

# Evolvability drives innovation in viral genomes

Guillaume Beslon<sup>1</sup>, Vincent Liard<sup>1</sup>, Santiago F. Elena<sup>2,3</sup>

<sup>1</sup> Université de Lyon, CNRS, INRIA, INSA-Lyon, UCB Lyon 1, LIRIS UMR5205, Beagle Team, F-69622 Lyon, France

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, E-46022 València, Spain

<sup>3</sup> Santa Fe Institute, Santa Fe NM 87501, USA

## Introduction

Viruses have very short and compact genomes and very high mutation rates [4]. Despite these characteristics, they are known to evolve by bursts, periods of high evolution rates alternating with periods of high stability [5,6]. The mechanisms that drive such switches from slow to fast evolutionary dynamics are mostly unknown although it has often been proposed that bursts of evolution may be triggered by environmental changes [7] or by the occurrence of deleterious events [8,9].

Here, we simulated the evolution of viral strains using the *aevo1* simulator [1,2]. By tracking down the long term evolution of 100 virus-like strains derived from 10 different wild-types, we were able to identify 16 strains undergoing bursts of fast evolutionary rates. By analyzing these 16 strains, we were able to identify the events triggering the bursts.

At first sight these events were very diverse, some being beneficial while others were (slightly) deleterious, some being large scale chromosomal rearrangements while others were local mutations. However, most of them share a common characteristic: they strongly increase the evolvability of the viral strains, thus opening the door for further favorable mutations to be fixed until this gained evolvability is exhausted and goes back to a very low value, marking the beginning of a new long period of evolutionary stasis.

## Methods

### Evolution experiments

All simulations were conducted with *aevo1* 5.0 (available at <http://www.aevo1.fr>). We designed a phenotypic target containing two functions and let 10 independent populations of 2500 individuals evolve, each during 200,000 generations under a very high mutation rate ( $1 \times 10^{-4}$  mutation  $\times$  bp<sup>-1</sup>  $\times$  generation<sup>-1</sup> for each kind of mutation). The genomes resulting from these 10 evolution experiments were later on called the wild-types (WT). Each WT was then cloned into 10 independent populations that were let to evolve during 30,000 more generations. At the end of this second phase of evolution, the resulting 100 evolved lineages were analyzed in detail.

## Overview of the analysis procedure

The analyses were conducted in 4 steps for each evolved lineage.

1. Using `aevol_misc_lineage` we identified the line of descent of the best final organism (*i.e.*, the best organism at generation 230,000) for each replicate.
2. Using `aevol_misc_ancestors_stats` we reconstructed the genomes of the ancestors of the best final organism as well as all the mutational events that occurred and were fixed in this lineage. The last 5,000 generations were removed from the analyses since they may contain individuals/events that were not fixed at generation 230,000.
3. We identified the replicates that had undergone a burst of evolution. This was done by isolating the replicates for which the fitness increase between generation 200,000 and generation 225,000 was clearly higher than the “bulk” of simulations (the bulk itself being centered on zero, showing that the other replicates did not significantly evolve during the additional 25,000 generations of the second phase of the evolution experiments).
4. For these replicates, we used `aevol_misc_ancestors_robustness` to measure the local curvature of the fitness landscape at the position of the ancestors of the best final organism (since this procedure is highly time consuming it was done every 10 generations).

## Measure of evolvability

`aevol_misc_ancestors_robustness` generates a very large set of offspring of the targeted individual (here the ancestor of the best final organism) and measures the proportion of these offspring that are better than, equivalent to, or worse than the starting individual. The measure of evolvability is then the sum of the improvement of all beneficial mutants. This measure thus takes into account both the probability that the individual undergoes a favorable mutation and the mean improvement that this mutation could provide.

## Results

### Evolution of the WTs

Owing to the very high mutation rates imposed during the first phase of the evolution experiments [3], the 10 WTs end-up with virus-like genome structures: genomes were very short (max: 607 bp, min: 232 bp, mean: 457 bp) with few genes (max: 13 genes, min: 5 genes, mean: 10 genes) densely packed.

### Evolution of the replicates

Only 16 of the 100 replicates showed a significant increase of fitness. The 84 other replicates showed no significant increase or even a decrease of fitness relative to

their corresponding WT. Interestingly, all the replicates that improved significantly did so very quickly during the second phase of the evolution experiment: all but one gained more than 75% of the fitness in less than 5,000 generations (the only one that needed more than 5,000 generations actually underwent multiple bursts during the 25,000 generations). Moreover, most clones fixed many non-neutral mutations during these 5,000 generations but much less during the remaining of the experiment. In other words, the evolution of these clones was generally due to cascades of mutational events occurring in a rather short time window. We called these cascades *innovations*. Figure 1 shows an example of such a cascade.

### Dynamic of innovation

In order to understand the dynamics of innovation in the 16 strains, we first measured the variation of evolvability during the cascade. As exemplified on figure 1, during the cascade, the evolvability was systematically increased by at least two orders of magnitude (an exception being the clones issued from WT6 that all evolved in a single event without showing significant increase in evolvability). This shows that cascades are triggered by events that moved the individual on the fitness landscape to a new position where the probability of improvement is strongly increased.

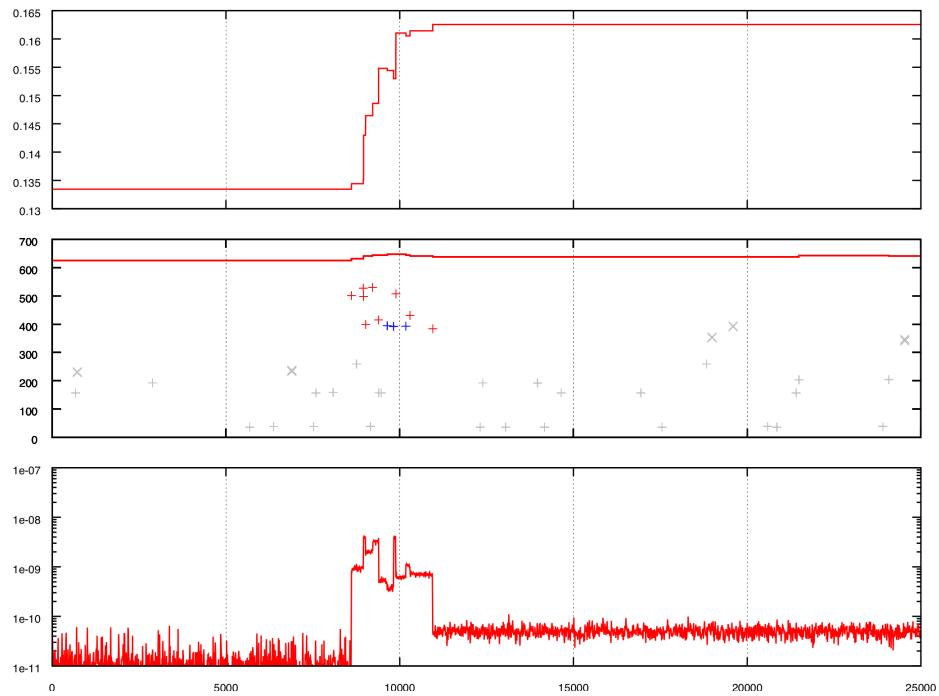
We then identified the first mutational event of the cascade, *i.e.* the ones that are most likely the triggers of the whole sequence of innovations through an increase in evolvability (*e.g.* the small insertion occurring at generation  $\tilde{8},500$  on figure 1). Interestingly the 16 events were:

- 3 favorable small insertions,
- 2 favorable inversions,
- 6 favorable chromosomal duplications,
- 1 deleterious substitutions,
- 3 deleterious small insertions,
- 1 deleterious chromosomal duplications.

Most of these events are favorable events. Strikingly, we never observed favorable substitution triggering an innovation while many favorable substitutions went to fixation during the innovation sequences.

### Discussion

Our experiments show that evolutionary bursts are a normal mode of evolution for short genomes like the ones we have simulated here. Importantly, the bursts observed here are not triggered by environmental variation (the environment remained the same throughout the entire experiment) nor by the systematic occurrence of a deleterious event (only 5 out of the 16 events triggering an innovation were deleterious). Rather, most of them were triggered by events



**Fig. 1.** WT 5, replicate 0: Top: evolution of fitness during the 25,000 generation of evolution of the replicate (*i.e.*, from generation 200,000 to generation 225,000). Middle: evolution of genome size and locus of fixed mutations (“+”) and rearrangements (“x”). Blue: deleterious mutations, gray: neutral mutations, red: favorable mutations. Bottom: evolution of evolvability (measured each 10 generations by the mean increase of fitness of the positive mutants among 1,000,000 replications - see methods).

that increased the evolvability of the lineage, thus opening the door for further favorable events.

An important question is whether this dynamic is specific to short and highly packed genomes, like those of viruses, or whether it could also be observed in longer ones (*e.g.* bacterial genomes). Although answering this question would require further experiments, we can at least argue that, in the conditions studied here (high mutation rates, short genomes, large effective population size), cascades are likely to be more visible since they occur on very short time scales and are surrounded by periods of apparent very low activity.

Last, a very simple explanation can be proposed to account for the kind of events that are the most likely to trigger innovations. Indeed, out of the 16 innovations observed here, 9 were triggered by chromosomal rearrangements and 6 were triggered by small insertions. If one considers the combinatorics of the different kinds of mutations simulated in *aevo*, one would get, for a 500 bp genomes:

- 500 possible point substitutions,
- 3,000 possible small deletions,
- 63,000 possible small insertions,
- 250,000 possible inversions/large chromosomal deletions,
- 125,000,000 possible translocations/large chromosomal duplications.

These values enlighten the observed dynamics. Indeed, the combinatorics of substitutions (and probably of small deletions) is so small that all possible beneficial substitutions are very quickly discovered in the population. Thus, in a “normal” regimen of evolution, populations quickly run out of possible beneficial substitution. When this happens, the only fuel for evolution relies on the combinatorics of small insertions and of all kinds of large chromosomal rearrangements, which is much larger. Moreover, these events strongly modify the genome sequence, by adding new loci to the genome, and creating a new set of possible beneficial mutations.

As a conclusion, our simulations show that the dynamics of innovation in viral genomes are driven by rare events that move the virus to a new position on the fitness landscape where the local curvature is steep enough as to trigger a cascade of favorable mutations (*i.e.*, increase of evolvability). Once these favorable mutations are exhausted, the virus goes back to stasis until another rare event triggers the next innovation.

## Acknowledgement

This work was supported by the European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427). The authors want to thank all the partners of the EvoEvo project for fruitful discussions.

## References

1. Knibbe C, Coulon A, Mazet O, Fayard JM, Beslon G. 2007. A long-term evolutionary pressure on the amount of noncoding DNA. *Mol Biol Evol* 24(10):2344-53.
2. Batut B, Parsons DP, Fischer S, Beslon G, Knibbe C. 2012. In silico experimental evolution: a tool to test evolutionary scenarios, *BMC Bioinformatics*, 14 (S15)
3. Fischer S, Bernard S, Beslon G, Knibbe C. 2014. A model for genome size evolution. *Bull Math Biol* 76(9):2249-91.
4. Belshaw R, Gardner A, Rambaut A, Pybus OG. 2007. Pacing a small cage: mutation and RNA viruses. *Trends Ecol Evol* 4, 188-93.
5. Bedford T, Cobey S, Pascual, M. 2011. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evol Biol* 11, 220.
6. Strelkova N, Lässig M. 2012. Clonal interference in the evolution of influenza. *Genetics* 192, 671-82.
7. Elena SF, Fraile A, García-Arenal F. 2014. Evolution and emergence of plant viruses. *Adv Virus Res* 88, 161-91.
8. Poon A, Chao L. 2005. The rate of compensatory mutation in the DNA bacteriophage  $\phi$ X174. *Genetics* 170, 989-99.
9. Torres-Barceló C, Daròs JA, Elena SF. 2010. Compensatory molecular evolution of HC-Pro, an RNA-silencing suppressor from a plant RNA virus. *Mol Biol Evol* 27, 543-51.