Modelling of Environmental Processes

Analysis of environmental data in the Past program

Gabriela Malec
Geospatial Computer Science

1. The aim of the project

The aim of the project was to analyze environmental data covering groups such as bacteria, plants and eukaryotes to find out whether there are differences in the structure of the groups mentioned and which groups vary the most. Another goal was to determine the taxons that make the communities different the most. Ultimately, it was also important to conduct analyses leading to the question of whether the chemical properties of soils differ among the three types of ecosystems. The entire analysis was performed in the PAST program.

2. Data preparation

The next step was data preparation, it required appropriate ordering of the data by creating sets of distinct groups – Birch, Pine and Tundra. As a result, it was possible to compare their relationships and check the similarities.

3. Analysis process

For all three groups individually, a set of functions was performed. First, a non-metric multidimensional scaling function was used with the Euclidean similarity index option selected. Novex hulls were additionally marked on the scatter plots to better illustrate the similarity and differences between the groups. Next, One-way ANOSIM analysis was performed, as well as one-way PERMANOVA. Also, the functoriality of SIMPER analysis was used. Then, for the chemical dataset, Principal Components Analysis was carried out.
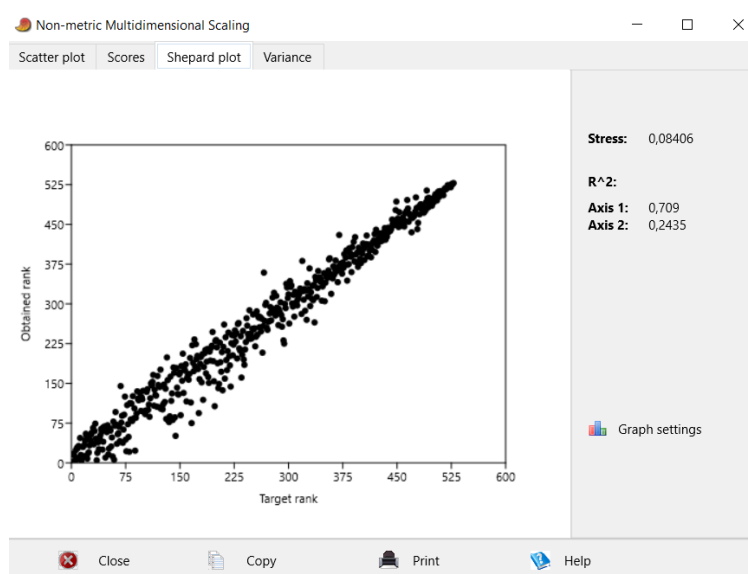
3.1. Bacteria
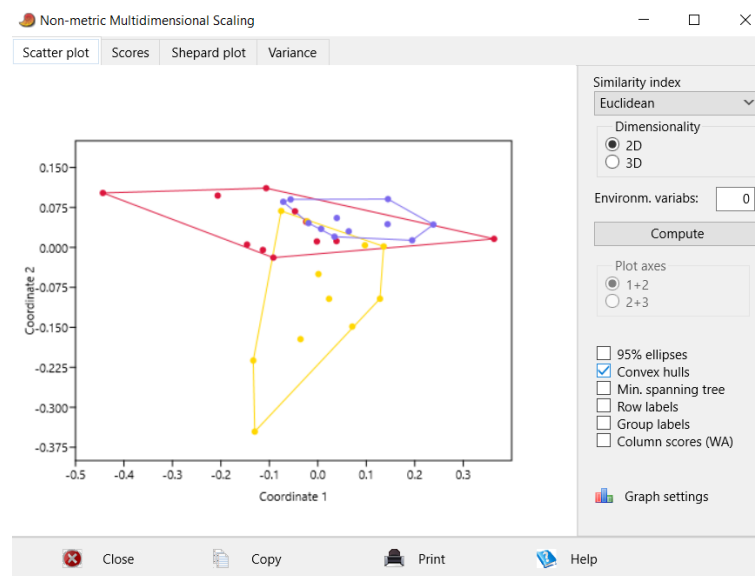
Pic. 1 NMDS – Bacteria – Scatter plot

Analyzing the graph, it can be seen that the Birch group, extends across the entire width of the graph. This suggests that there is considerable variability in the structure or composition of the bacteria in this group, which may indicate a diversity of species or subgroups. The second observation concerns the Tundra group, which is almost entirely contained within the Birch

group area. This may indicate that there is some degree of similarity in the composition or function of these bacteria, which are characteristic of both environments. When analyzing the Pine group, it extends in height and does not overlap much with the other groups. This may suggest that the bacteria in the Pine group are different or unique compared to those in the Birch and Tundra groups. This difference may be due to specific environmental factors, such as soil type, the presence of certain chemicals.

Pic. 2 NMDS – Bacteria – Shepard plot

Shepard Plot analysis provides information on the degree of match between the actual distances between points and the distances reconstructed by NMDS. In this case, the stress value is 0.08406, indicating a good match.

Below is also a table with the variances for the three groups analyzed.

| Birch	Pine	Tundra
0,083166	0,091248	0,13441
0,062153	0,018611	0,17328
0,37779	0,13187	0,078854
0,074883	0,12286	0,14128
0,047295	0,18442	0,089064
0,0634	0,14214	0,026681
0,14714	0,15989	0,084647
0,079103	0,28868	0,043212
0,035982	0,040919	0,020077
0,11303	0,1786	0,060648
0,43522	0,087272	0,12788

Pic. 3 NMDS – Bacteria – Variance

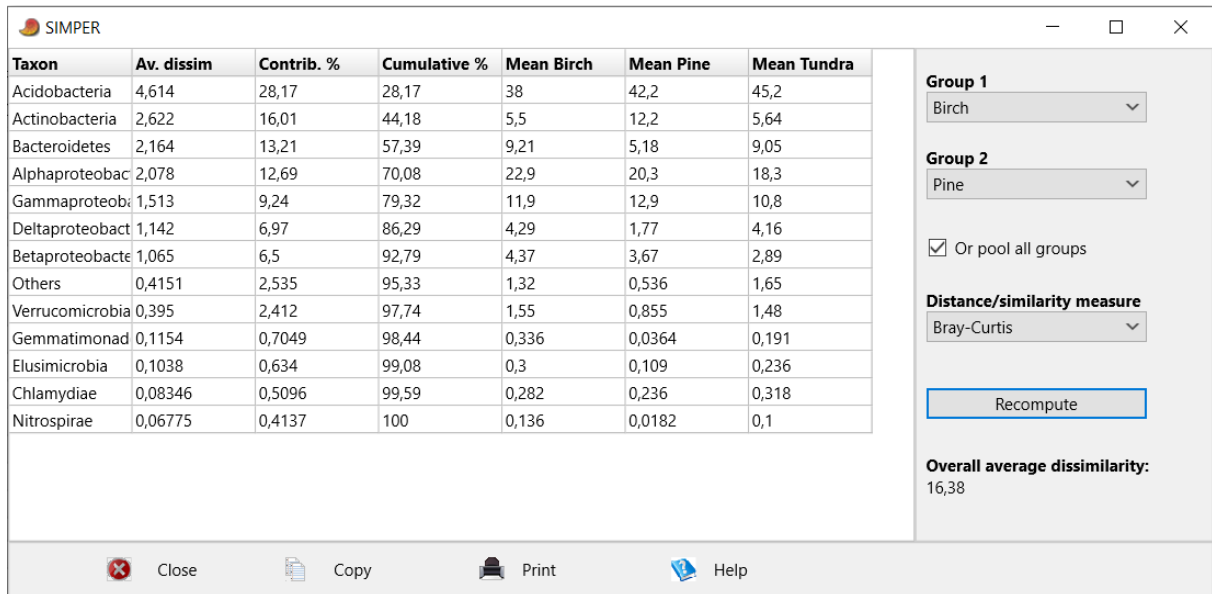The next step was one-way ANOSIM and one-way PERMANOVA analysis.
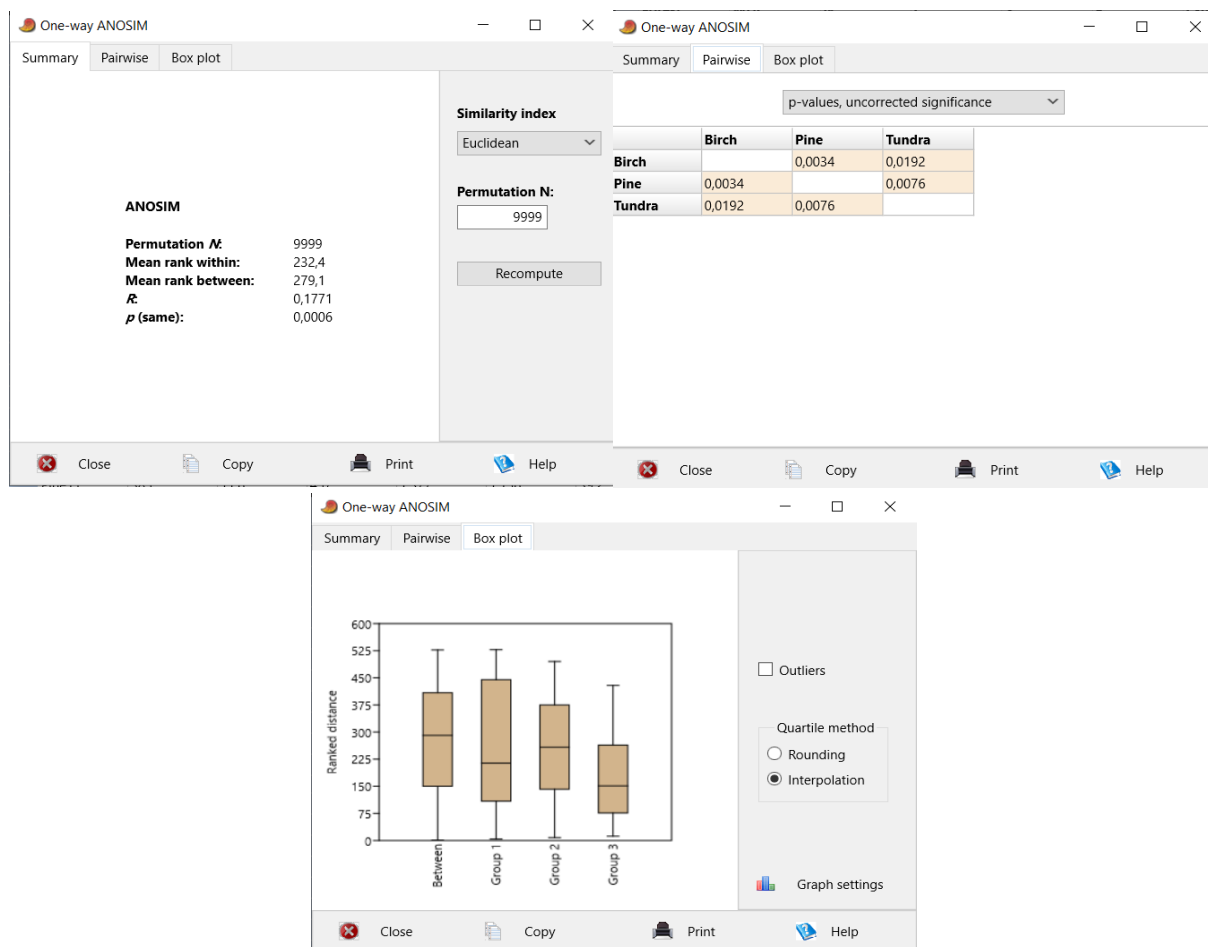
Pic. 4 Results of one way ANOSIM – Bacteria

The results indicate that the values of mean rank within equal to 232.4 and mean rank between equal to 279.1 show that the average similarity ranking within a group is lower than between groups. This means that within each group, scores are more similar to each other than to scores from other groups. In this case, an R-value of 0.1771 indicates some statistically significant differences between groups. A p-value of 0.0006 means that there is a very small chance that the observed differences between groups are due to chance. We can consider these differences statistically significant.
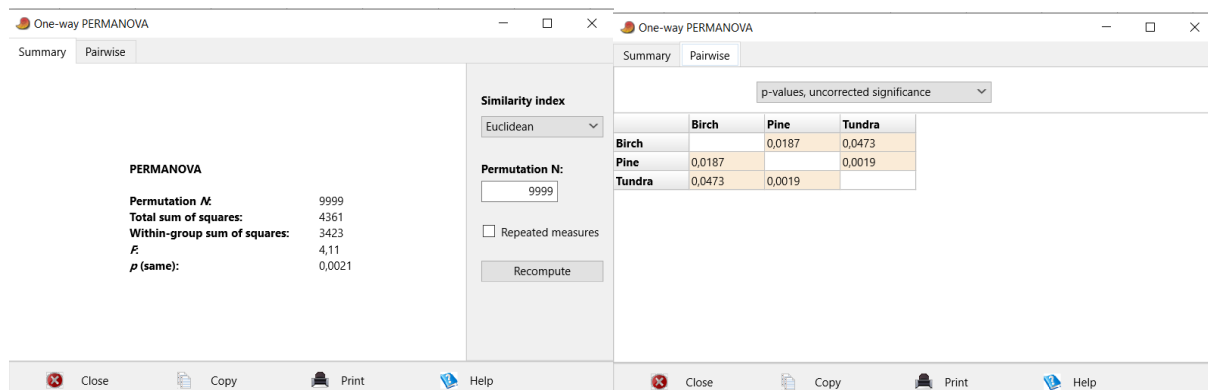
*Pic. 5 Results of one way PERMANOVA – Bacteria*

It can be concluded, the results of One-Way PERMANOVA analysis confirmed, statistically significant differences between groups of bacteria on the basis of euclidean similarity. There is a significant difference between groups, and this difference is greater than the within-group variability. This analysis further confirms that the differences between the bacterial groups are statistically significant.

Taxon	Av. dissim	Contrib. %	Cumulative %	Mean Birch	Mean Pine	Mean Tundra
Acidobacteria	4,614	28,17	28,17	38	42,2	45,2
Actinobacteria	2,622	16,01	44,18	5,5	12,2	5,64
Bacteroidetes	2,164	13,21	57,39	9,21	5,18	9,05
Alphaproteobac	2,078	12,69	70,08	22,9	20,3	18,3
Gammaproteoba	1,513	9,24	79,32	11,9	12,9	10,8
Deltaproteobact	1,142	6,97	86,29	4,29	1,77	4,16
Betaproteobacte	1,065	6,5	92,79	4,37	3,67	2,89
Others	0,4151	2,535	95,33	1,32	0,536	1,65
Verrucomicrobia	0,395	2,412	97,74	1,55	0,855	1,48
Gemmatimonad	0,1154	0,7049	98,44	0,336	0,0364	0,191
Elusimicrobia	0,1038	0,634	99,08	0,3	0,109	0,236
Chlamydiae	0,08346	0,5096	99,59	0,282	0,236	0,318
Nitrospirae	0,06775	0,4137	100	0,136	0,0182	0,1

Group 1
Birch

Group 2
Pine

☑ Or pool all groups

Distance/similarity measure
Bray-Curtis

Recompute

Overall average dissimilarity:
16,38

Close   Copy   Print   Help

Pic. 6 Simper - Bacteria

SIMPER analysis provides information on the specific bacterial species that contribute most to the differences between groups. The largest values for the percentage contribution of individual components to the total variation between groups were obtained by Acidobacteria. Overall average dissimilarity is 16,38. This value indicates that, on average, the groups differ from each other by 16,38%.
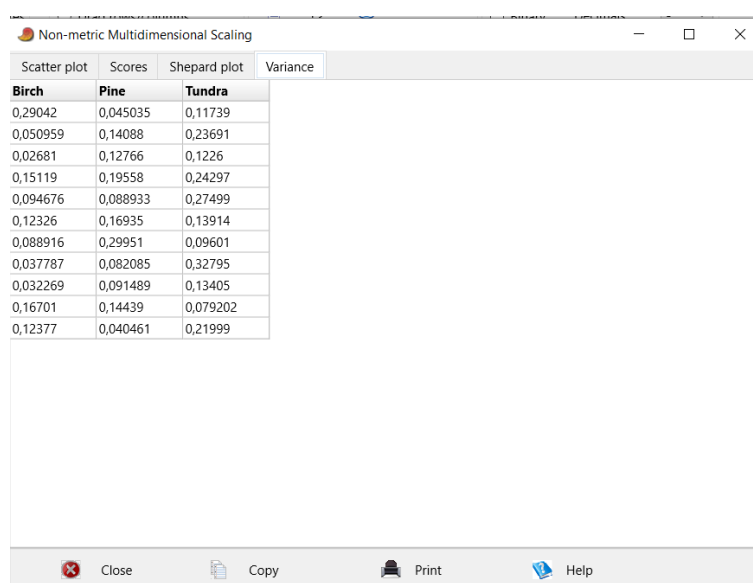
3.2. Eukaryotes
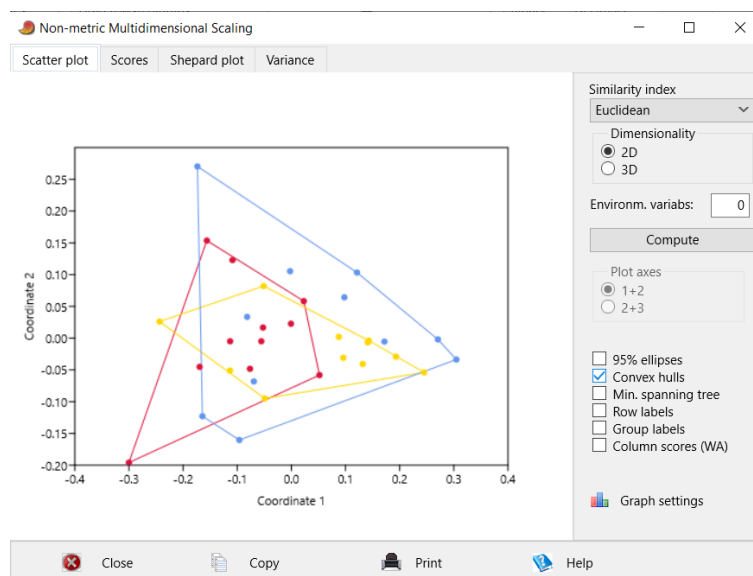
Pic. 7 NMDS – Eukaryotes – Scatter plot

The graph shows that the Tundra group is the most extensive, suggesting that the samples of this group differ from those of the other groups in terms of eukaryote composition. However, the convex hulls for all three groups overlap quite a bit, perhaps indicating similarity in eukaryote composition between the groups.
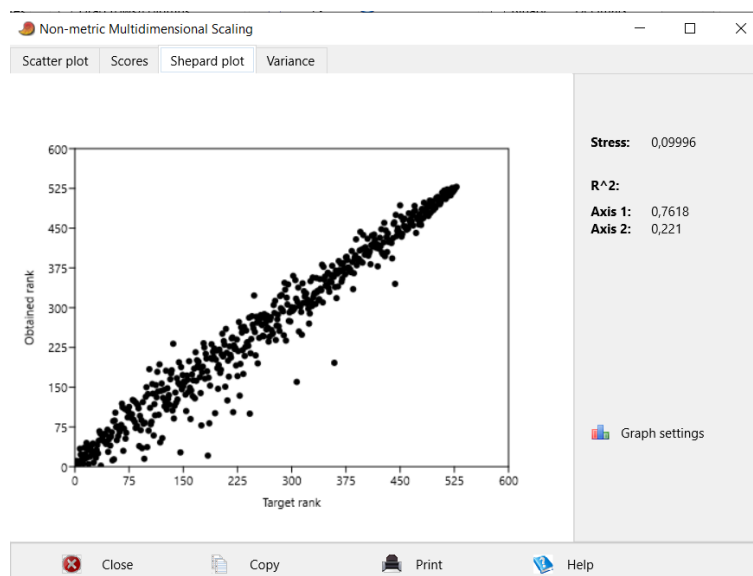

Pic. 8 NMDS – Eukaryotes – Shepard plot

Examining the Shepard Plot for the eukaryote data, based on stress and axis 1 and 2 values, suggests that the spatial representation of the data is good, but there is a certain amount of nonlinearity or noise. The points are distributed on the NMDS plot in a way that reflects their similarity in reality, although with some degree of deviation. Below again is a table with variance values.


Pic. 9 NMDS – Eukaryotes – Variance

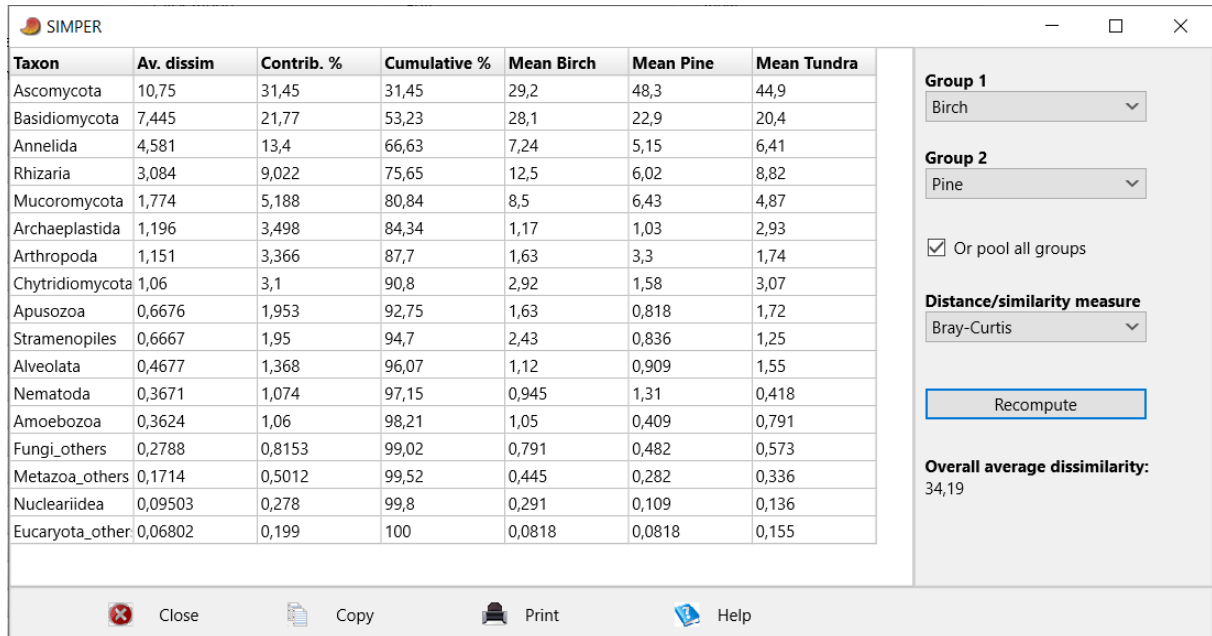The next step was one-way ANOSIM and one-way PERMANOVA analysis.
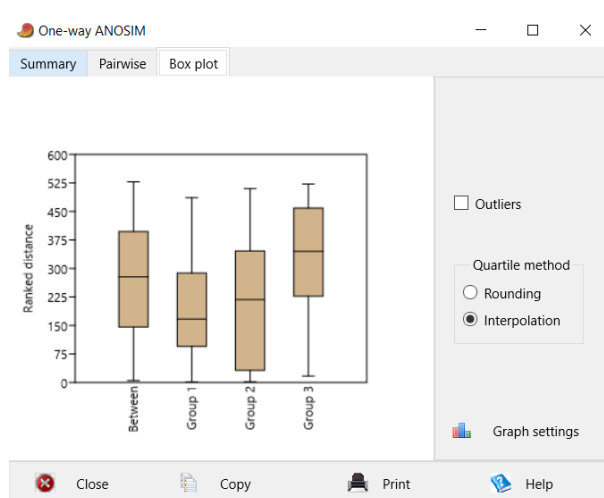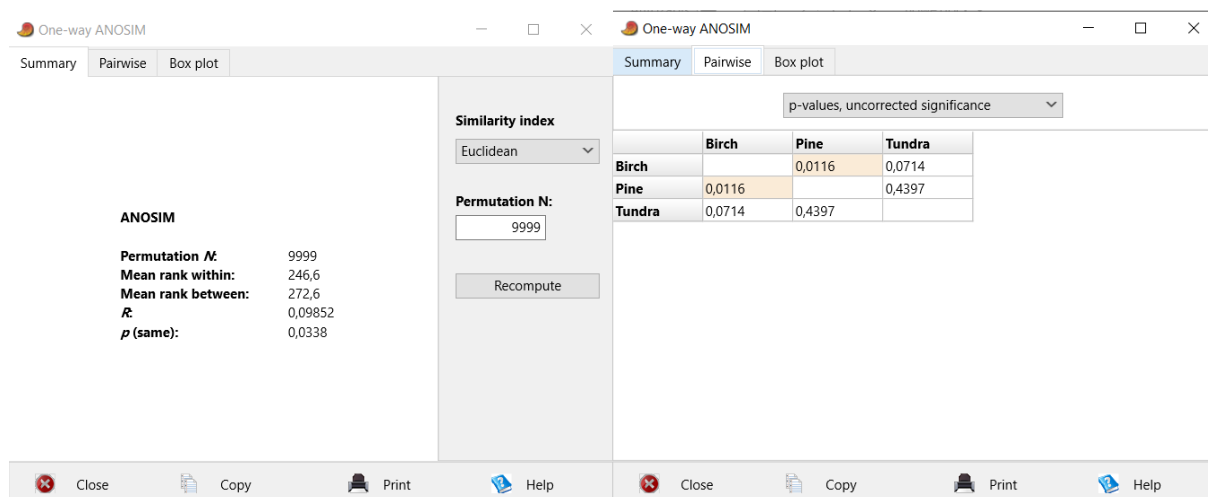
Pic. 10 Results of one way ANOSIM – Eukaryotes

Mean rank within refers to the average rank of samples within a group. In this case, it is 246.6, indicating that lower ranks were assigned to samples within the same group, suggesting similarity between samples within the group. In turn, mean rank between refers to the average rank of samples between groups. Here it is 272.6, suggesting that samples between groups have different ranks, indicating differences between groups. However, the R value is 0.09852, indicating that they are not large. The p-value indicates that the values are statistically significant.
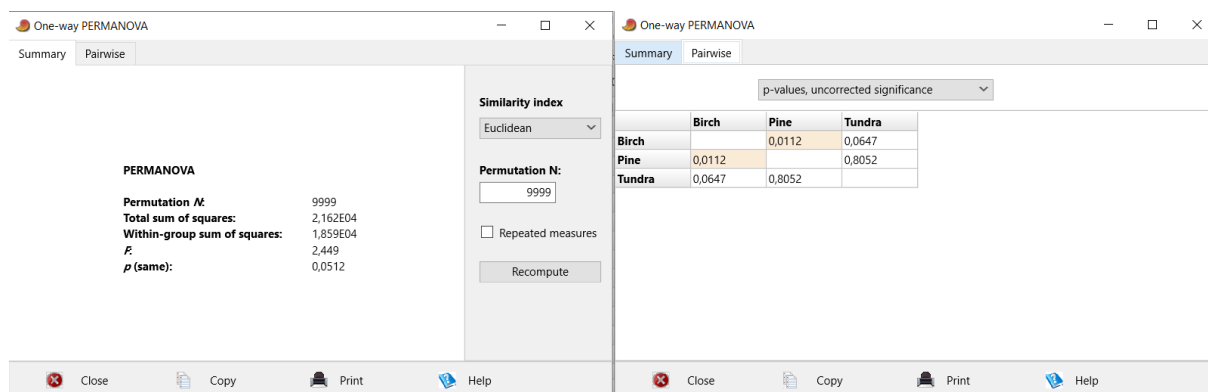
*Pic. 11 Results of one way PERMANOVA – Eukaryotes*

Based on the results of the one-way PERMANOVA analysis, we can confirm that there are statistically significant differences between groups of eukaryotes and that samples from the same group are statistically significantly more similar than samples from different groups.

Taxon	Av. dissim	Contrib. %	Cumulative %	Mean Birch	Mean Pine	Mean Tundra
Ascomycota	10,75	31,45	31,45	29,2	48,3	44,9
Basidiomycota	7,445	21,77	53,23	28,1	22,9	20,4
Annelida	4,581	13,4	66,63	7,24	5,15	6,41
Rhizaria	3,084	9,022	75,65	12,5	6,02	8,82
Mucoromycota	1,774	5,188	80,84	8,5	6,43	4,87
Archaeplastida	1,196	3,498	84,34	1,17	1,03	2,93
Arthropoda	1,151	3,366	87,7	1,63	3,3	1,74
Chytridiomycota	1,06	3,1	90,8	2,92	1,58	3,07
Apusozoa	0,6676	1,953	92,75	1,63	0,818	1,72
Stramenopiles	0,6667	1,95	94,7	2,43	0,836	1,25
Alveolata	0,4677	1,368	96,07	1,12	0,909	1,55
Nematoda	0,3671	1,074	97,15	0,945	1,31	0,418
Amoebozoa	0,3624	1,06	98,21	1,05	0,409	0,791
Fungi_others	0,2788	0,8153	99,02	0,791	0,482	0,573
Metazoa_others	0,1714	0,5012	99,52	0,445	0,282	0,336
Nucleariidea	0,09503	0,278	99,8	0,291	0,109	0,136
Eucaryota_other	0,06802	0,199	100	0,0818	0,0818	0,155

Group 1
Birch

Group 2
Pine

☑ Or pool all groups

Distance/similarity measure
Bray-Curtis

Recompute

Overall average dissimilarity:
34,19

Pic. 12 Simper - Eukaryotes

The largest values for the percentage contribution of individual components to the total variation between groups were obtained by Ascomycota. Overall average dissimilarity is 34.19. This value indicates that, on average, the groups differ from each other by 34.19%.

3.3. Plants
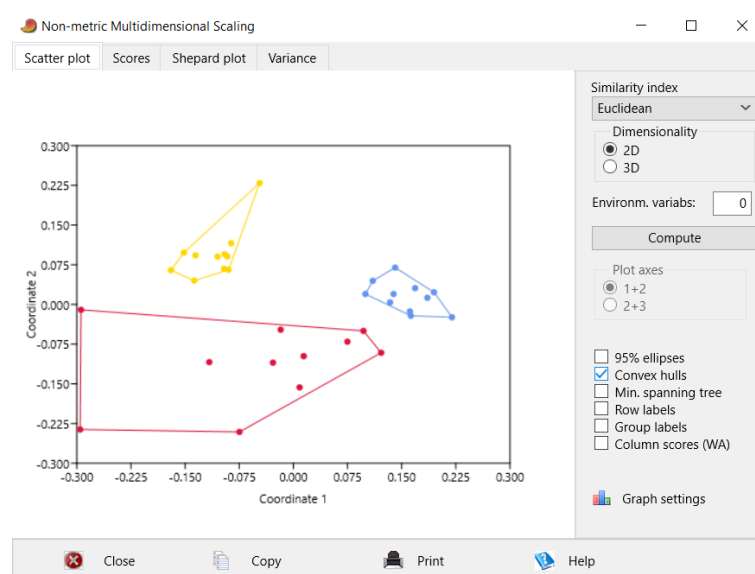
Pic. 13 NMDS – Plants – Scatter plot

Convex Hulls for the Plants group do not intersect at any point and are relatively distant from each other, this means that the groups are relatively separate and different from each other in terms of plant composition. The Birch group occupies the largest area, this suggests that the it has the greatest diversity and can be distinguished on the basis of plant composition.
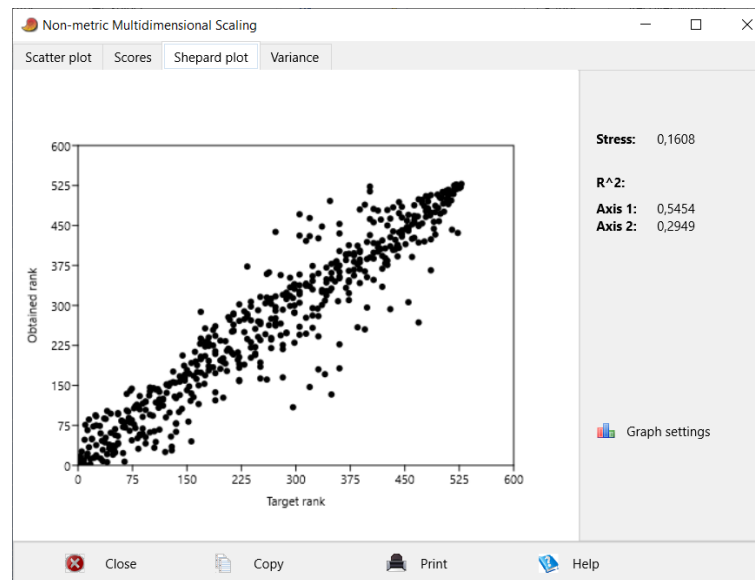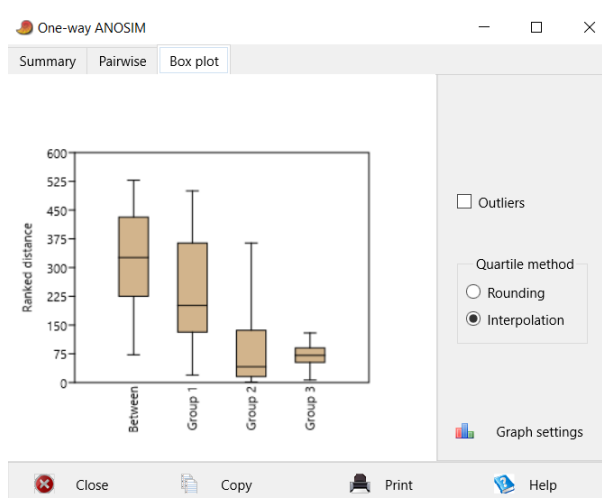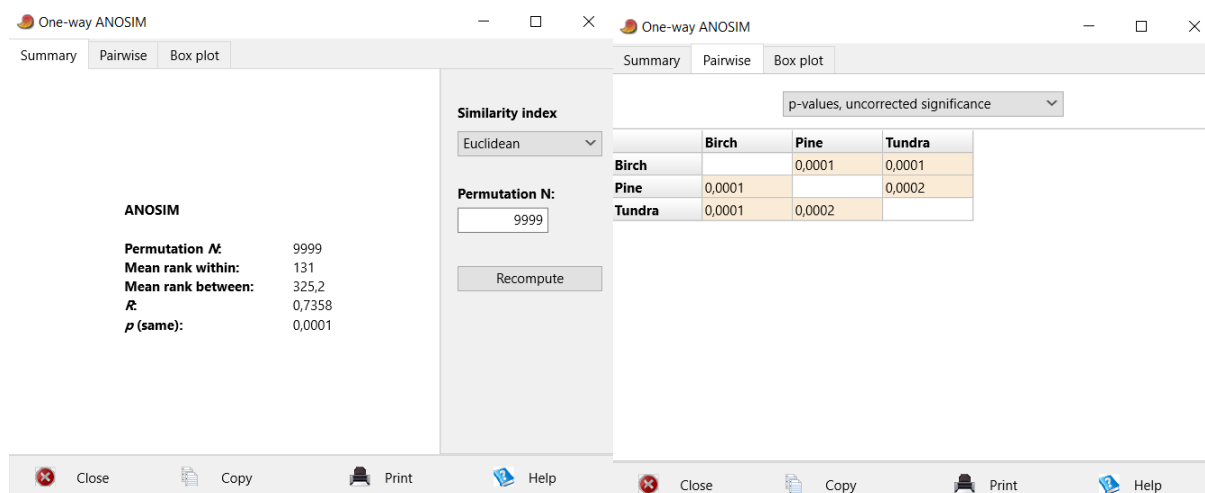

Pic. 14 NMDS – Plants – Shepard plot

The lower the stress value, the better the fit. In this case, the stress value is 0.1608, indicating a moderate distance fit on the graph. Although, there is a certain amount of nonlinearity or noise. In the table below it is possible to compare variance values.

Pic. 15 NMDS – Plants – Variance

Pic. 16 Results of one way ANOSIM – Plants

The mean rank within a group is 131. This indicates that samples within the same group are similar to each other. The mean rank between groups is 325.2. This indicates that there are differences between groups in terms of plant composition. The R-value is 0.7358. This means that there is a high degree of similarity between samples within the group. Based on these results, we can conclude that there is a statistically significant difference in plant composition between groups in the analyzed group.
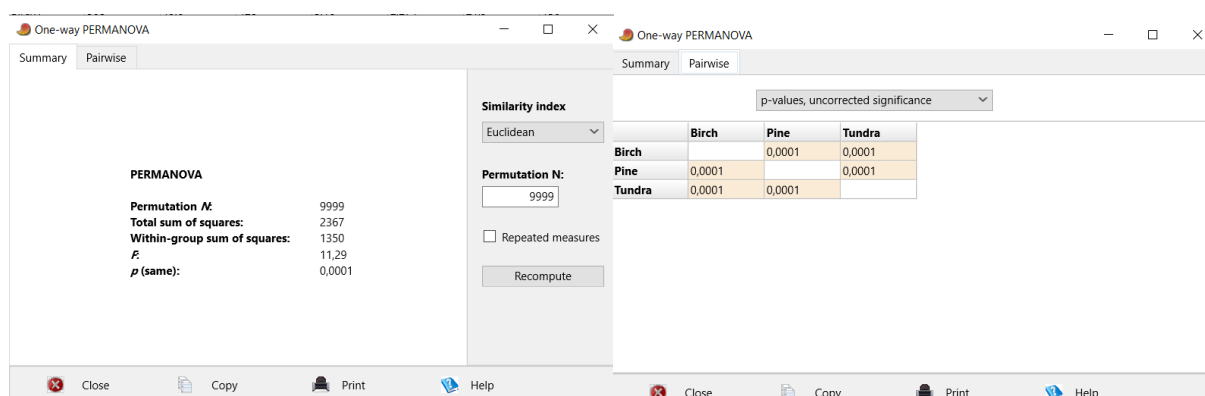
Pic. 17 Results of one way PERMANOVA – Plants

Again, through PERMANOVA analysis, it was confirmed that there is a statistically significant difference in plant composition between groups. The within-group variation is smaller than the between-group variation.

Taxon	Av. dissim	Contrib. %	Cumulative %	Mean Birch	Mean Pine	Mean Tundra
Empetrum nigru	5,707	6,808	6,808	2,18	0	4,91
Betula nana	5,292	6,313	13,12	0	0,364	4,73
Deschampsia ca	4,941	5,895	19,02	0,455	4	0
Salix lapponum	4,633	5,527	24,54	0,273	3,82	1
Cornus suecica	4,541	5,417	29,96	3,36	2,18	0,545
Deschampsia fle	4,274	5,099	35,06	2,64	0	3
Betula pubescen	4,063	4,847	39,91	3,73	0,182	0,364
Vaccinium vitis-i	3,855	4,599	44,5	2,09	0	2,64
Melampyrum sy	3,592	4,285	48,79	0,909	2,82	0
Rubus arcticus	3,133	3,737	52,53	0,909	2,45	0
Vaccinium myrtil	3,12	3,722	56,25	1,73	0	1,82
Vaccinium uligin	3,117	3,719	59,97	0,727	0	2,73
Rubus chamaem	2,207	2,633	62,6	1,27	0,545	1,09
Equisetum sylva	2,144	2,557	65,16	2	0	0,182
Solidago virgaur	1,858	2,216	67,38	1,18	0	0,909
Andromeda poli	1,747	2,085	69,46	0	0	1,64
Carex sp.	1,705	2,034	71,49	0	0	1,55
Lycopodium ann	1,529	1,824	73,32	1,18	0	0,455
Phyllodoce caer	1,452	1,733	75,05	0,182	0	1,27
Calamagrostis s	1,448	1,727	76,78	0,909	0,636	0
Goodyera repen	1,385	1,652	78,43	0,182	1,09	0
Trientalis europa	1,297	1,547	79,98	1,09	0	0,364
Arctostaphylos a	1,29	1,539	81,52	0,182	0	1,09
Juncus trifidus	1,251	1,492	83,01	0	0,182	1,09
Gymnocarpium	1,108	1,322	84,33	0,636	0,545	0
Sorbus aucupari	1,021	1,218	85,55	1	0	0
Polygonum vivip	0,9444	1,127	86,67	0	0,727	0,182
Huperzia selago	0,7886	0,9408	87,61	0	0,545	0,182
Geranium sylvat	0,7516	0,8967	88,51	0,636	0,182	0
Chamaenerion a	0,7343	0,876	89,39	0,727	0	0
Cirsium sp.	0,6423	0,7663	90,15	0,182	0,455	0
Hieracium muro	0,6332	0,7554	90,91	0	0,455	0,182
Juniperus comm	0,6283	0,7496	91,66	0,636	0	0
Linnaea borealis	0,6162	0,7351	92,39	0,636	0	0
Ledum palustris	0,6011	0,7171	93,11	0,182	0	0,364
Diphasiastrum a	0,5786	0,6903	93,8	0	0,364	0,182

SIMPER

Group 1
Birch

Group 2
Pine

☑ Or pool all groups

Distance/similarity measure
Bray-Curtis

Recompute

Overall average dissimilarity:
83,82

Close   Copy   Print   Help

Pic. 18 Simper - Plants

The largest values for the percentage contribution of individual components to the total variation between groups were obtained by Empetrum nigrum and Betula nana. Overall average dissimilarity is 83.82. This value indicates that, on average, the groups differ from each other by 83.82%.

3.4. Chemical data

For chemical data, principal components analysis was performed. PCA is a statistical technique used for dimensionality reduction and data exploration. Its results are included below.
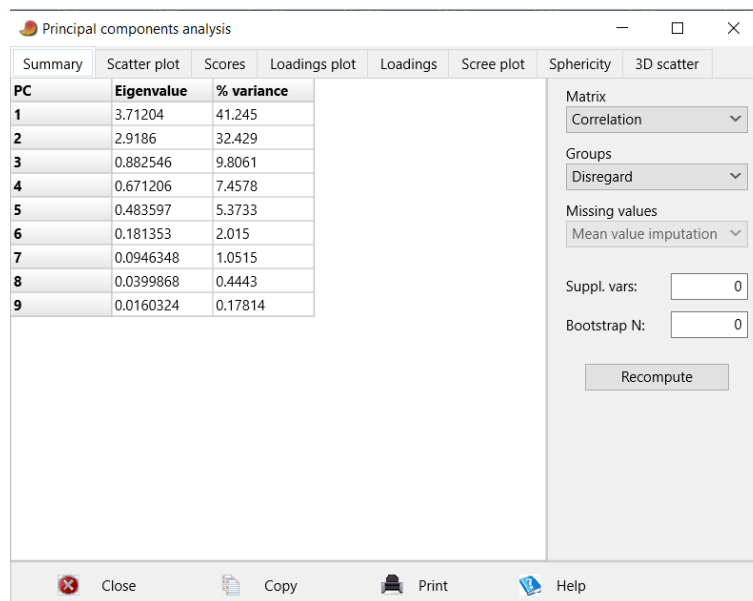

Pic. 19 PCA - Summary

Interpreting these results, it can be concluded that principal components with higher percentages of variance and larger eigenvalues are more significant in the analysis.
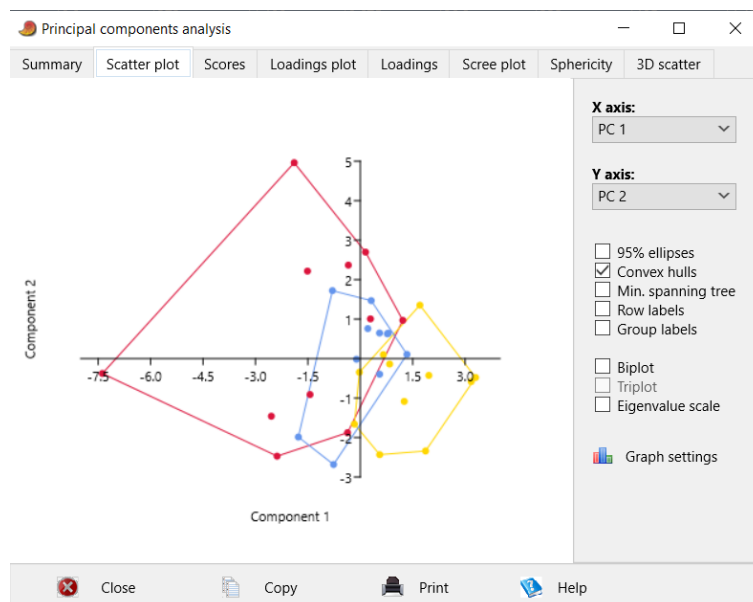

Pic. 20 PCA - Scatter plot

The "Birch" group is the largest this means that the observations belonging to it are more diverse and dispersed in the analysis space. The "Tundra" group partially overlaps with the "Birch" group. This suggests some similarity in features or characteristics between the two groups. This

may indicate some ecological proximity or similar environmental conditions that contribute to the similarity between the two groups. The "Pine" group form a separate area. This suggests that it has distinct features or characteristics that distinguish it from the other groups.
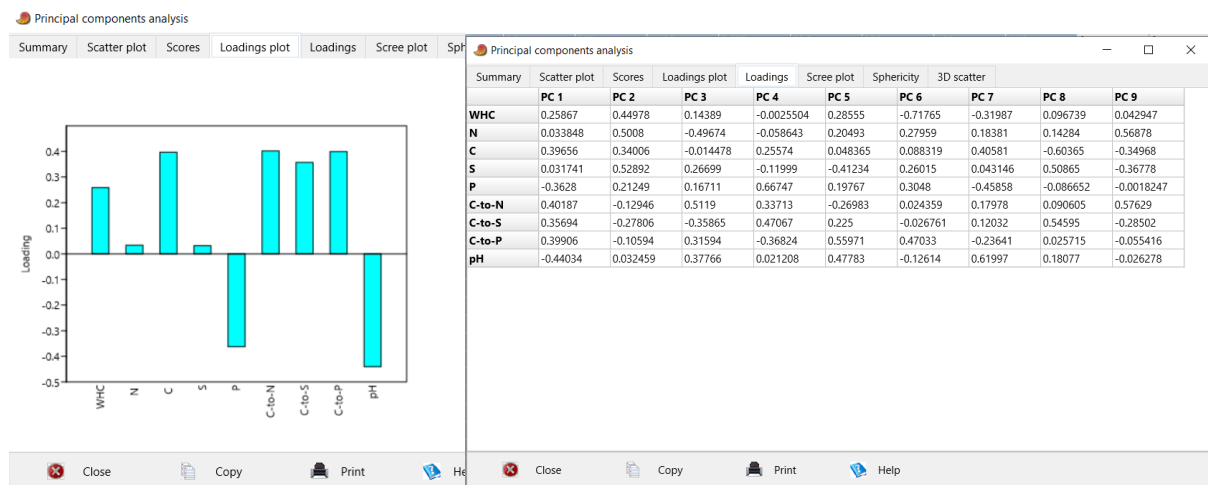
Pic. 21 PCA - Loadings plot

The loadings plot provides insights into the relationships between variables and principal components. Positive or negative loadings indicate the direction and strength of these relationships. Most variables show positive loadings, which suggests positive correlations. The coefficient of -0.45 for the variable pH suggests a negative correlation with one of the principal components.
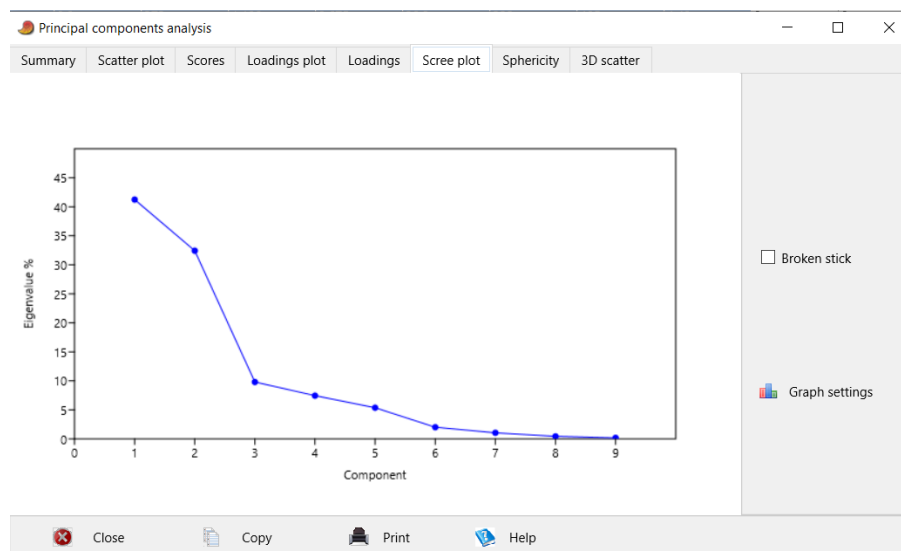
Pic. 22 PCA - Scree plot

Based on the scree plot, a decision can be made on the number of principal components to keep in the PCA analysis. Typically, those principal components that have values greater than the mean value are selected, and further principal components are omitted because they have little effect on the variance of the data. In this case, the values on the graph are decreasing, meaning

that each successive principal component explains a smaller and smaller percentage of the variance in the data. A value of 40% for the first principal component suggests that it is responsible for much of the variance in the data. After the third principal component, the values on the graph decrease more slowly, suggesting that subsequent principal components explain smaller and smaller amounts of variance in the data.
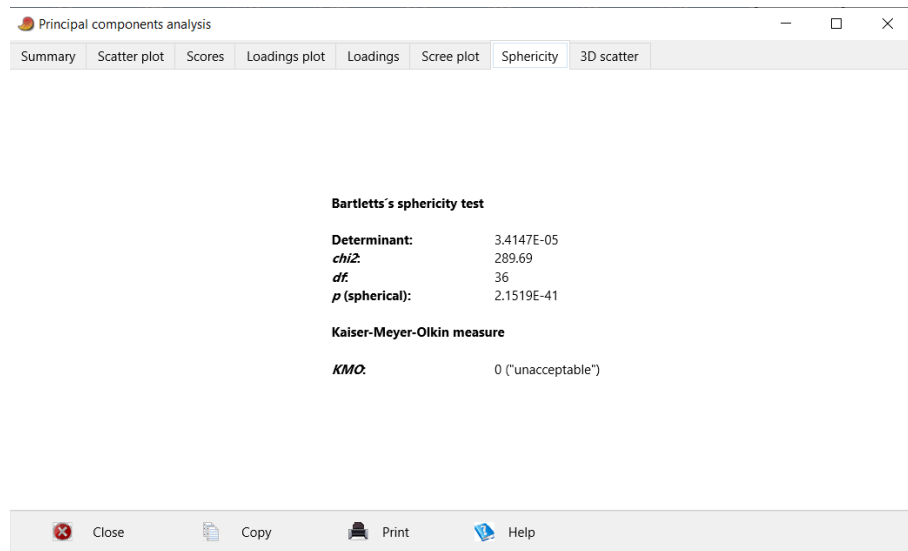
Pic. 23 PCA - Sphericity

Sphericity analysis provides information about the degree of interdependence between variables in the analyzed dataset. When the determinant has a low value (here 3.4147E-05), it means that the correlation matrix of the variables is close to the singularity and the variables are highly correlated. The chi-square (chi2) test result of 289.69, with degrees of freedom (df) of 36, indicates a significant difference between the observed correlations and the ideal sphericity model. A low p-value (2.1519E-41) confirms that the difference is statistically significant.
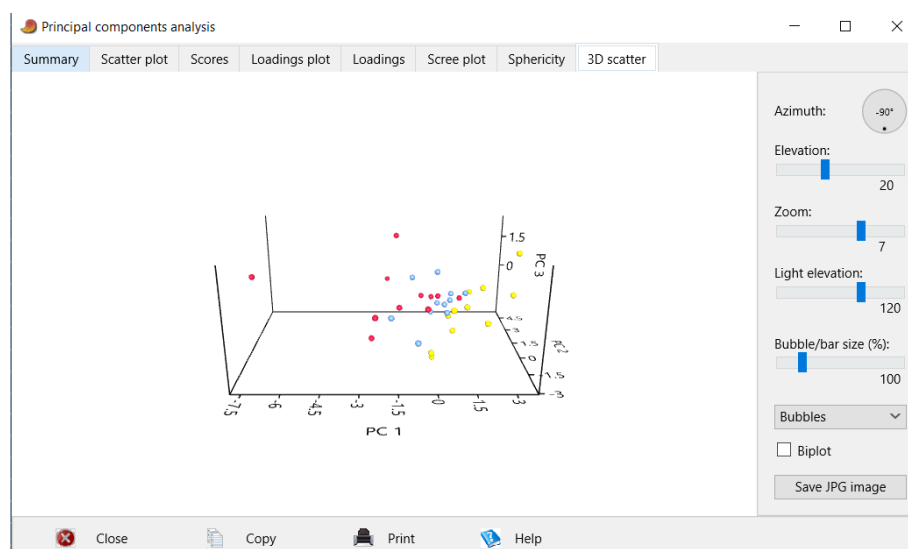
Pic. 24 PCA - 3D scatter

The analysis of the 3D scatter plot reveals differences in the distribution and interaction patterns among the different groups. The Birch group exhibits a higher level of variability in terms of composition or structure compared to the other groups, implying a greater diversity of species or subgroups within this group. The Pine group displays distinctive clusters or dissimilarities in bacterial composition. On the other hand, the Tundra group stands apart from the rest, indicating a unique bacterial composition specific to that group.

3.5 Mantel test

The Mantel test is used to evaluate the similarity between two distance or similarity matrices and test whether the similarity is statistically significant.

Firstly, a comparison was made between Bacteria and Eukaryotes.
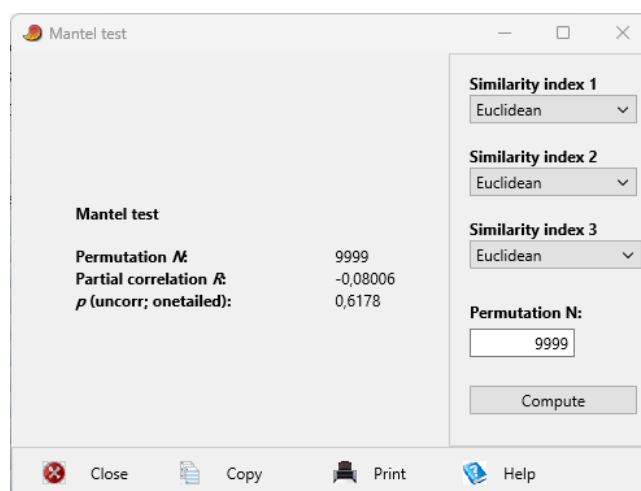
Pic. 25 Mantel test for Bacteria and Eukaryotes

The Mantel test results indicate a weak negative correlation between Bacteria and Eukaryotes, but this correlation is not statistically significant. It suggests that there is no strong evidence of a relationship between these groups in the analyzed dataset.

Secondly, a Mantet test was conducted between groups of Eukaryotes and Plants.
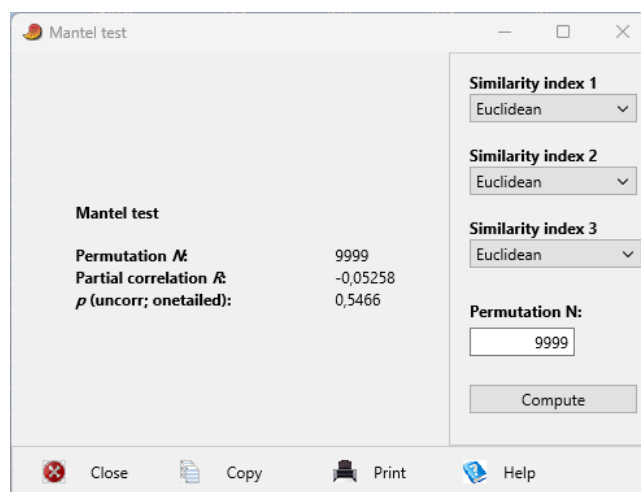
Pic. 26 Mantel test for Eukaryotes and Plants

Again, test results indicate a weak negative correlation between Eukaryotes and Plants, but this correlation is not statistically significant. It suggests that there is no strong evidence of a relationship between these groups in the analyzed dataset,

Finally, the same between groups of Bacteria and Plants.
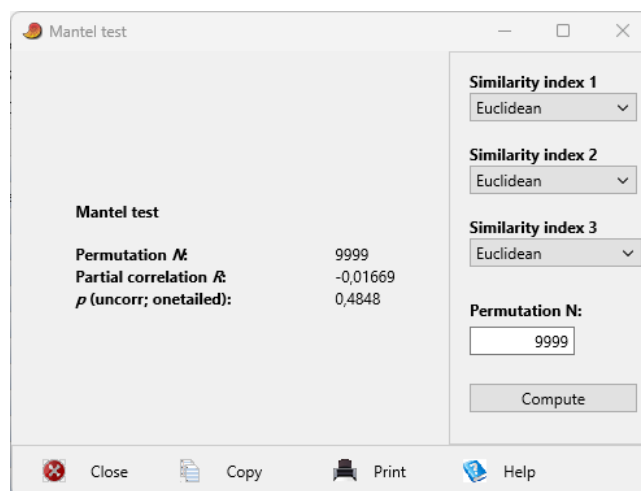
*Pic. 27 Mantel test for Bacteria and Plants*

As in previous cases, the Mantel test results indicate a very weak negative correlation between Bacteria and Plants, but this correlation is not statistically significant. It suggests that there is no strong evidence of a relationship between these groups in the analyzed dataset.
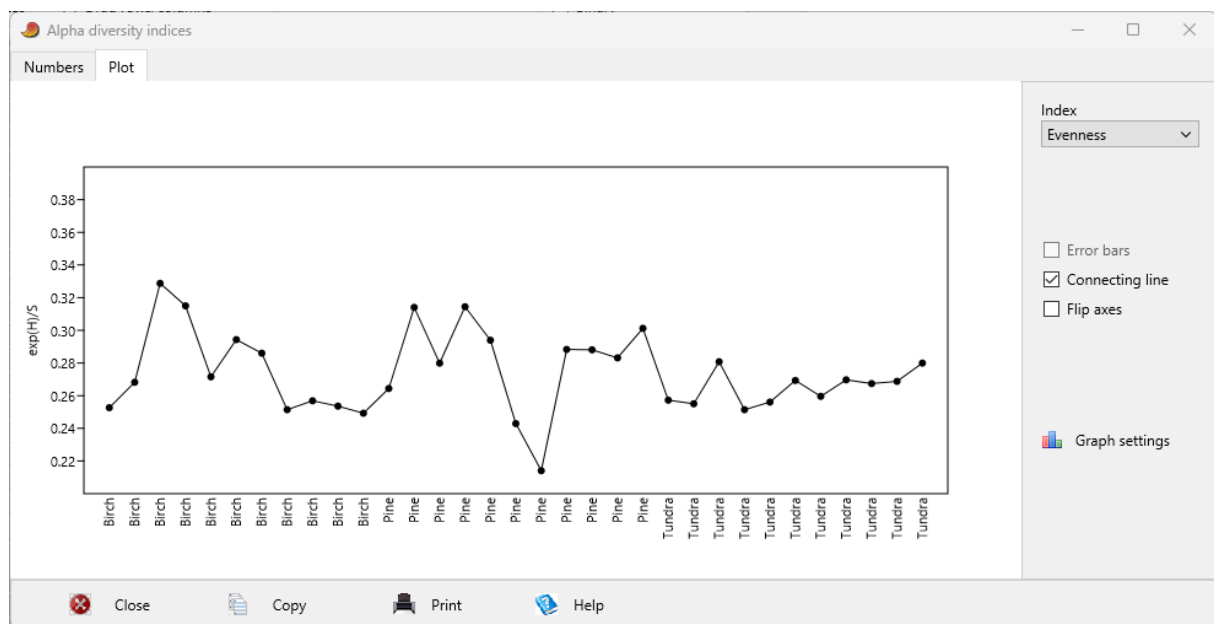
3.6 Shannon, Simpson, Piellou

In the last step, selected diversity indices (Shannon, Simpson, Piellou) were calculated for Plants, Eukaryotes and Bacteria in the three ecosystems studied. The first group analyzed was Bacteria.
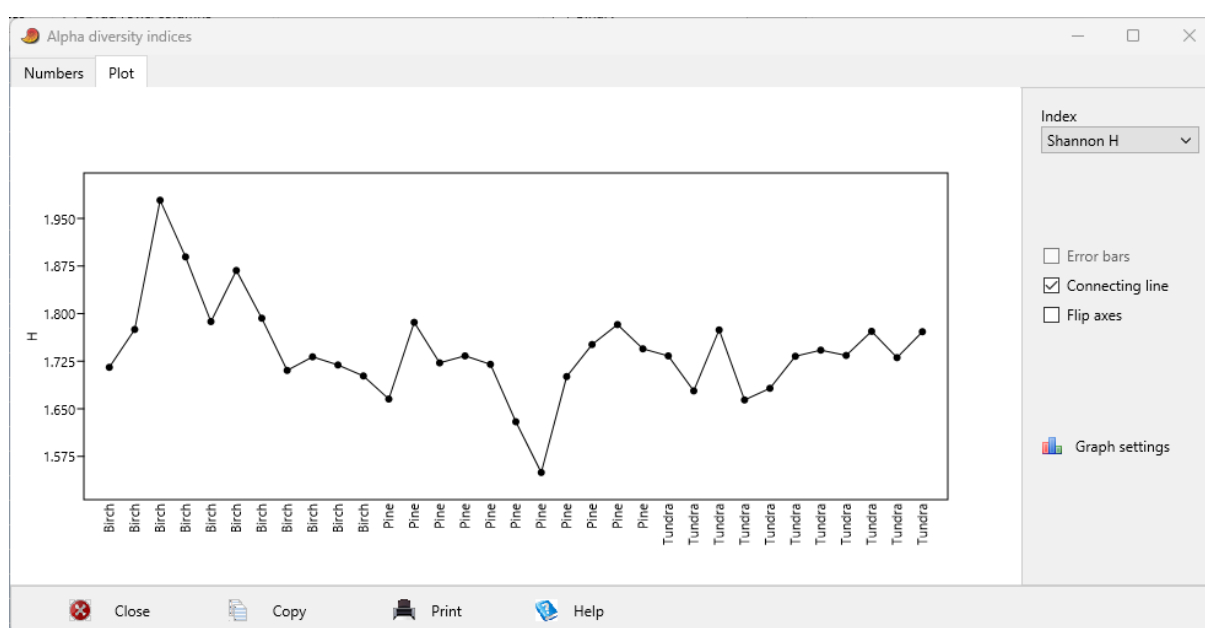

*Pic. 28 Shannon - Bacteria*
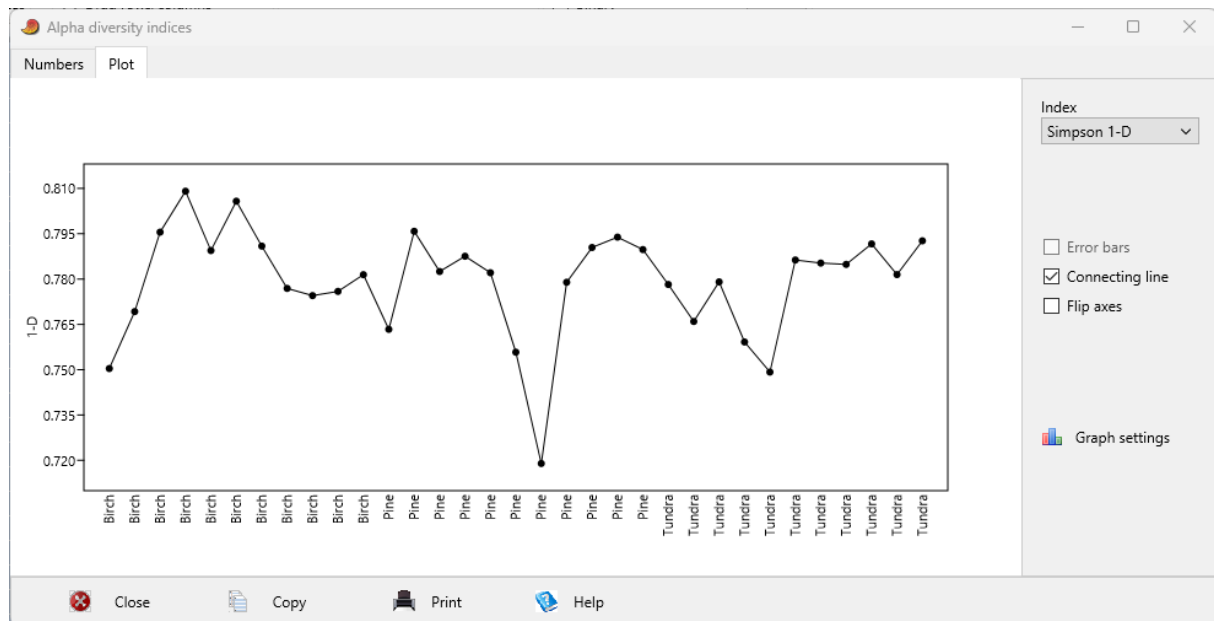
Pic. 29 Simpson - Bacteria

Pic. 30 Evenness - Bacteria

In summary, the Birch sector stands out with the highest values of diversity, indicating a more various and even community of bacteria species. The Pine sector has the lowest values, suggesting lower species richness and potentially less evenness. The Tundra sector shows relatively consistent values of diversity, indicating a relatively stable level of uniformity.
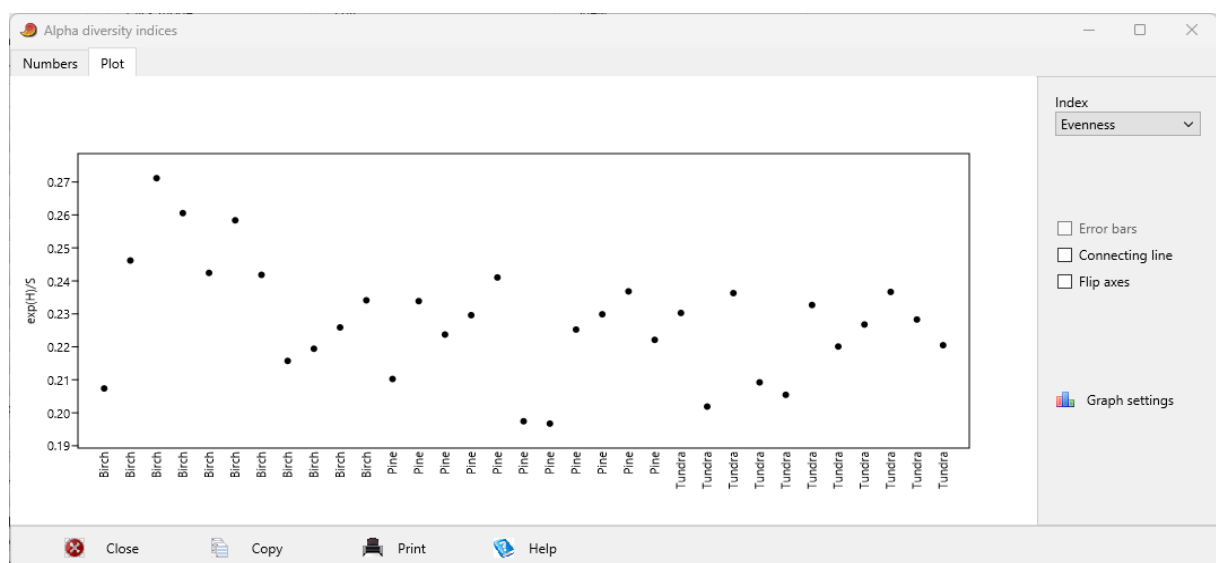
The second group analyzed were Eucaryotes.
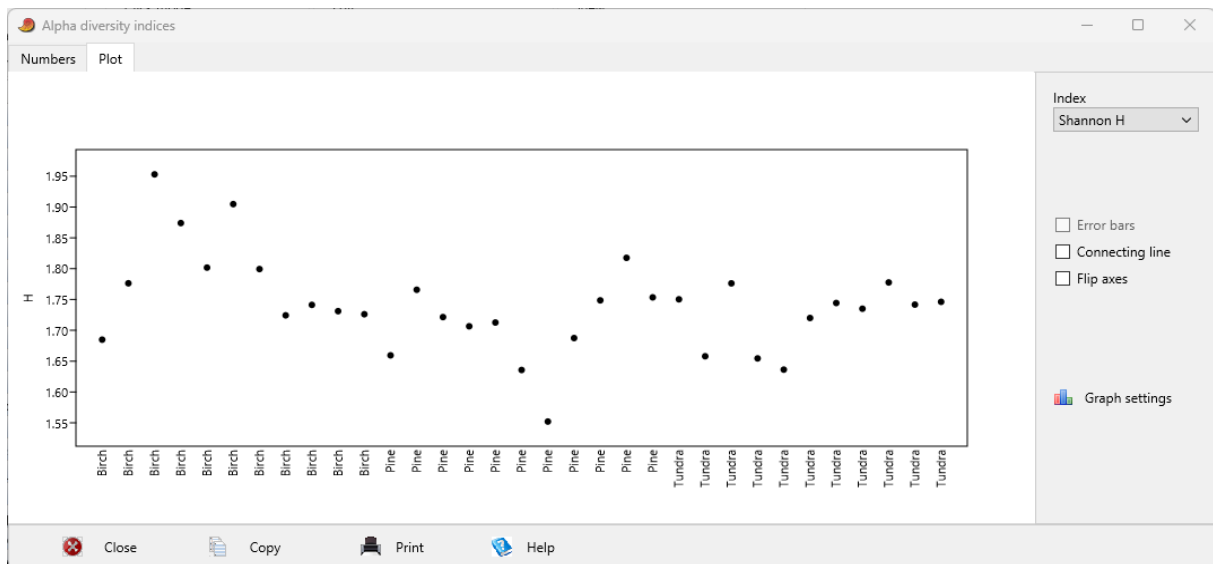
Pic. 31 Shannon - Eukaryotes
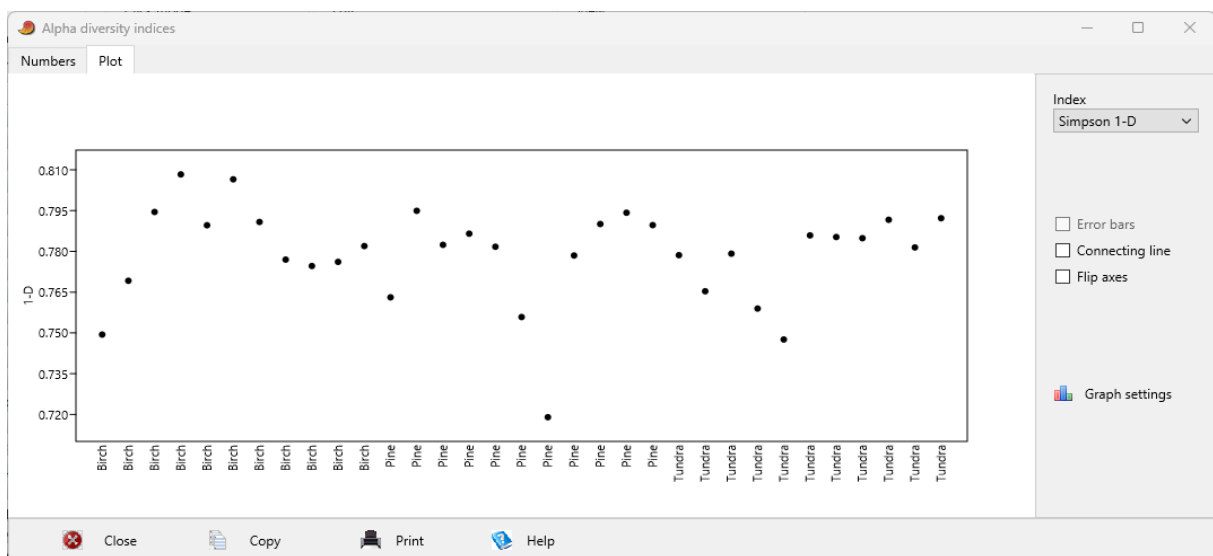


Pic. 32 Simpson – Eukaryotes

Pic. 33 Evenness - Eukaryotes

Overall, the analysis suggests that the Birch group has the highest diversity and evenness, indicating a more diverse and balanced community of eukaryotes species. The Simpson diversity values indicate high dominance in all groups, but with some variability, particularly in the Pine group. The evenness values do not show a strong pattern among the groups, suggesting that the distribution of eukaryores within the groups is more variable.
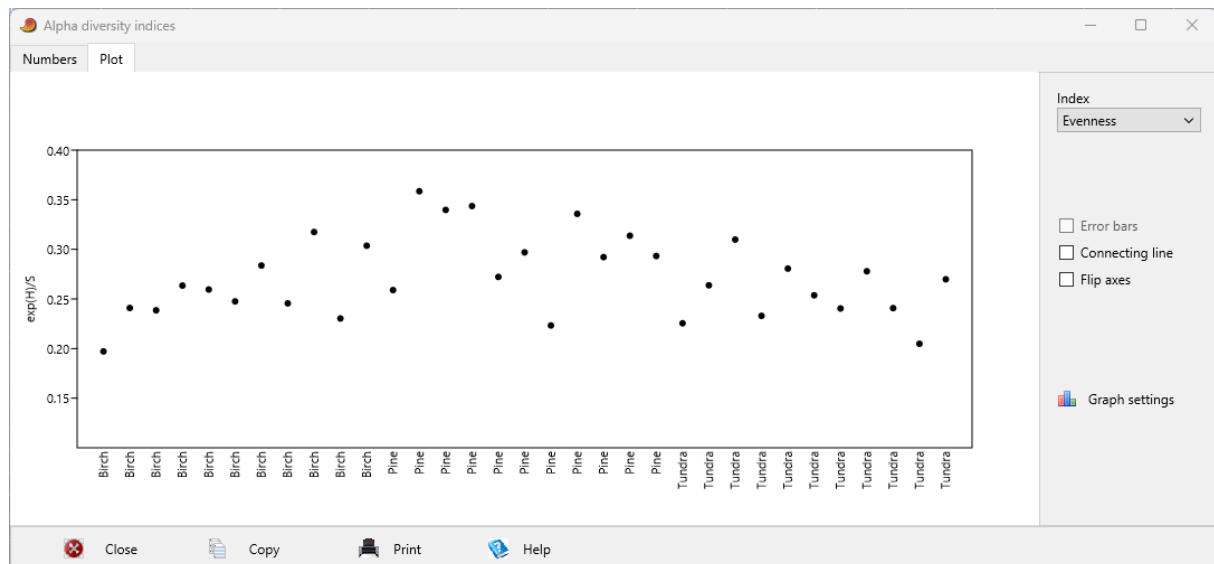
The third and final group analyzed was Plants.
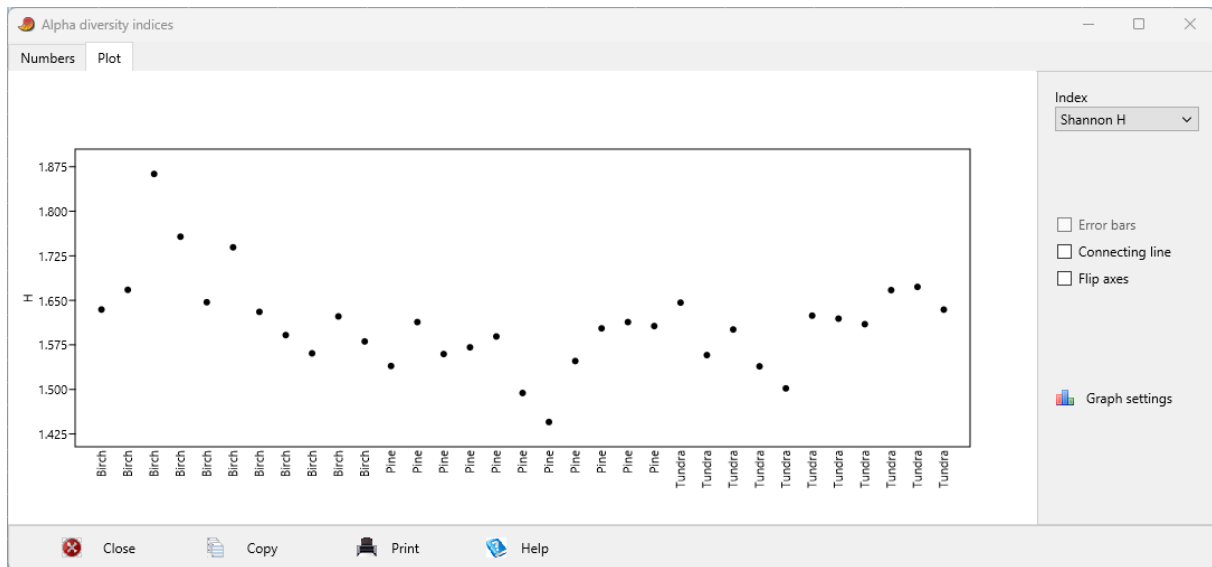
Pic. 34 Shannon - Plants
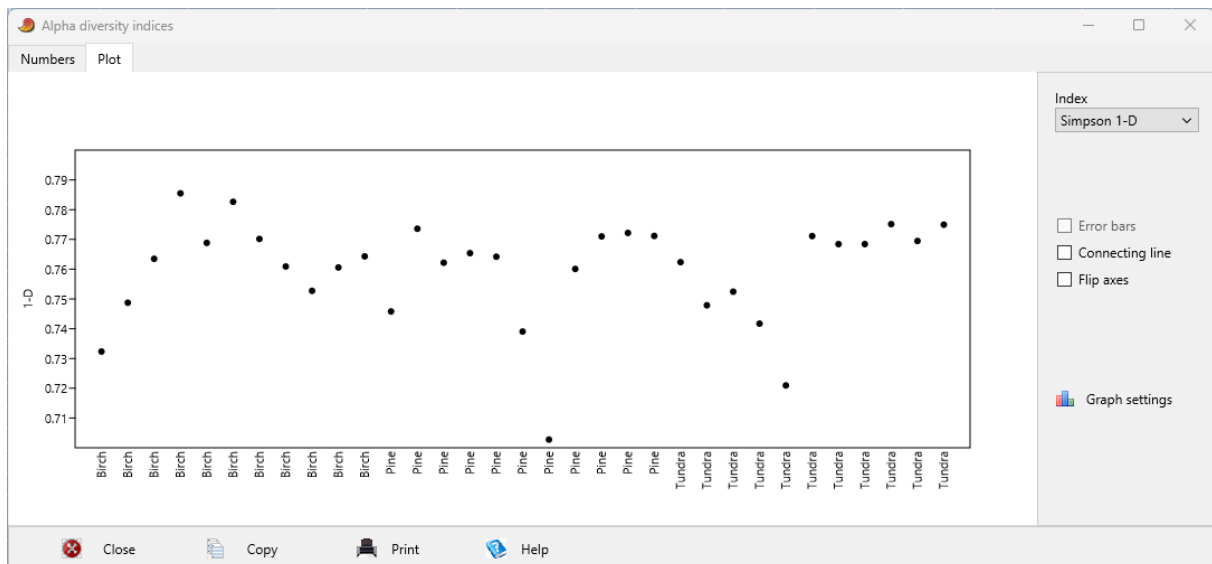
Pic. 35 Simpson - Plants

Pic. 36 Evenness – Plants

The Shannon diversity values are again the highest for the Birch group. This suggests that it is the most diverse and balanced community of vegetation species. The Pine and Tundra groups show similar values, but there are some notably low values observed in the Pine group. The Simpson diversity values, represented by the index 1-D, show some variability within each group but generally remain at relatively high levels. This indicates a high dominance of certain species within each group. However, there may be some differences in the dominance patterns within each group, with a few exceptions. The evenness values indicate that the Pine group has the highest evenness, suggesting a relatively equal distribution of species within the group.

4. Summary

In short, there are definitely differences in the structure of all the three mentioned communities. The smallest were observed in the case of eukaryotes, this can be seen by analyzing the results and in particular is evident in the scatter plot generated with NDMS where conex hulls overlap to some extent. In contrast, the greatest differences in structure can be seen for the herbal plants group, where the convex hulls are relatively far apart and none of the groups intersect. Using Simper, it was found that the largest contribution to dissimilarity between groups, showed taxons such as Acidobacteria in the Bacteria group, Ascomycota in the Eucaryotes group and for the Plants group - Empetrum nigrum and Betula nana. In the case of this project and the data set used, the relationship between communities cannot be clearly established. Moreover, it has been shown that soil properties are influenced by various factors, which vary across ecosystem types, leading to differences in soil chemical properties. By calculating selected diversity indices, it was discovered that, in general, the Birch group is the most diverse and balanced community of species considering bacteria, eukaryotes and vegetation.