa tggtatatta  
 27961 gagtaggagc tagaaaatca gcacctttaa ttgaattgtg cgtggatgag gctggttcta  
 28021 aatcacccat tcagtacatc gatatcggtt attatacagt ttccctgttta ccttttacaa

*Fig. 12B* *continued*

Fig. 12B  
continued

SUBSTITUTE SHEET (RULE 26)

28081 ttaattgcca ggaacctaaa ttgggtagtc ttgttagtgcg ttgttcgttc tatgaagact  
28141 ttttagagta tcatgacgtt cgtgtgttt tagatttcat ctaaacgaac aaacttaaat  
28201 gtctgataat ggaccccaaa atcagcgaaa tgcaactccgc attacgtttg gtggaccctc  
28261 agattcaact ggcagtaacc agaatggtgg ggcgcatca aaacaacgtc ggcccccaagg  
28321 tttacccaat aatactgcgt cttggttcac cgctctcaact caacatggca aggaagacct  
28381 taaattccct cgaggacaag gcgttccaat taacaccaat agcagtccag atgacccaaat  
28441 tggctactac cgaagagcta ccagacgaat tcgtggtgtt gacggtaaaa tgaaagatct  
28501 cagtccaaaga tggtatttct actacctagg aactgggcca gaagctggac ttccctatgg  
28561 tgctaacaaa gacggcatca tatgggttgc aactgagggga gccttgaata caccaaaaga  
28621 tcacattggc acccgcaatc ctgctaacaa tgctgcaatc gtgctacaac ttccctcaagg  
28681 aacaacattg ccoaaaggct tctacgcaga agggagcaga ggcggcagtc aagcctttc  
28741 tcgttcttca tcacgtagtc gcaacagttc aagaaattca actccaggca gcagtaaacg  
28801 aacttctcct gctagaatgg ctggcaatgg cggtgatgct gctttgctt tgctgctgct  
28861 tgacagattg aaccagcttgc agagcaaaat gtctggtaaa ggccaacaac aacaaggcca  
28921 aactgtcact aagaaatctg ctgctgaggc ttctaagaag cctcggcaaa aacgtactgc  
28981 cactaaagca tacaatgtaa cacaagttt cggcagacgt ggtccagaac aaaccccaagg  
29041 aaattttggg gaccaggaac taatcagaca aggaactgat tacaaacatt ggccgcaaatt  
29101 tgcacaattt gcccccaagcg cttcagcggtt cttcggaatg tcgcgcattt gcatggaaatg  
29161 cacaccttcg ggaacgtggc tgacctacac aggtgccatc aaattggatg acaaaagatcc  
29221 aaatttcaaa gatcaagtca ttttgcgtt gatcaatatt gacgcataca aaacattccc  
29281 accaacagag cttaaaaagg acaaaaagaa gaaggctgat gaaactcaag ctttaccgca

29341 gagacagaag aaacagcaaa ctgtgactct tcttcctgct gcagatttg atgatttctc  
29401 caaacaattg caacaatcca tgaggagtgc tgactcaact caggcctaaa ctcatgcaga  
29461 ccacacaagg cagatgggct atataaacgt tttcgcttt ccgtttacga tatatagtct  
29521 actcttgtgc agaatgaatt ctcgttaacta catagcacaa gtagatgtag ttaactttaa  
29581 tctcacatag caatctttaa tcagtgtgta acattaggaa ggacttgaaa gagccaccac  
29641 atttcacccg aggccacgcg gagtacgatc gagtgtacag tgaacaatgc tagggagagc  
29701 tgcttatatg gaagagccct aatgtgtaaa attaatttta gta

*Fig. 12B* *continued*

# Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC\_045512.2

## FASTA Graphics

Go to:

LOCUS	NC_045512	29903 bp ss-RNA	linear	VRL 18-JUL-
2020				
DEFINITION	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,			
	complete genome.			
ACCESSION	NC_045512			
VERSION	NC_045512.2			
DBLINK	BioProject: <a href="#">PRJNA485481</a>			
KEYWORDS	RefSeq.			
SOURCE	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)			
ORGANISM	<u>Severe acute respiratory syndrome coronavirus 2</u>			

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

	Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE	1 (bases 1 to 29903)
AUTHORS	Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z.
TITLE	A new coronavirus associated with human respiratory disease in China
JOURNAL	Nature 579 (7798), 265-269 (2020)
PUBMED	<u>32015508</u>
REMARK	Erratum: [Nature. 2020 Apr;580(7803):E7. PMID: 32296181]
REFERENCE	2 (bases 13476 to 13503)
AUTHORS	Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. and Howard, M.T.
TITLE	Programmed ribosomal frameshifting in decoding the SARS-CoV genome
JOURNAL	Virology 332 (2), 498-510 (2005)
PUBMED	<u>15680415</u>
REFERENCE	3 (bases 29728 to 29768)

*Fig. 12C* *continued*

SUBSTITUTE SHEET (RULE 26)

AUTHORS Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M. Jr.  
and  
Scott, W.G.

TITLE The structure of a rigorously conserved RNA element within the  
SARS  
virus genome

JOURNAL PLoS Biol. 3 (1), e5 (2005)  
15630477

REFERENCE 4 (bases 29609 to 29657)

AUTHORS Williams, G.D., Chang, R.Y. and Brian, D.A.

TITLE A phylogenetically conserved hairpin-type 3' untranslated region  
pseudoknot functions in coronavirus RNA replication

JOURNAL J. Virol. 73 (10), 8349-8355 (1999)  
10482585

REFERENCE 5 (bases 1 to 29903)

CONSRM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (17-JAN-2020) National Center for Biotechnology  
Information, NIH, Bethesda, MD 20894, USA

REFERENCE 6 (bases 1 to 29903)

*Fig. 12C continued*

SUBSTITUTE SHEET  
(RULE 26)

AUTHORS Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Hu, Y., Song, Z.-G.,  
Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.L., Zhang, Y.-L.,  
Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C.  
and  
Zhang, Y.-Z.

TITLE Direct Submission

JOURNAL Submitted (05-JAN-2020) Shanghai Public Health Clinical Center &  
School of Public Health, Fudan University, Shanghai, China

COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The  
reference sequence is identical to MN908947.  
On Jan 17, 2020 this sequence version replaced NC\_045512.1.  
Annotation was added using homology to SARSr-CoV NC\_004718.3. ###  
Formerly called 'Wuhan seafood market pneumonia virus.' If you  
have questions or suggestions, please email us at  
info@ncbi.nlm.nih.gov

and include the accession number NC\_045512.1.### Protein structures  
can be found at

<https://www.ncbi.nlm.nih.gov/structure/?term=sars-cov-2>.### Find  
all other Severe acute respiratory syndrome coronavirus 2  
(SARS-CoV-2) sequences at

*Fig. 12C continued*

*Fig. 12C continued*

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

```
##Assembly-Data-START##
Assembly Method      :: Megahit v. V1.1.3
Sequencing Technology :: Illumina
##Assembly-Data-END##
COMPLETENESS: full length.
```

FEATURES	Location/Qualifiers
source	1..29903 /organism="Severe acute respiratory syndrome coronavirus 2" /mol_type="genomic RNA" /isolate="Wuhan-Hu-1" /host="Homo sapiens" /db_xref="taxon: <u>2697049</u> " /country="China" /collection_date="Dec-2019"
5'UTR	1..265
gene	266..21555 /gene="ORF1ab" /locus_tag="GU280_gp01"

SUBSTITUTE SHEET (RULE 26)

CDS

```
/db_xref="GeneID:43740578"
join(266..13468,13468..21555)
/gene="ORF1ab"
/locus_tag="GU280_gp01"
/ribosomal_slippage
/note="pplab; translated by -1 ribosomal frameshift"
/codon_start=1
/product="ORF1ab polyprotein"
/protein_id="YP_009724389.1"
/db_xref="GeneID:43740578"

/translation="MESLVPGFNEKTHVQLSLPVQLQVRDVLVRGFGDSVEEVLSEARQ
HLKDGTGCLVEVEKGVLPOLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
```

*Fig. 12C continued*

LNSIIIKTIQPRVEKKLDGMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG  
DFVKATCEFCGTENLTKEGATTGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG  
LKТИLRKGGRТИAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGЛNDNL  
LEILQKEKVNIIVGDFKLNEEIAIILASFSASTSАFVETVKGLDYKAFKQIVESCGN  
FKVTKGKAKKGAWNIGEOKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA  
ITILDGISQYSLRLIDAMMFSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTIVYEKL  
KPVLWDLEEKFKEGVEFLRGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFKLV  
NKEFLALCADSIИIGGAKLКАLNЛGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII  
FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGMLLEIKDTEK  
YCALAPNMVTNNTFILKGGAPTKVTFGDDIVIEVQGYKSVNITFELDERIDKVLNEK

*Fig. 12C continued*

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF  
KLASHMYCFSFYPPDEDEEEGDCEEEEFPSTQYEGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQTVGQQDGSEDNQTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG  
YLKLTDNVYIKNADIVEEAKKVKP TVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTLEETKFLTENLLYIDINGN  
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPKKAGGTTEMLAKALRKV  
PTDNYITTYPGOGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM  
LAHAEETRKLMPCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

TLNDLNETLVTMP LGYVTHGLNLEEAARYMRSLKVPATSVSSPAVTAYNGYLTS  
KTPEEEHFIE TISLAGSYKDWSYSGQSTQLGIEFLRGDKSVYYT SNP TT FHLDGEVIT  
FDNLKTLLSLREVRTIKVFTVDNINLHTQVVDMMSMTYQQQFGPTYLDGADVT KIKPH  
NSHEGKTFYVLPNDDTLRVEAF EYYHTDPSFLGRYMSALNHTKKWKPQVNGLTSIK  
WADNNCYLATALLTLQQIELKF NPPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ  
IPCTCGKQATKYLVQQESPFVMMMSAPP A QYELKHGTETCASEYTGN YQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENS YTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVPNQPYPNASFDNFKEVCDNIKEADDLNQLTGYKKPASRELKV  
FFPD LNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

*Fig. 12C continued*

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD  
IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRLGLKTLATHGLAAVNS  
VPWDTIANYAKPFLNKVVSTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTSINSRI  
KASMPPTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLKVCLGSLIYSTA  
ALGVILMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSDLTYPSET  
IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL  
MWLIINLVQMAPISAMVRMYIFFASEFYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE  
CTTIVNGVRRSFYVYANGKGFCKLHNWNVCNCDFCAGSTFISDEVARDSLQFKRP  
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI  
NVIVFDGKSKEESSAKSASVYYSQLMCQPILLLDQALVSDVGD SAEVAVKMFDAYVN

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

TFSSTFNVPMEKLKTLVATAEAEELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC  
LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI  
WNVKDFMSLSEQLRKQIRSAAKNNLPFKLTCAATTRQVVNVVTTKIALKGKIVNNWL  
KQLIKVTLVFLFVAAIFYLITPVHVM SKHTDFSSEIIIGYKAIDGGVTRDIASTDT CFA  
NKHADEFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHF LP  
RVFSAVGNICYTPSKLIEYTDFATSACVLA AECTIFKDASGKP VP YCYDTNVLEGSVA  
YESLRPDTRYVLM DGSI IQFPNTYLEGSVRVVTTFDSEYCRHGT CERSEAGVCVSTSG  
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAI VVTCL  
AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVY SVIYLYLTFYLTN  
DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKH EYWFFSNYLKRRVFNGVSESTFE

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

EAALCTFLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL  
AKALNDFNSNSGSDVLYQPPQTSITSAVIQLQSGFRKMAFPSGKVEGCMVQVTCGTTLNG  
LWLDDVVYCPRHVIC TSEDMNP NYEDILLIRKSNNFLVQAGNVQLRVIGHSMQNCVL  
KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC  
GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFDVDRQTAQAAGTDTTITVN  
VLAWLYYAAVINGDRWFNLNRF TTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV  
LDMCASLKELLQNGMNGRTILGSALLEDEFPTFDVVRQCSGVTFQSAVKRTIKGTHHW  
LLLTI LTSLLVLVQSTQWSLFFFYENAEFLPFAMGI IAMSAFAMMFVKHKHAFLCLFL  
LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR  
TVYDDGARRVWTLMNVLTIVYKVYYGNALDQAISMWALII SVTSNYSGVVTVMFLAR

*Fig. 12C continued*

GIVFMCVEYCPPIFFITGNTLQCIMLVYCFGLGYFCTCYFGLFCLLNRYFRLTLGVYDYL  
VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL  
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL  
CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKSLNVAK  
SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN  
NIINNARDGCVPLNIIPLTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA  
DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA  
CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP  
KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANS TVLSFCFAVDAAK  
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

PNPKGFCDLKGKYVQIPTTCANDPGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA  
DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKEFLKTNCCRQEKD  
EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR  
LTKEYTMADLVYALRHFDEGNCDTLKEILVTYNCDDYFNKKDWYDFVENPDILRVYA  
NLGERVRQALLKTVQFCDAQRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV  
DSYYSLMPILTLTRALTAESHVDIDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDO  
TYHPNCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRE  
LGVVHNQDVNLHSSRLSFKELLVVAADPAMHAASGNLLLDKRTTCEFSVAALTNNVAEQ  
TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFQDGNAIASDYDYYRYNLPTMCDI  
RQLLFVVEVVDKYFDCYDGGCINANQVI VNNLDKSAGFPFNKGKARLYYDSMSYEDQ

*Fig. 12C continued*

DALFYTKRNVPIPTQMNLKYAISAKNRARTVAGVSICSTMNRQFHQKLILKSIAAT  
RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR  
KHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ  
AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHF SMM  
ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTEDLTKGPHEFCS  
QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH  
PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNNDNTSRYWEPEFYEAMYTPHTV  
LQAVGACVLCNSQTSLRGACIRRPFLCCKCCYDHVISTSHKLVL SVNPyVCNAPGCD  
VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW  
TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP

*Fig. 12C continued*

RPPLNRNYVFTGYRVTNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHEAIGLALYYPSARIIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDFKEVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHVYIGDPAQLPAPRTLLTKTLEPEYFNSVCRLMKTIGPDMFLGTCCRCPAEIVDTVSALVYDNKLKAHKDKSAQCFCFKMFYKGVIHDVSSAINRPQIGVVREFLTRNPawRKAVFISPYNQNAVASKILGLPTQTVDSQSQGSEYDYVIFTQTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATLQAENVTLKFDCSKVITGLHPTQAPTHLSVDTKFKT EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP

*Fig. 12C continued*

LMYKGLPWNVVRRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL  
CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA  
HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK  
FPVLHDIGNPKAIKCVQPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL  
FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLLQ  
LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN  
MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT  
VYTKVDGVDVELFENKTTLPVNVAFELWAKRNICKPVPEVKILNNLGVDIAANTVIWDY  
KRDAPAHISTIGVCSMTDIACKPTETICAPLTFFDGRVDGQVDLFRNARNGVLITEG  
SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYKKVDGVVQQLPETYFTQSRNLQEFK

*Fig. 12C continued*

PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESPR  
FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV  
TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD  
SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRCVIHFGAGSDKGVAPGTAVLRQWLP  
TGTLVDSLNDVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG  
FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGFIAWWTAFTVNVNASSSEAFL  
IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE  
GQINDMILSLLSKGRLLIRENNRVISSDVLVNN"

mat\_peptide

266..805

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="leader protein"

/note="nspl; produced by both ppl1 and pplab"

/protein\_id="YP\_009725297.1"

*Fig. 12C continued*

mat\_peptide 806..2719  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp2"  
/note="produced by both pp1a and pp1ab"  
/protein\_id="YP\_009725298.1"

mat\_peptide 2720..8554  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp3"  
/note="former nspl; conserved domains are: N-terminal acidic (Ac), predicted phosphoesterase, papain-like proteinase, Y-domain, transmembrane domain 1 (TM1), adenosine diphosphate-ribose 1''-phosphatase (ADRP); produced by both pp1a and pp1ab"  
/protein\_id="YP\_009725299.1"

mat\_peptide 8555..10054  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp4"

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

by

mat\_peptide /note="nsp4B\_TM; contains transmembrane domain 2 (TM2); produced by both ppla and pplab"  
/protein\_id="YP\_009725300.1"  
10055..10972  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="3C-like proteinase"  
/note="nsp5A\_3CLpro and nsp5B\_3CLpro; main proteinase (Mpro); mediates cleavages downstream of nsp4. 3D structure of the SARSr-CoV homolog has been determined (Yang et al., 2003); produced by both ppla and pplab"  
/protein\_id="YP\_009725301.1"  
mat\_peptide 10973..11842  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp6"  
/note="nsp6\_TM; putative transmembrane domain; produced by both ppla and pplab"  
/protein\_id="YP\_009725302.1"  
mat\_peptide 11843..12091

*Fig. 12C continued*

mat\_peptide /gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp7"  
/note="produced by both pplα and pplβ"  
/protein\_id="YP\_009725303.1"  
12092..12685  
mat\_peptide /gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp8"  
/note="produced by both pplα and pplβ"  
/protein\_id="YP\_009725304.1"  
12686..13024  
mat\_peptide /gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp9"  
/note="ssRNA-binding protein; produced by both pplα and  
pplβ"  
/protein\_id="YP\_009725305.1"  
13025..13441  
mat\_peptide /gene="ORF1ab"  
/locus\_tag="GU280\_gp01"

*Fig. 12C continued*

like

/product="nsp10"  
/note="nsp10\_CysHis; formerly known as growth-factor-

mat\_peptide

protein (GFL); produced by both pplα and pplβ"

/protein\_id="YP\_009725306.1"

join(13442..13468,13468..16236)

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="RNA-dependent RNA polymerase"

/note="nsp12; NiRAN and RdRp; produced by pplβ only"

/protein\_id="YP\_009725307.1"

mat\_peptide

16237..18039

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="helicase"

/note="nsp13\_ZBD, nsp13\_TB, and nsp\_HELcore; zinc-

binding

domain (ZD), NTPase/helicase domain (HEL), RNA

5'-triphosphatase; produced by pplβ only"

/protein\_id="YP\_009725308.1"

*Fig. 12C continued*

mat\_peptide 18040..19620  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="3'-to-5' exonuclease"  
/note="nsp14A2\_ExoN and nsp14B\_NMT; produced by pplab only"  
/protein\_id="YP\_009725309.1"

mat\_peptide 19621..20658  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="endoRNase"  
/note="nsp15-A1 and nsp15B-NendoU; produced by pplab only"  
/protein\_id="YP\_009725310.1"

mat\_peptide 20659..21552  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="2'-O-ribose methyltransferase"  
/note="nsp16\_OMT; 2'-o-MT; produced by pplab only"  
/protein\_id="YP\_009725311.1"

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

CDS 266..13483  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/note="ppla"  
/codon\_start=1  
/product="ORF1a polyprotein"  
/protein\_id="YP\_009725295.1"  
/db\_xref="GeneID: 43740578"  
  
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVILSEARQ  
HLKDGTGCGLVEVEKGVLPCLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE  
TLGVVLVPHVGETPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN  
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ  
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYLEQTPEIKLAKKFDTFNGECPNFVFP  
LNSIIKTIQPRVEKKLDGMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

*Fig. 12C continued*

DFVKATCEF CGTENLTKEGATT CGYL P QNAVVKIYCPA CHNSEV GPEHSLAEYHN ESG  
LKTILRKGGRTIAFGGCVF SYVGCHNK CAYWV PRASANIGCNHTGVV GEGSEG LNDNL  
LEILQKEKV NINIVGDFKLNEEIAII LASFSASTSAF VETVKGLDYKAFKQ IVE SCGN  
FKVTKGKA KKGAWNIGEQKSILSPL YAFASEAARVVRSIFSRTLETAQNSVRVLQKAA  
ITILDG ISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL  
KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKE SVQTFFKLV  
NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKC VKSREETGLLMP LKAPKEII  
FLEG ETLPTEVLTEEVVLKTGDLQPL QOPTSEA VEAPLVGTPVCINGMLLEIKDTEK  
YCALAPNM MVINNTFTLKGGA PTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK  
CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDES GEF

*Fig. 12C continued*

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE  
EEQEEDWLDDDSQQQTVGQQDGSEDNQTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG  
YLKLTDNVYIKNADIVEEAKKVKP TVVVNAANVYLKHGGGVAGALNKATNNAMQVESD  
DYIATNGPLKVGGSCLSGHNLAKHCLHVVGPNVNKGEDIQOLLKSAYENFNQHEVLLA  
PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA  
EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTLEETKFLTENLLYIDINGN  
LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV  
PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM  
LAHAEETRKLMPCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN  
TLNDLNETLVTMP LGYVTHGLNLEEAARYMRSLKVPATSVSSPAVTAYNGYLTS

*Fig. 12C continued*

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT  
FDNLKTLLSLREVRTIKVFTIVDNINLHTQVVDMMSMTYGOQFGPTYLDGADVTKIKPH  
NSHEGKTFYVLPNDDTLRVEAFEYYHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK  
WADNNCYLATALLTLQQIELKFNPALQDAYYRARAGEAANFCALILAYCNKTVGELG  
DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVO  
IPCTCGKQATKYLVQQESPVMMSSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK  
ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTIKPVTYKLDGVVCTEIDPKLDNYY  
KKDNSYFTEQPIDLVPNQPYPNASFNFKFVCDNIKEADDLNQLTGYKKPASRELKVT  
FFPDLNGDVVAIDYKHYTPSFKKGAKLHKPIVWHVNNTNKATYKPNTWCIRCLWST  
KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

*Fig. 12C continued*

IILKPANNSLKITEEVGHTDLMAAYVDNS SLTIKKPNELSRLGLKTLATHGLAAVNS  
VPWDTIANYAKPFLNKVVSTITNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTSNSRI  
KASMPPTIAKNTVKSVGKF CLEASFNYLKSPNF SKLINII IWFLLSVCLGS LIYSTA  
ALGVILMSNLGMP SYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSDLTYP SLET  
IQITISSEFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL  
MWLIINLVQMAP ISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE  
CTTIVNGVRRSFYVYANGKGFCKLHNWNCSVNCDFCAGSTFISDEVARDSLQFKRP  
INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI  
NVIVFDGKSKEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN  
TFSSSTFNVPMEKLKTLVATAEAEELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

LKLSHQSDIEVTGDSNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI  
WNVKDFMSLSEQLRKQIRSAAKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL  
KQLIKVTLVFLFVAAIFYLITPVHVM SKHTDFSSEIIIGYKAIDGGVTRDIASTDT CFA  
NKHADEFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHF LP  
RVFSAVGNICYTPSKLIEYTDFA TSACVLAAECTIFKDASGKPVPYC YDTNVLEG SVA  
YESLRPDTRYVLM DGSI IQFPNTYLEGSVRV VTFDSEYCRHGT CERSEAGVCVSTSG  
RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLI QPI GALDI SASIVAGGIVAI VVTCL  
AYYFMRFRRAFGEYSHVVAFNTLLELMSFTVLCLTPVYSFLPGVY SVIYLYLT FYL TN  
DVSFLAHIQWMVME TPLVPEWITIAYIICISTKHFWFFSNYLKRRVFNGVSFSTEE  
EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

AKALNDFNSNSGSDVLYQPPQTSITSAVLQSGFRKMAFP SGKVEGCMVQVTCGTTLNG  
LWLDDVVYCPRHVIC TSEDMNP NYEDLLIRKSNNFLVQAGNVQLRVIGHSMQNCVL  
KLKVDTANPKTPKYKFVRIQP GQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC  
GSVGFNIDYDCVSEFCYMHHMELPTGVHAGTDLEGNFYGP FVDRQTAQAAGTDTTITVN  
VLAWLYYAAVINGDRWFLNRFTTILNDFN LVAMKYNYEPLTQDHVDILGPLSAQTGIAV  
LDMCASLKELLQNGMNGRTI LGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW  
LLLTI LTSLLVLVQSTQWSLEFFFLYENAEFLPFAMGIIAMS AFAMMFVKHKHAFLCFL  
LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR  
TVYDDGARRWTLMNVLTLVYK VYYGNALDQAISMWALII SVTSNYSGVVTVMFLAR  
GIVFMCVEYCPIFFITGNTLQCIMLVYCF LGYFCTCYFGLFCLLNRYFRLTLGVYDYL

*Fig. 12C continued*

*Fig. 12C continued*

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL  
LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL  
CEEMLDNRATLQATIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKSLNVAK  
SEFDRDAAMQRKLEKMDQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN  
NIINNARDGCVPNIPILTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA  
DSKIVQLSEISMNDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA  
CTDDNALAYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP  
KGPKVKYLYFIKGLNNLNRMVLGSLAATVRLQAGNATEVPANSTVLSECAFADVAAK  
AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH  
PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA  
DAQSFNLNGFAV"

SUBSTITUTE SHEET (RULE 26)

mat\_peptide 266..805  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="leader protein"  
/note="nspl; produced by both pp1a and pp1ab"  
/protein\_id="YP\_009742608.1"  
  
mat\_peptide 806..2719  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp2"  
/note="produced by both pp1a and pp1ab"  
/protein\_id="YP\_009742609.1"  
  
mat\_peptide 2720..8554  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp3"  
/note="former nspl; conserved domains are: N-terminal  
acidic (Ac), predicted phosphoesterase, papain-like  
proteinase, Y-domain, transmembrane domain 1 (TM1),  
adenosine diphosphate-ribose 1''-phosphatase (ADRP); Fig. 12C continued

produced by both pplα and pplβ"  
/protein\_id="YP\_009742610.1"  
mat\_peptide 8555..10054  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp4"  
/note="nsp4B\_TM; contains transmembrane domain 2 (TM2); produced by both pplα and pplβ"  
/protein\_id="YP\_009742611.1"  
mat\_peptide 10055..10972  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="3C-like proteinase"  
/note="nsp5A\_3CLpro and nsp5B\_3CLpro; main proteinase (Mpro); mediates cleavages downstream of nsp4. 3D structure of the SARS-CoV homolog has been determined (Yang et al., 2003); produced by both pplα and pplβ"  
/protein\_id="YP\_009742612.1"  
mat\_peptide 10973..11842  
/gene="ORF1ab"

*Fig. 12C continued*

by

/locus\_tag="GU280\_gp01"  
/product="nsp6"  
/note="nsp6\_TM; putative transmembrane domain; produced  
mat\_peptide both ppla and pplab"  
/protein\_id="YP\_009742613.1"  
11843..12091  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp7"  
/note="produced by both ppla and pplab"  
/protein\_id="YP\_009742614.1"  
mat\_peptide 12092..12685  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp8"  
/note="produced by both ppla and pplab"  
/protein\_id="YP\_009742615.1"  
mat\_peptide 12686..13024  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"

*Fig. 12C continued*

/product="nsp9"  
/note="ssRNA-binding protein; produced by both pplα and pplβ"  
/protein\_id="YP\_009742616.1"  
mat\_peptide 13025..13441  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp10"  
/note="nsp10\_CysHis; formerly known as growth-factor-like  
protein (GFL); produced by both pplα and pplβ"  
/protein\_id="YP\_009742617.1"  
mat\_peptide 13442..13480  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/product="nsp11"  
/note="produced by pplα only"  
/protein\_id="YP\_009725312.1"  
stem\_loop 13476..13503  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"

*Fig. 12C continued*

/inference="COORDINATES:  
profile:Rfam-release-14.1:RF00507, Infernal:1.1.2"  
/function="Coronavirus frameshifting stimulation element  
stem-loop 1"  
stem\_loop 13488..13542  
/gene="ORF1ab"  
/locus\_tag="GU280\_gp01"  
/inference="COORDINATES:  
profile:Rfam-release-14.1:RF00507, Infernal:1.1.2"  
/function="Coronavirus frameshifting stimulation element  
stem-loop 2"  
gene 21563..25384  
/gene="S"  
/locus\_tag="GU280\_gp02"  
/gene\_synonym="spike glycoprotein"  
/db\_xref="GeneID:43740568"  
CDS 21563..25384  
/gene="S"  
/locus\_tag="GU280\_gp02"  
/gene\_synonym="spike glycoprotein"

*Fig. 12C continued*

/note="structural protein; spike protein"  
/codon\_start=1  
/product="surface glycoprotein"  
/protein\_id="YP\_009724390.1"  
/db\_xref="GeneID:43740568"

/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR  
SSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIR  
GWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHNNKSWMESEFRVY  
SSANNCTFEYVSQPFMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPO  
GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTEFL  
LKYNENGTTDAVDCALDPLSETKCTLKSFTVEKGIVQTSNFRVQPTESIVRFPNITN  
LCPEGEVFNAIREASVYAWNKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDCE  
TNVYADSEVIRGDEVROIAPGQTGKIADNYKLPPDDFTGCVIAWNNSNNLDSKVGGNYN

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

YLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY  
RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKF LPFQQFG  
RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQAVLYQDVNCTEVPVAI  
HADQLPTWRRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPR  
RARSVASQSIIAYTMSLGAENSVAYSNNSLAIPTNFTISVTTEILPVSMKTSVDCTM  
YICGDSTECNSNLLQYGSFCTQLNRALTGIAVEQDKNTQEVAQVKQIYKTPPIKDFG  
GFNFSQLPDP SKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLG DIAARDLICAQKFN  
GLTVIPLLTD EMTAQYTSALLAGTTISGWTFGAGAAALQIPFAMQMAYRFNGIGVTQN  
VLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA  
ISSVILNDILSRLDKVEAEVQIDRLITGRLOSLQTYVTQQLIRAAEIRASANLAATKMS

*Fig. 12C continued*

ECVLGQS KRVDF CGKGYHLM SFPQSAP HGVVFLHV TYVPAQ EKNFTTAPA ICHDGKAH  
FPREGVFV SNGTHWFVTQRNFYEPQIITTDNTFVGNC DVVIGIVNN TVYDPLQPELD  
SFKEELDKYFKNHTSPDV DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELG  
KYEQYIKWPWYIW LGFIAGLIAIVMVTIMLCCMTSCSCLKGCCSCGSCCKFDEDDSE  
PVLKGVKLHYT"

gene 25393..26220  
/gene="ORF3a"  
/locus\_tag="GU280\_gp03"  
/db\_xref="GeneID:43740569"

CDS 25393..26220  
/gene="ORF3a"  
/locus\_tag="GU280\_gp03"  
/codon\_start=1  
/product="ORF3a protein"  
/protein\_id="YP\_009724391.1"  
/db\_xref="GeneID:43740569"

*Fig. 12C continued*

/translation="MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFG  
WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE  
APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY  
NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQOLYSTQ  
LSTDGTGVEHVTEFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"

gene

26245..26472

/gene="E"

/locus\_tag="GU280\_gp04"

/db\_xref="GeneID:43740570"

CDS

26245..26472

/gene="E"

/locus\_tag="GU280\_gp04"

/note="ORF4; structural protein; E protein"

/codon\_start=1

/product="envelope protein"

/protein\_id="YP\_009724392.1"

*Fig. 12C continued*

/db\_xref="GeneID:43740570"

/translation="MYSFVSEETGTLIVNSVLLFLAFVVFLVTLAILTALRLCAYCC  
NIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV"

gene 26523..27191

/gene="M"

/locus\_tag="GU280\_gp05"

/db\_xref="GeneID:43740571"

CDS 26523..27191

/gene="M"

/locus\_tag="GU280\_gp05"

/note="ORF5; structural protein"

/codon\_start=1

/product="membrane glycoprotein"

/protein\_id="YP\_009724393.1"

/db\_xref="GeneID:43740571"

/translation="MADSNGTITVEELKKLLEQWNLVIGFLFLTWCICLLQFAYANRNR  
FLYIIKLIFLWLLWPVTLACEVLAAYRINWITGGIAIAMACLVGLMWLSYFIASFRL

*Fig. 12C continued*

FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD

IKDLPKEITVATSRTL<sup>SYY</sup>KLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSDNIA  
LLVQ"

gene 27202..27387  
 /gene="ORF 6"  
 /locus\_tag="GU280\_gp06"  
 /db\_xref="GeneID: 43740572"

CDS 27202..27387  
 /gene="ORF 6"  
 /locus\_tag="GU280\_gp06"  
 /codon\_start=1  
 /product="ORF6 protein"  
 /protein\_id="YP\_009724394.1"  
 /db\_xref="GeneID: 43740572"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIINKNSKSL  
TENKYSQLDEEQPMEID"

gene 27394..27759  
 /gene="ORF7a"

*Fig. 12C continued*

## SUBSTITUTE SHEET (RULE 26)

/locus\_tag="GU280\_gp07"  
/db\_xref="GeneID: 43740573"  
CDS 27394..27759  
/gene="ORF7a"  
/locus\_tag="GU280\_gp07"  
/codon\_start=1  
/product="ORF7a protein"  
/protein\_id="YP\_009724395.1"  
/db\_xref="GeneID: 43740573"  
  
/translation="MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNS  
PFHPLADNKFALTGFSTQFAFACPDGVKHVVQLRARSVSPKLFIRQEEVQELYSPIFL  
IVAAIVFITLCFTLKRKTE"  
gene 27756..27887  
/gene="ORF7b"  
/locus\_tag="GU280\_gp08"  
/db\_xref="GeneID: 43740574"  
CDS 27756..27887  
/gene="ORF7b"  
/locus\_tag="GU280\_gp08"

*Fig. 12C continued*

/codon\_start=1  
/product="ORF7b"  
/protein\_id="YP\_009725318.1"  
/db\_xref="GeneID:43740574"

/translation="MIELSLIDEYLCFLAFLFLVLIIMLITFWFSLELQDHNETCHA"

gene 27894..28259  
/gene="ORF8"  
/locus\_tag="GU280\_gp09"  
/db\_xref="GeneID:43740577"

CDS 27894..28259  
/gene="ORF8"  
/locus\_tag="GU280\_gp09"  
/codon\_start=1  
/product="ORF8 protein"  
/protein\_id="YP\_009724396.1"  
/db\_xref="GeneID:43740577"

/translation="MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK

*Fig. 12C continued*

WYIRVGARKSAPIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGS娄VVRC  
SFYEDFLEYHDVRVVLDFI"

gene 28274..29533  
/gene="N"  
/locus\_tag="GU280\_gp10"  
/db\_xref="GeneID: 43740575"

CDS 28274..29533  
/gene="N"  
/locus\_tag="GU280\_gp10"  
/note="ORF9; structural protein"  
/codon\_start=1  
/product="nucleocapsid phosphoprotein"  
/protein\_id="YP\_009724397.2"  
/db\_xref="GeneID: 43740575"

/translation="MSDNGPQNQRNAPRITFGGPDSTGSNONGERSGARSQRRPQG  
LPNNTASWFTALTQHGKEDLKFPRGQGVPIINTNSSPDDQIGYYRRATRRIRGGDGKMK  
DLSPRWYFYYLGTGPEAGLPYGANKDGLIWVATEGALNTPKDHIIGTRNPANNAAIVLQ

*Fig. 12C continued*

LPQGTTLPKGFYAECSRGGSQASSRSSRSRNSSRNSTPGSSRGTSARMAGNGDAA  
LALLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKPRQKRTATKAYNVTQAFGR  
RGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVIPS GTWLITYT  
GAIKLDDKDPNFKDQVILLNKHIDAYKTFFPTEPKDKKKKADETQALPQRQKKQQTV  
TLLPAADLDDFSKQLQQSMS SADSTQA"

gene 29558..29674  
/gene="ORF10"  
/locus\_tag="GU280\_gp11"  
/db\_xref="GeneID:43740576"

CDS 29558..29674  
/gene="ORF10"  
/locus\_tag="GU280\_gp11"  
/codon\_start=1  
/product="ORF10 protein"  
/protein\_id="YP\_009725255.1"  
/db\_xref="GeneID:43740576"

*Fig. 12C continued*

Fig. 12C  
continued

SUBSTITUTE SHEET (RULE 26)

stem\_loop

/translation="MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT"  
29609..29644  
/gene="ORF10"  
/locus\_tag="GU280\_gp11"  
/inference="COORDINATES:  
profile::Rfam-release-14.1:RF00165, Infernal:1.1.2"  
/function="Coronavirus 3' UTR pseudoknot stem-loop 1"  
stem\_loop  
29629..29657  
/gene="ORF10"  
/locus\_tag="GU280\_gp11"  
/inference="COORDINATES:  
profile::Rfam-release-14.1:RF00165, Infernal:1.1.2"  
/function="Coronavirus 3' UTR pseudoknot stem-loop 2"  
3'UTR  
stem\_loop  
29675..29903  
29728..29768  
/inference="COORDINATES:  
profile:Rfam-release-14.1:RF00164, Infernal:1.1.2"  
/note="basepair exception: alignment to the Rfam model  
implies coordinates 29740:29758 form a noncanonical C:T  
basepair, but the homologous positions form a highly  
conserved C:G basepair in other viruses, including SARS

(NC\_004718.3)"

/function="Coronavirus 3' stem-loop II-like motif (s2m)"

ORIGIN

1 attaaagggtt tataccttcc caggtaacaa accaaccaac ttgcgtatctc ttgttagatct  
61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcattgc ttagtgcact  
121 cacgcagttat aattaataac taattactgt cgttgcacagg acacgagtaa ctgcgtctatc  
181 ttctgcagge tgcttacggt ttgcgtccgtg ttgcagccga tcattcagcac atctagggtt  
241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgc cctggttca acgagaaaaac  
301 acacgtccaa ctcagtttgc ctgtttaca ggttcgccgac gtgcgttac gtggctttgg  
361 agactccgtg gaggaggtct tatcagaggg acgtcaacat cttaaagatg gcacttgtgg  
421 cttagtagaa gttgaaaaag gcgtttgcc tcaacttgaa cagccctatg tgttcatcaa  
481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact  
541 cgaaggcatt cagtacggtc gtagtggta gacacttgg gtccttgc ctcatgtgg  
601 cgaaatacca gtggcttacc gcaagggtct tcttcgttaag aacggtaata aaggagctgg  
661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga  
721 tccttatgaa gatttcaag aaaactggaa cactaaacat agcagtggta ttaccgtga  
781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg  
841 ccctgatggc tacccttttgc agtgcattaa agaccttcta gcacgtgctg gtaaagcttc  
901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg  
961 tgaacatgag catgaaatttgc ttggtaacac ggaacgttct gaaaagagct atgaatttgc

*Fig. 12C* continued

SUBSTITUTE SHEET (RULE 26)

1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa  
1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aaaagaaaaaa  
1141 gcttgatggc tttatggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg  
1201 caaccaaatg tgcccttcaa ctctcatgaa gtgtgatcat tgtggtaaaa cttcatggca  
1261 gacgggcgat tttgttaaag ccacttgcga attttgcgc actgagaatt tgactaaaga  
1321 aggtgccact acttgcgttt acttacccc aaatgctgtt gttaaaattt attgtccagc  
1381 atgtcacaat tcagaagtag gacctgagca tagtcttgcc gaataccata atgaatctgg  
1441 cttgaaaacc attcttcgta agggtggcgt cactattgcc tttggaggct gtgtgttctc  
1501 ttatgttggt tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg  
1561 ttgttaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga  
1621 aataactccaa aaagagaaaag tcaacatcaa tattgttggt gactttaaac ttaatgaaga  
1681 gatcgccatt attttggcat ctttttctgc ttccacaagt gctttgtgg aaactgtgaa  
1741 aggtttggat tataaagcat tcaaacaat ttttgcgtt ttttgcgtt ttttgcgtt  
1801 aaaaggaaaaa gctaaaaaaag gtgcctggaa tattggtaaa cagaaatcaa tactgagtcc  
1861 tctttatgca tttgcattcag aggctgcgt ttttgcgtt ttttgcgtt  
1921 tgaaaactgct caaaaattctg tgcgtttt acagaaggcc gctataacaa tactagatgg  
1981 aatttcacag tattcactga gactcattga tgctatgatg ttcacatctg atttggctac  
2041 taacaatcta gttgtatgg cttacattac aggtgggttt gttcgttga ctgcgttgcgt  
2101 gctaactaac atctttggca ctgtttatga aaaactcaaa cccgtccttgcgttgcgtt  
2161 agagaagttt aaggaagggtg tagagtttct tagagacgggt tggaaatttgcgttgcgtt  
2221 ctcaacctgt gcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt

Fig. 12C  
*continued*

SUBSTITUTE SHEET (RULE 26)

2281 ggagagtgtt cagacattct ttaagcttgt aaataaaattt ttggctttgt gtgctgactc  
2341 tatcattattt ggtggagcta aacttaaagc cttgaattta ggtgaaaacat ttgtcacgca  
2401 ctcaaaggga ttgtacagaa agtgtgttaa atccagagaa gaaactggcc tactcatgcc  
2461 tctaaaagcc cccaaaagaaa ttatcttctt agagggagaa acacttcccc cagaagtgtt  
2521 aacagagggaa gttgtcttga aaactggtga tttacaacca ttagaacaac ctactagtga  
2581 agctgttcaa gctccattgg ttggcacacc agtttgtatt aacgggctta tggtgctcga  
2641 aatcaaagac acagaaaaagt actgtgcct tgcacctaattt atgatggtaa caaacaatac  
2701 cttcacactc aaaggcggtg caccaacaaa ggttactttt ggtgatgaca ctgtgataga  
2761 agtgcaggt tacaagagt tgaatatcac ttttgaactt gatgaaagga ttgataaaagt  
2821 acttaatgag aagtgcctcg cctatacagt tgaactcggt acagaagtaa atgagtcgc  
2881 ctgtgttgtg gcagatgctg tcataaaaaac tttgcaacca gtatctgaat tacttacacc  
2941 actgggcatt gatttagatg agtggagttt ggctacatac tacttattt gatgactctgg  
3001 tgagtttaaa ttggcttcac atatgtattt ttctttctac cctccagatg aggatgaaga  
3061 agaagggtat tgtgaagaag aagagtttga gccatcaact caatatgagt atggtaactga  
3121 agatgattac caaggtaaac ctttggattt tgggccact tctgctgatc ttcaacactga  
3181 agaagagcaa gaagaagatt ggtagatga tgatagtcaa caaactgttgcgtcaacaaga  
3241 cggcagttagt gacaatcaga caactactat tcaaaacaattt gttgagggttc aacctcaatt  
3301 agagatggaa cttacaccag ttgttcagac tattgaagtg aatagtttta gtggttattt  
3361 aaaacttact gacaatgtat acattaaaaa tgcagacattt gtggaaagaag ctaaaaaggt  
3421 aaaaccaaca gtggttgttta atgcagccaa tttttttttt aaacatggag gaggtgttgc

Fig. 12C *continued*

SUBSTITUTE SHEET (RULE 26)

3481 aggagcctta aataaggcta ctaacaatgc catgcaagtt gaatctgatg attacatacg  
3541 tactaatgga ccacttaaag tgggtggtag ttgtgttttta agcggacaca atcttgctaa  
3601 acactgtctt catgttgtcg gcccaaatgt taacaaaggt gaagacattc aacttcttaa  
3661 gagtgcttat gaaaattttta atcagcacga agttctactt gcaccattat tatcagctgg  
3721 tatttttgtt gctgacccta tacattctt aagagttgt gtagatactg ttgcacaaaa  
3781 tgtctactta gctgtcttg ataaaaatct ctatgacaaa cttgtttcaa gctttttgga  
3841 aatgaagagt gaaaagcaag ttgaacaaaaa gatcgctgag attcctaaag aggaagttaa  
3901 gccatttata actgaaaagta aaccttcagt tgaacagaga aaacaagatg ataagaaaaat  
3961 caaagcttgt gttgaagaag ttacaacaac tctggaagaa actaagttcc tcacagaaaa  
4021 cttgttactt tatattgaca ttaatggcaa tcttcatcca gattctgcca ctcttggtag  
4081 tgacattgac atcactttct taaagaaaaga tgctccatat atagtgggtg atgttggca  
4141 agagggtgtt ttaactgctg tggttataacc tactaaaaag gctggggca ctactgaaat  
4201 gctagcgaaa gctttgagaa aagtgcacaa agacaattat ataaccattt acccggtca  
4261 gggtttaaat ggttacactg tagaggaggc aaagacagtg cttaaaaagt gtaaaagtgc  
4321 ctttacatt ctaccatcta ttatctctaa tgagaagcaa gaaattcttgc gaactgttcc  
4381 ttggaatttg cgagaaatgc ttgcacatgc agaagaaaaca cgcaaattaa tgctgtctg  
4441 tgtggaaact aaagccatag tttcaactat acagcgtaaa tataagggtt ttaaaataca  
4501 agagggtgtg gttgattatg gtgcttagatt ttacttttac accagtaaaa caactgttagc  
4561 gtcacttatac aacacactta acgatctaaa tgaaactctt gttacaatgc cacttggcta  
4621 tgtaacacat ggcttaaatt tggaagaagc tgctcggtat atgagatctc tcaaagtgcc  
4681 agctacagtt tctgtttctt cacctgatgc tgttacagcg tataatggtt atcttacttc

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

4741 ttcttctaaa acacctgaag aacattttat tgaaaccatc tcacttgctg gttcctataa  
4801 agattggcc tattctggac aatctacaca actaggata gaatttctta agagaggtga  
4861 taaaagtgtt tattacacta gtaatcctac cacattccac ctagatggtg aagttatcac  
4921 ctggacaat cttaagacac ttctttctt gagagaagtg aggactatta aggtgtttac  
4981 aacagtagac aacattaacc tccacacgca agttgtggac atgtcaatga catatggaca  
5041 acagtttgtt ccaacttatt tggatggagc tcatgttact aaaataaaac ctcataattc  
5101 acatgaaggt aaaacatttt atgttttacc taatgatgac actctacgtg ttgaggctt  
5161 tgagtactac cacacaactg atcctagttt tctggtagg tacatgtcag cattaaatca  
5221 cactaaaaag tggaaatacc cacaagttaa tggtttaact tctattaaat gggcagataa  
5281 caactgttat ctgccactg cattgttaac actccaacaa atagagttga agtttaatcc  
5341 acctgctcta caagatgctt attacagagc aagggttgtt gaagctgcta acttttgtc  
5401 acttatctta gcctactgta ataagacagt aggtgagttt ggtgatgtt gagaaacaat  
5461 gagttacttg tttcaacatg ccaatttgc ttcttgcaaa agagtcttga acgtgggtgt  
5521 taaaacttgc ggacaacagc agacaaccct taagggtgtt gaagctgtt tgcacatgg  
5581 cacactttct tatgaacaat ttaagaaagg tggcagata cttgtacgt gtggtaaaca  
5641 agctacaaaa tatctgttac aacaggagtc acctttgtt atgtgtcag caccacctgc  
5701 tcagtatgaa cttaagcatg gtacatttac ttgtgttagt gagtacactg gtaattacca  
5761 gtgtggtcac tataaacata taacttctaa agaaactttt tattgcatacg acgggtgttt  
5821 acttacaaaag tcotcagaat acaaagggtcc tattacggat gttttctaca aagaaaaacag  
5881 ttacacaaca accataaaac cagttactt taaattggat ggtgttgtt gtacagaaaat

*Fig. 12C continued*

SUBSTITUTE SHEET

(RULE 26)

5941 tgaccctaag ttggacaatt attataagaa agacaattct tatttcacag agcaaccaat  
6001 tgatcttgta ccaaaccaac catatccaaa cgcaagctc gataattttt agtttgtatg  
6061 tgataatatc aaatttgctg atgatttaaa ccagtttaact ggttataaga aacctgcttc  
6121 aagagagctt aaagttacat ttttccctga cttaaatggt gatgtggtgg ctattgatta  
6181 taaacactac acaccctttaaagg agctaaatttgc ttacataaaac ctattgtttg  
6241 gcatgttaac aatgcaacta ataaagccac gtataaacc aataacctggt gtatacgttg  
6301 tctttggagc acaaaaaccag ttgaaacatc aaattcgttt gatgtactga agtcagagga  
6361 cgcgccaggga atggataatc ttgcctgcga agatctaaaa ccagtctctg aagaagttagt  
6421 ggaaaaatcct accatacaga aagacgttct tgagtgtaat gtggaaaacta ccgaagttgt  
6481 aggagacatt atacttaaac cagcaaataa tagttaaaaa attacagaag aggttggcca  
6541 cacagatcta atggctgctt atgttagacaa ttcttagtctt actattaaaga aacctaataatga  
6601 attatctaga gtatttagtt tgaaaaaccct tgctactcat ggtttagctg ctgttaatag  
6661 tgtcccttgg gatactatacg ctaattatgc taagcctttt cttaacaaag ttgttagtac  
6721 aactactaac atagttcacac ggtgtttaaa ccgtgtttgt actaattata tgccttattt  
6781 cttaacttta ttgctacaat tgtgtacttt tactagaagt acaaattcta gaattaaagc  
6841 atctatgccg actactatacg caaagaatac tgttaagagt gtcggtaaat tttgtctaga  
6901 ggcttcattt aattattttga agtcacccaa tttttctaaa ctgataaaata ttataatttg  
6961 gttttacta ttaagtgttt gccttaggttc ttatctac tcaaccgctg cttaggtgt  
7021 tttaatgtct aatttaggca tgccttctta ctgtactggt tacagagaag gctatttgaa  
7081 ctctactaat gtcactattg caacctactg tactggttct ataccttgc gtgtttgtct  
7141 tagtggttta gattctttag acacctatcc ttcttttagaa actatacataaa ttaccatttc

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

7201 atcttttaaa tgggatttaa ctgctttgg cttagttgca gagtggttt tggcatatat  
 7261 tctttcact aggtttttct atgtacttgg attggctgca atcatgcaat tgaaaaatcag  
 7321 ctatttgca gtacattta ttagtaattc ttggcttatg tggtaataa ttaatcttgt  
 7381 acaaatggcc ccgatttcag ctatggtag aatgtacatc ttctttgcatttatttattt  
 7441 tgtatggaaa agttatgtgc atgtttaga cggtttagt tcataactt gtatgtatgt  
 7501 ttacaaacgt aatagagcaa caagagtcga atgtacaactt attgttaatg gtgttagaaag  
 7561 gtcctttat gtctatgcta atggaggtaa aggctttgc aaactacaca attggaaattt  
 7621 tgttaattgt gatacatttct gtgtgttag tacatttattt agtgtatgaag ttgcgagaga  
 7681 cttgtcacta cagtttaaaa gaccaataaa tcctactgac cagtcttctt acatcggttga  
 7741 tagtgttaca gtqaagaatg gttccatcca tctttacttt gataaaagctg gtcaaaaagac  
 7801 ttatgaaaga cattctctct ctcattttgt taacttagac aacctgagag ctaataaacac  
 7861 taaagggttca ttgccttatta atgttatagt ttttgtatgtt aaatcaaaaat gtgaagaatc  
 7921 atctgcaaaa tcagcgtctg tttactacag tcagcttatg tgtcaaccata tactgttact  
 7981 agatcaggca ttagtgtctg atgttggta tagtgcggaa gttgcagtta aaatgttga  
 8041 tgcttacgtt aatacgttt catcaacttt taacgtacca atggaaaaac tcaaaaacact  
 8101 agttgcaact gcagaagctg aacttgcaaa gaatgtgtcc ttagacaatg tcttatctac  
 8161 ttttatttca gcagctcggc aagggtttgt tgattcagat gtagaaaacta aagatgttgt  
 8221 tgaatgtctt aaattgtcac atcaatctga catagaagtt actggcgata gttgtataaa  
 8281 ctatatgctc acctataaca aagttggaaa catgacacccc cgtgacacctg gtgcttgtat  
 8341 tgactgttagt gcgcgtcata ttaatgcgca ggtgcaaaa agtcacaaca ttgctttgat

*Fig. 12C* *continued*

*Fig. 12C*  
*continued*

SUBSTITUTE SHEET (RULE 26)

```

8401 atggaacgtt aaagatttca tgtcattgtc tgaacaacta cgaaaacaaa tacgtagtgc
8461 tgctaaaaag aataacttac cttaaaggtt gacatgtgca actactagac aagttgttaa
8521 tgttgtaca acaaagatag cacttaaggg tggtaaaatt gttaataatt gggtgaagca
8581 gttaattaaa gttacacttg tgttcccttt tgttgctgct attttctatt taataaacacc
8641 tgttcatgtc atgtctaaac atactgactt ttcaagtgaa atcataggat acaaggctat
8701 tgatggtggt gtcactcggt acatagcattc tacagatact tggtttgcta acaaacatgc
8761 tgattttgac acatggtttta gccagcgtgg tggtagttat actaatgaca aagttgccc
8821 attgatttgct gcagtcataa caagagaagt gggttttgtc gtgcctgggtt tgccctggcac
8881 gatattacgc acaactaatg gtgacttttt gcatttttta cctagagttt ttagtgcagt
8941 tggttaacatc tgttacacac catcaaaaact tataagatgtac actgactttg caacatcagc
9001 ttgtgttttgcgtgctgaat gtacaatttt taaagatgct tctggtaagc cagtaccata
9061 ttgttatgtat accaatgtac tagaaggttc tgttgcttat gaaagtttac gccctgacac
9121 acgttatgtc ctcattggatg gctctattat tcaatttcct aacacctacc ttgaagggttc
9181 tgtagatgtg gtaacaactt ttgattctga gtactgttagg cacggcactt gtgaaagatc
9241 agaagctgggt gtttgtgtat ctactagtgg tagatgggtt cttaaacaatg attattacag
9301 atcttacca ggagttttct gtgggtgtaga tgctgtaaat ttacttacta atatgtttac
9361 accactaatt caacctatttgcgtttggatc catatcagca tctatagtag ctgggtgttat
9421 tgtagctatc gtagtaacat gccttgccta ctatattatg aggttttagaa gagctttgg
9481 tgaatacagt catgttagtttgcctttaatac tttacttacc cttatgtcat tcactgtact
9541 ctgtttaaca ccagtttact catttttacc tggtgtttat tctgttattt acttgtactt
9601 gacattttat cttactaatg atgtttcttt ttttagcacat attcagtggatggatggatgtt

```

9661 cacaccttta gtacctttct ggataacaat tgcttatatac atttgtattt ccacaaagca  
9721 tttctattgg ttcttttagta attacctaaa gagacgtgta gtcttaatg gtgttcctt  
9781 tagtactttt gaagaagctg cgctgtgcac cttttgtta aataaagaaa tgtatctaaa  
9841 gttgcgtagt gatgtgctat tacctcttac gcaatataat agataacttag ctctttataa  
9901 taagtacaag tatttttagtg gagcaatgga tacaacttagc tacagagaag ctgcttgg  
9961 tcacatcgca aaggctctca atgacttcag taactcaggt tctgatgttc tttaccaacc  
10021 accacaaaacc tctatcacct cagctgttt gcagagtgg tttagaaaaa tggcattccc  
10081 atctggtaaa gttgagggtt gtatggtaca agtaacttgc ggtacaacta cacttaacgg  
10141 tctttggctt gatgacgttag ttactgtcc aagacatgtg atctgcaccc ctgaagacat  
10201 qcttaaccct aattatgaag atttactcat tcgtaagtct aatcataatt tcttggtaca  
10261 ggctggtaat gttcaactca gggttattgg acattctatg caaaattgtg tacttaagct  
10321 taaggttgat acagccaatc ctaagacacc taagtataag tttgttcga ttcaaccagg  
(RULE 10381 acagactttt tcagtgttag cttgttacaa tggttcacca tctgggttt accaatgtgc  
10441 tatgaggccc aatttcacta ttaagggttc attccttaat gggtcatgtg gtatgtttgg  
26) 10501 tttaaacata gattatgact gtgtctttt ttgttacatg caccatatgg aattaccaac  
10561 tggagttcat gctggcacag acttagaagg taactttat ggaccttttgg ttgacaggca  
10621 aacagcacaa gcagctggta cggacacaac tattacagtt aatgttttag cttgggtgt  
10681 cgctgctgtt ataaatggag acaggtggtt tctcaatcga tttaccacaa ctcttaatga  
10741 cttaaacctt gtggctatga agtacaatta tgaacctcta acacaagacc atgttgacat  
10801 actaggaccc ttctctgctc aaactggaat tgccgtttt gatatgtgtg cttcattaaa

Fig. 12C *continued*

SUBSTITUTE SHEET  
(RULE 26)

10861 agaattactg caaaatggta tgaatggacg taccatattg ggtagtgtt tattagaaga  
10921 tgaatttaca cctttgatg ttgttagaca atgctcaggt gttactttcc aaagtgcagt  
10981 gaaaagaaca atcaaggta cacaccactg gttgttactc acaatttga ctcaacttt  
11041 agtttagtc cagagtactc aatggtctt gttcttttt ttgtatgaaa atgcctttt  
11101 acctttgct atgggtatta ttgctatgtc tgctttgca atgatgttg tcaaacadaa  
11161 gcatgcattt ctctgttgtt ttttgttacc ttctcttgcc actgttagctt attttaatat  
11221 ggtctatatg cctgctagtt gggtgatgcg tattatgaca tgggtggata tgggtgatac  
11281 tagttgtct ggttttaagc taaaagactg tgttatgtat gcatcagctg tagtgttact  
11341 aatccttatg acagcaagaa ctgtgtatga tgatggtgct aggagagtgt ggacacttat  
11401 gaatgtctg acactcgttt ataaagtttta ttatggtaat gcttagatc aagccatttc  
11461 catgtgggtt ctataaatct ctgttacttc taactactca ggtgttagtta caactgtcat  
11521 gttttggcc agaggtattg ttttatgtg tgttgagttat tgcccttattt ttttcataaac  
11581 tggtaataca cttcagtgtttaatgctatgtt ttattgtttc ttaggctattttt tttgtacttg  
11641 ttactttggc ctctttgtt tactcaaccg ctactttaga ctgacttttg gtgtttatgt  
11701 ttacttagtt tctacacagg agtttagata tatgaattca cagggactac tcccacccaa  
11761 gaatagcata gatgccttca aactcaacat taaattgttgg ggtgttgg gcaaaccttg  
11821 tatcaaagta gccactgtac agtctaaaat gtcagatgtt aagtgcacat cagtagtctt  
11881 actctcagtt ttgcaacaac tcagagtaga atcatcatct aaattgtggg ctcaatgtgt  
11941 ccagttcacac aatgacattc tcttagctaa agatactact gaagcctttg aaaaaatgg  
12001 ttcaactactt tctgtttgc tttccatgca gggtgctgtt gacataaaca agctttgtga  
12061 agaaatgctg gacaacaggg caaccttaca agctatagcc tcagagttt gttcccttcc

Fig. 12C  
continued

SUBSTITUTE SHEET (RULE 26)

12121 atcatatgca gctttgcta ctgctcaaga agcttatgag caggctgttgc  
 12181 ttctgaagtt gttcttaaaa agttgaagaa gtctttgaat gtggctaaat ctgaatttga  
 12241 ccgtgatgca gccatgcaac gtaagttgga aaagatggct gatcaagcta tgacccaaatt  
 12301 gtataaacag gtagatctg aggacaagag ggcaaaagt actagtgcta tgcagacaat  
 12361 gctttcact atgcttagaa agttggataa tgatgcactc aacaacatta tcaacaatgc  
 12421 aagagatggt tgtgtccct tgaacataat acctcttaca acagcagcca aactaatggt  
 12481 tgtcataccca gactataaca catataaaaa tacgtgtgat ggtacaacat ttacttatgc  
 12541 atcagcattt gggaaaatcc aacagggtgt agatgcagat agtaaaattt ttcaacttag  
 12601 taaaatttgt atggacaatt cacctaattt agcatggcct ttattgtaa cagctttaag  
 12661 ggccaatttc gotgtcaaatt tacagaataa ttagcttagt cctgttgcac tacgacagat  
 12721 gtttgtgct gccggtaacta cacaaactgc ttgcactgat gacaatgcgt tagcttacta  
 12781 caacacaaca aagggaggta gttttgtact tgcaactgtt tccgattttc aggatttgaa  
 12841 atgggctaga ttccctaaga gtgatggAAC tggtaacttac tatacagaac tggAACCC  
 12901 tttaggttt gttacagaca cacctaaagg tctaaagtgt aagtattttt actttattaa  
 12961 aggattaaac aacctaata gaggtatggt acttggtagt ttagctgccca cagtaagtct  
 13021 acaagctggt aatgcaacag aagtgcctgc caattcaact gtatttattt tctgtgtttt  
 13081 tgctgttagt gctgctaaag cttacaaaga ttatcttagt agtgggggac aaccaatcac  
 13141 taattgtgtt aagatgtgt gtacacacac tggtaactggt caggcaataa cagttacacc  
 13201 ggaagccaat atggatcaag aatcccttgg tggtaactcg tggtaactgt actgcogttg  
 13261 ccacatagat catccaaatc ctaaaggatt ttgtgactt aaggtaagt atgtacaaat

*Fig. 12C continued*

*Fig. 12C*  
*continued*

SUBSTITUTE SHEET  
(RULE 26)  
13321 acctacaact tgtgctaatt accctgtggg ttttacactt aaaaacacag tctgtaccgt  
13381 ctgcggtatg tggaaaggtt atggctgttag ttgtgatcaa ctccgcgaac ccatgcttca  
13441 gtcagctgat gcacaatcgt ttttaaacgg gtttgcggtg taagtgcagc ccgtcttaca  
13501 ccgtgcggca caggcactag tactgatgtc gtatacaggg cttttgacat ctacaatgat  
13561 aaagtagctg gtttgctaa attcctaaaa actaattgtt gtgcgttcca agaaaaggac  
13621 gaagatgaca atttaattga ttcttacttt gtagttaaga gacacacttt ctctaactac  
13681 caacatgaag aaacaattta taatttactt aaggattgtc cagctgtgc taaacatgac  
13741 ttctttaagt ttagaataga cggtgacatg gtaccacata tatcacgtca acgtcttact  
13801 aaatacacaat tggcagacct cgtctatgct ttaaggcatt ttgatgaagg taattgtgac  
13861 acattaaaag aaataacttgt cacatacaat tgttgtgatg atgattattt caataaaaag  
13921 gactggatg attttgtaga aaaccagat atattacgcg tatacgccaa cttaggtgaa  
13981 cgtgtacgcc aagctttgtt aaaaacagta caattctgtg atgccatgcg aaatgctggt  
14041 attgttggtg tactgacatt agataatcaa gatctcaatg gtaactggta tgatttcggt  
14101 gattcatac aaaccacgcc agtagtgga gttctgttg tagattctt ttattcattg  
14161 ttaatgccta tattaacctt gaccaggct ttaactgcag agtcacatgt tgacactgac  
14221 ttaacaaagc ttacattaa gtgggattt gtaaaaatatg acttcacgga agagaggtt  
14281 aaactcttt accgttattt taaatattgg gatcagacat accacccaaa ttgtgttaac  
14341 tgttggatg acagatgcat tctgcattgt gcaaacttta atgtttatt ctctacagt  
14401 ttcccaccta caagtttgg accactagtg agaaaaatat ttgttgatgg tgttccattt  
14461 gtagttcaa ctggatacca cttcagagag ctaggtgtt tacataatca ggatgtaaac  
14521 ttacatagct ctagacttag ttttaaggaa ttacttgcgt atgctgctga ccctgctatg

SUBSTITUTE SHEET (RULE 26)

14581 cacgctgctt ctggtaatct attactagat aaacgcacta cgtgctttc agtagctgca  
 14641 cttactaaca atgttgcttt tcaaactgtc aaacccggta attttaacaa agacttctat  
 14701 gactttgctg tgtctaaggg tttctttaag gaaggaagtt ctgttgaatt aaaacacttc  
 14761 ttctttgctc aggatggtaa tgctgctatc agcgattatg actactatcg ttataatcta  
 14821 ccaacaatgt gtgatatcag acaactacta tttgttagttg aagttgttga taagtacttt  
 14881 gattgttacg atggtggtcg tattaatgct aaccaagtca tcgtcaacaa cctagacaaa  
 14941 tcagctgggt ttccatttaa taaatggggt aaggctagac tttattatga ttcaatgagt  
 15001 tatgaggatc aagatgcact tttcgcatat acaaaaacgta atgtcatccc tactataact  
 15061 caaatgaatc ttaagtatgc cattagtgca aagaatagag ctcgcaccgt agctgggtgc  
 15121 tctatctgta gtactatgac caataagacag tttcatcaa aattattgaa atcaatagec  
 15181 gccactagag gagctactgt agtaattgga acaagcaaat tctatggtgg ttggcacaaac  
 15241 atgttaaaaaa ctgtttatag tgatgtagaa aaccctcacc ttatgggttg ggattatcct  
 15301 aaatgtgata gagccatgcc taacatgctt agaattatgg cctcaacttgt tcttgctcgc  
 15361 aaacatacaa cgtgtttag cttgtcacac cgtttctata gattagctaa tgagtgtgct  
 15421 caagtattga gtgaaatggt catgtgtggc ggttcactat atgttaaacc aggtggaaacc  
 15481 tcatacaggag atgccacaaac tgotttatgct aatagtgtt ttaacatttg tcaagctgtc  
 15541 acggccaatg ttaatgcact tttatctact gatggtaaca aaattgcccga taagtatgtc  
 15601 cgcaatttac aacacagact ttatgagtgt ctctatagaa atagagatgt tgacacacagac  
 15661 tttgtgaatg agttttacgc atattgcgt aaacatttct caatgatgat actctctgac  
 15721 gatgctgttg tgtgtttcaa tagcaattat gcatctcaag gtctagtggc tagcataaaag

*Fig. 12C continued*

SUBSTITUTE SHEET (RULE 26)

15781 aactttaagt cagttcttta ttatcaaaaac aatgtttta tgtctgaagc aaaatgttgg  
15841 actgagactg accttactaa aggacctcat gaattttgct ctcaacatac aatgcttagtt  
15901 aaacagggtg atgattatgt gtaccttcct taccagatc catcaagaat cctagggcc  
15961 ggctgtttg tagatgatat cgtaaaaaaca gatggtacac ttatgattga acggttcgtg  
16021 tcttagcta tagatgctta cccacttact aaacatccta atcaggagta tgctgatgtc  
16081 tttcatttgt acttacaata cataagaaaag ctacatgatg agttaacagg acacatgtta  
16141 gacatgtatt ctgttatgct tactaatgat aacacttcaa ggtattggga acctgagttt  
16201 tatgaggcta tgtacacacc gcatacagtc ttacaggctg ttggggcttg tgttcttgc  
16261 aattcacaga cttcattaag atgtggtgct tgcatacgta gaccattctt atgttgtaaa  
16321 tgctgttacg accatgtcat atcaacatca cataaattag tcttgtctgt taatccgtat  
16381 gtttgcaatg ctccagggttg tgatgtcaca gatgtgactc aactttactt aggaggtatg  
16441 agctattatt gtaaatcaca taaaccaccc attagtttc cattgtgtgc taatggacaa  
16501 gttttgggtt tatataaaaa tacatgtgtt ggtagcgata atgttactga cttaatgca  
16561 attgcaacat gtgactggac aaatgctggt gattacattt tagctaacac ctgtactgaa  
16621 agactcaagc ttttgcagc agaaacgctc aaagctactg aggagacatt taaactgtct  
16681 tatggtattt ctactgtacg tgaagtgctg tctgacagag aattacatct ttcatggaa  
16741 gttggtaaac ctagaccacc acttaaccga aattatgtct ttactggta tcgtgttaact  
16801 aaaaacagta aagtacaaaat aggagagtac acctttgaaa aaggtgacta tggtgatgt  
16861 gttgtttacc gaggtacaac aacttacaaa ttaaatgttg gtgattattt tggctgtaca  
16921 tcacatacag taatgccatt aagtgcaccc acactagtgc cacaagagca ctatgttaga  
16981 attactggct tatacccaac actcaatatc tcagatgagt tttctagcaa tgttgcaaat

Fig. 12C  
continued

SUBSTITUTE SHEET (RULE 26)

17041 tatcaaaaagg ttggtatgca aaagtattct acactccagg gaccacctgg tactggtaag  
 17101 agtcattttg ctattggcct agctctctac tacccttctg ctcgcatagt gtatacagct  
 17161 tgctctcatg ccgctgttga tgcactatgt gagaaggcat taaaatattt gcctatagat  
 17221 aaatgttagta gaattataacc tgcacgtgct cgtgttagagt gtttgataa attcaaagtg  
 17281 aattcaacat tagaacagta tgtctttgt actgtaaatg cattgcctga gacgacagca  
 17341 gatatagttg tccttgatga aatttcaatg gccacaaaatt atgatttgag tgttgtcaat  
 17401 gccagattac gtgctaagca ctatgtgtac attggcgacc ctgctcaatt acctgcacca  
 17461 cgcacattgc taactaaggg cacactagaa ccagaatatt tcaattcagt gtgttagactt  
 17521 atgaaaaacta taggtccaga catgttcctc ggaacttgtc ggcgttgc tgctgaaatt  
 17581 gttgacactg ttagtgctt ggtttatgt aataagctta aagcacataa agacaatca  
 17641 gctcaatgt taaaaatgtt ttataagggt gttatcacgc atgatgtttc atctgcaatt  
 17701 aacaggccac aaataggcgt ggtaagagaa ttcccttacac gtaaccctgc ttggagaaaa  
 17761 gctgtcttta ttccaccta taattcacag aatgctgtag cctcaaagat tttggacta  
 17821 ccaactcaaa ctgttgattc atcacaggc tcagaatatg actatgtcat attcaactcaa  
 17881 accactgaaa cagctcaactc ttgtaatgta aacagattt aatgttgctat taccagagca  
 17941 aaagtaggca tactttgcat aatgtctgtat agagacctt atgacaagtt gcaatttaca  
 18001 agtcttgaaa ttccacgtag gaatgtggca actttacaag ctgaaaatgt aacaggactc  
 18061 tttaaagatt gtagtaaggt aatcaactggg ttacatccta cacaggcacc tacacacctc  
 18121 agtgttgaca ctaaattcaa aactgaaggt ttatgtgttgc acatacctgg catacctaag  
 18181 gacatgacct atagaagact catctctatg atgggttttta aaatgaatta tcaagttaat

*Fig. 12C continued*

Fig. 12C  
continued

SUBSTITUTE SHEET (RULE 26)

18241 ggttacccta acatgtttat caccgcgaa gaagctataa gacatgtacg tgcatggatt  
18301 ggcttcgatg tcgaggggtg tcatgctact agagaagctg ttggcacca tttacctta  
18361 cagcttaggtt tttctacagg tgttaaccta gttgctgtac ctacaggta tggtgataca  
18421 cctaataata cagattttc cagagttgt gctaaaccac cgccctggaga tcaatttaaa  
18481 cacccatatac cacttatgtt caaaggactt ctttggaaatg tagtgcgtat aaagattgtt  
18541 caaatgttaa gtgacacact taaaaatctc tctgacagag tcgtatttgt cttatggca  
18601 catggcttg agttgacatc tatgaagtat tttgtgaaaaa taggacctga gcgcacctgt  
18661 tgtctatgtt atagacgtgc cacatgcttt tccactgctt cagacactta tgcctgttgg  
18721 catcattcta ttggatttga ttacgtctat aatccgtta tgattgtatgt tcaacaatgg  
18781 ggtttacag gtaacctaca aagcaaccat gatctgtatt gtcaagtcca tggtaatgca  
18841 catgttagcta gttgtgatgc aatcatgact aggtgtctag ctgtccacga gtgctttgtt  
18901 aagcgtgttg actggactat tgaatatcct ataattggtg atgaactgaa gattaatgcg  
18961 gctttagaa aggttcaaca catggttgtt aaagctgcatt tattagcaga caaattcccc  
19021 gttcttcacg acattggtaa ccctaaagct attaagtgtt tacctcaagc tgatgttagaa  
19081 tggaaagttct atgatgcaca gcctttagt gacaaagctt ataaaataga agaattattc  
19141 tattttatg ccacacattt tgacaaaattt acagatggtg tatgcctatt ttggaaattgc  
19201 aatgtcgata gatatcctgc taattccatt gttttagat ttgacactag agtgcstatct  
19261 aaccttaact tgcctggttt tgatggtggc agtttgtatg taaaataaaca tgcattccac  
19321 acaccagtt ttgataaaaag tgctttgtt aattttaaaac aattaccatt tttctattac  
19381 tctgacagtc catgtgagtc tcatggaaaaa caagtagtgt cagatataaga ttatgtacca  
19441 ctaaagtctg ctacgtgtat aacacgttgc aatttaggtt gtgctgtctg tagacatcat

SUBSTITUTE SHEET  
(RULE 26)

19501 gctaatgagt acagattgt a tctcgatgct tataacatga tgatctcagc tggcttttagc  
19561 ttgtgggtt acaaacaatt tgataacttat aacctctgga acactttac aagacttcag  
19621 agtttagaaa atgtggctt taatgttgt aataaggac actttgatgg acaacagggt  
19681 gaagtaccag tttctatcat taataaacact gtttacacaa aagttgatgg tttgtatgt  
19741 gaattgttg aaaataaaaac aacattacct gttaatgtag catttgagct ttgggctaag  
19801 cgcaacatta aaccagtacc agaggtgaaa atactcaata atttgggtgt ggacattgct  
19861 gctaatactg tgatctggg a ctacaaaaga gatgctccag cacatatatc tactatttgt  
19921 gttgttcta tgactgacat agccaagaaa ccaactgaaa cgatttgc accactcact  
19981 gtctttttg atggtagagt tgatggtaa gtagacttat tttagaaatgc ccgtaatgg  
20041 gttcttatta cagaaggtag tttttttttt ttacaaccat ctgttaggtcc caaacaagct  
20101 agtcttaatg gagtcacatt aattggagaa gccgtaaaaa cacagttcaa ttattataag  
20161 aaagttgatg gtgttgtcca acaattacct gaaacttaact ttactcagag tagaaattta  
20221 caagaattta aacccaggag tcaaattggaa attgatttct tagaatttagc tatggatgaa  
20281 ttcattgaac ggtataaatt agaaggctat gccttcgaac atatcgat tggagatttt  
20341 agtcatagtc agtttaggtgg tttacatcta ctgattggac tagctaaacg ttttaaggaa  
20401 tcacctttt aattagaaga ttttatttct atggacagta cagttaaaaa ctatttcata  
20461 acagatgogc aaacagggttc atctaagtgt gtgtgttctg ttattgattt attacttgc  
20521 gatttgttg aaataataaa atccccagat ttatctgttag tttctaaagggt tgtcaaagg  
20581 actattgact atacagaaat ttcatttatg ctgggtgt aagatggcca tgttagaaaca  
20641 ttttacccaa aattacaatc tagtcaagcgt tggcaacccgg gtgttgctat gcctaatctt

Fig. 12C *continued*

*Fig. 12C*  
continued

SUBSTITUTE SHEET  
(RULE 26)

20701 tacaaaatgc aaagaatgct attagaaaaag tgtgaccttc aaaattatgg tgatagtgc  
 20761 acattaccta aaggcataat gatgaatgtc gcaaaatata ctcaactgtg tcaatattta  
 20821 aacacattaa cattagctgt accctataat atgagagtta tacatttgg tgctgggtct  
 20881 gataaaggag ttgcaccagg tacagctgtt ttaagacagt gggtgcctac gggtaacgctg  
 20941 cttgtcgatt cagatcttaa tgactttgtc tctgatgcag attcaacttt gattgggtgat  
 21001 tgtgcaactg tacatacagc taataaatgg gatctcatia tttagtgatat gtacgaccct  
 21061 aagactaaaa atgttacaaa agaaaaatgac tctaaagagg gttttttcac ttacatttgt  
 21121 gggtttatac aacaaaagct agctcttggg ggttccgtgg ctataaaagat aacagaacat  
 21181 tcttggaaatg ctgatctta taagctcatg ggacacttcg catggtggac agcctttgtt  
 21241 actaatgtga atgcgtcatc atctgaagca ttttaattt gatgtaatta tcttggcaaa  
 21301 ccacgcgaac aaatagatgg ttatgtcatg catgcaaatt acatattttg gaggaataca  
 21361 aatccaattc agttgtcttc ctattctta tttgacatga gtaaatttcc ccttaaattt  
 21421 aggggtactg ctgttatgtc tttaaaagaa ggtcaaatca atgatatgtat tttatcttt  
 21481 cttagtaaag gtagacttat aatttagagaa aacaacagag ttgttatttc tagtgatgtt  
 21541 cttgttaaca actaaacgaa caatgtttgt ttttcttggtt ttattgccac tagtctctag  
 21601 tcagtgtgtt aatcttacaa ccagaactca attaccccct gcatacacta attctttcac  
 21661 acgtgggttt tattaccctg acaaagttt cagatctca gttttacatt caactcagga  
 21721 cttgttctta cttttctttt ccaatgttac ttggttccat gctatacatg tctctgggac  
 21781 caatggtaact aagaggtttg ataaccctgt cctaccattt aatgatggtg tttatttgc  
 21841 ttccactgag aagtctaaca taataagagg ctggatttt ggtactactt tagattcgaa  
 21901 gaccaggatcc ctacttattt ttaataacgc tactaatgtt gttattaaag tctgtgaatt

SUBSTITUTE SHEET (RULE 26)

21961 tcaattttgt aatgatccat tttgggtgt ttattaccac aaaaacaaca aaagtggat  
 22021 ggaaagttag ttcagagttt attctagtgc gaataattgc acttttgaat atgtctctca  
 22081 gcctttctt atggaccttg aaggaaaaca gggtaattc aaaaatcta gggaaatttgt  
 22141 gtttaagaat attgatggtt attttaaaat atattctaag cacacgccta ttaattttgt  
 22201 gcgtgatctc cctcagggtt tttcggttt agaaccattg gtagatttgc caataggtat  
 22261 taacatcaact aggtttcaaa ctttacttgc tttacataga agttatttga ctccctggta  
 22321 ttcttcttca ggttggacag ctggtgctgc agcttattat gtgggttatac ttcaacctag  
 22381 gactttcttca ttaaaatata atgaaaatgg aaccattaca gatgctgttag actgtgcact  
 22441 tgaccctctc tcagaaacaa agtgtacgtt gaaatccttc actgtagaaa aaggaatcta  
 22501 tcaaacttct aacttttagag tccaaccoaac agaatctatt gttagatttc ctaatattac  
 22561 aaacttgtgc ccttttggtg aagttttaa cgccaccaga tttgcattctg tttatgcttg  
 22621 gaacaggaag agaatcagca actgtgttgc tgattattct gtcctatata attccgcata  
 22681 attttccact tttaagtgtt atggagtgtc tcctactaaa ttaaatgatc tctgctttac  
 22741 taatgtctat gcagattcat ttgttaattag aggtgatgaa gtcagacaaa tcgctccagg  
 22801 gcaaactgga aagattgctg attataatta taaattacca gatgattttc caggctgcgt  
 22861 tatacgcttgg aattctaaca atcttgattc taagggttgtt ggtaattata attacctgtt  
 22921 tagattgttt aggaagtcta atctcaaacc ttttgagaga gatatttcaa ctgaaatcta  
 22981 tcaggccggt agcacacacctt gtaatgggtgt tgaaggtttt aattgttact ttcccttaca  
 23041 atcatatggt ttccaaaccca ctaatgggtgt tggttaccaa ccatacagag tagtagtact  
 23101 ttcttttggaa cttctacatg caccagcaac tggttggaa cctaaaaagt ctactaattt

*Fig. 12C continued*

*Fig. 12C*  
*continued*

## SUBSTITUTE SHEET (RULE 26)

23161 ggttaaaaac aaatgtgtca atttcaactt caatggttt acaaggcacag gtgttcttac  
23221 tgagtctaac aaaaagtttc tgcctttcca acaatttgc agagacattg ctgacactac  
23281 ttagtgcgtc cgtgatccac agacacttga gattcttgac attacaccat gttctttgg  
23341 tggtgtcagt gttataaacac caggaacaaa tacttctaac caggttgctg ttcttttatca  
23401 ggatgttaac tgcacagaag tccctgttgc tattcatgca gatcaactta ctctacttg  
23461 gcgtgtttat tctacagggtt ctaatgtttt tcaaacacgt gcaggctgtt taatagggc  
23521 tgaacatgtc aacaactcat atgagtgtga cataccatt ggtgcaggta tatgcgctag  
23581 ttatcagact cagactaatt ctccctcgcg ggcacgtgt gtagctagtc aatccatcat  
23641 tgcctacact atgtcacttg gtgcagaaaa tttagttgtt tactctaata actctattgc  
23701 catacccaca aattttacta tttagtggttac cacagaaatt ctaccagtgt ctatgacc  
23761 gacatcagta gattgtacaa tgtacatttg tggtagttca actgaatgca gcaatcttt  
23821 gttgcaatat ggcagttttt gtacacaatt aaaccgtgt ttaactggaa tagctgttga  
23881 acaagacaaa aacacccaag aagttttgc acaagtcaaa caaatttaca aaacaccacc  
23941 aattaaagat ttgggtggtt ttaatttttc acaaattatta ccagatccat caaaaccaag  
24001 caagaggtca tttattgaag atctactttt caacaaagtg acacttgcag atgctggctt  
24061 catcaaacaa tatggtgatt gccttggtga tattgtgtt agagaccta tttgtgcaca  
24121 aaagtttaac ggccttactg tttgccacc tttgctcaca gatgaaatga ttgctcaata  
24181 cacttctgca ctgttagcgg gtacaatcac ttctgggtgg acctttggtg caggtgctgc  
24241 attacaaata ccatttgcta tgcaaattggc ttataagggtt aatggtattg gagttacaca  
24301 gaatgttctc tatgagaacc aaaaattgat tgccaaaccaa tttaatagtg ctattggcaa  
24361 aattcaagac tcactttctt ccacagcaag tgcacttgga aaacttcaag atgtggtaaa

SUBSTITUTE SHEET (RULE 26)

24421 ccaaaaatgca caagctttaa acacgcttgc taaacaactt agctccaatt ttgggtgcaat  
24481 ttcaagtgtt ttaaatgata tccttcacg tcttgacaaa gttgaggctg aagtgcaaat  
24541 tgataggttg atcacaggca gacttcaaag tttgcagaca tatgtgactc aacaattaat  
24601 tagagctgca gaaatcagag cttctgctaa tcttgctgct actaaaaatgt cagagtgtgt  
24661 acttggacaa tcaaaaagag ttgatTTTg tggaaaggc tatcatctt tgtccttccc  
24721 tcagtcagca cctcatggtg tagtcttctt gcatgtgact tatgtccctg cacaagaaaa  
24781 gaacttcaca actgctcctg ccatttgtca tcatggaaaa gcacactttc ctctgtgaagg  
24841 tgtctttgtt tcaaatggca cacactggtt tctaacaaca aggaattttt atgaaccaca  
24901 aatcattact acagacaaca catttgtgtc tggtaactgt gatgttgtaa taggaattgt  
24961 caacaacaca gtttatgatc otttgcaacc tgaatttagac tcattcaagg aggagttaga  
25021 taaatatttt aagaatcata catcaccaga tgTTGATTt ggtgacatct ctggcattaa  
25081 tgcttcagtt gtAAACATTc aaaaagaaaat tgaccgcctc aatgaggTTg ccaagaattt  
25141 aaatgaatct ctcatcgatc tccaaagaact tggaaagtat gagcagtata taaaatggcc  
25201 atggtagatt tggcttaggtt ttatAGCTGG ctTGATTGCC atAGTAATGG tgacaattat  
25261 gctttgctgt atgaccagtt gctgttagttg tctcaaggc tggatTTCTT gtggatcctg  
25321 ctgcaaattt gatgaagacg actctgagcc aGTGCTCAAa ggagtcaaatt tacattacac  
25381 ataaacgaac ttatggattt gtttatgaga atcttcacaa ttggAACTGT aactttGAAG  
25441 caaggtgaaa tcaaggatgc tactccttca gatTTTGTc gCGCTACTGC aacgataaccg  
25501 atacaagcct cactccctt cggatggctt attgttggcg ttgcacttct tgctgtttt  
25561 cagagcgcTT ccaaaaatcat aaccctcaaa aagagatggc aacttagcact ctccaaagggt

Fig. 12C *continued*

SUBSTITUTE SHEET (RULE 26)

25621 gttcaacttg tttgcaactt gctgttgtg tttgtaacag ttactcaca cctttgctc  
 25681 gttgctgctg gccttgaagc ccctttctc tatcttatg cttagtcta cttcttgctg  
 25741 agtataaaact ttgtaagaat aataatgagg ctggctt gctggaaatg ccgttccaaa  
 25801 aaccatttac ttatgatgc caactatccc ctggctggc atactaattt ttacgactat  
 25861 tgtataccctt acaatagtgt aacttcttca attgtcatta cttcaggtga tggcacaaca  
 25921 agtcctatcc ctgaacatga ctaccagatt ggtggttata ctgaaaaatg ggaatctgga  
 25981 gtaaaagact gtgttgtatt acacagttac ttcaatttca actattacca gctgtactca  
 26041 actcaatttga gtacagacac tgggtgttcaa catgttaccc tcttcatttca caataaaatt  
 26101 gttgatgagc ctgaagaaca tgtccaaatt cacacaatcg acggtttcatc cggagttgtt  
 26161 aatccagtaa tggaaaccaat ttatgatgaa ccgacgacga ctactagcgt gccttggtaa  
 26221 gcacaagctg atgagttacga acttatgtac tcattcggtt cggaaagagac aggtacgtt  
 26281 atagttataa gcttacttct ttttcttgc ttcgtggat tcttgcttagt tacactagcc  
 26341 atccttactg cgcttcgatt gtgtgcgtac tgctgcaata ttgttaacgt gagtcttggta  
 26401 aaaccttctt ttacgttta ctctcggtt aaaaatctga attcttcttag agttctgtat  
 26461 cttctgggtct aaacgaacta aatattat tagttttctt gtttggaaact ttaatttttag  
 26521 ccatggcaga ttccaacggt actattaccg ttgaagagct taaaaagctc cttgaacaat  
 26581 ggaacctagt aataggtttc ctattcotta catggatttgc tcttctacaa tttgcctatg  
 26641 ccaacagggaa taggttttg tatataatta agttaattt cctctggctg ttatggccag  
 26701 taactttacgc ttgttttgtt cttgctgctg tttacagaat aaattggatc accgggtggaa  
 26761 ttgctatcgc aatggcttgtt cttgttaggct tgatgtggct cagctacttc attgcttctt  
 26821 tcagactgtt tgccgtacg cgttccatgt ggtcattcaa tccagaaact aacatttttc

*Fig. 12C*  
*continued*

SUBSTITUTE SHEET (RULE 26)

26881 tcaacgtgcc actccatggc actattctga ccagaccgct tctagaaaagt gaactcgtaa  
 26941 tcggagctgt gatccttcgt ggacatcttc gtattgctgg acaccatcta ggacgctgtg  
 27001 acatcaagga cctgcctaaa gaaatcactg ttgctacatc acgaacgctt tcttattaca  
 27061 aattgggagc ttgcgcagcgt gtagcaggtg actcaggttt tgctgcatac agtcgctaca  
 27121 ggattggcaa ctataaatta aacacagacc attccagtag cagtgacaat attgctttgc  
 27181 ttgtacagta agtgacaaca gatgtttcat ctcgttact ttcaaggtagtatacgag  
 27241 atattactaa ttattatgag gactttaaa gtttccattt ggaatcttga ttacatcata  
 27301 aacctcataa ttaaaaaattt atctaagtca ctaactgaga ataaatattc tcaatttagat  
 27361 gaagagcaac caatggagat tgattaaacg aacatgaaaa ttattcttt cttggcactg  
 27421 ataacactcg ctacttgtga gctttatcac taccaaaggt gtgttagagg tacaacagta  
 27481 ctttaaaag aaccttgctc ttctggaaaca tacgagggca attcaccatt tcatcctcta  
 27541 gctgataaca aatttgcaact gacttgcttt agcactcaat ttgctttgc ttgtcctgac  
 27601 ggcgtaaaac acgtctatca gttacgtgcc agatcagtt cacctaaact gttcatcaga  
 27661 caagaggaag ttcaagaact ttactctcca atttttctta ttgttgcggc aatagtgttt  
 27721 ataacacttt gttcacact caaaagaaaag acagaatgat tgaactttca ttaattgact  
 27781 tctatttgcg cttttagcc tttctgctat tccttgcattt aattatgctt attatcttt  
 27841 ggttctcaact tgaactgcaa gatcataatg aaacttgtca cgccctaaacg aacatgaaat  
 27901 ttcttgcattt cttaggaatc atcacaactg tagctgcatt tcaccaagaa tgttagttac  
 27961 agtcatgtac tcaacatcaa ccatatgtag ttgatgaccc gtgtcctatt cacttctatt  
 28021 ctaaatggta tattagagta ggagctagaa aatcagcacc tttaattgaa ttgtgcgtgg

*Fig. 12C continued*

Fig. 12C  
continued

## SUBSTITUTE SHEET (RULE 26)

28081 atgaggctgg ttctaaatca cccattcagt acatcgatat cggttaattat acagttcct  
28141 gtttaccttt tacaattaat tgccaggaac ctaaattggg tagtcttgta gtgcgttgtt  
28201 cgttctatga agactttta gagtatcatg acgttcgtgt tgtttagat ttcatctaaa  
28261 cgaacaaaact aaaatgtctg ataatggacc ccaaaatcag cgaaatgcac cccgcattac  
28321 gtttggtgga ccctcagatt caactggcag taaccagaat ggagaacgca gtggggcgcg  
28381 atcaaaaacaa cgtcgcccccc aaggtttacc caataatact gcgtcttggc tcaccgctct  
28441 cactcaacat ggcaaggaag accttaaatt ccctcgagga caaggcggttcaattAACAC  
28501 caatagcagt ccagatgacc aaattggcta ctaccgaaga gctaccagac gaattcgtgg  
28561 tggtgacggt aaaatgaaaag atctcagtcc aagatggtat ttctactacc taggaactgg  
28621 gccagaagct ggacttccct atggtgctaa caaagacggc atcatatggg ttgcaactga  
28681 gggagccttg aatacacccaa aagatcacat tggcacccgc aatcctgcta acaatgctgc  
28741 aatcgtgcta caacttcctc aaggaacaac attgccaaaa ggcttctacg cagaagggag  
28801 cagaggcggc agtcaagcct cttctcggtt ctcacatcacgt agtgcacaaca gttcaagaaaa  
28861 ttcaactcca ggcagcagta ggggaacttc tcctgctaga atggctggca atggcggtga  
28921 tgctgcttt gcttgctgc tgcttgacag attgaaccag cttgagagca aaatgtctgg  
28981 taaaggccaa caacaacaag gccaactgt cactaagaaa tctgctgctg aggcttctaa  
29041 gaagcctcggtt caaaaacgtt cttccactaa agcataacaat gtaacacaag ctttcggcag  
29101 acgtggtcca gaacaaaaccc aaggaaattt tggggaccag gaactaatca gacaagggAAC  
29161 tgattacaaa cattggccgc aaattgcaca atttgcccccc agcgcttcag cgttcttcgg  
29221 aatgtcgccgc attggcatgg aagtacaccc ttggggaaacg tggttgaccc acacaggtgc  
29281 catcaaattt gatgacaaaag atccaaattt caaagatcaa gtcattttgc tgaataagca

29341 tattgacgca tacaaaacat tcccaccaac agagcctaaa aaggacaaaaa agaagaaggc  
29401 tgatgaaact caagcttac cgcaagagaca gaagaaaacag caaactgtga ctcttcttcc  
29461 tgctgcagat ttggatgatt tctccaaaca attgcaacaa tccatgagca gtgctgactc  
29521 aactcaggcc taaactcatg cagaccacac aaggcagatg ggctatataa acgtttcgc  
29581 tttccgttt acgatatata gtctactttt gtgcagaatg aattctcgta actacatagc  
29641 acaagtagat gtagttaact ttaatctcac atagcaatct ttaatcagtg tgtaacatta  
29701 gggaggactt gaaagagcca ccacatttc accgaggcca cgccggagtac gatcgagtgt  
29761 acagtgaaca atgcttaggga gagctgccta tatggaagag ccctaatttg taaaattaat  
29821 ttttagtagtg ctatccccat gtgattttaa tagttctta ggagaatgac aaaaaaaaaaa  
29881 aaaaaaaaaaa aaaaaaaaaaa aaa

## CoV-2 RNA genome (29,903 bp)

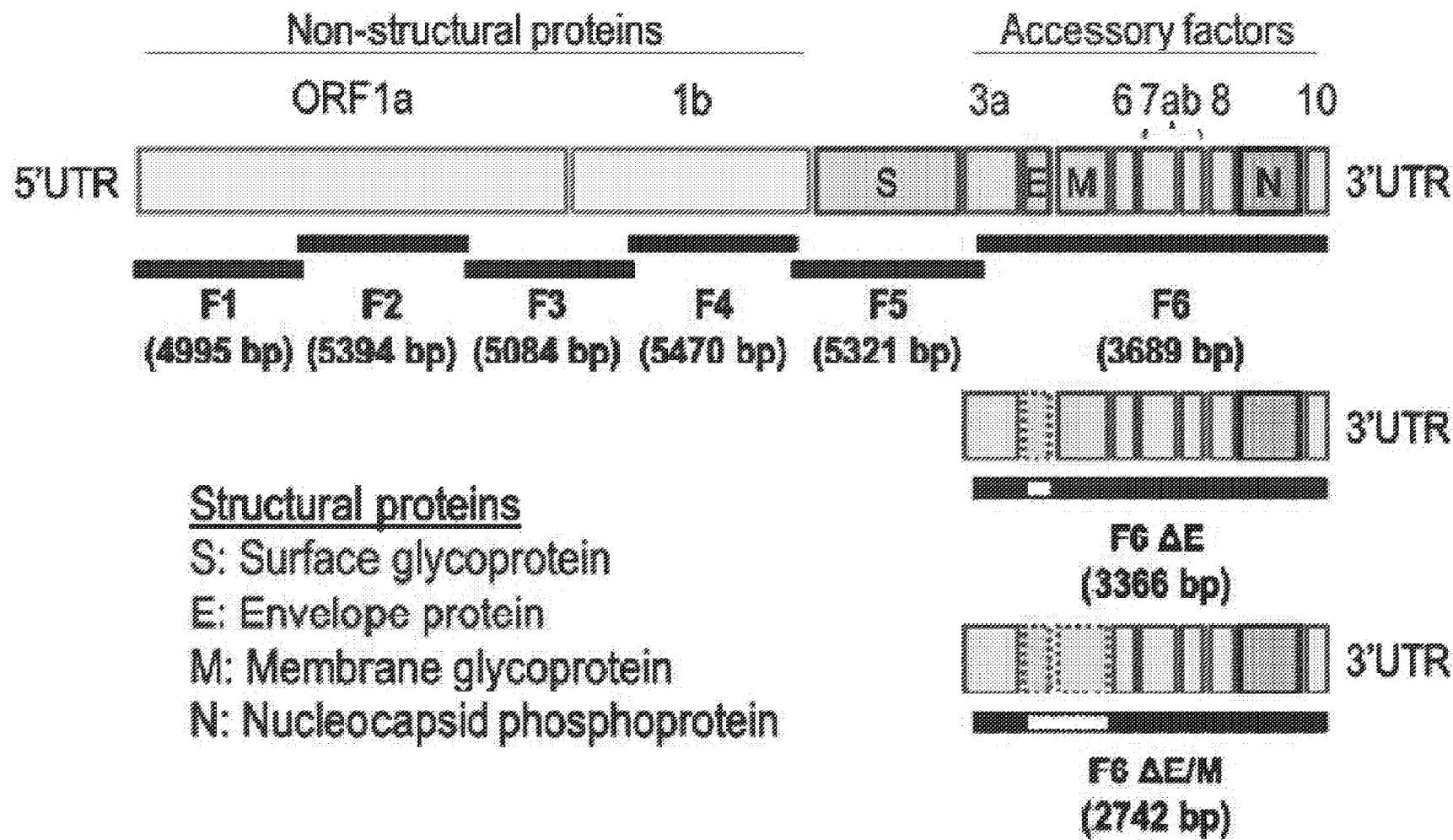


Fig. 13

SUBSTITUTE SHEET (RULE 26)

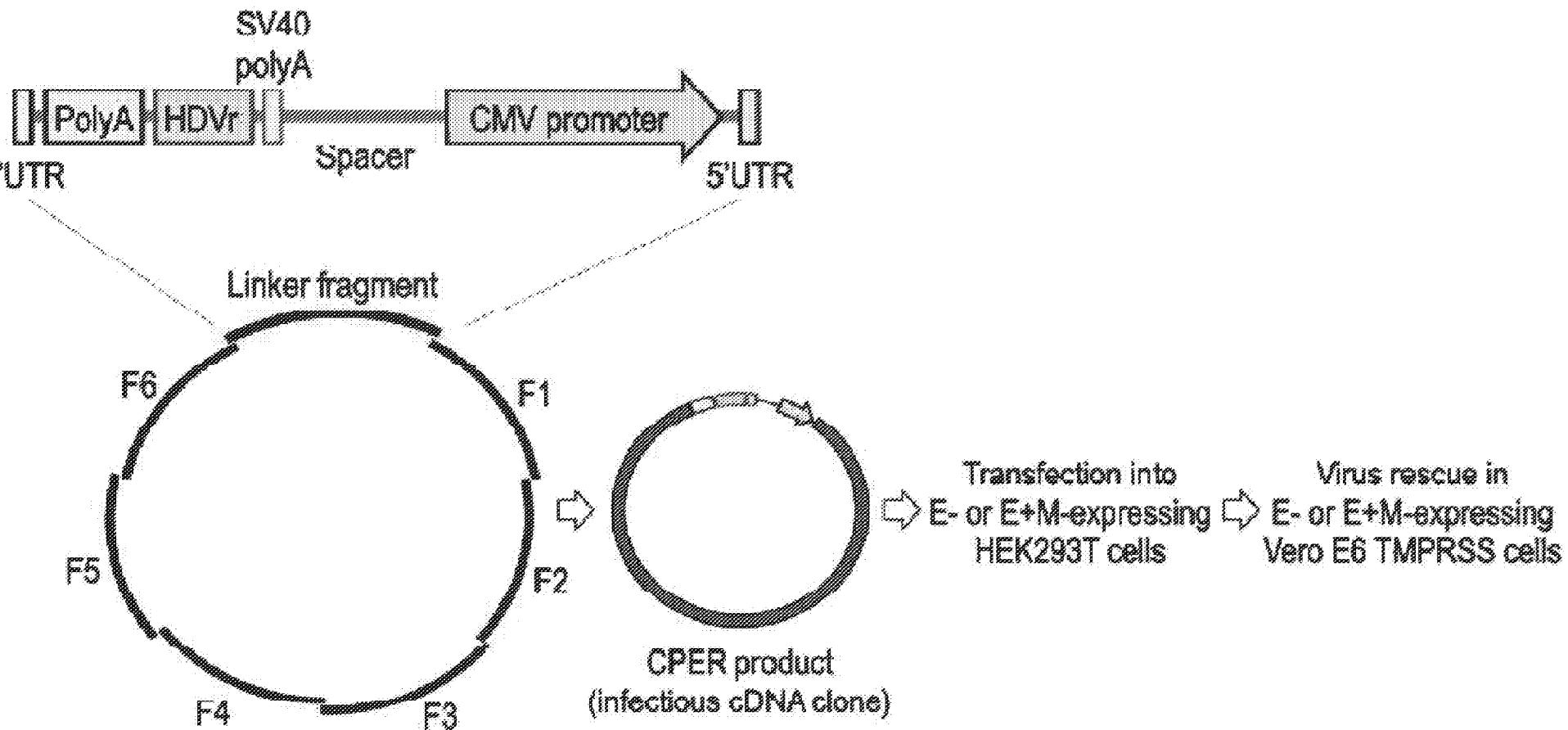


Fig. 14

*Codon-optimized CoV-2 E gene (Addgene)*

ATGTACTCTTCGTGAGCGAGGAAACCGGCACCCCTGATCGTGAACCTCCGTGCTGCTGTTCCCTGGCCCTCGTGGTGT  
 TCCTGCTGGTGACCCCTGGCTATCCTGACCGCTCTGAGACTGTGCGCTTACTGCTGCAACATCGTGAACGTGTCCCTG  
 GTGAAGCCCTCTTCTACGTGTACAGCCCGTGAAAGAACCTGAACAGCTCCAGGGTGCCTGACCTGCTGGTGTAA  
 (SEQ ID NO:13)

SUBSTITUTE

*Codon-optimized CoV-2 M gene (Addgene)*

ATGGCTGACTCTAACGGTACCATCACCGTGGAGGAACGTGAAGAACGCTGGAGCAGTGGAACCTGGTCATCGG  
 CTTCCCTGTTCTGACCTGGATCTGCCCTGCTGCAGTTGCCCTACGCTAACCGCAACAGGTTCCGTACATCATCAAGC  
 TGATCTTCCTGTGGCTGCTGTGGCCTGTGACCCCTGGCTTGCTCGTGGCTGCCGTGTACCGCATCAACTGGATC  
 (RULE 26) ACCGGCGGAATGCCATCGCTATGGCCTGCCTGGTGGGCCTGATGTGGCTGTCTTACITCATCGTAGCTCAGGC  
 TGTCGCCAGAACCCGTTCCATGTGGCTTCAACCCCGAGACCAACATCCTGCTGAACGTGCCCTGCACGGAAC  
 CATCCTGACCAGACCACTGCTGGAGAGCGAACCTGGTCATCGCGCTGTGATCCTGAGAGGGACACCTGCGTATCGC  
 CGGACACCAACCTGGTCGTTGCGACATCAAGGACCTGCCAAGGAAATCACCCTGGCTACCAGCCGCACCCCTGTC  
 CTACTACAAGCTGGGAGCTTCAGAGAGTGGCTGGTGAECTGGTTCTGCTGCTTACTCTCGCTACAGGATCGGT  
 AACTACAAGCTGAACACCGACCACAGCTCCTCTAGCGACAACATCGCCCTGCTGGTGCAGTAA (SEQ ID NO:14)

<https://www.nature.com/articles/s41467-021-23779-5>

*ΔE GENOME**Fig. 15.*

ATTAAGGTTTACCTCCAGGTACAAACCAACCAACTTCGATCTCTTAGATCTGTTCTAAACGAACCTT  
AAAATCTGTGGCTGCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAACTAATTACTGTCGTTG  
ACAGGACACGAGTAACCGTCTATCTGCAGGCTGCTACGGTTCGTCGTGTCAGCCGATCATCAGCACA  
TCTAGGTTCTGCCGGGTGTGACCGAAAGGTAAAGATGGAGAGCCTGTCCTGGTTCAACGAGAAAACACACGT  
CCAACTCAGTTGCCTGTTTACAGGTCGCGACGTGCTCGTACGTGGCTTGGAGACTCCGTGGAGGGAGGTCTTA  
TCAGAGGCACGTCAACATCTAAAGATGGACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTGCCTCAACTTG  
AACAGCCCTATGTGTTCATCAAACGTTGGATGCTCGAACCTCATGGTCATGTTATGGTTGAGCTGGTAGC  
SUBSTITUTE AGAAACTCGAAGGCATTCACTACGGTCGTAGTGGTGAGACACTGGTGTCCCTGTCCCTCATGTGGCGAAATACC  
AGTGGCTTACCGCAAGGTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCT  
SHEET AAAAGTCATTGACTTAGGCGACGAGCTGGCACTGATCCTTATGAAGATTTCAGAAAAGCTGGAACACTAAACAT  
AAGCAGTGGTGTACCCGTGAACTCATGCGTGAGCTAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTCT  
GTGGCCCTGATGGCTACCCCTCTGAGTCATTAAAGACCTCTAGCACGTGCTGGTAAAGCTTCATGCACTTGTCC  
GAACAACGGACTTTATTGACACTAAGAGGGGTGTAACTGCTGCCGTGAACATGAGCATGAAATTGCTTGGTAC  
ACGGAACGTTCTGAAAAGAGCTATGAATTGAGACACCTTGTAAATTGGCAAAGAAAATTGACACCTCTCA  
(RULE ATGGGGAAATGTCAAATTTCGTTTCCCTAAATTCCATAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAA  
GCTTGATGGCTTATGGGTAGAATTGATCTGTCTATCCAGTTGCGTCACCAAATGAATGCAACCAATGTGCCCTT  
26) CAACTCTCATGAAGTGTGATCATTGTTGGTAAACCTCATGGCAGACGGCGATTITGTTAAAGCCACTTGCGAATT  
TTGTGGCACTGAGAATTGACTAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCCAAATGCTGTTGTTAAAATT  
ATTGTCCAGCATGTCACAATTCAAGAAGTAGGACCTGAGCATAGTCTGCCGAATACCATAATGAATCTGGCTTGAA  
AACCATTCTCGTAAGGGTGGTCGCACTATTGCCCTTGGAGGGCTGTGTGTTCTTATGTTGGTTGCCATAACAAAGT  
GTGCCATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCATAACAGGTGTTGGAGAAGGGTCCGAAG

Fig. 15

GTCTTAATGACAACCTTCTGAAATACTCCAAAAAGAGAAAGTCACATCAATATTGTTGGTACTTTAAACTTAAT  
GAAGAGATGCCATTATTTGGCATCTTTCTGCTCCACAAGTGCTTTGTGGAAACTGTGAAAGGTTGGATTA  
TAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTAAAGTACAAAAGGAAAGCTAAAAAGGTGCCTGG  
AATATTGGTGAACAGAAATCAATACTGAGTCCTCTTATGCATTGCATCAGAGGCTGCTGTGTTACGATCAAT  
TTTCTCCGCACCTCTGAAACTGCTCAAATTCTGTGCGTGTACAGAAGGCGCTATAACAATACTAGATGGAA  
TTTCACAGTATTCACTGAGACTCATTGATGCTATGATGTTCACATCTGATTGGCTACTAACAACTAGTTGTAATG  
GCCTACATTACAGGTGGTGTGTTCAAGTGAACCTCGCAGTGGCTAACTAACATCTTGGCACTGTTATGAAAAACT  
CAAACCCGTCCTGATTGGCTTGAAGAGAAGTTAAGGAAGGTGTAGAGTTCTTAGAGACGGTTGGAAATTGT  
TAAATTATCTCAACCTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAGGAGAGT  
GTTCAAGACATTCTTAAGCTTGTAAATAAAATTGGCTTGTGCTGACTCTATCATTATTGGTGGAGCTAAACTT  
AAAGCCTTGAATTAGGTGAAACATTGTCACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAA  
GAAACTGGCCTACTCATGCCCTCTAAAGCCCCAAAGAAATTATCTTCAAGGGAGAAACACTCCCACAGAAG  
TGTAAACAGAGGAAGTTGTCTTGAAGGAACTGGTATTACAACCATTAGAACAAACCTACTAGTGAAGCTGTTGAAGC  
TCCATTGGTGGTACACCAGTTGTATTACCGGGCTTATGTTGCTCGAAATCAAAGACACAGAAAAGTACTGTGCC  
CTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTGGTG  
ATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTGAACCTGATGAAAGGATTGATAAAG  
TACTTAATGAGAAGTGCTCTGCCTATACAGTTGAACCTCGGTACAGAAGTAAATGAGTTGCCGTGTTGGCAGA  
TGCTGTATAAAACTTGTCAACCAAGTATCTGAATTACTTACACCACTGGCATTGATTAGATGAGTGGAGTATG  
GCTACATACTACTTATTGATGAGTCTGGTGAGTTAAATTGGCTTACATATGTATTGTTCTTACCCCTCCAGAT  
GAGGATGAAGAAGAAGGTGATTGTGAAGAAGAAGAGTTGAGCCATCAACTCAATATGAGTATGGTACTGAAGA  
TGATTACCAAGGTAAACCTTGGAAATTGGTGCCTCTGCTGCTTCAACCTGAAGAAGAGCAAGAAGAAGAT  
TGGTTAGATGATGATAGTCAACAAACTGTTGGTCAACAAGACGGCAGTGAGGGACAATCAGACAACACTATTCAA

Fig. 15

ACAATTGTTGAGGTCAACCTCAATTAGAGATGGAACCTACACCAGTTTCAGACTATTGAAGTGAATAGTTTA  
GTGGTTATTTAAAACCTACTGACAATGTATACTTAAACATGCAGACATTGTTGGAAGAAGCTAAAAGGTAAAACC  
AACAGTGGTTGTTAATGCAGCCAATGTTACCTTAAACATGGAGGAGGTGTCAGGAGCCTAAATAAGGCTAC  
TAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCACCTAAAGTGGTGGTAGTTGTT  
TTAACCGGACACAATCTTGCTAAACACTGTTCATGTTGTCGGCCAAATGTTAACAAAGGTGAAGACATTCAAC  
TTCTTAAGAGTGCTTATGAAAATTTAATCAGCACGAAGTTCTACTTGCACCATTATTATCAGCTGGTATTTGGTG  
CTGACCTATACATTCTTAAAGAGTTGTTAGATACTGTTGCACAAATGTTACTTAGCTGTTGATAAAAAATC  
TCTATGACAAACTGTTCAAGCTTTGGAAATGAAGAGTGAAAAGCAAGTTGAACAAAAGATCGCTGAGATTCC  
TAAAGAGGAAGTTAACGCCATTATAACTGAAAGTAAACCTTCAGTTGAACAGAGAAAACAAGATGATAAGAAAAT  
CAAAGCTTGTGTTAACAGTTACAACAACCTGGAAAGAAACTAACGTTCTCACAGAAAACCTGTTACTTTATATTG  
ACATTAATGGCAATCTTCAGATTCTGCCACTCTGTTAGTGACATTGACATCACTTCTAAAGAAAGATGCTC  
CATATATAGTGGGTGATGTTGTTCAAGAGGGTGTAACTGCTGTTGTTACCTACTAAAAAGGCTGGTGGCAC  
TACTGAAATGCTAGCGAAAGCTTGAGAAAAGTGCCAACAGACAATTATATAACCACTAACCGGGTCAGGGTTA  
AATGGTTACACTGTAGAGGAGGCAAAGACAGTGCTAAAAAGTGTAAAAGTGCCTTTACATTCTACCATCTATT  
TCTCTAATGAGAAGCAAGAAATTCTGGAACGTGTTCTGGAATTGCGAGAAATGCTGTCACATGCAGAAGAAAC  
ACGCAAATTAAATGCCGTGTTGAAACTAAAGCCATAGTTCAACTATACAGCGTAAATATAAGGGTATTAAA  
ATACAAGAGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG  
ACACACTTAACGATCTAAATGAAACTCTTGTACAATGCCACTGGCTATGTAACACATGGCTAAATTGGAAGAA  
GCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTTCTGTTCTCACCTGATGCTGTTACAGCGTATAA  
TGGTTATCTTACTTCTTCTTAAACACCTGAAGAACATTATTGAAACCATCTCACCTGCTGGTCTTCTTCTTCTTCTT  
TTGGTCTATTCTGGACAATCTACACAACTAGGTATAGAATTCTTAAGAGAGGTGATAAAAGTGTATATTACACTA  
GTAATCCTACCAACATTCCACCTAGATGGTGAAGTTACACCTTGTACAATCTTAAGACACCTCTTCTTCTTGTGGAGAGAA

Fig. 15

GTGAGGACTATTAAGGTGTTACAACAGTAGACAACATTAAACCTCACACGCAAGTTGTCATGACAT  
ATGGACAACAGTTGGTCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCATAATTACATGAAGG  
TAAAACATTTATGTTTACCTAATGATGACACTCTACGTGTTGAGGCTTTGAGTACTACCACACAACGTGATCCTA  
GTTTCTGGGTAGGTACATGTCAGCATTAAATCACACTAAAAAGTGGAAATACCCACAAGTTAATGGTTAACCTCT  
ATTAATGGGCAGATAACAACGTGTTATCTGCCACTGCATTGTTAACACTCCAACAAATAGAGTTGAAGTTAACCC  
ACCTGCTCTACAAGATGCTTATTACAGAGCAAGGGCTGGTGAAGCTGCTAACCTTGTCACTTATCTTAGCCTACT  
GTAATAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAAACAATGAGTTACTTGTTAACATGCCAATTAGATT  
TTGCAAAAGAGTCTGAACGTGGTGTGAAAACCTTGGAACAACAGCAGACAACCCCTAAGGGTGTAGAAGCTGT  
TATGTACATGGGCACACTTTCTTATGAACAATTAAAGAAAGTGTAGATAACCTTGTCAGTGTGGTAAACAAGCT  
ACAAAATATCTAGTACAACAGGAGTCACCTTGTTATGATGTCAGCACCACTGCTCAGTATGAACCTAACATG  
GTACATTACTTGCTAGTGAGTACACTGTAATTACCAAGTGTGGTCACTATAAACATATAACTTCTAAAGAAA  
TTGTATTGCATAGACGGTCTTACTTACAAAGTCCTCAGAATACAAAGGTCTTACCGGATGTTCTACAAAGA  
AAACAGTTACACAACACCATAAAACCAAGTTACTTATAAAATTGGATGGTGTGTTGTACAGAAATTGACCC  
TTGGACAATTATTATAAGAAAGACAATTCTTATTACAGAGCAACCAATTGATCTGTACCAAACCAACCATATCC  
AAACGCAAGCTCGATAATTAAAGTTGTATGTGATAATATCAAATTGCTGATGATTAAACCAAGTTA  
ACTGGTTATAAGAAACCTGCTTCAAGAGAGCTAAAGTTACATTTCCTGACTAAATGGTGTGGCTATTGATT  
AAACACTACACACCCTTTAAGAAAGGAGCTAAATTGTTACATAACCTATTGTTGGCATGTTAACATGCAAC  
TAATAAAGCCACGTATAAACCAAATACCTGGTGTACGTTGCTTGGAGCACAAACCAAGTTGAAACATCAA  
TCGTTGATGTACTGAAGTCAGAGGACGCCAGGGAAATGGATAATTGCTGCGAAGATCTAAACCAAGTCT  
GAAGAAGTAGTGGAAAATCCTACCATAACAGAAAGACGTTGAGTGTAAATGTGAAAACCAACCGAAGTTG  
GACATTATACTTAAACCAAGCAAATAATAGTTAAAAATTACAGAAGAGGTTGCCACACAGATCTAATGGCT  
ATGTAGACAATTCTAGTCTACTATTAAGAAACCTAATGAATTATCTAGAGTATTAGGTTGAAAACCC  
TTGCTACT

Fig. 15

CATGGTTAGCTGCTTAATAGTGTCCCTGGGACTATAGCTAATTATGCTAACGCTTTCTTAACAAAGTTGTT  
AGTACAACACTAACATAGTTACACGGTGTTAAACCGTGTTGACTAATTATATGCCCTATTTCTTACTTTATTG  
CTACAATTGTGTACTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAGCAAAGAAC  
TGTAAAGAGTGTGGTAAATTGTCTAGAGGCTTCATTAATTATTGAAGTCACCTAATTCTAAACTGATAAA  
TATTATAATTGGTTTTACTATTAAGTGTTCGCCTAGGTTCTTAATCTACTCAACCCTGCTTAGGTGTTAATG  
TCTAATTAGGCATGCCCTTACTGTACTGGTACAGAGAAGGCTATTGAACTCTACTAATGTCACTATTGCAAC  
CTACTGTACTGGTCTACCTTGTAGTGTCTAGTGGTTAGATTCTTAGACACCTATCCTCTTAGAAACT  
ATACAAATTACCATTTCATCTTTAAATGGGATTAACTGCTTTGGCTTAGTTGCAGAGTGGTTTTGGCATATATT  
CTTTCACTAGGTTTTCTATGTACTGGATTGGCTGCAATCATGCAATTGTTTCAGCTATTGCAGTACATTAA  
TTAGTAATTCTGGCTATGTGGTTAATAATTAAATCTTGTACAAATGGCCCCGATTCACTATGGTTAGAATGTAC  
ATCTTCTTGCACTTTATTATGTATGGAAAAGTTATGTGCATGTTAGACGGTTGTAATTCAACTTGTATG  
ATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACATTGTTAATGGTTAGAAGGTCTTTATG  
TCTATGCTAATGGAGGTAAAGGCTTTGCAAACACTACACAATTGGAATTGTTAATTGTGATACATTCTGTGCTGGT  
AGTACATTATTAGTGAAGTTGCGAGAGACTTGTCACTACAGTTAAAGACCAATAATCCTACTGACCACT  
CTTCTTACATCGTTGATAGTGTACAGTGAAGAATGGTCCATCCATCTTACTTTGATAAAGCTGGTCAAAAGACT  
TATGAAAGACATTCTCTCTCATTTGTTAACTTAGACAACCTGAGAGCTAATAACACTAAAGGTTCAATTGCCTATT  
AATGTTATAGTTTGATGGTAAATCAAATGTGAAGAATCATCTGAAAATCAGCGTCTGTTACTACAGTCAGCT  
TATGTGTCACCTATACTGTTACTAGATCAGGCATTAGTGTCTGATGTTGGTATAGTGCAGGAAAGTTGCAAGTTAAA  
ATGTTTGATGCTTACGTTAACGTTTACACTTAACTGACCAATTGAAAAACTCAAACACTAGTTGCAAC  
TGCAGAAGCTGAACCTGCAAAGAATGTGTCCCTAGACAATGTCTTATCTACTTTATTTCAGCAGCTCGGCAAGGG  
TTTGGTTGATTCAAGATGTAGAAAATCAAAGATGTTGTTGAATGTCACATCAATCTGACATAGAAGTTAC  
TGGCGATAGTTGTAATAACTATATGTCACCTATAACAAAGTTGAAAACATGACACCCCGTACCTTGGTGTGTTGT

Fig. 15

ATGACTGTAGTGCAGTCATATTAAATGCGCAGGTAGCAAAAAGTCACAACATTGCTTGATATGGAACGTTAAAG  
ATTTCATGTCAATTGTCTGAACAACTACGAAAACAAATACGTAGTGCCTGCTAAAAAGAATAACCTTACCTTTAAGTTG  
ACATGTGCAACTACTAGACAAGTTAACATGTTAACAAACAAAGATAGCACTTAAGGGTGGTAAAATTGTTAATA  
ATTGGTTGAAGCAGTTAACATTAAAGTTACACTTGTGTTCTTTGTTGCTGCTATTCTATTAAATAACACCTGTTCA  
TGTCAATGTCTAAACATACTGACTTTCAAGTGAAATCATAGGATAACAAGGCTATTGATGGTGGTGTCACTCGTGAC  
ATAGCATCTACAGATACTTGTGTTGCTAACAAACATGCTGATTGACACATGGTTAGCCAGCGTGGTGGTAGTTA  
TACTAATGACAAAGCTGCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTGTGCTGCCTGGTTGCCT  
GGCACGATATTACGCACAACATAATGGTACTTTGCATTCTACCTAGAGTTTAGTGCAGTTGGTAACATCTG  
TTACACACCATTAGAGTACACTGACTTGCAACATCAGCTTGTGTTGGCTGCTGAATGTACAATT  
TTAAAGATGCTCTGGTAAGCCAGTACCATATTGTTATGATAACCAATGTAAGTAGAAGGTTCTGTTGCTTATGAAAGT  
TTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTCAATTCTAACACCTACCTTGAAGGTTCTGTT  
AGAGTGGTAACAACTTTGATTCTGAGTACTGTAGGCACGGACTTGTGAAAGATCAGAAGCTGGTGTGTTGTT  
CTACTAGTGGTAGATGGTACTAACAAATGATTATTACAGATCTTACCAAGGAGTTCTGTGGTAGATGCTGTA  
AATTACTTAATATGTTACACCACTATTCAACCTATTGGTGTGTTGGACATATCAGCATCTATAGTAGCTGGT  
GGTATTGTAGCTATCGTAGTAACATGCCTGCCTACTATTATGAGGTTAGAAGAGCTTGGTGAATACAGTCA  
TGTAGTTGCCTTAATACTTTACTATTCTTATGTCATTCACTGTACTCTGTTAACACCCAGTTACTCATTCTTACCT  
GGTGTGTTATTCTGTTATTACTTGTACTTGACATTCTTACTAATGATGTTCTTTAGCACATATTCACTGGTGA  
TGGTTATGTTACACCTTAGTACCTTCTGGATAACAATTGCTTATATCATTGTATTCCACAAAGCATTTCTATTG  
GTTCTTGTAGTAATTACCTAAAGAGACGTGTAGTCATTAAATGGTGTGTTCTTACTTGTACTTGTGAAGAAGCTGCGCTGT  
GCACCTTGTAAATAAAAGAAATGTATCTAAAGTTGCGTAGTGTGCTATTACCTCTACGCAATATAATAGA  
TACTTAGCTCTTATAATAAGTACAAGTATTAGTGGAGCAATGGATAACAAGCTACAGAGAAGCTGCTTGT  
GTCATCTGCAAAGGCTCTCAATGACTCAGTTGATGTTCTTACCAACCACCAACCTCTATC

Fig. 15

ACCTCAGCTGTTTGCAGAGTGGTTAGAAAAATGGCATTCCATCTGGTAAAGTTGAGGGTTGTATGGTACAAG  
TAACTTGTGGTACAACACTAACGGTCTTGGCTTGATGACGTAGTTACTGTCCAAGACATGTGATCTGCACC  
TCTGAAGACATGCTTAACCTAATTATGAAGATTACTCATTGTAAGTCTAATCATAATTCTGGTACAGGCTGG  
TAATGTTCAACTCAGGGTTATTGGACATTCTATGCAAAATTGTGTACTTAAGCTTAAGGTTGATAACAGCCAATCCTA  
AGACACCTAAGTATAAGTTGTCGCATTCAACCAGGACAGACTTTTCAGTGTAGCTTACAATGGTTACCAATGGTTACCA  
TCTGGTGTACCAATGTGCTATGAGGCCAATTCACTATTAAAGGTTCATCCTTAATGGTTATGTGGTAGTGT  
TGGTTAACATAGATTATGACTGTGTCTTTTGTACATGCACCATATGGAATTACCAACTGGAGTTATGCTG  
GCACAGACTTAGAAGGTAACCTTATGGACCTTTGTTGACAGGCAAACAGCACAAGCAGCTGGTACGGACACAA  
CTATTACAGTTAACATGTTAGCTTGGTTGACGCTGCTGTTATAAATGGAGACAGGTGGTTCTCAATCGATTAC  
ACAACCTTAATGACTTAACCTTGTGGCATGAAGTACAATTATGAAACCTCTAACACAAGACCATGTTGACATACT  
AGGACCTCTTCTGCTCAAACCTGGAATTGCCGTTAGATATGTGTGCTTCAATTAAAGAATTACTGCAAATGGTA  
TGAATGGACGTACCATATTGGTAGTGCTTATTAGAAGATGAATTACACCTTTGATGTTAGACAATGCTCA  
GGTGTACTTCCAAAGTGCAGTGAAAAGAACATCAAGGGTACACACCCTGGTTACTCACAATTGACTT  
CACTTTAGTTAGTCCAGAGTACTCAATGGCTTGTCTTTTGTATGAAAATGCCTTTACCTTGTAT  
GGGTATTATTGCTATGTCTGCTTGTCAATGATGTTGTCACAAACATAAGCATGCATTCTCTGTTGTTGTTACC  
TTCTCTGCCACTGTAGCTTATTAAATATGGCTATATGCCCTGCTAGTTGGGTGATGCGTATTATGACATGGTTGG  
ATATGGTTGATACTAGTTGTCGGTTAAGCTAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTACTAATC  
CTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGGACACTTATGAATGCTTGTACACTCGTT  
ATAAAAGTTATTATGGTAATGCTTGTAGATCAAGCCATTCCATGTGGCTCTATAATCTCTGTTACTTCTAACTACT  
CAGGTGTAGTTACAACGTCAATGTTGGCCAGAGGTATTGTTATGTGTGTTGAGTATTGCCCTATTCTTCA  
TAACGGTAATACACTTCAGTGTATAATGCTAGTTATTGTTCTAGGCTATTGTTACTTGTACTTGGCCTCTT  
TTGTTTACTCAACCGCTACTTGTACTGACTCTGGTGTATGATTACTTAGTTCTACACAGGAGTTAGATATAT

Fig. 15

GAATTCACAGGGACTACTCCCACCAAGAACATAGCATAGTCCTCAAACACTAACATTAAATTGTTGGGTGTTGGT  
GGCAAACCTTGTATCAAAGTAGCCACTGTACAGTCTAAAATGTCAGATGAAAGTCACATCAGTAGTCTTACTCT  
CAGTTTGCAACAACTCAGAGTAGAATCATCATCTAAATTGTCAGGCTCAATGTCAGTTACACAATGACATTCTC  
TTAGCTAAAGATACTACTGAAGCCTTGAAAAAAATGGTTCACTACTTCTGTTGCTTCCATGCAGGGTGCTGT  
AGACATAAAACAAGCTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTAGTTCC  
CTTCCATCATATGCAGCTTTGCTACTGCTCAAGAAGCTTATGAGCAGGCTGCTAATGGTGATTCTGAAGTTGT  
TCTTAAAAGTTGAAGAAGTCTTGAATGTGGCTAAATCTGAATTGACCGTGATGCAAGCCATGCAACGTAAGTTG  
GAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTGAGGACAAGAGGGCAAAAGTTAC  
TAGTGCTATGCAGACAATGCTTTCACTATGCTAGAAAGTTGGATAATGATGCACTCAACAAACATTCAACAATG  
CAAGAGATGGTTGTTCCCTGAACATAATACCTCTTACAACAGCAGCAAACATAATGGTTGTCATACCAGACTAT  
AACACATATAAAATACGTGTGATGGTACAACATTACTTATGCATCAGCATTGTCAGGAAATCCAACAGGTTGTAG  
ATGCAGATAGTAAAATTGTTCAACTTAGTGAAATTAGTATGGACAATTACCTTAATTAGCATGGCTTATTGTA  
ACAGCTTAAAGGCCAATTCTGCTGCTCAAATTACAGAATAATGAGCTTAGTCTGTCAGCAGACAGATGTCTT  
GTGCTGCCGGTACTACACAAACTGCTTGCAGTACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTA  
GGTTTGTACTTGCACTGTTATCCGATTACAGGATTGAAATGGCTAGATTCCCTAACAGAGTGATGGAACGGTAC  
TATCTATACAGAACTGGAACCACCTTGTAGGTTACAGACACACCTAAAGGTCTAAAGTGAAAGTATTATACT  
TTATTAAAGGATTAAACACCTAAATAGAGGTATGGTACTTGGTAGTTAGCTGCCACAGTACGTCTACAAGCTGG  
TAATGCAACAGAAGTGCCTGCCAATTCAACTGTATTATCTTCTGCTTTGCTGTAGATGCTGCTAAAGCTTACA  
AAGATTATCTAGCTAGTGGGGACAACCAATCACTAATTGTGTTAACAGATGTTGTACACACACTGGTACTGGTCA  
GGCAATAACAGTTACACCGGAAGCCAATATGGATCAAGAACCTTGGTGGTGCATCGTGTCTGACTGCCGT  
TGCCACATAGATCATCCAAATCTAAAGGATTGACTTAAAGGTAAGTATGTACAAATACCTACAACCTGTGCA  
TAATGACCTGTGGTTTACACTAAAAACACAGTCTGTACCGTCTGCGGTATGTGAAAGGTTATGGCTGTAGT

Fig. 15

TGTGATCAACTCCCGAACCATGCTTCAGTCAGCTGATGCACAATCGTTAAACGGGTTGCGGTGAAGTGC  
AGCCC GTCTTACACCGTGCGGCACAGGC ACTAGTACTGATGTCGATA CAGGGCTTTGACATCTACAATGATAAA  
GTAGCTGGTTTGCTAAATTCTAAAAACTAATTGTTGTCGCTCCAAGAAAAGGACGAAGATGACAATTAAATTG  
ATTCTTACTTTGTAGTTAAGAGACACACTTCTCTAACTACCAACATGAAGAAACAATTATAATTACTTAAGGATT  
GTCCAGCTGTTGCTAAACATGACTTCTTAAGTTAGAATAGACGGTGACATGGTACCACTATATCACGTCAACGT  
CTTACTAAATACACAATGGCAGACCTCGTCTATGCTTAAGGCATTGATGAAGGTAAATTGTGACACATTAAAAG  
AAATACTTGTACATACAATTGTTGATGATGATTATTCAATAAAAAGGACTGGTATGATTGTTAGAAAACCA  
GATATATTACCGTATACGCCACTTAGGTGAACGTGTACGCCAGCTTGTAAAAACAGTACAATTCTGTGATG  
CCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGATAATCAAGATCTCAATGGTACTGGTATGATTCCG  
TGATTTCATACAAACCACGCCAGGTAGTGGAGTTCTGTTAGATTCTTATTATTCAATTGTTAATGCCTATATTAAC  
CTTGACCAGGGCTTAACTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTGTTA  
AAATATGACTTCACGGAAAGAGAGGTAAAACCTTTGACCGTTATTAAATATTGGGATCAGACATACCAACCAA  
ATTGTTAACTGTTGGATGACAGATGCATTCTGCATTGTGCAAACCTTAATGTTTATTCTCTACAGTGTCCAC  
CTACAAGTTGGACCACTAGTGAGAAAAATTGTTGATGGTGTCCATTGTTAGTTCAACTGGATACCACTTC  
AGAGAGCTAGGTGTTGACATAATCAGGATGTAACACTACATAGCTCTAGACTTAGTTAAGGAATTACTGTGT  
ATGCTGCTGACCTGCTATGCACGCTGCTCTGGTAATCTATTACTAGATAAACGCACTACGTGCTTCACTAGCT  
GCACTTACTAACATGTTGCTTCAAACACTGTCAAACCCGGTAATTAAACAAAGACTCTATGACTTTGCTGTGTCT  
AAGGGTTCTTAAAGGAAGGAAGTTCTGTTGAATTAAACACTTCTCTTGTGCTCAGGATGGTAATGCTGCTATCAG  
CGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAGACAACTACTATTGTTAGTTGAAGTTGTTG  
ATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAAATGCTAACCAAGTCATCGTCAACAAACCTAGACAAATCAGCT  
GGTTTCCATTAAATAAAATGGGTAAGGCTAGACTTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACTTT  
CGCATATACAAACGTAAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCACAAAGAATAGA

*Fig. 15*

GCTCGCACCGTAGCTGGTCTCTATCTGTAGTACTATGACCAATAGACAGTTCATCAAAAATTATTGAAATCAAT  
AGCCGCCACTAGAGGGAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTGGCACAACATGTTAAAAC  
TGTTTATAGTGTAGAAAACCCTCACCTTATGGGTTGGATTATCCTAAATGTGATAGAGCCATGCCTAACATG  
CTTAGAATTATGGCCTCACTTGTCTGCTCGCAAACATAACAACGTGTTAGCTTGTACACCGTTCTATAGATTA  
GCTAATGAGTGTGCTCAAGTATTGAGTGAATGGTCATGTGTGGCGTTCACTATATGTTAAACCAGGTGGAACCT  
CATCAGGAGATGCCACAAC TGCTTATGCTAATAGTGTAAAAACATTGTCAAGCTGTACCGCCAATGTTAATGCA  
CTTTTATCTACTGATGGTAACAAAATTGCCATAAGTATGTCGCAATTACAACACAGACTTTATGAGTGTCTCTA  
TAGAAATAGAGATGTTGACACAGACTTTGTGAATGAGTTTACGCATATTGCGTAAACATTCTCAATGATGATAC  
TCTCTGACGATGCTGTTGTGTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAG  
TCAGTTCTTATTATCAAACAAATGTTTATGTCGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACC  
TCATGAATTTCGCTCTCAACATAACATGCTAGTTAAACAGGGTGATGATTATGTGTACCTTCCTAACCGATCCAT  
CAAGAACCTAGGGGCGGCTGTTTGTAGATGATCGTAAAACAGATGGTACACTTATGATTGAACGGTTCGT  
GTCTTAGCTATAGATGCTTACCCACTTAAACATCCTAATCAGGAGTATGCTGATGTCTTCAATTGTACTTACA  
ATACATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGTTATGCTTACTAATGATAAC  
ACTTCAAGGTATTGGAACCTGAGTTTATGAGGCTATGTACACACCGCATAACAGTCTTACAGGCTGTTGGGCTT  
GTGTTCTTGCAATTCAAGACTTCAATTAGATGTGGTGCTTGCTACGTAGACCATTCTTATGTTGAAATGCTGT  
TACGACCATGTCATATCAACATCACATAAATTAGTCTGTCTGTTAACCGTATGTTGCAATGCTCCAGGTTGTGAT  
GTCACAGATGTGACTCAACTTACTTAGGAGGTATGAGCTATTATTGAAATCACATAAACCACCCATTAGTTTCC  
ATTGTTGCTAATGGACAAGTTTGGTTATATAAAAATACATGTTGGTAGCGATAATGTTACTGACTTTAATG  
CAATTGCAACATGTGACTGGACAAATGCTGGTATTACATTAGCTAACACCTGACTGAAAGACTCAAGCTTTT  
GCAGCAGAAACGCTCAAAGCTACTGAGGAGACATTAAACTGTTATGGTATTGCTACTGTAACGTGAAGTGCTGT  
CTGACAGAGAATTACATTTCATGGGAAGTTGGTAAACCTAGACCACCTAACCGAAATTATGTCCTTACTGGT

*Pig. 15*

TATCGTGTAACTAAAACAGTAAAGTACAAATAGGAGAGTACACCTTGAAAAAGGTGACTATGGTATGCCATT  
GTTTACCGAGGTACAACAACCTACAAATTAAATGTTGGTATTATTTGTGCTGACATCACATACAGTAATGCCATT  
AAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTACCCAAACACTCAATATCTCAGATG  
AGTTTCTAGCAATGTTGCAAATTCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGGACCACCTGGTAC  
TGGTAAGAGTCATTGCTATTGGCCTAGCTCTACTACCCCTCTGCTCGCATAGTGTATAACAGCTTGCTCTCATGC  
CGCTGTTGATGCACTATGTGAGAAGGCATTAATTCGCTATAGATAATGTAGTAGAATTACCTGCACGT  
GCTCGTAGAGTGTGATAAAATTCAAAGTGAATTCAACATTAGAACAGTATGTCGGTACTGTAATGCATT  
GCCTGAGACGACAGCAGATATAGTGTCTTGATGAAATTCAATGCCACAAATTATGATTTGAGTGTGTCAT  
GCCAGATTACGTGCTAACGCACTATGTGTACATTGGCGACCCCTGCTCAATTACCTGCACCGCACATTGCTAACTAA  
GGGCACACTAGAACCGAAATTTCAATTCAAGTGTGAGACTTATGAAAACATAGGTCCAGACATGTTCTCGGA  
ACTTGTGCGCGTTGCTGCTGAAATTGTTGACACTGTGAGTGTCTGGTTATGATAATAAGCTAAAGCACATAA  
AGACAAATCAGCTCAATGCTTAAATGTTTATAAGGGTGTATCACGCATGATGTTCATCTGCAAATTACAGGC  
CAAATAGGCGTGGTAAGAGAATTCTTACACGTAACCCCTGCTGGAGAAAAGCTGTCCTTATTCACCTTATAAT  
TCACAGAATGCTGTAGCCTCAAAGATTGGGACTACCAACTCAAACGTTGATTCTACAGGGCTCAGAATATG  
ACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTGTAATGTAACAGATTAAATGTTGCTATTACAGA  
GCAAAAGTAGGCATACTTGCTATAATGTCAGAGACCTTATGACAAGTTGCAATTACAAGTCTGAAATTCC  
ACGTAGGAATGTGGCAACTTACAAGCTGAAATGTAACAGGACTCTTAAAGATTGTTAGTAAGGTAATCACTGG  
GTTACATCCTACACAGGCACCTACACACCTCAGTGTGACACTAAATTCAAACGTTAAGGTTATGTTGACATAC  
CTGGCATACTAACAGGACATGACCTATAGAACACTCATCTCTATGATGGGTTAAATGAATTATCAAGTTAATGG  
TTACCTAACATGTTATCACCGCGAAGAAGCTATAAGACATGTACGTGCATGGATTGGCTCGATGTCAGGGG  
TGTCTAGCTAGAGAAGCTGTTGGTACCAATTACCTTACAGCTAGGTTTCTACAGGTGTTAACCTAGTTGC  
TGTACCTACAGGTTATGTTGATACACCTAATAATACAGATTTCAGAGTTAGTGCTAAACCACCGCCTGGAGATC

Fig. 15

AATTAAACACCTCATACCACTTATGTACAAAGGACTTCCTTGGAAATGTAGTGCCTATAAAGATTGTACAAATGTTA  
AGTGACACACTAAAAATCTCTGACAGAGTCGTATTTGTCTTATGGGCACATGGCTTGAGTTGACATCTATGAA  
GTATTTGTGAAAATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTCCACTGCTTCAG  
ACACTTATGCCTGTTGGCATCATTCTATTGGATTACGTCTATAATCCTTATGATTGATGTTCAACAATGGG  
GTTTACAGGTAACCTACAAAGCAACCAGTATCTGTATTGTCAAGTCCATGGTAATGCACATGTAGCTAGTTGTGA  
TGCAATCATGACTAGGTGTAGCTGTCACGAGTGCTTGTAAAGCGTGTGACTGGACTATTGAATATCCTATAA  
TTGGTGTGAACTGAAGATTAAATGCGGCTTGTAGAAAGGTTAACACACATGGTTGTTAAAGCTGCATTATTAGCAGA  
CAAATTCCCAGTTCTCACGACATTGTAACCTAAAGCTATTAAAGTGTGTACCTCAAGCTGATGTAGAATGGAAG  
TTCTATGATGCACAGCCTTGTAGTGACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGA  
CAAATTACAGATGGTGTATGCCTATTGGATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTGTAGAT  
TTGACACTAGAGTGCTATCTAACCTTAACCTGCCTGGTTGTGATGGTGGCAGTTGTATGTAACAAACATGCATT  
CACACACCAGCTTGTATAAAAGTGTCTTGTAAATTAAACAAATTACCACTTCTATTACTCTGACAGTCCATGT  
GAGTCTCATGGAAAACAAGTAGTGTCAGATAGATTATGTAACACTAAAGTCTGCTACGTGTATAACACGTTGCA  
ATTAGGTGGTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTGATGCTTATAACATGATGATCTCA  
GCTGGCTTAGCTTGTGGTTACAAACAATTGATACTTATAACCTCTGGAACACTTACAAGACTTCAGAGTT  
AGAAAATGTGGCTTAAATGTGTAAATAAGGGACACTTGTGATGGACAACAGGGTGAAGTACCAAGTCTTCTATCATT  
AATAACACTGTCTACACAAAGTTGATGGTGTGATGAGTAAATTGTTGAAATAACACATTACCTGTTAATGT  
AGCATTGAGCTTGGCTAACGCGAACATTAAACCAAGTAGTACCAAGAGGTGAAAATCTCAATAATTGGTGTGGA  
CATTGCTGCTAATACTGTGATCTGGACTACAAAGAGATGCTCCAGCACATATCTACTATTGGTGTCTA  
TGACTGACATAGCCAAGAAACCAACTGAAACGATTGTGCACTCACTGTCTTGTGATGGTAGAGTTGATGG  
TCAAGTAGACTTATTAGAAATGCCGTAATGGTGTCTTATTACAGAAGGTAGTGTAAAGGTTACAACCATCTG  
TAGGTCCCAAACAAGCTAGTCTTAATGGAGTCACATTAATTGGAGAAGCCGTAACACAGTTCAATTATTATAA

*Fig. 15*

GAAAGTTGATGGTGTCCAACAATTACCTGAAACTACTTCAGAGTAGAAATTACAAGAATTAAACCCA  
GGAGTCAAATGAAATTGATTCTAGAATTAGCTATGGATGAATTCACTGAACGGTATAAATTAGAAGGCTATGC  
CTTCGAACATATCGTTATGGAGATTTAGTCATAGTCAGTTAGGTGGTTACATCTACTGATTGGACTAGCTAAC  
GTTTAAGGAATCACCTTTGAATTAGAAGATTTATTCTATGGACAGTACAGTTAAAAACTATTCATAACAGAT  
GCGCAAACAGGTTCATCTAAGTGTGTCTGTTATTGATTTATTACTTGATGATTTGTTGAAATAATAAAATC  
CCAAGATTATCTGTAGTTCTAAGGTTGTCAAAGTGACTIONACTACAGAAATTCACTTATGCTTGGTGTA  
AAGATGGCCATGTAGAAACATTACCCAAAATTACAATCTAGTCAGCGTGGCAACCGGGTGGCTATGCCTAA  
TCTTACAAAATGCAAAGAATGCTATTAGAAAAGTGTGACCTCAAAATTATGGTGATAGTGCAACATTACCTAAA  
GGCATAATGATGAATGTCGCAAATATACTCAACTGTGTCAATTAAACACATTAAACATTAGCTGTACCCCTATAA  
TATGAGAGTTATACATTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTACAGCTGTTAAGACAGTGGTTG  
CCTACGGGTACGCTGCTGTCGATTAGCTTAATGACTTTGTCTGATGCGATTCAACTTGTGATTGGTGATTG  
TGCAACTGTACATACAGCTAATAATGGATCTCATTATTAGTATGATGACGACCTAAGACTAAAATGTTACA  
AAAGAAAATGACTCTAAAGAGGGTTTTCACTTACATTGTGGGTTATACAACAAAAGCTAGCTTGGAGGTT  
CCGTGGCTATAAAGATAACAGAACATTGGATGCTGATCTTATAAGCTCATGGGACACTCGCATGGTGGAC  
AGCCTTGTACTAATGTGAATGCGTCATCATCTGAAGCATTTAATTGGATGTAATTATCTGGCAAACCACGCG  
AACAAATAGATGGTTATGTCATGCGATGCAAATTACATATTGGAGGAATACAAATCCAATTAGCTGTCTCCTAT  
TCTTATTGACATGAGTAAATTCCCCCTAAATTAGGGTACTGCTGTTATGCTTTAAAAGAAGGTCAAATCAA  
TGATATGATTTATCTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATG  
TTCTTGTTAACAACTAAACGAACAATTGGTTTTCTGTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAACCTTA  
CAACCAGAACTCAATTACCCCTGCATACACTAATTCTTACACGTGGTGTATTACCCGTACAAAGTTTCACT  
CCTCAGTTTACATTCAACTCAGGACTTGTCTTACCTTCTTCAATGTTACTTGGGTCCATGCTATACATGTCTC  
TGGGACCAATGGTACTAAGAGGTTGATAACCCGTCTTACCATTAATGATGGTGTATTGCTTCCACTGAGA

*Fig. 15*

AGTCTAACATAATAAGAGGCTGGATTTGGTACTACTTAGATTGAAGACCCAGTOCCTACTTATTGTTAATAAC  
GCTACTAAATGTTGTTATTAAAGTCTGTGAATTCAATTGTAATGATCCATTGGGTGTTATTACCAACAAAAAC  
AACAAAAGTTGGATGGAAAGTGAGTTAGAGCTAGTGCAGATAATTGCACCTTGAAATATGTCAGC  
CTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTCAAAAATCTTAGGGAATTGTGTTAAGAATATTGATGG  
TTATTTAAAATATATTCTAACGACACGCCTATTAAATTAGTGCCTGATCTCCCTCAGGGTTTCGGCTTAGAACC  
ATTGGTAGATTGCCAATAGGTATTACATCACTAGGTTCAAACCTTACTTGCTTACATAGAAGTTATTGACTCC  
TGGTGATTCTTCAGGTGGACAGCTGGTGCAGCTTATTATGTGGGTTATCTTCAAOCTAGGACTTTCTAT  
TAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACCTGACCCCTCTCAGAAACAAAGTGTAC  
GTTGAAATCCTTCACTGTAGAAAAGGAATCTATCAAACCTCTAACCTTAGAGTCCAACCAACAGAATCTATTGTTA  
GATTTCTTAATATTACAAACTTGTGCCCTTGGTGAAGTTAACGCCACAGATTGCATCTGTTATGCTTGG  
ACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGCTTATATAATTCCGCATCATTTCCACTTTAAGTGT  
ATGGAGTGTCTCTACTAAATTAAATGATCTCTGCTTACTAATGTCTATGCAGATTCTGTAATTAGAGGTGAT  
GAAGTCAGACAAATCGCTCCAGGGCAAACCTGGAAAGATTGCTGATTATAATTATAATTACAGATGATTACAG  
GCTGCGTTAGCTGGAAATTCTAACAACTCTGATTCTAACAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTT  
AGGAAGTCTAACACCTTGGAGAGAGATATTCAACTGAAATCTATCAGGCCGGTAGCACACCTGTAAATG  
GTGTTGAAGGTTTAATTGTTACTTTCTTACAATCATATGGTTCCAACCCACTAACGGTGTGGTACCAACCAT  
ACAGAGTAGTAGTACTTTCTTGAACCTCACATGCACCACTGTTGGACCTAAAAGTCTACTAATTG  
GTTAAAAACAAATGTGTCAATTCAACTCAATGGTTAACAGGCACAGGTGTTACTGAGTCTAACAAAAAGTT  
TCTGCCTTCCAACAATTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATT  
TTGACATTACACCATGTTCTTGGTGGTGTAGTGTATAACACCAGGAACAAACTTCTAACCAACAGGTTGCTGTT  
CTTATCAGGATGTTACTGCACAGAAGTCCCTGTTGCTATTGCAACTTACTCCTACTGGCGTGTAT  
TCTACAGGTTCTAACGTCAGGCTGTTAACAGGCTGAACATGTCAACAAACTCATATGAGT

Fig. 15

GTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGT  
AGCTAGTCATCCATCATTGCCTACACTATGCACTTGGTCAGAAAATTCAAGTTGCTACTCTAATAACTCTATTGC  
CATACCCACAAATTCTACTATTAGTGTACCCACAGAAATTCAACAGTGTCTATGACCAAGACATCAGTAGATTGA  
CAATGTACATTGTGGTATTCAACTGAATGCAGCAATCTTGTCAATATGGCAGTTTGACACAATTAAAC  
CGTGCTTAACGGAAATAGCTGTTAACAGACAAAAACACCCAGAAGTTTGACAAAGTCAAACAAATTACA  
AAACACCACCAATTAAAGATTGGTGGTTAACAAATTACCAAGATCCATCAAACCAAGCAAGAGG  
TCATTATTGAAGATCTACTTTCAACAAAGTGACACTTGAGATGCTGGCTCATCAAACAATATGGTATTGCCT  
TGGTGTATTGCTGCTAGAGACCTCATTGTGCACAAAAGTTAACGGCCTACTGTTTGCCACCTTGCTCACAG  
ATGAAATGATTGCTCAATACACTCTGCACTGTTAGCGGGTACAATCACTCTGGTGGACCTTGTCAGGTGCT  
GCATTACAAATACCATITGCTATGCAAATGGCTTATAGGTITAATGGTATTGGAGTTACACAGAATGTTCTATGA  
GAACCAAAATTGATTGCCAACCAATTAAATAGTGTATTGGCAAATTCAAGACTCACTTCTTCCACAGCAAGTG  
CACTTGGAAAACCTCAAGATGTGGTCAACCAAAATGCACAAGCTTAAACACGCTTGTAAACAAACTTAGCTCAA  
TTTGTCGAATTCAAGTGTAAATGATATCCTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATA  
GGTTGATCACAGGCAGACTTCAAAGTTGCAGACATATGTGACTCAACAATTAGAGCTGCAGAAATCAGAG  
CTTCTGCTAATCTGCTGCTACTAAATGTCAGAGTGTGTACTTGGACAATCAAAAGAGTTGATTGTGGAAA  
GGGCTATCATCTTATGTCCTCCCTCAGTCAGCACCTCATGGTAGTCTTGTGACTTATGTCCTGCACA  
AGAAAAGAACCTCACAACTGCTCCTGCCATTGTGATGGAAAAGCACACTTCTCGTGAAGGTGTCTTGT  
CAAATGGCACACACTGGTTGTAACACAAAGGAATTATGAACCACAAATCATTACTACAGACAAACACATTGT  
GTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAAACACAGTTATGATCCTTGCAACCTGAATTAGACTCAT  
TCAAGGAGGAGTTAGATAAATATTAAAGAATCATACATCACCAAGATGTTGATTTAGGTGACATCTCTGGCATTAA  
TGCTTCAGTTGTAACATTCAAAAGAAATTGACCGCCTCAATGAGGTTGCCAGAAATTAAATGAATCTCTCATCG  
ATCTCCAAGAACCTGGAAAGTATGAGCAGTATATAAAATGCCATGGTACATTGGCTAGGTATTAGCTGGCTT

Fig. 15

GATTGCCATAGTAATGGTGACAATTATGCTTGTATGACCAGTTGCTGTAGTTGTCATAAGGGCTTTGTTCTT  
GTGGATCCTGCTGCAAATTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAAC  
GAACTTATGGATTITGTTATGAGAATCTTCACAATTGAACTGTAACTTGAAGCAAGGTGAAATCAAGGATGCTA  
CTCCTTCAGATTITGTTCGCGCTACTGCAACGATACGATAACGCCTCACTCCCTTCGGATGGCTTATTGTTGGC  
GTTGCACTTCTTGTGTTTCAAGAGCGCTTCCAAAATCATAACCCCTCAAAAGAGATGGCAACTAGCACTCTCAA  
GGGTGTTCACTTGTGCAACTTGCTGTTGTTGTAACAGTTACTCACACCTTGTCTGTTGCTGGCCT  
TGAAGCCCCCTTCTCTATCTTATGCTTAGTCTACTTGCAGAGTATAAAACTTGTAAAGAATAATAATGAGGCT  
TTGGCTTGTGGAAATGCCGTTCCAAAACCCATTACTTATGATGCCAATCTTGTGGCATACTAATTG  
TTACGACTATTGTATAACCTACAATAGTGTAACTTCTCAATTGTCATTACTTCAGGTGATGGCACAACAAGTCCTAT  
TTCTGAACATGACTACCAGATTGGTGGTTACTGAAAAATGGGAATCTGGAGTAAAAGACTGTGTTGATTACAC  
AGTTACTTCACCTCAGACTATTACCAAGCTGTAACACTCAATTGAGTACAGACACTGGTGTGAAACATGTTACCT  
CTTCATCTACAATAAAATTGTTGATGAGCCTGAAGAACATGTCCAATTCACACAATCGACGGTTCATCGGAGTT  
GTTAACATCCAGTAATGGAACCAATTATGATGAACCGACGACGACTACTAGCGTGCCTTGTAAAGCACAAGCTGATG  
AGTACGAACCTACGAACATAATTATATTAGTTTCTGTTGAACTTTAATTAGCCATGGCAGATTCCAACG  
GTACTATTACCGTTGAAGAGCTAAAAAGCTCTGAACATGGAACCTAGTAATAGGTTCTATTCTTACATGG  
ATTGTCCTACAATTGCTATGCCAACAGGAATAGGTTTGTATATAATTAGTTAATTTCCTCTGGCTGTTA  
TGGCCAGTAACCTAGTTGTGCTGCTGCTGTTACAGAATAAAATTGGATCACC GGTTGGAATTGCTATCGC  
AATGGCTTGTCTGTAGGCTTGATGTGGCTCAGCTACTTCATTGCTTCTTCAAGACTGTTGCGCGTACCGTTCCA  
TGTGGTCATTCAATCCAGAAAACTAACATTCTCTCAACGTGCCACTCCATGGCACTATTCTGACCAGACCGTTCTA  
GAAAGTGAACTCGAATCGGAGCTGTGATCCTCGTGGACATCTCGTATTGCTGGACACCCTAGGACGCTGTG  
ACATCAAGGACCTGCCTAAAGAAACTCACTGTTGCTACATCACGAACGCTTCTTATTACAAATTGGGAGCTTCGCA  
GCGTGTAGCAGGTGACTCAGGTTGCTGCATACAGTCGCTACAGGATTGGCAACTATAAATTAAACACAGACCAT

Fig. 15

TCCAGTAGCAGTGACAATATTGCTTGCTTACAGTAAGTGACAACAGATTTCATCTCGTTGACTTCAGGTTA  
CTATAGCAGAGATATTACTAATTATTATGAGGAATTAAAGTTCCATTGGAATCTGATTACATCATAAACCTCA  
TAATTAAAATTATCTAAGTCACTAACTGAGAATAAATATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGA  
TTAAACGAACATGAAAATTATTCTTCTGGCACTGATAACACTCGCTACTTGTGAGCTTATCACTACCAAGAGT  
GTGTTAGAGGTACAACAGTACTTTAAAAGAACCTGCTCTGGAACATACGAGGGCAATTACCAATTTCATCCT  
CTAGCTGATAACAAATTGCAC TGACTGCTTAGCACTCAATTGCTTGTGCTGACGGCGTAAAACACGT  
CTATCAGTTACGTGCCAGATCAGTTCACCTAAACTGTTCATCAGACAAGAGGAAGTCAAGAACATTACTCTCAA  
TTTTCTTATTGTTGGCAATAGTGTATAACACTTGCCTCACACTCAAAGAAAGACAGAACATGATTGAACCTTC  
ATTAATTGACTTCTATTGTGCTTTAGCCTTCTGCTATTGCTTAAATTATGCTTATTATCTTGGTTCTCAC  
TTGAACTGCAAGATCATAATGAAACTTGTACGCCTAACGAACATGAAATTCTTGTGTTCTTAGGAATCATCACA  
ACTGTAGCTGCATTCACCAAGAACATGTAGTTACAGTCATGTACTCAACATCAACCATAATGTAGTTGATGACCCGTG  
TCCTATTCACTCTATTCTAAATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTAATTGAATTGTGCGTGG  
ATGAGGCTGGTTCTAAATCACCCATTCACTGATATCGGTAAATTACAGTTCTGTTACCTTACAATTAA  
ATTGCCAGGAACCTAAATTGGTAGTCTTAGTGCCTGTTCTGTTATGAAGACTTTAGAGTATCATGACGTT  
CGTGTGTTAGATTCTAAACGAACAAACTAAAATGTCTGATAATGGACCCCCAAAATCAGCGAAATGCACC  
CCGCATTACGTTGGTGGACCCCTCAGATTCAACTGGCAGTAACCAGAACGGAGAACGCAGTGGGGCGCGATCAAA  
ACAACGTGGCCCCAAGGTTACCAATAACTGCGTCTGGTCAACCGCTCTCACTCAACATGGCAAGGAAGAC  
CTTAAATTCCCTGAGGACAAGGCGTCCAATTACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAA  
GAGCTACCAAGACGAATTGCGTGGTGGTACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTCTACTACCTAG  
GAACCTGGCCAGAAGCTGACTTCCCTATGGTGCTAACAAAGACGGCATCATGGGTTGCAACTGAGGGAGCCT  
TGAATACACAAAAGATCACATTGGCACCCTGCTAACAAATGCTGCAATCGTCTACAACACTCCTCAAGG  
AACAAACATTGCCAAAAGGCTTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCTTCTGTTCTCATCACG

Fig. 15

TAGTCGCAACAGTTCAAGAAAATTCAACTCCAGGCAGCAGTAGGGGAACCTCTCCTGCTAGAATGGCTGGCAATGG  
CGGTGATGCTGCTTGTGCTTGCTGCTGACAGATTGAACCAGCTTGAGAGCAAATGCTGGTAAAGGCCAA  
CAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTCTAAGAACGCCTCGGCAAAACGTACTGCC  
ACTAAAGCATACAATGTAACACAAGCTTCGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTGGGGACCAAG  
GAACTAATCAGACAAGGAACGTGATTACAAACATTGCCGCAAATTGCACAATTGCCCGAGCGCTCAGCGTTCT  
TCGGAATGTCGCGCATTGGCATGGAAGTCACACCTCGGAAACGTGGTTGACCTACACAGGTGCCATCAAATTGG  
ATGACAAAGATCCAATTCAAAGATCAAGTCATTGCTGAATAAGCATATTGACGCATACAAACATTCCCACCA  
ACAGAGCCTAAAAAGGACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTACCGCAGAGACAGAACAGCA  
AACTGTGACTCTTCTTGCTGCAGATTGGATGATTCTCAAACAATTGCAACAATCCATGAGCAGTGCTGACT  
CAACTCAGGCCTAAACTCATGCAGACCACACAAGGCAGATGGCTATATAAACGTTTCGCTTCCGTTACGAT  
ATATAGTCTACTCTTGTGCAGAATGAATTCTCGTAACTACATAGCACAAGTAGATGTAGTTAACTTTAATCTCACAT  
AGCAATCTTAATCAGTGTAAACATTAGGGAGGACTTGAAAGAGCCACCACATTTCACCGAGGCCACGCCGAG  
TACGATCGAGTGTACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAAATT  
TAGTAGTGCTATCCCCATGTGATTAAAGCTTCTTAGGAGAATGACAAAAAAAAAAAAAAAAAAAAAA  
AAAAAA (SEQ ID NO:15)

#### *ΔEM GENOME*

ATAAAGGTTATACCTTCCCAGGTAAACAAACCAACCAACTTCTGATCTCTGTAGATCTGTTCTCTAAACGAACCTT  
AAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAACTAATTACTGTCGTTG  
ACAGGACACGAGTAACCTCGTCTATCTCTGCAGGCTGCTACGGTTCGTCCGTGTTGCAGCCGATCATCAGCACA  
TCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAAGATGGAGAGCCTGTCCTGGTTCAACGAGAAAACACACGT

*Fig. 15*

CCAACTCAGTTGCCTGTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTGGAGACTCCGTGGAGGAGGTCTTA  
TCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTGCTCAACTTG  
AACAGGCCATATGTGTTCATCAAACGTTCGGATGCTCGAACCTCATGGTCATGTTATGGTTGAGCTGGTAGC  
AGAACTCGAAGGCATTCAAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCCTCATGTGGCGAAATACC  
AGTGGCTTACCGCAAGGTTCTTCGTAAGAACGGAATAAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCT  
AAAGTCATTGACTTAGGCGACGAGCTGGCACTGATCCTTATGAAGATTCAAGAAAATGGAACACTAAACAT  
AGCAGTGGTGTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGCATAACACTCGCTATGTCGATAACAACCTCT  
GTGGCCCTGATGGCTACCCCTCTGAGTCATTAAAGACCTCTAGCACGTGCTGGTAAAGCTTCATGCACTTGTCC  
GAACAACGGACTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTAAACATGAGCATGAAATTGCTTGGTAC  
ACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTGAAATTAAATTGGCAAAGAAATTGACACCTTCA  
ATGGGGAATGTCCAAATTGTATTCCCTTAAATTCCATAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAA  
GCTTGATGGCTTATGGGTAGAATTGATCTGTCTATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCTTT  
CAACTCTCATGAAGTGTGATCATTGTGGTAAAACCTCATGGCAGACGGGCATTGTTAAAGCCACTTGCATT  
TTGTGGCACTGAGAATTGACTAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCCAAATGCTGTTGTTAAATT  
ATTGTCCAGCATGTCACAATTAGAAGTAGGACCTGAGCATAGTCTGCCAATACCATATGAATCTGGCTTGAA  
AACCATTCTCGTAAGGGTGGTCGCACTATTGCCCTTGGAGGCTGTGTCTCTTATGTTGGTTGCCATAACAAGT  
GTGCCATTGGGTCCACGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGGAGAAGGTTCCGAAG  
GTCTTAATGACAACCTCTTGAAAATCTCAAAAGAGAAAGTCACATCAATATTGTTGGTACTTAAACTTAAT  
GAAGAGATGCCATTATTGGCATCTTCTGCTTCCACAAGTGTGTTGGAAACTGTGAAAGGTTGGATTAA  
TAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTAAAGTTACAAAAGGAAAAGCTAAAAAGGTGCCTGG  
AATATTGGTGAACAGAAATCAAAACTGAGTCCTCTTATGCATTGCACTCAGAGGCTGCTGTGTTGACGATCAAT  
TTTCTCCGCACTCTGAAACTGCTCAAAATTCTGTGCGTGTACAGAAGGCCGCTATAACAATACTAGATGGAA

Fig. 15

TITTCACAGTATTCACTGAGACTCATTGATGCTATGATGTTACATCTGATTGGCTACTAACAACTAGTTGTAATG  
GCCTACATTACAGGTGGTGTTCAGTTGACTTCGCACTGGCTAACTAACATCTTGGCACTGTTATGAAAAACT  
CAAACCCGCTTGTATTGGCTTGAAAGAGAAGTTAAGGAAGGTGTAGAGTTCTTAGAGACGGTTGGAAATTGT  
TAAATTATCTCACCTGTGCTTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAGGAGAGT  
GTTCAGACATTCTTAAGCTTGTAAATAAATTGGCTTGTGCTGACTCTATCATTATTGGTGGAGCTAAACTT  
AAAGCCTTGAATTAGGTGAAACATTGTCACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAA  
GAAACTGGCCTACTCATGCCCTAAAAGCCCCAAAGAAAATTATCTTCTTAGAGGGAGAAACACTCCCACAGAAG  
TGTAAACAGAGGAAGTTGTCTTGTAAAACCTGGTATTACAACCATTAGAACACACTAGTGAAGCTGTTGAAGC  
TCCATTGGTTGGTACACCAAGTTGTATTAAACGGGCTTATGTTGCTCGAAATCAAAGACACAGAAAAGTACTGTGCC  
CTTGCACCTAATATGATGGTAACAAACAATACCTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTGGTG  
ATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTGAACCTGATGAAAGGATTGATAAAG  
TACTTAATGAGAAGTGCTCTGCCATACAGTTGAACTCGGTACAGAAGTAAATGAGTTGCCGTGTTGTGGCAGA  
TGCTGTCATAAAACTTGCAACCAGTATCTGAATTACTACACCCTGGCATTGATTAGATGAGTGGAGTATG  
GCTACATACTACTTATTGATGAGTCTGGTGAAGTTAAATTGGCTTACATATGATTGTTCTTCTACCCCTCCAGAT  
GAGGATGAAGAAGAAGGTGATTGTGAAGAAGAAGAGTTGAGCCATCAACTCAATATGAGTATGGTACTGAAGA  
TGATTACCAAGGTAAACCTTGGAAATTGGGCCACTTCTGCTCTCAACCTGAAGAAGAGCAAGAAGAAGAT  
TGGTTAGATGATGATAGTCAACAACTGTTGGTCAACAAGACGGCAGTGAGGACAATCAGACAACACTATTCAA  
ACAATTGTTGAGGTTCAACCTCAATTAGAGATGGAACCTACACCAGTTGTTCAGACTATTGAAGTGAATAGTTTA  
GTGGTTATTAAAACCTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAAAAGGTAAAACC  
AACAGTGGTTGTTAATGCAGCCAATGTTACCTTAAACATGGAGGAGGTGTGCAGGAGCCTTAAATAAGGCTAC  
TAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCACCTAAAGTGGGTGGTAGTTGTT  
TTAAGCGGACACAATCTGCTTAAACACTGTCTTGTGCGGCCAAATGTTAACAAAGGTGAAGACATTCAAC

Fig. 15

TTCTTAAGAGTGCTTATGAAAATTAAATCAGCACGAAGTCTACTGCACCATTATTACAGCTGGTATTTGGTG  
CTGACCCATACTTCAAGAGTTGTAGATACTGTCGCACAAATGCTACTTAGCTGCTTGATAAAAATC  
TCTATGACAAACTGTTCAAGCTTTGGAAATGAAGAGTGAAAAGCAAGTGAACAAAAGATCGCTGAGATTCC  
TAAAGAGGAAGTTAACGCCATTATAACTGAAAGTAAACCTTCAGTTGAACAGAGAAAACAAGATGATAAGAAAAT  
CAAAGCTTGTGTTGAAGAAGTTACAACAACCTGGAAGAAACTAAGTTCTCACAGAAAACCTGTTACTTTATATTG  
ACATTAATGGCAATCTCATCCAGATTCTGCCACTCTGTTAGTGACATTGACATCACTTCTAAAGAAAGATGCTC  
CATATATAGTGGGTGATGTTCAAGAGGGTGTAACTGCTGTTACCTACTAAAAAGGCTGGTGGCAC  
TACTGAAATGCTAGCGAAAGCTTGAGAAAAGTGCCAACAGACAATTATATAACCACCTACCCGGTCAGGGTTA  
AATGGTTACACTGTAGAGGAGGCAAAGACAGTGCTAAAAAGTGTAAAAGTGCCTTACATTCTACCATCTATTA  
TCTCTAATGAGAAGCAAGAAATTCTTGAACGTGTTCTTGAATTGCGAGAAATGCTTGACATGCAGAAGAAC  
ACGCAAATTAAATGCCGTCTGTGGAAACTAAAGCCATAGTTCAACTATACAGCGTAAATATAAGGGTATTAAA  
ATACAAGAGGGTGTGGTGATTATGGTGCTAGATTACTTACACCAGTAAAACAACGTAGCGTCACTTATCA  
ACACACTTAACGATCTAAATGAAACTCTTGTACAATGCCACTGGCTATGTAACACATGGCTAAATTGGAAAGAA  
GCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTTCTGTTCTCACCTGATGCTTACAGCGTATAA  
TGGTTATCTTACTTCTTCTAAACACCTGAAGAACATTATTGAAACCATCTCACTGCTGGTCCCTATAAAGA  
TTGGTCTATTCTGGACAATCTACACAACTAGGTATAGAATTCTTAAGAGAGGTGATAAAAGTGTATATTACACTA  
GTAATCCTACCACATTCCACCTAGATGGTGAAGTTACACCTTGACAATTCTTAAGACACTCTTCTTGAGAGAA  
GTGAGGACTATTAAGGTGTTACAACAGTAGACAACATTAAACCTCCACACGCAAGTTGTGGACATGTCATGACAT  
ATGGACAACAGTTGGTCCAACCTATTGGATGGAGCTGATGTTACTAAAATAACCTCATAATTACATGAAGG  
TAAAACATTATGTTTACCTAATGATGACACTCTACGTGTTGAGGCTTGAGTACTACCACACAACGTACCTA  
GTTTCTGGTAGGTACATGTCAGCATTAAATCACACTAAAAGTGGAAATACCCACAAGTTAATGGTTAACCT  
ATTAAATGGCAGATAACAACGTATCTGCCACTGCATTGTTAACACTCCAACAAATAGAGTTGAAGTTAACCT

Fig. 15

ACCTGCTCTACAAGATGCTTATTACAGAGCAAGGGCTGGTGAAGCTGCTAACTTTGTGCACTTATCTTAGCCTACT  
GTAATAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAACAATGAGTTACTTGTTCACATGCCAATTAGATT  
TTGCAAAAGAGTCTGAACGTGGTGTAAAACCTGTGGACAACAGCAGACAACCCCTAACGGGTGAGAGCTGT  
TATGTACATGGGCACACTTTCTTATGAACAATTAAAGAAAGGTGTTAGATACCTGTACGTGTGGTAAACAAGCT  
ACAAAATATCTAGTACAACAGGAGTCACCTTTGTTATGATGTCAGCACCATGCTCAGTATGAACCTAACGATG  
GTACATTACTTGTCTAGTGAGTACACTGGTAATTACAGTGTGGTCACTATAAACATATAACTTCAAAGAAA  
TTGTATTGCATAGACGGTGTACTTACAAAGTCCTCAGAACATCAAAGGTCTTACGGATGTTTCTACAAAGA  
AAACAGTTACACAACAACCATAAAACCAGTTACTTATAAATTGGATGGTGTGTTGACAGAAATTGACCCCTAAG  
TTGGACAATTATTATAAGAAAGACAATTCTTATTTCACAGAGCAACCAATTGATCTGTACCAAACCAACCATATCC  
AAACGCAAGCTCGATAATTITAAGTTGTATGTGATAATATCAAATTGCTGATGATTAAACCAAGTTAACTGGTT  
ATAAGAAACCTGCTTCAAGAGAGCTTAAAGTTACATTTCCTGACTTAAATGGTATGTGGTGGCTATTGATTAT  
AAACACTACACACCCCTTTAAGAAAGGAGCTAAATTGTTACATAAACCTATTGTTGGCATGTTAACATGCAAC  
TAATAAAGCCACGTATAAACCAAATACCTGGTGTACGTTGTCTTGGAGCACAAACCAAGTTGAAACATCAAAT  
TCGTTTGATGTACTGAAGTCAGAGGACGCGCAGGGAAATGGATAATCTTGCTGCGAAGATCTAAACCAAGTCT  
GAAGAAGTAGTGGAAAATCCTACCATACAGAAAGACGTTCTGAGTGTAAATGTGAAAACCTACCGAAGTTGTAGGA  
GACATTATACTTAAACCAAGCAAATAATAGTTAAAAATTACAGAACAGGTTGGCCACACAGATCTAATGGCTGCTT  
ATGTAGACAATTCTAGTCTTACTTTAAAGAAACCTAATGAATTATCTAGAGTATTAGGTTGAAAACCCCTGCTACT  
CATGGTTAGCTGCTGTTAATAGTGTCCCTGGATACTATAGCTAATTATGCTAACGCTTTCTAACAAAGTTGTT  
AGTACAACACTAACATAGTTACACGGTGTAAACCGTGTGTTACTAATTATATGCTTATTCTTACTTTATTG  
CTACAATTGTGTACTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAGCAAAGAAC  
TGTAAAGAGTGTGGTAAATTGTCTAGAGGCTTCATTAAATTATTGAAGTCACCTAACCTAAACTGATAAA  
TATTATAATTGGTTTTACTATTAAAGTGTGTTGCCTAGGTTCTTAATCTACTCAACCGCTGCTTAGGTGTTAAATG

*Fig. 15*

TCTAATTAGGCATGCCTTACTGTACTGGTTACAGAGAAGGCTATTGAACCTACTAATGTCACTATTGCAAC  
CTACTGTACTGGTTCTACCTTGAGTGTCTTAGTGGTTAGATTCTTAGACACCTATCCTCTTAGAAACT  
ATACAAATTACCATTTCATCTTTAAATGGGATTAAC TGCTTGGCTTAGTTGCAGAGTGGTTGGCATATATT  
CTTTCACTAGGTTCTATGTACTGGATTGGCTGCAATCATGCAATTGTTTCAGCTATTGCAGTACATTIA  
TTAGTAATTCTGGCTTATGTGGTTAATAATTAAATCTGTACAAATGGCCCCGATTTCAGCTATGGTTAGAATGTAC  
ATCTTCTTGCACTATTATTATGTATGGAAAAGTTATGTGCATGTTAGACGGTTGAATTCAACTTGTATG  
ATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACATTGTTAATGGTGTAGAAGGTCTTTATG  
TCTATGCTAATGGAGGTAAAGGCTTTGCAAACACTACACAATTGGAATTGTGTTAATTGTGATACATTCTGTGCTGGT  
AGTACATTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTAAAAGACCAATAAACACTAAAGGTTATTGCCTATT  
CTTCTTACATCGTTGATAGTGTACAGTGAAGAATGGTCCATCCATCTTACTTTGATAAAAGCTGGTAAAAGACT  
TATGAAAGACATTCTCTCTCATTTGTTAACCTAGACAACCTGAGAGCTAATAACACTAAAGGTTATTGCCTATT  
AATGTTATAGTTTGATGGTAAATCAAATGTGAAGAATCATCTGCAAATCAGCGTCTGTTACTACAGTCAGCT  
TATGTGTCACCTATACTGTTACTAGATCAGGCATTAGTGTCTGATGTTGGTGAATGCGGAAGTTGAGTTAAA  
ATGTTGATGCTTACGTTAACGTTTCAACTTAAACGTACCAATGGAAAAACTCAAACACTAGTTGCAAC  
TGCAGAAGCTGAACCTGCAAAGAATGTGCTTAGACAATGTCTTACTTTATTTCAGCAGCTCGGCAAGGG  
TTTGGTATTGATGCTAATAACTATATGCTACCTATAACAAAGTTGAAAACATGACACCCCGTGAACCTTGGTCTTGT  
ATTGACTGTAGTGCACGTATTAATGCGCAGGTAGCAAAAGTCACAACATTGCTTGTATGGAACGTTAAAG  
ATTTCATGTCATTGTCTGAACAACTACGAAAACAAACGTAGTGTGCTAAAAAGAATAACTACCTTTAAGTTG  
ACATGTGCAACTACTAGACAAGTTGTTAACACAACAAAGATAGCACTTAAGGGTGGAAAATTGTTAATA  
ATTGGTTGAAGCAGTTAACATTAAAGTTACACTTGTGTTCTTTGTTGCTGCTATTCTATTAAATAACACCTGTTCA  
TGTCTATGCTAAACATACTGACTTTCAAGTGAATCATAGGATACAAGGCTATTGATGGTGGTGTCACTCGTGAC

*Fig. 15*

ATAGCATCTACAGATACTGTTTGCTAACAAACATGCTGATTTGACACATGGTTAGCCAGCGTGGTAGTTA  
TACTAATGACAAAGCTGCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTGTGCGCTGGTTGCCT  
GGCACGATATTACGCACAACATAATGGTACTTTGCATTCTACCTAGAGTTTACTGCAGTTGGTAACATCTG  
TTACACACCATAAAACTTATAGAGTACACTGACTTTGCAACATCAGCTGTGTTGGCTGCTGAATGTACAATT  
TTAAAGATGCTCTGGTAAGCCAGTACCATATTGTTATGATACCAATGTACTAGAAGGTTCTGCTTATGAAAGT  
TTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTATTCAATTCTAACACACCTACCTGAAGGTTCTGTT  
AGAGTGGTAACAACCTTGATTCTGAGTACTGTAGGCACGGCACTTGTGAAAGATCAGAAGCTGGTGTGTT  
CTACTAGTGGTAGATGGGTACTTAACAATGATTATTACAGATCTTACCAAGGAGTTCTGTGGTAGATGCTGTA  
AATTACTACTAATATGTTACACCACTAATTCAACCTATTGGTGCTTGGACATATCAGCATCTATAGTAGCTGGT  
GGTATTGTAGCTATCGTAGTAACATGCCTGCCTACTATTATGAGGTTAGAAGAGCTTGGTAATACAGTC  
TGTAGTTGCCCTTAATACTTACTATTCTTATGTCATTCACTGTACTCTGTTAACACCACTTACTCATTCTACCT  
GGTGTATTCTGTATTACTTGACTTGACATTCTTACTAATGATGTTCTTTAGCACATATTCACTGG  
TGGTTATGTTACACCTTAGTACCTTCTGGATAACAATTGCTTATATCATTGTATTCCACAAAGCATTCTATTG  
GTTCTTAGTAATTACCTAAAGAGACGTGAGTCTTAATGGTGTCTTCTTAGTACTTTGAAGAAGCTGCGCTGT  
GCACCTTTGTTAAATAAGAAATGTATCTAAAGTTGCGTAGTGATGTGCTTATTACCTTACGCAATATAATAGA  
TACTTAGCTCTTATAATAAGTACAAGTATTCTAGTGGAGCAATGGATAACAACAGTACAGAGAAGCTGCTTGT  
GTCATCTCGCAAAGGCTCTCAATGACTTCAGGTTCTGATGTTCTTACCAACCACCAACCTCTATC  
ACCTCAGCTGTTGCAGAGTGGTTAGAAAAATGGCATTCCATCTGGTAAAGTTGAGGGTTGTATGGTACAAG  
TAACCTGTGGTACAACACTAACCTAACGGCTTGGCTTGATGACGTAGTTACTGTCCAAGACATGTGATCTGCACC  
TCTGAAGACATGCTAACCTAATTATGAAGATTACTCATTGTAAGTCTAATCATAATTCTGGTACAGGCTGG  
TAATGTTCAACTCAGGGTATTGGACATTCTATGCAAATGTGTACTTAAGCTTAAGGTTGATACAGCCAATCCTA  
AGACACCTAAGTATAAGTTGTTGCATTCAACCAGGACAGACTTTCAAGTGTAGCTTACAATGGTACCCA

Fig. 15

TCTGGTGTACCAATGTGCTATGAGGCCAATTCACTATTAAGGGTCATCCTTAATGGTCATGTGGTAGTGT  
TGGTTAACATAGATTATGACTGTGTCTTTTGTACATGCACCATATGAAATTACCAACTGGAGTCATGCTG  
GCACAGACTTAGAAGGTAACCTTATGGACCTTTGTTGACAGGCAAACAGCACAAGCAGCTGGTACGGACACAA  
CTATTACAGTTAACAGTTAGCTTGGTACGCTGCTGTTAAAATGGAGACAGGTGGTCTCAATCGATTACC  
ACAACCTTAATGACTAACCTTGGCTATGAAGTACAATTATGAACCTCTAACACAAGACCATGTTGACATACT  
AGGACCTCTCTGCTCAAACGGATTGCCGTTAGATATGTGTGCTTCATTAAAAGAATTACTGCAAAATGGTA  
TGAATGGACGTACCATATTGGTAGTGCTTATTAGAAGATGAATTACACCTTTGATGTTAGACAATGCTCA  
GGTGTACTTCCAAAGTGCAGTGAAAAGAACATCAAGGGTACACACCCTGGTGTACTCACAATTGACTT  
CACTTTAGTTAGTCCAGAGTACTCAATGGCTTGTCTTTTGTATGAAAATGCCCTTACCTTGTAT  
GGGTATTATTGCTATGTCTGCTTGCATGATGTTGTCAAACATAAGCATGCATTCTCTGTTGTTGTTAC  
TTCTCTGCCACTGTAGCTTATTAAATATGGCTATATGCCTGCTAGTTGGGTGATGCGTATTATGACATGGTGG  
ATATGGTGTACTAGTTGTCTGGTTAACGCTAAAGACTGTGTTATGATGCATCAGCTGTAGTGTACTAATC  
CTTATGACAGCAAGAACTGTGATGATGGTCTAGGAGAGTGTGGACACTTATGAATGCTTGACACTCGTT  
ATAAAAGTTATTATGGTAATGCTTGTAGATCAAGCCATTCCATGTGGCTCTTATAATCTCTGTTACTTCTAACTACT  
CAGGTGTAGTTACAACGTCTATGTTGGCCAGAGGTATTGTTATGTTGAGTATTGCCCTATTCTTCA  
TAACGGTAATACACTTCAGTGTATAATGCTAGTTATTGTTCTAGGCTATTGTTACTTGGCCTCTT  
TTGTTACTCAACCGCTACTTAGACTGACTCTGGTGTATGATTACTTAGTTCTACACAGGAGTTAGATATAT  
GAATTACAGGGACTACTCCCACCAAGAACATGCATAGATGCCTCAAACACTCAACATTAAATTGTTGGGTGTTGGT  
GGCAAACCTGTATCAAAGTAGCCACTGTACAGTCTAAATGTCAGATGTCAGATGTCACATCAGTAGTCTTACTCT  
CAGTTTGCAACAACTCAGAGTAGAATCATCATCTAAATTGTTGGCTCAATGTGTCCAGTTACACAATGACATTCTC  
TTAGCTAAAGATACTACTGAAGCCTTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATGCCTCAGAGTTAGTTCC  
AGACATAAAACAAGCTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATGCCTCAGAGTTAGTTCC

Fig. 15

CTTCCATCATATGCAGTTTGCTACTGCTCAAGAAGCTTATGAGCAGGCTGGCTAATGGTATTCTGAAGTTGT  
TCTTAAAAAGTTGAAGAAGTCTTGAATGTGGCTAAATCTGAATTGACCGTGATGCAGCCATGCAACGTAAGTTG  
GAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGGCTAGATCTGAGGGACAAGAGGGCAAAAGTTAC  
TAGTGCTATGCAGACAATGCTTTCACTATGCTTAGAAAGTTGGATAATGATGCACTCAACAAACATTATCAACAATG  
CAAGAGATGGTTGTGTCCTTGAACATAATACCTCTTACAACAGCAGCCAAACTAATGGTTGTCAACCAGACTAT  
AACACATATAAAACACGTGTGATGGTACAACATTACTTATGCATCAGCATTGTTGGAAATCCAACAGGGTTGTAG  
ATGCAGATAGTAAAATTGTTCAACTTAGTGAAATTAGTATGGACAATTACCTAATTAGCATGGCCTTTATTGTA  
ACAGCTTAAGGGCCAATTCTGCTGCTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTT  
GTGCTGCCGGTACTACACAAACTGCTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTA  
GGTTTGTACTTGCACTGTTATCCGATTACAGGATTGAAATGGCTAGATTCCCTAAGAGTGATGGAACGGTAC  
TATCTATACAGAACTGGAACCACCTTGTAGGTTACAGACACACCTAAAGGTCTAAAGTGAAAGTATTATACT  
TTATTAAAGGATTAAACAACTAAATAGAGGTATGGTACTTGGTAGTTAGCTGCCACAGTACGTCTACAAGCTGG  
TAATGCAACAGAAGTGCCTGCCAATTCAACTGTATTATCTTCTGCTTTGCTGTAGATGCTGCTAAAGCTTACA  
AAGATTATCTAGCTAGTGGGGACAACCAACTAACTTGTGTTAAGATGTTGTGTACACACACTGGTACTGGTCA  
GGCAATAACAGTTACACCGGAAGCCAATATGGATCAAGAATCCTTGGTGGTCATCGTGGTCTGACTGCCGT  
TGCCACATAGATCATCCAAATCCTAAAGGATTTGTGACTTAAAGGTAAAGTGATGCTGCTAAAGCTTACA  
TAATGACCTGTGGGTTTACACTAAAAACACAGTCTGTACCGTCTGCGGTATGTGAAAGGTTATGGCTGTAGT  
TGTGATCAACTCCCGAACCCATGCTCAGTCAGCTGATGCACAATCGTTTAAACGGGTTGCGGTGTAAGTGC  
AGCCCGTCTTACACCGTGCAGGCACAGGCACTAGTACTGATGTCGTATACAGGGCTTTGACATCTACAATGATAAA  
GTAGCTGGTTTGTAAATTCTAAAAACTAATTGTTGTCGCTTCCAAGAAAAGGACGAAGATGACAATTAAATTG  
ATTCTTACTTTGTAGTTAAGAGACACACTTCTCAACTACCAACATGAAGAAACAATTATAATTACTTAAGGATT  
GTCCAGCTGTTGCTAACATGACTCTTAAGTTAGAATAGACGGTGACATGGTACCATATCACGTCAACGT

Fig. 15

CTTACTAAATACACAATGGCAGACCTCGTCTATGCTTAAGGCATTTGATGAAGGTAATTGTGACACATTAAG  
AAATACTTGTACACATACAATTGTTGTGATGATGATTATTCAATAAAAAGGACTGGTATGATTTGATGAAAACCCA  
GATATATTACCGTACGCCAACCTAGGTGAACGTGTACGCCAAGCTTGTAAAAACAGTACAATTCTGTGATG  
CCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTCGG  
TGATTTCATACAAACCACGCCAGGTAGTGGAGTCTGTTAGATTCTTATTTCATTGTTAATGCTATATTAAAC  
CTTGACCAGGGCTTAAC TG CAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTGTTA  
AAATATGACTTCACGGAAGAGAGGTTAAACTCTTGACCGTTATTAAATATTGGGATCAGACATACCAACCAA  
ATTGTTAAC TG TGGATGACAGATGCATTCTGCATTGTGCAAACCTTAATGTTTATTCTCACAGTGTCCCAC  
CTACAAGTTTGACCACTAGTGAGAAAAATTGTTGATGGTGTCCATTGTTAGTTCAACTGGATACCACTTC  
AGAGAGCTAGGTGTTGACATAATCAGGATGTAACATAGCTAGACTTAGTTAAGGAATTACTTGTT  
ATGCTGCTGACCTGCTATGCACGCTGCTCTGGTAATCTATTAGATAAACGCACTACGTGCTTCAGTAGCT  
GCACTTACTAACATGTTGCTTTCAAACGTCAAACCCGGTAATTAAACAAAGACTCTATGACTTTGCTGTCT  
AAGGGTTCTTAAGGAAGGAAGTTCTGTTGAATTAAACACTCTTCTTGCTCAGGATGGTAATGCTGCTATCAG  
CGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAGACAACTACTATTGTTAGTTGAAGTTGTTG  
ATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAATGCTAACCAAGTCATCGTCAACAAACCTAGACAAATCAGCT  
GGTTTCCATTAAATAAAATGGGTAAGGCTAGACTTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACTTT  
CGCATATACAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGA  
GCTCGCACCGTAGCTGGTGTCTATCTGTAGTACTATGACCAATAGACAGTTCATAAAAATTATTGAAATCAAT  
AGCCGCCACTAGAGGGAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTGGCACAACATGTTAAAAC  
TGTTTATAGTGTAGAAAACCTCACCTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCAACATG  
CTTAGAATTATGGCCTCACTTGTTGCTCGCAAACATACAAACGTGTTGATGCTGTACACCGTTCTATAGATTA  
GCTAATGAGTGTGCTCAAGTATTGAGTGAATGGTCAATGTGTGGGGTTCACTATATGTTAAACCAGGTGGAACCT

*Fig. 15*

CATCAGGAGATGCCACAACGTCTATGCTAATAGTGTAACTTACATTGTCAAGCTGTCACGGCCAATGTTAATGCA  
CTTTATCTACTGATGGTAACAAAATTGCCGATAAGTATGTCGCATTTACAACACAGACTTATGAGTGTCTCTA  
TAGAAATAGAGATGTTGACACAGACTTGTGAATGAGTTACGCATTTGCGTAAACATTCTCAATGATGATAC  
TCTCTGACGATGCTGTTGTGTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAG  
TCAGTTCTTATTATCAAACAAATGTTTATGTCGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACC  
TCATGAATTGCTCTCAACATACAATGCTAGTTAACACAGGGTGATGATTATGTCACCTTCCTACCCAGATCCAT  
CAAGAACCTAGGGGCCGGCTGTTGTAGATGATATCGTAAAAACAGATGGTACACTTATGATTGAACGGTCGT  
GTCTTAGCTATAGATGCTTACCCACTAAACATCCTAACAGGAGTATGCTGATGTCTTCATTGTACTTACA  
ATACATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGTTATGCTTACTAATGATAAC  
ACTTCAAGGTATTGGGAAACCTGAGTTTATGAGGCTATGTACACACCGCATACTGCTTACAGGCTGTTGGGCTT  
GTGTTCTTGCAATTACAGACTTCATTAAGATGTGGTCTGCATACTGAGACCATTCTATGTTGAAATGCTGT  
TAOGACCATGTCATATCAACATCACATAAATTAGTCTTGTCTGTTAACCGTATGTTGCAATGCTCCAGGTTGTGAT  
GTCACAGATGTGACTCAACTTACTTAGGAGGTATGAGCTATTATGTTAAATCACATAAACCAACCCATTAGTTTCC  
ATTGTGTGCTAATGGACAAGTTTGGTTATATAAAAATACATGTGTTGGTAGCGATAATGTTACTGACTTTAATG  
CAATTGCAACATGTGACTGGACAAATGCTGGTGATTACATTAGCTAACACCTGACTGAAAGACTCAAGCTTTT  
GCAGCAGAAACGCTCAAAGCTACTGAGGAGACATTAAACTGCTTATGGTATTGCTACTGTAAGTGCTGT  
CTGACAGAGAATTACATCTTCTGGAGTTGGTAAACCTAGACCACCACTAACCGAAATTATGTCCTTACTGGT  
TATCGTGTAACTAAAACAGTAAAGTACAATAGGAGAGTACACCTTGAAAGGTGACTATGGTATGCTGTT  
GTTTACCGAGGTACAACAACCTACAAATTAAATGTTGGTATTATTTGTGCTGACATCACATACAGTAATGCCATT  
AAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATACCCAAACACTCAATATCTCAGATG  
AGTTTCTAGCAATGTTGCAAATTCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGGACCACCTGGTAC  
TGGTAAGAGTCATTGCTATTGGCTAGCTCTACTACCCCTCTGCTCGCATAGTGTATACAGCTGCTCTCATGC

*Fig. 15*

CGCTGTTGATGCACTATGTGAGAAGGCATAAAATTTGCCTATAGATAAATGAGTAGAATTACCTGCACGT  
GCTCGTGTAGAGTGTGATCAAAGTGAATTCAACATTAGAACAGTATGTCCTTGACTGTAAATGCATT  
GCCTGAGACGACAGCAGATATAGTTGTCTTGATGAAATTCAATGCCACAAATTATGATTGAGTGTGTCAT  
GCCAGATTACGTGCTAACGCACTATGTCACATTGGCACCCTGCTCAATTACCTGCACCGCACATTGCTAACTAA  
GGGCACACTAGAACCGAAATTTCAATTCACTAGTGTAGACTTATGAAAACATAGGTCCAGACATGTTCTCGGA  
ACTTGTGGCGTTGCTGCTGAAATTGTTGACACTGTGAGTGCTTGGTTATGATAATAAGCTAAAGCACATAA  
AGACAAATCAGCTCAATGCTTAAATGTTTATAAGGGTGTATCACGCATGATGTTCATCTGCAATTAAACAGGC  
CACAAATAGGCGTGGTAAGAGAAATTCTTACACGTAACCCGCTGGAGAAAAGCTGCTTTATTCACCTTATAAT  
TCACAGAAATGCTGTAGCCTCAAAGATTGGACTACCAACTCAAACGCTACTCTGTAATGTAACAGATTAAATGTTGCTATTACAGA  
ACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTGTAATGTAACAGATTAAATGTTGCTATTACAGA  
GCAAAAGTAGGCATACTTGCTATAATGTCAGAGACCTTATGACAAGTTGCAATTACAAGTCTGAAATTCC  
ACGTAGGAATGTGGCAACTTACAAGCTGAAATGTAACAGGACTCTTAAAGATTGTAAGGTAATCACTGG  
GTTACATCCTACACAGGCACCTACACACCTCAGTGTGACACTAAATTCAAACGCTGAAGGTTATGTTGACATAC  
CTGGCATACTAACAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTAAATGAATTATCAAGTTAATGG  
TTACCTAACATGTTATCACCGCGAAGAAGCTATAAGACATGTCAGTGCTGGATTGGCTTCGATGTCGAGGGG  
TGTCTGCTACTAGAGAAGCTGTTGGTACCAATTACCTTACAGCTAGGTTTCTACAGGTGTTAACCTAGTTGC  
TGTACCTACAGGTTATGTTGATAACACCTAATAATACAGATTTCAGAGTTAGTGCTAAACCACCGCCTGGAGATC  
AATTAAACACCTACACCACTATGTACAAAGGACTCCTTGGATAGTGCGTATAAAGATTGACAAATGTTA  
AGTGACACACTAAAAATCTCTGACAGAGTCGTATTGCTTATGGCACATGGCTTGAGTTGACATCTATGAA  
GTATTTGTGAAAATAGGACCTGAGCGCACCTGTTGCTATGTGATAGACGTGCCACATGCTTTCCACTGCTTCAG  
ACACTTATGCTGTTGGCATCTTCTATTGGATTGATTACGTCTATAATCCGTTATGATTGATGTTCAACAATGGG  
GTTTACAGGTAACCTACAAAGCAACCAGATCTGTATTGTCAGTCCATGGTAATGCACATGTAGCTAGTTGTGA

*Fig. 15*

TGCAATCATGACTAGGTGTCTAGCTGCCACGAGTGCTTGTAAAGCGTGTGACTGGACTATTGAATATCCTATAA  
TTGGTGATGAAGTTAACCGGGCTTGTAGAAAGGTTCAACACATGGTGTAAAGCTGCATTATTAGCAGA  
CAAATTCCAGTTCTCACGACATTGGTAACCTAAAGCTATTAGTGTACCTCAAGCTGATGTAGAATGGAAG  
TTCTATGATGCACAGCCTTGTAGTGACAAAGCTTATAAAAATAGAAGAATTATTCTATTCTATGCCACACATTCTGA  
CAAATTACAGATGGTGTATGCCTATTTGGAATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTGTAGAT  
TTGACACTAGAGTGTCTAACCTAACCTGGTGTGATGGTGGCAGTTGTATGTAACATGCA  
CACACACCAGCTTGTATAAAAGTGTCTTGTAAATTAAAACAATTACCACTTCTATTACTCTGACAGTCCATGT  
GAGTCTCATGGAAAACAAGTAGTGTCAAGATAGATTGTACCACTAAAGTCTGCTACGTGTATAACACGTTGCA  
ATTAGGTGGTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTGATGCTTATAACATGATGATCTCA  
GCTGGCTTAGCTTGTGGTTACAAACAATTGATACTTATAACCTCTGGAACACTTTACAAGACTTCAGAGTT  
AGAAAATGTGGCTTTAACATGTTGAAATAAGGGACACTTGTGATGGGACAACAGGGTGAAGTACCACTTGTATCATT  
AATAACACTGTTACACAAAAGTTGATGGTGTGATGTTAGAATTGTTGAAAATAACACATTACCTGTTAATGT  
AGCATTGAGCTTGGGCTAACGCAACATTAAACCACTGGGACTACAAAGAGATGCTCCAGCACATATCTACTATTGGTGTG  
CATTGCTGCTAATACTGTGATCTGGGACTACAAAGAGATGCTCCAGCACATATCTACTATTGGTGTGTTCTA  
TGACTGACATAGCCAAGAAACCAACTGAAACGATTGTGACCACTCACTGCTTTGTGATGGTAGAGTTGATGG  
TCAAGTAGACTTATTAGAAATGCCGTAATGGTGTCTTATTACAGAAGGTAGTGTAAAGGTTACAACCATCTG  
TAGGTCCAAACAAGCTAGTCTTAATGGAGTCACATTAAATTGGAGAAGCCGAAAAACACAGTCAATTATTATAA  
GAAAGTTGATGGTGTGCAACAATTACCTGAAACTTACTTACTCAGAGTAGAAATTACAAGAATTAAACCA  
GGAGTCAAATGGAATTGATTCTAGAATTAGCTATGGATGAATTGACAGGTATAAAATTAGAAGGCTATGC  
CTTCGAACATATCGTTATGGAGATTAGTCAGTTAGGTGGTTACATCTACTGATTGGACTAGCTAAC  
GTTTAAGGAATCACCTTGTAAATTAGAAGATTATTCTATGGACAGTACAGTTAAAAACTATTCTATAACAGAT  
GCGCAAACAGGTTCATCTAAGTGTGTGTTATTGATTACTTGATGATTGTTGAAATAATAAAATC

Fig. 15

CCAAGATTATCTGTAGTTCTAAGGTTGTCAAAGTGACTATTGACTATACAGAAATTCAATTATGCTTGGTGTA  
AAGATGGCCATGTAGAACATTACCCAAAATTACAATCTAGTCAGCGTGGCAACCGGGTGTGCTATGCCTAA  
TCTTTACAAAATGCAAAGAATGCTATTAGAAAAGTGTGACCTC AAAATTATGGTGATAGTGCACACATTACCTAAA  
GGCATAATGATGAATGTCGAAAATATACTCAACTGTGTCAATATTAAACACATTAAACATTAGCTGTACCCATAA  
TATGAGAGTTATACATTGGTGCTGGTTCTGATAAAGGAGTTGCACCAAGGTACAGCTGTTTAAGACAGTGGTTG  
CCTACGGGTACGCTGCTTGTGATTAGATCTTAATGACTTTGTCTGTGATGCAGATTCAACTTGTGATTGGTATTG  
TGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGTACAGACCTAAGACTAAAAATGTTACA  
AAAGAAAATGACTCTAAAGAGGGTTTTCACTTACATTGTGGGTTTATACAACAAAAGCTAGCTTGGAGGTT  
CCGTGGCTATAAAGATAACAGAACATTCTTGAATGCTGATCTTATAAGCTCATGGGACACTTCGATGGTGGAC  
AGCCTTGTACTAATGTGAATGCGTCATCATCTGAAGCATTTAATTGGATGTAATTATCTTGGCAAACCACGCG  
AACAAATAGATGGTTATGTCATGCAAAATTACATATTGGAGGAATACAAATCCAATTCAAGTTGTCTTCCTAT  
TCTTATTGACATGAGTAAATTCCCCCTAAATTAAAGGGTACTGCTGTTATGTCTTAAAAGAAGGTCAAATCAA  
TGATATGATTATCTCTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTCTAGTGTG  
TTCTTGTAAACAACAAACGAAACATGTTGTTTCTTGTATTGCCACTAGTCTCTAGTCAAGTGTGTTAATCTTA  
CAACCAGAACTCAATTACCCCTGCATACTAACTATTCTTACACGTGGTGTATTACCCCTGACAAAGTTTCAGAT  
CCTCAGTTTACATTCAACTCAGGACTTGTCTTACCTTCTTCCAATGTTACTTGGTCCATGCTATACATGTCTC  
TGGGACCAATGGTACTAAGAGGTTGATAACCCCTGCTCCTACCATTTAATGATGGTGTATTCTGCTTCCACTGAGA  
AGTCTAACATAATAAGAGGCTGGATTTGGTACTACTTAAAGATTGAAAGACCCAGTCCCTACTTATTGTTAATAAC  
GCTACTAATGTTGTTATTAAAGTGTGAAATTCAATTGTAAATGATCCATTGGGTGTATTACCAACAAAAC  
AACAAAAGTTGGATGGAAAGTGAAGTTGAGAGTTATTCTAGTGCAGATAATTGCACTTTGAATATGTCTCTCAGC  
CTTCTTCTTATGGACCTTGAAGGAAAACAGGGTAATTCAAAAATCTTAGGGAATTGTGTTAAGAATATTGATGG  
TTATTAAATATTCTAACGACACGCCTATTAAATTAGTGCCTGATCTCCCTCAGGGTTTTCGGCTTGTAGAACC

*Fig. 15*

ATGGTAGATTGCCAATAGGTATTAACATCACTAGGTTCAAACCTTACTGCTTACATAGAAGTTATTGACTCC  
TGGTGATTCTTCAGGTTGGACAGCTGGGCTGCAGCTTATTATGTGGGTTATCTTCACCTAGGACTTTCTAT  
TAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCCTCTCTCAGAAACAAAGTGTAC  
GTTGAAATCCTCACTGTAGAAAAAGGAATCTATCAAACCTCTAACCTTAGAGTCCAACCAACAGAACATCTATTGTTA  
GATTTCTTAATATTACAAACTTGTGCCCTTGGTGAAGTTAACGCCACCAGATTGCATCTGTTATGCTTGG  
ACAGGAAGAGAACATCAGCAACTGTGTTGCTGATTATTCTGCTCTATATAATTCCGCATCATTTCCACTTTAAGTGT  
ATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTACTAATGTCTATGCAGATTCTTGTAAATTAGAGGTGAT  
GAAGTCAGACAAATCGCTCCAGGGCAAACCTGAAAGATTGCTGATTATAATTAAATTACAGATGATTACAG  
GCTCGTTATAGCTTGGAAATTCTAACAACTTGTGATTCTAACAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTT  
AGGAAGTCTAACACCTTTGAGAGAGATATTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAAATG  
GTGTTGAAGGTTTAATTGTTACTTTCTTACAATCATATGGTTCCAACCCACTAACGGTGTGGTTACCAACCAT  
ACAGAGTAGTAGTACTTTCTTGAACCTACATGCACCAAGCACTGTTGTGGACCTAAAAGTCTACTAATTG  
GTTAAAAACAAATGTGTCATTCAACTCAATGGTTAACAGGCACAGGTGTTACTGAGTCTAACAAAAAGTT  
TCTGCCTTCCAACAATTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTGAGATT  
TTGACATTACACCATGTTCTTGGTGGTCACTGTTATAACACCAGGAACAAACTTCTAACCGGTTGCTGTT  
CTTATCAGGATGTTACTGCACAGAAGTCCCTGTTGCTATTGATGCTGAGATCAACTTACTCCTACTGGCGTGT  
TCTACAGGTTCTAACGGTTCAAACACGTGCAGGCTGTTAACAGGGCTGAACATGTCAACAACTCATATGAGT  
GTGACATACCCATTGGTGCAGGTATATGCGCTAGTTACTGAGACTCAGACTAACCTCCTGGCGGGCACGTAGTGT  
AGCTAGTCAATCCATCATTGCTACACTATGTCACGGTGCAGAAAATTCACTGCTTACTCTAACAACTCTATTG  
CATACCCACAAATTACTATTAGTGTACCGACAGAAATTCTACCAAGTGTCTATGACCAAGACATCAGTAGATTGTA  
CAATGTACATTGTTGGTCAACTGAATGCAGCAATTGTTGCAATATGGCAGTTTGACACAAGTCAAACAAATTACA  
CGTGTAACTGGAATAGCTGTTGAAACAAGACAAAAACACCAAGAAGTTTGACACAAGTCAAACAAATTACA

*Fig. 15*

AAACACCACCAATTAAAGATTTGGTGGTTTAATTTCACAAATATTACCAAGATCCATCAAAACCAAGCAAGAGG  
TCATTATTGAAGATCTACTTTCAACAAAGTGACACTTGAGATGCTGGCTCATCAAACAATATGGTATTGCCT  
TGGTGTATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTAACGGCCTACTGTTTGCCACCTTGCTCACAG  
ATGAAATGATTGCTCAATAACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGACCTTGTCAGGTGCT  
GCATTACAAATACCATTGCTATGCAAATGGCTATAGGTTAATGGTATTGGAGTTACACAGAATGTTCTATGA  
GAACCAAAAATTGATTGCCAACCAATTAAATAGTGTATTGGCAAAATTCAAGACTCACTTCTTCCACAGCAAGTG  
CACTTGGAAAACCTCAAGATGTGGTCAACCAAAATGCACAAGCTTAAACACGCTTGTAAACAACCTTAGCTCAA  
TTTGGTGCATTCAAGTGTAAATGATATCCTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATA  
GGTTGATCACAGGCAGACTCAAAGTTGCAGACATATGTGACTCAACAATTAAATTAGAGCTGCAGAAATCAGAG  
CTTCTGCTAATCTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAGAGTTGATTTGTGGAAA  
GGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTAGTCTTGCATGTGACTTATGTCCTGCACA  
AGAAAAGAACCTCACAACTGCTCCTGCCATTGTCATGATGGAAAAGCACACTTICCTCGTGAAGGTGTCTTGT  
CAAATGGCACACACTGGTTGTAACACAAAGGAATTTTATGAACCACAAATCATTACTACAGACAACACATTGT  
GTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTATGATCCTTGCAACCTGAATTAGACTCAT  
TCAAGGAGGAGTTAGATAAAATATTTAAGAATCATACATCACAGATGTTGATTAGGTGACATCTCTGGCATTAA  
TGCTTCAGTTGTAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTAAATGAATCTCTCATCG  
ATCTCCAAGAACCTGGAAAGTATGAGCAGTATATAAAATGCCATGGTACATTGGCTAGGTTTATAGCTGGCTT  
GATTGCCATAGTAATGGTACAATTGCTTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGGTTCT  
GTGGATCCTGCTGCAAATTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAAC  
GAACCTATGGATTGTTATGAGAATCTCACAATTGGAACGTAACTTGAAGCAAGGTGAAATCAAGGATGCTA  
CTCCTTCAGATTGTTGCGCTACTGCAACGATAACGATAAGCCTCACTCCCTTGGATGGCTTATTGTTGGC  
GTTGCACTTCTGCTGTTTCAAGAGCGCTTCAAAATCATAACCTCAAAAGAGATGGCAACTAGCACTCTCAA

Fig. 15

GGGTGTTCACTTGTTGCAACTGCTGTTGTTGAACAGTTACTCACACCTTGCTCGTGTGCTGGCCT  
TGAAGCCCTTTCTATCTTATGCTTAGTCTACTTCGCAGAGTATAAACTTGTAAGAATAATAATGAGGCT  
TTGGCTTGCTGGAAATGCCGTTCCAAAAACCCATTACTTATGATGCCAACTATTTCTTGCTGGCATACTAATTG  
TTACGACTATTGTATAACCTACAATAGTGTAACTTCTCAATTGTCATTACTCAGGTGATGGCACAACAAGTCCTAT  
TTCTGAACATGACTACCAGATTGGTGGTTACTGAAAAATGGGAATCTGGAGTAAAAGACTGTGTTGATTACAC  
AGTTACTTCACCTCAGACTATTACCAAGCTGTACTCAACTCAATTGAGTACAGACACTGGTGTGAACATGTTACCTT  
CTTCATCTACAATAAAATTGTTGATGAGCCTGAAGAACATGTCCAATTACACAAATCGACGGTTATCCGGAGTT  
GTTAACCTCAGTAATGGAACCAATTATGATGAACCGACGACGACTACTAGCGTGCCTTGTAAGCACAAGCTGATG  
AGTACGAACATTGTGACAACAGATGTTCATCTCGTTGACTTTCAGGTTACTATAGCAGAGATAATTACTAATTATTAT  
GAGGACTTTAAAGTTCCATTGGAATCTTGATTACATCATAAACCTCATAATTAAAAATTATCTAAGTCACTAAC  
TGAGAATAAAATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGATTAAACGAACATGAAAATTATTCTTTC  
TTGGCACTGATAACACTCGCTACTTGTGAGCTTATCACTACCAAGAGTGTGTTAGAGGTACAACAGTACTTTAA  
AGAACCTTGCTCTGGAACATACGAGGGCAATTCAACATTCTAGCTGATAACAAATTGCACTGACT  
GCTTAGCACTCAATTGCTTTGCTTGCTGACGGCGTAAAACACGTCTATCAGTTACGTGCCAGATCAGTTCA  
CCTAAACTGTTCATCAGACAAGAGGAAGTTCAAGAACCTTACTCTCAATTCTTATTGTTGGCAATAGTGT  
TATAACACTTGTCTCACACTCAAAAGAAAGACAGAATGATTGAACCTTCAATTGACTTCTATTGTGCTTTTA  
GCCTTCTGCTATCCCTGTTTAATTATGCTTATTATCTTTGGTTCTCACTGAACTGCAAGATCATAATGAAACTT  
GTCACGCTAACGAACATGAAATTCTTGTGTTCTTAGGAATCATCACAACGTAGCTGCATTCAACAGAATGT  
AGTTTACAGTCATGACTCAACATCAACCATATGTAGTTGATGACCCGTGTCTTACTTCTATTCTAAATGGTAT  
ATTAGAGTAGGAGCTAGAAAATCAGCACCTTAATTGAATTGTGCGTGGATGAGGCTGGTCTAAATCACCCATT  
AGTACATCGATATCGGTAAATTACAGTTCTGTTACCTTTACAATTAAATTGCCAGGAACCTAAATTGGTAGT  
CTTGTAGTGCCTGTTCTATGAAGACTTTAGAGTATCATGACGTTCTGTTAGATTCTATCTAAACG

Fig. 15

AACAAACTAAAATGTCTGATAATGGACCCAAAATCAGCGAAATGCACCCCGCATTACGTTGGTGGACCCCTCAGA  
TTCAACTGGCAGTAACCAGAACGAGAACGAGTCAGTGGGGCGCATCAAAACACGTCGGCCCCAAGGTTACCCAA  
TAATACTGCGTCTTGGTTACCGCTCTCACTCAACATGGCAAGGAAGACCTAAATTCCCTCGAGGACAAGGCGTT  
CCAATTAAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAAGACGAATTGTGGTGGT  
GACGGTAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACGGGCCAGAACGCTGGACTCCCT  
ATGGTGCTAACAAAGACGGCATCATGGGTTGCAACTGAGGGAGCCTGAATAACACCAAAAGATCACATTGGCA  
CCCGCAATCCTGCTAACAAATGCTGCAATCGTGTACAACCTCCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGC  
AGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTCTCGTCTCATCACGTAGTCGAACAGTCAAGAAATTCAAC  
TCCAGGCAGCAGTAGGGGAACCTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTTGTGCTGCT  
GCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACACAACAAGGCCAAACTGTCACTAA  
GAAATCTGCTGCTGAGGCTCTAAGAACGCTCGGCAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCT  
TTGGCAGACGTGGTCCAGAACAAACCAAGGAATTGGGGACCCAGGAACATCAGACAAGGAACGTGATTAC  
AAACATTGGCCGCAAATTGCACAATTGGCCCCAGCGCTTCAGCGTTCTCGGAATGTCGCATGGCATGGAAG  
TCACACCTCGGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTCAAAGATCA  
AGTCATTGCTGAATAAGCATATTGACGCATAAAACATTCCCACCAACAGAGCCTAAAAGGACAAAAAGAA  
GAAGGCTGATGAAACTCAAGCCTACCGCAGAGACAGAACAGCAAACGCTAAACTGTGACTCTTCTCTGCTGAGA  
TTGGATGATTCTCAAACAATTGCAACAATTCCATGAGCAGTGACTCAACTCAGGCTAAACTCATGCAGACC  
ACACAAGGCAGATGGCTATATAACGTTTCGCTTCCGTTACGATATAGTCACTCTGTGAGAATGAAT  
TCTCGTAACATAGCACAAGTAGATGTAGTTAACCTTAATCTCACATAGCAATTAACTCAGTGTGAAACATTA  
GGGAGGACTTGAAAGAGCCACCATTTACCGAGGCCACGCGGAGTACGATCGAGTGTACAGTGAACAATGCT  
AGGGAGAGCTGCCTATATGGAAGAGCCCTAACGTGTAATTAGTAGTGCTATCCCCATGTGATTAAAT  
AGCTTCTTAGGAGAATGACAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO:16)

Fig. 15

*hACE2*

ATGTCAAGCTTCCCTGGCTCCTTCAGCCTGTTGCTGTAAC TGCTCAGTCCACC  
ATTGAGGAACAGGCCAAGACATTGGACAAGTTAACCAACGAAGCCGAAGACCTGTT  
TATCAAAGTTCACTTGCTTCTTGAATTATAACACCAATATTACTGAAGAGAAATGTCAA  
AACATGAATAATGCTGGGGACAAATGGTCTGCCTTTAAAGGAACAGTCCACACTTGCC  
CAAATGTATCCACTACAAGAAATTAGAATCTCACAGTCAAGCTTCAGCTGCAGGCTCTT  
CAGCAAAATGGGTCTTCAGTGCTCTCAGAAGACAAGAGCAAACGGTTGAACACAATTCTA  
AATACAATGAGCACCCTACAGTACTGGAAAAGTTGTAACCCAGATAATCCACAAGAA  
TGCTTATTACTTGAAACCAGGTTGAATGAAATAATGGCAAACAGTTAGACTACAATGAG  
AGGCTCTGGCTGGAAAGCTGGAGATCTGAGGTGGCAAGCAGCTGAGGCCATTATAT  
GAAGAGTATGTGGTCTTGAAAAATGAGATGGCAAGAGCAAATCATTATGAGGACTATGGG  
GATTATTGGAGAGGAGACTATGAAGTAAATGGGTAGATGGCTATGACTACAGCCGCGGC  
CAGTTGATTGAAGATGTGGAACATACCTTGAAGAGATTAACCATTATATGAACATCTT

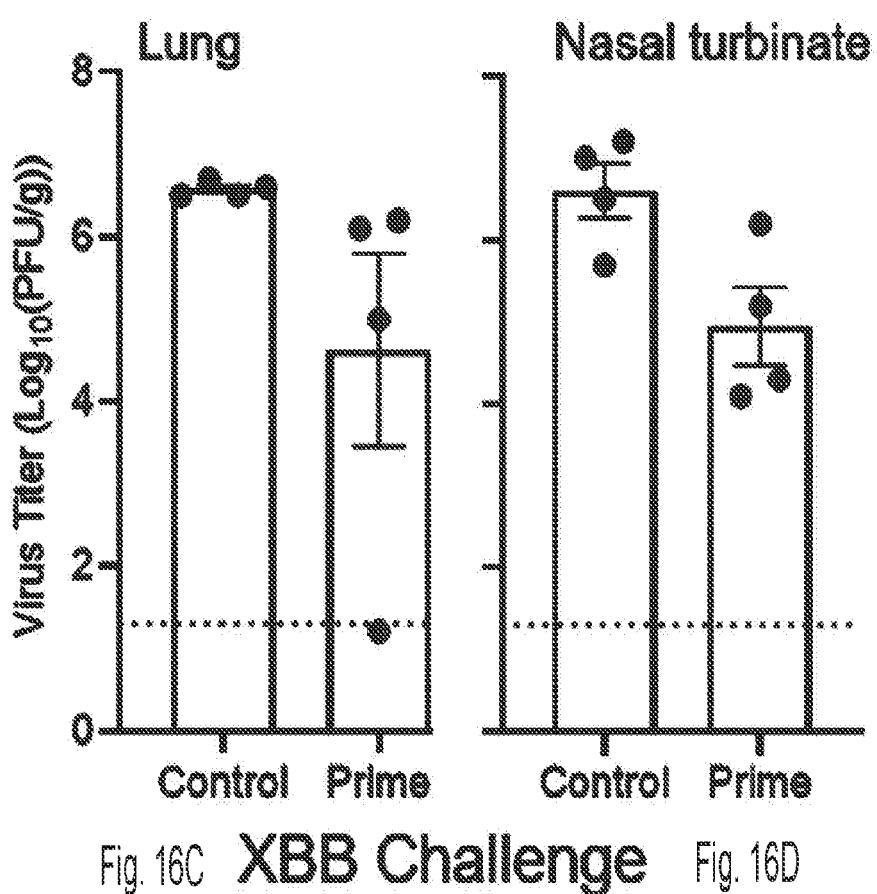
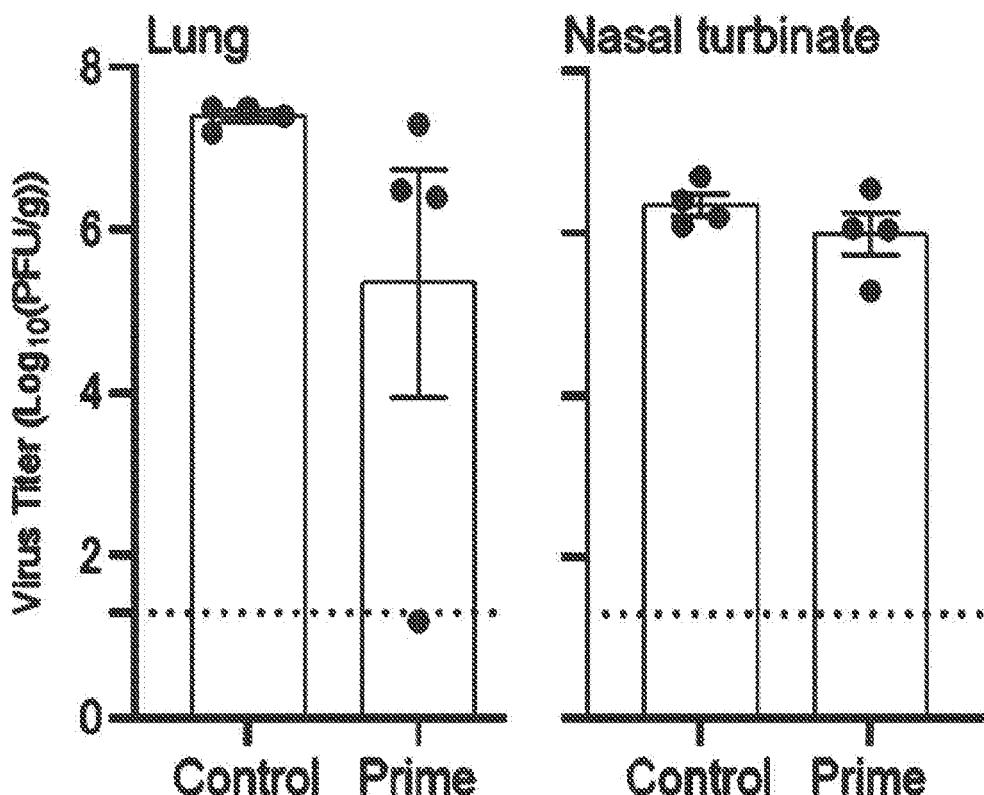
*Fig. 15*

CATGCCTATGTGAGGGCAAAGTTGATGAATGCCTATCCTTCTATATCAGTCCAATTGGA  
TGCCTCCCTGCTCATTGCTTGGTAGATGTGGGGTAGATTTGGACAAATCTGTACTCT  
TTGACAGTTCCCTTGGACAGAAACCAAACATAGATGTTACTGATGCAATGGTGGACCAG  
GCCTGGGATGCACAGAGAATATTCAAGGAGGCCGAGAAGTTCTTGATCTGTTGGTCTT  
CTTAATATGACTCAAGGATTCTGGAAAATTCCATGCTAACGGACCCAGGAAATGTTCAAG  
AAAGCAGTCTGCCATCCCACAGCTTGGACCTGGGAAGGGCGACTTCAGGATCCTTATG  
TGCACAAAGGTGACAATGGACGACTTCCTGACAGCTCATGAGATGGGCATATCCAG  
TATGATATGGCATATGCTGCACAACCTTCTGCTAAGAAATGGAGCTAATGAAGGATT  
CATGAAGCTGTTGGGAAATCATGTCACCTTCTGAGCCACACCTAACGATTTAAATCC  
ATTGGTCTTCTGTCACCGATTTCAAGAAGACAATGAAACAGAAATAACTTCCTGCTC  
AAACAAGCACTCACGATTGTTGGACTCTGCCATTTACTTACATGTTAGAGAAAGTGGAGG  
TGGATGGCTTAAAGGGAAATTCCAAAGACCAGTGGATGAAAAAGTGGTGGAGATG  
AAGCGAGAGATAGTTGGGTGGAACCTGTGCCCATGATGAAACATACTGTGACCC  
GCATCTCTGTTCCATGTTCTAATGATTACTCATTGATATTACACAAGGACCCCTT

*Fig. 15*

TACCAATTCCAGTTCAAGAAGCACTTGTCAAGCAGCTAACATGAAGGCCCTGCAC  
AAATGTGACATCTCAAACCTACAGAAGCTGGACAGAAACTGTTCAATATGCTGAGGCTT  
GGAAAATCAGAACCCCTGGACCCTAGCATTGGAAAATGTTGTAGGAGCAAAGAACATGAAT  
GTAAGGCCACTGCTCAACTACTTGAGCCCTTATTACCTGGCTGAAAGACCAGAACAAAG  
AATTCTTTGTGGGATGGAGTACCGACTGGAGTCCATATGCAGACCAAAGCATCAAAGTG  
AGGATAAGCCTAAAATCAGCTTGGAGATAAAGCATATGAATGGAACGACAATGAAATG  
TACCTGTTCCGATCATCTGTTGCATATGCTATGAGGCAGTACTTTAAAAGTAAAAAAT  
CAGATGATTCTTTGGGGAGGAGGATGTGCGAGTGGCTAATTGAAACCAAGAACATCTCC  
TTTAATTCTTGTCACTGCACCTAAAATGTGTCTGATATCATTCTAGAACTGAAGTT  
GAAAAGGCCATCAGGATGTCCCGGAGCCGTATCAATGATGCTTCCGTCTGAATGACAAC  
AGCCTAGAGTTCTGGGGATACAGCCAACACTTGGACCTCTAACCAAGCCCCCTGTTCC  
ATATGGCTGATTGTTTGGAGTTGTGATGGAGTGATAGTGGTGGCATTGTCATCCTG  
ATCTTCACTGGGATCAGAGATCGGAAGAAGAAAAATAAGCAAGAAGTGGAGAAAATCCT  
TATGCCTCCATCGATATTAGCAAAGGAGAAAATAATCCAGGATTCCAAAACACTGATGAT  
GTTCAGACCTCCTTGGTACCGAGACCTCCAGGTGGCGCCCGCTTAA  
(SEQ ID NO:17)

Fig. 15



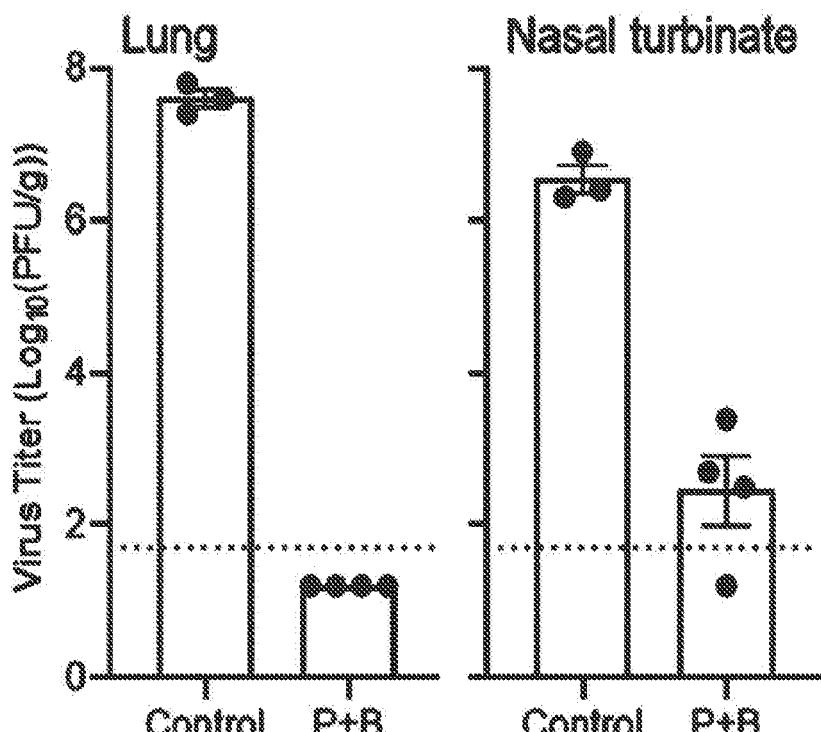
Fig. 17A **Delta Challenge**

Fig. 17B

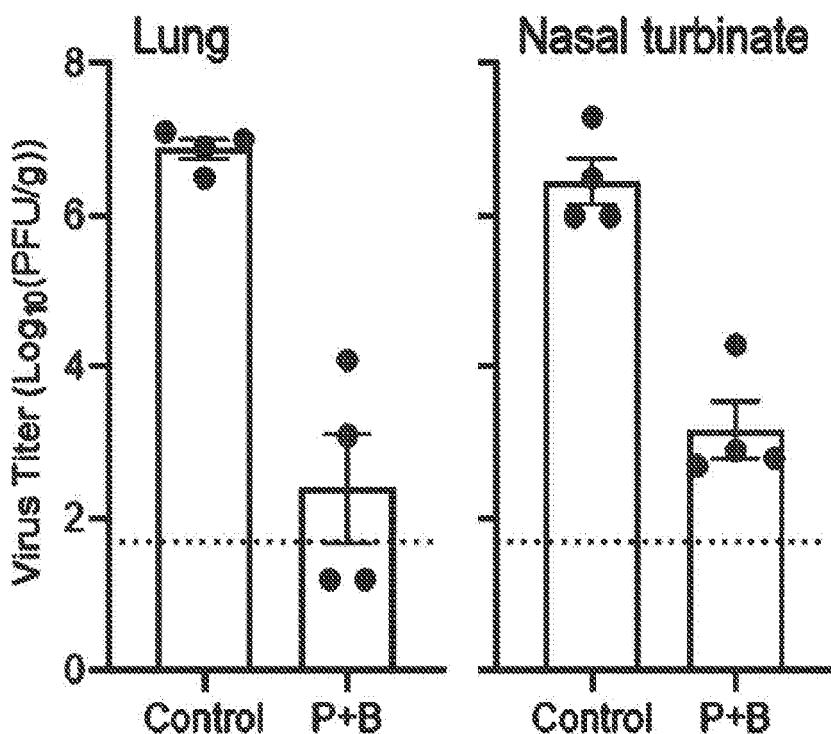
Fig. 17C **XBB Challenge**

Fig. 17D

*Fig. 18*

## SUBSTITUTE SHEET (RULE 26)

1 attaaagggtt tataccttcc caggttaacaa accaaccaac ttgcgtatctc ttgttagatct  
61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgtcact  
121 cacgcagtat aattaataac taattactgt cggtgacagg acacgagtaa ctgcgtctatc  
181 ttctgcaggc tgcttacggt ttgcgtccgtg ttgcgtccgtg tcatcagcac atctagggtt  
241 cgtccgggtg tgaccgaaag gtaagatgga gagcattgtc cctggttca acgagaaaaac  
301 acacgtccaa ctcagtttgc ctgtttaca gggtcgac gtgcgtcac gtggctttgg  
361 agactccgtg gaggaggct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg  
421 cttagtagaa gttgaaaaag gcgttttgc tcacttgaa cagccatgtg tgttcatcaa  
481 acgttcggat gtcgaactg cacccatgg tcatgttatg gttgagctgg tagcagaact  
541 cgaaggcatt cagtaacggc gtagtggta gacacttggt gtccttgcct ctcatgtgg  
601 cggaaatcca gtggcttacc gcaagggtct tcttcgttaa aacggtaata aaggagctgg  
661 tggccatagt tacggcccg atctaaagtct atttgactta ggcgacgagc ttggcactga  
721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttaccctgtga  
781 actcatgcgt gagottaacg gaggggcata cactcgctat gtcgataaca acttctgtgg  
841 ccctgatggc taccoctcttg agtgcattaa agaccttcta gcacgtgctg gttaaagcttc  
901 atgcactttg tccgaacaac tggactttat tgacacttaa aggggtgtat actgcgtccg  
961 tgaacatgag catgaaaatttgc ttgggtacac ggaacgttct gaaaagagct atgaattgca  
1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa  
1081 ttttgtatcc cccatcaaattt ccataatcaa gactattcaa ccaagggttg aaaagaaaaaa  
1141 gcttgatggc tttatggta gaattcgatc tqgttatcca gttgcgtcac caaatqaatg

SUBSTITUTE SHEET  
(RULE 26)

1201 caaccaaatg tgccttcaa ctctcatgaa gtgtgatcat tgtggtaaaa cttcatggca  
1261 gacggggat tttgttaaag ccacttgcga attttgtgge actgagaatt tgactaaaga  
1321 aggtgccact acttgtggtt acttacccca aaatgctgtt gttaaaattt attgtccagc  
1381 atgtcacaat tcagaagtag gacctgagca tagtcttgcc gaataccata atgaatctgg  
1441 cttgaaaaacc attcttcgta agggtggtcg cactattgcc tttggaggct gtgtgttctc  
1501 ttatgttggc tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg  
1561 ttgttaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga  
1621 aataactccaa aaagagaaaag tcaacatcaa tattgttggc gactttaaac ttaatgaaga  
1681 gatgcattt attttggcat cttttctgc ttccacaagt gctttgtgg aaactgtgaa  
1741 aggtttggat tataaagcat tcaaacaat ttgtaatcc tgtggtaatt ttaaagttac  
1801 aaaaggaaaa gctaaaaaaag gtgcctggaa tattggtaaa cagaaatcaa tactgagtc  
1861 tctttatgca tttgcattcag aggctgtcg tttgtacga tcaattttct cccgcactct  
1921 tgaaactgct caaaattctg tgcgtgtttt acagaaggcc gctataacaa tactagatgg  
1981 aatttcacag tattcactga gactcattga tgctatgtt ttcacatctg atttggctac  
2041 taacaatcta gttgtaatgg cctacattac aggtgggttt gttcagttga ctgcgcgtg  
2101 gctaactaac atctttggca ctgtttatgaa aaaactcaaa cccgtctttg attgtgttga  
2161 agagaagttt aaggaagggtg tagagtttct tagagacggt tggaaatttg ttaaatttat  
2221 ctcaacctgt gttgtgaaa ttgtcggtgg acaaattgtc acctgtgcaa agggaaattaa  
2281 ggagagtgtt cagacattct ttaagctgt aaataaaattt ttggcttgcgtgactc  
2341 tatcattatt ggtggagcta aacttaaaggc ttgtgttgcgtt ggtgaaacat ttgtcacgca

*Fig. 18*

# INTERNATIONAL SEARCH REPORT

International application No <b>PCT/US2023/027622</b>
--

**A. CLASSIFICATION OF SUBJECT MATTER**  
**INV. A61K39/12      A61P31/14**  
**ADD.**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

**A61K    A61P    C12N**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**EPO-Internal, Sequence Search**

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<b>X</b>	<p><b>NETLAND JASON ET AL:</b> "Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease",  <b>VIROLOGY</b>,  <b>vol. 399, no. 1,</b>  <b>27 January 2010 (2010-01-27), pages</b>  <b>120-128, XP085464129,</b>  <b>ISSN: 0042-6822, DOI:</b>  <b>10.1016/J.VIROL.2010.01.004</b></p> <p><b>abstract</b>  <b>page 121 – page 122; figures 2,3,5; table</b>  <b>2</b></p> <p>-----</p> <p style="text-align: center;">-/-</p>	<b>1,2,4-8,  17,22,  23,26,  30,31,  33,34,37</b>
<b>Y</b>		<b>10,  14-16,  21,  27-29,  32,35,  36,38-41</b>

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
---	--

**25 October 2023**

**07/11/2023**

Name and mailing address of the ISA/  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

**Renggli-Zulliger, N**

**INTERNATIONAL SEARCH REPORT**

International application No <b>PCT/US2023/027622</b>
--

**C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<b>ZHANG XIANWEN ET AL:</b> "A trans-complementation system for SARS-CoV-2 recapitulates authentic viral replication without virulence", <b>CELL</b> , ELSEVIER, AMSTERDAM NL, vol. 184, no. 8, 23 February 2021 (2021-02-23), page 2229, XP086538913, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2021.02.044 [retrieved on 2021-02-23] abstract; figures 1,3,4 page 2236, last paragraph	1,2,4-9, 11-13, 17,19, 20,22, 23,26
Y		10, 14-16, 21, 27-29, 32,35, 36,38-41
X	----- <b>LIU SHUFENG ET AL:</b> "Stable Cell Clones Harboring Self-Replicating SARS-CoV-2 RNAs for Drug Screen", <b>JOURNAL OF VIROLOGY</b> , vol. 96, no. 6, 23 March 2022 (2022-03-23), XP093027126, US ISSN: 0022-538X, DOI: 10.1128/jvi.02216-21 Retrieved from the Internet: URL: <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8941906/pdf/jvi.02216-21.pdf">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8941906/pdf/jvi.02216-21.pdf</a> >	1-3,5-8, 17,18, 22-26
Y	abstract; figure 1 -----	21,27-29