Inference with Viral Quasispecies

Methods for Individual Samples Comparative Analysis

Josep Gregori*

2025-01-09

Abstract

The study of viral quasispecies structure and diversity presents unique challenges in comparing samples, particularly when dealing with single experimental samples from different time points or conditions. Traditional statistical methods are often inapplicable in these scenarios, necessitating the use of resampling techniques to estimate diversity and variability. This paper discusses two proposed methods for comparing quasispecies samples: repeated rarefaction with z-test and permutation testing. The authors recommend the permutation test for its potential to reduce bias. The research highlights several key challenges in quasispecies analysis, including the need for high sequencing depth, limited clinical samples, technical inconsistencies leading to coverage disparities, and the sensitivity of diversity indices to sample size differences. To address these issues, the authors suggest using a combination of metrics with varying susceptibilities to large sample sizes, ranging from observed differences and ratios to multitest adjusted p-values. The paper emphasizes the importance of not relying solely on p-values, as the high statistical power resulting from large sample sizes can lead to very low p-values for small, potentially biologically insignificant differences. The authors also stress the need for multiple experimental replicates to account for stochastic variations and procedural inconsistencies, particularly when dealing with complex quasispecies populations.

Introduction

Quasispecies is the form that adopts an infection by RNA virus in a host, a complex and dynamic population of closely related but genetically diverse viral variants. This population structure arises due to the high mutation rates of RNA viruses, which can introduce multiple errors during genome replication. The quasispecies consists of a mutant spectrum or cloud of viral genomes that are continuously generated and subject to selection within the host [1–3].

In the study of quasispecies structure and diversity, we are often confronted with a situation where we need to compare two single quasispecies samples, possibly taken at two different infection or treatment time points. This situation poses statistic challenges, as traditional inference tests are not applicable in the comparison of only two quasispecies samples, each represented by a vector of haplotypes and corresponding frequencies. Any useful test will be based in estimates of expected diversity and corresponding variability for each of the two samples. Analytical approximations are available for some diversity indices [4–6], like Shannon entropy or Simpson index, but for others they are complex and depend on assumptions about the underlying haplotypes abundance distribution. Furthermore, we know that some diversity indices, which may be of interest, are sensitive to sample sizes differences, which means that unbalanced samples will require

*josep.gregori@gmail.com, VHIR, HUVH-UAB, Barcelona Campus, Spain

normalization. We focus on diversity indices which may be calculated directly from the vector of haplotype NGS read counts. In this situation we rely on tools based on resampling.

A resampling method commonly used to estimate variances is the bootstrap [7–9], where at each resampling cycle, we sample with replacement a number of reads equal to the sample size. At each bootstrap cycle diversity indices are computed, and at the end the estimated values and corresponding variances result from the mean and variance of the observed values in the B bootstrap cycles. However we know that the bootstrap suffers of the 0.632 limitation, which implies that in each resampling cycle with replacement at the full sample size, at most 63.2% of the reads in the sample are observed, meaning that 36.8% of the reads are replicates of already observed reads, and likewise that 36.8% of the reads in the quasispecies remain unobserved [10]. This limitation results in biases in the estimation of diversity indices sensitive to low and very low haplotype frequencies. Particularly affected are the fraction of singletons and rare haplotypes, the number of haplotypes and the Shannon entropy, or the Hill number at q=1 [10]. Note that any approach based on parametric multinomial resampling is submitted to the same limitations; resampling from a mutinomial given its parameters and a sample size, is equivalent to resampling with replacement.

In this situation, a valid alternative to the bootstrap is the jackknife [11], where H samples are created, with H the number of haplotypes, each with the exclusion of a single haplotype. For each subsample, calculate the diversity indices of interest, and from the distribution of the leave-one-out estimates, the expected value, the bias and the variance are estimated. However a strong requirement for the the jackknife is that the estimator has to be an smooth function and approximately linear. For most diversity indices this is not the case, and particularly for the estimator of the master haplotype frequency, an estimated maximum, which poses similar problems as the estimator of the median, lacks smoothness.

In this work, we suggest the use of two resampling methods with size normalization, involving resampling without replacement to a common size (rarefaction). The first represents a repeated rarefaction of both samples to a common size, being a fraction of the smaller. The second is a permutations test, where only the bigger sample is size normalized. Results of both tests on a series of in-vitro experiments are compared, and final recommendations are given.

Methods

Rarefaction

When the two samples to be compared are unbalanced, rarefaction is the preferred normalization method [10,12]. This technique primarily affects the larger sample, bringing it down to a reference size for comparison.

In single rarefaction, we estimate diversity indices and their variance by repeatedly rarefying the larger sample to the reference size. This process involves:

1. Conducting B rarefaction cycles
2. Obtaining rarefied haplotype counts in each cycle
3. Computing diversity indices based on these counts
4. Obtain estimates and variances from the set of resampled diversity values.

Each rarefaction cycle is a resampling without replacement to the reference size, which is typically the size of the smaller sample [10].

Double rarefaction

While bootstrap and jackknife methods are not suitable in our case, double rarefaction offers an alternative resampling approach. This method involves rarefying both samples to a common reference size that is lower than the smaller sample's size. We suggest using 80% the size of the smaller sample as the reference.

In the double rarefaction we obtain an empirical distribution of quasispecies indicators for each quasispecies, as a result of repeated rarefaction to a common reference size, and a corresponding distribution of their differences. From these empirical distributions we may compute 99% confidence intervals (CI) for the set of indicators in each quasispecies, and for their differences, estimate the standard error of the differences, normality p-values, and a measure of effect size like Cohen's d.

Our primary objective is to assess significant differences in the structure of two quasispecies, focusing on observed effect sizes of a minimum magnitude. The emphasis is on comparison rather than precise estimation. Double rarefaction to a fraction of the smaller sample size does involve some trade-offs, reduced precision in estimates and potential biases in results. However, these biases are balanced across both samples, maintaining the validity of the comparison. It's important to note that this approach prioritizes comparative analysis over absolute precision, making it suitable for detecting structural differences between quasispecies samples.

z-test

Besides the information provided by the empirical distributions obtained by double rarefaction, the estimated standard errors of the differences may be used in a t-test or a z-test to obtain approximated p-values. Given the coverages recommended for studying quasispecies, which are typically above $1. \times 10^4$ and preferably above $1. \times 10^5$ reads per amplicon and strand, the sample sizes are large enough to tend to approximate the distributions of the differences to a normal distribution due to the central limit theorem. Moreover, the degrees of freedom in a t-test with such large samples are also high enough to closely approximate the normal distribution.

Permutation tests

Alternatively, or in cases where the coverage is limited, a permutation test may be considered. In this case we approximate the null distribution, and assess how extreme are observed quasispecies diversity differences with respect to this estimated null distribution. In this work, a permutation test [8,13,14] is coded encompassing the following steps:

1. Calculate the indicators for the original two quasispecies samples, rarefying the bigger sample to the size of the smaller. Take the mean of R rarefaction cycles for each indicator. Compute the difference between these indicators as the observed test statistic T_{obs}.
2. Inside a resampling loop of B permutations to generate a null distribution of differences in diversity, at each cycle:
    i. Get rarefied haplotype counts of the bigger sample with a single cycle of rarefaction.
    ii. Assign labels to the haplotypes in each quasispecies based on their abundance, in descending order, independently for each quasispecies, and regardless of whether the haplotypes are the same or different between the samples. The focus is on haplotype abundance patterns rather than the specific haplotype identities.
    iii. Pool together the reads of both quasispecies.
    iv. Randomly shuffle the pooled reads and reassign half of them to each quasispecies.

v. Calculate the diversity indices for these permuted groups and their difference, T.

3. Compute 99% CI, mean and standard deviation of the generated null distribution, $\bar{T}_p$ and S_{null}.
4. Compute the p-value of T_{obs} with reference to this null distribution.
5. Compute a z statistic, Z_p, as:

$$Z_{\text{p}} = \frac{T_{\text{obs}} - \bar{T}_p}{SE_T} \approx \frac{T_{\text{obs}} - \bar{T}_p}{S_{\text{null}}}$$

where T_{obs} is the observed difference, $\bar{T}_p$ and S_{null} the mean and standard deviation of the null distribution. Note that the standard error of the difference, SE_T, is approximated by S_{null}, $SE_T \approx S_{\text{null}}$

6. Compute an approximated Cohen's d effect size as:

$$d_{\text{p}} = \frac{T_{\text{obs}}}{S_{\text{null}}\sqrt{N_p}}$$

where N_p is the number of permutations.

The Z statistic indicates how many standard deviations the mean of observed the differences is from the mean of the permutation distribution. The approximate Cohen's d provides a measure of the effect magnitude. Given the inherent limitations in the calculation of permutation p-values, which are bounded by the number of permutations and may result in values of 0, Z offers a continuous and more precise measure of how extreme is the observed difference with respect to the null distribution, allowing for easier interpretation. P-values of 0 can occur in permutation tests when the observed statistic is more extreme than all permuted statistics, a situation where Z remains informative. When 0, these p-values are reported as $< 1/N_p$, with N_p the number of permutations. Besides, the position of the observed difference with respect to the 99% CI of the null distribution provides an alternative measure to the p-value. The CI itself provides a measure of symmetry of the null distribution.

Note that in the double rarefaction method both samples were rarefied to a given fraction of the size of the smaller sample (80%), while in the permutations method only the bigger sample is rarefied to the size of the smaller sample within each permutation cycle.

Very large sample sizes and p-values

With sample sizes (N) so high, even a tiny difference in diversity measures between two quasispecies can appear statistically significant due to the extremely small standard errors. However, this statistical significance may not translate to biological or practical significance. In these cases, 99% confidence intervals are more appropriate. They assess the precision of the diversity estimates and the range of plausible values for the true difference between quasispecies.

The main focus should be on the magnitude of the difference (effect size) rather than just statistical significance. Effect size measures, like Cohen's d, can provide more meaningful information about the practical importance of observed differences, and are independent of sample sizes [15–18].

Effect size

Cohen's d is a standardized measure of effect size, independent of the sample size, that expresses the observed difference between two populations in units of the pooled standard deviation [15–18].

$$d = \frac{T_{\text{obs}}}{SE_T \sqrt{N}} \tag{1}$$

where, T_{obs} is the observed diference at a normalized size, SE_T the standard error of $\bar{T}$ obtained as the standard deviation of the distribution of differences, and N the number of differences generated by either method.

By contrast, the z-statistic is highly affected by the sample size, providing information about the statistical power of the test.

$$z = \frac{T_{\text{obs}} - \bar{T}}{SE_T} \tag{2}$$

These statistics are related as:

$$d = \frac{z}{\sqrt{N}} \tag{3}$$

When the two populations being compared show diversity values that are normally distributed, or when the sample size is large enough that the difference is approximately normally distributed, Cohen's d may be interpreted in probabilistic terms. Table 1 provides a guide for interpreting d values [15,18,19]:

Table 1: Cohen's d values interpretation. Effect magnitude and corresponding probabilistic interpretations, non-overlap (NOV) and probability of superiority (PS).

d	Magnitude	NOV (1)	PS (2)
0.0	Null	0	50.0
0.2	Small	8.0	55.6
0.5	Medium	19.7	63.8
0.8	Large	31.1	71.4
1.2	Very large	45.1	80.2

(1) NOV: Non-overlap between the two distributions.

$$NOV = \left[1 - 2\Phi\left(\frac{-\text{abs}(d)}{2}\right)\right] \times 100 \tag{4}$$

(2) PS: Probability of superiority. Probability that a randomly sampled individual from one group will have a higher score than a randomly sampled individual from the other group. It is a metric similar to the AUC.

$$PS = \Phi\left(\text{abs}(d)/\sqrt{2}\right) \times 100 \tag{5}$$

Where Φ represents the cumulative distribution function of the standard normal distribution. Note that in taking the absolute value of d we ensure a minimum PS of 50%. The sign of d will dictate which is the condition of superiority.

Quasispecies maturity indicators

A quasispecies matures as it explores the genetic space, primarily due to the low replication fidelity of RNA-dependent RNA polymerases. This process involves the incorporation of new, better-fitted variants into the quasispecies, while less-fitted variants are purged, both driven by fitness gradients, even in the absence of external selective pressures [20,21]. As the quasispecies 'matures', it grows in genetic complexity, following a natural tendency to evolve towards flat-like fitness landscapes with higher global quasispecies fitness [22].

In a recent work we reported two levels of in-host quasispecies evolution. A long-run deterministic trend with positive genetic enrichment, and progressive evenness increase, and limited short-term stochastic fluctuations in diveristy and structure. The concept of quasispecies maturity corresponds to this long-run deterministic trend, which involves a continuous decline in the frequency of dominant haplotypes [23], and a progressive increase in top haplotypes evenness, followed by an increase in global quasispecies evenness.

Table 2 shows the selected quasispecies structure indicators, expressing maturity, and their expected level in the two limiting cases, regular vs flat-like quasispecies [22]. These are the quasispecies features that will be compared in this paper.

Table 2: Selected maturity indicators.

Indicator	Description	Regular Qs	Flat-like Qs
Top25	Fraction of reads for top 25 hpl.	High	Low
Master	Dominant haplotype frequency	High	Low
Rare1	Fraction of reads for hpl $\leq 1\%$	Low	High
Rare2	Fraction of reads for hpl $\leq 0.1\%$	Low	High
RLE_1	Relative logarithmic evennes at $q = 1$	Low	High
RLE_2	Relative logarithmic evennes at $q = 2$	Low	High
RLE_∞	Relative logarithmic evennes at $q = \infty$	Low	High
R5	Evenness among top 5 haplotypes	Low	High
R10	Evenness among top 10 haplotypes	Low	High

Increased maturity involves a shrinkage in Top25 and Master besides increases in tho other indicators. Note that all these indicators show values in the range 0-1.

Relative logarithmic evenness

$$RLEq = RLE(p, q) = \log_{10}(D(p, q))/\log_{10}(D(p, 0)) \tag{6}$$

and $D(q, p)$ the Hill number of order q.

$$D(p, q) = \left(\sum_{i=1}^{H} p_i^q\right)^{1/(1-q)} \tag{7}$$

H number of haplotypes, $p = (p_1, p_2, \ldots, p_H)$ the set of haplotype frequencies.

Fraction of top haplotypes

On the set of haplotype frequencies sorted in decreasing order, $p_1 \geq p_2 \geq \cdots \geq p_H$, the following indicators are computed:

$$\mathrm{Top25}(p) = \left(\sum_{i=1}^{m} p_i \;\middle|\; m = \min(25, H) \right) \tag{8}$$

$$\mathrm{Top25R}(p) = \frac{\mathrm{Top25}(p)}{p_1} \tag{9}$$

Evenness on the set of top haplotypes

An indicator of evenness on the set of k top haplotypes is computed as:

$$R_k(p) = \left(\frac{l \cdot p_l}{\sum_{i=1}^{l} p_i} \;\middle|\; l = \min(k, H) \right) \tag{10}$$

Two limiting quasispecies states

Two extreme quasispecies states are considered in this study as limiting models of quasispecies structure: Quasispecies A, with a highly dominant haplotype approaching a frequency of 100% and a single variant genome at an infinitesimal frequency, and quasispecies Z, representing a fully flat quasispecies with an infinite number of haplotypes, all at equally infinitesimal frequencies [22–24]. These two quasispecies may be represented by the maturity indicator values shown in Table 3.

Table 3: Quasispecies maturity indicator values characterizing the two limiting quasispecies models A and Z.

Model	Top25	Master	Rare2	Rare1	RLE1	RLE2	RLEinf	R5	R10
A	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

A global quasispecies maturity score for a given quasispecies, Qs, may be computed as the distance from this quasispecies to Model A, d_A, in the multidimensional space represented by this set of indicators.

$$d_A(Qs) = \sqrt{\sum_i \left[(f_i(Qs) - f_i(A))^2 \right]}$$

where f_i is the set of maturity indicators. This score may be normalized to the range 0-1 as $d_N(Qs) = d_A(Qs)/d_A(Z)$. Whereas each indicator refers to a given aspect of quasispecies structure, d_N represents the combined effect of all changes in pushing the quasispecies towards its origin, A, or its fate, Z [23].

Samples description

The data used in this study is taken from [25]. The referenced study utilized serial virus passages in human hepatoma Huh-7.5 cells to cultivate HCV under controlled conditions. This experimental setup deliberately eliminated external selective pressures, enabling uninterrupted and extended viral replication. Despite the absence of external selective forces, viral evolution persisted due to two key factors: The emergence of new mutations resulting from the error-prone replication process of the NS5B polymerase, and the ongoing selection of viral variants driven by internal fitness differences within the quasispecies population [20,21]. A NS5B amplicon spanning positions 8552:8869 was sequenced for the quasispecies at passages 0 (p0.p0), 3 (p0.p3), 100 (p100.p0), 110 (p100.p10), and 200 (p200.p0).

Programs and libraries

All analyses were conducted using R [26], with the help of packages biostrings [27], knitr [28], tidyverse [29], ggplot2 [30]. R code of the main functions and the session info are collected in the appendix.

Results

Samples and NGS coverage

Table 4: Reads per amplicon sample and strand

Basal	Pass	FW	RV
p0	p0	124253	120939
p0	p3	186398	182387
p100	p0	179401	176168
p100	p10	136432	133209
p200	p0	123807	123946

In what follows only the fasta files for the forward strands of the amplicon are considered.

Quasispecies maturity indicators, raw values

Values of these indicators are first computed on the quasispecies as they are observed, without size normalization.

Table 5: Quasispecies maturity indicators. Raw values.

ID	nHpl	Master	Top25	Rare1	Rare2	dN
p0.p0	5858	0.72884	0.80131	0.24107	0.19969	0.18117
p0.p3	8325	0.73095	0.78525	0.26905	0.18898	0.18507
p100.p0	11948	0.49214	0.74294	0.33202	0.23125	0.27241
p100.p10	9744	0.39303	0.75431	0.28186	0.24154	0.29475
p200.p0	11490	0.43506	0.70577	0.37739	0.25735	0.31201

ID	R5	R10	RLE1	RLE2	RLEinf	dN
p0.p0	0.02398	0.02458	0.27677	0.07278	0.03646	0.18117
p0.p3	0.02070	0.02845	0.27276	0.06939	0.03472	0.18507
p100.p0	0.06210	0.06696	0.38164	0.14308	0.07552	0.27241
p100.p10	0.10277	0.03230	0.40176	0.17882	0.10168	0.29475
p200.p0	0.16119	0.08438	0.43750	0.17121	0.08902	0.31201

Maturity indicator values on rarefied quasispecies

Table 6: Quasispecies maturity indicators. Rarefied values to 100,000 reads.

ID	nHpl	Master	Top25	Rare1	Rare2	dN
p0.p0	4955	0.72881	0.80134	0.24112	0.19861	0.18187
p0.p3	5044	0.73091	0.78557	0.26909	0.18993	0.18752
p100.p0	7695	0.49214	0.74292	0.33210	0.23082	0.27520
p100.p10	7846	0.39302	0.75443	0.28181	0.24113	0.29621
p200.p0	9785	0.43510	0.70582	0.37738	0.25740	0.31312

ID	R5	R10	RLE1	RLE2	RLEinf	dN
p0.p0	0.02400	0.02455	0.28114	0.07421	0.03718	0.18187
p0.p3	0.02095	0.02854	0.28565	0.07348	0.03677	0.18752
p100.p0	0.06209	0.06577	0.39629	0.15012	0.07923	0.27520
p100.p10	0.10259	0.03237	0.40919	0.18314	0.10415	0.29621
p200.p0	0.16129	0.08435	0.44317	0.17418	0.09057	0.31312

Observed effects

Observed differences between consecutive time points, in rarefied maturity indicators. These values provide a reference scale for the magnitude of changes experienced by the quasispecies along the passages.

Table 7: Observed differences in quasispecies structure indicators. Quasispecies rarefied to 100,000 reads.

	nHpl	Master	Top25	Rare1	Rare2	dN
p0.p3 - p0.p0	89	0.0021	-0.0158	0.0280	-0.0087	0.0056
p100.p0 - p0.p0	2740	-0.2367	-0.0584	0.0910	0.0322	0.0933
p100.p10 - p100.p0	151	-0.0991	0.0115	-0.0503	0.0103	0.0210
p200.p0 - p100.p0	2090	-0.0570	-0.0371	0.0453	0.0266	0.0379

	R5	R10	RLE1	RLE2	RLEinf	dN
p0.p3 - p0.p0	-0.0030	0.0040	0.0045	-0.0007	-0.0004	0.0056
p100.p0 - p0.p0	0.0381	0.0412	0.1151	0.0759	0.0421	0.0933
p100.p10 - p100.p0	0.0405	-0.0334	0.0129	0.0330	0.0249	0.0210
p200.p0 - p100.p0	0.0992	0.0186	0.0469	0.0241	0.0113	0.0379

Table 8: Observed ratios in quasispecies structure indicators. Quasispecies rarefied to 100,000 reads.

	nHpl	Master	Top25	Rare1	Rare2	dN
p0.p3 / p0.p0	1.018	1.003	0.980	1.116	0.956	1.031
p100.p0 / p0.p0	1.553	0.675	0.927	1.377	1.162	1.513
p100.p10 / p100.p0	1.020	0.799	1.015	0.849	1.045	1.076
p200.p0 / p100.p0	1.272	0.884	0.950	1.136	1.115	1.138

	R5	R10	RLE1	RLE2	RLEinf	dN
p0.p3 / p0.p0	0.873	1.163	1.016	0.990	0.989	1.031
p100.p0 / p0.p0	2.588	2.679	1.410	2.023	2.131	1.513
p100.p10 / p100.p0	1.652	0.492	1.033	1.220	1.315	1.076
p200.p0 / p100.p0	2.598	1.282	1.118	1.160	1.143	1.138

Note that differences observed between p0 and p3, or between p100.p0 and p100.p10 are caused by stochastic variations in the short term, and that the differences between p100.p0 and p0.p0,

or between p200.p0 and p100.p0 are caused by increased quasispecies maturity, the evolution expected in the long run, with decreased Master and Top25, and increased rare fractions and evenness.

Despite very large NGS sample sizes, only a single experimental sample from each quasispecies is analyzed here. It's important to note that these experimental samples are susceptible to stochastic fluctuations, methodological discrepancies, and procedural inconsistencies When comparing single experimental samples, the inherent experimental variability may lead to unexpected or misleading results, particularly when the two quasispecies are very similar in structure.

Repeated rarefaction test

Test p0.p0 vs p0.p3

Table 9: Quasispecies coverage and rarefaction reference size

ID	Reads
p0.p3	186398
p0.p0	124253
RefSz	99402

Table 10: Observed values by repeated rarefaction.

	p0.p3	p0.p0	Diff	RelDiff	Ratio
Top25	0.785	0.801	-0.016	-0.020	0.980
Master	0.731	0.729	0.002	0.003	1.003
Rare1	0.269	0.241	0.028	0.116	1.116
Rare2	0.190	0.199	-0.009	-0.044	0.956
R5	0.021	0.024	-0.003	-0.123	0.877
R10	0.028	0.025	0.004	0.158	1.158
RLE1	0.286	0.281	0.005	0.016	1.016
RLE2	0.074	0.074	-0.001	-0.010	0.990
RLEinf	0.037	0.037	0.000	-0.011	0.989
dN	0.188	0.182	0.006	0.031	1.031

Table 11: Observed difference, and related statistics from the double rarefaction test

	Diff	RelDiff	Ratio	CI99.gap	z	d	NOV	PS
Top25	-0.01587	-0.01980	0.98020	0.01213	-15.66	-0.495	19.6	63.7
Master	0.00208	0.00286	1.00286	-0.00200	1.82	0.058	2.3	51.6
Rare1	0.02799	0.11612	1.11612	0.02383	25.01	0.791	30.7	71.2
Rare2	-0.00879	-0.04415	0.95585	0.00173	-4.33	-0.137	5.5	53.9
R5	-0.00295	-0.12331	0.87669	0.00001	-3.48	-0.110	4.4	53.1
R10	0.00389	0.15842	1.15842	0.00087	4.36	0.138	5.5	53.9
RLE1	0.00457	0.01624	1.01624	0.00104	4.69	0.148	5.9	54.2
RLE2	-0.00073	-0.00979	0.99021	-0.00052	-2.08	-0.066	2.6	51.9
RLEinf	-0.00041	-0.01098	0.98902	-0.00022	-2.33	-0.074	2.9	52.1
dN	0.00566	0.03110	1.03110	0.00285	7.52	0.238	9.5	56.7

All differences are small, smaller than 0.028 in absolute value, with ratios in the range 0.88 - 1.16. Effect sizes are also small except for Rare1, corresponding to the indicator with higher positive gap in 99% CIs. Rare1 represents the proliferation of new mutants as the quasispecies starts to replicate.

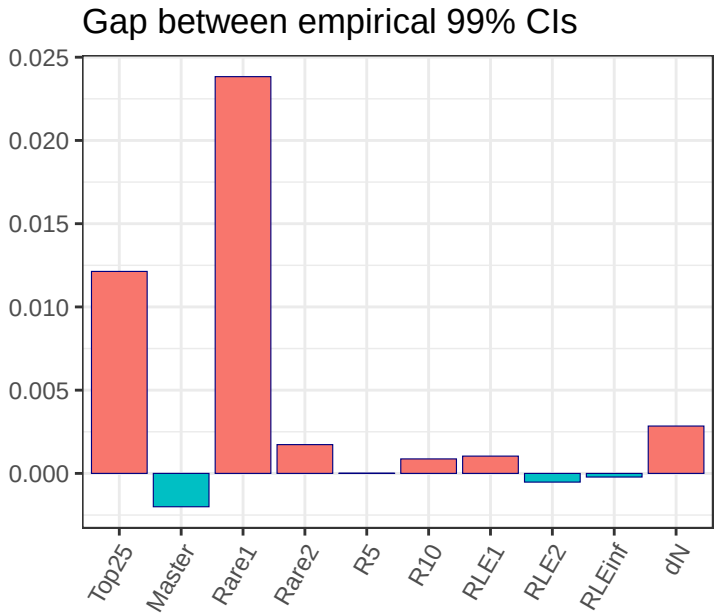

Figure 1: Gap between 99% CIs. Negative gaps show overlapped 99% CIs

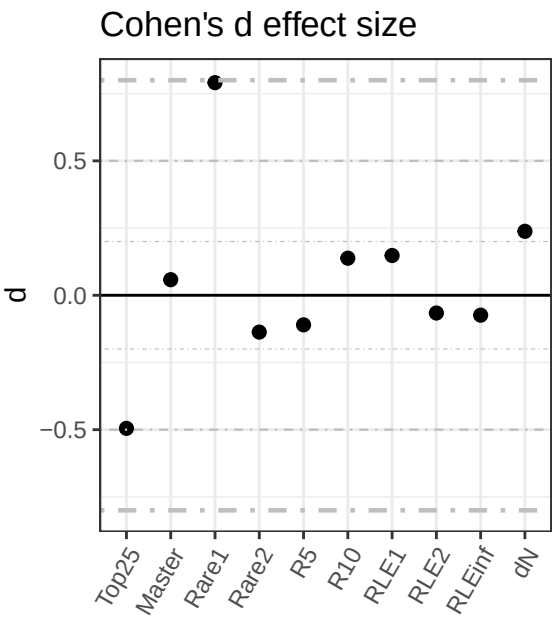

Figure 2: Effect size in double rarefaction test. Dash-dot line borders at 0.2, 0.5 and 0.8

Test p0.p0 vs p100.p0

Table 12: Quasispecies coverage and rarefaction reference size

ID	Reads
p100.p0	179401
p0.p0	124253
RefSz	99402

Table 13: Observed values by repeated rarefaction.

	p100.p0	p0.p0	Diff	RelDiff	Ratio
Top25	0.743	0.801	-0.058	-0.073	0.927
Master	0.492	0.729	-0.237	-0.325	0.675
Rare1	0.332	0.241	0.091	0.377	1.377
Rare2	0.231	0.199	0.032	0.161	1.161
R5	0.062	0.024	0.038	1.589	2.589
R10	0.066	0.025	0.041	1.677	2.677
RLE1	0.396	0.281	0.115	0.410	1.410
RLE2	0.150	0.074	0.076	1.023	2.023
RLEinf	0.079	0.037	0.042	1.132	2.132
dN	0.275	0.182	0.093	0.513	1.513

Table 14: Observed difference and related statistics from the double rarefaction test

	Diff	RelDiff	Ratio	CI99.gap	z	d	NOV	PS
Top25	-0.05839	-0.07287	0.92713	0.05480	-52.52	-1.661	59.4	88.0
Master	-0.23674	-0.32481	0.67519	0.23233	-188.37	-5.957	99.7	100.0
Rare1	0.09097	0.37736	1.37736	0.08682	76.37	2.415	77.3	95.6
Rare2	0.03203	0.16096	1.16096	0.02464	14.58	0.461	18.2	62.8
R5	0.03807	1.58896	2.58896	0.03315	25.05	0.792	30.8	71.2
R10	0.04121	1.67669	2.67669	0.03572	23.43	0.741	28.9	70.0
RLE1	0.11520	0.40963	1.40963	0.11185	117.65	3.720	93.7	99.6
RLE2	0.07595	1.02295	2.02295	0.07441	167.23	5.288	99.2	100.0
RLEinf	0.04209	1.13152	2.13152	0.04120	164.36	5.197	99.1	100.0
dN	0.09332	0.51297	1.51297	0.09059	118.53	3.748	93.9	99.6

This case represents 300 days of infection time, without external perturbations, all indicators show positive gaps between 99% empirical CIs. Observed differences are all in line with expected changes with quasispecies maturation [22,24], decreased frequency for master haplotype, decreased reads fraction for top haplotypes, and increased evenness, both among top haplotypes, and in the whole quasispecies. The magnitudes of the size effects are remarkable, showing this trend. Mainly for master, Rare1, and RLE values, and the global effect expressed by dN, indicating a flatter more even distribution of haplotypes, with decreased frequency in dominant haplotypes. Rare2 and R10 show a lower effect magnitude. Effect magnitudes align with inter-CI gap values.

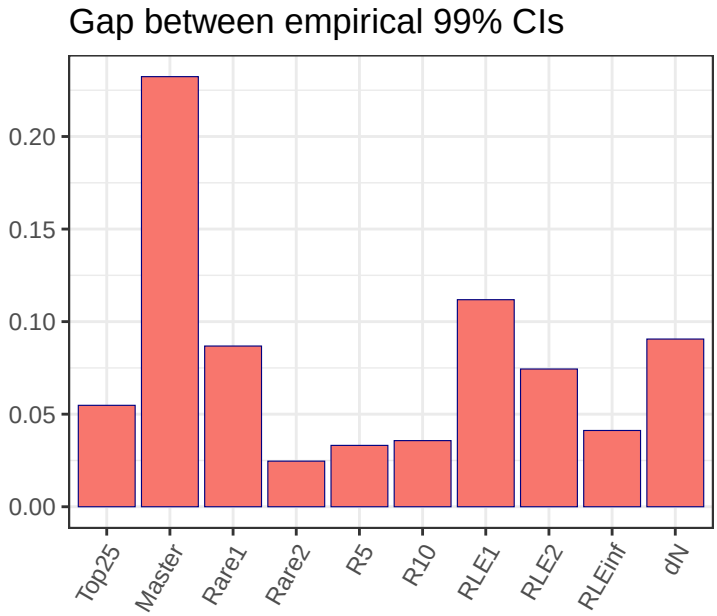

Figure 3: Gap between 99% CIs. Negative gaps show overlapped CIs

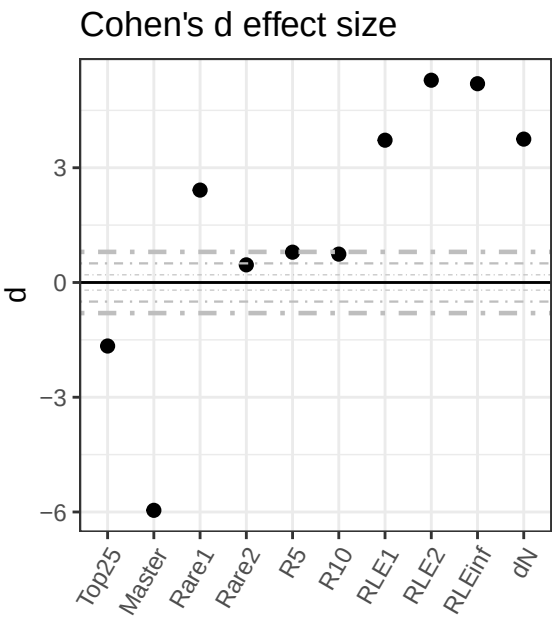

Figure 4: Effect size in double rarefaction test. Dash-dot line borders at 0.2, 0.5 and 0.8

Test p100.p0 vs p100.p10

Table 15: Quasispecies coverage and rarefaction reference size

ID	Reads
p100.p10	136432
p100.p0	179401
RefSz	109146

Table 16: Observed values by repeated rarefaction.

	p100.p10	p100.p0	Diff	RelDiff	Ratio
Top25	0.754	0.743	0.011	0.015	1.015
Master	0.393	0.492	-0.099	-0.201	0.799
Rare1	0.282	0.332	-0.050	-0.151	0.849
Rare2	0.242	0.231	0.010	0.045	1.045
R5	0.103	0.062	0.041	0.655	1.655
R10	0.032	0.066	-0.034	-0.510	0.490
RLE1	0.407	0.394	0.013	0.033	1.033
RLE2	0.182	0.149	0.033	0.221	1.221
RLEinf	0.103	0.079	0.025	0.315	1.315
dN	0.296	0.275	0.021	0.077	1.077

Table 17: Observed difference and related statistics from the double rarefaction test

	Diff	RelDiff	Ratio	CI99.gap	z	d	NOV	PS
Top25	0.01135	0.01527	1.01527	0.00786	11.42	0.361	14.3	60.1
Master	-0.09912	-0.20139	0.79861	0.09486	-83.45	-2.639	81.3	96.9
Rare1	-0.05012	-0.15097	0.84903	0.04603	-45.15	-1.428	52.5	84.4
Rare2	0.01048	0.04536	1.04536	0.00355	5.44	0.172	6.9	54.8
R5	0.04065	0.65455	1.65455	0.03427	23.81	0.753	29.3	70.3
R10	-0.03362	-0.51003	0.48997	0.02715	-19.32	-0.611	24.0	66.7
RLE1	0.01308	0.03320	1.03320	0.01027	15.59	0.493	19.5	63.6
RLE2	0.03289	0.22067	1.22067	0.03133	72.72	2.299	75.0	94.8
RLEinf	0.02478	0.31500	1.31500	0.02377	87.06	2.753	83.1	97.4
dN	0.02108	0.07671	1.07671	0.01875	30.45	0.963	37.0	75.2

Extra 30 days of infection, for a quasispecies, p100, showing incipient maturity. All indicators show positive gaps between empirical 99% CI. Effect sizes are relevant for Master, Rare1, RLE2 and RLEinf, and in lower measure dN. Observed differences and effect sizes for the other indicators are small. The direction of changes is opposite to that expected in the long run, for Top25, Rare1, and R10, which may be explained by quasispecies short term dynamics and stochastic variations.

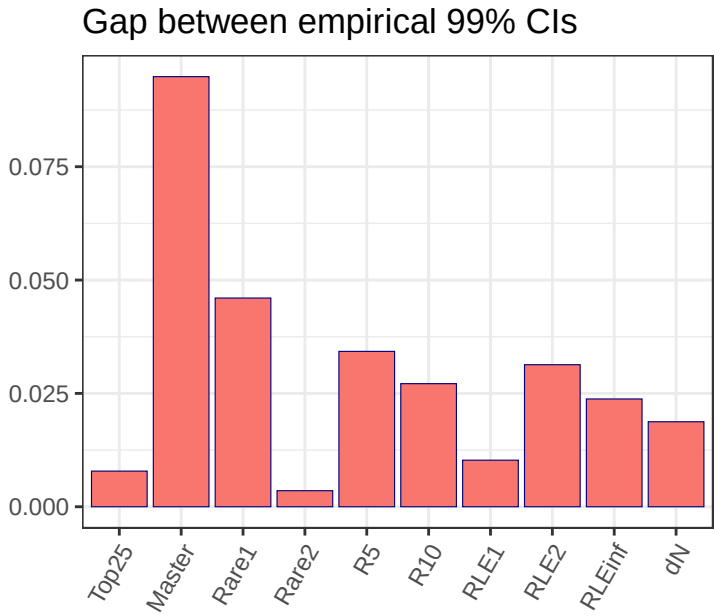

Figure 5: Gap between 99% CIs. Negative gaps show overlapped CIs

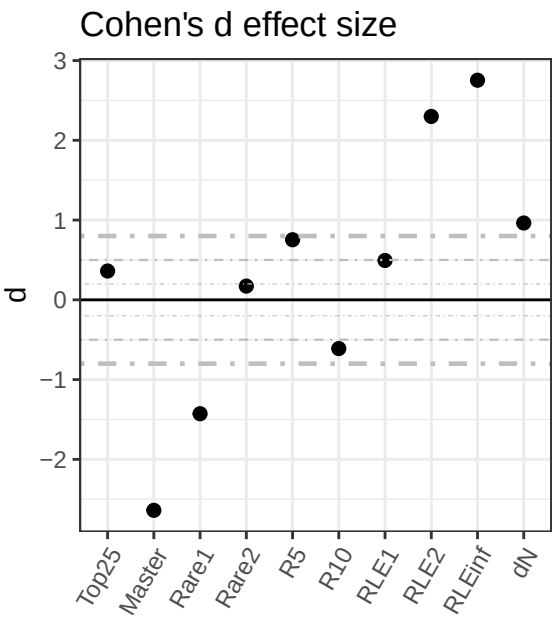

Figure 6: Effect size in double rarefaction test. Dash-dot line borders at 0.2, 0.5 and 0.8

Test p100.p0 vs p200.p0

Table 18: Quasispecies coverage and rarefaction reference size

ID	Reads
p200.p0	123807
p100.p0	179401
RefSz	99046

Table 19: Observed values by repeated rarefaction.

	p200.p0	p100.p0	Diff	RelDiff	Ratio
Top25	0.706	0.743	-0.037	-0.050	0.950
Master	0.435	0.492	-0.057	-0.116	0.884
Rare1	0.377	0.332	0.045	0.136	1.137
Rare2	0.258	0.231	0.027	0.116	1.116
R5	0.161	0.062	0.099	1.593	2.593
R10	0.084	0.066	0.019	0.283	1.283
RLE1	0.443	0.396	0.047	0.118	1.118
RLE2	0.174	0.150	0.024	0.160	1.160
RLEinf	0.091	0.079	0.011	0.143	1.143
dN	0.313	0.275	0.038	0.138	1.138

Table 20: Observed difference and related statistics from the double rarefaction test

	Diff	RelDiff	Ratio	CI99.gap	z	d	NOV	PS
Top25	-0.03713	-0.04997	0.95003	0.03301	-32.38	-1.024	39.1	76.5
Master	-0.05708	-0.11598	0.88402	0.05241	-43.80	-1.385	51.1	83.6
Rare1	0.04533	0.13651	1.13651	0.04110	36.40	1.151	43.5	79.2
Rare2	0.02686	0.11623	1.11623	0.01880	12.29	0.389	15.4	60.8
R5	0.09900	1.59258	2.59258	0.09182	47.17	1.492	54.4	85.4
R10	0.01859	0.28267	1.28267	0.00997	8.04	0.254	10.1	57.1
RLE1	0.04690	0.11829	1.11829	0.04357	49.01	1.550	56.2	86.3
RLE2	0.02410	0.16040	1.16040	0.02230	46.32	1.465	53.6	85.0
RLEinf	0.01135	0.14314	1.14314	0.01035	38.43	1.215	45.7	80.5
dN	0.03796	0.13790	1.13790	0.03532	48.46	1.533	55.6	86.1

All indicators show positive gaps between empirical 99% CI. Except for Rare2 and R10, all other indicators show an absolute Cohen's d above 1. The hundred passages after p100, resulted in a further reduced Master frequency and Top25 fraction, and increased values for all evenness indicators, showing the long term quasispecies dynamics observed in passing from p0 to p100. Top haplotype evenness indicators R5 and R10 show a ratio of 2.6 and 1.3 respectively. However the differences in all indicators are mitigated with respect to those observed in the first 100 passages, showing an asymptotic trend.

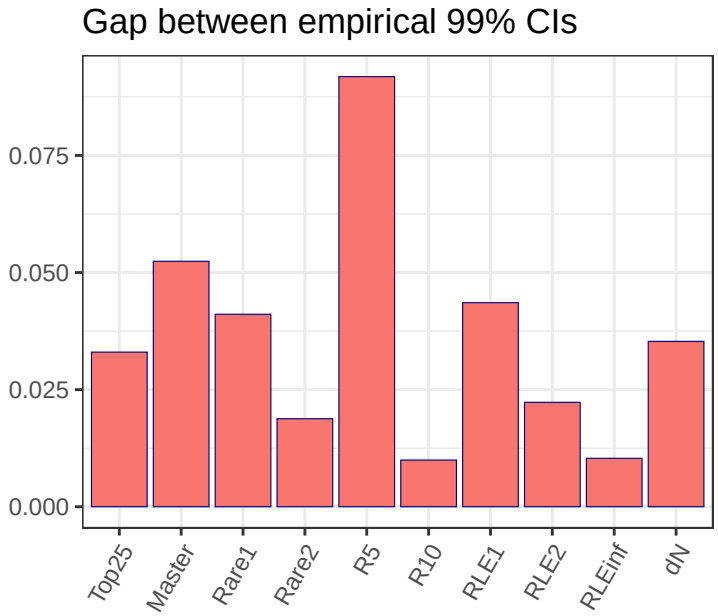

Figure 7: Gap between 99% CIs. Negative gaps show overlapped CIs

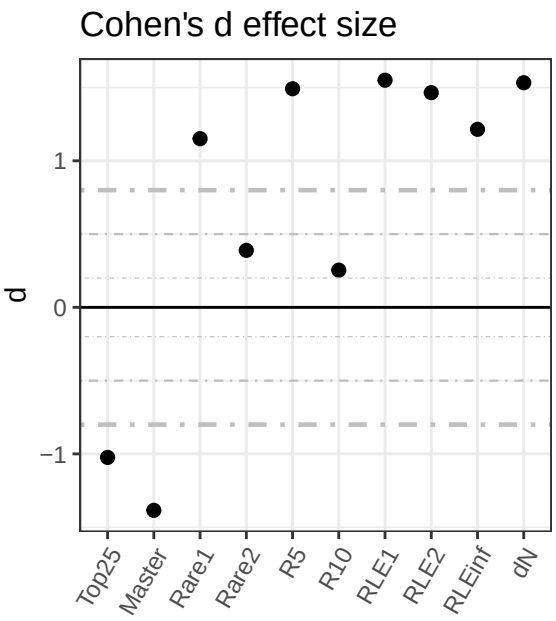

Figure 8: Effect size in double rarefaction test. Dash-dot line borders at 0.2, 0.5 and 0.8

Permutation tests

Test p0.p0 vs p0.p3

In a concise summary, the results for each tested indicator encompass several key statistics. The observed values for the two populations are denoted as $p0.p0$ and $p0.p3$, with their difference represented by *Obs.Diff*. To contextualize these observations within the permutation framework, *Perm.Diff* provides the mean of the differences between the indicators across all permutations under the null hypothesis of equality. The z indicates how many standard deviations the observed difference is from the mean of the permutation distribution, corresponding to the null hypothesis, that helps to interpret permutation p-values resulted as 0. Finally, the permutation *p.value*, and a logical indicating if the observed difference lays inside the 99% CI of the null diffrence. Together, these statistics offer a comprehensive assessment of the differences between the two quasispecies and their statistical relevance.

Table 21: Permutation test results for p3.p0-p0.p0

	p0.p0	p0.p3	Obs.Diff	Perm.Diff	d	z	In.99CI	p.value
Top25	0.8013	0.7854	-0.0159	0e+00	-0.314	-9.968	FALSE	0.000
Master	0.7288	0.7309	0.0020	0e+00	0.036	1.143	TRUE	0.253
Rare1	0.2411	0.2691	0.0281	0e+00	0.122	3.861	FALSE	0.000
Rare2	0.1997	0.1900	-0.0096	-1e-04	-0.118	-3.697	FALSE	0.000
R5	0.0240	0.0210	-0.0030	0e+00	-0.068	-2.159	TRUE	0.030
R10	0.0246	0.0285	0.0039	0e+00	0.098	3.095	FALSE	0.003
RLE1	0.2768	0.2812	0.0045	0e+00	0.090	2.854	FALSE	0.002
RLE2	0.0728	0.0721	-0.0007	0e+00	-0.042	-1.301	TRUE	0.199
RLEinf	0.0365	0.0361	-0.0004	0e+00	-0.047	-1.457	TRUE	0.142
dN	0.1812	0.1868	0.0056	0e+00	0.110	3.482	FALSE	0.000

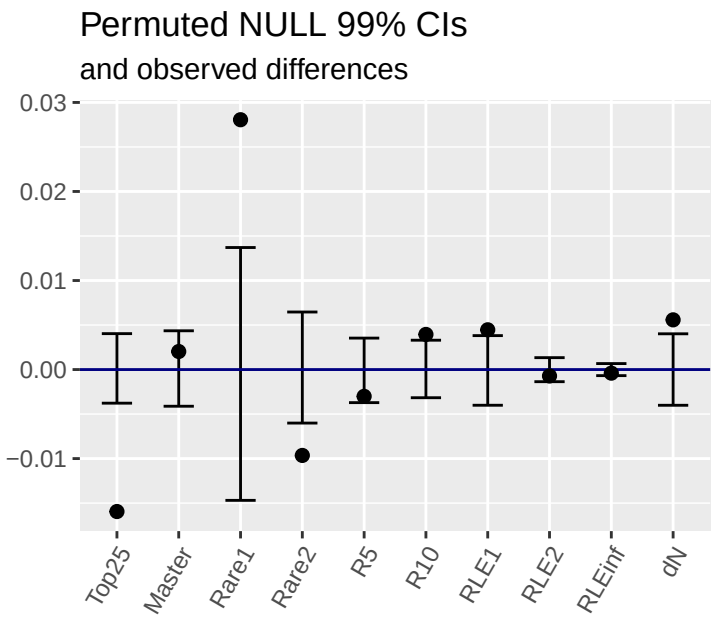

Figure 9: Permutation test results for p0.p3-p0.p0. 99% CI of the difference in the null distribution and observed differences.

Test p0.p0 vs p100.p0

Table 22: Permutation test results for p100.p0-p0.p0

	p0.p0	p100.p0	Obs.Diff	Perm.Diff	d	z	In.99CI	p.value
Top25	0.8013	0.7429	-0.0584	2e-04	-1.114	-35.311	FALSE	0
Master	0.7288	0.4921	-0.2367	2e-04	-3.871	-122.522	FALSE	0
Rare1	0.2411	0.3320	0.0910	-2e-04	1.618	51.266	FALSE	0
Rare2	0.1997	0.2311	0.0314	-2e-04	0.410	13.042	FALSE	0
R5	0.0240	0.0622	0.0382	0e+00	0.552	17.480	FALSE	0
R10	0.0246	0.0660	0.0414	1e-04	0.568	17.926	FALSE	0
RLE1	0.2768	0.3908	0.1141	-2e-04	2.375	75.194	FALSE	0
RLE2	0.0728	0.1475	0.0747	-1e-04	3.523	111.497	FALSE	0
RLEinf	0.0365	0.0778	0.0414	0e+00	3.734	118.170	FALSE	0
dN	0.1812	0.2742	0.0930	-1e-04	2.493	78.945	FALSE	0

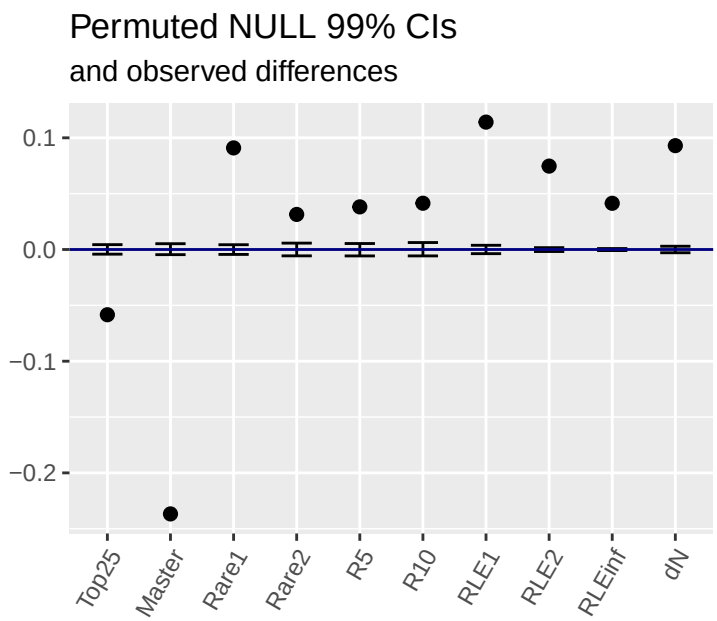

Figure 10: Permutation test results for p100.p0-p0.p0. 99% CI of the difference in the null distribution and observed differences.

Note that in p0.p0 vs p0.p3 permutations test we observed some p-values higher than 0, whereas in this test all p-values are observed as 0. These p-values are reported as <1.e-3. The resulting Z indicates how many standard deviations the observed difference is from the mean of the permutation distribution, corresponding to the null hypothesis.

Test p100.p0 vs p100.p10

Table 23: Permutation test results for p100.p10-p100.p0

	p100.p0	p100.p10	Obs.Diff	Perm.Diff	d	z	In.99CI	p.value
Top25	0.7429	0.7543	0.0115	1e-04	0.214	6.725	FALSE	0
Master	0.4921	0.3930	-0.0990	0e+00	-1.608	-50.818	FALSE	0
Rare1	0.3321	0.2819	-0.0503	0e+00	-0.898	-28.394	FALSE	0
Rare2	0.2310	0.2415	0.0105	-1e-04	0.157	4.991	FALSE	0
R5	0.0621	0.1028	0.0407	0e+00	0.436	13.781	FALSE	0
R10	0.0662	0.0323	-0.0339	1e-04	-0.432	-13.701	FALSE	0
RLE1	0.3885	0.4018	0.0132	0e+00	0.292	9.266	FALSE	0
RLE2	0.1463	0.1788	0.0325	0e+00	1.364	43.113	FALSE	0
RLEinf	0.0772	0.1017	0.0244	0e+00	1.610	50.902	FALSE	0
dN	0.2737	0.2947	0.0210	0e+00	0.602	19.044	FALSE	0

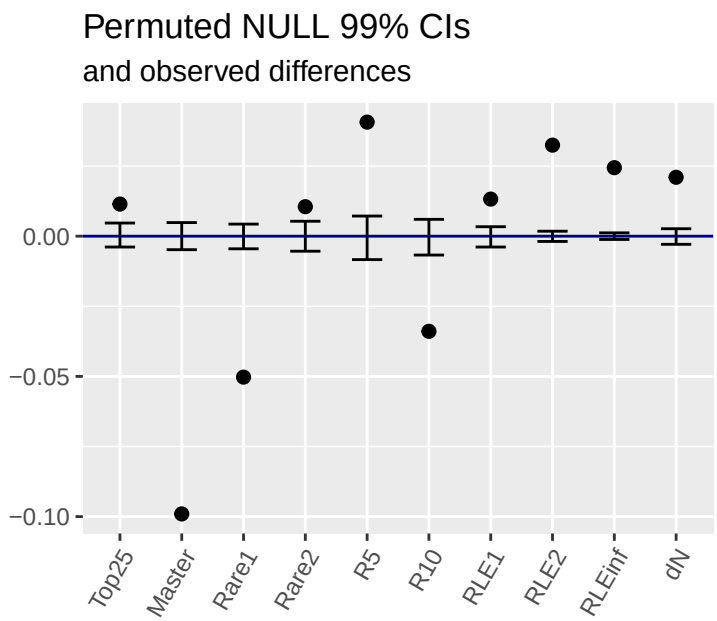

Figure 11: Permutation test results for p100.p10-p100.p0. 99% CI of the difference in the null distribution and observed differences.

Test p100.p0 vs p200.p0

Table 24: Permutation test results for p200.p0-p100.p0

	p100.p0	p200.p0	Obs.Diff	Perm.Diff	d	z	In.99CI	p.value
Top25	0.7429	0.7058	-0.0372	0e+00	-0.651	-20.591	FALSE	0
Master	0.4921	0.4351	-0.0571	0e+00	-0.891	-28.195	FALSE	0
Rare1	0.3321	0.3774	0.0453	0e+00	0.740	23.419	FALSE	0
Rare2	0.2308	0.2574	0.0266	0e+00	0.297	9.401	FALSE	0
R5	0.0622	0.1612	0.0990	0e+00	0.851	26.927	FALSE	0
R10	0.0661	0.0844	0.0183	2e-04	0.154	4.822	FALSE	0
RLE1	0.3909	0.4375	0.0466	0e+00	0.970	30.684	FALSE	0
RLE2	0.1475	0.1712	0.0237	0e+00	0.905	28.640	FALSE	0
RLEinf	0.0779	0.0890	0.0112	0e+00	0.750	23.725	FALSE	0
dN	0.2741	0.3120	0.0379	0e+00	0.973	30.778	FALSE	0

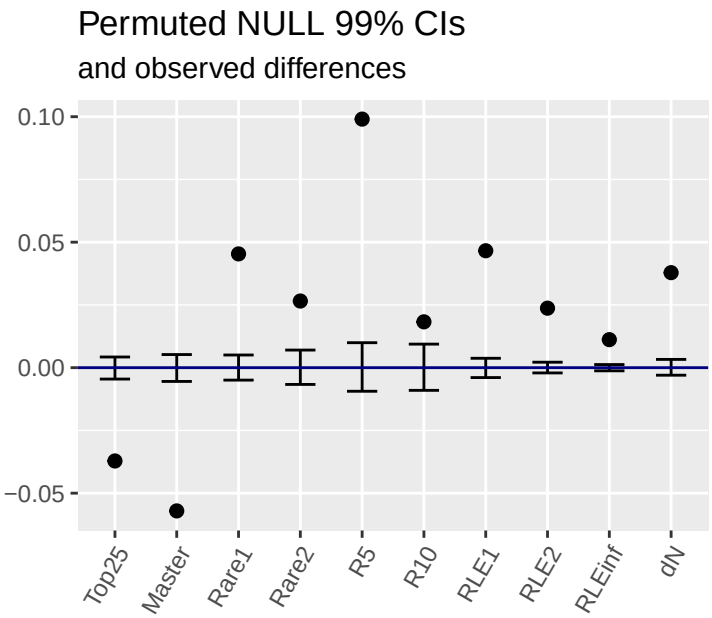

Figure 12: Permutation test results for p200.p0-p100.p0. 99% CI of the difference in the null distribution and observed differences.

Note that whereas all indicators behaved as expected in the long-run, the number of haplotypes remained at the same level. The differences correspond to a maintained diversity with increased evenness in the top haplotypes and globally in the quasispecies.

Observed differences between the two methods

The key distinction between the two methods lies in their rarefaction strategies:

1. Repeated Rarefaction and Z-test:
    - Reference size: A fraction of the smaller sample size.
    - Applied to both samples.
2. Permutation Test:
    - Reference size: The size of the smaller sample.
    - Applied only to the larger sample.

Due to the difference in the rarefaction approach, we anticipate that the permutation test may yield less biased and more precise results compared to the repeated rarefaction method. Besides, diversity values of quasispecies rarefied to an smaller size, are expected to be smaller and with lower variability. This effect may have a bigger impact on z and d values than on the observed difference, particularly for the indicators showing higher sensitivity to sample size variations.

p0.p0 vs p0.p3

This is the test comparing the two samples with lower diversity. The observed differences by the two methods are very similar (Figure 13), however both z and d are bigger in the double rarefaction method for all indicators (Figure 14). With Rare1, the indicator showing the higher difference.

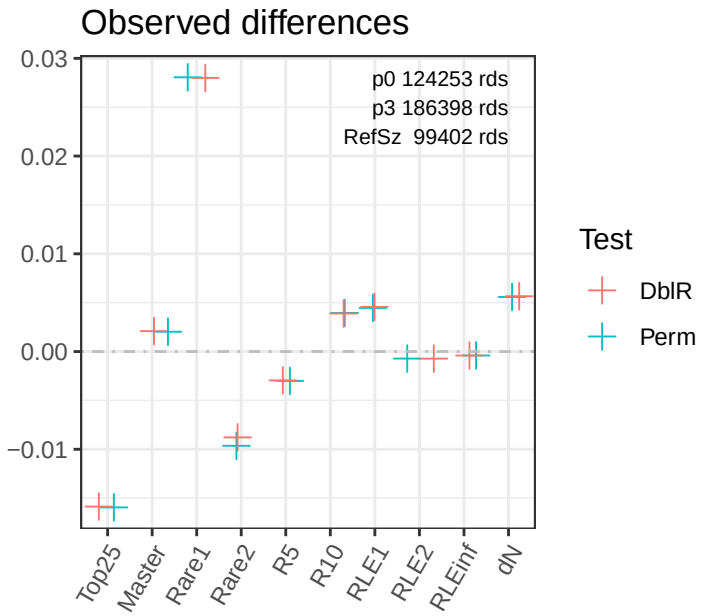

Figure 13: p0.p3 - p0.p0 observed differences. Points horizontally jittered.

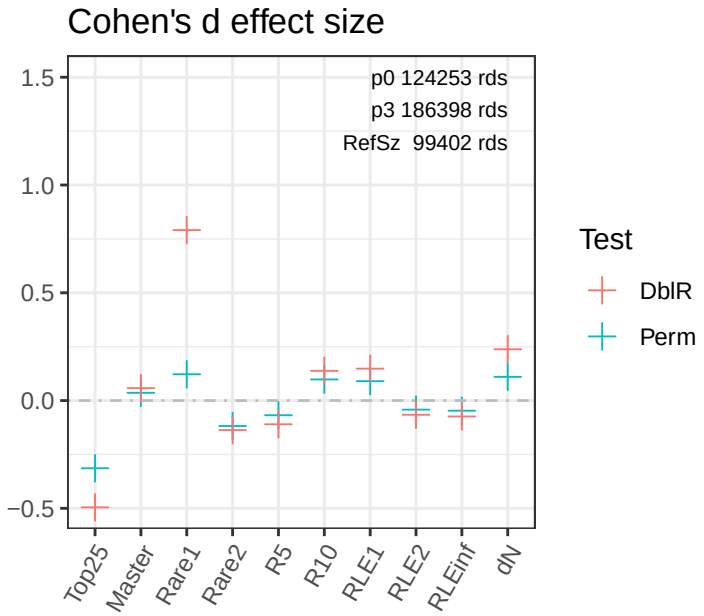

Figure 14: p0.p3 - p0.p0 effect size.

p0.p0 vs p100.p0

The two quasispecies being compared correspond to an evolution of 300 days in cell culture, and clearly show the variations expected on the long-run, with increased maturity. The differences in all indicators are very similar under both methods. Again, z and d values show higher differences under double rarefaction. The higher the effect, the higher the inflation under double rarefaction.

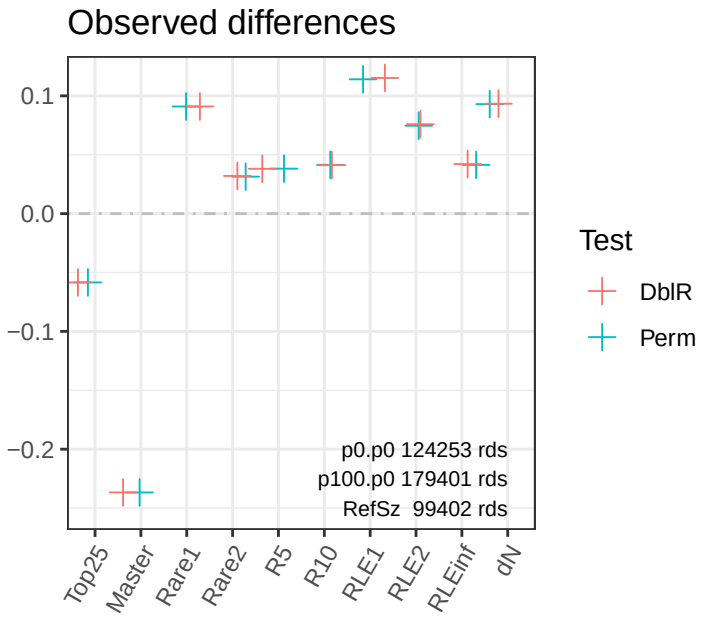

Figure 15: p100.p0 - p0.p0 observed differences. Points horizontally jittered.

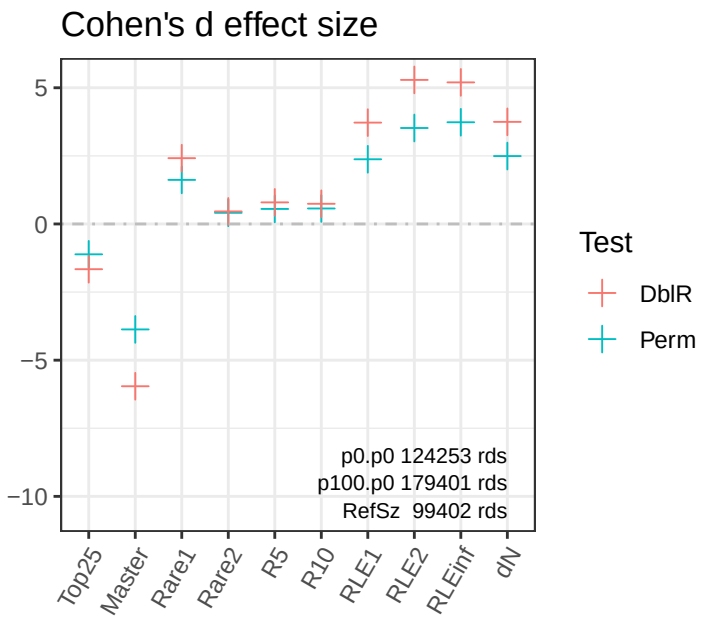

Figure 16: p100.p0 - p0.p0 effect size.

p100.p0 vs p100.p10

These represent the changes occurred in the quasispecies in ten passsages after p100, taht is about 30 days of infection. As in the previous comparison, the differences under both methods are very similar, although with inflated effect sizes under double rarefaction. As before, the higher the effect, the higher the inflation under double rarefaction.

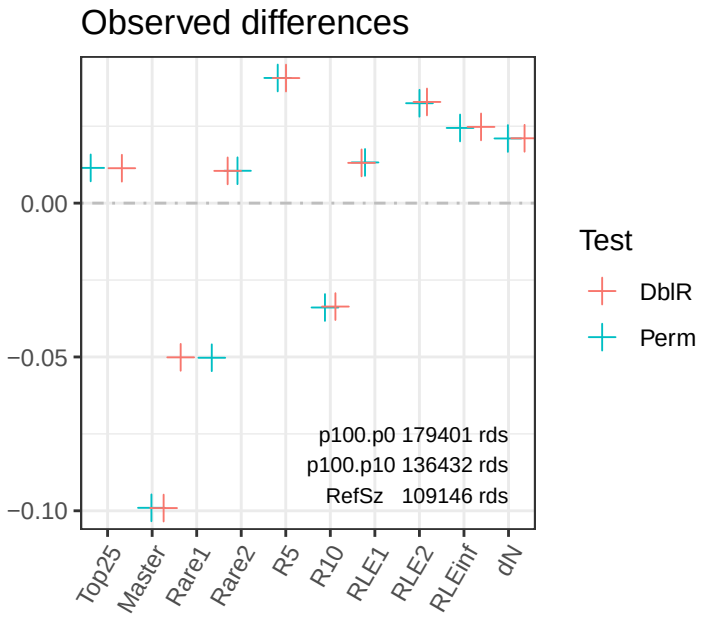

Figure 17: p100.p10 - p100.p0 observed differences. Points horizontally jittered.

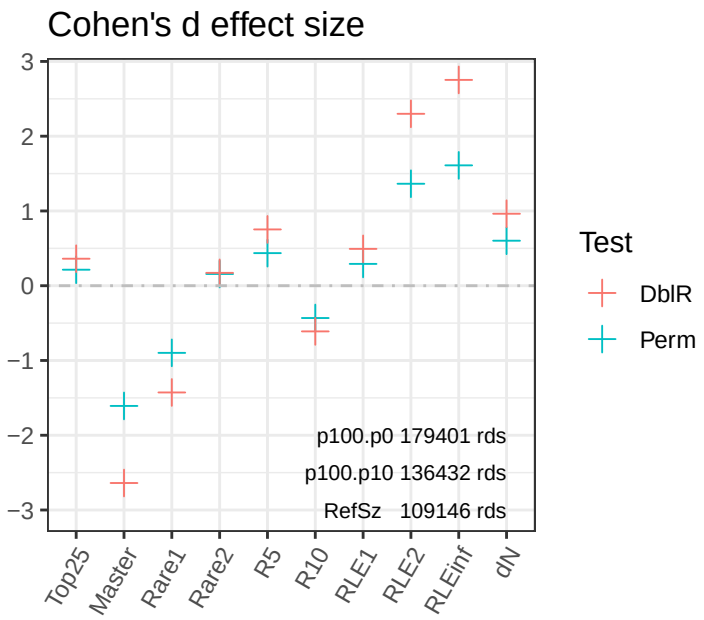

Figure 18: p100.p10 - p100.p0 effect size.

p100.p0 vs p200.p0

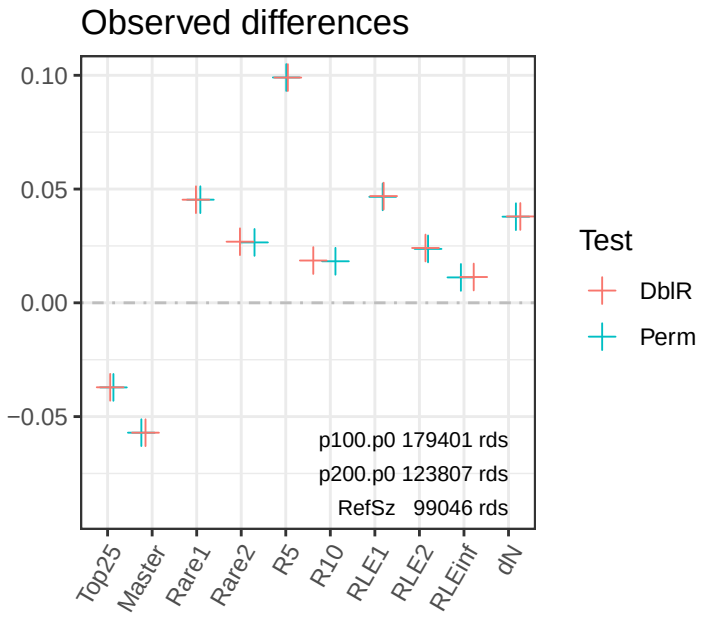

Figure 19: p200.p0 - p100.p0 observed differences. Points horizontally jittered.

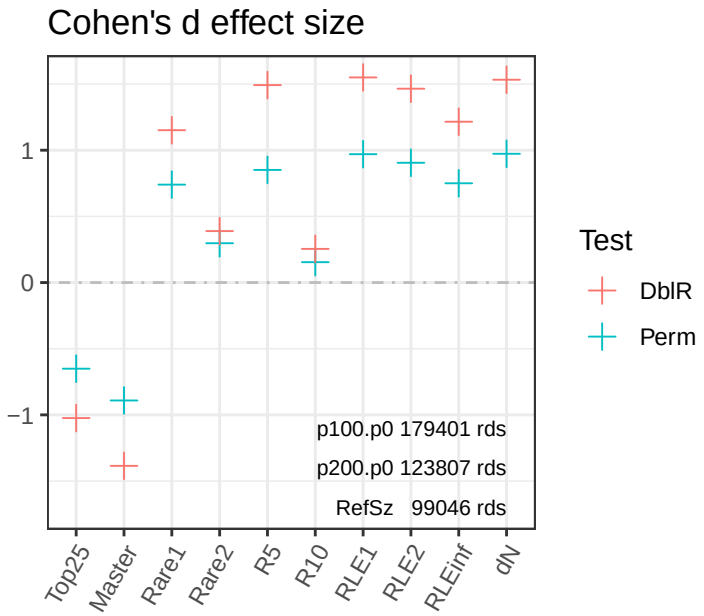

Figure 20: p200.p0 - p100.p0 effect size.

In summary, observed differences and ratios are very similar under both methods, but z and d values are higher in absolute value under the double rarefaction method, where higher biases in observed indicator values may be expected.

Interpretation of results

The interpretation of results should be based on several complementary metrics:

- Absolute observed difference
- Relative observed difference
- Observed ratio
- Cohen's d
- 99% empirical CIs
- z-statistic
- Adjusted p-values

The primary metrics are the observed difference, absolute and relative, and ratio. These are the main results, and often require expert interpretation to determine their biological or clinical significance. The importance of an difference of 1% or 10%, or a ratio of 1.1 or 0.75 depend of the context, and mainly of the level of maturity of the quasispecies being compared. Highly peaked fitness landscape quasispecies are expected to evolve with a sharp decline in Master and Top25, and corresponding increases in Rare1 and Rare2. Highly mature quasispecies in contrast evolve asymptotically in these indicators, with higher variations in the evenness indicators. Cohen's d provides a standardized measure of the magnitude of the effect, as a help to interpret the observed differences in statistical terms. The z-statistic indicates how extreme the observed difference is with respect to the null distribution. It provides information about statistical significance, showing how many standard deviations the observed difference is from the mean of the null distribution. When multiple comparisons are performed, adjusted p-values should be considered to control for the family-wise error rate or false discovery rate.

A comprehensive interpretation of results should consider all these metrics together:

1. Assess the magnitude and direction of the effect using the observed difference and ratio.
2. Evaluate the magnitude of the effect, in statistical terms, using Cohen's d.
3. Consider the statistical significance by means of empirical 99% CIs, z-statistics and adjusted p-values.
4. Contextualize these findings within the biological or clinical framework of the study, and the level of maturity of the involved quasispecies.

This multi-faceted approach provides a more robust interpretation of the differences between quasispecies, balancing practical significance with statistical considerations.

Discusion

Inference with viral quasispecies diversity faces several challenges:

- High sequencing depth is required to adequately capture the complexity of these viral populations.
- Due to the scarcity of clinical samples, and the costs associated with the full NGS lab process, typically only a single experimental sample is taken from a given biospecimen. This sample is then amplified and subjected to NGS to obtain a representation of the quasispecies.
- Small technical inconsistencies can lead to dissimilar coverages between the quasispecies to be compared.
- Several diversity indices and maturity indicators are sensitive to sample size differences, necessitating prior normalization of the samples to be compared.

Given these circumstances, the study of clinical quasispecies samples often requires the comparison of two single experimental samples, except when dealing with patient cohorts. In the later case, various traditional statistical methods are available for inferences between two or multiple groups. However, this is not applicable when comparing single samples in a longitudinal study, for instance.

Each quasispecies sample is represented by a set of NGS reads, which in a simplified form consist of a list of haplotypes and their corresponding read counts. With this sample structure, resampling methods are the most appropriated statistic tools for inference.

We proposed two generative methods:

- Double repeated rarefaction
- Quasispecies permutations including size normalization

Given its potential for reduced bias, we recommend the preferential use of the permutation test for comparing quasispecies structures. The repeated rarefaction method, while yielding similar observed differences and ratios, result in inflated effect sizes and reduced p-values.

In this study we observed further challenges, including:

1. The required depth leads to very large sample sizes, with very high statistical power, resulting in very low p-values for small differences that may have little biological or clinical significance.

2. Although less affected by sample size than the p-values, effect sizes can also be inflated.

These challenges were mitigated by not relying solely on p-values, but instead using a set of metrics with varying susceptibilities to large sample sizes:

- Observed differences and ratios (Minimally susceptible, clinically relevant)
- Cohen's d effect size (Moderately susceptible, clinically relevant)
- z-test (Highly susceptible, statistically relevant)
- 99% Confidence Intervals (Highly susceptible, statistically relevant)
- Multitest adjusted p-values (Highly susceptible, statistically relevant)

Despite very large NGS sample sizes, typically only a single experimental sample from each quasispecies is analyzed. However, these samples are subject to stochastic variations and procedural inconsistencies, meaning that sample replicates will result in values that differ to some extent. The more complex the quasispecies population, the higher the variability in sampling is expected . For critical analysis, the use of a minimum of three experimental replicates is strongly advised to ensure more reliable and representative results.

Short-term variations are not expected to be of biological, clinical, or practical importance except in the context of emerging variants resistant to current treatments. In these cases the study of reduced susceptibility mutations acquire precedence over quasispecies structural changes.

Acknowledgements

The authors employed Perplexity AI for English language correction and minor stylistic enhancements. Additionally, Perplexity AI assisted in retrieving literature references pertaining to resampling statistical methods and viral quasispecies [31].

References

1. Domingo, E., and Perales, C. (2019). Viral quasispecies. PLoS Genet *15*, e1008271. 10.1371/journal.pgen.1008271.

2. Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. Microbiology and Molecular Biology Reviews *76*, 159–216. 10.1128/mmbr.05023-11.

3. Domingo, E., García-Crespo, C., and Perales, C. (2021). Historical perspective on the discovery of the quasispecies concept. Annu Rev Virol *8*, 51–72. 10.1146/annurev-virology-091919-105900.

4. Salicru, M., Menendez, M.L., Morales, D., and Pardo, L. (1993). Asymptotic distributions of (h,fi)-entropies. Commun. Stat. Theory Methods *22*, 2015–2031. 10.1080/03610929308831131.

5. Pardo, l., Morales, D., Salicrú, M., and Menéndez, M.L. (1997). Large sample behavior of entropy measures when parameters are estimated. Communications in Statistics - Theory and Methods *26*, 483–501. 10.1080/03610929708831929.

6. Rey, A.A., Frery, A.C., Lucini, M., Gambini, J., Chagas, E.T.C., and Ramos, H.S. (2023). Asymptotic distribution of certain types of entropy under the multinomial law. Entropy *25*, 734. 10.3390/e25050734.

7. Efron, B., and Tibshirani, R.J. (1994). An introduction to the bootstrap 1st ed. (Chapman; Hall/CRC) 10.1201/9780429246593.

8. Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A., Epstein, R., and P., M.G. (2005). Bootstrap methods and permutation tests. In Introduction to the practice of statistics, W. H. McCabe, ed. (New York, USA: Freeman & Co.), pp. 14.1–14.70.

9. Ramachandran, K.M., and Tsokos, C.P. (2021). Chapter 13 - empirical methods. In Mathematical statistics with applications in r (third edition), K. M. Ramachandran and C. P. Tsokos, eds. (Academic Press), pp. 531–568. 10.1016/B978-0-12-817815-7.00013-0.

10. Gregori, J., Ibañez-Lligoña, M., Colomer-Castell, S., Campos, C., and Quer, J. (2024). Virus quasispecies rarefaction: Subsampling with or without replacement? Viruses *16*, 710. 10.3390/v16050710.

11. Efron, B., and Stein, C. (1981). The Jackknife Estimate of Variance. The Annals of Statistics *9*, 586–596. 10.1214/aos/1176345462.

12. Schloss, P.D. (2024). Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. MSphere *9*. 10.1128/msphere.00354-23.

13. Moore, J.H. (1999). Bootstrapping, permutation testing and the method of surrogate data. Physics in Medicine & Biology *44*, L11. 10.1088/0031-9155/44/6/101.

14. Holt, C.A., and Sullivan, S. (2023). Permutation tests for experimental data (U Iowa Legal Studies Research Paper, Experimental Economics) 10.2139/ssrn.3957609.

15. Kim, H.Y. (2015). Statistical notes for clinical researchers: Effect size. Restor Dent Endod. *40*, 328–31. 10.5395/rde.2015.40.4.328.

16. Nakagawa, S., and Cuthill, I.C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. Biol Rev Camb Philos Soc *82*, 591–605. 10.1111/j.1469-185X.2007.00027.x.

17. Kelley, K., and Preacher, K.J. (2012). On effect size. Psychol Methods *17*. 10.1037/a0028086.

18. Sawilowsky, S. (2009). New effect size rules of thumb. Journal of Modern Applied Statistical Methods. 10.22237/jmasm/1257035100.

19. Ellis, P.D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results (Cambridge University Press).

20. Domingo, E., Soria, M.E., Gallego, I., De Ávila, A.I., García-Crespo, C., Martínez-González, B., Gómez, J., and al., et (2020). A new implication of quasispecies dynamics: Broad virus diversification in absence of external perturbations. Infection, Genetics and Evolution *82*, 104278. 10.1016/j.meegid.2020.104278.

21. Gallego, I., Soria, M.E., García-Crespo, C., Chen, Q., Martínez-Barragán, P., Khalfaoui, S., Martínez-González, B., Sanchez-Martin, I., Palacios-Blanco, I., Ávila, A.I. de, *et al.* (2020). Broad and dynamic diversification of infectious hepatitis c virus in a cell culture environment. J Virol *94*. 10.1128/JVI.01856-19.

22. Gregori, J., Colomer-Castell, S., Ibañez-Lligoña, M., Garcia-Cehic, D., Campos, C., Buti, M., Riveiro-Barciela, M., Andrés, C., Piñana, M., González-Sánchez, A., *et al.* (2024). In-host flat-like quasispecies: Characterization methods and clinical implications. Microorganisms *12*, 1011. 10.3390/microorganisms12051011.

23. Gregori, J., Colomer-Castell, S., Campos C., Ibañez-Lligoña, M., García-Cehic, D., González-Camuesco, A., Cortese, M. F., Tabernero, D., Riveiro-Barciela, M., Buti, M., *et al.* (2025). Evolutionary measures for complex fast-evolving genetic systems. Quasispecies complexity measures - 2. Nature Comunications *submitted*.

24. Gregori, J., Ibañez-Lligoña, M., Colomer-Castell, S., Garcia-Cehic, D., Campos, C., and Quer, J. (2024). Association of liver damage and quasispecies maturity in chronic HCV patients: The fate of a quasispecies. Microorganisms *12*, 2213. 10.3390/microorganisms12112213.

25. Gregori, J., Soria, M.E., Gallego, I., Guerrero-Murillo, M., Esteban, J.I., Quer, J., Perales, C., and Domingo, E. (2018). Rare haplotype load as marker for lethal mutagenesis. PLoS One *13*, e0204877. 10.1371/journal.pone.0204877.

26. R Core Team (2022). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

27. Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2024). Biostrings: Efficient manipulation of biological strings.

28. Xie, Y. (2023). Knitr: A general-purpose package for dynamic report generation in r.

29. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., *et al.* (2019). Welcome to the tidyverse. Journal of Open Source Software *4*, 1686. 10.21105/joss.01686.

30. Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (Springer-Verlag New York).

31. Perplexity AI (2024). Perplexity AI.

Appendix A

R code of main parts

```
###  Diversity values to be tested
#####################################

div.nms <- c('nHpl',
             'Top25','Master','Rare1','Rare2',
             'R5','R10',
             'RLE1','RLE2','RLEinf')

###  Function computing selected diversity indices
diversity <- function(nr)
{
   p <- nr/sum(nr)
   q_0 <- nHpl <- length(p)      # Hill number for q=0
   q_1 <- exp(-sum(p*log(p)))    # Hill number for q=1
   q_2 <- 1/sum(p^2)             # Hill number for q=2
   q_inf <- 1/max(p)             # Hill number for q=Infinity
   RLE1 <- log10(q_1)/log10(q_0)      # Rel. log. evennes for q=1
   RLE2 <- log10(q_2)/log10(q_0)      # Rel. log. evennes for q=2
   RLEinf <- log10(q_inf)/log10(q_0)  # Rel. log. evennes for q=Infinity

   p <- sort(p,decreasing=TRUE)  # Frequencies in decreasing order
   Master <- p[1]                # Master frequency
   Top25 <- sum(p[1:25])         # Top 25 haplotypes, reads fraction
   Rare1 <- sum(p[p<0.01])       # Fraction of reads for all Hpl <1%
   Rare2 <- sum(p[p<0.001])      # Fraction of reads for all Hpl <0.1%
   R5 <- p[5]/mean(p[1:5])       # Evenness in top 5 haplotypes
   R10 <- p[10]/mean(p[1:10])    # Evenness in top 10 haplotypes

   vdiv <- c(round(nHpl,1),Top25,Master,Rare1,Rare2,R5,R10,RLE1,RLE2,RLEinf)
   names(vdiv) <- c('nHpl','Top25','Master','Rare1','Rare2','R5','R10',
                    'RLE1','RLE2','RLEinf')
   return(vdiv)
}

###  Rarefaction to estimate emprirical distributions of diversity indices
#########################################################################

## Rarefaction iterations
B <- 1000
### Factor of reference size
RefSz_f <- 0.8

# Function to perform counts rarefaction
rarefy_counts <- function(counts, size)
{
  raref <- sample(rep(seq_along(counts), counts),
                  size, replace = FALSE )
  tb <- tabulate(raref)
```

```r
  return(tb[tb!=0]) # purge hpl with 0 counts
}

# Repeated rarefaction to target_size
raref_divs <- function(counts, target_size, iterations = B) {
  rvals <- replicate(iterations,
                     diversity(rarefy_counts(counts, target_size)))
  return(list(n=iterations,
              rvals=rvals,
              est=apply(rvals,1,mean),
              var=apply(rvals,1,var),
              CI=apply(rvals,1,function(x) quantile(x,c(0.005,0.5,0.995)))))
}

### Gap between two CIs
ci_gap <- function(ci1, ci2) {
  idx.min <- which.min(c(ci1[1], ci2[1]))
  idx.max <- which.max(c(ci1[2], ci2[2]))
  # One inside the other
  if(idx.min==idx.max) return(NA)
  # Gap. Either + or -
  cm <- rbind(ci1,ci2)
  cm[idx.max,1]-cm[idx.min,2]
}

# CIs based on empirical distributions generated by rarefaction
#   and normality-based statistics.
dr_test <- function(data1,data2,n_perm=10000)
{
  obs_diff <- data2$est - data1$est
  dif <- (data2$rvals - data1$rvals)
  ### 99% difference CI
  CI <- apply(dif,1,function(x) quantile(x,c(0.005,0.995)))
  ### Gap between individual CIs
  gap <- sapply(1:ncol(data1$CI), function(j)
           ci_gap(c(data1$CI[1,j],data1$CI[3,j]),
                  c(data2$CI[1,j],data2$CI[3,j])))

  ### Standard error, z, and normal p-value
  se <- apply(dif,1,sd,na.rm=TRUE)
  z <- apply(dif,1,mean,na.rm=TRUE) / se
  N.p_value <- 2 * pnorm(-abs(z))

  ### Effect size
  sp <- se * sqrt(data1$n)
  d <- apply(dif,1,mean,na.rm=TRUE)/sp

  data.frame(#Ind=names(data1$est),
             data1_est = data1$est,
             data2_est = data2$est,
             dif = apply(dif,1,mean,na.rm=TRUE),
```

```
              CI99.gap = gap,
              CI99.low=CI[1,],
              CI99.upp=CI[2,],
              d=d,
              SE=se,
              z=z,
              N.p_value=N.p_value)
}

###  Permutation tests
#########################

# Number of permutations
Np <- 1000
# Number of sampling cycles in rarefaction
Nr <- 200

# Function to perform rarefaction
rarefy_counts <- function(counts, size) {
  raref <- sample( rep(seq_along(counts), counts),
                   size, replace = FALSE )
  tb <- tabulate(raref)
  return(tb[tb!=0]) # purge hpl with 0 counts
}

# One cycle of rarefaction
rarefy_sample <- function(counts,size)
{
   if(sum(counts) < size*1.05)
   {  return(counts)
   } else {  # Rarefy only if difference in size > 5%
     return(rarefy_counts(counts, size))
   }
}

# Diversity with rarefaction, return mean values.
observed_rarefied <- function(counts,size,n_raref=Nr)
{
  if(sum(counts) < size*1.05)
     return( diversity(counts) )

  rep.div <- replicate(n_raref,{ diversity(rarefy_sample(counts,size)) })
  apply(rep.div,1,mean)
}

# Function to perform permutation test
permutation_test <- function(counts1, counts2, n_permutations=Np, n_rare=Nr)
{
  # Calculate observed diversity with rarefaction
  H1 <- observed_rarefied(counts1,min(sum(counts1),sum(counts2)),n_rare)
  H2 <- observed_rarefied(counts2,min(sum(counts1),sum(counts2)),n_rare)
```

```r
  # Calculate observed difference in diversity indices with rarefaction
  observed_diff <- H2 - H1

  # Initialize a matrix to store the permutation differences
  inms <- div.nms
  perm_diffs <- matrix(0,n_permutations,length(inms))
  colnames(perm_diffs) <- inms

  # Perform permutations
  for (i in 1:n_permutations) {

    # Rarefy the bigger sample
    raref_counts1 <- rarefy_sample(counts1,min(sum(counts1),sum(counts2)))
    raref_counts2 <- rarefy_sample(counts2,min(sum(counts1),sum(counts2)))

    # Combine the reads from both samples to a pool of reads
    combined_reads <- c(rep(seq_along(raref_counts1),raref_counts1),
                        rep(seq_along(raref_counts2),raref_counts2))
    n1 <- sum(raref_counts1)
    n2 <- sum(raref_counts2)

    # Randomly shuffle the pooled reads
    permuted_reads <- sample(combined_reads)

    # Split the permuted counts into two new groups
    perm_reads1 <- permuted_reads[1:n1]
    perm_counts1 <- tabulate(perm_reads1)
    perm_counts1 <- perm_counts1[perm_counts1!=0]
    perm_reads2 <- permuted_reads[(n1+1):(n1+n2)]
    perm_counts2 <- tabulate(perm_reads2)
    perm_counts2 <- perm_counts2[perm_counts2!=0]

    # Calculate the diversity indices for the permuted groups
    perm_diffs[i,] <- diversity(perm_counts2) - diversity(perm_counts1)
  }

  perm_mean <- apply(perm_diffs,2,mean)
  perm_SD <- apply(perm_diffs,2,sd)
  perm_CI <- apply(perm_diffs,2,function(x) quantile(x,c(0.005,0.5,0.995)))

    # Return the test results
  list(
    H1=H1, H2=H2,
    observed_diff = observed_diff,
    perm_mean = perm_mean,
    perm_SD = perm_SD,
    perm_CI = perm_CI,
    perm_distr = perm_diffs
  )
}
```

Session info

R version 4.3.3 (2024-02-29 ucrt) Platform: x86_64-w64-mingw32/x64 (64-bit) Running under: Windows 11 x64 (build 22631)

Matrix products: default

locale: [1] LC_COLLATE=Catalan_Spain.utf8 LC_CTYPE=Catalan_Spain.utf8
[3] LC_MONETARY=Catalan_Spain.utf8 LC_NUMERIC=C
[5] LC_TIME=Catalan_Spain.utf8

time zone: Europe/Madrid tzcode source: internal

attached base packages: [1] stats4 stats graphics grDevices utils datasets methods
[8] base

other attached packages: [1] dqrng_0.4.1 gridExtra_2.3 lubridate_1.9.4
[4] forcats_1.0.0 stringr_1.5.1 dplyr_1.1.4
[7] purrr_1.0.2 readr_2.1.5 tidyr_1.3.1
[10] tibble_3.2.1 ggplot2_3.5.1 tidyverse_2.0.0
[13] Biostrings_2.70.3 GenomeInfoDb_1.38.8 XVector_0.42.0
[16] IRanges_2.36.0 S4Vectors_0.40.2 BiocGenerics_0.48.1

loaded via a namespace (and not attached): [1] utf8_1.2.4 generics_0.1.3 bitops_1.0-9
[4] stringi_1.8.4 hms_1.1.3 digest_0.6.37
[7] magrittr_2.0.3 evaluate_1.0.1 grid_4.3.3
[10] timechange_0.3.0 fastmap_1.2.0 tinytex_0.54
[13] fansi_1.0.6 scales_1.3.0 cli_3.6.3
[16] rlang_1.1.4 crayon_1.5.3 munsell_0.5.1
[19] withr_3.0.2 yaml_2.3.10 tools_4.3.3
[22] tzdb_0.4.0 colorspace_2.1-1 GenomeInfoDbData_1.2.11 [25] vctrs_0.6.5 R6_2.5.1 lifecycle_1.0.4
[28] zlibbioc_1.48.2 pkgconfig_2.0.3 pillar_1.9.0
[31] gtable_0.3.6 Rcpp_1.0.13-1 glue_1.8.0
[34] xfun_0.49 tidyselect_1.2.1 rstudioapi_0.17.1
[37] knitr_1.49 farver_2.1.2 htmltools_0.5.8.1
[40] labeling_0.4.3 rmarkdown_2.29 compiler_4.3.3
[43] RCurl_1.98-1.16