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

When using relative abundances there is a negative relation between two splits of the abundances, for example if the abundance of species 1 goes up then the other abundances will go down. However we want to investigate how the relative abundances of the 8 species are related to one another, this will be done 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).

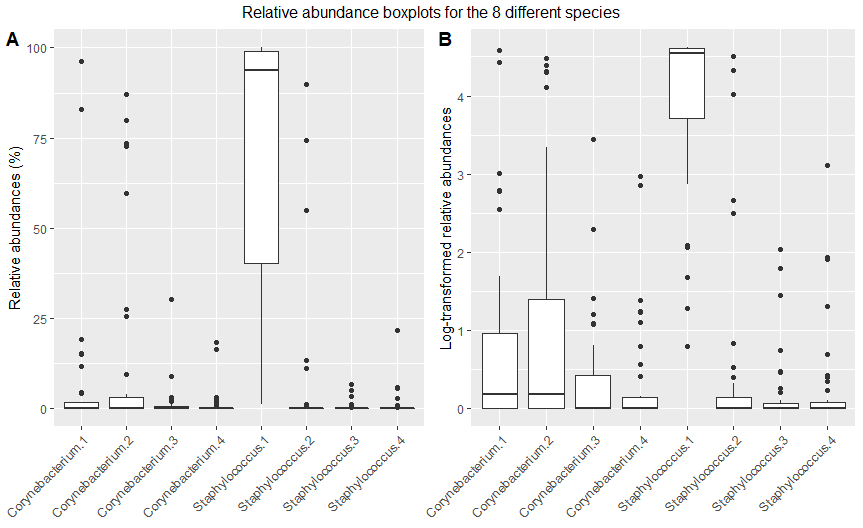


Figure 1: boxplots of the relative abundances (in % for A, Log-transformed for B) for the 8 different species.

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 as we have 40 samples per species.

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 using 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 (Figures 3 and 4). 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

Figure 2: Histogram (A) and QQ-plot (B) showing the permutation distribution with n = 10 000, between Corynebacterium 1 and Staphylococcus 1

Afbeelding met tekst, kaart

Automatisch gegenereerde beschrijving

Figure 3: Histogram (A) and QQ-plot (B) showing the permutation distribution with n = 10 000, between Staphylococcus 3 and Staphylococcus 4

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 heatmaps for show that overall the correlation coefficients are small, the six significant pairs in the non-adjusted heatmap are [Cor 1 with Cor 2, Cor 3], [Cor 2 with Cor 3], [Staph 1 with Cor 1, Cor 2, Cor 3]. Adjusting for multiple testing leaves three significant pairs [Cor 1 with Cor 2] and [Staph 1 with Cor 1, Cor 2] (Figure 4). The confidence intervals however are very wide, therefore we cannot conclude anything about the magnitude of the association, only about the direction (Figure 5). From the significant pairs, Staph 1 is negatively correlated with Cor 1 and 2, while Cor 1 and 2 are positively associated with each other. A possible cause for this is if Staph 1 grows in the same microbial environment as Cor 1 and 2 and are therefore competitors. Cor 1 and 2 are positively associated, this could be because of either a symbiotic relation or because the environment in which they thrive is slightly different from one another, which allows them to grow together and not compete.

Afbeelding met schermafbeelding

Automatisch gegenereerde beschrijving

Figure 4: Heatmaps showing the Spearman’s ρ in colour with it’s p-value as a number. A: p-values not adjusted B:p-values adjusted with the Benjamini-Yekutieli method

Afbeelding met schermafbeelding

Automatisch gegenereerde beschrijving

Figure 5: Confidence intervals Spearman’s ρ for each of the species pairs, adjusted with the BY method. Significant confidence intervals in red