## How are the relative abundances of the 8 species related to one another?

We investigate the relation between the relative abundances of the 8 species with the help of a correlation measure. Our search was limited by three correlation measures; Pearson, Spearman and Kendall. All three measures return the magnitude and direction of the association, however Pearson indicates a linear relation, while the other two indicate a monotonic relation as they are both ranked tests.

We decided on using the Spearman rank correlation to test the relations as this is less sensitive to outliers and because the relative abundance distributions of the species are not the same (Figure 1).

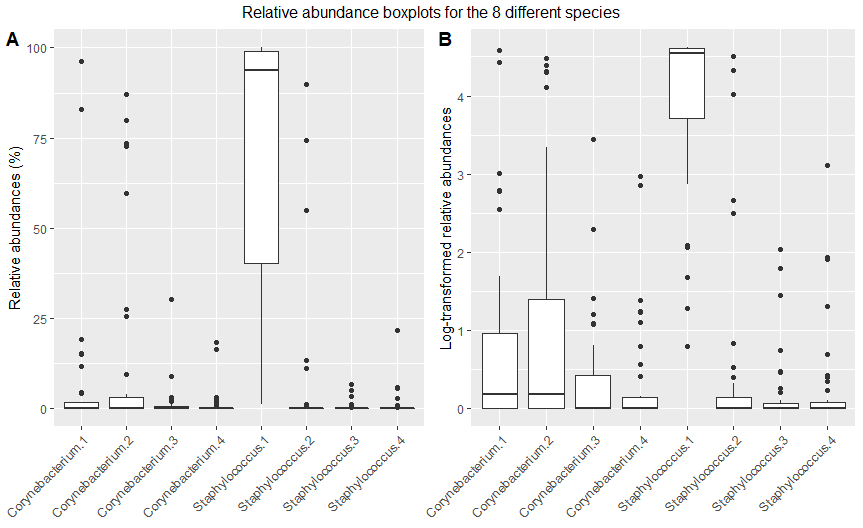


Figure 1: boxplots….

The null hypothesis states that the relative abundances are not related to each other and thus have a ρ value of 0. To be able to reject the null hypothesis we need to be able to test the significance of the ρ value.

This can be done with the help of a permutation test. For each combination of species a permutation distribution is made. This is done by permutating their relative abundances and calculating the Spearman correlation. As an example; to create a permutation distribution for Corynebacterium 1 and Corynebacterium 2 we permute the Corynebacterium 1 sequence and the Corynebacterium 2 sequence, then ρ is calculated. Then another permutation is made of each sequence and ρ calculated. If the sequence length is small, it might be possible to calculate ρ for each possible combination, however in our case this is not possible.

The library “Psych” was instead used, this calculates a p-value via the asymptotic t approximation with the formula.

To show that this test performs equivalent to a permutation test with n = 100 000 (higher than n in the figures, due to issues with loading an image with n = 100 000), we performed a permutation test for the correlations between [Corynebacterium 1, Staphylococcus 1] and [Staphylococcus 3, Staphylococcus 4], these were chosen as in the former their distributions are not alike, while in the latter case they are more similar. From the results we can see that both permutation distributions are normally distributed, aside from the permutation distribution between Staph 3 and 4 having a small skew. The p-values are in the same range for both tests, for Cor 1 vs Staph 1, the t-test returned a p-value of 0.000334 and 0.00048 for the permutation. In the second comparison the results are 0.06973 and 0.0701 respectively. The asymptotic t approximation was used as this is a quicker method and the package allows for easy creation of confidence intervals and adjusting for multiple testing.

Afbeelding met tekst, kaart

Automatisch gegenereerde beschrijving

Afbeelding met tekst, kaart

Automatisch gegenereerde beschrijving

To adjust for the multiple testing problem, we use the Benjamini-Yekutieli (BY) method, this is a less conservative correction than Bonferroni or Holms-Bonferroni (HB), which gives us more statistical power. However in our case, the amount of significant pairs does not change whether BY or HB was used.

The results of the correlation tests and complementary confidence intervals can be seen in Figure X.

Have to talk about result figures

Afbeelding met schermafbeelding

Automatisch gegenereerde beschrijving

Afbeelding met schermafbeelding

Automatisch gegenereerde beschrijving