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

(19) World Intellectual Property Organization
International Bureau

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

WIPO | PCT

(10) International Publication Number
WO 2024/015510 A1

(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).

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))*

(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.

WO 2024/015510 A1

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

CROSS-REFERENCE TO RELATED APPLICATIONS

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

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

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

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 vaccines.

SUMMARY

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 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 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 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); 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 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 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 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 partial orf7b (1–17 amino acid residues); the complete deletion of the orf8 protein; nsp3 having 494K, 579V, 763M, 793S, and/or1456I; 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, 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 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 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 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 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 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 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 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, 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, 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.

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, 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%v 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%v nucleotide sequence identity to SEQ ID NO:16.

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 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 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.

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 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 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 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 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 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 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 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 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 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 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. 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 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 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. 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 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

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 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 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 (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; 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 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.

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).

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.

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.

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), 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

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 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.

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., 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.

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 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 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 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 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 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 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 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 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. 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 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 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 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 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, 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 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 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 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 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 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 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 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. 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, 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 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.

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 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 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 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 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 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 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 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 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.

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 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 "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, 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) 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 (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 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 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, 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 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 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 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 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 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 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 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.

"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 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 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, 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 (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 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 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 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. 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 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. 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 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 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, 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 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 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 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.

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 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 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 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 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 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 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 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 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, 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 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 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 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 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 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 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 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., 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 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, 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 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 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, 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, 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 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 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. 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 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 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 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 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 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 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 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. 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 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.

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 $TCID_{50}$, 10^3 to 10^4 pfu or $TCID_{50}$, 10^4 to 10^5 pfu or $TCID_{50}$, 10^5 to 10^7 pfu or $TCID_{50}$, 10^6 to 10^8 pfu or $TCID_{50}$, 10^6 to 10^{10} pfu or $TCID_{50}$, or 10^8 to 10^{12} pfu or $TCID_{50}$, or more, or from about 10^6 to 10^8 pfu or $TCID_{50}$, 10^8 to 10^{10} pfu or $TCID_{50}$, or 10^{10} to 10^{12} pfu or $TCID_{50}$.

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.

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 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 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 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, 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 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 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 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 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 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 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. 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 nonglycoslyated 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 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 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 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 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 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 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.

Nucleotide Position	WA1		TS-11		Protein
	Nucleotide	Amino Acid	Nucleotide	Amino Acid	
344	CTC	L	TTC	F	nsp1
548	ATT	I	CTT	L	
2393	GTC	V	ATC	I	
4200	ATG	M	AAG	K	nsp3
4455	GCC	A	GTC	V	
5007	ACG	T	ATG	M	
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	S
22,296	CAT	H	CGT	R	
23,594–23,629		TNSPRRARSVAS	36-nt-Del	12-aa-Del	
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	M
26,907	CTG	L	TTG	L	
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	***nsp4***	***nsp5***	***nsp6***	***nsp9***	***nsp10***	***nsp12***	***nsp13***	***nsp14***	***nsp16***	***S***	***E***	***M***	***N***
D-37	T6763C T7201A*	-	C10073T*	-	-	-	-	C16238G*	-	-	A21965C C23926G G24822C	-	-	-
D-ca	T6763C	C8881T A8882C*	C10233T	-	G12600A G12754T*	C13861A* C13126T	T13285A	-	-	T20743C	G21622T* A22943C C23926G A24467C T24958C	-	C26471T	G28930A
D-B4	T6763C	A8882C*	C10233T	-	G12754N	C13126T	T13285A	-	-	T20743C C21033T	C21622T* A22943C C23926G C24174A* A24467C T24958C	G26213A	C26471T	-
D-D2	C3498S T6763C C7542T	C8881A	G10486T	T11299C	G12754N	-	T13285A	-	A16817T	A20810C*	C21616T* G21621A* A22943C C23926G A24467C T24958C T25041G*	G26246A* G26264T	C26471T	-

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 variants in relation to viruses deposited in GenBank.`*

Table 3.

	ORF1a				ORF1ab					Structural proteins			
	nsp3	***nsp4***	***nsp5***	***nsp6***	***nsp9***	***nsp10***	***nsp12***	***nsp13***	***nsp14***	***nsp16***	***S***	***E***	***M***	***N***
D-37	I2181T N2378K*	-	-	-	-	-	-	D5456E*	--	-	K1495 P803R	-	-	-
D-ca	I2181T	H2792Y I2921H*	L3338F	-	S4145N G4178D*	A4247D* T4302I	N4355K	-	-	-	G35V* E475D P803R Q893H F1147F	-	T7I	G243E
D-B4	I2181T	I2921H*	L3338F	-	G4178?	T4302I	N4355K	-	-	T6838I	G35V* E475D P803R Q884K* Q893H F1147F	V14I	T7I	-
D-D2	A1092G I2181T T2274I	Q3724A	C3433F	I3893T	G4178?	-	N4355K	-	-	D6897A*	T33I* G35S* E475D P803R Q893H F1147F C1241W*	T30I	T7I	-

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

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 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 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 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 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_2.

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	TCCCAGGTAACAAACCAACCAACTTTCG (SEQ ID NO:18)
	F1_reverse	CTTGCGTGTGGAGGTTAATGTTGTCTACTG (SEQ ID NO:19)
2	F2_forward	CATTAACCTCCACACGCAAGTTGTGGACATG (SEQ ID NO:20)
	F2_reverse	GTCTGTCCTGGTTGAATGCGAACAAACTTATAC (SEQ ID NO:21)
3	F3_forward	CGCATTCAACCAGGACAGACTTTTTCAGTG (SEQ ID NO:22)
	F3_reverse	GCCACACATGACCATTTCACTCAATACTTGAG (SEQ ID NO:23)
4	F4_forward	GTGAAATGGTCATGTGTGGCGGTTCACTATATG (SEQ ID NO:24)
	F4_reverse	CCTGGTGCAACTCCTTTATCAGAACCAG (SEQ ID NO:25)
5	F5_forward	GATAAAGGAGTTGCACCAGGTACAGCTGTTTTAAG (SEQ ID NO:26)
	F5_reverse	GTCGTCGTCGGTTCATCATAAATTGGTTCC (SEQ ID NO:27)
6	F6_forward	TATGATGAACCGACGACGACTACTAGCG (SEQ ID NO:28)
	F6_reverse	GTCATTCTCCTAAGAAGCTATTAAAATCACATGGGG (SEQ ID NO:29)
Linker	Linker_forward	CCATGTGATTTTAATAGCTTCTTAGGAGAATG (SEQ ID NO:30)
	Linker_reverse	CAAGAGATCGAAAGTTGGTTGGTTTGTTACCTGGG (SEQ ID NO:31)

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-µl 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 µl of a 50-µl reaction volume) was directly transfected into HEK293T stable cells (E or E/M cells) seeded in a 6-well plate (8.0×10^5 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 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 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 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. 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).

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 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 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

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 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 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 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).

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 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 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 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 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 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 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 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 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 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.

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 $1x10^6$ pfu/mL are prepared. The 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

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 $1x10^6$ 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. 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.

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. 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)

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.

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 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

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 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

(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.

(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 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)

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, 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)

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.

T-cell IFN-γ 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

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 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 vaccinated once by intranasal inoculation of 5 x 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 the lung and nasal turbinate 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 turbinate tissue of the same animals 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 turbinate 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 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 regiment 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 1,000-fold in the nasal turbinate 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 (Figure 17C). In the nasal turbinate 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

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).

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) (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). (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).

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)

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. 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). (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).

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).

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). (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).

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).
(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).

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).

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 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:

1. An isolated nucleic acid comprising a recombinant coronavirus genome having a genetic modification that inhibits or prevents expression of coronavirus envelope (E) protein.

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.

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.

4. The isolated nucleic acid of claim 1, 2 or 3 which comprises DNA.

5. The isolated nucleic acid of claim 1, 2 or 3 which comprises RNA.

6. An isolated cell comprising the isolated nucleic acid of any one of claims 1 to 5.

7. The isolated cell of claim 6 which is a mammalian cell.

8. The isolated cell of claim 7 which is a non-human primate cell.

9. The isolated cell of any one of claims 6 to 8 that stably expresses coronavirus E protein.

10. The isolated cell of any one of claims 6 to 8 that stably expresses hACE2 and optionally M protein.

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.

14. The isolated cell of any one of claims 11 to 13 that stably expresses hACE2.

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.

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:
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.

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.

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 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.

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 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.

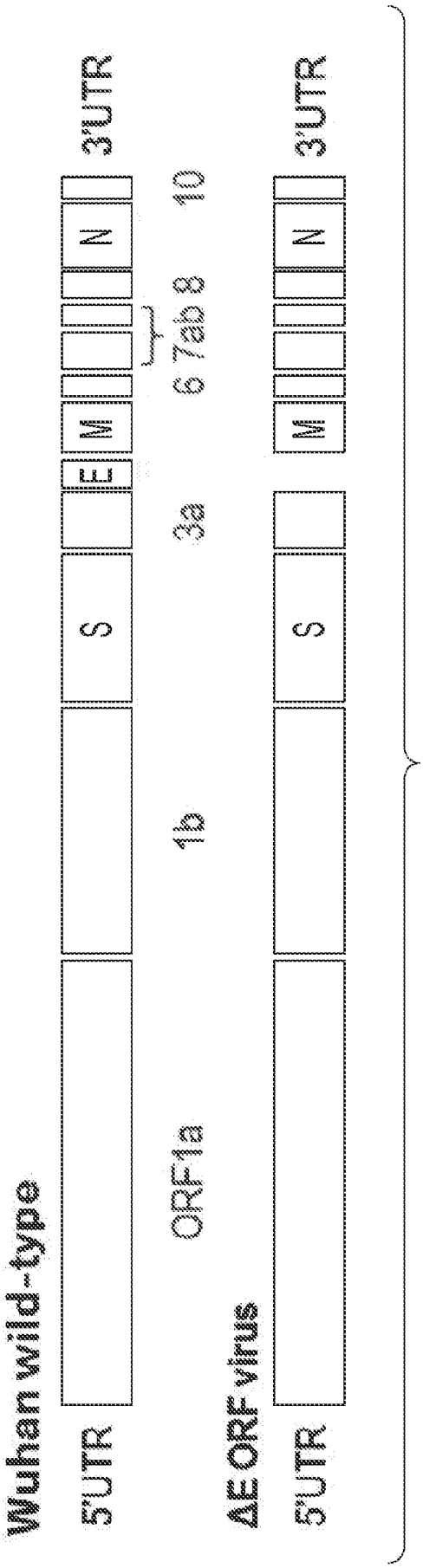

Wuhan wild-type
5'UTR
ORF1a
1b
S
3a
E
M
6
7ab
8
N
10
3'UTR
ΔE ORF virus
5'UTR
S
M
N
3'UTR

Fig. 1A

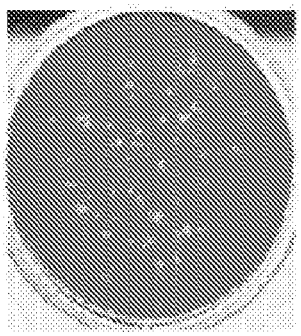


Fig. 1B

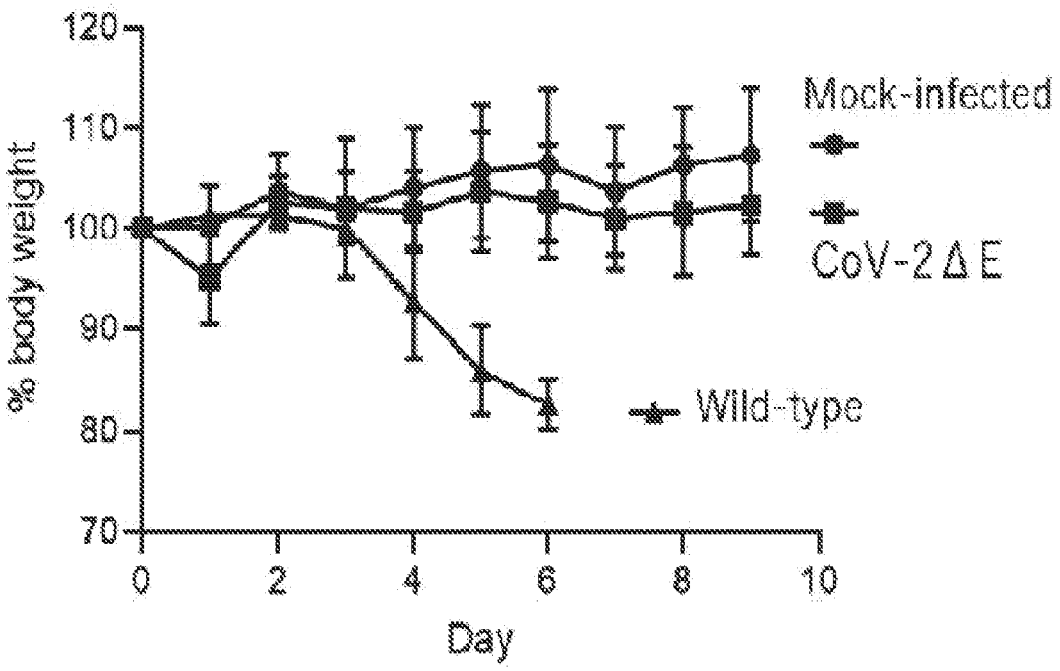

120
110
100
90
80
70
% body weight
0
2
4
6
8
10
Day
Mock-infected
CoV-2 Δ E
Wild-type

Fig. 2A

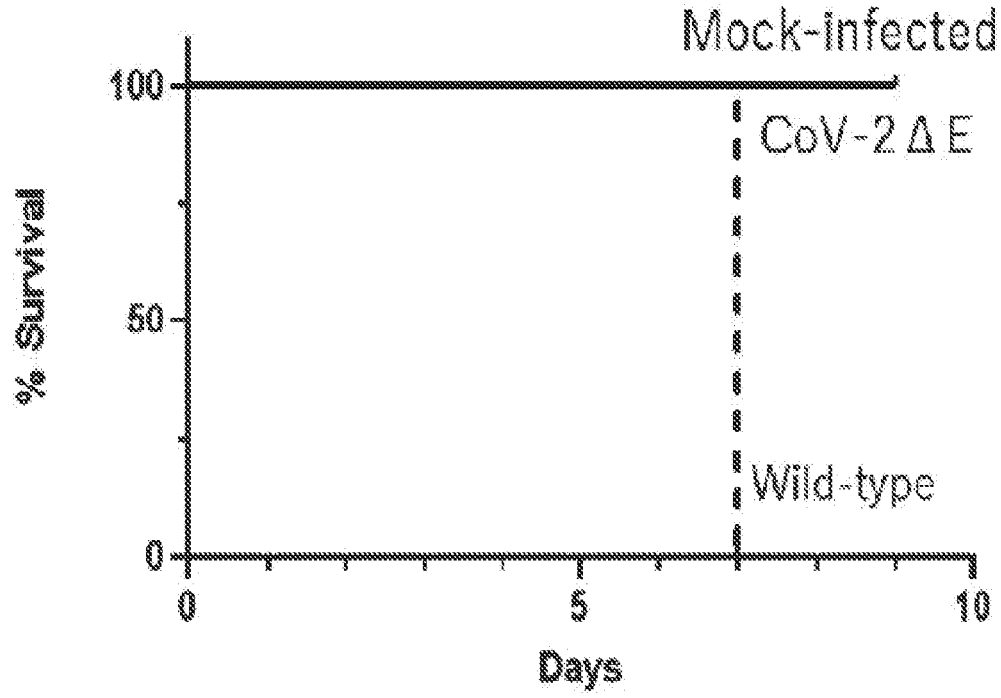

Mock-infected
100
50
0
% Survival
CoV-2 Δ E
Wild-type
0
5
10
Days

Fig. 2B

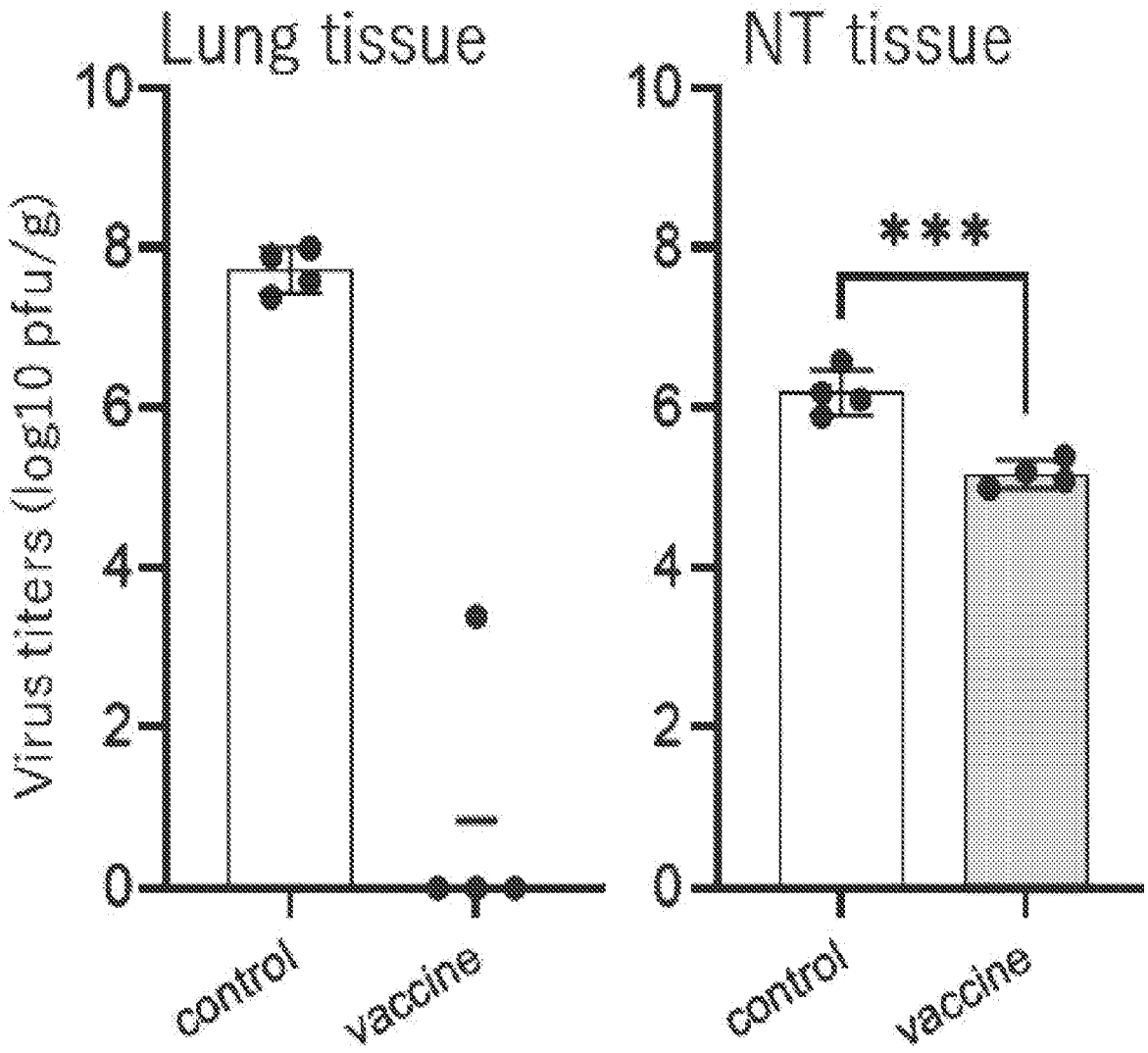

Lung tissue
NT tissue
Virus titers (log10 pfu/g)
10
8
6
4
2
0
***
control
vaccine
control
vaccine

Fig. 3

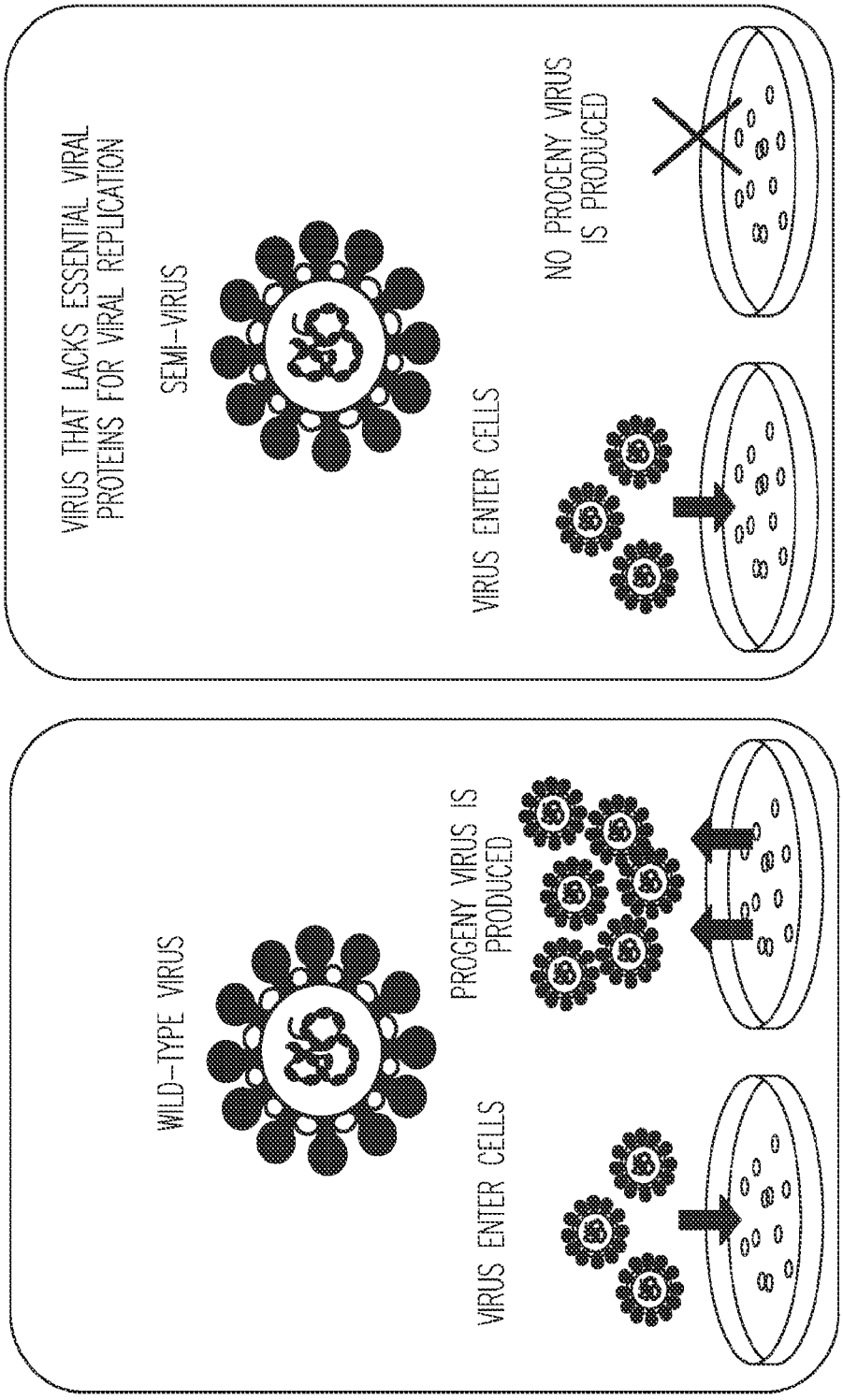

WILD-TYPE VIRUS
VIRUS ENTER CELLS
PROGENY VIRUS IS PRODUCED
VIRUS THAT LACKS ESSENTIAL VIRAL PROTEINS FOR VIRAL REPLICATION
SEMI-VIRUS
VIRUS ENTER CELLS
NO PROGENY VIRUS IS PRODUCED

Fig. 4

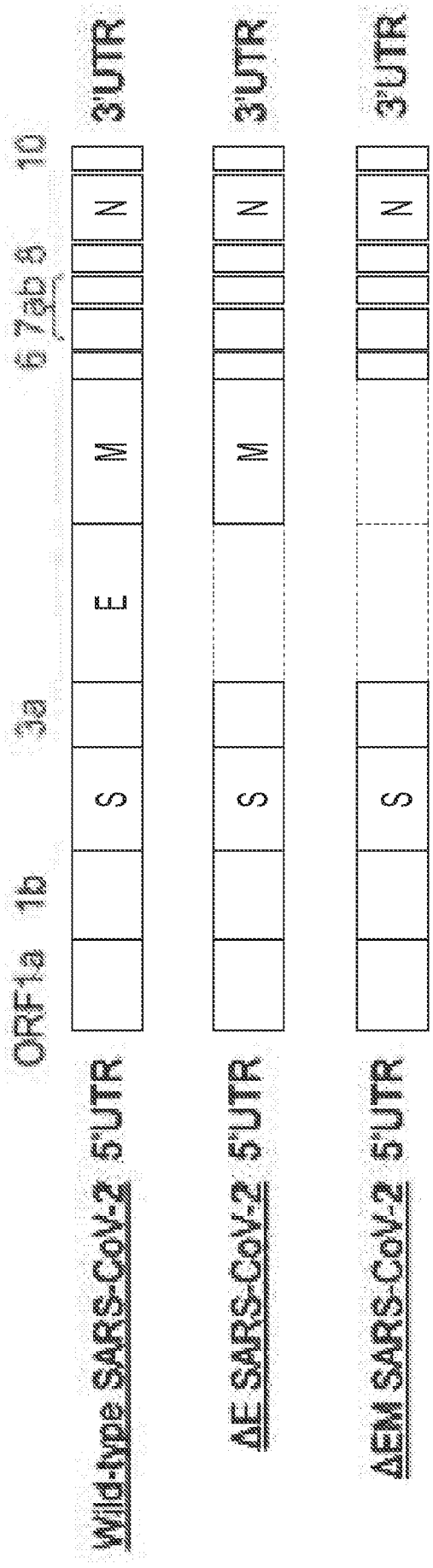

ORF1a
1b
S
3a
E
M
6 7ab 8
10
5'UTR
3'UTR
Wild-type SARS-CoV-2
ΔE SARS-CoV-2
ΔEM SARS-CoV-2
N

Fig. 5

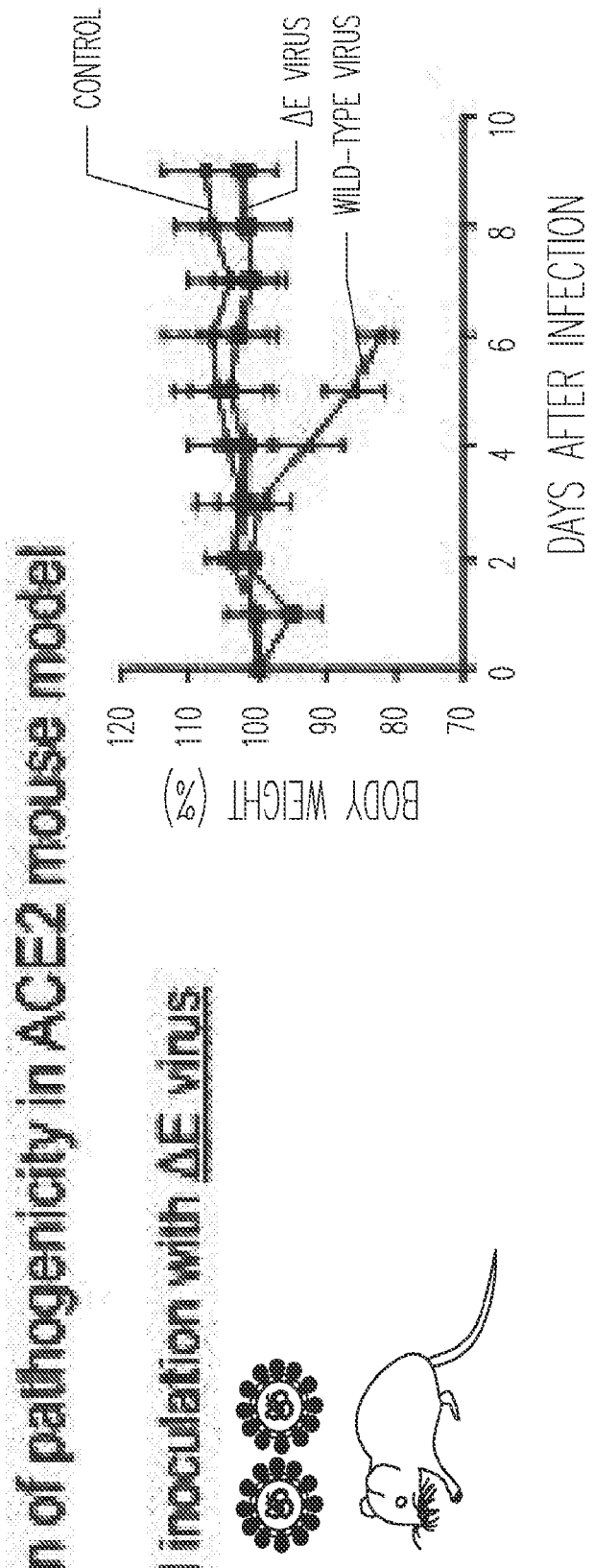

Evaluation of pathogenicity in ACE2 mouse model
Intranasal inoculation with ΔE virus
BODY WEIGHT (%)
120
110
100
90
80
70
0
2
4
6
8
10
DAYS AFTER INFECTION
CONTROL
ΔE VIRUS
WILD-TYPE VIRUS

Fig. 6

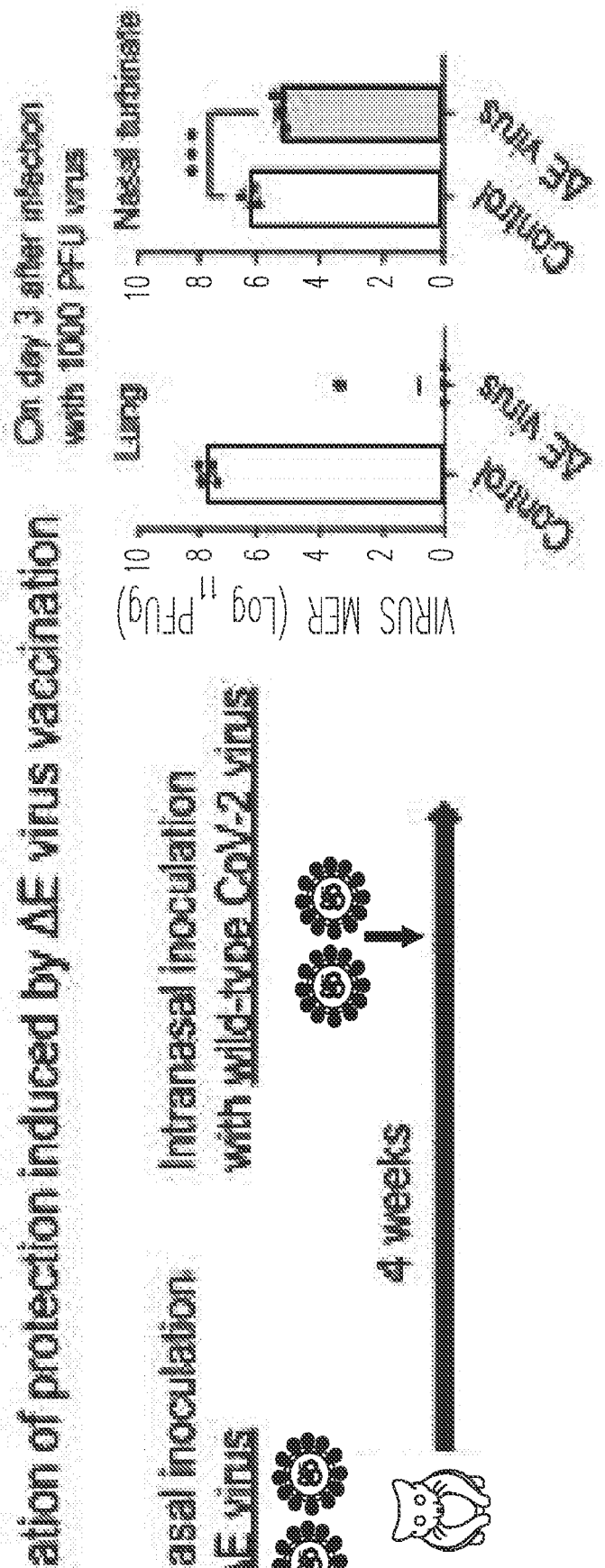


Fig. 6 (continued)

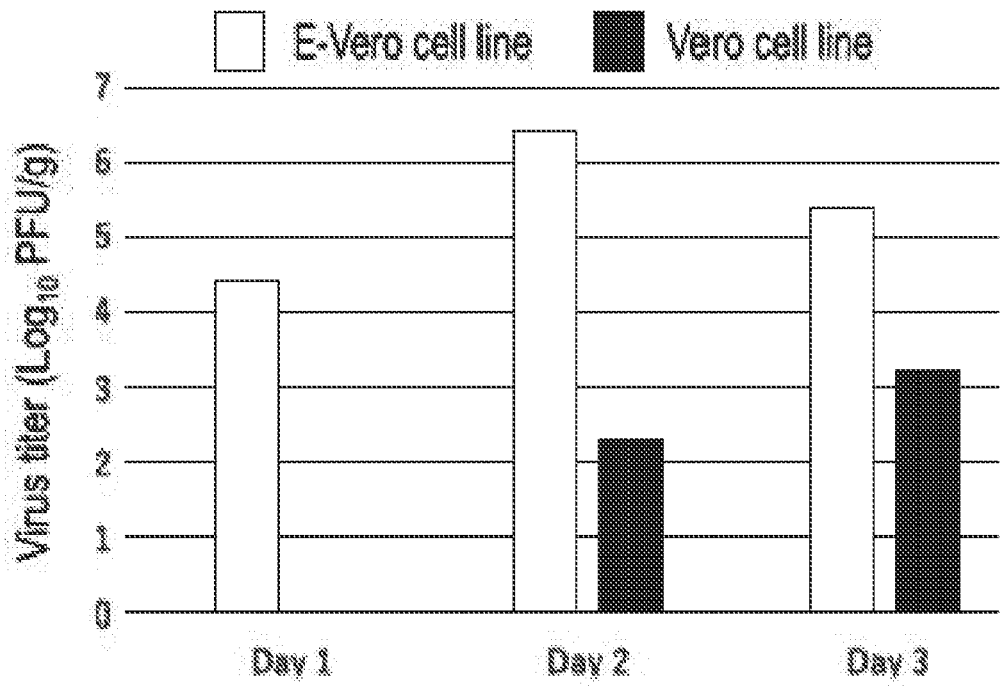

E-Vero cell line
Vero cell line
Virus titer (Log_{10} PFU/g)
7
6
5
4
3
2
1
0
Day 1
Day 2
Day 3

Fig. 7

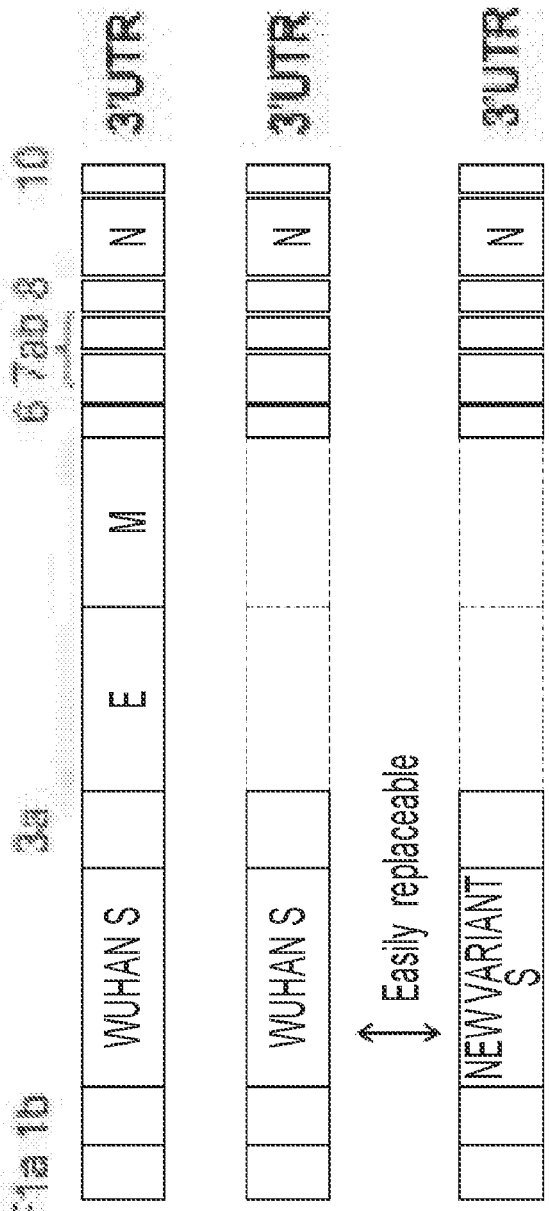

ORF1a 1b
3a
E
M
6 7ab 8
10
5'UTR
WUHAN S
3'UTR
N
Wuhan SARS-CoV-2
Wuhan ΔEM SARS-CoV-2
Easily replaceable
NEW VARIANT S
ΔEM SARS-CoV-2 with spike from new variants

Fig. 8

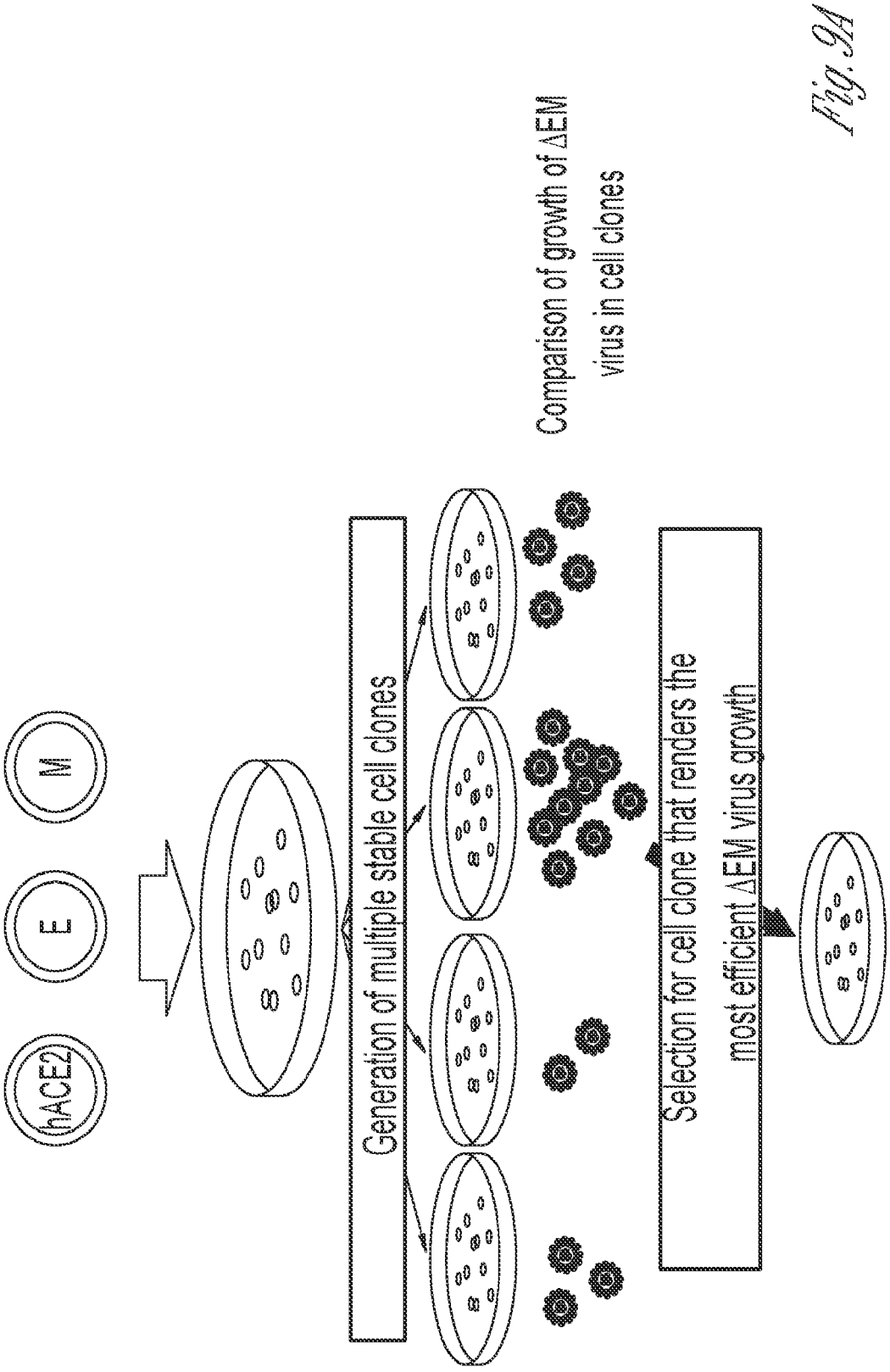

hACE2
E
M
Generation of multiple stable cell clones
Comparison of growth of ΔEM virus in cell clones
Selection for cell clone that renders the most efficient ΔEM virus growth

Fig. 9A

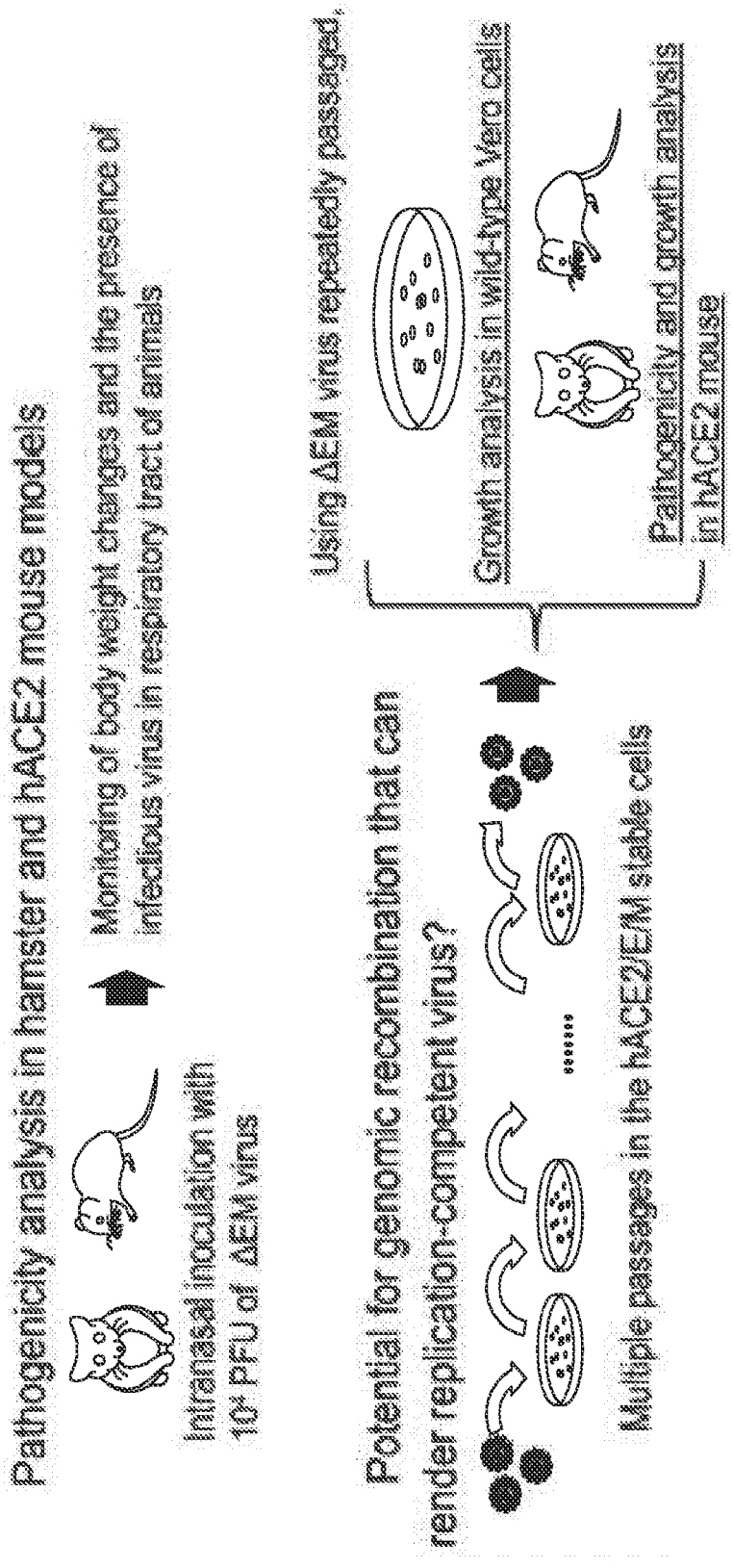
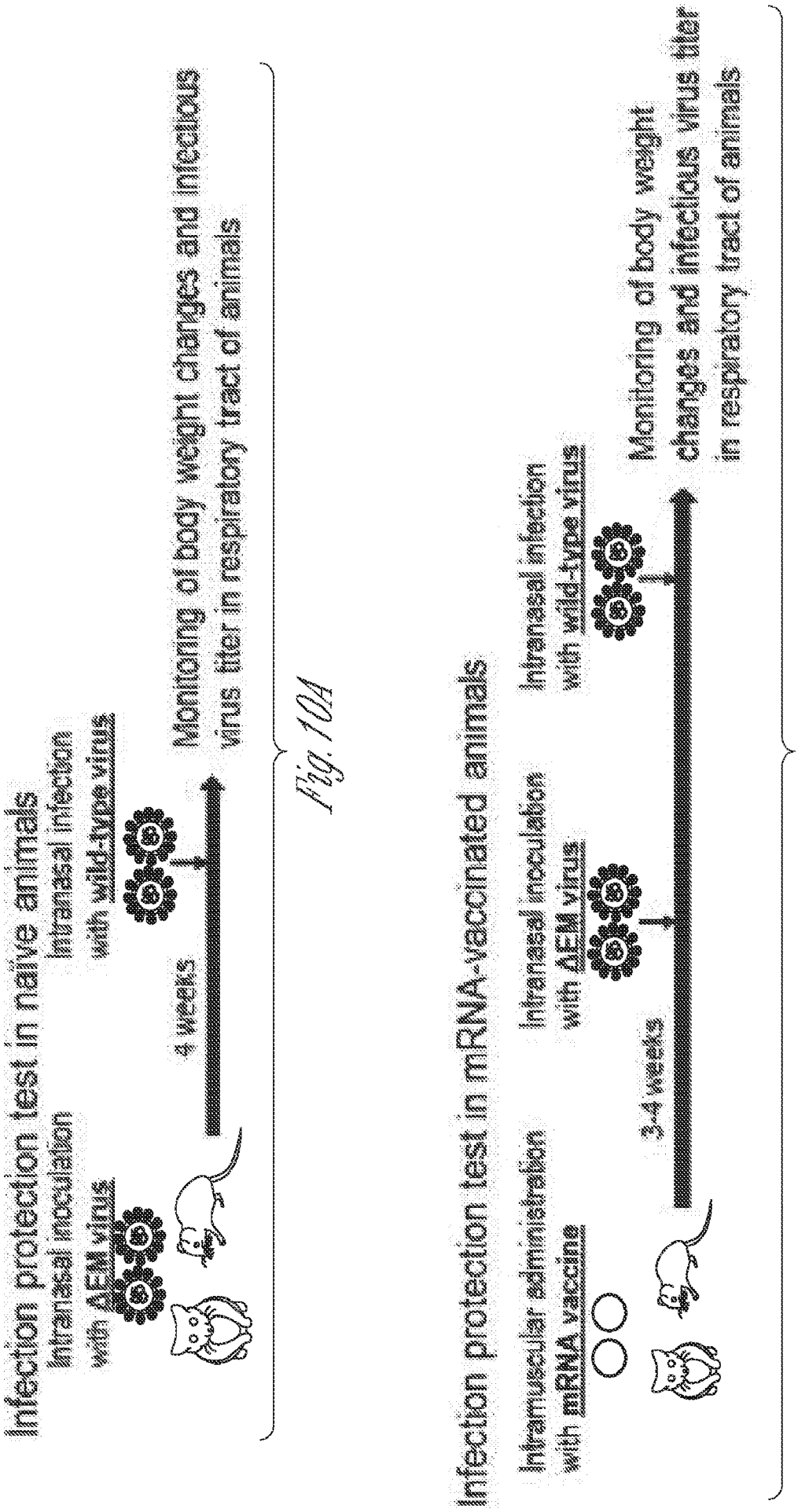

Pathogenicity analysis in hamster and hACE2 mouse models
Intranasal inoculation with 10⁴ PFU of ΔEM virus
Monitoring of body weight changes and the presence of infectious virus in respiratory tract of animals
Potential for genomic recombination that can render replication-competent virus?
Multiple passages in the hACE2/EM stable cells
Using ΔEM virus repeatedly passaged,
Growth analysis in wild-type Vero cells
Pathogenicity and growth analysis in hACE2 mouse

Fig. 9B


Infection protection test in naive animals
Intranasal inoculation with ΔEM virus
Intranasal infection with wild-type virus
4 weeks
Monitoring of body weight changes and infectious virus titer in respiratory tract of animals
Fig. 10A
Infection protection test in mRNA-vaccinated animals
Intramuscular administration with mRNA vaccine
Intranasal inoculation with ΔEM virus
Intranasal infection with wild-type virus
3-4 weeks
Monitoring of body weight changes and infectious virus titer in respiratory tract of animals
Fig. 10B

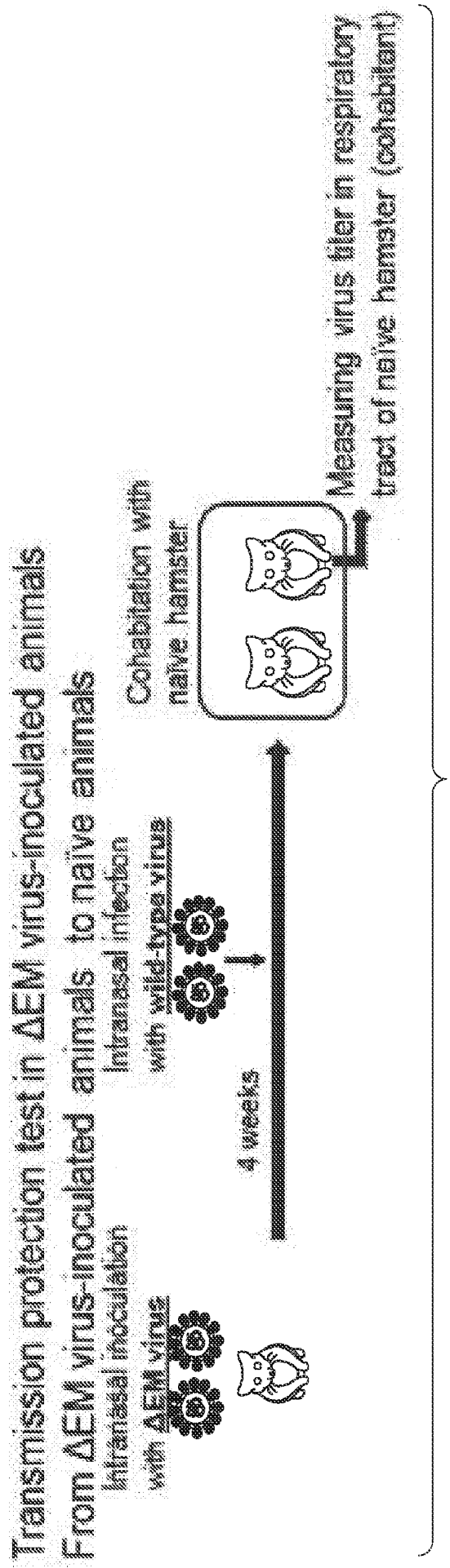

Transmission protection test in ΔEM virus-inoculated animals
From ΔEM virus-inoculated animals to naive animals
Intranasal inoculation
with ΔEM virus
Intranasal infection
with wild-type virus
4 weeks
Cohabitation with
naive hamster
Measuring virus titer in respiratory
tract of naive hamster (cohabitant)

Fig. 10C

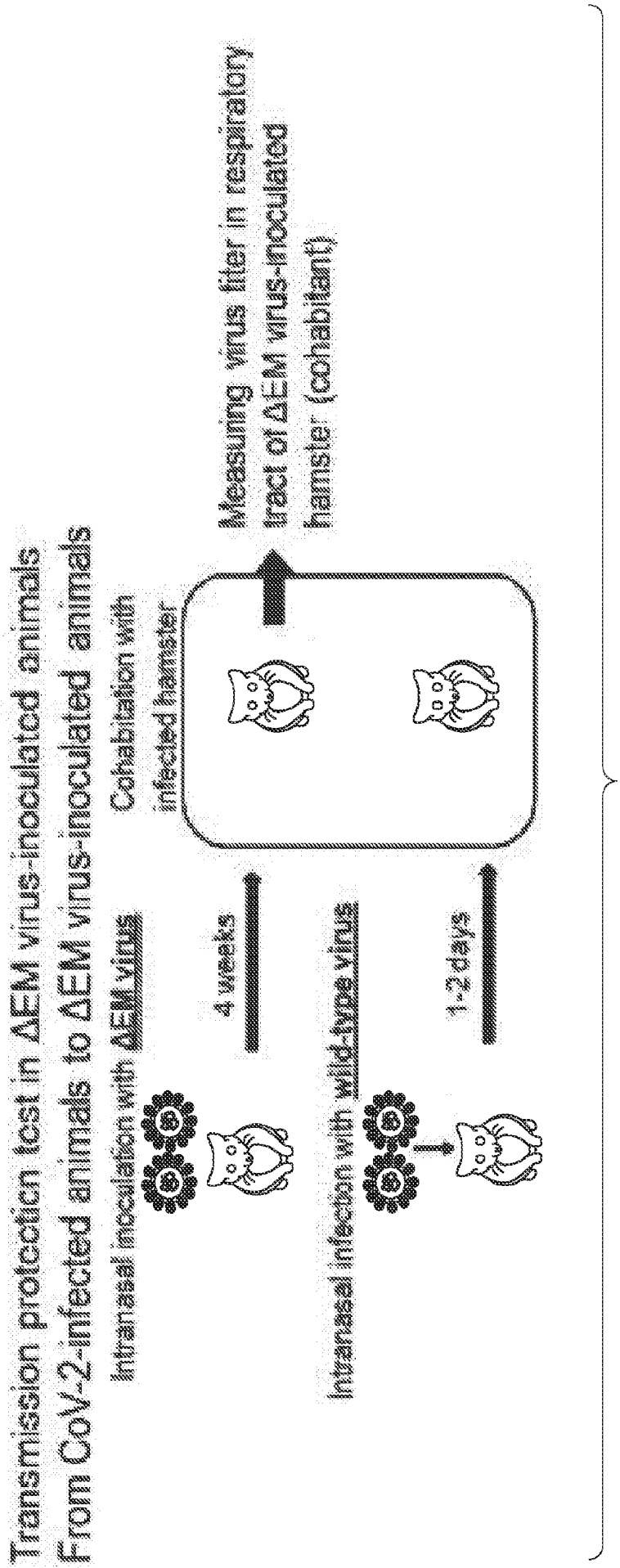

Transmission protection test in ΔEM virus-inoculated animals
From CoV-2-infected animals to ΔEM virus-inoculated animals
Intranasal inoculation with ΔEM virus
4 weeks
Cohabitation with infected hamster
Measuring virus titer in respiratory tract of ΔEM virus-inoculated hamster (cohabitant)
Intranasal infection with wild-type virus
1-2 days

Fig. 10D

	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="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
```

Fig. 12A continued

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

Fig. 12A continued

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKSGGTTEMLAKALRKV

Fig. 12A continued

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNAVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

Fig. 12A continued

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWXX

XXXXXXXXXXVLKSEDAQGMDNLACEDLKLVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSISCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

Fig. 12A continued

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPLPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

Fig. 12A continued

AYYFMRFRRAFGEYSHVVAFNILLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

AKALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

Fig. 12A continued

LPSLAAVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

Fig. 12A continued

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD

EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR

LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA

NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV

DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ

TYHPNCVNCLDDRCILHCANFNVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRE

LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ

Fig. 12A continued

TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI

RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ

DALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT

RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR

KHTTCCSLSHRFYRLANECAQVLSEMVMCGSSLYVKPGGTSSGDATTAYANSVFNICQ

AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM

ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS

QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH

PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV

LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD

Fig. 12A continued

VTDVTQLYLGGMSYYCKSHKLPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW

TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP

RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH

TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK

SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF

KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ

LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK

AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA

VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD

RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT

Fig. 12A continued

EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG

CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP

LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL

CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA

HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK

FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL

FWNCNVDRYPVNSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ

LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN

MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT

VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY

Fig. 12A continued

KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG

SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK

PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP

FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV

TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD

SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP

TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG

FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL

IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE

GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat_peptide 239..778

Fig. 12A continued

```
                /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"
                /product="nsp7"
mat_peptide     12065..12658
```

Fig. 12A continued

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

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
```

Fig. 12A continued

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

Fig. 12A continued

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKSGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

Fig. 12A continued

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNAVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWXX

XXXXXXXXXXVLKSEDAQGMDNLACEDLKLVSEEVVENPTIQKDVLECNVKTTEVVGD

Fig. 12A continued

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSISCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

Fig. 12A continued

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPLPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

Fig. 12A continued

AKALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

LPSLAAVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

Fig. 12A continued

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVAIVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

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

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

SSVLHSTQDLFLPFFSNVTWFHAIHVSGINGTKRFDNPVLPFNDGVYFASXEKSNIIR

GWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDVYYHKNNKSWMESXXXVY

SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ

GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFL

LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN

Fig. 12A continued

LCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF

TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN

YRYRLFRKSNLKPFERDISTEIYQAGSKPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY

RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG

RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAI

HADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSRR

RARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTM

YICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG

GFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFN

GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQN

Fig. 12A continued

VLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQNVVNQNAQALNTLVKQLSSNFGA

ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS

ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAH

FPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD

SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELG

KYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSE

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="MDLFMRIFTIGTVTLKQGEIKDATPLDFVRATATIPIQASLPFG

WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE

APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY

NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ

LSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"

gene 26218..26445
/gene="E"

CDS 26218..26445
/gene="E"
/codon_start=1
/product="envelope protein"
/protein_id="UAL04649.1"

/translation="MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCC
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="MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNR

FLYIIKLIFLWLLWPVTLACFVLAAVYRINWITGGIAXAMACLVGLMWLSYFIASFRL

FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD

IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIA
                     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="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSL
                    TENKYSQLDEEQPMEID"
     gene           27367..27732
                    /gene="ORF7a"
     CDS            27367..27732
                    /gene="ORF7a"
                    /codon_start=1
                    /product="ORF7a protein"
                    /protein_id="UAL04652.1"

/translation="MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNS

PFHPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSASPKLFIRQEEVQELYSPIFL
                    IVAAIVFITLCFTLKRKTE"
     gene           27729..27860
                    /gene="ORF7b"
     CDS            27729..27860
```

Fig. 12A continued

```
                    /gene="ORF7b"
                    /codon_start=1
                    /product="ORF7b"
                    /protein_id="UAL04653.1"

/translation="MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNEICHA"
     gene           27867..28232
                    /gene="ORF8"
     CDS            27867..28232
                    /gene="ORF8"
                    /codon_start=1
                    /product="ORF8 protein"
                    /protein_id="UAL04654.1"

/translation="MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK
WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRC
                    SFYEDFLEYHDVRVVLDLI"
     gene           28247..29506
                    /gene="N"
     CDS            28247..29506
```

Fig. 12A continued

```
                     /gene="N"
                     /codon_start=1
                     /product="nucleocapsid phosphoprotein"
                     /protein_id="UAL04655.1"
/translation="MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQG

LPNNTASWFTALTQHGKEGLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMK

DLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQ

LPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSMGTSPARMAGNGCDAA

LALLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGR

RGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYT

GAIKLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKAYETQALPQRQKKQQTV
                     TLLPAADLDDFSKQLQQSMSSADSTQA"
     gene            29531..29647
                     /gene="ORF10"
```

Fig. 12A continued

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

Fig. 12A continued

```
 301 acaggttcgc gacgtgctcg tacgtggctt tggagactcc gtggaggagg tcttatcaga
 361 ggcacgtcaa catcttaaag atggcacttg tggcttagta gaagttgaaa aaggcgtttt
 421 gcctcaactt gaacagccct atgtgttcat caaacgttcg gatgctcgaa ctgcacctca
 481 tggtcatgtt atggttgagc tggtagcaga actcgaaggc attcagtacg gtcgtagtgg
 541 tgagacactt ggtgtccttg tccctcatgt gggcgaaata ccagtggctt accgcaaggt
 601 tcttcttcgt aagaacggta ataaaggagc tggtggccat agttacggcg ccgatctaaa
 661 gtcatttgac ttaggcgacg agcttggcac tgatccttat gaagattttc 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 gcagacacct tttgaaatta aattggcaaa
1021 gaaatttgac accttcaatg gggaatgtcc aaattttgta tttcccttaa attccataat
1081 caagactatt caaccaaggg ttgaaaagaa aaagcttgat ggctttatgg gtagaattcg
1141 atctgtctat ccagttgcgt caccaaatga atgcaaccaa atgtgccttt caactctcat
1201 gaagtgtgat cattgtggtg aaacttcatg gcagacgggc gattttgtta aagccacttg
1261 cgaattttgt ggcactgaga atttgactaa agaaggtgcc actacttgtg gttacttacc
1321 ccaaaatgct gttgttaaaa tttattgtcc agcatgtcac aattcagaag taggacctga
1381 gcatagtctt gccgaatacc ataatgaatc tggcttgaaa accattcttc gtaagggtgg
1441 tcgcactatt gcctttggag gctgtgtgtt ctcttatgtt ggttgccata acaagtgtgc
```

Fig. 12A continued

```
1501 ctattgggtt ccacgtgcta gcgctaacat aggttgtaac catacaggtg ttgttggaga
1561 aggttccgaa ggtcttaatg acaaccttct tgaaatactc caaaaagaga aagtcaacat
1621 caatattgtt ggtgacttta aacttaatga agagatcgcc attattttgg catcttttc
1681 tgcttccaca agtgcttttg tggaaactgt gaaaggtttg gattataaag cattcaaaca
1741 aattgttgaa tcctgtggta attttaaagt tacaaaagga aaagctaaaa aaggtgcctg
1801 gaatattggt gaacagaaat caatactgag tcctctttat gcatttgcat cagaggctgc
1861 tcgtgttgta cgatcaattt tctcccgcac tcttgaaact gctcaaaatt ctgtgcgtgt
1921 tttacagaag gccgctataa caatactaga tggaatttca cagtattcac tgagactcat
1981 tgatgctatg atgttcacat ctgatttggc tactaacaat ctagttgtaa tggcctacat
2041 tacaggtggt gttgttcagt tgacttcgca gtggctaact aacatctttg gcactgttta
2101 tgaaaaactc aaacccgtcc ttgattggct tgaagagaag tttaaggaag gtgtagagtt
2161 tcttagagac ggttgggaaa ttgttaaatt tatctcaacc tgtgcttgtg aaattgtcgg
2221 tggacaaatt gtcacctgtg caaaggaaat taaggagagt gttcagacat tctttaagct
2281 tgtaaataaa ttttttggctt tgtgtgctga ctctatcatt attggtggag ctaaacttaa
2341 agccttgaat ttaggtgaaa catttgtcac gcactcaaag ggattgtaca gaaagtgtgt
2401 taaatccaga gaagaaactg gcctactcat gcctctaaaa gccccaaaag aaattatctt
2461 cttagaggga gaaacacttc ccacagaagt gttaacagag gaagttgtct tgaaaactgg
2521 tgatttacaa ccattagaac aacctactag tgaagctgtt gaagctccat tggttggtac
2581 accagtttgt attaacgggc ttatgttgct cgaaatcaaa gacacagaaa agtactgtgc
2641 ccttgcacct aatatgatgg taacaaacaa taccttcaca ctcaaaggcg gtgcaccaac
2701 aaaggttact tttggtgatg acactgtgat agaagtgcaa ggttacaaga gtgtgaata
```

Fig. 12A continued

```
2761 cacttttgaa cttgatgaaa ggattgataa agtacttaat gagaagtgct ctgcctatac
2821 agttgaactc ggtacagaag taaatgagtt cgcctgtgtt gtggcagatg ctgtcataaa
2881 aactttgcaa ccagtatctg aattacttac accactgggc attgatttag atgagtggag
2941 tatggctaca tactacttat ttgatgagtc tggtgagttt aaattggctt cacatatgta
3001 ttgttctttt taccctccag atgaggatga agaagaaggt gattgtgaag aagaagagtt
3061 tgagccatca actcaatatg agtatggtac tgaagatgat taccaaggta aacctttgga
3121 atttggtgcc acttctgctg ctcttcaacc tgaagaagag caagaagaag attggttaga
3181 tgatgatagt caacaaactg ttggtcaaca agacggcagt gaggacaatc agacaactac
3241 tattcaaaca attgttgagg ttcaacctca attagagatg gaacttacac cagttgttca
3301 gactattgaa gtgaatagtt ttagtggtta tttaaaactt actgacaatg tatacattaa
3361 aaatgcagac attgtggaag aagctaaaaa ggtaaaacca acagtggttg ttaatgcagc
3421 caatgtttac cttaaacatg gaggaggtgt tgcaggagcc ttaaataagg ctactaacaa
3481 tgccatgcaa gttgaatctg atgattacat agctactaat ggaccactta aagtgggtgg
3541 tagttgtgtt ttaagcggac acaatcttgc taaacactgt cttcatgttg tcggcccaaa
3601 tgttaacaaa ggtgaagaca ttcaacttct taagagtgct tatgaaaatt ttaatcagca
3661 cgaagttcta cttgcaccat tattatcagc tggtattttt ggtgctgacc ctatacattc
3721 tttaagagtt tgtgtagata ctgttcgcac aaatgtctac ttagctgtct ttgataaaaa
3781 tctctatgac aaacttgttt caagcttttt ggaaatgaag agtgaaaagc aagttgaaca
3841 aaagatcgct gagattccta aagaggaagt taagccattt ataactgaaa gtaaaccttc
3901 agttgaacag agaaaacaag atgataagaa aatcaaagct tgtgttgaag aagttacaac
```

Fig. 12A continued

```
3961 aactctggaa gaaactaagt tcctcacaga aaacttgtta ctttatattg acattaatgg
4021 caatcttcat ccagattctg ccactcttgt tagtgacatt gacatcactt tcttaaagaa
4081 agatgctcca tatatagtgg gtgatgttgt tcaagagggt gttttaactg ctgtggttat
4141 acctactaaa aagtctggtg gcactactga aatgctagcg aaagctttga gaaaagtgcc
4201 aacagacaat tatataacca cttacccggg tcagggttta aatggttaca ctgtagagga
4261 ggcaaagaca gtgcttaaaa agtgtaaaag tgccttttac attctaccat ctattatctc
4321 taatgagaag caagaaattc ttggaactgt ttcttggaat ttgcgagaaa tgcttgcaca
4381 tgcagaagaa acacgcaaat taatgcctgt ctgtgtggaa actaaagcca tagtttcaac
4441 tatacagcgt aaatataagg gtattaaaat acaagagggt gtggttgatt atggtgctag
4501 attttacttt tacaccagta aaacaactgt agcgtcactt atcaacacac ttaacgatct
4561 aaatgaaact cttgttacaa tgccacttgg ctatgtaaca catggcttaa atttggaaga
4621 agctgctcgg tatatgagat ctctcaaagt gccagctaca gtttctgttt cttcacctga
4681 tgctgttaca gcgtataatg gttatcttac ttcttcttct aaaacacctg aagaacattt
4741 tattgaaacc atctcacttg ctggttccta taaagattgg tcctattctg gacaatctac
4801 acaactaggt atagaatttc ttaagagagg tgataaaagt gtatattaca ctagtaatcc
4861 taccacattc cacctagatg gtgaagttat caccttttgac aatcttaaga cacttctttc
4921 tttgagagaa gtgaggacta ttaaggtgtt tacaacagta gacaacatta acctccacac
4981 gcaagttgtg gacatgtcaa tgacatatgg acaacagttt ggtccaactt atttggatgg
5041 agctgatgtt actaaaataa aacctcataa ttcacatgaa ggtaaaacat tttatgtttt
5101 acctaatgat gacactctac gtgttgaggc ttttgagtac taccacacaa ctgatcctag
5161 ttttctgggt aggtacatgt cagcattaaa tcacactaaa aagtggaaat atccacaag
```

Fig. 12A continued

```
5221 taatggttta acttctatta aatgggcaga taacaactgt tatcttgcca ctgcattgtt
5281 aacactccaa caaatagagt tgaagtttaa tccacctgct ctacaagatg cttattacag
5341 agcaagggct ggtgaagctg ctaacttttg tgcacttatc ttagcctact gtaataagac
5401 agtaggtgag ttaggtgatg ttagagaaac aatgagttac ttgtttcaac atgccaattt
5461 agattcttgc aaaagagtct tgaacgcggt gtgtaaaact tgtggacaac agcagacaac
5521 ccttaagggt gtagaagctg ttatgtacat gggcacactt tcttatgaac aatttaagaa
5581 aggtgttcag ataccttgta cgtgtggtaa acaagctaca aaatatctag tacaacagga
5641 gtcacctttt gttatgatgt cagcaccacc tgctcagtat gaacttaagc atggtacatt
5701 tacttgtgct agtgagtaca ctggtaatta ccagtgtggt cactataaac atataacttc
5761 taaagaaact ttgtattgca tagacggtgc tttacttaca aagtcctcag aatacaaagg
5821 tcctattacg gatgttttct acaaagaaaa cagttacaca acaaccataa aaccagttac
5881 ttataaattg gatggtgttg tttgtacaga aattgaccct aagttggaca attattataa
5941 gaaagacaat tcttatttca cagagcaacc aattgatctt gtaccaaacc aaccatatcc
6001 aaacgcaagc ttcgataatt ttaagtttgt atgtgataat atcaaatttg ctgatgattt
6061 aaaccagtta actggttata agaaacctgc ttcaagagag cttaaagtta catttttccc
6121 tgacttaaat ggtgatgtgg tggctattga ttataaacac tacacaccct cttttaagaa
6181 aggagctaaa ttgttacata aacctattgt ttggcatgtt aacaatgcaa ctaataaagc
6241 cacgtataaa ccaaatacct ggtgtatacg ttgtctttgg annnnnnnnn nnnnnnnnnn
6301 nnnnnnnnnn nnnnatgtac tgaagtcaga ggacgcgcag ggaatggata atcttgcctg
6361 cgaagatcta aaactagtct ctgaagaagt agtggaaaat cctaccatac agaaagacgt
```

Fig. 12A continued

```
6421 tcttgagtgt aatgtgaaaa ctaccgaagt tgtaggagac attatactta aaccagcaaa
6481 taatagttta aaaattacag aagaggttgg ccacacagat ctaatggctg cttatgtaga
6541 caattctagt cttactatta agaaacctaa tgaattatct agagtattag gtttgaaaac
6601 ccttgctact catggtttag ctgctgttaa tagtgtccct tgggatacta tagctaatta
6661 tgctaagcct tttcttaaca aagttgttag tacaactact aacatagtta cacggtgttt
6721 aaaccgtgtt tgtactaatt atatgcctta tttctttact ttattgctac aattgtgtac
6781 ttttactaga agtacaaatt ctagaattaa agcatctatg ccgactacta tagcaaagaa
6841 tactgttaag agtgtcggta aattttgtct agaggcttca tttaattatt tgaagtcacc
6901 taatttttct aaactgataa atattataat ttggttttta ctattaagtg tttgcctagg
6961 ttctttaatc tactcaaccg ctgctttagg tgttttaatg tctaatttag gcatgccttc
7021 ttactgtact ggttacagag aaggctattt gaactctact aatgtcacta ttgcaaccta
7081 ctgtactggt tctatatctt gtagtgtttg tcttagtggt ttagattctt tagacaccta
7141 tccttcttta gaaactatac aaattaccat ttcatctttt aaatgggatt taactgcttt
7201 tggcttagtt gcagagtggt ttttggcata tattcttttc actaggtttt tctatgtact
7261 tggattggct gcaatcatgc aattgttttt cagctatttt gcagtacatt ttattagtaa
7321 ttcttggctt atgtggttaa taattaatct tgtacaaatg gccccgattt cagctatggt
7381 tagaatgtac atcttctttg catcatttta ttatgtatgg aaaagttatg tgcatgttgt
7441 agacggttgt aattcatcaa cttgtatgat gtgttacaaa cgtaatagag caacaagagt
7501 cgaatgtaca actattgtta atggtgttag aaggtccttt tatgtctatg ctaatggagg
7561 taaaggcttt tgcaaactac acaattggaa ttgtgttaat tgtgatacat tctgtgctgg
7621 tagtacattt attagtgatg aagttgcgag agacttgtca ctacagttta aaagaccaa
```

Fig. 12A continued

```
7681 aaatcctact gaccagtctt cttacatcgt tgatagtgtt acagtgaaga atggttccat
7741 ccatctttac tttgataaag ctggtcaaaa gacttatgaa agacattctc tctctcattt
7801 tgttaactta gacaacctga gagctaataa cactaaaggt tcattgccta ttaatgttat
7861 agtttttgat ggtaaatcaa aatgtgaaga atcatctgca aaatcagcgt ctgtttacta
7921 cagtcagctt atgtgtcaac ctatactgtt actagatcag gcattagtgt ctgatgttgg
7981 tgatagtgcg gaagttgcag ttaaaatgtt tgatgcttac gttaatacgt tttcatcaac
8041 ttttaacgta ccaatggaaa aactcaaaac actagttgca actgcagaag ctgaacttgc
8101 aaagaatgtg tccttagaca atgtcttatc tacttttatt tcagcagctc ggcaagggtt
8161 tgttgattca gatgtagaaa ctaaagatgt tgttgaatgt cttaaattgt cacatcaatc
8221 tgacatagaa gttactggcg atagttgtaa taactatatg ctcacctata acaaagttga
8281 aaacatgaca ccccgtgacc ttggtgcttg tattgactgt agtgcgcgtc atattaatgc
8341 gcaggtagca aaaagtcaca acattgcttt gatatggaac gttaaagatt tcatgtcatt
8401 gtctgaacaa ctacgaaaac aaatacgtag tgctgctaaa aagaataact taccttttaa
8461 gttgacatgt gcaactacta gacaagttgt taatgttgta acaacaaaga tagcacttaa
8521 gggtggtaaa attgttaata attggttgaa gcagttaatt aaagttacac ttgtgttcct
8581 ttttgttgct gctatttttct atttaataac acctgttcat gtcatgtcta aacatactga
8641 cttttcaagt gaaatcatag gatacaaggc tattgatggt ggtgtcactc gtgacatagc
8701 atctacagat acttgttttg ctaacaaaca tgctgatttt gacacatggt ttagccagcg
8761 tggtggtagt tatactaatg acaaagcttg cccattgatt gctgcagtca taacaagaga
8821 agtgggtttt gtcgtgcctg gtttgcctgg cacgatatta cgcacaacta atggtgactt
```

Fig. 12A continued

```
 8881 tttgcatttc ttacctagag tttttagtgc agttggtaac atctgttaca caccatcaaa
 8941 acttatagag tacactgatt ttgcaacatc agcttgtgtt ttggctgctg aatgtacaat
 9001 ttttaaagat gcttctggta agccattacc atattgttat gataccaatg tactagaagg
 9061 ttctgttgct tatgaaagtt tacgccctga cacacgttat gtgctcatgg atggctctat
 9121 tattcaattt cctaacacct accttgaagg ttctgttaga gtggtaacaa cttttgattc
 9181 tgagtactgt aggcacggca cttgtgaaag atcagaagct ggtgtttgtg tatctactag
 9241 tggtagatgg gtacttaaca atgattatta cagatcttta ccaggagttt tctgtggtgt
 9301 agatgctgta aatttactta ctaatatgtt tacaccacta attcaaccta ttggtgcttt
 9361 ggacatatca gcatctatag tagctggtgg tattgtagct atcgtagtaa catgccttgc
 9421 ctactatttt atgaggttta gaagagcttt tggtgaatac agtcatgtag ttgcctttaa
 9481 tactttacta ttccttatgt cattcactgt actctgttta acaccagttt actcattctt
 9541 acctggtgtt tattctgtta tttacttgta cttgacattt tatcttacta atgatgtttc
 9601 ttttttagca catattcagt ggatggttat gttcacacct ttagtacctt tctggataac
 9661 aattgcttat atcatttgta tttccacaaa gcatttctat tggttcttta gtaattacct
 9721 aaagagacgt gtagtcttta atggtgtttc ctttagtact tttgaagaag ctgcgctgtg
 9781 caccttttttg ttaaataaag aaatgtatct aaagttgcgt agtgatgtgc tattacctct
 9841 tacgcaatat aatagatact tagctcttta taataagtac aagtatttta gtggagcaat
 9901 ggatacaact agctacagag aagctgcttg ttgtcatctc gcaaaggctc tcaatgactt
 9961 cagtaactca ggttctgatg ttctttacca accaccacaa atctctatca cctcagctgt
10021 tttgcagagt ggttttagaa aaatggcatt cccatctggt aaagttgagg gttgtatggt
10081 acaagtaact tgtggtacaa ctacacttaa cggtctttgg cttgatgacg tagtttag
```

Fig. 12A continued

```
10141 tccaagacat gtgatctgca cctctgaaga catgcttaac cctaattatg aagatttact
10201 cattcgtaag tctaatcata atttcttggt acaggctggt aatgttcaac tcagggttat
10261 tggacattct atgcaaaatt gtgtacttaa gcttaaggtt gatacagcca atcctaagac
10321 acctaagtat aagtttgttc gcattcaacc aggacagact ttttcagtgt tagcttgtta
10381 caatggttca ccatctggtg tttaccaatg tgctatgagg cccaatttca ctattaaggg
10441 ttcattcctt aatggttcat gtggtagtgt tggttttaac atagattatg actgtgtctc
10501 tttttgttac atgcaccata tggaattacc aactggagtt catgctggca cagacttaga
10561 aggtaacttt tatggacctt ttgttgacag gcaaacagca caagcagctg gtacggacac
10621 aactattaca gttaatgttt tagcttggtt gtacgctgct gttataaatg gagacaggtg
10681 gtttctcaat cgatttacca caactcttaa tgactttaac cttgtggcta tgaagtacaa
10741 ttatgaacct ctaacacaag accatgttga catactagga cctctttctg ctcaaactgg
10801 aattgccgtt ttagatatgt gtgcttcatt aaaagaatta ctgcaaaatg gtatgaatgg
10861 acgtaccata ttgggtagtg ctttattaga agatgaattt acacctttg atgttgttag
10921 acaatgctca ggtgttactt tccaaagtgc agtgaaaaga acaatcaagg gtacacacca
10981 ctggttgtta ctcacaattt tgacttcact tttagtttta gtccagagta ctcaatggtc
11041 tttgttcttt tttttgtatg aaaatgcctt tttacctttt gctatgggta ttattgctat
11101 gtctgctttt gcaatgatgt ttgtcaaaca taagcatgca tttctctgtt tgtttttgtt
11161 accttctctt gccgctgtag cttattttaa tatggtctat atgcctgcta gttgggtgat
11221 gcgtattatg acatggttgg atatggttga tactagtttg tctggtttta agctaaaaga
11281 ctgtgttatg tatgcatcag ctgtggtgtt actaatcctt atgacagcaa gaactgtgta
```

Fig. 12A continued

```
11341 tgatgatggt gctaggagag tgtggacact tatgaatgtc ttgacactcg tttataaagt
11401 ttattatggt aatgctttag atcaagccat ttccatgtgg gctcttataa tctctgttac
11461 ttctaactac tcaggtgtag ttacaactgt catgtttttg gccagaggta ttgtttttat
11521 gtgtgttgag tattgcccta ttttcttcat aactggtaat acacttcagt gtataatgct
11581 agtttattgt ttcttaggct atttttgtac ttgttacttt ggcctctttt gtttactcaa
11641 ccgctacttt agactgactc ttggtgttta tgattactta gtttctacac aggagtttag
11701 atatatgaat tcacagggac tactcccacc caagaatagc atagatgcct tcaaactcaa
11761 cattaaattg ttgggtgttg gtggcaaacc ttgtatcaaa gtagccactg tacagtctaa
11821 aatgtcagat gtaaagtgca catcagtagt cttactctca gttttgcaac aactcagagt
11881 agaatcatca tctaaattgt gggctcaatg tgtccagtta cacaatgaca ttctcttagc
11941 taaagatact actgaagcct ttgaaaaaat ggtttcacta ctttctgttt tgctttccat
12001 gcagggtgct gtagacataa acaagctttg tgaagaaatg ctggacaaca gggcaacctt
12061 acaagctata gcctcagagt ttagttccct tccatcatat gcagcttttg ctactgctca
12121 agaagcttat gagcaggctg ttgctaatgg tgattctgaa gttgttctta aaaagttgaa
12181 gaagtctttg aatgtggcta aatctgaatt tgaccgtgat gcagccatgc aacgtaagtt
12241 ggaaaagatg gctgatcaag ctatgaccca aatgtataaa caggctagat ctgaggacaa
12301 gagggcaaaa gttactagtg ctatgcagac aatgcttttc actatgctta gaaagttgga
12361 taatgatgca ctcaacaaca ttatcaacaa tgcaagagat ggttgtgttc ccttgaacat
12421 aatacctctt acaacagcag ccaaactaat ggttgtcata ccagactata acacatataa
12481 aaatacgtgt gatggtacaa catttactta tgcatcagca ttgtgggaaa tccaacaggt
12541 tgtagatgca gatagtaaaa ttgttcaact tagtgaaatt agtatggaca attcacctaa
```

Fig. 12A continued

```
12601 tttagcatgg cctcttattg taacagcttt aagggccaat tctgctgtca aattacagaa
12661 taatgagctt agtcctgttg cactacgaca gatgtcttgt gctgccggta ctacacaaac
12721 tgcttgcact gatgacaatg cgttagctta ctacaacaca acaaagggag gtaggtttgt
12781 acttgcactg ttatccgatt tacaggattt gaaatgggct agattcccta agagtgatgg
12841 aactggtact atctatacag aactggaacc accttgtagg tttgttacag acacacctaa
12901 aggtcctaaa gtgaagtatt tatactttat taaaggatta aacaacctaa atagaggtat
12961 ggtacttggt agtttagctg ccacagtacg tctacaagct ggtaatgcaa cagaagtgcc
13021 tgccaattca actgtattat ctttctgtgc ttttgctgta gatgctgcta aagcttacaa
13081 agattatcta gctagtgggg gacaaccaat cactaattgt gttaagatgt tgtgtacaca
13141 cactggtact ggtcaggcaa taacagttac accggaagcc aatatggatc aagaatcctt
13201 tggtggtgca tcgtgttgtc tgtactgccg ttgccacata gatcatccaa atcctaaagg
13261 attttgtgac ttaaaaggta agtatgtaca aatacctaca acttgtgcta atgaccctgt
13321 gggttttaca cttaaaaaca cagtctgtac cgtctgcggt atgtggaaag gttatggctg
13381 tagttgtgat caactccgcg aacccatgct tcagtcagct gatgcacaat cgtttttaaa
13441 cgggtttgcg gtgtaagtgc agcccgtctt acaccgtgcg gcacaggcac tagtactgat
13501 gtcgtataca gggcttttga catctacaat gataaagtag ctggttttgc taaattccta
13561 aaaactaatt gttgtcgctt ccaagaaaag gacgaagatg acaatttaat tgattcttac
13621 tttgtagtta agagacacac tttctctaac taccaacatg aagaaacaat ttataattta
13681 cttaaggatt gtccagctgt tgctaaacat gacttcttta agtttagaat agacggtgac
13741 atggtaccac atatatcacg tcaacgtctt actaaataca caatggcaga cctcgtctat
```

Fig. 12A continued

```
13801 gctttaaggc attttgatga aggtaattgt gacacattaa aagaaatact tgtcacatac
13861 aattgttgtg atgatgatta tttcaataaa aaggactggt atgattttgt agaaaaccca
13921 gatatattac gcgtatacgc caacttaggt gaacgtgtac gccaagcttt gttaaaaaca
13981 gtacaattct gtgatgccat gcgaaatgct ggtattgttg gtgtactgac attagataat
14041 caagatctca atggtaactg gtatgatttc ggtgatttca tacaaaccac gccaggtagt
14101 ggagttcctg ttgtagattc ttattattca ttgttaatgc ctatattaac cttgaccagg
14161 gctttaactg cagagtcaca tgttgacact gacttaacaa agccttacat taagtgggat
14221 ttgttaaaat atgacttcac ggaagagagg ttaaaactct ttgaccgtta ttttaaatat
14281 tgggatcaga cataccaccc aaattgtgtt aactgtttgg atgacagatg cattctgcat
14341 tgtgcaaact ttaatgtttt attctctaca gtgttcccac ttacaagttt tggaccacta
14401 gtgagaaaaa tatttgttga tggtgttcca tttgtagttt caactggata ccacttcaga
14461 gagctaggtg ttgtacataa tcaggatgta aacttacata gctctagact tagttttaag
14521 gaattacttg tgtatgctgc tgaccctgct atgcacgctg cttctggtaa tctattacta
14581 gataaacgca ctacgtgctt ttcagtagct gcacttacta acaatgttgc ttttcaaact
14641 gtcaaacccg gtaattttaa caaagacttc tatgactttg ctgtgtctaa gggtttcttt
14701 aaggaaggaa gttctgttga attaaaacac ttcttctttg ctcaggatgg taatgctgct
14761 atcagcgatt atgactacta tcgttataat ctaccaacaa tgtgtgatat cagacaacta
14821 ctatttgtag ttgaagttgt tgataagtac tttgattgtt acgatggtgg ctgtattaat
14881 gctaaccaag tcatcgtcaa caacctagac aaatcagctg gttttccatt taataaatgg
14941 ggtaaggcta gactttatta tgattcaatg agttatgagg atcaagatgc acttttcgca
15001 tatacaaaac gtaatgtcat ccctactata actcaaatga atcttaagta tgccattac
```

Fig. 12A continued

```
15061 gcaaagaata gagctcgcac cgtagctggt gtctctatct gtagtactat gaccaataga
15121 cagtttcatc aaaaattatt gaaatcaata gccgccacta gaggagctac tgtagtaatt
15181 ggaacaagca aattctatgg tggttggcac aacatgttaa aaactgttta tagtgatgta
15241 gaaaaccctc accttatggg ttgggattat cctaaatgtg atagagccat gcctaacatg
15301 cttagaatta tggcctcact tgttcttgct cgcaaacata caacgtgttg tagcttgtca
15361 caccgtttct atagattagc taatgagtgt gctcaagtat tgagtgaaat ggtcatgtgt
15421 ggcagttcac tatatgttaa accaggtgga acctcatcag gagatgccac aactgcttat
15481 gctaatagtg tttttaacat ttgtcaagct gtcacggcca atgttaatgc acttttatct
15541 actgatggta acaaaattgc cgataagtat gtccgcaatt tacaacacag actttatgag
15601 tgtctctata gaaatagaga tgttgacaca gactttgtga atgagtttta cgcatatttg
15661 cgtaaacatt tctcaatgat gatactctct gacgatgctg ttgtgtgttt caatagcact
15721 tatgcatctc aaggtctagt ggctagcata aagaacttta agtcagttct ttattatcaa
15781 aacaatgttt ttatgtctga agcaaaatgt tggactgaga ctgaccttac taaaggacct
15841 catgaatttt gctctcaaca tacaatgcta gttaaacagg gtgatgatta tgtgtacctt
15901 ccttacccag atccatcaag aatcctaggg gccggctgtt ttgtagatga tatcgtaaaa
15961 acagatggta cacttatgat tgaacggttc gtgtctttag ctatagatgc ttacccactt
16021 actaaacatc ctaatcagga gtatgctgat gtctttcatt tgtacttaca atacataaga
16081 aagctacatg atgagttaac aggacacatg ttagacatgt attctgttat gcttactaat
16141 gataacactt caaggtattg ggaacctgag ttttatgagg ctatgtacac accgcataca
16201 gtcttacagg ctgttggggc ttgtgttctt tgcaattcac agacttcatt aagatgtggt
16261 gcttgcatac gtagaccatt cttatgttgt aaatgctgtt acgaccatgt catatcaac
```

Fig. 12A continued

```
16321 tcacataaat tagtcttgtc tgttaatccg tatgtttgca atgctccagg ttgtgatgtc
16381 acagatgtga ctcaacttta cttaggaggt atgagctatt attgtaaatc acataaacta
16441 cccattagtt ttccattgtg tgctaatgga caagtttttg gtttatataa aaatacatgt
16501 gttggtagcg ataatgttac tgactttaat gcaattgcaa catgtgactg gacaaatgct
16561 ggtgattaca ttttagctaa cacctgtact gaaagactca agctttttgc agcagaaacg
16621 ctcaaagcta ctgaggagac atttaaactg tcttatggta ttgctactgt acgtgaagtg
16681 ctgtctgaca gagaattaca tctttcatgg gaagttggta aacctagacc accacttaac
16741 cgaaattatg tctttactgg ttatcgtgta actaaaaaca gtaaagtaca aataggagag
16801 tacacctttg aaaaaggtga ctatggtgat gctgttgttt 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 tggtactggt aagagtcatt ttgctattgg cctagctctc
17101 tactaccctt ctgctcgcat agtgtataca gcttgctctc atgccgctgt tgatgcacta
17161 tgtgagaagg cattaaaata tttgcctata gataaatgta gtagaattat acctgcacgt
17221 gctcgtgtag agtgttttga taaattcaaa gtgaattcaa cattagaaca gtatgtcttt
17281 tgtactgtaa atgcattgcc tgagacgaca gcagatatag ttgtctttga tgaaatttca
17341 atggccacaa attatgattt gagtgttgtc aatgccagat tacgtgctaa gcactatgtg
17401 tacattggcg accctgctca attacctgca ccacgcacat tgctaactaa gggcacacta
17461 gaaccagaat atttcaattc agtgtgtaga cttatgaaaa ctataggtcc agacatgttc
```

Fig. 12A continued

```
17521 ctcggaactt gtcggcgttg tcctgctgaa attgttgaca ctgtgagtgc tttggtttat
17581 gataataagc ttaaagcaca taaagacaaa tcagctcaat gctttaaaat gttttataag
17641 ggtgttatca cgcatgatgt ttcatctgca attaacaggc cacaaatagg cgtggtaaga
17701 gaattcctta cacgtaaccc tgcttggaga aaagctgtct ttatttcacc ttataattca
17761 cagaatgctg tagcctcaaa gattttggga ctaccaactc aaactgttga ttcatcacag
17821 ggctcagaat atgactatgt catattcact caaaccactg aaacagctca ctcttgtaat
17881 gtaaacagat ttaatgttgc tattaccaga gcaaaagtag gcatactttg cataatgtct
17941 gatagagacc tttatgacaa gttgcaattt acaagtcttg aaattccacg taggaatgtg
18001 gcaactttac aagctgaaaa tgtaacagga ctctttaaag attgtagtaa ggtaatcact
18061 gggttacatc ctacacaggc acctacacac ctcagtgttg acactaaatt caaaactgaa
18121 ggtttatgtg ttgacatacc tggcatacct aaggacatga cctatagaag actcatctct
18181 atgatgggtt ttaaaatgaa ttatcaagtt aatggttacc ctaacatgtt tatcacccgc
18241 gaagaagcta taagacatgt acgtgcatgg attggcttcg atgtcgaggg gtgtcatgct
18301 actagagaag ctgttggtac caatttacct ttacagctag gtttttctac aggtgttaac
18361 ctagttgctg tacctacagg ttatgttgat acacctaata atacagattt ttccagagtt
18421 agtgctaaac caccgcctgg agatcaattt aaacacctca taccacttat gtacaaagga
18481 cttccttgga atgtagtgcg tataaagatt gtacaaatgt taagtgacac acttaaaaat
18541 ctctctgaca gagtcgtatt tgtcttatgg gcacatggct ttgagttgac atctatgaag
18601 tattttgtga aaataggacc tgagcgcacc tgttgtctat gtgatagacg tgccacatgc
18661 ttttccactg cttcagacac ttatgcctgt tggcatcatt ctattggatt tgattacgtc
18721 tataatccgt ttatgattga tgttcaacaa tggggtttta caggtaacct acaaagca
```

Fig. 12A continued

```
18781 catgatctgt attgtcaagt ccatggtaat gcacatgtag ctagttgtga tgcaatcatg
18841 actaggtgtc tagctgtcca cgagtgcttt gttaagcgtg ttgactggac tattgaatat
18901 cctataattg gtgatgaact gaagattaat gcggcttgta gaaaggttca acacatggtt
18961 gttaaagctg cattattagc agacaaattc ccagttcttc acgacattgg taaccctaaa
19021 gctattaagt gtgtacctca agctgatgta gaatggaagt tctatgatgc acagccttgt
19081 agtgacaaag cttataaaat agaagaatta ttctattctt atgccacaca ttctgacaaa
19141 ttcacagatg gtgtatgcct attttggaat tgcaatgtcg atagatatcc tgttaattcc
19201 attgtttgta gatttgacac tagagtgcta tctaacctta acttgcctgg ttgtgatggt
19261 ggcagtttgt atgtaaataa acatgcattc cacacaccag cttttgataa aagtgctttt
19321 gttaatttaa aacaattacc atttttctat tactctgaca gtccatgtga gtctcatgga
19381 aaacaagtag tgtcagatat agattatgta ccactaaagt ctgctacgtg tataacacgt
19441 tgcaatttag gtggtgctgt ctgtagacat catgctaatg agtacagatt gtatctcgat
19501 gcttataaca tgatgatctc agctggcttt agcttgtggg tttacaaaca atttgatact
19561 tataacctct ggaacacttt tacaagactt cagagtttag aaaatgtggc ttttaatgtt
19621 gtaaataagg gacactttga tggacaacag ggtgaagtac cagtttctat cattaataac
19681 actgtttaca caaaagttga tggtgttgat gtagaattgt ttgaaaataa aacaacatta
19741 cctgttaatg tagcatttga gctttgggct aagcgcaaca ttaaaccagt accagaggtg
19801 aaaatactca ataatttggg tgtggacatt gctgctaata ctgtgatctg ggactacaaa
19861 agagatgctc cagcacatat atctactatt ggtgtttgtt ctatgactga catagccaag
19921 aaaccaactg aaacgatttg tgcaccactc actgtctttt ttgatggtag agttgatggt
```

Fig. 12A continued

```
19981 caagtagact tatttagaaa tgcccgtaat ggtgttctta ttacagaagg tagtgttaaa
20041 ggtttacaac catctgtagg tcccaaaacaa gctagtctta atggagtcac attaattgga
20101 gaagccgtaa aaacacagtt caattattat aagaaagttg atggtgttgt ccaacaatta
20161 cctgaaactt actttactca gagtagaaat ttacaagaat ttaaacccag gagtcaaatg
20221 gaaattgatt tcttagaatt agctatggat gaattcattg aacggtataa attagaaggc
20281 tatgccttcg aacatatcgt ttatggagat tttagtcata gtcagttagg tggtttacat
20341 ctactgattg gactagctaa acgttttaag gaatcacctt ttgaattaga agattttatt
20401 cctatggaca gtacagttaa aaactatttc ataacagatg cgcaaacagg ttcatctaag
20461 tgtgtgtgtt ctgttattga tttattactt gatgattttg ttgaaataat aaaatcccaa
20521 gatttatctg tagtttctaa ggttgtcaaa gtgactattg actatacaga aatttcattt
20581 atgctttggt gtaaagatgg ccatgtagaa acattttacc caaaattaca atctagtcaa
20641 gcgtggcaac cgggtgttgc tatgcctaat ctttacaaaa tgcaaagaat gctattagaa
20701 aagtgtgacc ttcaaaatta tggtgatagt gcaacattac ctaaaggcat aatgatgaat
20761 gtcgcaaaat atactcaact gtgtcaatat ttaaacacat taacattagc tgtaccctat
20821 aatatgagag ttatacattt tggtgctggt tctgataaag gagttgcacc aggtacagct
20881 gttttaagac agtggttgcc tacgggtacg ctgcttgtcg attcagatct taatgacttt
20941 gtctctgatg cagattcaac tttgattggt gattgtgcaa ctgtacatac agctaataaa
21001 tgggatctca ttattagtga tatgtacgac cctaagacta aaaatgttac aaaagaaaat
21061 gactctaaag agggtttttt cacttacatt tgtgggttta tacaacaaaa gctagctctt
21121 ggaggttccg tggctataaa gataacagaa cattcttgga atgctgatct ttataagctc
21181 atgggacact tcgcatggtg gacagccttt gttactaatg tgaatgcgtc atcatctga
```

Fig. 12A continued

```
21241 gcatttttaa ttggatgtaa ttatcttggc aaaccacgcg aacaaataga tggttatgtc
21301 atgcatgcaa attacatatt ttggaggaat acaaatccaa ttcagttgtc ttcctattct
21361 ttatttgaca tgagtaaatt tcccccttaaa ttaaggggta ctgctgttat gtctttaaaa
21421 gaaggtcaaa tcaatgatat gattttatct cttcttagta aaggtagact tataattaga
21481 gaaaaacaaca gagttgttat ttctagtgat gttcttgtta acaactaaac gaacaatgtt
21541 tgtttttctt gttttattgc cactagtctc tagtcagtgt gttaatctta gaaccagaac
21601 tcaattaccc cctgcataca ctaattcttt cacacgtggt gtttattacc ctgacaaagt
21661 tttcagatcc tcagttttac attcaactca ggacttgttc ttacctttct tttccaatgt
21721 tacttggttc catgctatac atgtctctgg gaccaatggt actaagaggt ttgataaccc
21781 tgtcctacca tttaatgatg gtgtttattt tgcttccayt gagaagtcta acataataag
21841 aggctggatt tttggtacta ctttagattc gaagacccag tccctactta ttgttaataa
21901 cgctactaat gttgttatta aagtctgtga atttcaattt tgtaatgatc catttttgga
21961 tgtttattac cacaaaaaca acaaaagttg gatggaaagt gnnnnnngag tttattctag
22021 tgcgaataat tgcacttttg aatatgtctc tcagcctttt cttatggacc ttgaaggaaa
22081 acagggtaat ttcaaaaatc ttagggaatt tgtgtttaag aatattgatg gttatttttaa
22141 aatatattct aagcacacgc ctattaattt agtgcgtgat ctccctcagg gtttttcggc
22201 tttagaacca ttggtagatt tgccaatagg tattaacatc actaggtttc aaactttact
22261 tgctttacat agaagttatt tgactcctgg tgattcttct tcaggttgga cagctggtgc
22321 tgcagcttat tatgtgggtt atcttcaacc taggactttt ctattaaaat ataatgaaaa
22381 tggaaccatt acagatgctg tagactgtgc acttgaccct ctctcagaaa caaagtgtac
```

Fig. 12A continued

```
22441 gttgaaatcc ttcactgtag aaaaaggaat ctatcaaact tctaacttta gagtccaacc
22501 aacagaatct attgttagat ttcctaatat tacaaacttg tgcccttttg gtgaagtttt
22561 taacgccacc agatttgcat ctgtttatgc ttggaacagg aagagaatca gcaactgtgt
22621 tgctgattat tctgtcctat ataattccgc atcattttcc acttttaagt gttatggagt
22681 gtctcctact aaattaaatg atctctgctt tactaatgtc tatgcagatt catttgtaat
22741 tagaggtgat gaagtcagac aaatcgctcc agggcaaact ggaaagattg ctgattataa
22801 ttataaatta ccagatgatt ttacaggctg cgttatagct tggaattcta acaatcttga
22861 ttctaaggtt ggtggtaatt ataattaccg gtatagattg tttaggaagt ctaatctcaa
22921 accttttgag agagatattt caactgaaat ctatcaggcc ggtagcaaac cttgtaatgg
22981 tgttgaaggt tttaattgtt actttccttt acaatcatat ggtttccaac ccactaatgg
23041 tgttggttac caaccataca gagtagtagt actttctttt gaacttctac atgcaccagc
23101 aactgtttgt ggacctaaaa agtctactaa tttggttaaa aacaaatgtg tcaatttcaa
23161 cttcaatggt ttaacaggca caggtgttct tactgagtct aacaaaaagt ttctgccttt
23221 ccaacaattt ggcagagaca ttgctgacac tactgatgct gtccgtgatc cacagacact
23281 tgagattctt gacattacac catgttcttt tggtggtgtc agtgttataa caccaggaac
23341 aaatacttct aaccaggttg ctgttcttta tcagggtgtt aactgcacag aagtccctgt
23401 tgctattcat gcagatcaac ttactcctac ttggcgtgtt tattctacag gttctaatgt
23461 ttttcaaaca cgtgcaggct gtttaatagg ggctgaacat gtcaacaact catatgagtg
23521 tgacataccc attggtgcag gtatatgcgc tagttatcag actcagacta attctcgtcg
23581 gcgggcacgt agtgtagcta gtcaatccat cattgcctac actatgtcac ttggtgcaga
23641 aaattcagtt gcttactcta ataactctat tgccataccc acaaatttta ctattagtgt
```

Fig. 12A continued

```
23701 taccacagaa attctaccag tgtctatgac caagacatca gtagattgta caatgtacat
23761 ttgtggtgat tcaactgaat gcagcaatct tttgttgcaa tatggcagtt tttgtacaca
23821 attaaaccgt gctttaactg gaatagctgt tgaacaagac aaaaacaccc aagaagtttt
23881 tgcacaagtc aaacaaattt acaaaacacc accaattaaa gattttggtg gttttaattt
23941 ttcacaaata ttaccagatc catcaaaacc aagcaagagg tcatttattg aagatctact
24001 tttcaacaaa gtgacacttg cagatgctgg cttcatcaaa caatatggtg attgccttgg
24061 tgatattgct gctagagacc tcatttgtgc acaaaagttt aacggcctta ctgttttgcc
24121 acctttgctc acagatgaaa tgattgctca atacacttct gcactgttag cgggtacaat
24181 cacttctggt tggacctttg gtgcaggtgc tgcattacaa ataccatttg ctatgcaaat
24241 ggcttatagg tttaatggta ttggagttac acagaatgtt ctctatgaga accaaaaatt
24301 gattgccaac caatttaata gtgctattgg caaaattcaa gactcacttt cttccacagc
24361 aagtgcactt ggaaaacttc aaaatgtggt caaccaaaat gcacaagctt taaacacgct
24421 tgttaaacaa cttagctcca attttggtgc aatttcaagt gttttaaatg atatcctttc
24481 acgtcttgac aaagttgagg ctgaagtgca aattgatagg ttgatcacag gcagacttca
24541 aagtttgcag acatatgtga ctcaacaatt aattagagct gcagaaatca gagcttctgc
24601 taatcttgct gctactaaaa tgtcagagtg tgtacttgga caatcaaaaa gagttgattt
24661 ttgtggaaag ggctatcatc ttatgtcctt ccctcagtca gcacctcatg gtgtagtctt
24721 cttgcatgtg acttatgtcc ctgcacaaga aaagaacttc acaactgctc ctgccatttg
24781 tcatgatgga aaagcacact ttcctcgtga aggtgtcttt gtttcaaatg gcacacactg
24841 gtttgtaaca caaaggaatt tttatgaacc acaaatcatt actacagaca acacatttgt
```

Fig. 12A continued

```
24901 gtctggtaac tgtgatgttg taataggaat tgtcaacaac acagtttatg atcctttgca
24961 acctgaatta gactcattca aggaggagtt agataaatat tttaagaatc atacatcacc
25021 agatgttgat ttaggtgaca tctctggcat taatgcttca gttgtaaaca ttcaaaaaga
25081 aattgaccgc ctcaatgagg ttgccaagaa tttaaatgaa tctctcatcg atctccaaga
25141 acttggaaag tatgagcagt atataaaatg gccatggtac atttggctag gttttatagc
25201 tggcttgatt gccatagtaa tggtgacaat tatgctttgc tgtatgacca gttgctgtag
25261 ttgtctcaag ggctgttgtt cttgtggatc ctgctgcaaa tttgatgaag acgactctga
25321 gccagtgctc aaaggagtca aattacatta cacataaacg aacttatgga tttgtttatg
25381 agaatcttca caattggaac tgtaactttg aagcaaggtg aaatcaagga tgctactcct
25441 ttagattttg ttcgcgctac tgcaacgata ccgatacaag cctcactccc tttcggatgg
25501 cttattgttg gcgttgcact tcttgctgtt tttcagagcg cttccaaaat cataaccctc
25561 aaaaagagat ggcaactagc actctccaag ggtgttcact ttgtttgcaa cttgctgttg
25621 ttgtttgtaa cagtttactc acacctttg ctcgttgctg ctggccttga agccccttt
25681 ctctatcttt atgctttagt ctacttcttg cagagtataa actttgtaag aataataatg
25741 aggctttggc tttgctggaa atgccgttcc aaaaacccat tactttatga tgccaactat
25801 tttctttgct ggcatactaa ttgttacgac tattgtatac cttacaatag tgtaacttct
25861 tcaattgtca ttacttcagg tgatggcaca acaagtccta tttctgaaca tgactaccag
25921 attggtggtt atactgaaaa atgggaatct ggagtaaaag actgtgttgt attacacagt
25981 tacttcactt cagactatta ccagctgtac tcaactcaat tgagtacaga cactggtgtt
26041 gaacatgtta ccttcttcat ctacaataaa attgttgatg agcctgaaga acatgtccaa
26101 attcacacaa tcgacggttc atccggagtt gttaatccag taatggaacc aatttatgat
```

Fig. 12A continued

```
26161 gaaccgacga cgactactag cgtgcctttg taagcacaag ctgatgagta cgaacttatg
26221 tactcattcg tttcggaaga gacaggtacg ttaatagtta atagcgtact tctttttctt
26281 gctttcgtgg tattcttgct agttacacta gccatcctta ctgcgcttcg attgtgtgcg
26341 tactgctgca atattgttaa cgtgagtctt gtaaaacctt ctttttacgt ttactctcgt
26401 gttaaaaatc tgaattcttc tagagttcct gatcttctgg tctaaacgaa ctaaatatta
26461 tattagtttt tctgtttgga actttaattt tagccatggc agattccaac ggtactatta
26521 ccgttgaaga gcttaaaaag ctccttgaac aatggaacct agtaataggt ttcctattcc
26581 ttacatggat ttgtcttcta caatttgcct atgccaacag gaataggttt ttgtatataa
26641 ttaagttaat tttcctctgg ctgttatggc cagtaacttt agcttgtttt gtgcttgctg
26701 ctgtttacag aataaattgg atcaccggtg gaattgctay cgcaatggct tgtcttgtag
26761 gcttgatgtg gctcagctac ttcattgctt ctttcagact gtttgcgcgt acgcgttcca
26821 tgtggtcatt caatccagaa actaacattc ttctcaacgt gccactccat ggcactattc
26881 tgaccagacc gcttctagaa agtgaactcg taatcggagc tgtgatcctt cgtggacatc
26941 ttcgtattgc tggacaccat ctaggacgct gtgacatcaa ggacctgcct aaagaaatca
27001 ctgttgctac atcacgaacg ctttcttatt acaaattggg agcttcgcag cgtgtagcag
27061 gtgactcagg ttttgctgca tacagtcgct acaggattgg caactataaa ttaaacacag
27121 accattccag tagcagtgac aatattgctt tgcttgtaca gtaagtgaca acagatgttt
27181 catctcgttg actttcaggt tactatagca gagatattac taattattat gaggactttt
27241 aaagtttcca tttggaatct tgattacatc ataaacctca taattaaaaa tttatctaag
27301 tcactaactg agaataaata ttctcaatta gatgaagagc aaccaatgga gattgattaa
```

Fig. 12A continued

```
27361 acgaacatga aaattattct tttcttggca ctgataacac tcgctacttg tgagctttat
27421 cactaccaag agtgtgttag aggtacaaca gtacttttaa aagaaccttg ctcttctgga
27481 acatacgagg gcaattcacc atttcatcct ctagctgata acaaatttgc actgacttgc
27541 tttagcactc aatttgcttt tgcttgtcct gacggcgtaa aacacgtcta tcagttacgt
27601 gccagatcag cttcacctaa actgttcatc agacaagagg aagttcaaga actttactct
27661 ccaatttttc ttattgttgc ggcaatagtg tttataacac tttgcttcac actcaaaaga
27721 aagacagaat gattgaactt tcattaattg acttctattt gtgcttttta gcctttctgc
27781 tattccttgt tttaattatg cttattatct tttggttctc acttgaactg caagatcata
27841 atgaaatttg tcacgcctaa acgaacatga aatttcttgt tttcttagga atcatcacaa
27901 ctgtagctgc atttcaccaa gaatgtagtt tacagtcatg tactcaacat caaccatatg
27961 tagttgatga cccgtgtcct attcacttct attctaaatg gtatattaga gtaggagcta
28021 gaaaatcagc acctttaatt gaattgtgcg tggatgaggc tggttctaaa tcacccattc
28081 agtacatcga tatcggtaat tatacagttt cctgtttacc ttttacaatt aattgccagg
28141 aacctaaatt gggtagtctt gtagtgcgtt gttcgttcta tgaagacttt ttagagtatc
28201 atgacgttcg tgttgtttta gatttaatct aaacgaacaa actnaaatgt ctgataatgg
28261 accccaaaat cagcgaaatg cacccgcat tacgtttggt ggaccctcag attcaactgg
28321 cagtaaccag aatggagaac gcagtggggc gcgatcaaaa caacgtcggc cccaaggttt
28381 acccaataat actgcgtctt ggttcaccgc tctcactcaa catggcaagg aaggccttaa
28441 attccctcga ggacaaggcg ttccaattaa caccaatagc agtccagatg accaaattgg
28501 ctactaccga agagctacca gacgaattcg tggtggtgac ggtaaaatga aagatctcag
28561 tccaagatgg tatttctact acctaggaac tgggccagaa gctggacttc cctatggtgc
```

Fig. 12A continued

```
28621 taacaaagac ggcatcatat gggttgcaac tgagggagcc ttgaatacac caaaagatca
28681 cattggcacc cgcaatcctg ctaacaatgc tgcaatcgtg ctacaacttc ctcaaggaac
28741 aacattgcca aaaggcttct acgcagaagg gagcagaggc ggcagtcaag cctcttctcg
28801 ttcctcatca cgtagtcgca acagttcaag aaattcaact ccaggcagca gtatgggaac
28861 ttctcctgct agaatggctg gcaatggctg tgatgctgct cttgctttgc tgctgcttga
28921 cagattgaac cagcttgaga gcaaaatgtc tggtaaaggc caacaacaac aaggccaaac
28981 tgtcactaag aaatctgctg ctgaggcttc taagaagcct cggcaaaaac gtactgccac
29041 taaagcatac aatgtaacac aagctttcgg cagacgtggt ccagaacaaa cccaaggaaa
29101 ttttggggac caggaactaa tcagacaagg aactgattac aaacattggc cgcaaattgc
29161 acaatttgcc cccagcgctt cagcgttctt cggaatgtcg cgcattggca tggaagtcac
29221 accttcggga acgtggttga cctacacagg tgccatcaaa ttggatgaca aagatccaaa
29281 tttcaaagat caagtcattt tgctgaataa gcatattgac gcatacaaaa cattcccacc
29341 aacagagcct aaaaaggaca aaaagaagaa ggcttatgaa actcaagcct taccgcagag
29401 acagaagaaa cagcaaactg tgactcttct tcctgctgca gatttggatg atttctccaa
29461 acaattgcaa caatccatga gcagtgctga ctcaactcag gcctaaactc atgcagacca
29521 cacaaggcag atgggctata taaacgtttt cgcttttccg tttacgatat atagtctact
29581 cttgtgcaga atgaattctc gtaactacat agcacaagta gatgtagtta actttaatct
29641 cacatagcaa tctttaatca gtgtgtaaca ttagggagga cttgaaagag ccaccacatt
29701 ttcaccgagg ccactcggag tacgatcgag tgtacagtga acaatgctag ggagagctgc
29761 ctatatggaa gagccctaat gtgtaaaatt aattttagta gtgctatccc catgtgattt
```

Fig. 12A continued

Fig. 12A continued

29821 taatagcttc ttagga

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
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate
            SARS-CoV-2/human/NLD/EMC-Omicron-1/2021, complete genome.
ACCESSION   OM287553
VERSION     OM287553.1
KEYWORDS    .
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
```

Fig. 12B continued

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, eabo2202 (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

```
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

/codon_start=1
/product="ORF1ab polyprotein"
/protein_id="UJD17627.1"
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

Fig. 12B continued

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNER

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

Fig. 12B continued

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

Fig. 12B continued

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLIKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNIKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNSV

PWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRIK

ASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTAA

Fig. 12B continued

LGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLETI

QITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLM

WLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVEC

TTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRPI

NPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPIN

VIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNT

FSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVECL

KLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNITLIW

NVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWLK

QLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFAN

Fig. 12B continued

KHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLPR

VFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAY

ESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGR

WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLA

YYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTND

VSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEE

AALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLA

KALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGL

WLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLK

LKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRHNFTIKGSFLNGSCG

Fig. 12B continued

SVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNV

LAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVL

DMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWL

LLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFLL

PSLATVAYFNMVYMPASWVMRIMTWLDMVDTSFKLKDCVMYASAVVLLILMTARTVYD

DGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLARGVVF

MCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQ

EFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVLLSVL

QQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKLCEEM

LDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFD

Fig. 12B continued

RDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIIN

NARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDADSKI

VQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDD

NALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGIGTIYTELEPPCRFVTDTPKGPK

VKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAKAYKD

YLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPK

GFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSADAQS

FLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKDEDDN

LIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQRLTKY

TMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGE

Fig. 12B continued

RVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSYY

SLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHP

NCVNCLDDRCILHCANFNVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRELGVV

HNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKP

GNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLL

FVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALF

AYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGAT

VVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTT

CCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTA

NVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSD

Fig. 12B continued

DAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTM

LVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQE

YADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAV

GACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV

TQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAG

DYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPL

NRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMP

LSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFA

IGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNS

TLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAP

Fig. 12B continued

RILLIKGILEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKD

KSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASK

ILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLY

DKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLC

VDVPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHAT

REAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIPLMYK

GLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCLCDRR

ATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNAHVAS

CDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADKFPVL

HDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCLFWNC

Fig. 12B continued

NVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQLPFF

YYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYNMMIS

AGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNTVYTK

VDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDYKRDA

PAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEGSVKG

LQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFKPRSQ

MEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESPFELE

DFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKVTIDY

TEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATL

PKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTL

Fig. 12B continued

LVDSDLNDFVSDADSTLIGDCATVHTANKWDFIISDMYDPKTKNVTKENDSKEGFFTY

ICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFLIGCN

YLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKEGQIN

DMILSLLSKGRLIIRENNRVVISSDVLVNN"

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"

mat_peptide 9998..10915
/gene="ORF1ab"

Fig. 12B continued

```
                /product="3C-like proteinase"
mat_peptide     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

```
     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="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
```

Fig. 12B continued

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

Fig. 12B *continued*

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNER

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

Fig. 12B continued

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

Fig. 12B continued

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNIKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNSV

PWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRIK

ASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTAA

LGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLETI

QITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLM

WLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVEC

Fig. 12B continued

TTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRPI

NPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPIN

VIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVNT

FSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVECL

KLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNITLIW

NVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWLK

QLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFAN

KHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLPR

VFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAY

ESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGR

Fig. 12B continued

WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLA

YYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTND

VSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEE

AALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLA

KALNDFSNSGSDVLYQPPQISITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGL

WLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLK

LKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRHNFTIKGSFLNGSCG

SVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNV

LAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVL

DMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHWL

Fig. 12B continued

LLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFLL

PSLATVAYFNMVYMPASWVMRIMTWLDMVDTSFKLKDCVMYASAVVLLILMTARTVYD

DGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLARGVVF

MCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYLVSTQ

EFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVLLSVL

QQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKLCEEM

LDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFD

RDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIIN

NARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDADSKI

VQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTACTDD

Fig. 12B continued

NALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPK

VKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAKAYKD

YLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPK

GFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSADAQS

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

```
mat_peptide     9998..10915
                /gene="ORF1ab"
                /product="3C-like proteinase"
mat_peptide     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     13376..13414
```

Fig. 12B continued

```
                    /gene="ORF1ab"
                    /product="nsp11"
     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="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR

SSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRFDNPVLPFNDGVYFASIEKSNIIRGW
```

Fig. 12B continued

IFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDHKNNKSWMESEFRVYSSANN

CTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVREPEDLPQGFS

ALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKY

NENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCP

FDEVFNATRFASVYAWNRKRISNCVADYSVLYNLAPFFTFKCYGVSPTKLNDLCFTNV

YADSFVIRGDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNKLDSKVSGNYNYLY

RLFRKSNLKPFERDISTEIYQAGNKPCNGVAGFNCYFPLRSYSFRPTYGVGHQPYRVV

VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLKGTGVLTESNKKFLPFQQFGRDI

ADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHAD

QLTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTKSHRRAR

Fig. 12B continued

SVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYIC

GDSTECSNLLLQYGSFCTQLKRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKYFGGFN

FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFKGLT

VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY

ENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNHNAQALNTLVKQLSSKFGAISS

VLNDIFSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV

LGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPR

EGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFK

EELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE

QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVL

KGVKLHYT"

Fig. 12B continued

```
     gene            25318..26145
                     /gene="ORF3a"
     CDS             25318..26145
                     /gene="ORF3a"
                     /codon_start=1
                     /product="ORF3a protein"
                     /protein_id="UJD17630.1"

/translation="MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFG

WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE

APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY

NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ

LSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"
     gene            26170..26397
                     /gene="E"
     CDS             26170..26397
```

Fig. 12B continued

```
                    /gene="E"
                    /codon_start=1
                    /product="envelope protein"
                    /protein_id="UJD17631.1"

/translation="MYSFVSEEIGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCC
                    NIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV"
     gene           26448..27116
                    /gene="M"
     CDS            26448..27116
                    /gene="M"
                    /codon_start=1
                    /product="membrane glycoprotein"
                    /protein_id="UJD17632.1"

/translation="MAGSNGTITVEELKKLLEEWNLVIGFLFLTWICLLQFAYANRNR

FLYIIKLIFLWLLWPVTLTCFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRL

FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD
```

Fig. 12B continued

```
IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIA
                    LLVQ"
     gene           27127..27312
                    /gene="ORF6"
     CDS            27127..27312
                    /gene="ORF6"
                    /codon_start=1
                    /product="ORF6 protein"
                    /protein_id="UJD17633.1"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSL
                    TENKYSQLDEEQPMEID"
     gene           27319..27684
                    /gene="ORF7a"
     CDS            27319..27684
                    /gene="ORF7a"
                    /codon_start=1
                    /product="ORF7a protein"
                    /protein_id="UJD17634.1"
```

Fig. 12B continued

```
/translation="MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNS

PFHPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFL
                    IVAAIVFITLCFTLKRKTE"
     gene           27681..27812
                    /gene="ORF7b"
     CDS            27681..27812
                    /gene="ORF7b"
                    /codon_start=1
                    /product="ORF7b"
                    /protein_id="UJD17635.1"

/translation="MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA"
     gene           27819..28184
                    /gene="ORF8"
     CDS            27819..28184
                    /gene="ORF8"
                    /codon_start=1
                    /product="ORF8 protein"
                    /protein_id="UJD17636.1"
```

Fig. 12B continued

```
/translation="MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK

WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRC
                    SFYEDFLEYHDVRVVLDFI"
     gene           28199..29449
                    /gene="N"
     CDS            28199..29449
                    /gene="N"
                    /codon_start=1
                    /product="nucleocapsid phosphoprotein"
                    /protein_id="UJD17637.1"

/translation="MSDNGPQNQRNALRITFGGPSDSTGSNQNGGARSKQRRPQGLPN

NTASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS

PRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQ

GTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSKRTSPARMAGNGGDAALAL
```

Fig. 12B continued

```
LLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGP

EQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAI

KLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLL

                    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

```
     stem_loop        29644..29684
                      /note="Coronavirus 3' stem-loop II-like motif (s2m)"
ORIGIN
        1 agatctgttc tctaaacgaa ctttaaaatc tgtgtggctg tcactcggct gcatgcttag
       61 tgcactcacg cagtataatt aataactaat tactgtcgtt gacaggacac gagtaactcg
      121 tctatcttct gcaggctgct tacggtttcg tccgtgttgc agccgatcat cagcacatct
      181 aggttttgtc cgggtgtgac cgaaaggtaa gatggagagc cttgtccctg gtttcaacga
      241 gaaaacacac gtccaactca gtttgcctgt tttacaggtt cgcgacgtgc tcgtacgtgg
      301 ctttggagac tccgtggagg aggtcttatc agaggcacgt caacatctta aagatggcac
      361 ttgtggctta gtagaagttg aaaaaggcgt tttgcctcaa cttgaacagc cctatgtgtt
      421 catcaaacgt tcggatgctc gaactgcacc tcatggtcat gttatggttg agctggtagc
      481 agaactcgaa ggcattcagt acggtcgtag tggtgagaca cttggtgtcc ttgtccctca
      541 tgtgggcgaa ataccagtgg cttaccgcaa ggttcttctt cgtaagaacg gtaataaagg
      601 agctggtggc catagttacg gcgccgatct aaagtcattt gacttaggcg acgagcttgg
      661 cactgatcct tatgaagatt ttcaagaaaa ctggaacact aaacatagca gtggtgttac
      721 ccgtgaactc atgcgtgagc ttaacggagg ggcatacact cgctatgtcg ataacaactt
      781 ctgtggccct gatggctacc ctcttgagtg cattaaagac cttctagcac gtgctggtaa
      841 agcttcatgc actttgtccg aacaactgga ctttattgac actaagaggg gtgtatactg
      901 ctgccgtgaa catgagcatg aaattgcttg gtacacggaa cgttctgaaa agagctatga
      961 attgcagaca cctttttgaaa ttaaattggc aaagaaattt gacaccttca atggggaatg
```

Fig. 12B continued

```
1021 tccaaatttt gtatttccct taaattccat aatcaagact attcaaccaa gggttgaaaa
1081 gaaaaagctt gatggcttta tgggtagaat tcgatctgtc tatccagttg cgtcaccaaa
1141 tgaatgcaac caaatgtgcc tttcaactct catgaagtgt gatcattgtg gtgaaacttc
1201 atggcagacg ggcgattttg ttaaagccac ttgcgaattt tgtggcactg agaatttgac
1261 taaagaaggt gccactactt gtggttactt accccaaaat gctgttgtta aaatttattg
1321 tccagcatgt cacaattcag aagtaggacc tgagcatagt cttgccgaat accataatga
1381 atctggcttg aaaaccattc ttcgtaaggg tggtcgcact attgcctttg gaggctgtgt
1441 gttctcttat gttggttgcc ataacaagtg tgcctattgg gttccacgtg ctagcgctaa
1501 cataggttgt aaccatacag gtgttgttgg agaaggttcc gaaggtctta atgacaacct
1561 tcttgaaata ctccaaaaag agaaagtcaa catcaatatt gttggtgact ttaaacttaa
1621 tgaagagatc gccattattt tggcatcttt ttctgcttcc acaagtgctt ttgtggaaac
1681 tgtgaaaggt ttggattata aagcattcaa acaaattgtt gaatcctgtg gtaattttaa
1741 agttacaaaa ggaaaagcta aaaaaggtgc ctggaatatt ggtgaacaga aatcaatact
1801 gagtcctctt tatgcatttg catcagaggc tgctcgtgtt gtacgatcaa ttttctcccg
1861 cactcttgaa actgctcaaa attctgtgcg tgttttacag aaggccgcta taacaatact
1921 agatggaatt tcacagtatt cactgagact cattgatgct atgatgttca catctgattt
1981 ggctactaac aatctagttg taatggccta cattacaggt ggtgttgttc agttgacttc
2041 gcagtggcta actaacatct ttggcactgt ttatgaaaaa ctcaaacccg tccttgattg
2101 gcttgaagag aagtttaagg aaggtgtaga gtttcttaga gacggttggg aaattgttaa
2161 atttatctca acctgtgctt gtgaaattgt cggtggacaa attgtcacct gtgcaaagga
2221 aattaaggag agtgttcaga cattctttaa gcttgtaaat aaatttttgg ctttgtgtgc
```

Fig. 12B continued

```
2281 tgactctatc attattggtg gagctaaact taaagccttg aatttaggtg aaacatttgt
2341 cacgcactca aagggattgt acagaaagtg tgttaaatcc agagaagaaa 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 cctaatatga tggtaacaaa
2641 caataccttc acactcaaag gcggtgcacc aacaaaggtt acttttggtg atgacactgt
2701 gatagaagtg caaggttaca agagtgtgaa tatcactttt gaacttgatg aaaggattga
2761 taaagtactt aatgagaggt gctctgccta tacagttgaa ctcggtacag aagtaaatga
2821 gttcgcctgt gttgtggcag atgctgtcat aaaaactttg caaccagtat 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 gtaaaccttt ggaatttggt gccacttctg ctgctcttca
3121 acctgaagaa gagcaagaag aagattggtt agatgatgat agtcaacaaa ctgttggtca
3181 acaagacggc agtgaggaca atcagacaac tactattcaa acaattgttg aggttcaacc
3241 tcaattagag atggaactta caccagttgt tcagactatt gaagtgaata gttttagtgg
3301 ttatttaaaa cttactgaca atgtatacat taaaaatgca gacattgtgg aagaagctaa
3361 aaaggtaaaa ccaacagtgg ttgttaatgc agccaatgtt taccttaaac atggaggagg
3421 tgttgcagga gccttaaata aggctactaa caatgccatg caagttgaat ctgatgatta
```

Fig. 12B continued

```
3481 catagctact aatggaccac ttaaagtggg tggtagttgt gttttaagcg gacacaatct
3541 tgctaaacac tgtcttcatg ttgtcggccc aaatgttaac aaaggtgaag acattcaact
3601 tcttaagagt gcttatgaaa attttaatca gcacgaagtt ctacttgcac cattattatc
3661 agctggtatt tttggtgctg accctataca ttctttaaga gtttgtgtag atactgttcg
3721 cacaaatgtc tacttagctg tctttgataa aaatctctat gacaaacttg tttcaagctt
3781 tttggaaatg aagagtgaaa agcaagttga acaaaagatc gctgagattc ctaaagagga
3841 agttaagcca tttataactg aaagtaaacc ttcagttgaa cagagaaaac aagatgataa
3901 gaaaatcaaa gcttgtgttg aagaagttac aacaactctg gaagaaacta agttcctcac
3961 agaaaacttg ttactttata ttgacattaa tggcaatctt catccagatt ctgccactct
4021 tgttagtgac attgacatca ctttcttaaa gaaagatgct ccatatatag tgggtgatgt
4081 tgttcaagag ggtgttttaa ctgctgtggt tatacctact aaaaaggctg gtggcactac
4141 tgaaatgcta gcgaaagctt tgagaaaagt gccaacagac aattatataa ccacttaccc
4201 gggtcagggt ttaaatggtt acactgtaga ggaggcaaag acagtgctta aaaagtgtaa
4261 aagtgccttt tacattctac catctattat ctctaatgag aagcaagaaa ttcttggaac
4321 tgtttcttgg aatttgcgag aaatgcttgc acatgcagaa gaaacacgca aattaatgcc
4381 tgtctgtgtg gaaactaaag ccatagtttc aactatacag cgtaaatata agggtattaa
4441 aatacaagag ggtgtggttg attatggtgc tagattttac ttttacacca gtaaaacaac
4501 tgtagcgtca cttatcaaca cacttaacga tctaaatgaa actcttgtta caatgccact
4561 tggctatgta acacatggct taaatttgga agaagctgct cggtatatga gatctctcaa
4621 agtgccagct acagtttctg tttcttcacc tgatgctgtt acagcgtata atggttatct
4681 tacttcttct tctaaaacac ctgaagaaca ttttattgaa accatctcac ttgctggttc
```

Fig. 12B
continued

```
4741 ctataaagat tggtcctatt ctggacaatc tacacaacta ggtatagaat ttcttaagag
4801 aggtgataaa agtgtatatt acactagtaa tcctaccaca ttccacctag atggtgaagt
4861 tatcaccttt gacaatctta agacacttct ttctttgaga gaagtgagga ctattaaggt
4921 gtttacaaca gtagacaaca ttaacctcca cacgcaagtt gtggacatgt caatgacata
4981 tggacaacag tttggtccaa cttatttgga tggagctgat gttactaaaa taaaacctca
5041 taattcacat gaaggtaaaa cattttatgt tttacctaat gatgacactc tacgtgttga
5101 ggcttttgag tactaccaca caactgatcc tagttttctg ggtaggtaca tgtcagcatt
5161 aaatcacact aaaaagtgga aatacccaca agttaatggt ttaacttcta ttaaatgggc
5221 agataacaac tgttatcttg ccactgcatt gttaacactc caacaaatag agttgaagtt
5281 taatccacct gctctacaag atgcttatta cagagcaagg gctggtgaag cggctaactt
5341 ttgtgcactt atcttagcct actgtaataa gacagtaggt gagttaggtg atgttagaga
5401 aacaatgagt tacttgtttc aacatgccaa tttagattct tgcaaaagag tcttgaacgt
5461 ggtgtgtaaa acttgtggac aacagcagac aacccttaag ggtgtagaag ctgttatgta
5521 catgggcaca ctttcttatg aacaatttaa gaaaggtgtt cagatacctt gtacgtgtgg
5581 taaacaagct acaaaatatc tagtacaaca ggagtcacct tttgttatga tgtcagcacc
5641 acctgctcag tatgaactta agcatggtac atttacttgt gctagtgagt acactggtaa
5701 ttaccagtgt ggtcactata aacatataac ttctaaagaa actttgtatt gcatagacgg
5761 tgctttactt acaaagtcct cagaatacaa aggtcctatt acggatgttt tctacaaaga
5821 aaacagttac acaacaacca taaaaccagt tacttataaa ttggatggtg ttgtttgtac
5881 agaaattgac cctaagttgg acaattatta taagaaagac aattcttatt tcacagagca
```

Fig. 12B continued

```
5941 accaattgat cttgtaccaa accaaccata tccaaacgca agcttcgata attttaagtt
6001 tgtatgtgat aatatcaaat ttgctgatga tttaaaccag ttaactggtt ataagaaacc
6061 tgcttcaaga gagcttaaag ttacattttt ccctgactta aatggtgatg tggtggctat
6121 tgattataaa cactacacac cctcttttaa gaaaggagct aaattgttac ataaacctat
6181 tgtttggcat gttaacaatg caactaataa agccacgtat aaaccaaata cctggtgtat
6241 acgttgtctt tggagcacaa aaccagttga aacatcaaat tcgtttgatg tactgaagtc
6301 agaggacgcg cagggaatgg ataatcttgc ctgcgaagat ctaaaaccag tctctgaaga
6361 agtagtggaa aatcctacca tacagaaaga cgttcttgag tgtaatgtga aaactaccga
6421 agttgtagga gacattatac ttaaaccagc aaataatata aaaattacag aagaggttgg
6481 ccacacagat ctaatggctg cttatgtaga caattctagt cttactatta agaaacctaa
6541 tgaattatct agagtattag gtttgaaaac ccttgctact catggtttag ctgctgttaa
6601 tagtgtccct tgggatacta tagctaatta tgctaagcct tttcttaaca aagttgttag
6661 tacaactact aacatagtta cacggtgttt aaaccgtgtt tgtactaatt atatgcctta
6721 tttctttact ttattgctac aattgtgtac ttttactaga agtacaaatt ctagaattaa
6781 agcatctatg ccgactacta tagcaaagaa tactgttaag agtgtcggta aattttgtct
6841 agaggcttca tttaattatt tgaagtcacc taatttttct aaactgataa atattataat
6901 ttggttttta ctattaagtg tttgcctagg ttctttaatc tactcaaccg ctgctttagg
6961 tgttttaatg tctaatttag gcatgccttc ttactgtact ggttacagag aaggctattt
7021 gaactctact aatgtcacta ttgcaaccta ctgtactggt tctatacctt gtagtgtttg
7081 tcttagtggt ttagattctt tagacaccta tccttctttа gaaactatac aaattaccat
7141 ttcatctttt aaatgggatt taactgcttt tggcttagtt gcagagtggt ttttggcata
```

Fig. 12B

continued

```
7201 tattcttttc actaggtttt tctatgtact tggattggct gcaatcatgc aattgttttt
7261 cagctatttt gcagtacatt ttattagtaa ttcttggctt atgtggttaa taattaatct
7321 tgtacaaatg gccccgattt cagctatggt tagaatgtac atcttctttg catcatttta
7381 ttatgtatgg aaaagttatg tgcatgttgt agacggttgt aattcatcaa cttgtatgat
7441 gtgttacaaa cgtaatagag caacaagagt cgaatgtaca actattgtta atggtgttag
7501 aaggtccttt tatgtctatg ctaatggagg taaaggcttt tgcaaactac acaattggaa
7561 ttgtgttaat tgtgatacat tctgtgctgg tagtacattt attagtgatg aagttgcgag
7621 agacttgtca ctacagttta aaagaccaat aaatcctact gaccagtctt cttacatcgt
7681 tgatagtgtt acagtgaaga atggttccat ccatctttac tttgataaag ctggtcaaaa
7741 gacttatgaa agacattctc tctctcattt tgttaactta gacaacctga gagctaataa
7801 cactaaaggt tcattgccta ttaatgttat agtttttgat ggtaaatcaa aatgtgaaga
7861 atcatctgca aaatcagcgt ctgtttacta cagtcagctt atgtgtcaac ctatactgtt
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

```
8401 tgctgctaaa aagaataact tacctttaa gttgacatgt gcaactacta gacaagttgt
8461 taatgttgta acaacaaaga tagcacttaa gggtggtaaa attgttaata attggttgaa
8521 gcagttaatt aaagttacac ttgtgttcct ttttgttgct gctatttct atttaataac
8581 acctgttcat gtcatgtcta aacatactga cttttcaagt gaaatcatag gatacaaggc
8641 tattgatggt ggtgtcactc gtgacatagc atctacagat acttgttttg ctaacaaaca
8701 tgctgatttt gacacatggt ttagccagcg tggtggtagt tatactaatg acaaagcttg
8761 cccattgatt gctgcagtca taacaagaga agtgggtttt gtcgtgcctg gtttgcctgg
8821 cacgatatta cgcacaacta atggtgactt tttgcatttc ttacctagag tttttagtgc
8881 agttggtaac atctgttaca caccatcaaa acttatagag tacactgact ttgcaacatc
8941 agcttgtgtt ttggctgctg aatgtacaat ttttaaagat gcttctggta agccagtacc
9001 atattgttat gataccaatg tactagaagg ttctgttgct tatgaaagtt tacgccctga
9061 cacacgttat gtgctcatgg atggctctat tattcaattt cctaacacct accttgaagg
9121 ttctgttaga gtggtaacaa cttttgattc tgagtactgt aggcacggca cttgtgaaag
9181 atcagaagct ggtgtttgtg tatctactag tggtagatgg gtacttaaca atgattatta
9241 cagatcttta ccaggagttt tctgtggtgt agatgctgta aatttactta ctaatatgtt
9301 tacaccacta attcaaccta ttggtgcttt ggacatatca gcatctatag tagctggtgg
9361 tattgtagct atcgtagtaa catgccttgc ctactatttt atgaggttta gaagagcttt
9421 tggtgaatac agtcatgtag ttgcctttaa tactttacta ttccttatgt cattcactgt
9481 actctgttta acaccagttt actcattctt acctggtgtt tattctgtta tttacttgta
9541 cttgacattt tatcttacta atgatgtttc ttttttagca catattcagt ggatggttat
9601 gttcacacct ttagtacctt tctggataac aattgcttac atcatttgta tttccacaaa
```

Fig. 12B continued

```
 9661 gcatttctat tggttcttta gtaattacct aaagagacgt gtagtcttta atggtgtttc
 9721 ctttagtact tttgaagaag ctgcgctgtg caccttttttg ttaaataaag aaatgtatct
 9781 aaagttgcgt agtgatgtgc tattacctct tacgcaatat aatagatact tagctcttta
 9841 taataagtac aagtatttta gtggagcaat ggatacaact agctacagag aagctgcttg
 9901 ttgtcatctc gcaaaggctc tcaatgactt cagtaactca ggttctgatg ttctttacca
 9961 accaccacaa atctctatca cctcagctgt tttgcagagt ggttttagaa aaatggcatt
10021 cccatctggt aaagttgagg gttgtatggt acaagtaact tgtggtacaa ctacacttaa
10081 cggtctttgg cttgatgacg tagtttactg tccaagacat gtgatctgca cctctgaaga
10141 catgcttaac cctaattatg aagatttact cattcgtaag tctaatcata atttcttggt
10201 acaggctggt aatgttcaac tcagggttat tggacattct atgcaaaatt gtgtacttaa
10261 gcttaaggtt gatacagcca atcctaagac acctaagtat aagtttgttc gcattcaacc
10321 aggacagact ttttcagtgt tagcttgtta caatggttca ccatctggtg tttaccaatg
10381 tgctatgagg cacaatttca ctattaaggg ttcattcctt aatggttcat gtggtagtgt
10441 tggttttaac atagattatg actgtgtctc tttttgttac atgcaccata tggaattacc
10501 aactggagtt catgctggca cagacttaga aggtaacttt tatggacctt ttgttgacag
10561 gcaaacagca caagcagctg gtacggacac aactattaca gttaatgttt tagcttggtt
10621 gtacgctgct gttataaatg gagacaggtg gtttctcaat cgatttacca caactcttaa
10681 tgactttaac cttgtggcta tgaagtacaa ttatgaacct ctaacacaag accatgttga
10741 catactagga cctctttctg ctcaaactgg aattgccgtt ttagatatgt gtgcttcatt
10801 aaaagaatta ctgcaaaatg gtatgaatgg acgtaccata ttgggtagtg ctttattaga
```

Fig. 12B continued

```
10861 agatgaattt acacctttg atgttgttag acaatgctca ggtgttactt tccaaagtgc
10921 agtgaaaaga acaatcaagg gtacacacca ctggttgtta ctcacaattt tgacttcact
10981 tttagtttta gtccagagta ctcaatggtc tttgttcttt tttttgtatg aaaatgcctt
11041 tttacctttt gctatgggta ttattgctat gtctgctttt gcaatgatgt ttgtcaaaca
11101 taagcatgca tttctctgtt tgtttttgtt accttctctt gccactgtag cttattttaa
11161 tatggtctat atgcctgcta gttgggtgat gcgtattatg acatggttgg atatggttga
11221 tactagtttt aagctaaaag actgtgttat gtatgcatca gctgtagtgt tactaatcct
11281 tatgacagca agaactgtgt atgatgatgg tgctaggaga gtgtggacac ttatgaatgt
11341 cttgacactc gtttataaag tttattatgg taatgcttta gatcaagcca tttccatgtg
11401 ggctcttata atctctgtta cttctaacta ctcaggtgta gttacaactg tcatgttttt
11461 ggccagaggt gttgttttta tgtgtgttga gtattgccct attttcttca taactggtaa
11521 tacacttcag tgtataatgc tagtttattg tttcttaggc tatttttgta cttgttactt
11581 tggcctcttt tgtttactca accgctactt tagactgact cttggtgttt atgattactt
11641 agtttctaca caggagttta gatatatgaa ttcacaggga ctactcccac ccaagaatag
11701 catagatgcc ttcaaactca acattaaatt gttgggtgtt ggtggcaaac cttgtatcaa
11761 agtagccact gtacagtcta aaatgtcaga tgtaaagtgc acatcagtag tcttactctc
11821 agttttgcaa caactcagag tagaatcatc atctaaattg tgggctcaat gtgtccagtt
11881 acacaatgac attctcttag ctaaagatac tactgaagcc tttgaaaaaa tggtttcact
11941 actttctgtt ttgctttcca tgcagggtgc tgtagacata aacaagcttt gtgaagaaat
12001 gctggacaac agggcaacct tacaagctat agcctcagag tttagttccc ttccatcata
12061 tgcagctttt gctactgctc aagaagctta tgagcaggct gttgctaatg gtgattctga
```

Fig. 12B

continued

```
12121 agttgttctt aaaaagttga agaagtcttt gaatgtggct aaatctgaat ttgaccgtga
12181 tgcagccatg caacgtaagt tggaaaagat ggctgatcaa gctatgaccc aaatgtataa
12241 acaggctaga tctgaggaca agagggcaaa agttactagt gctatgcaga caatgctttt
12301 cactatgctt agaaagttgg ataatgatgc actcaacaac attatcaaca atgcaagaga
12361 tggttgtgtt cccttgaaca taatacctct tacaacagca gccaaactaa tggttgtcat
12421 accagactat aacacatata aaaatacgtg tgatggtaca acatttactt atgcatcagc
12481 attgtgggaa atccaacagg ttgtagatgc agatagtaaa attgttcaac ttagtgaaat
12541 tagtatggac aattcaccta atttagcatg gcctcttatt gtaacagctt taagggccaa
12601 ttctgctgtc aaattacaga ataatgagct tagtcctgtt gcactacgac agatgtcttg
12661 tgctgccggt actacacaaa ctgcttgcac tgatgacaat gcgttagctt actacaacac
12721 aacaaaggga ggtaggtttg tacttgcact gttatccgat ttacaggatt tgaaatgggc
12781 tagattccct aagagtgatg gaactggtac tatctataca gaactggaac caccttgtag
12841 gtttgttaca gacacaccta aaggtcctaa agtgaagtat ttatacttta ttaaaggatt
12901 aaacaaccta aatagaggta tggtacttgg tagtttagct gccacagtac gtctacaagc
12961 tggtaatgca acagaagtgc ctgccaattc aactgtatta tctttctgtg cttttgctgt
13021 agatgctgct aaagcttaca aagattatct agctagtggg ggacaaccaa tcactaattg
13081 tgttaagatg ttgtgtacac acactggtac tggtcaggca ataacagtca caccggaagc
13141 caatatggat caagaatcct ttggtggtgc atcgtgttgt ctgtactgcc gttgccacat
13201 agatcatcca aatcctaaag gattttgtga cttaaaaggt aagtatgtac aaatacctac
13261 aacttgtgct aatgaccctg tgggttttac acttaaaaac acagtctgta ccgtctgcgg
```

Fig. 12B continued

```
13321 tatgtggaaa ggttatggct gtagttgtga tcaactccgc gaacccatgc ttcagtcagc
13381 tgatgcacaa tcgtttttaa acgggtttgc ggtgtaagtg cagcccgtct tacaccgtgc
13441 ggcacaggca ctagtactga tgtcgtatac agggcttttg acatctacaa tgataaagta
13501 gctggttttg ctaaattcct aaaaactaat tgttgtcgct tccaagaaaa ggacgaagat
13561 gacaatttaa ttgattctta ctttgtagtt aagagacaca ctttctctaa ctaccaacat
13621 gaagaaacaa tttataattt acttaaggat tgtccagctg ttgctaaaca tgacttcttt
13681 aagtttagaa tagacggtga catggtacca catatatcac gtcaacgtct tactaaatac
13741 acaatggcag acctcgtcta tgctttaagg cattttgatg aaggtaattg tgacacatta
13801 aaagaaatac ttgtcacata caattgttgt gatgatgatt atttcaataa aaaggactgg
13861 tatgattttg tagaaaaccc agatatatta cgcgtatacg ccaacttagg tgaacgtgta
13921 cgccaagctt tgttaaaaac agtacaattc tgtgatgcca tgcgaaatgc tggtattgtt
13981 ggtgtactga cattagataa tcaagatctc aatggtaact ggtatgattt cggtgatttc
14041 atacaaacca cgccaggtag tggagttcct gttgtagatt cttattattc attgttaatg
14101 cctatattaa ccttgaccag ggctttaact gcagagtcac atgttgacac tgacttaaca
14161 aagccttaca ttaagtggga tttgttaaaa tatgacttca cggaagagag gttaaaactc
14221 tttgaccgtt attttaaata ttgggatcag acataccacc caaattgtgt taactgtttg
14281 gatgacagat gcattctgca ttgtgcaaac tttaatgttt tattctctac agtgttccca
14341 cttacaagtt ttggaccact agtgagaaaa atatttgttg atggtgttcc atttgtagtt
14401 tcaactggat accacttcag agagctaggt gttgtacata atcaggatgt aaacttacat
14461 agctctagac ttagttttaa ggaattactt gtgtatgctg ctgaccctgc tatgcacgct
14521 gcttctggta atctattact agataaacgc actacgtgct tttcagtagc tgcacttact
```

Fig. 12B

continued

```
14581 aacaatgttg cttttcaaac tgtcaaaccc ggtaatttta acaaagactt ctatgacttt
14641 gctgtgtcta agggtttctt taaggaagga agttctgttg aattaaaaca cttcttcttt
14701 gctcaggatg gtaatgctgc tatcagcgat tatgactact atcgttataa tctaccaaca
14761 atgtgtgata tcagacaact actatttgta gttgaagttg ttgataagta ctttgattgt
14821 tacgatggtg gctgtattaa tgctaaccaa gtcatcgtca acaacctaga caaatcagct
14881 ggttttccat ttaataaatg gggtaaggct agactttatt atgattcaat gagttatgag
14941 gatcaagatg cacttttcgc atatacaaaa cgtaatgtca tccctactat aactcaaatg
15001 aatcttaagt atgccattag tgcaaagaat agagctcgca ccgtagctgg tgtctctatc
15061 tgtagtacta tgaccaatag acagtttcat caaaaattat tgaaatcaat agccgccact
15121 agaggagcta ctgtagtaat tggaacaagc aaattctatg gtggttggca caatatgtta
15181 aaaactgttt atagtgatgt agaaaaccct caccttatgg gttgggatta tcctaaatgt
15241 gatagagcca tgcctaacat gcttagaatt atggcctcac ttgttcttgc tcgcaaacat
15301 acaacgtgtt gtagcttgtc acaccgtttc tatagattag ctaatgagtg tgctcaagta
15361 ttgagtgaaa tggtcatgtg tggcggttca ctatatgtta aaccaggtgg aacctcatca
15421 ggagatgcca caactgctta tgctaatagt gtttttaaca tttgtcaagc tgtcacggcc
15481 aatgttaatg cacttttatc tactgatggt aacaaaattg ccgataagta tgtccgcaat
15541 ttacaacaca gactttatga gtgtctctat agaaatagag atgttgacac agactttgtg
15601 aatgagtttt acgcatattt gcgtaaacat ttctcaatga tgatactctc tgacgatgct
15661 gttgtgtgtt tcaatagcac ttatgcatct caaggtctag tggctagcat aaagaacttt
15721 aagtcagttc tttattatca aaacaatgtt tttatgtctg aagcaaaatg ttggactgag
```

Fig. 12B continued

```
15781 actgacctta ctaaaggacc tcatgaattt tgctctcaac atacaatgct agttaaacag
15841 ggtgatgatt atgtgtacct tccttaccca gatccatcaa gaatcctagg ggccggctgt
15901 tttgtagatg atatcgtaaa aacagatggt acacttatga ttgaacggtt cgtgtcttta
15961 gctatagatg cttacccact tactaaacat cctaatcagg agtatgctga tgtctttcat
16021 ttgtacttac aatacataag aaagctacat gatgagttaa caggacacat gttagacatg
16081 tattctgtta tgcttactaa tgataacact tcaaggtatt gggaacctga gttttatgag
16141 gctatgtaca caccgcatac agtcttacag gctgttgggg cttgtgttct ttgcaattca
16201 cagacttcat taagatgtgg tgcttgcata cgtagaccat tcttatgttg taaatgctgt
16261 tacgaccatg tcatatcaac atcacataaa ttagtcttgt ctgttaatcc gtatgtttgc
16321 aatgctccag gttgtgatgt cacagatgtg actcaacttt acttaggagg tatgagctat
16381 tattgtaaat cacataaacc acccattagt tttccattgt gtgctaatgg acaagttttt
16441 ggtttatata aaaatacatg tgttggtagc gataatgtta ctgactttaa tgcaattgca
16501 acatgtgact ggacaaatgc tggtgattac attttagcta acacctgtac tgaaagactc
16561 aagctttttg cagcagaaac gctcaaagct actgaggaga catttaaact gtcttatggt
16621 attgctactg tacgtgaagt gctgtctgac agagaattac atctttcatg ggaagttggt
16681 aaacctagac caccacttaa ccgaaattat gtctttactg gttatcgtgt aactaaaaac
16741 agtaaagtac aaataggaga gtacaccttt gaaaaaggtg actatggtga tgctgttgtt
16801 taccgaggta caacaactta caaattaaat gttggtgatt attttgtgct gacatcacat
16861 acagtaatgc cattaagtgc acctacacta gtgccacaag agcactatgt tagaattact
16921 ggcttatacc caacactcaa tatctcagat gagttttcta gcaatgttgc aaattatcaa
16981 aaggttggta tgcaaaagta ttctacactc cagggaccac ctggtactgg taagagtcat
```

Fig. 12B

continued

```
17041 tttgctattg gcctagctct ctactaccct tctgctcgca tagtgtatac agcttgctct
17101 catgccgctg ttgatgcact atgtgagaag gcattaaaat atttgcctat agataaatgt
17161 agtagaatta tacctgcacg tgctcgtgta gagtgttttg ataaattcaa agtgaattca
17221 acattagaac agtatgtctt ttgtactgta aatgcattgc ctgagacgac agcagatata
17281 gttgtctttg atgaaatttc aatggccaca aattatgatt tgagtgttgt caatgccaga
17341 ttacgtgcta agcactatgt gtacattggc gaccctgctc aattacctgc accacgcaca
17401 ttgctaacta agggcacact agaaccagaa tatttcaatt cagtgtgtag acttatgaaa
17461 actataggtc cagacatgtt cctcggaact tgtcggcgtt gtcctgctga aattgttgac
17521 actgtgagtg ctttggttta tgataataag cttaaagcac ataaagacaa atcagctcaa
17581 tgctttaaaa tgttttataa gggtgttatc acgcatgatg tttcatctgc aattaacagg
17641 ccacaaatag gcgtggtaag agaattcctt acacgtaacc ctgcttggag aaaagctgtc
17701 tttatttcac cttataattc acagaatgct gtagcctcaa agattttggg actaccaact
17761 caaactgttg attcatcaca gggctcagaa tatgactatg tcatattcac tcaaaccact
17821 gaaacagctc actcttgtaa tgtaaacaga tttaatgttg ctattaccag agcaaaagta
17881 ggcatacttt gcataatgtc tgatagagac ctttatgaca agttgcaatt tacaagtctt
17941 gaaattccac gtaggaatgt ggcaacttta caagctgaaa atgtaacagg actctttaaa
18001 gattgtagta aggtaatcac tgggttacat cctacacagg cacctacaca cctcagtgtt
18061 gacactaaat tcaaaaactga aggtttatgt gttgacgtac ctggcatacc taaggacatg
18121 acctatagaa gactcatctc tatgatgggt tttaaaatga attatcaagt taatggttac
18181 cctaacatgt ttatcacccg cgaagaagct ataagacatg tacgtgcatg gattggcttc
```

Fig. 12B continued

```
18241 gatgtcgagg ggtgtcatgc tactagagaa gctgttggta ccaatttacc tttacagcta
18301 ggtttttcta caggtgttaa cctagttgct gtacctacag gttatgttga tacacctaat
18361 aatacagatt tttccagagt tagtgctaaa ccaccgcctg gagatcaatt taaacacctc
18421 ataccactta tgtacaaagg acttccttgg aatgtagtgc gtataaagat tgtacaaatg
18481 ttaagtgaca cacttaaaaa 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 acaggtaacc tacaaagcaa ccatgatctg tattgtcaag tccatggtaa tgcacatgta
18781 gctagttgtg atgcaatcat gactaggtgt ctagctgtcc acgagtgctt tgttaagcgt
18841 gttgactgga ctattgaata tcctataatt ggtgatgaac tgaagattaa tgcggcttgt
18901 agaaaggttc aacacatggt tgttaaagct gcattattag cagacaaatt cccagttctt
18961 cacgacattg gtaaccctaa agctattaag tgtgtacctc aagctgatgt agaatggaag
19021 ttctatgatg cacagccttg tagtgacaaa gcttataaaa tagaagaatt attctattct
19081 tatgccacac attctgacaa attcacagat ggtgtatgcc tattttggaa ttgcaatgtc
19141 gatagatatc ctgctaattc cattgtttgt agatttgaca ctagagtgct atctaacctt
19201 aacttgcctg gttgtgatgg tggcagtttg tatgtaaata aacatgcatt ccacacacca
19261 gcttttgata aaagtgcttt tgttaattta aaacaattac catttttcta ttactctgac
19321 agtccatgtg agtctcatgg aaaacaagta gtgtcagata tagattatgt accactaaag
19381 tctgctacgt gtataacacg ttgcaattta ggtggtgctg tctgtagaca tcatgctaat
19441 gagtacagat tgtatctcga tgcttataac atgatgatct cagctggctt tagcttgtgg
```

Fig. 12B

continued

```
19501 gtttacaaac aatttgatac ttataacctc tggaacactt ttacaagact tcagagttta
19561 gaaaatgtgg cttttaatgt tgtaaataag ggacactttg atggacaaca gggtgaagta
19621 ccagtttcta tcattaataa cactgtttac acaaaagttg atggtgttga tgtagaattg
19681 tttgaaaata aaacaacatt acctgttaat gtagcatttg agctttgggc taagcgcaac
19741 attaaaccag taccagaggt gaaaatactc aataatttgg gtgtggacat tgctgctaat
19801 actgtgatct gggactacaa aagagatgct ccagcacata tatctactat tggtgtttgt
19861 tctatgactg acatagccaa gaaaccaact gaaacgattt gtgcaccact cactgtcttt
19921 tttgatggta gagttgatgg tcaagtagac ttatttagaa atgcccgtaa tggtgttctt
19981 attacagaag gtagtgttaa aggtttacaa ccatctgtag gtcccaaaca agctagtctt
20041 aatggagtca cattaattgg agaagccgta aaaacacagt tcaattatta taagaaagtt
20101 gatggtgttg tccaacaatt acctgaaact tactttactc agagtagaaa tttacaagaa
20161 tttaaaccca ggagtcaaat ggaaattgat ttcttagaat tagctatgga tgaattcatt
20221 gaacggtata aattagaagg ctatgccttc gaacatatcg tttatggaga ttttagtcat
20281 agtcagttag gtggtttaca tctactgatt ggactagcta aacgttttaa ggaatcacct
20341 tttgaattag aagattttat tcctatggac agtacagtta aaaactattt cataacagat
20401 gcgcaaacag gttcatctaa gtgtgtgtgt tctgttattg atttattact tgatgatttt
20461 gttgaaataa taaaatccca agatttatct gtagtttcta aggttgtcaa agtgactatt
20521 gactatacag aaatttcatt tatgctttgg tgtaaagatg gccatgtaga aacattttac
20581 ccaaaattac aatctagtca agcgtggcaa ccgggtgttg ctatgcctaa tctttacaaa
20641 atgcaaagaa tgctattaga aaagtgtgac cttcaaaatt atggtgatag tgcaacatta
20701 cctaaaggca taatgatgaa tgtcgcaaaa tatactcaac tgtgtcaata tttaaacaca
```

Fig. 12B

continued

```
20761 ttaacattag ctgtacccta taatatgaga gttatacatt ttggtgctgg ttctgataaa
20821 ggagttgcac caggtacagc tgttttaaga cagtggttgc ctacgggtac gctgcttgtc
20881 gattcagatc ttaatgactt tgtctctgat gcagattcaa ctttgattgg tgattgtgca
20941 actgtacata cagctaataa atgggatttc attattagtg atatgtacga ccctaagact
21001 aaaaatgtta caaaagaaaa tgactctaaa gagggttttt tcacttacat ttgtgggttt
21061 atacaacaaa agctagctct tggaggttcc gtggctataa agataacaga acattcttgg
21121 aatgctgatc tttataagct catgggacac ttcgcatggt ggacagcctt tgttactaat
21181 gtgaatgcgt catcatctga agcattttta attggatgta attatcttgg caaaccacgc
21241 gaacaaatag atggttatgt catgcatgca aattacatat tttggaggaa tacaaatcca
21301 attcagttgt cttcctattc tttatttgac atgagtaaat ttccccttaa attaaggggt
21361 actgctgtta tgtctttaaa agaaggtcaa atcaatgata tgattttatc tcttcttagt
21421 aaaggtagac ttataattag agaaaacaac agagttgtta tttctagtga tgttcttgtt
21481 aacaactaaa cgaacaatgt ttgtttttct tgttttattg ccactagtct ctagtcagtg
21541 tgttaatctt acaaccagaa ctcaattacc ccctgcatac actaattctt tcacacgtgg
21601 tgtttattac cctgacaaag ttttcagatc ctcagtttta cattcaactc aggacttgtt
21661 cttacctttc ttttccaatg ttacttggtt ccatgttatc tctgggacca atggtactaa
21721 gaggtttgat aaccctgtcc taccatttaa tgatggtgtt tattttgctt ccattgagaa
21781 gtctaacata ataagaggct ggatttttgg tactacttta gattcgaaga cccagtccct
21841 acttattgtt aataacgcta ctaatgttgt tattaaagtc tgtgaatttc aattttgtaa
21901 tgatccattt ttggaccaca aaaacaacaa aagttggatg gaaagtgagt tcagagttta
```

Fig. 12B continued

```
21961 ttctagtgcg aataattgca cttttgaata tgtctctcag cctttttctta tggaccttga
22021 aggaaaaacag ggtaatttca aaaatcttag ggaatttgtg tttaagaata ttgatggtta
22081 ttttaaaata tattctaagc acacgcctat tatagtgcgt gagccagaag atctccctca
22141 gggtttttcg gctttagaac cattggtaga tttgccaata ggtattaaca tcactaggtt
22201 tcaaacttta cttgctttac atagaagtta tttgactcct ggtgattctt cttcaggttg
22261 gacagctggt gctgcagctt attatgtggg ttatcttcaa cctaggactt ttctattaaa
22321 atataatgaa aatggaacca ttacagatgc tgtagactgt gcacttgacc ctctctcaga
22381 aacaaagtgt acgttgaaat ccttcactgt agaaaaagga atctatcaaa cttctaactt
22441 tagagtccaa ccaacagaat ctattgttag atttcctaat attacaaact tgtgcccttt
22501 tgatgaagtt tttaacgcca ccagatttgc atctgtttat gcttggaaca ggaagagaat
22561 cagcaactgt gttgctgatt attctgtcct atataatctc gcaccatttt tcacttttaa
22621 gtgttatgga gtgtctccta ctaaattaaa tgatctctgc tttactaatg tctatgcaga
22681 ttcatttgta attagaggtg atgaagtcag acaaatcgct ccagggcaaa ctggaaatat
22741 tgctgattat aattataaat taccagatga ttttacaggc tgcgttatag cttggaattc
22801 taacaagctt gattctaagg ttagtggtaa ttataattac ctgtatagat tgtttaggaa
22861 gtctaatctc aaacctttttg agagagatat ttcaactgaa atctatcagg ccggtaacaa
22921 accttgtaat ggtgttgcag gttttaattg ttactttcct ttacgatcat atagtttccg
22981 acccacttat ggtgttggtc accaaccata cagagtagta gtactttctt ttgaacttct
23041 acatgcacca gcaactgttt gtggacctaa aaagtctact aatttggtta aaaacaaatg
23101 tgtcaatttc aacttcaatg gtttaaaagg cacaggtgtt cttactgagt ctaacaaaaa
```

Fig. 12B continued

```
23161 gtttctgcct ttccaacaat ttggcagaga cattgctgac actactgatg ctgtccgtga
23221 tccacagaca cttgagattc ttgacattac accatgttct tttggtggtg tcagtgttat
23281 aacaccagga acaaatactt ctaaccaggt tgctgttctt tatcagggtg ttaactgcac
23341 agaagtccct gttgctattc atgcagatca acttactcct acttggcgtg tttattctac
23401 aggttctaat gtttttcaaa cacgtgcagg ctgtttaata ggggctgaat atgtcaacaa
23461 ctcatatgag tgtgacatac ccattggtgc aggtatatgc gctagttatc agactcagac
23521 taagtctcat cggcgggcac gtagtgtagc tagtcaatcc atcattgcct acactatgtc
23581 acttggtgca gaaaattcag ttgcttactc taataactct attgccatac ccacaaattt
23641 tactattagt gttaccacag aaattctacc agtgtctatg accaagacat cagtagattg
23701 tacaatgtac atttgtggtg attcaactga atgcagcaat cttttgttgc aatatggcag
23761 tttttgtaca caattaaaac gtgctttaac tggaatagct gttgaacaag acaaaaacac
23821 ccaagaagtt tttgcacaag tcaaacaaat ttacaaaaca ccaccaatta aatattttgg
23881 tggttttaat ttttcacaaa tattaccaga tccatcaaaa ccaagcaaga ggtcatttat
23941 tgaagatcta cttttcaaca aagtgacact tgcagatgct ggcttcatca aacaatatgg
24001 tgattgcctt ggtgatattg ctgctagaga cctcatttgt gcacaaaagt ttaaaggcct
24061 tactgttttg ccacctttgc tcacagatga aatgattgct caatacactt ctgcactgtt
24121 agcgggtaca atcacttctg gttggacctt tggtgcaggt gctgcattac aaataccatt
24181 tgctatgcaa atggcttata ggtttaatgg tattggagtt acacagaatg ttctctatga
24241 gaaccaaaaa ttgattgcca accaatttaa tagtgctatt ggcaaaattc aagactcact
24301 ttcttccaca gcaagtgcac ttggaaaact tcaagatgtg gtcaaccata atgcacaagc
24361 tttaaacacg cttgttaaac aacttagctc caaatttggt gcaatttcaa gtgttttaaa
```

Fig. 12B

continued

```
24421 tgatatcttt tcacgtcttg acaaagttga ggctgaagtg caaattgata ggttgatcac
24481 aggcagactt caaagtttgc agacatatgt gactcaacaa ttaattagag ctgcagaaat
24541 cagagcttct gctaatcttg ctgctactaa aatgtcagag tgtgtacttg gacaatcaaa
24601 aagagttgat ttttgtggaa agggctatca tcttatgtcc ttccctcagt cagcacctca
24661 tggtgtagtc ttcttgcatg tgacttatgt ccctgcacaa gaaaagaact tcacaactgc
24721 tcctgccatt tgtcatgatg gaaaagcaca ctttcctcgt gaaggtgtct ttgtttcaaa
24781 tggcacacac tggtttgtaa cacaaaggaa tttttatgaa ccacaaatca ttactacaga
24841 caacacattt gtgtctggta actgtgatgt tgtaatagga attgtcaaca acacagttta
24901 tgatcctttg caacctgaat tagattcatt caaggaggag ttagataaat attttaagaa
24961 tcatacatca ccagatgttg atttaggtga catctctggc attaatgctt cagttgtaaa
25021 cattcaaaaa gaaattgacc gcctcaatga ggttgccaag aatttaaatg aatctctcat
25081 cgatctccaa gaacttggaa agtatgagca gtatataaaa tggccatggt acatttggct
25141 aggttttata gctggcttga ttgccatagt aatggtgaca attatgcttt gctgtatgac
25201 cagttgctgt agttgtctca agggctgttg ttcttgtgga tcctgctgca aatttgatga
25261 agacgactct gagccagtgc tcaaaggagt caaattacat tacacataaa cgaacttatg
25321 gatttgttta tgagaatctt cacaattgga actgtaactt tgaagcaagg tgaaatcaag
25381 gatgctactc cttcagattt tgttcgcgct actgcaacga taccgataca agcctcactc
25441 cctttcggat ggcttattgt tggcgttgca cttcttgctg tttttcagag cgcttccaaa
25501 atcataactc tcaaaaagag atggcaacta gcactctcca agggtgttca ctttgtttgc
25561 aacttgctgt tgttgtttgt aacagtttac tcacaccttt tgctcgttgc tgctggcctt
```

Fig. 12B continued

```
25621 gaagcccctt ttctctatct ttatgcttta gtctacttct tgcagagtat aaactttgta
25681 agaataataa tgaggctttg gctttgctgg aaatgccgtt ccaaaaaccc attactttat
25741 gatgccaact atttttctttg ctggcatact aattgttacg actattgtat accttacaat
25801 agtgtaactt cttcaattgt cattacttca ggtgatggca caacaagtcc tatttctgaa
25861 catgactacc agattggtgg ttatactgaa aaatgggaat ctggagtaaa agactgtgtt
25921 gtattacaca gttacttcac ttcagactat taccagctgt actcaactca attgagtaca
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 gagataggta cgttaatagt taatagcgta
26221 cttcttttttc ttgctttcgt ggtattcttg ctagttacac tagccatcct tactgcgctt
26281 cgattgtgtg cgtactgctg caatattgtt aacgtgagtc ttgtaaaacc ttctttttac
26341 gtttactctc gtgttaaaaa tctgaattct tctagagttc ctgatcttct ggtctaaacg
26401 aactaaatat tatattagtt tttctgtttg gaactttaat tttagccatg gcaggttcca
26461 acggtactat taccgttgaa gagcttaaaa agctccttga agaatggaac ctagtaatag
26521 gtttcctatt ccttacatgg atttgtcttc tacaatttgc ctatgccaac aggaataggt
26581 ttttgtatat aattaagtta attttcctct ggctgttatg gccagtaact ttaacttgtt
26641 ttgtgcttgc tgctgtttac agaataaatt ggatcaccgg tggaattgct atcgcaatgg
26701 cttgtcttgt aggcttgatg tggctcagct acttcattgc ttctttcaga ctgtttgcgc
26761 gtacgcgttc catgtggtca ttcaatccag aaactaacat tcttctcaac gtgccactcc
26821 atggcactat tctgaccaga ccgcttctag aaagtgaact cgtaatcgga gctgtgatcc
```

Fig. 12B
continued

```
26881 ttcgtggaca tcttcgtatt gctggacacc atctaggacg ctgtgacatc aaggacctgc
26941 ctaaagaaat cactgttgct acatcacgaa cgctttctta ttacaaattg ggagcttcgc
27001 agcgtgtagc aggtgactca ggttttgctg catacagtcg ctacaggatt ggcaactata
27061 aattaaacac agaccattcc agtagcagtg acaatattgc tttgcttgta cagtaagtga
27121 caacagatgt ttcatctcgt tgactttcag gttactatag cagagatatt actaattatt
27181 atgcggactt ttaaagtttc catttggaat cttgattaca tcataaacct cataattaaa
27241 aatttatcta agtcactaac tgagaataaa tattctcaat tagatgaaga gcaaccaatg
27301 gagattgatt aaacgaacat gaaaattatt cttttcttgg cactgataac actcgctact
27361 tgtgagcttt atcactacca agagtgtgtt agaggtacaa cagtacttct aaaagaacct
27421 tgctcttctg gaacatacga gggcaattca ccatttcatc ctctagctga taacaaattt
27481 gcactgactt gctttagcac tcaatttgct tttgcttgtc ctgacggcgt aaaacacgtc
27541 tatcagttac gtgccagatc agtttcacct aaactgttca tcagacaaga ggaagttcaa
27601 gaactttact ctccaatttt tcttattgtt gcggcaatag tgtttataac actttgcttc
27661 acactcaaaa gaaagacaga atgattgaac tttcattaat tgacttctat ttgtgctttt
27721 tagcctttct gttattcctt gttttaatta tgcttattat cttttggttc tcacttgaac
27781 tgcaagatca taatgaaact tgtcacgcct aaacgaacat gaaatttctt gttttcttag
27841 gaatcatcac aactgtagct gcatttcacc aagaatgtag tttacagtca tgtactcaac
27901 atcaaccata tgtagttgat gacccgtgtc ctattcactt ctattctaaa tggtatatta
27961 gagtaggagc tagaaaatca gcacctttaa ttgaattgtg cgtggatgag gctggttcta
28021 aatcacccat tcagtacatc gatatcggta attatacagt ttcctgttta ccttttacaa
```

Fig. 12B continued

```
28081 ttaattgcca ggaacctaaa ttgggtagtc ttgtagtgcg ttgttcgttc tatgaagact
28141 ttttagagta tcatgacgtt cgtgttgttt tagatttcat ctaaacgaac aaacttaaat
28201 gtctgataat ggaccccaaa atcagcgaaa tgcactccgc attacgtttg gtggacctc
28261 agattcaact ggcagtaacc agaatggtgg ggcgcgatca aaacaacgtc ggccccaagg
28321 tttacccaat aatactgcgt cttggttcac cgctctcact caacatggca aggaagacct
28381 taaattccct cgaggacaag gcgttccaat taacaccaat agcagtccag atgaccaaat
28441 tggctactac cgaagagcta ccagacgaat tcgtggtggt gacggtaaaa tgaaagatct
28501 cagtccaaga tggtatttct actacctagg aactgggcca gaagctggac ttccctatgg
28561 tgctaacaaa gacggcatca tatgggttgc aactgaggga gccttgaata caccaaaaga
28621 tcacattggc acccgcaatc ctgctaacaa tgctgcaatc gtgctacaac ttcctcaagg
28681 aacaacattg ccaaaaggct tctacgcaga agggagcaga ggcggcagtc aagcctcttc
28741 tcgttcctca tcacgtagtc gcaacagttc aagaaattca actccaggca gcagtaaacg
28801 aacttctcct gctagaatgg ctggcaatgg cggtgatgct gctcttgctt tgctgctgct
28861 tgacagattg aaccagcttg agagcaaaat gtctggtaaa ggccaacaac aacaaggcca
28921 aactgtcact aagaaatctg ctgctgaggc ttctaagaag cctcggcaaa aacgtactgc
28981 cactaaagca tacaatgtaa cacaagcttt cggcagacgt ggtccagaac aaacccaagg
29041 aaattttggg gaccaggaac taatcagaca aggaactgat tacaaacatt ggccgcaaat
29101 tgcacaattt gcccccagcg cttcagcgtt cttcggaatg tcgcgcattg gcatggaagt
29161 cacaccttcg ggaacgtggt tgacctacac aggtgccatc aaattggatg acaaagatcc
29221 aaatttcaaa gatcaagtca ttttgctgaa taagcatatt gacgcataca aaacattccc
29281 accaacagag cctaaaaagg acaaaaagaa gaaggctgat gaaactcaag ccttaccgca
```

Fig. 12B

continued

```
29341 gagacagaag aaacagcaaa ctgtgactct tcttcctgct gcagatttgg atgatttctc
29401 caaacaattg caacaatcca tgagcagtgc tgactcaact caggcctaaa ctcatgcaga
29461 ccacacaagg cagatgggct atataaacgt tttcgctttt ccgtttacga tatatagtct
29521 actcttgtgc agaatgaatt ctcgtaacta catagcacaa gtagatgtag ttaactttaa
29581 tctcacatag caatctttaa tcagtgtgta acattaggga ggacttgaaa gagccaccac
29641 attttcaccg aggccacgcg gagtacgatc gagtgtacag tgaacaatgc tagggagagc
29701 tgcctatatg 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: PRJNA485481
KEYWORDS    RefSeq.
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
```

Fig. 12C continued

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 32015508

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 15680415

REFERENCE 3 (bases 29728 to 29768)

Fig. 12C continued

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)
PUBMED 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)
PUBMED 10482585
REFERENCE 5 (bases 1 to 29903)
CONSRTM 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

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.### 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

```
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:2697049"
                     /country="China"
                     /collection_date="Dec-2019"
     5'UTR           1..265
     gene            266..21555
                     /gene="ORF1ab"
                     /locus_tag="GU280_gp01"
```

Fig. 12C continued

```
                     /db_xref="GeneID:43740578"
     CDS             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="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
```

Fig. 12C continued

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

Fig. 12C continued

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

Fig. 12C continued

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

Fig. 12C continued

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

Fig. 12C continued

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

Fig. 12C continued

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

Fig. 12C continued

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

Fig. 12C continued

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD

EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR

LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA

NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV

DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ

TYHPNCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRE

LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ

TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI

RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ

Fig. 12C continued

DALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT

RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR

KHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ

AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM

ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS

QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH

PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV

LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD

VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW

TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP

Fig. 12C continued

RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH

TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK

SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF

KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ

LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK

AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA

VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD

RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT

EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG

CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP

Fig. 12C continued

LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL

CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA

HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK

FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL

FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ

LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN

MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT

VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY

KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG

SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK

Fig. 12C continued

PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP

FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV

TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD

SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP

TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG

FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL

IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE

GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

mat_peptide 266..805

/gene="ORF1ab"

/locus_tag="GU280_gp01"

/product="leader protein"

/note="nsp1; produced by both pp1a and pp1ab"

/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 nsp1; 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

```
                /note="nsp4B_TM; contains transmembrane domain 2 (TM2);
                produced by both pp1a and pp1ab"
                /protein_id="YP_009725300.1"
     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 SARSr-CoV homolog has been determined
                (Yang et al., 2003); produced by both pp1a and pp1ab"
                /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 pp1a and pp1ab"
                /protein_id="YP_009725302.1"
     mat_peptide    11843..12091
```

Fig. 12C continued

```
                /gene="ORF1ab"
                /locus_tag="GU280_gp01"
                /product="nsp7"
                /note="produced by both pp1a and pp1ab"
                /protein_id="YP_009725303.1"
mat_peptide     12092..12685
                /gene="ORF1ab"
                /locus_tag="GU280_gp01"
                /product="nsp8"
                /note="produced by both pp1a and pp1ab"
                /protein_id="YP_009725304.1"
mat_peptide     12686..13024
                /gene="ORF1ab"
                /locus_tag="GU280_gp01"
                /product="nsp9"
                /note="ssRNA-binding protein; produced by both pp1a and
                pp1ab"
                /protein_id="YP_009725305.1"
mat_peptide     13025..13441
                /gene="ORF1ab"
                /locus_tag="GU280_gp01"
```

Fig. 12C continued

```
                    /product="nsp10"
                    /note="nsp10_CysHis; formerly known as growth-factor-
like
                    protein (GFL); produced by both pp1a and pp1ab"
                    /protein_id="YP_009725306.1"
     mat_peptide    join(13442..13468,13468..16236)
                    /gene="ORF1ab"
                    /locus_tag="GU280_gp01"
                    /product="RNA-dependent RNA polymerase"
                    /note="nsp12; NiRAN and RdRp; produced by pp1ab 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_HEL1core; zinc-
binding
                    domain (ZD), NTPase/helicase domain (HEL), RNA
                    5'-triphosphatase; produced by pp1ab 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 pp1ab
                     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 pp1ab
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 pp1ab only"
                     /protein_id="YP_009725311.1"
```

Fig. 12C continued

```
     CDS             266..13483
                     /gene="ORF1ab"
                     /locus_tag="GU280_gp01"
                     /note="pp1a"
                     /codon_start=1
                     /product="ORF1a polyprotein"
                     /protein_id="YP_009725295.1"
                     /db_xref="GeneID:43740578"

/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
```

Fig. 12C continued

DFVKAICEFCGTENLIKEGAIICGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

Fig. 12C continued

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKITVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

Fig. 12C continued

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

Fig. 12C continued

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

Fig. 12C continued

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

Fig. 12C continued

AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

Fig. 12C continued

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

DAQSFLNGFAV"

Fig. 12C continued

```
mat_peptide     266..805
                /gene="ORF1ab"
                /locus_tag="GU280_gp01"
                /product="leader protein"
                /note="nsp1; 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 nsp1; 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 pp1a and pp1ab"
                /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 pp1a and pp1ab"
                /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 SARSr-CoV homolog has been determined
                (Yang et al., 2003); produced by both pp1a and pp1ab"
                /protein_id="YP_009742612.1"
mat_peptide     10973..11842
                /gene="ORF1ab"
```

Fig. 12C continued

```
                    /locus_tag="GU280_gp01"
                    /product="nsp6"
                    /note="nsp6_TM; putative transmembrane domain; produced
by
                    both pp1a and pp1ab"
                    /protein_id="YP_009742613.1"
     mat_peptide    11843..12091
                    /gene="ORF1ab"
                    /locus_tag="GU280_gp01"
                    /product="nsp7"
                    /note="produced by both pp1a and pp1ab"
                    /protein_id="YP_009742614.1"
     mat_peptide    12092..12685
                    /gene="ORF1ab"
                    /locus_tag="GU280_gp01"
                    /product="nsp8"
                    /note="produced by both pp1a and pp1ab"
                    /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 pp1a and
                    pp1ab"
                    /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 pp1a and pp1ab"
                    /protein_id="YP_009742617.1"
     mat_peptide    13442..13480
                    /gene="ORF1ab"
                    /locus_tag="GU280_gp01"
                    /product="nsp11"
                    /note="produced by pp1a 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

GWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY

SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ

GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFL

LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN

LCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF

TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN

Fig. 12C continued

YLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY

RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG

RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAI

HADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPR

RARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTM

YICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG

GFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFN

GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQN

VLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA

ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS

Fig. 12C continued

ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAH

FPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD

SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELG

KYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSE

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

NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ

LSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"
     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="MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCC
                     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="MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNR

FLYIIKLIFLWLLWPVTLACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRL
```

Fig. 12C continued

FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD

IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIA
LLVQ"

gene 27202..27387
/gene="ORF6"
/locus_tag="GU280_gp06"
/db_xref="GeneID:43740572"

CDS 27202..27387
/gene="ORF6"
/locus_tag="GU280_gp06"
/codon_start=1
/product="ORF6 protein"
/protein_id="YP_009724394.1"
/db_xref="GeneID:43740572"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSL
TENKYSQLDEEQPMEID"

gene 27394..27759
/gene="ORF7a"

Fig. 12C continued

```
                    /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

PFHPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFL
                    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="MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA"
     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

```
WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRC
                    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="MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQG

LPNNTASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMK

DLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQ
```

Fig. 12C continued

LPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAA

LALLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGR

RGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYT

GAIKLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTV

TLLPAADLDDFSKQLQQSMSSADSTQA"

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

```
                /translation="MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT"
stem_loop       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           29675..29903
stem_loop       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
```

Fig. 12C

continued

(NC_004718.3)"
/function="Coronavirus 3' stem-loop II-like motif (s2m)"

ORIGIN

```
      1 attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct
     61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
    121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc
    181 ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt
    241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acgagaaaac
    301 acacgtccaa ctcagtttgc ctgttttaca ggttcgcgac gtgctcgtac gtggctttgg
    361 agactccgtg gaggaggtct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg
    421 cttagtagaa gttgaaaaag gcgttttgcc tcaacttgaa cagccctatg tgttcatcaa
    481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact
    541 cgaaggcatt cagtacggtc gtagtggtga gacacttggt gtccttgtcc ctcatgtggg
    601 cgaaatacca gtggcttacc gcaaggttct tcttcgtaag aacggtaata aaggagctgg
    661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga
    721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttacccgtga
    781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg
    841 ccctgatggc taccctcttg agtgcattaa agaccttcta gcacgtgctg gtaaagcttc
    901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg
    961 tgaacatgag catgaaattg cttggtacac ggaacgttct gaaaagagct atgaattgca
```

Fig. 12C continued

```
1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa
1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aaaagaaaaa
1141 gcttgatggc tttatgggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg
1201 caaccaaatg tgcctttcaa ctctcatgaa gtgtgatcat tgtggtgaaa cttcatggca
1261 gacgggcgat tttgttaaag ccacttgcga attttgtggc actgagaatt tgactaaaga
1321 aggtgccact acttgtggtt acttacccca aaatgctgtt gttaaaattt attgtccagc
1381 atgtcacaat tcagaagtag gacctgagca tagtcttgcc gaataccata atgaatctgg
1441 cttgaaaacc attcttcgta agggtggtcg cactattgcc tttggaggct gtgtgttctc
1501 ttatgttggt tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg
1561 ttgtaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga
1621 aatactccaa aaagagaaag tcaacatcaa tattgttggt gactttaaac ttaatgaaga
1681 gatcgccatt attttggcat ctttttctgc ttccacaagt gcttttgtgg aaactgtgaa
1741 aggtttggat tataaagcat tcaaacaaat tgttgaatcc tgtggtaatt ttaaagttac
1801 aaaaggaaaa gctaaaaaag gtgcctggaa tattggtgaa cagaaatcaa tactgagtcc
1861 tctttatgca tttgcatcag aggctgctcg tgttgtacga tcaatttttct cccgcactct
1921 tgaaactgct caaaattctg tgcgtgtttt acagaaggcc gctataacaa tactagatgg
1981 aatttcacag tattcactga gactcattga tgctatgatg ttcacatctg atttggctac
2041 taacaatcta gttgtaatgg cctacattac aggtggtgtt gttcagttga cttcgcagtg
2101 gctaactaac atctttggca ctgtttatga aaaactcaaa cccgtccttg attggcttga
2161 agagaagttt aaggaaggtg tagagtttct tagagacggt tgggaaattg ttaaatttat
2221 ctcaacctgt gcttgtgaaa ttgtcggtgg acaaattgtc acctgtgcaa aggaaattaa
```

Fig. 12C
continued

```
2281 ggagagtgtt cagacattct ttaagcttgt aaataaattt ttggctttgt gtgctgactc
2341 tatcattatt ggtggagcta aacttaaagc cttgaatttа ggtgaaacat ttgtcacgca
2401 ctcaaaggga ttgtacagaa agtgtgttaa atccagagaa gaaactggcc tactcatgcc
2461 tctaaaagcc ccaaaagaaa ttatcttctt agagggagaa acacttccca cagaagtgtt
2521 aacagaggaa gttgtcttga aaactggtga tttacaacca ttagaacaac ctactagtga
2581 agctgttgaa gctccattgg ttggtacacc agtttgtatt aacgggctta tgttgctcga
2641 aatcaaagac acagaaaagt actgtgccct tgcacctaat atgatggtaa caaacaatac
2701 cttcacactc aaaggcggtg caccaacaaa ggttactttt ggtgatgaca ctgtgataga
2761 agtgcaaggt tacaagagtg tgaatatcac ttttgaactt gatgaaagga ttgataaagt
2821 acttaatgag aagtgctctg cctatacagt tgaactcggt acagaagtaa atgagttcgc
2881 ctgtgttgtg gcagatgctg tcataaaaac tttgcaacca gtatctgaat tacttacacc
2941 actgggcatt gatttagatg agtggagtat ggctacatac tacttatttg atgagtctgg
3001 tgagtttaaa ttggcttcac atatgtattg ttctttctac cctccagatg aggatgaaga
3061 agaaggtgat tgtgaagaag aagagtttga gccatcaact caatatgagt atggtactga
3121 agatgattac caaggtaaac ctttggaatt tggtgccact tctgctgctc ttcaacctga
3181 agaagagcaa gaagaagatt ggttagatga tgatagtcaa caaactgttg gtcaacaaga
3241 cggcagtgag gacaatcaga caactactat tcaaacaatt gttgaggttc aacctcaatt
3301 agagatggaa cttacaccag ttgttcagac tattgaagtg aatagtttta gtggttattt
3361 aaaacttact gacaatgtat acattaaaaa tgcagacatt gtggaagaag ctaaaaaggt
3421 aaaaccaaca gtggttgtta atgcagccaa tgtttacctt aaacatggag gaggtgttgc
```

Fig. 12C continued

```
3481 aggagcctta aataaggcta ctaacaatgc catgcaagtt gaatctgatg attacatagc
3541 tactaatgga ccacttaaag tgggtggtag ttgtgtttta agcggacaca atcttgctaa
3601 acactgtctt catgttgtcg gcccaaatgt taacaaaggt gaagacattc aacttcttaa
3661 gagtgcttat gaaaatttta atcagcacga agttctactt gcaccattat tatcagctgg
3721 tatttttggt gctgacccta tacattcttt aagagtttgt gtagatactg ttcgcacaaa
3781 tgtctactta gctgtctttg ataaaaatct ctatgacaaa cttgtttcaa gctttttgga
3841 aatgaagagt gaaaagcaag ttgaacaaaa gatcgctgag attcctaaag aggaagttaa
3901 gccatttata actgaaagta aaccttcagt tgaacagaga aaacaagatg ataagaaaat
3961 caaagcttgt gttgaagaag ttacaacaac tctggaagaa actaagttcc tcacagaaaa
4021 cttgttactt tatattgaca ttaatggcaa tcttcatcca gattctgcca ctcttgttag
4081 tgacattgac atcactttct taaagaaaga tgctccatat atagtgggtg atgttgttca
4141 agagggtgtt ttaactgctg tggttatacc tactaaaaag gctggtggca ctactgaaat
4201 gctagcgaaa gctttgagaa aagtgccaac agacaattat ataaccactt acccgggtca
4261 gggtttaaat ggttacactg tagaggaggc aaagacagtg cttaaaaagt gtaaaagtgc
4321 cttttacatt ctaccatcta ttatctctaa tgagaagcaa gaaattcttg gaactgtttc
4381 ttggaatttg cgagaaatgc ttgcacatgc agaagaaaca cgcaaattaa tgcctgtctg
4441 tgtggaaact aaagccatag tttcaactat acagcgtaaa tataagggta ttaaaataca
4501 agagggtgtg gttgattatg gtgctagatt ttacttttac accagtaaaa caactgtagc
4561 gtcacttatc aacacactta acgatctaaa tgaaactctt gttacaatgc cacttggcta
4621 tgtaacacat ggcttaaatt tggaagaagc tgctcggtat atgagatctc tcaaagtgcc
4681 agctacagtt tctgtttctt cacctgatgc tgttacagcg tataatggtt atcttacttc
```

Fig. 12C continued

```
4741 ttcttctaaa acacctgaag aacattttat tgaaaccatc tcacttgctg gttcctataa
4801 agattggtcc tattctggac aatctacaca actaggtata gaatttctta agagaggtga
4861 taaaagtgta tattacacta gtaatcctac cacattccac ctagatggtg aagttatcac
4921 ctttgacaat cttaagacac ttctttcttt gagagaagtg aggactatta aggtgtttac
4981 aacagtagac aacattaacc tccacacgca agttgtggac atgtcaatga catatggaca
5041 acagtttggt ccaacttatt tggatggagc tgatgttact aaaataaaac ctcataattc
5101 acatgaaggt aaaacatttt atgttttacc taatgatgac actctacgtg ttgaggcttt
5161 tgagtactac cacacaactg atcctagttt tctgggtagg tacatgtcag cattaaatca
5221 cactaaaaag tggaaatacc cacaagttaa tggtttaact tctattaaat gggcagataa
5281 caactgttat cttgccactg cattgttaac actccaacaa atagagttga agtttaatcc
5341 acctgctcta caagatgctt attacagagc aagggctggt gaagctgcta acttttgtgc
5401 acttatctta gcctactgta ataagacagt aggtgagtta ggtgatgtta gagaaacaat
5461 gagttacttg tttcaacatg ccaatttaga ttcttgcaaa agagtcttga acgtggtgtg
5521 taaaacttgt ggacaacagc agacaaccct taagggtgta gaagctgtta tgtacatggg
5581 cacactttct tatgaacaat ttaagaaagg tgttcagata ccttgtacgt gtggtaaaca
5641 agctacaaaa tatctagtac aacaggagtc acctttttgtt atgatgtcag caccacctgc
5701 tcagtatgaa cttaagcatg gtacatttac ttgtgctagt gagtacactg gtaattacca
5761 gtgtggtcac tataaacata taacttctaa agaaactttg tattgcatag acggtgcttt
5821 acttacaaag tcctcagaat acaaaggtcc tattacggat gttttctaca aagaaaacag
5881 ttacacaaca accataaaac cagttactta taaattggat ggtgttgttt gtacagaaat
```

Fig. 12C continued

```
5941 tgaccctaag ttggacaatt attataagaa agacaattct tatttcacag agcaaccaat
6001 tgatcttgta ccaaaccaac catatccaaa cgcaagcttc gataatttta agtttgtatg
6061 tgataatatc aaatttgctg atgatttaaa ccagttaact ggttataaga aacctgcttc
6121 aagagagctt aaagttacat ttttccctga cttaaatggt gatgtggtgg ctattgatta
6181 taaacactac acaccctctt ttaagaaagg agctaaattg ttacataaac ctattgtttg
6241 gcatgttaac aatgcaacta ataaagccac gtataaacca aatacctggt gtatacgttg
6301 tctttggagc acaaaaccag ttgaaacatc aaattcgttt gatgtactga agtcagagga
6361 cgcgcaggga atggataatc ttgcctgcga agatctaaaa ccagtctctg aagaagtagt
6421 ggaaaatcct accatacaga aagacgttct tgagtgtaat gtgaaaacta ccgaagttgt
6481 aggagacatt atacttaaac cagcaaataa tagtttaaaa attacagaag aggttggcca
6541 cacagatcta atggctgctt atgtagacaa ttctagtctt actattaaga aacctaatga
6601 attatctaga gtattaggtt tgaaaaccct tgctactcat ggtttagctg ctgttaatag
6661 tgtcccttgg gatactatag ctaattatgc taagcctttt cttaacaaag ttgttagtac
6721 aactactaac atagttacac ggtgtttaaa ccgtgtttgt actaattata tgccttattt
6781 ctttacttta ttgctacaat tgtgtacttt tactagaagt acaaattcta gaattaaagc
6841 atctatgccg actactatag caaagaatac tgttaagagt gtcggtaaat tttgtctaga
6901 ggcttcattt aattatttga agtcacctaa ttttttctaaa ctgataaata ttataatttg
6961 gtttttacta ttaagtgttt gcctaggttc tttaatctac tcaaccgctg ctttaggtgt
7021 tttaatgtct aatttaggca tgccttctta ctgtactggt tacagagaag gctatttgaa
7081 ctctactaat gtcactattg caacctactg tactggttct ataccttgta gtgtttgtct
7141 tagtggttta gattctttag acacctatcc ttctttagaa actatacaaa ttaccatttc
```

Fig. 12C continued

```
7201 atcttttaaa tgggatttaa ctgcttttgg cttagttgca gagtggtttt tggcatatat
7261 tcttttcact aggtttttct atgtacttgg attggctgca atcatgcaat tgtttttcag
7321 ctattttgca gtacatttta ttagtaattc ttggcttatg tggttaataa ttaatcttgt
7381 acaaatggcc ccgatttcag ctatggttag aatgtacatc ttctttgcat cattttatta
7441 tgtatggaaa agttatgtgc atgttgtaga cggttgtaat tcatcaactt gtatgatgtg
7501 ttacaaacgt aatagagcaa caagagtcga atgtacaact attgttaatg gtgttagaag
7561 gtccttttat gtctatgcta atggaggtaa aggcttttgc aaactacaca attggaattg
7621 tgttaattgt gatacattct gtgctggtag tacatttatt agtgatgaag ttgcgagaga
7681 cttgtcacta cagtttaaaa gaccaataaa tcctactgac cagtcttctt acatcgttga
7741 tagtgttaca gtgaagaatg gttccatcca tctttacttt gataaagctg gtcaaaagac
7801 ttatgaaaga cattctctct ctcattttgt taacttagac aacctgagag ctaataacac
7861 taaaggttca ttgcctatta atgttatagt ttttgatggt aaatcaaaat gtgaagaatc
7921 atctgcaaaa tcagcgtctg tttactacag tcagcttatg tgtcaaccta tactgttact
7981 agatcaggca ttagtgtctg atgttggtga tagtgcggaa gttgcagtta aaatgtttga
8041 tgcttacgtt aatacgtttt catcaacttt taacgtacca atggaaaaac tcaaaacact
8101 agttgcaact gcagaagctg aacttgcaaa gaatgtgtcc ttagacaatg tcttatctac
8161 ttttatttca gcagctcggc aagggtttgt tgattcagat gtagaaacta aagatgttgt
8221 tgaatgtctt aaattgtcac atcaatctga catagaagtt actggcgata gttgtaataa
8281 ctatatgctc acctataaca aagttgaaaa catgacaccc cgtgaccttg gtgcttgtat
8341 tgactgtagt gcgcgtcata ttaatgcgca ggtagcaaaa agtcacaaca ttgctttgat
```

Fig. 12C continued

```
8401 atggaacgtt aaagatttca tgtcattgtc tgaacaacta cgaaaacaaa tacgtagtgc
8461 tgctaaaaag aataacttac cttttaagtt gacatgtgca actactagac aagttgttaa
8521 tgttgtaaca acaaagatag cacttaaggg tggtaaaatt gttaataatt ggttgaagca
8581 gttaattaaa gttacacttg tgttcctttt tgttgctgct attttctatt taataacacc
8641 tgttcatgtc atgtctaaac atactgactt ttcaagtgaa atcataggat acaaggctat
8701 tgatggtggt gtcactcgtg acatagcatc tacagatact tgttttgcta acaaacatgc
8761 tgattttgac acatggttta gccagcgtgg tggtagttat actaatgaca aagcttgccc
8821 attgattgct gcagtcataa caagagaagt gggttttgtc gtgcctggtt tgcctggcac
8881 gatattacgc acaactaatg gtgacttttt gcatttctta cctagagttt ttagtgcagt
8941 tggtaacatc tgttacacac catcaaaact tatagagtac actgactttg caacatcagc
9001 ttgtgttttg gctgctgaat gtacaatttt taaagatgct tctggtaagc cagtaccata
9061 ttgttatgat accaatgtac tagaaggttc tgttgcttat gaaagtttac gccctgacac
9121 acgttatgtg ctcatggatg gctctattat tcaatttcct aacacctacc ttgaaggttc
9181 tgttagagtg gtaacaactt ttgattctga gtactgtagg cacggcactt gtgaaagatc
9241 agaagctggt gtttgtgtat ctactagtgg tagatgggta cttaacaatg attattacag
9301 atctttacca ggagttttct gtggtgtaga tgctgtaaat ttacttacta atatgtttac
9361 accactaatt caacctattg gtgctttgga catatcagca tctatagtag ctggtggtat
9421 tgtagctatc gtagtaacat gccttgccta ctattttatg aggtttagaa gagcttttgg
9481 tgaatacagt catgtagttg cctttaatac tttactattc cttatgtcat tcactgtact
9541 ctgtttaaca ccagtttact cattcttacc tggtgtttat tctgttattt acttgtactt
9601 gacattttat cttactaatg atgtttcttt tttagcacat attcagtgga tggttatgtt
```

Fig. 12C
continued

```
 9661 cacaccttta gtacctttct ggataacaat tgcttatatc atttgtattt ccacaaagca
 9721 tttctattgg ttctttagta attacctaaa gagacgtgta gtctttaatg gtgtttcctt
 9781 tagtactttt gaagaagctg cgctgtgcac ctttttgtta aataaagaaa tgtatctaaa
 9841 gttgcgtagt gatgtgctat tacctcttac gcaatataat agatacttag ctctttataa
 9901 taagtacaag tattttagtg gagcaatgga tacaactagc tacagagaag ctgcttgttg
 9961 tcatctcgca aaggctctca atgacttcag taactcaggt tctgatgttc tttaccaacc
10021 accacaaacc tctatcacct cagctgtttt gcagagtggt tttagaaaaa tggcattccc
10081 atctggtaaa gttgagggtt gtatggtaca agtaacttgt ggtacaacta cacttaacgg
10141 tctttggctt gatgacgtag tttactgtcc aagacatgtg atctgcacct ctgaagacat
10201 gcttaaccct aattatgaag atttactcat tcgtaagtct aatcataatt tcttggtaca
10261 ggctggtaat gttcaactca gggttattgg acattctatg caaaattgtg tacttaagct
10321 taaggttgat acagccaatc ctaagacacc taagtataag tttgttcgca ttcaaccagg
10381 acagactttt tcagtgttag cttgttacaa tggttcacca tctggtgttt accaatgtgc
10441 tatgaggccc aatttcacta ttaagggttc attccttaat ggttcatgtg gtagtgttgg
10501 ttttaacata gattatgact gtgtctcttt ttgttacatg caccatatgg aattaccaac
10561 tggagttcat gctggcacag acttagaagg taacttttat ggacctttg ttgacaggca
10621 aacagcacaa gcagctggta cggacacaac tattacagtt aatgttttag cttggttgta
10681 cgctgctgtt ataaatggag acaggtggtt tctcaatcga tttaccacaa ctcttaatga
10741 ctttaacctt gtggctatga agtacaatta tgaacctcta acacaagacc atgttgacat
10801 actaggacct ctttctgctc aaactggaat tgccgtttta gatatgtgtg cttcattaaa
```

Fig. 12C continued

```
10861 agaattactg caaaatggta tgaatggacg taccatattg ggtagtgctt tattagaaga
10921 tgaatttaca ccttttgatg ttgttagaca atgctcaggt gttactttcc aaagtgcagt
10981 gaaaagaaca atcaagggta cacaccactg gttgttactc acaattttga cttcactttt
11041 agttttagtc cagagtactc aatggtcttt gttcttttct ttgtatgaaa atgccttttt
11101 acctttttgct atgggtatta ttgctatgtc tgcttttgca atgatgtttg tcaaacataa
11161 gcatgcattt ctctgtttgt ttttgttacc ttctcttgcc actgtagctt attttaatat
11221 ggtctatatg cctgctagtt gggtgatgcg tattatgaca tggttggata tggttgatac
11281 tagtttgtct ggttttaagc taaaagactg tgttatgtat gcatcagctg tagtgttact
11341 aatccttatg acagcaagaa ctgtgtatga tgatggtgct aggagagtgt ggacacttat
11401 gaatgtcttg acactcgttt ataaagttta ttatggtaat gctttagatc aagccatttc
11461 catgtgggct cttataatct ctgttacttc taactactca ggtgtagtta caactgtcat
11521 gtttttggcc agaggtattg tttttatgtg tgttgagtat tgccctattt tcttcataac
11581 tggtaataca cttcagtgta taatgctagt ttattgtttc ttaggctatt tttgtacttg
11641 ttactttggc ctcttttgtt tactcaaccg ctactttaga ctgactcttg gtgtttatga
11701 ttacttagtt tctacacagg agtttagata tatgaattca cagggactac tcccacccaa
11761 gaatagcata gatgccttca aactcaacat taaattgttg ggtgttggtg gcaaaccttg
11821 tatcaaagta gccactgtac agtctaaaat gtcagatgta aagtgcacat cagtagtctt
11881 actctcagtt ttgcaacaac tcagagtaga atcatcatct aaattgtggg ctcaatgtgt
11941 ccagttacac aatgacattc tcttagctaa agatactact gaagcctttg aaaaaatggt
12001 ttcactactt tctgttttgc tttccatgca gggtgctgta gacataaaca agctttgtga
12061 agaaatgctg gacaacaggg caaccttaca agctatagcc tcagagttta gttcccttcc
```

Fig. 12C

continued

```
12121 atcatatgca gcttttgcta ctgctcaaga agcttatgag caggctgttg ctaatggtga
12181 ttctgaagtt gttcttaaaa agttgaagaa gtctttgaat gtggctaaat ctgaatttga
12241 ccgtgatgca gccatgcaac gtaagttgga aaagatggct gatcaagcta tgacccaaat
12301 gtataaacag gctagatctg aggacaagag ggcaaaagtt actagtgcta tgcagacaat
12361 gcttttcact atgcttagaa agttggataa tgatgcactc aacaacatta tcaacaatgc
12421 aagagatggt tgtgttccct tgaacataat acctcttaca acagcagcca aactaatggt
12481 tgtcatacca gactataaca catataaaaa tacgtgtgat ggtacaacat ttacttatgc
12541 atcagcattg tgggaaatcc aacaggttgt agatgcagat agtaaaattg ttcaacttag
12601 tgaaattagt atggacaatt cacctaattt agcatggcct cttattgtaa cagctttaag
12661 ggccaattct gctgtcaaat tacagaataa tgagcttagt cctgttgcac tacgacagat
12721 gtcttgtgct gccggtacta cacaaactgc ttgcactgat gacaatgcgt tagcttacta
12781 caacacaaca aagggaggta ggtttgtact tgcactgtta tccgatttac aggatttgaa
12841 atgggctaga ttccctaaga gtgatggaac tggtactatc tatacagaac tggaaccacc
12901 ttgtaggttt gttacagaca cacctaaagg tcctaaagtg aagtatttat actttattaa
12961 aggattaaac aacctaaata gaggtatggt acttggtagt ttagctgcca cagtacgtct
13021 acaagctggt aatgcaacag aagtgcctgc caattcaact gtattatctt tctgtgcttt
13081 tgctgtagat gctgctaaag cttacaaaga ttatctagct agtgggggac aaccaatcac
13141 taattgtgtt aagatgttgt gtacacacac tggtactggt caggcaataa cagttacacc
13201 ggaagccaat atggatcaag aatcctttgg tggtgcatcg tgttgtctgt actgccgttg
13261 ccacatagat catccaaatc ctaaaggatt ttgtgactta aaaggtaagt atgtacaaat
```

Fig. 12C continued

```
13321 acctacaact tgtgctaatg accctgtggg ttttacactt aaaaaacacag tctgtaccgt
13381 ctgcggtatg tggaaaggtt atggctgtag ttgtgatcaa ctccgcgaac ccatgcttca
13441 gtcagctgat gcacaatcgt ttttaaacgg gtttgcggtg taagtgcagc ccgtcttaca
13501 ccgtgcggca caggcactag tactgatgtc gtatacaggg cttttgacat ctacaatgat
13561 aaagtagctg gttttgctaa attcctaaaa actaattgtt gtcgcttcca agaaaaggac
13621 gaagatgaca atttaattga ttcttacttt gtagttaaga gacacacttt ctctaactac
13681 caacatgaag aaacaattta taatttactt aaggattgtc cagctgttgc taaacatgac
13741 ttctttaagt ttagaataga cggtgacatg gtaccacata tatcacgtca acgtcttact
13801 aaatacacaa tggcagacct cgtctatgct ttaaggcatt ttgatgaagg taattgtgac
13861 acattaaaag aaatacttgt cacatacaat tgttgtgatg atgattattt caataaaaag
13921 gactggtatg attttgtaga aaacccagat atattacgcg tatacgccaa cttaggtgaa
13981 cgtgtacgcc aagctttgtt aaaaacagta caattctgtg atgccatgcg aaatgctggt
14041 attgttggtg tactgacatt agataatcaa gatctcaatg gtaactggta tgatttcggt
14101 gatttcatac aaaccacgcc aggtagtgga gttcctgttg tagattctta ttattcattg
14161 ttaatgccta tattaacctt gaccagggct ttaactgcag agtcacatgt tgacactgac
14221 ttaacaaagc cttacattaa gtgggatttg ttaaaatatg acttcacgga agagaggtta
14281 aaactctttg accgttattt taaatattgg gatcagacat accacccaaa ttgtgttaac
14341 tgtttggatg acagatgcat tctgcattgt gcaaacttta atgttttatt ctctacagtg
14401 ttcccaccta caagttttgg accactagtg agaaaaatat ttgttgatgg tgttccattt
14461 gtagtttcaa ctggatacca cttcagagag ctaggtgttg tacataatca ggatgtaaac
14521 ttacatagct ctagacttag ttttaaggaa ttacttgtgt atgctgctga ccctgctatg
```

Fig. 12C

continued

```
14581 cacgctgctt ctggtaatct attactagat aaacgcacta cgtgcttttc 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 tttgtagttg aagttgttga taagtacttt
14881 gattgttacg atggtggctg tattaatgct aaccaagtca tcgtcaacaa cctagacaaa
14941 tcagctggtt ttccatttaa taaatggggt aaggctagac tttattatga ttcaatgagt
15001 tatgaggatc aagatgcact tttcgcatat acaaaacgta atgtcatccc tactataact
15061 caaatgaatc ttaagtatgc cattagtgca aagaatagag ctcgcaccgt agctggtgtc
15121 tctatctgta gtactatgac caatagacag tttcatcaaa aattattgaa atcaatagcc
15181 gccactagag gagctactgt agtaattgga acaagcaaat tctatggtgg ttggcacaac
15241 atgttaaaaa ctgtttatag tgatgtagaa aaccctcacc ttatgggttg ggattatcct
15301 aaatgtgata gagccatgcc taacatgctt agaattatgg cctcacttgt tcttgctcgc
15361 aaacatacaa cgtgttgtag cttgtcacac cgtttctata gattagctaa tgagtgtgct
15421 caagtattga gtgaaatggt catgtgtggc ggttcactat atgttaaacc aggtggaacc
15481 tcatcaggag atgccacaac tgcttatgct aatagtgttt ttaacatttg tcaagctgtc
15541 acggccaatg ttaatgcact tttatctact gatggtaaca aaattgccga taagtatgtc
15601 cgcaatttac aacacagact ttatgagtgt ctctatagaa atagagatgt tgacacagac
15661 tttgtgaatg agttttacgc atatttgcgt aaacatttct caatgatgat actctctgac
15721 gatgctgttg tgtgtttcaa tagcacttat gcatctcaag gtctagtggc tagcataaag
```

Fig. 12C continued

15781 aactttaagt cagttcttta ttatcaaaac aatgttttta tgtctgaagc aaaatgttgg
15841 actgagactg accttactaa aggacctcat gaattttgct ctcaacatac aatgctagtt
15901 aaacagggtg atgattatgt gtaccttcct tacccagatc catcaagaat cctaggggcc
15961 ggctgttttg tagatgatat cgtaaaaaca gatggtacac ttatgattga acggttcgtg
16021 tctttagcta tagatgctta cccacttact aaacatccta atcaggagta tgctgatgtc
16081 tttcatttgt acttacaata cataagaaag ctacatgatg agttaacagg acacatgtta
16141 gacatgtatt ctgttatgct tactaatgat aacacttcaa ggtattggga acctgagttt
16201 tatgaggcta tgtacacacc gcatacagtc ttacaggctg ttggggcttg tgtttctttgc
16261 aattcacaga cttcattaag atgtggtgct tgcatacgta gaccattctt atgttgtaaa
16321 tgctgttacg accatgtcat atcaacatca cataaattag tcttgtctgt taatccgtat
16381 gtttgcaatg ctccaggttg tgatgtcaca gatgtgactc aactttactt aggaggtatg
16441 agctattatt gtaaatcaca taaaccaccc attagttttc cattgtgtgc taatggacaa
16501 gtttttggtt tatataaaaa tacatgtgtt ggtagcgata atgttactga ctttaatgca
16561 attgcaacat gtgactggac aaatgctggt gattacattt tagctaacac ctgtactgaa
16621 agactcaagc tttttgcagc agaaacgctc aaagctactg aggagacatt taaactgtct
16681 tatggtattg ctactgtacg tgaagtgctg tctgacagag aattacatct ttcatgggaa
16741 gttggtaaac ctagaccacc acttaaccga aattatgtct ttactggtta tcgtgtaact
16801 aaaaaacagta aagtacaaat aggagagtac acctttgaaa aaggtgacta tggtgatgct
16861 gttgtttacc gaggtacaac aacttacaaa ttaaatgttg gtgattattt tgtgctgaca
16921 tcacatacag taatgccatt aagtgcacct acactagtgc cacaagagca ctatgttaga
16981 attactggct tatacccaac actcaatatc tcagatgagt tttctagcaa tgttgcaaat

Fig. 12C
continued

```
17041 tatcaaaagg ttggtatgca aaagtattct acactccagg gaccacctgg tactggtaag
17101 agtcatttttg ctattggcct agctctctac tacccttctg ctcgcatagt gtatacagct
17161 tgctctcatg ccgctgttga tgcactatgt gagaaggcat taaaatattt gcctatagat
17221 aaatgtagta gaattatacc tgcacgtgct cgtgtagagt gttttgataa attcaaagtg
17281 aattcaacat tagaacagta tgtcttttgt actgtaaatg cattgcctga gacgacagca
17341 gatatagttg tctttgatga aatttcaatg gccacaaatt atgatttgag tgttgtcaat
17401 gccagattac gtgctaagca ctatgtgtac attggcgacc ctgctcaatt acctgcacca
17461 cgcacattgc taactaaggg cacactagaa ccagaatatt tcaattcagt gtgtagactt
17521 atgaaaacta taggtccaga catgttcctc ggaacttgtc ggcgttgtcc tgctgaaatt
17581 gttgacactg tgagtgcttt ggtttatgat aataagctta aagcacataa agacaaatca
17641 gctcaatgct ttaaaatgtt ttataagggt gttatcacgc atgatgtttc atctgcaatt
17701 aacaggccac aaataggcgt ggtaagagaa ttccttacac gtaaccctgc ttggagaaaa
17761 gctgtcttta tttcacctta taattcacag aatgctgtag cctcaaagat tttgggacta
17821 ccaactcaaa ctgttgattc atcacagggc tcagaatatg actatgtcat attcactcaa
17881 accactgaaa cagctcactc ttgtaatgta aacagattta atgttgctat taccagagca
17941 aaagtaggca tactttgcat aatgtctgat agagaccttt atgacaagtt gcaatttaca
18001 agtcttgaaa ttccacgtag gaatgtggca actttacaag ctgaaaatgt aacaggactc
18061 tttaaagatt gtagtaaggt aatcactggg ttacatccta cacaggcacc tacacacctc
18121 agtgttgaca ctaaattcaa aactgaaggt ttatgtgttg acatacctgg catacctaag
18181 gacatgacct atagaagact catctctatg atgggtttta aaatgaatta tcaagttaat
```

Fig. 12C continued

```
18241 ggttacccta acatgtttat cacccgcgaa gaagctataa gacatgtacg tgcatggatt
18301 ggcttcgatg tcgaggggtg tcatgctact agagaagctg ttggtaccaa tttaccttta
18361 cagctaggtt tttctacagg tgttaaccta gttgctgtac ctacaggtta tgttgataca
18421 cctaataata cagatttttc cagagttagt gctaaaccac cgcctggaga tcaatttaaa
18481 cacctcatac cacttatgta caaaggactt ccttggaatg tagtgcgtat aaagattgta
18541 caaatgttaa gtgacacact taaaaatctc tctgacagag tcgtatttgt cttatgggca
18601 catggctttg agttgacatc tatgaagtat tttgtgaaaa taggacctga gcgcacctgt
18661 tgtctatgtg atagacgtgc cacatgcttt tccactgctt cagacactta tgcctgttgg
18721 catcattcta ttggatttga ttacgtctat aatccgttta tgattgatgt tcaacaatgg
18781 ggttttacag gtaacctaca aagcaaccat gatctgtatt gtcaagtcca tggtaatgca
18841 catgtagcta gttgtgatgc aatcatgact aggtgtctag ctgtccacga gtgctttgtt
18901 aagcgtgttg actggactat tgaatatcct ataattggtg atgaactgaa gattaatgcg
18961 gcttgtagaa aggttcaaca catggttgtt aaagctgcat tattagcaga caaattccca
19021 gttcttcacg acattggtaa ccctaaagct attaagtgtg tacctcaagc tgatgtagaa
19081 tggaagttct atgatgcaca gccttgtagt gacaaagctt ataaaatagа agaattattc
19141 tattcttatg ccacacattc tgacaaattc acagatggtg tatgcctatt ttggaattgc
19201 aatgtcgata gatatcctgc taattccatt gtttgtagat ttgacactag agtgctatct
19261 aaccttaact tgcctggttg tgatggtggc agtttgtatg taaataaaca tgcattccac
19321 acaccagctt ttgataaaag tgcttttgtt aatttaaaac aattaccatt tttctattac
19381 tctgacagtc catgtgagtc tcatggaaaa caagtagtgt cagatataga ttatgtacca
19441 ctaaagtctg ctacgtgtat aacacgttgc aatttaggtg gtgctgtctg tagacatcat
```

Fig. 12C

continued

```
19501 gctaatgagt acagattgta tctcgatgct tataacatga tgatctcagc tggctttagc
19561 ttgtgggttt acaaacaatt tgatacttat aacctctgga acacttttac aagacttcag
19621 agtttagaaa atgtggcttt taatgttgta aataagggac actttgatgg acaacagggt
19681 gaagtaccag tttctatcat taataacact gtttacacaa aagttgatgg tgttgatgta
19741 gaattgtttg aaaataaaac aacattacct gttaatgtag catttgagct ttgggctaag
19801 cgcaacatta aaccagtacc agaggtgaaa atactcaata atttgggtgt ggacattgct
19861 gctaatactg tgatctggga ctacaaaaga gatgctccag cacatatatc tactattggt
19921 gtttgttcta tgactgacat agccaagaaa ccaactgaaa cgatttgtgc accactcact
19981 gtctttttg atggtagagt tgatggtcaa gtagacttat ttagaaatgc ccgtaatggt
20041 gttcttatta cagaaggtag tgttaaaggt ttacaaccat ctgtaggtcc caaacaagct
20101 agtcttaatg gagtcacatt aattggagaa gccgtaaaaa cacagttcaa ttattataag
20161 aaagttgatg gtgttgtcca acaattacct gaaacttact ttactcagag tagaaattta
20221 caagaattta aacccaggag tcaaatggaa attgatttct tagaattagc tatggatgaa
20281 ttcattgaac ggtataaatt agaaggctat gccttcgaac atatcgttta tggagatttt
20341 agtcatagtc agttaggtgg tttacatcta ctgattggac tagctaaacg ttttaaggaa
20401 tcacctttg aattagaaga ttttattcct atggacagta cagttaaaaa ctatttcata
20461 acagatgcgc aaacaggttc atctaagtgt gtgtgttctg ttattgattt attacttgat
20521 gattttgttg aaataataaa atcccaagat ttatctgtag tttctaaggt tgtcaaagtg
20581 actattgact atacagaaat ttcatttatg ctttggtgta aagatggcca tgtagaaaca
20641 ttttacccaa aattacaatc tagtcaagcg tggcaaccgg gtgttgctat gcctaatctt
```

Fig. 12C continued

```
20701 tacaaaatgc aaagaatgct attagaaaag tgtgaccttc aaaattatgg tgatagtgca
20761 acattaccta aaggcataat gatgaatgtc gcaaaatata ctcaactgtg tcaatattta
20821 aacacattaa cattagctgt accctataat atgagagtta tacattttgg tgctggttct
20881 gataaaggag ttgcaccagg tacagctgtt ttaagacagt ggttgcctac gggtacgctg
20941 cttgtcgatt cagatcttaa tgactttgtc tctgatgcag attcaacttt gattggtgat
21001 tgtgcaactg tacatacagc taataaatgg gatctcatta ttagtgatat gtacgaccct
21061 aagactaaaa atgttacaaa agaaaatgac tctaaagagg gtttttcac ttacatttgt
21121 gggtttatac aacaaaagct agctcttgga ggttccgtgg ctataaagat aacagaacat
21181 tcttggaatg ctgatcttta taagctcatg ggacacttcg catggtggac agcctttgtt
21241 actaatgtga atgcgtcatc atctgaagca tttttaattg gatgtaatta tcttggcaaa
21301 ccacgcgaac aaatagatgg ttatgtcatg catgcaaatt acatattttg gaggaataca
21361 aatccaattc agttgtcttc ctattcttta tttgacatga gtaaatttcc ccttaaatta
21421 aggggtactg ctgttatgtc tttaaaagaa ggtcaaatca atgatatgat tttatctctt
21481 cttagtaaag gtagacttat aattagagaa aacaacagag ttgttatttc tagtgatgtt
21541 cttgttaaca actaaacgaa caatgtttgt ttttcttgtt ttattgccac tagtctctag
21601 tcagtgtgtt aatcttacaa ccagaactca attaccccct gcatacacta attctttcac
21661 acgtggtgtt tattaccctg acaaagtttt cagatcctca gttttacatt caactcagga
21721 cttgttctta cctttctttt ccaatgttac ttggttccat gctatacatg tctctgggac
21781 caatggtact aagaggtttg ataaccctgt cctaccattt aatgatggtg tttattttgc
21841 ttccactgag aagtctaaca taataagagg ctggattttt ggtactactt tagattcgaa
21901 gacccagtcc ctacttattg ttaataacgc tactaatgtt gttattaaag tctgtgaatt
```

Fig. 12C

continued

```
21961 tcaattttgt aatgatccat ttttgggtgt ttattaccac aaaaaacaaca aaagttggat
22021 ggaaaagtgag ttcagagttt attctagtgc gaataattgc acttttgaat atgtctctca
22081 gccttttctt atggaccttg aaggaaaaca gggtaatttc aaaaaatctta gggaatttgt
22141 gtttaagaat attgatggtt attttaaaat atattctaag cacacgccta ttaatttagt
22201 gcgtgatctc cctcagggtt tttcggcttt agaaccattg gtagatttgc caataggtat
22261 taacatcact aggtttcaaa ctttacttgc tttacataga agttatttga ctcctggtga
22321 ttcttcttca ggttggacag ctggtgctgc agcttattat gtgggttatc ttcaacctag
22381 gacttttcta ttaaaatata atgaaaatgg aaccattaca gatgctgtag actgtgcact
22441 tgacccctctc tcagaaacaa agtgtacgtt gaaatccttc actgtagaaa aaggaatcta
22501 tcaaaacttct aactttagag tccaaccaac agaatctatt gttagatttc ctaatattac
22561 aaacttgtgc ccttttggtg aagtttttaa cgccaccaga tttgcatctg tttatgcttg
22621 gaacaggaag agaatcagca actgtgttgc tgattattct gtcctatata attccgcatc
22681 attttccact tttaagtgtt atggagtgtc tcctactaaa ttaaatgatc tctgctttac
22741 taatgtctat gcagattcat ttgtaattag aggtgatgaa gtcagacaaa tcgctccagg
22801 gcaaactgga aagattgctg attataatta taaattacca gatgatttta caggctgcgt
22861 tatagcttgg aattctaaca atcttgattc taaggttggt ggtaattata attacctgta
22921 tagattgttt aggaagtcta atctcaaacc ttttgagaga gatatttcaa ctgaaatcta
22981 tcaggccggt agcacacctt gtaatggtgt tgaaggtttt aattgttact ttcctttaca
23041 atcatatggt ttccaaccca ctaatggtgt tggttaccaa ccatacagag tagtagtact
23101 ttcttttgaa cttctacatg caccagcaac tgtttgtgga cctaaaaagt ctactaattt
```

Fig. 12C continued

```
23161 ggttaaaaac aaatgtgtca atttcaactt caatggttta acaggcacag gtgttcttac
23221 tgagtctaac aaaaagtttc tgcctttcca acaatttggc agagacattg ctgacactac
23281 tgatgctgtc cgtgatccac agacacttga gattcttgac attacaccat gttcttttgg
23341 tggtgtcagt gttataacac caggaacaaa tacttctaac caggttgctg ttctttatca
23401 ggatgttaac tgcacagaag tccctgttgc tattcatgca gatcaactta ctcctacttg
23461 gcgtgtttat tctacaggtt ctaatgtttt tcaaacacgt gcaggctgtt taatagggggc
23521 tgaacatgtc aacaactcat atgagtgtga catacccatt ggtgcaggta tatgcgctag
23581 ttatcagact cagactaatt ctcctcggcg ggcacgtagt gtagctagtc aatccatcat
23641 tgcctacact atgtcacttg gtgcagaaaa ttcagttgct tactctaata actctattgc
23701 catacccaca aattttacta ttagtgttac cacagaaatt ctaccagtgt ctatgaccaa
23761 gacatcagta gattgtacaa tgtacatttg tggtgattca actgaatgca gcaatctttt
23821 gttgcaatat ggcagttttt gtacacaatt aaaccgtgct ttaactggaa tagctgttga
23881 acaagacaaa aacacccaag aagtttttgc acaagtcaaa caaatttaca aaacaccacc
23941 aattaaagat tttggtggtt ttaatttttc acaaatatta ccagatccat caaaaccaag
24001 caagaggtca tttattgaag atctactttt caacaaagtg acacttgcag atgctggctt
24061 catcaaacaa tatggtgatt gccttggtga tattgctgct agagacctca tttgtgcaca
24121 aaagtttaac ggccttactg ttttgccacc tttgctcaca gatgaaatga ttgctcaata
24181 cacttctgca ctgttagcgg gtacaatcac ttctggttgg acctttggtg caggtgctgc
24241 attacaaata ccatttgcta tgcaaatggc ttataggttt aatggtattg gagttacaca
24301 gaatgttctc tatgagaacc aaaaattgat tgccaaccaa tttaatagtg ctattggcaa
24361 aattcaagac tcactttctt ccacagcaag tgcacttgga aaacttcaag atgtggtcaa
```

Fig. 12C
continued

```
24421 ccaaaatgca caagctttaa acacgcttgt taaacaactt agctccaatt ttggtgcaat
24481 ttcaagtgtt ttaaatgata tcctttcacg tcttgacaaa gttgaggctg aagtgcaaat
24541 tgataggttg atcacaggca gacttcaaag tttgcagaca tatgtgactc aacaattaat
24601 tagagctgca gaaatcagag cttctgctaa tcttgctgct actaaaatgt cagagtgtgt
24661 acttggacaa tcaaaaagag ttgatttttg tggaaagggc tatcatctta tgtccttccc
24721 tcagtcagca cctcatggtg tagtcttctt gcatgtgact tatgtccctg cacaagaaaa
24781 gaacttcaca actgctcctg ccatttgtca tgatggaaaa gcacactttc ctcgtgaagg
24841 tgtctttgtt tcaaatggca cacactggtt tgtaacacaa aggaattttt atgaaccaca
24901 aatcattact acagacaaca catttgtgtc tggtaactgt gatgttgtaa taggaattgt
24961 caacaacaca gtttatgatc ctttgcaacc tgaattagac tcattcaagg aggagttaga
25021 taaatatttt aagaatcata catcaccaga tgttgattta ggtgacatct ctggcattaa
25081 tgcttcagtt gtaaacattc aaaaagaaat tgaccgcctc aatgaggttg ccaagaattt
25141 aaatgaatct ctcatcgatc tccaagaact tggaaagtat gagcagtata taaaatggcc
25201 atggtacatt tggctaggtt ttatagctgg cttgattgcc atagtaatgg tgacaattat
25261 gctttgctgt atgaccagtt gctgtagttg tctcaagggc tgttgttctt gtggatcctg
25321 ctgcaaattt gatgaagacg actctgagcc agtgctcaaa ggagtcaaat tacattacac
25381 ataaacgaac ttatggattt gtttatgaga atcttcacaa ttggaactgt aactttgaag
25441 caaggtgaaa tcaaggatgc tactccttca gattttgttc gcgctactgc aacgataccg
25501 atacaagcct cactccctttt cggatggctt attgttggcg ttgcacttct tgctgttttt
25561 cagagcgctt ccaaaatcat aaccctcaaa aagagatggc aactagcact ctccaagggt
```

Fig. 12C continued

```
25621 gttcactttg tttgcaactt gctgttgttg tttgtaacag tttactcaca ccttttgctc
25681 gttgctgctg gccttgaagc ccctttttctc tatctttatg ctttagtcta cttcttgcag
25741 agtataaact ttgtaagaat aataatgagg ctttggcttt gctggaaatg ccgttccaaa
25801 aacccattac tttatgatgc caactatttt ctttgctggc atactaattg ttacgactat
25861 tgtatacctt acaatagtgt aacttcttca attgtcatta cttcaggtga tggcacaaca
25921 agtcctattt ctgaacatga ctaccagatt ggtggttata ctgaaaaatg ggaatctgga
25981 gtaaaagact gtgttgtatt acacagttac ttcacttcag actattacca gctgtactca
26041 actcaattga gtacagacac tggtgttgaa catgttacct tcttcatcta caataaaatt
26101 gttgatgagc ctgaagaaca tgtccaaatt cacacaatcg acggttcatc cggagttgtt
26161 aatccagtaa tggaaccaat ttatgatgaa ccgacgacga ctactagcgt gcctttgtaa
26221 gcacaagctg atgagtacga acttatgtac tcattcgttt cggaagagac aggtacgtta
26281 atagttaata gcgtacttct ttttcttgct ttcgtggtat tcttgctagt tacactagcc
26341 atccttactg cgcttcgatt gtgtgcgtac tgctgcaata ttgttaacgt gagtcttgta
26401 aaaccttctt tttacgttta ctctcgtgtt aaaaatctga attcttctag agttcctgat
26461 cttctggtct aaacgaacta aatattatat tagtttttct gtttggaact ttaattttag
26521 ccatggcaga ttccaacggt actattaccg ttgaagagct taaaaagctc cttgaacaat
26581 ggaacctagt aataggtttc ctattcctta catggatttg tcttctacaa tttgcctatg
26641 ccaacaggaa taggtttttg tatataatta agttaatttt cctctggctg ttatggccag
26701 taactttagc ttgttttgtg cttgctgctg tttacagaat aaattggatc accggtggaa
26761 ttgctatcgc aatggcttgt cttgtaggct tgatgtggct cagctacttc attgcttctt
26821 tcagactgtt tgcgcgtacg cgttccatgt ggtcattcaa tccagaaact aacattcttc
```

Fig. 12C

continued

```
26881 tcaacgtgcc actccatggc actattctga ccagaccgct tctagaaagt gaactcgtaa
26941 tcggagctgt gatccttcgt ggacatcttc gtattgctgg acaccatcta ggacgctgtg
27001 acatcaagga cctgcctaaa gaaatcactg ttgctacatc acgaacgctt tcttattaca
27061 aattgggagc ttcgcagcgt gtagcaggtg actcaggttt tgctgcatac agtcgctaca
27121 ggattggcaa ctataaatta aacacagacc attccagtag cagtgacaat attgctttgc
27181 ttgtacagta agtgacaaca gatgtttcat ctcgttgact ttcaggttac tatagcagag
27241 atattactaa ttattatgag gacttttaaa gtttccattt ggaatcttga ttacatcata
27301 aacctcataa ttaaaaattt atctaagtca ctaactgaga ataaatattc tcaattagat
27361 gaagagcaac caatggagat tgattaaacg aacatgaaaa ttattctttt cttggcactg
27421 ataacactcg ctacttgtga gctttatcac taccaagagt gtgttagagg tacaacagta
27481 cttttaaaag aaccttgctc ttctggaaca tacgagggca attcaccatt tcatcctcta
27541 gctgataaca aatttgcact gacttgcttt agcactcaat ttgcttttgc ttgtcctgac
27601 ggcgtaaaac acgtctatca gttacgtgcc agatcagttt cacctaaact gttcatcaga
27661 caagaggaag ttcaagaact ttactctcca attttttctta ttgttgcggc aatagtgttt
27721 ataacacttt gcttcacact caaaagaaag acagaatgat tgaactttca ttaattgact
27781 tctatttgtg ctttttagcc tttctgctat tccttgtttt aattatgctt attatctttt
27841 ggttctcact tgaactgcaa gatcataatg aaacttgtca cgcctaaacg aacatgaaat
27901 ttcttgtttt cttaggaatc atcacaactg tagctgcatt tcaccaagaa tgtagtttac
27961 agtcatgtac tcaacatcaa ccatatgtag ttgatgaccc gtgtcctatt cacttctatt
28021 ctaaatggta tattagagta ggagctagaa aatcagcacc tttaattgaa ttgtgcgtgg
```

Fig. 12C continued

```
28081 atgaggctgg ttctaaatca cccattcagt acatcgatat cggtaattat acagtttcct
28141 gtttaccttt tacaattaat tgccaggaac ctaaattggg tagtcttgta gtgcgttgtt
28201 cgttctatga agacttttta gagtatcatg acgttcgtgt tgttttagat ttcatctaaa
28261 cgaacaaact aaaatgtctg ataatggacc ccaaaatcag cgaaatgcac cccgcattac
28321 gtttggtgga ccctcagatt caactggcag taaccagaat ggagaacgca gtggggcgcg
28381 atcaaaaacaa cgtcggcccc aaggtttacc caataatact gcgtcttggt tcaccgctct
28441 cactcaacat ggcaaggaag accttaaatt ccctcgagga caaggcgttc caattaacac
28501 caatagcagt ccagatgacc aaattggcta ctaccgaaga gctaccagac gaattcgtgg
28561 tggtgacggt aaaatgaaag atctcagtcc aagatggtat ttctactacc taggaactgg
28621 gccagaagct ggacttccct atggtgctaa caaagacggc atcatatggg ttgcaactga
28681 gggagccttg aatacaccaa aagatcacat tggcacccgc aatcctgcta acaatgctgc
28741 aatcgtgcta caacttcctc aaggaacaac attgccaaaa ggcttctacg cagaagggag
28801 cagaggcggc agtcaagcct cttctcgttc ctcatcacgt agtcgcaaca gttcaagaaa
28861 ttcaactcca ggcagcagta ggggaacttc tcctgctaga atggctggca atggcggtga
28921 tgctgctctt gctttgctgc tgcttgacag attgaaccag cttgagagca aaatgtctgg
28981 taaaggccaa caacaacaag gccaaactgt cactaagaaa tctgctgctg aggcttctaa
29041 gaagcctcgg caaaaacgta ctgccactaa agcatacaat gtaacacaag ctttcggcag
29101 acgtggtcca gaacaaaccc aaggaaattt tggggaccag gaactaatca gacaaggaac
29161 tgattacaaa cattggccgc aaattgcaca atttgccccc agcgcttcag cgttcttcgg
29221 aatgtcgcgc attggcatgg aagtcacacc ttcgggaacg tggttgacct acacaggtgc
29281 catcaaattg gatgacaaag atccaaattt caaagatcaa gtcattttgc tgaataagca
```

Fig. 12C
continued

```
29341 tattgacgca tacaaaacat tcccaccaac agagcctaaa aaggacaaaa agaagaaggc
29401 tgatgaaact caagccttac cgcagagaca gaagaaacag caaactgtga ctcttcttcc
29461 tgctgcagat ttggatgatt tctccaaaca attgcaacaa tccatgagca gtgctgactc
29521 aactcaggcc taaactcatg cagaccacac aaggcagatg ggctatataa acgttttcgc
29581 ttttccgttt acgatatata gtctactctt gtgcagaatg aattctcgta actacatagc
29641 acaagtagat gtagttaact ttaatctcac atagcaatct ttaatcagtg tgtaacatta
29701 gggaggactt gaaagagcca ccacatttcc accgaggcca cgcggagtac gatcgagtgt
29761 acagtgaaca atgctaggga gagctgccta tatggaagag ccctaatgtg taaaattaat
29821 tttagtagtg ctatccccat gtgattttaa tagcttctta ggagaatgac aaaaaaaaaa
29881 aaaaaaaaaa aaaaaaaaaa aaa
```

Fig. 12C continued

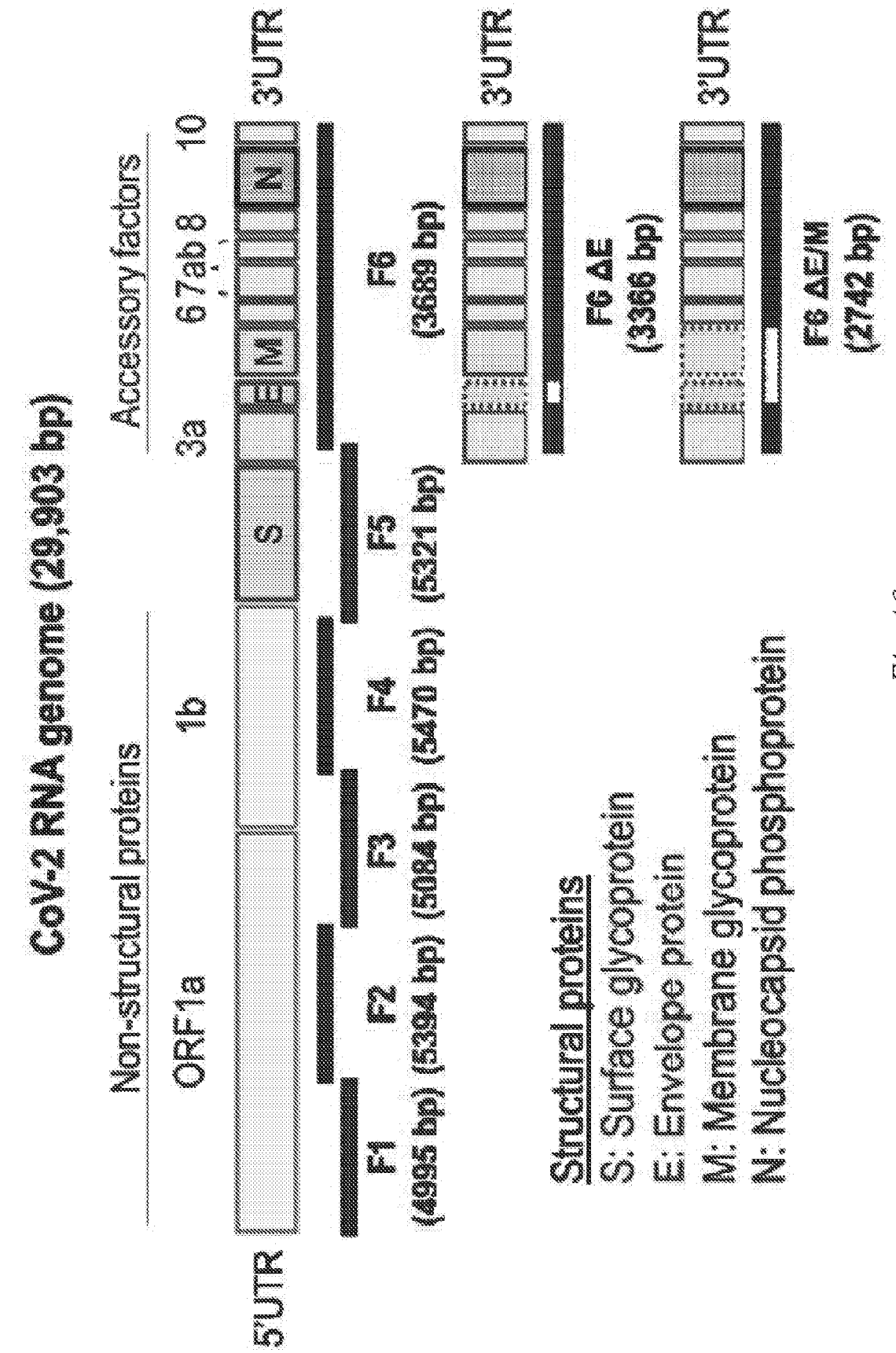

CoV-2 RNA genome (29,903 bp)
Non-structural proteins
Accessory factors
ORF1a
1b
3a
6 7ab 8
10
5'UTR
S
E
M
N
3'UTR
F1
F2
F3
F4
F5
F6
(4995 bp) (5394 bp) (5084 bp) (5470 bp) (5321 bp) (3689 bp)
F6 ΔE
(3366 bp)
F6 ΔE/M
(2742 bp)
Structural proteins
S: Surface glycoprotein
E: Envelope protein
M: Membrane glycoprotein
N: Nucleocapsid phosphoprotein

Fig. 13

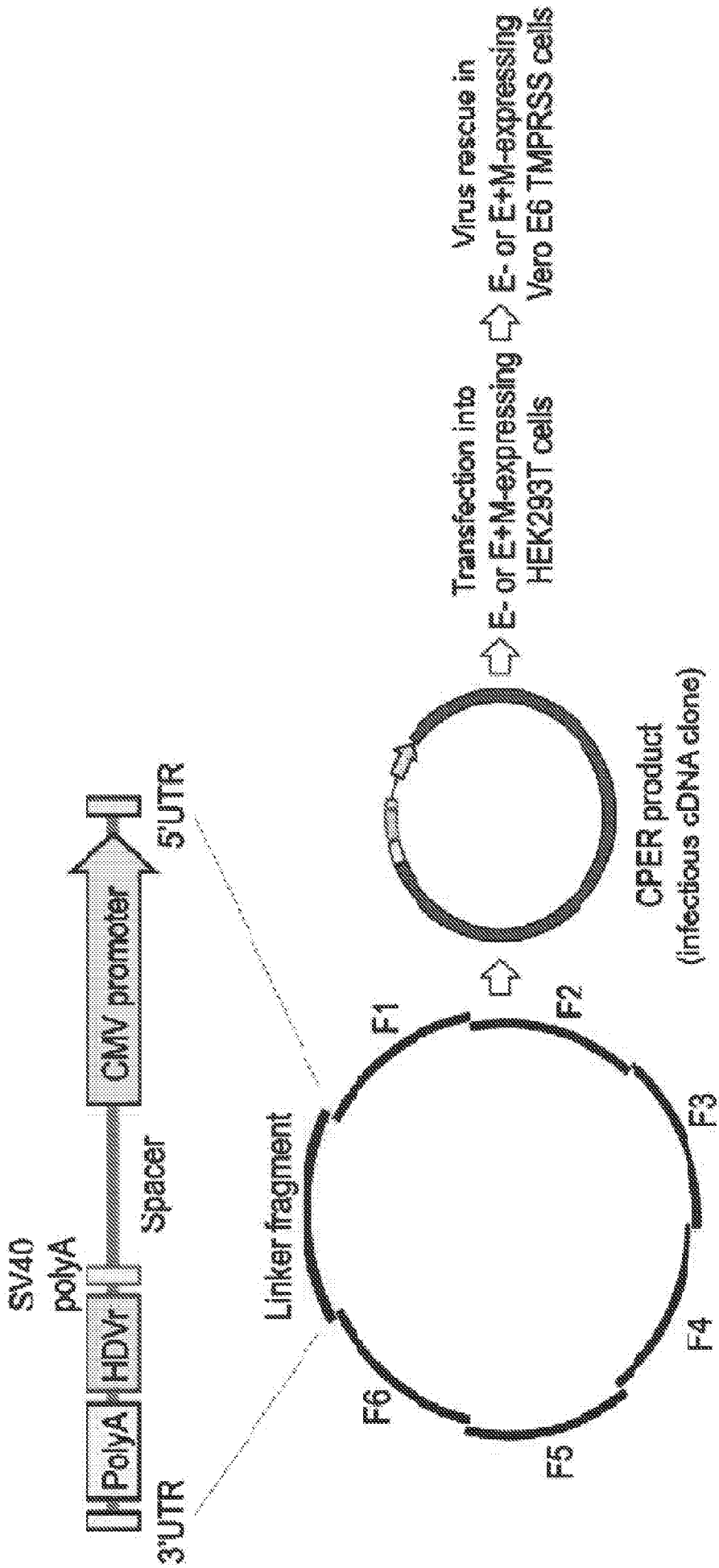

Virus rescue in
E- or E+M-expressing
Vero E6 TMPRSS cells
Transfection into
E- or E+M-expressing
HEK293T cells
CPER product
(infectious cDNA clone)
5'UTR
CMV promoter
Spacer
SV40
polyA
HDVr
PolyA
3'UTR
Linker fragment
F1
F2
F3
F4
F5
F6

Fig. 14

Codon-optimized CoV-2 E gene (Addgene)

ATGTACTCTTTCGTGAGCGAGGAAACCGGCACCCTGATCGTGAACTCCGTGCTGCTGTTCCTGGCCTTCGTGGTGT
TCCTGCTGGTGACCCTGGCTATCCTGACCGCTCTGAGACTGTGCGCTTACTGCTGCAACATCGTGAACGTGTCCCTG
GTGAAGCCCTCTTTCTACGTGTACAGCCGCGTGAAGAACCTGAACAGCTCCAGGGTGCCTGACCTGCTGGTGTAA
(SEQ ID NO:13)

Codon-optimized CoV-2 M gene (Addgene)

ATGGCTGACTCTAACGGTACCATCACCGTGGAGGAACTGAAGAAGCTGCTGGAGCAGTGGAACCTGGTCATCGG
CTTCCTGTTCCTGACCTGGATCTGCCTGCTGCAGTTCGCCTACGCTAACCGCAACAGGTTCCTGTACATCATCAAGC
TGATCTTCCTGTGGCTGCTGTGGCCTGTGACCCTGGCTTGCTTCGTGCTGGCTGCCGTGTACCGCATCAACTGGATC
ACCGGCGGAATCGCCATCGCTATGGCCTGCCTGGTGGGCCTGATGTGGCTGTCTTACTTCATCGCTAGCTTCAGGC
TGTTCGCCAGAACCCGTTCCATGTGGTCTTTCAACCCCGAGACCAACATCCTGCTGAACGTGCCTCTGCACGGAAC
CATCCTGACCAGACCACTGCTGGAGAGCGAACTGGTCATCGGCGCTGTGATCCTGAGAGGACACCTGCGTATCGC
CGGACACCACCTGGGTCGTTGCGACATCAAGGACCTGCCCAAGGAAATCACCGTGGCTACCAGCCGCACCCTGTC
CTACTACAAGCTGGGAGCTTCTCAGAGAGTGGCTGGTGACTCTGGTTTCGCTGCTTACTCTCGCTACAGGATCGGT
AACTACAAGCTGAACACCGACCACAGCTCCTCTAGCGACAACATCGCCCTGCTGGTGCAGTAA (SEQ ID NO:14)

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

ΔE GENOME

Fig. 15

ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTT
AAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTG
ACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACA
TCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGT
CCAACTCAGTTTGCCTGTTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTA
TCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTG
AACAGCCCTATGTGTTCATCAAACGTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGC
AGAACTCGAAGGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACC
AGTGGCTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCT
AAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACACTAAACAT
AGCAGTGGTGTTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACTTCT
GTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTGTCC
GAACAACTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTGCTTGGTAC
ACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAAATTAAATTGGCAAAGAAATTTGACACCTTCA
ATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAA
GCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCCTTT
CAACTCTCATGAAGTGTGATCATTGTGGTGAAACTTCATGGCAGACGGGCGATTTTGTTAAAGCCACTTGCGAATT
TTGTGGCACTGAGAATTTGACTAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTT
ATTGTCCAGCATGTCACAATTCAGAAGTAGGACCTGAGCATAGTCTTGCCGAATACCATAATGAATCTGGCTTGAA
AACCATTCTTCGTAAGGGTGGTCGCACTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAGT
GTGCCTATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGTTGGAGAAGGTTCCGAAG

Fig. 15

GTCTTAATGACAACCTTCTTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGACTTTAAACTTAAT
GAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCTTCCACAAGTGCTTTTGTGGAAACTGTGAAAGGTTTGGATTA
TAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTTTAAAGTTACAAAAGGAAAAGCTAAAAAAAGGTGCCTGG
AATATTGGTGAACAGAAATCAATACTGAGTCCTCTTTATGCATTTGCATCAGAGGCTGCTCGTGTTGTACGATCAAT
TTTCTCCCGCACTCTTGAAACTGCTCAAAATTCTGTGCGTGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAA
TTTCACAGTATTCACTGAGACTCATTGATGCTATGATGTTCACATCTGATTTGGCTACTAACAATCTAGTTGTAATG
GCCTACATTACAGGTGGTGTTGTTCAGTTGACTTCGCAGTGGCTAACTAACATCTTTGGCACTGTTTATGAAAAACT
CAAACCCGTCCTTGATTGGCTTGAAGAGAAGTTTAAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTGT
TAAATTTATCTCAACCTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGGAGAGT
GTTCAGACATTCTTTAAGCTTGTAAATAAATTTTTGGCTTTGTGTGCTGACTCTATCATTATTGGTGGAGCTAAACTT
AAAGCCTTGAATTTAGGTGAAACATTTGTCACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAA
GAAACTGGCCTACTCATGCCTCTAAAAGCCCCAAAAGAAATTATCTTCTTAGAGGGAGAAACACTTCCCACAGAAG
TGTTAACAGAGGAAGTTGTCTTGAAAACTGGTGATTTACAACCATTAGAACAACCTACTAGTGAAGCTGTTGAAGC
TCCATTGGTTGGTACACCAGTTTGTATTAACGGGCTTATGTTGCTCGAAATCAAAGACACAGAAAAGTACTGTGCC
CTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTTGGTG
ATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTTGAACTTGATGAAAGGATTGATAAAG
TACTTAATGAGAAGTGCTCTGCCTATACAGTTGAACTCGGTACAGAAGTAAATGAGTTCGCCTGTGTTGTGGCAGA
TGCTGTCATAAAAACTTTGCAACCAGTATCTGAATTACTTACACCACTGGGCATTGATTTAGATGAGTGGAGTATG
GCTACATACTACTTATTTGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTCTACCCTCCAGAT
GAGGATGAAGAAGAAGGTGATTGTGAAGAAGAAGAGTTTGAGCCATCAACTCAATATGAGTATGGTACTGAAGA
TGATTACCAAGGTAAACCTTTGGAATTTGGTGCCACTTCTGCTGCTCTTCAACCTGAAGAAGAGCAAGAAGAAGAT
TGGTTAGATGATGATAGTCAACAAACTGTTGGTCAACAAGACGGCAGTGAGGACAATCAGACAACTACTATTCAA

Fig. 15

ACAATTGTTGAGGTTCAACCTCAATTAGAGATGGAACTTACACCAGTTGTTCAGACTATTGAAGTGAATAGTTTTA
GTGGTTATTTAAAACTTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAAAAAGGTAAAACC
AACAGTGGTTGTTAATGCAGCCAATGTTTACCTTAAACATGGAGGAGGTGTTGCAGGAGCCTTAAATAAGGCTAC
TAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCACTTAAAGTGGGTGGTAGTTGTGTT
TTAAGCGGACACAATCTTGCTAAACACTGTCTTCATGTTGTCGGCCCAAATGTTAACAAAGGTGAAGACATTCAAC
TTCTTAAGAGTGCTTATGAAAATTTTAATCAGCACGAAGTTCTACTTGCACCATTATTATCAGCTGGTATTTTTGGTG
CTGACCCTATACATTCTTTAAGAGTTTGTGTAGATACTGTTCGCACAAATGTCTACTTAGCTGTCTTTGATAAAAATC
TCTATGACAAACTTGTTTCAAGCTTTTTGGAAATGAAGAGTGAAAAGCAAGTTGAACAAAAGATCGCTGAGATTCC
TAAAGAGGAAGTTAAGCCATTTATAACTGAAAGTAAACCTTCAGTTGAACAGAGAAAACAAGATGATAAGAAAAT
CAAAGCTTGTGTTGAAGAAGTTACAACAACTCTGGAAGAAACTAAGTTCCTCACAGAAAACTTGTTACTTTATATTG
ACATTAATGGCAATCTTCATCCAGATTCTGCCACTCTTGTTAGTGACATTGACATCACTTTCTTAAAGAAAGATGCTC
CATATATAGTGGGTGATGTTGTTCAAGAGGGTGTTTTAACTGCTGTGGTTATACCTACTAAAAAGGCTGGTGGCAC
TACTGAAATGCTAGCGAAAGCTTTGAGAAAAGTGCCAACAGACAATTATATAACCACTTACCCGGGTCAGGGTTTA
AATGGTTACACTGTAGAGGAGGCAAAGACAGTGCTTAAAAAGTGTAAAAGTGCCTTTTACATTCTACCATCTATTA
TCTCTAATGAGAAGCAAGAAATTCTTGGAACTGTTTCTTGGAATTTGCGAGAAATGCTTGCACATGCAGAAGAAAC
ACGCAAATTAATGCCTGTCTGTGTGGAAACTAAAGCCATAGTTTCAACTATACAGCGTAAATATAAGGGTATTAAA
ATACAAGAGGGTGTGGTTGATTATGGTGCTAGATTTTACTTTTACACCAGTAAAACAACTGTAGCGTCACTTATCA
ACACACTTAACGATCTAAATGAAACTCTTGTTACAATGCCACTTGGCTATGTAACACATGGCTTAAATTTGGAAGAA
GCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTTTCTGTTTCTTCACCTGATGCTGTTACAGCGTATAA
TGGTTATCTTACTTCTTCTTCTAAAACACCTGAAGAACATTTTATTGAAACCATCTCACTTGCTGGTTCCTATAAAGA
TTGGTCCTATTCTGGACAATCTACACAACTAGGTATAGAATTTCTTAAGAGAGGTGATAAAAGTGTATATTACACTA
GTAATCCTACCACATTCCACCTAGATGGTGAAGTTATCACCTTTGACAATCTTAAGACACTTCTTTCTTTGAGAGAA

Fig. 15

GTGAGGACTATTAAGGTGTTTACAACAGTAGACAACATTAACCTCCACACGCAAGTTGTGGACATGTCAATGACAT
ATGGACAACAGTTTGGTCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCATAATTCACATGAAGG
TAAAACATTTTATGTTTTACCTAATGATGACACTCTACGTGTTGAGGCTTTTGAGTACTACCACACAACTGATCCTA
GTTTTCTGGGTAGGTACATGTCAGCATTAAATCACACTAAAAAGTGGAAATACCCACAAGTTAATGGTTTAACTTCT
ATTAAATGGGCAGATAACAACTGTTATCTTGCCACTGCATTGTTAACACTCCAACAAATAGAGTTGAAGTTTAATCC
ACCTGCTCTACAAGATGCTTATTACAGAGCAAGGGCTGGTGAAGCTGCTAACTTTTGTGCACTTATCTTAGCCTACT
GTAATAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAACAATGAGTTACTTGTTTCAACATGCCAATTTAGATTC
TTGCAAAAGAGTCTTGAACGTGGTGTGTAAAACTTGTGGACAACAGCAGACAACCCTTAAGGGTGTAGAAGCTGT
TATGTACATGGGCACACTTTCTTATGAACAATTTAAGAAAGGTGTTCAGATACCTTGTACGTGTGGTAAACAAGCT
ACAAAATATCTAGTACAACAGGAGTCACCTTTTGTTATGATGTCAGCACCACCTGCTCAGTATGAACTTAAGCATG
GTACATTTACTTGTGCTAGTGAGTACACTGGTAATTACCAGTGTGGTCACTATAAACATATAACTTCTAAAGAAACT
TTGTATTGCATAGACGGTGCTTTACTTACAAAGTCCTCAGAATACAAAGGTCCTATTACGGATGTTTTCTACAAAGA
AAACAGTTACACAACAACCATAAAACCAGTTACTTATAAATTGGATGGTGTTGTTTGTACAGAAATTGACCCTAAG
TTGGACAATTATTATAAGAAAGACAATTCTTATTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCC
AAACGCAAGCTTCGATAATTTTAAGTTTGTATGTGATAATATCAAATTTGCTGATGATTTAAACCAGTTAACTGGTT
ATAAGAAACCTGCTTCAAGAGAGCTTAAAGTTACATTTTTCCCTGACTTAAATGGTGATGTGGTGGCTATTGATTAT
AAACACTACACACCCTCTTTTAAGAAAGGAGCTAAATTGTTACATAAACCTATTGTTTGGCATGTTAACAATGCAAC
TAATAAAGCCACGTATAAACCAAATACCTGGTGTATACGTTGTCTTTGGAGCACAAAACCAGTTGAAACATCAAAT
TCGTTTGATGTACTGAAGTCAGAGGACGCGCAGGGAATGGATAATCTTGCCTGCGAAGATCTAAAACCAGTCTCT
GAAGAAGTAGTGGAAAATCCTACCATACAGAAAGACGTTCTTGAGTGTAATGTGAAAACTACCGAAGTTGTAGGA
GACATTATACTTAAACCAGCAAATAATAGTTTAAAAATTACAGAAGAGGTTGGCCACACAGATCTAATGGCTGCTT
ATGTAGACAATTCTAGTCTTACTATTAAGAAACCTAATGAATTATCTAGAGTATTAGGTTTGAAAACCCTTGCTACT

Fig. 15

CATGGTTTAGCTGCTGTTAATAGTGTCCCTTGGGATACTATAGCTAATTATGCTAAGCCTTTTCTTAACAAAGTTGTT
AGTACAACTACTAACATAGTTACACGGTGTTTAAACCGTGTTTGTACTAATTATATGCCTTATTTCTTTACTTTATTG
CTACAATTGTGTACTTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAGCAAAGAATAC
TGTTAAGAGTGTCGGTAAATTTTGTCTAGAGGCTTCATTTAATTATTTGAAGTCACCTAATTTTTCTAAACTGATAAA
TATTATAATTTGGTTTTTACTATTAAGTGTTTGCCTAGGTTCTTTAATCTACTCAACCGCTGCTTTAGGTGTTTTAATG
TCTAATTTAGGCATGCCTTCTTACTGTACTGGTTACAGAGAAGGCTATTTGAACTCTACTAATGTCACTATTGCAAC
CTACTGTACTGGTTCTATACCTTGTAGTGTTTGTCTTAGTGGTTTAGATTCTTTAGACACCTATCCTTCTTTAGAAACT
ATACAAATTACCATTTCATCTTTTAAATGGGATTTAACTGCTTTTGGCTTAGTTGCAGAGTGGTTTTTGGCATATATT
CTTTTCACTAGGTTTTTCTATGTACTTGGATTGGCTGCAATCATGCAATTGTTTTTCAGCTATTTTGCAGTACATTTTA
TTAGTAATTCTTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCGATTTCAGCTATGGTTAGAATGTAC
ATCTTCTTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGACGGTTGTAATTCATCAACTTGTATG
ATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCCTTTTATG
TCTATGCTAATGGAGGTAAAGGCTTTTGCAAACTACACAATTGGAATTGTGTTAATTGTGATACATTCTGTGCTGGT
AGTACATTTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTTAAAAGACCAATAAATCCTACTGACCAGT
CTTCTTACATCGTTGATAGTGTTACAGTGAAGAATGGTTCCATCCATCTTTACTTTGATAAAGCTGGTCAAAAGACT
TATGAAAGACATTCTCTCTCTCATTTTGTTAACTTAGACAACCTGAGAGCTAATAACACTAAAGGTTCATTGCCTATT
AATGTTATAGTTTTTGATGGTAAATCAAAATGTGAAGAATCATCTGCAAAATCAGCGTCTGTTTACTACAGTCAGCT
TATGTGTCAACCTATACTGTTACTAGATCAGGCATTAGTGTCTGATGTTGGTGATAGTGCGGAAGTTGCAGTTAAA
ATGTTTGATGCTTACGTTAATACGTTTTCATCAACTTTTAACGTACCAATGGAAAAACTCAAAACACTAGTTGCAAC
TGCAGAAGCTGAACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCTACTTTTATTTCAGCAGCTCGGCAAGGG
TTTGTTGATTCAGATGTAGAAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCAATCTGACATAGAAGTTAC
TGGCGATAGTTGTAATAACTATATGCTCACCTATAACAAAGTTGAAAACATGACACCCCGTGACCTTGGTGCTTGT

Fig. 15

ATTGACTGTAGTGCGCGTCATATTAATGCGCAGGTAGCAAAAAGTCACAACATTGCTTTGATATGGAACGTTAAAG
ATTTCATGTCATTGTCTGAACAACTACGAAAACAAATACGTAGTGCTGCTAAAAAGAATAACTTACCTTTTAAGTTG
ACATGTGCAACTACTAGACAAGTTGTTAATGTTGTAACAACAAAGATAGCACTTAAGGGTGGTAAAATTGTTAATA
ATTGGTTGAAGCAGTTAATTAAAGTTACACTTGTGTTCCTTTTTGTTGCTGCTATTTTCTATTTAATAACACCTGTTCA
TGTCATGTCTAAACATACTGACTTTTCAAGTGAAATCATAGGATACAAGGCTATTGATGGTGGTGTCACTCGTGAC
ATAGCATCTACAGATACTTGTTTTGCTAACAAACATGCTGATTTTGACACATGGTTTAGCCAGCGTGGTGGTAGTTA
TACTAATGACAAAGCTTGCCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTTGTCGTGCCTGGTTTGCCT
GGCACGATATTACGCACAACTAATGGTGACTTTTTGCATTTCTTACCTAGAGTTTTTAGTGCAGTTGGTAACATCTG
TTACACACCATCAAAACTTATAGAGTACACTGACTTTGCAACATCAGCTTGTGTTTTGGCTGCTGAATGTACAATTT
TTAAAGATGCTTCTGGTAAGCCAGTACCATATTGTTATGATACCAATGTACTAGAAGGTTCTGTTGCTTATGAAAGT
TTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTATTCAATTTCCTAACACCTACCTTGAAGGTTCTGTT
AGAGTGGTAACAACTTTTGATTCTGAGTACTGTAGGCACGGCACTTGTGAAAGATCAGAAGCTGGTGTTTGTGTAT
CTACTAGTGGTAGATGGGTACTTAACAATGATTATTACAGATCTTTACCAGGAGTTTTCTGTGGTGTAGATGCTGTA
AATTTACTTACTAATATGTTTACACCACTAATTCAACCTATTGGTGCTTTGGACATATCAGCATCTATAGTAGCTGGT
GGTATTGTAGCTATCGTAGTAACATGCCTTGCCTACTATTTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCA
TGTAGTTGCCTTTAATACTTTACTATTCCTTATGTCATTCACTGTACTCTGTTTAACACCAGTTTACTCATTCTTACCT
GGTGTTTATTCTGTTATTTACTTGTACTTGACATTTTATCTTACTAATGATGTTTCTTTTTTAGCACATATTCAGTGGA
TGGTTATGTTCACACCTTTAGTACCTTTCTGGATAACAATTGCTTATATCATTTGTATTTCCACAAAGCATTTCTATTG
GTTCTTTAGTAATTACCTAAAGAGACGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGT
GCACCTTTTTGTTAAATAAAGAAATGTATCTAAAGTTGCGTAGTGATGTGCTATTACCTCTTACGCAATATAATAGA
TACTTAGCTCTTTATAATAAGTACAAGTATTTTAGTGGAGCAATGGATACAACTAGCTACAGAGAAGCTGCTTGTT
GTCATCTCGCAAAGGCTCTCAATGACTTCAGTAACTCAGGTTCTGATGTTCTTTACCAACCACCACAAACCTCTATC

Fig. 15

ACCTCAGCTGTTTTGCAGAGTGGTTTTAGAAAAATGGCATTCCCATCTGGTAAAGTTGAGGGTTGTATGGTACAAG
TAACTTGTGGTACAACTACACTTAACGGTCTTTGGCTTGATGACGTAGTTTACTGTCCAAGACATGTGATCTGCACC
TCTGAAGACATGCTTAACCCTAATTATGAAGATTTACTCATTCGTAAGTCTAATCATAATTTCTTGGTACAGGCTGG
TAATGTTCAACTCAGGGTTATTGGACATTCTATGCAAAATTGTGTACTTAAGCTTAAGGTTGATACAGCCAATCCTA
AGACACCTAAGTATAAGTTTGTTCGCATTCAACCAGGACAGACTTTTTCAGTGTTAGCTTGTTACAATGGTTCACCA
TCTGGTGTTTACCAATGTGCTATGAGGCCCAATTTCACTATTAAGGGTTCATTCCTTAATGGTTCATGTGGTAGTGT
TGGTTTTAACATAGATTATGACTGTGTCTCTTTTTGTTACATGCACCATATGGAATTACCAACTGGAGTTCATGCTG
GCACAGACTTAGAAGGTAACTTTTATGGACCTTTTGTTGACAGGCAAACAGCACAAGCAGCTGGTACGGACACAA
CTATTACAGTTAATGTTTTAGCTTGGTTGTACGCTGCTGTTATAAATGGAGACAGGTGGTTTCTCAATCGATTTACC
ACAACTCTTAATGACTTTAACCTTGTGGCTATGAAGTACAATTATGAACCTCTAACACAAGACCATGTTGACATACT
AGGACCTCTTTCTGCTCAAACTGGAATTGCCGTTTTAGATATGTGTGCTTCATTAAAAGAATTACTGCAAAATGGTA
TGAATGGACGTACCATATTGGGTAGTGCTTTATTAGAAGATGAATTTACACCTTTTGATGTTGTTAGACAATGCTCA
GGTGTTACTTTCCAAAGTGCAGTGAAAAGAACAATCAAGGGTACACACCACTGGTTGTTACTCACAATTTTGACTT
CACTTTTAGTTTTAGTCCAGAGTACTCAATGGTCTTTGTTCTTTTTTTTGTATGAAAATGCCTTTTTACCTTTTGCTAT
GGGTATTATTGCTATGTCTGCTTTTGCAATGATGTTTGTCAAACATAAGCATGCATTTCTCTGTTTGTTTTTGTTACC
TTCTCTTGCCACTGTAGCTTATTTTAATATGGTCTATATGCCTGCTAGTTGGGTGATGCGTATTATGACATGGTTGG
ATATGGTTGATACTAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACTAATC
CTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGGACACTTATGAATGTCTTGACACTCGTTT
ATAAAGTTTATTATGGTAATGCTTTAGATCAAGCCATTTCCATGTGGGCTCTTATAATCTCTGTTACTTCTAACTACT
CAGGTGTAGTTACAACTGTCATGTTTTTGGCCAGAGGTATTGTTTTTATGTGTGTTGAGTATTGCCCTATTTTCTTCA
TAACTGGTAATACACTTCAGTGTATAATGCTAGTTTATTGTTTCTTAGGCTATTTTTGTACTTGTTACTTTGGCCTCTT
TTGTTTACTCAACCGCTACTTTAGACTGACTCTTGGTGTTTATGATTACTTAGTTTCTACACAGGAGTTTAGATATAT

Fig. 15

GAATTCACAGGGACTACTCCCACCCAAGAATAGCATAGATGCCTTCAAACTCAACATTAAATTGTTGGGTGTTGGT
GGCAAACCTTGTATCAAAGTAGCCACTGTACAGTCTAAAATGTCAGATGTAAAGTGCACATCAGTAGTCTTACTCT
CAGTTTTGCAACAACTCAGAGTAGAATCATCATCTAAATTGTGGGCTCAATGTGTCCAGTTACACAATGACATTCTC
TTAGCTAAAGATACTACTGAAGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGT
AGACATAAACAAGCTTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTTAGTTCC
CTTCCATCATATGCAGCTTTTGCTACTGCTCAAGAAGCTTATGAGCAGGCTGTTGCTAATGGTGATTCTGAAGTTGT
TCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGAATTTGACCGTGATGCAGCCATGCAACGTAAGTTG
GAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTGAGGACAAGAGGGCAAAAGTTAC
TAGTGCTATGCAGACAATGCTTTTCACTATGCTTAGAAAGTTGGATAATGATGCACTCAACAACATTATCAACAATG
CAAGAGATGGTTGTGTTCCCTTGAACATAATACCTCTTACAACAGCAGCCAAACTAATGGTTGTCATACCAGACTAT
AACACATATAAAAATACGTGTGATGGTACAACATTTACTTATGCATCAGCATTGTGGGAAATCCAACAGGTTGTAG
ATGCAGATAGTAAAATTGTTCAACTTAGTGAAATTAGTATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTA
ACAGCTTTAAGGGCCAATTCTGCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTT
GTGCTGCCGGTACTACACAAACTGCTTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTA
GGTTTGTACTTGCACTGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGATGGAACTGGTAC
TATCTATACAGAACTGGAACCACCTTGTAGGTTTGTTACAGACACACCTAAAGGTCCTAAAGTGAAGTATTTATACT
TTATTAAAGGATTAAACAACCTAAATAGAGGTATGGTACTTGGTAGTTTAGCTGCCACAGTACGTCTACAAGCTGG
TAATGCAACAGAAGTGCCTGCCAATTCAACTGTATTATCTTTCTGTGCTTTTGCTGTAGATGCTGCTAAAGCTTACA
AAGATTATCTAGCTAGTGGGGGACAACCAATCACTAATTGTGTTAAGATGTTGTGTACACACACTGGTACTGGTCA
GGCAATAACAGTTACACCGGAAGCCAATATGGATCAAGAATCCTTTGGTGGTGCATCGTGTTGTCTGTACTGCCGT
TGCCACATAGATCATCCAAATCCTAAAGGATTTTGTGACTTAAAAGGTAAGTATGTACAAATACCTACAACTTGTGC
TAATGACCCTGTGGGTTTTACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGT

Fig. 15

TGTGATCAACTCCGCGAACCCATGCTTCAGTCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTGTAAGTGC
AGCCCGTCTTACACCGTGCGGCACAGGCACTAGTACTGATGTCGTATACAGGGCTTTTGACATCTACAATGATAAA
GTAGCTGGTTTTGCTAAATTCCTAAAAACTAATTGTTGTCGCTTCCAAGAAAAGGACGAAGATGACAATTTAATTG
ATTCTTACTTTGTAGTTAAGAGACACACTTTCTCTAACTACCAACATGAAGAAACAATTTATAATTTACTTAAGGATT
GTCCAGCTGTTGCTAAACATGACTTCTTTAAGTTTAGAATAGACGGTGACATGGTACCACATATATCACGTCAACGT
CTTACTAAATACACAATGGCAGACCTCGTCTATGCTTTAAGGCATTTTGATGAAGGTAATTGTGACACATTAAAAG
AAATACTTGTCACATACAATTGTTGTGATGATGATTATTTCAATAAAAAGGACTGGTATGATTTTGTAGAAAACCCA
GATATATTACGCGTATACGCCAACTTAGGTGAACGTGTACGCCAAGCTTTGTTAAAAACAGTACAATTCTGTGATG
CCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTTCGG
TGATTTCATACAAACCACGCCAGGTAGTGGAGTTCCTGTTGTAGATTCTTATTATTCATTGTTAATGCCTATATTAAC
CTTGACCAGGGCTTTAACTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTTGTTA
AAATATGACTTCACGGAAGAGAGGTTAAAACTCTTTGACCGTTATTTTAAATATTGGGATCAGACATACCACCCAA
ATTGTGTTAACTGTTTGGATGACAGATGCATTCTGCATTGTGCAAACTTTAATGTTTTATTCTCTACAGTGTTCCCAC
CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGATGGTGTTCCATTTGTAGTTTCAACTGGATACCACTTC
AGAGAGCTAGGTGTTGTACATAATCAGGATGTAAACTTACATAGCTCTAGACTTAGTTTTAAGGAATTACTTGTGT
ATGCTGCTGACCCTGCTATGCACGCTGCTTCTGGTAATCTATTACTAGATAAACGCACTACGTGCTTTTCAGTAGCT
GCACTTACTAACAATGTTGCTTTTCAAACTGTCAAACCCGGTAATTTTAACAAAGACTTCTATGACTTTGCTGTGTCT
AAGGGTTTCTTTAAGGAAGGAAGTTCTGTTGAATTAAAACACTTCTTCTTTGCTCAGGATGGTAATGCTGCTATCAG
CGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAGACAACTACTATTTGTAGTTGAAGTTGTTG
ATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAATGCTAACCAAGTCATCGTCAACAACCTAGACAAATCAGCT
GGTTTTCCATTTAATAAATGGGGTAAGGCTAGACTTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACTTTT
CGCATATACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGA

Fig. 15

GCTCGCACCGTAGCTGGTGTCTCTATCTGTAGTACTATGACCAATAGACAGTTTCATCAAAAATTATTGAAATCAAT
AGCCGCCACTAGAGGAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTTGGCACAACATGTTAAAAAC
TGTTTATAGTGATGTAGAAAACCCTCACCTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAACATG
CTTAGAATTATGGCCTCACTTGTTCTTGCTCGCAAACATACAACGTGTTGTAGCTTGTCACACCGTTTCTATAGATTA
GCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTATATGTTAAACCAGGTGGAACCT
CATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTTTTTAACATTTGTCAAGCTGTCACGGCCAATGTTAATGCA
CTTTTATCTACTGATGGTAACAAAATTGCCGATAAGTATGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTA
TAGAAATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGCATATTTGCGTAAACATTTCTCAATGATGATAC
TCTCTGACGATGCTGTTGTGTGTTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAG
TCAGTTCTTTATTATCAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACC
TCATGAATTTTGCTCTCAACATACAATGCTAGTTAAACAGGGTGATGATTATGTGTACCTTCCTTACCCAGATCCAT
CAAGAATCCTAGGGGCCGGCTGTTTTGTAGATGATATCGTAAAAACAGATGGTACACTTATGATTGAACGGTTCGT
GTCTTTAGCTATAGATGCTTACCCACTTACTAAACATCCTAATCAGGAGTATGCTGATGTCTTTCATTTGTACTTACA
ATACATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGTTATGCTTACTAATGATAAC
ACTTCAAGGTATTGGGAACCTGAGTTTTATGAGGCTATGTACACACCGCATACAGTCTTACAGGCTGTTGGGGCTT
GTGTTCTTTGCAATTCACAGACTTCATTAAGATGTGGTGCTTGCATACGTAGACCATTCTTATGTTGTAAATGCTGT
TACGACCATGTCATATCAACATCACATAAATTAGTCTTGTCTGTTAATCCGTATGTTTGCAATGCTCCAGGTTGTGAT
GTCACAGATGTGACTCAACTTTACTTAGGAGGTATGAGCTATTATTGTAAATCACATAAACCACCCATTAGTTTTCC
ATTGTGTGCTAATGGACAAGTTTTTGGTTTATATAAAAATACATGTGTTGGTAGCGATAATGTTACTGACTTTAATG
CAATTGCAACATGTGACTGGACAAATGCTGGTGATTACATTTTAGCTAACACCTGTACTGAAAGACTCAAGCTTTTT
GCAGCAGAAACGCTCAAAGCTACTGAGGAGACATTTAAACTGTCTTATGGTATTGCTACTGTACGTGAAGTGCTGT
CTGACAGAGAATTACATCTTTCATGGGAAGTTGGTAAACCTAGACCACCACTTAACCGAAATTATGTCTTTACTGGT

Fig. 15

TATCGTGTAACTAAAAACAGTAAAGTACAAATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGTGATGCTGTT
GTTTACCGAGGTACAACAACTTACAAATTAAATGTTGGTGATTATTTTGTGCTGACATCACATACAGTAATGCCATT
AAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATACCCAACACTCAATATCTCAGATG
AGTTTTCTAGCAATGTTGCAAATTATCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGGACCACCTGGTAC
TGGTAAGAGTCATTTTGCTATTGGCCTAGCTCTCTACTACCCTTCTGCTCGCATAGTGTATACAGCTTGCTCTCATGC
CGCTGTTGATGCACTATGTGAGAAGGCATTAAAATATTTGCCTATAGATAAATGTAGTAGAATTATACCTGCACGT
GCTCGTGTAGAGTGTTTTGATAAATTCAAAGTGAATTCAACATTAGAACAGTATGTCTTTTGTACTGTAAATGCATT
GCCTGAGACGACAGCAGATATAGTTGTCTTTGATGAAATTTCAATGGCCACAAATTATGATTTGAGTGTTGTCAAT
GCCAGATTACGTGCTAAGCACTATGTGTACATTGGCGACCCTGCTCAATTACCTGCACCACGCACATTGCTAACTAA
GGGCACACTAGAACCAGAATATTTCAATTCAGTGTGTAGACTTATGAAAACTATAGGTCCAGACATGTTCCTCGGA
ACTTGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAGCTTAAAGCACATAA
AGACAAATCAGCTCAATGCTTTAAAATGTTTTATAAGGGTGTTATCACGCATGATGTTTCATCTGCAATTAACAGGC
CACAAATAGGCGTGGTAAGAGAATTCCTTACACGTAACCCTGCTTGGAGAAAAGCTGTCTTTATTTCACCTTATAAT
TCACAGAATGCTGTAGCCTCAAAGATTTTGGGACTACCAACTCAAACTGTTGATTCATCACAGGGCTCAGAATATG
ACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTTGTAATGTAAACAGATTTAATGTTGCTATTACCAGA
GCAAAAGTAGGCATACTTTGCATAATGTCTGATAGAGACCTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCC
ACGTAGGAATGTGGCAACTTTACAAGCTGAAAATGTAACAGGACTCTTTAAAGATTGTAGTAAGGTAATCACTGG
GTTACATCCTACACAGGCACCTACACACCTCAGTGTTGACACTAAATTCAAAACTGAAGGTTTATGTGTTGACATAC
CTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTTAAAATGAATTATCAAGTTAATGG
TTACCCTAACATGTTTATCACCCGCGAAGAAGCTATAAGACATGTACGTGCATGGATTGGCTTCGATGTCGAGGGG
TGTCATGCTACTAGAGAAGCTGTTGGTACCAATTTACCTTTACAGCTAGGTTTTTCTACAGGTGTTAACCTAGTTGC
TGTACCTACAGGTTATGTTGATACACCTAATAATACAGATTTTTCCAGAGTTAGTGCTAAACCACCGCCTGGAGATC

Fig. 15

AATTTAAACACCTCATACCACTTATGTACAAAGGACTTCCTTGGAATGTAGTGCGTATAAAGATTGTACAAATGTTA
AGTGACACACTTAAAAATCTCTCTGACAGAGTCGTATTTGTCTTATGGGCACATGGCTTTGAGTTGACATCTATGAA
GTATTTTGTGAAAATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCTTCAG
ACACTTATGCCTGTTGGCATCATTCTATTGGATTTGATTACGTCTATAATCCGTTTATGATTGATGTTCAACAATGGG
GTTTTACAGGTAACCTACAAAGCAACCATGATCTGTATTGTCAAGTCCATGGTAATGCACATGTAGCTAGTTGTGA
TGCAATCATGACTAGGTGTCTAGCTGTCCACGAGTGCTTTGTTAAGCGTGTTGACTGGACTATTGAATATCCTATAA
TTGGTGATGAACTGAAGATTAATGCGGCTTGTAGAAAGGTTCAACACATGGTTGTTAAAGCTGCATTATTAGCAGA
CAAATTCCCAGTTCTTCACGACATTGGTAACCCTAAAGCTATTAAGTGTGTACCTCAAGCTGATGTAGAATGGAAG
TTCTATGATGCACAGCCTTGTAGTGACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGA
CAAATTCACAGATGGTGTATGCCTATTTTGGAATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTTGTAGAT
TTGACACTAGAGTGCTATCTAACCTTAACTTGCCTGGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTC
CACACACCAGCTTTTGATAAAAGTGCTTTTGTTAATTTAAAACAATTACCATTTTTCTATTACTCTGACAGTCCATGT
GAGTCTCATGGAAAACAAGTAGTGTCAGATATAGATTATGTACCACTAAAGTCTGCTACGTGTATAACACGTTGCA
ATTTAGGTGGTGCTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTCGATGCTTATAACATGATGATCTCA
GCTGGCTTTAGCTTGTGGGTTTACAAACAATTTGATACTTATAACCTCTGGAACACTTTTACAAGACTTCAGAGTTT
AGAAAATGTGGCTTTTAATGTTGTAAATAAGGGACACTTTGATGGACAACAGGGTGAAGTACCAGTTTCTATCATT
AATAACACTGTTTACACAAAAGTTGATGGTGTTGATGTAGAATTGTTTGAAAATAAAACAACATTACCTGTTAATGT
AGCATTTGAGCTTTGGGCTAAGCGCAACATTAAACCAGTACCAGAGGTGAAAATACTCAATAATTTGGGTGTGGA
CATTGCTGCTAATACTGTGATCTGGGACTACAAAAGAGATGCTCCAGCACATATATCTACTATTGGTGTTTGTTCTA
TGACTGACATAGCCAAGAAACCAACTGAAACGATTTGTGCACCACTCACTGTCTTTTTTGATGGTAGAGTTGATGG
TCAAGTAGACTTATTTAGAAATGCCCGTAATGGTGTTCTTATTACAGAAGGTAGTGTTAAAGGTTTACAACCATCTG
TAGGTCCCAAACAAGCTAGTCTTAATGGAGTCACATTAATTGGAGAAGCCGTAAAAACACAGTTCAATTATTATAA

Fig. 15

GAAAGTTGATGGTGTTGTCCAACAATTACCTGAAACTTACTTTACTCAGAGTAGAAATTTACAAGAATTTAAACCCA
GGAGTCAAATGGAAATTGATTTCTTAGAATTAGCTATGGATGAATTCATTGAACGGTATAAATTAGAAGGCTATGC
CTTCGAACATATCGTTTATGGAGATTTTAGTCATAGTCAGTTAGGTGGTTTACATCTACTGATTGGACTAGCTAAAC
GTTTTAAGGAATCACCTTTTGAATTAGAAGATTTTATTCCTATGGACAGTACAGTTAAAAACTATTTCATAACAGAT
GCGCAAACAGGTTCATCTAAGTGTGTGTGTTCTGTTATTGATTTATTACTTGATGATTTTGTTGAAATAATAAAATC
CCAAGATTTATCTGTAGTTTCTAAGGTTGTCAAAGTGACTATTGACTATACAGAAATTTCATTTATGCTTTGGTGTA
AAGATGGCCATGTAGAAACATTTTACCCAAAATTACAATCTAGTCAAGCGTGGCAACCGGGTGTTGCTATGCCTAA
TCTTTACAAAATGCAAAGAATGCTATTAGAAAAGTGTGACCTTCAAAATTATGGTGATAGTGCAACATTACCTAAA
GGCATAATGATGAATGTCGCAAAATATACTCAACTGTGTCAATATTTAAACACATTAACATTAGCTGTACCCTATAA
TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTACAGCTGTTTTAAGACAGTGGTTG
CCTACGGGTACGCTGCTTGTCGATTCAGATCTTAATGACTTTGTCTCTGATGCAGATTCAACTTTGATTGGTGATTG
TGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGATATGTACGACCCTAAGACTAAAAATGTTACA
AAAGAAAATGACTCTAAAGAGGGTTTTTTCACTTACATTTGTGGGTTTATACAACAAAAGCTAGCTCTTGGAGGTT
CCGTGGCTATAAAGATAACAGAACATTCTTGGAATGCTGATCTTTATAAGCTCATGGGACACTTCGCATGGTGGAC
AGCCTTTGTTACTAATGTGAATGCGTCATCATCTGAAGCATTTTTAATTGGATGTAATTATCTTGGCAAACCACGCG
AACAAATAGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTCCTAT
TCTTTATTTGACATGAGTAAATTTCCCCTTAAATTAAGGGGTACTGCTGTTATGTCTTTAAAAGAAGGTCAAATCAA
TGATATGATTTTATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATG
TTCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTA
CAACCAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTTCAGAT
CCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTGGTTCCATGCTATACATGTCTC
TGGGACCAATGGTACTAAGAGGTTTGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCACTGAGA

Fig. 15

AGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGACCCAGTCCCTACTTATTGTTAATAAC
GCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAAC
AACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGC
CTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTTGTGTTTAAGAATATTGATGG
TTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACC
ATTGGTAGATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCC
TGGTGATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTAT
TAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTAC
GTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGTTA
GATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCATCTGTTTATGCTTGGA
ACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCATTTTCCACTTTTAAGTGTT
ATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGAT
GAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAG
GCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTT
AGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATG
GTGTTGAAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCAT
ACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTG
GTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTT
TCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATTC
TTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAACAAATACTTCTAACCAGGTTGCTGTT
CTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTTTAT
TCTACAGGTTCTAATGTTTTTCAAACACGTGCAGGCTGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGT

Fig. 15

GTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGT
AGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGC
CATACCCACAAATTTTACTATTAGTGTTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTA
CAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATCTTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAAC
CGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTACA
AAACACCACCAATTAAAGATTTTGGTGGTTTTAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGG
TCATTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGCCT
TGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCCTTACTGTTTTGCCACCTTTGCTCACAG
ATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCT
GCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGA
GAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAATTCAAGACTCACTTTCTTCCACAGCAAGTG
CACTTGGAAAACTTCAAGATGTGGTCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAA
TTTTGGTGCAATTTCAAGTGTTTTAAATGATATCCTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATA
GGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAG
CTTCTGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAA
GGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACA
AGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTT
CAAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGT
GTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTCAT
TCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACCAGATGTTGATTTAGGTGACATCTCTGGCATTAA
TGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCG
ATCTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTT

Fig. 15

GATTGCCATAGTAATGGTGACAATTATGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTT
GTGGATCCTGCTGCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAAC
GAACTTATGGATTTGTTTATGAGAATCTTCACAATTGGAACTGTAACTTTGAAGCAAGGTGAAATCAAGGATGCTA
CTCCTTCAGATTTTGTTCGCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTGGC
GTTGCACTTCTTGCTGTTTTTCAGAGCGCTTCCAAAATCATAACCCTCAAAAAGAGATGGCAACTAGCACTCTCCAA
GGGTGTTCACTTTGTTTGCAACTTGCTGTTGTTGTTTGTAACAGTTTACTCACACCTTTTGCTCGTTGCTGCTGGCCT
TGAAGCCCCTTTTCTCTATCTTTATGCTTTAGTCTACTTCTTGCAGAGTATAAACTTTGTAAGAATAATAATGAGGCT
TTGGCTTTGCTGGAAATGCCGTTCCAAAAACCCATTACTTTATGATGCCAACTATTTTCTTTGCTGGCATACTAATTG
TTACGACTATTGTATACCTTACAATAGTGTAACTTCTTCAATTGTCATTACTTCAGGTGATGGCACAACAAGTCCTAT
TTCTGAACATGACTACCAGATTGGTGGTTATACTGAAAAATGGGAATCTGGAGTAAAAGACTGTGTTGTATTACAC
AGTTACTTCACTTCAGACTATTACCAGCTGTACTCAACTCAATTGAGTACAGACACTGGTGTTGAACATGTTACCTT
CTTCATCTACAATAAAATTGTTGATGAGCCTGAAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTT
GTTAATCCAGTAATGGAACCAATTTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAGCTGATG
AGTACGAACTTACGAACTAAATATTATATTAGTTTTTCTGTTTGGAACTTTAATTTTAGCCATGGCAGATTCCAACG
GTACTATTACCGTTGAAGAGCTTAAAAAGCTCCTTGAACAATGGAACCTAGTAATAGGTTTCCTATTCCTTACATGG
ATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGGTTTTTGTATATAATTAAGTTAATTTTCCTCTGGCTGTTA
TGGCCAGTAACTTTAGCTTGTTTTGTGCTTGCTGCTGTTTACAGAATAAATTGGATCACCGGTGGAATTGCTATCGC
AATGGCTTGTCTTGTAGGCTTGATGTGGCTCAGCTACTTCATTGCTTCTTTCAGACTGTTTGCGCGTACGCGTTCCA
TGTGGTCATTCAATCCAGAAACTAACATTCTTCTCAACGTGCCACTCCATGGCACTATTCTGACCAGACCGCTTCTA
GAAAGTGAACTCGTAATCGGAGCTGTGATCCTTCGTGGACATCTTCGTATTGCTGGACACCATCTAGGACGCTGTG
ACATCAAGGACCTGCCTAAAGAAATCACTGTTGCTACATCACGAACGCTTTCTTATTACAAATTGGGAGCTTCGCA
GCGTGTAGCAGGTGACTCAGGTTTTGCTGCATACAGTCGCTACAGGATTGGCAACTATAAATTAAACACAGACCAT

Fig. 15

TCCAGTAGCAGTGACAATATTGCTTTGCTTGTACAGTAAGTGACAACAGATGTTTCATCTCGTTGACTTTCAGGTTA
CTATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGTTTCCATTTGGAATCTTGATTACATCATAAACCTCA
TAATTAAAAATTTATCTAAGTCACTAACTGAGAATAAATATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGA
TTAAACGAACATGAAAATTATTCTTTTCTTGGCACTGATAACACTCGCTACTTGTGAGCTTTATCACTACCAAGAGT
GTGTTAGAGGTACAACAGTACTTTTAAAAGAACCTTGCTCTTCTGGAACATACGAGGGCAATTCACCATTTCATCCT
CTAGCTGATAACAAATTTGCACTGACTTGCTTTAGCACTCAATTTGCTTTTGCTTGTCCTGACGGCGTAAAACACGT
CTATCAGTTACGTGCCAGATCAGTTTCACCTAAACTGTTCATCAGACAAGAGGAAGTTCAAGAACTTTACTCTCCAA
TTTTTCTTATTGTTGCGGCAATAGTGTTTATAACACTTTGCTTCACACTCAAAAGAAAGACAGAATGATTGAACTTTC
ATTAATTGACTTCTATTTGTGCTTTTTAGCCTTTCTGCTATTCCTTGTTTTAATTATGCTTATTATCTTTTGGTTCTCAC
TTGAACTGCAAGATCATAATGAAACTTGTCACGCCTAAACGAACATGAAATTTCTTGTTTTCTTAGGAATCATCACA
ACTGTAGCTGCATTTCACCAAGAATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTG
TCCTATTCACTTCTATTCTAAATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGG
ATGAGGCTGGTTCTAAATCACCCATTCAGTACATCGATATCGGTAATTATACAGTTTCCTGTTTACCTTTTACAATTA
ATTGCCAGGAACCTAAATTGGGTAGTCTTGTAGTGCGTTGTTCGTTCTATGAAGACTTTTTAGAGTATCATGACGTT
CGTGTTGTTTTAGATTTCATCTAAACGAACAAACTAAAATGTCTGATAATGGACCCCAAAATCAGCGAAATGCACC
CCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAA
ACAACGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGAC
CTTAAATTCCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAA
GAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAG
GAACTGGGCCAGAAGCTGGACTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCT
TGAATACACCAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGG
AACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTCCTCATCACG

Fig. 15

TAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCAATGG
CGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAA
CAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTGCC
ACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTTTGGGGACCAG
GAACTAATCAGACAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTTGCCCCCAGCGCTTCAGCGTTCT
TCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGG
ATGACAAAGATCCAAATTTCAAAGATCAAGTCATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCA
ACAGAGCCTAAAAAGGACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCA
AACTGTGACTCTTCTTCCTGCTGCAGATTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGACT
CAACTCAGGCCTAAACTCATGCAGACCACACAAGGCAGATGGGCTATATAAACGTTTTCGCTTTTCCGTTTACGAT
ATATAGTCTACTCTTGTGCAGAATGAATTCTCGTAACTACATAGCACAAGTAGATGTAGTTAACTTTAATCTCACAT
AGCAATCTTTAATCAGTGTGTAACATTAGGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAG
TACGATCGAGTGTACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAATTT
TAGTAGTGCTATCCCCATGTGATTTTAATAGCTTCTTAGGAGAATGACAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAA (SEQ ID NO:15)

ΔEM GENOME

ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTT
AAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTG
ACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACA
TCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGT

Fig. 15

CCAACTCAGTTTGCCTGTTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTA
TCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTG
AACAGCCCTATGTGTTCATCAAACGTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGC
AGAACTCGAAGGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACC
AGTGGCTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCT
AAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACACTAAACAT
AGCAGTGGTGTTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACTTCT
GTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTGTCC
GAACAACTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTGCTTGGTAC
ACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAAATTAAATTGGCAAAGAAATTTGACACCTTCA
ATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAA
GCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCCTTT
CAACTCTCATGAAGTGTGATCATTGTGGTGAAACTTCATGGCAGACGGGCGATTTTGTTAAAGCCACTTGCGAATT
TTGTGGCACTGAGAATTTGACTAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTT
ATTGTCCAGCATGTCACAATTCAGAAGTAGGACCTGAGCATAGTCTTGCCGAATACCATAATGAATCTGGCTTGAA
AACCATTCTTCGTAAGGGTGGTCGCACTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAGT
GTGCCTATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGTTGGAGAAGGTTCCGAAG
GTCTTAATGACAACCTTCTTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGACTTTAAACTTAAT
GAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCTTCCACAAGTGCTTTTGTGGAAACTGTGAAAGGTTTGGATTA
TAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTTTAAAGTTACAAAAGGAAAAGCTAAAAAAGGTGCCTGG
AATATTGGTGAACAGAAATCAATACTGAGTCCTCTTTATGCATTTGCATCAGAGGCTGCTCGTGTTGTACGATCAAT
TTTCTCCCGCACTCTTGAAACTGCTCAAAATTCTGTGCGTGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAA

Fig. 15

TTTCACAGTATTCACTGAGACTCATTGATGCTATGATGTTCACATCTGATTTGGCTACTAACAATCTAGTTGTAATG
GCCTACATTACAGGTGGTGTTGTTCAGTTGACTTCGCAGTGGCTAACTAACATCTTTGGCACTGTTTATGAAAAACT
CAAACCCGTCCTTGATTGGCTTGAAGAGAAGTTTAAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTGT
TAAATTTATCTCAACCTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGGAGAGT
GTTCAGACATTCTTTAAGCTTGTAAATAAATTTTTGGCTTTGTGTGCTGACTCTATCATTATTGGTGGAGCTAAACTT
AAAGCCTTGAATTTAGGTGAAACATTTGTCACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAA
GAAACTGGCCTACTCATGCCTCTAAAAGCCCCAAAAGAAATTATCTTCTTAGAGGGAGAAACACTTCCCACAGAAG
TGTTAACAGAGGAAGTTGTCTTGAAAACTGGTGATTTACAACCATTAGAACAACCTACTAGTGAAGCTGTTGAAGC
TCCATTGGTTGGTACACCAGTTTGTATTAACGGGCTTATGTTGCTCGAAATCAAAGACACAGAAAAGTACTGTGCC
CTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTTGGTG
ATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTTGAACTTGATGAAAGGATTGATAAAG
TACTTAATGAGAAGTGCTCTGCCTATACAGTTGAACTCGGTACAGAAGTAAATGAGTTCGCCTGTGTTGTGGCAGA
TGCTGTCATAAAAACTTTGCAACCAGTATCTGAATTACTTACACCACTGGGCATTGATTTAGATGAGTGGAGTATG
GCTACATACTACTTATTTGATGAGTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTCTACCCTCCAGAT
GAGGATGAAGAAGAAGGTGATTGTGAAGAAGAAGAGTTTGAGCCATCAACTCAATATGAGTATGGTACTGAAGA
TGATTACCAAGGTAAACCTTTGGAATTTGGTGCCACTTCTGCTGCTCTTCAACCTGAAGAAGAGCAAGAAGAAGAT
TGGTTAGATGATGATAGTCAACAAACTGTTGGTCAACAAGACGGCAGTGAGGACAATCAGACAACTACTATTCAA
ACAATTGTTGAGGTTCAACCTCAATTAGAGATGGAACTTACACCAGTTGTTCAGACTATTGAAGTGAATAGTTTTA
GTGGTTATTTAAAACTTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAAAAAGGTAAAACC
AACAGTGGTTGTTAATGCAGCCAATGTTTACCTTAAACATGGAGGAGGTGTTGCAGGAGCCTTAAATAAGGCTAC
TAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCACTTAAAGTGGGTGGTAGTTGTGTT
TTAAGCGGACACAATCTTGCTAAACACTGTCTTCATGTTGTCGGCCCAAATGTTAACAAAGGTGAAGACATTCAAC

Fig. 15

TTCTTAAGAGTGCTTATGAAAATTTTAATCAGCACGAAGTTCTACTTGCACCATTATTATCAGCTGGTATTTTTGGTG
CTGACCCTATACATTCTTTAAGAGTTTGTGTAGATACTGTTCGCACAAATGTCTACTTAGCTGTCTTTGATAAAAATC
TCTATGACAAACTTGTTTCAAGCTTTTTGGAAATGAAGAGTGAAAAGCAAGTTGAACAAAAGATCGCTGAGATTCC
TAAAGAGGAAGTTAAGCCATTTATAACTGAAAGTAAACCTTCAGTTGAACAGAGAAAACAAGATGATAAGAAAAT
CAAAGCTTGTGTTGAAGAAGTTACAACAACTCTGGAAGAAACTAAGTTCCTCACAGAAAACTTGTTACTTTATATTG
ACATTAATGGCAATCTTCATCCAGATTCTGCCACTCTTGTTAGTGACATTGACATCACTTTCTTAAAGAAAGATGCTC
CATATATAGTGGGTGATGTTGTTCAAGAGGGTGTTTTAACTGCTGTGGTTATACCTACTAAAAAGGCTGGTGGCAC
TACTGAAATGCTAGCGAAAGCTTTGAGAAAAGTGCCAACAGACAATTATATAACCACTTACCCGGGTCAGGGTTTA
AATGGTTACACTGTAGAGGAGGCAAAGACAGTGCTTAAAAAGTGTAAAAGTGCCTTTTACATTCTACCATCTATTA
TCTCTAATGAGAAGCAAGAAATTCTTGGAACTGTTTCTTGGAATTTGCGAGAAATGCTTGCACATGCAGAAGAAAC
ACGCAAATTAATGCCTGTCTGTGTGGAAACTAAAGCCATAGTTTCAACTATACAGCGTAAATATAAGGGTATTAAA
ATACAAGAGGGTGTGGTTGATTATGGTGCTAGATTTTACTTTTACACCAGTAAAACAACTGTAGCGTCACTTATCA
ACACACTTAACGATCTAAATGAAACTCTTGTTACAATGCCACTTGGCTATGTAACACATGGCTTAAATTTGGAAGAA
GCTGCTCGGTATATGAGATCTCTCAAAGTGCCAGCTACAGTTTCTGTTTCTTCACCTGATGCTGTTACAGCGTATAA
TGGTTATCTTACTTCTTCTTCTAAAACACCTGAAGAACATTTTATTGAAACCATCTCACTTGCTGGTTCCTATAAAGA
TTGGTCCTATTCTGGACAATCTACACAACTAGGTATAGAATTTCTTAAGAGAGGTGATAAAAGTGTATATTACACTA
GTAATCCTACCACATTCCACCTAGATGGTGAAGTTATCACCTTTGACAATCTTAAGACACTTCTTTCTTTGAGAGAA
GTGAGGACTATTAAGGTGTTTACAACAGTAGACAACATTAACCTCCACACGCAAGTTGTGGACATGTCAATGACAT
ATGGACAACAGTTTGGTCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCATAATTCACATGAAGG
TAAAACATTTTATGTTTTACCTAATGATGACACTCTACGTGTTGAGGCTTTTGAGTACTACCACACAACTGATCCTA
GTTTTCTGGGTAGGTACATGTCAGCATTAAATCACACTAAAAAGTGGAAATACCCACAAGTTAATGGTTTAACTTCT
ATTAAATGGGCAGATAACAACTGTTATCTTGCCACTGCATTGTTAACACTCCAACAAATAGAGTTGAAGTTTAATCC

Fig. 15

ACCTGCTCTACAAGATGCTTATTACAGAGCAAGGGCTGGTGAAGCTGCTAACTTTTGTGCACTTATCTTAGCCTACT
GTAATAAGACAGTAGGTGAGTTAGGTGATGTTAGAGAAACAATGAGTTACTTGTTTCAACATGCCAATTTAGATTC
TTGCAAAAGAGTCTTGAACGTGGTGTGTAAAACTTGTGGACAACAGCAGACAACCCTTAAGGGTGTAGAAGCTGT
TATGTACATGGGCACACTTTCTTATGAACAATTTAAGAAAGGTGTTCAGATACCTTGTACGTGTGGTAAACAAGCT
ACAAAATATCTAGTACAACAGGAGTCACCTTTTGTTATGATGTCAGCACCACCTGCTCAGTATGAACTTAAGCATG
GTACATTTACTTGTGCTAGTGAGTACACTGGTAATTACCAGTGTGGTCACTATAAACATATAACTTCTAAAGAAACT
TTGTATTGCATAGACGGTGCTTTACTTACAAAGTCCTCAGAATACAAAGGTCCTATTACGGATGTTTTCTACAAAGA
AAACAGTTACACAACAACCATAAAACCAGTTACTTATAAATTGGATGGTGTTGTTTGTACAGAAATTGACCCTAAG
TTGGACAATTATTATAAGAAAGACAATTCTTATTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCC
AAACGCAAGCTTCGATAATTTTAAGTTTGTATGTGATAATATCAAATTTGCTGATGATTTAAACCAGTTAACTGGTT
ATAAGAAACCTGCTTCAAGAGAGCTTAAAGTTACATTTTTCCCTGACTTAAATGGTGATGTGGTGGCTATTGATTAT
AAACACTACACACCCTCTTTTAAGAAAGGAGCTAAATTGTTACATAAACCTATTGTTTGGCATGTTAACAATGCAAC
TAATAAAGCCACGTATAAACCAAATACCTGGTGTATACGTTGTCTTTGGAGCACAAAACCAGTTGAAACATCAAAT
TCGTTTGATGTACTGAAGTCAGAGGACGCGCAGGGAATGGATAATCTTGCCTGCGAAGATCTAAAACCAGTCTCT
GAAGAAGTAGTGGAAAATCCTACCATACAGAAAGACGTTCTTGAGTGTAATGTGAAAACTACCGAAGTTGTAGGA
GACATTATACTTAAACCAGCAAATAATAGTTTAAAAATTACAGAAGAGGTTGGCCACACAGATCTAATGGCTGCTT
ATGTAGACAATTCTAGTCTTACTATTAAGAAACCTAATGAATTATCTAGAGTATTAGGTTTGAAAACCCTTGCTACT
CATGGTTTAGCTGCTGTTAATAGTGTCCCTTGGGATACTATAGCTAATTATGCTAAGCCTTTTCTTAACAAAGTTGTT
AGTACAACTACTAACATAGTTACACGGTGTTTAAACCGTGTTTGTACTAATTATATGCCTTATTTCTTTACTTTATTG
CTACAATTGTGTACTTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCTATGCCGACTACTATAGCAAAGAATAC
TGTTAAGAGTGTCGGTAAATTTTGTCTAGAGGCTTCATTTAATTATTTGAAGTCACCTAATTTTTCTAAACTGATAAA
TATTATAATTTGGTTTTTACTATTAAGTGTTTGCCTAGGTTCTTTAATCTACTCAACCGCTGCTTTAGGTGTTTTAATG

Fig. 15

TCTAATTTAGGCATGCCTTCTTACTGTACTGGTTACAGAGAAGGCTATTTGAACTCTACTAATGTCACTATTGCAAC
CTACTGTACTGGTTCTATACCTTGTAGTGTTTGTCTTAGTGGTTTAGATTCTTTAGACACCTATCCTTCTTTAGAAACT
ATACAAATTACCATTTCATCTTTTAAATGGGATTTAACTGCTTTTGGCTTAGTTGCAGAGTGGTTTTTGGCATATATT
CTTTTCACTAGGTTTTTCTATGTACTTGGATTGGCTGCAATCATGCAATTGTTTTTCAGCTATTTTGCAGTACATTTTA
TTAGTAATTCTTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCGATTTCAGCTATGGTTAGAATGTAC
ATCTTCTTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGACGGTTGTAATTCATCAACTTGTATG
ATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCCTTTTATG
TCTATGCTAATGGAGGTAAAGGCTTTTGCAAACTACACAATTGGAATTGTGTTAATTGTGATACATTCTGTGCTGGT
AGTACATTTATTAGTGATGAAGTTGCGAGAGACTTGTCACTACAGTTTAAAAGACCAATAAATCCTACTGACCAGT
CTTCTTACATCGTTGATAGTGTTACAGTGAAGAATGGTTCCATCCATCTTTACTTTGATAAAGCTGGTCAAAAGACT
TATGAAAGACATTCTCTCTCTCATTTTGTTAACTTAGACAACCTGAGAGCTAATAACACTAAAGGTTCATTGCCTATT
AATGTTATAGTTTTTGATGGTAAATCAAAATGTGAAGAATCATCTGCAAAATCAGCGTCTGTTTACTACAGTCAGCT
TATGTGTCAACCTATACTGTTACTAGATCAGGCATTAGTGTCTGATGTTGGTGATAGTGCGGAAGTTGCAGTTAAA
ATGTTTGATGCTTACGTTAATACGTTTTCATCAACTTTTAACGTACCAATGGAAAAACTCAAAACACTAGTTGCAAC
TGCAGAAGCTGAACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCTACTTTTATTTCAGCAGCTCGGCAAGGG
TTTGTTGATTCAGATGTAGAAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCAATCTGACATAGAAGTTAC
TGGCGATAGTTGTAATAACTATATGCTCACCTATAACAAAGTTGAAAACATGACACCCCGTGACCTTGGTGCTTGT
ATTGACTGTAGTGCGCGTCATATTAATGCGCAGGTAGCAAAAAGTCACAACATTGCTTTGATATGGAACGTTAAAG
ATTTCATGTCATTGTCTGAACAACTACGAAAACAAATACGTAGTGCTGCTAAAAAGAATAACTTACCTTTTAAGTTG
ACATGTGCAACTACTAGACAAGTTGTTAATGTTGTAACAACAAAGATAGCACTTAAGGGTGGTAAAATTGTTAATA
ATTGGTTGAAGCAGTTAATTAAAGTTACACTTGTGTTCCTTTTTGTTGCTGCTATTTTCTATTTAATAACACCTGTTCA
TGTCATGTCTAAACATACTGACTTTTCAAGTGAAATCATAGGATACAAGGCTATTGATGGTGGTGTCACTCGTGAC

Fig. 15

ATAGCATCTACAGATACTTGTTTTGCTAACAAACATGCTGATTTTGACACATGGTTTAGCCAGCGTGGTGGTAGTTA
TACTAATGACAAAGCTTGCCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTTGTCGTGCCTGGTTTGCCT
GGCACGATATTACGCACAACTAATGGTGACTTTTTGCATTTCTTACCTAGAGTTTTTAGTGCAGTTGGTAACATCTG
TTACACACCATCAAAACTTATAGAGTACACTGACTTTGCAACATCAGCTTGTGTTTTGGCTGCTGAATGTACAATTT
TTAAAGATGCTTCTGGTAAGCCAGTACCATATTGTTATGATACCAATGTACTAGAAGGTTCTGTTGCTTATGAAAGT
TTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTATTCAATTTCCTAACACCTACCTTGAAGGTTCTGTT
AGAGTGGTAACAACTTTTGATTCTGAGTACTGTAGGCACGGCACTTGTGAAAGATCAGAAGCTGGTGTTTGTGTAT
CTACTAGTGGTAGATGGGTACTTAACAATGATTATTACAGATCTTTACCAGGAGTTTTCTGTGGTGTAGATGCTGTA
AATTTACTTACTAATATGTTTACACCACTAATTCAACCTATTGGTGCTTTGGACATATCAGCATCTATAGTAGCTGGT
GGTATTGTAGCTATCGTAGTAACATGCCTTGCCTACTATTTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCA
TGTAGTTGCCTTTAATACTTTACTATTCCTTATGTCATTCACTGTACTCTGTTTAACACCAGTTTACTCATTCTTACCT
GGTGTTTATTCTGTTATTTACTTGTACTTGACATTTTATCTTACTAATGATGTTTCTTTTTTAGCACATATTCAGTGGA
TGGTTATGTTCACACCTTTAGTACCTTTCTGGATAACAATTGCTTATATCATTTGTATTTCCACAAAGCATTTCTATTG
GTTCTTTAGTAATTACCTAAAGAGACGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGT
GCACCTTTTTGTTAAATAAAGAAATGTATCTAAAGTTGCGTAGTGATGTGCTATTACCTCTTACGCAATATAATAGA
TACTTAGCTCTTTATAATAAGTACAAGTATTTTAGTGGAGCAATGGATACAACTAGCTACAGAGAAGCTGCTTGTT
GTCATCTCGCAAAGGCTCTCAATGACTTCAGTAACTCAGGTTCTGATGTTCTTTACCAACCACCACAAACCTCTATC
ACCTCAGCTGTTTTGCAGAGTGGTTTTAGAAAAATGGCATTCCCATCTGGTAAAGTTGAGGGTTGTATGGTACAAG
TAACTTGTGGTACAACTACACTTAACGGTCTTTGGCTTGATGACGTAGTTTACTGTCCAAGACATGTGATCTGCACC
TCTGAAGACATGCTTAACCCTAATTATGAAGATTTACTCATTCGTAAGTCTAATCATAATTTCTTGGTACAGGCTGG
TAATGTTCAACTCAGGGTTATTGGACATTCTATGCAAAATTGTGTACTTAAGCTTAAGGTTGATACAGCCAATCCTA
AGACACCTAAGTATAAGTTTGTTCGCATTCAACCAGGACAGACTTTTTCAGTGTTAGCTTGTTACAATGGTTCACCA

Fig. 15

TCTGGTGTTTACCAATGTGCTATGAGGCCCAATTTCACTATTAAGGGTTCATTCCTTAATGGTTCATGTGGTAGTGT
TGGTTTTAACATAGATTATGACTGTGTCTCTTTTTGTTACATGCACCATATGGAATTACCAACTGGAGTTCATGCTG
GCACAGACTTAGAAGGTAACTTTTATGGACCTTTTGTTGACAGGCAAACAGCACAAGCAGCTGGTACGGACACAA
CTATTACAGTTAATGTTTTAGCTTGGTTGTACGCTGCTGTTATAAATGGAGACAGGTGGTTTCTCAATCGATTTACC
ACAACTCTTAATGACTTTAACCTTGTGGCTATGAAGTACAATTATGAACCTCTAACACAAGACCATGTTGACATACT
AGGACCTCTTTCTGCTCAAACTGGAATTGCCGTTTTAGATATGTGTGCTTCATTAAAAGAATTACTGCAAAATGGTA
TGAATGGACGTACCATATTGGGTAGTGCTTTATTAGAAGATGAATTTACACCTTTTGATGTTGTTAGACAATGCTCA
GGTGTTACTTTCCAAAGTGCAGTGAAAAGAACAATCAAGGGTACACACCACTGGTTGTTACTCACAATTTTGACTT
CACTTTTAGTTTTAGTCCAGAGTACTCAATGGTCTTTGTTCTTTTTTTTGTATGAAAATGCCTTTTTACCTTTTGCTAT
GGGTATTATTGCTATGTCTGCTTTTGCAATGATGTTTGTCAAACATAAGCATGCATTTCTCTGTTTGTTTTTGTTACC
TTCTCTTGCCACTGTAGCTTATTTTAATATGGTCTATATGCCTGCTAGTTGGGTGATGCGTATTATGACATGGTTGG
ATATGGTTGATACTAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACTAATC
CTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGGACACTTATGAATGTCTTGACACTCGTTT
ATAAAGTTTATTATGGTAATGCTTTAGATCAAGCCATTTCCATGTGGGCTCTTATAATCTCTGTTACTTCTAACTACT
CAGGTGTAGTTACAACTGTCATGTTTTTGGCCAGAGGTATTGTTTTTATGTGTGTTGAGTATTGCCCTATTTTCTTCA
TAACTGGTAATACACTTCAGTGTATAATGCTAGTTTATTGTTTCTTAGGCTATTTTTGTACTTGTTACTTTGGCCTCTT
TTGTTTACTCAACCGCTACTTTAGACTGACTCTTGGTGTTTATGATTACTTAGTTTCTACACAGGAGTTTAGATATAT
GAATTCACAGGGACTACTCCCACCCAAGAATAGCATAGATGCCTTCAAACTCAACATTAAATTGTTGGGTGTTGGT
GGCAAACCTTGTATCAAAGTAGCCACTGTACAGTCTAAAATGTCAGATGTAAAGTGCACATCAGTAGTCTTACTCT
CAGTTTTGCAACAACTCAGAGTAGAATCATCATCTAAATTGTGGGCTCAATGTGTCCAGTTACACAATGACATTCTC
TTAGCTAAAGATACTACTGAAGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGT
AGACATAAACAAGCTTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCTATAGCCTCAGAGTTTAGTTCC

Fig. 15

CTTCCATCATATGCAGCTTTTGCTACTGCTCAAGAAGCTTATGAGCAGGCTGTTGCTAATGGTGATTCTGAAGTTGT
TCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGAATTTGACCGTGATGCAGCCATGCAACGTAAGTTG
GAAAAGATGGCTGATCAAGCTATGACCCAAATGTATAAACAGGCTAGATCTGAGGACAAGAGGGCAAAAGTTAC
TAGTGCTATGCAGACAATGCTTTTCACTATGCTTAGAAAGTTGGATAATGATGCACTCAACAACATTATCAACAATG
CAAGAGATGGTTGTGTTCCCTTGAACATAATACCTCTTACAACAGCAGCCAAACTAATGGTTGTCATACCAGACTAT
AACACATATAAAAATACGTGTGATGGTACAACATTTACTTATGCATCAGCATTGTGGGAAATCCAACAGGTTGTAG
ATGCAGATAGTAAAATTGTTCAACTTAGTGAAATTAGTATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTA
ACAGCTTTAAGGGCCAATTCTGCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCTT
GTGCTGCCGGTACTACACAAACTGCTTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGGGAGGTA
GGTTTGTACTTGCACTGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGATGGAACTGGTAC
TATCTATACAGAACTGGAACCACCTTGTAGGTTTGTTACAGACACACCTAAAGGTCCTAAAGTGAAGTATTTATACT
TTATTAAAGGATTAAACAACCTAAATAGAGGTATGGTACTTGGTAGTTTAGCTGCCACAGTACGTCTACAAGCTGG
TAATGCAACAGAAGTGCCTGCCAATTCAACTGTATTATCTTTCTGTGCTTTTGCTGTAGATGCTGCTAAAGCTTACA
AAGATTATCTAGCTAGTGGGGGACAACCAATCACTAATTGTGTTAAGATGTTGTGTACACACACTGGTACTGGTCA
GGCAATAACAGTTACACCGGAAGCCAATATGGATCAAGAATCCTTTGGTGGTGCATCGTGTTGTCTGTACTGCCGT
TGCCACATAGATCATCCAAATCCTAAAGGATTTTGTGACTTAAAAGGTAAGTATGTACAAATACCTACAACTTGTGC
TAATGACCCTGTGGGTTTTACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGT
TGTGATCAACTCCGCGAACCCATGCTTCAGTCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTGTAAGTGC
AGCCCGTCTTACACCGTGCGGCACAGGCACTAGTACTGATGTCGTATACAGGGCTTTTGACATCTACAATGATAAA
GTAGCTGGTTTTGCTAAATTCCTAAAAACTAATTGTTGTCGCTTCCAAGAAAAGGACGAAGATGACAATTTAATTG
ATTCTTACTTTGTAGTTAAGAGACACACTTTCTCTAACTACCAACATGAAGAAACAATTTATAATTTACTTAAGGATT
GTCCAGCTGTTGCTAAACATGACTTCTTTAAGTTTAGAATAGACGGTGACATGGTACCACATATATCACGTCAACGT

Fig. 15

CTTACTAAATACACAATGGCAGACCTCGTCTATGCTTTAAGGCATTTTGATGAAGGTAATTGTGACACATTAAAAG
AAATACTTGTCACATACAATTGTTGTGATGATGATTATTTCAATAAAAAGGACTGGTATGATTTTGTAGAAAACCCA
GATATATTACGCGTATACGCCAACTTAGGTGAACGTGTACGCCAAGCTTTGTTAAAAACAGTACAATTCTGTGATG
CCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTTCGG
TGATTTCATACAAACCACGCCAGGTAGTGGAGTTCCTGTTGTAGATTCTTATTATTCATTGTTAATGCCTATATTAAC
CTTGACCAGGGCTTTAACTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTTGTTA
AAATATGACTTCACGGAAGAGAGGTTAAAACTCTTTGACCGTTATTTTAAATATTGGGATCAGACATACCACCCAA
ATTGTGTTAACTGTTTGGATGACAGATGCATTCTGCATTGTGCAAACTTTAATGTTTTATTCTCTACAGTGTTCCCAC
CTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGATGGTGTTCCATTTGTAGTTTCAACTGGATACCACTTC
AGAGAGCTAGGTGTTGTACATAATCAGGATGTAAACTTACATAGCTCTAGACTTAGTTTTAAGGAATTACTTGTGT
ATGCTGCTGACCCTGCTATGCACGCTGCTTCTGGTAATCTATTACTAGATAAACGCACTACGTGCTTTTCAGTAGCT
GCACTTACTAACAATGTTGCTTTTCAAACTGTCAAACCCGGTAATTTTAACAAAGACTTCTATGACTTTGCTGTGTCT
AAGGGTTTCTTTAAGGAAGGAAGTTCTGTTGAATTAAAACACTTCTTCTTTGCTCAGGATGGTAATGCTGCTATCAG
CGATTATGACTACTATCGTTATAATCTACCAACAATGTGTGATATCAGACAACTACTATTTGTAGTTGAAGTTGTTG
ATAAGTACTTTGATTGTTACGATGGTGGCTGTATTAATGCTAACCAAGTCATCGTCAACAACCTAGACAAATCAGCT
GGTTTTCCATTTAATAAATGGGGTAAGGCTAGACTTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACTTTT
CGCATATACAAAACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGA
GCTCGCACCGTAGCTGGTGTCTCTATCTGTAGTACTATGACCAATAGACAGTTTCATCAAAAATTATTGAAATCAAT
AGCCGCCACTAGAGGAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTTGGCACAACATGTTAAAAAC
TGTTTATAGTGATGTAGAAAACCCTCACCTTATGGGTTGGGATTATCCTAAATGTGATAGAGCCATGCCTAACATG
CTTAGAATTATGGCCTCACTTGTTCTTGCTCGCAAACATACAACGTGTTGTAGCTTGTCACACCGTTTCTATAGATTA
GCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTTCACTATATGTTAAACCAGGTGGAACCT

Fig. 15

CATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTTTTTAACATTTGTCAAGCTGTCACGGCCAATGTTAATGCA
CTTTTATCTACTGATGGTAACAAAATTGCCGATAAGTATGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTA
TAGAAATAGAGATGTTGACACAGACTTTGTGAATGAGTTTTACGCATATTTGCGTAAACATTTCTCAATGATGATAC
TCTCTGACGATGCTGTTGTGTGTTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAG
TCAGTTCTTTATTATCAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACC
TCATGAATTTTGCTCTCAACATACAATGCTAGTTAAACAGGGTGATGATTATGTGTACCTTCCTTACCCAGATCCAT
CAAGAATCCTAGGGGCCGGCTGTTTTGTAGATGATATCGTAAAAACAGATGGTACACTTATGATTGAACGGTTCGT
GTCTTTAGCTATAGATGCTTACCCACTTACTAAACATCCTAATCAGGAGTATGCTGATGTCTTTCATTTGTACTTACA
ATACATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGTTATGCTTACTAATGATAAAC
ACTTCAAGGTATTGGGAACCTGAGTTTTATGAGGCTATGTACACACCGCATACAGTCTTACAGGCTGTTGGGGCTT
GTGTTCTTTGCAATTCACAGACTTCATTAAGATGTGGTGCTTGCATACGTAGACCATTCTTATGTTGTAAATGCTGT
TACGACCATGTCATATCAACATCACATAAATTAGTCTTGTCTGTTAATCCGTATGTTTGCAATGCTCCAGGTTGTGAT
GTCACAGATGTGACTCAACTTTACTTAGGAGGTATGAGCTATTATTGTAAATCACATAAACCACCCATTAGTTTTCC
ATTGTGTGCTAATGGACAAGTTTTTGGTTTATATAAAAATACATGTGTTGGTAGCGATAATGTTACTGACTTTAATG
CAATTGCAACATGTGACTGGACAAATGCTGGTGATTACATTTTAGCTAACACCTGTACTGAAAGACTCAAGCTTTTT
GCAGCAGAAACGCTCAAAGCTACTGAGGAGACATTTAAACTGTCTTATGGTATTGCTACTGTACGTGAAGTGCTGT
CTGACAGAGAATTACATCTTTCATGGGAAGTTGGTAAACCTAGACCACCACTTAACCGAAATTATGTCTTTACTGGT
TATCGTGTAACTAAAAACAGTAAAGTACAAATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGTGATGCTGTT
GTTTACCGAGGTACAACAACTTACAAATTAAATGTTGGTGATTATTTTGTGCTGACATCACATACAGTAATGCCATT
AAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATACCCAACACTCAATATCTCAGATG
AGTTTTCTAGCAATGTTGCAAATTATCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGGACCACCTGGTAC
TGGTAAGAGTCATTTTGCTATTGGCCTAGCTCTCTACTACCCTTCTGCTCGCATAGTGTATACAGCTTGCTCTCATGC

Fig. 15

CGCTGTTGATGCACTATGTGAGAAGGCATTAAAATATTTGCCTATAGATAAATGTAGTAGAATTATACCTGCACGT
GCTCGTGTAGAGTGTTTTGATAAATTCAAAGTGAATTCAACATTAGAACAGTATGTCTTTTGTACTGTAAATGCATT
GCCTGAGACGACAGCAGATATAGTTGTCTTTGATGAAATTTCAATGGCCACAAATTATGATTTGAGTGTTGTCAAT
GCCAGATTACGTGCTAAGCACTATGTGTACATTGGCGACCCTGCTCAATTACCTGCACCACGCACATTGCTAACTAA
GGGCACACTAGAACCAGAATATTTCAATTCAGTGTGTAGACTTATGAAAACTATAGGTCCAGACATGTTCCTCGGA
ACTTGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAGCTTAAAGCACATAA
AGACAAATCAGCTCAATGCTTTAAAATGTTTTATAAGGGTGTTATCACGCATGATGTTTCATCTGCAATTAACAGGC
CACAAATAGGCGTGGTAAGAGAATTCCTTACACGTAACCCTGCTTGGAGAAAAGCTGTCTTTATTTCACCTTATAAT
TCACAGAATGCTGTAGCCTCAAAGATTTTGGGACTACCAACTCAAACTGTTGATTCATCACAGGGCTCAGAATATG
ACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTTGTAATGTAAACAGATTTAATGTTGCTATTACCAGA
GCAAAAGTAGGCATACTTTGCATAATGTCTGATAGAGACCTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCC
ACGTAGGAATGTGGCAACTTTACAAGCTGAAAATGTAACAGGACTCTTTAAAGATTGTAGTAAGGTAATCACTGG
GTTACATCCTACACAGGCACCTACACACCTCAGTGTTGACACTAAATTCAAAACTGAAGGTTTATGTGTTGACATAC
CTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTTAAAATGAATTATCAAGTTAATGG
TTACCCTAACATGTTTATCACCCGCGAAGAAGCTATAAGACATGTACGTGCATGGATTGGCTTCGATGTCGAGGGG
TGTCATGCTACTAGAGAAGCTGTTGGTACCAATTTACCTTTACAGCTAGGTTTTTCTACAGGTGTTAACCTAGTTGC
TGTACCTACAGGTTATGTTGATACACCTAATAATACAGATTTTTCCAGAGTTAGTGCTAAACCACCGCCTGGAGATC
AATTTAAACACCTCATACCACTTATGTACAAAGGACTTCCTTGGAATGTAGTGCGTATAAAGATTGTACAAATGTTA
AGTGACACACTTAAAAATCTCTCTGACAGAGTCGTATTTGTCTTATGGGCACATGGCTTTGAGTTGACATCTATGAA
GTATTTTGTGAAAATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCTTCAG
ACACTTATGCCTGTTGGCATCATTCTATTGGATTTGATTACGTCTATAATCCGTTTATGATTGATGTTCAACAATGGG
GTTTTACAGGTAACCTACAAAGCAACCATGATCTGTATTGTCAAGTCCATGGTAATGCACATGTAGCTAGTTGTGA

Fig. 15

TGCAATCATGACTAGGTGTCTAGCTGTCCACGAGTGCTTTGTTAAGCGTGTTGACTGGACTATTGAATATCCTATAA
TTGGTGATGAACTGAAGATTAATGCGGCTTGTAGAAAGGTTCAACACATGGTTGTTAAAGCTGCATTATTAGCAGA
CAAATTCCCAGTTCTTCACGACATTGGTAACCCTAAAGCTATTAAGTGTGTACCTCAAGCTGATGTAGAATGGAAG
TTCTATGATGCACAGCCTTGTAGTGACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGA
CAAATTCACAGATGGTGTATGCCTATTTTGGAATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTTGTAGAT
TTGACACTAGAGTGCTATCTAACCTTAACTTGCCTGGTTGTGATGGTGGCAGTTTGTATGTAAATAAACATGCATTC
CACACACCAGCTTTTGATAAAAGTGCTTTTGTTAATTTAAAACAATTACCATTTTTCTATTACTCTGACAGTCCATGT
GAGTCTCATGGAAAACAAGTAGTGTCAGATATAGATTATGTACCACTAAAGTCTGCTACGTGTATAACACGTTGCA
ATTTAGGTGGTGCTGTCTGTAGACATCATGCTAATGAGTACAGATTGTATCTCGATGCTTATAACATGATGATCTCA
GCTGGCTTTAGCTTGTGGGTTTACAAACAATTTGATACTTATAACCTCTGGAACACTTTTACAAGACTTCAGAGTTT
AGAAAATGTGGCTTTTAATGTTGTAAATAAGGGACACTTTGATGGACAACAGGGTGAAGTACCAGTTTCTATCATT
AATAACACTGTTTACACAAAAGTTGATGGTGTTGATGTAGAATTGTTTGAAAATAAAACAACATTACCTGTTAATGT
AGCATTTGAGCTTTGGGCTAAGCGCAACATTAAACCAGTACCAGAGGTGAAAATACTCAATAATTTGGGTGTGGA
CATTGCTGCTAATACTGTGATCTGGGACTACAAAAGAGATGCTCCAGCACATATATCTACTATTGGTGTTTGTTCTA
TGACTGACATAGCCAAGAAACCAACTGAAACGATTTGTGCACCACTCACTGTCTTTTTTGATGGTAGAGTTGATGG
TCAAGTAGACTTATTTAGAAATGCCCGTAATGGTGTTCTTATTACAGAAGGTAGTGTTAAAGGTTTACAACCATCTG
TAGGTCCCAAACAAGCTAGTCTTAATGGAGTCACATTAATTGGAGAAGCCGTAAAAACACAGTTCAATTATTATAA
GAAAGTTGATGGTGTTGTCCAACAATTACCTGAAACTTACTTTACTCAGAGTAGAAATTTACAAGAATTTAAACCCA
GGAGTCAAATGGAAATTGATTTCTTAGAATTAGCTATGGATGAATTCATTGAACGGTATAAATTAGAAGGCTATGC
CTTCGAACATATCGTTTATGGAGATTTTAGTCATAGTCAGTTAGGTGGTTTACATCTACTGATTGGACTAGCTAAAC
GTTTTAAGGAATCACCTTTTGAATTAGAAGATTTTATTCCTATGGACAGTACAGTTAAAAACTATTTCATAACAGAT
GCGCAAACAGGTTCATCTAAGTGTGTGTGTTCTGTTATTGATTTATTACTTGATGATTTTGTTGAAATAATAAAATC

Fig. 15

CCAAGATTTATCTGTAGTTTCTAAGGTTGTCAAAGTGACTATTGACTATACAGAAATTTCATTTATGCTTTGGTGTA
AAGATGGCCATGTAGAAACATTTTACCCAAAATTACAATCTAGTCAAGCGTGGCAACCGGGTGTTGCTATGCCTAA
TCTTTACAAAATGCAAAGAATGCTATTAGAAAAGTGTGACCTTCAAAATTATGGTGATAGTGCAACATTACCTAAA
GGCATAATGATGAATGTCGCAAAATATACTCAACTGTGTCAATATTTAAACACATTAACATTAGCTGTACCCTATAA
TATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTACAGCTGTTTTAAGACAGTGGTTG
CCTACGGGTACGCTGCTTGTCGATTCAGATCTTAATGACTTTGTCTCTGATGCAGATTCAACTTTGATTGGTGATTG
TGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGATATGTACGACCCTAAGACTAAAAATGTTACA
AAAGAAAATGACTCTAAAGAGGGTTTTTTCACTTACATTTGTGGGTTTATACAACAAAAGCTAGCTCTTGGAGGTT
CCGTGGCTATAAAGATAACAGAACATTCTTGGAATGCTGATCTTTATAAGCTCATGGGACACTTCGCATGGTGGAC
AGCCTTTGTTACTAATGTGAATGCGTCATCATCTGAAGCATTTTTAATTGGATGTAATTATCTTGGCAAACCACGCG
AACAAATAGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTCCTAT
TCTTTATTTGACATGAGTAAATTTCCCCTTAAATTAAGGGGTACTGCTGTTATGTCTTTAAAAGAAGGTCAAATCAA
TGATATGATTTTATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATG
TTCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTA
CAACCAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTTCAGAT
CCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTGGTTCCATGCTATACATGTCTC
TGGGACCAATGGTACTAAGAGGTTTGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCACTGAGA
AGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGACCCAGTCCCTACTTATTGTTAATAAC
GCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAAC
AACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGC
CTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTTGTGTTTAAGAATATTGATGG
TTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACC

Fig. 15

ATTGGTAGATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCC
TGGTGATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTAT
TAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTAC
GTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGTTA
GATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCATCTGTTTATGCTTGGA
ACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCATTTTCCACTTTTAAGTGTT
ATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGAT
GAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAG
GCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTT
AGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATG
GTGTTGAAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCAT
ACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTG
GTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTT
TCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATTC
TTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAACAAATACTTCTAACCAGGTTGCTGTT
CTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTTTAT
TCTACAGGTTCTAATGTTTTTCAAACACGTGCAGGCTGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGT
GTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGT
AGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGC
CATACCCACAAATTTTACTATTAGTGTTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTA
CAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATCTTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAAC
CGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTACA

Fig. 15

AAACACCACCAATTAAAGATTTTGGTGGTTTTAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGG
TCATTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGCCT
TGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCCTTACTGTTTTGCCACCTTTGCTCACAG
ATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCT
GCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGA
GAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAATTCAAGACTCACTTTCTTCCACAGCAAGTG
CACTTGGAAAACTTCAAGATGTGGTCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAA
TTTTGGTGCAATTTCAAGTGTTTTAAATGATATCCTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATA
GGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAG
CTTCTGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAA
GGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACA
AGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTT
CAAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGT
GTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTCAT
TCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACCAGATGTTGATTTAGGTGACATCTCTGGCATTAA
TGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCG
ATCTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTT
GATTGCCATAGTAATGGTGACAATTATGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTT
GTGGATCCTGCTGCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAAC
GAACTTATGGATTTGTTTATGAGAATCTTCACAATTGGAACTGTAACTTTGAAGCAAGGTGAAATCAAGGATGCTA
CTCCTTCAGATTTTGTTCGCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTGGC
GTTGCACTTCTTGCTGTTTTTCAGAGCGCTTCCAAAATCATAACCCTCAAAAAGAGATGGCAACTAGCACTCTCCAA

Fig. 15

GGGTGTTCACTTTGTTTGCAACTTGCTGTTGTTGTTTGTAACAGTTTACTCACACCTTTTGCTCGTTGCTGCTGGCCT
TGAAGCCCCTTTTCTCTATCTTTATGCTTTAGTCTACTTCTTGCAGAGTATAAACTTTGTAAGAATAATAATGAGGCT
TTGGCTTTGCTGGAAATGCCGTTCCAAAAACCCATTACTTTATGATGCCAACTATTTTCTTTGCTGGCATACTAATTG
TTACGACTATTGTATACCTTACAATAGTGTAACTTCTTCAATTGTCATTACTTCAGGTGATGGCACAACAAGTCCTAT
TTCTGAACATGACTACCAGATTGGTGGTTATACTGAAAAATGGGAATCTGGAGTAAAAGACTGTGTTGTATTACAC
AGTTACTTCACTTCAGACTATTACCAGCTGTACTCAACTCAATTGAGTACAGACACTGGTGTTGAACATGTTACCTT
CTTCATCTACAATAAAATTGTTGATGAGCCTGAAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTT
GTTAATCCAGTAATGGAACCAATTTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAGCTGATG
AGTACGAACTTGTGACAACAGATGTTTCATCTCGTTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTAT
GAGGACTTTTAAAGTTTCCATTTGGAATCTTGATTACATCATAAACCTCATAATTAAAAATTTATCTAAGTCACTAAC
TGAGAATAAATATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGATTAAACGAACATGAAAATTATTCTTTTC
TTGGCACTGATAACACTCGCTACTTGTGAGCTTTATCACTACCAAGAGTGTGTTAGAGGTACAACAGTACTTTTAAA
AGAACCTTGCTCTTCTGGAACATACGAGGGCAATTCACCATTTCATCCTCTAGCTGATAACAAATTTGCACTGACTT
GCTTTAGCACTCAATTTGCTTTTGCTTGTCCTGACGGCGTAAAACACGTCTATCAGTTACGTGCCAGATCAGTTTCA
CCTAAACTGTTCATCAGACAAGAGGAAGTTCAAGAACTTTACTCTCCAATTTTTCTTATTGTTGCGGCAATAGTGTT
TATAACACTTTGCTTCACACTCAAAAGAAAGACAGAATGATTGAACTTTCATTAATTGACTTCTATTTGTGCTTTTTA
GCCTTTCTGCTATTCCTTGTTTTAATTATGCTTATTATCTTTTGGTTCTCACTTGAACTGCAAGATCATAATGAAACTT
GTCACGCCTAAACGAACATGAAATTTCTTGTTTTCTTAGGAATCATCACAACTGTAGCTGCATTTCACCAAGAATGT
AGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTCTATTCTAAATGGTAT
ATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGGATGAGGCTGGTTCTAAATCACCCATTC
AGTACATCGATATCGGTAATTATACAGTTTCCTGTTTACCTTTTACAATTAATTGCCAGGAACCTAAATTGGGTAGT
CTTGTAGTGCGTTGTTCGTTCTATGAAGACTTTTTAGAGTATCATGACGTTCGTGTTGTTTTAGATTTCATCTAAACG

Fig. 15

AACAAACTAAAATGTCTGATAATGGACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGA
TTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCCAAGGTTTACCCAA
TAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCCTCGAGGACAAGGCGTT
CCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGT
GACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCT
ATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCAAAAGATCACATTGGCA
CCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGC
AGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAAC
TCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTCTTGCTTTGCTGCT
GCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAA
GAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCT
TTCGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTAC
AAACATTGGCCGCAAATTGCACAATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAG
TCACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCA
AGTCATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAGAA
GAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCTTCTTCCTGCTGCAGA
TTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACTCAGGCCTAAACTCATGCAGACC
ACACAAGGCAGATGGGCTATATAAACGTTTTCGCTTTTCCGTTTACGATATATAGTCTACTCTTGTGCAGAATGAAT
TCTCGTAACTACATAGCACAAGTAGATGTAGTTAACTTTAATCTCACATAGCAATCTTTAATCAGTGTGTAACATTA
GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGTACAGTGAACAATGCT
AGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAATTTTAGTAGTGCTATCCCCATGTGATTTTAAT
AGCTTCTTAGGAGAATGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO:16)

Fig. 15

hACE2

ATGTCAAGCTCTTCCTGGCTCCTTCTCAGCCTTGTTGCTGTAACTGCTGCTCAGTCCACC

ATTGAGGAACAGGCCAAGACATTTTTGGACAAGTTTAACCACGAAGCCGAAGACCTGTTC

TATCAAAGTTCACTTGCTTCTTGGAATTATAACACCAATATTACTGAAGAGAATGTCCAA

AACATGAATAATGCTGGGGACAAATGGTCTGCCTTTTTAAAGGAACAGTCCACACTTGCC

CAAATGTATCCACTACAAGAAATTCAGAATCTCACAGTCAAGCTTCAGCTGCAGGCTCTT

CAGCAAAATGGGTCTTCAGTGCTCTCAGAAGACAAGAGCAAACGGTTGAACACAATTCTA

AATACAATGAGCACCATCTACAGTACTGGAAAAGTTTGTAACCCAGATAATCCACAAGAA

TGCTTATTACTTGAACCAGGTTTGAATGAAATAATGGCAAACAGTTTAGACTACAATGAG

AGGCTCTGGGCTTGGGAAAGCTGGAGATCTGAGGTCGGCAAGCAGCTGAGGCCATTATAT

GAAGAGTATGTGGTCTTGAAAAATGAGATGGCAAGAGCAAATCATTATGAGGACTATGGG

GATTATTGGAGAGGAGACTATGAAGTAAATGGGGTAGATGGCTATGACTACAGCCGCGGC

CAGTTGATTGAAGATGTGGAACATACCTTTGAAGAGATTAAACCATTATATGAACATCTT

Fig. 15

CATGCCTATGTGAGGGCAAAGTTGATGAATGCCTATCCTTCCTATATCAGTCCAATTGGA

TGCCTCCCTGCTCATTTGCTTGGTGATATGTGGGGTAGATTTTGGACAAATCTGTACTCT

TTGACAGTTCCCTTTGGACAGAAACCAAACATAGATGTTACTGATGCAATGGTGGACCAG

GCCTGGGATGCACAGAGAATATTCAAGGAGGCCGAGAAGTTCTTTGTATCTGTTGGTCTT

CCTAATATGACTCAAGGATTCTGGGAAAATTCCATGCTAACGGACCCAGGAAATGTTCAG

AAAGCAGTCTGCCATCCCACAGCTTGGGACCTGGGGAAGGGCGACTTCAGGATCCTTATG

TGCACAAAGGTGACAATGGACGACTTCCTGACAGCTCATCATGAGATGGGGCATATCCAG

TATGATATGGCATATGCTGCACAACCTTTTCTGCTAAGAAATGGAGCTAATGAAGGATTC

CATGAAGCTGTTGGGGAAATCATGTCACTTTCTGCAGCCACACCTAAGCATTTAAAATCC

ATTGGTCTTCTGTCACCCGATTTTCAAGAAGACAATGAAACAGAAATAAACTTCCTGCTC

AAACAAGCACTCACGATTGTTGGGACTCTGCCATTTACTTACATGTTAGAGAAGTGGAGG

TGGATGGTCTTTAAAGGGGAAATTCCCAAAGACCAGTGGATGAAAAAGTGGTGGGAGATG

AAGCGAGAGATAGTTGGGGTGGTGGAACCTGTGCCCCATGATGAAACATACTGTGACCCC

GCATCTCTGTTCCATGTTTCTAATGATTACTCATTCATTCGATATTACACAAGGACCCTT

Fig. 15

TACCAATTCCAGTTTCAAGAAGCACTTTGTCAAGCAGCTAAACATGAAGGCCCTCTGCAC

AAATGTGACATCTCAAACTCTACAGAAGCTGGACAGAAACTGTTCAATATGCTGAGGCTT

GGAAAATCAGAACCCTGGACCCTAGCATTGGAAAATGTTGTAGGAGCAAAGAACATGAAT

GTAAGGCCACTGCTCAACTACTTTGAGCCCTTATTTACCTGGCTGAAAGACCAGAACAAG

AATTCTTTTGTGGGATGGAGTACCGACTGGAGTCCATATGCAGACCAAAGCATCAAAGTG

AGGATAAGCCTAAAATCAGCTCTTGGAGATAAAGCATATGAATGGAACGACAATGAAATG

TACCTGTTCCGATCATCTGTTGCATATGCTATGAGGCAGTACTTTTTAAAAGTAAAAAAT

CAGATGATTCTTTTTGGGGAGGAGGATGTGCGAGTGGCTAATTTGAAACCAAGAATCTCC

TTTAATTTCTTTGTCACTGCACCTAAAAATGTGTCTGATATCATTCCTAGAACTGAAGTT

GAAAAGGCCATCAGGATGTCCCGGAGCCGTATCAATGATGCTTTCCGTCTGAATGACAAC

AGCCTAGAGTTTCTGGGGATACAGCCAACACTTGGACCTCCTAACCAGCCCCCTGTTTCC

ATATGGCTGATTGTTTTTGGAGTTGTGATGGGAGTGATAGTGGTTGGCATTGTCATCCTG

ATCTTCACTGGGATCAGAGATCGGAAGAAGAAAAATAAAGCAAGAAGTGGAGAAAATCCT

TATGCCTCCATCGATATTAGCAAAGGAGAAAATAATCCAGGATTCCAAAACACTGATGAT

GTTCAGACCTCCTTTGGTACCGAGACCTCCCAGGTGGCGCCCGCTTAA

(SEQ ID NO:17)

Fig. 15

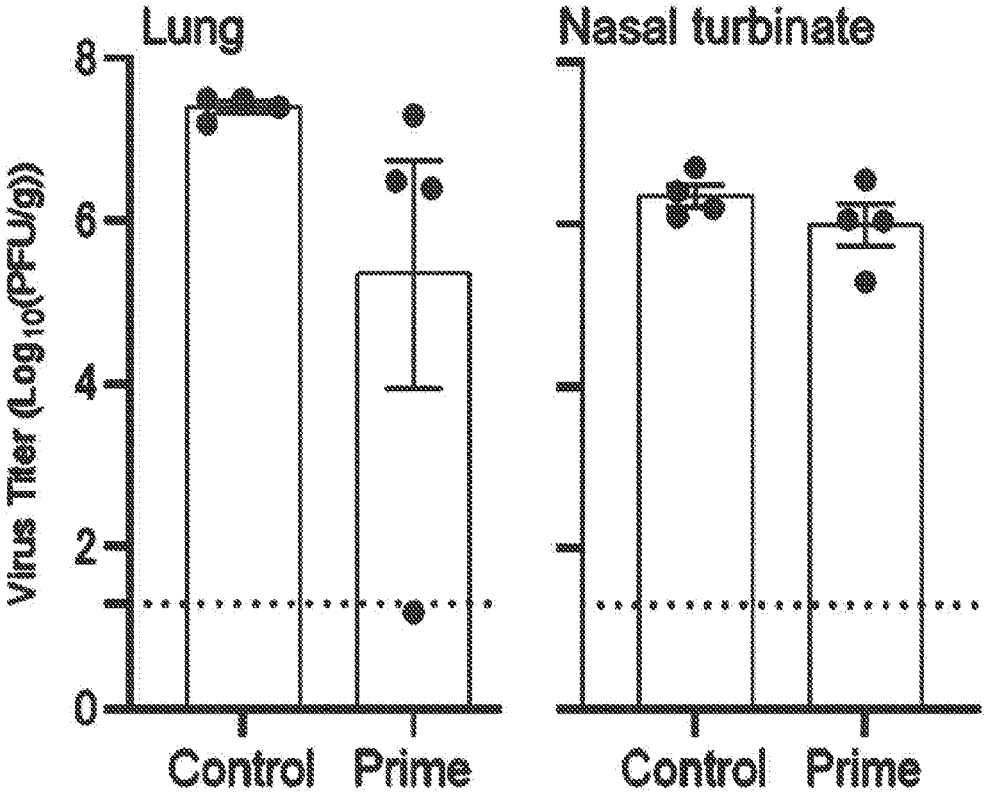


Fig. 16A Delta Challenge *Fig. 16B*

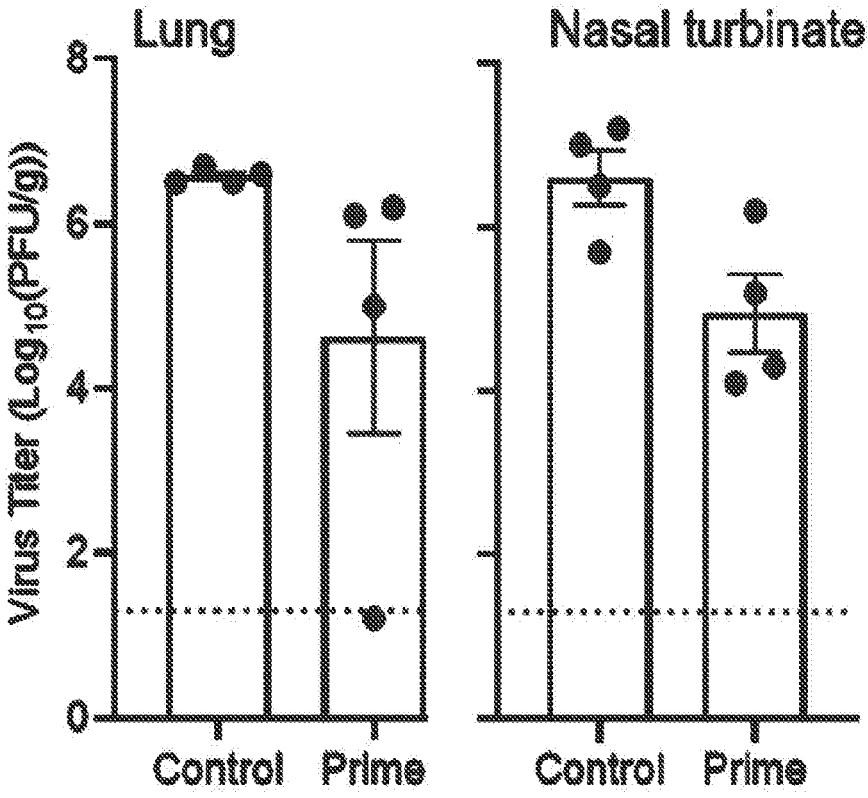


Fig. 16C XBB Challenge Fig. 16D

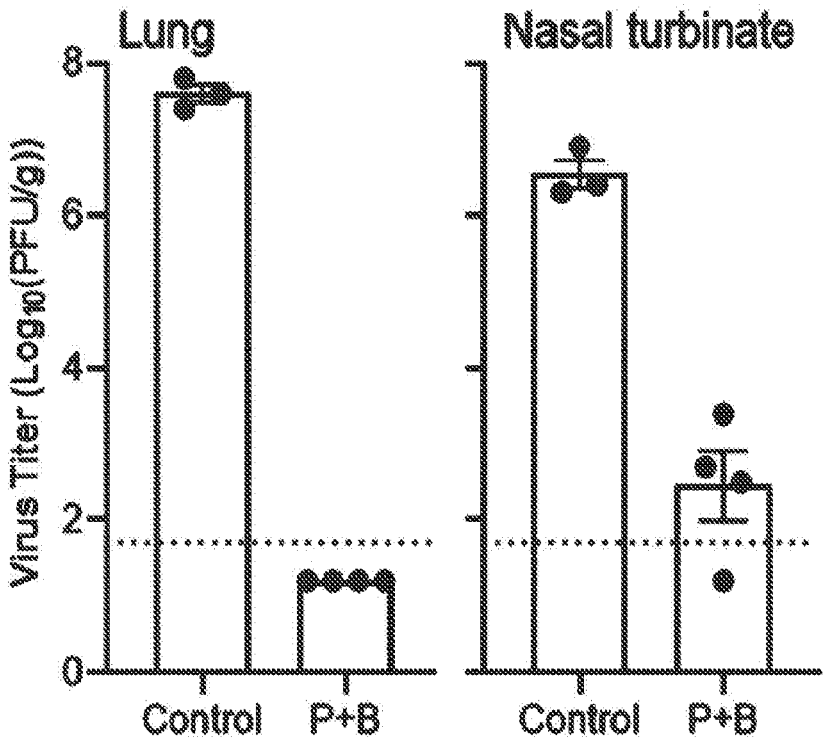


Fig. 17A **Delta Challenge** Fig. 17B

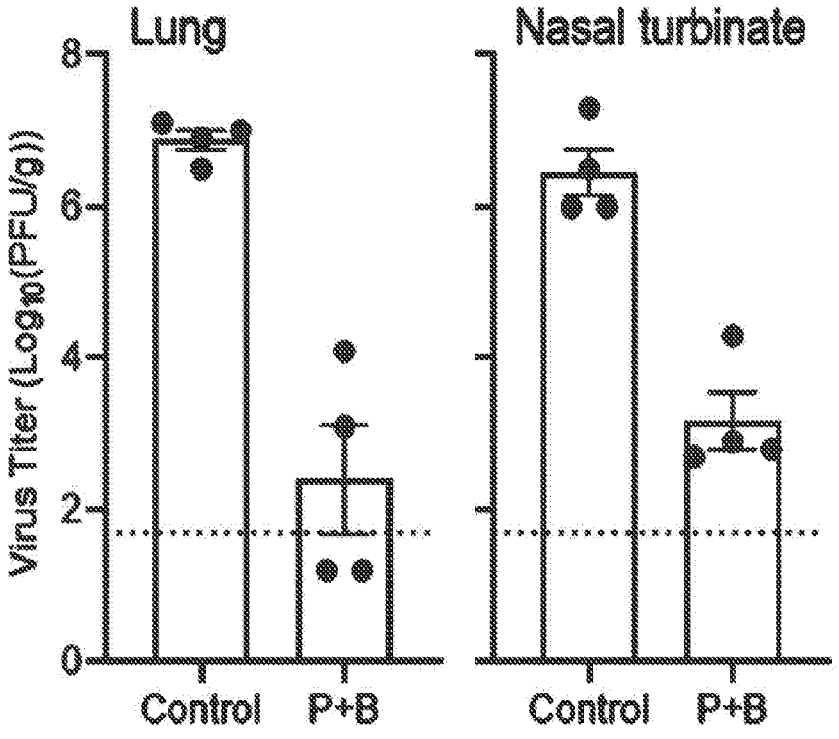


Fig. 17C **XBB Challenge** Fig. 17D

```
   1 attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct
  61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
 121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc
 181 ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt
 241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acgagaaaac
 301 acacgtccaa ctcagtttgc ctgttttaca ggttcgcgac gtgctcgtac gtggctttgg
 361 agactccgtg gaggaggtct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg
 421 cttagtagaa gttgaaaaag gcgttttgcc tcaacttgaa cagccctatg tgttcatcaa
 481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact
 541 cgaaggcatt cagtacggtc gtagtggtga gacacttggt gtccttgtcc ctcatgtggg
 601 cgaaatacca gtggcttacc gcaaggttct tcttcgtaag aacggtaata aaggagctgg
 661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga
 721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttacccgtga
 781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg
 841 ccctgatggc taccctcttg agtgcattaa agaccttcta gcacgtgctg gtaaagcttc
 901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg
 961 tgaacatgag catgaaattg cttggtacac ggaacgttct gaaaagagct atgaattgca
1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa
1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aaaagaaaaa
1141 gcttgatggc tttatgggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg
```

Fig. 18

```
1201 caaccaaatg tgcctttcaa ctctcatgaa gtgtgatcat tgtggtgaaa cttcatggca
1261 gacgggcgat tttgttaaag ccacttgcga attttgtggc actgagaatt tgactaaaga
1321 aggtgccact acttgtggtt acttacccca aaatgctgtt gttaaaattt attgtccagc
1381 atgtcacaat tcagaagtag gacctgagca tagtcttgcc gaataccata atgaatctgg
1441 cttgaaaacc attcttcgta agggtggtcg cactattgcc tttggaggct gtgtgttctc
1501 ttatgttggt tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg
1561 ttgtaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga
1621 aatactccaa aaagagaaag tcaacatcaa tattgttggt gactttaaac ttaatgaaga
1681 gatcgccatt attttggcat ctttttctgc ttccacaagt gcttttgtgg aaactgtgaa
1741 aggtttggat tataaagcat tcaaacaaat tgttgaatcc tgtggtaatt ttaaagttac
1801 aaaaggaaaa gctaaaaaag gtgcctggaa tattggtgaa cagaaatcaa tactgagtcc
1861 tctttatgca tttgcatcag aggctgctcg tgttgtacga tcaatttttct cccgcactct
1921 tgaaactgct caaaattctg tgcgtgtttt acagaaggcc gctataacaa tactagatgg
1981 aatttcacag tattcactga gactcattga tgctatgatg ttcacatctg atttggctac
2041 taacaatcta gttgtaatgg cctacattac aggtggtgtt gttcagttga cttcgcagtg
2101 gctaactaac atctttggca ctgtttatga aaaactcaaa cccgtccttg attggcttga
2161 agagaagttt aaggaaggtg tagagtttct tagagacggt tgggaaattg ttaaatttat
2221 ctcaacctgt gcttgtgaaa ttgtcggtgg acaaattgtc acctgtgcaa aggaaattaa
2281 ggagagtgtt cagacattct ttaagcttgt aaataaattt ttggctttgt gtgctgactc
2341 tatcattatt ggtggagcta aacttaaagc cttgaattta ggtgaaacat ttgtcacgca
```

Fig. 18

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/027622

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.
X	NETLAND JASON ET AL: "Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease", VIROLOGY, vol. 399, no. 1, 27 January 2010 (2010-01-27), pages 120-128, XP085464129, ISSN: 0042-6822, DOI: 10.1016/J.VIROL.2010.01.004	1,2,4-8, 17,22, 23,26, 30,31, 33,34,37
Y	abstract page 121 - page 122; figures 2,3,5; table 2	10, 14-16, 21, 27-29, 32,35, 36,38-41

-/--

[X] 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/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Renggli-Zulliger, N

3

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/027622

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHANG XIANWEN ET AL: "A trans-complementation system for SARS-CoV-2 recapitulates authentic viral replication without virulence", CELL, 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]	1,2,4-9, 11-13, 17,19, 20,22, 23,26
Y	abstract; figures 1,3,4 page 2236, last paragraph	10, 14-16, 21, 27-29, 32,35, 36,38-41

X	LIU SHUFENG ET AL: "Stable Cell Clones Harboring Self-Replicating SARS-CoV-2 RNAs for Drug Screen", JOURNAL OF VIROLOGY, 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:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8941906/pdf/jvi.02216-21.pdf>	1-3,5-8, 17,18, 22-26
Y	abstract; figure 1	21,27-29

3