

Beta Diversity and Unconstrained/Constrained Ordination

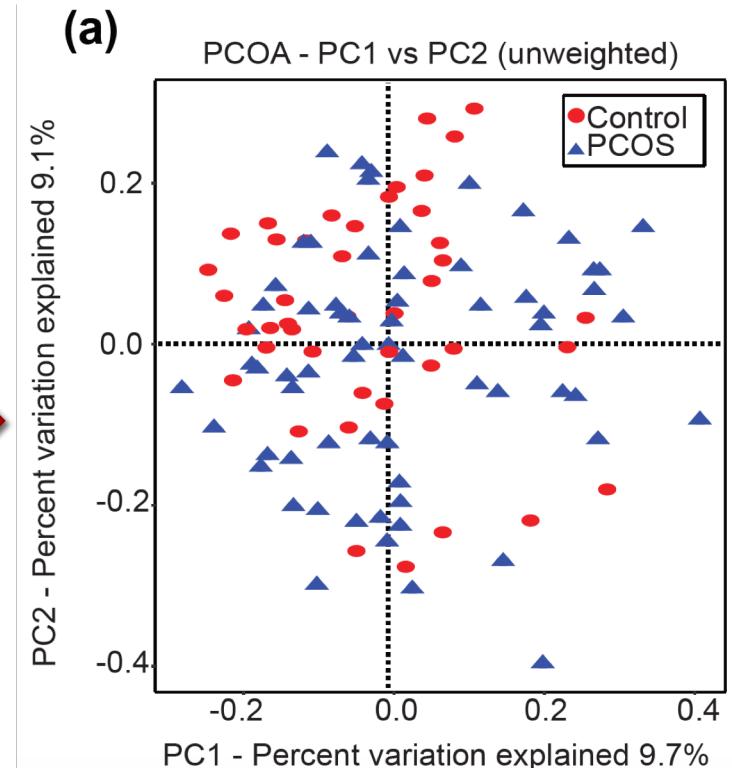
PEDRO J. TORRES
SDSU BB CLUB 2018

Analyzing multivariate Data

Description	g_Bifidobacterium	g_Adlercreutzia	g_Bacteroides	g_Parabacteroides	g_Prevotella
49.T1	0.029688273	0.004136566	0.006927264	0.001979218	0.013894112
49.T2	0.000491517	0.000523228	0.008086253	0.004835897	0.003440622
49.T4	9.79E-05	0.001598278	0.01141627	0.004990541	0.030302042
49.T5	0.000425743	0.002057759	0.015326758	0.004221954	0.012914213
54.T3	0.000208195	0.003580946	0.01486509	0.00341439	0.040972685
54.T4	0.0002342	0.001059476	0.020330779	0.003022294	0.036613247
54.T5	0.000502487	0.003668157	0.015526858	0.005250992	0.054318878
54.T2	0.001465668	0.000318624	0.023163932	0.002676438	0.042950454
54.T1	0.000590846	0.000398478	0.02590104	0.007818404	0.02067961
51.T1	0.000449573	0.004570658	0.026899446	0.008953994	0.017383486
56.T4	0.000323158	0.0250123	0.02594149	0.00133376	0.237190884
56.T5	0.00032808	0.02515512	0.01511228	0.00152540	0.033323067
49.T3	0.000140549	0.01289861	0.024305259	0.008796032	0.02850787
56.T1	0.000592098	0.001453332	0.037266301	0.011788137	0.028905157
56.T2	0.000581024	0.002091686	0.002963221	0.001568764	0.012375806
56.T3	0.000647591	0.005156741	0.03106037	0.005084786	0.021490418
52.T1	0.003072076	0.000393856	0.004883813	0.004174872	0.031744781
52.T2	0.100924328	0.000793474	0.021576925	0.004217941	0.007948661
52.T3	0.000857375	0.001089097	0.037492759	0.011331248	0.010079944
52.T4	0.000239794	0.001408788	0.021461543	0.003596907	0.022900306
52.T5	0.00030994	0.0040735	0.017843702	0.005401815	0.016338278
55.T1	0.000358749	0.001638286	0.015749067	0.003886444	0.013453076
55.T2	0.001227948	0.003192665	0.020936515	0.004666203	0.012218084
55.T5	0.000177167	0.004657110	0.005402400	0.005545210	0.003222001

COMPLICATED

Beta
Diversity

Analyzing multivariate Data

Want to know which samples are the most similar and perhaps which are behaving in a similar way.

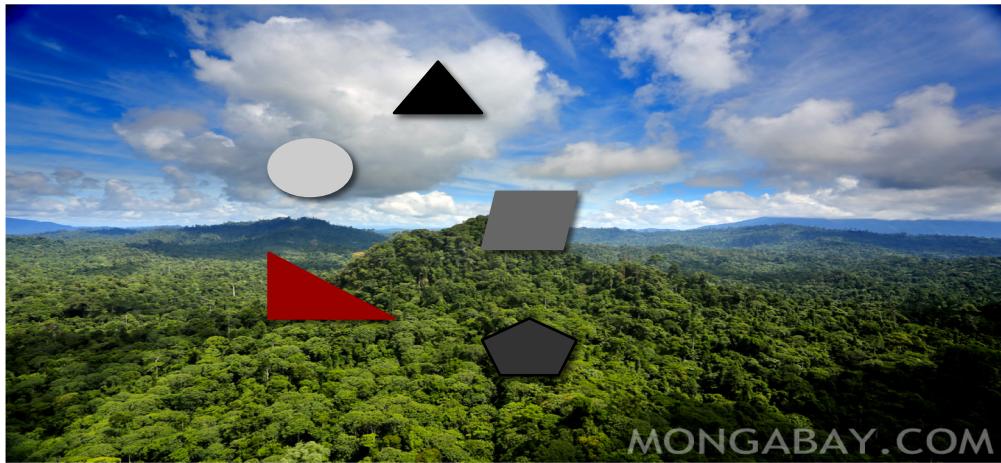
Outline

1. Get Data
2. Get beta diversity dissimilarity matrix
3. Use unconstrained ordination to explore data cloud
4. Statistics to look for potential underlying factors contributing to patterns (seen or unseen)
5. Constrained Ordination to test hypothesis

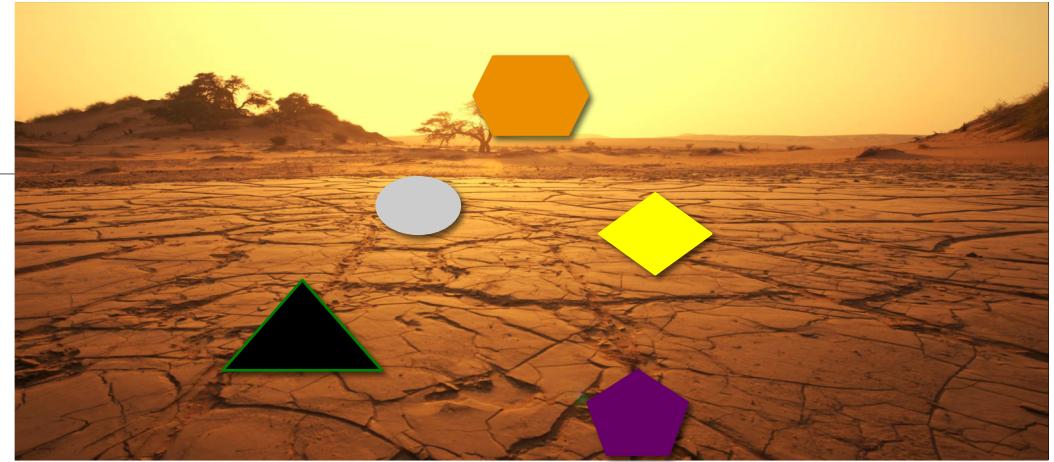
Beta Diversity metrics

- ❖ Operated of a pair of samples (i.e., between sample diversity)

Beta Diversity



Community A



Community B

Beta Diversity metrics

- ❖ Operated of a pair of samples (i.e., between sample diversity)
 - ❖ **Qualitative diversity metrics** – only considers the presence/absence of features
 - ❖ **Quantitative diversity metrics** - considers abundance of features
 - ❖ **Phylogenetic diversity metrics** - incorporate evolutionary relationships between taxa (where they lay of a tree)
 - ❖ **Non-phylogenetic diversity metrics** – assume all taxa are equally related, so makes no assumption of evolutionary relationship

Beta Diversity metrics

- ❖ Jaccard distance – a qualitative, non-phylogenetic approach
- ❖ Bray Curtis – a quantitative, non phylogenetic approach
- ❖ Unweighted UniFrac – a qualitative, phylogenetic approach
- ❖ Weighted UniFrac – a quantitative

Dissimilarity Matrix Will Depend on Metric Used

	gut1	left. palm1	right. palm1	tongue1
gut1	0.0	0.19	0.15	0.65
left.palm1	0.19	0.0	0.07	0.69
right.palm1	0.15	0.07	0.0	0.70
tongue1	0.65	0.69	0.70	0.0

As you add more samples you start to increase your comparisons/matrix leading to an increase in dimensions as well.

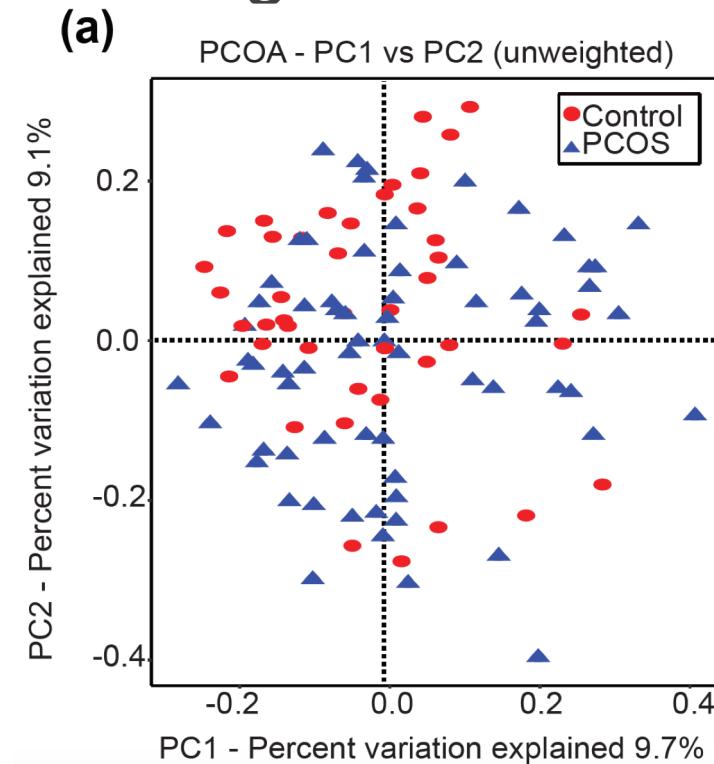
Ordination

Goal – to represent sample relationships in a low-dimensional space

Great for data exploration and are considered hypothesis generators.

What is ordination?

- ❖ In an ordination graph, sites/samples are plotted so that distances between them in a graph reflect the ecological differences between them.



Types of Ordination

- ❖ Unconstrained – Exploratory/Hypothesis generating
- ❖ Does not use a priori hypothesis in any way, reduces dimensions on the basis of some general criterion (minimizing residual variance and minimizing stress function)

Unconstrained

- ❖ Includes: PCoA and NMDS
- ❖ Extremely useful to visualize broad patterns

- ❖ Problem: Tend to look at patterns of overall dispersion which can sometimes mask differences in multivariate location among groups

Statistics

- ❖ PERMANOVA – a non parametric multivariate statistical test
 - ❖ Compares groups of objects to test the null hypothesis that the centroids and dispersion of groups are equivalent for all groups.
- ❖ Once an important factor is identified we can move on to...

Types of Ordination

- ❖ Unconstrained – Exploratory/Hypothesis generating
 - ❖ Does not use a priori hypothesis in any way, reduces dimensions on the basis of some general criterion (minimizing residual variance and minimizing stress function)
- ❖ Constrained - Hypothesis testing
 - ❖ Use a priori hypothesis to produce the plot
 - ❖ Data matrix is subjected to a weighted regression to my constraining variable and the fitted values are then plotted.

Constrained Analysis

- ❖ Includes: Canonical correspondence analysis (CCA) and Canonical analysis of principal coordinates (CAP)
- ❖ Allows you to uncover important patterns in the multivariate data by reference to a relevant hypothesis aka you want to explicitly plot against certain gradients/factors
- ❖ Simply put, CCA can be used as a direct gradient analysis (think linear regression) use with continuous variables
- ❖ CAP can be used for categorical or continuous variables
- ❖ Use both and see what works best for your dataset, both have different ways of calculating your distance matrix in regards to your constraining variables

Conclusion

These are all useful tools for data exploration and hypothesis testing.

They all have their own usefulness depending on the question being asked and the data at hand.

One approach does not completely replace another, understand what each step does and each can add something useful to your data.