"Cheater" particles render the VEGAS platform unsuitable 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 the VEGAS system has major limitations precluding its use for directed evolution. The primary technical issue with the VEGAS system is an immediate contamination with "cheater" particles that bypass directed evolution circuits. By sequencing we find these cheater particles contain Sindbis structural genes instead of the intended directed evolution target transgene. These cheaters outcompete the VEGAS transgenes within 2 rounds of transduction but cannot themselves activate synthetic circuits that drive expression of Sindbis structural genes, preventing directed evolution campaigns. Similar results have been obtained in independent labs. Taken together, the VEGAS system does not work as described and, without significant redesign to suppress cheaters, cannot be used for mammalian directed evolution campaigns.

43 **INTRODUCTION**

44 A new generation of mammalian directed evolution systems have been reported, which exploit
45 viral life cycles to couple mutagenesis of a virally packaged transgene-of-interest with selection
46 for higher fitness variants (Berman et al. 2018; English et al. 2019). In these systems, essential
47 structural components are removed from the viral genome and placed under the control of
48 synthetic gene expression circuits within the host cell and regulated by the transgene-of-
49 interest. Selection operates on natural mutations arising in the transgene during error-prone
50 viral replication that then impact the expression of the transgene-regulated viral structural
51 genes. Each host cell should be infected with a single virus particle each round (i.e. MOI ≤ 1)
52 so that only the mutation(s) conferring a selective advantage are enriched. Variants that more
53 efficiently drive the synthetic circuit are then overrepresented in the packaged viral particles
54 used for further rounds of evolution (Berman et al. 2018; English et al. 2019). The English, et
55 al. (2019) viral evolution of genetically actuating sequences (VEGAS) system offered a
56 simplified platform for simultaneous mutagenesis and selection using Sindbis virus (SINV).
57 The VEGAS system was designed to link the activity of a transgene to expression of the Sindbis
58 Structural Genome (SSG; comprising Capsid, E3, E2, 6K and E1 genes) (English et al. 2019).
59 For the VEGAS system to support directed evolution, a SINV particle must contain the target
60 transgene, and after transduction should drive expression of the SSG in the next round of
61 evolution.

62

63 Here, we find that the VEGAS system does not support viral propagation when performed as
64 described. Similar to English et al. and others (Shapiro et al. 2010; Fayzulin et al. 2005), we
65 found Sindbis virus can efficiently package a transgene and transduce cells, however in the
66 context of the VEGAS system, these particles do not then support the multiple rounds of
67 propagation required for directed evolution. Mechanistically, particles generated using the
68 VEGAS system are immediately contaminated with "cheaters" that contain Sindbis structural
69 genes. These cheaters flood the system, dilute the target transgene, and do not allow directed
70 evolution campaigns.

71

72

RESULTS AND DISCUSSION

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)). To generate Sindbis particles we electroporated BHK-21 cultures with pSinHelper, pSinCapsid and pTSin-EGFP seed mRNA (**Figure S1A)**. After 24 h, we detected ~10^9 genome copies (gc)/mL in the supernatant (Round (R)0; **Figure 1A**) and observed bright EGFP fluorescence in most packaging cells (R0; **Figure 1B**). English et al. initiated directed evolution campaigns by transducing target BHK-21 cells with R0 virus at a calculated multiplicity of infection (MOI) of 1 gc/cell (English et al. 2019). Thus, we next transduced target BHK-21 cells in the context of constitutively expressed Sindbis structural genome (CMV-SSG plasmid; +SSG) or control DNA at a calculated MOI of 1 (R1; **Figure 1A**). Importantly, R1 SINV particles were inefficiently generated ($<10^6$, >1000x lower titer than R0) and titers were equivalent in samples that received SSG or control DNA (-SSG; $7.35 \times 10^6 \pm 2.30 \times 10^5$ gc/mL; +SSG, $7.33 \pm 4.77 \times 10^5$ gc/mL; **Figure 1A**), indicating this system has no dependency on SSG. Moreover, based on GFP expression, SINV particles did not efficiently transduce target cells (**Figure 1B**), and both SSG dependency and transgene delivery are essential requirements for VEGAS directed evolution campaigns.

One possible cause for the low viral titers observed at R1 is that transfected host cells failed to express the Sindbis structural genes provided via the CMV-SSG plasmid. Thus, we next examined host cell expression of CMV-SSG at R1. To avoid the possibility of transfected plasmids contributing to high background signal during qPCR, we developed a barcoding method to discriminate between plasmid DNA (pDNA) and expressed mRNA (**Figure S2A**). Using this approach, we confirmed that the CMV-SSG plasmid was highly expressed (>30% of GAPDH, **Figure S2B,C**).

We next addressed possible explanations for the low GFP expression when transduced at an MOI of 1 (**Figure 1B**). We suspected that residual seed mRNA in the supernatant following R0 electroporation could contaminate the R0 output and inflate titers. To test this, we degraded unpackaged seed mRNA by treating R0 and R1 supernatants with RNase A prior to titration. Indeed, RNase A treatment of R0 samples decreased the measured virus titer by 93%, however R1 samples were much less RNase-sensitive (**Figure S3**). This differential sensitivity indicated

105 that residual unpackaged seed RNA was present in the output from R0 confounding accurate
106 MOI calculations.

107

108 To determine if inaccurate MOI calculations can explain why the VEGAS system does not
109 support viral propagation after R0, we added undiluted (neat) R0 virus (SINV-EGFP) to CMV-
110 SSG-expressing cells. Transduction with neat virus resulted in 5-10% GFP-positive cells at R1
111 (arrowheads, **Figure 1D**), indicating that Sindbis packaging and delivery of the target transgene
112 was successful. We then transduced R2 target cells with SINV-EGFP particles at an MOI of 1,
113 however, at R2 we did not observe GFP-positive cells (**Figure 1D**), and output from this
114 transduction was low and not dependent on SSG (**Figure 1C**). These data suggest that
115 bypassing unreliable R1 MOI calculations with neat transductions (and then controlling MOI
116 beyond R1) was still not sufficient to support viral propagation at subsequent rounds of
117 transduction. Overall, the VEGAS system does not support SINV propagation at an MOI of 1,
118 precluding its use for directed evolution campaigns.

119

120 To investigate the integrity of the VEGAS system over successive rounds of transduction we
121 performed long-read nanopore sequencing of viral RNA (**Figure 1E**). Using the approach for
122 transgene recovery as described (English et al. 2019), all reads should map to the transgene
123 subjected to directed evolution. Surprisingly, we found only 53.4% of reads mapped to the
124 EGFP transgene in the initial packaged virus (R0; **Figure 1E**) and transgene inclusion was
125 reduced to 30.3% by R2. The remaining reads were from cheater particles that contain envelope
126 and capsid sequences derived from seed RNA provided at R0. We note that the *in vitro*-
127 transcribed VEGAS seed mRNAs (pSinHelper and pSinCapsid) retain components of the
128 SINV NSP1 packaging signal and replication sequences (Weiss, Geigenmüller-Gnirke, and
129 Schlesinger 1994) (**Figure S4**), which would allow seed mRNA to compete with the SINV-
130 EGFP transgene for packaging. These data show that the VEGAS system preferentially
131 packages SINV structural elements, and these cheater particles then outcompete VEGAS
132 transgenes and block directed evolution efforts.

133

134 Previous work has shown Sindbis virus can retain a neutral GFP transgene for at least six viral
135 passages at low MOIs (0.1 PFU/cell) (Thomas et al. 2003), however it is possible that in the

136 context of a split VEGAS system, selective pressure for a transgene is required to maintain it
137 across rounds of propagation. To test this possibility, we built a simple synthetic circuit using
138 the DNA-binding domain of serum response factor (SRF, fused to VP64) to drive transgene
139 expression under the control of the serum response element (SRE). We first validated this
140 circuit using SRF-NLS-VP64/SRE-dependent expression of a firefly luciferase reporter
141 (SRE_LUC), which drove luciferase (**Figure 2A)** at levels comparable to (English et al. 2019)
142 (see English et al, **Figure S3B**). Next, we subcloned SRF-NLS-VP64 into the VEGAS pTSin
143 vector and packaged it in a 1:1 ratio with pTSin-EGFP to provide an evolutionarily neutral
144 competitor. Under selection, the EGFP transgene should drop out while the higher fitness SRF-
145 NLS-VP64-containing viruses, capable of driving expression of the Sindbis structural genome,
146 should rapidly dominate the culture. As predicted, the proportion of GFP-positive cells
147 decreased with serial passage (**Figure 2B),** however, this did not translate into efficient SRF-
148 NLS-VP64 viral production at R2 (**Figure 2C**).

149

150 Importantly, the circuit-inducing SRF-NLS-VP64 transgene did not outcompete the control
151 EGFP transgene (3.2% SRF-NLS-VP64 vs 21.7% GFP by R2), and both transgenes were
152 outcompeted by cheater sequences containing Sindbis structural elements (75.1%, **Figure 2D**).
153 Together, these data show that even under selective pressure, the VEGAS system does not
154 support productive rounds of replication required for directed evolution campaigns. Instead,
155 VEGAS campaigns are flooded with cheater particles containing Sindbis structural genes that
156 prevent the circuit-driven selection required for directed evolution to occur.

157

158 While it is possible that undocumented methodological subtleties could explain why VEGAS
159 Sindbis particles fail to propagate across rounds of transduction, we have received feedback
160 from English et al. on the results described here and have thus far not identified any. To
161 generate the Sindbis particles required to initiate VEGAS campaigns, English *et al.* used the
162 NEON electroporator, whereas we used the Amaxa electroporator which also delivers RNA to
163 the cytosol (Ekstrom and Dean 2011) and efficiently generates functional Sindbis particles
164 (Shapiro et al. 2010). Importantly, using a different electroporator does not explain our inability
165 to run VEGAS directed evolution campaigns, since both machines produced Sindbis particles
166 that transduced GFP to R1 target cells with equivalent efficiency (~5-10%). Instead, our data

167 suggests that the VEGAS system is non-functional due to circuit-independent cheater particles
168 that dominate the system.

169

170 Our experiments involved VEGAS components generated and made publicly available by
171 English *et al.* when provided (Addgene plasmids #127692–127695). However, we could not
172 obtain the CMV-SSG plasmid from English *et al.* and generated this component ourselves,
173 which we validated and used in **Figure 1**. Moreover, English *et al.* similarly did not provide
174 the VEGAS circuits used in their original work, and thus we generated our own serum response
175 factor circuit used in **Figure 2**. Again, while we confirmed this circuit is functional and could
176 drive expression from the SRE comparable to the circuit used by English *et al.*, we found this
177 circuit could not support viral propagation past R2. A similar inability to use the VEGAS
178 system has been observed independently across our separate laboratories over the three years
179 since the platform was first described.

180

181 Importantly, Sindbis virus is pathogenic to humans and has the potential to cause meningitis
182 (Meno et al. 2021; Griffin 2005). Extreme caution should be used when trying to implement
183 the published VEGAS system in a BSL2 environment since a single recombination event
184 between the transgenic SINV genome and the host cell SSG expression plasmid (which share
185 homologous sequences) could produce a replication-competent virus. In our hands we tested
186 for and did not detect replication-competent virus.

187

188 While we believe there is a future for Sindbis-based mammalian directed evolution systems,
189 this will require substantial efforts to address critical technical challenges that limit VEGAS
190 utility and, optimally, any repairs or upgrades should be confirmed by independent groups. In
191 summary, although we confirm that packaging and transduction of SINV particles efficiently
192 delivers transgenes to fresh host cells as previously documented (English et al. 2019; Fayzulin
193 et al. 2005; Shapiro et al. 2010), we conclude that due to the prevalence of cheater particles
194 that carry Sindbis structural components instead of target transgenes, the VEGAS system is not
195 suitable for use as a mammalian directed evolution platform.

196

197 **METHODS**

198 **Cell Culture**

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

206 **Molecular Biology and Plasmid Construction**

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

221 **mRNA synthesis**

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

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

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

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

237 **Viral Titration**

238 Viral supernatants were titrated as outlined by English and colleagues (English et al. 2019).
239 Following collection and clarification, undiluted SINV-containing supernatants were
240 combined with the TaqMan™ Fast Virus 1-Step Master Mix (ThermoFisher, #4444434) and
241 the appropriate primer-probe sets in technical triplicates. Plates were run on a QuantStudio™
242 7 Flex Real-Time PCR System (ThermoFisher). As described previously, serially diluted *in*
243 *vitro*-transcribed pTSin mRNA was used as the standard curve (ranging between 10^3-10^7
244 genome copies (gc) per mL) (English et al. 2019). Standard curves were used to determine viral
245 titers in gc/mL. Primer sequences are listed in Supplementary File 2 – Primer List.

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

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

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

262 **RNase Digestion**

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

269 **Barcoded qPCR**

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

281 **Luciferase Assay**

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

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

290 **Transgene Isolation**

291 Sextuplicate viral supernatants from each round of viral replication were pooled in equal ratios.
292 Viral RNA from 400 µl aliquots was isolated with the MagMAX™ Viral RNA Isolation Kit
293 (ThermoFisher, #AM1939) as per the manufacturer's recommendations. Transgene sequences,
294 situated between nsP4 and the viral 3' untranslated region (UTR) were reverse-transcribed and
295 PCR-amplified using the SuperScript™ IV One-Step RT-PCR System (ThermoFisher,
296 #12594025) and the 26S-F primer and pooled SinRev primer. Primer sequences are listed in
297 Supplementary File 2 – Primer List. Amplicons of the appropriate size range for the transgene
298 of interest were then gel extracted for nanopore sequencing using the ISOLATE II PCR and
299 Gel Kit (Bioline, #BIO-52060) (**Figure S5**).

300 **Nanopore Sequencing**

301 *Sample Processing*

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

309 *Sequencing and Basecalling*

310 Up to 40 fmol of DNA library was loaded onto ONT Flongle flow cells (R9.4.1) in a MinION
311 Mk1B Sequencer fitted with a Flongle Adapter (ONT). Sequencing was performed using
312 MinKNOW (ONT, version 4.2.8) under default parameters and the following specified inputs:
313 kit used, SQK-LSK109; 0.5 hours between MUX scans; basecalling, disabled. A minimum of
314 120,000 raw reads were obtained for each sample. Raw FAST5 files were processed using

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

317 *Alignment*

318 Quality-filtered base-called reads (in FASTQ format) were processed using EPI2ME Desktop
319 Agent (ONT, version 3.3.0.1031). Reference files, consisting of complete transgene-coding
320 sequences or helper- and capsid-coding sequences, were uploaded using the Fasta Reference
321 Upload workflow (v2021.07.15). For each sample, 100,000 reads were aligned to reference
322 files using the Fastq Custom Alignment workflow (v2021.03.25) using default parameters.
323

324 **Microscopy**

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

329 **Statistics**

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

336

ACKNOWLEDGMENTS

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

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.

359 **FIGURE LEGENDS**

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

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

375

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

377 A. Activation of an SRE-regulated luciferase reporter (SRE_LUC) by SRF-NLS-VP64.
378 Error bars represent mean ± SEM (N = 3, with 3 technical replicates). A Brown-
379 Forsythe and Welch ANOVA test was used to determine statistical significance. ****
380 $p < 0.0001$.
381 B. Phase contrast and EGFP fluorescence images of BHK-21 cells at R0-R2 for an SRF-
382 NLS-VP64/SRE_SSG circuit. Scale bars represent 200 µm.
383 C. Viral titers of an SRF-NLS-VP64 transgene-carrying virus packaged in a 1:1 ratio with
384 an EGFP-carrying virus (N = 6).
385 D. Nanopore sequencing of viral RNA isolated from pooled samples (N = 6). Reads
386 (~100,000 per sample) were aligned to viral reference sequences.
387
388
389

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.

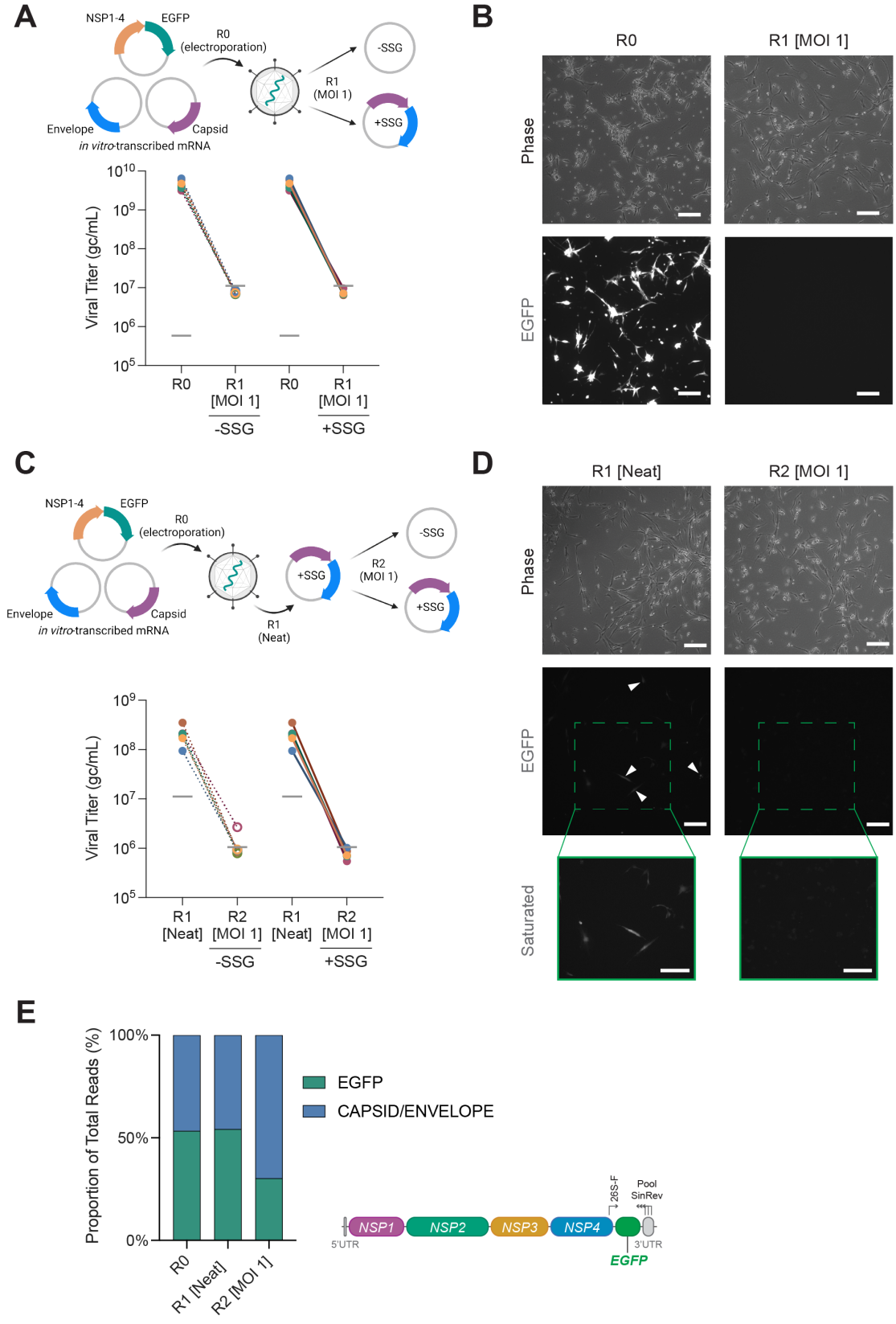
Schematic showing the position of the SINV packaging signal within VEGAS seed mRNAs (asterisks). Genes and truncations drawn to scale.

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.

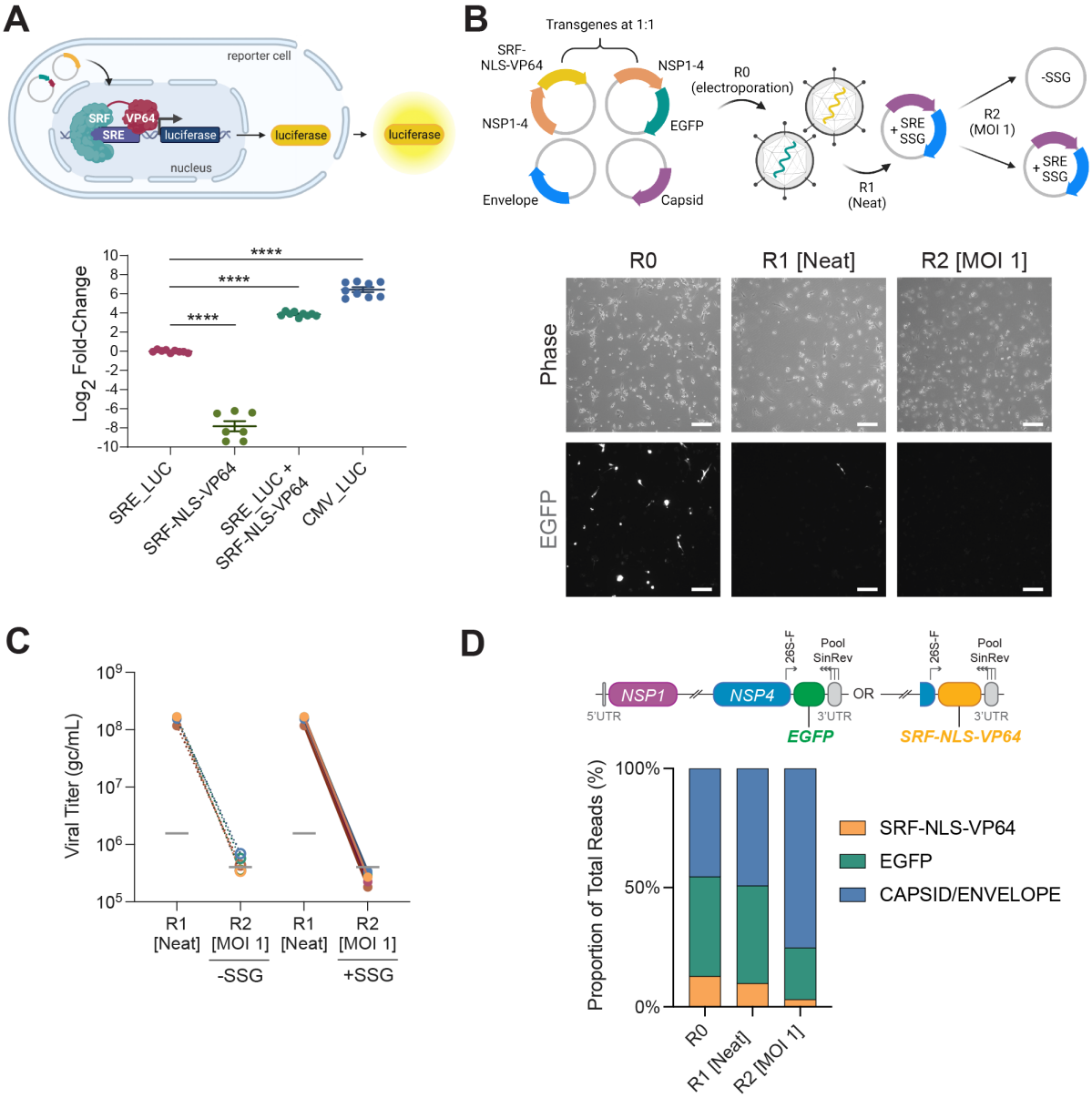
MAIN FIGURES

Figure 1



437 **Figure 2**

438



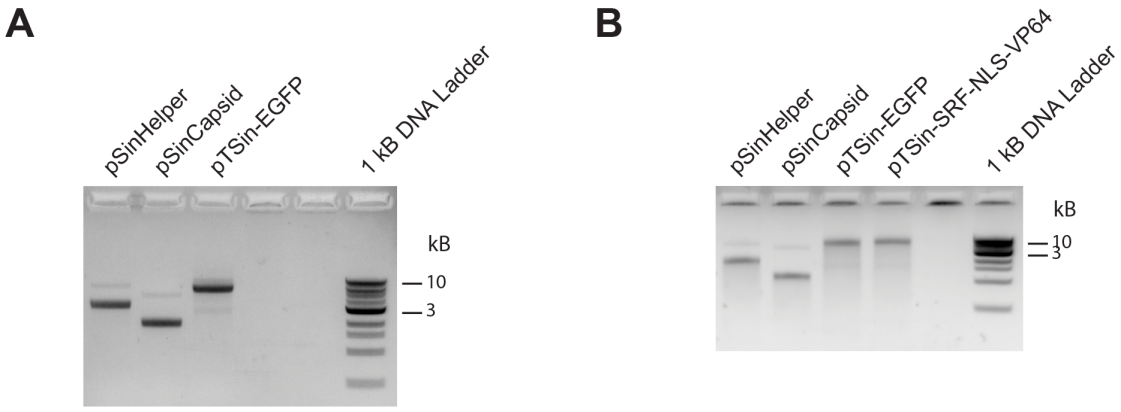
439

440 **SUPPLEMENTAL FIGURES**

441

442 **Supplementary Figure 1**

443



444

445

446

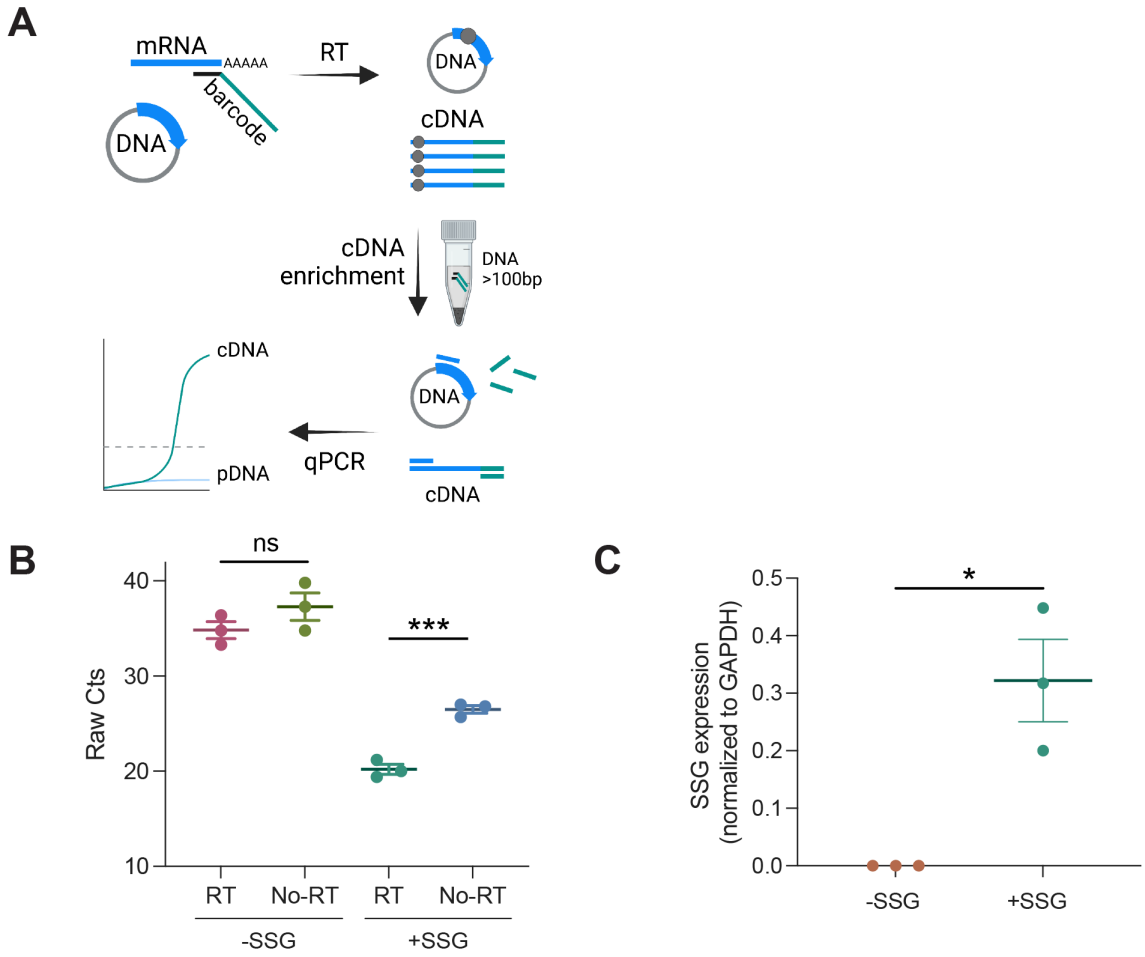
447

448

449

450 **Supplementary Figure 2**

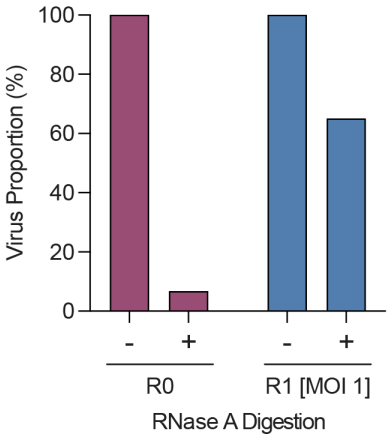
451



452

453 **Supplementary Figure 3**

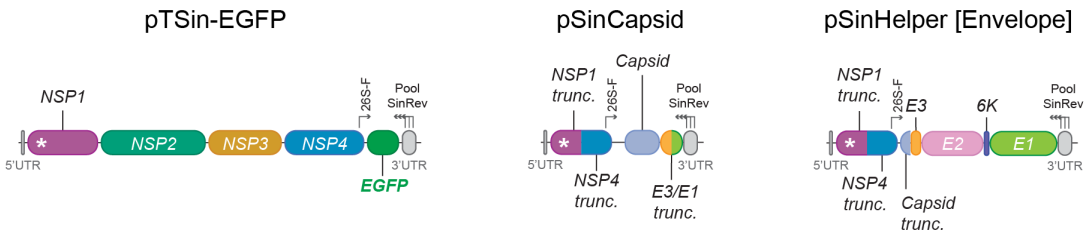
454



455
456
457
458
459

460 **Supplementary Figure 4**

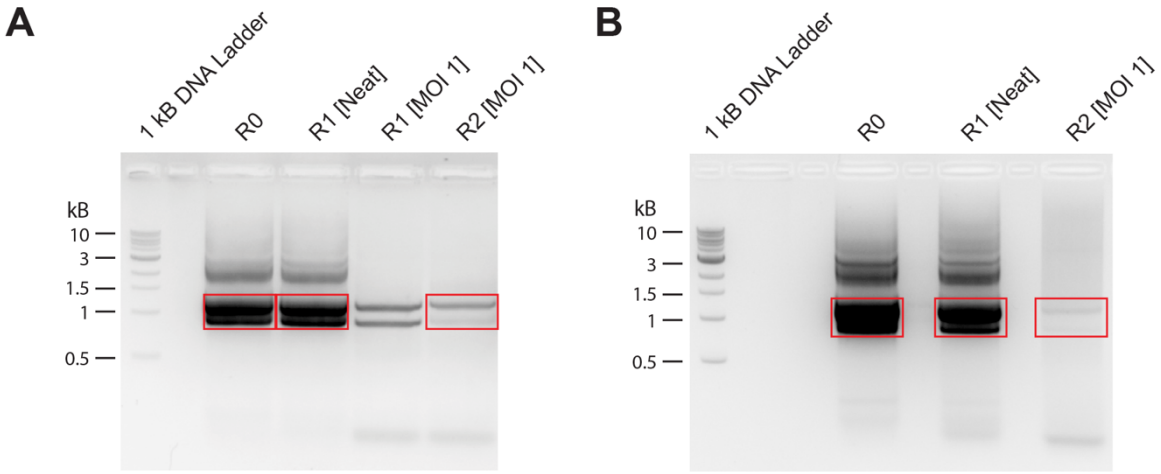
461



462
463
464
465
466
467

468 **Supplementary Figure 5**

469



470

471 **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

472

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.

Griffin, D. E. 2005. "Neuronal Cell Death in Alphavirus Encephalomyelitis." *Current Topics in Microbiology and Immunology* 289: 57–77.

Meno, Kgothatso, Clarence Yah, Adriano Mendes, and Marietjie Venter. 2021. "Incidence of Sindbis Virus in Hospitalized Patients With Acute Fevers of Unknown Cause in South Africa, 2019-2020." *Frontiers in Microbiology* 12: 798810.

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.