

# Attenuation of RNA viruses by redirecting their evolution in sequence space

Gonzalo Moratorio<sup>1</sup>, Rasmus Henningsson<sup>1,2,3†</sup>, Cyril Barbezange<sup>1†</sup>, Lucia Carrau<sup>1,4</sup>, Antonio V. Bordería<sup>1,2</sup>, Hervé Blanc<sup>1</sup>, Stephanie Beaucourt<sup>1</sup>, Enzo Z. Poirier<sup>1,4</sup>, Thomas Vallet<sup>1</sup>, Jeremy Boussier<sup>2,5,6</sup>, Bryan C. Mounce<sup>1</sup>, Magnus Fontes<sup>2,3</sup> and Marco Vignuzzi<sup>1\*</sup>

**RNA viruses pose serious threats to human health. Their success relies on their capacity to generate genetic variability and, consequently, on their adaptive potential. We describe a strategy to attenuate RNA viruses by altering their evolutionary potential. We rationally altered the genomes of Coxsackie B3 and influenza A viruses to redirect their evolutionary trajectories towards detrimental regions in sequence space. Specifically, viral genomes were engineered to harbour more serine and leucine codons with nonsense mutation targets: codons that could generate Stop mutations after a single nucleotide substitution. Indeed, these viruses generated more Stop mutations both *in vitro* and *in vivo*, accompanied by significant losses in viral fitness. *In vivo*, the viruses were attenuated, generated high levels of neutralizing antibodies and protected against lethal challenge. Our study demonstrates that cornering viruses in 'risky' areas of sequence space may be implemented as a broad-spectrum vaccine strategy against RNA viruses.**

Vaccines remain the most successful means of controlling morbidity and mortality caused by RNA viruses, yet relatively few viral vaccines exist. In recent years, several severe outbreaks have occurred: chikungunya and Zika viruses in the Americas, coronaviruses in the Middle East, and Ebola virus in West Africa. Consequently, there is a need for rationally designed and broadly applicable vaccine strategies. Given their high mutation rates, large population sizes and short generation times, RNA viruses can evolve rapidly, and strategies to control RNA viruses should take into account this adaptive potential. Due to their mutation rates, RNA viruses generate networks of closely related genetic variants, linked through mutation<sup>1</sup>, that allow them to escape from selective pressures and adapt to different environments<sup>2</sup>. Ultradeep characterization of single nucleotide polymorphisms of viral populations reveals thousands of variants<sup>3</sup> heterogeneously distributed throughout the genome. This 'genetic architecture' suggests that certain mutations might be more or less accessible depending on the original nucleotide or codon, thereby defining different mutational neighbourhoods within the same sequence space<sup>4</sup>. In evolutionary biology, sequence space refers to every combination of a given sequence and theoretically is a vast multidimensional hypercube connecting all possible combinations. The localization of a virus population in sequence space, defined by its starting sequence, should then determine which mutational neighbourhoods are accessible. It is thus proposed that access to certain neighbourhoods will determine the potential of reaching beneficial mutations to facilitate adaptation<sup>4,5</sup>.

However, even the best of mutational neighbourhoods is not without risk in terms of impact of mutation on fitness. Most studies addressing how organisms explore sequence space are theoretical and tested *in silico* using digital organisms<sup>6,7</sup>. Limited empirical data support the notion that the 'viable' sequence space is significantly smaller than the theoretical. For RNA viruses, most

mutations are deleterious<sup>8,9</sup>, and up to 40% of non-synonymous changes are lethal<sup>3,10</sup>. This is further evidenced in lethal mutagenesis, where antiviral treatment with mutagenic compounds leads to extinction<sup>11,12</sup>. Consequently, viruses have probably established a balance between generating beneficial mutations and tolerating detrimental ones. This trait, termed 'mutational robustness', is defined as 'phenotypic conservation in light of genetic variation'<sup>13</sup>. Given the biological constraints that limit the viable sequence space occupied by RNA viruses, we asked whether their capacity to explore sequence space (and, ultimately, their fitness) could be restricted by placing them closer to detrimental mutational neighbourhoods.

To address this question, we genetically engineered Coxsackie B3 and influenza A viruses to present altered synonymous codon architectures that would change their starting positions in sequence space and therefore limit their access to mutational neighbourhoods. Specifically, we rewired leucine and serine codons to constrain the expansion of viral populations towards detrimental mutational neighbourhoods, where the product of mutation would favour nonsense mutation targets resulting in Stop mutations. These engineered viruses were attenuated *in vivo*, with an increased number of Stop mutations in viral progeny. Animals immunized with these vaccine candidates were protected against lethal infection. We thus show that RNA viruses can be rationally attenuated by redirecting their evolutionary trajectories towards detrimental areas of sequence space.

## Results

**Reprogramming a viral genome to have enhanced proclivity for non-sense mutations.** Our goal was to assess the effect of shifting the location of a virus in sequence space towards less 'hospitable' regions that increase its propensity to generate non-sense mutations. However, altering location in sequence space requires changes in nucleotide sequence, which can result in confounding factors such

<sup>1</sup>Viral Populations and Pathogenesis Unit, Institut Pasteur, CNRS UMR 3569, 28 rue du Dr. Roux, 75724 Paris cedex 15, France. <sup>2</sup>International Group for Data Analysis, Institut Pasteur, C3BI, USR 3756 IP CNRS, 28 rue du Dr. Roux, 75724 Paris cedex 15, France. <sup>3</sup>Centre for Mathematical Sciences, Lund University, 22100 Lund, Sweden. <sup>4</sup>Sorbonne Paris Cité, Université Paris Diderot, Cellule Pasteur, 75013 Paris, France. <sup>5</sup>Unité d'Immunobiologie des Cellules Dendritiques, Institut Pasteur, Inserm 1223, 25 rue du Dr. Roux, 75724 Paris cedex 15, Paris, France. <sup>6</sup>Ecole doctorale Frontières du vivant, Université Paris Diderot, 75013 Paris, France. <sup>†</sup>These authors contributed equally to this work. \*e-mail: [marco.vignuzzi@pasteur.fr](mailto:marco.vignuzzi@pasteur.fr)

as changes in the amino-acid sequence or RNA structure, or the introduction of nucleotide and codon biases<sup>14,15</sup>. To minimize these factors, we introduced only synonymous changes, so that viral proteins retained the same amino-acid sequence and the same functions as wild-type virus. Furthermore, we only changed the codons for two amino acids with the highest codon redundancy (leucine and serine) to limit the overall change in nucleotide sequence to less than 5% and to focus on codons on which mutations would have the greatest impact on viability. Among the Leu and Ser codons, we defined a category termed '1-to-Stop', because point mutations on these codons could result in Stop mutations (Fig. 1a). We also defined a 'NoStop' category, which in contrast requires two nucleotide changes to reach Stop mutations. For Coxsackie virus B3 (CVB3), we targeted the P1 structural protein-coding region (Fig. 1a), which lacks RNA structural elements required for replication and translation. Because this viral RNA is translated into a single polyprotein that is cleaved into individual viral proteins, a Stop mutation appearing in this region will inactivate the virus. We thus generated CVB3 in which all 117 Ser/Leu codons in P1 were synonymously changed to either 1-to-Stop or NoStop categories. To investigate if this strategy can be applied broadly, we also targeted a very different RNA virus, the influenza A virus, which has a segmented, negative sense genome. In this case, two genomic segments were targeted independently: the PA polymerase subunit protein (111 Ser/Leu codons) or the haemagglutinin glycoprotein HA (94 Ser/Leu codons) (Fig. 1a).

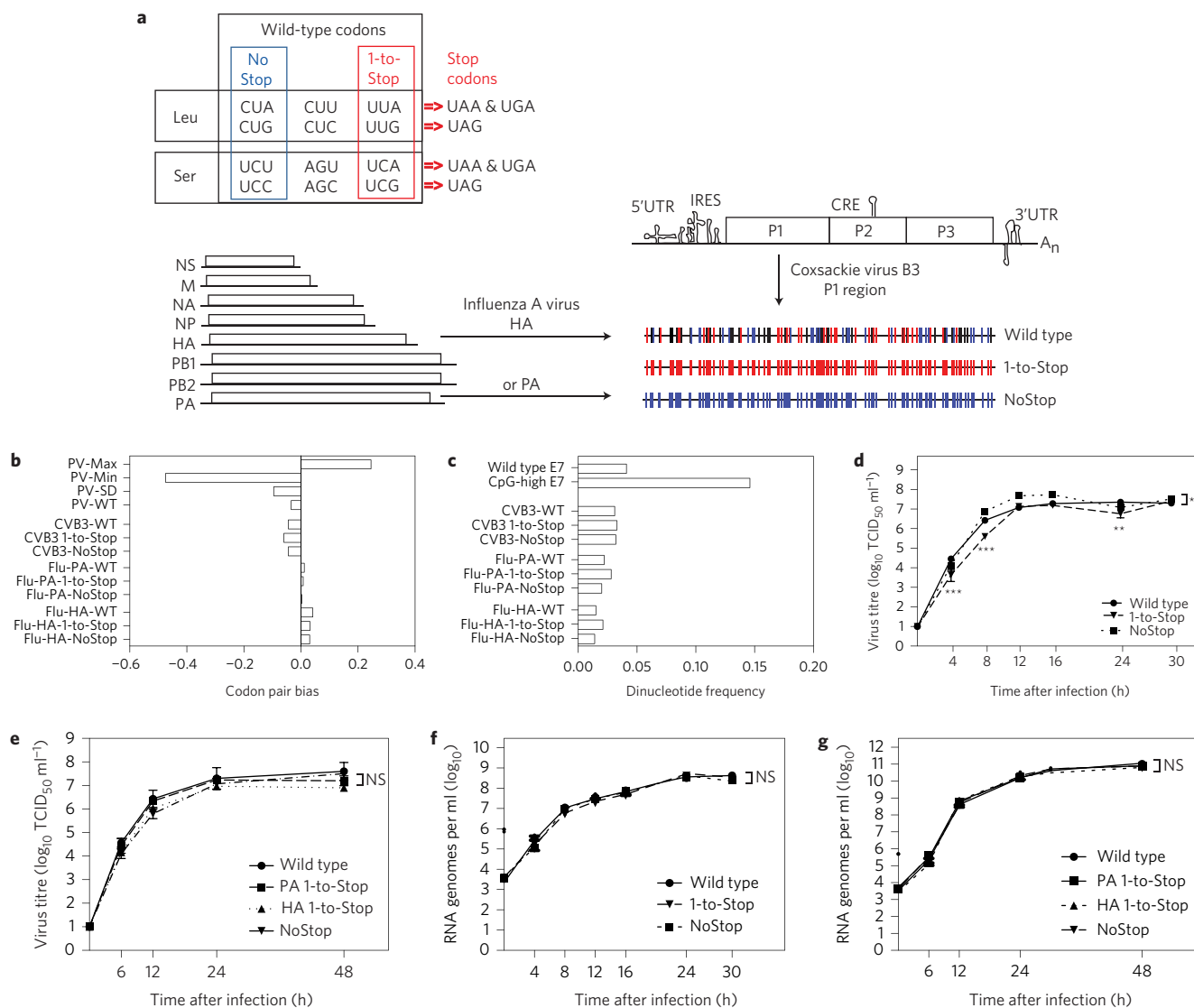
Large-scale codon reshuffling may attenuate viruses by introducing a bias in codon pairs that are under-represented in the human genome<sup>14</sup>, so we verified that this potentially confounding effect did not occur in our design. We compared our viruses (Fig. 1b) with those used by Coleman *et al.*<sup>14</sup> (PV-Min, PV-Max and PV-SD) to link codon pair bias and attenuation. These poliovirus (a related enterovirus) constructs were codon-shuffled in the same P1 region. The change in codon pair bias in our viruses was significantly less than for the PV-Min virus, which had 1,000-fold reduced replication<sup>14</sup>. Even PV-SD, which was not attenuated in their study, presented more bias than our constructs. The attenuation obtained from codon reshuffling can also result from dinucleotide bias, where higher CpG content decreases replication<sup>15</sup>. Compared to the dinucleotide frequency described for the highly attenuated echovirus 7 by Tulloch *et al.*<sup>15</sup> (CpG-high E7), no significant changes in CpG dinucleotide frequency were introduced in our 1-to-Stop and NoStop constructs (Fig. 1c). Subsequently, we investigated the production of infectious particles for these viruses at high (Fig. 1d,e) or low (Supplementary Fig. 1a,b) multiplicities of infection (MOI). All viruses grew to comparable final titres, although fewer infectious progeny were produced at some time points for the CVB3 1-to-Stop construct (Fig. 1d). To test whether this was due to defects in replication, we quantified RNA genome synthesis. All viruses generated the same amounts of RNA at every time point (Fig. 1f,g). We further confirmed these results in an *in vitro* RNA replication assay using replication complexes purified from infected cells, where yields for wild-type and 1-to-Stop viruses were equal (Supplementary Fig. 2). Thus, our data showed that, for both CVB3 and influenza A virus, the altered Leu/Ser codons had no discernible impact on RNA synthesis and replication kinetics. Finally, genetic and phenotypic stabilities were evaluated after 10 passages. No reversion in Ser/Leu altered positions was observed, and each virus retained its phenotype (growth titres, relative number of Stop mutations) (Supplementary Fig. 3).

**1-to-Stop viruses have lower fitness and are highly sensitive to mutation.** The relative fitness of wild-type, 1-to-Stop and NoStop viruses was measured under normal and mutagenic conditions

against a neutral, genetically marked competitor<sup>16,17</sup>. For CVB3 (Fig. 2a), five mutagenic conditions were used: three base analogues (ribavirin, 5-fluorouracil (5-FU) and 5-azacytidine (5-AZC)), amiloride (which perturbs intracellular Mg<sup>2+</sup> and Mn<sup>2+</sup> concentrations, essential cofactors of the viral polymerase<sup>18</sup>) and Mn<sup>2+</sup> itself, which increases the polymerase error rate. In all cases, the 1-to-Stop virus presented significantly lower fitness than wild-type CVB3, while the NoStop virus presented the same or higher fitness (Fig. 2a). Using an alternative method to evaluate viral fitness, we measured mean plaque size for each viral population treated with the three base analogues. The 1-to-Stop virus produced significantly smaller plaques, while the NoStop virus produced larger plaques (Fig. 2b). The progeny viruses obtained during mutagenic treatments were then deep-sequenced, and sequence data were computationally treated to reduce noise (see Methods) and mathematically modelled to estimate error (see Supplementary Section 'Mathematical assessment of background noise'). The number of reads presenting Stop mutations was then calculated. As expected, 1-to-Stop virus samples contained significantly more reads with Stop mutations than wild-type virus, whereas NoStop had significantly fewer (Fig. 2c). Finally, to further support that the 1-to-Stop viruses were sensitive to the increased mutational load on these codons, we coupled the 1-to-Stop virus with a high-fidelity polymerase<sup>19</sup>. This virus would have an intrinsically lower error rate to counter the extrinsically higher error resulting from mutagen treatment. Indeed, 1-to-Stop high-fidelity viruses recovered the wild-type virus phenotype (Supplementary Fig. 4).

For influenza A constructs, viruses were treated with three concentrations of ribavirin, 5-AZC or Mn<sup>2+</sup> (Fig. 3a–c, left panels). Under each mutagenic condition, the 1-to-Stop PA and HA viruses were more sensitive to mutation, presenting lower titres than the wild type and their NoStop counterparts. In fact, 1-to-Stop viruses exhibited up to 50-fold reduction in viral titres with ribavirin treatment (Fig. 3a, left) and a 100-fold decrease after high concentrations of Mn<sup>2+</sup> (Fig. 3c, left). We then quantified the number of Stop mutations in the mutagen-treated progeny viruses for each replicate and at each concentration. The 1-to-Stop PA populations presented a dose-dependent and significantly higher increase in the number of Stop mutations along the PA gene compared to wild-type virus (Fig. 3a–c, middle panels). As a control, the 1-to-Stop HA virus, which has a wild type-like PA sequence, did not present more Stop mutations in the PA gene. Instead, the 1-to-Stop HA virus presented more Stop mutations in the HA gene (Fig. 3a–c, right panels). Together, these data confirm for both viruses that relocalizing the position of an RNA virus in sequence space to increase its likelihood of generating non-sense mutations resulted in a higher sensitivity to increased mutational load.

**1-to-Stop viruses are attenuated *in vivo*.** To evaluate this attenuation strategy *in vivo*, mice were given a sublethal dose of wild-type, 1-to-Stop or NoStop CVB3, and virus titres were determined over seven days. Although the 1-to-Stop virus replicated with wild type-like kinetics during the first five days in most mice, it was no longer detectable in pancreata (Fig. 4a) and in most hearts (Fig. 4b) by day 7. Importantly, the NoStop CVB3 virus retained the same virulence phenotype as wild-type virus (Fig. 4a,b). By deep sequencing, we observed threefold more Stop mutations in 1-to-Stop virus in these tissues (Fig. 4c), as was observed in tissue culture (Fig. 2c,b and Supplementary Fig. 3c). The NoStop virus control, on the other hand, presented significantly fewer Stop mutations. To further evaluate the effect of mutation on each virus, we determined each population's specific infectivity (the ratio of total genomes versus infectious genomes produced) (Fig. 4d). Although all CVB3 viruses

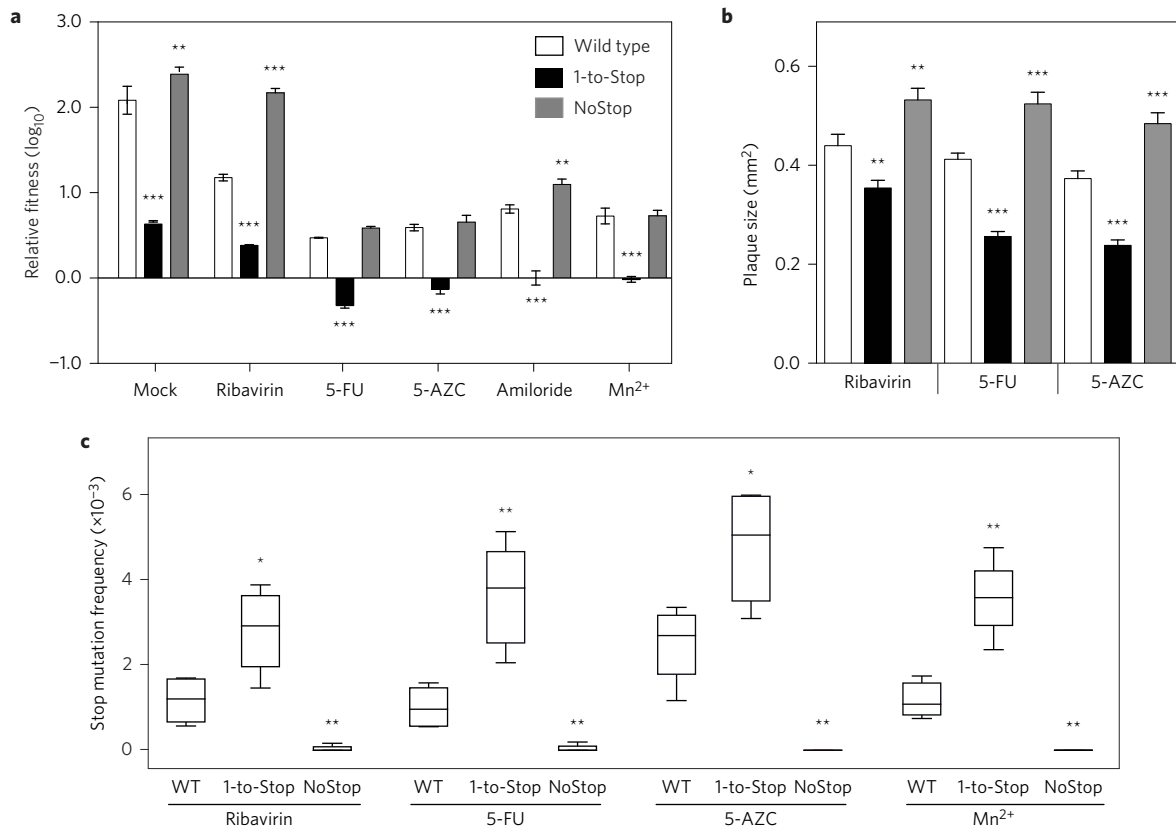


**Figure 1 | Construction of 1-to-Stop and NoStop RNA viruses.** **a**, Schematic representation of the six Leu and six Ser codons that are similarly represented in wild-type virus genomes, with codons belonging to the 1-to-Stop (red) and NoStop categories (blue) shown in colour. The stop mutations that can occur after a single point mutation are indicated. The Coxsackie virus B3 (CVB3) genome illustrates RNA structures required for replication (5'untranslated region (UTR), IRES, CRE and 3'UTR) and the single open reading frame encoding structural proteins (P1 region) and non-structural proteins (P2, P3 regions). The 117 Ser/Leu codons of the P1 region were altered to construct the 1-to-Stop viruses and NoStop viruses. The influenza A virus genome is composed of eight gene segments. The HA and PA genes were altered at 94 and 111 Ser/Leu codons, respectively, to generate the 1-to-Stop and NoStop viruses.  $A_n$ , polyA tail. **b**, Codon pair bias of wild-type, 1-to-Stop and NoStop CVB3 and influenza A viruses compared to previously published wild-type poliovirus (PV-WT) and constructs engineered to attenuate viruses through codon pair deoptimization: PV-Max, in which codon pairs over-represented in the human genome were maximized (no reduction in replication); PV-Min, in which codon pair bias was maximized by using codon pairs under-represented in the human genome (1,000-fold reduction in replication); and PV-SD, with randomly shuffled codons (no reduction in replication). **c**, CpG dinucleotide frequencies in wild-type, 1-to-Stop and NoStop CVB3 and influenza A viruses, relative to previously published wild-type Echovirus 7 (E7) and its high-CpG content construct shown to be attenuated. **d**, Production of infectious viral progeny over time of wild-type, 1-to-Stop and NoStop CVB3 viruses in HeLa cells infected at MOI = 1. **e**, Production of infectious viral progeny over time of wild-type, 1-to-Stop PA and HA, and NoStop PA influenza A viruses in MDCK cells infected at MOI = 1. **f**, Replication kinetics of wild-type, 1-to-Stop and NoStop CVB3 viruses in HeLa cells infected at MOI = 1. **g**, Replication kinetics of wild type, 1-to-Stop PA and HA, and NoStop PA influenza A viruses in MDCK cells infected at MOI = 1. In **d-g**, graphs show mean and s.e.m.;  $n = 3$  per group. NS, non-significant; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-way analysis of variance with a Bonferroni post-test, comparing wild type to each mutant).

presented high infectivity at three days of infection, a significantly larger proportion of 1-to-Stop genomes were non-infectious at seven days.

For influenza, mice were infected intranasally with  $1 \times 10^5$  plaque-forming units (p.f.u.) of wild type, 1-to-Stop PA or HA (PAS and HAS), or NoStop PA, and virus titres were determined in the lungs. Both 1-to-Stop viruses were attenuated (10- to 50-fold reduction in titres), with a larger decrease for the HA

construct (Fig. 4e). Once more, the NoStop virus control was as pathogenic as the wild-type virus. The number of Stop mutations present in progeny genomes in mouse lungs was quantified for the PA (Fig. 4f) and HA (Fig. 4g) genes. In both cases, the 1-to-Stop viruses presented a fourfold increase in Stop mutations compared with wild-type virus. Attenuation was further evaluated by monitoring daily weight loss (Fig. 4h). Mice infected with wild-type virus lost a mean of 12.5% of their weight by day 5,



**Figure 2 | The 1-to-Stop CVB3 virus is highly sensitive to mutation.** **a**, Relative fitness by a direct competition assay in the absence (mock) or presence of 200  $\mu$ M ribavirin, 5-fluorouracil (5-FU), 5-azacytidine (5-AZC), amiloride or 1 mM manganese ( $Mn^{2+}$ ). Wild type (white), 1-to-Stop (black) and NoStop (grey) competed against a neutral, marked wild-type CVB3. Graphs show mean and s.e.m.;  $n = 6$  per group.  $**P < 0.01$ ,  $***P < 0.001$  (two-tailed unpaired  $t$ -test with Bonferroni correction, comparing wild type to each mutant). **b**, Viruses were grown in the presence of 200  $\mu$ M of three different mutagens. Graphs show mean and s.e.m.;  $n = 1,000$  per group.  $**P < 0.01$ ,  $***P < 0.001$  (Mann-Whitney test). **c**, Frequency of Stop mutations observed in deep-sequencing reads from wild type and 1-to-Stop populations passed in 50  $\mu$ M RNA mutagens: ribavirin, 5-FU, 5-AZC and 0.5 mM  $Mn^{2+}$ . Boxes show median and interquartile range, whiskers show range or 1.5 interquartile range in the case of outliers (indicated by dots);  $n = 10$ .  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$  (two-tailed unpaired  $t$ -test with Bonferroni correction comparing wild type to each mutant).

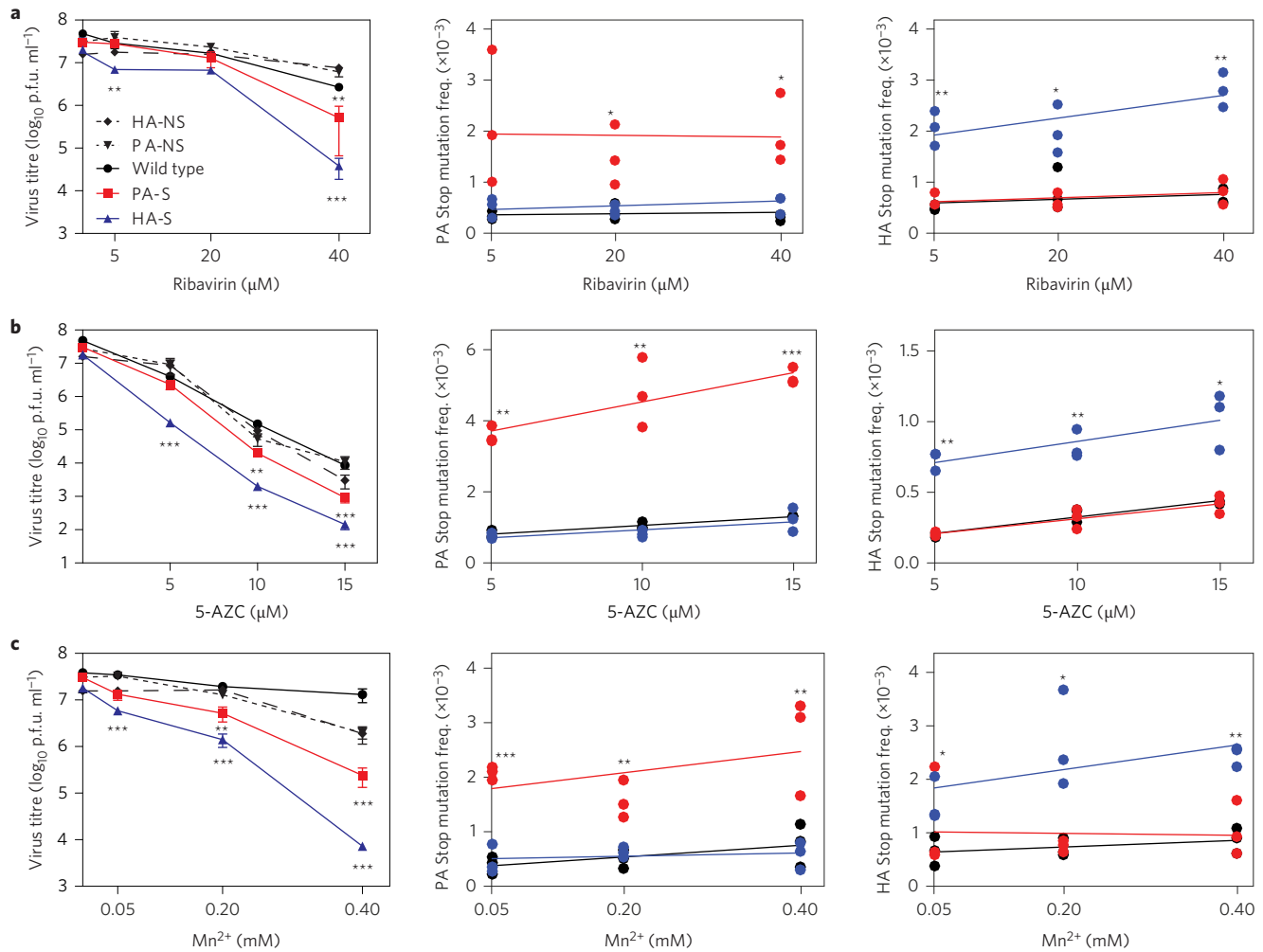
whereas those infected with 1-to-Stop variants lost 6.5% (HAS) and 7.5% (PAS).

**1-to-Stop influenza viruses are immunogenic and protect against challenge.** To investigate the vaccine potential of 1-to-Stop viruses, mice were immunized with either the 1-to-Stop viruses (HAS and PAS) or phosphate-buffered saline (PBS) and, after 21 days, were challenged with wild-type virus. All infected animals seroconverted, with antibody titres ranging from 320 to 1,280, as determined by haemagglutination inhibition assays (Fig. 4i). Similar titres were obtained from animals infected with wild-type virus. Following challenge infection, no virus was detected in the lungs of HAS- and PAS-immunized mice, compared to high titres in PBS-immunized mice (Fig. 4j). By 14 days after challenge, mice immunized with 1-to-Stop variants returned to normal weight, whereas PBS-immunized mice only recovered 50% of the weight loss (Fig. 4k), a profile similar to naive mice infected with wild-type virus (Fig. 4h).

**1-to-Stop virus coupled with a low-fidelity polymerase is optimally attenuated.** Our results demonstrate that relocating a virus in an unfavourable region of sequence space, where a copy error has a higher likelihood of generating non-sense mutations, can attenuate viruses. Moreover, the treatment of these viruses with mutagens to extrinsically increase error rates resulted in even greater loss of infectivity. Previously, we have described CVB3

polymerase variants with intrinsically increased error rates that resemble mutagenic treatment<sup>17</sup>. We therefore engineered the low-fidelity viral polymerase I230F into the 1-to-Stop virus to generate a ‘SpeedyStop’ virus. We infected mice with wild-type, 1-to-Stop or SpeedyStop viruses and quantified virus titres in pancreata (Fig. 5a) and hearts (Fig. 5b). The degree of attenuation significantly increased for SpeedyStop virus compared to its normal-fidelity counterpart. Virus was undetectable in some mouse organs as early as three days after infection, and no longer detectable in any organ by day 7. Accordingly, a survival curve of mice receiving a lethal dose of wild type and equivalent doses of 1-to-Stop and SpeedyStop viruses revealed the latter to be completely attenuated (Fig. 5c). Finally, we deep-sequenced virus from infected tissues and confirmed that SpeedyStop presented a higher number of Stop mutations in sequencing reads than the other viruses (Fig. 5d).

**1-to-Stop and SpeedyStop viruses induce high levels of neutralizing antibodies and protect against lethal challenge.** To evaluate the immunogenicity and protective efficacy of the 1-to-Stop viruses, mice were immunized with  $1 \times 10^5$  p.f.u. of each virus, or with PBS, and blood was collected after three weeks. Mice immunized with 1-to-Stop or SpeedyStop viruses produced high levels of antibody able to neutralize 1,000 p.f.u. of wild-type CVB3 (Fig. 5e). These same mice were challenged with a lethal dose (10 LD<sub>50</sub>, that is, 10 times the dose lethal to 50% of animals



**Figure 3 | The 1-to-Stop influenza A viruses are highly sensitive to mutation.** **a–c**, Left: sensitivity of wild-type, 1-to-Stop (PA-S and HA-S) and NoStop (PA-NS and HA-NS) influenza A viruses to increasing concentrations of ribavirin (**a**), 5-azacytidine (5-AZC) (**b**) and  $Mn^{2+}$  (**c**). Graphs show mean and s.e.m.;  $n = 3$ .  $**P < 0.01$ ,  $***P < 0.001$  (two-tailed unpaired  $t$ -test with Bonferroni correction, comparing wild type to each mutant). Middle: frequency of Stop mutations observed in deep-sequencing reads from wild-type (black) and 1-to-Stop (PA-S, red and HA-S, blue) progeny in the PA gene. Right: frequency of Stop mutations observed in deep-sequencing reads from wild-type (black) and 1-to-Stop (PA-S in red and HA-S in blue) progeny in the HA gene. Bars show mean and s.e.m.;  $n = 3$  per group.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$  (two-tailed unpaired  $t$ -test with Bonferroni correction, comparing wild type to each mutant).

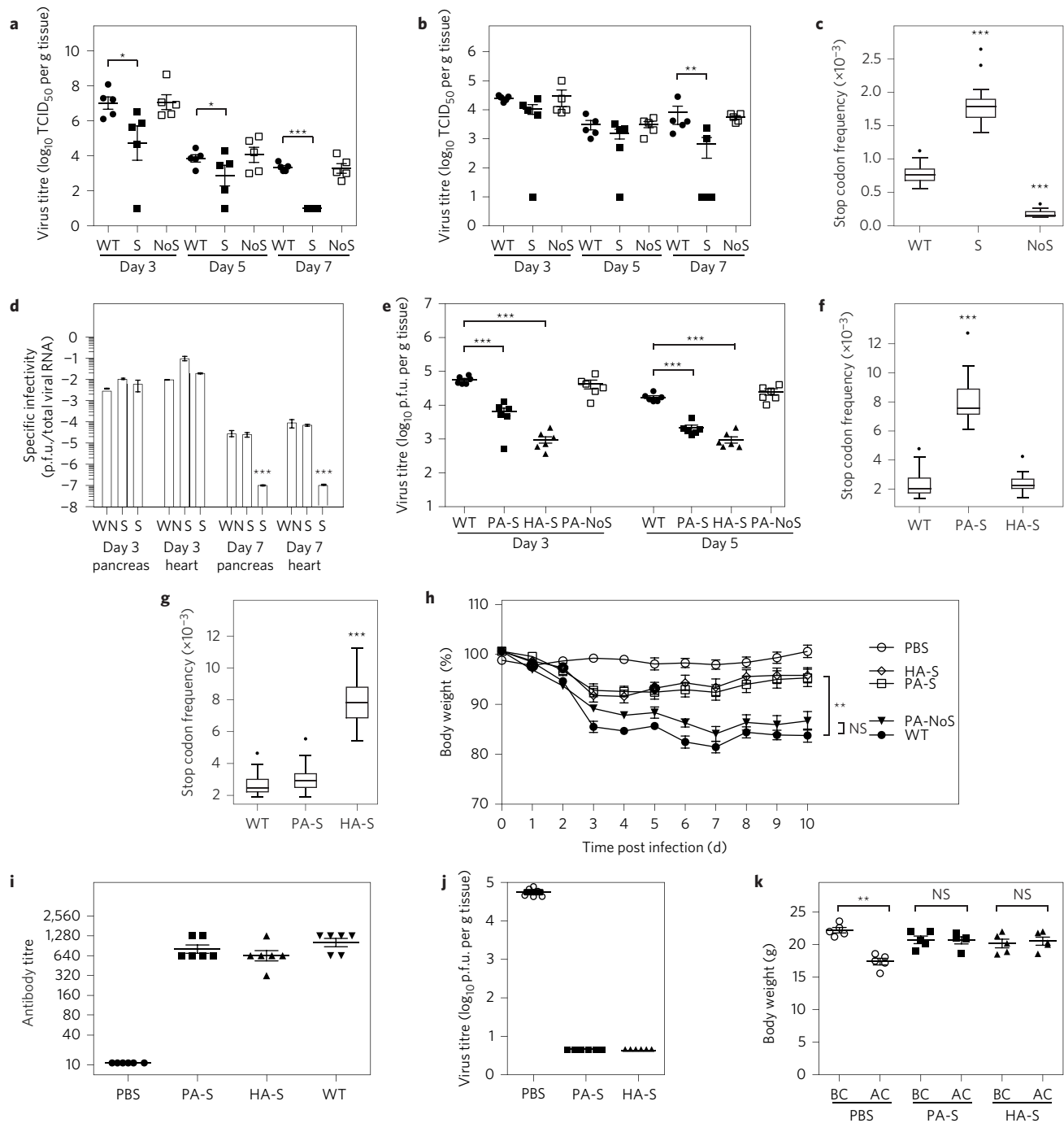
tested) of wild-type CVB3. Most control mice succumbed to infection after eight days, whereas all of the 1-to-Stop or SpeedyStop immunized mice were protected (Fig. 5f).

**Empirical fitness distributions and landscape model.** To further investigate how 1-to-Stop populations may be constrained in their exploration of sequence space and the fitness landscape, we measured the relative fitness of wild-type and 1-to-Stop CVB3 in tissue culture in biological triplicates under five mutagenic conditions, as well as under normal conditions. A range of low to high mutagenic conditions was used to accelerate evolution in sequence space and to increase the mutational load. Under normal conditions, both viruses presented mostly positive fitness values (Fig. 6a). Under mutagenic conditions, a greater proportion of wild-type samples presented positive fitness compared to 1-to-Stop virus (Fig. 6a), and the relative decrease in fitness for 1-to-Stop virus was significant under every growth condition. In fact, only 3 populations out of 45 presented positive fitness values with respect to wild-type (Fig. 6b). We then characterized the mutant spectra of each sample by whole-genome deep sequencing to calculate the mean entropy of each

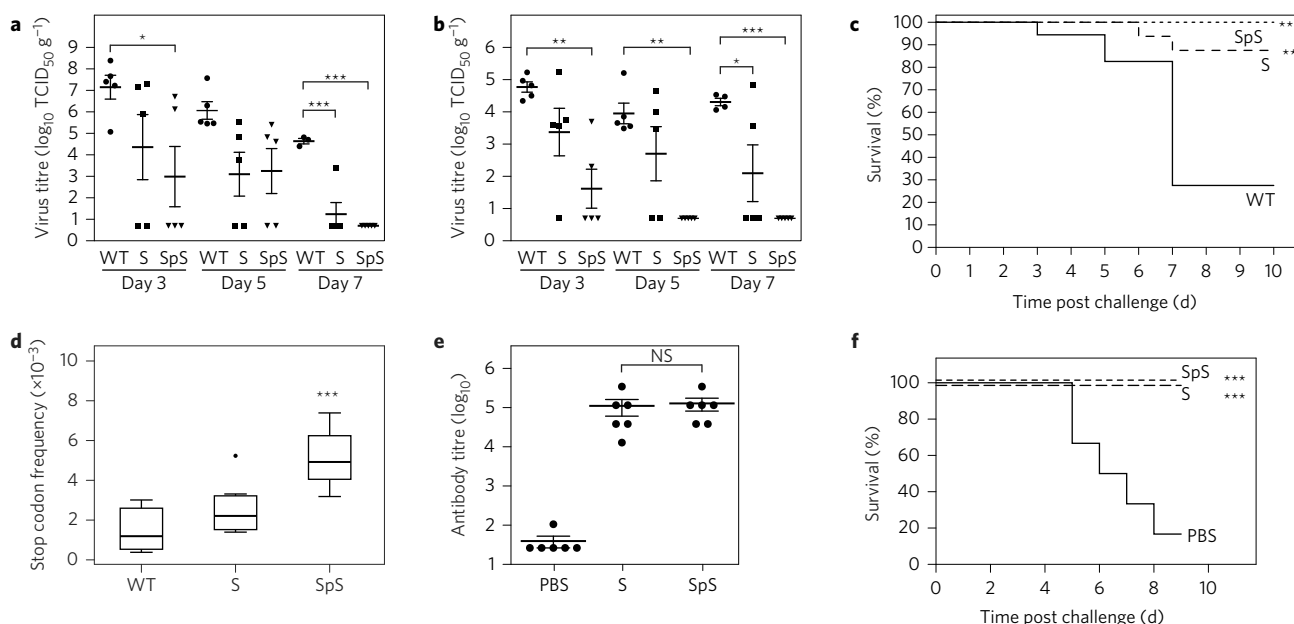
population, as a general measure of sequence space exploration. For illustrative purposes, we generated a fitness landscape-like model based on the entropy values coupled with their empirical fitness values (Fig. 6c). The landscape reveals that for similar mean entropy values, 1-to-Stop populations consistently present a lower fitness and are unable to ‘climb back up’ to the fitness landscape occupied by wild-type virus.

## Discussion

In this work, we present a strategy to attenuate viruses based on evolutionary principles, by restricting their evolutionary potential. Our experimental design is based on the rational relocalization of viral populations in sequence space to redirect them towards detrimental regions of the fitness landscape. Sequence space is a conceptual framework that can help monitor adaptive walks and evolutionary trajectories. For RNA viruses rapidly expanding in sequence space, emerging minority mutations can foretell the directionality of evolution well before the entire population shifts<sup>2,20</sup>. In theory, however, sequence space is immensely larger than the subregions occupied by biologically viable genotypes. For RNA viruses, these constraints derive from the compact nature of



**Figure 4 | The 1-to-Stop viruses are attenuated *in vivo*.** **a, b**, Virus titres in pancreata (**a**) and hearts (**b**) of mice infected with  $1 \times 10^5$  TCID<sub>50</sub> of wild-type (WT), 1-to-Stop (S) and NoStop (NoS) CVB3. Bars show mean and s.e.m., and data are representative of three experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni correction). The limit of detection was  $10^1$  TCID<sub>50</sub> per ml. **c**, Frequency of Stop mutations in CVB3 populations from infected tissues (hearts and pancreata combined). Boxes show median and interquartile range, whiskers show range or 1.5 interquartile range in the case of outliers (indicated by dots);  $n = 62$ . \*\*\* $P < 0.0001$  (two-tailed unpaired *t*-test with Bonferroni correction, comparing wild type to each mutant). **d**, Specific infectivity (TCID<sub>50</sub>/RNA genomes) of CVB3 viruses from infected tissues. Bars show mean and s.e.m.;  $n = 3$ . \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni, comparing wild type to each mutant). **e**, Virus titres in lungs of mice infected with either wild-type, 1-to-Stop (PA-S and HA-S) or NoStop (PA-NoS) influenza A viruses. Bars show mean and s.e.m.;  $n = 6$ , and data are representative of three experiments. \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni correction). **f, g**, Frequency of Stop mutations in the PA (**f**) or HA (**g**) genes of influenza A viruses from infected lungs. Boxes show median and interquartile range, whiskers show range or 1.5 interquartile range in the case of outliers (indicated by dots);  $n = 20$ . \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni correction, comparing wild type to each mutant). **h**, Weight loss in mice infected with  $1 \times 10^5$  TCID<sub>50</sub> of influenza A variants. Graphs show mean and s.e.m.;  $n = 6$ . NS, non-significant; \*\*\* $P < 0.001$  (two-way analysis of variance). **i, j**, Mice were immunized with influenza 1-to-Stop variants (PA-S and HA-S), wild type or PBS. After 21 days, serum antibody titres were determined ( $n = 6$ ) (**i**), mice were challenged with a lethal dose of virus ( $1 \times 10^6$  TCID<sub>50</sub>) and lung titres were determined 4 days after challenge ( $n = 6$ ) (**j**). Limit of detection  $< 10$  p.f.u.  $g^{-1}$ . Bars show mean and s.e.m. **k**, Weights of immunized mice were compared before challenge (BC) and 14 days after challenge (AC) infection;  $n = 5$  per group. NS, non-significant; \*\* $P < 0.01$  (paired *t*-test with Bonferroni correction).



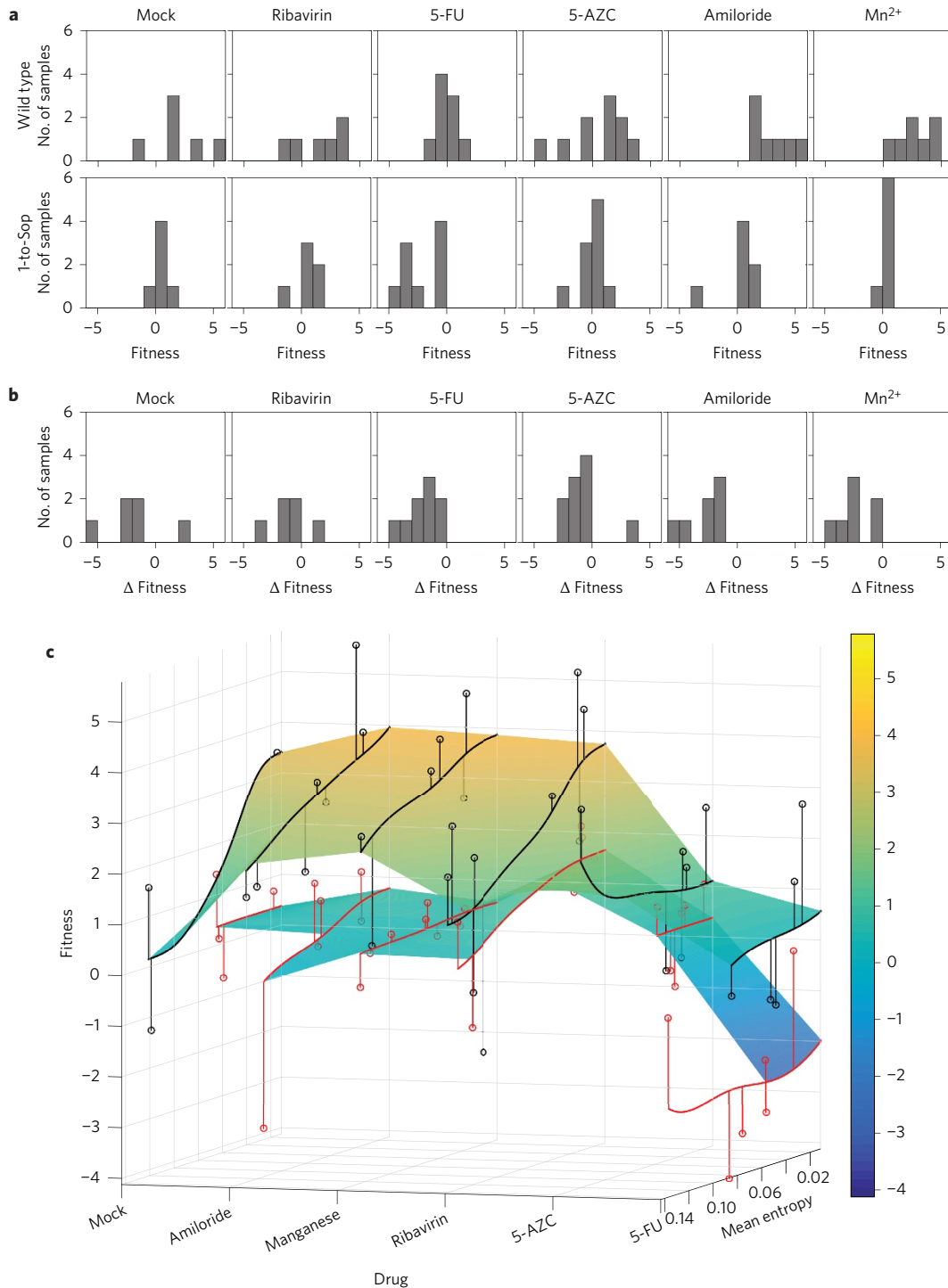
**Figure 5 | Attenuation of 'SpeedyStop' virus, 1-to-Stop CVB3 coupled with a mutator polymerase.** **a,b**, Virus titres (TCID<sub>50</sub> per g) in pancreata (**a**) and hearts (**b**) of mice infected with  $1 \times 10^5$  TCID<sub>50</sub> of wild type (WT), 1-to-Stop (S) or 1-to-Stop coupled with the low-fidelity polymerase mutation I230F, SpeedyStop (SpS) viruses. Bars show mean and s.e.m., and data are representative of three experiments. Day 7 values are set at the limit of detection. \* $P < 0.05$ , \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni correction) **c**, Survival curve of mice infected with either  $1 \times 10^6$  TCID<sub>50</sub> of wild-type (WT, solid line), 1-to-Stop (S, long dashes) or SpeedyStop (SpS, short dashes) viruses;  $n = 17$ . \*\*\* $P < 0.001$  (Mantel-Cox test). **d**, Frequency of Stop mutations observed in deep-sequencing reads from wild-type (WT), 1-to-Stop (S) and SpeedyStop (SpS) CVB3 populations from infected tissues (hearts and pancreata combined). Boxes show median and interquartile range, whiskers show range or 1.5 interquartile range in the case of outliers, and individual dots indicate outliers;  $n = 27$ . \*\*\* $P < 0.001$  (two-tailed unpaired *t*-test with Bonferroni correction, comparing wild type to each mutant). **e**, Neutralizing antibody titres (inverse dilution of sera able to neutralize 1,000 p.f.u. of wild type CVB3) in mice immunized with PBS, or  $10^5$  p.f.u. of 1-to-Stop (S) or SpeedyStop (SpS) viruses;  $n = 6$ . NS, non-significant (two-tailed unpaired *t*-test). **f**, Protection of mice in **e** immunized with PBS, 1-to-Stop (S, long dashes) or SpeedyStop (SpS, short dashes) and challenged with 10 LD<sub>50</sub> of wild-type CVB3;  $n = 6$ . \*\*\* $P < 0.001$  (Mantel-Cox test).

their genomes, and the likelihood that a mutation will hit an essential function is high. Indeed, the majority of mutations are deleterious, as evidenced in many studies that intrinsically or extrinsically increase mutation rates<sup>3,10,21–23</sup>. Despite this, viruses retain high mutation rates to facilitate adaptation. Consequently, mutational robustness has been suggested as an important counter-mechanism<sup>24</sup>. Although better characterized in theoretical and *in silico* work<sup>25–27</sup>, some of the best experimental data have been obtained with non-coding RNA structures that assessed the retention of folding capacity or ribozyme activity in light of mutation<sup>28</sup>. Some evidence for mutational robustness has also been observed in RNA viruses<sup>29–31</sup>. A more recent study that analysed the large-scale codon reshuffling of poliovirus reported the impact of altering mutational robustness<sup>4</sup>. However, the extent of codon changes was such that robustness could not be decoupled from other effects, such as CpG dinucleotide bias<sup>15,32</sup> or codon pair bias<sup>14,33</sup>.

Redesigning the genomic architecture of an RNA virus, as performed here, relies on using its evolutionary potential to its own detriment. Other studies have recoded all amino acids to achieve viral attenuation, but we focused on two amino acids—leucine and serine—for two main reasons. The first was to minimize the confounding effects of manipulating the genome. In contrast to previous strategies<sup>14,34</sup>, we targeted less than 5% of the genome. These viruses were designed to minimize or exclude the effects of codon deoptimization<sup>35–38</sup>, codon pair deoptimization<sup>14,33,34</sup> and CpG/UpA dinucleotide bias<sup>15,32,39</sup>. We thus modified a minimum number and class of codons to retain as much as possible of the wild-type identity. However, it should be noted that although the aforementioned effects were minimized, we cannot be certain that some of these factors cannot, at least partially, account for some of the observed attenuation. The second reason for focusing only

on leucine and serine was to take advantage of the unique properties of the Leu and Ser codons. These codons are the most redundant, with the greatest range of exploration of sequence space. This feature was addressed in the mathematical framework designed by Archetti<sup>40</sup> and based on McLachlan's chemical similarity matrix<sup>41</sup>, which predicts the potential effect of point mutations over synonymous codons. These codons could be clustered into three groups, among which two were of particular interest for our design. The 1-to-Stop group presented the highest likelihood of changes into non-sense mutational targets (NSMTs). The other set contained the NoStop Ser/Leu codons, which are two mutations away from becoming a Stop codon, and made ideal controls with the same number of modifications.

By placing 1-to-Stop populations closer to detrimental mutational neighbourhoods, we conferred to them a lower mutational robustness that would be prone to the most detrimental type of mutation, the Stop mutation. Consequently, proximity to 'hostile' regions in sequence space may drive viruses to regions of lower fitness than a normal, more evolvable population<sup>42</sup>. It should be noted that, in addition to affecting mutational robustness, changes in adaptability and evolvability may also play into the attenuated phenotypes observed here. In addition, our findings experimentally support the notion that viruses may avoid volatile codons that neighbour Stop codons, as was proposed by Plotkin and Dushoff for influenza HA genes<sup>43</sup>. Moreover, deep sequencing supported the proposed mechanism of attenuation: an increase of three- to sixfold in Stop mutation frequencies. It should be noted that the absolute number of Stop mutations made during infection is probably higher than that captured at the moment of sampling, because such genomes cannot replicate and would thus be degraded and cleared.



**Figure 6 | Re-localizing viruses to an inhospitable area of the fitness landscape. a**, Distribution of fitness values. The proportion (y axis, number of samples) of individual fitness values (x axis,  $\log_{10}$  fitness) of wild-type and 1-to-Stop populations derived from mock or mutagenic conditions. **b**, Relative change in fitness of 1-to-Stop compared to wild type under each growth condition. The differences between wild type and 1-to-Stop are significant ( $P = 9.66 \times 10^{-8}$ , two-tailed *t*-test,  $n = 45$ ). **c**, Illustrative fitness landscape model based on empirical data. Wild-type (black) and 1-to-Stop (red) samples are shown as circles. Fitness values ( $\log_{10}$  fitness) are shown on the y axis and assigned colour according to the colour scale on the right, mutagenic conditions on the x axis and mean entropy calculations from whole-genome deep sequencing data on the z axis. Smoothed curves show trends in how mean entropy affects fitness for each construct and drug treatment. Lines connecting samples to the smoothed curves show deviations from the model. The smoothed curves are connected by linear interpolation to create sequence and fitness landscape surfaces, separately, for wild-type and 1-to-stop viruses.

Regarding *in vivo* studies, the 1-to-Stop viruses exhibited clear, attenuated phenotypes. We have demonstrated that a 1-to-Stop vaccine stock could be readily produced in cell culture with genetic and phenotypic stability, is both immunogenic and

protective against lethal infection, and that coupling a mutator polymerase to this construct enhances attenuation without compromising immunogenicity. *In vivo* attenuation was also associated with higher frequencies of Stop mutations in target organs and a



higher loss of infectivity. Additionally, translation of RNAs containing Stop codons would result in truncated proteins that may contribute to a better activation of the immune system<sup>44,45</sup>. Moreover, although non-viable genomes do not replicate, their presence could have further impact through defective interference<sup>46,47</sup> or through an adjuvant effect, as observed for defective interfering genomes<sup>48,49</sup>.

It is important to emphasize that our measure of Stop mutations relies on deep sequencing technology that presents background noise. We took several measures to reduce this concern. From a mathematical standpoint, we modelled the noise in our samples and confirmed that our data were unlikely to be affected (see Supplementary Information). In addition, the background noise was further reduced by the maximum likelihood estimation of the frequencies, which takes base Phred quality scores into account. From an experimental and biological standpoint, we implemented many controls, such as the NoStop viral genomes that are genetically modified as the 1-to-Stop genomes. We also increased sample sizes and biological replicates, performed work in two different RNA viruses, intrinsically (fidelity variants) and extrinsically (mutagenic treatment) altered mutation pressure, and performed work in different cell types and animals. In total, 25 independent experiments, *in vitro* and *in vivo*, produced 420 sequencing samples to determine Stop mutations. In all cases, 1-to-Stop generated more Stop mutations than wild-type virus and NoStop controls. Nevertheless, one must consider the noise within the data and must not rely on the absolute values of the measures, but rather their relative values within each experiment with respect to the controls.

We developed this approach in two viral families, the Picornaviridae and Orthomyxoviridae, to cover the broad range of RNA virus biology. CVB3 is a non-enveloped, positive-sense, non-segmented, single-stranded RNA virus, whereas influenza A is an enveloped, negative-sense, segmented, single-stranded RNA virus. Finally, by modelling empirical data into an illustrative fitness landscape, we demonstrated that 1-to-Stop populations cannot escape the detrimental regions of sequence space associated with lower fitness. Our data suggest that this strategy could be broadly applied to potentially any RNA virus. In summary, our results are a proof of concept that viral genomes can be re-engineered to change their starting position in sequence space and redirect them towards detrimental mutational neighbourhoods, to generate 'suicidal', self-limiting vaccine strains.

## Methods

**Cells and viruses.** HeLa and MDCK cells were obtained from the American Type Culture Collection without further authentication by our laboratory, but were confirmed to be mycoplasma-free. HeLa cells were maintained in DMEM medium with 10% new born calf serum (NBCS), and MDCK cells were maintained in MEM medium with 5% fetal calf serum (FCS). Wild-type Coxsackie virus B3 (Nancy strain), 1-to-Stop and No-Stop variants were generated from a pCB3-Nancy infectious cDNA plasmid. Wild-type influenza A virus (A/Paris/2590/2009 (H1N1pdm09)), 1-to-Stop and No-Stop variants were generated from bidirectional reverse genetics plasmids (provided by S. van der Werf at the Institut Pasteur). We generated 1-to-Stop and No-Stop viruses of Coxsackie and influenza A that bear 117 and 111/94 different synonymous codons, respectively, by *de novo* synthetic gene technology (Eurogentec). All newly generated DNA plasmids were Sanger sequenced in full (GATC Biotech) to confirm each of the 117/111/94 positions. A detailed list of all codon changes introduced is provided in Supplementary Table 1. The low-fidelity 1-to-Stop virus, named SpeedyStop, was generated in CVB3 by insertion of the I230F mutation in the viral polymerase three-dimensional gene by site-directed mutagenesis of the 1-to-Stop CVB3 infectious clone.

**Generation of Coxsackie virus stocks by *in vitro* transcription and transfection.** CVB3 cDNA plasmids were linearized with *Sal I*. Linearized plasmids were purified with the Macherey-Nagel PCR purification kit. Linearized plasmid (5 µg) was *in vitro* transcribed using T7 RNA polymerase (Fermentas). Transcript (10 µg) was electroporated into HeLa cells, which were washed twice in PBS (w/o Ca<sup>2+</sup> and Mg<sup>2+</sup>) and resuspended in PBS (w/o Ca<sup>2+</sup> and Mg<sup>2+</sup>) at 1 × 10<sup>7</sup> cells per ml. Electroporation conditions were as follows: 0.4 mm cuvette, 25 mF, 700 V, maximum resistance, exponential decay in a Biorad GenePulser XCell

electroporator. Cells were recovered in DMEM. A 500 µl volume of p0 virus stocks was used to infect fresh HeLa cell monolayers for three more passages. For each passage, virus was collected by three freeze–thaw cycles and clarified by spinning at 10,000 r.p.m. for 10 min. Three independent stocks were generated for each virus. Consensus sequencing of virus stocks used in downstream experiments confirmed the stability of the engineered mutations and did not detect any additional mutations across the genome.

**Generation of influenza A virus stocks by reverse genetics.** Using 35 mm plates and DMEM supplemented with 10% FCS, co-cultures of 293T (4 × 10<sup>5</sup> per well) and MDCK (3 × 10<sup>5</sup> per well) cells were transfected with the eight bidirectional plasmids both driving protein expression and directing vRNA template synthesis, using 0.5 µg of each plasmid and 18 µl of FUGENE HD (Roche). DNA and transfection reagents were first mixed, then incubated at room temperature for 15 min and finally added to cells, which were then incubated at 35 °C. Sixteen hours later, the DNA-transfection reagent mix was removed, cells were washed twice in DMEM, and 2 ml of DMEM containing 1 µg ml<sup>-1</sup> of L-1-tosylamido-2-phenyl chloromethyl ketone treated trypsin (TPCK-trypsin, Sigma-Aldrich) was added. Cells were incubated at 35 °C for two more days, supernatants were collected and clarified, and virus was titrated by median tissue culture infectious doses (TCID<sub>50</sub>) as described below. Three independent stocks were generated for each virus. Consensus sequencing of virus stocks used in downstream experiments confirmed the stability of the engineered mutations and did not detect any additional mutations across the genome.

**Genetic stability of viruses.** To evaluate its genetic stability, all generated viruses were passaged ten times in HeLa cells (CVB3) or in MDCK cells (influenza A) at low MOI (0.01), and passages 1, 3, 5, 7 and 10 were sequenced.

**Viral titres by TCID<sub>50</sub>.** Tenfold serial dilutions of virus were prepared in serum-free DMEM media. Dilutions were performed in 12 replicates, and 100 µl of dilution was transferred to 1 × 10<sup>4</sup> Vero-E6 (Coxsackie virus) or MDCK (influenza A virus) cells plated in 100 µl DMEM. After five days, living cell monolayers were fixed and stained with crystal violet 0.2%.

**Viral titres by plaque assay.** Vero-E6 (Coxsackie virus) or MDCK-SIAT (influenza A virus) cells were seeded into six-well plates and virus preparations were serially diluted (tenfold) in DMEM serum-free medium. Cells were washed twice with PBS and infected with 250 µl dilution for 30 min at 37 °C, after which a solid overlay comprising DMEM medium and 1% wt/vol agarose (Invitrogen) was added. Two days after infection, cells were fixed and stained with crystal violet 0.2%, and plaques were enumerated.

**Replication kinetics and quantification of viral genomes.** For growth kinetics, HeLa (Coxsackie virus) or MDCK (influenza A virus) cells were infected at an MOI of 1 or 0.1, frozen at different time points after infection, and later titred. Coxsackie viruses were collected by three freeze–thaw cycles, and influenza A viruses were collected in clarified supernatant. For real-time reverse transcription polymerase chain reaction (qRT-PCR) analysis of Coxsackie virus, total RNA was extracted by TRIzol reagent (Invitrogen) and purified. The TaqMan RNA-to-cDNA one-step RT-PCR kit (Applied Biosystems) was used to quantify viral RNA. Each 25 µl reaction contained 5 µl RNA, 100 µM each primer (forward 5'-GCATATGGTGATGATGTGATCGTAGC-3' and reverse 5'-GGGTACTGTTCATCTGCTCTAAA-3') and 25 pmol probe 5'-[6-Fam] GGTTACGGGCTGATCATG-3' in an ABI 7000 machine. Reverse transcription was performed at 50 °C for 30 min and 95 °C for 10 min, and was followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. A standard curve ( $y = -0.2837x + 12.611$ ,  $R^2 = 0.99912$ ) was generated using *in vitro* transcribed genomic RNA. For influenza A virus, a similar Taqman methodology was used based on the WHO-recommended M-segment detection method. The qRT-PCR protocol consisted of an initial reverse transcription step (45 °C for 15 min), followed by an activation step of 3 min at 95 °C, 50 amplification cycles with 10 s at 95 °C, 10 s at 55 °C and 20 s at 72 °C and a final cooling step of 30 s at 40 °C. The primers used were forward: 5' CTT CTA ACC GAG GTC GAA ACG TA 3' and reverse: 5' GGT GAC AGG ATT GGT CTT GTC TTT A 3'. The probe was (HEX): 5' TCA GGC CCC CTC AAA GCC GAG 3'. The Ct values were converted into numbers of vRNA copies using a standard curve obtained with a serial dilution of a quantified synthetic M-segment RNA transcript.

**Viral passages under mutagenic conditions.** The mutagenic compounds (Sigma Aldrich) used were:

Ribavirin IUPAC 1-[(2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxy-methyl)oxolan-2-yl]-1H-1,2,4-triazole-3-carboxamide): 50, 100 and 200 µM for Coxsackie viruses and 5 and 20 µM for influenza A viruses;  
5-Fluorouracil IUPAC 5-fluoro-1H-pyrimidine-2,4-dione: 50, 100 and 200 µM for Coxsackie viruses and 5 and 30 µM for influenza A viruses;  
5-Azacytidine IUPAC 4-amino-1-β-D-ribofuranosyl-1,3,5-tria-zin-2(1H)-one: 50, 100 and 200 µM for Coxsackie viruses and 5 and 15 µM for influenza A viruses;

Amiloride IUPAC 3,5-diamino-6-chloro-*N*-(diaminomethylene) pyrazine-2-carboxamide: 100 and 200  $\mu\text{M}$  for Coxsackie viruses; Manganese ( $\text{Mn}^{2+}$ ), 0.5 mM and 1 mM for Coxsackie viruses.

HeLa (Coxsackie virus B3) or MDCK (influenza A virus) cell monolayers in six-well plates were pretreated for 4 h with ribavirin, AZC, 5-FU,  $\text{MnCl}_2$  or amiloride compounds at different concentrations. Cells were then infected at an MOI of 0.1 for Coxsackie and 0.001 for influenza A virus with passage 2 viruses. At 48 h post-infection, Coxsackie viruses were collected by three freeze–thaw cycles, and influenza A viruses were collected in clarified supernatant. Virus titres (TCID<sub>50</sub> or plaque assay) were determined. The same procedure was repeated for five passages under each different mutagenic condition in three biological replicates, except for influenza A viruses, which were passaged only in low mutagenic conditions in ribavirin, 5-FU and 5-AZC.

**Measurement of plaque size.** Coxsackie virus plaque measurements were performed on subconfluent monolayers of  $1 \times 10^7$  Vero-E6 cells in 10 cm dishes. To ensure non-overlapping plaques the amount of virus was determined empirically (40–70 per dish for Coxsackie). Each plate was scanned individually after 30 h post-infection at 300 d.p.i. Sixteen-bit image files were analysed using ImageJ. The same protocol was used to measure the plaque phenotype of mutagen pre-treated viral populations. Wild-type and 1-to-Stop viruses were submitted to high concentrations of ribavirin, 5-FU and AZC and time post-infection was increased to 40 h to better recover viral viability to perform plaque measures.

**Highly quantitative direct competition assay for empirical fitness measures.** For Coxsackie virus, relative fitness values were obtained by competing wild-type, NoStop and 1-to-Stop virus, obtained from different passages under each mutagen/compound, with a marked reference virus that contained four adjacent silent mutations in the polymerase region introduced by direct mutagenesis. Co-infections were performed in triplicate at an MOI of 0.01 using a 1:1 mixture of each variant with the reference virus. After 24 h, supernatants were collected and one volume of TRIzol reagent (Invitrogen) was added to extract the viral RNA. The proportion of each virus was determined by qRT–PCR on extracted RNA using a mixture of Taqman probes labelled with two different fluorescent reporter dyes. MGB\_CVB3\_WT detects wild-type and 1-to-Stop viruses with the sequence CGCATCGTACCCATGG, and was labelled at the 5' end with a 6FAM dye; MGB\_CVB3\_Ref contains the four silent mutations, CGCTAGTACCCATGG, and was labelled with a 5' VIC dye. Each 25  $\mu\text{l}$  reaction contained 5  $\mu\text{l}$  RNA, 900 nM of each primer (forward primer, 5'-GATCGCATATGTTGATGATGTGA-3'; reverse primer, 5'-AGCTTCAGCAGGAGTAAAGATGCA-3') and 150 nM of each probe. Using a known standard for the wild-type and reference virus during the qRT–PCR, we were able to calculate the RNA concentration for each viral variant with high sensitivity. The relative fitness was determined by the method described in the work by Carrasco *et al.*<sup>16,17</sup>, using the RNA determinations for each virus. Briefly, the formula  $W = [R(t)/R(0)]^{1/t}$  represents the fitness  $W$  of each mutant genotype relative to the common competitor reference sequence, where  $R(0)$  and  $R(t)$  represent the ratio of mutant to reference virus densities in the inoculation mixture and at  $t$  days post-inoculation (1 day in this case), respectively. The fitness of the normal wild type to reference virus was 1.019, indicating no significant differences in fitness caused by the silent mutations engineered in the reference virus (competitor).

**In vitro replication assays in crude membranes.** Three confluent T25 flasks of HeLa cells were infected at an MOI of 3 with wild-type or 1-to-Stop CVB3 viruses. After 16 h of infection, cells were trypsinized, collected and washed with ice-cold PBS, then resuspended in 1 ml swelling buffer made of 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 1.5 mM MgCl<sub>2</sub>, one tablet of protease inhibitor (Complete Mini EDTA-free, Roche) diluted in autoclaved water. Cells were stored for 15 min on ice, then Dounce-homogenized with 30–40 strokes using a 7 ml Dounce All-Glass tissue grinder (Kimble-Chase). The unbroken cells and nuclei were removed by centrifugation at 500g for 5 min and the supernatant fraction was centrifuged at 12,000g for 10 min at 4 °C. The pellet was suspended in 1 ml of buffer (10 mM Tris hydrochloride pH 8.0, 10 mM NaCl, 15% glycerol). This wash step was repeated three times and the pellet was resuspended in 200  $\mu\text{l}$  of storage buffer made of 250 mM sucrose, 10 mM Tris-HCl pH 7.4, 10 mM NaCl, and one tablet of protease inhibitor diluted in autoclaved water. Protein quantity was estimated by Bradford assay. Pellets were diluted at 10 mg ml<sup>-1</sup>, aliquoted, and stored at –80 °C. To perform the *in vitro* replication assay, 25  $\mu\text{l}$  of membrane extract were mixed with 25  $\mu\text{l}$  replication solution made of 1 $\times$  *in vitro* transcription buffer (SP6 mMESSAGE mMACHINE kit, Ambion), 10 mM dithiothreitol (Invitrogen), 10  $\mu\text{g}$  ml<sup>-1</sup> actinomycin D, 5 mM creatine phosphate, 25  $\mu\text{g}$  ml<sup>-1</sup> creatine phosphokinase, 1 mM ATP, 1 mM GTP, 1 mM CTP, 50  $\mu\text{M}$  UTP, 1  $\mu\text{l}$  RNase out recombinant ribonuclease inhibitor (Invitrogen) in autoclaved water, 2  $\mu\text{g}$  *in vitro* transcribed viral RNA and 20  $\mu\text{Ci}$  UTP [ $\alpha$ -<sup>32</sup>P] (PerkinElmer). Samples were incubated for 2 h at 37 °C. RNA extraction was performed using phenol-chloroform, then samples were purified on Illustra MicroSpin S200 HR columns. Samples were run on a 1% agarose gel, dried in a gel-drier machine, and imaged using a Typhoon FLA9500 (GE Healthcare).

**Mouse husbandry and ethics.** Mice were kept in the Pasteur Institute animal facilities under Biosafety Level 2 conditions, with water and food supplied *ad libitum*, and they were handled in accordance with the Animal Committee regulations of the Institut Pasteur in Paris, France, in accordance with the 2010/63 EU directive adopted on 22 September 2010 by the European Parliament and the European Union Council. Mouse protocols 2013–0101 and 2013–0021 were evaluated and approved by the Ethics Committee on Animal Experimentation CETEA no. 89 (Institut Pasteur), working under the French national Ministère de l'Enseignement supérieur et de la Recherche (MESR). All studies were carried out in BALB/c male mice (between five and six weeks old) from Charles River.

**Coxsackie virus infections in vivo.** Mice were infected intraperitoneally with  $1 \times 10^5$  TCID<sub>50</sub> wild-type or 1-to-Stop viruses in 0.20 ml volumes. For tissue tropism studies, we collected whole organs (pancreas and heart) at 3, 5 and 7 days post-infection, and these were homogenized in PBS using a Precellys 24 tissue homogenizer (Bertin Technologies). Viral RNA was extracted using TRIzol reagent (Invitrogen). Full genome PCR, viral titres by TCID<sub>50</sub> and qRT–PCR were performed as described above. Survival curves were generated by injecting four-week-old mice ( $n = 8$  mice per virus) with  $5 \times 10^6$  TCID<sub>50</sub> of virus and monitoring morbidity and mortality for 10 days after infection. For protection studies, mice were immunized with PBS or  $5 \times 10^5$  TCID<sub>50</sub> of 1-to-Stop or SpeedyStop virus. At 21 days after immunization, serum was collected to quantify the production of neutralizing antibodies. Mice were then challenged with  $1 \times 10^6$  of wild-type virus (hyper-virulent strain 372V of Coxsackie virus B3) and survival was monitored over the following 10 days.

**Neutralization assay.** At 3 weeks after immunization, serum was collected, heat-inactivated at 56 °C for 30 min, and serially diluted in DMEM, and the CVB3 stock was diluted to a working concentration of  $3 \times 10^3$  TCID<sub>50</sub>. Neutralizing antibody titres were determined by TCID<sub>50</sub> reduction assay in Vero-E6 cells, and 50  $\mu\text{l}$  of each diluted serum sample was mixed with 50  $\mu\text{l}$  of CVB3 at a working concentration and added to 96-well plates for incubation at 37 °C for 2 h. Following incubation, eight replicates of each dilution were used to infect  $1 \times 10^4$  Vero-E6 cells seeded in a 96-well plate. At 6 days post-infection, the cells were observed under a microscope for the presence of cytopathic effect (CPE). Neutralization titres were determined as the highest serum dilution that could prevent CPE in more than 50% of cells.

**Influenza virus infection in vivo.** Mice were infected intra-nasally with  $1 \times 10^5$  TCID<sub>50</sub> wild-type, 1-to-Stop or NoStop viruses as a 20  $\mu\text{l}$  volume (diluted in PBS). Lungs were collected at three and five days post-infection and were homogenized in PBS using a Precellys 24 tissue homogenizer (Bertin Technologies). Infectious virus within homogenized tissues was titrated by plaque assay, and titres were expressed as p.f.u. per g organ (p.f.u. g<sup>-1</sup>). Viral RNA was extracted using TRIzol reagent (Invitrogen). Virus genomic variability was evaluated by deep sequencing, as described in the following.

**Serum antibody titre by haemagglutination inhibition assay.** Mice were infected intra-nasally with  $1 \times 10^4$  p.f.u. of wild-type, 1-to-Stop or NoStop viruses and bled for serum on day 21 post-infection. Antibody titres correspond to the maximum dilution able to inhibit agglutination of red blood cells in the presence of influenza virus under standardised conditions, as previously described<sup>50</sup>.

**Full genome analysis by deep sequencing.** To estimate the population diversity of variants by deep sequencing, Coxsackie virus cDNA libraries were performed using the kit Maxima H Minus First Strand cDNA Synthesis (ThermoFisher) and oligo dT as a primer from RNA extracted from virus generated in HeLa cells or different mouse organs. The viral genome was amplified using a high-fidelity polymerase (Phusion) to generate 1 amplicon of 7.2 kb in length (full-length genome). The primers and PCR were designed and optimized in the laboratory (5' GAAAACGCGGGGAGGGTCAA3' and 5' ACCCCCTCCCCAACTGTAA 3'). For influenza A virus, the viral RNA genome was extracted from infected-cell supernatants (Macherey-Nagel), reverse-transcribed with an Accuscript High Fidelity 1st strand cDNA Synthesis kit (Agilent) using 5'-AGCRAAAGCAGG-3' primer and amplified by PCR using a high-fidelity polymerase (Phusion). Eight PCRs were designed to cover the coding regions of the eight genomic segments (primer sequences are available upon request). For mouse organs, RNA was extracted with TRIzol reagent (Invitrogen) and PA and HA segments were targeted by PCR. The PCR products were purified and fragmented (Fragmentase), multiplexed, clustered on cBot for sequencing in GAIIX, or clustered and sequenced on NextSeq500, Illumina technology and analysed with established deep-sequencing data analysis tools and in-house scripts.

**Codon frequencies.** The sequenced reads for each sample were aligned to their respective reference genomes using BWA. Per-site codon frequencies were estimated for each sample by considering the reads covering the given site. Only nucleotides with Phred base quality scores  $\geq 30$  were used. The observed codon frequencies were modelled by a multinomial distribution, observed under noise. The noise model is given directly by the Phred base quality scores, which describe the probability of a read error at each nucleotide in each read. Finally, the ML (maximum likelihood)

estimates of the codon frequencies were computed numerically using this model. Under perfect quality scores, the model would simplify to a multinomial model and each estimated codon frequency would correspond to the proportion of reads with that codon. However, with actual quality scores, the impact of read errors is reduced, because the model corrects for the read error rate. The mathematical model for background error is further described in the Supplementary Information.

**Stop codons.** Per-sample Stop codon frequencies were computed by summing the Stop codon frequencies over all modified codon sites, giving a number approximately equal to the frequency of viral genomes that have been rendered unviable by incorporating a Stop codon at one of the modified sites. The computations were done for all samples, and next-generation sequencing batch effects were avoided by only comparing samples obtained on the same sequencing runs. Box plots and linear regression plots were used to visualize the frequency distributions for relevant groups and covariates.

**Fitness distribution graphs.** Histograms were generated showing empirical fitness values with the samples grouped by construct and mutagenic conditions. The difference in fitness ( $\Delta$ Fitness) between pairs of wild-type and 1-to-Stop codons from the same experimental conditions were also computed and shown in histograms, again grouped by mutagen.

**Entropy calculation from deep sequencing data.** The entropy  $-\sum_{i=1}^{64} P(x_i) \log(P(x_i))$ , at a given codon site in a given sample, was computed directly from the ML estimates of the codon frequencies. Then, for each sample, the mean entropy was computed over all codon sites in the entire genome. Smoothed curves, capturing trends for the mapping of mean entropy to empirical fitness, were created for each construct and mutagen. The curves were then linearly interpolated between mutagens (roughly ordered by mutagen characteristics) to create an illustrative landscape surface.

**Statistical methods.** No statistical methods were used to predetermine the sample size. All experiments were performed three times and  $n$  values represent biological replicates. Equal variance was assumed.  $P$  values  $\geq 0.05$  were considered non-significant. For deep-sequencing analysis, when outliers were identified they were indicated in the figures and legends. For animal studies, mice were randomly allocated to different cages before experiments, and no mice were excluded from analyses. The investigator was blinded to group allocation when virus was titred from collected tissues.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request. In-house codes are also available at any time upon request to the authors.

Received 22 July 2016; accepted 27 April 2017;  
published 5 June 2017

## References

- Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* **51**, 151–178 (1997).
- Bordería, A. V. *et al.* Group selection and contribution of minority variants during virus adaptation determines virus fitness and phenotype. *PLoS Pathog.* **11**, e1004838 (2015).
- Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).
- Lauring, A. S., Acevedo, A., Cooper, S. B. & Andino, R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe* **12**, 623–632 (2012).
- Hall, A. R., Griffiths, V. F., MacLean, R. C. & Colegrave, N. Mutational neighbourhood and mutation supply rate constrain adaptation in *Pseudomonas aeruginosa*. *Proc. Biol. Sci.* **277**, 643–650 (2010).
- Wilke, C. O. Adaptive evolution on neutral networks. *Bull. Math. Biol.* **63**, 715–730 (2001).
- Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).
- Sanjuán, R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. Lond. B* **365**, 1975–1982 (2010).
- Cuevas, J. M., Domingo-Calap, P. & Sanjuán, R. The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol.* **29**, 17–20 (2012).
- Sanjuán, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401 (2004).
- Anderson, J. P., Daifuku, R. & Loeb, L. A. Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.* **58**, 183–205 (2004).
- Perales, C., Martín, V. & Domingo, E. Lethal mutagenesis of viruses. *Curr. Opin. Virol.* **1**, 419–422 (2011).
- Elena, S. F. RNA virus genetic robustness: possible causes and some consequences. *Curr. Opin. Virol.* **2**, 525–530 (2012).
- Coleman, J. R. *et al.* Virus attenuation by genome-scale changes in codon pair bias. *Science* **320**, 1784–1787 (2008).
- Tulloch, F. *et al.* RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife* **3**, e04531 (2015).
- Carrasco, P., Daròs, J. A., Agudelo-Romero, P. & Elena, S. F. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J. Virol. Methods* **139**, 181–188 (2007).
- Gnädig, N. F. *et al.* Coxsackievirus B3 mutator strains are attenuated *in vivo*. *Proc. Natl Acad. Sci. USA* **109**, E2294–E2303 (2012).
- Levi, L. I. *et al.* Fidelity variants of RNA dependent RNA polymerases uncover an indirect, mutagenic activity of amiloride compounds. *PLoS Pathog.* **6**, e1001163 (2010).
- McDonald, S. *et al.* Design of a genetically stable high fidelity Coxsackievirus B3 polymerase that attenuates virus growth *in vivo*. *J. Biol. Chem.* **291**, 13999–14011 (2016).
- Isakov, O. *et al.* Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics* **31**, 2141–2150 (2015).
- Loeb, L. A. *et al.* Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl Acad. Sci. USA* **96**, 1492–1497 (1999).
- Crotty, S. & Andino, R. Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin. *Microbes Infect.* **4**, 1301–1307 (2002).
- Bordería, A. V., Rozen-Gagnon, K. & Vignuzzi, M. Fidelity variants and RNA quasispecies. *Curr. Top. Microbiol. Immunol.* **392**, 1–20 (2015).
- Stern, A. *et al.* Costs and benefits of mutational robustness in RNA viruses. *Cell Rep.* **8**, 1026–1036 (2014).
- Wilke, C. O. Evolution of mutational robustness. *Mutat. Res.* **522**, 3–11 (2003).
- Sanjuán, R., Forment, J. & Elena, S. F. *In silico* predicted robustness of viroid RNA secondary structures. II. Interaction between mutation pairs. *Mol. Biol. Evol.* **23**, 2123–2130 (2006).
- Sanjuán, R., Forment, J. & Elena, S. F. *In silico* predicted robustness of viroids RNA secondary structures. I. The effect of single mutations. *Mol. Biol. Evol.* **23**, 1427–1436 (2006).
- Schultes, E. A. & Bartel, D. P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448–452 (2000).
- Sanjuán, R., Cuevas, J. M., Furió, V., Holmes, E. C. & Moya, A. Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* **3**, e93 (2007).
- Graci, J. D. *et al.* Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. *J. Virol.* **86**, 2869–2873 (2012).
- Codoñer, F. M., Daròs, J.-A., Solé, R. V. & Elena, S. F. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* **2**, e136 (2006).
- Atkinson, N. J., Witteveldt, J., Evans, D. J. & Simmonds, P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.* **42**, 4527–4545 (2014).
- Le Nouën, C. *et al.* Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc. Natl Acad. Sci. USA* **111**, 13169–13174 (2014).
- Mueller, S. *et al.* Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.* **28**, 723–726 (2010).
- Burns, C. C. *et al.* Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J. Virol.* **80**, 3259–3272 (2006).
- Aragónès, L., Guix, S., Ribes, E., Bosch, A. & Pintó, R. M. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* **6**, e1000797 (2010).
- Nougairède, A. *et al.* Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. *PLoS Pathog.* **9**, e1003172 (2013).
- Meng, J., Lee, S., Hotard, A. L. & Moore, M. L. Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. *mBio* **5**, e01704–14 (2014).
- Gaunt, E. *et al.* Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *eLife* **5**, e12735 (2016).
- Archetti, M. Genetic robustness at the codon level as a measure of selection. *Gene* **443**, 64–69 (2009).
- McLachlan, A. D. Tests for comparing related amino-acid sequences. Cytochrome *c* and cytochrome *c* 551. *J. Mol. Biol.* **61**, 409–424 (1971).
- Burch, C. L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625–628 (2000).
- Plotkin, J. B. & Dushoff, J. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc. Natl Acad. Sci. USA* **100**, 7152–7157 (2003).
- Kovac, A. *et al.* Misfolded truncated protein  $\tau$  induces innate immune response via MAPK pathway. *J. Immunol.* **187**, 2732–2739 (2011).

45. Dabaghian, M., Latifi, A. M., Tebianian, M., Dabaghian, F. & Ebrahimi, S. M. A truncated C-terminal fragment of *Mycobacterium tuberculosis* HSP70 enhances cell-mediated immune response and longevity of the total IgG to influenza A virus M2e protein in mice. *Antiviral Res.* **120**, 23–31 (2015).
46. Crowder, S. & Kirkegaard, K. Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nat. Genet.* **37**, 701–709 (2005).
47. Gonzalez-Lopez, C., Arias, A., Pariente, N., Gómez-Mariano, G. & Domingo, E. Preextinction viral RNA can interfere with infectivity. *J. Virol.* **78**, 3319–3324 (2004).
48. Yount, J. S., Kraus, T. A., Horvath, C. M., Moran, T. M. & López, C. B. A novel role for viral-defective interfering particles in enhancing dendritic cell maturation. *J. Immunol.* **177**, 4503–4513 (2006).
49. Tapia, K. *et al.* Defective viral genomes arising *in vivo* provide critical danger signals for the triggering of lung antiviral immunity. *PLoS Pathog.* **9**, e1003703 (2013).
50. Boge, T. *et al.* A probiotic fermented dairy drink improves antibody response to influenza vaccination in the elderly in two randomised controlled trials. *Vaccine* **27**, 5677–5684 (2009).

### Acknowledgements

The authors thank M. Declercq, M. Barbet, S. Van der Werf, C. Saleh and A. Pizzorno for assistance and discussions. This work was supported by an Institut Pasteur Innovative Ideas

in Vaccinology - GPF Vaccinology grant, an ERC PoC grant no. 727758 and the Roux-Cantarini Fellowship (G.M.).

### Author contributions

G.M., C.B. and M.V. designed the experiments. G.M., C.B., L.C., A.V.B., H.B., E.Z.P., S.B., T.V. and B.C.M. performed experiments. G.M., R.H., J.B., C.B., M.F. and M.V. analysed the data. G.M. and M.V. wrote the paper.

### Additional information

**Supplementary information** is available for this paper.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to M.V.

**How to cite this article:** Moratorio, G. *et al.* Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.* **2**, 17088 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Competing interests

The Institut Pasteur has filed a European Patent Application that covers the alteration of RNA virus sequence space and in particular the manipulation of codons to skew towards non-sense mutation targets such as Stop mutations, as a means of attenuating viruses. G.M., C.B. and M.V. are listed inventors (World Patent Application no. WO2016120412).