Taylor's power law rules the dynamics of allele frequencies during viral evolution in response to host changes

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

[‡]Present address: Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil

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

Author for correspondence: Santiago F. Elena. Email: santiago.elena@csic.es

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

Subject Category: Life Sciences – Physics interface
Subject Areas: Evolution

Abstract: Sudden and gradual changes in host species composition —transitioning from more permissive to less permissive hosts— have been shown to influence viral fitness, virulence, and rates of molecular evolution. In this study, we analyzed the roles of stochasticity and selection using 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 framework revealed a reduction in effective population size by approximately half in scenarios involving sudden host changes. These scenarios were also associated with fewer and generally weaker beneficial mutations. Second, genetic distance analysis between populations at consecutive time points indicated that populations undergoing gradual host changes evolved steadily until the original host was no longer present. Distances to the ancestral sequence in these cases exhibited occasional leapfrog phenomena, where the rise of certain haplotypes is not predictable based solely on their relatedness to previously dominant ones. In contrast, populations exposed to sudden host changes displayed less stable compositions and diverged from the ancestral sequence at a more constant rate. Third, we observed that allele frequencies distribution followed a power law consistent with Taylor's law. Both treatments exhibited high levels of allele aggregation with significant fluctuations, where neutral, beneficial, and deleterious alleles could be distinguished by their general behavior and their position on Taylor's plot. Finally, we found evidence that the host replacement regime influences the temporal distribution of mutations across the genome.

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

1. Introduction

The molecular dynamics of adaptation are heavily influenced by environmental heterogeneity. Pathogens adapt to new hosts by fixing mutations that enhance fitness in the novel host, even if these mutations are neutral or detrimental in the original host. Consequently, the availability of novel hosts in the environment affects both the rate and direction of molecular adaptation. For example, the adaptation dynamics of the Sindbis virus (SINV) to a less permissive cell type in laboratory tissue culture are determined by the rate at which these novel host cells "invade" the environment [1,2]. In a series of evolution experiments, Morley et al. [1] introduced a novel host cell type at varying rates, ranging from a gradual increase in the proportion of the novel cell type with each passage to an abrupt shift to an environment composed entirely of the novel host. Gradual changes were associated with virus populations achieving higher fitness in both novel and original hosts, as well as greater convergence among populations that fixed the same adaptive mutations [1].

Allele frequencies typically fluctuate over time in both natural and experimental populations due to a variety of factors, including purely stochastic processes (*e.g.*, genetic drift, fluctuating selection, migration, or unpredictable ecological changes) and deterministic processes (*e.g.*, directional selection). Indeed, temporal fluctuations are ubiquitous in physical systems, raising the question of whether they follow universal laws. Various power-law relationships have been found to be pervasive in physical and biological systems. One such relationship, known as Taylor's power law —or the fluctuation scaling law— posits that the variance of a system's, σ^2, elements scales as a power of the mean, μ: $\sigma^2 = V\mu^\beta$ [3]. Originally described in ecology, this law naturally arises in many complex systems [4–9]. Its parameters capture both the amplitude of the noise level (V) and the degree of temporal aggregation (β) in the fluctuations observed within the system [5,6,8]. Notably, in the context of infectious diseases, Taylor's law has been applied to temporal data from the human microbiota, revealing that an individual's negative health status is associated with increased noise and system instability [8]. Similarly, recent studies have found that during severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, the transcripts dynamics in cells from human intestinal organoids, but not pulmonary cells, exhibit an increase-decrease-increase pattern in system noise and instability as infection progresses and the virus accumulates [9].

In the context of temporal variations in allele frequencies during virus adaptation to novel hosts, we propose that Taylor's power law can be used to model how the variance of allele frequencies changes over time. In populations influenced solely by stochastic processes (*i.e.*, Exponential), it is expected that $\beta = 2$, meaning that variance scales quadratically with the mean allele frequency. In contrast, a $\beta > 2$ indicates nonrandom processes that amplify variability, such as migration or environmental heterogeneity, whereas a $\beta < 2$ suggests processes that constrain variability, such as stabilizing selection or density dependence (with $\beta = 1$ corresponding to a Poisson process).

Many natural processes exhibit long-term memory or persistent behavior (autocorrelation), meaning that a high value is likely to be followed by another high value (and similarly for low values) [10-15]. This behavior can be quantified using the Hurst exponent (denoted as H throughout this work), where: $0 < H < 0.5$ indicates anti-persistent (negative autocorrelation) behavior, $H = 0.5$ indicates a random walk, and $0.5 < H < 1$ indicates persistent behavior. Interestingly, many studied processes have an estimated $H \approx 0.7$, a phenomenon known as the Hurst phenomenon [10]. Although Taylor's law and the Hurst phenomenon both describe variability and scaling behavior, they focus on different aspects of the process and are indirectly related. Specifically, when Taylor's law is applied to the variance of fluctuations at different time scales, the simple relationship $\beta = 2H$ must hold [11]. This connection arises because both laws describe the fractal or scaling properties of purely stochastic processes.

Here, we estimate population genetic parameters —such as the selection coefficient per allele (s) and effective population size (N_e)— and describe allele dynamics in experimentally evolving SINV populations under two different temporal schemes of host replacement. SINV populations that experienced a sudden replacement of a highly susceptible host with a less permissive one exhibited smaller N_e values, indicating stronger bottlenecks and genetic drift. This observation prompted us to investigate how noise and selection affect the dynamics of virus molecular adaptation to novel hosts. First, we show that under gradual replacement of the highly susceptible host, the genetic composition of the viral populations changed steadily until the host was completely removed. In contrast, under the sudden treatment, the shift in population composition was more pronounced and less consistent. Next, we characterized Taylor's power law for the temporal variation in allele frequency. This power model allowed us to examine in greater detail how selection and noise resulting from genetic drift influence the dynamics of molecular adaptation. Finally, we investigate the persistent behavior

characteristic of the Hurst phenomenon in the distribution of mutations along SINV genomes, finding evidence that, in the sudden treatment, mutations become more randomly distributed earlier in time.

2. Methods

2.1. 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 the less-susceptible host CHO cells, which were genetically modified (pgsD-677 ATCC CRL-2244) to be more resistant to SINV infection [1]. 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 [1,2] were chosen for an in-depth analysis to investigate the link between population parameters and system dynamics of molecular adaptation to a novel host. The initial stock was prepared by expression of an infectious clone in BHK-21 cells for 24 h [16]. In the gradual treatment, the proportion of CHO cells in the cell cultures increased at each passage, reaching 100% at the last passage. In contrast, in the sudden treatment, cell cultures were entirely composed of CHO cells from the first passage. For each treatment, nine populations were evolved and sequenced at passages 4, 7, 10, 13, 16, 19, 22, and 25. Following RNA extraction, two technical replicates were prepared for each sample from the reverse transcription step onward. Sample preparation is fully explained in [2].

The final allele frequency tables for each sample were obtained from [2]. Briefly, reads were trimmed with cutadapt version 1.8.3 [17], aligned to the consensus sequence of the original stock with BWA version 0.7.10 [18] and variant calling was conducted with QUASR version 7.01 [19] and VarScan version 2.3.9 [20]. Only variants with a frequency higher than 1% and present in both technical replicates were kept.

2.2. Wright-Fisher approximate Bayesian computation estimates of selection coefficients and effective population sizes

The selection coefficient (s) and effective population size (N_e) were estimated for each allele with approxwf [21], which uses a discrete approximation of the Wright-Fisher diffusion model and Bayesian inference to estimate population parameters. The implementation assumes a log-Uniform(1, 5) prior distribution on N_e and a Normal(0, 0.05) prior distribution on s [21]. The algorithm runs an MCMC using 51 states for 25,000 interactions, discarding the first 2000 interactions as burn-in. The mean values of s and N_e from their posterior distributions were used in downstream analyses.

2.3. Genetic diversity within evolving viral populations

To measure the genetic distance between two viral populations from consecutive time points, the allele frequency difference (AFD) [22] was calculated at each polymorphic site and averaged by the richness of the sample with custom R scripts, where n represents the total number of different alleles observed at the polymorphism and the f_i terms are the proportion of allele i in the two viral populations.

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

As a second measure of within-sample genetic diversity, we used allele richness, defined as the number of unique alleles per locus, adjusted for sample size.

2.4. Taylor's fluctuations power law

Mean and variance of allele frequencies (p) across time ($\langle p \rangle$ 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: $\log\sigma_p^2 = \log V + \beta \log\langle p \rangle$ [3]. 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 Normal distribution with a mean of zero and standard deviations 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.

2.5. 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 version 2.4.2 [23]. 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.

2.6. Statistical analyses

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

3. Results and Discussion

3.1. A sudden host transition is associated to stronger genetic drift

The selection coefficient (s) and effective population size (N_e) were estimated for each allele of the evolving viral populations (figure 1). Regardless of the cistron, the sudden host transition treatment consistently resulted in ~2 times smaller N_e values. Because s and N_e can influence each other, we tested for the effects of cistron, treatment, population and their interactions on both s and N_e parameters. All independent variables and their interactions, with the exception of the interaction between treatment and cistron, were significant (MANOVA; $P < 0.0001$ for treatment, cistron and population, and $P = 0.0012$ for the interaction between cistron and population). By analyzing s and N_e independently, treatment had the relatively greater effect on N_e but very small effect on s ($\eta_P^2 = 0.99$ for N_e and $\eta_P^2 = 0.01$ for s). However, *post hoc* pairwise tests show that, on average, alleles observed in the sudden transition at the *nsp1*, *nsp2* and *nsp3* cistrons are significantly less deleterious that those found under the gradual transition. The proteins encoded by these three cistrons are involved in the formation of the viral replication complex [24], thus

being expected to be under strong purifying selection. In particular, the product of *nsp3* is involved in host specificity and virulence.

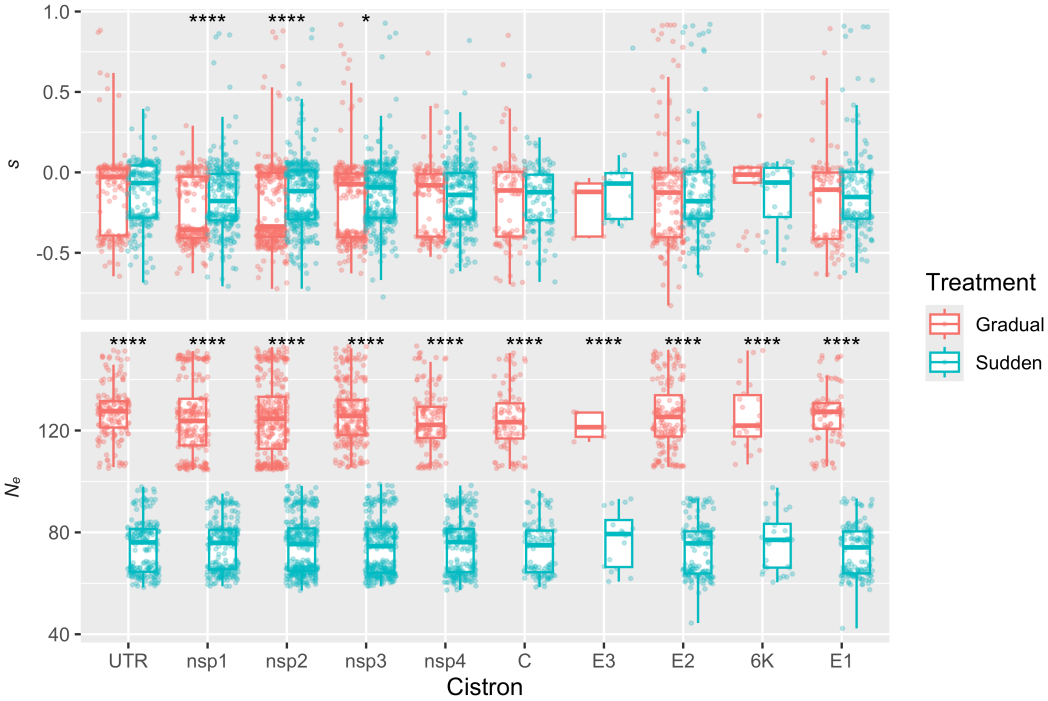

Figure 1. Population parameters s and N_e for each allele across all populations. Boxplots of the mean values of the posterior distribution of s and N_e for each allele, displayed by treatment and cistron. Asterisks represent the significance of Mann-Whitney two-samples tests: * $P < 0.05$ and **** $P < 0.0001$.

Sudden host transition was associated with stronger (more extreme) bottlenecks, that led to reduced N_e, which in turn increased the effects of genetic drift. Interestingly, in contrast to the gradual treatment, cistron had no effect on s (ANOVA; $P < 0.0001$ for the 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 the effect 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 ($\eta_P^2 = 0.03$).

3.2. A sudden host transition is associated to more instability population allele composition

Changes in the composition of the populations between consecutive passages were evaluated by summing the allele frequency differences (AFD) [22] 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 [25], happens when one haplotype supersedes another and both have 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. Five not mutually exclusive mechanisms can be brought forward to explain this leapfrog effect. First, balancing selection that might favor diversity at certain loci, allowing rare haplotypes to gain an advantage under changing environmental or selective conditions [26]. Second, epistasis and hitchhiking in which specific combinations of alleles in rare haplotypes may confer a fitness advantage (epistasis) that becomes beneficial over time, hitchhiking them to dominance [27-29]. Third, environmental fluctuations can alter the fitness landscape, favoring haplotypes that were previously at a disadvantage, even if they are less related to the dominant haplotypes [30-32]. Fourth, genetic drift in small population allows rare haplotypes to randomly increase in frequency, contributing to their rise in dominance over time [33]. And, fifth, lineage sorting, by which low-frequency alleles unrelated to previously dominant ones but linked through deeper evolutionary events [34], and recombination that breaks apart haplotypes and creates new advantageous combinations that are not present in the previously dominant group [35]. Since we have observed the leapfrog phenomena in the gradual transition regime (larger N_e; figure 1), we propose that genetic drift may contribute less than the other four mechanisms.

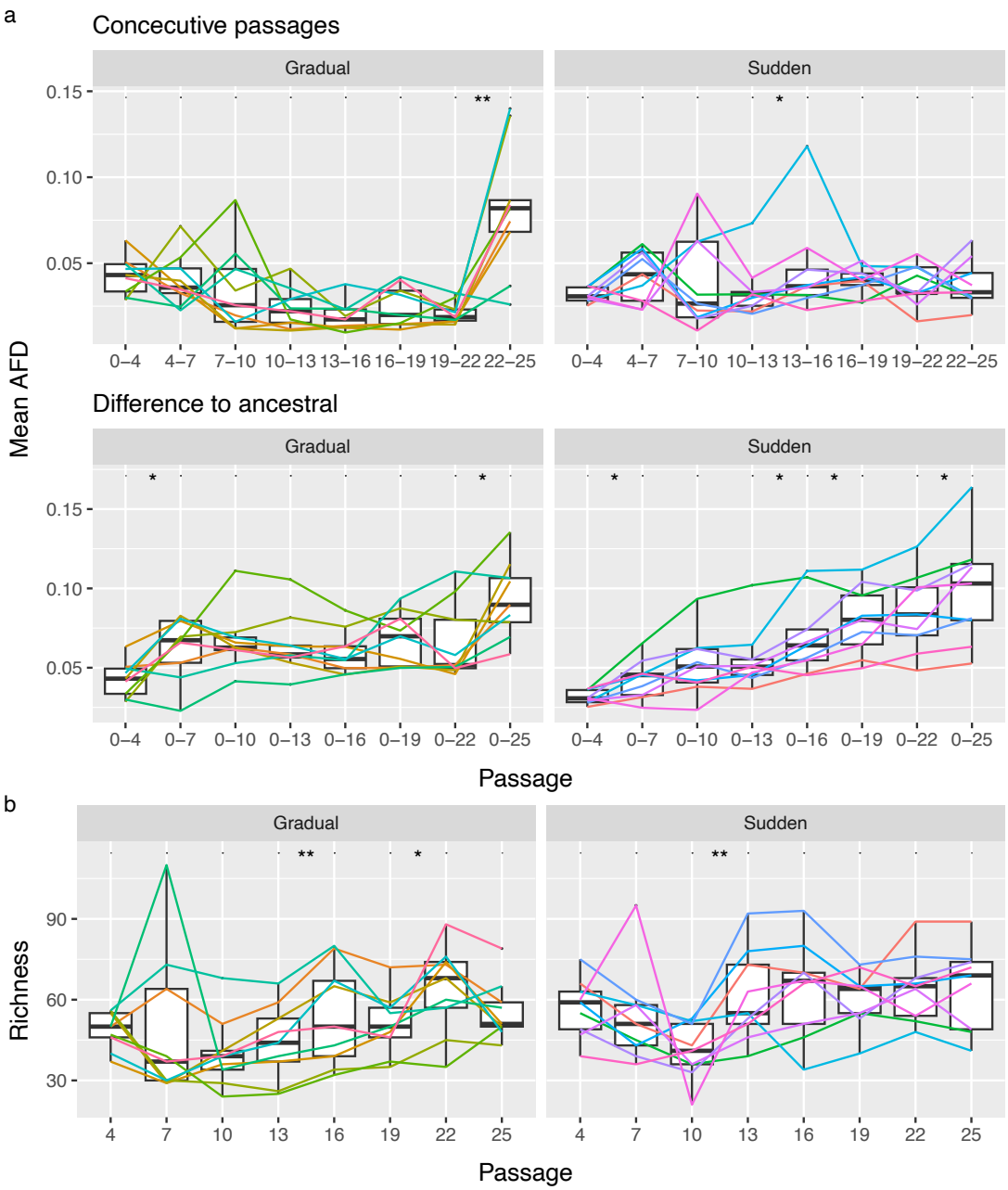

Figure 2. Evolution of the populations' composition. (a) Boxplots of the mean AFD between two consecutive passages and between each passage and ancestral sequence, also showing the mean AFD for each population represented by lines. (b) Boxplot of the richness of each passage. Lines represent the richness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

3.3. BHK-21-specialists persisted until the complete elimination of BHK-21 cells from the environment

267 We hypothesize the co-occurrence of BHK-21-specialist, CHO-specialist and generalist
268 haplotypes in the gradual treatment. Clonal interference and competition are expected
269 between and within generalists and specialists, but less so between haplotypes specialized
270 in different host cells. Clonal interference occurs when beneficial mutations are lost due
271 to the competition with other beneficial mutations that are present in other haplotypes
272 [25,36,37]. Here, it is more likely that beneficial mutations on generalists will be lost due
273 to competition with specialists, although the opposite may also happen, especially if there
274 is no tradeoff in the evolution of generalists [38]. The large shift in population
275 composition seen at the last passage in the gradual treatment suggests that BHK-21-
276 specialists became extinct once BHK-21 cells were completely absent from the
277 environment. We noted, however, that two populations did not seem to follow this trend,
278 at least as strongly, suggesting that in those cases, a large fraction of high-fitness
279 generalists could have emerged (figure 2a).

280 To better understand the evolution in the composition of the populations, the richness
281 of each sample, defined as the number of polymorphic sites, was analyzed (figure 2b). In
282 particular, we were interested to investigate whether the elimination of BHK-21-
283 specialists was, as expected, associated with a loss in richness. Interestingly, we noted
284 that in the sudden treatment, a loss of richness is initially seen followed by a relatively
285 high richness that is maintained from passage 13 onward. This observation, together with
286 the fact that at the same time populations are evolving with high variability, suggests that,
287 between passages, there is a high number of mutations that are lost but are being
288 replenished by new ones. Focusing again on the gradual treatment, while a drop in
289 richness can be seen at the last point, it is not significant. Thus, we found that the apparent
290 loss of BHK-21-specialists does not fully explain the large shift in population
291 composition observed at the last passage in the gradual treatment.

292 We reanalyzed data of the fitness of the evolved populations on BHK-21 and CHO
293 cells to better contextualize the results obtained above (figure 3). The fitness of
294 populations evolved in the sudden treatment is higher in CHO cells than in BHK-21 cells,
295 consistent with populations majorly composed by CHO-specialist haplotypes. However,
296 no significant difference in fitness was seen for populations evolved in the gradual
297 replacement regime. Given that there is no tradeoff between hosts in these populations
298 [1], our results suggested that, for the gradual treatment, the fitness of CHO-specialists
299 plus generalists in CHO cells is similar to the fitness of generalists in BHK-21 cells. Thus,

300 gradual host replacement is associated not only with the emergence of generalist
301 populations but also of high-fitness generalist haplotypes [1].
302

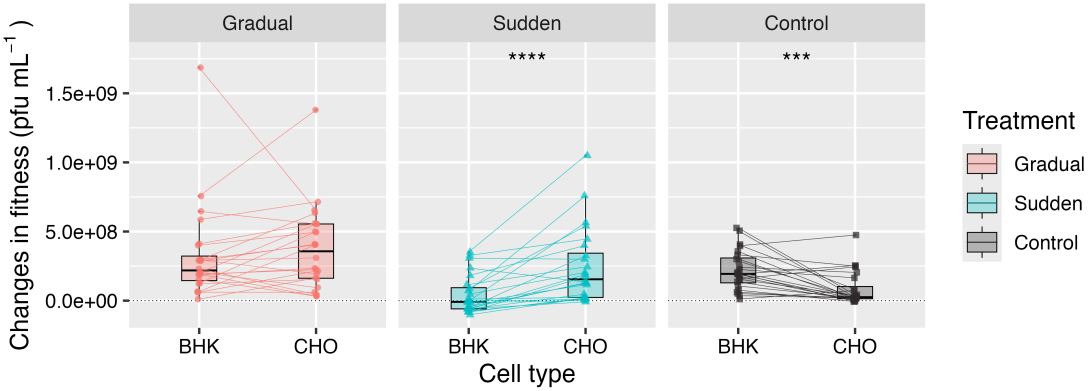

Figure 3. Effect of host replacement regime on fitness. Boxplots of the change in fitness relative to the ancestral for control (lineages evolved in BHK-21 cells), sudden and gradual treatments. Lines represent lineages, and asterisks represent the significance of paired Wilcoxon tests: *** $P < 0.001$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

303

304 **3.4. Allele frequencies display large fluctuations and aggregation behavior**

305 Morley et al. [2] observed that some positively selected mutations were lost after reaching
306 intermediate allele frequencies, suggesting a clonal interference effect, which was
307 particularly pronounced in the gradual treatment lineages. Here, we sought to determine
308 whether temporal fluctuations in allele frequencies followed Taylor's power law. We
309 hypothesize that due to the larger effect of drift on smaller populations, alleles in the
310 sudden treatment will fluctuate closer to what is expected by the Poisson distribution, and
311 thus, present smaller β than alleles in the gradual treatment. Additionally, large
312 fluctuations due to clonal interference, especially for those alleles that reached
313 intermediate or high frequencies, are also expected in the gradual treatment. Inspection
314 of the Taylor's plots showed that the fraction of zeros had a drastic impact on the fit
315 (figure 4), which is to be expected given that only eight time points are present. Here, it
316 is important to note that the presence of zeros might be either due to biological reasons,
317 *i.e.*, the allele was not present at that time point, or due to technical noise (measurement
318 error). The latter is more likely for low frequency alleles that fluctuate close to the
319 detection limit.
320

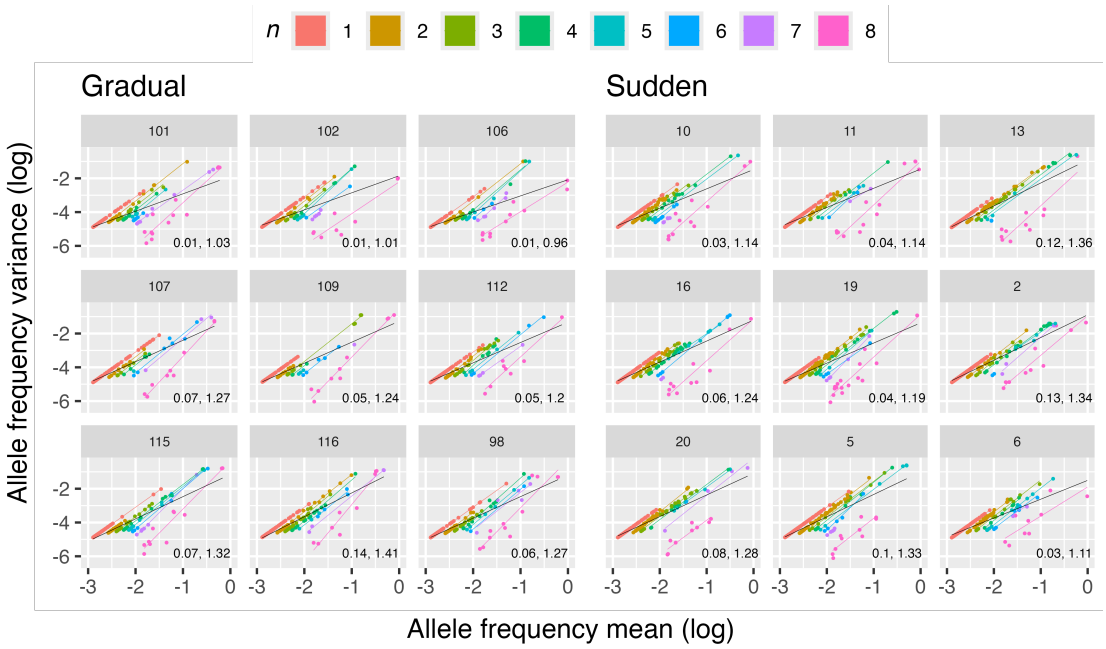

Figure 4. Taylor's law plots. Each panel corresponds to one population, where each point corresponds to the mean and standard deviation of an allele frequency across all passages. Alleles are colored by their number of non-zero occurrences (n), and for each n, a colored line represents the fit to Taylor's law. A black line represents the fit to Taylor's law for all alleles together. 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 able to capture how these values are distributed and measure the strength of aggregation, with an important caveat. Due to the bounded measures of allele frequency in the interval (0, 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 drive β_n down, making it to resemble a Poisson distribution. Late fixation will conversely drive β_n up. Thus, this parameter must be analyzed in respect to its deviation from 2, which is the expected for random fluctuations without fixation. In other words, mean frequencies and their fluctuations will occupy different regions in respect to $\beta_n = 2$ (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 deviations of 0.2, 0.05 and 0.01, respectively (figure S1b), where the higher the standard deviation, the 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 (figure S1c). As expected, when fitting to Taylor's law without adjusting for n, random matrices behave close to a Poisson distribution ($\beta \sim 1$), where parameter V increases with larger fluctuations.

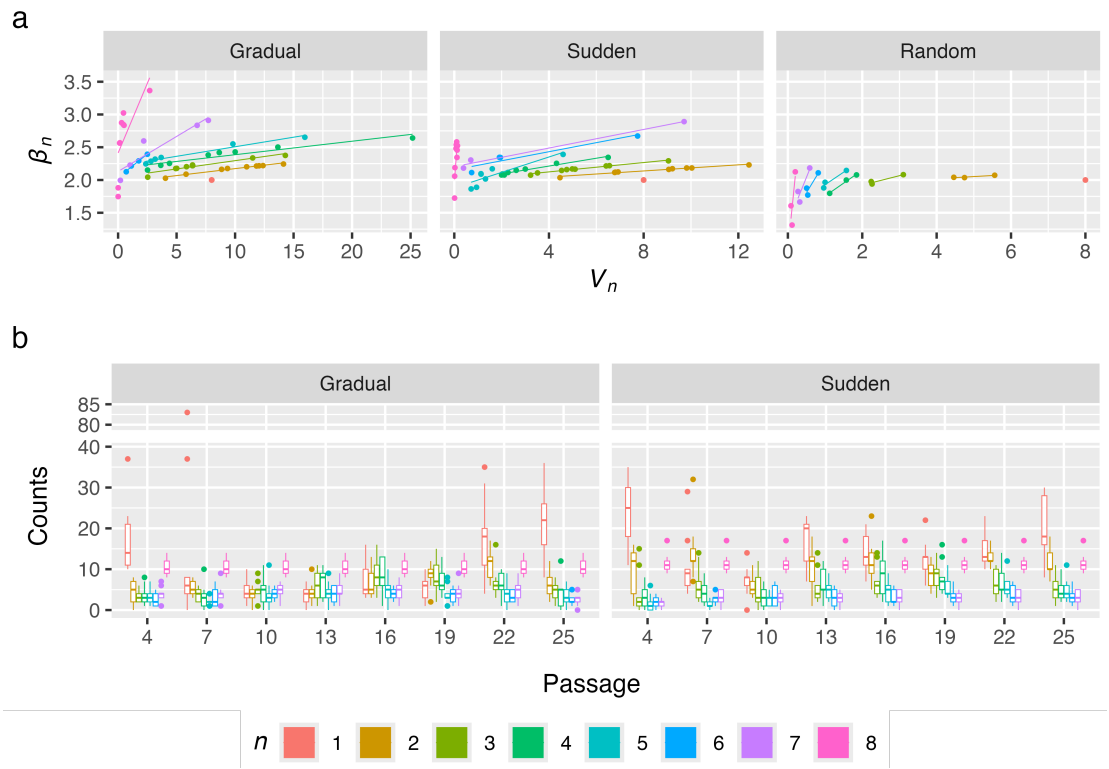

Figure 5. Distribution of alleles frequencies across passages by their n. (a) Taylor's parameters estimated from alleles frequencies for each n for the gradual and sudden treatments and random walk matrices. (b) For each passage, boxplots of the distribution of alleles at each category of n.

To better understand whether frequencies are distributed across the sequenced passages, the number of alleles separated by n at each passage and plotted (figure 5b). In

352 the sudden treatment, a higher concentration of alleles with an n of one or two at the first
353 passages is detected, in agreement with a strong bottleneck at the initial passages as the
354 virus populations are introduced to a completely new host CHO. The number of alleles
355 with an n of one or two starts to rise again in the sudden treatment at passage 13; and
356 overall, a higher number of alleles with an n of one or two across all passages is seen in
357 comparison with the gradual treatment (figure 5b). This is in accordance with the
358 expected effects of genetic drift, and also with the continuous emergence and loss of new
359 mutations in these shifting populations that was discussed above. In the gradual
360 treatment, alleles with an n of one or two are also concentrated at the first and last
361 passages. However, in comparison to the sudden treatment, they are less concentrated at
362 the first passages and their number only starts to rise again at passage 22 (figure 5b).
363
364 **3.5. Effects of treatment and selection coefficient on alleles' fluctuation**
365 Next, we tested the effect of treatment, population and n on the distribution of alleles
366 accounting for all interactions between these variables. The number of non-zero
367 occurrences n had the highest influence on Taylor's parameters V_n and β_n (MANOVA; P
368 < 0.0001; $\eta_P^2 = 0.83$). Both the main effect treatment and the interaction between
369 treatment and n had a significant influence on Taylor's parameter (MANOVA;
370 respectively, $P = 0.0019$ and $\eta_P^2 = 0.14$, and $P = 0.0080$ and $\eta_P^2 = 0.11$), which indicates
371 that host-change rate has a significant and large impact on the distribution of allele
372 frequencies. The main effect population and the interaction between n and population
373 were not significant, despite large effect sizes (MANOVA; respectively, $P = 0.0531$ and
374 $\eta_P^2 = 0.23$, and $P = 0.0837$ and $\eta_P^2 = 0.22$).
375 Alleles were divided into three categories, neutral, deleterious and beneficial, to
376 determine whether strength of selection has an impact on Taylor's parameters. By
377 visualizing the distribution of s (figure 6a) it becomes clearer that s follows a mostly
378 bimodal distribution with a peak for negatively selected alleles and another for neutrally
379 selected ones around zero, with only a fraction being positively selected. Interestingly,
380 negatively selected alleles on the gradual treatment seem to be subjected to stronger
381 negative selection. Based on these distributions, we set a threshold of $s < -0.16$ for alleles
382 to be considered as negatively selected and $s > 0.16$ for them to be considered positively
383 selected, whereas alleles with $s \in (-0.16, 0.16)$ are considered as neutral.
384

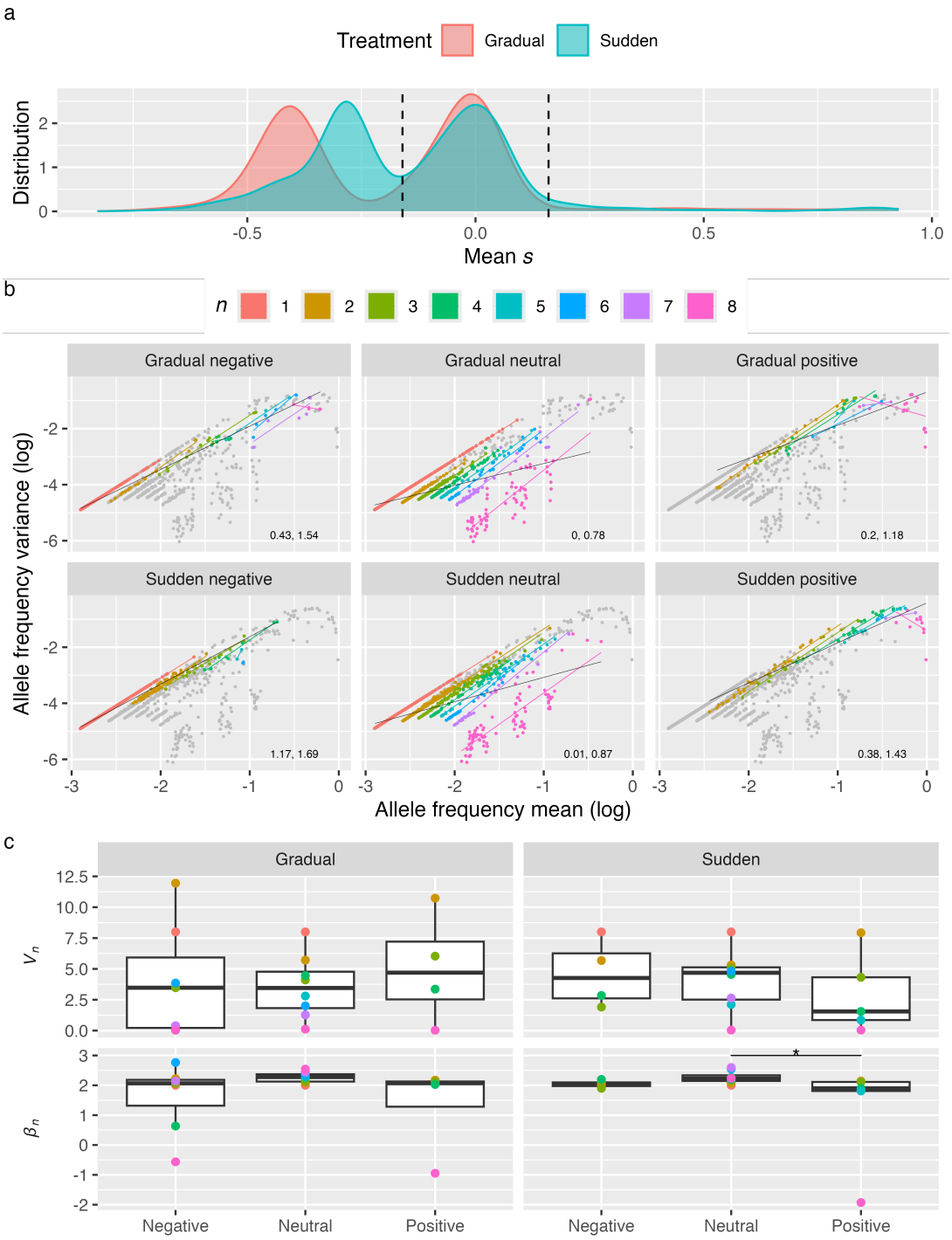

Figure 6. Effects of selection on allele fluctuations. (a) Distribution of *s* per treatment. Dashed lines represent the thresholds for categorizing neutral, negatively and positively selected alleles. Taylor's parameters *V* and *β*, respectively, for the fit without adjusting for *n*, are shown. (b) Taylor's plots of combined allele frequencies across populations by treatment and *s*. All the alleles of each treatment are represented as dots and are colored by their number of non-zero occurrences (*n*) in their selection classification. For each *n* and *s*, a colored line represents the fit to Taylor's law. A black line

represents the fit to Taylor's law for all alleles classified as the corresponding s in the panel. 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: * $P < 0.05$.

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; $P < 0.0001$; $\eta_P^2 = 0.75$), followed by the interaction between n and s ($P = 0.0011$; $\eta_P^2 = 0.31$), and s ($P = 0.0299$; $\eta_P^2 = 0.2$). 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 neutral 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 $n = 8$ that have low mean frequency and standard deviation are neutral. Due to this, if the data is fitted to Taylor's law without adjusting for n, neutral alleles will be the ones closer to a Poisson distribution. Also, neutral alleles exhibit the highest values of β_n (figure 6c). Secondly, positively selected alleles with an n = 8 have negative β_n and their standard deviation decreases as their mean frequency increase. This is due to 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 selected ones. Additionally, positively selected alleles occupy a region of the Taylor's plot close to the upper limits of the system, especially for higher n. Thirdly, negatively selected alleles occupy a region of the

414 Taylor's plot between neutral and positively selected alleles. This difference is clearer
415 for the sudden treatment.
416
417 **3.6. Host replacement regime affects the distribution of mutations along SINV**
418 **genome**
419 Lastly, it is likely that host change dynamics may also have an effect not only on the
420 temporal aggregation dynamics of allele frequencies but also on genomic location of
421 mutations due to the interplay between drift and selection and linkage. Due to genetic
422 hitchhiking [2,27-29], neutral or low fitness mutations may benefit from being strongly
423 linked to high fitness ones. In addition, the gradual host replacement is associated with
424 greater clonal interference [2], in which genomes containing multiple beneficial
425 mutations outcompete those containing only one, promoting the fixation of linked
426 mutations [2,39-41]. Given that the fixation of sweeps of mutations is greater in the
427 gradual treatment [2], we sought to investigate whether mutations in the gradual treatment
428 are less evenly distributed along the genome, meaning if they are aggregating.
429 For this, we estimated the empirical Hurst's exponent H to measure persistent
430 behavior in the placement of mutations at each passage for each population (figure 7).
431 Here, H was estimated from binary sequences with the same length as SINV genome,
432 where ones represent sites with any disagreement to the ancestral sequence. Despite high
433 variability in the estimation, especially for the sudden treatment, evidence of persistent
434 behavior was found in most cases, were the median value at each passage for both
435 treatments were always above 0.5 (figure 7). At the first two initial passages, median H
436 is higher for the sudden treatment, follow by two passages where H is more or less similar
437 between treatments, and finally, H decreases and becomes smaller on the sudden
438 treatment (despite these differences are only significant in passage 22, $P = 0.040$). This
439 suggests that the continuous emergence and loss of mutations on the sudden treatment
440 that was previously observed at the final passages is associated with more random
441 scattering of mutations along the genome, and might be a big factor on why mutations are
442 more randomly positioned on this treatment. However, we found that treatment did not
443 have a significant influence on H (linear model with passage and treatment as orthogonal
444 fixed effects and population as a random effect nested within treatment; sudden
445 coefficient estimate = –0.043, $P = 0.100$).
446

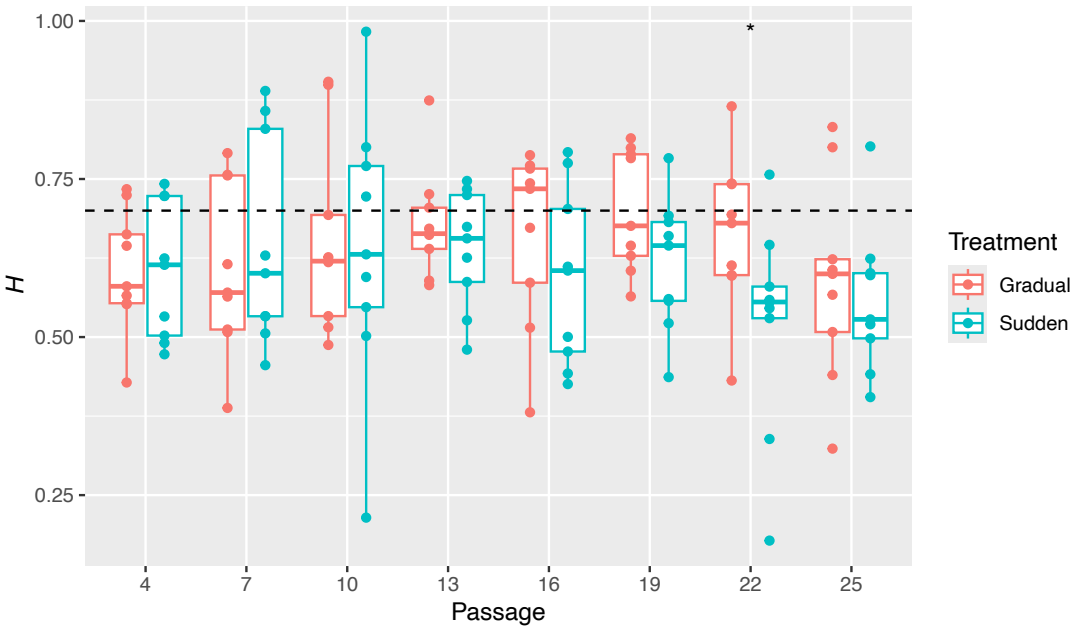

Figure 7. Persistent behavior in the placement of mutations. Hurst exponents (H) estimated from binary sequences (where 1 represent a mutant allele) for each population at each passage. Dashed line represents the 0.7 Hurst phenomenon threshold. Asterisks represent the significance of unpaired Wilcoxon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

4. Conclusions

This study explores the molecular dynamics of virus adaptation to novel hosts. By reanalyzing the genomic data from experimental evolution of SINV in populations subjected to sudden versus gradual host replacements, we investigated how environmental heterogeneity, genetic drift, and selection influenced temporal fluctuations in allele frequency and genome-wide mutation distribution. Using two mathematical tools from the complex systems toolbox, Taylor's power law and the Hurst exponent, we characterized whether the observed fluctuations were consistent with random noise or governed by more complex underlying laws.

Our findings emphasize the critical role of environmental heterogeneity in shaping viral adaptation. Gradual host replacement enables more stable and convergent evolutionary trajectories, enhancing viral fitness across environments. This regime promotes clonal interference, facilitating the fixation of linked beneficial mutations and time-persistent genomic mutation patterns. Conversely, sudden host transitions reduce effective population sizes, amplifying genetic drift, destabilizing allele frequencies, and causing more scattered mutation distributions along the viral genomes. These results

underscore the importance of the rate and nature of environmental change in determining evolutionary outcomes.

Additionally, applying Taylor's power law and the Hurst exponent provided insights into the scaling behavior and temporal aggregation of allele frequency fluctuations. Neutral alleles exhibited greater variability, while positively selected alleles tended to fix earlier. The interaction of genetic drift, selection, and linkage influenced mutation distribution, with gradual host replacement favoring aggregated mutation patterns due to hitchhiking of linked mutations.

This study advances our understanding of virus evolution and provides a novel framework for predicting pathogen adaptation to changing environments, such as shifts in host availability or immune pressures.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Supplementary figure S1 (pdf) and 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.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. J.M.F.S.: conceptualization, formal analyses, investigation, methodology, software, visualization, writing – original draft, writing – review and editing; M.J.O-U.: conceptualization, investigation, methodology, writing – review and editing; V.J.M.: data curation, investigation; P.E.T.: data curation, investigation, supervision, writing – review and editing; S.F.E.: conceptualization, formal analyses, investigation, methodology, writing – original draft, writing – review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

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. The authors would like to 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.

References

1. Morley VJ, Mendiola SY, Turner PE. 2015 Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proc. R. Soc. B* **282**, 20150801. (doi:10.1098/rspb.2015.0801)
2. Morley VJ, Turner PE. 2017 Dynamics of molecular evolution in RNA virus populations depend on sudden versus gradual environmental change. *Evolution* **71**, 872–883. (doi:10.1111/evo.13193)
3. Taylor LR. 1961 Aggregation, variance and the mean. *Nature* **189**, 732–735. (doi:10.1038/189732a0)
4. Eisler Z, Bartos I, Kertész J. 2008 Fluctuation scaling in complex systems: Taylor's law and beyond. *Advan. Phys.* **57**, 89–142. (doi:10.1080/00018730801893043)
5. Fronczak A, Fronczak P. 2010 Origins of Taylor's power law for fluctuation scaling in complex systems. *Phys. Rev. E* **81**, 066112. (doi:10.1103/PhysRevE.81.066112)
6. Kendal WS., Jørgensen B. 2011 Tweedie convergence: A mathematical basis for Taylor's power law, 1/*f* noise, and multifractality. *Phys. Rev. E* **84**, 066120. (doi:10.1103/PhysRevE.84.066120)
7. Lazzardi S, Valle F, Mazzolini A, Scialdone A, Caselle M, Osella M. 2023 Emergent statistical laws in single-cell transcriptomic data. *Phys. Rev. E* **107**, 044403. (doi:10.1103/PhysRevE.107.044403)
8. Martí JM, Martínez-Martínez D, Rubio T, Gracia C, Peña M, Latorre A, Moya A, Garay CP. 2017 Health and disease imprinted in the time variability of the human microbiome. *mSystems* **2**, e00144-16. (doi:10.1128/mSystems.00144-16)
9. Silva JMF, Oteo JA, Garay CP, Elena SF. 2024 System and transcript dynamics of cells infected with severe acute respiratory syndrome virus 2 (SARS-CoV-2). *PLoS Complex Syst.* **1**, e000016. (doi:10.1371/journal.pcsy.0000016)
10. Hurst HE. 1951 Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.* **116**, 770–799. (doi:10.1061/TACEAT.0006518)

11. Mandelbrot BB, Wallis JR. 1969 Robustness of the rescaled range *R*/*S* in the measurement of noncyclic long run statistical dependence. *Water Resour. Res.* **5**, 967–988 (doi:10.1029/WR005i005p00967).
12. Feder J. 1988 *Fractals*. New York: Plenum Press. (doi:10.1007/978-1-4899-2124-6)
13. Blumm N, Ghoshal G, Forró Z, Schich M, Bianconi G, Bouchaud JP, Barabási AL. 2012 Dynamics of ranking processes in complex systems. *Phys. Rev. Lett.* **109**, 128701. (doi:10.1103/PhysRevLett.109.128701)
14. Ghorbani M, Jonckheere EA, Bogdan P. 2018 Gene expression is not random: scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. *Front. Physiol.* **9**, 1446. (doi:10.3389/fphys.2018.01446)
15. Oteo JA, Oteo-García G. 2022 Mutations along human chromosomes: How randomly scattered are they? *Phys. Rev. E* **106**, 064404. (doi:10.1103/PhysRevE.106.064404)
16. Rice CM, Levis R, Strauss JH, Huang HV. 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. *J. Virol.* **61**, 3809–3819. (doi:10.1128/JVI.61.12.3809-3819.1987)
17. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12. (doi:10.14806/ej.17.1.200)
18. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
19. Gaidatzis D, Lerch A, Hahne F, Stadler MB. 2015 QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132. (doi:10.1093/bioinformatics/btu781)
20. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009 VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285. (doi:10.1093/bioinformatics/btp373)
21. Ferrer-Admetlla A, Leuenberger C, Jensen JD, Wegmann D. 2016 An approximate Markov model for the Wright-Fisher diffusion and its application to time serie data. *Genetics* **203**, 831–846. (doi:10.1534/genetics.115.184598)
22. Berner D. 2019 Allele frequency difference AFD–an intuitive alternative to F_{ST} for quantifying genetic population differentiation. *Genes* **10**, 308. (doi:10.3390/genes10040308)

23. Borchers H. 2023 pracma: Practical numerical math functions. Version 2.4.42 [R package]. (doi:10.32614/CRAN.package.pracma)
24. Kääriänen L, Ahola T. 2002 Functions of alphavirus nonstructural proteins in RNA replication. *Prog. Nuclic Acids Res. Mol. Biol*. **71**, 187-222. (doi:10.1016/s0079-6603(02)71044-1)
25. Gerrish PJ, Lenski RE. 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127–144. (doi:10.1023/A:1017067816551)
26. Desai MM, Fisher DS. 2007 Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798. (doi:10.1534/genetics.106.067678)
27. Smith JM, Haigh J. 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35. (doi:10.1017/S0016672308009579)
28. Lang GI, Rice DP, Hickman MJ, Sodergre E, Weinstock GM, Botstein D, Desai MM. 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574. (doi:10.1038/nature12344)
29. Burskirk SE, Peace RE, Lang GI. 2017 Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proc. Natl. Acad. Sci. USA* **114**, 8330–8335. (doi:10.1073/pnas.1702314114)
30. Flynn KM, Cooper TF, Moore FBG, Cooper VS. 2013 The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet*. **9**, e1003426. (doi:10.1371/journal.pgen.1003426)
31. Melbinger A, Vergassola M. 2015 The impact of environmental fluctuations on evolutionary fitness functions. *Sci. Rep*. **5**, 15221. (doi:10.1038/srep15211)
32. Cervera H, Lalić J, Elena SF. 2016 Effect of host species on topography of the fitness landscape for a plant RNA virus. *J. Virol*. **22**, 10160-10169 (doi:10.1128/JVI.01243-16)
33. Gillespie JH. 2000 Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**, 2909-2919. (doi:10.1093/genetics/155.2.909)
34. Brunet E, Derrida B, Mueller AH, Munier S. 2007 Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76**, 041104. (doi:10.1103/PhysRevE.76.041104)
35. Cohen E, Kessler DA, Levine H. 2005 Recombination dramatically speeds up evolution of finite populations. *Phys. Rev. Lett*. **94**, 098102. (doi:10.1103/PhysRevLett.94.098102)

36. Miralles R, Gerrish PJ, Moya A, Elena SF. 1999 Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745-1747. (doi:10.1126/science.285.5434.1745)
37. Park S, Krug J. 2007 Clonal interference in large populations. *Proc. Natl. Acad. Sci. USA* **46**, 18135–18140. (doi:10.1073/pnas.0705778104)
38. Remold SK. 2012 Understanding specialism when the jack-of-all-trades can be the master of all. *Proc. R. Soc. B* **1749**, 4861-4869. (doi:10.1098/rspb.2012.1990)
39. Gillespie JH. 1984 Molecular evolution over the mutational landscape. *Evolution* **38**, 1116–1129. (doi:10.1111/j.1558-5646.1984.tb00380.x)
40. Schiffels S, Szollosi GJ, Mustonen V, Lässig M. 2011 Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375. (doi:10.1534/genetics.111.132027)
41. Good BH, Desai MM. 2014 Deleterious passengers in adapting populations. *Genetics* **198**, 1183-1208. (doi:10.1534/genetics.114.170233)