



(51) International Patent Classification:

C12N 15/10 (2006.01) C12N 15/86 (2006.01)
C12N 7/00 (2006.01) C40B 40/08 (2006.01)

(21) International Application Number:

PCT/AU2024/050763

(22) International Filing Date:

17 July 2024 (17.07.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2023904160 20 December 2023 (20.12.2023) AU

(71) Applicants: **THE UNIVERSITY OF SYDNEY** [AU/AU]; Parramatta Road, The University of Sydney, New South Wales 2006 (AU). **CENTENARY INSTITUTE OF CANCER MEDICINE AND CELL BIOLOGY** [AU/AU]; Building 93, Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050 (AU).

(72) Inventors: **HESSELSOON, Daniel**; c/o Centenary Institute of Cancer Medicine and Cell Biology, Building 93, Royal Prince Alfred Hospital, Missenden Road, Camperdown,

New South Wales 2050 (AU). **NEELY, Graham Gregory**; c/o The University of Sydney, Parramatta Road, Sydney, New South Wales 2006 (AU). **COLE, Alexander**; c/o Centenary Institute of Cancer Medicine and Cell Biology, Building 93, Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050 (AU). **DENES, Christopher**; c/o The University of Sydney, Parramatta Road, Sydney, New South Wales 2006 (AU).

(74) Agent: **SPRUSON & FERGUSON**; GPO BOX 3898, Sydney, New South Wales 2001 (AU).

(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.

(54) Title: A DIRECTED EVOLUTION SYSTEM

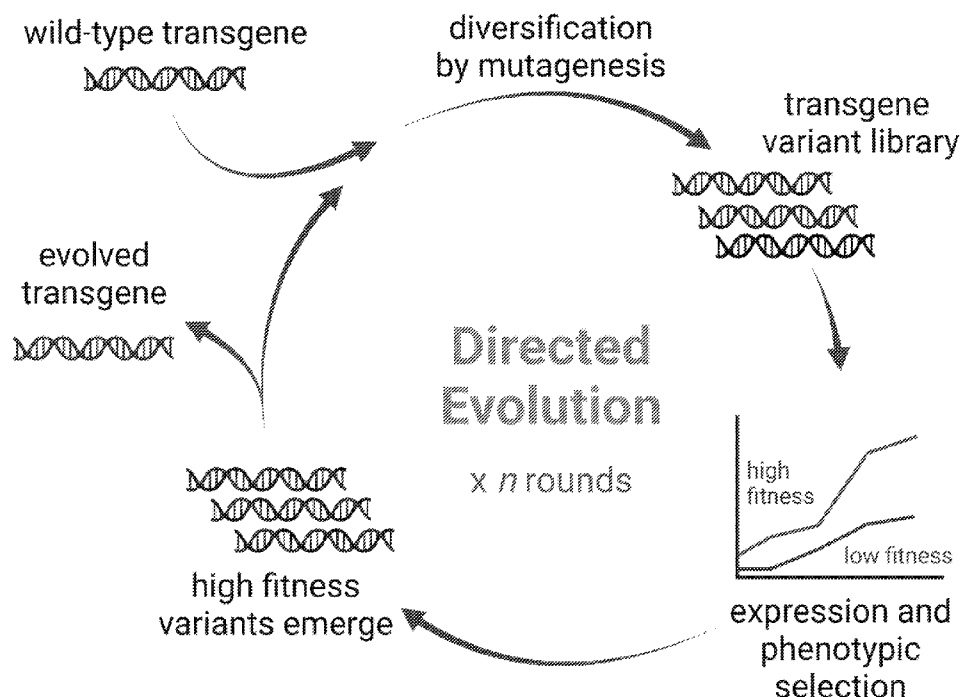


FIG. 1

(57) Abstract: The present invention relates to a system for evolving a target biomolecule, methods of using such system and composition comprising said system.

(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))*
- *with sequence listing part of description (Rule 5.2(a))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

A DIRECTED EVOLUTION SYSTEM

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from Australian Provisional Application No. 2023904160,
5 filed on 20 December 2023, which is hereby incorporated by reference in its entirety herein.

SEQUENCE LISTING

The instant application contains a sequence listing which has been submitted
electronically as an XML document in the ST26 format and is hereby incorporated by reference
10 in its entirety. Said XML copy, created on 15 July 2024, is named "P0063048PCT" and is 66,946
bytes in size.

FIELD

The present invention relates to a system for evolving a target biomolecule, methods of
15 using such system and composition comprising said system. The invention has been developed
primarily as a directed evolution system and will be described hereinafter with reference to this
application. However, it will be appreciated that the invention is not limited to this particular field
of use.

20 BACKGROUND

Using iterative rounds of diversification, selection, and amplification, directed evolution
(DE) can produce biomolecules with new or improved functions (FIG. 1, (1, 2, 4-7)). While this
approach has been widely used to evolve molecules in simple prokaryotic and eukaryotic systems
(8, 9), these environments lack the full complement of post-translational modifications, protein-
25 protein interactions, and signaling networks found in mammalian cells (3).

Ideally, proteins destined for mammalian applications would be evolved directly in
mammalian cells. Historically, this has been achieved using ex mammalia mutagenesis
techniques combined with phenotypic screening in mammalian cells (10). More recently, targeted
mutagenesis has linked protein function to selectable or screenable markers allowing target
30 diversification and variant selection in the same mammalian cell (11, 12). However, cell-based
approaches that link an integrated target molecule to cellular fitness can be derailed by mutations

in the host genome (13). Placing the target in a viral genome can mitigate this issue since naive host cells can be provided for each round of DE. However, existing virus-based mammalian DE systems are limited by safety concerns (14), low mutational rates (15), are target-specific (16, 17), or lack functionality (18, 19).

5

SUMMARY OF THE INVENTION

The present invention broadly relates to a novel DE system which overcomes or ameliorates at least one of the disadvantages of the prior art, or to provide a useful alternative. In one or more aspects, the present invention drives evolution using a virus-like particle (VLP) based on an RNA virus.

10

In a first aspect, the present invention provides a directed evolution system, comprising:

- (i) a first nucleic acid sequence encoding one or more non-structural genes of a virus and at least one evolution candidate,
- (ii) a second nucleic acid sequence encoding one or more packaging genes, and
- 15 (iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

wherein the first nucleic acid sequence and the second nucleic acid sequence generates a first virus-like particle (VLP) after expression in a host cell, and wherein said first VLP and the third nucleic acid sequence generates a second VLP after transduction of the first VLP and expression of the third nucleic acid sequence in a host cell,

20

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and

wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

25 Preferably, the one or more packaging gene is different from the virus. In one embodiment, the one or more packaging genes is from a different virus. In a particular embodiment, the one or more packaging genes is a vesicular stomatitis virus (VSV) G DNA encoding a VSV G protein.

The virus preferably is an RNA virus, more preferably an alphavirus. In a particular embodiment, the RNA virus is Semliki Forest Virus (SFV).

30 Preferably, the one or more non-structural genes comprises non-structural protein (NSP) 1, NSP2, NSP3 and NSP4 of SFV non-structural genes.

Preferably, the one or more non-structural genes further comprises a 5' untranslated region and a 3' untranslated region of SFV non-structural genes operably linked to NSP1, NSP2, NSP3 and NSP4.

In one embodiment, the VLP is a virus-like vesicle (VLV).

5 Preferably, the one or more non-structural genes comprises at least one mutation that increases virus titer production.

10 In one embodiment, the at least one mutation is selected from G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834- A, T-8859-A, T-8864-C, G-9211-A , A-10427-G, G-11560-A, or any other modifications that improve packaging or functionality of the system. In one particular embodiment, the at least one mutation comprises G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834- A, T-8859-A, T-8864-C, G-9211-A , A-10427-G, and G-11560-A.

15 In another embodiment, the at least one mutation is selected from G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, E2393K or any other modifications that improve packaging or functionality of the system. In one particular embodiment, the at least one mutation comprises G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, and E2393K.

20 Preferably, the one or more packaging genes comprises at least one mutation that increases virus titer production. Preferably, the at least one mutation is selected from A-11871-G, T-11978-C and other modifications that improve packaging or functionality of the system.

Preferably, the first nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to a DNA sequence encoding the one or more non-structural genes of the RNA virus, operably linked to a DNA sequence encoding the at least one evolution candidate.

25 Preferably, the first nucleic acid sequence further comprises an alphavirus subgenomic RNA promoter (SGP) operably linked between the one or more non-structural genes of the RNA virus and the at least one evolution candidate. In one embodiment, the alphavirus SGP is an SFV promoter.

30 Preferably, the second nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to the one or more packaging genes.

In one embodiment, the promoter sequence is a cytomegalovirus immediate early promoter (CMV promoter) sequence.

Preferably, the host cell is a mammalian cell, such as BHK-21 cell or HEK293T cell.

Preferably, the evolution candidate comprises a sequence of at least one transgene of interest. The at least one transgene of interest preferably encodes a protein, a protein domain, a protein-coding region, a non-protein-coding region, a truncated protein, a truncated protein domain, a truncated protein coding region or a truncated non-protein-coding region. In one embodiment, the evolution candidate is selected from the group consisting of Tet transactivator (tTa), reverse Tet transactivator (rtTa), and anti-p53 nanobody.

Preferably, the second VLP and the third nucleic acid sequence generates further VLPs after transduction of the second VLP and expression of the third nucleic acid sequence in a host cell, wherein the generation of the further VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. Preferably, the system further generates one or more additional VLPs through expression of the third nucleic acid and transduction of the further VLPs in a host cell, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. In one embodiment, the system is a continuous system.

In a second aspect, the present invention provides a method of performing directed evolution of at least one evolution candidate, comprising the directed evolution system according to the first aspect.

In a third aspect, the present invention provides a method of performing directed evolution of at least one evolution candidate, comprising:

(i) expressing in a host cell

(a) a first nucleic acid sequence encoding one or more non-structural genes of a virus and the at least one evolution candidate, , and

(b) a second nucleic acid sequence encoding one or more packaging genes,

(ii) generating a first virus-like particle (VLP) by the host cell, and

(iii) transducing a host cell with the first VLP and expressing a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes to generate a second VLP,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate in step (iii), and

wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

Preferably, the method further comprises screening the host cells for VLPs comprising a desirable evolved candidate.

In one embodiment, the method further comprises transducing a host cell with the second VLP and expressing the third nucleic acid sequence to generate further VLPs, wherein the generation of the further VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. Preferably, the method further comprises transducing a host cell with the further VLPs and expressing the third nucleic acid sequence to generate one or more additional VLPs, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

In a fourth aspect, the present invention provides a virus-like particle produced by the method according to the second or the third aspect.

In a fifth aspect, the present invention provides a host cell comprising the virus-like particle according to the fourth aspect.

In a sixth aspect, the present invention provides a composition comprising the virus-like particle according to the fourth aspect.

In a seventh aspect, the present invention provides a kit for performing directed evolution, comprising.

(i) a first nucleic acid sequence encoding one or more non-structural genes of an virus and at least one evolution candidate,

(ii) a second nucleic acid sequence encoding one or more packaging genes, and

(iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

wherein the first nucleic acid sequence and the second nucleic acid sequence generates a first virus-like particle (VLP) after expression in a host cell, and wherein said first VLP and the third nucleic acid sequence generates a second VLP after transduction of the first VLP and expression of the third nucleic acid sequence in a host cell,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and

wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

In one embodiment, the kit comprises the directed evolution system according to the first aspect. In another embodiment, the kit is for use in the method according to the second or the third aspect.

5

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1. Schematic diagram of Directed Evolution

FIG. 2. Host-dependent propagation of VLVs. (a) SFV VLVs are initially packaged using SFV-DE DNA replicons that encode the DE target transgene in cells that constitutively express VSVG. Infectious VLVs are propagated for evolution in host cells that express VSVG under the control of a circuit that is directly or indirectly regulated by the DE target. Titers (b) and amplification factors (c) of eGFP-LUC VLVs propagated in cells constitutively expressing CMV_VSVG (+VSVG) for all rounds (R) RP-R5 or control DNA (-VSVG; for R4 and R5 presented within the gray box) ($N = 6$). (d) Allele frequency of mutations in a neutral eGFP transgene. Dotted line represents detection limit for viral variants (0.3%). (e) Mutational spectrum of viral variants from R2 ($N = 95$). (f) Circuit design for tTA mediated activation of TRE3G. (g) Dox-dependent repression of a TRE3G-regulated LUC reporter ($N = 4$). (h) Titers of serially diluted tTA VLVs propagated on cells expressing VSVG under the control of TRE3G in the absence of doxycycline ($N = 3$). Colors are for illustrative purposes and do not reflect different membrane compositions.

FIG. 3. Benchmarking the SFV-VSVG VLV DE system platform. (a) Schematic representation of the SFV DNA replicon genome. FP/RP indicate the positions of the primers used for transgene isolation/sequencing. Propagation (b) and amplification (c) of WT and ADA>RLE attenuated eGFP-LUC VLVs in CMV_VSVG-expressing BHK-21 cells ($N = 4$). (d) Brightfield and epifluorescence microscopy of +VSVG BHK-21 cells at RP or following transduction with eGFP-LUC VLVs at RA-R4 (representative of $N = 4$). Scale bars, 200 μm . (e) Microscopy of -VSVG BHK-21 cells transduced at R3 and R4 (representative of $N = 4$). Scale bars, 200 μm . (f) RT-PCR and DNA gel electrophoresis of SFV-DE eGFP-LUC VLV transgenes ($N = 4$ pooled).

FIG. 4. RT-PCR and DNA gel electrophoresis of isolated transgenes from serially diluted SFV-DE tTA : eGFP-LUC VLVs (R1-R3; $N = 3$ pooled).

FIG. 5. VLV propagation is dependent on host expression of VSVG. (a) Circuit design for SRF-VP64 mediated activation of SRE. (b) FBS dose-dependent SRE_LUC reporter activation by endogenous factors ($N = 3$). (c) SRF-VP64-mediated induction of an SRE-regulated LUC reporter ($N = 3$). Titers (d) and amplification factors (e) of SRF-VP64 and neutral eGFP-LUC VLVs propagated on cells expressing VSVG under the control of an SRE promoter in 1% FBS-supplemented growth medium (gray box) ($N = 3$). (f) Nanopore sequencing of transgene RNA isolated from pooled co-packaged VLV samples ($N = 6$) aligned to reference sequences (>10,000

reads per sample). For E1-E4, VLVs were propagated in 1% FBS-supplemented growth medium. (g) Brightfield and epifluorescence microscopy of BHK-21 cells at RP or following transduction with a 1:1 packaged cohort of SRF-VP64:eGFP-LUC VLVs at RA-E4 (representative of $N = 6$). Scale bars, 200 μm . (h) RT-PCR and DNA gel electrophoresis of isolated transgenes from a 1:1 packaged cohort of SRF-VP64:eGFP-LUC VLVs ($N = 6$ pooled). (i) The DNA bands from E3 and E4 marked as 'Truncated' in (h) were cloned and sequenced ($n=11$ clones per sample). (j) Reads from (f) ($>10,000$ per sample) were aligned to a VSVG reference sequence. For E1-E4, VLVs were propagated in 1% FBS-supplemented growth medium.

FIG. 6. The SFV-VSVG VLV DE system generates authentic evolution products. (a) Titers and amplification factors of VLVs propagated on cells expressing VSVG under the control of TRE3G (gray box) ($N = 4$). RP (packaging) and RA (amplification) indicate rounds of VLV propagation under constitutive VSVG expression, while E1-X labelling indicates rounds of evolution under transgene-regulated VSVG expression. (b) Allele frequency of the major variants identified in Campaign 1. (c) Dox-resistance of evolved tTA variants ($N = 5$). (d) Variant-induced structural changes in tTA modelled with AlphaFold2 (red, mutated residues; blue, displaced functional groups; dashed green circle, drug binding pocket).

FIG. 7. Directed evolution of tTA (Campaign 2). (a) DNA gel electrophoresis of isolated tTA transgenes following RT-PCR (from independent Campaigns 1 and 2). Allele frequency of the major variants individually (b) and combined (c) identified in Campaign 2. A similar analysis of mutation combinations for Campaign 1 cannot be extracted from short-read Illumina sequencing because of the distance between the identified Q32R and R158G residues. Note, a synonymous passenger mutation (E159E) rose to similar levels as each of the R158G, D178G, H179R, Q180R mutations during this campaign. Analysis in (c) permitted synonymous substitutions at E159 to capture all reads analyzed in (b). (d) Dox-resistance of evolved tTA variants ($N = 5$). (e) Isolated effects of aa178-180 single mutants ($N = 5$). (f) Alignment of crystal structure 4AC0 (blue) with AlphaFold2-modeled tTA (gray). (g) Variant-induced structural changes in tTA modelled with AlphaFold2 (red, mutated residues; blue, displaced functional groups; dashed green circle, drug binding pocket).

FIG. 8. Enhancing drug-inducible transcriptional control. (a) Circuit design for rtTA-mediated activation of TRE3G. (b) Dox-dependent activation of a TRE3G-regulated LUC reporter ($N = 3$). (c) Titers of VLVs propagated on cells expressing VSVG under the control of TRE3G at 100 ng/mL dox (gray box) ($N = 4$). (d) Allele frequency of the major variants identified during long-term propagation on minimal concentrations of dox. (e) Dox-sensitivity of evolved tTA variants ($N = 3$).

FIG. 9. Directed evolution of rtTA-3G. (a) Amplification factors of VLVs propagated on cells expressing VSVG under the control of TRE3G at 100 ng/mL dox (gray box) ($N = 4$). (b) Dox concentration and AF for each round of rtTA-3G evolution. (c) DNA gel electrophoresis of isolated

rtTA transgenes following RT-PCR. (d) Allele frequencies of single and double mutant variants during long-term propagation on minimal concentrations of dox. (e) Basal activity of the evolved variants in the absence of dox ($N = 3$). (f) Alignment of top 5 ranked AlphaFold2 predictions for rtTA-3G (tan) and the D5N/M59I variant (blue; red, mutated residues).

- 5 **FIG. 10.** Directed evolution of a nanobody with the SFV-VSVG VLV DE system. (a) Schematic of Nb139-VP64 and Nb139-eGFP fusions. (b) Effect of 16 μ M cisplatin on parental Nb139-eGFP nuclear localization ($N = 18$). (c) p53 immunofluorescence (red) and Hoechst nuclear staining (gray); Quantification of nuclear p53 accumulation. Scale bars, 50 μ m. ($N \geq 272$ cells analyzed per replicate). (d) Amplification factors of VLVs propagated on cells expressing VSVG under the control of a p53 2-hybrid circuit (gray box) ($N = 4$). (e) DNA gel electrophoresis of isolated Nb139-VP64 transgenes following RT-PCR. (f) Allele frequencies of single and double mutant variants during long-term propagation on the 2-hybrid circuit.

- FIG. 11.** Directed evolution of an intracellular nanobody. (a) 2-hybrid circuit design for nanobody-p53 interactions. (b) Recruitment of Nb139-VP64 to a p53 bait activates the 2-hybrid circuit (SV40, positive interaction control (PMID: 9043710, $N = 4$)). Normalized to the ZF43RE_LUC + p53-ZF43 bait incomplete circuit. (c) Titers of VLVs propagated on cells expressing VSVG under the control of a p53 2-hybrid circuit (gray box) ($N = 4$). (d) Allele frequency of the major variants identified during long-term propagation on the 2-hybrid circuit. (e) Crystal structure 4QO1 showing Nb139 (gray; green, Nb139 complementarity-determining regions; red, evolved variant positions) in a complex with p53 (blue). (f) Circuit activation by Nb139-VP64 variants in cells expressing a p53 bait and LUC reporter ($N = 4$). (g) Timelapse of cisplatin-treated cells that express Nb139[S26P]-eGFP fusion or GFP alone (red, nuclei labelled with mCherry; scale bar, 25 μ m). White arrows indicate cells of interest; asterisks indicate foci formation. (h) Quantification of G. ($N > 1000$ tracked cells).

- 25 **FIG. 12.** Individual channels from timelapse in Fig. 10G. Nuclei labeled with mCherry (red). Nb139[S26P]-eGFP biosensor or turboGFP control (green). White arrows indicate cells of interest; asterisks indicate foci formation. Scale bars, 25 μ m.

DESCRIPTION OF EMBODIMENTS

- 30 In one or more aspects, the present invention drives evolution using a VLP based on the alphavirus Semliki Forest Virus (SFV). Chimeric in design, the capsid-free VLPs according to one or more aspect of the present invention are packaged and passaged with nucleic acids that possess sequences with reduced homology or no homologous sequence, and thus cannot recombine to produce pathogenic virus. The use of an RNA viral backbone which lacks proof-

reading capacity according to one or more aspects of the present invention allows for a naturally high mutation rate that can be exploited for transgene diversification.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein may be used in the practice for testing of the present invention, the preferred materials and methods are described herein. In describing and claiming the present invention, the following terminology will be used.

It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

The indefinite articles 'a' and 'an' are used here to refer to or encompass singular or plural elements or features and should not be taken as meaning or defining "one" or a "single" element or feature. For example, "a" cell includes one cell, one or more cells and a plurality of cells.

Unless the context requires otherwise, the terms "comprise", "comprises" and "comprising", or similar terms are intended to mean a non-exclusive inclusion, such that a recited list of elements or features does not include those stated or listed elements solely, but may include other elements or features that are not listed or stated.

As used herein, unless otherwise specified the use of the ordinal adjectives "first", "second", "third", etc., to describe a common object, merely indicate that different instances of like objects are being referred to and are not intended to imply that the objects so described must be in a given sequence, either temporally, spatially, in ranking, or in any other manner.

As used herein, when referring to a measurable value such as an amount, a temporal duration, and the like, the term "about" is meant to encompass variations of $\pm 20\%$ or $\pm 10\%$, more preferably $\pm 5\%$, even more preferably $\pm 1\%$, and still more preferably $\pm 0.1\%$ from the specified value, as such variations are appropriate to perform the disclosed methods.

Ranges: throughout this disclosure, various aspects of the invention can be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, and 6. This applies regardless of the breadth of the range.

The present invention employs methods and techniques that are known to a person having ordinary skill in the art. Such methods and techniques as employed in the present invention include conventional molecular biology, microbiology, and recombinant DNA methods and techniques as disclosed and explained fully in the relevant literature, for example only, Becker's
5 World of the Cell, 9th edition, Hardin, J., et al., Pearson (2015); Essential Cell Biology, 5th edition, Alberts, B, et al., T&F/Garland (2019); Freshney's Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, Freshney, R.I., and Capes-Davis, A., Wiley-Blackwell (2021); Gel Electrophoresis: Nucleic Acids, Martin, R., Garland Science (2020); Karp's Cell and Molecular Biology, 9th edition, Karp, G., et al., Wiley (2020); Lewin's Genes, 12th edition, Krebs,
10 J.E., et al., Jones & Bartlett Learning (2017); Molecular Biology of the Cell, 6th edition, Alberts, B., et al., Garland Science (2014); Molecular Biology of the Gene, 7th edition, Watson, J., et al., Pearson (2013); Molecular Biology, 5th edition, Weaver, R., McGraw-Hill Education (2011); Molecular Biology: Principles of Genome Function, 2nd edition, Craig, N., et al., Oxford University Press (2014); Molecular Cell Biology, 8th edition, Lodish, H, W.H. Freeman (2016); Molecular
15 Cloning: A Laboratory Manual, Volumes 1, 2, and 3, 4th edition, Green, M.R., and Sambrook, J., Cold Spring Harbour Laboratory Press (2014); and Nucleic Acid Hybridization, Anderson, M.L.M., Garland Science (2020).

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood
20 to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides. ESTs, chromosomes, cDNAs, mRNAs, and rRNAs are representative examples of molecules that may be referred to as nucleic acids.

The term "equivalent," when used in reference to nucleotide sequences, is understood to
25 refer to nucleotide sequences encoding functionally equivalent polypeptides. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the nucleic acids described herein due to the degeneracy of the genetic code.

The term "variant," when used in the context of a polynucleotide sequence, may
30 encompass a polynucleotide sequence related to that of a gene or the coding sequence thereof. This definition may also include, for example, "allelic," "splice," "species," or "polymorphic" variants. The polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between
35 individuals of a given species. Polymorphic variants may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence

of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

As used herein, the terms "peptide," "polypeptide," and "protein" are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide must contain at least two amino acids, and no limitation is placed on the maximum number of amino acids that may comprise a protein or peptide's sequence. Polypeptides include any peptide or protein comprising two or more amino acids joined to each other by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. "Polypeptides" include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides, or a combination thereof.

A "fusion protein" as used herein refers to a protein wherein the protein comprises two or more proteins linked together by peptide bonds or other chemical bonds. The proteins can be linked together directly by a peptide or other chemical bond, or with one or more amino acids between the two or more proteins, referred to herein as a spacer.

Abbreviations of amino acids and nucleic acids, and analogs and derivatives of amino acids and nucleic acids will be known to a person having ordinary skill in the art. Such abbreviations may be found as published as the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB) recommendation.

In the context of the present invention, the following abbreviations for the commonly occurring nucleic acid bases are used. "A" refers to adenosine, "C" refers to cytosine, "G" refers to guanosine, "T" refers to thymidine, and "U" refers to uridine. "Transform", "transforming", and "transformation" is used herein to refer to a process of introducing an isolated nucleic acid into the interior of an organism.

The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the macromolecule. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. An

"isolated cell" or "isolated population of cells" is a cell or population of cells that is not present in its natural environment.

The term "evolved" or "evolve" as used herein refers to the change (i.e., the evolution) in the inherited characteristics of biological populations over successive generations. Evolutionary processes give rise to diversity at every level of biological organization, including species, individual organisms and molecules such as DNA and proteins (Hall and Hallgrimsson, eds. 2008, Strickberger's Evolution (4th ed.), Jones & Bartlett).

In one aspect, the present invention provides a directed evolution system, comprising:

- (i) a first nucleic acid sequence encoding one or more non-structural genes of a virus and at least one evolution candidate,
- (ii) a second nucleic acid sequence encoding one or more packaging genes, and
- (iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

wherein the first nucleic acid sequence and the second nucleic acid sequence generates a first VLP after expression in a host cell, and wherein said first VLP and the third nucleic acid sequence generates a second VLP after transduction of the first VLP and expression of the third nucleic acid sequence in a host cell,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and

wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

The virus preferably is a RNA virus, more preferably an alphavirus. As defined herein, an "Alphavirus" is a member of the Group IV Togaviridae family of viruses. Alphaviruses include, but are not limited to Aura virus, Babanki virus, Barmah Forest virus, Bebaru virus, Cabassou virus, Chikungunya virus, Eastern equine encephalitis virus, Everglades virus, Fort Morgan virus, Getah virus, Highlands virus, Kyzylagach virus, Mayaro virus, Me Tri virus, Middelburg virus, Mosso das Pedras virus, Mucambo virus, Ndumu virus, O'nyong'nyong virus, Pixuna virus, Rio Negro virus, Ross River virus, Sagiama virus, Salmon pancreas disease virus, Semliki Forest virus, Sindbis virus, Southern elephant seal virus, Tonate virus, Trocara virus, Una virus, Venezuelan equine encephalitis virus, Western equine encephalitis virus and Whataroa virus.

In a particular embodiment, the RNA virus is Semliki Forest Virus (SFV). In one embodiment, the SFV comprises SEQ ID NO: 1 or 2.

As defined herein, an alphavirus "non-structural protein" (NSP) can be selected from the group consisting of NSP1, NSP2, NSP3 and NSP4. Preferably, the one or more non-structural genes comprises NSP1, NSP2, NSP3 and NSP4 of SFV non-structural genes. The NSP1-4 proteins are translated from the first two-thirds of the viral genomic RNA. These proteins form a complex which directs replication of the genomic RNA to form antigenomic RNA. In some embodiments, the one or more non-structural genes is any one or more of SEQ ID NO. 3, 5, 7, and 9. In some embodiments, the one or more non-structural genes is any one or more of SEQ ID NO. 4, 6, 8, and 10.

A "mutation" as used therein is a change in a DNA sequence resulting in an alteration from its natural state. The mutation can comprise a deletion and/or insertion and/or duplication and/or substitution of at least one deoxyribonucleic acid base such as a purine (adenine and/or thymine) and/or a pyrimidine (guanine and/or cytosine).

Mutations may or may not produce discernible changes in the observable characteristics (phenotype) of an organism.

As used herein, "increase" or "greater" refers to expression levels which are at least 10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%, 90% higher or more, and/or 1.1 fold, 1.2 fold, 1.4 fold, 1.6 fold, 1.8 fold, 2.0 fold higher or more, and any and all whole or partial increments therebetween, than a control.

As used herein, the terms "control," or "reference" are used interchangeably, and refer to a value that is used as a standard of comparison.

"Titers" are numerical measures of the concentration of a virus or viral vector compared to a reference sample, where the concentration is determined either by the activity of the virus, or by measuring the number of viruses in a unit volume of buffer. The titer of viral stocks are determined, e.g., by measuring the infectivity of a solution or solutions (typically serial dilutions) of the viruses, e.g., on HeLa cells using the soft agar method (see, Graham & Van Der Eb (1973) Virology 52:456-467) or by monitoring resistance conferred to cells, e.g., G418 resistance encoded by the virus or vector, or by quantitating the viruses by UV spectrophotometry (see, Chardonnet & Dales (1970) Virology 40:462-477).

Preferably, the one or more packaging gene is different from the virus. In one embodiment, the one or more packaging genes is from a different virus. In a particular embodiment, the one or more packaging genes is a vesicular stomatitis virus (VSV) G DNA encoding a VSV G protein. In one embodiment, the VSV encoding the VSV G protein can be from any VSV serotype known in the art. Non-limiting examples of VSV serotypes include the Indiana (IND-VSV) serotype and New Jersey (NJ-VSV) serotype. In some embodiments, the VSV G sequence is SEQ ID NO: 11 or 12.

In one embodiment, the evolution is further driven by a chemical compound. Preferably, the chemical compound can increase mutation rate of the at least one evolution candidate. Preferably, the chemical compound is a drug, a prodrug, an analog, or a derivative thereof. In one embodiment, the drug, prodrug, analog, or derivative thereof is selected from (and not limiting to) one or more of 5-Azacytidine, 5-hydroxymethyl-2'-deoxycytidine, Decitabine, Gemcitabine, 5-hydroxydeoxycytidine, KP1212 (5,6-dihydro-5-aza-2'-deoxycytidine), 5-hydroxymethyl-2'-deoxyuridine, 5-fluorouracil, KP1461 (N4-heptyloxycarbonyl-5,6-dihydro-5-aza-2'-deoxycytidine), Ribavirin, Favipiravir, and Molnupiravir.

In one embodiment, the evolution is further driven by one or more mutations that increases the error-prone viral genome replication. Preferably, the one or more mutations that increases the error-prone viral genome replication is included in a gene encoding a replicase or a replication complex for the virus. In one embodiment, the one or more mutations that increases the error-prone viral genome replication is included in the one or more non-structural genes of the virus.

Preferably, the one or more non-structural genes further comprises a 5' untranslated region and a 3' untranslated region of SFV non-structural genes operably linked to NSP1, NSP2, NSP3 and NSP4.

In some embodiments, the 5'UTR sequence is SEQ ID NO: 13. In some embodiment, the 3' UTR is SEQ ID NO: 15.

In one embodiment, the VLP is a virus-like vesicle (VLV). VLV are SFV-based self-propagating infectious particles that propagate in the cytoplasm and produce infectious spherules containing only one structural protein, the VSV-G glycoprotein, which promotes vesicle budding and spread from infected cells.

Preferably, the one or more non-structural genes comprises at least one mutation that increases virus titer production. In one embodiment, the at least one mutation is selected from G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834-A, T-8859-A, T-8864-C, G-9211-A, A-10427-G, G-11560-A, or other modifications that improve packaging or functionality of the system. In one particular embodiment, the at least one mutation comprises G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834-A, T-8859-A, T-8864-C, G-9211-A, A-10427-G, and G-11560-A.

In another embodiment, the at least one mutation is selected from G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, E2393K or any other modifications that improve packaging or functionality of the system. In one particular embodiment, the at least one mutation comprises G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, and E2393K.

Preferably, the one or more packaging genes comprises at least one mutation that increases virus titer production. Preferably, the at least one mutation is selected from A-11871-G, T-11978-C and other modifications that improve packaging or functionality of the system.

As would be understood by the skilled person, the number indicated in the abovementioned mutations refers to the position of the nucleotide of a nucleic acid sequence, for example, the nucleic acid sequence of NSP1-4 of SFV and VSVG as disclosed in *Rose et al 2014* (15) and in PCT/US2015/030102, or the position of the amino acid of a protein sequence. In one embodiment, the VLPs disclosed herein accumulated beneficiary mutations and produced VLPs with 1000 times higher titers after 50 passages in culture than the VLPs without the mutations.

Preferably, the first nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to a DNA sequence encoding the one or more non-structural genes of the RNA virus, operably linked to a DNA sequence encoding the at least one evolution candidate. Preferably, the first nucleic acid sequence further comprises an alphavirus subgenomic RNA promoter (SGP) operably linked between the one or more non-structural genes of the RNA virus and the at least one evolution candidate. In one embodiment, the alphavirus SGP is a SFV promoter. In one embodiment, the SGP sequence is SEQ ID NO: 14.

Preferably, the second nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to the one or more packaging genes.

The term "promoter" as used herein is defined as a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a polynucleotide sequence.

As used herein, the term "promoter/regulatory sequence" means a nucleic acid sequence which is required for expression of a gene product operably linked to the promoter/regulatory sequence. In some instances, this sequence may be the core promoter sequence and in other instances, this sequence may also include an enhancer sequence and other regulatory elements which are required for expression of the gene product. The promoter/regulatory sequence may, for example, be one which expresses the gene product in a tissue specific manner.

A "constitutive" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a cell under most or all physiological conditions of the cell. In one embodiment, the promoter sequence is a cytomegalovirus immediate early promoter (CMV promoter) sequence.

An "inducible" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a cell substantially only when an inducer which corresponds to the promoter is present in the cell.

In one embodiment, the first nucleic acid sequence, the second nucleic acid sequence and/or the third nucleic acid sequence are included in at least one DNA construct. In one embodiment, the first nucleic acid sequence, the second nucleic acid sequence and the third nucleic acid sequence are each included in a separate DNA construct. Preferably, the DNA construct comprises a vector.

A "vector" is a composition of matter which comprises an isolated nucleic acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, polynucleotides associated with ionic or amphiphilic compounds, plasmids, and viruses.

In another embodiment, the first nucleic acid sequence, the second nucleic acid sequence and/or the third nucleic acid sequence are integrated into a host cell genome for stable expression.

Preferably, the evolution candidate comprises a sequence of at least one transgene of interest. The term "transgene of interest" as used herein should be understood to refer to a nucleotide sequence encoding a gene or truncated gene product of interest. Such a gene or truncated gene product of interest should be understood to refer to a gene or truncated gene product intended to be evolved in a directed evolution process as disclosed herein. It will be appreciated that the term "transgene of interest" includes variations of the transgene of interest that are a result of the directed evolution process disclosed herein. A person of ordinary skill will appreciate that a transgene of interest could be any nucleic acid encoding a gene or truncated gene product to be evolved. Accordingly, in one embodiment, the at least one transgene of interest encodes a protein, a protein domain, a protein-coding region, a non-protein-coding, a truncated protein, a truncated protein domain, a truncated-protein coding region or a truncated non-protein-coding region. In one embodiment, the evolution candidate is selected from the group consisting of Tet transactivator (tTa), reverse Tet transactivator (rtTa), and anti-p53 nanobody.

Preferably, the second VLP and the third nucleic acid sequence generates further VLPs after transduction of the second VLP and expression of the third nucleic acid sequence in a host cell, wherein the generation of the further VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. Preferably, the system further generates one or more additional VLPs through expression of the third nucleic acid and transduction of the further VLPs in a host cell, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. In one embodiment, the system is a continuous system.

As described herein, the "further VLPs" and "additional VLPs" are VLPs that contain an evolved candidate. Both the "further VLPs" and the "additional VLPs" can be used to further transduce a host cell with the expression of the third nucleic acid to generate one or more "further

VLPs” and “additional VLPs” until a desirable evolved candidate is obtained, wherein the generation of the further VLPs and the additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

5 In another aspect, the present invention provides a method of performing directed evolution of at least one evolution candidate, comprising the directed evolution system as disclosed herein.

In another aspect, the present invention provides a method of performing directed evolution of at least one evolution candidate, comprising:

- (i) expressing in a host cell with
 - 10 (a) a first nucleic acid sequence encoding one or more non-structural genes of a virus and the at least one evolution candidate, and
 - (b) a second nucleic acid sequence encoding one or more packaging genes,
- (ii) generating a first virus-like particle (VLP) by the host cell, and
- (iii) transducing a host cell with the first VLP and expressing a third nucleic acid sequence
15 encoding a target element for the at least one evolution candidate and the one or more packaging genes to generate a second VLP,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate in step (iii), and

20 wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

In one embodiment, “expression” of the first, second and/or third nucleic acid sequence in a host cell can be done by transfecting and expressing one or more DNA constructs that includes the first, second and/or third nucleic acid sequence. In another embodiment, the first nucleic acid sequence, the second nucleic acid sequence and/or the third nucleic acid sequence are
25 integrated into a host cell genome for stable expression.

Preferably, the method further comprises screening the host cells for VLPs comprising a desirable evolved candidate.

In one embodiment, the method further comprises transducing a host cell with the second VLP and expressing the third nucleic acid sequence to generate further VLPs, wherein the
30 generation of the further VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate. Preferably, the method further comprises transducing a host cell with the further VLPs and expressing the third nucleic acid sequence to generate one or

more additional VLPs, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

In another aspect, the present invention provides a virus-like particle produced by the method disclosed herein.

5 In another aspect, the present invention provides a host cell comprising the virus-like particle disclosed herein.

The term "host cell," as used herein, should be understood to refer to a cell that can host a nucleic acid sequence, a nucleic acid construct, a vector, a VLP and/or a virus. A cell can host a nucleic acid sequence, a nucleic acid construct, a vector, a VLP and/or a virus if it supports
10 expression of the nucleic acid sequence, the nucleic acid construct, the vector, the VLP and/or the virus, replication of the nucleic acid sequence, the nucleic acid construct, the vector, and/or the genome of the VLP or the virus, and/or the generation of viral particles. Preferably, the host cell is a mammalian cell, such as BHK-21 cell or HEK293T cell.

In another aspect, the present invention provides a composition comprising the virus-like
15 particle disclosed herein.

As used herein, the term "composition" or "pharmaceutical composition" refers to a mixture of at least one compound useful within the invention with other chemical components, such as carriers, stabilizers, diluents, adjuvants, dispersing agents, suspending agents, thickening agents, and/or excipients. The composition or the pharmaceutical composition facilitates administration
20 of the compound to an organism. Multiple techniques of administering a compound exist in the art including, but not limited to: intravenous, oral, aerosol, parenteral, ophthalmic, pulmonary and topical administration.

The language "pharmaceutically acceptable carrier" includes a pharmaceutically acceptable salt, pharmaceutically acceptable material, composition or carrier, such as a liquid or
25 solid filler, diluent, excipient, solvent or encapsulating material, involved in carrying or transporting a compound(s) of the present invention within or to the subject such that it may perform its intended function. Typically, such compounds are carried or transported from one organ, or portion of the body, to another organ, or portion of the body. Each salt or carrier must be "acceptable" in the sense of being compatible with the other ingredients of the formulation, and
30 not injurious to the subject. Some examples of materials that may serve as pharmaceutically acceptable carriers include: sugars, such as lactose, glucose and sucrose; starches, such as corn starch and potato starch; cellulose, and its derivatives, such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatin; talc; excipients, such as cocoa butter and suppository waxes; oils, such as peanut oil, cottonseed oil, safflower oil, sesame
35 oil, olive oil, corn oil and soybean oil; glycols, such as propylene glycol; polyols, such as glycerin,

sorbitol, mannitol and polyethylene glycol; esters, such as ethyl oleate and ethyl laurate; agar; buffering agents, such as magnesium hydroxide and aluminum hydroxide; alginic acid; pyrogen-free water; isotonic saline; Ringer's solution; ethyl alcohol; phosphate buffer solutions; diluent; granulating agent; lubricant; binder; disintegrating agent; wetting agent; emulsifier; coloring agent; 5 release agent; coating agent; sweetening agent; flavoring agent; perfuming agent; preservative; antioxidant; plasticizer; gelling agent; thickener; hardener; setting agent; suspending agent; surfactant; humectant; carrier; stabilizer; and other non-toxic compatible substances employed in pharmaceutical formulations, or any combination thereof. As used herein, "pharmaceutically acceptable carrier" also includes any and all coatings, antibacterial and antifungal agents, and 10 absorption delaying agents, and the like that are compatible with the activity of the compound, and are physiologically acceptable to the subject.

Supplementary active compounds may also be incorporated into the compositions.

As used herein, the term "effective amount" or "therapeutically effective amount" means the amount of the virus like particle generated from vector of the invention which is required to prevent 15 the particular disease condition, or which reduces the severity of and/or ameliorates the disease condition or at least one symptom thereof or condition associated therewith.

The term "ameliorating" or "treating" means that the clinical signs and/or the symptoms associated with a disease are lessened as a result of the actions performed. The signs or symptoms to be monitored will be well known to the skilled clinician.

20 The term "treatment" as used within the context of the present invention is meant to include therapeutic treatment as well as prophylactic, or suppressive measures for the disease or disorder. As used herein, the term "treatment" and associated terms such as "treat" and "treating" means the reduction of the progression, severity and/or duration of a disease condition or at least one symptom thereof. The term 'treatment' therefore refers to any regimen that can benefit a 25 subject. The treatment may be in respect of an existing condition or may be prophylactic (preventative treatment). Treatment may include curative, alleviative or prophylactic effects. References herein to "therapeutic" and "prophylactic" treatments are to be considered in their broadest context. The term "therapeutic" does not necessarily imply that a subject is treated until total recovery. Similarly, "prophylactic" does not necessarily mean that the subject will not 30 eventually contract a disease condition. Thus, for example, the term treatment includes the administration of an agent prior to or following the onset of a disease or disorder thereby preventing or removing all signs of the disease or disorder. As another example, administration of the agent after clinical manifestation of the disease to combat the symptoms of the disease comprises "treatment" of the disease.

35 As used herein, by "combination therapy" is meant that a first agent is administered in conjunction with another agent. "In combination with" or "In conjunction with" refers to

administration of one treatment modality in addition to another treatment modality. As such, "in combination with" refers to administration of one treatment modality before, during, or after delivery of the other treatment modality to the individual. Such combinations are considered to be part of a single treatment regimen or regime.

5 A "subject" or "patient," as used therein, may be a human or non-human mammal. Non-human mammals include, for example, livestock and pets, such as ovine, bovine, porcine, canine, feline and murine mammals. Preferably, the subject is human.

 The term "biological" or "biological sample" refers to a sample obtained from an organism or from components (e.g., cells) of an organism. The sample may be of any biological tissue or
10 fluid. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, bone marrow, cardiac tissue, sputum, blood, lymphatic fluid, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. Biological samples may also include sections of tissues such as frozen sections taken for histological purposes.

15 In another aspect, the present invention provides a kit for performing directed evolution, comprising.

- (i) a first nucleic acid sequence encoding one or more non-structural genes of a virus and at least one evolution candidate,
- (ii) a second nucleic acid sequence encoding one or more packaging genes, and
- 20 (iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

 wherein the first nucleic acid sequence and the second nucleic acid sequence generates a first virus-like particle (VLP) after expression in a host cell, and wherein said first VLP and the third nucleic acid sequence generates a second VLP after transduction of the first VLP and
25 expression of the third nucleic acid in a host cell,

 wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and

 wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

30 In some embodiments, a kit is provided that includes reagents, vectors, cells, software, systems, and/or apparatuses for carrying out the methods provided herein. Typically, the kit will also include instructions for performing the method as disclosed herein.

In one embodiment, the kit comprises the directed evolution system according to the first aspect. In another embodiment, the kit is for use in the method according to the second or the third aspect.

5 In some embodiments, a kit is provided for treating, preventing, or ameliorating an a given disease, disorder or condition, or a symptom thereof, as described herein wherein the kit comprises: a) a compound or compositions as described herein; and optionally b) an additional agent or therapy as described herein. The kit can further include instructions or a label for using the kit to treat, prevent, or ameliorate the disease, disorder or condition. In yet other embodiments, the invention extends to kits assays for a given disease, disorder or condition, or a symptom
10 thereof, as described herein. Such kits may, for example, contain the reagents from PCR or other nucleic acid hybridization technology (microarrays) or reagents for immunologically based detection techniques (e.g., ELISpot, ELISA).

In one particular aspect, the present invention provides an SFV-based DE system for evolving a target protein, comprising:

- 15 (i) a first plasmid comprising non-structural genes of the SFV genome and an evolution candidate, wherein the non-structural SFV genes include one or more mutations that increases virus titer when using a VSVG envelope protein instead of proteins encoded by the endogenous SFV structural genes;
- (ii) a second plasmid comprising a gene sequence encoding the VSVG envelope protein;
20 and
- (iii) a synthetic circuit comprising a target element that is responsive to the evolution candidate and a sequence encoding the VSVG envelope protein gene.

Preferably, a VLP containing the SFV non-structural genes and the evolution candidate is first produced by transfecting the first and second plasmid in a host cell, and the VLP is further
25 amplified. The evolution candidate is then subjected to direct evolution through selective pressure by transducing the VLP with the synthetic circuit. The expression of the VSVG envelope protein from the synthetic circuit (and thus production of VLP comprising the evolution target with increased activity) is controlled by the target element under the regulation of the evolution candidate.

30

Table 1 - Sequences

SEQUENCE NAME	Seq ID NO.	NUCLEOTIDE SEQUENCE (5' to 3')	Seq ID NO.	PROTEIN SEQUENCE	NOTES
SFV	1	TGGCGGATGTGTGACATACACGACGCGCAAAA GATTTTGTCCAGCTCCTGCCACCTCCGCTA CGCAGAGATTAAACACCCACGATGGCCGC CAAAGTGATGTTGATATTGAGGCTGACAGC CCATTATCAAGTCTTTGCAGAAAGGCATTCC GTCGTTGAGGTGGAGTCATTGCAGGTCACA CCAAATGACCATGCAAAATGCCAGAGCATTTT CGCACCTGGCTACCAAAATTGATCGAGCAGGA GACTGACAAAGACACACTCATCTTGGATATC GGCAGTGGCCTTCCAGGAGAAATGATGCTA CGACAAATACCACTGCGTATGCCCTATGCG CAGCGCAGAAAGACCCCGAAAGGCTCGTATG CTACGCAAGAAACTGGCAGCGGCCTCCGA GAAAGTGCTGGATAGAGAGATCGCAGGAAA AATCACCGACCTGCAGACCGTCACTGGCTACG CCAGACCGTGAATCTCCTACCTTTTGCTGCG ATACAGACGTACGCTGCTGACGGCAGCCG AAGTGGCCGTATACCGAGCGTGTATGCTGT ACATGCACCAACATCGCTGTACCATCAGGCG ATGAAAGGTGTCAGAACGGCGTATTGGATTG GGTTTGACACCCCGTTTATGTTTGACGC GCTAGCAGGCGGTATCCAACCTACGCCAC AACTGGGCCGACGAGCAGGTGTACAGGC CAGGAACATAGGACTGTGTGACGATCCTTG ACTGAGGGAAGACTCGGCAAACTGTCCATTG TCCGCAAGAGCAATTGAAACCTTGCAGACAC AGTCATGTTCTCGGTAGTACTACATTGTACA CTGAGAGCAGAAAGTACTGAGGAGCTGGC ACTTACCTCCGTATCCACCTGAAAGGTAA ACAACTCTTACCTGTAGGTGCGATACCATC GTATCATGTGAAGGTCACGTAGTTAAGAAA TCACATGTGCCCCGGCCTGTACGGTAAAC GGTAGGGTACGCCGTGACGTATCACGCGGA GGGATTCCTAGTGTGCAAGACCAAGACACT GTCAAGGAGAAAGAGTCTCATTCCTTGAT	2	MAAKVHVDIEADSPFIKSLQKAFSEVESLQVTPND HANARAFSLATKLEQETDKDTLIDIGSAPSRMM STHKYHCVCPMRSAEDPERLVYAKKLAAS E KVLD REIAGKITDLQTVMA TPADESPTFCLHTDVTCTAAE VAVYQDVYAVHAPTSLYHQAMKGVRTAYWIGFDTT PFMFDALAGAYPTYATNWADEQVLQARNIGLCAASL TEGRGLKLSILRKKQLKPCDVTMFVSGSTLYTESRKL LRSWHLPSVFLKKGQSFTCRCDTIVSCEGYVVKKIT MCPGLYGKTVGYAVTYHAEGFLVCKTDTTVKGERV SFPVCTYVPSTICDQMTG I LATDITPEDAQKLLVGLN QRIVNGRTQRNTNTMKNYLLPIVAVAFSKWAREYK ADLDEKPLGVRESLTCCCLWAFKTRKMHTMYKK PDTQTIVKVPSEFNSFVPSLWSTGLAIPVRSRIKMILL AKTKRESIPVLDASSARDAEEQEERLEAELTREAL PPLVP I APAEAGVVDVDEELEYHAGAGVETPRSA LKVTAQPN G VLLGNYVWVSPQTVLKSSKLAPVHPLA EQVKIITHNGRAGRYQVDGYDGRVLLPCGSAIPVE FQALSESATMVYNEREFVNRKLYHIAVHGPSLNTDE ENYEKVRARTDAEYVFDVKKCCVKREEASGLVLV GELTNPPFHEFAYEGLKIRPSAPYKTTVGVFGVPG SGKSAIKSLVT K HDLVTSKKKENCQEIVNDVKKHRG LDIAKTVDLSILLNGCRRVADILYVDEAFACHSGT L LA LIALVKPRSKVLCGDPKQCQGFENMMQLKVNFNHNI CTEVCHKSISSRRCRTPVTAIVS I LHYGGKMRNTNPC NKP I IIDTTGQTKPKPGDIVLTCFRGWKQLQLDYRG HEVMTAAASQGLTRKGVYAVRQKVNENPLYAPASE HVNVL I TRTEDRLVWKTLAGDPWIKVLSNIPQGNFT ATLEWQEEHDKMKVIEGPAAPVDAFQNKANVCW AKSLVPVLDTAGIRLTAEWSTIITAFKEDRAYSPVA LNEICTKYGVVDLSGLFSAPKVSLEYENHNWDNRP GGRMYGFNAATAARLEARHTFLKGQWHTGKQAVIA ERKIQLSVLDNVIPINRRLPHALVTEYKTVKGSRVE WLVNKKVRGYHVLVSEYNLALPRRRTVWLSPLNVTG ADRCYDLSGLP R LEGREFDLVFVNIHTEFRIHHYQQC VDHAMKQLMLGGDALRLLKPGGSLLMRAYGYADKIS	[...] = transgene sequence position Note: bold and underlined text indicates mutation. Long stretches of bold underlined text indicate a feature: CGGCTGGAA is the A674R/D675L/A676E attenuation mutation aATTCCTCCA CCaCGCCCaA AGAGAGCTGC ATACCTTGCC TCCCGCGCtG CGGAGCGGAC CGGTGCCctGC GCctAGAAAG CCt is the stretch with codon

		GCACCTACGTCCTCCCTCAACCATCTGTGATCA AATGACTGGCATACTGGCGACCGACATCACAA CCGAGGACGCACAGAAAGTTGTTAGTGGGA TTGAATCAGAGGATAGTTGTGAACGGAAGAA CACAGCAAAACACTAACACGATGAAGAACTA TCTGCTCCGATTGCGCCGTCGATTAGC AAGTGGCGAGGGAATACAAAGCAGACCTT GATGATGAAAAACCTCTGGGTGTCGAGAGA GGTCACTTACTTGCTGCTGCTTGTTGGGCATT TAAACGAGGAAGATGCACACCATGTACAAG AAACAGACACCCAGACAAATAGTGAAGGTGC CTTCAGAGTTTAACCTCGTTCGTCATCCCGAG CCTATGCTACAGGCCCTCGCAATCCCGATC AGATCAGCATTAAAGATGCTTTTGGCCAAGA AGACCAAGCGAGAGTCAATACCTGTTCTCGA CGCTGCTCAGCCAGGATGCTGAACAAGA GGAGAGGAGAGGTTGGAGCCCGAGCTGAC TAGAGAGCCCTTACACCCCTCGTCCCGCCG GCGCCGGCGGAGACGGGAGTCGTCGACGT CGACGTTGAAGAACTAGAGTATCACGACGT GCAGGGGCTGTGGAACACCTCGCAGCGCG TTGAAAGTCAACGCACAGCCGAACGGCGTA CTACTAGAAATTACGTAGTTCTGTCCCGC AGACCGTCTCAAGAGCTCCAAGTTGGCCC CCGTGACCCCTCTAGCAGAGCAGGTGAAAT AATAACACATAACGGGAGGGCGGCGCGTTA CCAGGTCAGCGGATATGACGCGAGGTCCT ACTACCATGTGGATCGGCCATTCGCGTCCT GAGTTTCAAGCTTTGAGCGAGAGCGCCACTA TGGGTACAACGAAAGGGAGTTCGTCAACAG GAACTATACCATATTGCGGTTACCGGACCG TCGCTGAACACCGCAGGAGGAACTACGAG AAAGTCAGAGCTGAAAGAACTGACGCGGAGT ACGTGTTTCGAGTAGATAAAAAATGCTGCGT CAAGAGAGAGGAGCGTCGGGTTTGGTGT GGTGGGAGAGCTAACCAACCCCGGTTCCA TGAAATCGCCTACGAAGGGCTGAAGATCAGG CCGTCCGACCATATAAGACTACAGTAGTAG GAGTCTTTGGGGTTCCGGGATCAGGCAAGT CTGCTATTATTAAAGAGCCTCGTGACCAACA	EAVSSLSRKFFSSARVLRPDCVTNTEVFLFSNFD NGKRPSTLHQMNKLSAVYAGEAMHTAGCAPSYRV KRADIATCTEAAVNAANARGTVGDGVCRAVAKKW PSAFGEATPVGTIKTVMCGSYPIHVAVPNFSAATTE AEGDRELAAVYRAVAEVRNLSVAIPLLSGVFS GGDRLLQSLNHLFTAMDATDADVTYCRDKSWEK KIQEADJRTTAVELLNDDVELTTDLVRVHPDSSLVGR KGYSTTDGSLYSYFEGTKFNQAIDMAEILTWPRLLQ EANEQICLYALGETMDNIRSKPVNDSSTPPRTV PCLCRYAMTAERITRLRSHQVKSMVVCSSFPPLPKYH VDGVQVKVCEKVLFDPTVPSVWSPRKYAASSTTDS DRSLRGFDLDWTTDSSSTASDTMSLPSLQSCDIDS YEPMAPIVATADVHPEPAGIADLAADVHPEPADHVDL ENIPPPRPKRAAYLASRAAERPVPAPRKPTAPRT AFRNKLPLTFGDFDEHEVDALASGITFGDFDVLRLG RAGAYIFSSDTGSGHLQKQSVRQHNLCQAQLDAVE EEKMYPPKLDTEREKLLLLKMQMHPSEANKSRYQS RKVENMKATVVDRLTSGARLYTGADVGRIPITYAVRY PRPVYPTVIERFSSPDVAIAACNEYL SRNYPTVAS QITDEYDAYLDMVDGSDSCLDRATFCPAKLRCYPKH HAYHQPTVRSAPSPFQNTLQSVLAAATKRNCNV QMRLEPTMDSAVFNVECFKRYACSGEYWEYAKQP IRITTENITTYTKLKGPKAAALFAKTHNLVPLQEVPM DRFTVDMKRDVKVTPGTHTEERPKVQVQAAEPLA TAYLCGIHRELVRRLNAVLRPNVHTLFDMSAEDFDAI ASHFHPGDPVLETDIASFDSKQDDSLALTGLMILEDL GVDQYLLDLIEAAFGIESSCHLPTGTRFKFGAMMKS GMFLTINTVNLITIASRVLEQRLTDSACAAFIGDDNI VHGVIDKLMACERCAWVNMVEKIDAVMGEKPPYF CGGFIVDSVTQTACRVSDPLKRLFKLGKPLTAEDK QDEDRRRALSDEVSKWFRITGLGAELEVALTSRYKV EGCKSILIAMATLARDIKAFKLRGPVIHLYGGPRLVR *	optimisations, where lowercase letters indicate the mutant nucleotide.
--	--	---	--	---

			CGATCTGGTCACACGCGGCAAGAGGAGAA CTGCCAGGAAATAGTCAACGACGTGAAGAAG CACCGCGGACTGGACATCCAGGCAAAAACA GTGGACTCCATCCTGCTAAACGGGTGTCGTC GTGCCGTGGACATCCTATATGTGGACGAGG CTTTCGCTTGCCATTCGGTACTCTGCTAGC CCTAATTGCTCTTGTAAACCTCGGAGCAAA GTGGTGTATCGGAGACCCCAAGCAATGC GGATTCTTCAATATGATGCAGCTTAAGGTGA ACTTCAACCACAACATCTGCACTGAAGTATGT CATAAAAGTATATCCAGACGTTGCACGCGTC CAGTCACGGCCATCGTGTCTACATTGCACATA CGGAGGCAAGATGCGCACGACCAACCCCGTG CAACAAACCCATAATCATAGACACCCACAGGA CAGACCAAGCCCAAGCCAGGAGACATCGTG TTAACATGCTTCCGAGGCTGGTAAAGCAGC TGCAGTTGGACTACCGTGGACACGAAGTCAT GACAGCAGCAGCATCTCAGGCGCTCACCCG CAAAGGGGTATACGCCGTAAGGCAGAAAGT GAATGAAAATCCCCTGTATGCCCTGCCGTGC GAGCAGTGAATGTACTGCTGACGCGCACT GAGGATAGGCTGCTGTGGAACACGCTGGCC GGCGATCCCTGGATTAAGGTCTCTATCAAAACA TCCACAGGGTAACTTACGGCCACATTGGA AGAATGGCAAGAAGAACACGACAAAAATAATG AAGGTGATTGAAGGACCGGCTGCCCTGTG GACGCGTCCAGAACAAAGCGAACGTGTGTT GGCGAAAAGCCTGGTGCCTGTCTCTGGACA CTGCCGGAATCAGATTGACAGCAGAGGAGT GGAGCACCATTAATTACAGCATTTAAGGAGGA CAGAGCTTACTCTCCAGTGGTGGCCTTGAT GAAATTTGCACCAAGTACTATGGAGTTGACC TGGACAGTGGCCTGTTTTCTGCCCCGAAGGT GTCCCTGTATTACGAGAACCAACCACTGGGAT AACAGACCTGGTGAAGGATGTATGGATTCA ATGCCGCAACAGCTGCCAGGCTGGAAGCTA GACATACCTTCTGAAGGGGAGTGGCATAAC GGGCAAGCAGGCAGTTATCGCAGAAAAGAAA AATCCAAACCGCTTCTGTGCTGGACAATGTA ATTCTTATCAACCGCAGGCTGCCCGCACGCCCC

				TGGTG <u>A</u> CTGAGTACAAGACGGTTAAAGGCAG TAGGGTTGAGTGGCTGGTCAATAAAGTAAGA GGTACCACGTCTCTGCTGGTGAGTGAGTAC AACCTGGCTTTGCCTCGACGCAGGGTCAC TT GGTTGTCACCGCTGAATGTCACAGGCGCCG ATAGTGCTACGACCTAAGTTTAGGACTGCC <u>GCGCIGGA</u> GGCAGGTCGACTTGGTCTT TGTGAACATTACACGGAATTCAGATCCAC CACTACGACGAGTGTGCGACCACGCCATGA AGCTGCAGATGCTTGGGGAGATGCGCTAC GACTGCTAAACCCGGCGGAGCCTCTTGAT GAGAGCTTACGGATACGCCGATAAAATCAGC GAAGCCGTTGTTCTCTTAAGCAGAAAGT TCTCGTCTGCAAGAGTGTGCGCCCGGATTG TGTCAACAGCAATACAGAAAGTTCCTTGCTG TTCTCCAACTTTGACAACGGAAAGAGACCCT CTACGCTACACCAGATGAATACCAAGCTGAG TGCCGTGTATGCCGGAGAACCATGCACAC GGCCGGGTGTGCACCATCCTACAGAGTTAA GAGAGCAGACATAGCCACGTGCACAGAAGC GGCTGTGTTAACGCAGCTAACGCCCGTGG AACTGTAGGGGATGCGTATGCAGGGCCGT GGCGAAGAAATGGCCGTACGCCTTTAAAGG AGAAGCAACACCAGTGGGCACAAATTAACA GTATGTGCGGCTCGTACCCCGTCATCCACG CTGTAGCGCCTAATTTCTCTGCCACGACTGA AGCGGAAGGGGACCGCGAATTGGCCGCTGT CTACCGGCAGTGGCCGCCGAAGTAAACAG ACTGTCACTGAGCAGCGTAGCCATCCCCTG CTGTCCACAGGAGTGTTCAGCGCGGGAAGA GATAGGCTGCAGCAATCCCTCAACCATCTAT TCACAGCAATGGACGCCACGGACGCTGACG TGACCATCTACTGCAGAGACAAAAGTTGGGA GAAGAAAATCCAGGAAGCCATAGACACGAG GACGGCTGTGGAGTTGCTCAATGATGACGTG GAGCTGACCACAGACTTGGTGAGAGTGCAC CCGGACAGAGCCTGGTGGGTCGTAAGGGC TACAGTACCAC TGACGGGTCGCTGACTCGT ACTTTGAAGGTACGAAATTCACCAGGCTGC TATTGATATGCCAGAGATAC TGACGTTGTGG

				CCCAGACTGCAAGAGGCAACGAACAGATAT GCCTATACGCGCTGGCGAAACAATGGACA ACATCAGATCCAAATGTCCGGTGAACGATT CGATTCAACAACACCTCCAGGACAGTGCCC TGCTGTGCCGCTACGCAATGACAGCAGAAC GGATCAGCCCGCTTAGGTACACCAAGTTAA AAGCATGGTGGTTGCTCATCTTTCCCTC CCGAAATACCATGTAGATGGGTGCAGAAG GTAAAGTGCAGAAAGTTCTCCTGTTGACC CGACGGTACCTTCAGTGGTTAGTCCGCGAA GTATGCCGCACTACGACGGACCACTCAGAT CGGTCGTACGAGGTTTACTTGGACTGGA CCACCGACTCGTCTCCACTGCCAGCGATAC CATGTGCTACCCAGTTTGCAGTCGTGTGAC ATCGACTCGATCTACGAGCCAATGGCTCCCA TAGTAGTGACGGCTGACGTACACCCTGAACC CGCAGGCATCGCGACCTGGCGGCAGATGT GCATCCTGAACCCGACAGCCATGTGGACCTC GAGAACCcaATTCCTCCACCcCGCCcaAAGA <u>GAGCTGCATACCTTGCC</u> <u>ICCCGCGCTGCGG</u> <u>AGCGACCGGTGCC</u> <u>CGGCTAGAAAGCCTAC</u> GCC TGCCCAAGGACTGCGTTTAGGAACAA GCTGCTTTGACGTTGGCGACTTTGACGAG CACGAGGTCGATGCGTTGGCTCCGGGATT ACTTCGGAGACTTCGACGACGTCCTGCGAC TAGGCCGCGGGTGCAATATTTTCTCCTC GGACACTGGCAGCGGACATTTACAACAAAA TCCGTTAGGCAGCAATCTCCAGTGCGCAC AACTGGATGCGGTCGAGGAGGAGAAAATGT ACCCGCCAAAATTGGATACTGAGAGGGAGAA GCTGTTGCTGCTGAAAATGCAGATGCACCCA TCGGAGGCTAATAAGAGTCGATACCAGTCTC GCAAAGTGAGAACATGAAAGCCACGGTGG TGACAGGCTCACATCGGGGGCCAGATTGT ACACGGGAGCGGACGTAGGCCGCAATACCAA CATACGCGGTCGGTACCCCCCCCCCGTGT ACTCCCCTACCGTGATCGAAAGATTCTCAAG CCCCGATGTAGCAATCGCAGCGTGCAACGA ATACCTATCCAGAAATTACCCAACAGTGGCG TCGTACCAGATAACAGATGAATACGACGCAT

				ACTTGGACATGGTTGACGGGTCCGGATAGTTG CTTGACAGAGCGACATTCTGCCCGCGGAA GCTCCGGTGCTACCCGAAACATCATGCGTAC CACACCCGACTGTACGCAGTGCCGTCCCG TCACCCTTTCAGAACACACTACAGAGCCGTGC TAGCGCCGCCACCAAGAGAAACTGCAACG TCACGCAATGCGAGAACTACCCACCATGGA CTCGCAGTGTCAACGTGGAGTGCITCAAG CGCTATGCCTGCTCCGGAGAAATATGGGAAG AATATGCTAAACAACCTATCCGGATAACCACT GAGAACATCACTACCTATGTGACCAAAATTGA AAGCCCCGAAAGCTGCTGCCTTGTTCCGTAA GACCCACAACTTGGTCCGCTGCAGGAGGTT CCCATGGACAGATTACCGGTCGACATGAAAC GAGATGTCAAAGTCACGCCAGGGACGAAACA CACAGAGGAAAGACCCAAAGTCCAGGTAATT CAAGCAGCGGAGCCATTGGCGACCGCTTAC CTGTGCGGCATCCACAGGGAATTAGTAAGGA GACTAAATGCTGTGTACGCCCTAACGTGCA CACATTGTTTGATATGTCGCCCGAAGACTTT GACGCGATCATGCCCTCTCACITCCACCCAG GAGACCGGTTCTAGAGACGGACATTGCATC ATTCGACAAAGCCAGGACGACTCCTTGGCT CTTACAGGTTTAAATGATCCTCGAAGATCTAG GGTGGATCAGTACCTGCTGGACTTGATCGA GGCAGCCCTTGGGGAATATCCAGCTGTAC CTACCAACTGGCACGCGCTTCAAGTTCGGAG CTATGATGAAATCGGGCATGTTTCTGACTTTG TTTATTAACACTGTTTTGAACATCACCATAGC AAGCAGGGTACTGGAGCAGAGACTCACTGA CTCCGCTGTCCGCCCTTCATCGGCGACGA CAACATCGTTCACGGAGTGATCTCCGACAAG CTGATGGCGGAGAGGTGCGCGTCGTGGGTC AACATGGAGGTGAAAGATCATTTGACGCTGTCA TGGCGAAAAACCCCATATTTTTGTGGGG ATTATAGTTTTTGACAGCGTCACACAGACC GCCTGCCGTGTTTCAGACCCACTTAAGCGCC TGTTCAAGTTGGGTAAGCCGCTAACAGCTGA AGACAAGCAGGACGAAGACAGGCGACGAGC ACTGAGTGACGAGGTTAGCAAGTGGTTCCG

		GACAGGCTTGGGGCCGAACCTGGAGGTGGC ACTAACATCTAGGTAT A GGTAGAGGGCTGC AAAAGTATCCTCATAGCCATGGCCACCTTGG CGAGGGACATTAAAGCGTTTAAAGAAATTGAG AGGACCTGTTATACACCTCTACGGCGGTCCT AGATTGGTGCGTTAATACACAGAAATCTGATT ATAGCGCACTATTATAGCACCC[...] G TTAGGGT AGCAATGGCATTGATATAGCAAGAAATTTG AAAAAGAAAAAGTTAGGGTAAGCAATGGCA TATAACCATAACTGTATAACTTGTAAACAAAGC GCAACAAGACCTGCGCAATTGGCCCCGTGG TCCGCCTCACGGAACCTCGGGCAACTCAT TTGACACATTAAATTGGCAATAATTGGAAGCTT ACATAAGCTTAAATTCGACGAATAATTGGATT TTATTTTATTTTGCAATTGGTTTTTAATATTTCC			
5'UTR	13	TGGCGGATGTGTGACATACACGACGCCAAAA GATTTTGTTCAGCTCCTGCCACCTCCGCTA CGCGAGAGATTAAACCCACG		N/A	N/A
NSP1	3	ATGGCCGCCAAAGTGCATGTTGATATTGAGG CTGACAGCCCATTCATCAAGTCTTTGCAGAA GGCATTTCCGTTCGAGGTGGAGTCATTG CAGGTCACACCAATGACCATGCAAAATGCCA GAGCATTTTCGACCTGGCTACCAAAATTGAT CGAGCAGGAGACTGACAAAGACACACTCATC TTGGATATCGCAGTGGCCTTCCAGGAGAA TGATGCTACGCACAAATACCACTGCGTATG CCCTATCGCAGCGCAGAAAGACCCCGAAAG GCTCGTATGCTACGCCAAAGAAACTGGCAGCG GCCTCCG A GAAGGTGCTGGATAGAGAGATC GCAGGAAAAATCACCGACCTGCAGACCCGTCA TGGCTACGCCAGACGCTGAATCTCCTACCTT TTGCTGCTACAGACGTCACGTGTCGTACG GACGCCGAAGTGCGCGTATACAGGACGCTG TATGCTGTACATGCACCAACATCGCTGTACC ATCAGCGGATGAAAGGTGTCAGAACGCGCT ATTGATTGGGTTTGACACCCCGCTTTTAT GTTTGACGCGCTAGCAGGCGCGTATCCAAC CTACGCCACAAACTGGGCCGACGAGCAGGT GTTACAGGCCAGGAACATAGGACTGTGTGCA	4	MAAKVHVHDIADSPFIKSLQKAFPSFEVESLQVTPND HANARAFSHLATKLEQETDKDTLILDIGSAPRRMM STHKYHCVCPMRSAEDPERLVCYAKKLAASAEKVLD REIAGKITDLQTVMATPDAESPTFCLHTDVTCTRAE VAVYQDVYAVHAPTSLYHQAMKGVRTAYWIGFDTT PFMFDALAGAYPTYATNWADEQVLQARNIGLCAASL TEGRLGKLSILRKQLKPCDVMFVSGSTLYTESRKL LRSWHLPSVFLKQKQSFCTCRCDTIVSCEGYVVKKIT MCPGLYKTVGYAVTYHAEGLVCKTDTTVKGERV SFPVCTYVSPSTICDQMTG L ATDITPEDAQKLLVGLN QRIVNGRTQRNTNMTKNYLLPIVAVAFSKWAREYK ADLDDEKPLGVREERSLTCCLWAFKTRKMHMTMYKK PDTQTIVKVPSEFNSFVPSLWSTGLAIPVRSRIKMLL AKKTKRES I PVLDASSARDAEEKEERLEAELTREAL PPLVPTAPAEETGVVDVDVVEELEYHAGA	Note: bold and underlined text indicates mutation. Contains G106E, L347L, V351I, L481S, I516T mutations (mutation numbering used to describe NSP1-4 as a polypeptide, so numbers begin at start of NSP1).

			GCATCCTTGACTGAGGGAAGACTCGGCAAAC TGTCATTCTCCGCAAGAAAGCAATTGAAACC TTGCGACACAGTCATGTTCTCGGTAGGATCT ACATTGTACACTGAGAGCAGAAAGCTACTGA GGAGCTGGCATTACCCCTCCGTATTCACCT GAAAGGTAACAATCCTTTACCTGTAGGTGC GATACCATCGTATCATGTGAAGGGTACGTAG TTAAGAAAAATCACTATGTGCCCCGGCCTGTA CGGTAAACGGTAGGGTACGCCGTGACGTA TCACGGGAGGATTCTTAGTGTGCAAGAC CACAGACACTGTCAAAGGAGAAAAGTCTCA TCCCTGTATGCACCTACGTCCCCTCAACCA TCTGTATCAAAATGACTGGCATACTGGCGAC CGACATCACACCGGAGGACGCACAGAAGTT GTTAGTGGGATTGAATCAGAGGATAGTTGTG AACGGAAGAACACACGCGAAACACTAACACGA TGAAGAACTATCTGCTTCCGATTGTGGCCGT CGCATTTAGCAAGTGGCGAGGGAATACAA GGCAGACCTTGATGATGAAAAACCTCTGGGT GTCCGAGAGAGGTCACTTACTTGCTGCTGCT TGTGGGCATTTAAACGAGGAAGATGCACAC CATGTACAAGAAACCAGACACCCAGACAATA GTGAAGGTGCTTCAGAGTTTAACTCGTTGG TCATCCCGAGCCTATGGTCTACAGGCCCTCGC AATCCCAGTCAGATCACGCATTAAGATGCTTT TGGCCAAGAAGACCAAGCGAGAGTCAATACC TGTTCTCGACGCTGTCAGCCAGGGATGCT GAACAAGAGGAGAAAGGAGGTTGGAGGCC GAGCTGACTAGAGAAGCCTTACCACCCCTCG TCCCCACCGCGCGGCGGAGACGGGAGTC GTCGACGTCGACGTTGAAGAACTAGAGTATC ACGCAGGTGCA

NSP2	5	GGGGTCGTGGAAACACCTCGCAGCGCGTTG AAAGTCACCGCACAGCCGAACGCGCTACTA CTAGGAAATTACGTAGTTCTGTCGCCGCGAGA CCGTGCTCAAGAGCTCCAAGTTGGCCCCCG TGCACCTCTAGCAGAGCAGGTGAAAAATAAT AACACATAACGGGAGGGCGCGCGCTTACCA GGTCGACGGATATGACGGCAGGGCTCTACT ACCATGTGGATCGGCCATTCCGGTCCCTGAG TTTCAAGCTTTGAGCGAGAGCGCCACTATGG TGTAACAAGAAAGGGAGTTCTGCAACAGGAA ACTATACCATATTGCCGTTACGGACCGTCG CTGAACACCGCAGGAGGAACTACGAGAAA GTCAGAGCTGAAAGAACTGACGCCCGAGTAC GTGTTGACGTAGATAAAAAATGCTGCGTCA AGAGAGAGGAAGCGTCGGTTTGGTGTGG TGGAGAGCTAACCAACCCCGGTTCCATGA ATTCCGCTACGAAGGGCTGAAGATCAGGCC GTCCGCCACCATATAAGACTACAGTAGTAGGA GTCCTTTGGGTTCCGGGATCAGGCAAGCTCG CTATTATTAAGAGCCTCGTGACCAACACGA TCTGTTACCCAGCGGCAAGAAAGGAGAACTG CCAGGAAATAGTCAACGACGTGAAGAAGCAC CGCGGACTGGACATCCAGGCAAAAACAGTG GACTCCATCCTGCTAAACGGGTGTCGTCGTG CCGTGGACATCCTATATGTGGACGAGGCTTT CGCTTGCCATTCCGGTACTCTGCTAGCCCTA ATTGCTCTTGTAAACCTCGGAGCAAGTGG TGTTATGCGGAGACCCCAAGCAATGCGGATT CTTCAATATGATGCAGCTTAAGGTGAACCTCA ACCACAACTCTGCAGCTGAAGTATGTCATAA AGTATATCCAGACGTTGCACCGCTCCAGTCA CGGCCATCGTGTCTACATTGCACTACGGAGG CAAGATGCGCACGACCAACCCGTGCAACAAA CCCATAAATCATAGACACCAAGGACAGACCA AGCCCAAGCCAGGAGACATCGTGTTAACATG CTTCCGAGGCTGGGTAAAGCAGCTGCAGTT GGACTACCGTGGACACGAAGTCATGACAGC AGCAGCATCTCAGGGCCTCACCCGCAAGG GGTATACGCCGTAAAGGCAGAAAGGTGAATGAA AATCCCCTTGTATGCCCCCTGCGTCGGAGCACG	6	GWETPRALKVTAQPN <u>G</u> VLLGNVYVLSPQTVLKSS KLAPVHPLAEQVKIITHNGRAGRYQVDGYDGRVLLP CGSAIPVEFQALSESATMVYNEREFVNRKLYHIAVH GPSLNTDEENYEKVAERDAEYVFDVKKCCVKR EEASGLVLVGENLTNPPFEHFAYEGKIRPSAPYKTTV VGVFGVPGSGSAIKSLVTKHDLVTSKKKENCQEIV NDVKHHRGLDIQAKTVDSILLNGCRRADVILYVDEAF ACHSGTLLALIALVKPRSKVLCGDPKQCGFFNMMQ LKVNFHNICTEVCHKSISRRCTRPVTAIVSILHYGG KMRTTNPCKPIIDTTGQTKPKPGDIVLTCFRGWVK QLQDYRGHEVMTAAASQGLTRKGVYAVRQKVNEN PLYAPASEHVNLLTRTEDRLVWKTLAGDPWIKVLS NIPQGNFTATLEEWQEEHDKIMKVEGPAAPVDAFQ NKANVCWAKSLVPVLDTAGIRLTAEEWSTIITAFKED RAYSPWALNEICTKYYGVDLDSGLFSAPKVSLEYEN NHWDRPGRMYGFNAATAARLEARHTFLKGQWH TGKQAVIAERKIPLSVLDNVIPINRRLPHALVTEYKT VKGSRVLEWLNKVRGYHVLLVSEYNLALPRRRVTW LSPLNVTGADRCYDLISGLPRL <u>E</u> GRFDLVFVNIHTEF RIHHYQQCVDHAMKLMQLGGDALRLKPGGSLLMR AYGYADKISEAVVSSLSRKFSSARVLRPDCVTSNTE VFLFSNFDNGKRPSTLHQMNMTKLSAVYAGEAMHTA GC	Note: bold and underlined text indicates mutation. Long stretches of bold underlined text indicate a feature. Contains D555G, L880L, T860T, A1151T mutations (mutation numbering used to describe NSP1-4 as a polypeptide, so numbers begin at start of NSP1). Also contains A674R/D675L/A 676E triple mutation for attenuation. <u>CGGCTGGAA</u> is the A674R/D675L/A 676E attenuation mutation
------	---	--	---	---	--

			TGAATGTACTGCTGACGCGCACTGAGGATAG GCTGGTGTGAAAAACGCTGGCCGGCGATCC CTGGATTAAAGGTCCTATCAACATTCACAG GGTAACTTTACGGCCACATTGGAAGAAATGGC AAGAAACACACGACAAAATAATGAAGGTGAT TGAAGACCGGCTGCGCCTGTGGACGCGTT CCAGAACAAAGCGAACGTGTGTTGGCGAA AAGCCTGGTGCCTGTCTGGACACTGCCCG AATCAGATTGACAGCAGAGGAGTGGAGCAC CATAATTACAGCATTTAAGGAGACAGAGCT TACTCTCAGTGGTGGCCTTGAATGAAATTT GCACCAAGTACTATGGAGTTGACCTGGACAG TGGCCTGTTTCTGCCCGAAGGTGTCCTG TATTACGAGAACAACCACTGGGATAACAGAC CTGGTGAAGGATGTATGGATTCAATGCCGC AACAGCTGCCAGGCTGGAAGCTAGACATACC TTCCTGAAGGGGCACTGGCATACGGGCAAG CAGGCAGTTATCGCAGAAAGAAAAATCCAAC CGCTTCTGTGCTGGACAATGTAATTCCTATC AACCGCAGGCTGCCGACGCCCTGGT <u>ACT</u> GAGTACAAGACGGTTAAAGGCAGTAGGGTTG AGTGCTGGTCAATAAGTAAGAGGGTACCA CGTCTGCTGGTGGTGGTGGTACAACTGGCT TTGCCTCGACGCGAGGTCACTTGGTTGTAC CGCTGAATGTCACAGGCGCCGATAGGTGCT ACGACCTAAGTTTAGGACTGCCCG <u>CGGCTGG</u> <u>AA</u> GGCAGGTTGACTTGGTCTTTGTGAACAT TCACACGGAAATTCAGAAATCCACCCTACCAG CAGTGTGTCGACCAAGCCATGAAGCTGCAG ATGCTTGGGGAGATGCGCTACGACTGCTAA AACCGGGCGCAGCCTCTTGATGAGAGCTTA CGGATACGCCGATAAAATCAGCGAAGCCGTT GTTTCTCCTTAAGCAGAAAGTTCTCGTCTG CAAGAGTGTGCGCCGGATTGTGCACCAG CAATACAGAAAGTTCTTGCTGTTCTCCAAC TTGACAAACGGAAAGAGACCCCTCTACGCTACA CCAGATGAATACCAAGCTGAGTGCCGCTGAT GCCGGAGAAAGCCATGCACACGGCCGGGTGT

NSP3	7	GCACCATCCTACAGAGTTAAGAGAGCAGACA TAGCCACGTGCACAGAAAGCGGCTGTGGTTAA CGCAGCTAACGCCCGTGGAACCTGTAGGGGA TGGCGTATGCAGGCGCGTGGCGAAGAAATG GCCGTAGCCCTTTAAGGGAGAAACCAACCA GTGGGCACAAATTAACACAGTCACTGTGGCGCT CGTACCCCGTCACTCCACGCTGTAGCGCCTAA TTTCTCTGCCACGACTGAAGCGGAAGGGGA CCGCGAATTGGCCGCTGTCTACCGGGCAGT GGCCGCCGAAGTAACACAGACTGTCACTGAG CAGCGTAGCCATCCCGCTGCTGTCCACAGG AGTTTCAGCGGGCGGAAGAGATAGGCTGCA GCAATCCCTCAACCATCTATTACACAGCAATG GACGCCACGGACGCTGACGTGACCATCTAC TGCAGAGACAAAAGTTGGGAGAAAGAAATCC AGGAAGCCA T AGACAGGAGACGGCTGTGG AGTTGCTCAATGATGACGTGGAGCTGACCAC AGACTTGGTGAGAGTGCACCCCGGACAGCACT CCTGGTGGTCGTAAGGGCTACAGTACCAGT GACGGGTCGCTGACTCGTACTTTGAAGGTA CGAAATTCAACCCAGGCTGCTATTGATATGGC AGAGATACTGACGTTGTGGCCACAGCTGCAA GAGGCAACGAAACAGATATGCCCTATACGGCG TGGCGAAACAATGGACAACATCAGATCCAA ATGTCCGGTGAACGATTCCGATTCTATCAACA CCTCCAGGACAGTGCCCTGCTGCTGCGCGC TACGCAATGACAGCAGAACGGATC A CCCGC CTTAGTCAACCAAGTTAAAGCATGGTGG TTTGCTCATCTTTTCCCTCCCGAAATACCAT GTAGATGGGTGCAGAAAGGTAAAGTGCGAG AAGGTTCTCCTGTTGACCCCGACGGTACCTT CAGTGGTTAGTCCGCGGAAGTATGCCGCATC TACGACGGACCACTCAGATCGGTCGTTACGA GGGTTTGACTTGACTGGACCCACCGACTCGT CTTCCACTGCCAGCGATACCATGTCGCTACC CAGTTTGAGTGTGTGACATCGACTCGATC TACGAGCCCAATGGCTCCCATAGTAGTGACGG CTGACGTACACCCCTGAACCCCGCAGGCATCG CGGACCTGGCGGCAGATGTGCATCCTGAAC CCGACAGCCATGTGGACCTCGAGAACCC aAI	8	APSYRVKRA ~~DI~~ ATCTEAAVVNAANARGTVGDGVCRA VAKKWPSAFKGEATPVGKTVMCGSYPIHAVA ~~PN~~ FSATTEAEGDRELAAYRAVAEEVNRSLSSVAIPLL STGVFSGGRDRLQQSLNHLFTAMDATDADVTIYCRD KSWKKIQEADIRTA ~~VEL~~ LNDDVELTTDLVRVHPDS SLVGRKGYSTTDGSLYSYFEGTKFNQAADMAEILTL WPRLQEANEQICLYALGETMDNIRSKCPVNDSDST PPRTVPCLCRYAMTAERITRLRSHQVKSMVVCSSFP LPKYHVDGVQKVKEKVELLLFDPTVPSWSPRKYAAS TTDHS ~~DR~~ SLRGFDLDWTTDSSSTASDTMSLPSLQS CDIDSIYEPMAPIVVTADVHPEPAGIADLAADVHPEPA DHVDLEN PIPPRPKRAAYLASRAAERPVPAPRKPT PAPRTAFRNKLPITFGDFDEHEVDALASGITFGDFD DVLRLGRAGA	Note: bold and underlined text indicates mutation. Long stretches of bold underlined text indicate a feature: Contains I1492I, M1494T, A1610T (mutation numbering used to describe NSP1-4 as a polypeptide, so numbers begin at start of NSP1). Contains a stretch of sequence codon-optimised for synthesis requirements. <u>aATTCCTCCA</u> <u>CCaCGCCCaA</u> <u>AGAGAGCTGC</u> <u>ATACCTTGCC</u> <u>TCCCGCGCtG</u> <u>CGGAGCGGAC</u> <u>CGGTGCctGC</u> <u>GCctAGAAAG</u>
------	---	--	---	---	--

		<u>ICCTCCACCaCGCCCCaAAGAGAGCTGCATA</u> <u>CCTTGCCCTCCCGCGCtGCGGAGCGACCGGT</u> <u>GCCtGCGCtAGAAAGCCtACGCCTGCCCA</u> AGGACTGCGTTTAGGAACAAGCTGCCCTTTGA CGTTCCGGCGACTTTGACGAGCACGAGGTCG ATGCGTTGGCCTCCGGGATTACTTTCCGGAGA CTTCGACGACGTCTCGGACTAGGCCCGCGC GGGTGCA		<u>CCt</u> is the stretch with codon optimisations to reduce GC-content for synthesis, where lowercase letters indicate the changed nucleotide.
NSP4	9	TATATTTTCTCTCGGACACTGGCAGCGGAC ATTACAAACAAAATCCGTTAGCGAGCACAAAT CTCCAGTCCGACAACTGGATGCGGTCGAG GAGAGAAAATGTACCCGCCAAAATTGGATA CTGAGAGGGAGAAAGCTGTTGCTGCTGAAAT GCAGATGCACCCATCGGAGGCTAATAAGAGT CGATACCACTCTCGCAAGTGGAACATGA AAGCCACGGTGGTGACAGGCTCACATCGG GGGCCAGATTGTACACGGGAGCGGACGTAG GCCGCATACCAACATACGCGGTTCCGGTACCC CCGCCCGGTGTACTCCCTACCGTGATCGAA AGATTCTCAAGCCCCGATGTAGCAATCGCAG CGTGCAACGAATACCTATCCAGAAATTACCC AACAGTGGCGTCGTACCAGATAACAGATGAA TACGACGCATACCTGGACATGGTTGACGGGT CGGATAGTTGCTGGACAGAGCGACATTCTG CCCGCGGAAGCTCCGGTGCTACCCGAAACA TCATGCGTACCACCGACCGACTGTACGCAGT GCCGTCCCGTCACCCTTTCAGAACACACTAC AGAGCGGTGCTAGCGGCCGCCACCAAGAGAA ACTGCAACGTACCGCAATGCGAGAACTACC CACCATGGACTCGGCAGTGTCAACGTGGA	10	Note: bold and underlined text indicates mutation. Contains N2015S, E2393K mutations (mutation numbering used to describe NSP1-4 as a polypeptide, so numbers begin at start of NSP1).

				GTGCTTCAAGCGCTATGCCTGCTCGGGAGAA TATTGGGAAGAAATATGCTAAACAACCTATCC GGATAACCACTGAGAACATCACTACCTATGT GACCAAATTGAAAGGCCCGAAAGCTGCTGCC TTGTCGCTAAGACCCACAACTTGGTCCGC TGCAGGAGGTTCCCATGGACAGATTCACTGGT CGACATGAAACGAGATGTCAAAGTCACTCCA GGACGAAACACACAGAGGAAAGACCCAAA GTCCAGGTAAITCAAGCAGCGGAGCCATTGG CGACCGCTTACCTGTGCGGCAATCCACAGGG AATTAGTAAGGAGACTAAATGCTGTGTTACG CCCTAACGTGCACACATTGTTTGATATGTCG GCCGAAGACTTTGACCGCATCATGCCCTCTC ACTCCACCCAGGAGACCCGGTTCTAGAGAC GGACATTGCATCATTCGACAAAAGCCAGGAC GACTCCTTGGCTCTTACAGGTTTAAATGATCCT CGAAGATCTAGGGTGGATCAGTACCTGCTG GACTTGATCGAGGCGAGCCTTTGGGAAATAT CCAGCTGTCACTACCAACTGGACGCGCTT CAAGTTCGGAGCTATGATGAAATCGGGCATG TTCTGACTTTGTTTATTAACTGTTTGAAC ATCACCATAGCAAGCAGGGTACTGGAGCAGA GACTCACTGACTCCGCTGTGCGCCTTCAT CGCGACGACAAACATCGTTACGGAGTGAT CTCCGACAAGCTGATGGCGGAGAGGTGCGC GTCGTGGTCAACATGGAGGTGAAGATCATT GACGCTGTCATGGCGAAAAACCCCATATT TTGTGGGGATTTCATAGTTTTTGACAGCGT CACACAGACCGCTGCCGTGTTTCAGACCCA CTTAAGCGCTGTTCAAGTTGGGTAAGCCGC TAACAGCTGAAGACAAAGCAGGACGAAGACA GGCGACGAGCACTGAGTGACGAGGTTAGCA AGTGTTCCGGACAGGCTTGGGGCCGAAC TGGAGGTGGCACTAACATCTAGGTATAGGT AGAGGGCTGCAAAAGTATCCTCATAGCCATG GCCACCTTGGCGAGGACATTAAAGGCGTTTA AGAAATTGAGAGGACCTGTTATACACCTCTA CGCGCGTCCCTAGATTGGTGGCTTAA

Subgenomic Promoter (SGP)	14	TACACAGAAATCTGATTATAGCGCACTATTAT AGCACC		N/A	N/A	N/A
3'UTR	15	GTTAGGGTAGGCAATGGCATTGATATAGCAA GAAAAATTGAAAAACAGAAAAAGTTAGGGTAAG CAATGGCATATAACCAATAACTGTATAACTTGT AACAAAGCGCAACAAGACCTGCGCAATTGGC CCCGTGGTCCGCTCACGGAACCTCGGGGC AACTCATATTGACACATTAATTGGCAATAATT GGAAGCTTACATAAGCTTAATTGACGACGAATAA TTGGATTTTATTTTATTTTGCAATTGGTTTTT AATATTTCC		N/A	N/A	N/A
VSVG	11	ATGAAGTGCCCTTTGTACTTAGCCTTTTATT CATTGGGGTGAAATTGCAAGTTCACCATAGTT TTCCACACAAACCAAAAAAGAACTGGAAAA ATGTTCCCTTCTAAATTACCAATTATTGCCCGTCA AGCTCAGATTTAAATTGGCATAATGACTTAAT AGGCACAGCCTTACAAGTCAAAATGCCCAAG AGTCACAAGGCTATTCAAGCAGACGCGTTGGA TGTTGATGCTTCCAAATGGGTCACTACTTGT GATTTCCGCTGGTATGACCGGAAGTATATAA CACATTCCATCCGATCCTTCACTCCACTCTGTA GAACAATGCAAGGAAAGCATTGAACAAACGA AACAAGGAACCTGGCTGAATCCAGGCTTCCC TCCTCAAAAGTTGTGGATATGCAACTGTGACG GATGCCGAAGCAGTGATTGTCCAGGTGACTC CTCACCATGTGCTGGTTGATGAATACACAGG AGAAATGGGTTGATTCACAGTTTCATCAACGGA AAATGCAGCAATTACATATGCCCCCACTGTCC ATAACTCTACAACCTGGCATTCTGACTATAAG GTCAAAAGGCTATGTGATTCTAACCTCATTTG CATGGACATCACCTTCTTCTCAGAGGACGGA GAGCTATCATCCCTGGGAAAGGAGGGCACA GGGTTCAGAAAGTAACACTCTTTGCTTATGAAAC TGGAGGCAAGGCCCTGCAAAATGCAATACTGC AAGCATTGGGGAGTCAGACTCCCATCAGGTG TCTGGTTCGAGATGGCTGATAAGGATCTCTT TGCTGCAGCCAGATTCCCTGAAATGCCAGAA GGGTCAAGTATCTCTGCTCCATCTCAGACCT CAGTGGATGTAAGTCTAAATTCAGGACGTTGA	12	N/A	N/A	N/A

			GAGGATCTTGGATTATTCCCTCTGCCAAGAA ACCTGGAGCAAAATCAGAGCGGGTCTTCCAA TCTCTCCAGTGGATCTCAGCTATCTTGCTCCT AAAAACCCAGGAACCGGCTCTTTCACCA TAATCAATGGTACCCTAAATACTTTGAGACC AGATACATCAGAGTCGATATTGCTGCTCCAA TCCTCTCAAGAAATGGTCGGAATGATCAGTGG AACTACCCACAGAAAGGGAACGTGGGATGAC TGGCACCATATGAAGACGTGGAAATTGGAC CCAATGGAGTCTGAGGACCAGTTCAGGATA TAAGTTTCCCTTATACATGATTGGACATGGTA TGTTGGACTCCGATCTTCATCTTAGCTCAAAG GCTCAGGTGTTCGAACATCCTCACATTCAAG ACGCTGCTTCGCAACTTCTCTGATGATGAGAG TTTATTTTTTGGTGATACTGGGCTATCCAAA ATCCAATCGAGCTTGTAAGAGTTGGTTCAG TAGTTGGAAAAGCTCTATTGCCCTCTTTTTCT TTATCATAGGGTTAATCATTGGACTATTCTTG GTTCTCCGAGTTGGTATCCATCTTTGCATTAA ATTAAAGCACACCAAGAAAGACAGATTTATA CAGACATAGAGATGAACCGACTTGGAAAGTA A		
--	--	--	--	--	--

So that preferred embodiments of the invention may be fully understood and put into practical effect, reference is made to the following non-limiting examples.

EXAMPLES

5 Materials and Methods

Cell Culture

BHK-21 [C-13] cells (#CCL-10) were sourced from the American Type Culture Collection (ATCC). Cells were maintained in a humidified 37°C (5% CO₂) atmosphere in MEM α (ThermoFisher, #32571101) supplemented with 5% fetal bovine serum (FBS) (Bovogen, #SFBS-F, French origin) and 10% tryptose phosphate broth (TPB) (ThermoFisher, #CM0283B), hereafter
10 defined as 'BHK-21 Growth Medium'. Fresh aliquots of cells were thawed for directed evolution campaigns once in-house passage counts reached 28.

Molecular Biology and Plasmid Construction

All SFV-based DE system and synthetic circuit plasmids were designed in SnapGene®
15 (Version 6.1). Sequences of interest (e.g. promoter sequences, transgene inserts, plasmid vector backbones etc.) were isolated by high-fidelity PCR amplification with Velocity DNA Polymerase (Bioline, #BIO-21099) or Q5® High-Fidelity 2X Mastermix (NEB, #M0492) using primers synthesized by Integrated DNA Technologies (IDT). Assembly of PCR amplicons was performed using the NEBuilder HiFi DNA Assembly Master Mix (NEB, #E2621) following manufacturer's
20 recommendations. Assembled products were transformed into NEB® 10-beta Competent *E. coli* (NEB, #C3019) and selected on LB agar plates (ThermoFisher, #22700025) supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, #A9518). Single colonies grown overnight in liquid LB broth (ThermoFisher, #12795-084) supplemented with 100 µg/ml ampicillin were processed with the ISOLATE II Plasmid Mini Kit (Bioline, #BIO-52057) for sequence verification. Plasmid constructs
25 were verified by restriction digestion and Sanger sequencing at the Australian Genome Research Facility (AGRF). Once successful assembly had been confirmed, fresh streaks of confirmed positive constructs were generated on antibiotic-supplemented LB agar plates and cultures generated for plasmid isolation using the PureYield™ Plasmid Maxiprep System (Promega, #A2393), producing the yields and purity necessary for transfection applications.

30 - *Semliki Forest Virus (SFV) System Constructs*

The initial SFV DNA plasmid construct (pSFV-DE_eGFP-LUC) was designed for expression from a constitutive CMV promoter in the mammalian expression vector pcDNA3.1(+). This parental plasmid was linearized by PCR and the amplicon purified by gel extraction. SFV

elements (5'UTR, NSP1-4, 3'UTR), the sequences of which were derived from pSFV3 (Addgene #92072) (28), were synthesized as a series of gBlocks HiFi Gene Fragments by IDT and used for NEBuilder HiFi assembly. We built fourteen point mutations into the NSP genome sequence which evolved in a capsid-deficient SFV strain for production of high-titer VLVs enveloped with VSVG (15). Two intermediates containing separate halves of the NSP1-4_eGFP-LUC_3'UTR insert were generated from which the final assembly was obtained by PCR and HiFi assembly.

Subsequently, the NSP2⁶⁷⁴ADA⁶⁷⁶ codon Vloop was mutated to ⁶⁷⁴RLE⁶⁷⁶ by HiFi assembly to attenuate cytopathic effects (29). This attenuated VLV (pSFV-DE) forms the basis of the SFV DNA replicon system presented within this study, and all transgenes for directed evolution were inserted as a direct replacement of the eGFP-LUC coding sequence by PCR and HiFi assembly.

Packaging and Amplification of SFV-VSVG DE VLVs

- Packaging (RP)

SFV-VSVG DE VLVs were packaged in BHK-21 cells following transfection with plasmid DNA. BHK-21 cells were seeded in 6-well plates at a density of 1.95×10^5 cells/well in 2 mL BHK-21 Growth Medium and incubated for 24 hours. A total of 1 µg of plasmid DNA (2:1:1 of pCMV_VSVG, pSFV-DE_[transgene] and pSFV-DE_eGFP-LUC) was diluted in Opti-MEM Reduced Serum Medium with GlutaMAX Supplement (ThermoFisher, #51985034) and transfected into cells using TransIT-2020 Transfection Reagent (Mirus Bio, #MIR5400) following the manufacturer's recommendations (employing a 3:1 transfection reagent:DNA ratio). 6 hours post-transfection, cells were rinsed twice with DPBS (Sigma-Aldrich, #D8537) and 1.2 mL of BHK-21 Growth Medium was added. VLV-containing supernatant was collected 24 hours post-transfection and centrifuged at 1000 g for 5 minutes to pellet cellular debris. Clarified supernatants were collected for titration and subsequent transduction experiments.

- Amplification (RA)

Packaged VLVs from RP were propagated further in constitutive VSVG-expressing BHK-21 cells to amplify infectious, enveloped particles. BHK-21 cells were seeded in 6-well plates (1.95×10^5 cells/well in 2 mL) in BHK-21 Growth Medium and incubated for 24 hours. A total of 1 µg of pCMV_VSVG DNA was transfected into cells using TransIT-2020 Transfection Reagent. 6 hours post-transfection, cells were rinsed twice with DPBS and neat RP VLVs added (500 µL for 6-well plates) in the presence of 8 µg/ml polybrene (Sigma-Aldrich, #H9268). Cells were incubated with VLVs for 1 hour and rinsed twice with DPBS before 1.2 mL BHK-21 Growth Medium was added for a further 23 hours of incubation. VLV-containing supernatants were collected and processed as described in *Packaging (RP)*.

VLV Titration

VLV-containing supernatants were titrated as per (11) using an NSP2-specific primer-probe set. Following collection and clarification, undiluted VLV-containing supernatants were combined with the TaqMan™ Fast Virus 1-Step Master Mix (ThermoFisher, #4444434) in a minimum of technical duplicates. Plates were run on a QuantStudio™ 7 Flex or QuantStudio™ 6 Pro Real-Time PCR System (ThermoFisher) using 'Fast' protocol parameters. Serially diluted pSFV_eGFP-LUC plasmid DNA was used to generate a standard curve for absolute quantification (ranging between 10^3 - 10^7 genome copies (gc) per reaction). Plotted standard curves were used to determine VLV titers in gc/mL. Detection thresholds were determined with no template control reactions and typically ranged between 10^2 - 10^3 gc/mL. For subsequent VLV transductions, these values were subtracted from calculated titers. Primer sequences are listed in Table 2 below.

Table 2. List of primers.

SEQ ID NO	PRIMER NAME	SEQUENCE (5' to 3')	PURPOSE	TARGET
	qPCR TITRATION			
16	CD_SFVTitre_qPCR_FWD	CAAGCCAGGAGACATCGTG	qPCR titration of PROTEUS SFV genomes	pSFV (NSP2)
17	CD_SFVTitre_qPCR_PROBE	56-FAM/AACATGCTT/ZEN/CCGAGGCTGGGTAAA/3IABkFQ/		
18	CD_SFVTitre_qPCR_REV	GCTGTGTCATGACTTCGTGTCC		
	TRANSGENE ISOLATION/AMPLIFICATION			
19	CD_SFVtransgene_F1	CGGTCCTAGATTGGTGCG	To amplify transgene sequence from the pSFV backbone by RT-PCR. F1 binds at the 3' end of NSP4; R1 binds within the viral 3'UTR.	pSFV (NSP4)
20	CD_SFVtransgene_R1	GCCATTGCTTACCCTAACTTTTTC		pSFV (3'UTR)

VLV Transduction (Evolution Round 1 (E1)-onwards)

For evolution, amplified VLVs from RA were propagated in synthetic circuit-expressing cells that require VLV-encoded transgene functionality to activate transcription of the VSVG packaging element. BHK-21 cells were seeded in 6-well plates at a density of 1.95×10^5 cells/well in 2 mL BHK-21 Growth Medium, incubated for 24 hours, and transfected using *TransIT*-2020 Transfection Reagent (Mirus Bio, #MIR5400) following the manufacturer's recommendations with a total of 1 µg of synthetic circuit-encoding plasmids. The SRF circuit used 1 µg of pSRE_VSVG; the tTA/rTA circuits used 1 µg of pTRE3G_VSVG; the Nb139 circuit supplied two plasmids at a 1:1 ratio (p53-ZF43:pZF43-VSVG), 0.5 µg/plasmid. Growth medium was replaced prior to transfection if the circuit of interest required chemical additives or adjusted media (e.g. reduced [FBS]). 6 hours post-transfection, a mock-transfected well was trypsinized and counted to calculate the volume of titrated VLV inoculum needed to achieve an MOI of 1 gc/cell. Typical counts ranged between 3 - 10×10^5 cells/well. An MOI of 1 is used to restrict circuit activation to the activity of a single transgene variant per cell to minimize crosstalk. Cells were rinsed once with DPBS before VLV was applied in a 500 µL volume of BHK-21 Growth Medium supplemented with 8 µg/ml polybrene. Cells were incubated with VLVs for 1 hour and rinsed twice with DPBS before 1.2 mL BHK-21 Growth Medium was added for a further 23 hours of incubation (supplemented

with chemical additives (e.g. doxycycline or reduced [FBS] where relevant)). VLV-containing supernatants were collected and processed as described in *Packaging (RP)*.

Each subsequent round uses the titrated VLVs produced in the preceding round to iteratively diversify, select and amplify variants of improved fitness.

5 Notes: the number of cells transduced, and hence the number of variant transgene-carrying VLVs screened per round, can be increased by proportionally upscaling to larger flask footprints. All evolution campaigns presented within this article represent experiments performed in 6-well plates. An MOI >1 could be used to achieve higher circuit activation, but dominance of high fitness mutations may be delayed by the piggybacking of low fitness variants following co-
10 transduction of a single cell. The concentration of dox used to initiate tTA and rtTA campaigns was optimized to permit propagation of VLVs.

Luciferase/Resazurin Assay

BHK-21 cells were seeded in 2 x 96-well plates at 6.6×10^3 cells/well. Cells were transfected with a total 34 ng of plasmid constructs using *TransIT-2020* Transfection Reagent
15 following the manufacturer's recommendations. At 6 hours post-transfection, cells were washed twice with DPBS, fresh BHK-21 Growth Medium added and the plates incubated for 24 hours. Cell viability (as a proxy for cell density) was determined by replacement of growth medium with 30 μ g/ml resazurin (Sigma-Aldrich, #R7017)-supplemented medium and incubation for approximately 30 min. Resorufin fluorescence was measured on an Infinite M1000 PRO
20 microplate reader (Tecan). Supplemented medium was removed before cells were rinsed once with room temperature DPBS and replaced with 30 μ L growth medium. Luciferase activity was assessed using the Steady-Glo® Luciferase Assay System (Promega, E2510) following manufacturer's recommendations in black-bottom plates. Firefly luminescence was measured on an Infinite M1000 PRO microplate reader (Tecan). Raw firefly luminescence values were
25 background-subtracted and then normalized against well-matched resazurin fluorescence prior to analyses.

Transgene Isolation

SFV RNA was extracted from 140 μ L culture supernatant using the QIAamp Viral RNA Mini Kit (Qiagen, #52904) following the manufacturer's recommendations. SFV RNA was reverse
30 transcribed into cDNA and all SFV transgene sequences, positioned between NSP4 and the viral 3' untranslated region (UTR), were PCR-amplified using the OneTaq® One-Step RT-PCR Kit (E5315S). PCR was conducted using an NSP4-specific forward primer and a viral 3'UTR-specific reverse primer (listed in Table 2 above). Amplified transgenes were resolved by gel electrophoresis using standard protocols and amplicons of the appropriate size range for the
35 transgene of interest were purified by gel extraction using the ISOLATE II PCR and Gel Kit

(Bioline, #BIO-52060). Amplicons were sequenced through long-read nanopore, short-read Illumina, or single read Sanger sequencing approaches. For Sanger sequencing, amplicon transgene DNA was blunt cloned into the linear plasmid backbone of the PCR Cloning Kit (NEB, #E1202S) using manufacturer's recommendations. Following transformation, single colonies
5 were cultured and sequenced as described in **Molecular Biology and Plasmid Construction**.

Long-read (Oxford Nanopore) Sequencing

- Sample Processing

RT-PCR-isolated transgene DNA sequences were processed to generate the libraries required for full-length transgene sequencing with Oxford Nanopore Technologies (ONT) Flongle flow cells (ONT, #FLO-FLG001). Sequencing samples were prepared following the
10 recommendations found within the ONT protocol 'Amplicons by Ligation' (version ACDE_9110_v110_revV_10Nov2020). DNA concentrations were determined using the Qubit™ dsDNA HS Assay Kit (ThermoFisher, #Q32851) and 200 fmol of DNA was processed using the NEBNext Companion Module for ONT Ligation Sequencing (NEB, #E7180). Sequencing
15 adapters were ligated onto sequences with the Ligation Sequencing Kit (ONT, #SQK-LSK110).

- Sequencing and Basecalling

Up to 40 fmol of each DNA library was loaded on ONT Flongle flow cells (R9.4.1 chemistry, #FLO-FLG001) in a MinION Mk1B Sequencer (ONT, #MIN-101B) fitted with a Flongle Adapter (ONT, #ADP-FLG001). Sequencing was performed using MinKNOW (ONT, version 4.2.8) using default
20 parameters and the following inputs: kit used, SQK-LSK110; 0.5 hours between MUX scans; basecalling, disabled. A minimum 2×10^4 raw reads were obtained per flow cell. Raw FAST5 files were basecalled using Guppy (version 4.5.2) with the minimum q-score filter set to 7.0. 'Basecalled FASTQ reads have been deposited at the Gene Expression Omnibus (GEO; GSE250502).

- Alignment

Quality-filtered basecalled reads (in FASTQ format) were processed using EPI2ME Desktop Agent (ONT, version 3.3.0.1031). Sequence reference files were uploaded into the program with the Fasta Reference Upload workflow (v2021.07.15). For each sample, all reads were aligned to reference files using the Fastq Custom Alignment workflow (v2021.03.25) using default
30 parameters.

Short-read (Illumina) Sequencing

- Sample Processing

RT-PCR-isolated transgene DNA sequences were processed further to generate the yield of DNA required for short-read Illumina sequencing with NovogeneAIT Genomics Singapore (Novogene). Extracted DNA was further amplified by high-fidelity PCR amplification with Q5® High-Fidelity 2X Mastermix (NEB, #M0492) using the same primer set as for initial transgene isolation. Amplicons were resolved by DNA gel electrophoresis and purified to have a minimum 1.5 µg for sequencing.

- *Sequencing*

Samples were sequenced with Novogene using their 'Microbial PCR Product Whole Genome Sequencing' strategy. Amplicons underwent fragmentation prior to PCR-free library preparation and sequencing using NovaSeq PE250 technology. Raw FASTQ reads have been deposited at the Gene Expression Omnibus (GEO; GSE250502).

- *Variant Nucleotide Analysis*

To identify variant nucleotides, raw FASTQ files were aligned to transgene reference sequences (FASTA) with *minimap2*. Any mate-pair issues were fixed in the output .sam files with the *samtools fixmate* command before reads were sorted (*samtools sort*) and reformatted (*samtools mpileup*). VarScan 2.3.9 (30) was used to call for variants with command *mpileup2snp* with arguments *--min-coverage 10000 --min-reads2 15 --min-var-freq 0.00025 --p-value 0.01*. Further annotation was performed with *vcf-annotator* using default commands, aligning to transgene reference sequences (GenBank format).

Immunocytochemistry

Cells were fixed with 4% PFA for 20 minutes at room temperature. Blocking was performed in PBS with 5% Normal Goat Serum, 1% BSA, 0.05% Triton-X, 0.3M Glycine for 1 hour at room temperature. Cells were then incubated with 1:400 anti-P53 mAb (Clone 7F5) (Cell Signaling Technology, #2527) in blocking solution for one hour at room temperature, followed by five washes with PBS-0.05% Triton-X. A goat anti-rabbit Alexa647 secondary antibody (Invitrogen, #A21245) and Hoechst 33142 in blocking solution were then used to incubate cells for one hour at room temperature. Cells were then washed 5 times with PBS-0.05% Triton-X, and imaged in PBS.

Epifluorescence Microscopy

Phase contrast and eGFP fluorescence images were obtained on an Axio Vert.A1 FL (Zeiss) microscope fitted with an AxioCam ICM1 camera (Zeiss 60N-C 2/3" 0.63X adapter) at 5X magnification. eGFP images were captured with a BP475/40 excitation and BP530/50 emission filter (FT500 beam splitter). Images were collected with Zen 2 Blue Edition (Zeiss, version 2.0.0.0).

High-throughput Imaging

- *Fixed*

Cells were imaged using the Opera Phenix Plus High-Content Screening system. High-throughput image analysis was performed using Harmony Software V5.1 (Perkin Elmer).

- *Live*

- 5 Cells were imaged using the Opera Phenix Plus High-Content Screening system. Live cell tracking was performed by detecting nuclear fluorescence across a period of 24 hours (~11 min intervals) using Harmony Software V5.1 (Perkin Elmer). Tracked objects with detectable nuclear GFP signal were exported.

Protein Structure Prediction

- 10 The AlphaFold2_mmseqs2 Google Colab notebook from ColabFold (v1.5.2-patch; <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>; <https://github.com/sokrypton/ColabFold>) was used to predict protein structures using default settings. The top-ranked prediction by average pLDDT was used for annotation and visualization with UCSF Chimera (Version 1.17.3), developed by the Resource for Biocomputing, Visualization,
15 and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (31).

Statistics

- Statistical analyses were performed in GraphPad Prism 9.2.0, GraphPad Software, San Diego, California USA, www.graphpad.com. All data were plotted as mean $\pm$ SEM. Cisplatin
20 treatment assays, VLV titer plots and AF plots were analyzed with two-tailed unpaired t-tests. Fold-changes in luciferase activity were statistically analyzed with a non-parametric Kruskal-Wallis test with a Dunn's multiple comparisons test (FIG. 11b) or a repeated measures one-way ANOVA with the Geisser-Greenhouse correction with a Dunn's multiple comparisons test (FIG. 11f). Means were compared to the control baseline mean. Changes in nuclear eGFP over time
25 were analyzed using a two-way ANOVA with Šídák's multiple comparisons test. p -values < 0.05 were considered significant (*, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001).

Example 1 - Capsid-deficient VLVs support host-dependent propagation of the SFV genome

- Alphavirus genomic RNA (gRNA) is recognized by a strain-specific capsid protein that
30 packages gRNA into infectious particles. While cognate packaging signals encoded in the gRNA are sufficient for encapsidation (20), there are additional redundant packaging sequences distributed throughout the alphavirus genome (21). In the context of DE, these interactions generate "cheater" particles that interfere with viral replication and contribute to a failure to recover

authentic DE products (19). While an intact capsid is essential for the pathogenicity of blood-borne viruses, the capsid protein is dispensable for *in vitro* propagation of VLVs (22).

To explore whether eliminating capsid-gRNA interactions enables robust host-dependent viral propagation, we designed a chimeric two-component system based on a self-replicating Semliki Forest Virus (SFV) replicon (23), where the infectivity of SFV VLVs is determined by the expression level of the Indiana vesiculovirus G (VSVG) coat protein in a host mammalian cell permissive to SFV replication (BHK-21, FIG. 2a). Importantly, VSVG does not encapsidate viral RNA and there are no regions of significant homology between the RNA encoding VSVG and the SFV genome, reducing opportunities for recombination events that could restore replication competence. We generated an SFV DNA replicon incorporating fourteen point mutations (four synonymous) in the Non-Structural Proteins (NSPs 1-4), which were reported to produce high-titer SFV VLVs pseudotyped with VSVG (FIG. 3a, (24)). Further, to reduce the cytopathic effects of SFV transduction, we exchanged a three amino acid loop within NSP2 with an attenuated variant (A674R/D675L/A676E (25)) to generate the pSFV-DE replicon construct (FIG. 2a, FIG. 3a). Attenuation did not affect SFV-DE/VSVG VLV titer or amplification factor (the ratio of VLVs released per VLV transduced) (FIG. 3b and 3c), indicating that reduced cytotoxicity was achieved without compromising VLV fitness.

Implementation of chimeric VLVs for mammalian DE campaigns requires complementation by host cell expression of VSVG. To confirm host-dependence, SFV-DE VLVs carrying an eGFP-P2A-Luciferase (eGFP-LUC) reporter were propagated for multiple rounds at high titer ($>10^8$ genome copies (gc)/mL; FIG. 2b, R1-R3) and then used to infect VSVG- or mock-transfected host cells (FIG. 2b, R4-R5). VSVG-transfected cells supported SFV-DE VLV titers leading to an amplification factor >1000 (FIG. 2c). Without VSVG, however, VLV titers drop below the limit of detection and these mock-transfected (VSVG negative) cells showed an amplification factor <1 .

While eGFP expression was maintained in cells that constitutively expressed VSVG (FIG. 3d), no eGFP expression was detectable after two rounds of transduction on mock-transfected cells (FIG. 3e). Of note, VSVG-expressing cells showed a gradual reduction in eGFP levels over 5 rounds of transduction with SFV-DE VLVs (FIG. 3d). Since VSVG levels were constitutive and not dependent on the viral transgene, and since smaller viral genomes have a replicative advantage (26), we hypothesized that the reduced eGFP signal was due to truncation of the GFP transgene over rounds of transduction. Indeed, rounds of transduction were accompanied by progressive truncation of the viral transgene (FIG. 3f), suggesting that selective pressure is required to maintain a full-length viral transgene.

DE requires diversification of the target transgene to produce variants with increased fitness. Alphaviruses are error-prone with reported mutation frequencies $>10^{-4}$ per nucleotide in

each round of replication (27). Using non-viral DNA and RNA templates, we established a detection limit of 0.3% for new mutations by amplicon deep sequencing (dotted line, FIG. 2d). We observed an accumulation of bona fide mutations (~1 per 10⁴ transduced cells) during the propagation of VLVs carrying the unselected eGFP-LUC reporter (FIG. 2d, RP-R2). Multiple substitution types were observed with a strong A-to-G (and complementary U-to-C) transition mutational bias (FIG. 2e, Table 3 (28)).

Table 3. eGFP-LUC variants. All substitutions in transgene RNA detected at $\geq 0.3\%$ over three rounds of VLV propagation.

eGFP-LUC				RP	R1	R2
Protein Domain/Element	Amino Acid Substitution	Nucleotide Substitution	Substitution Type	Allele Frequency (%)		
eGFP	K4M	A>U	Transversion		1.64%	2.11%
	G5D	G>A	Transition	0.62%	1.10%	1.47%
	F9Y	U>A	Transversion	0.48%	0.76%	1.06%
	V17D	U>A	Transversion	0.96%	1.54%	2.22%
	F72L	U>C	Transition			0.39%
	S73G	A>G	Transition			0.31%
	F84L	U>C	Transition			0.34%
	Y107C	A>G	Transition			0.33%
	K108R	A>G	Transition		0.33%	0.43%
	K108E	A>G	Transition			0.37%
	K114R	A>G	Transition		0.31%	0.33%
	T119A	A>G	Transition			0.34%
	N122S	A>G	Transition			0.31%
	K127R	A>G	Transition		0.32%	0.70%
	I129V	A>G	Transition		0.33%	1.60%
	K132R	A>G	Transition			0.32%
	N136S	A>G	Transition			0.45%
	K141R	A>G	Transition	0.95%	0.56%	0.30%
	E143Q	G>C	Transversion			0.30%
	N145S	A>G	Transition			0.34%
	N150S	A>G	Transition			0.32%
	K157R	A>G	Transition			0.30%
	K159R	A>G	Transition		0.36%	0.46%
	K163R	A>G	Transition			0.43%
	F166L	U>C	Transition		0.30%	0.31%
	K167R	A>G	Transition			0.34%
	V177A	U>C	Transition			0.31%
	K210K [*]	A>G	Transition		0.30%	0.37%
	T226A	A>G	Transition			0.30%
	I230T	U>C	Transition			0.34%
Linker-P2A	S9S [*]	U>C	Transition	0.33%		
	Q13Q [*]	A>G	Transition	0.31%		
	G15G [*]	A>G	Transition		0.31%	
	D16G	A>G	Transition	0.36%		
	E18E [*]	A>G	Transition	0.30%	0.30%	0.32%
LUC	N20D	A>G	Transition			0.30%
	E2E [*]	A>G	Transition			0.30%
	E18E [*]	A>G	Transition	0.30%	0.30%	0.32%

Q25L	A>U	Transversion		0.42%	0.32%
H27R	A>G	Transition		0.30%	
K31R	A>G	Transition			0.34%
Y33C	A>G	Transition			0.39%
D44G	A>G	Transition			0.30%
A45A^	A>G	Transition			0.30%
T52A	A>G	Transition			0.41%
E65E^	A>G	Transition			0.30%
Y70C	A>G	Transition			0.30%
N75D	A>G	Transition		0.32%	0.33%
C81Y	G>A	Transition			0.40%
S82G	A>G	Transition			0.31%
N84D	A>G	Transition			0.32%
F89S	U>C	Transition		0.59%	0.61%
F97L	U>C	Transition	0.34%		
G99G^	U>C	Transition	0.38%		
N106D	A>G	Transition			0.34%
N106S	A>G	Transition			0.61%
D107G	A>G	Transition	0.37%		
Y109C	A>G	Transition			0.50%
N110S	A>G	Transition			0.31%
L115P	U>C	Transition	0.39%		
N116S	A>G	Transition			0.54%
M118V	A>G	Transition			0.59%
S121G	A>G	Transition			0.33%
S129R	C>A	Transversion	0.31%		
S129G	A>G	Transition			0.38%
K130R	A>G	Transition	0.61%	0.41%	0.51%
K131E	A>G	Transition	0.51%	0.37%	0.41%
K131K^	A>G	Transition	0.86%	0.46%	0.99%
K131R	A>G	Transition			0.54%
L133P	U>C	Transition			0.33%
Q134Q^	A>G	Transition			0.71%
K135R	A>G	Transition	0.36%	0.53%	1.41%
K135E	A>G	Transition			0.57%
I136V	A>G	Transition			0.31%
N138S	A>G	Transition		0.33%	0.57%
N138D	A>G	Transition			0.50%
Q140R	A>G	Transition	0.33%		
Q140Q^	A>G	Transition		0.32%	0.79%
K141R	A>G	Transition	0.95%	0.56%	0.30%
K141E	A>G	Transition			0.48%
K142E	A>G	Transition	0.68%	0.35%	0.42%
K142R	A>G	Transition	0.57%	0.43%	0.43%
I146M	A>G	Transition			0.46%
K148R	A>G	Transition			0.94%
I149V	A>G	Transition	0.33%		0.64%
I149T	U>C	Transition	0.32%		
I151T	U>C	Transition	0.30%		
S154G	A>G	Transition	0.30%		0.58%
K155E	A>G	Transition	0.31%		0.33%
K155R	A>G	Transition	0.42%	0.39%	2.13%
T156A	A>G	Transition		0.31%	
Q159R	A>G	Transition			0.30%
Q162Q^	A>G	Transition			0.31%

	S163G	A>G	Transition			0.31%
	M164T	U>C	Transition	0.39%		
	Y165C	A>G	Transition			0.32%
	T169A	A>G	Transition			0.52%
	P173P [^]	A>G	Transition			0.45%
	N177S	A>G	Transition			0.41%
	Y179C	A>G	Transition			0.65%
	S185G	A>G	Transition	0.40%	0.34%	0.30%
	D189G	A>G	Transition	0.41%		
	K190K [^]	A>G	Transition			0.38%
	T191A	A>G	Transition		0.32%	0.32%
	H192T	U>C	Transition	0.54%		
	M196V	A>G	Transition			0.40%
	N197D	A>G	Transition	0.38%		
	N197S	A>G	Transition	0.34%	0.31%	0.33%
	S198G	A>G	Transition	0.32%		
	S199G	A>G	Transition			0.43%
	T202A	A>G	Transition			0.43%
	K206R	A>G	Transition		0.31%	0.72%
	G207D	G>A	Transition	0.30%		
	V208V [^]	A>G	Transition	0.36%	0.36%	0.92%
	V208A	U>C	Transition			0.30%
	H212R	A>G	Transition	0.35%		

[^]Synonymous mutations.

Application of selective pressure to a target transgene requires a tight link between the resulting expression level of VSVG and viral infectivity (i.e. fitness). To test whether transgene-dependent VSVG induction gives VLVs a selective advantage, we used a circuit that is activated by the tetracycline-controlled transactivator (tTA; FIG. 2f) in the absence of doxycycline (dox) (FIG. 2g). VLVs carrying the circuit-activating tTA transgene were serially diluted with VLVs carrying a neutral eGFP-LUC transgene and propagated in the absence of dox (FIG. 2h). Within 3 rounds, the titers of populations containing tTA VLVs diverged from the non-activating eGFP-LUC control (FIG. 2h). This was consistent with the propagation-associated enrichment of the tTA transgene at dilutions up to 1:1000 (FIG. 4). We next tested a serum-responsive circuit (FIG. 5 a and b, (19)) in which expression of the serum response factor DNA binding domain (SRF[DBD]) fused to the VP64 activation domain (SRF-VP64) further enhanced circuit activity (FIG. 5c). VLVs carrying the circuit-activating SRF-VP64 transgene exhibited a small proliferative advantage over those carrying eGFP-LUC (FIG. 5d and e), which correlated with increasing prevalence in a direct competition experiment within four rounds (FIG. 5f-h). Selection rapidly favored a shorter SRF transgene that retained a minimal DNA-binding domain (29) (FIG. 5i). Chimeric VLVs do not indiscriminately package significant amounts of VSVG RNA, and none was detected by R4 (FIG. 5j), indicating that VLV propagation is completely dependent on host cell expression of VSVG. The SFV-VSVG VLV DE system, therefore, resolves problems identified in other virus-based mammalian DE systems (19).

Example 2 – An SFV-VSVG VLV DE system generates authentic evolution products

To test whether a circuit linking a VLV-encoded transgene to VSVG production can evolve a new protein function, we used a tTA-regulated circuit (FIG. 2f) to select for doxycycline (dox) resistance in tTA. Saturating clonal selections in bacteria have identified point mutations distributed across the tTA protein that confer dox resistance (30), some of which are also sufficient to provide dox resistance in mammalian cells (15). Having confirmed that tTA activity was suppressed by dox in BHK-21 cells using a luciferase reporter under the control of an optimized tetracycline response element (TRE3G (31); Fig. 2g), we asked whether tTA activity could support VLV propagation on a TRE3G-regulated VSVG circuit under mild selective pressure. tTA and eGFP-LUC VLVs were independently packaged and amplified in cells constitutively expressing VSVG before switching to a TRE3G-regulated circuit exposed to minimally inhibiting dox (0.1 ng/mL, E1-E3, gray box), where tTA provided a large selective advantage over the neutral eGFP-LUC transgene (FIG. 6a and b). Note, all VLVs amplified in cells constitutively expressing VSVG show efficient transduction at E1 so the exponential selective advantage of circuit activation is observed from round E2.

We conducted two independent ten-round DE campaigns to isolate dox-resistant tTA variants (FIG. 7a). A R158G variant (resulting from an A-to-G transition (FIG. 2e)) was detected in both campaigns by E4 (Table 4), with different second-site mutations appearing by E6: Q32R (Campaign 1, FIG. 6c) and a triple mutant D178G/H179R/Q180R (Campaign 2, FIG. 7b and c). Both variants from Campaign 1 have been previously identified as dox resistance mutants (30, 32). In isolation, R158G provided modest dox resistance, while Q32R had minimal effect (FIG. 6d, Table 5). However, the Q32R/R158G double mutant showed strong resistance to fully inhibitory concentrations of dox (FIG. 6d), consistent with their synchronized increase in frequency once both appeared in the population (FIG. 6c, Table 4). Similarly, in Campaign 2, the DHQ-to-GRR mutation enhanced the dox resistance of the initial R158G mutation (FIG. 7d, Table 5). Within this cluster of mutations at positions 178-180, all of the dox resistance was associated with the previously identified D178G variant (FIG. 7e, (30)), indicating that H179R/Q180R were passengers arising from a complex mutational event. Structural modelling of the parental tTA protein with AlphaFold2 (33) was congruent with the drug-bound crystal structure (FIG. 7f). Therefore, we modelled the dox resistant mutations from both campaigns, revealing side chain rearrangements in a linker region distal to the drug-binding site (FIG. 6e, FIG. 7g). Together, these campaigns recovered authentic evolution products in tTA, validating SFV-DE VLVs as an efficient platform for DE.

Table 4. tTA variants. Non-synonymous substitutions in transgene RNA detected at $\geq 1\%$ during two SFV-VSVG VLV DE system campaigns.

tTA Campaign 1			RA	E2	E4	E6	E8	E10
Protein Domain	Amino Acid Substitution	Nucleotide Substitution	Allele Frequency (%)					
TetR	Q32R	A>G				1.25	8.10	36.66
	K48R	A>G						1.21
	H100R	A>G					0.50	1.44
	H151Y	C>T					1.81	2.83
	R158G	A>G			0.42	0.87	12.36	42.72
	P161L	C>T				1.68		
VP64	D262N	G>A		2.23	0.34			
[dox] (ng/mL)			0	0.1	0.1	1	0.5	0.5

tTA Campaign 2			RA	E2	E4	E6	E8	E10
Protein Domain	Amino Acid Substitution	Nucleotide Substitution	Allele Frequency (%)					
TetR	K33R	A>G		0.52	0.47	1.48		
	K48R	A>G					1.27	
	S92G	A>G					1.24	
	T103A	A>G					1.27	
	H151R	A>G					1.29	
	R158G	A>G			0.33	0.81	56.92	95.99
	T160A	A>G			0.49	1.57		
	D178G	A>G				0.65	55.60	95.92
	H179R	A>G				0.59	55.41	95.96
	Q180R	A>G				0.58	55.93	95.97
	K199R	A>G					1.17	
VP64	S205G	A>G					1.29	0.30
	S264G	A>G			0.70	1.50		
[dox] (ng/mL)			0	0.1	0.1	1	0.5	0.5

Table 5. Analysis of tTA and rtTA-3G mutants. Half maximal inhibitory/effective concentrations (IC₅₀/EC₅₀) and 95% confidence intervals (95% CI) for tTA and rtTA-3G figures.

Figure 6D (tTA)	Variant					
	WT	R158G	Q32R	Q32R/R158G		
IC ₅₀	0.27	1.22	0.48	4.53		
95% CI	0.21 to 0.35	0.84 to 1.74	0.35 to 0.66	2.85 to 7.05		
Figure 8E (rtTA-3G)	Variant					
	WT	M59I	D5N	D5N/M59I		
EC ₅₀	38.65	12.1	15.94	7.02		
95% CI	28.14 to 55.24	7.61 to 19.56	10.56 to 24.55	3.92 to 12.29		
Figure 97D (tTA)	Variant					
	WT	R158G	D178G/H179R/ Q180R	R158G/D178G/ H179R/Q180R		
IC ₅₀	0.25	0.78	1.56	9.5		
95% CI	0.19 to 0.34	0.56 to 1.07	1.05 to 2.36	3.57 to 22.66		
Figure 97E (tTA)	Variant					
	WT	R158G	D178G	H179R	Q180R	R158G/D178G
IC ₅₀	0.25	1.16	4.5	0.27	0.27	20.35
95% CI	0.21 to 0.29	0.84 to 1.6	3.26 to 6.15	0.2 to 0.38	0.18 to 0.42	6.57 to 335.6

Example 3 - Enhancing drug-inducible transcriptional control

We next asked if the SFV-VSVG VLV DE system can be used to improve existing molecular tools. Here, we focused on increasing the dox sensitivity of the third-generation reverse tetracycline-controlled transactivator (rtTA-3G), which has been extensively optimized using other methods (14, 34). Using a TRE3G-regulated circuit (FIG. 8a), the parental rtTA-3G protein had an EC₅₀ of 39 ng/mL dox (FIG. 8b). On the TRE3G-regulated circuit, rtTA-3G VLVs had a large fitness advantage over neutral eGFP-LUC VLVs at 100 ng/mL dox (FIG. 8c, FIG. 9a). We propagated rtTA-3G VLVs for 30 rounds, adjusting the dox concentration to maintain strong selective pressure for increased dox sensitivity (FIG. 9b). The VLVs were sequenced every 5 rounds (FIG. 9c), revealing an M59I variant that appeared early in the campaign and reached fixation by E30 (FIG. 8d, Table 6). A second D5N variant appeared by E20 and the majority of rtTA-3G transgenes (57.75%) carried both mutations by E30 (FIG. 8d, FIG. 9d, Table 6). No additional variants were detected by E60 at 3 ng/mL dox suggesting that this campaign reached a local fitness peak. Both single mutations individually enhanced the dox sensitivity of rtTA-3G and the D5N/M59I double mutant was further improved (EC₅₀ 7 ng/mL; FIG. 8e, Table 5), without increased leakiness in the absence of dox (FIG. 9e). Both mutations were distal to the dox binding site and were not predicted to cause any major structural changes (FIG. 9f). Overall, the D5N/M59I variant has a superior dox response profile and will find broad utility as a fourth generation rtTA tool (rtTA-4G) in challenging applications that require tight control of gene expression.

Table 6. rtTA-3G variants. Non-synonymous substitutions in transgene RNA detected at $\geq 1\%$ of the population during an SFV-VSVG VLV DE system campaign.

rtTA			RA	E5	E10	E15	E20	E25	E30
Protein Domain	Amino Acid Substitution	Nucleotide Substitution	Allele Frequency (%)						
rTetR	D5N	G>A					9.02	36.10	51.75
	M59I	G>A		0.72	26.16	35.44	85.13	88.71	99.81
	I194M	A>G			3.59	1.09			
	G203R	T>C						3.04	0.33
VP64	T209P	A>C		1.64					
	F228L	T>C				1.10			
[dox] (ng/mL)			0	25	30	50	10	15	12.5

Example 4 - Directed evolution of an intracellular nanobody

Intracellular nanobodies (Nb) have potential for interrogating or modulating “undruggable” targets but are prone to instability when localized inside mammalian cells (35). For example, a p53 biosensor based on Nb139 (FIG. 10a, (36)) does not localize to the nucleus in response to cisplatin (FIG. 10b) despite robust nuclear p53 accumulation (FIG. 10c). To improve Nb139 interaction with p53, we conducted a DE campaign regulated by Nb139 activating a p53-

dependent synthetic circuit. Using a 2-hybrid circuit design with a p53 bait (FIG. 11a), the parental Nb139-VP64 fusion successfully activated a reporter circuit (FIG. 11b), and Nb139-VP64 VLVs outcompeted neutral eGFP-LUC VLVs (FIG. 11c, FIG. 10d). Long-term propagation of Nb139-VP64 VLVs (FIG. 10e) led to the accumulation of S26P and Y60C mutations that evolved to fixation in the population (FIG. 11d, FIG. 10f, Table 7). S26P and Y60C map to framework region (FR) 1 and FR2, respectively, indicating that these variant residues do not interact directly with p53 (FIG. 11e). However, only the S26P variant increased reporter activity at early and late timepoints (FIG. 11f), with no further increase observed in the S26P/Y60C double mutant. To test whether S26P improved the sensitivity of a p53 biosensor (FIG. 10a), we expressed Nb139-eGFP fusions (parental and variants) in BHK-21 cells. In response to cisplatin, Nb139[S26P]-eGFP translocated to the nucleus and labeled nuclear puncta prior to cell death, in contrast to uniform expression throughout the cell in tGFP controls (FIG. 11g, FIG. 12). The S26P single mutant exhibited the largest response to cisplatin, while Y60C also modestly improved sensitivity in the biosensor format (FIG. 11h). This campaign demonstrates that the intracellular function of Nb139 (crystal structure 4QO1), which was already classified as stable (35), can be further improved through evolution within a mammalian cell. Notably, the evolved variants did not directly alter the Nb139-p53 binding interface. Overall, our application of the SFV-VSVG VLV DE system here has generated a novel p53 biosensor that allows the visualization of p53 nuclear recruitment *in vivo*.

Table 7. Nb139 variants. Non-synonymous substitutions in transgene RNA detected at $\geq 1\%$ during a SFV-VSVG VLV DE system campaign.

Nb139-VP64			RA	E5	E10	E15	E20	E26	E30	E34
Protein Domain	Amino Acid Substitution	Nucleotide Substitution	Allele Frequency (%)							
Nb139 Scaffold	L5F	C>T		15.19						
	S26P	T>C					3.09	73.68	94.14	99.36
Nb139 CDR [^]	Y33C	A>G			37.96	20.44				
	A34V	C>T				2.09				
Nb139 Scaffold	Q40R	A>G		2.75	35.93	22.29	0.98	0.33	2.48	0.54
	R44G	A>G		2.63	34.21	21.24				
	Y60C	A>G		37.88	63.99	77.12	99.92	99.88	99.90	99.90
	R68K	G>A		1.30						
	R68G	A>G								1.60
	R73G	A>G							1.69	0.96
	V80A	T>C								2.06
Tags	I136V	A>G						9.15	11.41	2.65
	N136S	A>G						9.15	11.32	2.64
	L141S	T>C					1.86	13.92	2.43	0.32
	T146A	A>G		0.32	14.77	3.36				
	T148A	A>G			6.29			8.81	11.17	2.81
	H153R	A>G							2.07	
Linker	S158G	A>G							2.85	0.46
VP64	F183S	T>C				2.19				
	D184G	A>G						9.66	11.99	2.89

[^]CDR = complementarity-determining region

Discussion

Mammalian directed evolution aims to generate biomolecules that are optimized *in situ* for a desired activity. This is dependent on a high mutation rate to diversify the target and a robust link between function and fitness to maintain the integrity of the system, technical challenges that have not yet been overcome. To this end, we have developed a novel DE platform based on a chimeric SFV design, which takes advantage of the high error rate inherent in alphavirus replication while solving major issues with previous strategies (11).

Natural viral infections generate a diversity of virions encoding incomplete genomes that collectively compete with host defenses (37). Our work was motivated by the recent observation that capsid-genome interactions are distributed throughout the alphavirus genome (21), which contribute to the production of non-functional "cheater" particles containing incomplete genomes that encode structural proteins and contaminate DE campaigns (19). The capsidless VLVs used in the present DE system utilize the cellular exosome pathway to escape the host cell (24), eliminating the co-evolved and highly redundant links between the viral genome and its cognate capsid. This approach maintained system integrity and enabled long-term propagation (>30 rounds) on a synthetic DE circuit. These results pave the way for the adaptation of the present DE system to other mammalian cell types to provide tissue- or disease-specific environments for DE campaigns. The effectiveness of the present DE campaigns using VLVs depends on how tightly the transgene function is coupled to VLV fitness. We confirmed that the present DE system generates authentic evolution products by recovering well-characterised doxycycline resistance mutations in tTA. While single mutations appeared in the VLV population at low frequency, they were rapidly outcompeted by double mutant variants that exhibited greater dox resistance, highlighting the power of the present DE system to generate complex mutations that would be difficult to generate through existing technologies. For example, sampling the complete double mutant sequence space of tTA would require testing 20^{247} combinations, which is beyond any existing experimental approach. In a more challenging test, the sensitivity of the highly optimized rtTA-3G protein was further increased by long-term propagation on limiting amounts of doxycycline. This shows that selective pressure can be maintained over extended DE campaigns. Finally, by optimizing the intracellular function of a nanobody, we generated a p53 biosensor that responds to DNA damage. We anticipate that the present DE system will be suitable for the generation or optimization of diverse biomolecules designed to function in complex mammalian systems.

Table 8. Other sequences

Sequence Name	SEQ ID NO.	Sequence
pSFV_eGF P-P2A- luc2pre (3'UTR)_RL Eloop	21	1 gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc
		61 catatatgga gtccgcgtt acataactta cggtaaatgg ccgcctggc tgaccgcca
		121 acgaccccc ccattgaag tcaataatga cgtatgttcc catagtaacg ccaataggga
		181 ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtaacac
		241 aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaaa tggccgcct
		301 ggcattatgc ccagtacatg accttatggg accttctac ttggcagtac atctacgtat
		361 tagtcacgc tattaccatg glgaltcgtt ttggcagla calcaatggg cgtggatagc
		421 ggttgactc acggggattt ccaagtctcc acccattga cgtcaatggg agttgtttt
		481 ggcacaaaa tcaacgggac ttccaaaaat gtctaaacaa ctccgcccc ttgacgcaaa
		541 tgggcggtag cgtgtacgg tgggaggtct atataagcag agctctctgg ctaactagag
		601 aaccactgc ttactggctt atcgaaatta atagactca ctatagggag acccaagctg
		661 gctagcatgg cggatgtgtg acatacacga cgcaaaaaga ttttgttcca gtcctgcca
		721 cctccgtac gcgagagatt aaccaccac gatggccgc aaagtgcacg ttgatatga
		781 ggctgacagc ccattcatca agtcttttga gaaggcattt ccgtcgttcc aggtggagtc
		841 attgcaggtc acaccaaatg accatgcaaa tgccagagca ttttcgcacc tggctaccaa
		901 attgatcgag caggagactg acaaaagcac actcatcttg gatatcgga gtgcgccttc
		961 caggagaatg atgtctacgc acaaatacca ctgcgtatgc cctatgcga ggcagaaga
		1021 cccgaaaag ctcgtatgct acgcaaaaga actggcagcg gctcccgaga aggtgctgga
		1081 tagagagatc gcaggaaaaa tcaccgacct gcagaccgtc atggctacgc cagacgtga
		1141 atctcctacc ttttgctctg atacagacgt cacgtgtcgt acggcagccg aagtggccgt
		1201 ataccaggac gtgtatgctg tacatgcacc aacatcgctg taccatcagg cgatgaaagg
		1261 tgtcagaacg gcgtattgga ttgggtttga caccacccc ttatatgttg acgcgtagc
		1321 aggcgcgtat ccaacctacg ccacaaactg ggcgcagcag caggtgttac aggccaggaa
		1381 cataggactg tgtgcagcat ccttgactga ggaagactc ggcaaacctg ccattctccg
		1441 caagaagcaa ttgaaacctt gcgacacagt catgttctcg ttaggatcta cattgtacac
		1501 tgagagcaga aagctactga ggagctggca cttaccctcc gtattccacc tgaaaggtaa
		1561 acaatccttt acctgtaggt gcgataccat cgtatcatgt gaagggtacg tagttaagaa
		1621 aatcactatg tgcccggcc tgtacggtaa aacggtaggg tacgccgtga cgtatcacgc
		1681 ggagggattc ctagtgtgca agaccacaga cactgtcaaa ggagaaagag tctcattccc
		1741 tgtatgcacc tacgtcccct caaccatctg tgaataaatg actggcatac tggcgaccga
		1801 catcacaccg gaggagcac agaagttgtt agtgggattg aatcagagga tagttgtgaa
		1861 cggaagaaca cagcgaataa ctaacacgat gaagaactat ctgcttccga ttgtggccgt
		1921 cgcatttagc aagtggcgga gggaatacaa ggcagacctt gatgatgaaa aacctctggg
		1981 tgtccgagag aggtcactta ctgtcgtctg cttgtgggca tttaaacga ggaagatgca
		2041 caccatgtac aagaaaccag acaccagac aatagtgaag gtgccttcag agtttaactc

2101	gttcgtcatc	ccgagcctat	ggtctacagg	cctcgcaatc	ccagtcagat	cacgcattaa
2161	gatgcttttg	gccaaaga	ccaagcgaga	gtcaatacct	gttctcgacg	cgtcgtcagc
2221	cagggatgct	gaacaagagg	agaaggagag	gttgaggcc	gagctgacta	gagaagcctt
2281	accaccctc	gtcccacccg	cgccggcgga	gaaggagtc	gtcgacgtcg	acgttgaaga
2341	actagagtat	cacgaggtg	caggggtcgt	gaaacacct	cgcagcgtg	tgaagtcac
2401	cgcacagccg	aacggcggtac	tactaggaaa	ttacgtagtt	ctgtccccg	agaccgtgct
2461	caagagctcc	aagttggccc	ccgtgcacc	tctagcagag	caggtgaaaa	taataacaca
2521	taacgggagg	gccggccgtt	accaggtcga	cggtatgac	ggcagggtcc	tactaccatg
2581	tygatcgcc	attccggtcc	ctgagttcca	agctttgagc	gagagcgcca	ctatggtgta
2641	caacgaaagg	gagttcgtca	acaggaacct	ataccatatt	gccgttcacg	gaccgtcgct
2701	gaacaccgac	gaggagaact	acgagaaagt	cagagctgaa	agaactgacg	ccgagtacgt
2761	gttcgacgta	gataaaaaat	gctgcgtcaa	gagagaggaa	gcgtcgggtt	tggtgttggt
2821	gggagagcta	accacccccc	cgttccatga	attcgctac	gaagggtcga	agatcaggcc
2881	gtcggcacca	tataagacta	cagtagtagg	agcttttggg	gttcgggat	caggcaagtc
2941	tgctattatt	aagagcctcg	tgacaaaaca	cgatctggtc	accaggggca	agaaaggagaa
3001	ctgccaggaa	atagtcaacg	acgtgaagca	gcaccgggga	ctggacatcc	aggcaaaaaac
3061	agtggactcc	atcctgctaa	acgggtgctg	tctgcccgtg	gacatcctat	atgtggacga
3121	ggctttcgt	tgccattccg	gtactctgt	agccctaatt	gctcttgta	aacctcggag
3181	caaatgggtg	ttatgcggag	acccaagca	atgggattc	ttcaatatga	tgcagcttaa
3241	ggtgaacttc	aaccacaaca	tctgcactga	agtatgtcat	aaaagtatat	ccagacgttg
3301	cacggtcca	gtcacggcca	tctgtctac	attgcactac	ggaggcaaga	tgcgcacgac
3361	caaccctgc	aacaaaccca	taatcataga	caccacagga	cagaccaagc	ccaagccagg
3421	agacatcgtg	ttaacatgct	tccgagctg	ggtaaagcag	ctgcagttgg	actaccgtgg
3481	acacgaagtc	atgacagcag	cagcatctca	gggctcacc	cgcgaagggg	tatacgccgt
3541	aaggcagaag	tgaaatgaaa	atcccttgta	tgcctcgtg	tcggagcacg	tgaatgtact
3601	gctgacgcgc	actgagata	ggctgggtgtg	gaaaacgctg	gccggcgatc	cctggattaa
3661	ggtcctatca	aacattccac	aggtaactt	tacggccaca	ttggaagaat	ggcaagaaga
3721	acacgacaaa	ataatgaagg	tgattgaagg	accgctgcg	cctgtggacg	cgttccagaa
3781	caaaagcaac	gtgtgttggg	cgaaaagcct	ggtgcctgtc	ctggacacgt	cgggaatcag
3841	attgacagca	gaggagtga	gcaccataat	tacagcattt	aaggaggaca	gagcttactc
3901	tccagtgggtg	gccttgaatg	aaatttgac	caagtactat	ggagttgacc	tggacagtgg
3961	cctgttttct	gccccgaagg	tgtccctgta	ttacgagaac	aaccactggg	ataacagacc
4021	tggtggaagg	atgtatggat	tcaatgccgc	aacagctgcc	aggctggaag	ctagacatac
4081	cttcctgaag	gggcagtggc	atacgggcaa	gcaggcagtt	atcgcagaaa	gaaaaatcca
4141	accgcttct	gtctggaca	atgtaattcc	latcaaccgc	aggctgcgc	acgcccctgg
4201	gactgagtac	aagacggbta	aaggcagtag	ggttgagtgg	ctggtcaata	aagtgaagag
4261	gtaccacgtc	ctgctgtgta	gtgagtacaa	ctggcctttg	cctcgacgca	gggtcaacttg
4321	gttgtcacccg	ctgaatgtca	caggcgccga	tagtgtctac	gacctaaagt	taggactgcc
4381	gcggctggaa	ggcaggttcc	acttggtctt	tgtgaacatt	cacacggaat	tcagaatcca
4441	ccactaccag	cagtggtgctg	accacgccat	gaagtgccag	atgcttgggg	gagatgcgt
4501	acgactgcta	aaacccggcg	gcagcctctt	gatgagagct	tacggatacg	ccgataaaat

4561	cagcgaagcc	gttgttttct	ccttaagcag	aaagtctctg	tctgcaagag	tggtgagccc
4621	ggattgtgtc	accagcaata	cagaagtgtt	cttgtgtgtc	tccaactttg	acaaaggaaa
4681	gagacctct	acgtacacc	agatgaatac	caagtgtagt	gccgtgtatg	coggagaagc
4741	catgcacacg	gccgggtgtg	caccatcta	agaggttaag	agagcagaca	tagccacgtg
4801	cacagaagcg	gctgtgttta	acgcagctaa	cggcgttggg	actgtagggg	atggcgtatg
4861	cagggccgtg	gcgaagaaat	ggcgttcagc	ctttaaggga	gaagcaaac	cagtgggcac
4921	aattaaaaca	gtcatgtgcg	gctcgtacc	cgtcatccac	gctgtagcgc	ctaatctctc
4981	tgccacgact	gaagcggaag	gggaccgcga	attggccgct	gtctaccggg	cagtggccgc
5041	cgaaagtaaac	agactgtcac	tgagcagcgt	agccatcccc	ctgctgtcca	caggagtgtt
5101	cagcggcgga	agagataggc	tgagcaatc	cctcaaccat	ctattcacag	caatggacgc
5161	cacggacgct	gacgtgacca	tctactgcag	agacaaaagt	tgggagaaga	aaatccaggga
5221	agccatagac	acgaggacgg	ctgtggagtt	gctcaatgat	gacgtggagc	tgaccacaga
5281	cttggtgaga	gtgcaccggg	acagcagcct	ggtgggtcgt	aagggtctaca	gtaccactga
5341	cgggtcgctg	tactgttact	ttgaaggtac	gaaattcaac	caggctgcta	ttgatatggc
5401	agagatactg	acgttgtggc	ccagactgca	agaggcaaac	gaacagatat	gcctatacgc
5461	gctgggcgaa	acaatggaca	acatcagatc	caaatgtccg	gtgaacgatt	ccgatttcac
5521	aacacctccc	aggacagtgc	cctgcctgtg	ccgtacgcga	atgacagcag	aacggatcac
5581	ccgccttagg	tcacaccaag	ttaaaagcat	ggtggtttgc	tcattctttc	ccctcccga
5641	ataccatgta	gatgggtgtc	agaaggtaaa	gtggagaaag	gttctcctgt	tcgacccgac
5701	ggtaccttca	gtggttagtc	cgcggaaagta	tgcgcgatct	acgacggacc	actcagatcg
5761	gtcgtttacga	gggttttgact	tggactggac	cacgcactcg	ctttccactg	ccagcgatcac
5821	catgtcgcta	cccagtttgc	agtcgtgtga	catcgactcg	atctacagac	caatggctcc
5881	catagtagtg	acggctgacg	tacacctga	acccgcagcc	atccgggacc	tggcgggcaga
5941	tgtgcatcct	gaaccgcag	accatgtgga	cctcgagaac	ccaattcttc	caccacgccc
6001	aaagagagct	gcataccttg	cctccgcgcg	tgcggagcga	ccggtgcctg	cgcctagaaa
6061	gcctacgcct	gcccaaggga	ctgcgttttag	gaacaagctg	cttttgacgt	tcggcgactt
6121	tgacgagcac	gaggtcgatg	cgttggcctc	cgggattact	ttcggagact	tcgacgacgt
6181	cctgcgacta	ggcgcgcggg	gtgcataatat	tttctcctcg	gacactggca	ggggacattt
6241	acaacaaaaa	tcctgttaggc	agcacaatct	ccagtgcgca	caactgggatg	cggctcgaggga
6301	ggagaaaaatg	taccgcgcaa	aattggatac	tgagaggggag	aagctgttgc	tgctgaaaaat
6361	gcagatgcac	ccatcgaggg	ctaataagag	tcgataccag	ctcgcgaaaag	tggagaacat
6421	gaaagccacg	gtggttgaca	ggctcacatc	ggggggccaga	ttgtacacgg	gagcggacgt
6481	aggccgcata	ccaacatacg	cggttcggta	ccccgcgcc	gtgtactccc	ctaccgtgat
6541	cgaaagattc	tcaagccccg	atgtagcaat	cgcagcgtgc	aacgaataacc	tatccagaaa
6601	ttacccaaca	gtggcgtcgt	accagataac	agatgaatac	gacgcatact	tggacatgggt
6661	tgacgggtcg	gatatgttgt	tggacagagc	gacatctctg	ccggcgaaagc	tcgggtgcta
6721	ccgaaacat	catgcgtacc	accagccgac	tgtacgcagt	gccgtcccg	cacctttca
6781	gaacacacta	cagagcgtgc	tagcggccgc	caccaagaga	aactgcaacg	tcacgcaaat
6841	gcgagaacta	cccacatgg	actcggcagt	gttcaacgtg	gagtgccttca	agcgtatgc
6901	ctgctccgga	gaatatggg	aagaatatgc	taacaacct	atccggataa	ccactgagaa
6961	catcactacc	tatgtgacca	aattgaaaag	cccgaaagct	gctgccttgt	tcgctaagac

7021	ccacaacttg	gttcgcctgc	aggaggttcc	catggacaga	ttcacggctc	acatgaaacg
7081	agatgtcaaa	gtcactccag	ggacgaaca	cacagaggaa	agacccaaag	tcaggtaaat
7141	tcaagcagcg	gagccattgg	cgaccgtta	cctgtgcggc	atccacaggg	aattagtaag
7201	gagactaaat	gctgtgttac	gccttaacgt	gcacacattg	tttgatatgt	cggccgaaga
7261	ctttgacgcg	atcatgcctt	ctcactttcc	ccaggagagc	ccggttctag	agacggacat
7321	tgcatcattc	gacaaaagcc	agacgactc	cttggctctt	acaggtttta	tgatcctcga
7381	agatctagg	gtggatcagt	acctgctgga	cttgatcgag	gcagcctttg	gggaaatata
7441	cagctgtcac	ctaccaactg	gcacgcgtt	caagtctgga	gctatgatga	aatcggggcat
7501	gtttctgact	ttgtttatta	acactgtttt	gaacatoccc	atagcaagca	gggtactgga
7561	gcagagactc	actgactccg	cctgtgcggc	cttcactggc	gacgacaaca	tcgttcacgg
7621	agtgatctcc	gacaaagctga	tggcgagag	gtgcgcgtcg	tgggtcaaca	tggaggtgaa
7681	gatacttgac	gctgtcatgg	gcgaaaaacc	cccatatttt	tgtgggggat	tcatagtttt
7741	tgacagcgtc	acacagaccg	cctgccgtgt	ttcagaccca	cttaagcgcc	tgttcaagtt
7801	gggtaagccg	ctaacaagctg	aagacaagca	ggacgaagac	aggcgacgag	cactgagtga
7861	cgaggttagc	aagtgggtcc	ggacaggctt	ggggcccgaa	ctggaggtgg	cactaacatc
7921	taggtataag	gtagagggct	gcaaaagtat	ctcctatagcc	atggccacct	tggcgagggg
7981	cattaaggcg	tttaagaaat	tgagaggacc	tgttatatac	ctctacggcg	gtcctagatt
8041	ggtgcgttaa	tacacagaat	tctgattata	gcgcactatt	atagcaccat	ggtgagcaag
8101	ggcgaggagc	tgttcacgg	ggtggtgccc	atcctggctc	agctggacgg	cgacgtaaac
8161	ggccacaagt	tcagcgtgtc	cggcgagggc	gagggcgatg	ccacctacgg	caagctgacc
8221	ctgaagttca	tctgcaccac	cggaagctg	cccgctccct	ggccccacct	cgtgaccacc
8281	ctgacctacg	gcgtgcagtg	cttcagccgc	taccocgacc	acatgaagca	gcacgacttc
8341	ttcaagtccg	ccatgcccg	aggctacgtc	caggagcgca	ccatcttctt	caaggaacgac
8401	ggcaactaca	agaccgcgc	cgaggtgaag	ttcgagggcg	acacctgggt	gaaccgcac
8461	gagctgaagg	gcctcgactt	caaggaggac	ggcaacatcc	tggggcaca	gctggagtac
8521	aactacaaca	gccacaacgt	ctatatcatg	gcgacaacgg	agaagaacgg	catcaagggtg
8581	aacttcaaga	tccgccacaa	catcgaggac	ggcagcgtgc	agctcgccga	ccactaccag
8641	cagaacaccc	ccatcgcgga	cgccccctg	ctgctgccc	acaaccacta	cctgagcacc
8701	cagtcgcgcc	tgagcaaa	ccccaacgag	aagcgcgac	acatggtcct	gctggagttc
8761	gtgaccgccg	ccgggatcac	tctcggcatg	gacgagctgt	acaagtcagg	atccggagcc
8821	acgaactttc	ctctgttaaa	gcaagcagga	gacgtggaag	aaaaccccg	gcccatggaa
8881	gatgccaaaa	acattaaaga	gggccacg	ccattctacc	cactcgaaga	cgggaccggc
8941	ggcgagcagc	tgacaaaagc	catgaagcgc	tacgccccgg	tgccccggc	catcgccctt
9001	accgacgcac	atctcgaggt	ggacattacc	tacgcgagat	acttcgagat	gagcgttcgg
9061	ctggcagaag	ctatgaagcg	clatgggctg	aatacaaac	atcggtatct	gggtgtgcagc
9121	gagaatagct	tgcagttctt	catgccccgt	ttggtgtccc	tggtcatcgg	tgtggctgtg
9181	gccccagcta	acgacatcta	caacgagcgc	gagctgtgta	acagcatggg	catcagccag
9241	cccaccgtcg	tattcgtgag	caagaaaggg	ctgcaaaaga	tcctcaacgt	gcaaaaagaag
9301	ctaccgatca	tacaaaagat	catcatcatg	gatacgaaga	ccgactacca	gggcttccaa
9361	agcatgtaca	ccttcgtgac	ttccatttg	ccaccggct	tcaacgagta	cgacttcgtg
9421	cccgagagct	tcgaccggga	caaaaccatc	ggcctgatca	tgaacagtag	tggcagttacc

9481	ggattgcccc	aggcgtagc	cctaccgcac	cgacccgctt	gtgtccgatt	cagtcattgcc
9541	cgcgaccccc	tcttcggcaa	ccagatcatc	cccgacaccg	ctatcctcag	cgtggtgccca
9601	tttcaccacg	gcttcggcat	gttcaccacg	ctgggctact	tgatctcgcg	ctttcgggtc
9661	gtgctcatgt	accgttcga	ggaggagcta	ttcttgcgca	gcttgcaaga	ctataagatt
9721	caatctgccc	tgctgtgccc	cacactatct	gaccttctcg	ctaaagacac	tctcatcgac
9781	aagtacgacc	taagcaactt	gcacgagatc	gcacgcgccg	ggcgccgct	cagcaaggag
9841	gtaggtgagg	ccgtggccaa	acgcttcac	ctaccaggca	tccgccagg	ctacggcctg
9901	acagaaacaa	ccagcgccat	tctgataccc	cccgaggggg	acgacaagcc	tggcgcagta
9961	ggcaaggtgg	tgccctctct	cgaggctaa	gtgtgggact	tggaacccg	taagacactg
10021	ggtgtgaacc	agcgcgccga	gctgtgcgtc	cgtggcccca	tgatcatgag	cggctacgtt
10081	aaaaaccccc	aggctacaaa	cgtctctatc	gacaaggacg	gctggctgca	cagcggcgac
10141	atcgctact	gggacgagga	cgagcacttc	ttcatcgtgg	accggctgaa	gagcctgata
10201	aaatacaagg	gctaccaggt	agccccagcc	gaactggaga	gcatcctgct	gcaacacccc
10261	aacatcttcg	acgcgggggt	cgcggcctg	cccgacgacg	atgcggcgga	gctgcccgc
10321	gcagtcgtcg	tgctggaaca	cggtaaaacc	atgaaccgaga	aggagatcgt	ggactatgtg
10381	gccagccagg	ttacaaccgc	caagaagctg	cgcggtggtg	ttgtgttcgt	ggacgaggtg
10441	cctaaaggac	tgaccggcaa	gttgagccg	cgcgaagatcc	cgagattct	cattaaggcc
10501	aagaaggcgg	gcaagatcgc	cgtgtcgact	gtttaagtta	gggtaggcaa	tggtcattgat
10561	atagcaagaa	aattgaaaac	agaaaagtt	agggtaaagca	atggcatata	accataactg
10621	tataacttgt	aacaaagcgc	aacaaagacct	gogcaattgg	ccccgtggtc	cgctcacgg
10681	aaactcgggg	caactcatat	tgacacatta	atlggcaata	atlggaagct	tacataagct
10741	taattcgacg	aataattgga	tttttatctt	atlttgcaat	tggtttttta	tatttccaaa
10801	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa
10861	aaaaaaaaacta	gtctgcatta	atgaatcggc	tcgagtctag	agggcccggt	taaacccgct
10921	gacagcctc	gactgtgcoo	tctagtggcc	agccatctgt	tgtttgccc	tccccgtgc
10981	cttccttgac	cctggaagg	gccactccca	ctgtcctttc	ctaataaaat	gaggaaattg
11041	catcgcatcg	tctgagtagg	tgctattcta	ttctgggggg	tggggtgggg	caggacagca
11101	agggggagga	ttgggaagac	aatagcaggc	atgctgggga	tgccgtgggc	tctatggcct
11161	ctgaggcgga	aagaaccagc	tggggctcta	gggggtatcc	ccacgcgcc	tgtagcggcg
11221	cattaagcgc	ggcgggtgtg	gtggttacgc	gcaggtgac	cgctacactt	gccagcgccc
11281	tagcgccgc	tcctttcgct	ttcttccttc	ctttctcgc	cagcttcgcc	ggctttcccc
11341	gtcaagctct	aaatcggggg	ctccctttag	ggttccgatt	tagtgcttta	cggcacctcg
11401	accccaaaa	acttgattag	ggtgatggtt	cacgtagtgg	gccatcgccc	tgatagacgg
11461	tttttcgccc	tttgacgttg	gagtcacagt	tctttaatag	tggactcttg	ttccaaactg
11521	gaacaacact	caaccctatc	tcggtctalt	cttttgattt	alaaggat	ttgcccattt
11581	cggcctattg	gttaaaaaat	gagctgattt	aacaaaaatt	taacgcgaat	taattctgtg
11641	gaatgtgtgt	cagtttaggg	gtggaaagtc	ccagggctcc	ccagcagca	gaagtatgca
11701	aagcatgcat	ctcaattagt	cagcaaccag	gtgtggaag	tccccaggct	ccccagcagg
11761	cagaagtatg	caaagcatgc	atctcaatta	gtcagcaacc	atagtcocgc	ccctaactcc
11821	gcccccccg	cccctaactc	cgccagttc	cgcctattct	ccgccccatg	gctgactaat
11881	tttttttatt	tatgcagagg	ccgaggccgc	ctctgcctct	gagctattcc	agaagtatgtg

11941	aggaggcctt	tttgaggcc	taggctttt	caaaaagctc	ccggagcctt	gtatatccat
12001	tttcggatct	gatcaagaga	caggatgagg	atcgtttcgc	atgattgaac	aagatggatt
12061	gcacgcaggt	tctcggccg	cttggttga	gaggtatttc	ggctatgact	gggcacaaca
12121	gacaatcggc	tgcctgagt	ccgcctgtt	ccggtgtca	gcgaggggc	gcccgttct
12181	ttttgtcaag	accgactgt	ccggtgcct	cagtgaactg	caggacagg	cagcgggct
12241	atcgtggctg	gccacgacgg	cggttccct	cgcagctgtg	ctcgacgttg	tcactgaagc
12301	gggaaggac	tggctgctat	tggcggaagt	gccggggcag	gatctcctgt	catctcacct
12361	tgctcctgcc	gagaaagtat	ccatcatggc	tgatgcaatg	cggcggtcgc	atacgttga
12421	tcggctacc	tgccattcg	accaccaagc	gaaacatcgc	atcgagcgag	cacgtactcg
12481	gatggaagcc	ggtcttgtcg	atcaggatga	tctggacgaa	gagcatcagg	ggctcgcgcc
12541	agccgaactg	ttcgccaggc	tcaaggcgcg	catgcccgac	ggcgaggatc	tcgtcgtgac
12601	ccatggcgat	gcctgcttgc	cgaatatcat	ggtggaaaaat	ggccgctttt	ctggattcat
12661	cgactgtggc	cggctgggtg	tggcggaccg	ctatcaggac	atagcgttgg	ctacccgtga
12721	tattgctgaa	gagcttggcg	gcgaatggcg	tgaccgcttc	ctcgtgcttt	acggtatcgc
12781	cgctcccgat	tgcgacgca	tgccttcta	tgcccttctt	gacgagttct	tctgagcggg
12841	actctggggt	tcgaaatgac	cgaccaagcg	acgcccaccc	tgccatcacg	agattttcgat
12901	tcacccgccg	ccttctatga	aaggttggcg	ttcggaatcg	ttttccggga	cgcggcgtgg
12961	atgacctcc	agcgcgggga	tctcatgctg	gagttcttcg	cccaccccaa	cttgtttatt
13021	gcagcttata	atggttacaa	ataaagcaat	agcatcacaa	atttcacaaa	taaagcattt
13081	ttttcactgc	attctagtgt	tggtttgttc	aaactaatca	atgtatatcta	toatgtctgt
13141	ataccgtcga	cctctagcta	gagcttggcg	taatcatggt	catagctgtt	tccgtgtlga
13201	aattgttatt	cgtccacaat	tcacacaaac	atacagagccg	gaagcataaa	gtgtaaagcc
13261	tggggtgcct	aatgagttag	ctaactcaca	ttaattgctg	tgcgctcaat	gcccgttttc
13321	cagtcgggaa	acctgtcgtg	ccagctgcat	taatgaatcg	gccaacgcgc	ggggagagggc
13381	ggtttgcgta	tgggcgctc	ttcgccttc	tcgctcaactg	actcgtcgc	ctcggctcgtt
13441	cggctgcggc	gagcgttata	agctcactca	aagcggttaa	tacggtttatc	cacagaatca
13501	ggggataaacg	caggaaaagaa	catgtgagca	aaagcccgag	aaaagcccgag	gaaccgtataa
13561	aaggccgcgt	tgcctggcgtt	tttccatagg	ctcgcgcgc	ctgacgagca	tcacaaaaat
13621	cgacgctcaa	gtcagaggtg	gcgaaacccg	acagacttat	aaagatacca	ggcgtttccc
13681	cctggaagct	ccctcgtgcg	ctctcctgtt	ccgacctgcg	cgtttaccgg	atacctgtcc
13741	gcctttctcc	cttcgggaaag	cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt
13801	tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aacccccgt	tcagcccgac
13861	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtcacaac	cggtaagaca	cgacttatcg
13921	ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggtatgtagg	cgggtgctaca
13981	gagtlcctga	agtgttgccc	taactacggc	tacactagaa	gaacagtatt	tggtatctgc
14041	gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	cggcaaaacaa
14101	accaccgctg	gtagcgggtt	ttttgtttgc	aagcagcaga	tlacgcgcag	aaaaaaagga
14161	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaaactca
14221	cgttaaggga	ttttggtcat	gagattatca	aaaaggatct	tcacctagat	ccttttaaat
14281	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	tgacagttac
14341	caatgccttaa	tcagtgaggc	acctatctca	ggatctgttc	tatttcgttc	atccatagtt

		14401 gcctgactcc ccgtcgtgta gataactacg ataccgggagg gcttaccatc tggcccccagt 14461 gctgcaatga tacccgcgaga cccacgctca ccggtctccag atttatcagc aataaaaccag 14521 ccagccggaa ggccgcgagcg cagaagtggc cctgcaactt tatccgcctc catccagtct 14581 attaatgtgt gccgggaagc tagagtaagt agtccgcccag ttaatagttt gcgcaacggt 14641 gttgccattg ctacaggcat cgtggtgtca cgtcgtcgtg ttggtatggc ttcattcagc 14701 tccggttccc aacgatcaag gcgagttaca tgcattcctga cgcagtggtt atcactcatg 14761 agtcctctcg gtcctccgat cactgcataa ttctcttact gtcattgccc cgttaagatg cttttctgtg 14821 gttatggcag cactgcataa gtcattctga gaatagtgtg tgcggcgacc gagttgctct 14881 actggtgagt actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct 14941 tgcgcggcgt caatacggga taataccgcy ccacatagca gaactttaaa agtgcctcatc 15001 attgaaaaac gttcttcggg gcgaaaaactc tcaaggatct taccgctgtt gagatccagt 15061 tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt 15121 tctgggtgag caaaaacagg aaggcaaaat gcgcacaaa agggaataag ggcgacacgg 15181 aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta tcagggttat 15241 tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccc 15301 cgcacatttc ccgaaaaagt gccacctgac gtcgacggat cgggagatct cccgatcccc 15361 tatggtgcac tctcagtaca atctgctctg atgcgcgata gttaaagccag tatctgctcc 15421 ctgcttgtgt gttggaggtc gctgagtagt gcgcgagcaa aatttaagct acaacaaggc 15481 aaggcttgac cgacaattgc atgaagaatc tgcttagggt taggcgtttt gcgctgcttc 15541 gcgatgtacy ggccagatat acgcgtt
enhanced GFP	22	MVSKGEELFTGVVPIILVELDGVNKGKFSVSGEGEGDATYKGLTLKFTCTTGKLPVPWPTLVTLTYGVCFSRYPDHMKQHDFE KSAMPEGYVQERTIFFKDDGNYKTRAEVKEFEGDTIVNRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQKNIGIKVNFKIRH NIEDGSVQLADHYQNTPIGDGPVLLPDNHYLSTQALSKEPNKRDHMLLEFVTAAGITIGMDELYK

synthetic luc2 version of the luciferase	23	MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVRLAEAMKRYGLNTNHRIVVCSSENS LQFFMPVLGALFIGVAVAPANDIYNRELINSMGISOPTVVFVSKKGLQKILNVQKLPILIQKILIMDSKTDYQGFQSMYTFVTS HLPFGGENYDFVPESFDRDKTIALINMSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVFPFHGFGMFTTLGY LICGFRVLMYRFEELFLRSLQDYKIQSALLVPTLFSSFFAKSTLLDKYDLNLSHEIASGGAPLSKEVGEAVAKRFLHPGIRQGY GLTETTSAILITPEGDDKPGAVGVVPEAKVVDLDTGKTLGVNQRGELCVRGPMIMSGYNNPEATNALIDKDGWLHSGDIAIY WDEDEHFFIVDRKSLIKYKGQYVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAAVVVLEHGKMTTEKEIVDYVASQVTTAK KLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAVSTV
pHCMV-G	24	1 gagcttgccc cattgcatac gttgtatcca tatcataata tgtacattta tattggctca 61 tgtccaacat taccgccatg ttgacattga ttattgacta gttatttaata gtaatcaatt 121 acgggggtcat tagttcatag cccatataatg gagttccgag ttacataact tacggtaaat 181 gggccgcctg gctgaccgcc caacgacccc gcccattga cgtcaataat gacgtatggt 241 cccatagtaa cgccaatagg gactttccat tgacgtcaat ggtggagta ttacgggtaa 301 actgcccact tggcagtaca tcaagtgtat ccatgcccac gtacgcccc tattgacgtc 361 aatgacggta aatggccgcg ctggcattat gccagtlaca tgaccttat ggactttcct 421 acttggcagt acatctacgt attagtcact gctattacca tgggtgatcg gttttggcag 481 tacatcaatg ggcgtggata gcggtttgac tcacggggat ttccaagtct ccaccccatt 541 gacgtcaatg ggagtttgtt ttggcaccaa aatcaacggg actttccaaa atgtogtaac 601 aactccgcc ccatgacgca aatgggcggt aggcgtgtac ggtgggaggt ctatataagc 661 agagctcgt tagtgaaccg tcagatcgcc tggagacgcc atccacgtg ttttgacctc 721 catagaagac accgggaccg atccagcctc cggtcgacgg atcctgagaa ctccaggggtg 781 agtttgggga ccttgatattg ttctttcttt ttgcctattg taaaattcat gttatatgga 841 gggggcaaa ttttcagggt gttgtttaga atgggaagat gtcccttgta tcaccatgga 901 cctcatgat aattttgttt ctttcacttt ctactctgtt gacaaccatt gtctcctctt 961 attttcttt cattttctgt aacttttctg ttaaaacttta gcttgcat tttaacgaatt 1021 ttlaaatcca cttttgttta ttgtcagat tgtaaglacl ttcttaata actttttttt 1081 caaggcaatc aggtatatt atattgtact tcagcacagt tttagagaa aattgttata 1141 attaaatgat aaggtagaat atttctgcat ataaattctg gctggcgtgg aaatatctt 1201 attggtagaa acaactacac cctggtcac atcctgcctt tctctttatg gttacaatga 1261 tatacactgt ttgagatgag gataaaatac tctgagtcca aaccggggcc ctctgctaac 1321 catgttcacg ccttctctc ttctctacag ctctgggca acgtgctggt tgttgtgctg 1381 tctcatcatt ttggcaaaaga attcctcgac ggaatcctcg aggaattctg acactatgaa 1441 gtgccttttg tacttagcct ttttattcat tgggtggaat tgcaagttca ccatagtttt 1501 tccacacaac caaaaaggaa actggaaaaa tgttctctct aattaccatt attgcccgtc 1561 aagctcagat ttaaatggc ataatgactt aataggcaca gccttacaag tcaaaatgcc 1621 caagagtcac aaggctattc aagcagacgg ttggtatgtt catgcttcca aatgggtcac 1681 tacttgtgat ttccgttggt atggaccgaa gtatataaca cattccatcc gatccctcac 1741 tccatctgta gaacaatgca aggaagcat tgaacaaaga aaacaaggaa cttggctgaa 1801 tccaggcttc cctcctcaaa gttgtggata tgcaactgtg acggatgccg aagcagtgat 1861 tgtccagggt actcctcacc atgtgctggt tgaataac acaggagaat ggttgattc 1921 acagttcatc aacggaaaat gcagcaatta catatgccc actgtccata actctacaac

1981	ctggcattct	gactataaag	tcaaagggct	atgtgattct	aacctcattt	ccatggacat
2041	caccttcttc	tcagaggacg	gagagctatc	atccctggga	aaggagggca	cagggttcag
2101	aagtaactac	tttgcttatg	aaactggagg	caagcctgc	aaaatgcaat	actgcaagca
2161	ttggggagtc	agactcccat	caggtgtctg	gttcagatg	gctgataaag	atctctttgc
2221	tgagccaga	ttccctgaat	gcccagaag	gtcaagatc	tctgctccat	ctcagacctc
2281	agtggatgta	agtctaattc	aggacgttga	gagagatctg	gattattccc	tctgccaaga
2341	aacctggagc	aaaatcagag	cgggtcttcc	aatctctcca	gtggatctca	gctatcttgc
2401	tcctaaaaac	ccagggaaccg	gtcctgcttt	caccataatc	aatggtaccc	taaaatactt
2461	tgagaccaga	tacatcagag	togatattgc	tgtccaatc	ctctcaagaa	tggtcggaat
2521	gacagtgga	actaccacag	aaagggaact	gtgggatgac	tgggcaccat	atgaagacgt
2581	ggaatttga	cccaatggag	ttctgaggac	cagttcagga	tataagtttc	ctttatacat
2641	gattggacat	ggtatgttgg	actccgatct	tcatcttagc	tcaaaggctc	aggtgttcga
2701	acatcctcac	attcaagacg	ctgcttcgca	acttctgat	gatgagagtt	tattttttgg
2761	tgatactggg	ctatccaaaa	atccaatcga	gcttgtagaa	ggttggttca	gtagttggaa
2821	aagctctatt	gcctcttttt	tccttatcat	agggttaact	atggactat	tcctggttct
2881	ccgagtttgt	atccatcttt	gcattaaatt	aaagcacacc	aagaaaagac	agattttatc
2941	agacatagag	atgaacogac	ttggaagta	actcaaatcc	tgcaacaacg	attcttcatg
3001	tttggaccac	atcaacttgt	gataccatgc	tcaagagggc	ctcaattata	tttgagtttt
3061	taatttttat	gaaaaaaaa	aaaaaaaacg	gaattctctg	agggatccgt	cgagggaattc
3121	actctcaggg	tgcaggctgc	ctatcagaag	gtgtgggctg	gtgtggccaa	tgccttggct
3181	cacaaatacc	actgagatct	ttttccctct	gccaaaaaatt	atggggacat	catgaagccc
3241	cttgagcatc	tgactttctg	ctaataaagg	aaatttattt	tcattgcaat	agtggtgttg
3301	aattttttgt	gtctctcact	cggaaaggaca	tatggccata	caaatcattt	aaaacatcag
3361	aatgagtatt	tggtttagag	tttggcaaca	tatgcccata	tgctggctgc	catgaacaaa
3421	ggttggtctat	aaagaggtca	tcagtatatg	aaacagcccc	ctgctgtcca	ttccttattc
3481	catagaaaag	ccttgacttg	aggttagatt	ttttttatat	tttgttttgt	gttattttttt
3541	tctttaacat	ccctaaaatt	ttccttacat	gttttactag	ccagattttt	cctcctctcc
3601	tgactactcc	cagtcatagc	tgtccctctt	ctcttatgga	gatccctcga	cggatcggcc
3661	gcaattcgtg	atcatgtcat	agctgtttcc	tgttgaaat	tgttatccgc	tcacaattcc
3721	acacaacata	cgaagccgga	gcataaaagt	taaagccttg	ggtgcctaatt	gagtgaagta
3781	actcacatta	attgctgtgc	gctcactgcc	cgtttccag	tcgggaaacc	tgtcgtgcca
3841	gctgcattaa	tgaatcggcc	aacgcgcggg	gagagggcgt	ttgcgtattg	ggcgtctctc
3901	cgtctcctcg	ctcactgact	cgtgcgcctc	ggtcgttcgg	ctgcggcgag	cgggtatcagc
3961	tcactcaaa	gcggtaatac	ggttatccac	agaaatcagg	gataacgcag	gaaagaacat
4021	gtgagcaaaa	ggccagcaaa	aggccaggaa	cgttaaaaa	gccgcgttgc	tggcgttttt
4081	ccataggctc	ggccccctcg	acgagcatca	caaaaatcga	cgtcgaagtc	agaggtggcg
4141	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	tcgtgcgctc
4201	tcctgttccg	accctgccgc	ttaccggata	cctgtccccc	tttctccctt	cgggaaagcgt
4261	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtagggtg	ttcgtctcaa
4321	gctgggctgt	gtgcacgaac	cccccgctca	gccgcaccgc	tgcgcttat	cgggtaacta
4381	tcgtcttgag	tccaaaccgg	taagacacga	cttatcgcca	ctggcagcag	ccactggtaa

4441	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	ggtggcctaa
4501	ctacggctac	actagaagaa	cagtatttgg	tatctgcgct	ctgctgaagc	cagttacctt
4561	cggaaaaaga	gttggtagct	ttgatccgg	caacaaacc	accgctggtg	goggtgggtt
4621	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaggga	tctcaagaag	atcctttgat
4681	cttttctacg	gggtctgacg	ctcagtgtaa	cgaaactca	cgttaaggga	ttttggtcat
4741	gagattatca	aaaagatct	tcacctagat	cgttttaaat	taaaaatgaa	gttttaaatc
4801	aatctaaagt	atatatgagt	aaacttggtc	tgacagttac	caatgcttaa	tcagtgaggc
4861	acctatctca	gcgatctgtc	tatttcgttc	atccatagtt	gcctgactcc	cgtctcgtga
4921	gataactacg	atacgggagg	gcttaccatc	tggcccaagt	gctgcaatga	tacgcgaga
4981	cccacgctca	cggctccag	atttatcagc	aataaaccag	ccagccggaa	gggcccagcg
5041	cagaagtggc	cctgcacatt	tatccgcctc	caccagctct	attaattggt	gccggggaagc
5101	tagagtaagt	agttcgccag	ttaatatggt	gcgcaacgtt	gttgccattg	ctacaggcat
5161	cgtggtgtca	cgctcgtcgt	ttggtatggc	ttcattcagc	tccgggtccc	aacgatcaag
5221	gcgagttaca	tgatccccc	tggtgtgcaa	aaaagcgggt	agtcctctcg	gtcctccgat
5281	cgttgtcaga	agtaagtbtg	ccgcagtggt	atcactcatg	gttatggcag	cactgcataa
5341	ttctcttact	gtcatgccat	ccgtaagatg	ctttctctgt	actggtgagt	actcaaccaa
5401	gtcattctga	gaatagtgtg	tgcggcgacc	gagttgctct	tgcccgcgct	caatacggga
5461	taataccgcg	ccacatagca	gaacttttaa	agtctcatc	attggaaaac	gttcttcggg
5521	gcgaaaactc	tcaaggatct	taccgctgtt	gagatccagt	tcgatgtaac	ccactcgtgc
5581	acccaaactga	tcttcagcat	cttttacttt	caccagcgtt	tctgggtgag	caaaaacaggg
5641	aaggcaaaat	gccgcaaaaa	agggaaataag	ggcgacacgg	aaatgttgaa	tactcatact
5701	cttccttttt	caatattatt	gaagcattta	tcagggttat	tgtctcatga	gcggatacat
5761	atttgaatgt	atttagaaaa	ataaacaaat	aggggttccg	cgcacatttc	cccgaaaaagt
5821	gccacctaata	ttgtaagcgt	taatatattg	ttaaaattcg	cgttaaat	ttgttaaatc
5881	agctcatttt	ttaaccaata	ggcgaatac	ggcaaaatcc	ctataaatc	aaaagaatag
5941	accgagatag	ggttgagtgt	tggtccagtt	tggaacaaga	gtccactatt	aaagaacgtg
6001	gactccaaacg	tcaaaaggcg	aaaaaccgtc	tatcaggggc	atggccact	acgtgaacca
6061	tcaccctaata	caagtttttt	gggtctgagg	tgcgtaaaag	cactaaatcg	gaacctaaaa
6121	gggagccccc	gatttagagc	ttgacgggga	aagccggcga	acgtggcgag	aaaggaaagg
6181	aagaaagcga	aaggagcggg	cgttagggcg	ctggcaagt	tagcggtcac	gctgcgcgta
6241	accaccacac	cgcgcgcgct	taatgcgcg	ctacagggcg	cgtccattc	gccattcagg
6301	ctgcgcaact	gttggaagg	gcgatacgtg	cgggcctctt	cgtatttac	ccagctggcg
6361	aaagggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	cagggttttc	ccagtcacga
6421	cgttgtaaaa	cgacggccag	tgagcgcgcg	taatacgact	cactatagg	cgaattggag
6481	ctccaccgcg	gtggcgggcg	ctctaga			

REFERENCES

1. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
2. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
3. Hendel, S. J. & Shoulders, M. D. Directed evolution in mammalian cells. *Nat. Methods* **18**, 346–357 (2021).
4. Eigen, M. & Gardiner, W. Evolutionary molecular engineering based on RNA replication. *J. Macromol. Sci. Part A Pure Appl. Chem.* **56**, 967–978 (1984).
5. Liao, H., McKenzie, T. & Hageman, R. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 576–580 (1986).
6. Chen, K. Q. & Arnold, F. H. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnology* **9**, 1073–1077 (1991).
7. Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
8. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
9. Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
10. Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831–843 (2003).
11. Park, H. & Kim, S. Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.* **49**, e32 (2021).
12. Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
13. Molina, R. S. *et al.* In vivo hypermutation and continuous evolution. *Nat Rev Methods Primers* **2**, (2022).
14. Das, A. T. *et al.* Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* **279**, 18776–18782 (2004).

15. Berman, C. M. *et al.* An Adaptable Platform for Directed Evolution in Human Cells. *J. Am. Chem. Soc.* **140**, 18093–18103 (2018).
16. Jewel, D. *et al.* Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells. *Nat. Methods* **20**, 95–103 (2023).
- 5 17. Klenk, C. *et al.* A Vaccinia-based system for directed evolution of GPCRs in mammalian cells. *Nat. Commun.* **14**, 1770 (2023).
18. English, J. G. *et al.* VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell* **178**, 748–761.e17 (2019).
19. Denes, C. E. *et al.* The VEGAS Platform Is Unsuitable for Mammalian Directed Evolution.
10 *ACS Synth. Biol.* **11**, 3544–3549 (2022).
20. Kim, D. Y., Firth, A. E., Atasheva, S., Frolova, E. I. & Frolov, I. Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *J. Virol.* **85**, 8022–8036 (2011).
21. Brown, R. S., Anastasakis, D. G., Hafner, M. & Kielian, M. Multiple capsid protein binding
15 sites mediate selective packaging of the alphavirus genomic RNA. *Nat. Commun.* **11**, 4693 (2020).
22. Ruiz-Guillen, M. *et al.* Capsid-deficient alphaviruses generate propagative infectious microvesicles at the plasma membrane. *Cell. Mol. Life Sci.* **73**, 3897–3916 (2016).
23. Rolls, M. M., Webster, P., Balba, N. H. & Rose, J. K. Novel infectious particles generated by
20 expression of the vesicular stomatitis virus glycoprotein from a self-replicating RNA. *Cell* **79**, 497–506 (1994).
24. Rose, N. F. *et al.* In vitro evolution of high-titer, virus-like vesicles containing a single structural protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16866–16871 (2014).
25. Akhrymuk, I., Lukash, T., Frolov, I. & Frolova, E. I. Novel Mutations in nsP2 Abolish Chikungunya Virus-Induced Transcriptional Shutoff and Make the Virus Less Cytopathic without Affecting Its Replication Rates. *J. Virol.* **93**, (2019).
26. Belshaw, R., Pybus, O. G. & Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* **17**, 1496–1504 (2007).
27. Patterson, E. I. *et al.* Measuring Alphavirus Fidelity Using Non-Infectious Virus Particles.
30 *Viruses* **12**, (2020).
28. O'Hara, P. J., Nichol, S. T., Horodyski, F. M. & Holland, J. J. Vesicular stomatitis virus

defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell* **36**, 915–924 (1984).

29. Drewett, V. *et al.* Serum response factor cleavage by caspases 3 and 7 linked to apoptosis in human BJAB cells. *J. Biol. Chem.* **276**, 33444–33451 (2001).

5 30. Hecht, B., Müller, G. & Hillen, W. Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J. Bacteriol.* **175**, 1206–1210 (1993).

31. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* **10**, 81 (2010).

10 32. Müller, G. *et al.* Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nat. Struct. Biol.* **2**, 693–703 (1995).

33. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

34. Zhou, X., Vink, M., Klaver, B., Berkhout, B. & Das, A. T. Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther.* **13**, 1382–1390 (2006).

15 35. Dingus, J. G., Tang, J. C. Y., Amamoto, R., Wallick, G. K. & Cepko, C. L. A general approach for stabilizing nanobodies for intracellular expression. *Elife* **11**, (2022).

36. Bethuyn, J. *et al.* A nanobody modulates the p53 transcriptional program without perturbing its functional architecture. *Nucleic Acids Res.* **42**, 12928–12938 (2014).

20 37. Vignuzzi, M. & López, C. B. Defective viral genomes are key drivers of the virus-host interaction. *Nat Microbiol* **4**, 1075–1087 (2019).

38. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).

39. Tian, R. *et al.* Establishing a synthetic orthogonal replication system enables accelerated evolution in. *Science* **383**, 421–426 (2024).

25

Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. It will therefore be appreciated by those of skill in the art that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present invention.

30

All computer programs, algorithms, patent and scientific literature referred to herein is incorporated herein by reference.

Any reference to publications cited in this specification is not an admission that the disclosures constitute common general knowledge in Australia.

CLAIMS

1. A directed evolution system, comprising:
- (i) a first nucleic acid sequence encoding one or more non-structural genes of a virus
5 and at least one evolution candidate,
 - (ii) a second nucleic acid sequence encoding one or more packaging genes, and
 - (iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

wherein the first nucleic acid sequence and the second nucleic acid sequence generates
10 a first virus-like particle (VLP) after expression in a host cell, and wherein said first VLP and the third nucleic acid sequence generates a second VLP after transduction of the first VLP and expression of the third nucleic acid sequence in a host cell,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and
15 wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

2. The directed evolution of claim 1, wherein the one or more packaging gene is different from the virus
20

3. The directed evolution system of claim 1, wherein the one or more packaging genes is from a different virus.

4. The directed evolution system of claim 1, wherein the one or more packaging genes is a vesicular stomatitis virus (VSV) G DNA encoding a VSV G protein.
25

5. The directed evolution system of claim 1, wherein the virus is a RNA virus.

6. The directed evolution system of claim 1, wherein the virus is an alphavirus.
30

7. The directed evolution system of claim 1, wherein the virus is Semliki Forest Virus (SFV).

8. The directed evolution system of claim 7, wherein the one or more non-structural genes comprises NSP1, NSP2, NSP 3 and NSP4 of SFV non-structural genes.
35

9. The directed evolution system of claim 7, wherein the one or more non-structural genes further comprises a 5' untranslated region and a 3' untranslated region of SFV non-structural genes operably linked to NSP1, NSP2, NSP3 and NSP4.

10. The directed evolution system of claim 1, wherein the VLP is a virus-like vesicle (VLV)

11. The directed evolution system of any one of claims 1 to 10, wherein the one or more non-structural genes comprises at least one mutation that increases virus titer production.

12. The directed evolution system of claim 11, wherein the at least one mutation is selected from G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834-A, T-8859-A, T-8864-C, G-9211-A, A-10427-G, G-11560-A, and other modifications that improve packaging or functionality of the system.

13. The directed evolution system of claim 11, wherein the at least one mutation comprises G-4700-A, A-5424-G, G-5434-A, T-5825-C, T-5930-C, A-6047-G, G-6783-A, G-6963-A, G-7834-A, T-8859-A, T-8864-C, G-9211-A, A-10427-G, and G-11560-A.

14. The directed evolution system of claim 11, wherein the at least one mutation is selected from G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, E2393K or any other modifications that improve packaging or functionality of the system.

15. The directed evolution system of claim 11, wherein the at least one mutation comprises G106E, L347L, V351I, L481S, I516T, D555G, L880L, T860T, A1151T, 1492I, M1494T, A1610T, N2015S, and E2393K.

16. The directed evolution system of any one of claims 1 to 15, wherein the one or more packaging genes comprises at least one mutation that increases virus titer production selected from A-11871-G, T-11978-C and other modifications that improve packaging or functionality of the system.

17. The directed evolution system of claim 1, wherein the first nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to a DNA sequence encoding the one or more non-structural genes of the RNA virus, operably linked to a DNA sequence encoding the at least one evolution candidate.

18. The directed evolution system of claim 17, wherein the first nucleic acid sequence further comprises an alphavirus subgenomic RNA promoter (SGP) operably linked between the one or more non-structural genes of the RNA virus and the at least one evolution candidate.

19. The directed evolution system of claim 18, wherein the alphavirus SGP is a SFV promoter.

20. The directed evolution system of claim 1, wherein the second nucleic acid sequence comprises a DNA sequence comprising a promoter sequence operably linked to the one or more
5 viral packaging genes.

21. The directed evolution system of any one of claims 17 to 20, wherein the promoter sequence is a cytomegalovirus immediate early promoter (CMV promoter) sequence.

10 22. The directed evolution system of claim 1, wherein the host cell is a mammalian cell.

23. The directed evolution system of claim 22, wherein the mammalian cell is a BHK-21 cell or a HEK293T cell.

15 24. The directed evolution system of claim 1, wherein the evolution candidate comprises a sequence of at least one transgene of interest.

25. The directed evolution system of claim 24, wherein the at least one transgene of interest encodes a protein, a protein domain, a protein-coding region, a non-protein-coding, a truncated
20 protein, a truncated protein domain, a truncated-protein coding region or a truncated non-protein-coding region.

26. The directed evolution system of claim 24 or 25, wherein the evolution candidate is selected from the group consisting of Tet transactivator (tTa), reverse Tet transactivator (rtTa),
25 and anti-p53 nanobody.

27. The directed evolution system of claim 1, wherein the second VLP and the third nucleic acid sequence generates further VLPs after transduction of the second VLP and expression of the third nucleic acid sequence in a host cell, wherein the generation of the further VLPs are
30 dependent on the target element regulated directly or indirectly by the at least one evolution candidate

28. The directed evolution system of claim 27, wherein the system further generates one or more additional VLPs through expression of the third nucleic acid and transduction of the further
35 VLPs in a host cell, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

29. A method of performing directed evolution of at least one evolution candidate, comprising the directed evolution system of claim 1.

30. A method of performing directed evolution of at least one evolution candidate, comprising:

(i) expressing in a host cell with

(a) a first nucleic acid sequence encoding one or more non-structural genes of a virus and the at least one evolution candidate, and

(b) a second nucleic acid sequence encoding one or more viral packaging genes,

(ii) generating a first virus-like particle (VLP) by the host cell, and

(iii) transducing a host cell with the first VLP and expressing a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes to generate a second VLP,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate in step (iii), and

wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

31. The method of claim 30, further comprises screening the host cells for VLPs comprising a desirable evolved candidate.

32. The method of claim 30, further comprises transducing a host cell with the second VLP and expressing the third nucleic acid sequence to generate further VLPs, wherein the generation of the further VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

33. The method of claim 32, further comprises transducing a host cell with the further VLPs and expressing the third nucleic acid sequence to generate one or more additional VLPs, wherein the generation of the one or more additional VLPs are dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

34. A virus-like particle produced by the method of any one of claims 29 to 33.

35. A host cell comprising the virus-like particle of claim 34.

36. A composition comprising the virus-like particle of claim 34.

37. A kit for performing directed evolution, comprising

- (i) a first nucleic acid sequence encoding one or more non-structural genes of a virus and at least one evolution candidate,
- (ii) a second nucleic acid sequence encoding one or more packaging genes, and
- (iii) a third nucleic acid sequence encoding a target element for the at least one evolution candidate and the one or more packaging genes,

wherein the first nucleic acid sequence and the second nucleic acid sequence generates a first virus-like particle (VLP) after expression in a host cell, and wherein said first VLP and the third plasmid generates a second VLP after transduction of the first VLP and expression of the third nucleic acid sequence in a host cell,

wherein said evolution is driven by at least one round of error-prone viral genome replication of the at least one evolution candidate after transduction of the first VLP, and wherein the generation of the second VLP is dependent on the target element regulated directly or indirectly by the at least one evolution candidate.

38. The kit of claim 37, comprising the directed evolution system of claim 1.

39. The kit of claim 37, for use in the method of any one of claims 29 to 33.

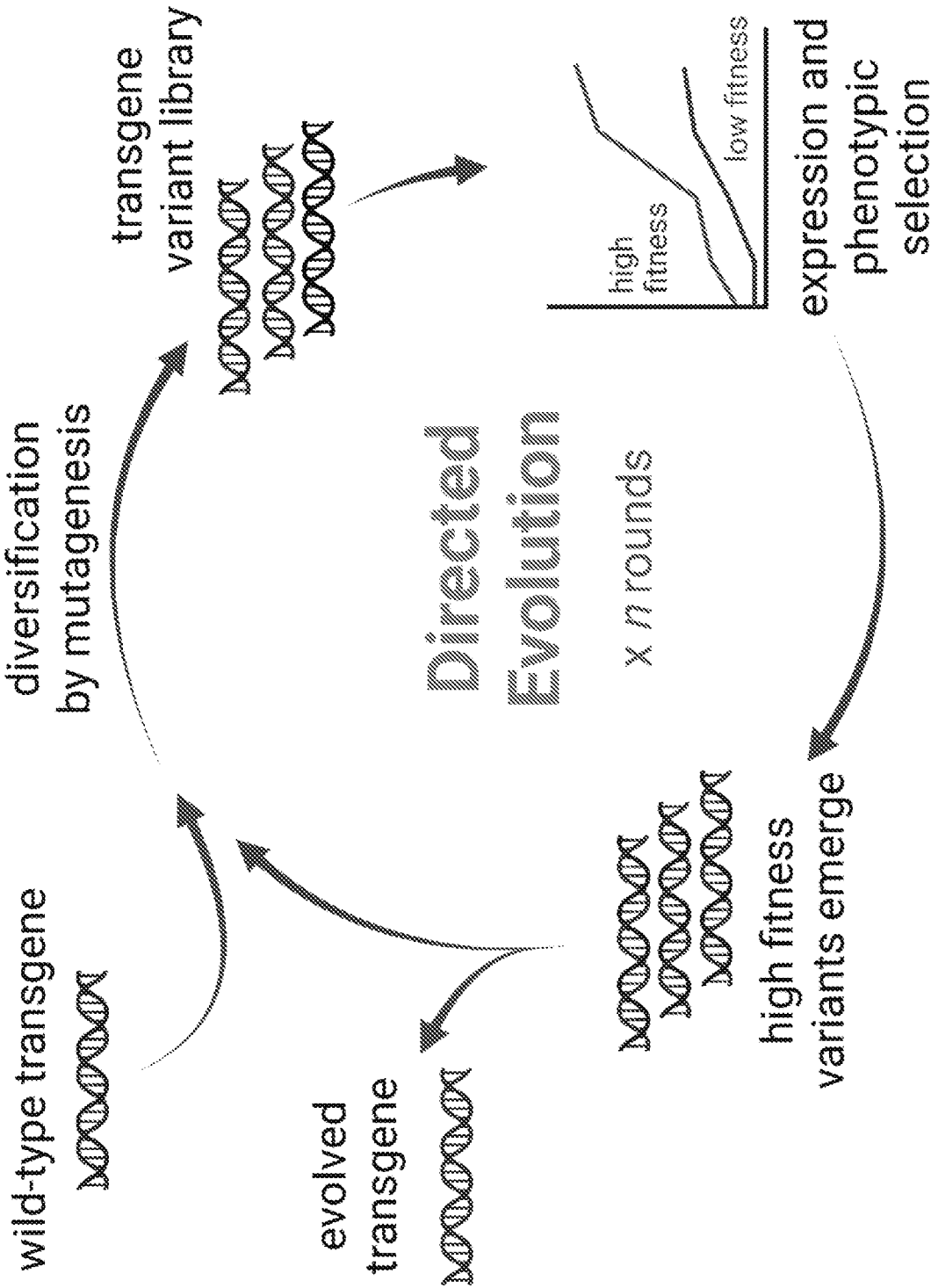


FIG. 1

2/12

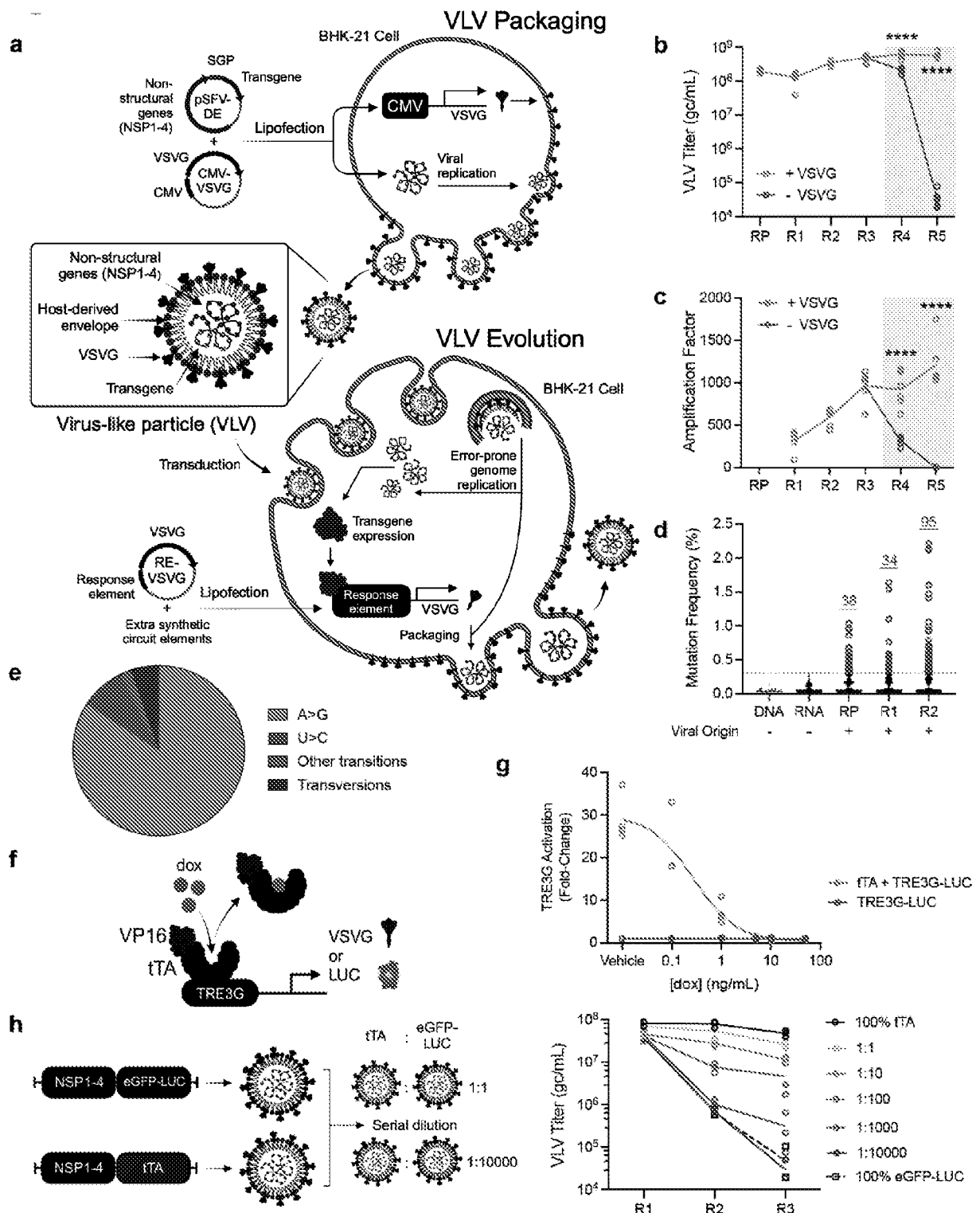


FIG. 2

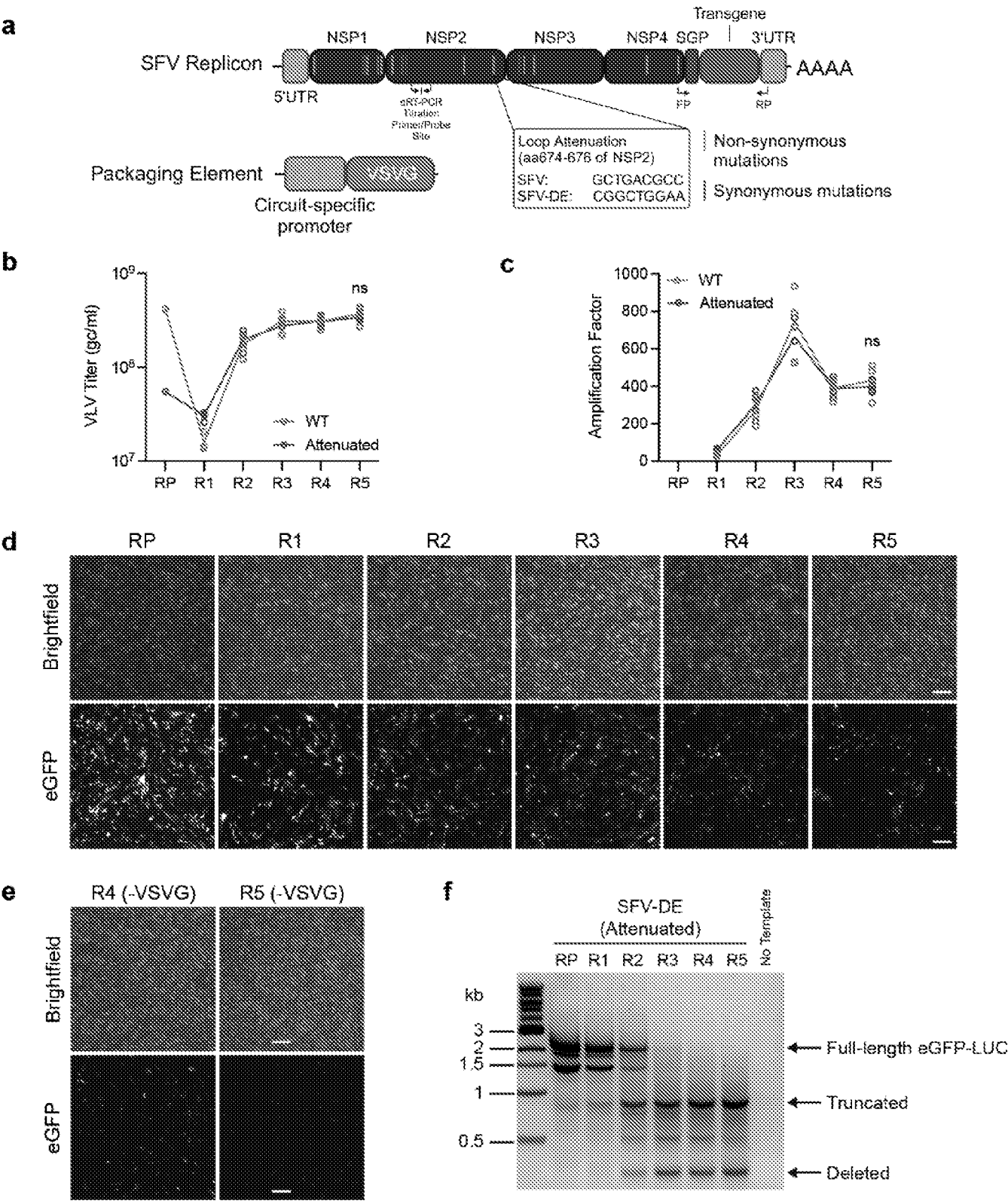


FIG. 3

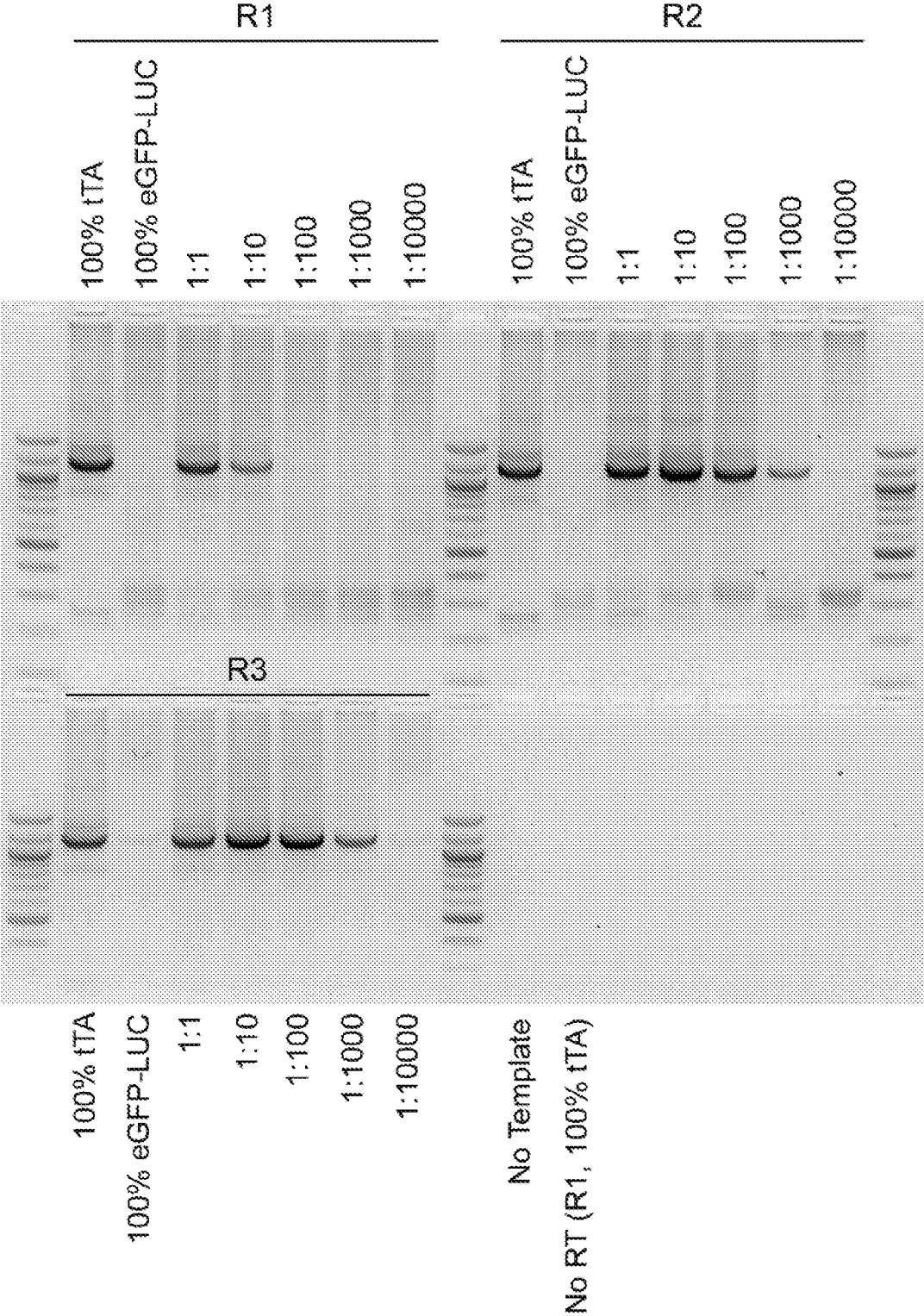


FIG. 4

5/12

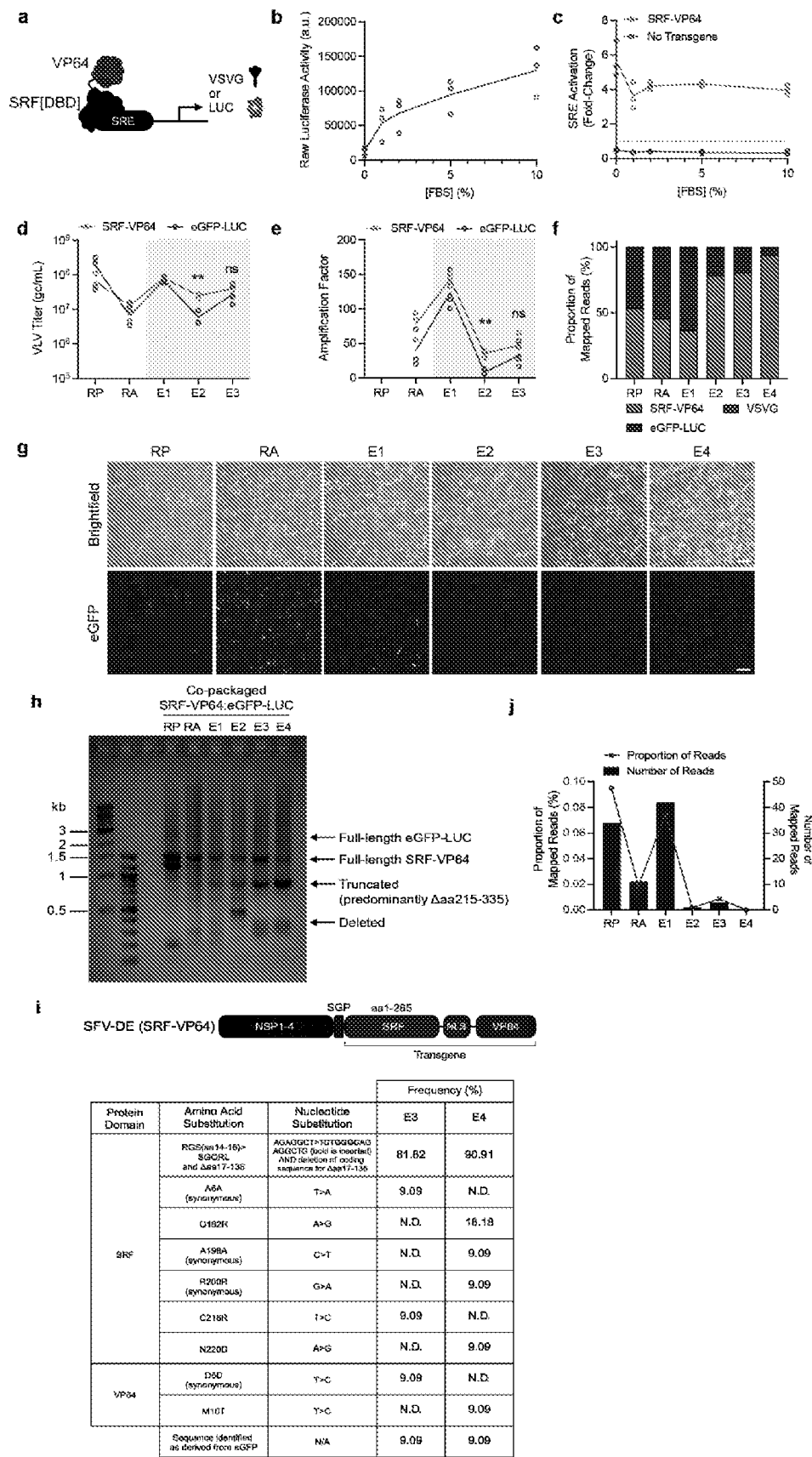


FIG. 5

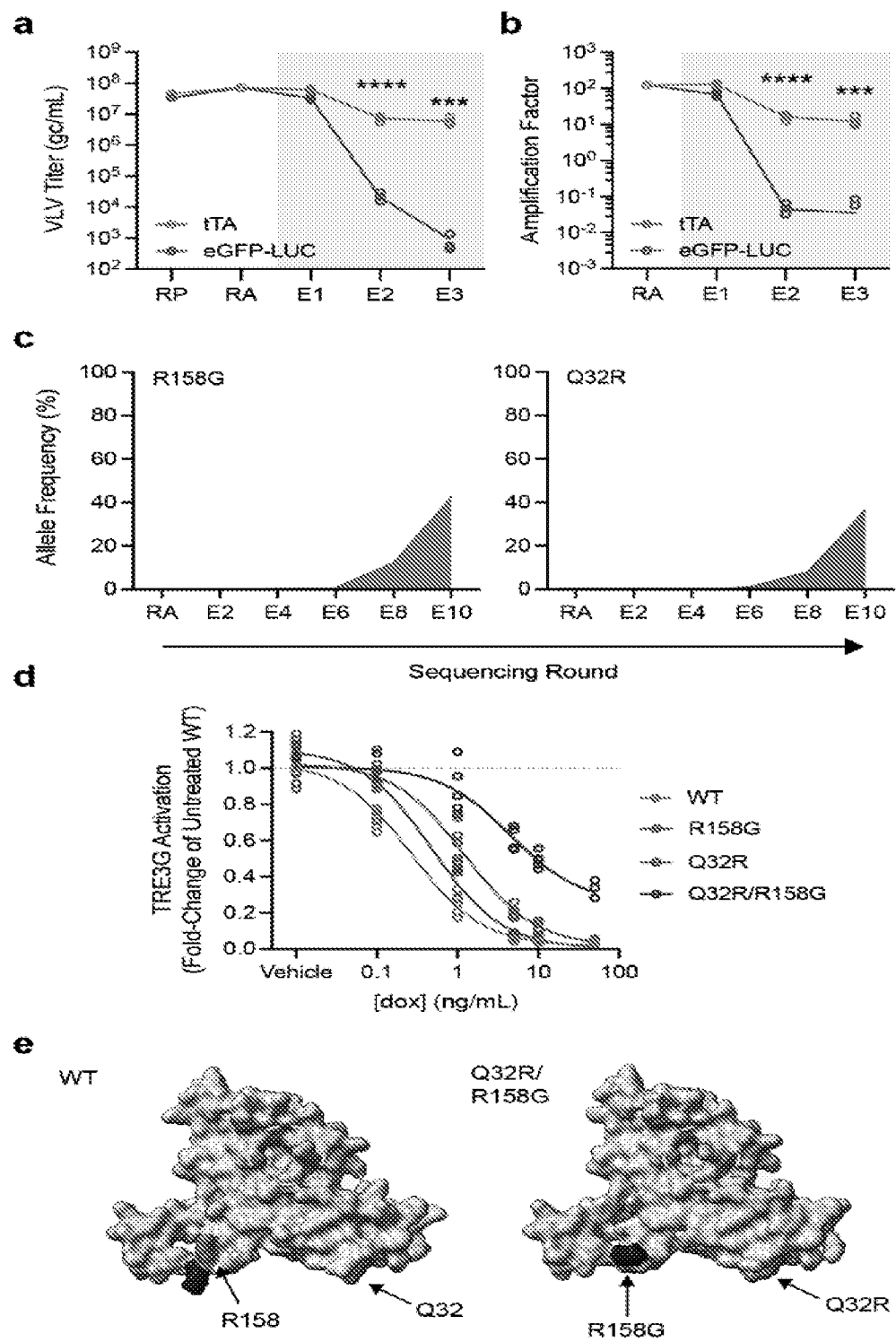


FIG. 6

7/12

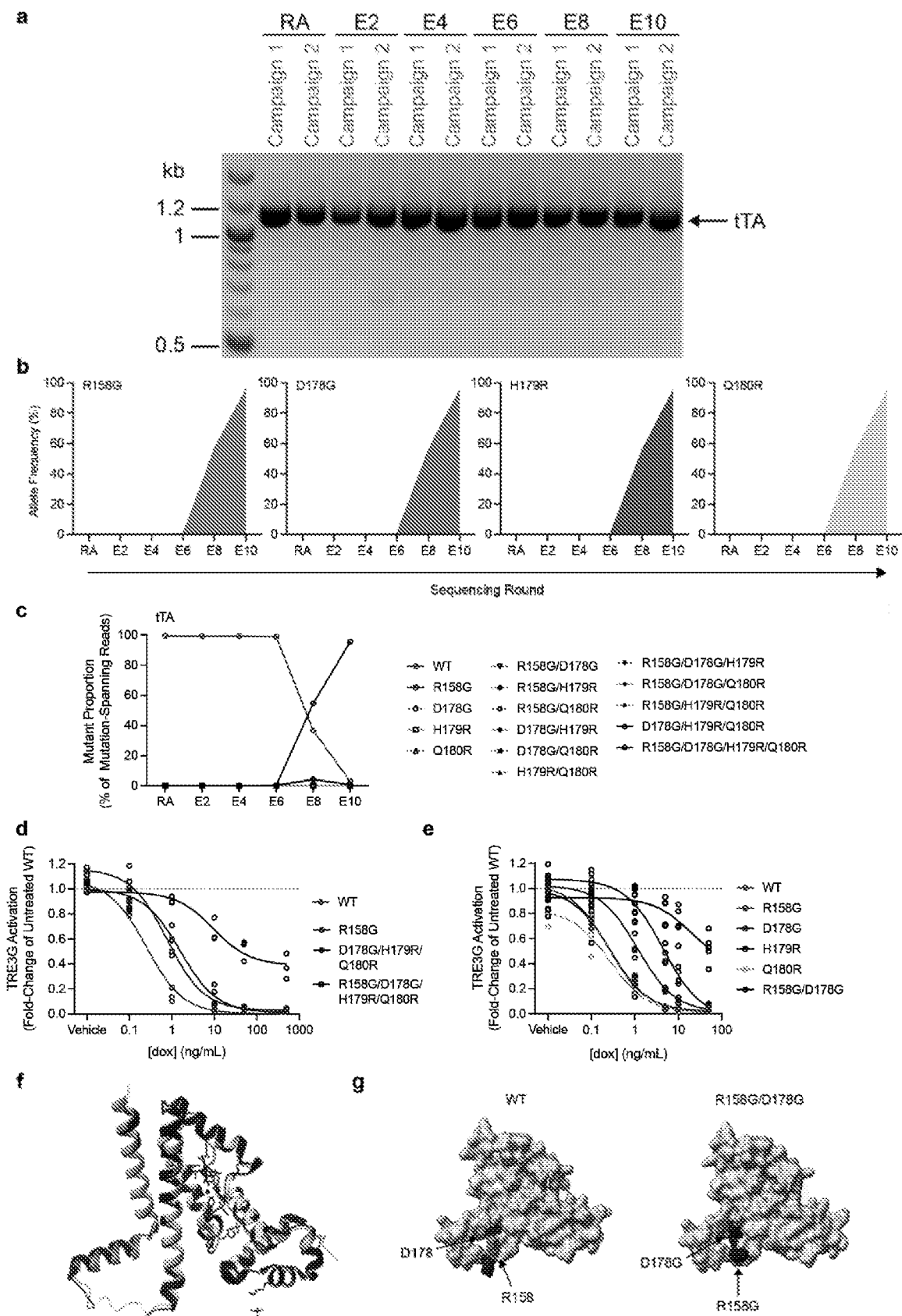
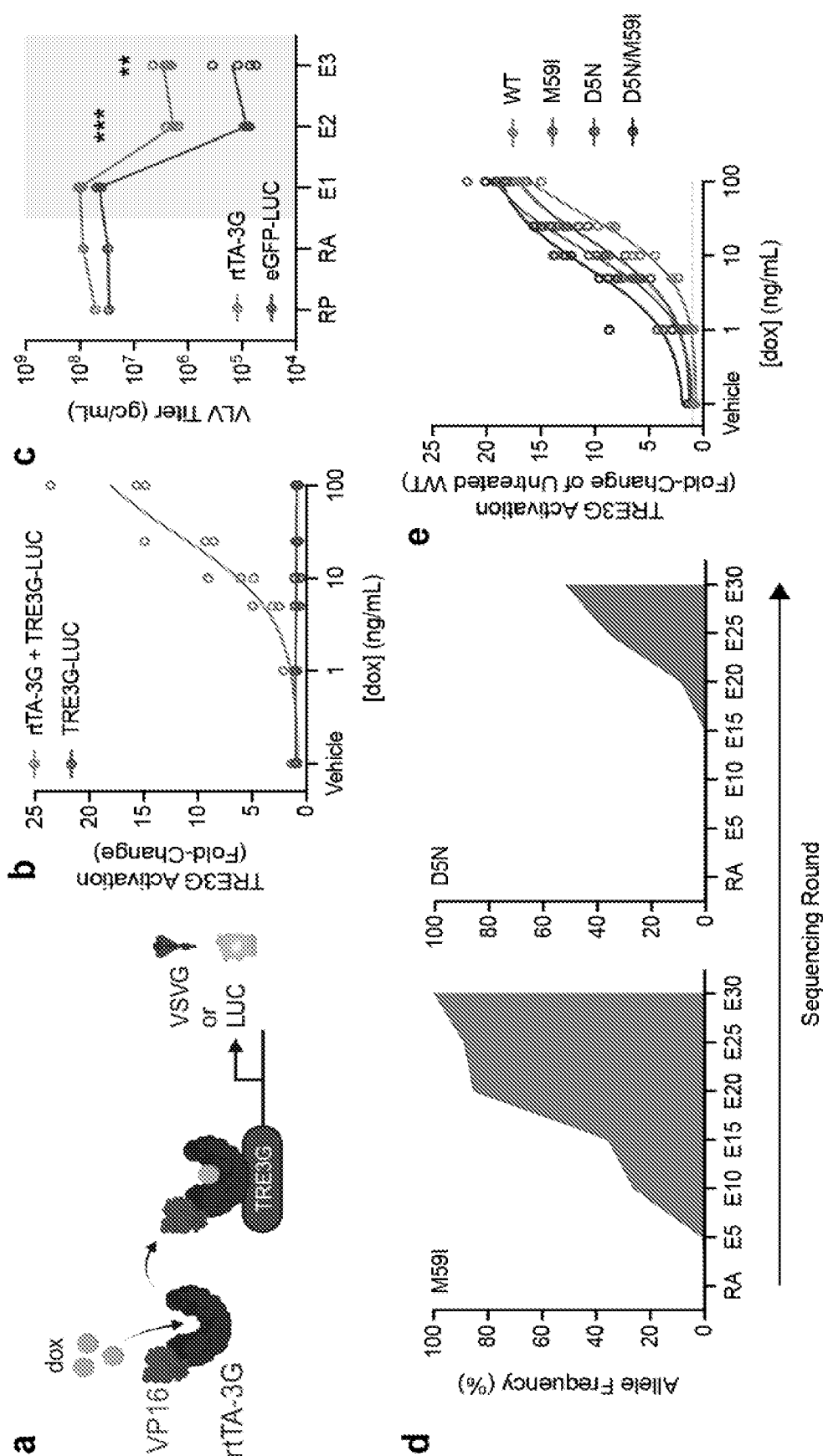


FIG. 7



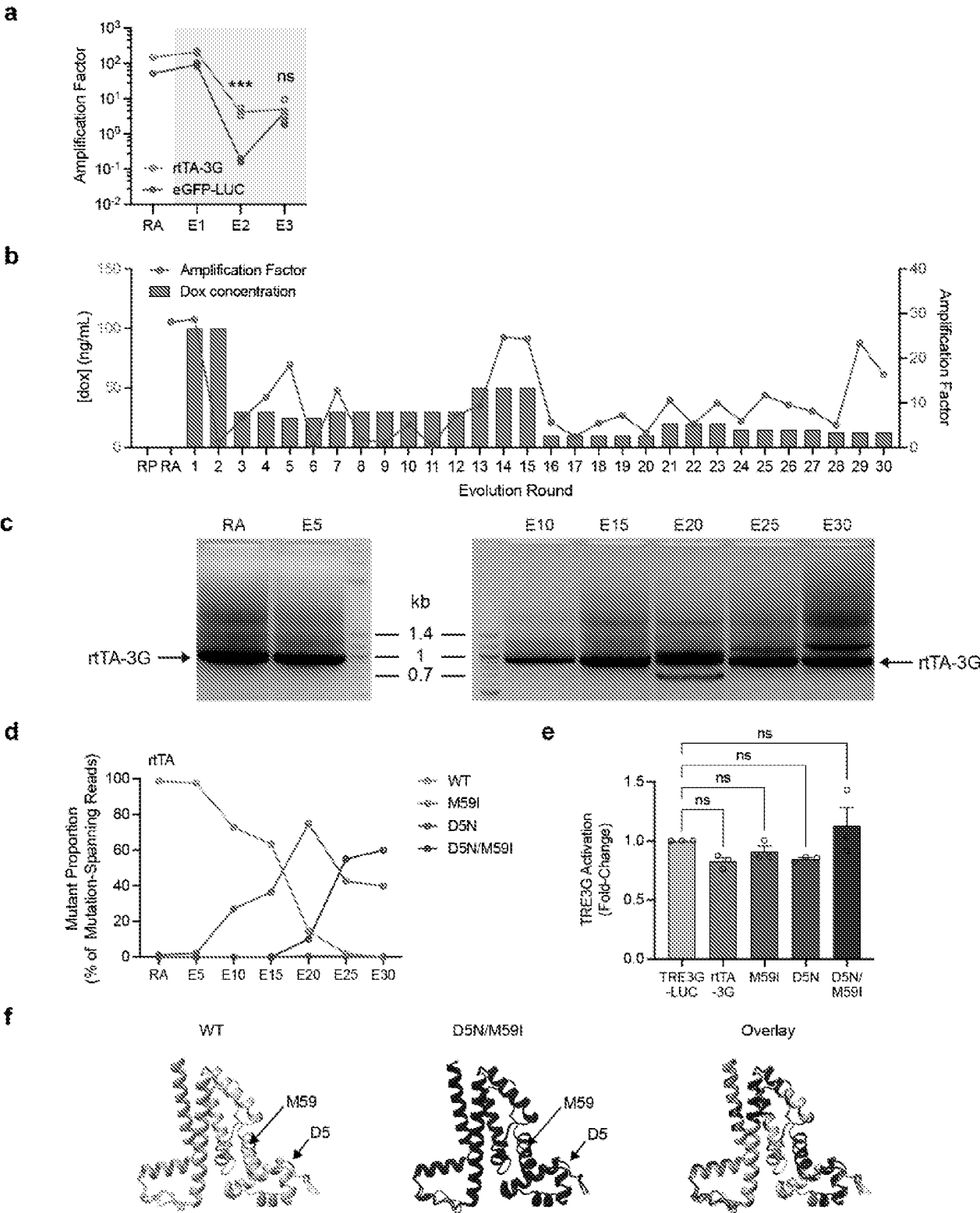


FIG. 9

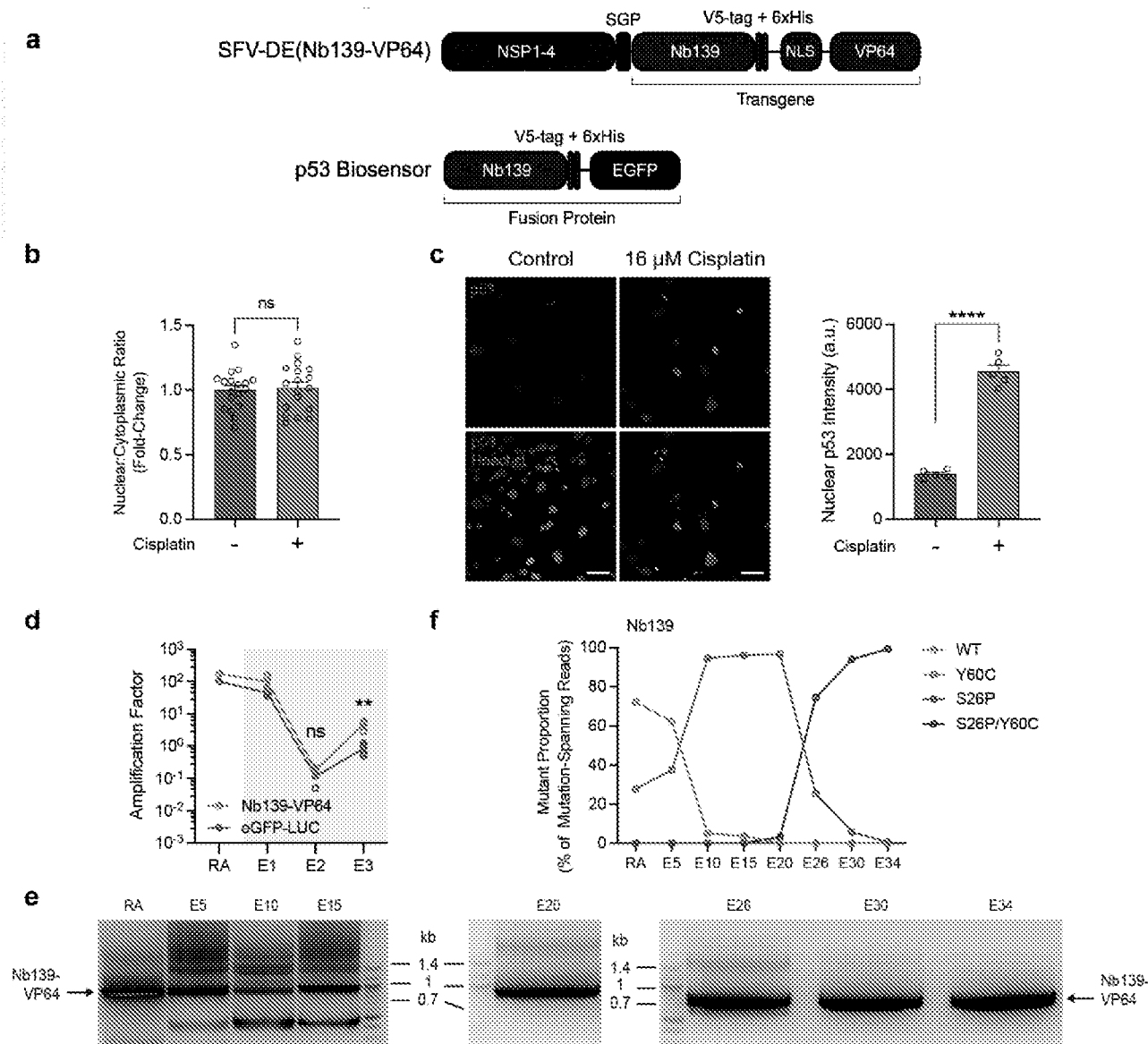


FIG. 10

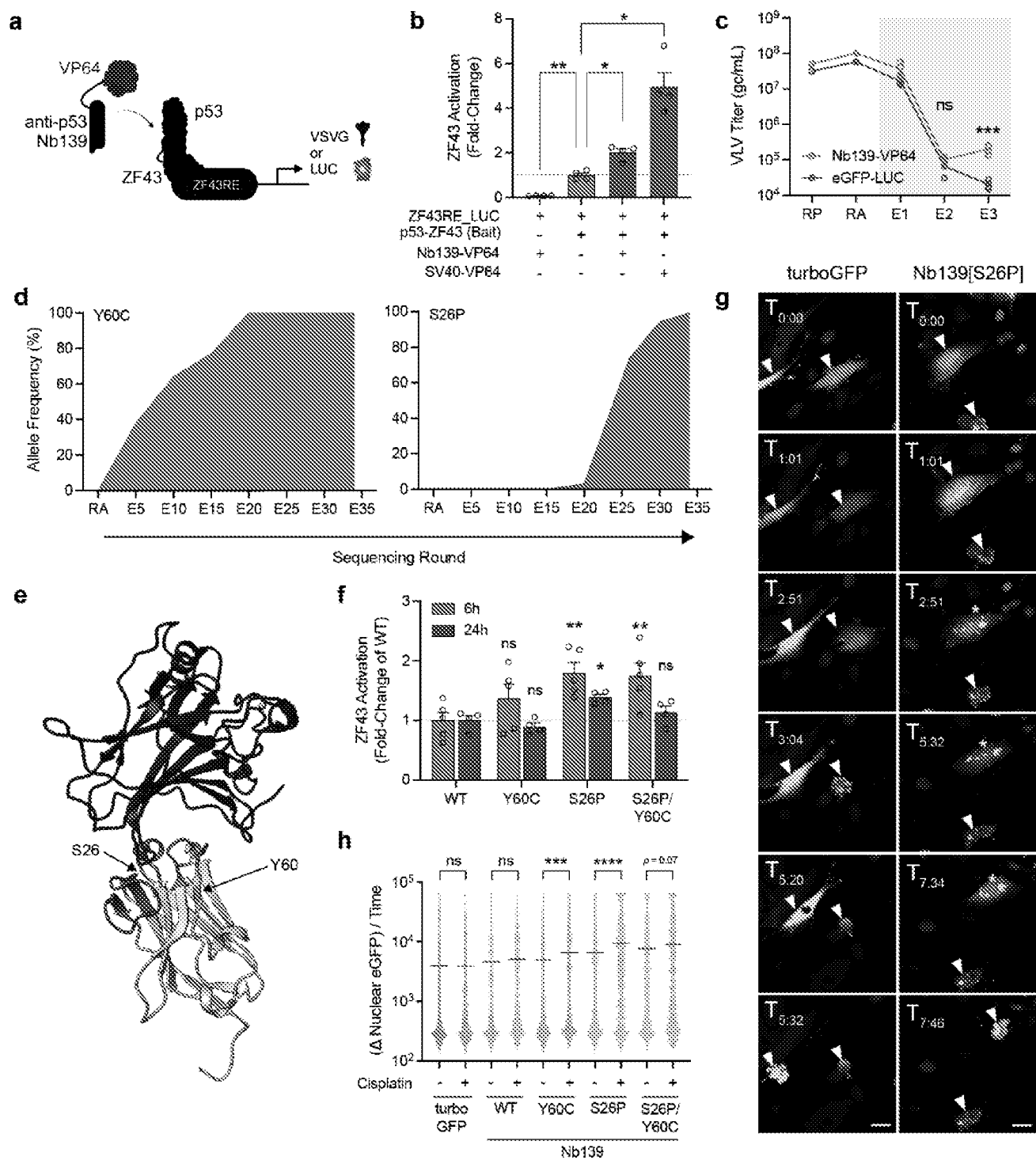


FIG. 11

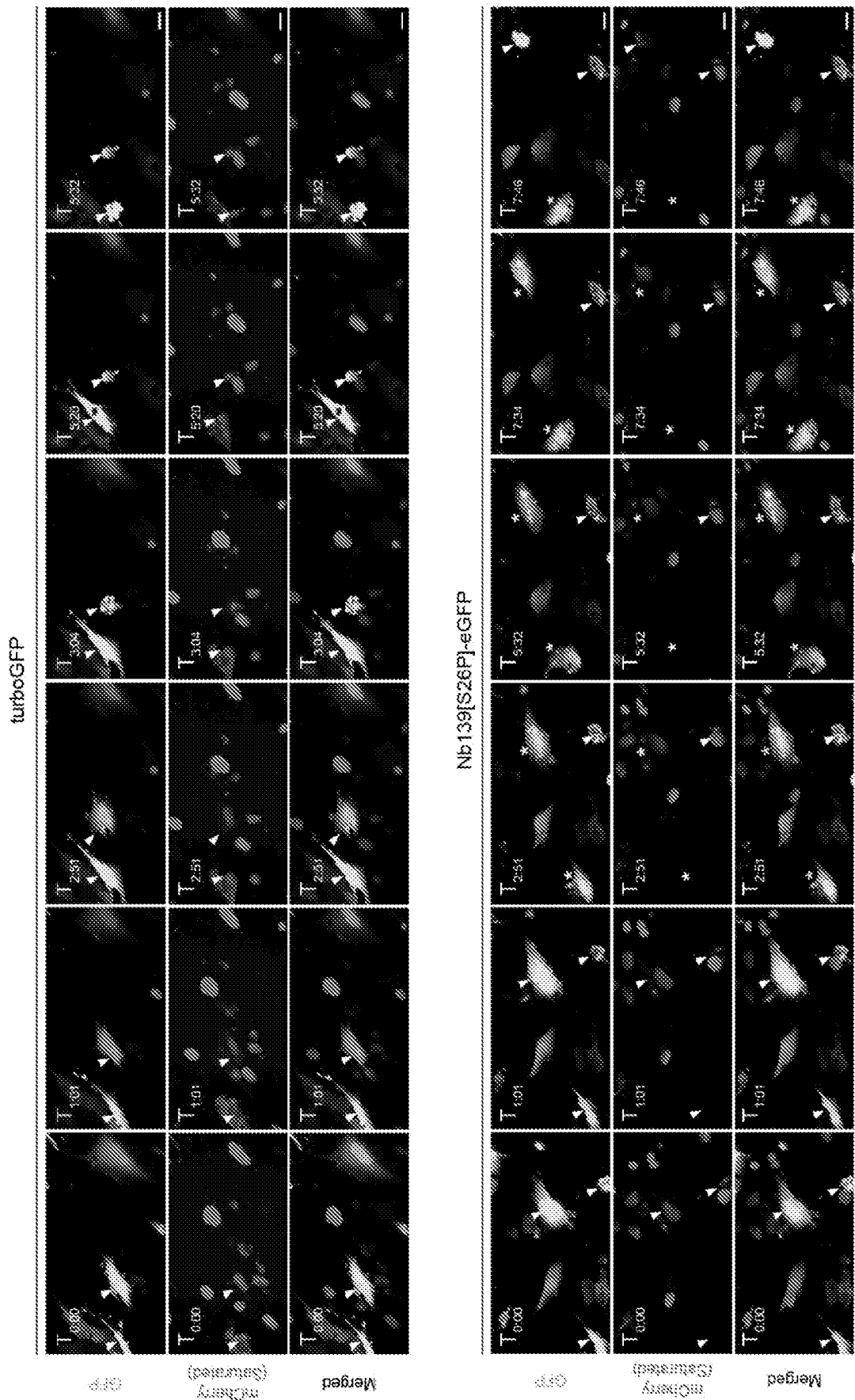


FIG.12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2024/050763

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/10 (2006.01) C12N 7/00 (2006.01) C12N 15/86 (2006.01) C40B 40/08 (2006.01)

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)

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)

STN (MEDLINE, CAPLUS, BIOSIS, EMBASE): directed evolution, virus particle, cheater, capsid-free, VSV G, packaging, target element controlled propagation, alphavirus, SFV, tTA, SRE, SRF, VP64, nanobody and like terms; **EPOQUE** (PATENW English language): C12N15/1058, C12N2770/36111, directed evolution, virus particle, cheater, capsid-free, VSV G, packaging, target element controlled propagation, alphavirus, SFV, tTA, SRE, SRF, VP64, nanobody and like terms; **Google**: directed evolution, response element, alphavirus, mammalian, mPACE; **Google Scholar**: directed evolution, Semliki Forest Virus, inventors names; **IP Australia's internal databases**: applicants and inventors names.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

☒ 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	"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
"D" document cited by the applicant in the international application	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search
17 September 2024

Date of mailing of the international search report
17 September 2024

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
Email address: pct@ipaustalia.gov.au

Authorised officer

Joan Holgate
AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)
Telephone No. +61 2 6283 2671

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU2024/050763

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed.
 - b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).
☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without the furnished WIPO Standard ST.26 compliant sequence listing.

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/AU2024/050763
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	English JG, et al. VEGAS as a platform for facile directed evolution in mammalian cells. Cell. 2019 Jul 25;178(3):748-761. Abstract; Directed Evolution of Transcription Factors with Sindbis Virus, pages 749-751; VEGAS for the Evolution of GPCRs, pages 751-754; VEGAS for the Evolution of Active-State Nanobodies, page 754; Table S6; Data S2; Figures 2-3, S3	1, 5-6, 17-18, 20-39
X Y	Denes CE, et al. The VEGAS platform is unsuitable for mammalian directed evolution. ACS Synthetic Biology. 2022 Oct 11;11(10):3544-3549. Reference 2 Abstract; second last and third last paragraphs, page 3548	1, 5-6, 17-18, 20-39 2-4, 7-16, 19
Y	Rose NF, et al. In vitro evolution of high-titer, virus-like vesicles containing a single structural protein. Proceedings of the National Academy of Sciences. 2014 Nov 25;111(47):16866-16871. Abstract; Figures 1, 3, 6; Table 1	2-4, 7-16, 19
Form PCT/ISA/210 (fifth sheet) (July 2019)		

INTERNATIONAL SEARCH REPORT		International application No.	
Information on patent family members		PCT/AU2024/050763	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
End of Annex			

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.
Form PCT/ISA/210 (Family Annex)(July 2019)