Genetic shifts in viral populations: how sudden *vs* gradual changes in host species composition affect the stability and dynamics of genetic variability in viral populations

João M. F. Silva[1], María J. Olmo-Uceda[1], Valerie J. Morley[2,$], Paul E. Turner[2,3], and Santiago F. Elena[1,4,*]

[1]Instituto de Biología Integrativa de Sistemas (I[2]SysBio), CSIC – Universitat de València, Paterna, Valencia, Spain

[2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

[3]QBio Institute, Yale University, New Haven, CT, USA

[4]The Santa Fe Institute, Santa Fe, NM, USA

[*]Corresponding author: I[2]SysBio, CSIC-UV, Catedrático Agustín Escardino 9, Paterna, 46980 València, Spain. Email: santiago.elena@csic.es

[$]Present address: Gingko Bioworks Inc., Boston, MA, USA

ORCID:

JMFS: 0000-0002-4880-031X

MJOU: 0000-0001-8570-424X

VJM: 0000-0001-6805-7562

PET: 000-0003-3490-7498

SFE: 0000-0001-8249-5593

Abstract

Sudden *vs* gradual changes in host species composition —from more to less permissive– have been shown to affect viral fitness, virulence, and rates of molecular evolution. In this study, we characterized the role of noise and selection in previously published time series data from evolving populations of Sindbis virus under different rates of host replacement. First, an approximate Markov model within the Wright-Fisher diffusion analytical framework indicated a reduction in effective population size by about half, along with the co-occurrence of fewer and generally weaker beneficial mutations in the scenario of sudden host composition change. Second, analysis of genetic distance between populations at consecutive time points showed that populations under gradual host change evolved steadily until the original host was no longer present, whereas distances to the ancestral sequence showed evidence of leapfrog phenomenon at some points. In contrast, the sudden treatment was associated with relatively less stable population compositions that diverged from the ancestral sequence at a more constant rate. Third, we found that allele frequencies and their dispersion follow a power law known as Taylor's law. Both treatments showed a high degree of allele aggregation with large fluctuations, where neutral, beneficial and deleterious alleles can be discerned by their general behavior and the region they occupy on Taylor's plot. Finally, we found evidence that host replacement regime affects how the mutations are distributed across the genome by time.

Keywords: Approximate Wright-Fisher diffusion model, effective population size, experimental evolution, selection coefficient, Sindbis virus, Taylor's temporal fluctuation scaling, Hurst's exponent, virus evolution

The molecular dynamics of adaptation is largely influenced by the heterogeneity of the environment. Seeing that pathogens adapt to new hosts by the fixation of mutations that increase fitness in the novel host but may otherwise have a neutral or even negative effect in the original host, the availability of novel hosts in the environment influences both the rate and direction of molecular adaptation. For instance, the adaptation dynamics of Sindbis virus (SINV) to a less permissive cell type in laboratory tissue-culture is dictated by the rate at which these novel host-cells "invade" the environment (Morley et al., 2015; Morley & Turner, 2017). In this series of evolution experiments, a novel host cell type was introduced into the environment at different rates, varying from a gradual increase in the proportion of the novel cell type at each passage to a sudden change to an environment composed entirely of the novel host. More-gradual changes were associated with virus populations reaching higher fitness in both novel and original hosts, as well as greater convergence among populations fixing the same adaptative mutations (Morley et al., 2015).

Power law relationships are pervasive in physical and biological systems. One such relationship, known as Taylor's power law, or fluctuation scaling law, states that there is a power relationship between the mean and the dispersion (traditionally the variance) of the elements in a system (Taylor, 1961). This law, first described in ecology, arises naturally in many complex systems (Eisler et al., 2008; Fronczak & Fronczak, 2010; Kendal & Jørgensen, 2011; Lazzardi et al., 2023), and its parameters are able to capture both noise (parameter V) and the distribution (or aggregation; parameter β) of a system (Martí et al., 2017; Silva et al., 2024). Notably, Taylor's law has been used to analyze temporal data of the human microbiota, where an individual's negative health status was related to noise and system instability (Martí et al., 2017). Similarly, it was recently found that throughout the course of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the dynamics of transcripts in cells from human intestinal organoids, but not pulmonary cells, are characterized by an increase-decrease-increase pattern of system noise and instability (Silva et al., 2024).

Statistical laws also arise naturally in many other time series processes. For instance, several natural processes exhibit persistent behavior (autocorrelation), which can be seen as how likely a high value is to be followed by another high value (likewise for low values) (Hurst, 1951; Mandelbrot & Wallis, 1969; Feder, 1988; Ghorbani et al., 2018; Oteo & Oteo-García 2022). This can be measured by a parameter known as the Hurst exponent (denoted H throughout this work), where $0 < H < 0.5$ indicates anti-persistent

(negative autocorrelation); $H = 0.5$ indicates the process is a random walk; and $0.5 < H < 1$ indicates the presence of persistent behavior. Interestingly, many natural processes have an estimated $H \sim 0.7$, which is termed as the Hurst phenomenon (Hurst, 1951).

Here we estimate population genetic parameters such as the selection coefficient (s) and effective population size (N_e) and describe alleles dynamics in experimentally evolving SINV populations under two different temporal schemes of host replacement. Viral populations experiencing a sudden replacement of a highly susceptible host by a less permissive one showed smaller N_e values that indicate stronger bottlenecks and genetic drift, prompting us to investigate how noise and selection affect the dynamics of virus molecular adaptation to novel hosts. First, we show that in the gradual replacement of the highly susceptible host regime, the genetic composition of the viral populations changed steadily until the host was completely removed from the environment. In contrast, in the sudden treatment, the shift in population composition was overall greater and less steady. Then we characterize Taylor's law for allele frequency and a power relationship between alleles' entropy and their absolute difference to the next time point. These power models allowed us to investigate more closely how selection and noise caused by genetic drift affect the dynamics of molecular adaptation. And lastly, we characterize the persistent behavior of the placement of mutations in SINV genomes and found evidence that, in the sudden treatment, mutations become more randomly distributed earlier.

Materials and methods

Description of the study system and data acquisition

We analyzed data from an experimental evolution study that tracked the molecular adaptation of SINV (species *Alphavirus sindbis*, genus *Alphavirus*, family *Togaviridae*) typically cultured on a highly susceptible host, BHK-21 cells, while challenged to infect a less-susceptible host: CHO cells, which were genetically modified (pgsD-677 ATCC CRL-2244) to be more resistant to SINV infection (Morley et al., 2015). Briefly, SINV populations were evolved through 25 passages (~100 virus generations) in monolayer cell cultures with a multiplicity of infection (MOI) of ~0.01 plaque-forming units (pfu) per cell at each passage. Two different treatments from the original study (Morley et al., 2015; Morley & Turner, 2017) were chosen for an in-depth analysis to investigate the

118 link between population parameters and system dynamics of molecular adaptation to a
119 novel host. The initial stock was prepared by expression of an infectious clone in BHK-
120 21 cells for 24 h (Rice et al., 1987). In the gradual treatment, the proportion of CHO cells
121 in the cell cultures increased at each passage, reaching 100% at the last passage. In
122 contrast, in the sudden treatment, cell cultures were entirely composed of CHO cells from
123 the first passage. For each treatment, nine populations were evolved and sequenced at
124 passages 4, 7, 10, 13, 16, 19, 22, and 25. Following RNA extraction, two technical
125 replicates were prepared for each sample from the reverse transcription step onward.
126 Sample preparation is fully explained in Morley & Turner (2017).
127 The final allele frequency tables for each sample were obtained from Morley &
128 Turner (2017). Briefly, reads were trimmed with cutadapt v1.8.3 (Martin, 2011), aligned
129 to the consensus sequence of the original stock with BWA v0.7.10 (Li & Durbin, 2009)
130 and variant calling was conducted with QUASR v7.01 (Gaidatzis et al., 2015) and
131 VarScan v2.3.9 (Koboldt et al., 2009). Only variants with a frequency higher than 1%
132 and present in both technical replicates were kept.
133
134 **Wright-Fisher approximate Bayesian computation estimates of selection coefficients**
135 **and effective population sizes**
136 The selection coefficient (s) and effective population size (N_e) were estimated for each
137 allele with approxwf (Ferrer-Admetlla et al., 2016), which uses a discrete approximation
138 of the Wright-Fisher diffusion model and Bayesian inference to estimate population
139 parameters. The mean values of s and N_e from their posterior distributions were used in
140 downstream analyses.
141
142 **Genetic diversity within evolving viral populations**
143 To measure the genetic distance between two viral populations from consecutive time
144 points, the allele frequency difference (AFD) (Berner, 2019) was calculated at each
145 polymorphic site and averaged by the richness of the sample with custom R scripts, where
146 n represents the total number of different alleles observed at the polymorphism and the f_i
147 terms are the proportion of allele i in the two viral populations.

148 $$AFD = \frac{1}{2}\sum_{i=1}^{n} |(f_{i1} - f_{i2})|$$

149

Taylor's fluctuations power law

Mean and variance of allele frequencies (p) across time ($\bar{p}$ and σ_p^2, respectively) were computed and used to estimate Taylor's parameters V and β by linear regressions in the log-log space: $\log(\sigma_p^2) = \log(V) + \beta\log(\bar{p})$ (Taylor, 1961). Given that the proportion of zeros had a significant influence on the fits, Taylor's parameters were estimated independently for alleles grouped by their number of non-zero occurrences. Multivariate analysis of variance (MANOVA) of Taylor's parameters were performed with parameters estimated from at least five observations. The magnitude of effects was evaluated using the η_P^2 statistic. Conventionally, $0.01 < \eta_P^2 \leq 0.14$ are considered as medium effects and $\eta_P^2 > 0.14$ as large ones. Three random walk matrices were constructed based on the cumulative sum of random variables drawn from a Gaussian distribution with a mean of zero and standard deviation of 0.2, 0.05 and 0.01, respectively. For each matrix, 1000 random walks with eight time points were generated, and values below 0.01 were set to zero and those above 0.99 were set to one. After eliminating walks with only zeros, 792, 757 and 587 random walks were present in the matrices with standard deviation 0.2, 0.05 and 0.01, respectively.

Long-range dependence of mutation sites

To characterize the long-range dependence behavior of mutations along the SINV genome, rescaled-range analyses were performed on the placement of mutations on each time point with the R package pracma v2.4.2 (Borchers, 2022). For this analysis, binary sequences of length 11,703 (which is the genome length) with ones in the position of mutations were used to estimate the empirical Hurst's exponent (H). The minimum window size was set to 2000 to avoid windows with only zeros. Indels were excluded from this analysis.

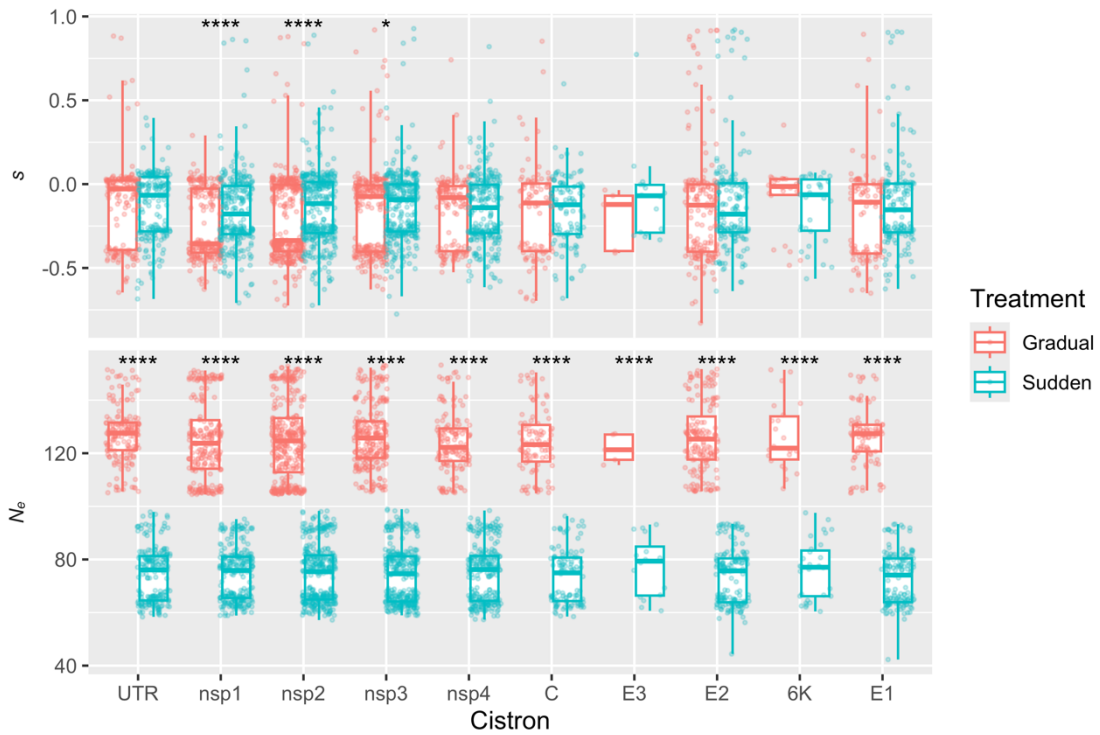
Statistical analyses

All the statistical analyses indicated above were performed with R version 4.4.0 under RStudio version 2024.04.2+764.

Results and Discussion

Sudden host transition is associated to stronger bottlenecks and more drift

183 The selection coefficient (s) and effective population size (N_e) were estimated for each
184 allele of the evolving viral populations (Figure 1). Regardless of the cistron, the sudden
185 host transition treatment consistently resulted in ~ 2 times smaller N_e values. Because s
186 and N_e can influence each other, we tested for the effects of cistron, treatment, population
187 and their interactions on both s and N_e parameters. All independent variables and their
188 interactions, with the exception of the interaction between treatment and cistron, were
189 significant (MANOVA; $P < 0.0001$ for treatment, cistron and population, and $P = 0.0012$
190 for the interaction between cistron and population). By analyzing s and N_e independently,
191 treatment had the relatively greater effect on N_e but very small effect on s ($\eta_P^2 = 0.99$ for
192 N_e and $\eta_P^2 = 0.01$ for s).
193



194
195 **Figure 1.** Population parameters s and N_e for each allele across all populations. Boxplots
196 of the mean values of the posterior distribution of s and N_e for each allele, displayed by
197 treatment and cistron. Asterisks represent the significance of Mann-Whitney two-
198 samples tests: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001.
199
200 Sudden host transition was associated with stronger (more extreme) bottlenecks, that
201 led to reduced N_e, which in turn increased the effects of genetic drift. Interestingly, in
202 contrast to the gradual treatment, cistron had no effect on s (ANOVA; $P < 0.0001$ for the
203 gradual and $P = 0.3400$ for the sudden treatments, respectively). An explanation could

be that under the bottleneck, the effect of selection across different regions may be weaker or less discernible, while was more apparent when the population was less affected by drift as in the gradual treatment. However, despite this difference, the effect size of cistron on *s* in the gradual treatment was still moderate (η_P^2 = 0.03).

Sudden treatment is associated to more instability in the composition of populations and BHK-21-specialists persisted until the complete elimination of BHK-21 cells from the environment in the gradual treatment

Changes in the composition of the populations between consecutive passages were evaluated by summing the allele frequency differences (AFD) (Berner, 2019) of all alleles. Overall, the composition of the populations was more variable (less steady) in the sudden treatment (Figure 2), most likely due to greater effects of genetic drift. In the gradual treatment, most populations changed steadily until the last passage. At this point, a large shift in the composition of the populations was seen (Figure 2a).

The rate at which the populations diverged from the ancestral sequence also differed between treatments (Figure 2a). While in sudden treatment a trend of constant differentiation from the ancestral is seen, in the gradual treatment, we have that, at some points, populations do not diverge from or even become more similar to the ancestral. This phenomenon, known as leapfrog effect (Gerrish & Lenski, 1998), happens when one haplotype supersedes another and both are evolved from the same common ancestor. Since both haplotypes are more similar to their ancestor than they are to one another, the population can become more similar to the ancestral population as it evolves.

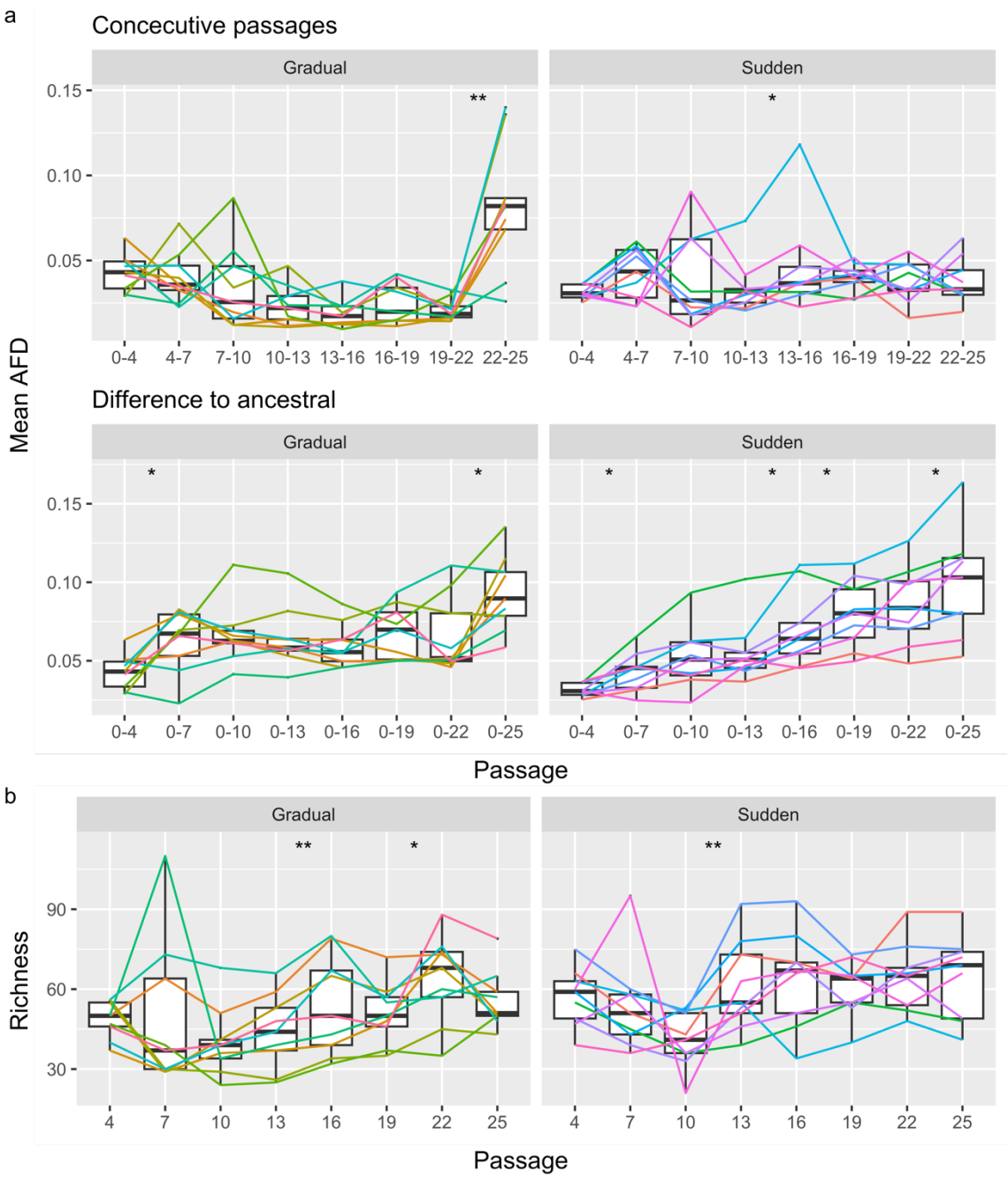


Figure 2. Evolution of the populations' composition. (a) Boxplots of the mean AFD between two consecutive passages and between each passage and ancestral sequence, also showing the mean AFD for each population represented by lines. (b) Boxplot of the richness of each passage. Lines represent the richness of each population, and asterisks represent the significance of paired samples Wilcoxon tests: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001.

We hypothesize the co-occurrence of BHK-21-specialist, CHO-specialist and generalist haplotypes in the gradual treatment. Clonal interference and competition are

237 expected between and within generalists and specialists, but less so between haplotypes
238 specialized in different host cells. Clonal interference occurs when beneficial mutations
239 are lost due to the competition with other beneficial mutations that are present in other
240 haplotypes. Here, it is more likely that beneficial mutations on generalists will be lost
241 due to competition with specialists, although the opposite may also happen specially if
242 there is no trade-off in the evolution of generalists. The large shift in population
243 composition seen at the last passage in the gradual treatment suggests that BHK-21-
244 specialists became extinct once BHK-21 cells were completely absent from the
245 environment. We noted, however, that two populations did not seem to follow this trend,
246 at least as strongly, suggesting that in those cases, a large fraction of high-fitness
247 generalists could have emerged (Figure 2a).
248 To better understand the evolution in the composition of the populations, the richness
249 of each sample, defined as the number of polymorphic sites, was analyzed (Figure 2b).
250 In particular, we were interested to investigate whether the elimination of BHK-21-
251 specialists was, as expected, associated with a loss in richness. Interestingly, we noted
252 that in the sudden treatment, a loss of richness is initially seen followed by a relatively
253 high richness that is maintained from passage 13 onward. This observation, together with
254 the fact that at the same time populations are evolving with high variability, suggests that,
255 between passages, there is a high number of mutations that are lost but are being
256 replenished by new ones. Focusing again on the gradual treatment, while a drop in
257 richness can be seen at the last point, it is not significant. Thus, we found that the apparent
258 loss of BHK-21-specialists does not fully explain the large shift in population
259 composition observed at the last passage in the gradual treatment.
260 We reanalyzed data of the fitness of the evolved populations on BHK-21 and CHO
261 cells to better contextualize the results obtained above (Figure 3). The fitness of
262 populations evolved in the sudden treatment is higher in CHO cells than in BHK-21 cells,
263 consistent with populations majorly composed by CHO-specialist haplotypes. However,
264 no significant difference in fitness was seen for populations evolved in the gradual
265 replacement regime. Given that there is no trade-off between hosts in these populations
266 (Morley et al., 2015), our results suggested that, for the gradual treatment, the fitness of
267 CHO-specialists plus generalists in CHO cells is similar to the fitness of generalists in
268 BHK-21 cells. Thus, gradual host replacement is associated not only with the emergence
269 of generalist populations but also of high-fitness generalist haplotypes.
270

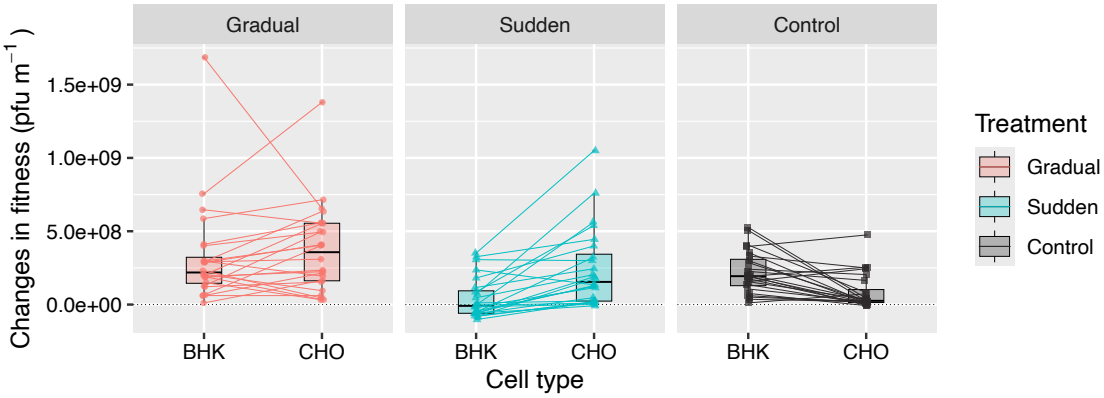


Figure 3. Effect of host replacement regime on fitness. Boxplots of the change in fitness relative to the ancestral for control (lineages evolved in BHK-21 cells), sudden and gradual treatments. Lines represent lineages, and asterisks represent the significance of paired Wilcoxon tests: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001.

Allele frequencies display large fluctuations and aggregation behavior

Morley et al. (2017) observed that some positively selected mutations were lost after reaching intermediate allele frequencies, suggesting a clonal interference effect, which was particularly pronounced in the gradual treatment lineages. Here, we sought to determine whether temporal fluctuations in allele frequencies followed Taylor's power law. We hypothesize that due to the larger effect of drift on smaller populations, alleles in the sudden treatment will fluctuate closer to what is expected by the Poisson distribution, and thus, present smaller β than alleles in the gradual treatment. Additionally, large fluctuations due to clonal interference, especially for those alleles that reached intermediate or high frequencies, are also expected in the gradual treatment. Inspection of the Taylor's plots showed that the fraction of zeros had a drastic impact on the fit (Figure 4), which is to be expected given that only eight time points are present. Here, it is important to note that the presence of zeros might be either due to biological reasons, *i.e.*, the allele was not present at that time point, or due to technical noise (measurement error). The latter is more likely for low frequency alleles that fluctuate close to the detection limit.

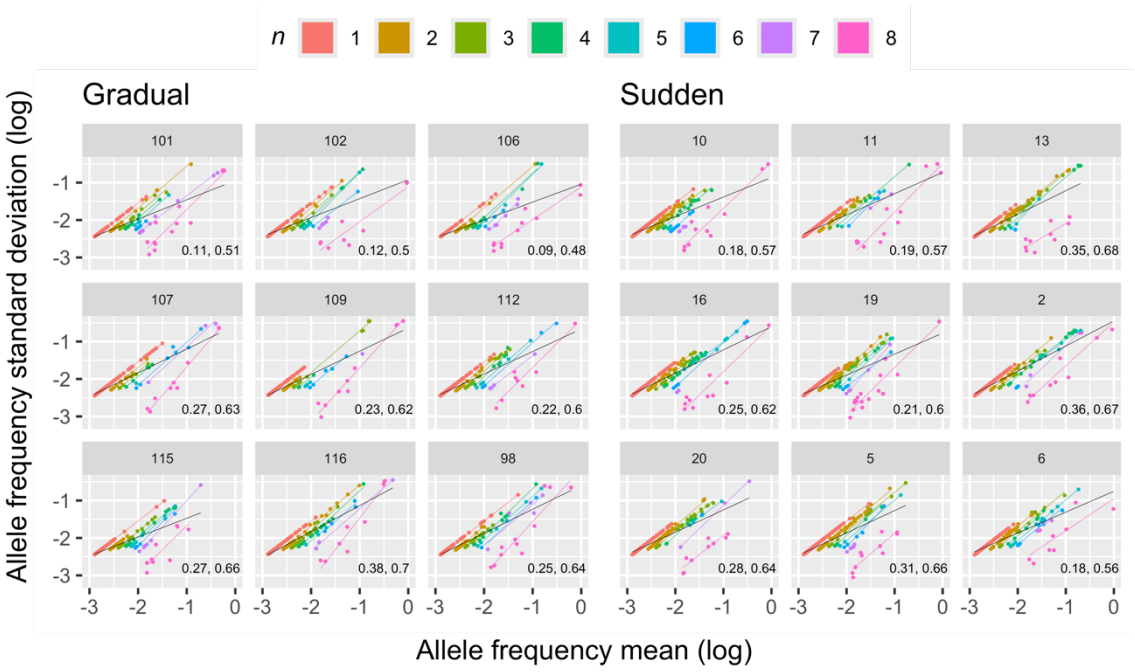


Figure 4. Taylor's law plots. Each panel corresponds to one population, where each point corresponds to the mean and standard deviation of an allele frequency across all passages. Alleles are colored by their number of non-zero occurrences (n), and for each n, a colored line represents the fit to Taylor's law. A black line represents the fit to Taylor's law for all alleles. Taylor's parameters V and β, respectively, for the fit without adjusting for n, are shown.

Fits to Taylor's law were then performed using alleles grouped by their number of non-zero occurrences (n) to analyze Taylor's parameters V and β, henceforth V_n and β_n, and how they scale by host-change regime. In all cases, β_n was estimated to be close to or higher than one (Figure 5a). When performing fits by the number of zeros, we expect that, as the mean allele frequency increases, we tend to observe aggregation since values will be distributed across non-zero spaces. However, parameter β_n is still be able to capture how these values are distributed and measure the strength of aggregation, with an important caveat. Due to the limits of the system (maximum frequency = 1), early fixation of an allele will make it so that it will exhibit high mean frequency and relatively low standard deviation, given that most time points will be ones. This, in turn, will make drive β_n down, making it resemble a Poisson distribution. Late fixation (and clonal interference) will conversely drive β_n up. Thus, this parameter must be analyzed in respect to its deviation from 1, which is the expected for random fluctuations without

fixation. In other words, mean frequencies and their fluctuations will occupy different regions in respect to $\beta_n = 1$ (Supplementary Figure S1a)

To test this, three simulated matrices of random walks were generated based on the cumulative sum of random variables drawn from a normal distribution with mean zero and standard deviation of 0.2, 0.05 and 0.01, respectively (Supplementary Figure S1b), where the higher the standard deviation, higher the fluctuation between time points. We then performed fits to the simulated random walks and observed that, in this case, β_n tends to be lower than one when the frequency of zeros reduces (Figure 5a). Additionally, β_n decrease as fluctuation increases, indicating that higher fluctuations are associated to lower β_n. This is due to alleles occupying different regions of the Taylor's plot depending on the strength of their fluctuations (Supplementary Figure S1c). As expected, when fitting to Taylor's law without adjusting for n, random matrices behave close to a Poisson distribution ($\beta \sim 0.5$), where parameter V increases with larger fluctuations.

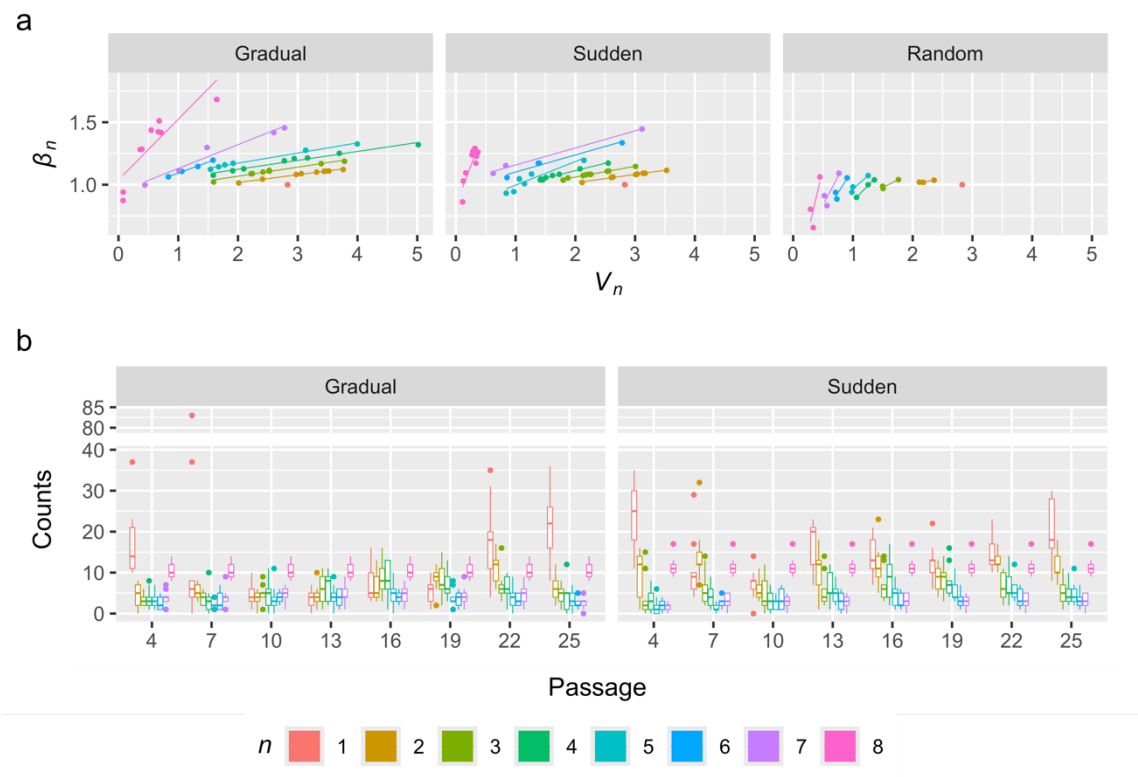


Figure 5. Distribution of alleles frequencies across passages by their n. (a) Taylor's parameters estimated from alleles frequencies for each n for the gradual and sudden treatments and random walk matrices. (b) For each passage, boxplots of the distribution of alleles at each category of n.

To better understand whether frequencies are distributed across the sequenced passages, the number of alleles separated by *n* at each passage and plotted (Figure 4b). In the sudden treatment, a higher concentration of alleles with an *n* of one or two at the first passages is detected, in agreement with a strong bottleneck at the initial passages as the virus populations are introduced to a completely new host CHO. The number of alleles with an *n* of one or two starts to rise again in the sudden treatment at passage 13; and overall, a higher number of alleles with an *n* of one or two across all passages is seen in comparison with the gradual treatment (Figure 5b). This is in accordance with the expected effects of genetic drift, and also with the continuous emergence and loss of new mutations in these shifting populations that was discussed above. In the gradual treatment, alleles with an *n* of one or two are also concentrated at the first and last passages. However, in comparison to the sudden treatment, they are less concentrated at the first passages and their number only starts to rise again at passage 22 (Figure 4b).

Effects of treatment and selection coefficient on alleles' fluctuation

Next, we tested the effect of treatment, population and *n* on the distribution of alleles accounting for all interactions between these variables. The number of non-zero occurrences *n* had the highest influence on Taylor's parameters V_n and β_n (MANOVA; $P < 0.0001$; $\eta_P^2 = 0.92$). Both the main effect treatment and the interaction between treatment and *n* had a significant influence on Taylor's parameter (MANOVA; respectively, $P = 0.0023$ and $\eta_P^2 = 0.14$, and $P = 0.0133$ and $\eta_P^2 = 0.10$), which indicates that host-change rate has a significant and large impact on the distribution of allele frequencies. The interaction between *n* and population was also significant (MANOVA; $P = 0.0212$ and $\eta_P^2 = 0.25$), but not the main effect population alone ($P = 0.0641$ and $\eta_P^2 = 0.22$).

Alleles were divided into three categories, neutral, deleterious and beneficial, to determine whether strength of selection has an impact on Taylor's parameters. By visualizing the distribution of *s* (Figure 6a) it becomes clearer that *s* follows a mostly bimodal distribution with a peak for negatively selected alleles and another for neutrally selected ones around zero, with only a fraction being positively selected. Interestingly, negatively selected alleles on the gradual treatment seem to be subjected to stronger negative selection. Based on these distributions, we set a threshold of -0.16 for alleles to be considered as negatively selected and 0.16 for them to be considered positively selected, whereas alleles with *s* between -0.16 and 0.16 are considered as neutral.

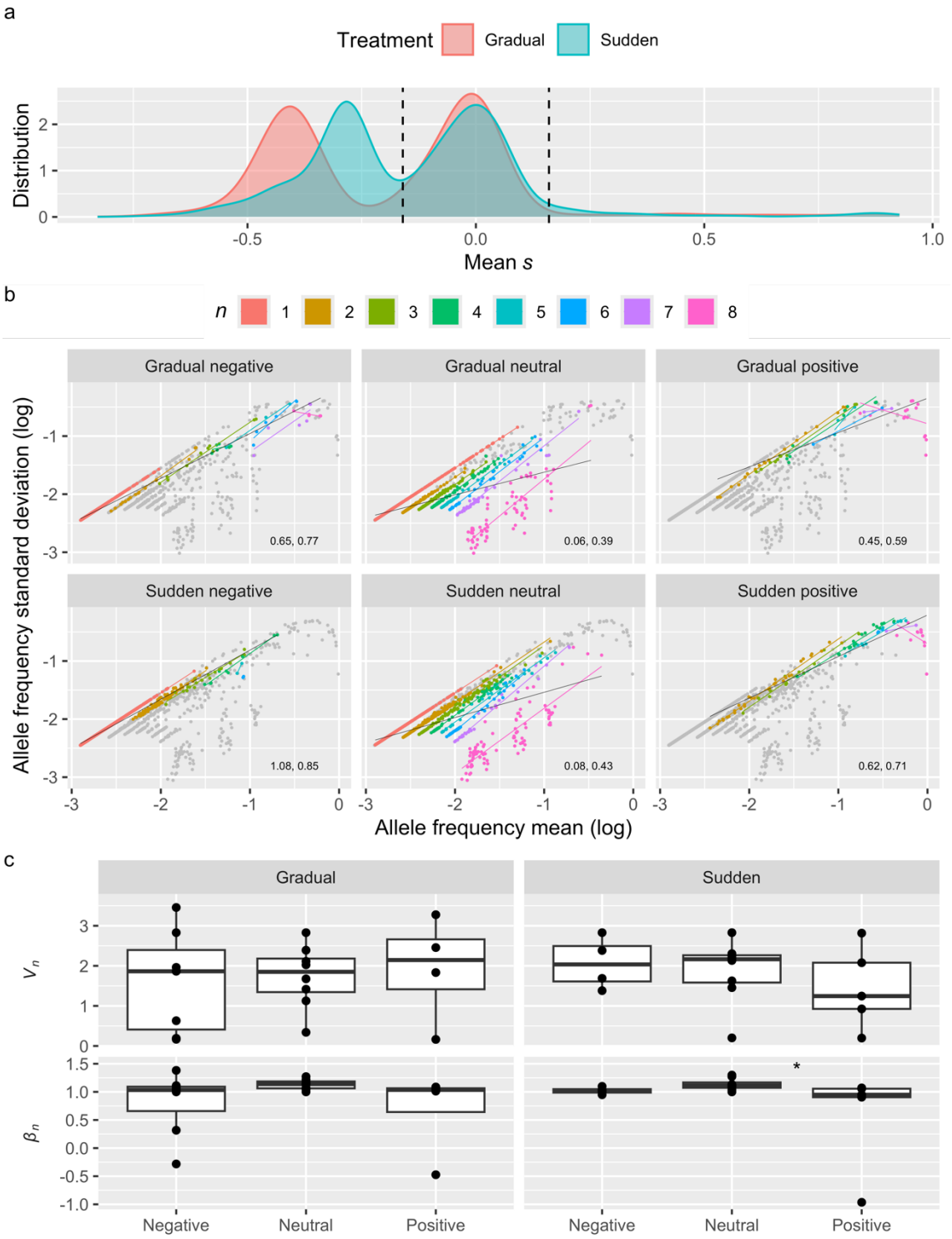


Figure 6. Effects of selection on allele fluctuations. (a) Distribution of *s* per treatment. Dashed lines represent the thresholds for categorizing neutrally, negatively and positively selected alleles. Taylor's parameters *V* and *β*, respectively, for the fit without adjusting for *n*, are shown. (b) Taylor's plots of combined allele frequencies across populations by treatment and *s*. Alleles are colored by their number of non-zero occurrences (*n*), and for

each n, a colored line represents the fit to Taylor's law. A black line represents the fit to Taylor's law for all alleles. Taylor's parameters V and β, respectively, for the fit without adjusting for n, are shown. (c) Boxplots of V_n and β_n by treatment and s category. Asterisks represent the significance of unpaired Wilcoxon tests: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001.

Given that the separation of alleles by s drastically reduces sample size in some populations and n combinations, fits to Taylor's law were performed by treatment, (where frequencies from all population were combined), selection and n (Figure 6b). Still, sample size drastically affected the fits, especially in cases where alleles with low mean were not present (*e.g.*, positively selected alleles with n = 8) to accurately scale how standard deviation varies with the mean. Additionally, some n categories are missing entirely for either positively or negatively selected alleles depending on treatment. Again, n had the highest influence on Taylor's parameters V_n and β_n (MANOVA; $P < 0.0001$; $\eta_P^2 = 0.76$), followed by the interaction between n and s ($P = 0.0017$; $\eta_P^2 = 0.3$), and s ($P = 0.0309$; $\eta_P^2 = 0.2$). Here, treatment did not have a significant impact on Taylor's parameters. It is likely that by combining the allele frequencies across populations the analysis became confounded. This is supported by the fact that the interaction between n and population was significant when Taylor's parameters were estimated separately for each population.

Despite not finding a significant influence of treatment on Taylor's, some very interesting observations can be made from these Taylor's plots in regard to s. Firstly, points are far more dispersed for neutrally selected alleles, and occupy a region composed by alleles that are present at several time points but have low mean frequency and low standard deviation. Most notably, all alleles with an n of 8 that have low mean frequency and standard deviation are neutrally selected. Due to this, if the data is fitted to Taylor's law without adjusting for n, neutrally selected alleles will be the ones closer to a Poisson distribution. Also, neutrally selected alleles exhibit the highest values of β_n (Figure 6c). Secondly, positively selected alleles with an n of 8 have negative β_n and their standard deviation decreases as their mean frequency increase. This is caused by early fixation causing most time points to be ones, and consequently, their standard deviation will be low. When fitting the data to Taylor's law without adjusting for n, this drives β down, and as a result, β is lower for positively selected alleles in comparison with negatively

409 selected ones. Additionally, positively selected alleles occupy a region of the Taylor's
410 plot close to the upper limits of the system, especially for higher *n*. Thirdly, negatively
411 selected alleles occupy a region of the Taylor's plot between neutrally and positively
412 selected alleles. This difference is clearer for the sudden treatment.
413
414 **Host replacement regime affects the distribution of mutations along SINV genome**
415 Lastly, it is likely that host change dynamics may also have an effect not only on the
416 aggregation dynamics of allele frequencies but also on the placement of mutations due to
417 the interplay between drift and selection and linkage. Due to genetic hitchhiking (Smith
418 & Haigh, 1974; Gillespie, 2000; Schiffels et al., 2011; Morley & Turner, 2017), neutral
419 or low fitness mutations may benefit from being in the proximity of high fitness ones. In
420 addition, the gradual host replacement is associated with greater clonal interference
421 (Morley & Turner, 2017), in which genomes containing multiple beneficial mutations
422 outcompete those containing only one, promoting the fixation of linked mutations (Park
423 & Krug, 2007; Good et al., 2012; Morley & Turner, 2017). Given that the fixation of
424 sweeps of mutations is greater in the gradual treatment (Morley & Turner, 2017), we
425 sought to investigate whether mutations in the gradual treatment are less evenly
426 distributed along the genome, meaning if they are aggregating.
427 For this, we estimated the empirical Hurst's exponent *H* to measure persistent
428 behavior in the placement of mutations at each passage for each population (Figure 7).
429 Here, *H* was estimated from binary sequences with the same length as SINV genome,
430 where ones represent sites with any disagreement to the ancestral sequence. Despite high
431 variability in the estimation, especially for the sudden treatment, evidence of persistent
432 behavior was found in most cases, were the median value at each passage for both
433 treatments were always above 0.5 (Figure 7). At the first two initial passages, median *H*
434 is higher for the sudden treatment, follow by two passages where *H* is more or less similar
435 between treatments, and finally, *H* decreases and becomes smaller on the sudden
436 treatment (despite these differences are only significant in passage 22, $P = 0.040$). This
437 suggests that the continuous emergence and loss of mutations on the sudden treatment
438 that was previously observed at the final passages is associated with more random
439 scattering of mutations along the genome, and might be a big factor on why mutations are
440 more randomly positioned on this treatment. However, we found that treatment did not
441 have a significant influence on *H* (linear model with passage and treatment as orthogonal

fixed effects and population as a random effect nested within treatment; sudden coefficient estimate = –0.043, $P = 0.100$).

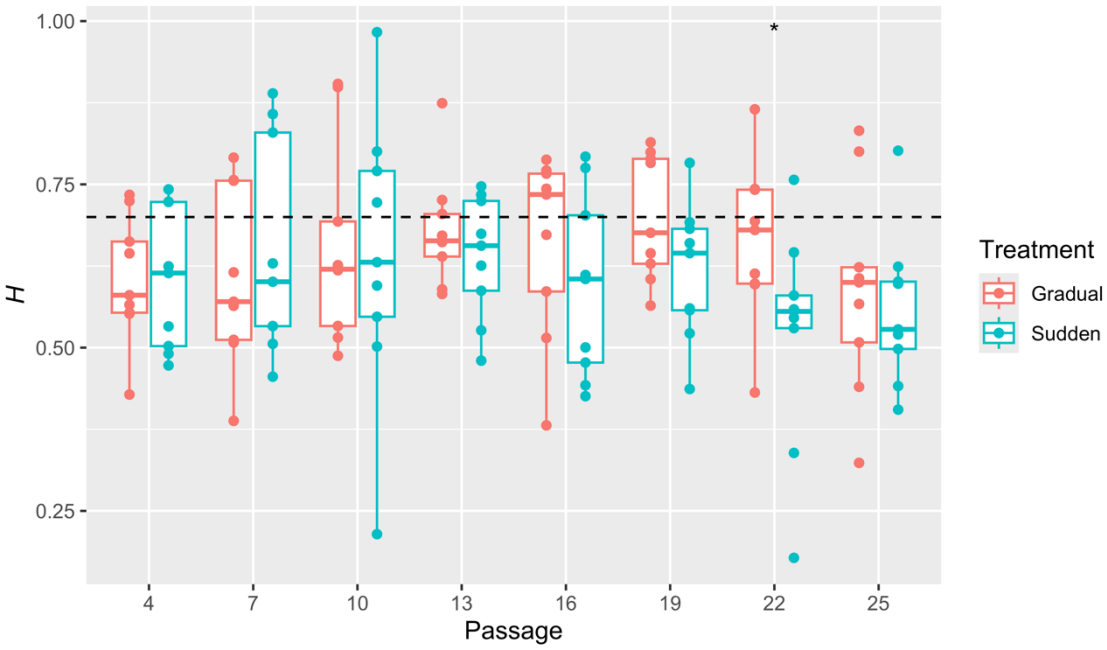


Figure 7. Persistent behavior in the placement of mutations. Hurst exponents (*H*) estimated from binary sequences (where 1 represent a mutant allele) for each population at each passage. Dashed line represents the 0.7 Hurst phenomenon threshold. Asterisks represent the significance of unpaired Wilcoxon tests: * < 0.05, ** < 0.01, *** <0.001 and **** < 0.0001.

Conclusions

In conclusion, our study demonstrates that the dynamics of viral adaptation are heavily influenced by the rate of environmental change, specifically the pace of host replacement. Gradual changes in host composition foster more stable evolutionary trajectories, with higher fitness convergence and more persistent mutation patterns, while sudden transitions lead to greater genetic drift and instability. Most interesting is how genetic drift is affecting the populations evolved in the sudden host replacement, where instability seems to be associated with a large number of new mutations arising and being lost after an initial bottleneck, which could be related to the quasispecies structure of these populations. These findings highlight the critical role of current versus novel host availability in shaping the adaptive landscape of viral populations, offering insights into

how pathogens might evolve in response to varying degrees of ecological disruption. Future research could further explore these dynamics across different viral species and host systems to generalize our understanding of viral evolution in changing environments.

Supplementary material

Supplementary material is available online at *Molecular Biology and Evolution*.

Data availability

The results of approxwf and alleles frequencies (Excel file) are available at the Zenodo repository https://doi.org/10.5281/zenodo.14557803. R scripts used to perform the analyses are available at https://www.github.com/jmfagundes/SINVevo.

Author contributions

J.M.F.S. conducted the analyses, and wrote and edited the article with input from the other authors. M.J.O.U participated in the conception of methodological approaches, conducted the analyses and edited drafts of the paper. V.J.M. collected the data and edited drafts of the paper. P.E.T. collected the data and edited drafts of the paper. S.F.E. conceived the study, conducted the analyses, and edited drafts of the paper.

Funding

This work was supported by grant PID2022-136912NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF a way of making Europe" and by Generalitat Valenciana grant CIPROM/2022/59 to S.F.E. M.J.O.U. was supported by grant FPU2019/05246 funded by MCIN/AEI/10.13039/501100011033 and by "ESF investing in your future".

Acknowledgements

We thank Dr. Carlos P. Garay and Dr. José A. Oteo for very fruitful discussions on the meaning of Taylor's power law and Hurst's exponent in evolving complex systems.

Conflict of interests: The authors declare no conflict of interest.

References

Berner, D. (2019). Allele frequency difference AFD–an intuitive alternative to F_{ST} for quantifying genetic population differentiation. *Genes*, *10*, 308.

Blumm, N., Ghoshal, G., Forró, Z., Schich, M., Bianconi, G., Bouchaud, J. P., & Barabási, A. L. (2012). Dynamics of ranking processes in complex systems. *Physical Review Letters, 109*, 128701.

Borchers, H. (2022). pracma: Practical Numerical Math Functions. Version 2.4.2 [R package]. Available from: https://CRAN.R-project.org/package=pracma

Eisler, Z., Bartos, I., & Kertész, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics*, *57*, 89–142.

Feder, J. (1988). *Fractals*. New York: Plenum Press

Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D., & Wegmann, D. (2016). An approximate Markov model for the Wright-Fisher diffusion and its application to time serie data. *Genetics*, *203*, 831–846.

Foll, M., Shim, H., & Jensen, J. D. (2015). WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, *15*, 87–98.

Fronczak, A., & Fronczak, P. (2010). Origins of Taylor's power law for fluctuation scaling in complex systems. *Physical Review E*, *81*, 066112.

Gaidatzis, D., Lerch, A., Hahne, F., & Stadler, M. B. (2015). QuasR: quantification and annotation of short reads in R. *Bioinformatics*, *31*, 1130–1132.

Gerrish, P. J., & Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica, 102*, 127–144.

Ghorbani, M., Jonckheere, E. A., & P. Bogdan. (2018). Gene expression is not random: scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. *Frontiers in Physiology*, *9*, 1446.

Gillespie, J. H. (1984). Molecular evolution over the mutational landscape. *Evolution*, *38*, 1116–1129.

Hurst, H. E. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770–799.

Jørgensen, B., Martínez, J. R., & Tsao, M. (1994). Asymptotic behavior of the variance function. *Scandinavian Journal of Statistics*, *21*, 23–243.

Kendal, W. S., & Jørgensen, B. (2011). Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Physical Review E*, *83*, 066115.

Kendal, W. S., & Jørgensen, B. (2011). Tweedie convergence: A mathematical basis for Taylor's power law, 1/*f* noise, and multifractality. *Physical Review E*, *84*, 066120.

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., & Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, *25*, 2283–2285.

Lazzardi, S., Valle, F., Mazzolini, A., Scialdone, A., Caselle, M., & Osella, M. (2023). Emergent statistical laws in single-cell transcriptomic data. *Physical Review E*, *107*, 044403.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, *5*, 967–988.

Martí, J. M. (2018). Robust analysis of time series in virome metagenomics. In *The Human Virome: Methods and Protocols*, Moya A. and V. Pérez-Brocal (eds). *Methods in Molecular Biology* vol. 1838, chap. 17, pp. 245–260.

Martí, J. M., Martínez-Martínez, D., Rubio, T., Gracia, C., Peña, M., Latorre, A., Moya, A., & Garay, C. P. (2017). Health and disease imprinted in the time variability of the human microbiome. *mSystems*, *2*, e00144-16.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*, 10–12.

Morley, V. J., Mendiola, S. Y., & Turner, P. E. (2015). Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proceedings of the Royal Society Series B: Biological Sciences*, *282*, 20150801.

Morley, V. J., & Turner, P. E. (2017). Dynamics of molecular evolution in RNA virus populations depend on sudden versus gradual environmental change. *Evolution*, *71*, 872–883.

Oteo, J. A., & Oteo-García, G. (2022). Mutations along human chromosomes: How randomly scattered are they? *Physical Review E*, *106*, 064404.

Rice, C. M., Levis, R., Strauss, J. H., & Huang, H. V. (1987). Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and *in vitro* mutagenesis to generate defined mutants. *Journal of Virology*, *61*, 3809–3819.

Schiffels, S., Szollosi, G. J., Mustonen, V., & Lässig, M. (2011). Emergent neutrality in adaptive asexual evolution. *Genetics*, *189*, 1361–1375.

Silva, J. M. F., Oteo, J. A., Garay, C. P., & Elena, S. F. (2024). System and transcript dynamics of cells infected with severe acute respiratory syndrome virus 2 (SARS-CoV-2). *PLoS Complex Systems*, *1*, e000016.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*, 23–35.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, *189*, 732–735.