MB590-012
Microbiome Analysis
Beta Diversity

Dr. Christine Hawkes

NC STATE UNIVERSITY

Today’s outline

- Beta-diversity
  - Overview
  - Distance/dissimilarity metrics
- Phylogenetic beta-diversity
  - Overview
  - Transformation
  - Distance
- Ordination

- Reminder: HPC tutorial certificates **due by Tue 2/22** (3 of you still have not logged in!)

HPC prep for next week (before class!)

See: MicrobiomeAnalysis2022/dada2-on-hpc

Log in and go to the scratch directory:

```
cd /share/mb590s22
ls
```

There should be a directory with your UnityID.

```
cd $USER
```

If not:

```
mkdir $USER
cd $USER
```

Check the groups – you should be in ***mb590s22*** and ***bioinfo*** :

```
groups
```

Before logging out, do:

```
module load conda
conda init tcsh
```

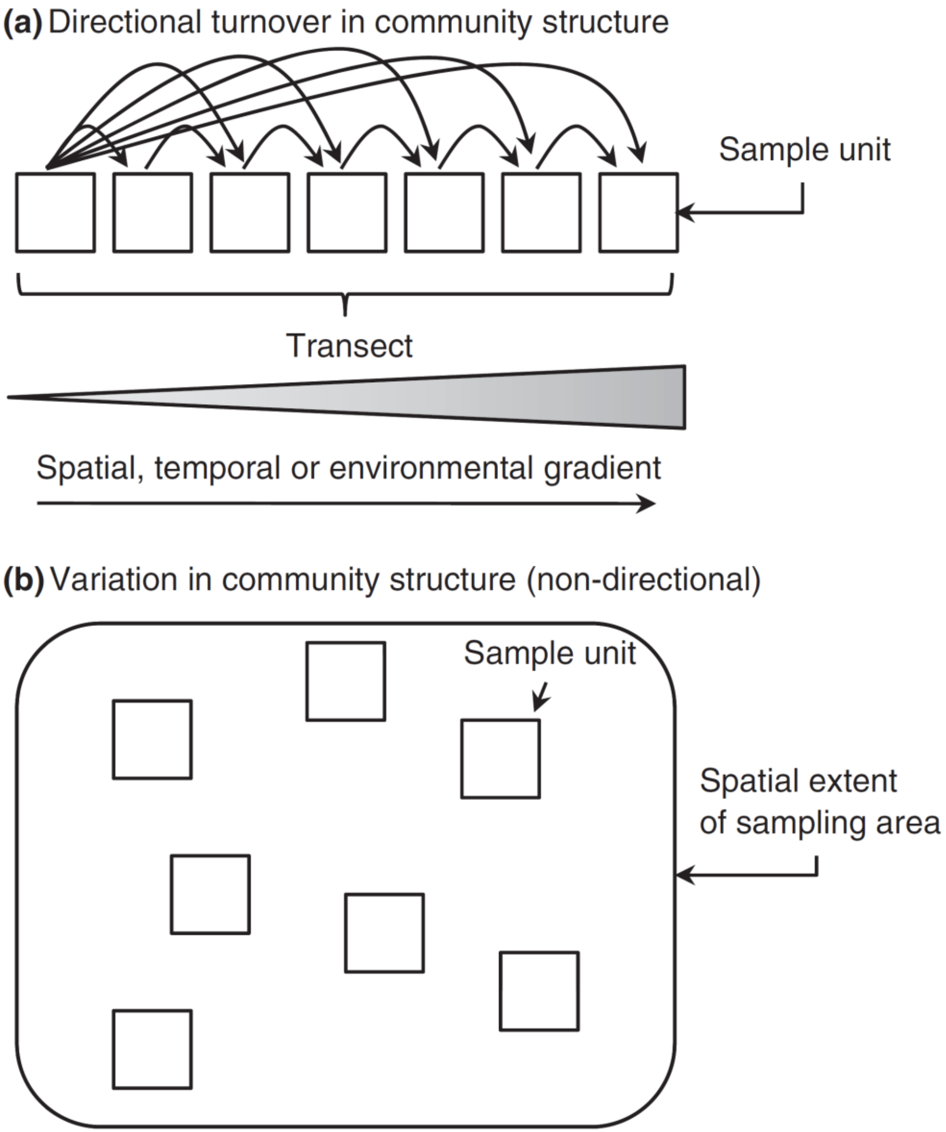
Scale of diversity

1. **alpha** diversity = local, single community

2. **beta** diversity = between two communities

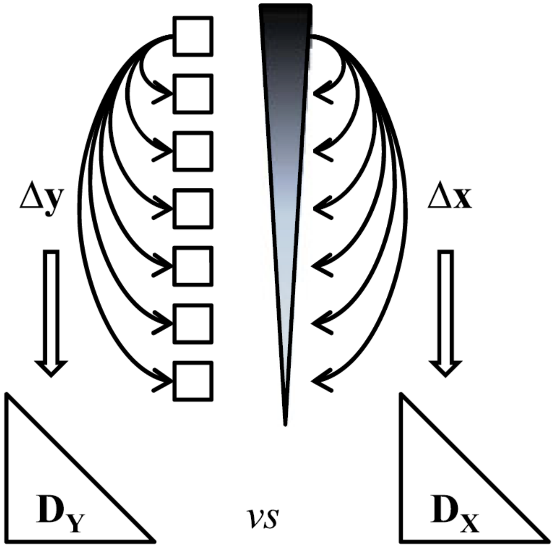
Beta-diversity

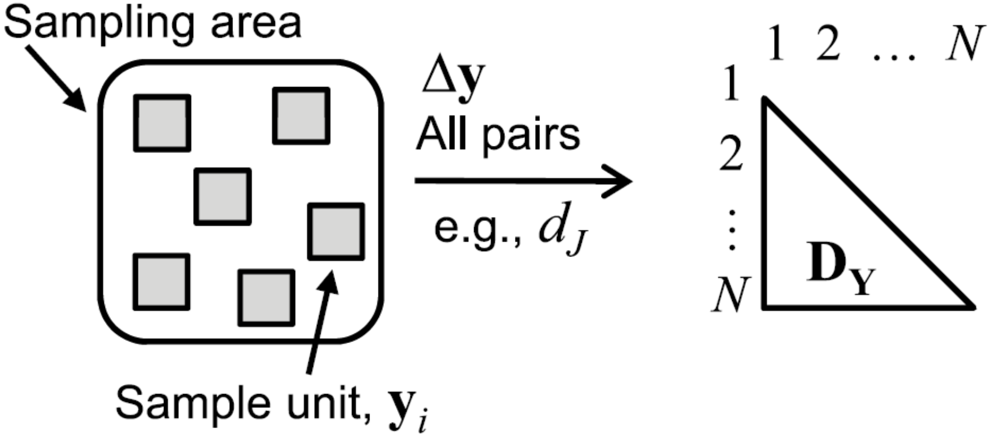
- Variation in species across samples
- Measure change in community structure between samples
- Change can be in identity or abundance of taxa


(a) Directional turnover in community structure
Sample unit
Transect
Spatial, temporal or environmental gradient
(b) Variation in community structure (non-directional)
Sample unit
Spatial extent of sampling area

Anderson et al. 2011 Eco Lett doi: 10.1111/j.1461-0248.2010.01552.x

Beta-diversity – distance/ dissimilarity comparisons

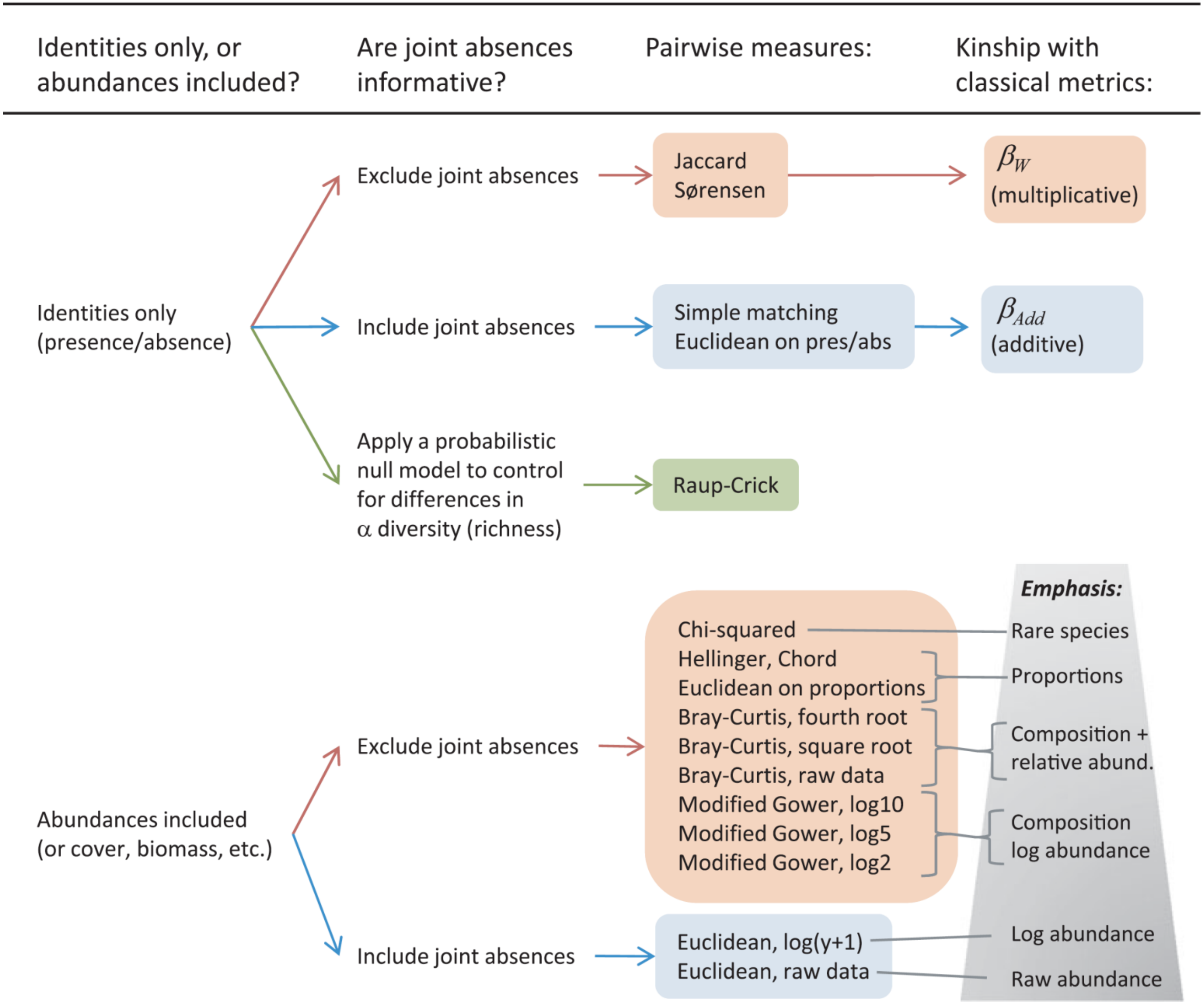




- Whether gradient turnover or spatial variation, goal is to define a community matrix of pairwise dissimilarities between samples
- Can also examine relationship of those dissimilarities to environmental data

Anderson et al. 2011 Eco Lett doi: 10.1111/j.1461-0248.2010.01552.x

Beta-diversity – workflow



Identities only, or abundances included?

Are joint absences informative?

Pairwise measures:

Kinship with classical metrics:

Identities only (presence/absence)

Exclude joint absences → Jaccard, Sørensen → β_W (multiplicative)

Include joint absences → Simple matching, Euclidean on pres/abs → β_{Add} (additive)

Apply a probabilistic null model to control for differences in α diversity (richness) → Raup-Crick

Abundances included (or cover, biomass, etc.)

Exclude joint absences → Chi-squared; Hellinger, Chord; Euclidean on proportions; Bray-Curtis, fourth root; Bray-Curtis, square root; Bray-Curtis, raw data; Modified Gower, log10; Modified Gower, log5; Modified Gower, log2

Include joint absences → Euclidean, log(y+1); Euclidean, raw data

Emphasis:

Rare species

Proportions

Composition + relative abund.

Composition log abundance

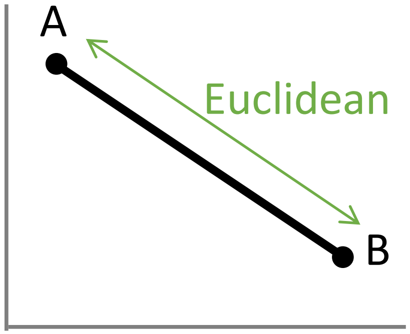
Log abundance

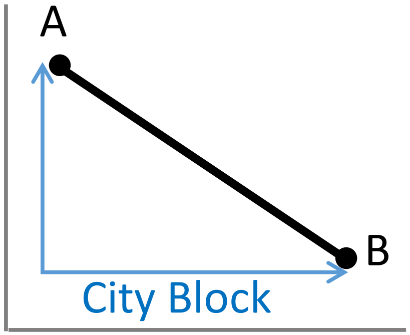
Raw abundance

Anderson et al. 2011 Eco Lett doi: 10.1111/j.1461-0248.2010.01552.x

Beta-diversity – distance / dissimilarity

- Turnover and variation are measured as pairwise distance in species composition
- Some distances can be thought of geometrically
- Common examples:



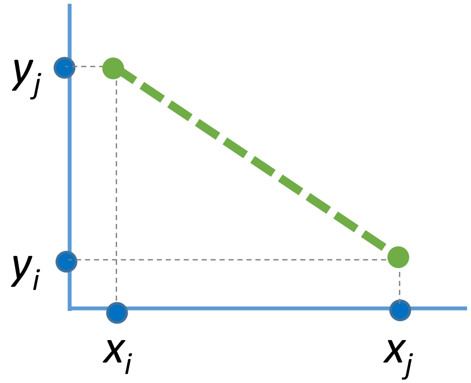


Anderson et al. 2011 Eco Lett doi: 10.1111/j.1461-0248.2010.01552.x

Beta-diversity – Euclidean distance

- Euclidean distance between two points

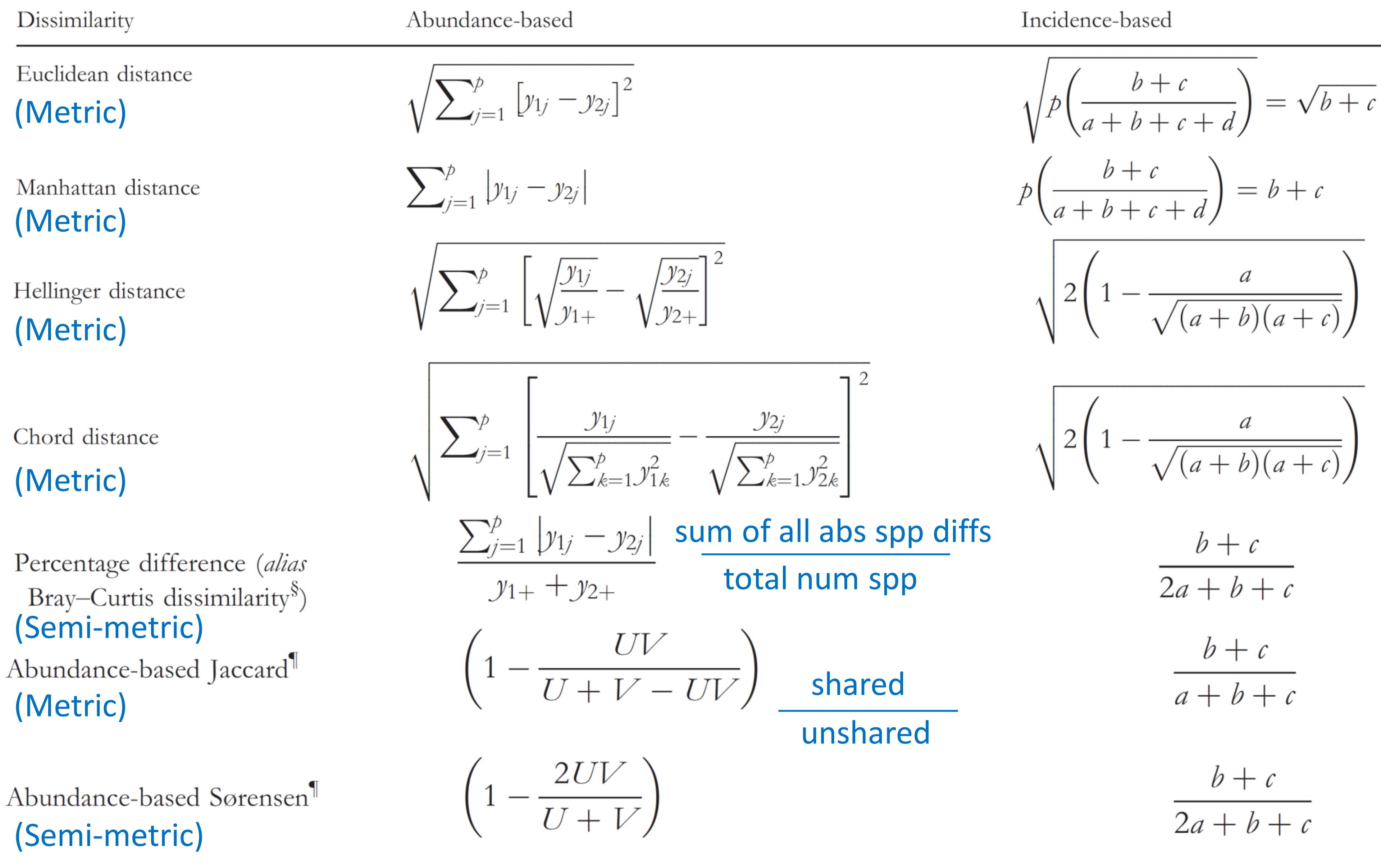
$$D^{ij}_{Eucl} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$



- Extended to many pairs of samples (i,j) with many ASVs (k = 1 to n)

$$D^{ij}_{Eucl} = \sqrt{\Sigma^{n}_{k=1}(\underset{\underset{\text{\# ASV k in sample i}}{\uparrow}}{x_{k,i}} - \underset{\underset{\text{\# ASV k in sample j}}{\uparrow}}{x_{k,j}})^2}$$

Beta-diversity – common distances



Dissimilarity	Abundance-based	Incidence-based
Euclidean distance (Metric)	$\sqrt{\sum_{j=1}^{p}\left[y_{1j}-y_{2j}\right]^{2}}$	$\sqrt{p\left(\frac{b+c}{a+b+c+d}\right)}=\sqrt{b+c}$
Manhattan distance (Metric)	$\sum_{j=1}^{p}\left\lvert y_{1j}-y_{2j}\right\rvert$	$p\left(\frac{b+c}{a+b+c+d}\right)=b+c$
Hellinger distance (Metric)	$\sqrt{\sum_{j=1}^{p}\left[\sqrt{\frac{y_{1j}}{y_{1+}}}-\sqrt{\frac{y_{2j}}{y_{2+}}}\right]^{2}}$	$\sqrt{2\left(1-\frac{a}{\sqrt{(a+b)(a+c)}}\right)}$
Chord distance (Metric)	$\sqrt{\sum_{j=1}^{p}\left[\frac{y_{1j}}{\sqrt{\sum_{k=1}^{p}y_{1k}^{2}}}-\frac{y_{2j}}{\sqrt{\sum_{k=1}^{p}y_{2k}^{2}}}\right]^{2}}$	$\sqrt{2\left(1-\frac{a}{\sqrt{(a+b)(a+c)}}\right)}$
Percentage difference (*alias* Bray–Curtis dissimilarity[§]) (Semi-metric)	$\frac{\sum_{j=1}^{p}\left\lvert y_{1j}-y_{2j}\right\rvert}{y_{1+}+y_{2+}}$ $\frac{\text{sum of all abs spp diffs}}{\text{total num spp}}$	$\frac{b+c}{2a+b+c}$
Abundance-based Jaccard[¶] (Metric)	$\left(1-\frac{UV}{U+V-UV}\right)$ $\frac{\text{shared}}{\text{unshared}}$	$\frac{b+c}{a+b+c}$
Abundance-based Sørensen[¶] (Semi-metric)	$\left(1-\frac{2UV}{U+V}\right)$	$\frac{b+c}{2a+b+c}$

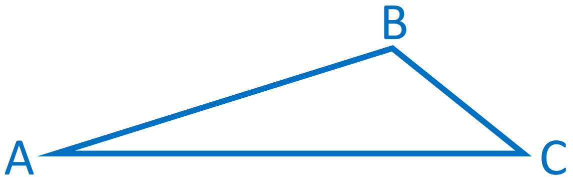
Legendre & Caceras 2013 Eco Lett 16: 951-963 doi: 10.1111/ele.12141; Chao et al. 2004 Eco Lett 8: 148-159 doi 10.1111/j.1461-0248.2004.00707.x

Beta-diversity – distance properties

Metric distances

1. Minimum value is zero when two samples are identical
2. When two samples differ, distance is positive (no negative distances allowed)
3. Distances are symmetric (A to B is same as B to A)
4. Triangle inequality axiom: with three samples, distance between one pair cannot be larger than the sum of the other two distances

   ((A to B) + (B to C)) > (A to C)



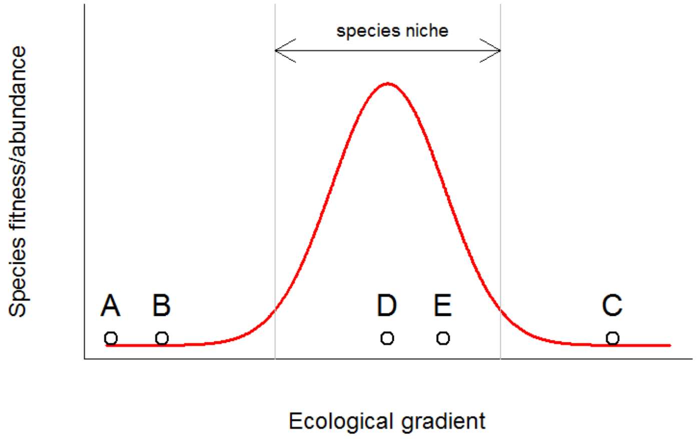
Semi-metric distances violate #4

Non-metric distances can also violate #1-3 (not used in ecology)

Gottelli & Ellison (2004) A Primer of Ecological Statistics. Sinauer Assoc.

Beta-diversity – joint/double zeros

- Double zeros typically do not provide ecological insight
- Symmetrical coefficients
  - treat double zeros the same as shared presences
  - consider samples with shared zeros as more similar
  - Euclidean, Manhattan
- Asymmetrical coefficients
  - distance does not change with double zeros
  - consider only shared presences
  - Bray-Curtis, Jaccard, Sorensen, Hellinger, Chord



Anderson et al. 2011 EcoLett doi: 10.1111/j.1461-0248.2010.01552.x

Distance properties

Double-zero asymmetry

Largest distances between samples with no shared taxa

Distance increases with unique taxa within sites

Replication invariance

Unit invariance

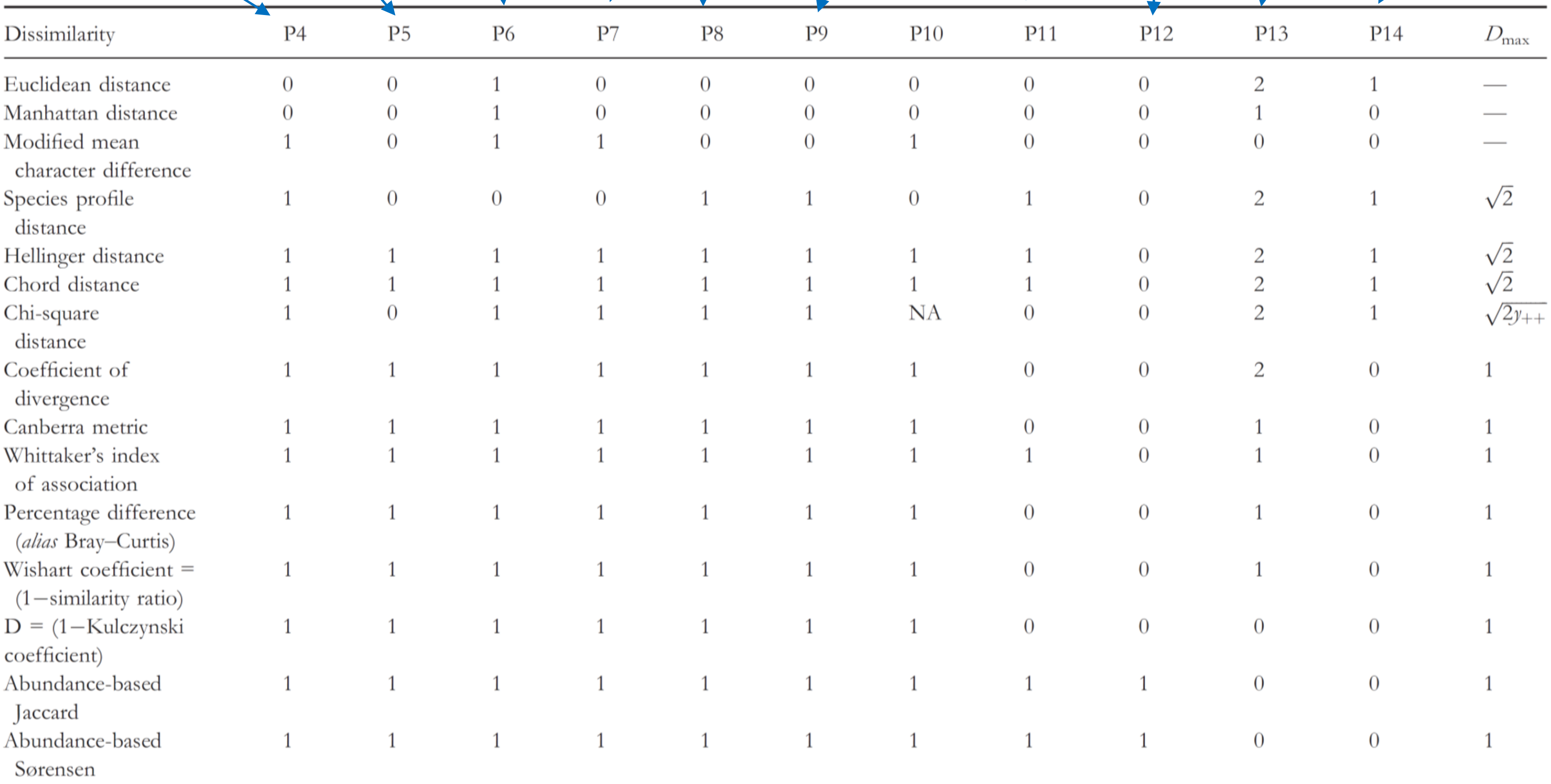
Fixed upper bound

Invariance to abundance

Corrections for undersampling

Euclidean properties

Emulated by transform + Euclidean distance



Dissimilarity	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	$D_{\max}$
Euclidean distance	0	0	1	0	0	0	0	0	0	2	1	—
Manhattan distance	0	0	1	0	0	0	0	0	0	1	0	—
Modified mean character difference	1	0	1	1	0	0	1	0	0	0	0	—
Species profile distance	1	0	0	0	1	1	0	1	0	2	1	$\sqrt{2}$
Hellinger distance	1	1	1	1	1	1	1	1	0	2	1	$\sqrt{2}$
Chord distance	1	1	1	1	1	1	1	1	0	2	1	$\sqrt{2}$
Chi-square distance	1	0	1	1	1	1	NA	0	0	2	1	$\sqrt{2y_{++}}$
Coefficient of divergence	1	1	1	1	1	1	1	0	0	2	0	1
Canberra metric	1	1	1	1	1	1	1	0	0	1	0	1
Whittaker's index of association	1	1	1	1	1	1	1	1	0	1	0	1
Percentage difference (*alias* Bray–Curtis)	1	1	1	1	1	1	1	0	0	1	0	1
Wishart coefficient = (1−similarity ratio)	1	1	1	1	1	1	1	0	0	1	0	1
D = (1−Kulczynski coefficient)	1	1	1	1	1	1	1	0	0	0	0	1
Abundance-based Jaccard	1	1	1	1	1	1	1	1	1	0	0	1
Abundance-based Sørensen	1	1	1	1	1	1	1	1	1	0	0	1

Legendre & Caceras 2013 Eco Lett 16: 951-963 doi: 10.1111/ele.12141

Choice of dissimilarity metrics affects distances

	ASV 1	ASV2	ASV3
Plot 1	0	1	1
Plot 2	1	0	0
Plot 3	0	4	8

Distances for plots 1 and 2

$$\text{Euclidean} = \sqrt{(0-1)^2+(1-0)^2+(1-0)^2} = 1.732$$

$$\text{Bray-Curtis} = \frac{(|0-1|+|1-0|+|1-0|)}{(2+1)} = 0.100 * 100 = 100\% \text{ dissimilar}$$

Distances for plots 1 and 3

$$\text{Euclidean} = \sqrt{(0-0)^2+(1-4)^2+(1-8)^2} = 7.615$$

$$\text{Bray-Curtis} = \frac{(|0-0|+|1-4|+|1-8|)}{(2+12)} = 0.714 * 100 = 71.4\% \text{ dissimilar}$$

Choice of dissimilarity metrics affects distances

	ASV 1	ASV2
Plot 1	1	0
Plot 2	1	1
Plot 3	10	0
Plot 4	10	10

Example calcs for Plot3 & Plot4

Euclidean

$$=\sqrt{(10-10)^2+(10-0)^2} = 10$$

Bray-Curtis

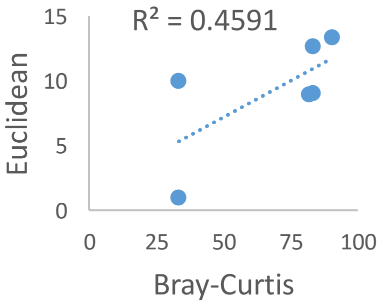
$$= \frac{(|10-10|+|10-0|)}{(10+20)} = 0.333 * 100 = 33.3\%$$

Bray Curtis Distance Matrix

	Plot 1	Plot 2	Plot 3	Plot 4
Plot 1	0			
Plot 2	33.3	0		
Plot 3	81.8	83.3	0	
Plot 4	90.5	83.3	33.3	0

Euclidean Curtis Distance Matrix

	Plot 1	Plot 2	Plot 3	Plot 4
Plot 1	0			
Plot 2	1.0	0		
Plot 3	9.0	9.1	0	
Plot 4	13.4	12.7	10.0	0



Phylogenetic Beta diversity

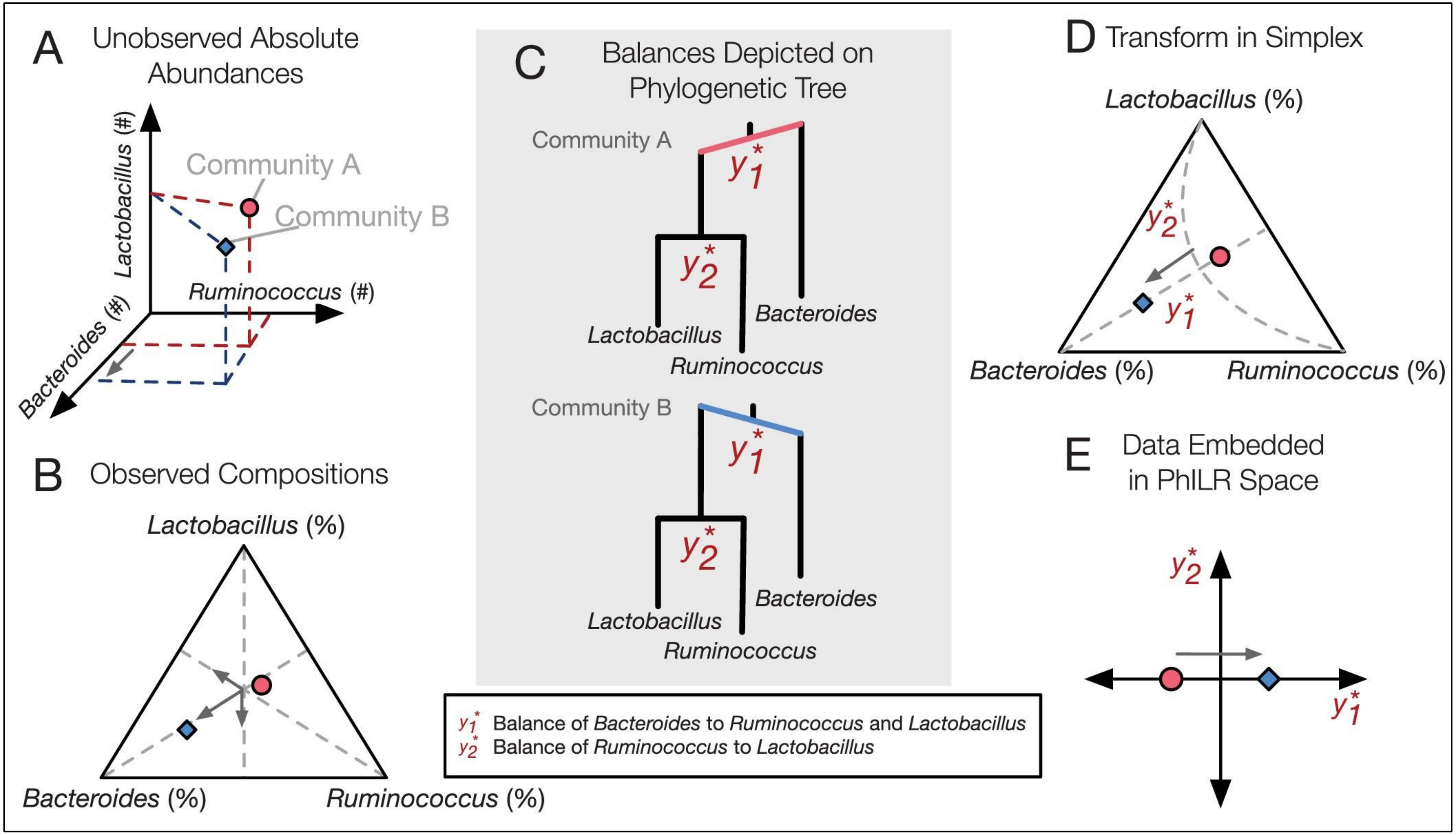
- Data transformation: PhILR
- Distances: Unifrac (unweighted and weighted)
- Distance + ordination: DPCoA

Phylogenetic Isometric Log Ratio (PhILR) Transformation for Compositional Data

- ILR transforms represent the ratio (“balance”) of relative abundances on each side of a binary partition
- PhILR defines those partitions by the phylogeny
  - For each internal node, transforms data as the log ratio of the geometric mean relative abundances of taxa in the two clades descending from that node
  - Uses unit-length branches, but can include branch length weights
  - Option to include weights for taxa abundances (used to down-weight influence of taxa with many zero or near-zero counts)
- Results in coordinates that capture evolutionary relationships between clades
- Use with Euclidean distances for PCoA

Silverman et al. 2017 eLife DOI: 10.7554/eLife.21887.001

Phylogenetic Isometric Log Ratio Transformation for Compositional Data



y_1^* Balance of *Bacteroides* to *Ruminococcus* and *Lactobacillus*
y_2^* Balance of *Ruminococcus* to *Lactobacillus*

Silverman et al. 2017 eLife DOI: 10.7554/eLife.21887.001

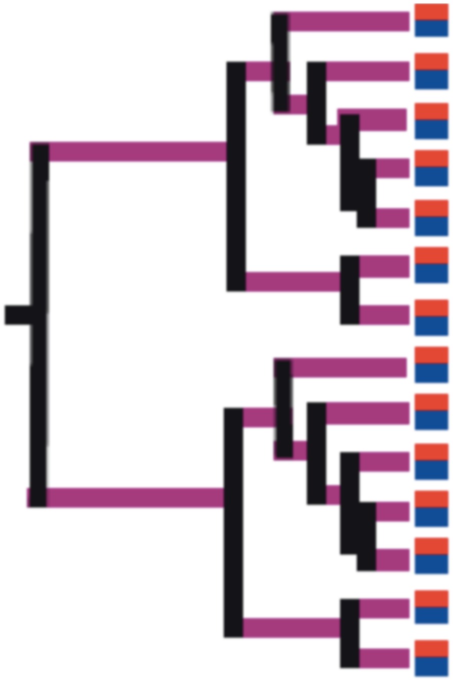
Phylogeny-aware distances

- Unifrac – unweighted
  - Phylogenetic extension of Jaccard index
  - Qualitative – based on presence/absence
  - Defines distances between pairs of samples as proportion of branch lengths unique to one sample or the other
  - Relies more heavily on shallow tree structure (tips)
  - Sensitive to sequencing depth (rarefaction used to get around this problem, but sensitive to specific rarefactions)
- Unifrac – weighted
  - Quantitative
  - Each branch length is weighted by the difference in proportional abundance of taxa between two samples
  - Relies more heavily on deep structure of the tree

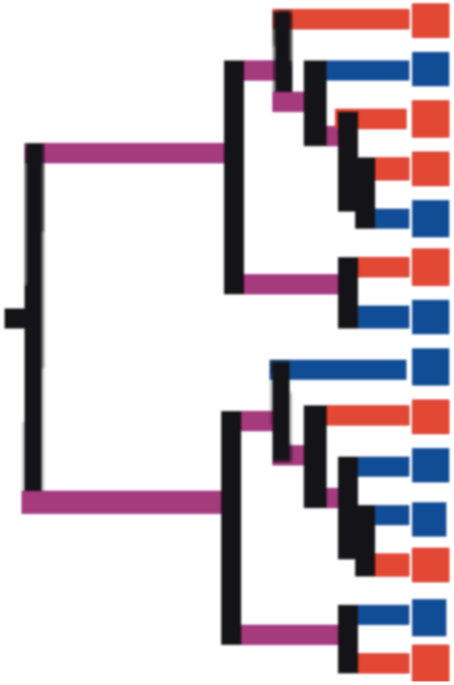
Wong et al. 2016 PLoS1 DOI: 10.1371/journal.pone.0161196; Fukuyama 2019 Genome Bio DOI: 10.1186/s13059-019-1735-y

Unweighted UniFrac

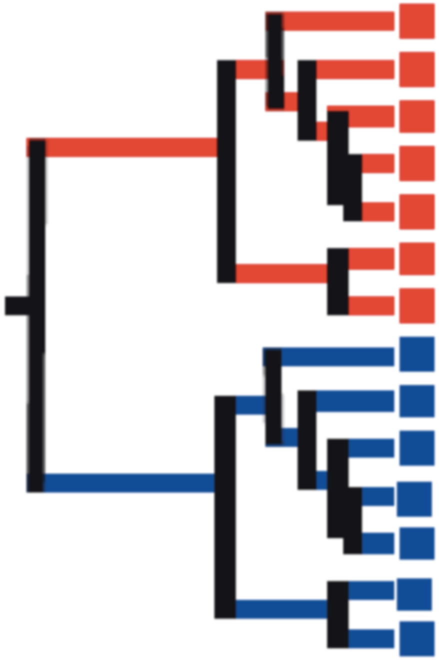
Dist(x, y) = (🟥 + 🟦) / (🟥 + 🟦 + 🟪)



UF = (0+0)/(0+0+26) = **0**
All branches shared
(identical communities)



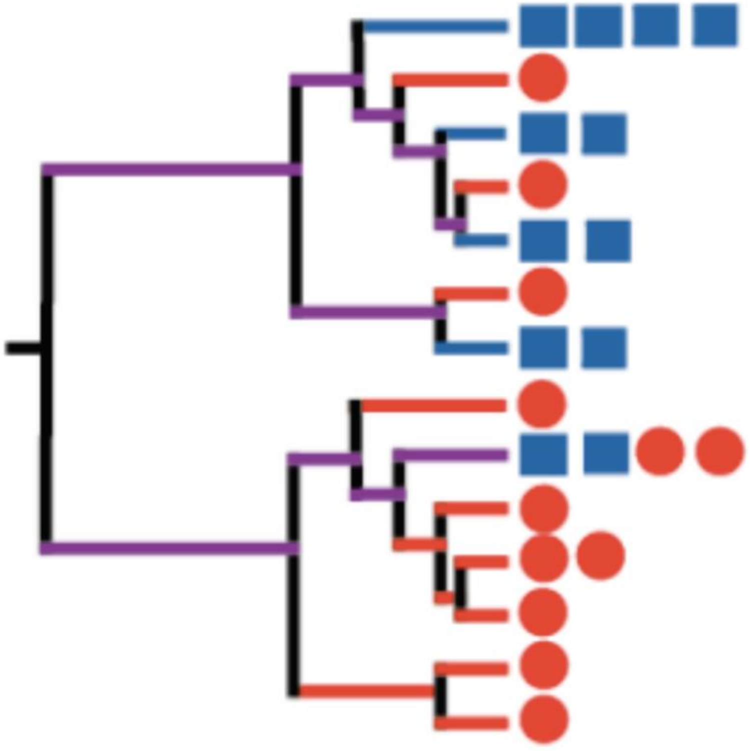
UF = (7+7)/(7+7+12) = **0.53**
Approx. half of branches shared
(similar communities)



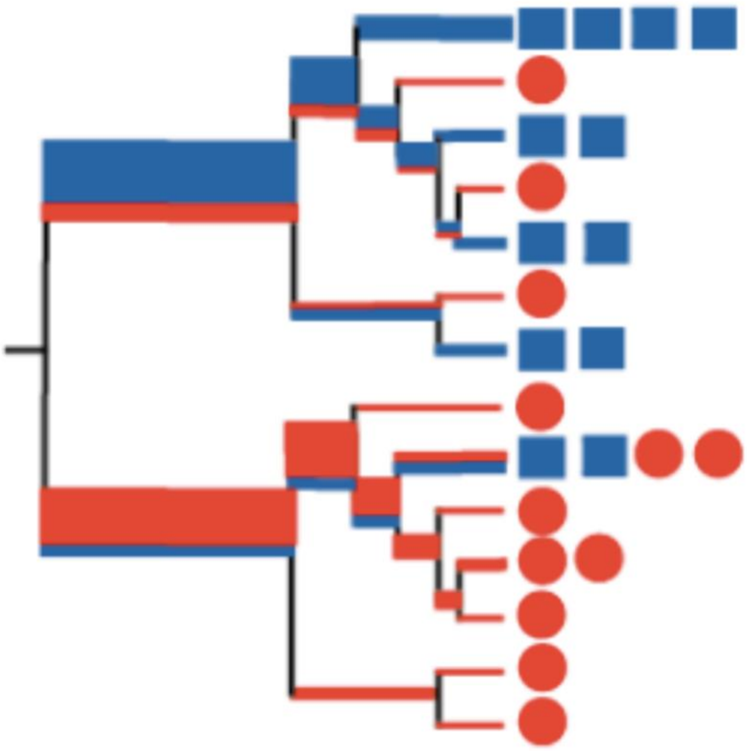
UF = (13+13)/(13+13+0) = **1**
No branches shared
(unique communities)

Unweighted vs. weighted UniFrac

Based on presence/absence only



Weights branch lengths by different abundances of red/blue



Note that there are multiple approaches to weighting – if you decide to use this approach for your data, be sure to read up on the latest methods at the time of your analysis.

Additional phylogenetic distance options

- Generalized Unifrac (Chen et al. 2012)
    - https://academic.oup.com/bioinformatics/article/28/16/2106/324465
    - https://cran.r-project.org/web/packages/GUniFrac/index.html
- Information Unifrac, Ratio Unifrac (Wong et al. 2016)
    - https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161196
    - https://search.r-project.org/CRAN/refmans/abdiv/html/unifrac.html
- Evodiss (Pavoine et al. 2016)
    - https://onlinelibrary.wiley.com/doi/full/10.1111/oik.03262
    - https://search.r-project.org/CRAN/refmans/adiv/html/evodiss.html

Ordination

- Goal is to represent sample and species relationships in low-dimensional space
- Distance between points in the ordination reflects the underlying distance measure and are proportional to their dissimilarity
- Each axis represents some amount of variation in the data

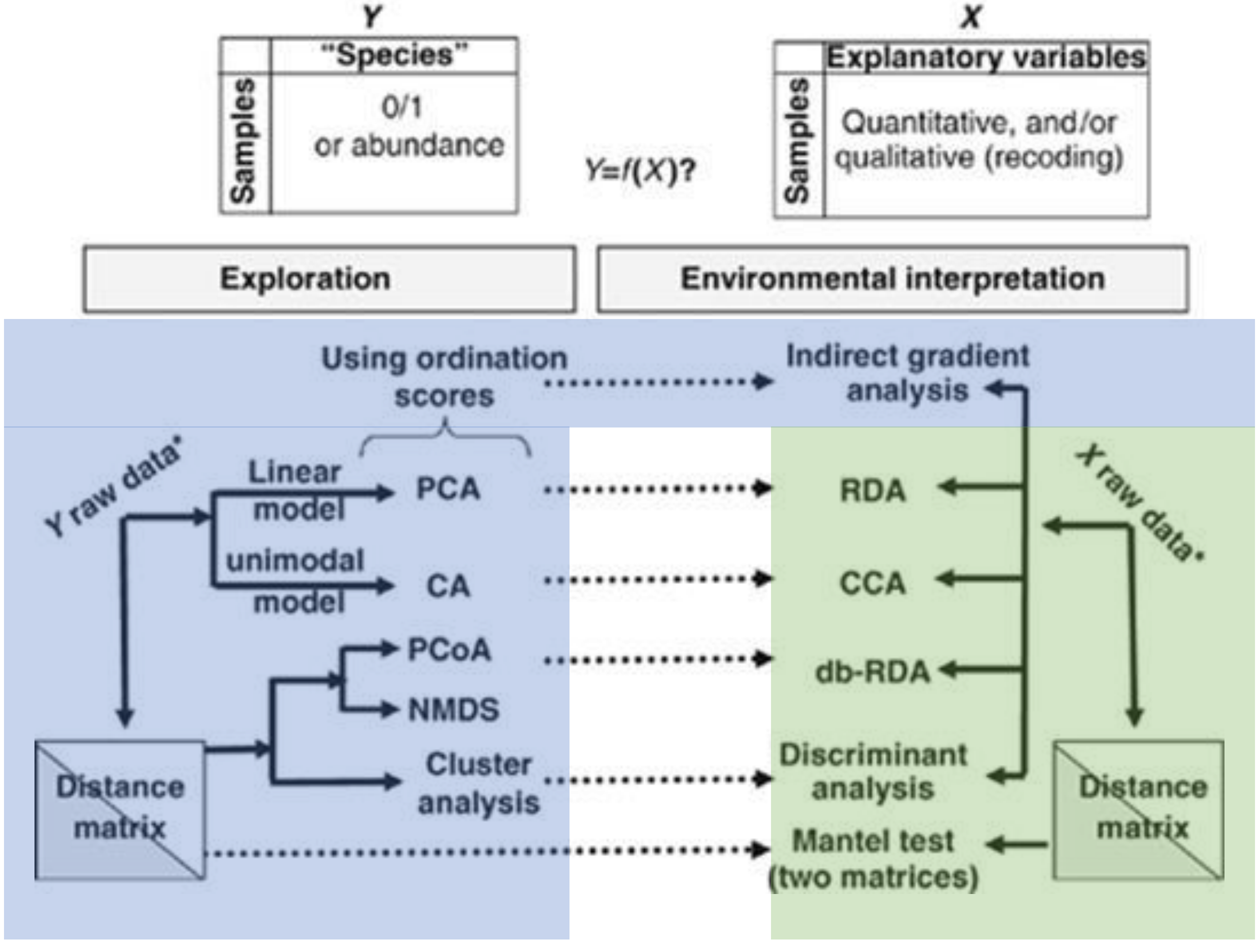
- Ordination is on samples, species can be overlain
- Can correlate or constrain ordination axes with environmental variables

- Primarily used for data reduction, exploration, and visualization
- Ordination is NOT a statistical test of significant differences among treatments

Constrained vs. Unconstrained

Unconstrained (indirect gradient): includes the species table, and resulting axes can be correlated with environmental variables

Constrained (direct gradient): considers only the variation in the species table that can be explained by environmental variables



Ramette (2007) FEMS Microbiology Ecology 62: 142-160 doi: 10.1111/j.1574-6941.2007.00375.x

Common Unconstrained Ordination Methods

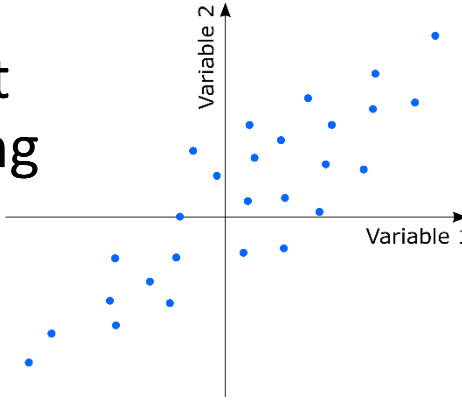
- Principal components analysis (PCA)

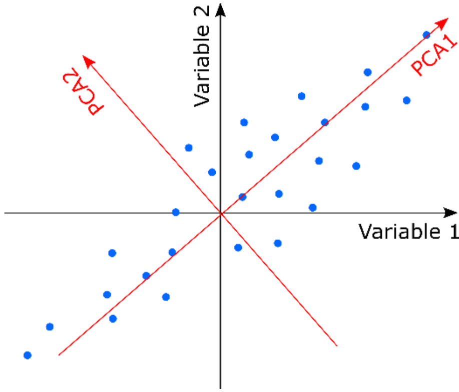
- Principal coordinates analysis (PCoA)
    - metric multidimensional scaling (MMDS)

- Non-metric multidimensional scaling (NMDS)

PCA

- Used for data reduction: reduce n cases by p variables to synthetic variables (axes) while preserving Euclidean distance – provides unique solution
- Finds successive orthogonal (uncorrelated) axes that have the strongest linear correlation structure among variables
- Requires linearity and multivariate normality
- All data must be in same units or standardized to be unitless
- Highly affected by outliers
- Heterogeneous data results in horseshoe pattern (ends of gradients pulled together) because PCA interprets shared zeros as a positive relationship
- Not useful for microbial community data, works for environmental or trait data





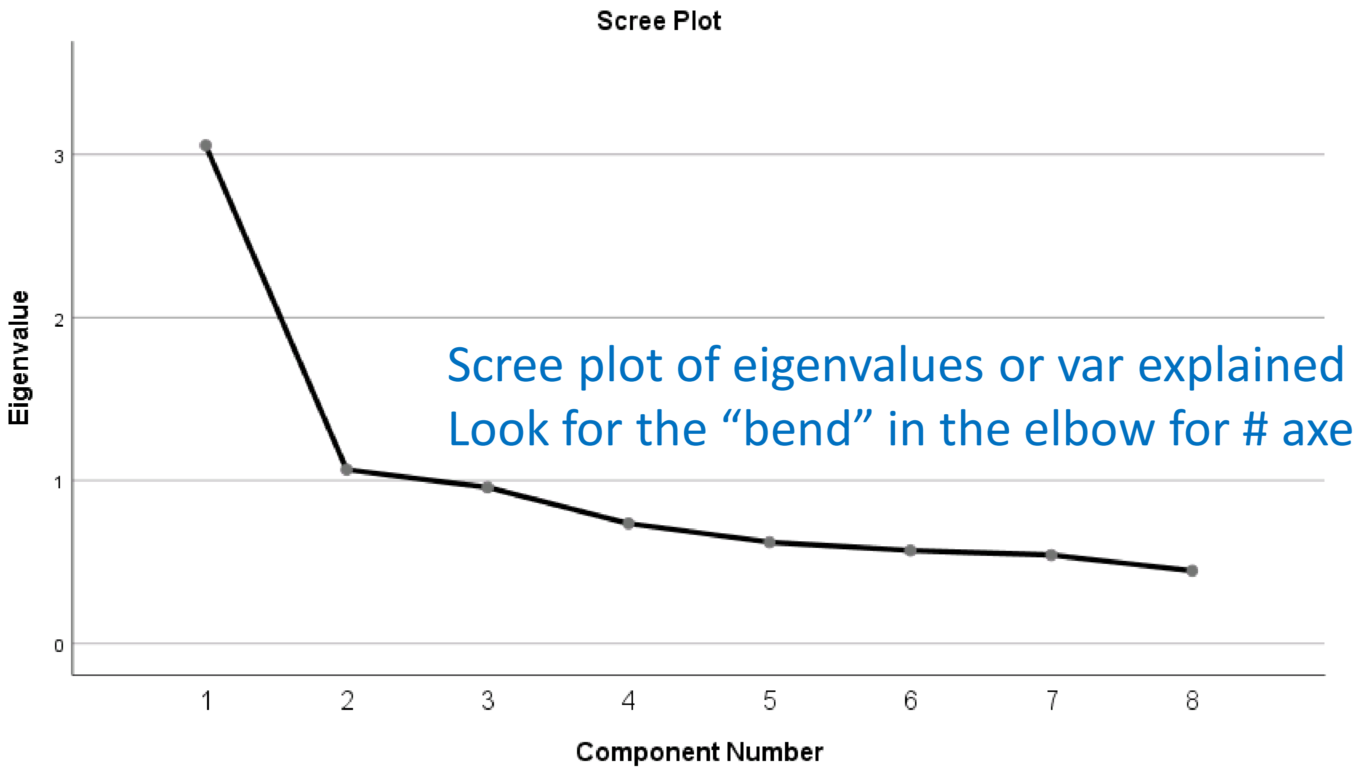
PCoA

- Preserves Euclidean distances between samples
- Finds successive orthogonal axes that best explain variability in the data
- Solution is unique and will always be the same

- Works with any dissimilarity measure
- Can handle quantitative, qualitative, or mixed variables
- Samples with high variability can strongly influence the solution

- Same as PCA when Euclidean distances are used

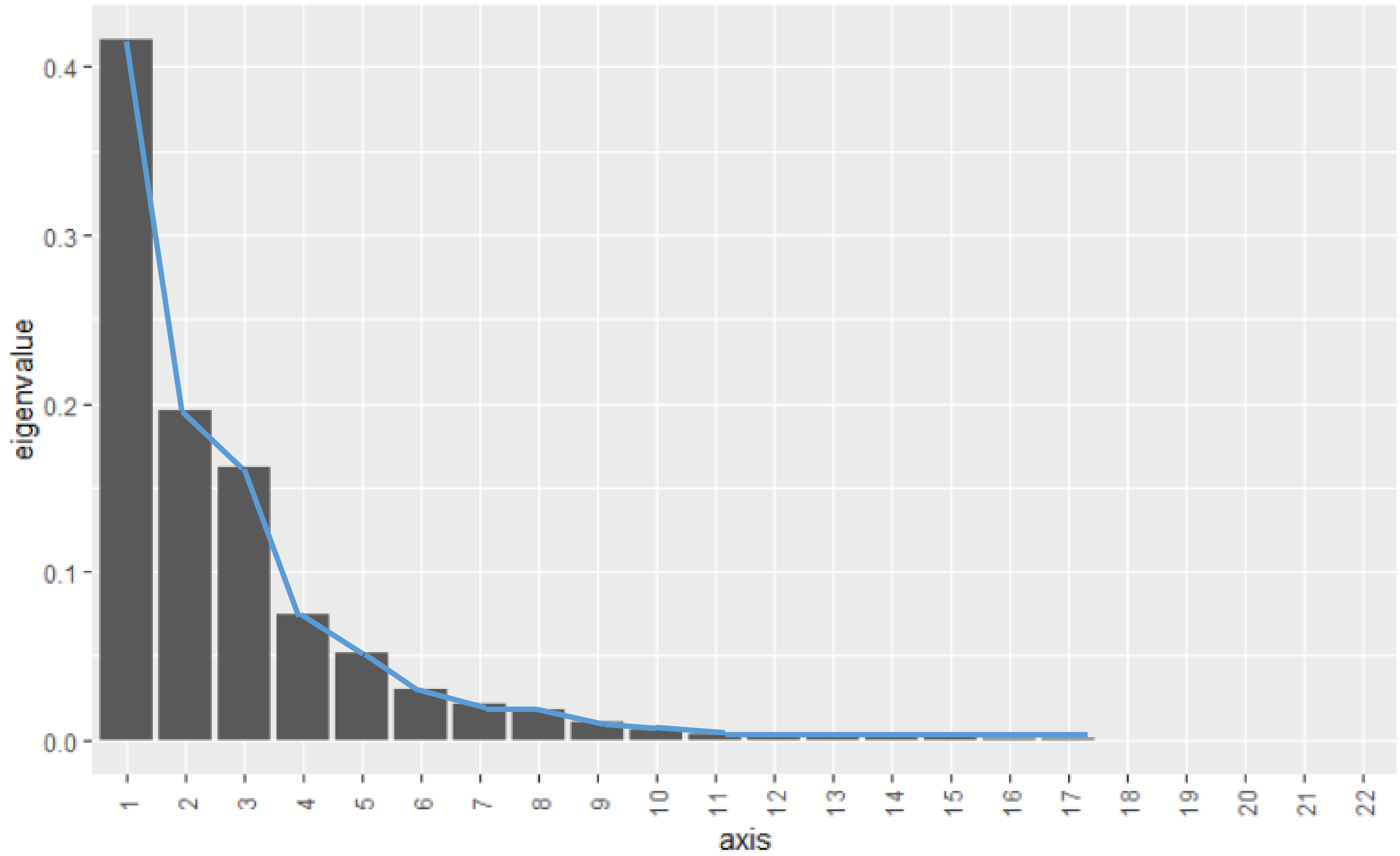
PCoA and PCA – fit and # axes



Scree plot of eigenvalues or var explained by PC
Look for the “bend” in the elbow for # axes

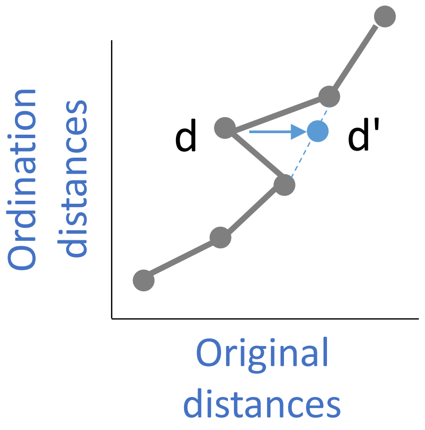
Total variance explained also used as a criterion

PCoA and PCA – fit and # axes

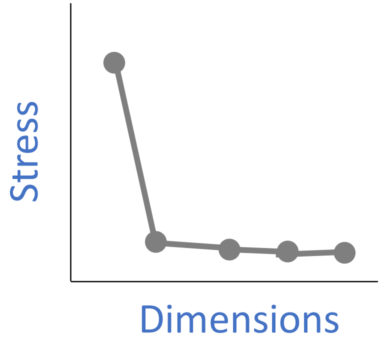


NMDS

- Non-parametric – uses ranked distances
- Can handle data that are non-normal, non-linear, heterogeneous, and have many zeros
- Iteratively searches for axes (k dimensions) that minimize the stress of the configuration
    - Stress = goodness of fit; departure from monotonicity in the original distances vs. in the ordination space (i.e., how well is the original preserved?)



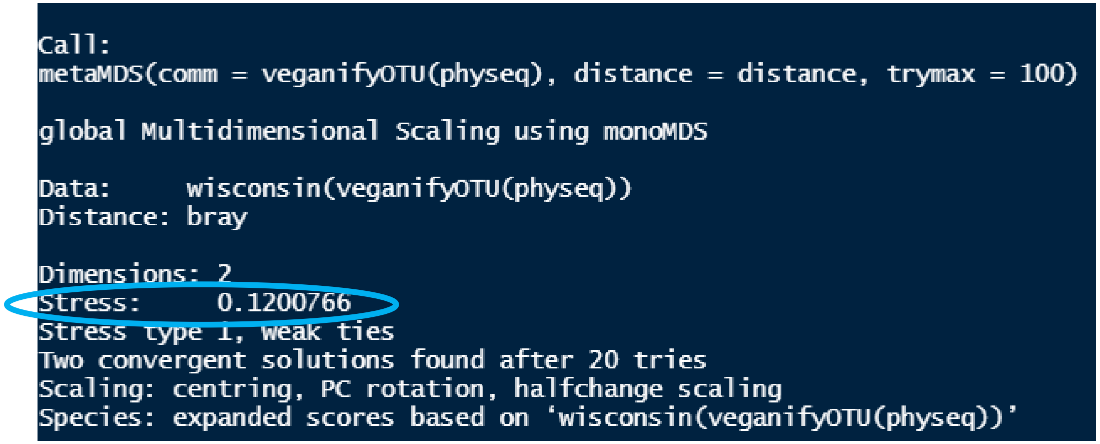
- Solution depends on starting configuration
    - Random iterations vs. prior ord start
    - Might be local not global



- Axes not necessarily orthogonal
- Dissimilarities can be distorted by ranks

NMDS – rules of thumb

Stress	Representation
< 0.05	Excellent
< 0.10	Good
< 0.20	Acceptable
> 0.20	Unsatisfactory - rerun



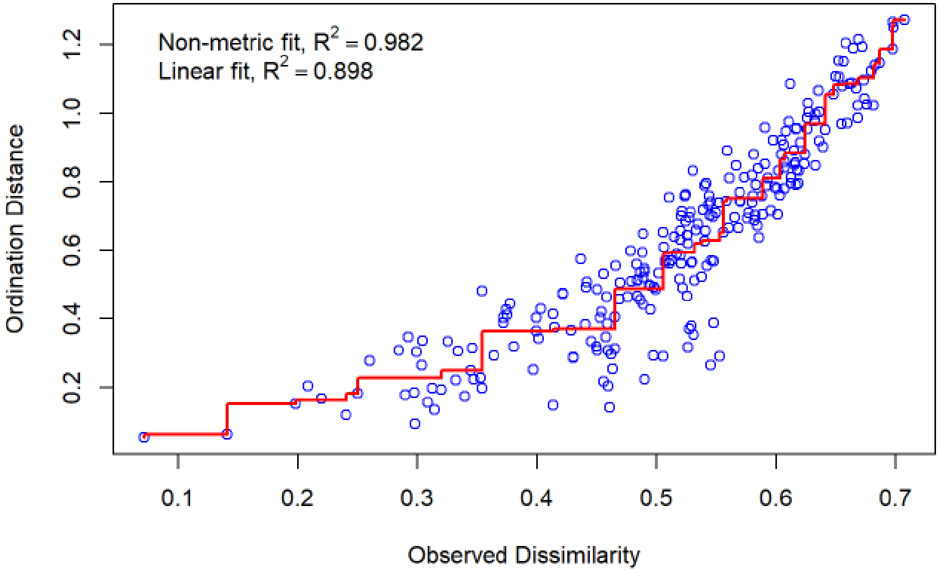
```
Call:
metaMDS(comm = veganifyOTU(physeq), distance = distance, trymax = 100)

global Multidimensional Scaling using monoMDS

Data:     wisconsin(veganifyOTU(physeq))
Distance: bray

Dimensions: 2
Stress:     0.1200766
Stress type 1, weak ties
Two convergent solutions found after 20 tries
Scaling: centring, PC rotation, halfchange scaling
Species: expanded scores based on 'wisconsin(veganifyOTU(physeq))'
```

Also need good fit between ordination distances and true dissimilarities ($r^2 > 0.9$)



High stress or poor fit?

- Increase # of axes (k)
- Rerun with new start configuration
- Use different distance metric or transformation

Double Principal Coordinates Analysis (DPCoA)

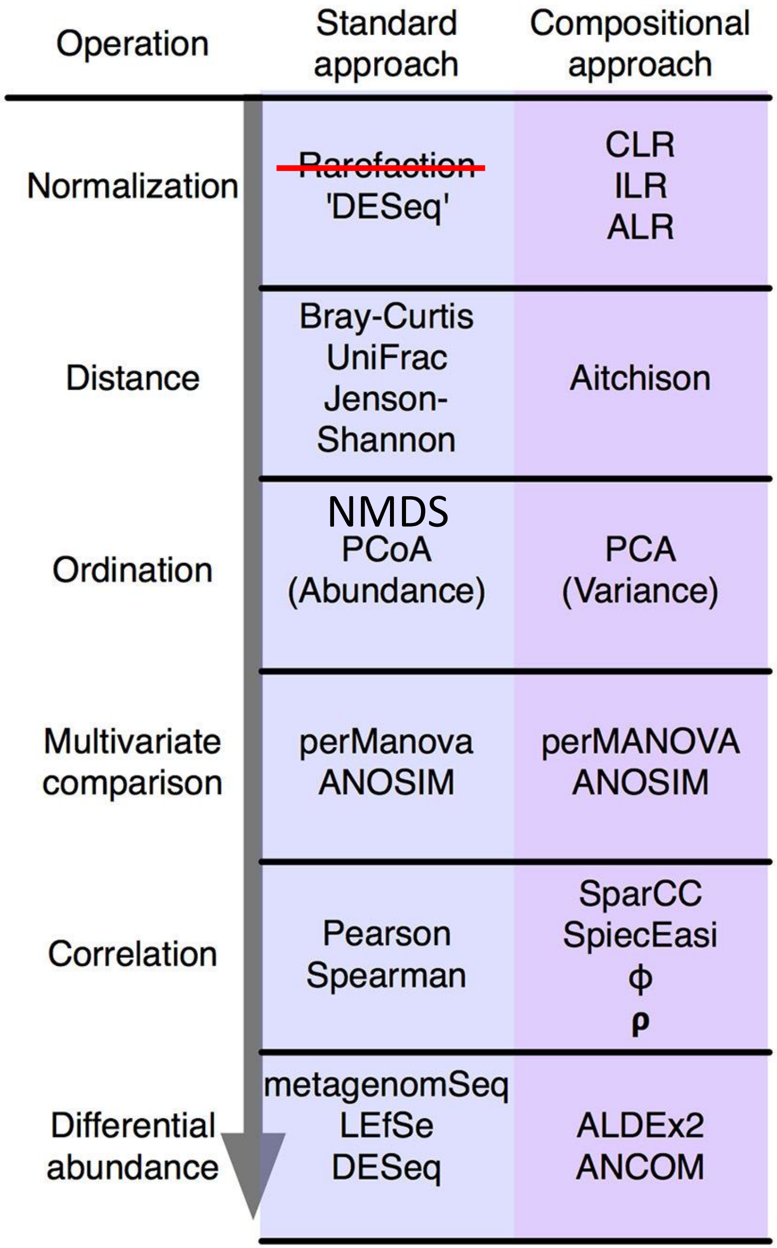
- Aims to obtain low-dimensional representation of taxa abundance data accounting for relationships among taxa
- Based on Rao’s diversity index
  - Taxa first positioned in high-dim space so that distances match phylogenetic distances
  - Each community/sample then positioned at center of a cloud of its taxa and weighted by taxa abundance in the community
  - PCoA is used to find low-dim representation of the species points for each community
- Similarity measure given by phylogeny (but can be other)
- Largely robust to noise in the data
- In generalized DPCoA, use of tuning parameter, r, gives full family of options (e.g., R::adaptiveGPCA)
  - r=0, DPCoA = Euclidean distance PCA (no phylog)
  - r=1, DPCoA = Rao’s DPCoA distance (phylog)

Pavoine et al. 2004 J Theor Bio DOI: 10.1016/j.jtbi.2004.02.014; Fukuyama 2019 Genome Bio DOI: 10.1186/s13059-019-1735-y

Other phylogenetic ordination approaches

- evoPCoA (Pavoine et al. 2016)
  - https://onlinelibrary.wiley.com/doi/full/10.1111/oik.03262
- Edge PCA (Matson & Evans 2013)
  - https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0056859

Recall contrast between standard and compositional approaches to analysis of beta diversity



Operation	Standard approach	Compositional approach
Normalization	~~Rarefaction~~ 'DESeq'	CLR ILR ALR
Distance	Bray-Curtis UniFrac Jenson-Shannon	Aitchison
Ordination	NMDS PCoA (Abundance)	PCA (Variance)
Multivariate comparison	perManova ANOSIM	perMANOVA ANOSIM
Correlation	Pearson Spearman	SparCC SpiecEasi ϕ ρ
Differential abundance	metagenomSeq LEfSe DESeq	ALDEx2 ANCOM

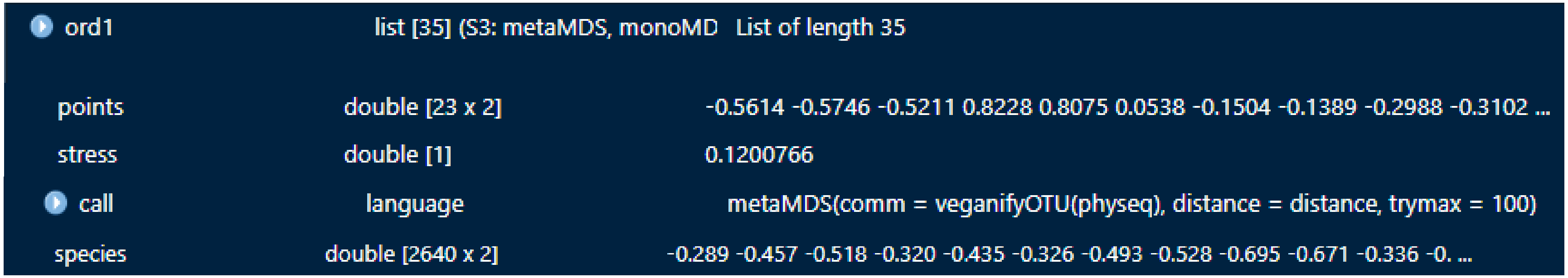
Gloor et al. 2017 Frontiers Micro 8:2224 doi: 10.3389/fmicb.2017.02224

Some practical issues

- Accessing results in list objects
- “Phylo” warnings
- Plotting options

Accessing results: lists

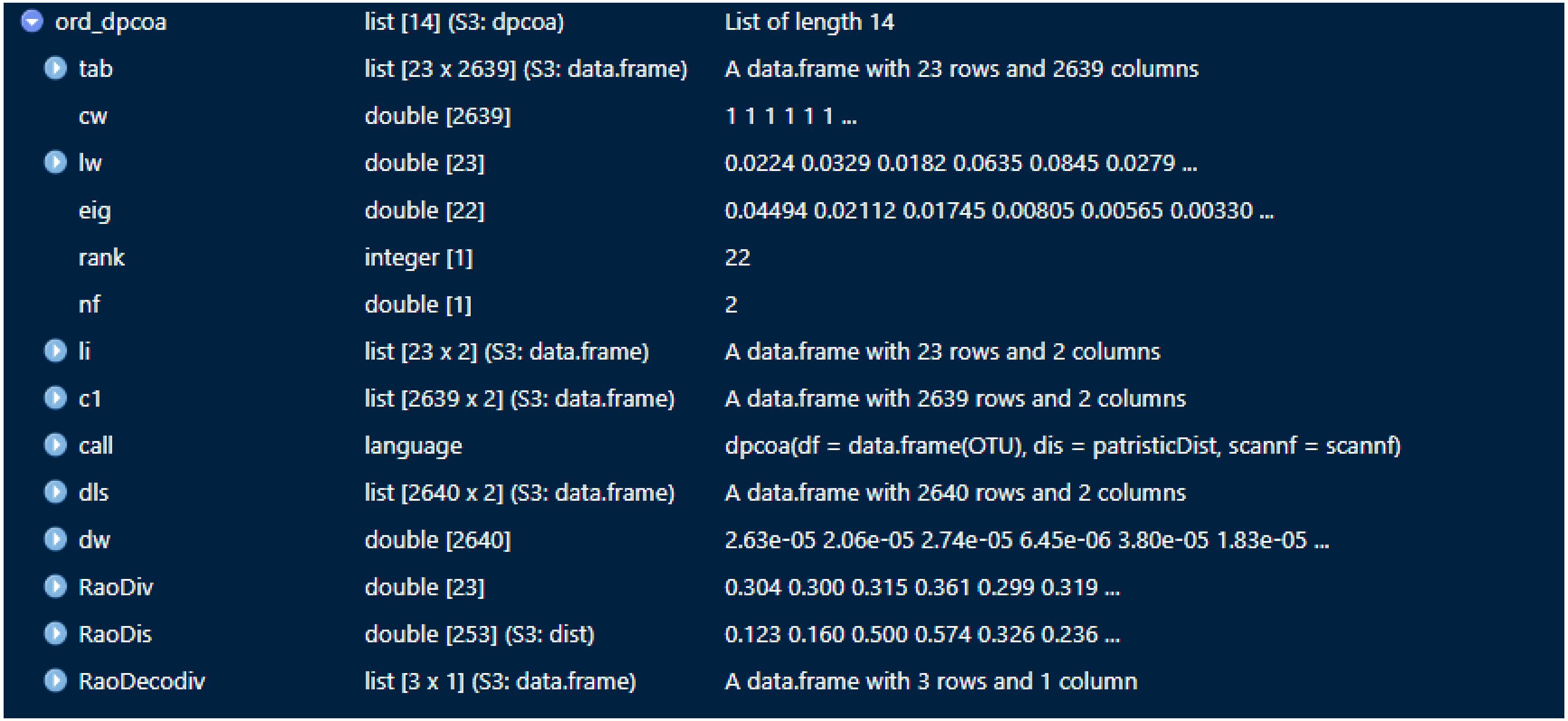
NMDS



```
ord1                  list [35] (S3: metaMDS, monoMD   List of length 35

  points            double [23 x 2]          -0.5614 -0.5746 -0.5211 0.8228 0.8075 0.0538 -0.1504 -0.1389 -0.2988 -0.3102 ...
  stress            double [1]               0.1200766
  call              language                 metaMDS(comm = veganifyOTU(physeq), distance = distance, trymax = 100)
  species           double [2640 x 2]        -0.289 -0.457 -0.518 -0.320 -0.435 -0.326 -0.493 -0.528 -0.695 -0.671 -0.336 -0. ...
```

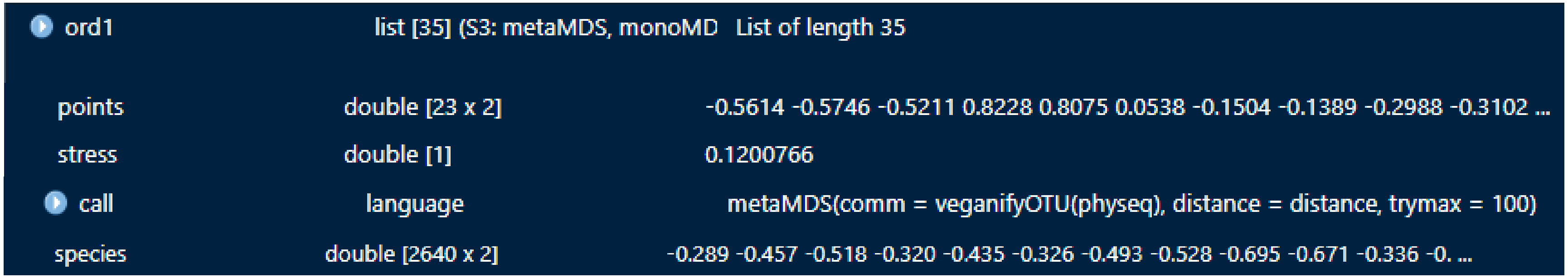
DPCoA



```
ord_dpcoa             list [14] (S3: dpcoa)                List of length 14
  tab                 list [23 x 2639] (S3: data.frame)    A data.frame with 23 rows and 2639 columns
  cw                  double [2639]                        1 1 1 1 1 1 ...
  lw                  double [23]                          0.0224 0.0329 0.0182 0.0635 0.0845 0.0279 ...
  eig                 double [22]                          0.04494 0.02112 0.01745 0.00805 0.00565 0.00330 ...
  rank                integer [1]                          22
  nf                  double [1]                           2
  li                  list [23 x 2] (S3: data.frame)       A data.frame with 23 rows and 2 columns
  c1                  list [2639 x 2] (S3: data.frame)     A data.frame with 2639 rows and 2 columns
  call                language                             dpcoa(df = data.frame(OTU), dis = patristicDist, scannf = scannf)
  dls                 list [2640 x 2] (S3: data.frame)     A data.frame with 2640 rows and 2 columns
  dw                  double [2640]                        2.63e-05 2.06e-05 2.74e-05 6.45e-06 3.80e-05 1.83e-05 ...
  RaoDiv              double [23]                          0.304 0.300 0.315 0.361 0.299 0.319 ...
  RaoDis              double [253] (S3: dist)              0.123 0.160 0.500 0.574 0.326 0.236 ...
  RaoDecodiv          list [3 x 1] (S3: data.frame)        A data.frame with 3 rows and 1 column
```

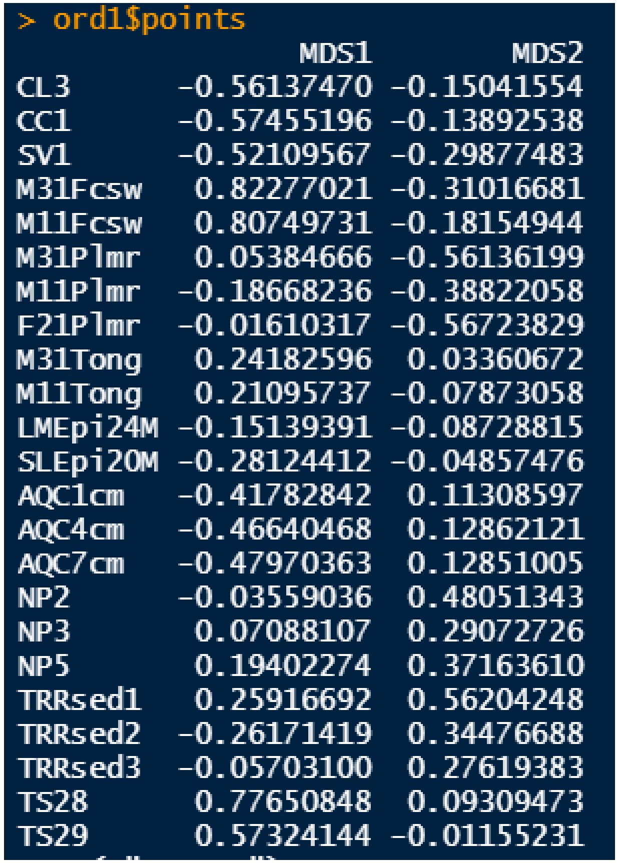
Accessing results: lists level 1

NMDS



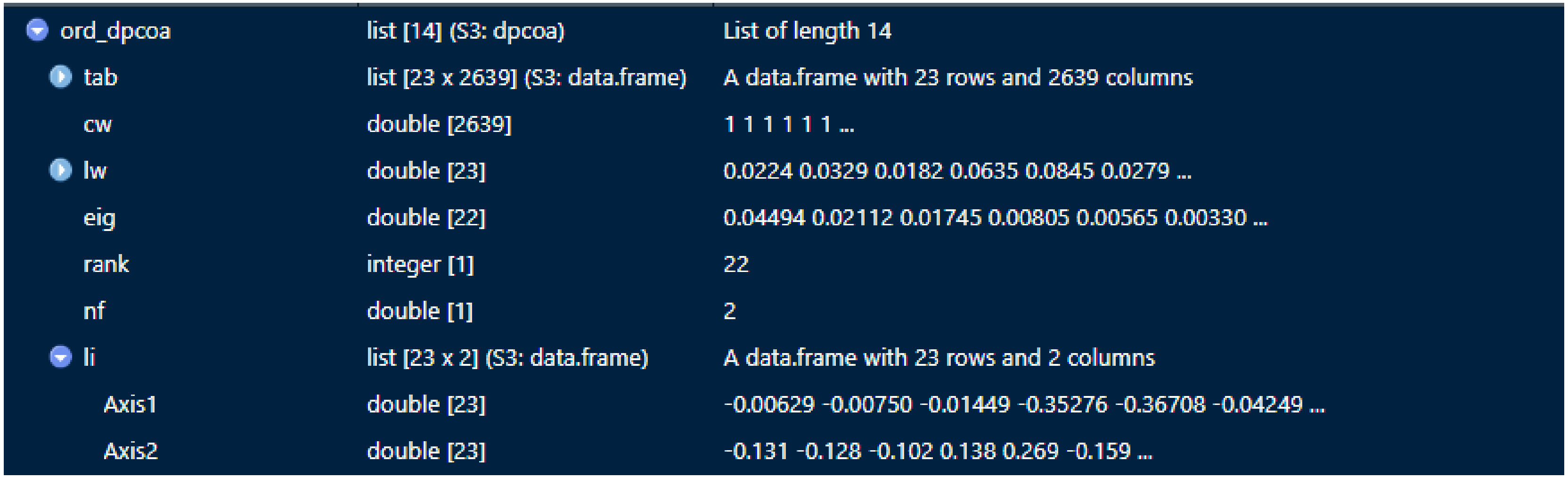
```
ord1        list [35] (S3: metaMDS, monoMD   List of length 35

  points    double [23 x 2]     -0.5614 -0.5746 -0.5211 0.8228 0.8075 0.0538 -0.1504 -0.1389 -0.2988 -0.3102 ...
  stress    double [1]          0.1200766
  call      language            metaMDS(comm = veganifyOTU(physeq), distance = distance, trymax = 100)
  species   double [2640 x 2]   -0.289 -0.457 -0.518 -0.320 -0.435 -0.326 -0.493 -0.528 -0.695 -0.671 -0.336 -0. ...
```

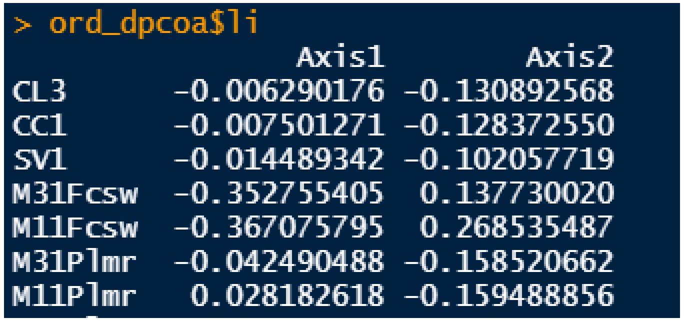


```
> ord1$points
                 MDS1        MDS2
CL3       -0.56137470 -0.15041554
CC1       -0.57455196 -0.13892538
SV1       -0.52109567 -0.29877483
M31Fcsw    0.82277021 -0.31016681
M11Fcsw    0.80749731 -0.18154944
M31Plmr    0.05384666 -0.56136199
M11Plmr   -0.18668236 -0.38822058
F21Plmr   -0.01610317 -0.56723829
M31Tong    0.24182596  0.03360672
M11Tong    0.21095737 -0.07873058
LMEpi24M  -0.15139391 -0.08728815
SLEpi20M  -0.28124412 -0.04857476
AQC1cm    -0.41782842  0.11308597
AQC4cm    -0.46640468  0.12862121
AQC7cm    -0.47970363  0.12851005
NP2       -0.03559036  0.48051343
NP3        0.07088107  0.29072726
NP5        0.19402274  0.37163610
TRRsed1    0.25916692  0.56204248
TRRsed2   -0.26171419  0.34476688
TRRsed3   -0.05703100  0.27619383
TS28       0.77650848  0.09309473
TS29       0.57324144 -0.01155231
```

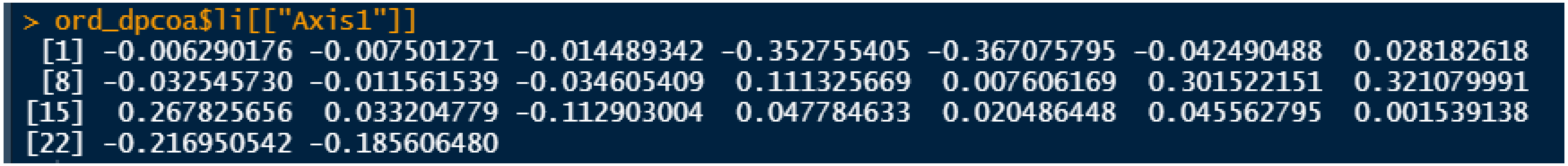
Accessing results: lists level 2



```
ord_dpcoa        list [14] (S3: dpcoa)               List of length 14
  tab            list [23 x 2639] (S3: data.frame)   A data.frame with 23 rows and 2639 columns
  cw             double [2639]                       1 1 1 1 1 1 ...
  lw             double [23]                         0.0224 0.0329 0.0182 0.0635 0.0845 0.0279 ...
  eig            double [22]                         0.04494 0.02112 0.01745 0.00805 0.00565 0.00330 ...
  rank           integer [1]                         22
  nf             double [1]                          2
  li             list [23 x 2] (S3: data.frame)      A data.frame with 23 rows and 2 columns
    Axis1        double [23]                         -0.00629 -0.00750 -0.01449 -0.35276 -0.36708 -0.04249 ...
    Axis2        double [23]                         -0.131 -0.128 -0.102 0.138 0.269 -0.159 ...
```

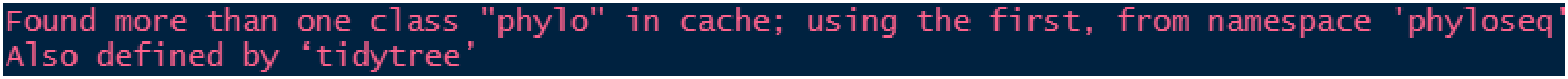


```
> ord_dpcoa$li
                Axis1        Axis2
CL3      -0.006290176 -0.130892568
CC1      -0.007501271 -0.128372550
SV1      -0.014489342 -0.102057719
M31Fcsw  -0.352755405  0.137730020
M11Fcsw  -0.367075795  0.268535487
M31Plmr  -0.042490488 -0.158520662
M11Plmr   0.028182618 -0.159488856
```



```
> ord_dpcoa$li[["Axis1"]]
 [1] -0.006290176 -0.007501271 -0.014489342 -0.352755405 -0.367075795 -0.042490488  0.028182618
 [8] -0.032545730 -0.011561539 -0.034605409  0.111325669  0.007606169  0.301522151  0.321079991
[15]  0.267825656  0.033204779 -0.112903004  0.047784633  0.020486448  0.045562795  0.001539138
[22] -0.216950542 -0.185606480
```

Phylo conflict warnings



```
Found more than one class "phylo" in cache; using the first, from namespace 'phyloseq'
Also defined by 'tidytree'
```

- This warning indicates that both phyloseq and tidytree share class “phylo”
  - shows up when you’re accessing the tree in the ps object
- Can be ignored here, but it’s a good idea to specify your package use when these arise
- In Rmd, can be suppressed in code chunk with message=FALSE

Many, many plotting options

- phyloseq::plot_ordination
- vegan::ordiplot
  - vegan::ordiellipse, vegan::ordihull, vegan::ordispider
- can also modify many plot options with ggplot2
  - ggplot2::geom_line, ggplot2::geom_point
- ggordiplots has lots of built-in options
- ggvegan is another option

- Good tutorial on diy versions in ggplot2:
  - https://rstudio-pubs-static.s3.amazonaws.com/694016_e2d53d65858d4a1985616fa3855d237f.html

Practice!

- Wk6_betadiv.html from GitHub

```
download.file(url = "https://rawgithubfilelocation",
destfile = ".../yourlocalfolder/filename")
```

- Working with GlobalPatterns data again
  - Filtered and VST transformed for class
- Two NMDS examples phyloseq and vegan+envfit
- Phylogentic ILR transform and PCoA
- PCoA with Unifrac distances - unweighted and weighted
- Coding exercises: clr transformed data, PCoA in phyloseq, PCoA in other package, DPCoA