

Metagenomics and metatranscriptomics workshop

Biodiversity analysis in R



Forestry and Agricultural
Biotechnology Institute

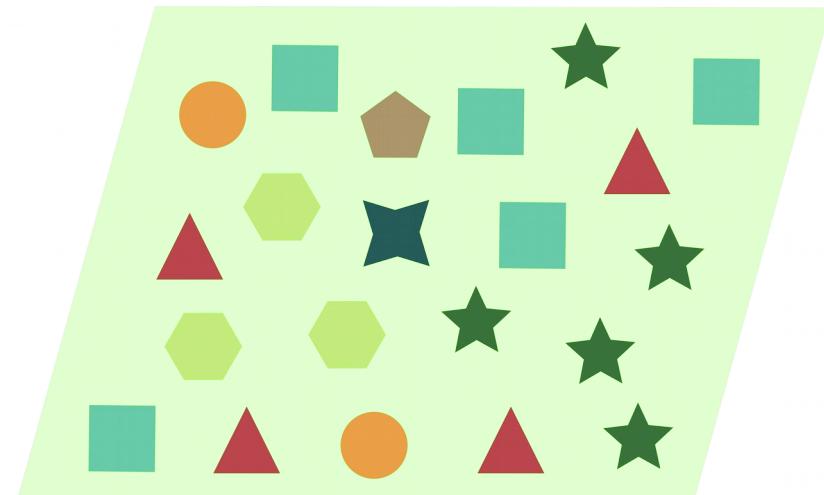


UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Biodiversity

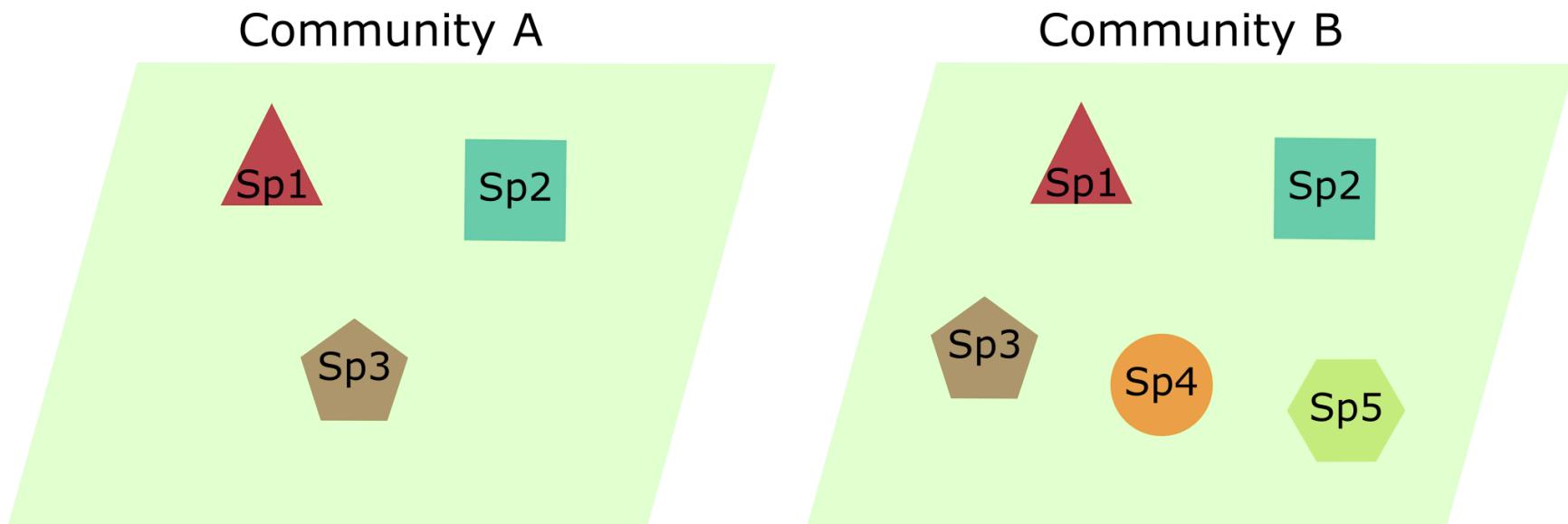
Variety of life, at all levels of organization

Genetic
Species
Ecosystems

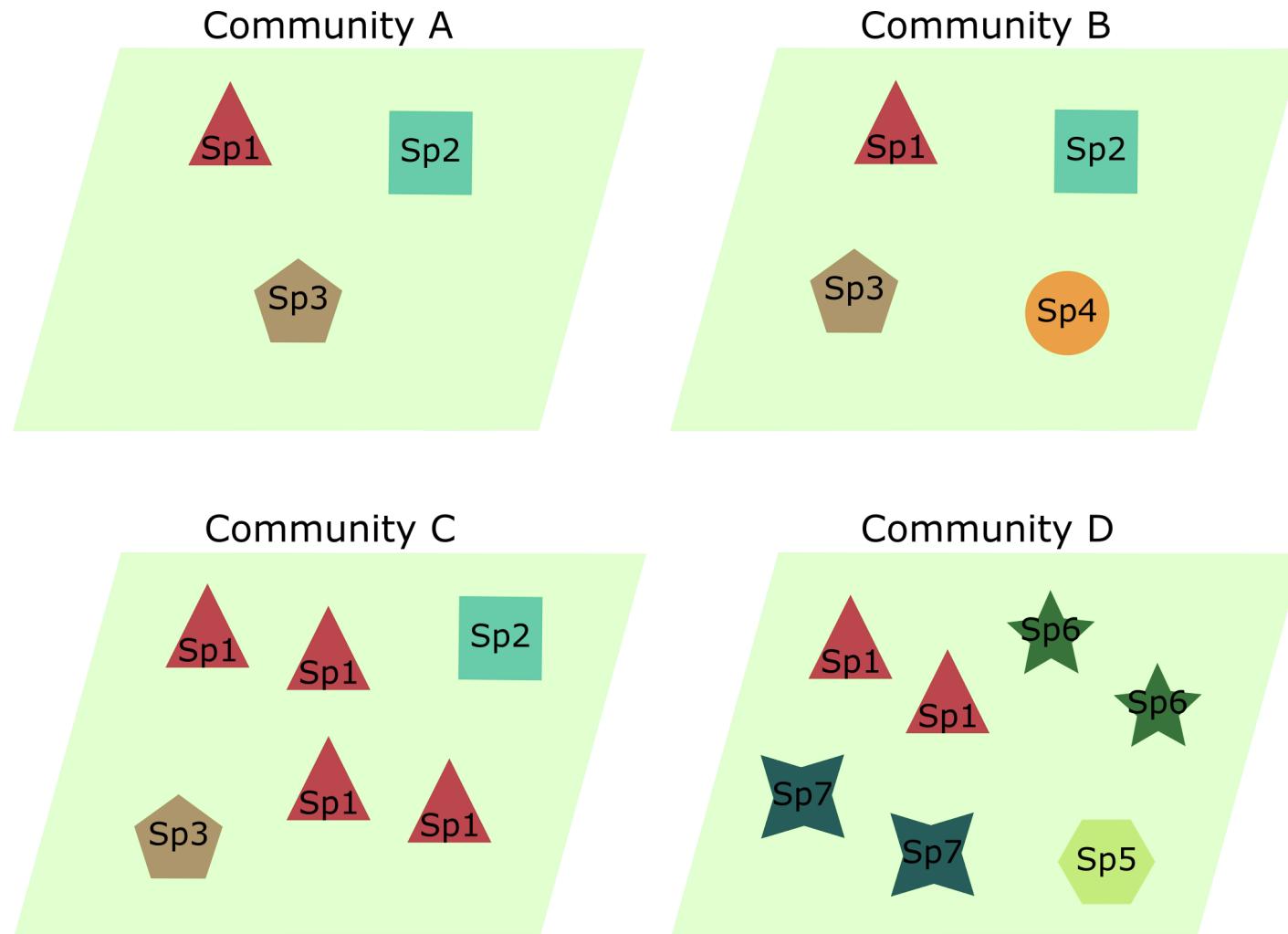


Number - Abundance - Proportions

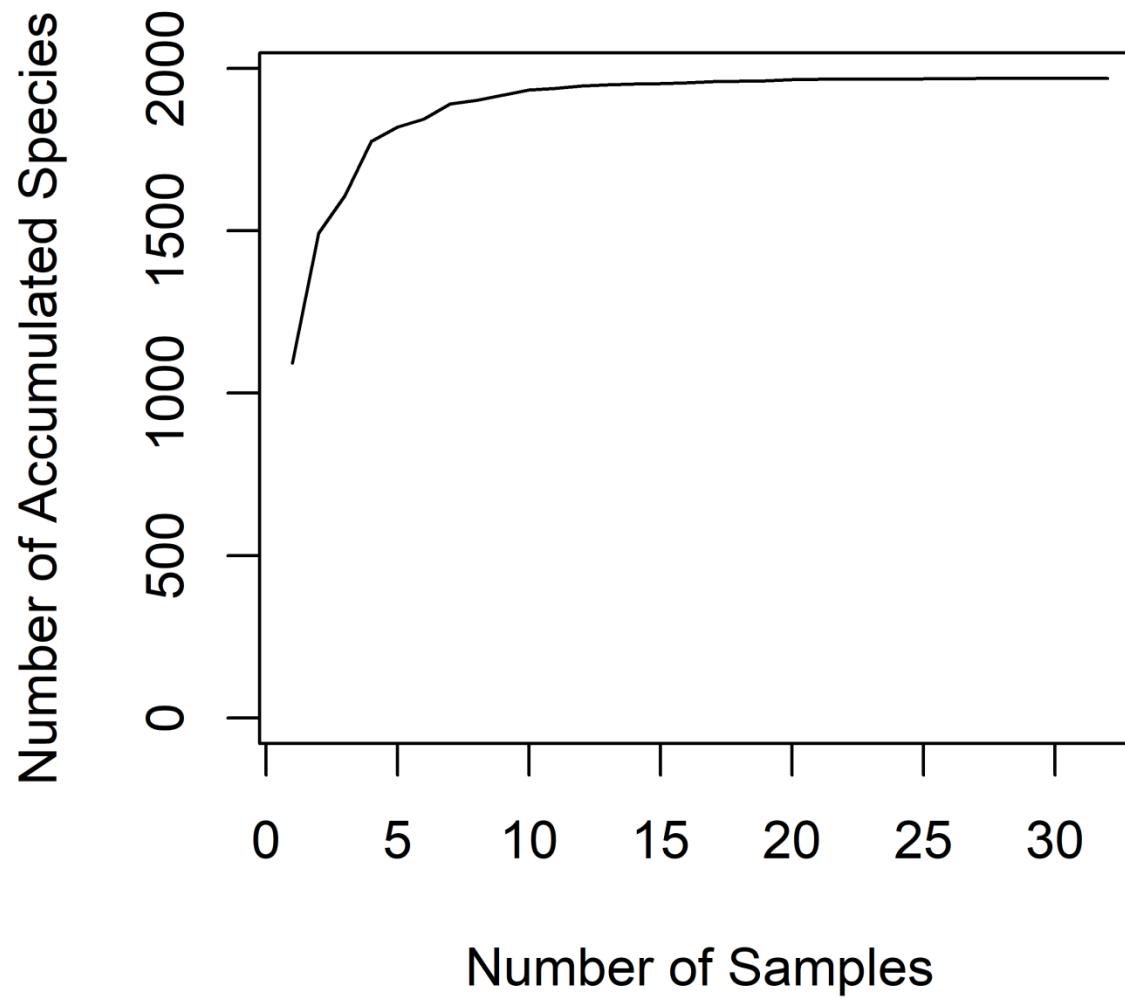
Species richness = # species



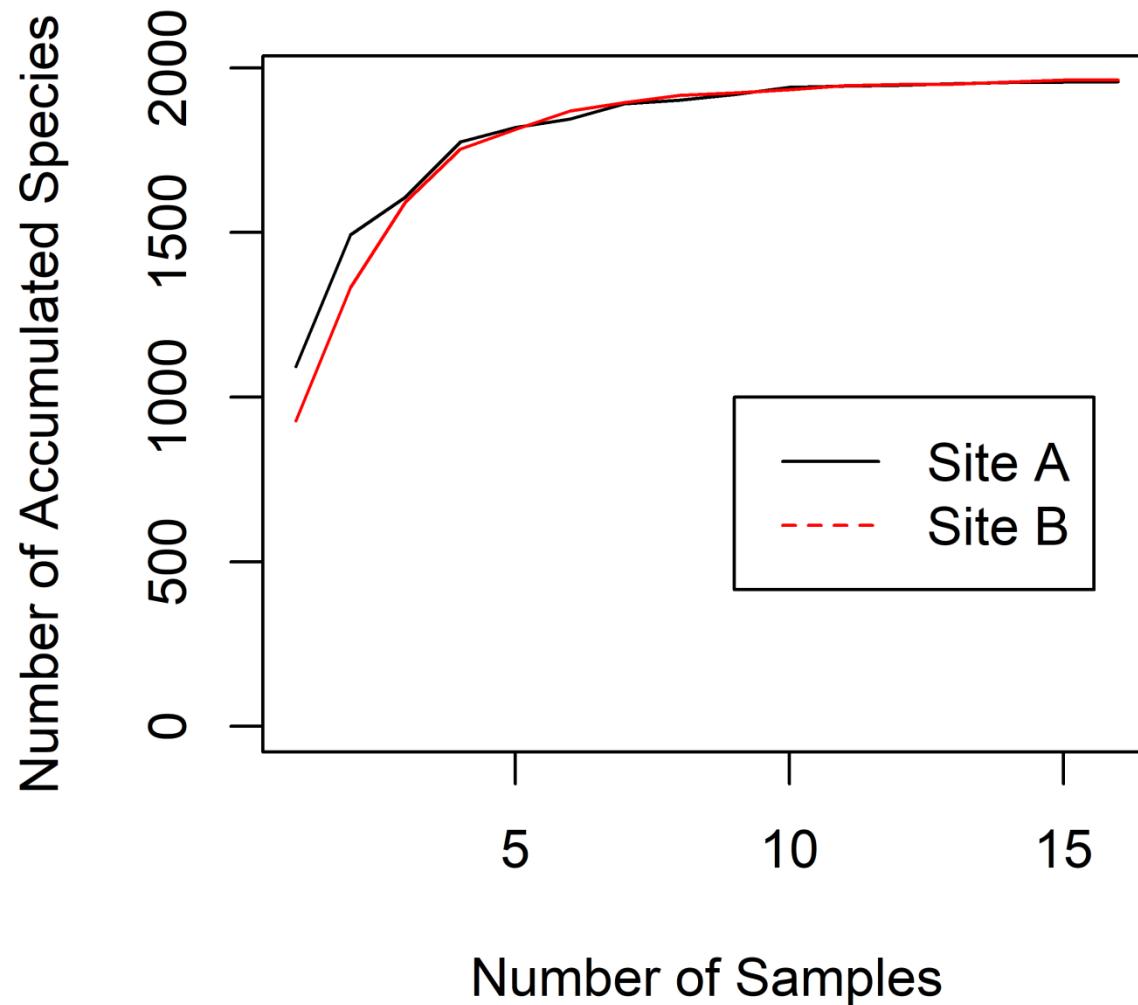
Species richness = # species



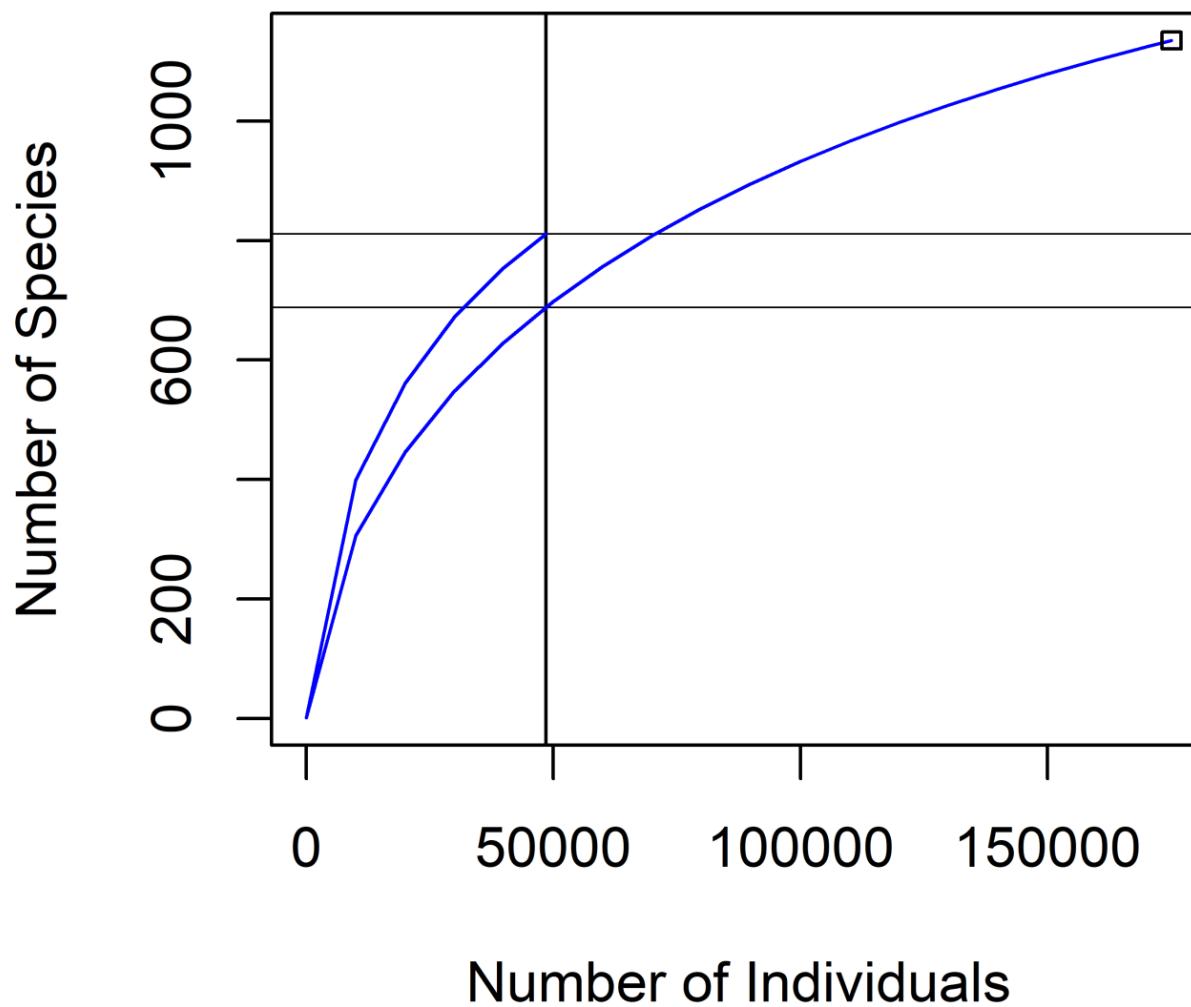
Species accumulation curve



Species accumulation curve

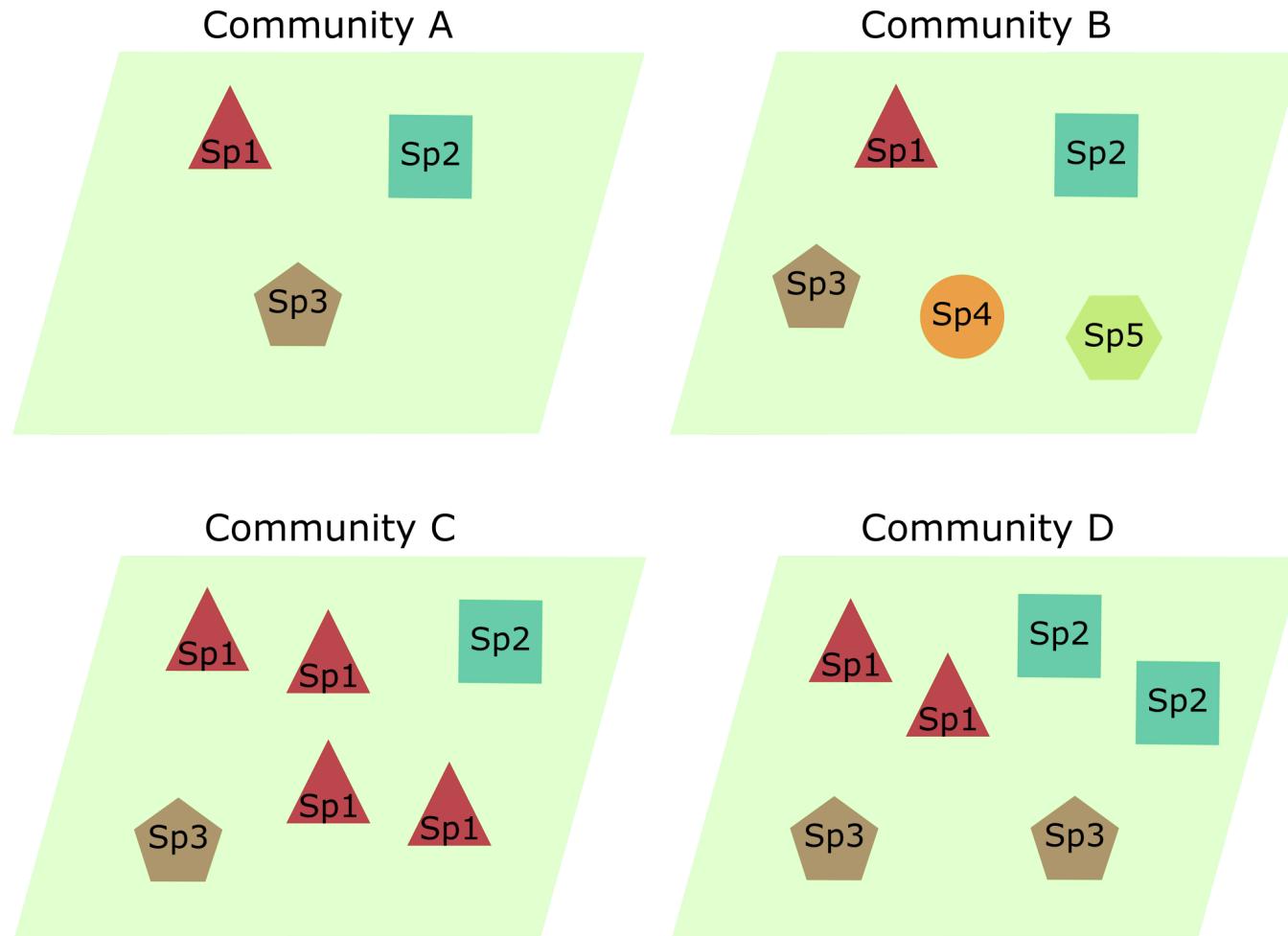


Rarefaction curve



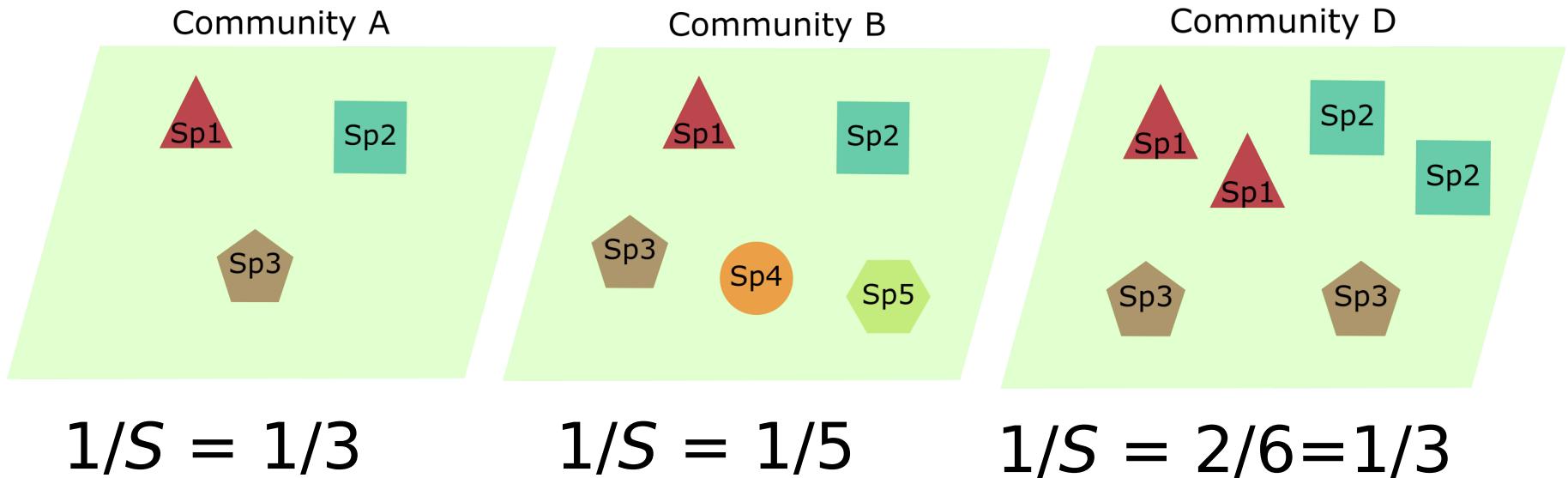
Species richness = # spp.

Evenness = equality # ind./sp.



Maximum evenness

Completely Homogeneous



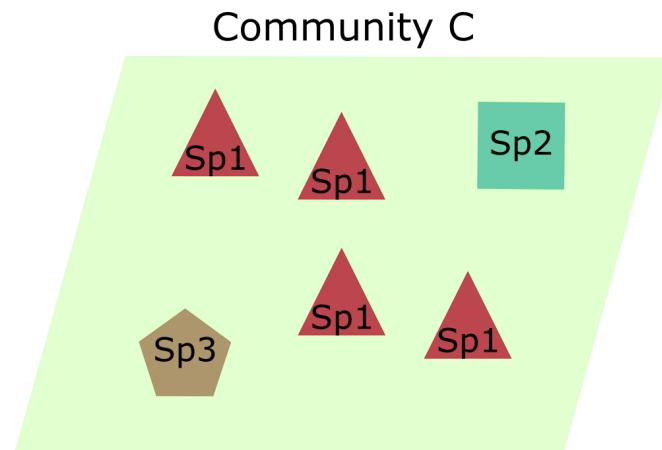
S - species richness

Minimum evenness

Completely Heterogeneous

Less frequent specie:
1 individual

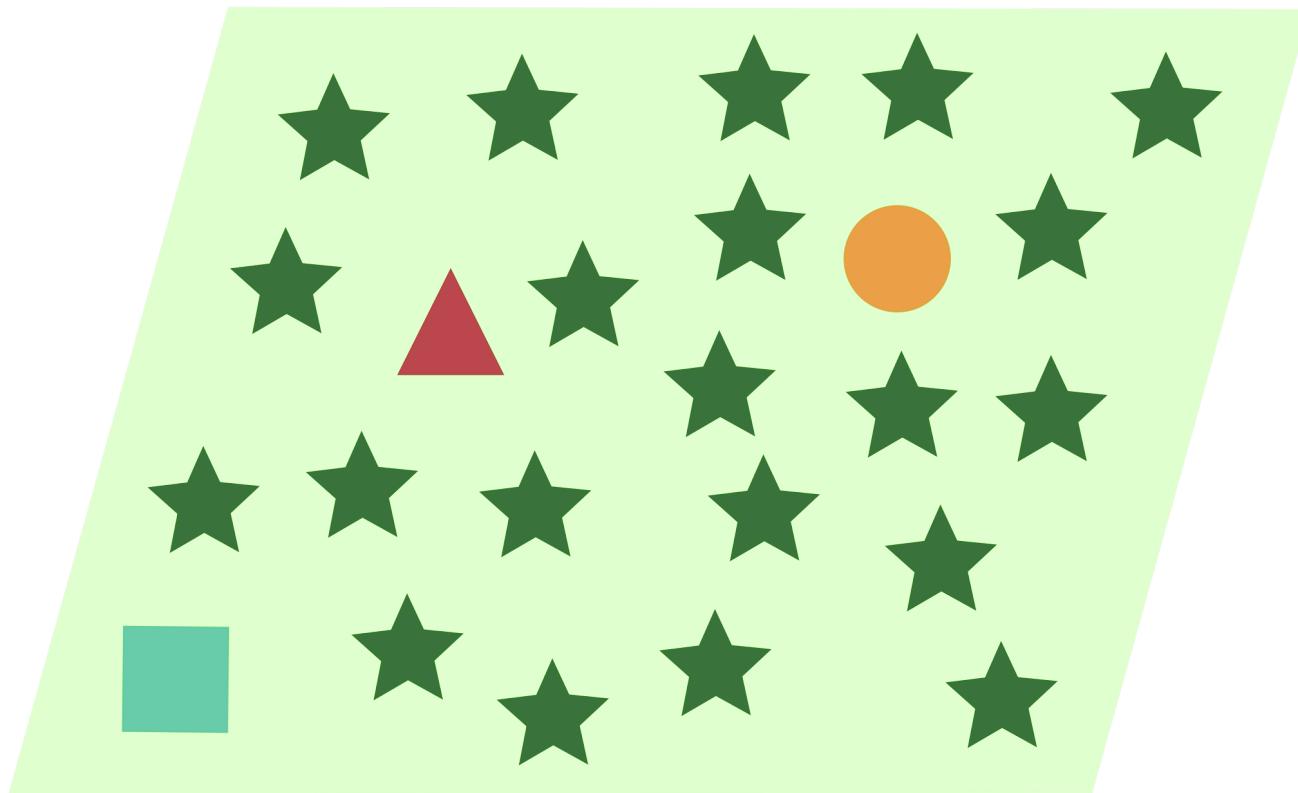
Dominant specie:
 $Tot-(S-1)$ individuals



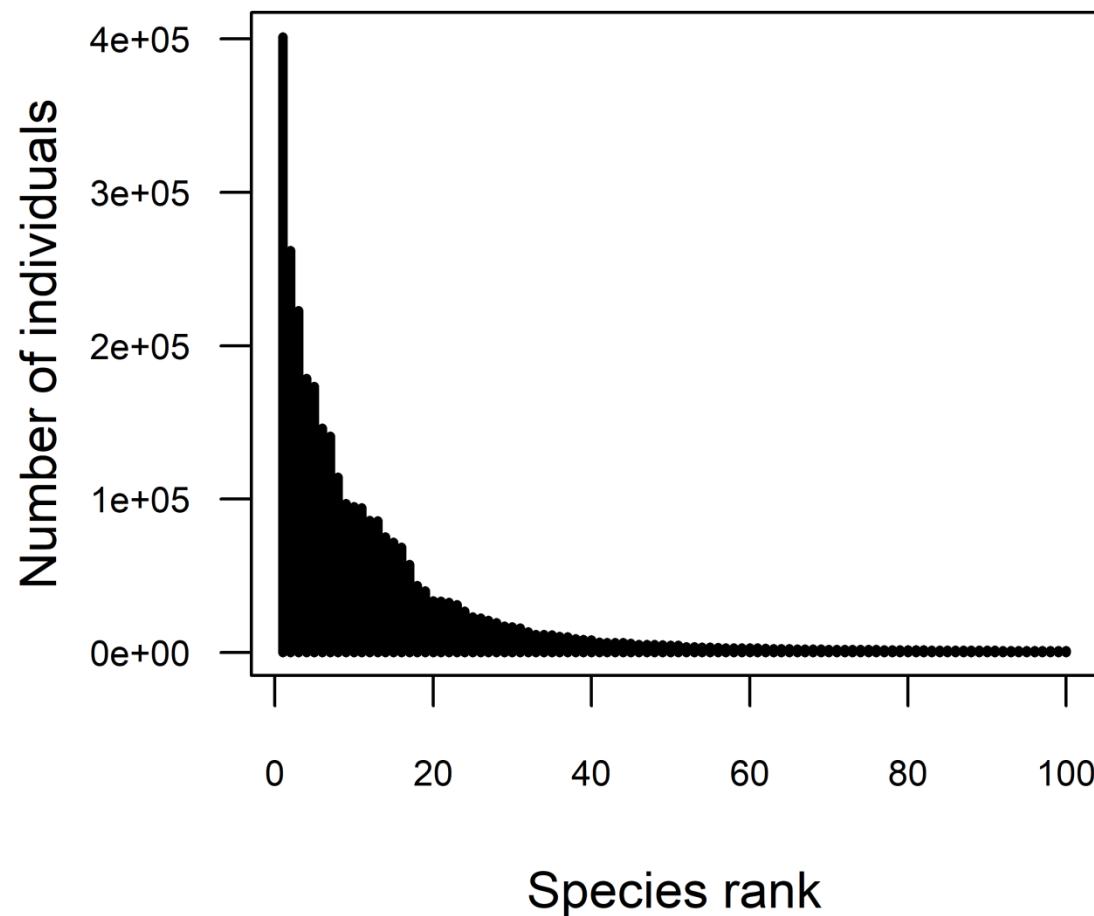
$$Tot-(S-1) = 6-(3-1) = 4$$

S – species richness
Tot – total number of individuals

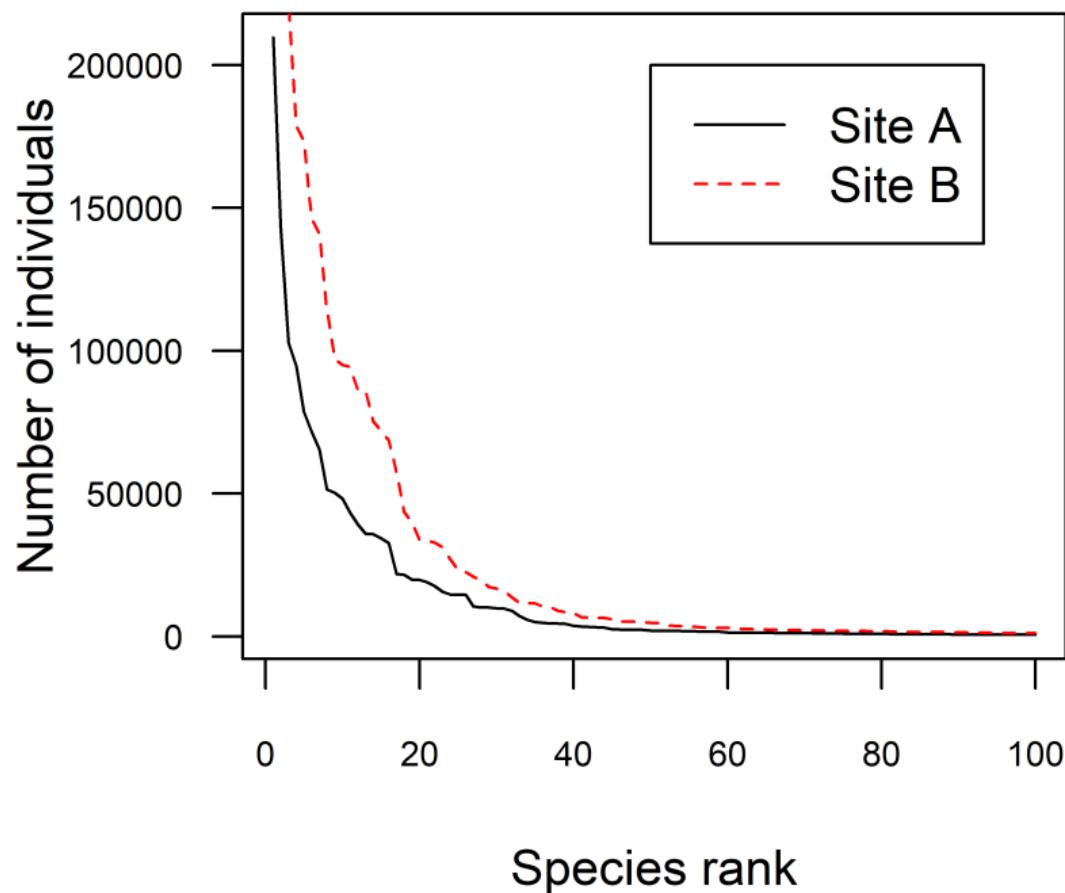
Relative abundance: common & rare spp.



Relative abundance: rank-abundance curve



Relative abundance: rank-abundance curve



Diversity indices

Diversity index - A mathematical expression that combines species richness and evenness as a measure of diversity.

Whittaker, R. H. (1960):

- α diversity of a local community or habitat
- β difference in diversity associated with differences in habitat or spatial scale
- γ total diversity of a region or other spatial unit

adiversity

Shannon-Weaver

$$H = - \sum_{i=1}^S p_i \log_b p_i$$

Simpson

$$D_1 = 1 - \sum_{i=1}^S p_i^2$$

Inverse Simpson

$$D_2 = \frac{1}{\sum_{i=1}^S p_i^2}$$

p_i - proportion of species i
 S - number of species

adiversity

Shannon-Weaver

$$H = - \sum_{i=1}^S p_i \log_b p_i$$

Simpson

$$D_1 = 1 - \sum_{i=1}^S p_i^2$$

Inverse Simpson

$$D_2 = \frac{1}{\sum_{i=1}^S p_i^2}$$

p_i - proportion of species i
 S - number of species

adiversity

Shannon-Weaver

$$H = - \sum_{i=1}^S p_i \log_b p_i$$

Simpson

$$D_1 = 1 - \sum_{i=1}^S p_i^2$$

Inverse Simpson

$$D_2 = \frac{1}{\sum_{i=1}^S p_i^2}$$

p_i - proportion of species i
 S - number of species

adiversity

Shannon-Weaver

$$H = - \sum_{i=1}^S p_i \log_b p_i$$

Simpson

$$D_1 = 1 - \sum_{i=1}^S p_i^2$$

Inverse Simpson

$$D_2 = \frac{1}{\sum_{i=1}^S p_i^2}$$

p_i - proportion of species i
 S - number of species

adiversity

Shannon-Weaver

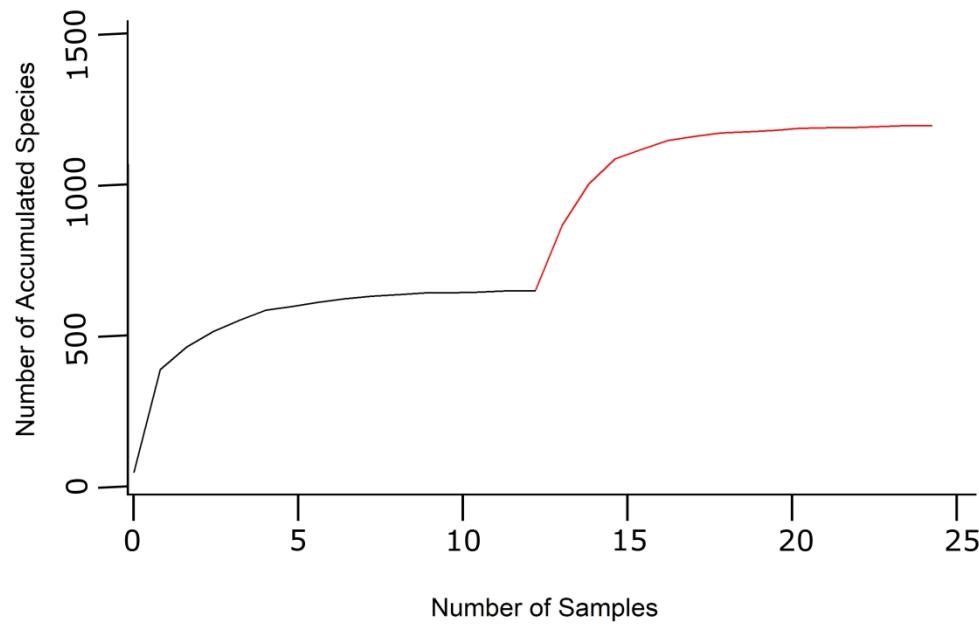
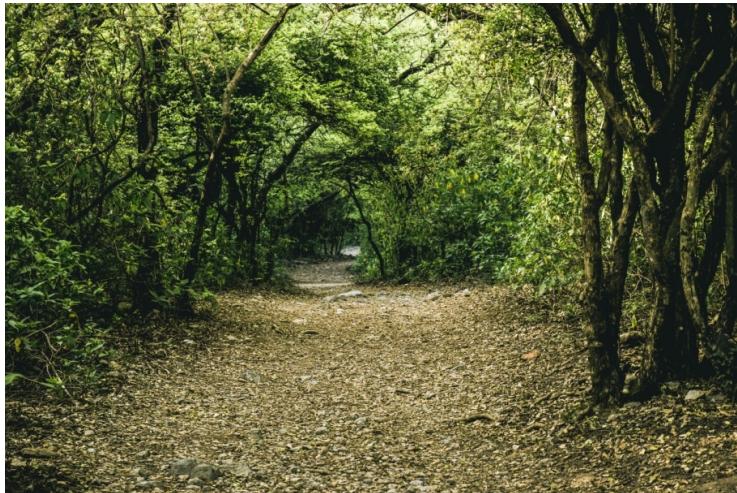
Simpson

Sample size

Inverse Simpson

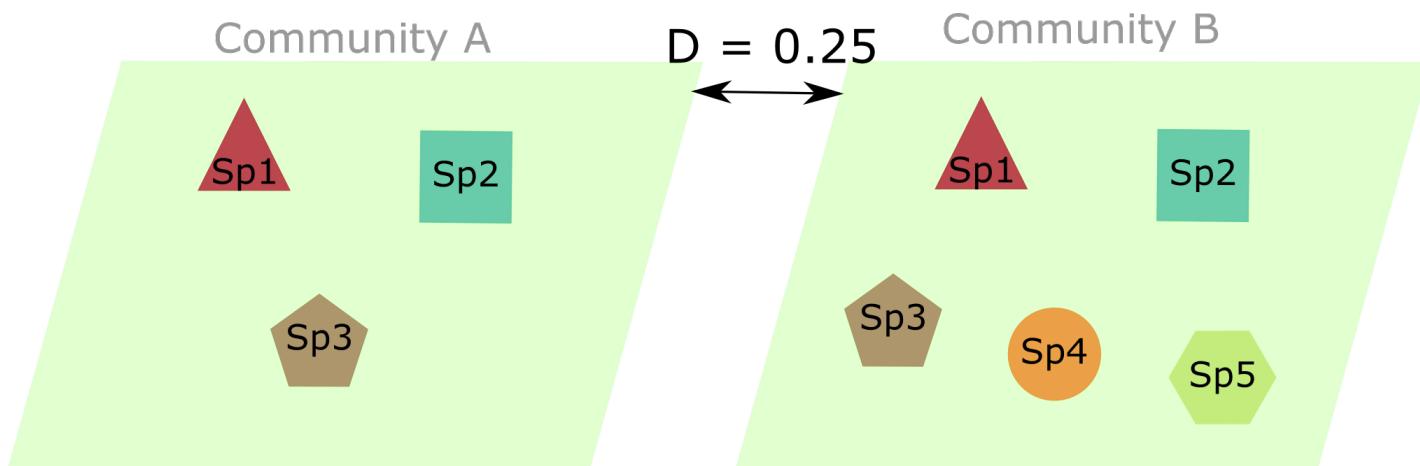
Others...

Species composition



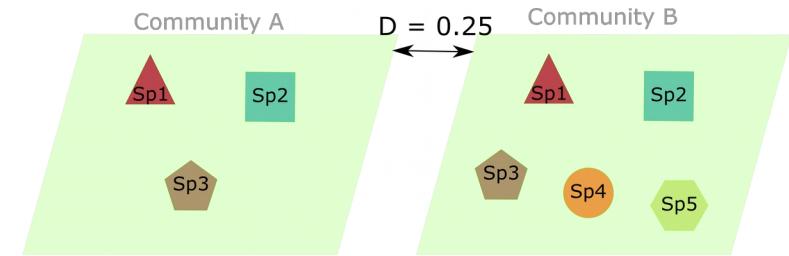
Analysis of differences in species composition (β)

Ecological distance



Analysis of differences in species composition (β)

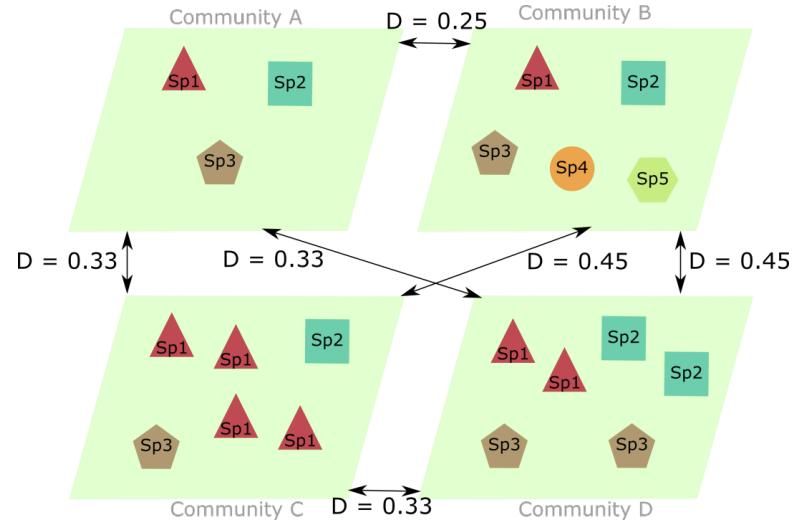
Ecological distance



Differences in species composition can be summarized with a single statistic

Information on the identities of the species is not available any longer

Ecological distances



Euclidean distance

Chi-square distance

Bray-Curtis dissimilarity

Jaccard and Sorenson distance

Euclidean distance

Sample	Sp 1	Sp 2	Sp 3
A	1	1	0
B	5	5	0
C	0	0	1

	A	B	C
A	0		
B	5.66	0	
C	1.73	7.14	0

Abundance - Species shared

Chi-square distance

Sample	Sp 1	Sp 2	Sp 3
A	1	1	0
B	5	5	0
C	0	0	1

	A	B	C
A	0		
B	1.57	0	
C	3.75	3.75	0

Proportions of species

Smaller Abundance - Larger abundance

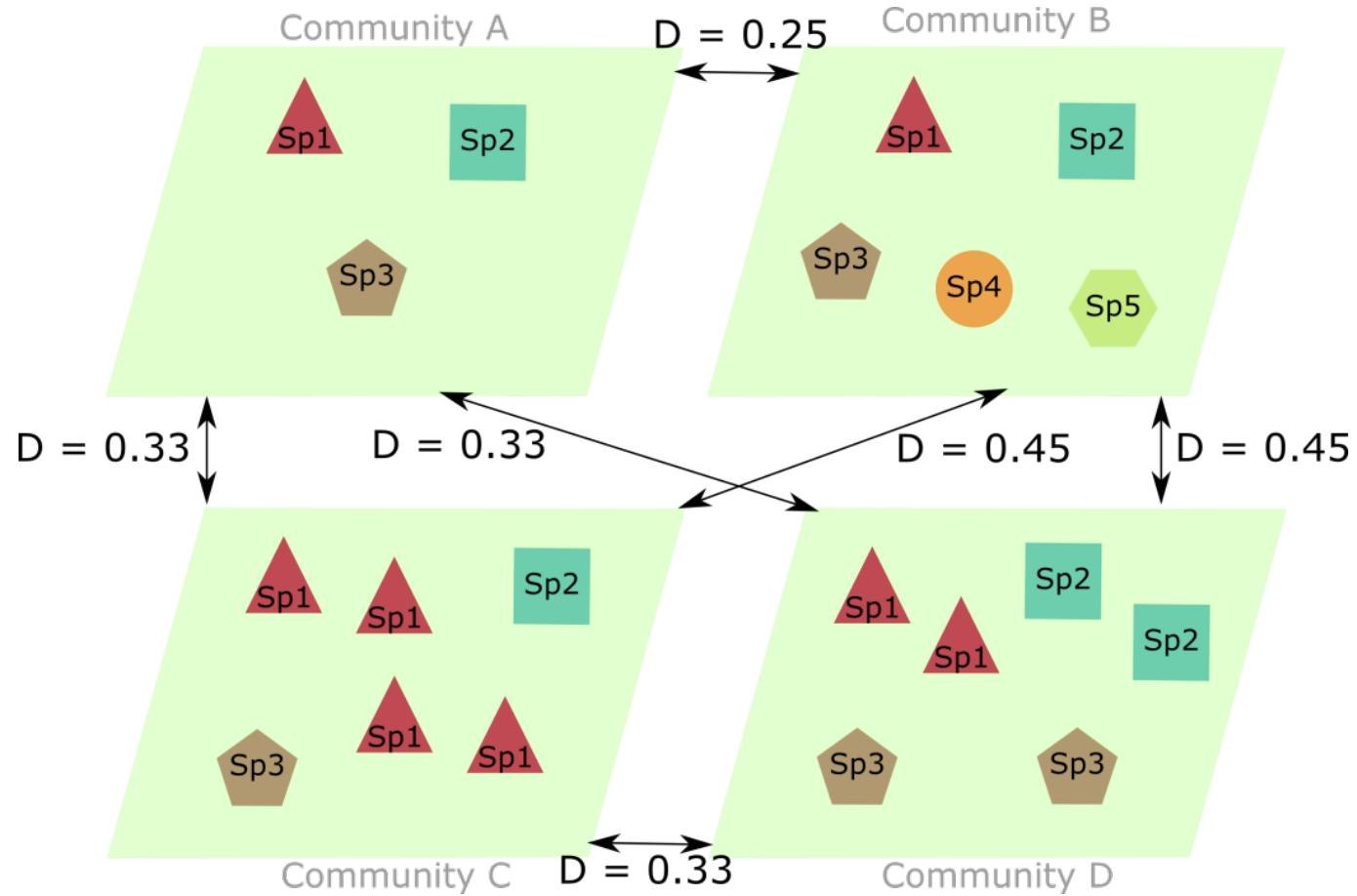
Bray-Curtis dissimilarity

Sample	Sp 1	Sp 2	Sp 3
A	1	1	0
B	5	5	0
C	0	0	1

	A	B	C
A	0		
B	0.67	0	
C	1	1	0

Abundance - Species shared

Bray-Curtis dissimilarity



Jaccard distance and Sorenson distance

Presence / absence

Sorenson - Bray-Curtis

Choice of a distance measure

Euclidean distance

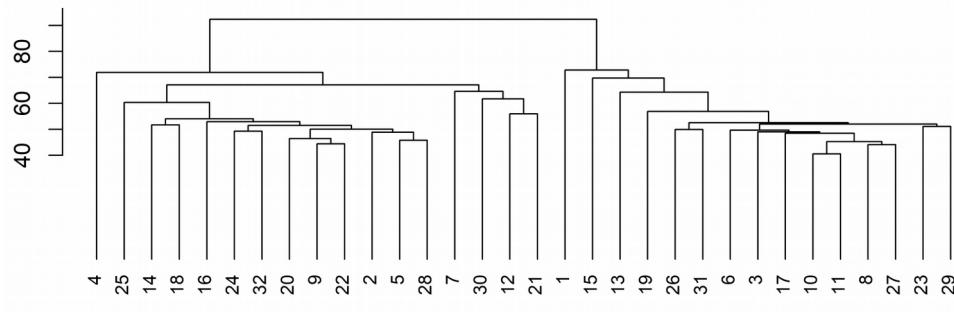
Chi-square distance

Bray-Curtis dissimilarity

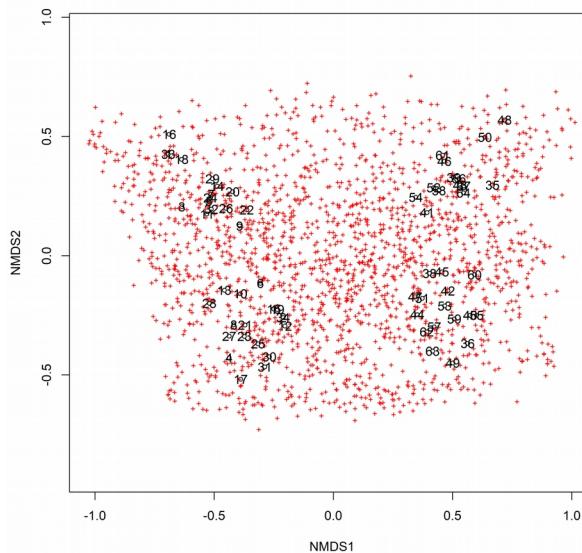
Presence/absence:

Jaccard and Sorenson distance

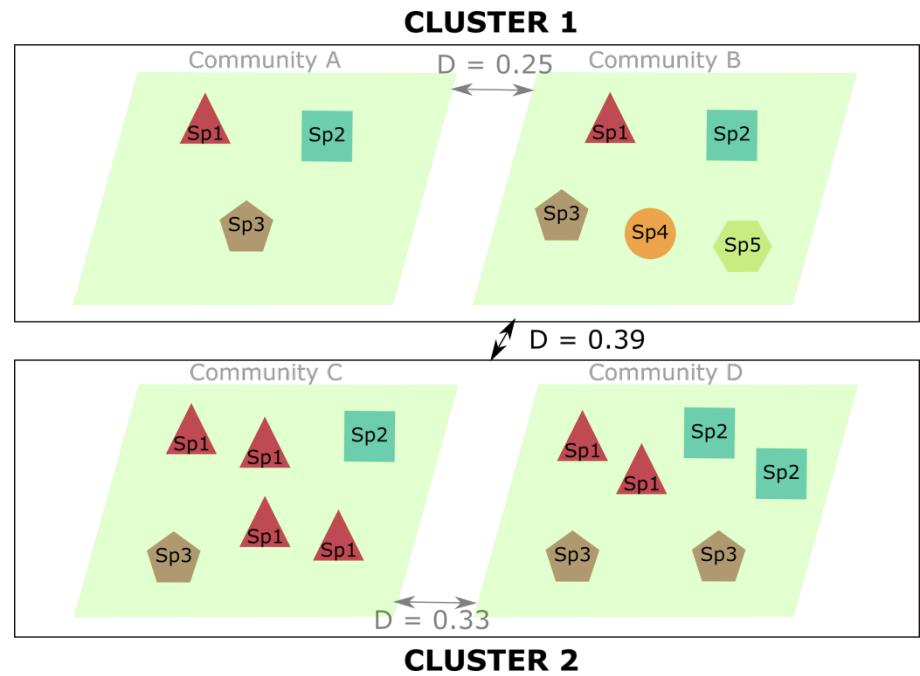
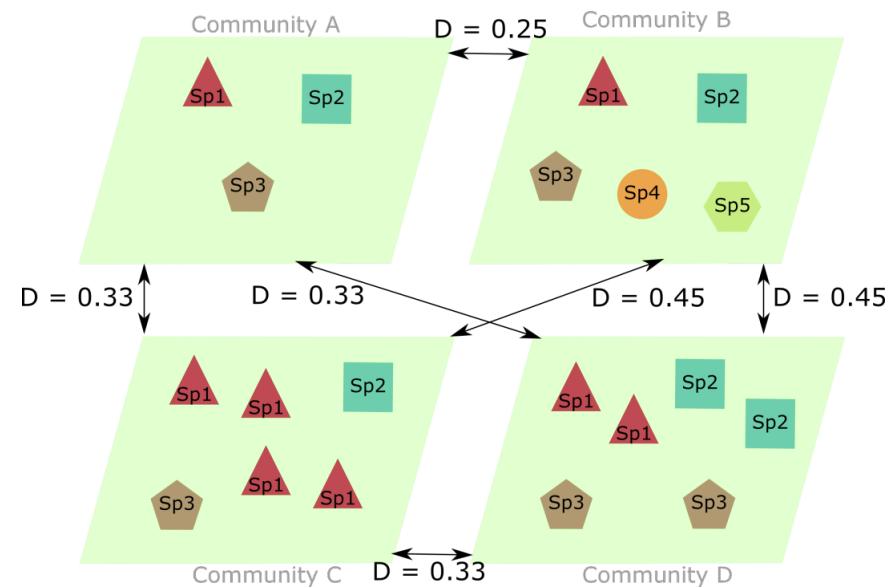
Analysis of ecological distance by clustering:



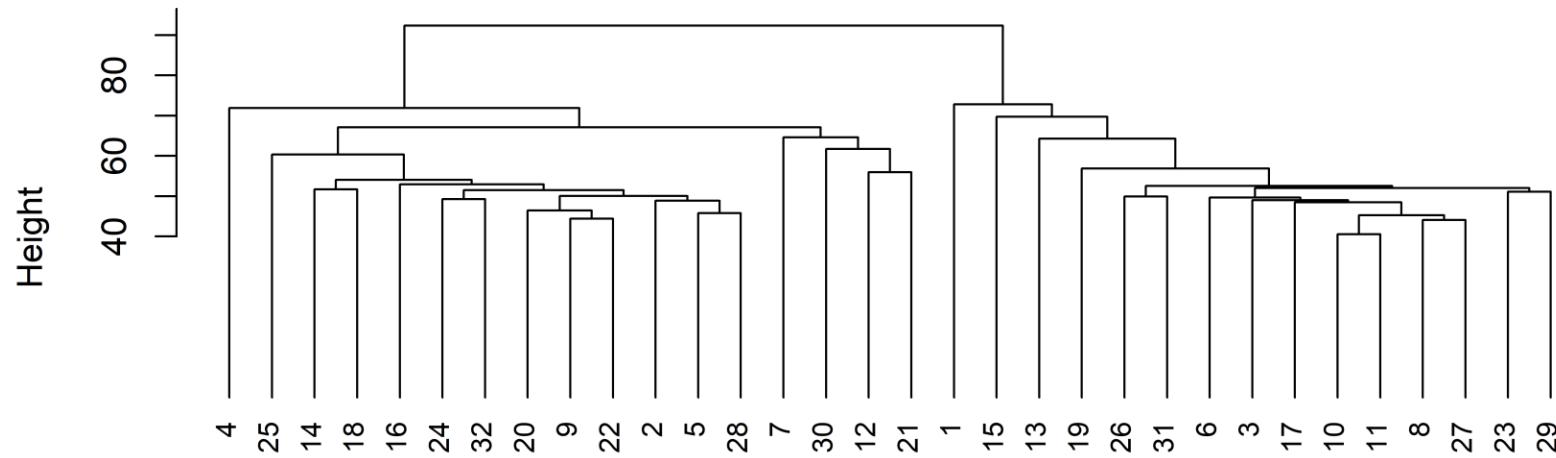
by ordination:



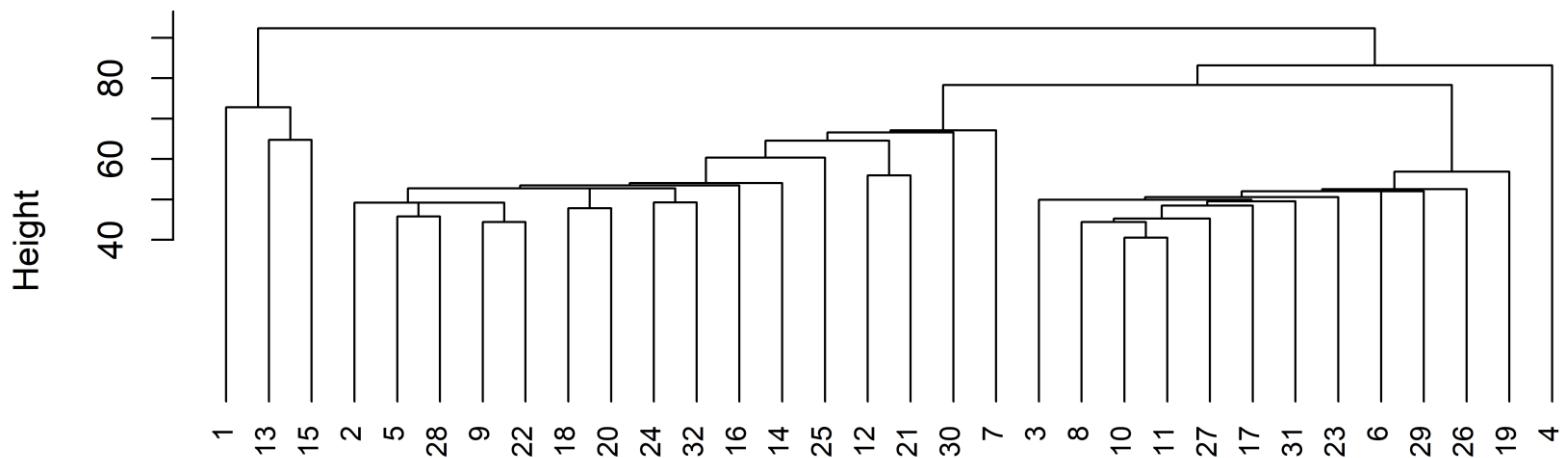
Analysis of ecological distance by clustering



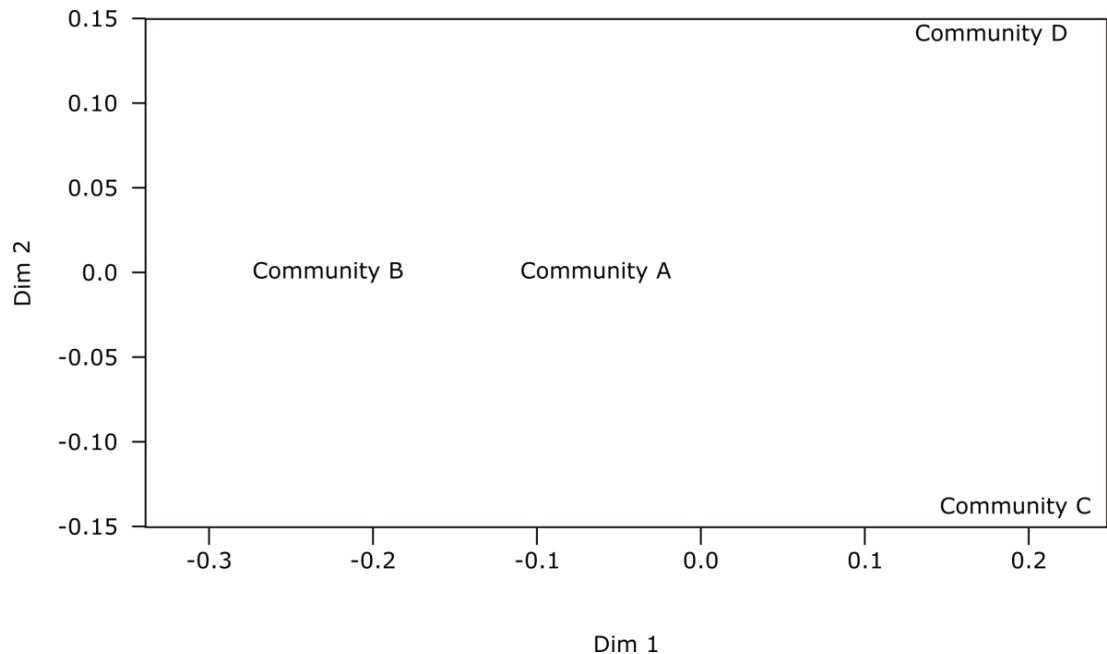
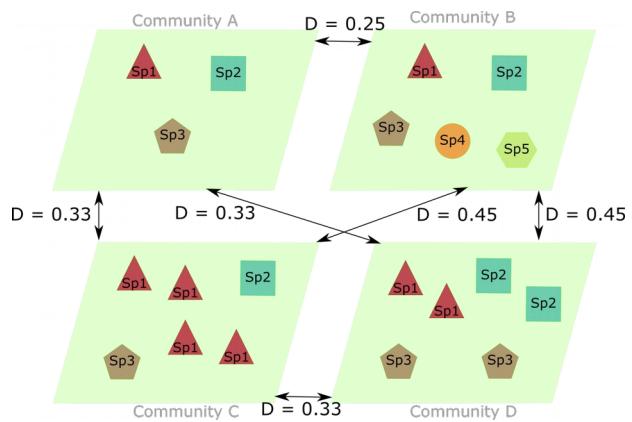
Agglomerative Hierarchical Clustering



Divisive Hierarchical Clustering



Analysis of ecological distance by ordination



Analysis of ecological distance by ordination

Unconstrained:

PCA

PCoA

NMDS

CA

Constrained:

RDA

CCA

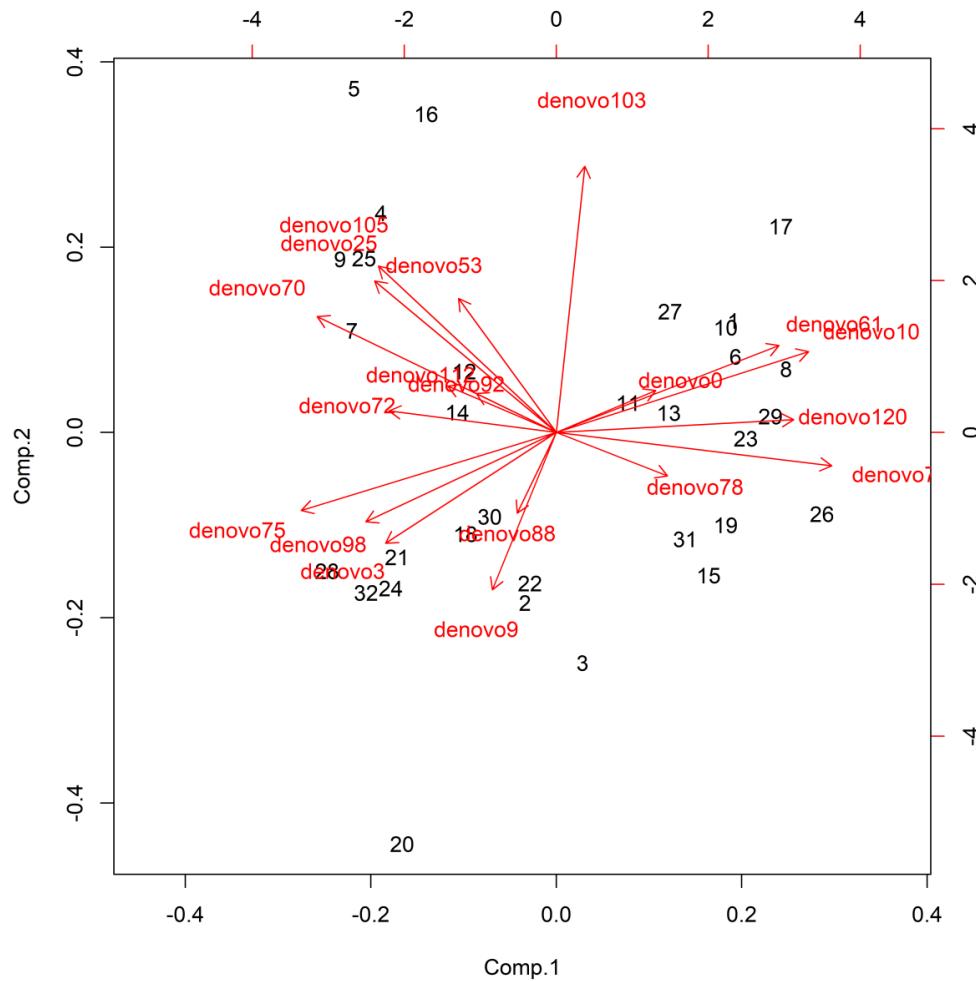
db-RDA

Only use ecological distance

Use environmental variables to guide the ordination

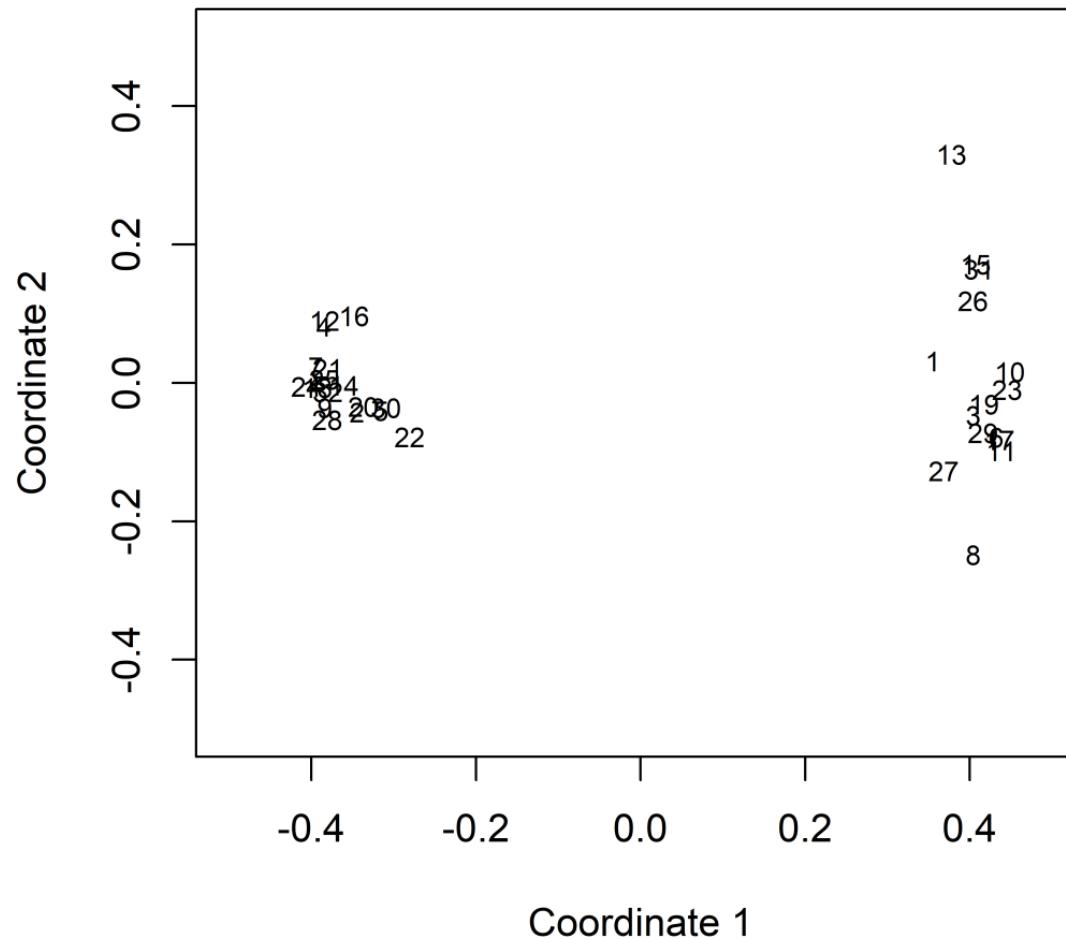
Principal Component Analysis

PCA - Euclidean distance



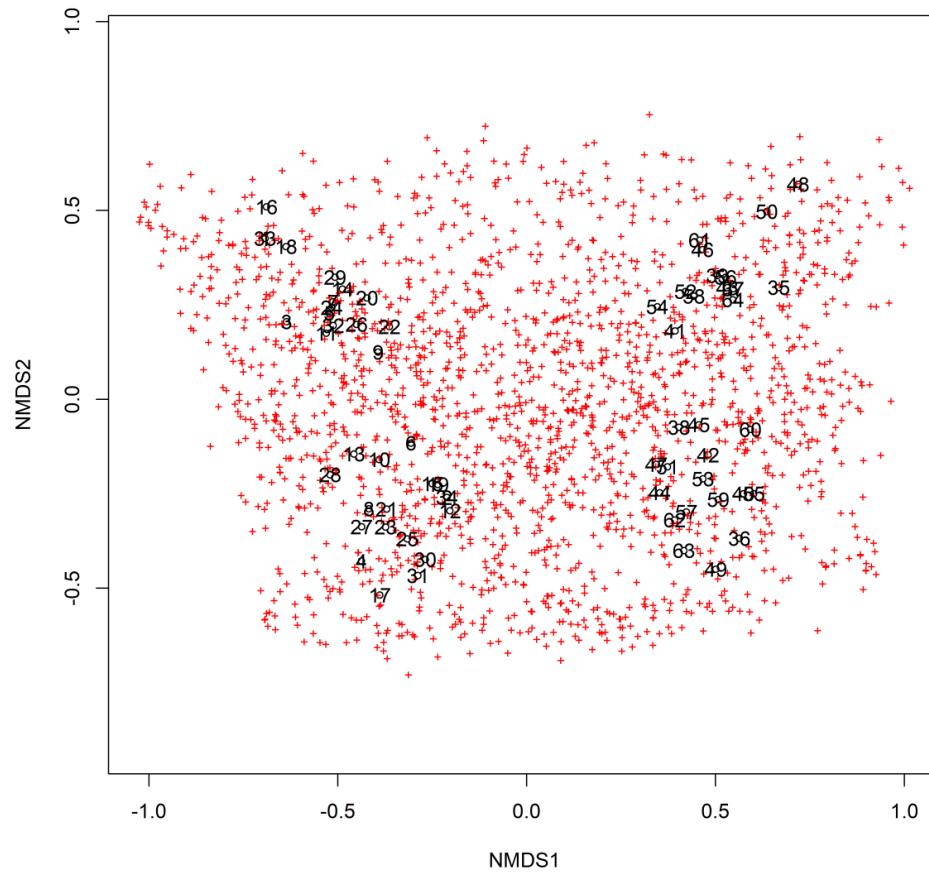
Principal Coordinates Analysis

PCoA - any ecological distance



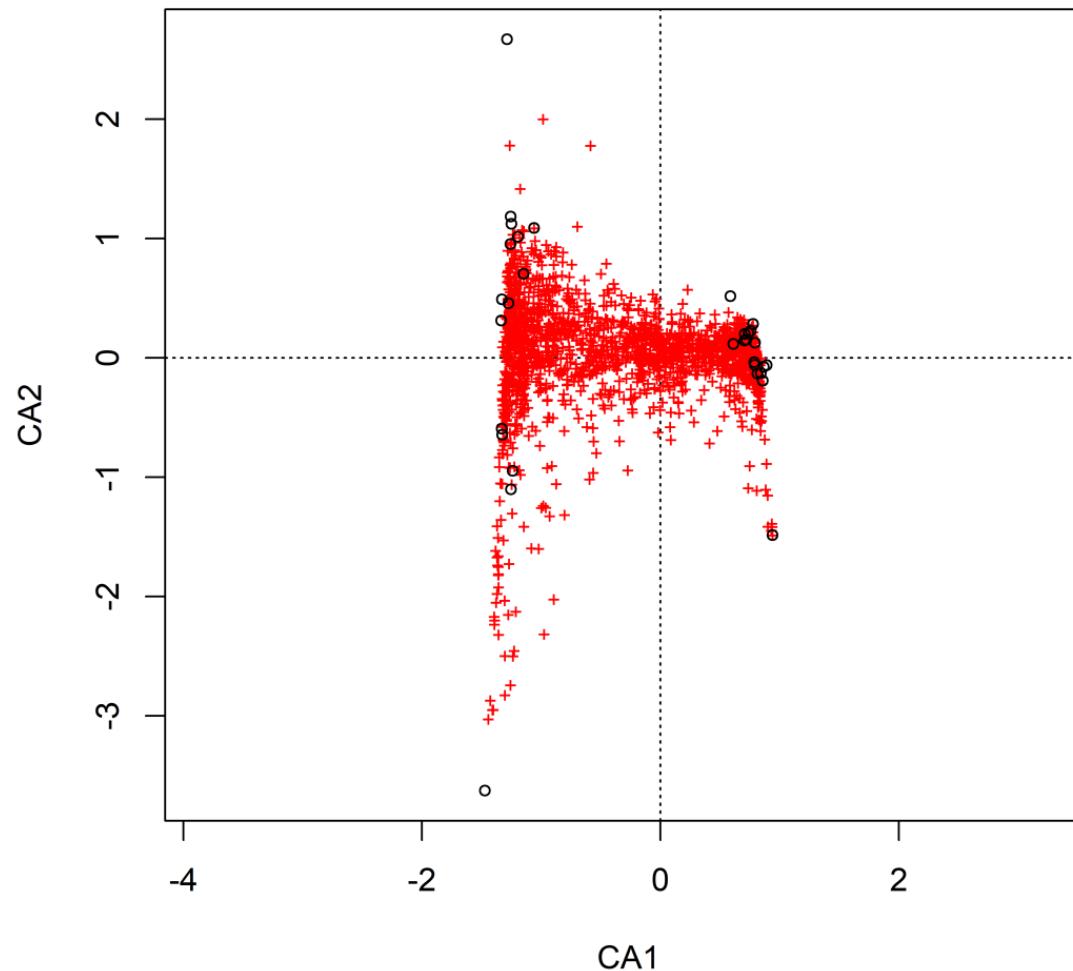
Non-metric Multidimensional Scaling

NMDS - any ecological distance
- rank-based approach
- random process



Correspondence Analysis

CA - chi-square distance



Analysis of ecological distance by ordination

Unconstrained:

PCA

PCoA

NMDS

CA

Constrained:

RDA

CCA

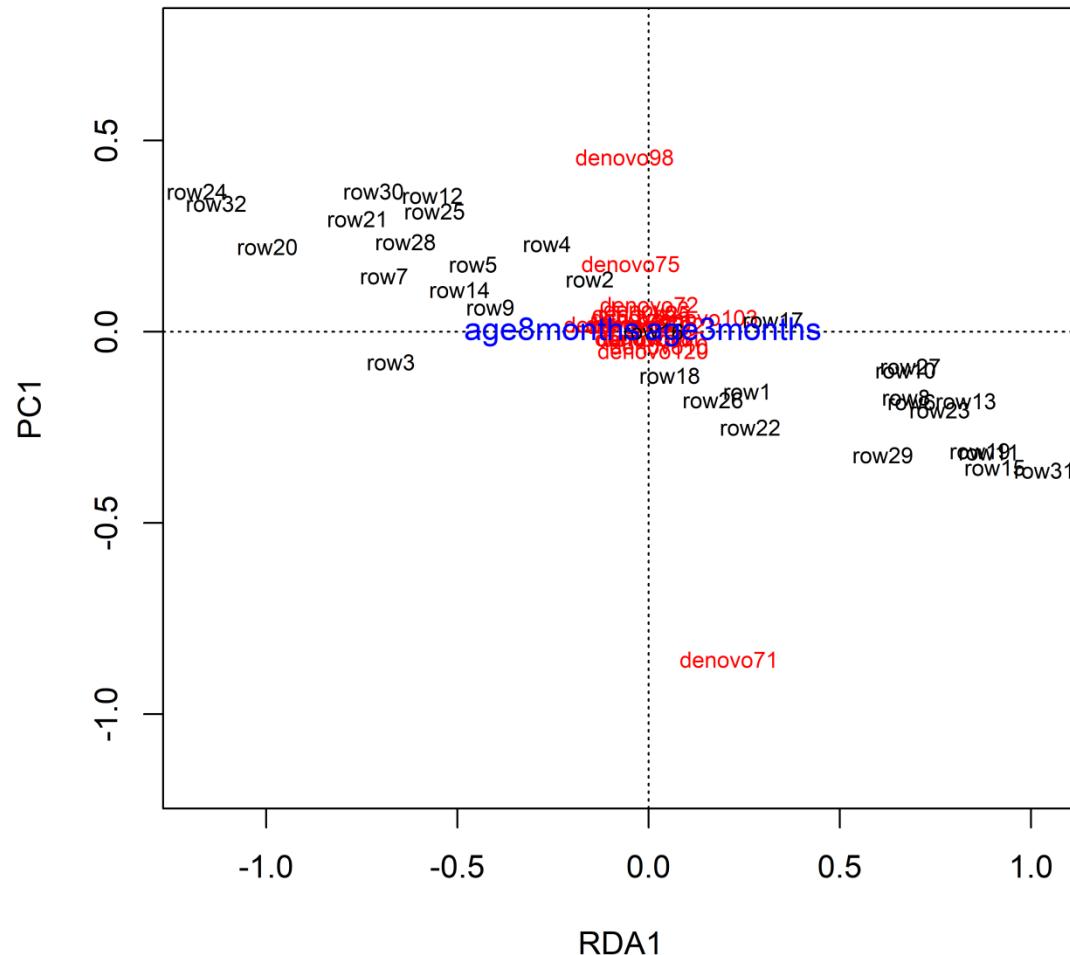
db-RDA

Only use ecological
distance

Use environmental
variables to guide
the ordination

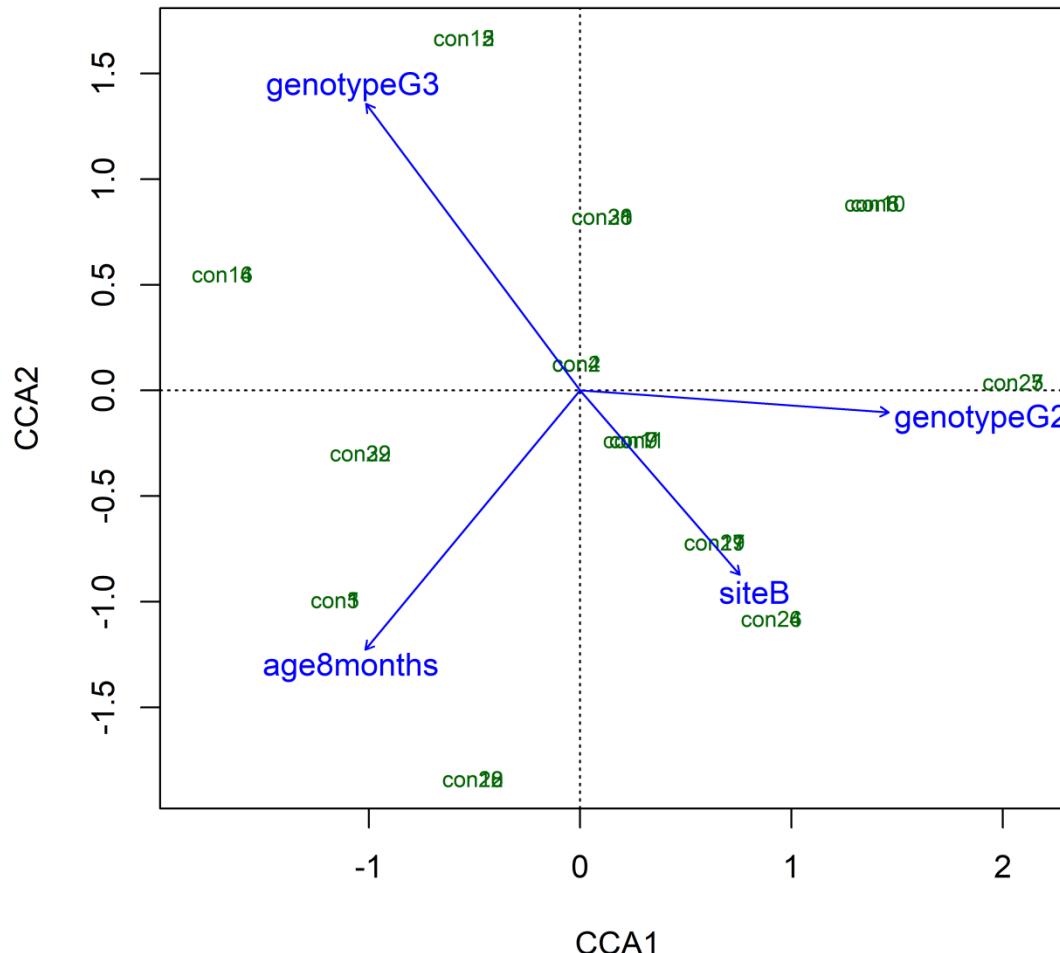
Redundancy Analysis

RDA - Euclidean distance



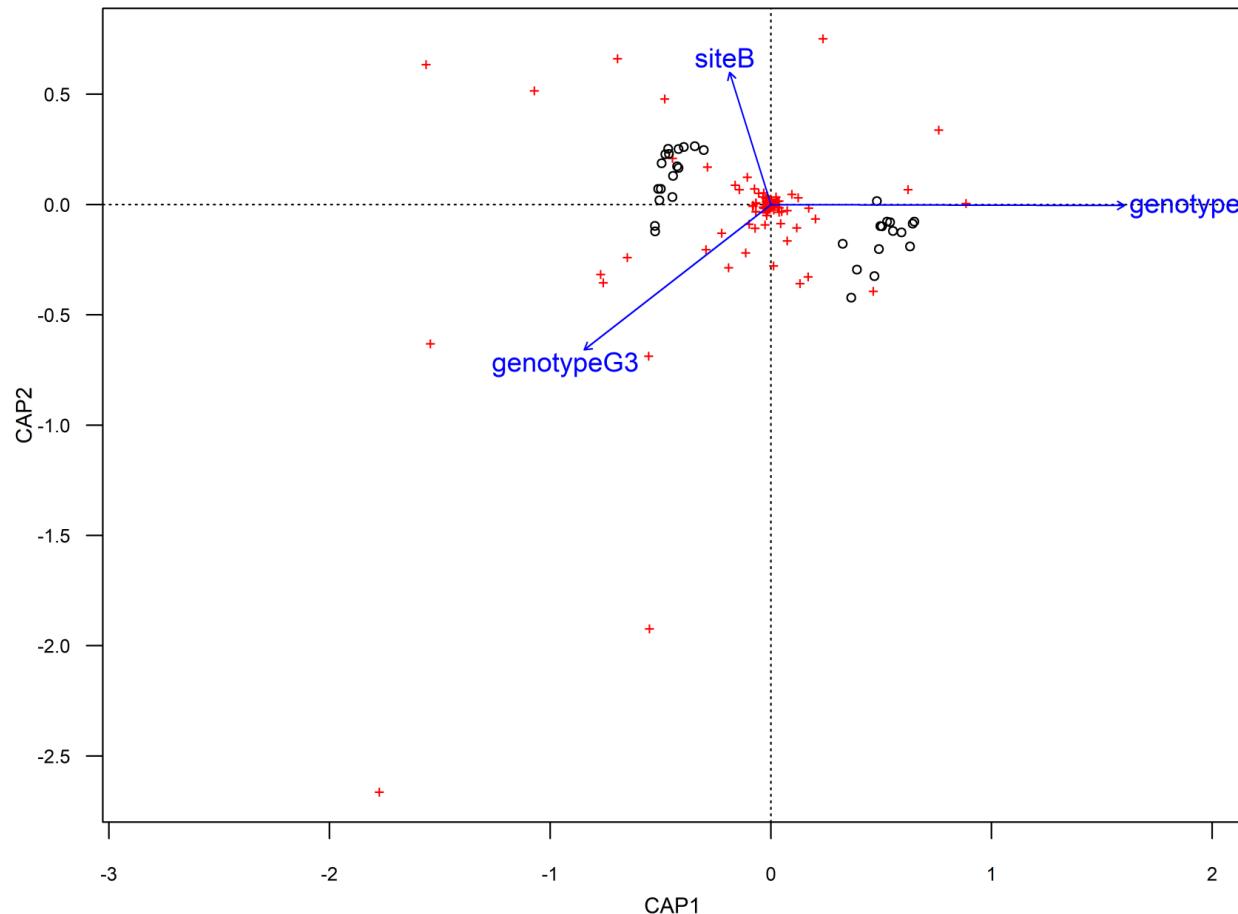
Canonical Correspondence Analysis

CCA – chi-square distance



Distance-based Redundancy Analysis

db-RDA – any ecological distance



Analysis of ecological distance by ordination

Unconstrain:

PCA

CA

PCoA

NMDS

Constrain:

RDA

CCA

db-RDA

Only use ecological
distance

Use environmental
variables to guide
the ordination