The VEGAS platform is not suitable for mammalian directed evolution

Christopher E. Denes[1,4], Alexander J. Cole[2,4], Minh Thuan Nguyen Tran[3], Mohd Khairul Nizam Mohd Khalid[3], Alex W. Hewitt[3], Daniel Hesselson[2,5], G. Gregory Neely[1,5]

[1]The Dr. John and Anne Chong Lab for Functional Genomics, Charles Perkins Centre and School of Life & Environmental Sciences, The University of Sydney, Sydney, NSW, 2006, Australia

[2]Centenary Institute and Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, 2006, Australia

[3]Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Tasmania, 7000, Australia

[4]These authors contributed equally

[5]These authors contributed equally

*Correspondence: d.hesselson@centenary.org.au (D.H.), greg.neely@sydney.edu.au (G.G.N.)

ORCIDs:

C.E.D.: 0000-0002-8669-8088

A.J.C.: 0000-0001-9019-6884

M.T.N.T.: 0000-0002-9554-6449

M.K.N.M.K.: 0000-0001-7200-3102

A.W.H.: 0000-0002-5123-5999

D.H.: 0000-0002-8675-9426

G.G.N.: 0000-0002-1957-9732

Keywords: VEGAS, Sindbis virus, directed evolution, replicability, reproducibility

Abstract

Directed evolution uses cycles of gene diversification and selection to generate proteins with novel properties. While traditionally directed evolution is performed in prokaryotic systems, recently a mammalian directed evolution system (viral evolution of genetically actuating sequences, or "VEGAS") has been described. Here we report that this platform has major technical issues precluding its use for directed evolution. These issues include a rapid loss of VEGAS system integrity, an inability to propagate VEGAS Sindbis particles across rounds of transduction, and widespread prevalence of circuit "cheaters". Similar results have been obtained in independent labs. It may be possible to use Sindbis virus for mammalian directed evolution in the future, however, in its reported form, VEGAS is not suitable for use as a mammalian directed evolution platform.

INTRODUCTION

A new generation of mammalian directed evolution systems have been reported, which exploit viral life cycles to couple mutagenesis of a virally packaged transgene-of-interest with selection for higher fitness variants (Berman et al. 2018; English et al. 2019). In these systems, essential structural components are removed from the viral genome and placed under the control of synthetic gene expression circuits within the host cell and regulated by the transgene-of-interest. Selection operates on natural mutations arising in the transgene during error-prone viral replication that then impact the expression of the transgene-regulated viral structural genes. Variants that more efficiently drive the synthetic circuit are then overrepresented in the packaged viral particles used for further rounds of evolution (Berman et al. 2018; English et al. 2019). The English, et al. (2019) viral evolution of genetically actuating sequences (VEGAS) system offered a simplified platform for simultaneous mutagenesis and selection using Sindbis virus (SINV). The VEGAS system was designed to link the activity of a transgene to expression of the Sindbis Structural Genome (SSG; comprising Capsid, E3, E2, 6K and E1 genes) (English et al. 2019). Thus, to apply the selective pressure required for directed evolution, SINV replication must be dependent on SSG-expression in the mammalian host cell.

Here, we find that the VEGAS system does not result in viral propagation across rounds of replication when performed as described. Investigating why the VEGAS system failed, we found a rapid loss of system integrity during rounds of replication. While SINV efficiently packaged a transgene and transduced cells as previously reported (Shapiro et al. 2010; Fayzulin et al. 2005), we found that propagation beyond this initial transduction event was hijacked by "cheaters" that packaged SSG components, thereby short-circuiting transgene-dependent evolution. We conclude that contrary to its initial description, the VEGAS system is not suitable for mammalian directed evolution.

RESULTS

In our efforts to implement the VEGAS system, we first attempted to reproduce SSG-dependent replication of SINV (Figure 1B from (English et al. 2019)). We electroporated six independent BHK-21 cultures with pSinHelper, pSinCapsid and pTSin-EGFP seed mRNA (**Figure S1A)** to package SSG-deficient SINV-EGFP particles. After 24 h, we detected 4.72 x $10^9 \pm 5.16 \times 10^8$ (SEM) genome copies (gc)/mL in the supernatant (Round (R)0; **Figure 1A**) and observed bright EGFP fluorescence in most packaging cells (R0; **Figure 1B**). Fresh BHK-

76 21 cultures were transfected with a CMV-SSG plasmid (+SSG) or control DNA (-SSG) and
77 SINV-EGFP particles were added at a calculated multiplicity of infection (MOI) of 1 gc/cell
78 (R1; **Figure 1A**). After 24h, viral titers did not show SSG-dependence (-SSG; 7.35 x 10^6 ±
79 2.30 x 10^5 gc/mL; +SSG, 7.33 ± 4.77 x 10^5 gc/mL; **Figure 1A**) and GFP fluorescence was
80 rarely observed in the transduced cells (**Figure 1B**). Furthermore, these low titers precluded
81 adding R1 SINV-EGFP particles to R2 cells at an apparent MOI of 1 (see below). Together,
82 these data show that SINV particles produced using the VEGAS protocol are not generated at
83 quantities required to perform further rounds of transduction.

84

85 One possible cause for the rapid loss of viral titers observed between R0 and R1 is that
86 transfected host cells failed to express the Sindbis structural genes provided via DNA plasmid.
87 Thus, we next examined host cell expression of the CMV-SSG plasmid used in R1. To avoid
88 the possibility of transfected plasmids contributing to high background signal during qPCR,
89 we developed a barcoding method to discriminate between plasmid DNA (pDNA) and
90 expressed mRNA (**Figure S2A**). Using this approach, we confirmed that the CMV-SSG
91 plasmid was highly expressed (**Figure S2B,C**). We next addressed possible explanations for
92 the low expression of the SINV-EGFP transgene when transduced at an MOI of 1. We
93 suspected that residual seed mRNA in the supernatant following R0 transfection could have
94 been carried over in the R0 output and inflated titers, impeding accurate calculations of MOIs.
95 To test this, we treated packaged R0 virus samples with RNase A to degrade seed mRNA that
96 was not protected by encapsidation. Indeed, RNase A treatment of R0 samples decreased the
97 measured virus titer by 93%, whereas R1 samples, which were not directly exposed to seed
98 mRNA, were much less RNase-sensitive (**Figure S3**). This differential sensitivity indicated
99 that residual packaging RNA was present in the output from R0. To determine whether the
100 inaccurate SINV-EGFP R0 titers were responsible for the lack of SSG-dependent replication,
101 we added neat (undiluted) packaged SINV-EGFP to CMV-SSG-expressing BHK-21 cells. This
102 approach produced 5-10% GFP-positive cells at R1 (arrowheads, **Figure 1D**), demonstrating
103 initial Sindbis packaging was successful, and generated sufficiently high viral titers to test
104 SSG-dependence at R2 (**Figure 1C**). However, at R2 we still did not observe GFP-positive
105 cells (**Figure 1D**), consistent with a dramatic drop in viral titers and lack of SSG-dependence
106 (**Figure 1C**). Overall, initial R0 titer calculations are inflated by contaminating seed mRNA
107 making R1 MOI calculations unreliable. Regardless of this issue, efficient transduction of neat

108 SINV preparations still leads to unproductive replication, indicating that directed evolution
109 campaigns can not be performed with the VEGAS system in its reported form.

110

111 To investigate the integrity of the VEGAS system over successive rounds of transduction we
112 performed long-read nanopore sequencing of viral RNA (**Figure 1E**). Using the approach for
113 transgene recovery as described (English et al. 2019), all reads should map to the transgene
114 subjected to directed evolution. Surprisingly, we found only 53.4% of reads mapped to the
115 EGFP transgene in the initial packaged virus (R0; **Figure 1E**) and transgene inclusion was
116 reduced to 30.3% by R2. The remaining reads mapped to envelope and capsid sequences
117 derived from the pSinHelper and pSinCapsid plasmids. We note that the *in vitro*-transcribed
118 VEGAS pSinHelper and pSinCapsid mRNAs retain the SINV NSP1 packaging signal (Weiss,
119 Geigenmüller-Gnirke, and Schlesinger 1994) (**Figure S4**), which could allow seed mRNA to
120 compete with the SINV-EGFP transgene for packaging. These data show that on application,
121 the VEGAS system rapidly loses integrity and preferentially packages SINV structural
122 elements (cheater particles), further impacting the utility of the VEGAS system as an efficient
123 tool for mammalian directed evolution.

124

125 Although replication-competent SINV can retain a GFP transgene for at least six viral passages
126 at low MOIs (0.1 PFU/cell) (Thomas et al. 2003), we next considered the possibility that the
127 VEGAS system requires selective pressure within a circuit to produce sufficient viral titers to
128 propagate past R2 and maintain system integrity. To test this, we first validated a simple circuit
129 whereby a serum response factor DNA-binding domain (SRF)-NLS-VP64 fusion protein
130 activates a serum response element (SRE)-regulated firefly luciferase reporter (SRE_LUC)
131 (**Figure 2A**). In our hands, SRF-NLS-VP64 activated SRE_LUC similarly to the maximal
132 activation of SRE-Luciferase previously reported (Figure S3B from (English et al. 2019)),
133 confirming the integrity of this circuit. Next, we subcloned SRF-NLS-VP64 into the VEGAS
134 pTSin vector and packaged it in a 1:1 ratio with pTSin-EGFP to provide an evolutionarily
135 neutral competitor. Under selection, the EGFP transgene should drop out while the higher
136 fitness SRF-NLS-VP64-containing viruses, capable of driving expression of the Sindbis
137 structural genome, should rapidly dominate the culture. As predicted, the percentage of GFP-
138 positive cells decreased with serial passage (**Figure 2B),** however by R2 we still did not detect
139 SSG-dependent replication (**Figure 2C**). Thus, even this simple VEGAS circuit is not capable

140 of driving the viral replication required for a directed evolution campaign. Importantly, the
141 circuit-inducing SRF-NLS-VP64 transgene did not outcompete the control EGFP transgene
142 (3.2% SRF-NLS-VP64 vs 21.7% GFP by R2), and both were rapidly outcompeted by "cheater"
143 sequences containing the Sindbis structural elements (75.1%, **Figure 2D**). Together, these data
144 show that even under selective pressure, simple VEGAS feedback circuit components do not
145 show a selective advantage nor lead to the productive rounds of replication required for directed
146 evolution campaigns. Instead, these strategies rapidly become "short-circuited" by cheater
147 particles containing Sindbis structural genes.

148

149 **DISCUSSION**

150 While it is possible that undocumented methodological subtleties could explain why VEGAS
151 Sindbis particles fail to propagate across rounds of transduction, we have thus far not identified
152 any that could account for the loss of VEGAS system integrity. To generate the Sindbis
153 particles used to initiate VEGAS campaigns, English *et al.* used the NEON electroporator,
154 whereas we used the Amaxa electroporator that has previously been used successfully to
155 deliver RNA to the cytosol (Ekstrom and Dean 2011) and efficiently generate functional
156 Sindbis particles (Shapiro et al. 2010). It is unlikely that using a different electroporator would
157 explain our inability to run VEGAS directed evolution campaigns, since both machines
158 produced Sindbis particles that transduced R1 target cells with equivalent efficiency (~5-10%).
159 Instead, the VEGAS system is non-functional due to technical limitations at R2 and beyond.

160

161 Furthermore, our experiments involved VEGAS components generated and made publicly
162 available by English *et al.* when provided (Addgene plasmids #127692–127695). Of note, we
163 could not obtain the CMV-SSG plasmid from English *et al.* and were required to generate this
164 component ourselves, which we validated and used in Figure 1. Moreover, English *et al.*
165 similarly did not provide the original VEGAS circuits required to reproduce their original work,
166 and thus we generated our own serum response factor circuit which we used in Figure 2. Again,
167 while we confirmed this circuit is functional and could drive expression from the SRE as in
168 English *et al.*, we found this circuit could not support viral propagation past R2. A similar
169 inability to use the VEGAS system as described has been observed independently across our
170 separate laboratories over the 3 years since the platform was first described (English et al.
171 2019). We have consulted with English et al. but did not receive advice that resolved these

172 critical technical issues. Together, we conclude that the VEGAS system in its published form
173 is not suitable for directed evolution campaigns.

174

175 Importantly, as Sindbis virus is pathogenic to humans and has the potential to cause meningitis,
176 extreme caution should be used when trying to implement the published VEGAS system in a
177 BSL2 environment since a single recombination event between the transgenic SINV genome
178 and the host cell SSG expression plasmid (which share homologous sequences) could produce
179 a replication-competent virus. In our hands we tested for and did not detect competent virus,
180 however the VEGAS system is almost immediately hijacked by "cheaters" that packaged SSG
181 components, making it currently unsuitable for directed evolution. While we believe there is a
182 future for Sindbis-based mammalian directed evolution systems, this will require substantial
183 efforts to address critical design flaws that limit VEGAS utility. Optimally, any repairs or
184 upgrades should be confirmed by independent groups prior to publication.

185

186 In summary, although we confirm that packaging and transduction of SINV particles efficiently
187 delivers transgenes to fresh host cells as previously documented (English et al. 2019; Fayzulin
188 et al. 2005; Shapiro et al. 2010), we conclude that due to fundamental biological limitations
189 (lack of virus propagation across rounds, loss of VEGAS system integrity, and widespread
190 prevalence of circuit "cheaters"), the VEGAS system as published is not suitable for use as a
191 mammalian directed evolution platform.

192

193 **METHODS**

194 **Cell Culture**

195 BHK-21 [C-13] cells were purchased from the American Type Culture Collection (#CCL-10).
196 Cells were grown in a humidified 37°C (5% CO_2) atmosphere in MEM α (ThermoFisher,
197 #32571101) supplemented with 5% HyClone fetal bovine serum (FBS) (Cytiva Life Sciences,
198 #SH30084.03, AU origin) and 10% tryptose phosphate broth (TPB) (ThermoFisher,
199 #CM0283B), referred to as BHK-21 Growth Medium. During transduction and recovery, cells
200 were maintained in serum-free MEM α with 10% TPB, referred to as BHK-21 Recovery
201 Medium (Serum-Free).

202 **Molecular Biology and Plasmid Construction**

203 All plasmids were designed in SnapGene® (version 5.3.2). Plasmids were generated by PCR
204 amplification of sequences of interest with Velocity DNA Polymerase (Bioline, #BIO-21099)
205 using primers synthesized by IDT or restriction enzyme digestion with NEB High-Fidelity
206 enzymes. Assembly of amplicons was performed using the NEBuilder HiFi DNA Assembly
207 Master Mix (NEB, #E2621). Assembled products were transformed into NEB® 10-beta
208 Competent *E. coli* (NEB, #C3019) and selected on LB agar plates (ThermoFisher, #22700025)
209 supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, #A9518). Individual colonies grown
210 overnight in liquid LB broth (ThermoFisher, #12795-084) supplemented with 100 µg/ml
211 ampicillin were processed with either the ISOLATE II Plasmid Mini Kit (Bioline, #BIO-
212 52057) for sequence verification, or the PureYield™ Plasmid Maxiprep System (Promega,
213 #A2393) for transfection and *in vitro* transcription applications. Plasmid constructs were
214 verified by restriction digestion and Sanger sequencing at the Australian Genome Research
215 Facility (AGRF). A list of plasmids used and generated in this study is available in
216 Supplementary File 1 – Plasmid List.

217 **mRNA synthesis**

218 mRNA for electroporation and transfection was produced using the mMESSAGE
219 mMACHINE™ SP6 Transcription Kit (ThermoFisher, #AM1340) from XbaI-linearized SP6-
220 driven plasmids as outlined previously (English et al. 2019). mRNA concentrations were
221 calculated using the Qubit™ RNA BR Assay Kit (ThermoFisher, #Q10210). mRNA integrity

222 was assessed by gel electrophoresis (**Figure S1**). mRNA was frozen at -80°C immediately after
223 transcription as per (English et al. 2019)**.**

224 **Packaging of SINV Particles (referred to as Round 0 (R0))**

225 1 x 10^6 BHK-21 cells were electroporated with a total of 7.8 µg of mRNA (1:1:1 of pSinHelper,
226 pSinCapsid and pTSin-EGFP/pTSin-SRF-NLS-VP64) using Amaxa 2B (Lonza) as per the
227 manufacturer's instructions for BHK-21 cells. The Amaxa electroporator efficiently delivers
228 SINV RNA into the cell cytosol (Ekstrom and Dean 2011), allowing for the packaging of
229 mature SINV particles (Shapiro et al. 2010). Electroporated cells were plated in BHK-21
230 Recovery Medium (Serum-Free). Virus-containing supernatants were collected 24 hours after
231 mRNA delivery and centrifuged at 500 *g* for 5 minutes to pellet cellular debris. Clarified
232 supernatants were collected for titration and subsequent transduction experiments.

233 **Viral Titration**

234 Viral supernatants were titrated as outlined by English and colleagues (English et al. 2019).
235 Following collection and clarification, undiluted SINV-containing supernatants were
236 combined with the TaqMan™ Fast Virus 1-Step Master Mix (ThermoFisher, #4444434) and
237 the appropriate primer-probe sets in technical triplicates. Plates were run on a QuantStudio™
238 7 Flex Real-Time PCR System (ThermoFisher). As described previously, serially diluted *in*
239 *vitro*-transcribed pTSin mRNA was used as the standard curve (ranging between 10^3-10^7
240 genome copies (gc) per mL) (English et al. 2019). Standard curves were used to determine viral
241 titers in gc/mL. Primer sequences are listed in Supplementary File 2 – Primer List.

242 **Transduction of Viral Particles (Round 1 (R1)-onwards)**

243 BHK-21 cells were seeded in 6-well plates at 6.5 x 10^4 cells/well, incubated for 24 hours, and
244 transfected with a total of 2.5 µg of SSG-encoding DNA/well using *Trans*IT-2020 Transfection
245 Reagent (Mirus Bio, #MIR5400) following the manufacturer's recommendations. 6 hours post-
246 transfection, cells were rinsed twice with DPBS before viral inoculum (either undiluted or
247 diluted to a calculated MOI of 1 gc/cell) was applied in a 380 µl volume of BHK-21 Recovery
248 Medium (Serum-Free). Cells were incubated with virus for 1 hour and rinsed twice with DPBS
249 before 1.5 mL BHK-21 Recovery Medium (Serum-Free) was added for a further 23 hours of
250 incubation. Virus-containing supernatants were collected and processed for titration and
251 transduction experiments as described in *Packaging of SINV Particles*.

252 We note that recombination between the SINV genome and homologous sequences in the SSG
253 3'UTR (**Figure S4)** could produce competent viral populations. While this has not been
254 observed in our independent labs, we urge responsible research practices and approvals for
255 working in appropriate Biosafety Level 2 facilities, and the application of appropriate testing
256 protocols for determining replication competency. Assaying viral titer amplification on cells ±
257 SSG elements in parallel can be used to quickly ascertain structural element dependency status.

258 **RNase Digestion**

259 To test if inflated R0 viral titers were a result of residual packaging mRNA, pooled viral
260 supernatants were digested with 0 or 400 ng RNase A (Macherey-Nagel, #740505.50) for 4
261 hours at 37°C. Samples were then incubated for 20 minutes at room temperature with 80 units
262 of RNaseOUT (ThermoFisher, #10777019). Samples were titrated as described in *Viral*
263 *Titration*. RNase A treatment was not applied to any samples used for subsequent transduction
264 assays.

265 **Barcoded qPCR**

266 RNA was extracted using the FavorPrep™ Blood/Cultured Cell Total RNA Mini Kit (Fisher
267 Biotec, #FABRK 001-1) and 400 ng of cDNA was synthesized using the iScript™ Select
268 cDNA Synthesis Kit (Bio-Rad Laboratories, #1708897) on a Mastercycler® nexus PCR
269 Thermal Cycler (Eppendorf): 42°C for 60 min, 85°C for 5 min. cDNA was produced using two
270 primers: a specific primer for hamster GAPDH, and a SINV structural genome primer
271 containing a barcode. To remove residual primers, cDNA was bound to AMPure XP magnetic
272 beads (Beckman Coulter, #A63881) in 0.75 M LiCl, 20% PEG 8000 buffer at a 1:0.9
273 DNA:buffer volume ratio, and washed twice with 70% ethanol. qPCR was performed using
274 SYBR™ Select Master Mix (ThermoFisher, #4472908) following manufacturer's
275 recommendations in triplicate on a QuantStudio™ 7 Flex Real-Time PCR System
276 (ThermoFisher). Primer sequences are listed in Supplementary File 2 – Primer List.

277 **Luciferase Assay**

278 BHK-21 cells were seeded in 96-well plates at 2200 cells/well. Cells were transfected with a
279 total 90 ng of plasmid constructs using *Trans*IT-2020 Transfection Reagent following the
280 manufacturer's recommendations. After 6 hours, cells were washed twice with DPBS and
281 switched to BHK-21 Recovery Medium. At 24 hours post-transfection, luciferase activity was

282 assessed using the Dual-Glo Luciferase Assay System (Promega, #E2940) following
283 manufacturer's recommendations in black-bottom plates. Both Renilla and firefly
284 luminescence were measured on an Infinite M1000 PRO microplate reader (Tecan). Raw
285 firefly luminescence values were normalized to Renilla luminescence and log2-transformed.

286 **Transgene Isolation**

287 Sextuplicate viral supernatants from each round of viral replication were pooled in equal ratios.
288 Viral RNA from 400 µl aliquots was isolated with the MagMAX™ Viral RNA Isolation Kit
289 (ThermoFisher, #AM1939) as per the manufacturer's recommendations. Transgene sequences,
290 situated between nsP4 and the viral 3' untranslated region (UTR) were reverse-transcribed and
291 PCR-amplified using the SuperScript™ IV One-Step RT-PCR System (ThermoFisher,
292 #12594025) and the 26S-F primer and pooled SinRev primer. Primer sequences are listed in
293 Supplementary File 2 – Primer List. Amplicons of the appropriate size range for the transgene
294 of interest were then gel extracted for nanopore sequencing using the ISOLATE II PCR and
295 Gel Kit (Bioline, #BIO-52060) (**Figure S5**).

296 **Nanopore Sequencing**

297 *Sample Processing*

298 Isolated transgene DNA was processed to generate libraries for full-length transgene
299 sequencing using Oxford Nanopore Technologies (ONT) Flongle flow cells (ONT, FLO-
300 FLG001). Samples were prepared for sequencing according to the ONT protocol 'Amplicons
301 by Ligation' (version ACDE_9064_V109_REVP_14AUG2019). 200 fmol of DNA
302 (determined following quantification with the Qubit™ dsDNA HS Assay Kit) was processed
303 using the NEBNext Companion Module for ONT Ligation Sequencing (NEB, #E7180).
304 Sequencing adapters were added using the Ligation Sequencing Kit (ONT, #SQK-LSK109).

305 *Sequencing and Basecalling*

306 Up to 40 fmol of DNA library was loaded onto ONT Flongle flow cells (R9.4.1) in a MinION
307 Mk1B Sequencer fitted with a Flongle Adapter (ONT). Sequencing was performed using
308 MinKNOW (ONT, version 4.2.8) under default parameters and the following specified inputs:
309 kit used, SQK-LSK109; 0.5 hours between MUX scans; basecalling, disabled. A minimum of
310 120,000 raw reads were obtained for each sample. Raw FAST5 files were processed using

311 Guppy (version 4.5.2) with the minimum q-score set to 7.0. Raw FAST5 and basecalled
312 FASTQ reads were deposited at the European Nucleotide Archive (PRJEB47639).

313 *Alignment*

314 Quality-filtered base-called reads (in FASTQ format) were processed using EPI2ME Desktop
315 Agent (ONT, version 3.3.0.1031). Reference files, consisting of complete transgene-coding
316 sequences or helper- and capsid-coding sequences, were uploaded using the Fasta Reference
317 Upload workflow (v2021.07.15). For each sample, 100,000 reads were aligned to reference
318 files using the Fastq Custom Alignment workflow (v2021.03.25) using default parameters.
319

320 **Microscopy**

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

325 **Statistics**

326 Statistical analyses were performed in GraphPad Prism 9.2.0 for Mac, GraphPad Software, San
327 Diego, California USA, www.graphpad.com. Fold-changes in luciferase activity were
328 statistically analysed with a Brown-Forsythe and Welch ANOVA test, assuming Gaussian
329 distribution and unequal standard deviations. Means were compared to the control baseline
330 mean. qPCR data was analysed using multiple student t-test. p-values < 0.05 were considered
331 significant. All data were plotted as mean ± SEM.

332 **ACKNOWLEDGMENTS**

333 A.W.H. was supported by National Health and Medical Research Council (NHMRC)
334 Practitioner Fellowship GNT1150144. D.H. is funded by NHMRC grant GNT1185002 and the
335 Centenary Institute. G.G.N. is funded by NHMRC grants GNT1107514, GNT1158164,
336 GNT1158165, GNT1185002, the NSW Ministry of Health, and a kind donation from Dr. John
337 and Anne Chong. Figure schematics were created with BioRender.com.

338

339

AUTHOR CONTRIBUTIONS

Conceptualization, CED, AJC, DH, GGN; Methodology, CED, AJC, MTNT, DH; Investigation, CED, AJC; Resources, AWH, GGN; Data analysis, CED, AJC, MKNMK, MTNT, DH; Writing - original draft, CED, AJC, MTNT, DH; Writing - review & editing, CED, AJC, MTNT, MKNMK, AWH, DH, GGN; Supervision and project administration, AWH, DH, GGN; Funding acquisition, AWH, DH, GGN.

DECLARATION OF INTEREST

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community.

355 **FIGURE LEGENDS**

356 **Figure 1. VEGAS generates unproductive transductions.**

357 A. Viral titers after packaging and transduction of pTSin-EGFP into control or SSG-
358 expressing BHK-21 cells (N = 6). Horizontal gray bars indicate the batch-specific
359 thresholds for viral detection.
360 B. Phase contrast and EGFP fluorescence images of BHK-21 cells used for packaging (R0)
361 and transduction (R1 [MOI 1]). Scale bars represent 200 µm.
362 C. Viral titers after transduction with undiluted R0 virus into SSG-expressing BHK-21
363 cells. At R2, viruses were transduced at MOI 1 into control or SSG-expressing BHK-
364 21 cells (N = 6).
365 D. Phase contrast and EGFP fluorescence images of BHK-21 cells transduced with
366 undiluted R0 virus (R1 [Neat]) and MOI 1 at R2. Scale bars represent 200 µm. White
367 arrowheads indicate EGFP-positive cells. Green boxes denote a magnified section of
368 the image with enhanced brightness and contrast to highlight GFP-positive cells.
369 E. Nanopore sequencing of viral RNA isolated from pooled samples in Panels C and D (N
370 = 6). Reads (~100,000 per sample) were aligned to viral reference sequences.

371

372 **Figure 2. Selective pressure does not rescue system integrity.**

373 A. Activation of an SRE-regulated luciferase reporter (SRE_LUC) by SRF-NLS-VP64.
374 Error bars represent mean ± SEM (N = 3, with 3 technical replicates). A Brown-
375 Forsythe and Welch ANOVA test was used to determine statistical significance. ****
376 $p < 0.0001$.
377 B. Phase contrast and EGFP fluorescence images of BHK-21 cells at R0-R2 for an SRF-
378 NLS-VP64/SRE_SSG circuit. Scale bars represent 200 µm.
379 C. Viral titers of an SRF-NLS-VP64 transgene-carrying virus packaged in a 1:1 ratio with
380 an EGFP-carrying virus (N = 6).
381 D. Nanopore sequencing of viral RNA isolated from pooled samples (N = 6). Reads
382 (~100,000 per sample) were aligned to viral reference sequences.
383
384
385

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Integrity of *in vitro*-transcribed mRNA.

A. Gel electrophoresis of VEGAS *in vitro*-transcribed mRNA.
B. Gel electrophoresis of *in vitro*-transcribed mRNA used for SRF-NLS-VP64 selection circuit.

Figure S2. Expression of the SSG in transfected BHK-21 cells.

A. Schematic of the barcoded qPCR method used to discriminate mRNA from pDNA during transient transfection. A barcoded SSG-specific primer was used for reverse transcription. Unincorporated barcoded primers were depleted by bead clean-up (indicated by gray circles). qPCR was performed using an SSG-specific forward primer and a barcode-specific reverse primer to discriminate between SSG cDNA and residual pDNA.
B. Raw Ct values from barcoded qPCR for SSG expression of the CMV-SSG plasmid in BHK-21 cells, ± reverse transcription (RT). t-tests were used to determine statistical significance. ***$p < 0.001$, ns = non-significant.
C. SSG expression was calculated relative to GAPDH. Error bars represent mean ± SEM (N = 3). t-tests were used to determine statistical significance. *$p < 0.05$.

Figure S3. RNase A-digested R0 and R1 viruses.

Pooled SINV from R0 and R1 (N = 6, produced on +SSG BHK-21 cells) were treated with RNAse A and titrated.

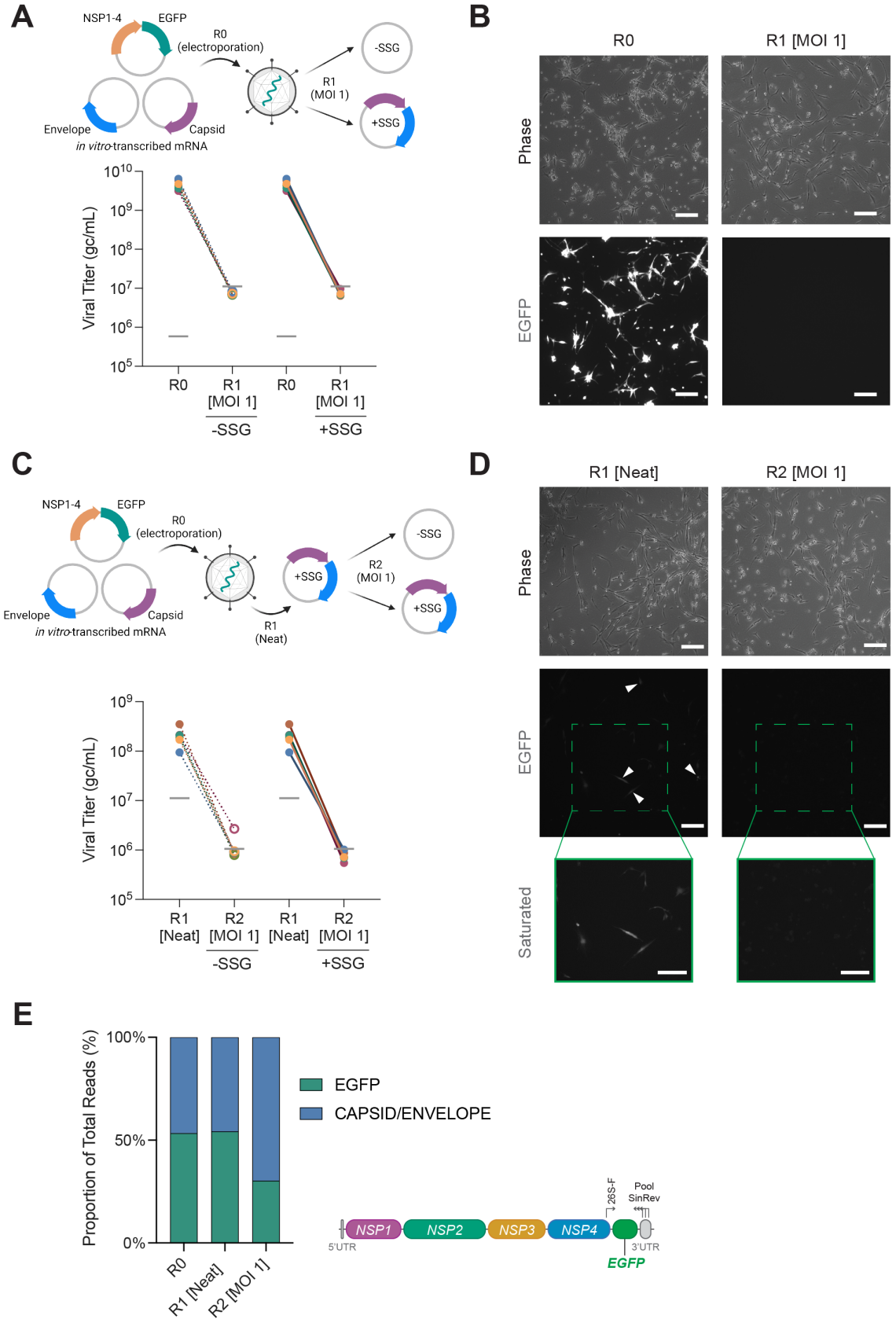
Figure S4. Packaging signal alignment in seed mRNA sequences.

Schematic showing the position of the SINV packaging signal within VEGAS seed mRNAs (asterisks). Genes and truncations drawn to scale.

Figure S5. Isolated transgene RT-PCR amplicons for nanopore sequencing.

A. Transgenes from electroporated VEGAS viruses (from Figure 1) were amplified by RT-PCR (boxed in red) and gel extracted for nanopore sequencing.
B. Transgenes from the SRF-NLS-VP64/SRE_SSG circuit (from Figure 3C) were amplified by RT-PCR (boxed in red) and gel extracted for nanopore sequencing.

428 **MAIN FIGURES**



429

430 **Figure 1**

431

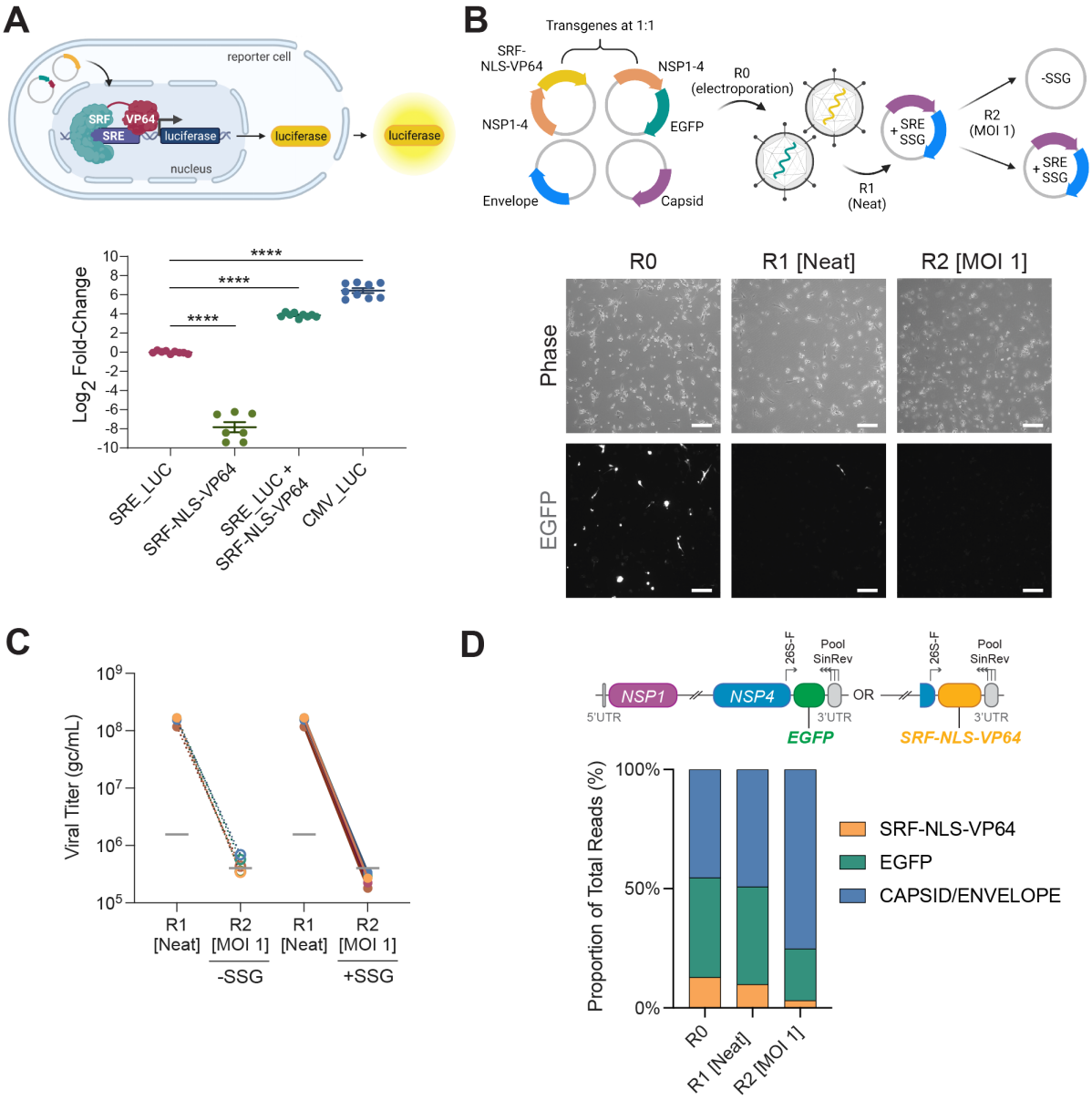
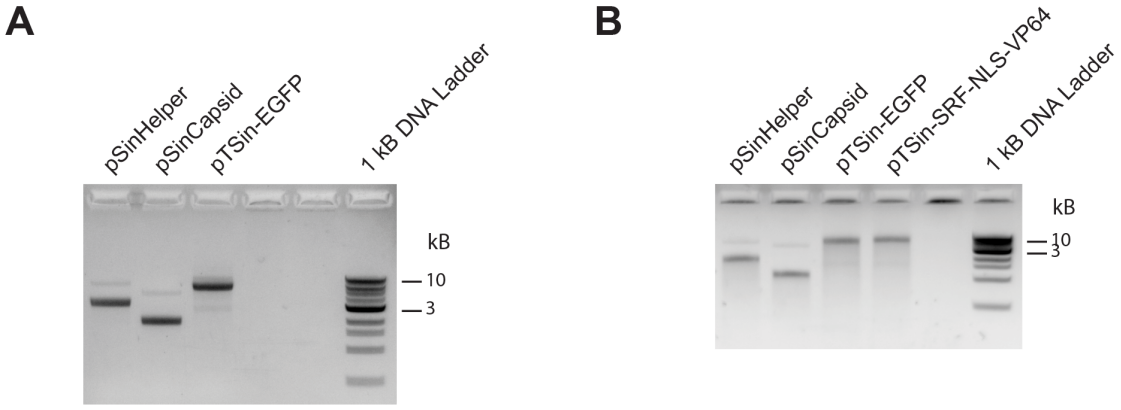

A
reporter cell
SRF
VP64
SRE
luciferase
luciferase
luciferase
nucleus
B
Transgenes at 1:1
SRF-NLS-VP64
NSP1-4
NSP1-4
EGFP
Envelope
Capsid
R0 (electroporation)
R1 (Neat)
+ SRE SSG
R2 (MOI 1)
-SSG
+ SRE SSG
Log2 Fold-Change
****
****
****
SRE_LUC
SRF-NLS-VP64
SRE_LUC + SRF-NLS-VP64
CMV_LUC
R0
R1 [Neat]
R2 [MOI 1]
Phase
EGFP
C
Viral Titer (gc/mL)
R1 [Neat]
R2 [MOI 1]
R1 [Neat]
R2 [MOI 1]
-SSG
+SSG
D
26S-F
Pool SinRev
NSP1
NSP4
OR
5'UTR
3'UTR
EGFP
SRF-NLS-VP64
Proportion of Total Reads (%)
100%
50%
0%
SRF-NLS-VP64
EGFP
CAPSID/ENVELOPE
R0
R1 [Neat]
R2 [MOI 1]

Figure 2

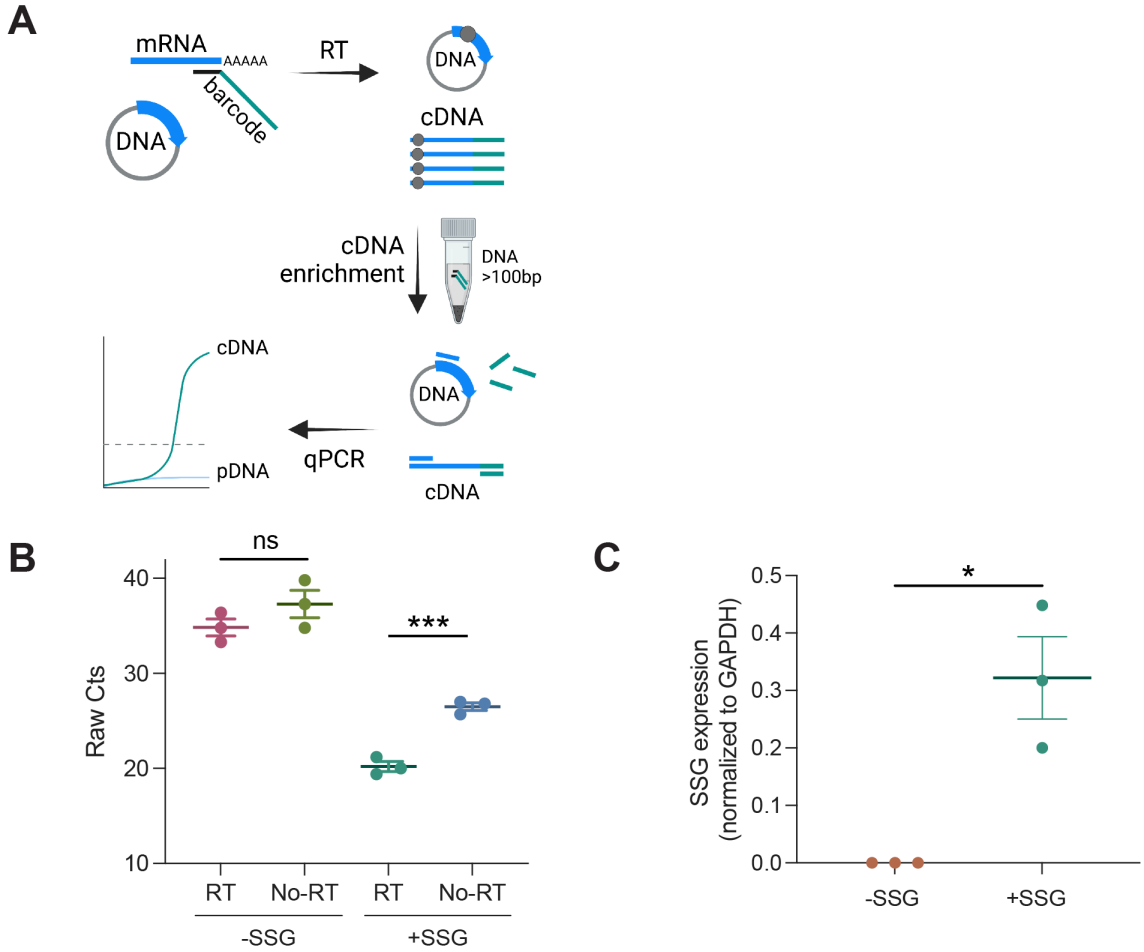
434 **SUPPLEMENTAL FIGURES**



435

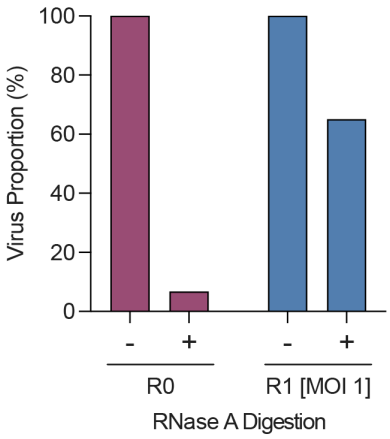
436 **Supplementary Figure 1**

437



438

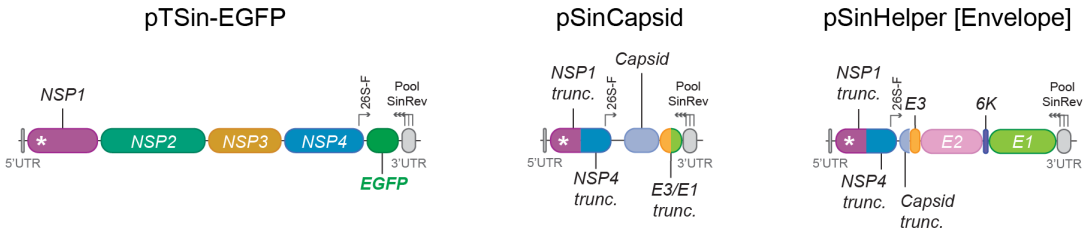
439 **Supplementary Figure 2**



440

441 **Supplementary Figure 3**

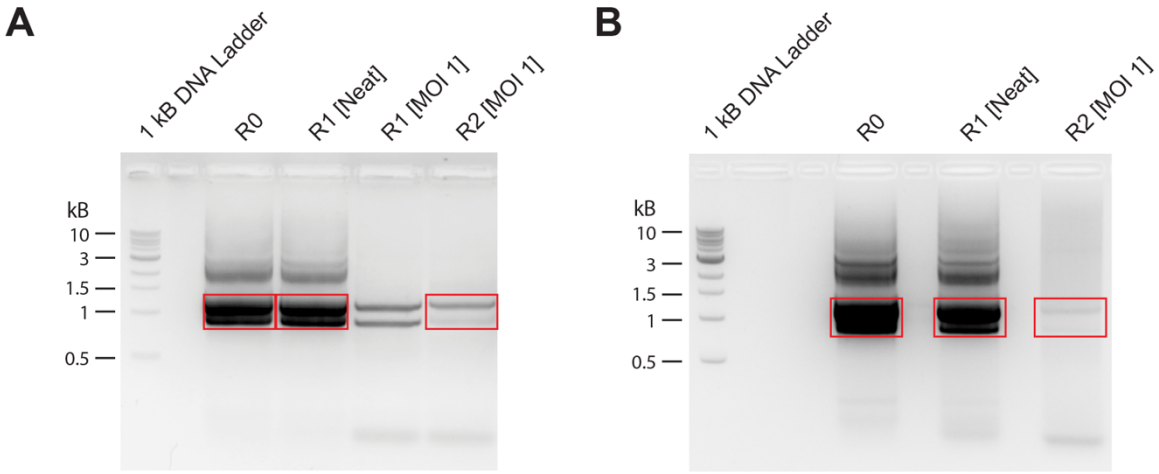
442



443

444 **Supplementary Figure 4**

445



446

447

448 **Supplementary Figure 5**

449 **KEY RESOURCES TABLE:**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
NEB® 10-beta Competent *E. coli* (High Efficiency)	NEB	C3019
Chemicals, Peptides, and Recombinant Proteins		
Ampicillin sodium salt	Sigma-Aldrich	A9518
RNase A	Macherey-Nagel	740505.50
Critical Commercial Assays		
Dual-Glo® Luciferase Assay Kit	Promega	E2940
FavorPrep™ Blood/Cultured Cell Total RNA Mini Kit	Fisher Biotec	FABRK 001-1
iScript™ Select cDNA Synthesis Kit	Bio-Rad Laboratories	1708897
ISOLATE II PCR and Gel Kit	Bioline	BIO-52060
ISOLATE II Plasmid Mini Kit	Bioline	BIO-52057
Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-LSK109
MagMAX™ Viral RNA Isolation Kit	ThermoFisher	AM1939
mMESSAGE mMACHINE™ SP6 Transcription Kit	ThermoFisher	AM1340
NEBNext® Companion Module for ONT Ligation Sequencing	NEB	E7180
NEBuilder® HiFi DNA Assembly Master Mix	NEB	E2621
PureYield™ Plasmid Maxiprep System	Promega	A2393

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Qubit™ dsDNA HS Assay Kit	ThermoFisher	Q32851
Qubit™ RNA BR Assay Kit	ThermoFisher	Q10210
SuperScript™ IV One-Step RT-PCR System	ThermoFisher	12594025
SYBR™ Select Master Mix	ThermoFisher	4472908
TaqMan™ Fast Virus 1-Step Master Mix	ThermoFisher	4444434
*Trans*IT®-2020 Transfection Reagent	Mirus Bio	MIR5400
Velocity DNA Polymerase	Bioline	BIO-21099
Deposited Data		
European Nucleotide Archive	This paper	PRJEB47639
Experimental Models: Cell Lines		
Hamster: BHK-21 [C-13]	ATCC	CCL-10
Software and Algorithms		
EPI2ME Desktop Agent (Version 3.3.0.1031)	Oxford Nanopore Technologies	https://www.nanoporetech.com
GraphPad Prism (Version 9.2.0)	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Guppy (Version 4.5.2)	Oxford Nanopore Technologies	https://www.nanoporetech.com
MinKNOW (Version 4.2.8)	Oxford Nanopore Technologies	https://www.nanoporetech.com
SnapGene® (Version 5.3.2)	Insightful Science	https://www.snapgene.com

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UMIC-seq	GitHub	https://github.com/fhlab/UMIC-seq
Zen 2 Blue Edition (Version 2.0.0.0)	Zeiss	https://www.zeiss.com
Recombinant DNA		
See Supplementary File 1 - Plasmid List. Plasmid maps are available upon request.	N/A	N/A
Oligonucleotides		
See Supplementary File 2 - Primer List.	N/A	N/A
Other		
0.45 µm filter	Merck Millipore	SLHV033RS
6-well plate	Corning	3516
96-well plate	Corning	3596
96-well plate, black-bottom	Interpath	655209
AMPure XP Magnetic Beads	Beckman Coulter	A63881
DMEM, high glucose	ThermoFisher	11965118
DPBS	Sigma-Aldrich	D8537
Fetal Bovine Serum (HyClone), AU origin	Cytiva Life Sciences	SH30084.03
Flongle Flow Cell (R9.4.1)	Oxford Nanopore Technologies	FLO-FLG001
LB Agar	ThermoFisher	22700025
LB Broth	ThermoFisher	12795-084

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MEM α with nucleosides	ThermoFisher	32571101
Opti-MEM Reduced Serum Medium (GlutaMAX Supplement)	ThermoFisher	51985034
RNaseOUT Recombinant Ribonuclease Inhibitor	ThermoFisher	10777019
Tryptose Phosphate Broth	ThermoFisher	CM0283B
XbaI	NEB	R0145

450

REFERENCES

Berman, Chet M., Louis J. Papa 3rd, Samuel J. Hendel, Christopher L. Moore, Patreece H. Suen, Alexander F. Weickhardt, Ngoc-Duc Doan, et al. 2018. "An Adaptable Platform for Directed Evolution in Human Cells." *Journal of the American Chemical Society* 140 (51): 18093–103.

Ekstrom, Marta H., and David A. Dean. 2011. "The Amaxa Nucleofector System and Standard Electroporation: A Comparison of the Efficiency of DNA Nuclear Localization." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 19 (May): S269–70.

English, Justin G., Reid H. J. Olsen, Katherine Lansu, Michael Patel, Karoline White, Adam S. Cockrell, Darshan Singh, Ryan T. Strachan, Daniel Wacker, and Bryan L. Roth. 2019. "VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells." *Cell* 178 (3): 748–61.e17.

Fayzulin, Rafik, Rodion Gorchakov, Olga Petrakova, Evgenia Volkova, and Ilya Frolov. 2005. "Sindbis Virus with a Tricomponent Genome." *Journal of Virology* 79 (1): 637–43.

Shapiro, Jillian S., Andrew Varble, Alissa M. Pham, and Benjamin R. Tenoever. 2010. "Noncanonical Cytoplasmic Processing of Viral microRNAs." *RNA* 16 (11): 2068–74.

Thomas, John M., William B. Klimstra, Kate D. Ryman, and Hans W. Heidner. 2003. "Sindbis Virus Vectors Designed to Express a Foreign Protein as a Cleavable Component of the Viral Structural Polyprotein." *Journal of Virology* 77 (10): 5598–5606.

Weiss, B., U. Geigenmüller-Gnirke, and S. Schlesinger. 1994. "Interactions between Sindbis Virus RNAs and a 68 Amino Acid Derivative of the Viral Capsid Protein Further Defines the Capsid Binding Site." *Nucleic Acids Research* 22 (5): 780–86.