Inference with Viral Quasispecies

Methods for Individual Samples Comparative Analysis

Josep Gregori*

2024-12-30

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

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. 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 the variance of some diversity indices, 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 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.

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

Methods

A resampling method commonly used to estimate variances is the bootstrap (Efron and Tibshirani 1994; Hesterberg et al. 2005; Ramachandran and Tsokos 2021), 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 (Gregori, Ibañez-Lligoña, Colomer-Castell, Campos, et al. 2024). 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 (Gregori, Ibañez-Lligoña, Colomer-Castell, Campos, et al. 2024).

In this situation, a valid alternative to the bootstrap is the jackknife (Efron and Stein 1981), 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.

Repeated rarefaction

When the two samples to be compared are unbalanced, rarefaction is the preferred normalization method (Schloss 2024; Gregori, Ibañez-Lligoña, Colomer-Castell, Campos, et al. 2024). 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 (Gregori, Ibañez-Lligoña, Colomer-Castell, Campos, et al. 2024).

While bootstrap and jackknife methods are not suitable for 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 half the size of the smaller sample as the reference.

Our primary objective is to assess significant differences in the structure and composition of two quasispecies, focusing on observed effect sizes of a minimum magnitude. The emphasis is on comparison rather than precise estimation. Double rarefaction to half the size of the smaller sample 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 and compositional differences between quasispecies samples.

z-test

With diversity estimates and corresponding variances, obtained from repeated rarefaction, we may use a t-test or a z-test to assess differences through p-values and confidence intervals. 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 distribution of the difference 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

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 (Moore 1999; Hesterberg et al. 2005; Holt and Sullivan 2023) 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. Compute the difference between these indicators as the observed test statistic.
2. Inside a resampling loop for B permutations to generate a null distribution of differences in diversity, at each cycle:
   i. Get rarefied haplotype counts of the bigger sample.
   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.
3. Compare the observed difference to this null distribution to obtain a p-value.
4. Calculate the Standardized Permutation Statistic (SPS), Z, as the observed difference minus the mean of the permuted differences, divided by the standard deviation of the permuted differences, the estimated null distribution.

$$Z_{\text{perm}} = \frac{T_{\text{obs}} - T_{\text{perm}}}{S_{\text{perm}}}$$

5. Calculate an approximated Cohen's d effect size as:

$$d_{\text{perm}} = \frac{Z_{\text{perm}}}{\sqrt{N_p}}$$

   where N_p is the number of permutations.

The Z statistic indicates how many standard deviations the observed difference 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.

Note that in the double rarefaction method both samples were rarefied to half the size of the smaller sample, while in the permutations method only the bigger sample is rarefied to the size of the smaler sample. Before each permutation cycle a rarefaction cycle is performed.

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 (Kim 2015; Nakagawa and Cuthill 2007; Kelley and Preacher 2012; Sawilowsky 2009).

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 (Kim 2015; Nakagawa and Cuthill 2007; Kelley and Preacher 2012; Sawilowsky 2009).

$$d = \frac{\bar{D}_1 - \bar{D}_2}{S_{\text{pool}}} \tag{1}$$

where,

$$S_{\text{pool}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{2}$$

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

$$z = \frac{\bar{D}_1 - \bar{D}_2}{S_{\text{pool}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3}$$

These statistics are related as:

$$d = z \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{4}$$

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 (Sawilowsky 2009; Kim 2015; Ellis 2010):

Table 1: Cohen's d vlues 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.

$$NOV = \left[1 - 2\Phi\left(\frac{-\text{abs}(d)}{2}\right)\right] \times 100 \tag{5}$$

(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.

$$PS = \Phi\left(\text{abs}(d)/\sqrt{2}\right) \times 100 \tag{6}$$

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

Table 2 shows the expected level of each selected quasispecies indicator in the two limiting cases, regular vs flat-like quasispecies (Gregori, Colomer-Castell, et al. 2024):

Table 2: Selected maturity indicators.

Indicator	Description	Regular Qs	Flat-like Qs
Master	Dominant haplotype frequency	High	Low
Top25	Fraction of reads for top 25 hpl.	High	Low
Rare1	Fraction of reads for hpl $\leq 1\%$	Low	High
Rare2	Fraction of reads for hpl $\leq 0.1\%$	Low	High
Singl	Fraction of singletons	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
R25	Evenness among top 25 haplotypes	Low	High

Relative logarithmic evenness

$$RLEq = RLE(p, q) = \log_{10}(D(p, q))/\log_{10}(D(p, 0)) \tag{7}$$

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

$$D(p,q) = \left(\sum_{i=1}^{H} p_i^q \right)^{1/(1-q)} \tag{8}$$

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:

$$\text{Top25}(p) = \left(\sum_{i=1}^{m} p_i \;\middle|\; m = \min(25, H) \right) \tag{9}$$

$$\text{Top25R}(p) = \frac{\text{Top25}(p)}{p_1} \tag{10}$$

Evenness on the set of top haplotypes

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

$$R_k(p) = \left(\frac{l \cdot p_l}{\sum_{i=1}^{l} p_i} \;\middle|\; l = \min(k, H) \right) \tag{11}$$

Singletons and very rare haplotypes

The fraction of quasispecies reads for singleton haplotypes is an indicator of quasispecies maturity (Gregori, Colomer-Castell, et al. 2024), and represents the fraction of rarest haplotypes that has been observed. It may be also used as a marker of mutagenic treatments (Gregori et al. 2018), however it shows poor behavior under rarefaction. Suppose that n/N is the fraction of singleton reads in the full sample, before rarefaction, and suppose that it has to be rarefied to half this size. The expected number of singletons in the rarefied sample is $n/2$, however the observed number of singletons is usually higher, and much higher in some cases, like mutagenic treatments or mature quasispecies. The reason is that when drawing reads without replacement from the full sample, all singletons will be observed as singletons, and expected to be n/2. However, a number of haplotypes represented by 2, 3, 4 reads, and above, will also be observed as singletons, contributing to an inflation in the observed number of singletons in the rarefied quasispecies. As a result, the fraction of singletons in the rarefied quasispecies is biased to higher values.

Samples description

The data used in this study is taken from (Gregori et al. 2018). 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 (Domingo et al. 2020; Gallego et al. 2020). 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).

Results

Samples and NGS coverage

Table 3: 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 4: Quasispecies maturity indicators. Raw values.

ID	nHpl	Master	Top25	Rare1	Rare2	Singl
p0.p0	5858	0.72884	0.80131	0.24107	0.19969	0.03675
p0.p3	8325	0.73095	0.78525	0.26905	0.18898	0.03709
p100.p0	11948	0.49214	0.74294	0.33202	0.23125	0.05103
p100.p10	9744	0.39303	0.75431	0.28186	0.24154	0.05034
p200.p0	11490	0.43506	0.70577	0.37739	0.25735	0.06982

ID	R5	R10	R25	RLE1	RLE2	RLEinf
p0.p0	0.02398	0.02458	0.03114	0.27677	0.07278	0.03646
p0.p3	0.02070	0.02845	0.04714	0.27276	0.06939	0.03472
p100.p0	0.06210	0.06696	0.06227	0.38164	0.14308	0.07552
p100.p10	0.10277	0.03230	0.03571	0.40176	0.17882	0.10168
p200.p0	0.16119	0.08438	0.07582	0.43750	0.17121	0.08902

Maturity indicator values on rarefied quasispecies

Table 5: Quasispecies maturity indicators. Rarefied values to 100,000 reads.

ID	nHpl	Master	Top25	Rare1	Rare2	Singl
p0.p0	4955	0.72881	0.80134	0.24112	0.19861	0.03775
p0.p3	5044	0.73091	0.78557	0.26909	0.18993	0.03926
p100.p0	7695	0.49214	0.74292	0.33210	0.23082	0.05691
p100.p10	7846	0.39302	0.75443	0.28181	0.24113	0.05422
p200.p0	9785	0.43510	0.70582	0.37738	0.25740	0.07320

ID	R5	R10	R25	RLE1	RLE2	RLEinf
p0.p0	0.02400	0.02455	0.03122	0.28114	0.07421	0.03718
p0.p3	0.02095	0.02854	0.04711	0.28565	0.07348	0.03677
p100.p0	0.06209	0.06577	0.06199	0.39629	0.15012	0.07923
p100.p10	0.10259	0.03237	0.03610	0.40919	0.18314	0.10415
p200.p0	0.16129	0.08435	0.07627	0.44317	0.17418	0.09057

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.

	nHpl	Master	Top25	Rare1	Rare2
p0.p3 - p0.p0	89	0.00210	-0.01577	0.02797	-0.00868
p100.p0 - p0.p0	2740	-0.23667	-0.05843	0.09098	0.03221
p100.p10 - p100.p0	151	-0.09912	0.01151	-0.05029	0.01031
p200.p0 - p100.p0	2090	-0.05704	-0.03709	0.04528	0.02658

	R5	R10	R25	RLE1	RLE2	RLEinf
p0.p3 - p0.p0	-0.00304	0.00400	0.01589	0.00451	-0.00073	-0.00041
p100.p0 - p0.p0	0.03810	0.04122	0.03077	0.11514	0.07590	0.04206
p100.p10 - p100.p0	0.04050	-0.03340	-0.02588	0.01291	0.03302	0.02492
p200.p0 - p100.p0	0.09919	0.01858	0.01428	0.04688	0.02406	0.01133

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.

Diversity values to be tested

```
div.nms <- c('Master','Top25','Rare2','R10','RLE1','RLEinf')

###  Function computing selected diversity indices
diversity <- function(nr)
{
   p <- nr/sum(nr)
   q_0 <- 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
   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
   Rare2 <- sum(p[p<0.001])      # Fraction of reads for all Hpl <0.1%
   R10 <- p[10]/mean(p[1:10])    # Evenness in top 10 haplotypes

   vdiv <- c(Master,Top25,Rare2,R10,RLE1,RLEinf)
   names(vdiv) <- c('Master','Top25','Rare2','R10','RLE1','RLEinf')
   return(vdiv)
}
```

Repeated rarefaction to estimate mean and variance of diversity indices

```
# Function to perform counts 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
}

# Rarefaction for larger sample
raref_divs <- function(counts, target_size, iterations = 1000) {
  rvals <- replicate(iterations,
                     diversity(rarefy_counts(counts, target_size)))
  return(list(n=iterations,
              est=apply(rvals,1,mean),
              var=apply(rvals,1,var)))
}
```

Normal z-test

```
# z-test based on normality
z_test <- function(data1,data2)
{ dif <- (data2$est - data1$est)
```

```r
  sp <- sqrt(((data1$n-1)*sqrt(data1$var) + (data2$n-1)*sqrt(data2$var))/
              (data1$n+data2$n-2))
  se <- sp * sqrt(1/data1$n+1/data2$n)
  z <- dif / se
  p_value <- 2 * pnorm(-abs(z))
  delta.CI <- qnorm(0.995) * se
  data.frame(#Ind=names(data1$est),
             data1_est = data1$est,
             data2_est = data2$est,
             dif = dif,
             Sp=sp,
             d=dif/sp,
             CI99.lower = dif-delta.CI,
             CI99.upper = dif+delta.CI,
             SE=se,
             z=z,
             p_value=p_value)
}
```

Test p0.p0 vs p0.p3

```
##      ID                  FlNm Basal Pass  Reads nHpl
## 2 p0.p3 p0.Ctl.p3.A2.PrFW.fna    p0   p3 186398 8325
## 1 p0.p0 p0.Ctl.p0.A2.PrFW.fna    p0   p0 124253 5858
```

```
##
## Rarefaction size: 62126
```

```
set.seed(9876)

idx1 <- pairs[iTest,2]
ID1 <- dict$ID[idx1]
fasta1 <- dict$FlNm[idx1]
seqs <- readDNAStringSet(file.path(ntDir,fasta1))
parts <- t(sapply(names(seqs),function(str) strsplit(str,split='\\|')[[1]]))
rds1 <- sort(as.integer(parts[,2]),decreasing=TRUE)

idx2 <- pairs[iTest,1]
ID2 <- dict$ID[idx2]
fasta2 <- dict$FlNm[idx2]
seqs <- readDNAStringSet(file.path(ntDir,fasta2))
parts <- t(sapply(names(seqs),function(str) strsplit(str,split='\\|')[[1]]))
rds2 <- sort(as.integer(parts[,2]),decreasing=TRUE)

n1 <- sum(rds1)
n2 <- sum(rds2)

data1 <- raref_divs(rds1, ref.sz, iterations = 1000)
data2 <- raref_divs(rds2, ref.sz, iterations = 1000)
```

Repeated rarefaction

```
qs.rare <- data.frame(Q1=data1$est,Var1=data1$var) %>%
  cbind(data.frame(Q2=data2$est,Var2=data2$var)) %>%
  mutate(across(where(is.numeric),~signif(.,4)))
colnames(qs.rare) <- c(ID1,'Var1',ID2,'Var2')
qs.rare
```

```
##          p0.p0      Var1   p0.p3      Var2
## Master 0.72890 1.605e-06 0.73100 2.338e-06
## Top25  0.80150 1.279e-06 0.78580 1.873e-06
## Rare2  0.19860 4.038e-06 0.19050 4.224e-06
## R10    0.02450 9.451e-07 0.02851 1.044e-06
## RLE1   0.29050 1.257e-06 0.29540 1.861e-06
## RLEinf 0.03878 4.080e-08 0.03837 5.869e-08
```

Effect size

```
cat(pairs.ID$Test[iTest])
```

```
## p0.p3 - p0.p0
```

```r
rrt.p0vsp3_z <- res_z <- z_test(data1,data2)
colnames(res_z)[1:3] <- c(ID1,ID2,paste(c(ID2,ID1),collapse='-'))

PS <- round(pnorm(abs(res_z$d)/sqrt(2))*100,1)
data.frame(Diff=data2$est-data1$est,
           RelDiff=(data2$est-data1$est)/data1$est,
           Ratio=data2$est/data1$est,
           d=round(res_z$d,3),
           NOV=round((1-2*pnorm(-abs(res_z$d)/2))*100,1),
           PS=PS) %>%
  knitr::kable(digits=5)
```

	Diff	RelDiff	Ratio	d	NOV	PS
Master	0.00212	0.00290	1.00290	0.057	2.3	51.6
Top25	-0.01577	-0.01968	0.98032	-0.446	17.7	62.4
Rare2	-0.00816	-0.04107	0.95893	-0.181	7.2	55.1
R10	0.00401	0.16358	1.16358	0.127	5.1	53.6
RLE1	0.00490	0.01686	1.01686	0.139	5.5	53.9
RLEinf	-0.00040	-0.01041	0.98959	-0.027	1.1	50.8

z-test

```r
res_z %>%
  #select(-c(SE,Sp,z)) %>%
  mutate(adj.p=p.adjust(p_value,method='BH')) %>%
  mutate(across(where(is.numeric),~signif(.,4))) %>%
  select(1:3,z,p_value,adj.p,d) %>%
  knitr::kable()
```

	p0.p0	p0.p3	p0.p3-p0.p0	z	p_value	adj.p	d
Master	0.72890	0.73100	0.0021160	1.2650	0.2058000	0.2469000	0.05659
Top25	0.80150	0.78580	-0.0157700	-9.9760	0.0000000	0.0000000	-0.44610
Rare2	0.19860	0.19050	-0.0081590	-4.0470	0.0000519	0.0001558	-0.18100
R10	0.02450	0.02851	0.0040080	2.8380	0.0045350	0.0068020	0.12690
RLE1	0.29050	0.29540	0.0048990	3.1080	0.0018850	0.0037710	0.13900
RLEinf	0.03878	0.03837	-0.0004038	-0.6058	0.5446000	0.5446000	-0.02709

Top25, Rare2, R10, and RLE1 are statistically significant with adjusted p-values below 0.01. As noted above, this result may be missleading as no major differences where expected in quasispecies structure in only three passages, representing 9 days of infection (72hrs each passage). However, in the short term quasispecies dynamics are expected to cause limited stochastic fluctuations in quasispecies diversity. Note that observed differences and effect sizes, d, are small, except for Top25, which represents the aggregated frequencies of the 25 most abundant haplotypes. If any concern is raised about a similar result, the use of experimental sample replicates from the same biospecimen may be indicated.

Test p0.p0 vs p100.p0

```
##        ID                    FlNm Basal Pass  Reads  nHpl
## 3 p100.p0 p100.Ctl.p0.A2.PrFW.fna  p100   p0 179401 11948
## 1   p0.p0   p0.Ctl.p0.A2.PrFW.fna    p0   p0 124253  5858
```

```
##
## Rarefaction size: 62126
```

Repeated rarefaction

```
##          p0.p0      Var1 p100.p0      Var2
## Master 0.72890 1.605e-06 0.49220 2.695e-06
## Top25  0.80150 1.279e-06 0.74310 1.964e-06
## Rare2  0.19860 4.038e-06 0.23110 4.387e-06
## R10    0.02450 9.451e-07 0.06521 6.627e-06
## RLE1   0.29050 1.257e-06 0.40820 1.441e-06
## RLEinf 0.03878 4.080e-08 0.08243 1.418e-07
```

Effect size

```
## p100.p0 - p0.p0
```

	Diff	RelDiff	Ratio	d	NOV	PS
Master	-0.23671	-0.32476	0.67524	-6.207	99.8	100.0
Top25	-0.05843	-0.07289	0.92711	-1.642	58.8	87.7
Rare2	0.03244	0.16329	1.16329	0.716	28.0	69.4
R10	0.04071	1.66145	2.66145	0.967	37.1	75.3
RLE1	0.11768	0.40502	1.40502	3.454	91.6	99.3
RLEinf	0.04365	1.12562	2.12562	2.566	80.1	96.5

z-test

	p0.p0	p100.p0	p100.p0-p0.p0	z	p_value	adj.p	d
Master	0.72890	0.49220	-0.23670	-138.80	0	0	-6.2070
Top25	0.80150	0.74310	-0.05843	-36.72	0	0	-1.6420
Rare2	0.19860	0.23110	0.03244	16.01	0	0	0.7160
R10	0.02450	0.06521	0.04071	21.62	0	0	0.9668
RLE1	0.29050	0.40820	0.11770	77.23	0	0	3.4540
RLEinf	0.03878	0.08243	0.04365	57.38	0	0	2.5660

This case represents 300 days of infection time without external perturbations, all indicators are statistically significant, with adjusted p-values <1.e-6. Observed differences are all in line with expected changes with quasispecies maturation (Gregori, Colomer-Castell, et al. 2024; Gregori, Ibañez-Lligoña, Colomer-Castell, Garcia-Cehic, et al. 2024), and the magnitudes of the size effects are remarkable. Mainly for master and RLE1, and RLEinf, indicating a flatter more even distribution of haplotypes, with decreased frequency in dominant haplotypes.

Test p100.p0 vs p100.p10

```
##         ID                     FlNm Basal Pass  Reads  nHpl
## 4 p100.p10 p100.Ctl.p10.A2.PrFW.fna  p100  p10 136432  9744
## 3  p100.p0  p100.Ctl.p0.A2.PrFW.fna  p100   p0 179401 11948
```

```
##
## Rarefaction size: 68216
```

Repeated rarefaction

```
##        p100.p0      Var1 p100.p10      Var2
## Master 0.49210 2.306e-06  0.39310 1.886e-06
## Top25  0.74300 1.855e-06  0.75450 1.346e-06
## Rare2  0.23120 4.424e-06  0.24180 3.331e-06
## R10    0.06527 5.524e-06  0.03219 2.688e-06
## RLE1   0.40590 1.296e-06  0.41840 9.584e-07
## RLEinf 0.08180 1.183e-07  0.10730 1.535e-07
```

Effect size

```
## p100.p10 - p100.p0
```

	Diff	RelDiff	Ratio	d	NOV	PS
Master	-0.09905	-0.20127	0.79873	-2.605	80.7	96.7
Top25	0.01148	0.01545	1.01545	0.323	12.8	59.0
Rare2	0.01061	0.04589	1.04589	0.239	9.5	56.7
R10	-0.03308	-0.50681	0.49319	-0.741	28.9	70.0
RLE1	0.01243	0.03063	1.03063	0.382	15.2	60.6
RLEinf	0.02547	0.31136	1.31136	1.328	49.3	82.6

z-test

	p100.p0	p100.p10	p100.p10-p100.p0	z	p_value	adj.p	d
Master	0.49210	0.39310	-0.09905	-58.250	0e+00	0e+00	-2.6050
Top25	0.74300	0.75450	0.01148	7.230	0e+00	0e+00	0.3233
Rare2	0.23120	0.24180	0.01061	5.352	1e-07	1e-07	0.2393
R10	0.06527	0.03219	-0.03308	-16.560	0e+00	0e+00	-0.7406
RLE1	0.40590	0.41840	0.01243	8.544	0e+00	0e+00	0.3821
RLEinf	0.08180	0.10730	0.02547	29.690	0e+00	0e+00	1.3280

Extra 30 days of infection, for a quasispecies, p100, showing incipient maturity. All indicators are statistically significant with adjusted p-values <1.-e7. Effect sizes are relevant for Master and RLEinf. Note that both indicators are highly correlated. Observed differences and effect sizes for the other indicators are small. The direction of changes is opposite for Top25 and R10, which may be explained by quasispecies dynamics in the short term.

Test p100.p0 vs p200.p0

```
##        ID                    FlNm Basal Pass  Reads  nHpl
## 5 p200.p0 p200.Ctl.p0.A2.PrFW.fna  p200   p0 123807 11490
## 3 p100.p0 p100.Ctl.p0.A2.PrFW.fna  p100   p0 179401 11948
```

```
##
## Rarefaction size: 61904
```

Repeated rarefaction

```
##        p100.p0      Var1 p200.p0      Var2
## Master 0.49220 2.782e-06 0.43510 2.017e-06
## Top25  0.74310 2.124e-06 0.70590 1.692e-06
## Rare2  0.23050 4.383e-06 0.25710 4.224e-06
## R10    0.06517 6.173e-06 0.08509 7.059e-06
## RLE1   0.40830 1.550e-06 0.45600 1.106e-06
## RLEinf 0.08245 1.435e-07 0.09423 1.287e-07
```

Effect size

```
## p200.p0 - p100.p0
```

	Diff	RelDiff	Ratio	d	NOV	PS
Master	-0.05709	-0.11600	0.88400	-1.453	53.2	84.8
Top25	-0.03714	-0.04998	0.95002	-1.000	38.3	76.0
Rare2	0.02657	0.11525	1.11525	0.583	22.9	66.0
R10	0.01992	0.30560	1.30560	0.393	15.6	60.9
RLE1	0.04769	0.11678	1.11678	1.407	51.8	84.0
RLEinf	0.01178	0.14285	1.14285	0.613	24.1	66.8

z-test

	p100.p0	p200.p0	p200.p0-p100.p0	z	p_value	adj.p	d
Master	0.49220	0.43510	-0.05709	-32.490	0	0	-1.4530
Top25	0.74310	0.70590	-0.03714	-22.360	0	0	-1.0000
Rare2	0.23050	0.25710	0.02657	13.040	0	0	0.5833
R10	0.06517	0.08509	0.01992	8.784	0	0	0.3928
RLE1	0.40830	0.45600	0.04769	31.460	0	0	1.4070
RLEinf	0.08245	0.09423	0.01178	13.710	0	0	0.6133

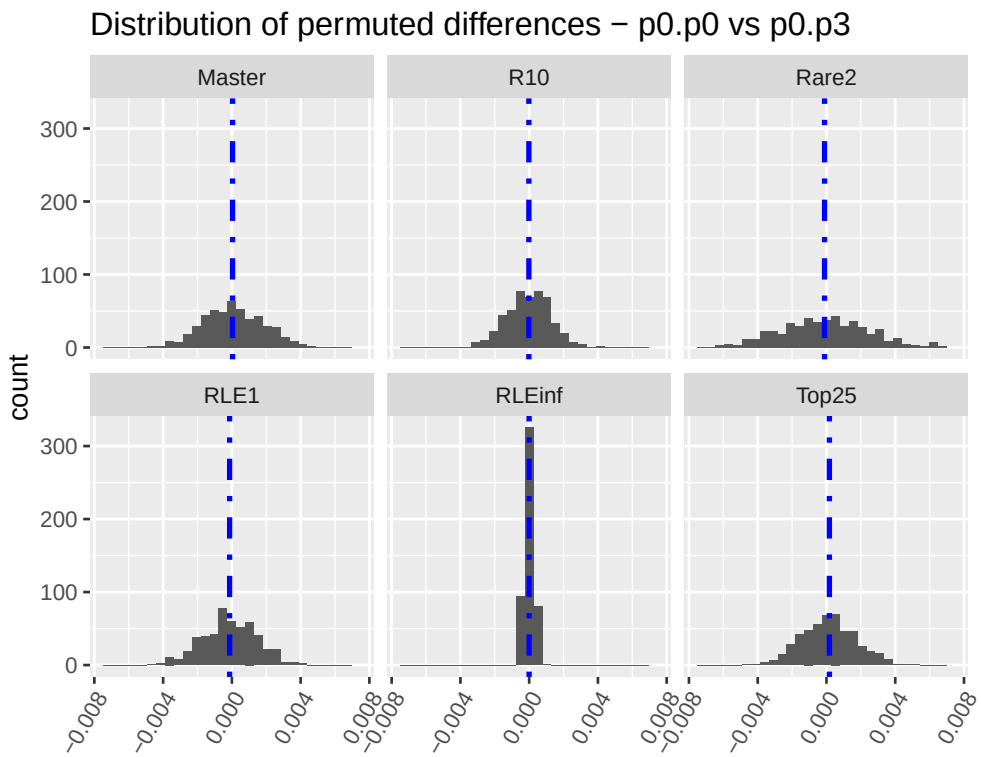
All indicators are statistically significant, with high z absolute values, and adjusted p-values <1.e-6. Master, RLE1 and Top25 show the the higher effect sizes with absolute d values above 1. The hundred passages after p100, resulted in a further reduced Master frequency and Top25 fraction, and increased values for all evenness indicators. Still showing the long term quasispecies dynamics observed in passing from p0 to p100. However the differences in all indicators are mitigated with respect to those observed along the first 100 passages.

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.Med* provides the median 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 calculated as 0. Finally, the permutation *p.value*, and the Benjamini-Hochberg multitest adjusted p-values. Together, these statistics offer a comprehensive assessment of the differences between the two quasispecies and their statistical relevance.

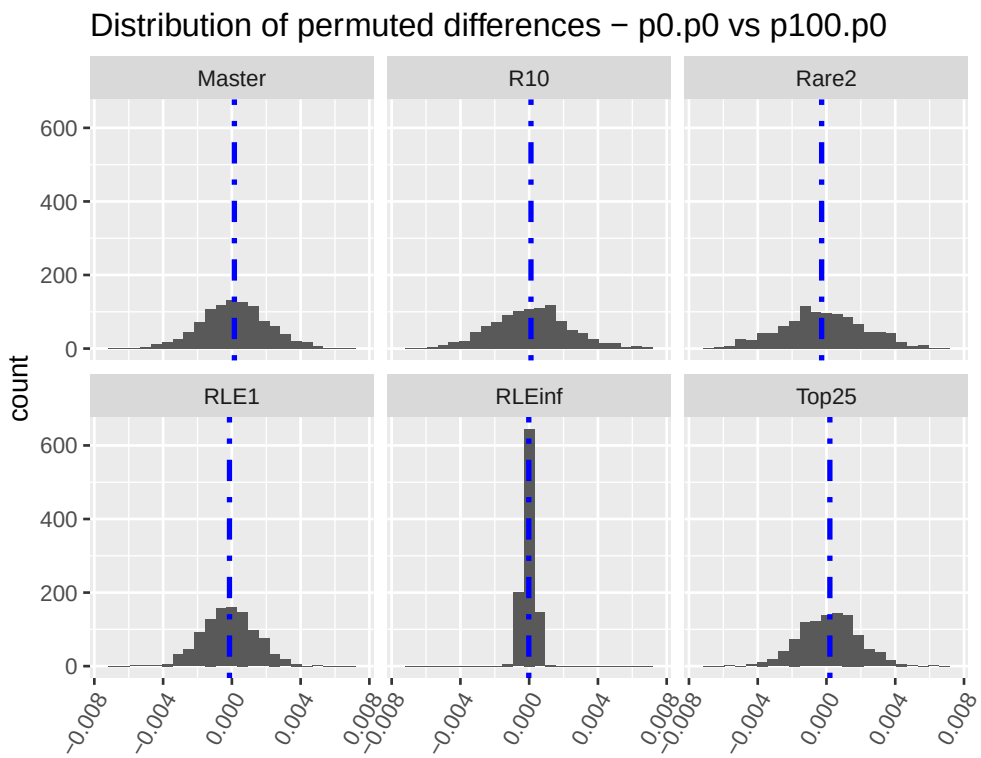
	p0.p0	p0.p3	Obs.Diff	Perm.Diff	d	z	p.value	p.adj.BH
Master	0.72884	0.73089	0.00204	0.00010	0.04947	1.10622	0.254	0.2540
Top25	0.80131	0.78537	-0.01594	0.00014	-0.45899	-10.26344	0.000	0.0000
Rare2	0.19969	0.19014	-0.00955	-0.00013	-0.15928	-3.56163	0.000	0.0000
R10	0.02458	0.02851	0.00393	0.00000	0.13551	3.03009	0.004	0.0060
RLE1	0.27677	0.28121	0.00444	-0.00012	0.13007	2.90844	0.002	0.0040
RLEinf	0.03646	0.03605	-0.00041	-0.00001	-0.06458	-1.44399	0.136	0.1632



Test p0.p0 vs p100.p0

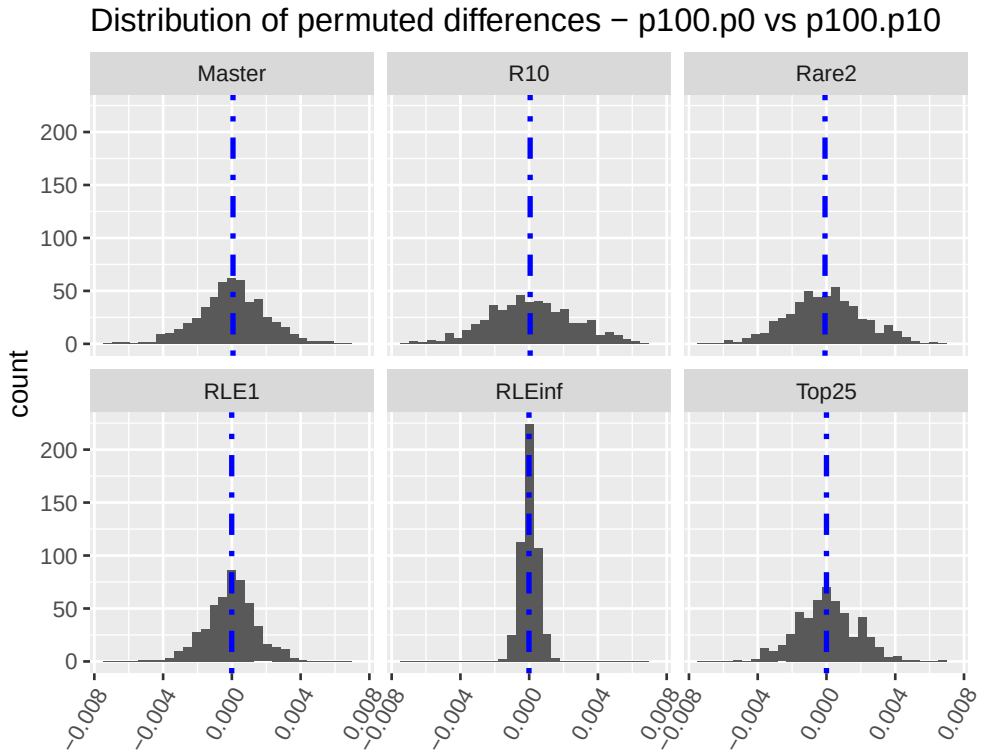
	p0	p100	Obs.Diff	Perm.Diff	d	z	p.value	p.adj.BH
Master	0.72884	0.49213	-0.23671	0.00018	-5.47871	-122.50760	0	0
Top25	0.80131	0.74291	-0.05840	0.00016	-1.57891	-35.30553	0	0
Rare2	0.19969	0.23110	0.03141	-0.00019	0.58329	13.04285	0	0
R10	0.02458	0.06609	0.04151	0.00009	0.80331	17.96260	0	0
RLE1	0.27677	0.39084	0.11406	-0.00016	3.36271	75.19254	0	0
RLEinf	0.03646	0.07783	0.04137	-0.00003	5.28371	118.14731	0	0

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.



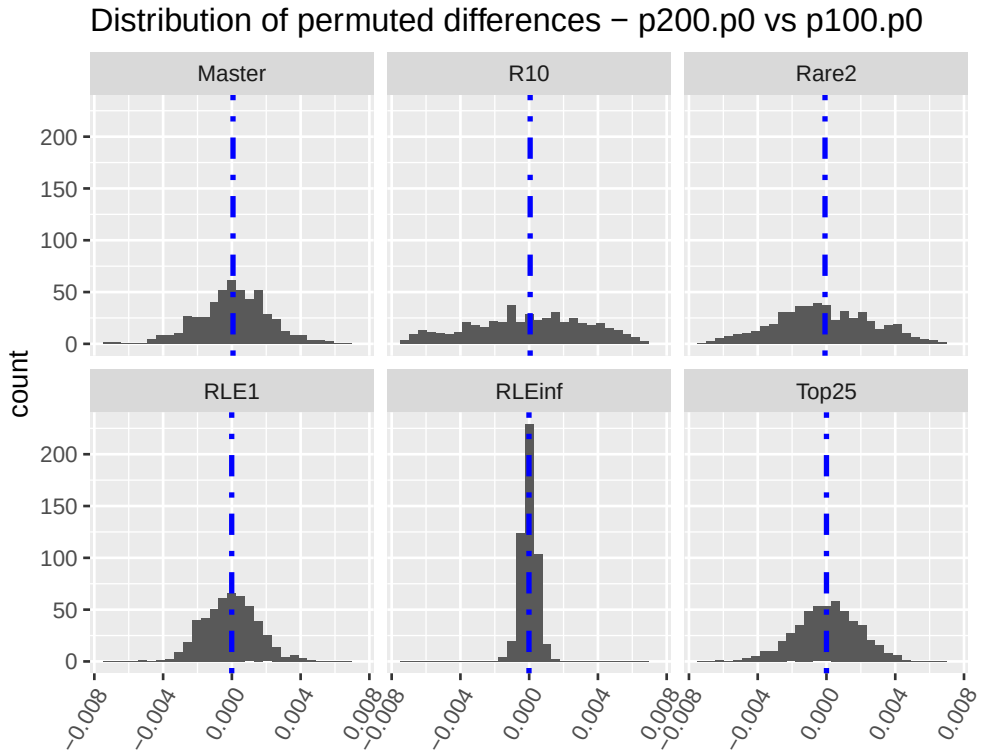
Test p100.p0 vs p100.p10

	p100.p0	p100.p10	Obs.Diff	Perm.Diff	d	z	p.value	p.adj.BH
Master	0.49210	0.39303	-0.09907	0.00002	-2.30879	-51.62607	0	0
Top25	0.74285	0.75431	0.01146	0.00005	0.29271	6.54528	0	0
Rare2	0.23102	0.24154	0.01052	-0.00005	0.22070	4.93501	0	0
R10	0.06631	0.03230	-0.03401	0.00012	-0.61966	-13.85604	0	0
RLE1	0.38849	0.40176	0.01326	-0.00003	0.40897	9.14483	0	0
RLEinf	0.07724	0.10168	0.02444	-0.00001	2.30342	51.50614	0	0



Test p100.p0 vs p200.p0

	p100.p0	p200.p0	Obs.Diff	Perm.Diff	d	z	p.value	p.adj.BH
Master	0.49211	0.43506	-0.05705	0.00012	-1.27492	-28.50810	0	0
Top25	0.74291	0.70577	-0.03715	0.00004	-0.90593	-20.25722	0	0
Rare2	0.23073	0.25735	0.02662	-0.00010	0.41779	9.34212	0	0
R10	0.06619	0.08438	0.01819	0.00008	0.21561	4.82125	0	0
RLE1	0.39095	0.43750	0.04656	-0.00007	1.37338	30.70970	0	0
RLEinf	0.07786	0.08902	0.01116	-0.00003	1.07093	23.94663	0	0



Equivalences between the two tests

Despite being based on different principles, we may compare the results of the two tests, double repeated rarefaction (RRT) test, and permutation test (PT), for each quasispecies structure indicator, by the statistics z and d.

p0.p0 vs p0.p3

	p0.p0	p0.p3	Obs.Diff	Ratio	pt.z	rrt.z	pt.d	rrt.d
Master	0.7288	0.7309	0.0020	1.0028	1.1062	1.2653	0.0495	0.0566
Top25	0.8013	0.7854	-0.0159	0.9801	-10.2634	-9.9762	-0.4590	-0.4461
Rare2	0.1997	0.1901	-0.0095	0.9522	-3.5616	-4.0467	-0.1593	-0.1810
R10	0.0246	0.0285	0.0039	1.1599	3.0301	2.8383	0.1355	0.1269
RLE1	0.2768	0.2812	0.0044	1.0160	2.9084	3.1077	0.1301	0.1390
RLEinf	0.0365	0.0360	-0.0004	0.9888	-1.4440	-0.6058	-0.0646	-0.0271

p0.p0 vs p0.p100

	p0	p100	Obs.Diff	Ratio	pt.z	rrt.z	pt.d	rrt.d
Master	0.7288	0.4921	-0.2367	0.6752	-122.5076	-138.7980	-5.4787	-6.2072
Top25	0.8013	0.7429	-0.0584	0.9271	-35.3055	-36.7160	-1.5789	-1.6420
Rare2	0.1997	0.2311	0.0314	1.1573	13.0429	16.0113	0.5833	0.7160
R10	0.0246	0.0661	0.0415	2.6886	17.9626	21.6179	0.8033	0.9668
RLE1	0.2768	0.3908	0.1141	1.4121	75.1925	77.2321	3.3627	3.4539
RLEinf	0.0365	0.0778	0.0414	2.1348	118.1473	57.3821	5.2837	2.5662

p100.p0 vs p100.p10

	p100.p0	p100.p10	Obs.Diff	Ratio	pt.z	rrt.z	pt.d	rrt.d
Master	0.4921	0.3930	-0.0991	0.7987	-51.6261	-58.2467	-2.3088	-2.6049
Top25	0.7428	0.7543	0.0115	1.0154	6.5453	7.2296	0.2927	0.3233
Rare2	0.2310	0.2415	0.0105	1.0455	4.9350	5.3517	0.2207	0.2393
R10	0.0663	0.0323	-0.0340	0.4871	-13.8560	-16.5599	-0.6197	-0.7406
RLE1	0.3885	0.4018	0.0133	1.0341	9.1448	8.5436	0.4090	0.3821
RLEinf	0.0772	0.1017	0.0244	1.3164	51.5061	29.6907	2.3034	1.3278

p100.p0 vs p200.p0

	p100.p0	p200.p0	Obs.Diff	Ratio	pt.z	rrt.z	pt.d	rrt.d
Master	0.4921	0.4351	-0.0571	0.8841	-28.5081	-32.4881	-1.2749	-1.4529
Top25	0.7429	0.7058	-0.0371	0.9500	-20.2572	-22.3613	-0.9059	-1.0000
Rare2	0.2307	0.2574	0.0266	1.1154	9.3421	13.0432	0.4178	0.5833
R10	0.0662	0.0844	0.0182	1.2748	4.8213	8.7837	0.2156	0.3928
RLE1	0.3909	0.4375	0.0466	1.1191	30.7097	31.4648	1.3734	1.4072
RLEinf	0.0779	0.0890	0.0112	1.1433	23.9466	13.7149	1.0709	0.6133

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

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

Due to this difference in rarefaction approach, we anticipate that the permutation test will yield less biased and more precise results compared to the repeated rarefaction method.

Remarkably, both tests produced similar values for z and Cohen's d. The most substantial differences were expected in indicators more sensitive to sample size variations, particularly the RLEinf (Relative Logarithmic Evenness at q=infinity), defined as $RLE_{\text{inf}} = \log_{10}(1/\text{master})/\log_{10}(H)$ Here 'master', the frequency of the most abundant haplotype, and H the number of haplotypes.

The high similarity in results between the two tests empirically validates the proposed permutation test scheme. Given its potential for reduced bias, we recommend the preferential use of the permutation test for comparing quasispecies diversities. The repeated rarefaction method, while yielding similar results, may be redundant in this context.

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
- 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. A the importance of an difference of 1% or 10%, or a ratio of 1.1 or 0.75 depend of the context. Cohen's d is an effect size measure that represents the observed difference in units of the pooled standard deviation of the two quasispecies. It provides a standardized measure of the magnitude of the effect. 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 using Cohen's d.
3. Consider the statistical significance using the z-statistic and adjusted p-values.
4. Contextualize these findings within the biological or clinical framework of the study.

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 methods:

- Repeated rarefaction + z-test
- Permutation test

The high similarity in results between the two methodss empirically validates the proposed permutation test scheme. Given its potential for reduced bias, we recommend the preferential use of the permutation test for comparing quasispecies diversities. The repeated rarefaction method, while yielding similar results, may be redundant in this context.

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.

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 (Perplexity AI 2024).

References

Domingo, E., M. E. Soria, I. Gallego, A. I. De Ávila, C. García-Crespo, B. Martínez-González, J. Gómez, and et al. 2020. "A New Implication of Quasispecies Dynamics: Broad Virus Diversification in Absence of External Perturbations." *Infection, Genetics and Evolution* 82: 104278. https://doi.org/10.1016/j.meegid.2020.104278.

Efron, B., and C. Stein. 1981. "The Jackknife Estimate of Variance." *The Annals of Statistics* 9 (3): 586–96. https://doi.org/10.1214/aos/1176345462.

Efron, B., and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap.* 1st ed. Chapman; Hall/CRC. https://doi.org/10.1201/9780429246593.

Ellis, P. D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results.* Cambridge University Press.

Gallego, I., M. E. Soria, C. García-Crespo, Q. Chen, P. Martínez-Barragán, S. Khalfaoui, B. Martínez-González, et al. 2020. "Broad and Dynamic Diversification of Infectious Hepatitis c Virus in a Cell Culture Environment." *J Virol* 94 (6). https://doi.org/10.1128/JVI.01856-19.

Gregori, J., S. Colomer-Castell, M. Ibañez-Lligoña, D. Garcia-Cehic, C. Campos, M. Buti, M. Riveiro-Barciela, et al. 2024. "In-Host Flat-Like Quasispecies: Characterization Methods and Clinical Implications." *Microorganisms* 12 (5): 1011. https://doi.org/10.3390/microorganisms12051011.

Gregori, J., M. Ibañez-Lligoña, S. Colomer-Castell, C. Campos, and J. Quer. 2024. "Virus Quasispecies Rarefaction: Subsampling with or Without Replacement?" *Viruses* 16 (5): 710. https://doi.org/10.3390/v16050710.

Gregori, J., M. Ibañez-Lligoña, S. Colomer-Castell, D. Garcia-Cehic, C. Campos, and J. Quer. 2024. "Association of Liver Damage and Quasispecies Maturity in Chronic HCV Patients: The Fate of a Quasispecies." *Microorganisms* submitted (to appear).

Gregori, J., M. E. Soria, I. Gallego, M. Guerrero-Murillo, J. I. Esteban, J. Quer, C. Perales, and E. Domingo. 2018. "Rare Haplotype Load as Marker for Lethal Mutagenesis." *PLoS One* 13 (10): e0204877. https://doi.org/10.1371/journal.pone.0204877.

Hesterberg, T., D. S. Moore, S. Monaghan, A. Clipson, R. Epstein, and McCabe G. P. 2005. "Bootstrap Methods and Permutation Tests." In *Introduction to the Practice of Statistics*, edited by W. H. McCabe, 5th ed., 14.1–70. New York, USA: Freeman & Co.

Holt, C. A., and S. Sullivan. 2023. "Permutation Tests for Experimental Data." 2022-31. U Iowa Legal Studies Research Paper, Experimental Economics. https://doi.org/10.2139/ssrn.3957609.

Kelley, K., and K. J. Preacher. 2012. "On Effect Size." *Psychol Methods* 17 (2). https://doi.org/10.1037/a0028086.

Kim, H. Y. 2015. "Statistical Notes for Clinical Researchers: Effect Size." *Restor Dent Endod.* 40 (4): 328–31. https://doi.org/10.5395/rde.2015.40.4.328.

Moore, Jason H. 1999. "Bootstrapping, Permutation Testing and the Method of Surrogate Data." *Physics in Medicine & Biology* 44 (6): L11. https://doi.org/10.1088/0031-9155/44/6/101.

Nakagawa, S., and I. C. Cuthill. 2007. "Effect Size, Confidence Interval and Statistical Significance: A Practical Guide for Biologists." *Biol Rev Camb Philos Soc* 82 (4): 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x.

Perplexity AI. 2024. "Perplexity AI." https://www.perplexity.ai.

Ramachandran, K. M., and C. P. Tsokos. 2021. "Chapter 13 - Empirical Methods." In

Mathematical Statistics with Applications in r (Third Edition), edited by K. M. Ramachandran and C. P. Tsokos, Third Edition, 531–68. Academic Press. https://doi.org/10.1016/B978-0-12-817815-7.00013-0.

Sawilowsky, S. 2009. "New Effect Size Rules of Thumb." *Journal of Modern Applied Statistical Methods*, no. 2. https://doi.org/10.22237/jmasm/1257035100.

Schloss, P. D. 2024. "Rarefaction Is Currently the Best Approach to Control for Uneven Sequencing Effort in Amplicon Sequence Analyses." *MSphere* 9 (2). https://doi.org/10.1128/msphere.00354-23.