Targeted Metagenomics (cont) and Metagenomics

Slides thank to Jenn DeBruyn

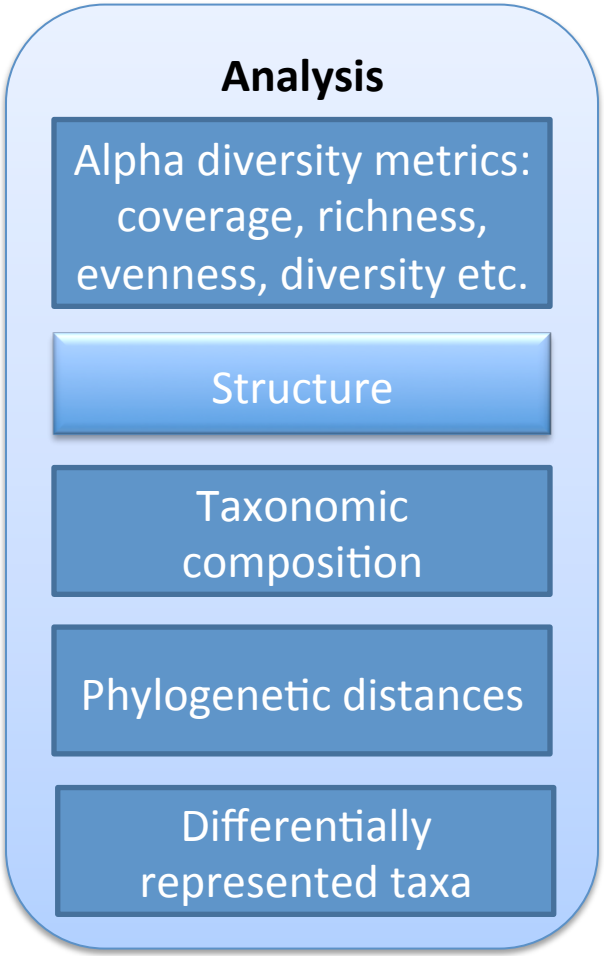
Associate Professor

BIOSYSTEMS ENGINEERING & SOIL SCIENCE

UTIA INSTITUTE OF AGRICULTURE

THE UNIVERSITY OF TENNESSEE

Structure



- Structure refers to the distribution of taxa in a given community

Comparing community structures

OTU table: Samples x species

	Community			
	A	**B**	**C**	**D**
OTU1	1	1	1	0
OTU2	0	2	2	1
OTU3	0	0	1	1
OTU4	0	0	0	1
OTU5	2	4	4	1
OTU6	3	1	1	3
OTU7	1	1	0	1
OTU8	1	1	0	1
OTU9	0	0	0	0
OTU10	1	1	1	1

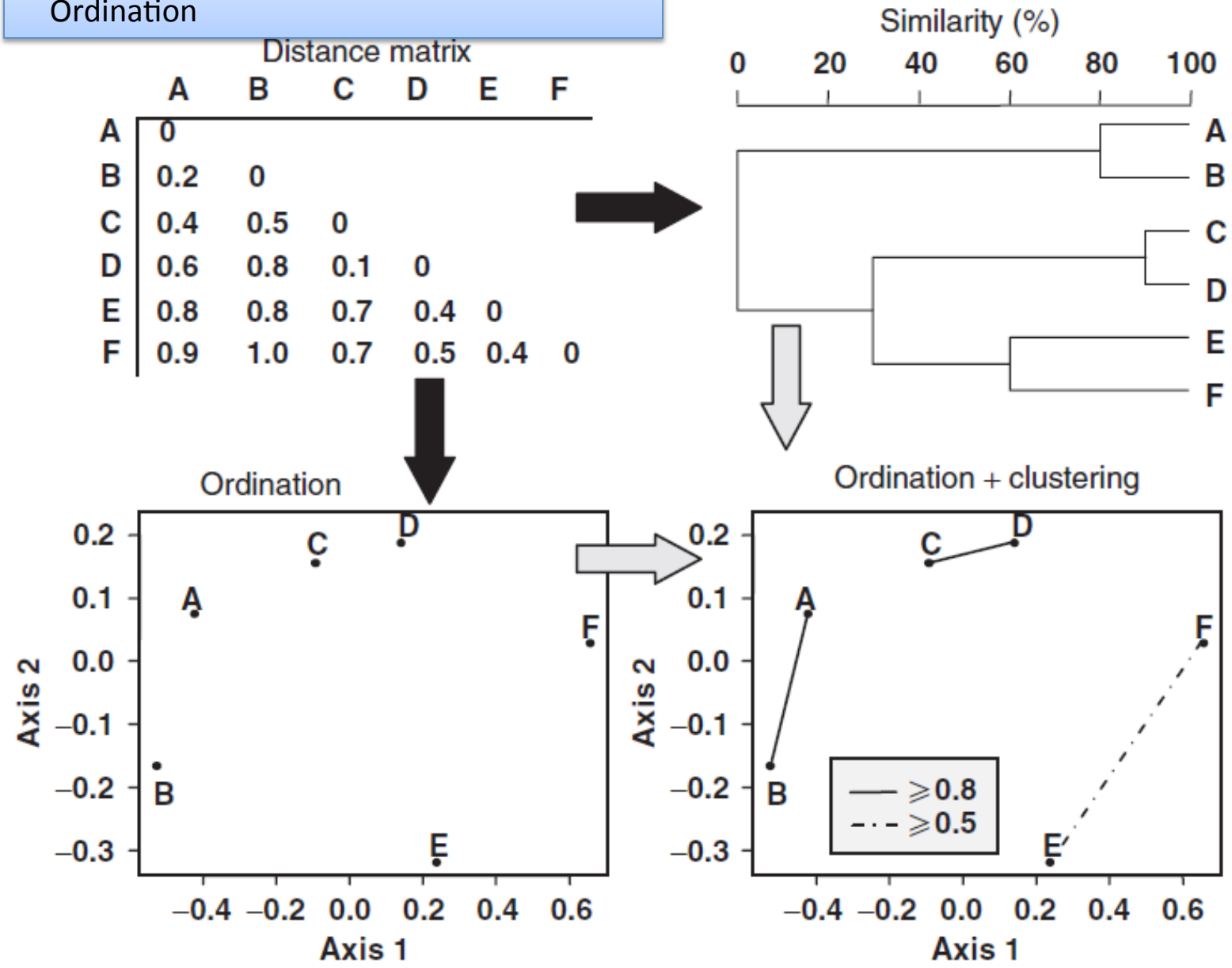
OTU = operational taxonomic unit ("species")

Calculate distance between each pair
Dissimilarity index (or **1 - similarity** index)
Options: Bray Curtis, Sorenson, Euclidean, Jaccard etc.

	A	**B**	**C**	**D**
A	0			
B	d(A,B)	0		
C	d(A,C)	d(B,C)	0	
D	d(A,D)	d(B,D)	d(C,D)	0

Multivariate analysis:
- Clustering (dedograms)
- Ordination (NMDS, PCA, CCA etc.)

Multivariate analysis:
Classification/ Clustering (dendograms)
Ordination



Example Ordination

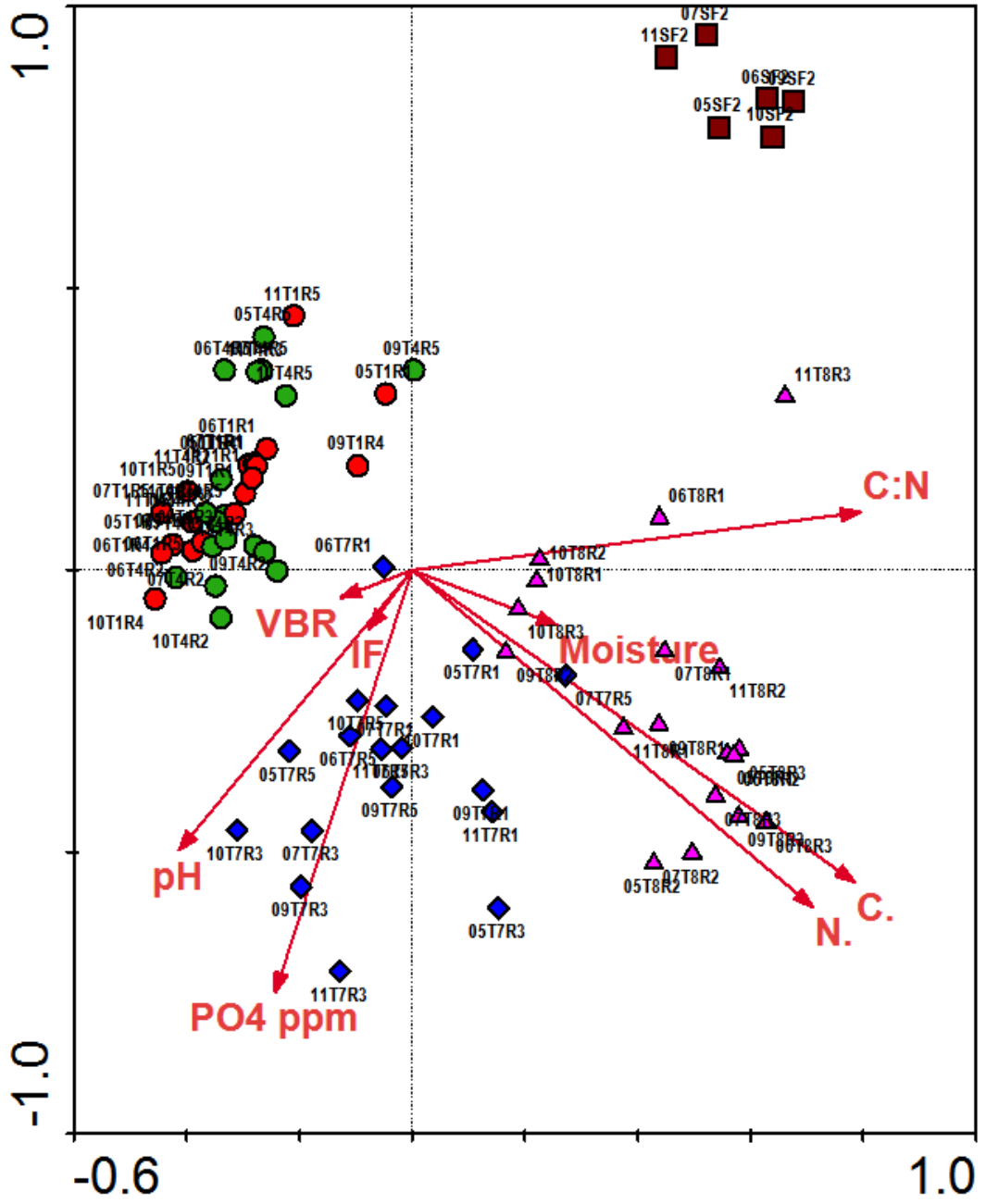
T1 Corn/soybean/ wheat - conventional till and inputs

T4 Corn/ soybean/ wheat - organic (reduced inputs, winter cover)

T7 Early successional community - annually burned

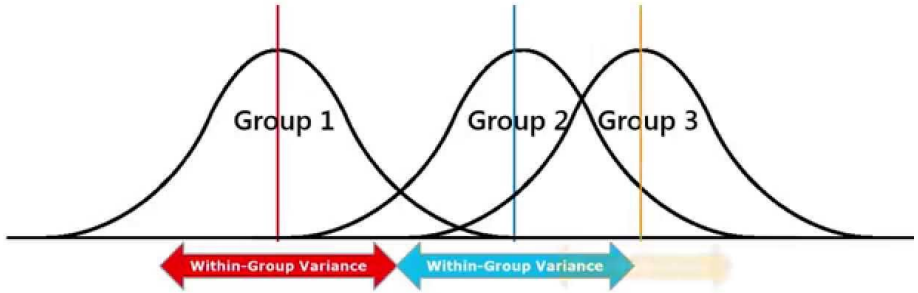
T8 Early successional community - annually mowed, never tilled

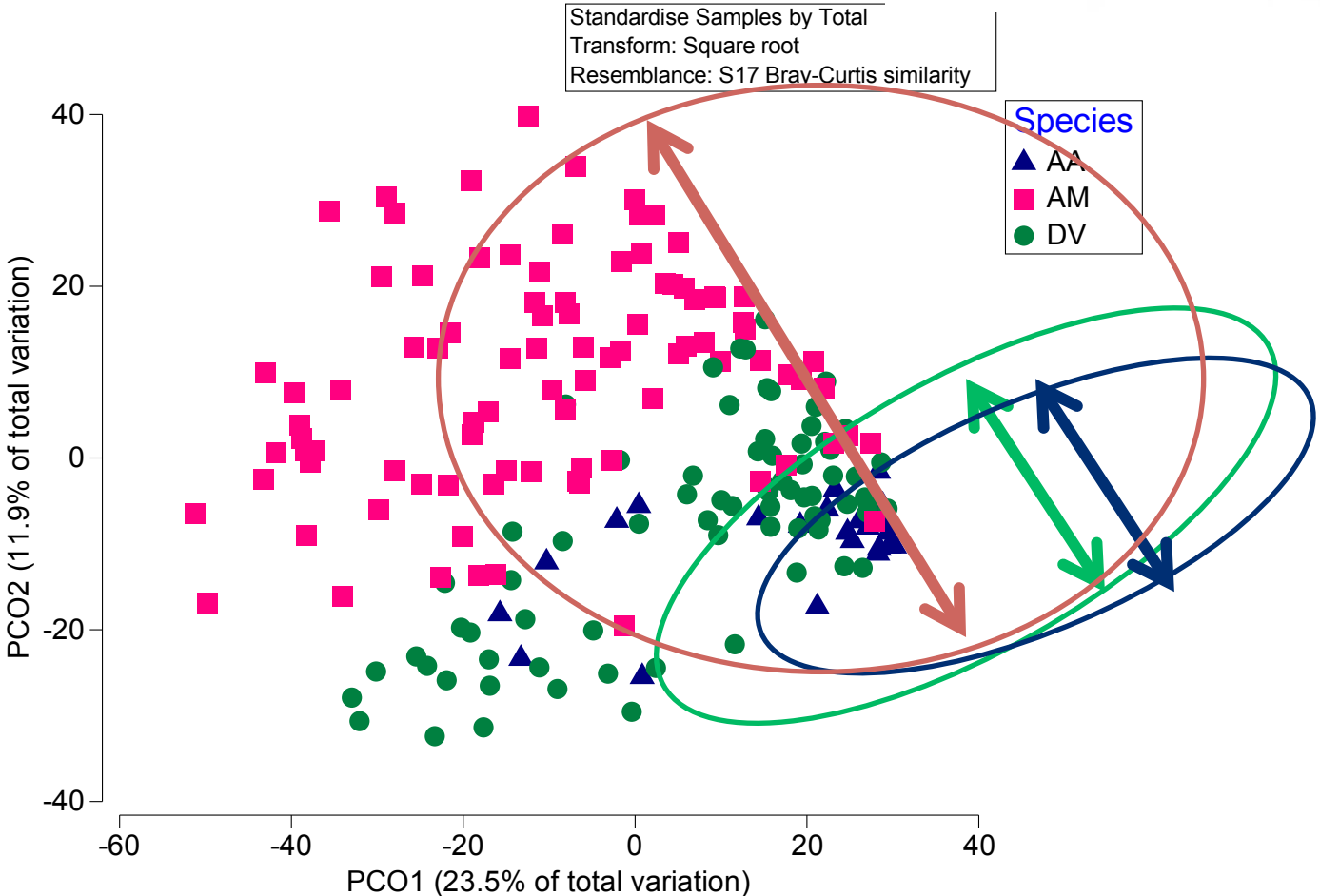
SF2 Mid successional forest (40 – 60 years post agriculture)



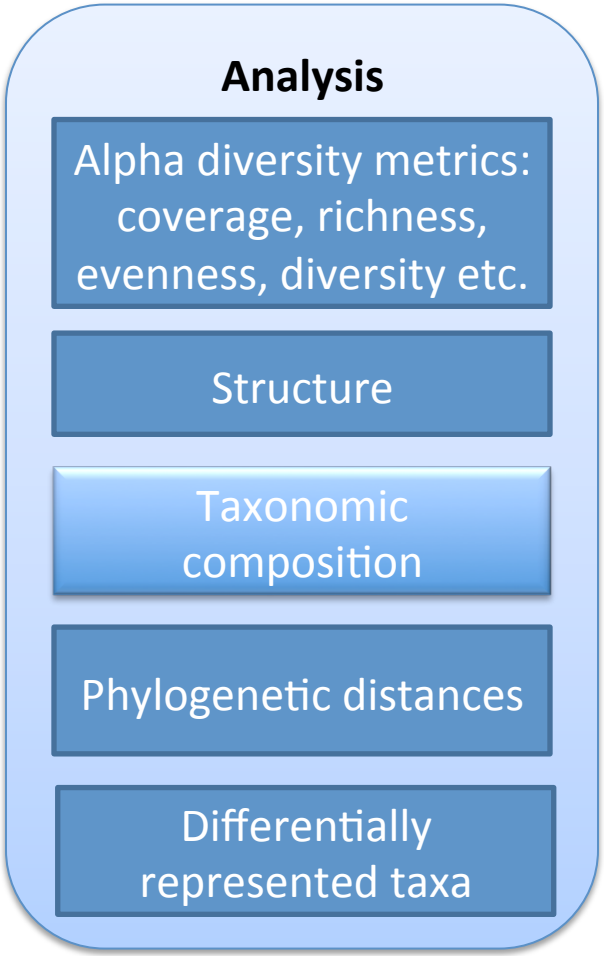
Statistical Tests for Determining Differences in Structure

- Multivariate versions of ANOVA
  - ANOSIM
  - PERMANOVA



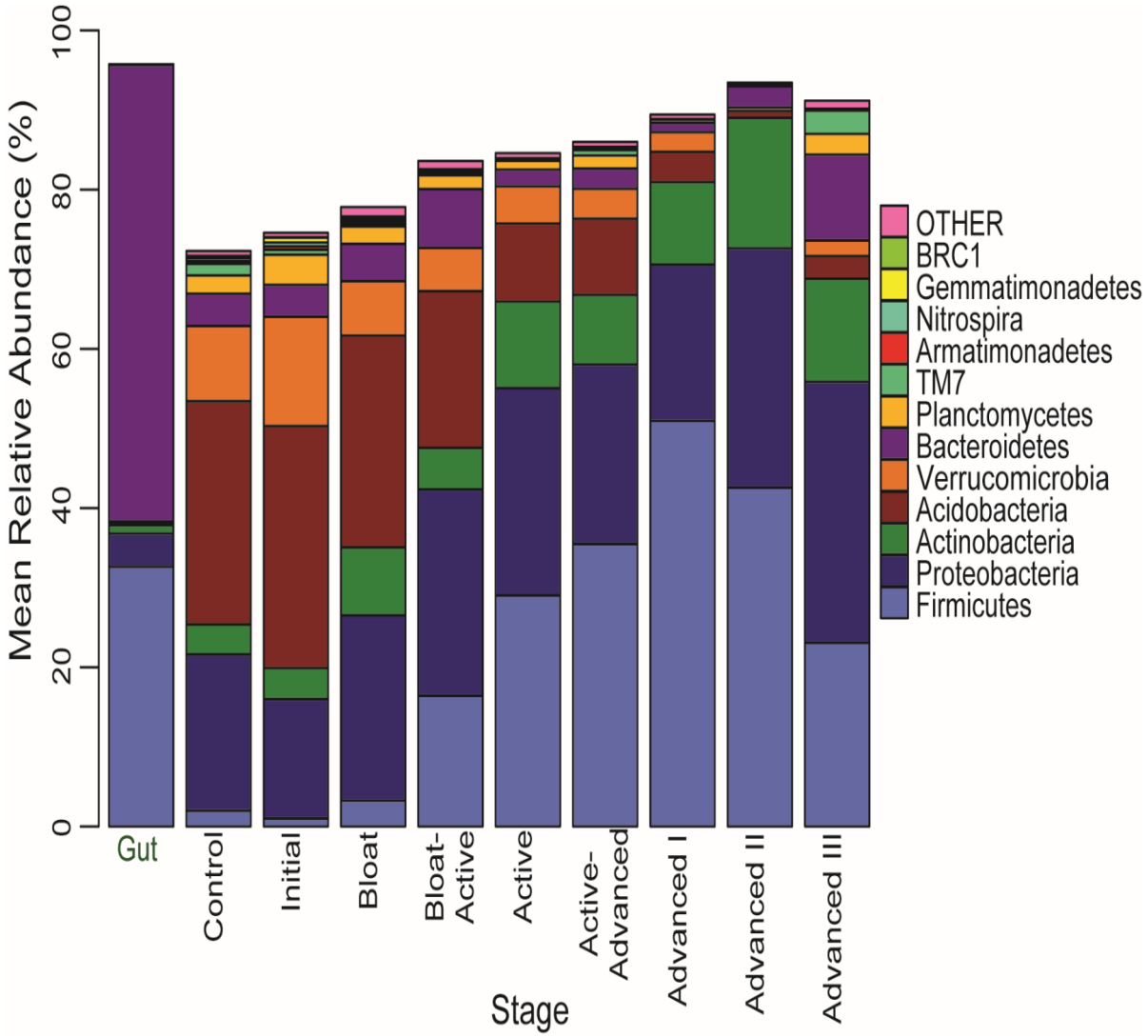


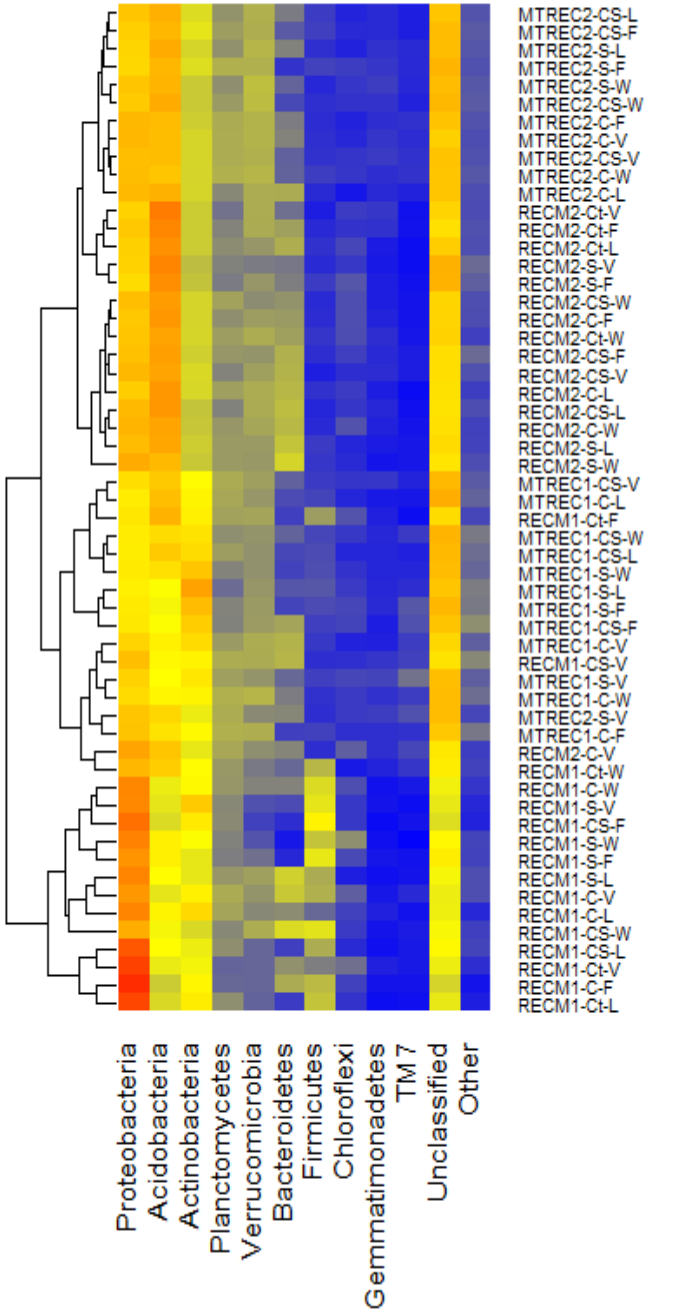
Taxonomic Composition



- Who's there?
- Taxa of importance (e.g. pathogens, nitrogen fixers)

Example: Soil bacterial communities

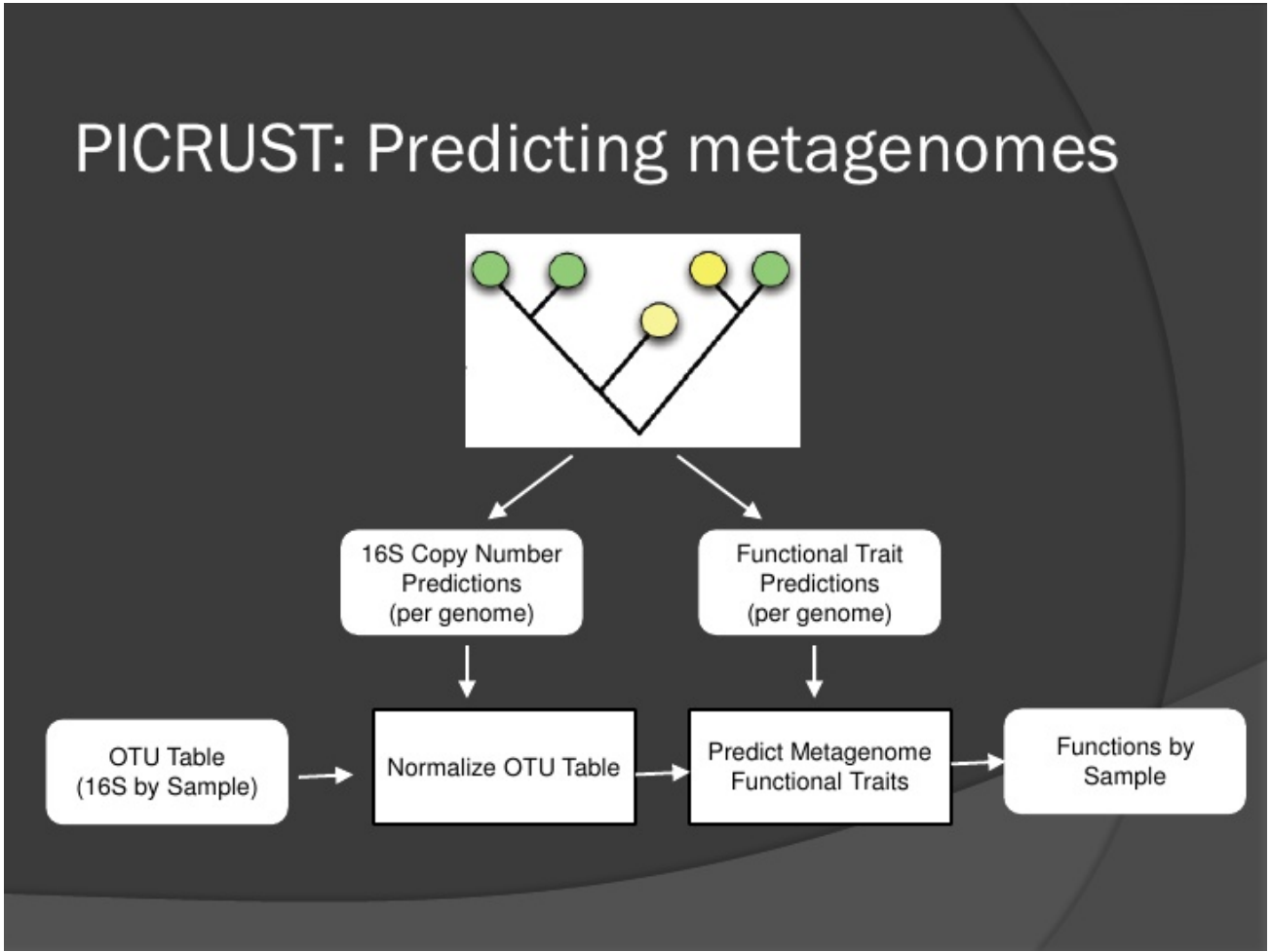




Cobaugh et al. *PLoS ONE*. 2015;10: e0130201

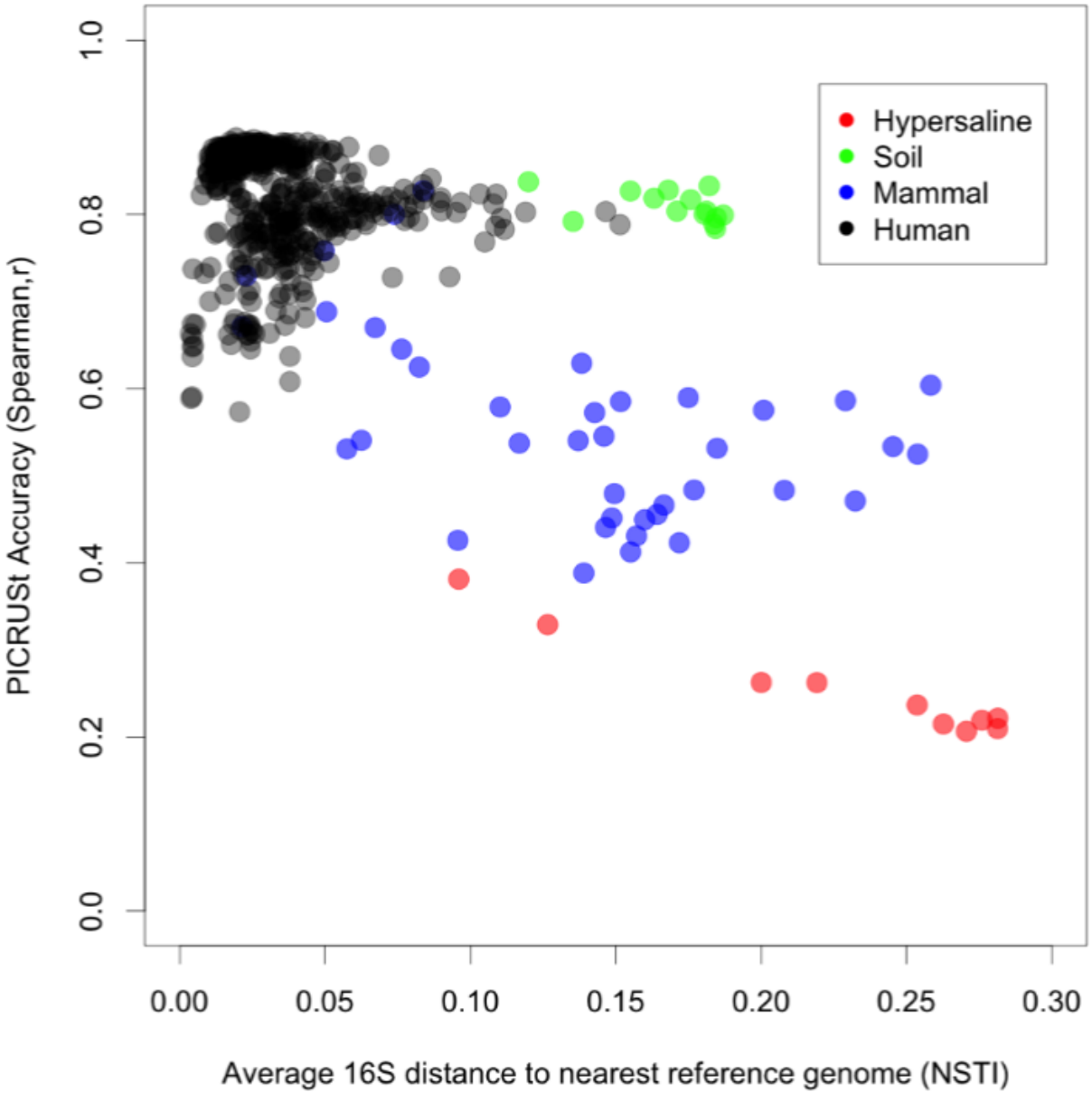
Functional Prediction Tools

- Packages:
  - PiCRUST: Phylogenetic investigations of communities by reconstruction of unobserved states (Langille et al. 2013)
  - Tax4FUN (R package)



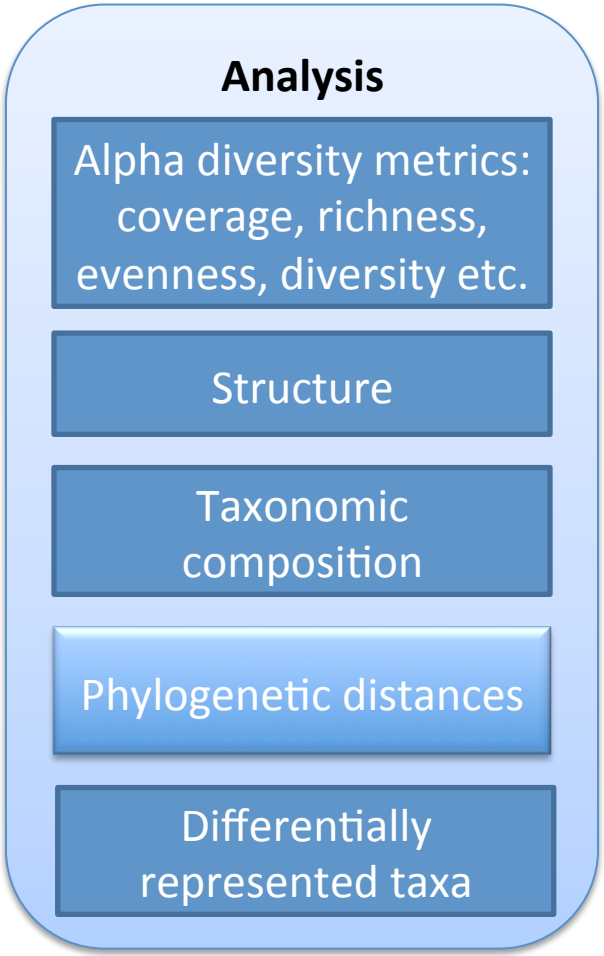
Functional Prediction Tools:
Database introduces phylogenetic bias

Most common soil phyla (usually >1% relative abundance)	# of Genomes IMG, Sept 2016	% of genomes in database (total: 42,289)
Proteobacteria	19458	46.0%
Acidobacteria	**53**	**0.1%**
Actinobacteria	5282	29.3%
Verrucomicrobia	**85**	**0.2%**
Bacteroidetes	1626	3.8%
Planctomycetes	**76**	**0.2%**
Firmicutes	12370	29.3%
Armatimonadetes	**9**	**0.02%**
Chloroflexi	**99**	**0.2%**
Gemmatimonadetes	**19**	**0.04%**
Nitrospirae	**26**	**0.1%**



Take home message:
PiCrust/ Tax4Fun are cool tools for hypothesis generation, but DON'T OVERSELL!

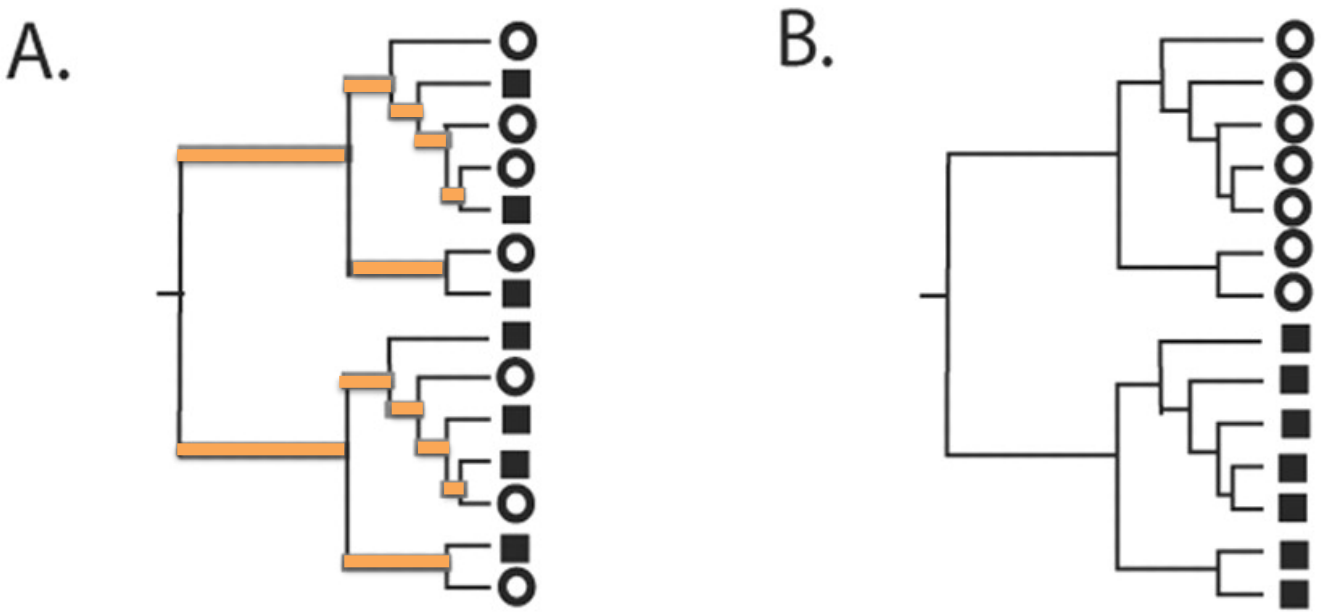
Phylogenetic distances



UNIFRAC

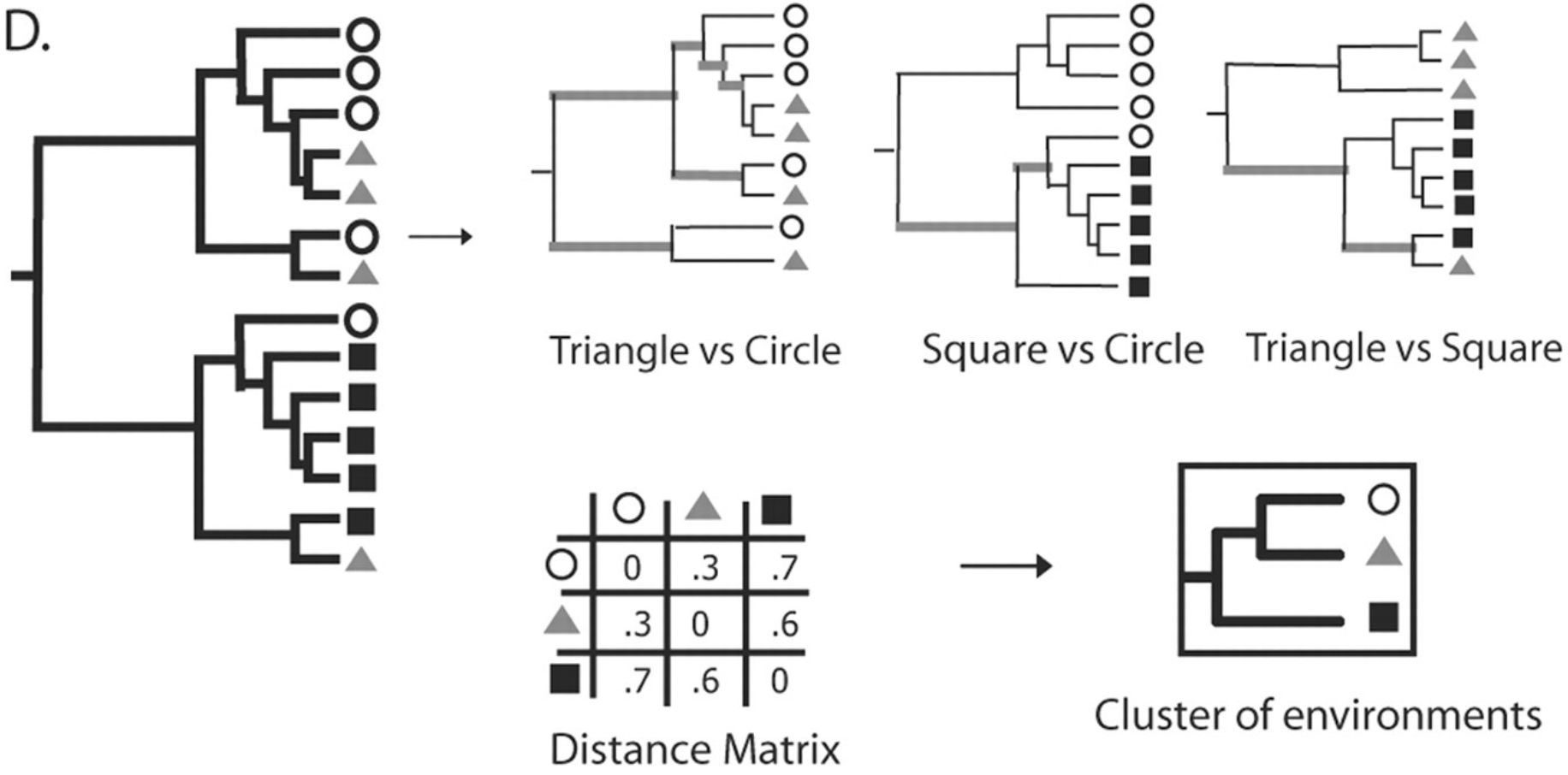
- How related communities are
- A distance measurement between communities
- Incorporates information about phylogenetic relatedness of OTUs

Phylogenetic distances: UNIFRAC



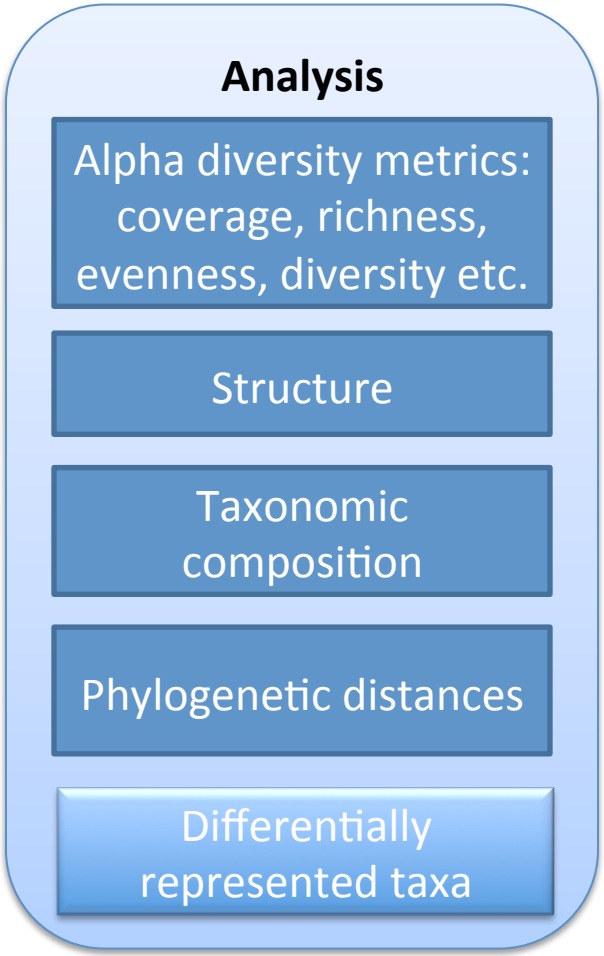
Example: Comparing two samples (Square and Circle)
Calculate the fraction of shared branches

Applied and Environmental Microbiology, December 2005, p. 8228-8235, Vol. 71, No. 12
UniFrac: a New Phylogenetic Method for Comparing Microbial Communities
Catherine Lozupone[1], and Rob Knight[2*]



Applied and Environmental Microbiology, December 2005, p. 8228-8235, Vol. 71, No. 12
UniFrac: a New Phylogenetic Method for Comparing Microbial Communities
Catherine Lozupone[1], and Rob Knight[2*]

Differentially represented taxa

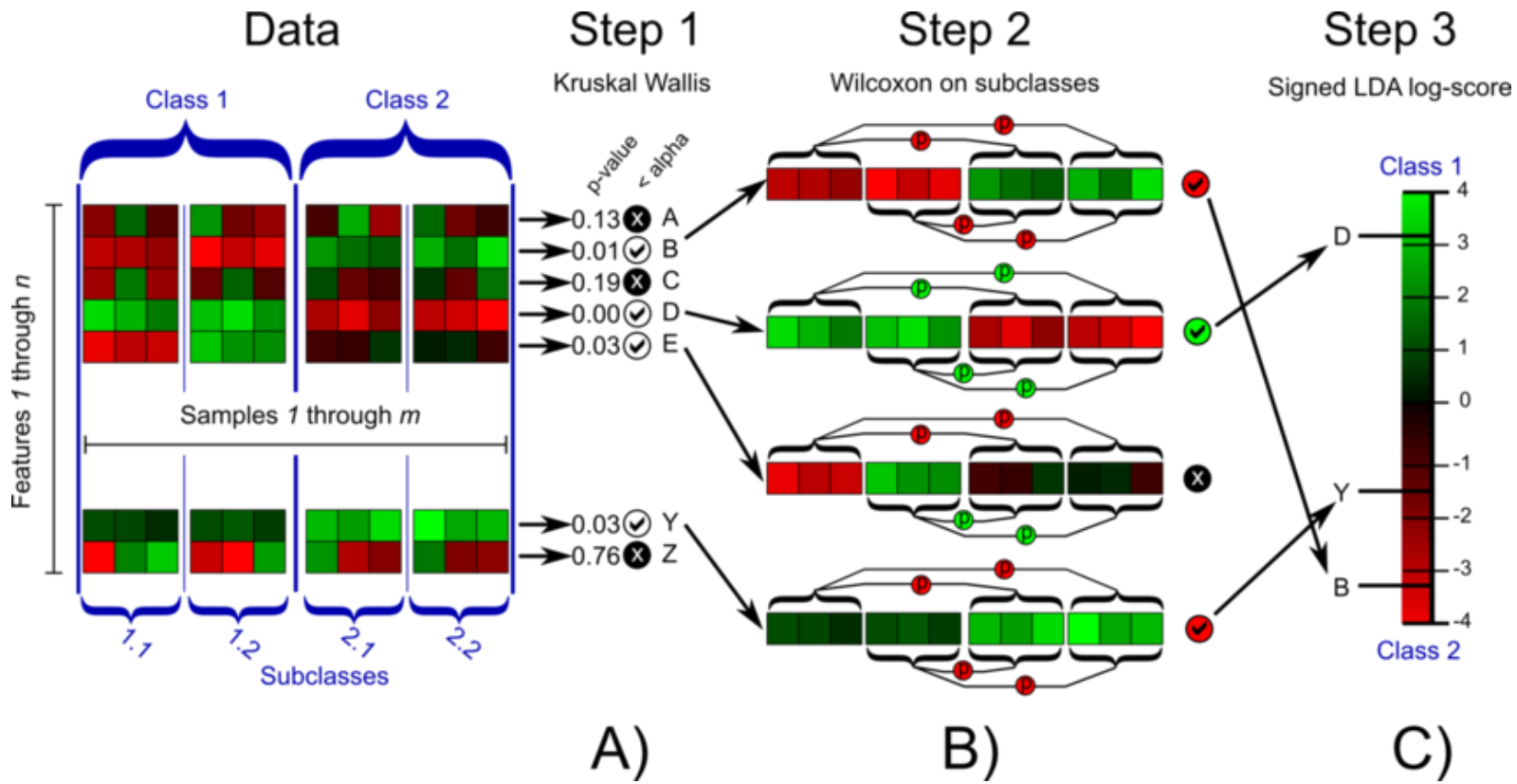


- Who's driving the differences?
- Can we find discriminatory biomarkers of different states?

Discriminatory Species

- LEfSe: Linear Discriminant Effect Size Analysis
- SIMPER: Similarity Percentages
  - Contribution of species to dissimilarity between treatments
- STAMP: Statistical Analysis of Metageonmic Profiles
  - Software package that helps identify statistically relevant features in metagenomes

LEfSe

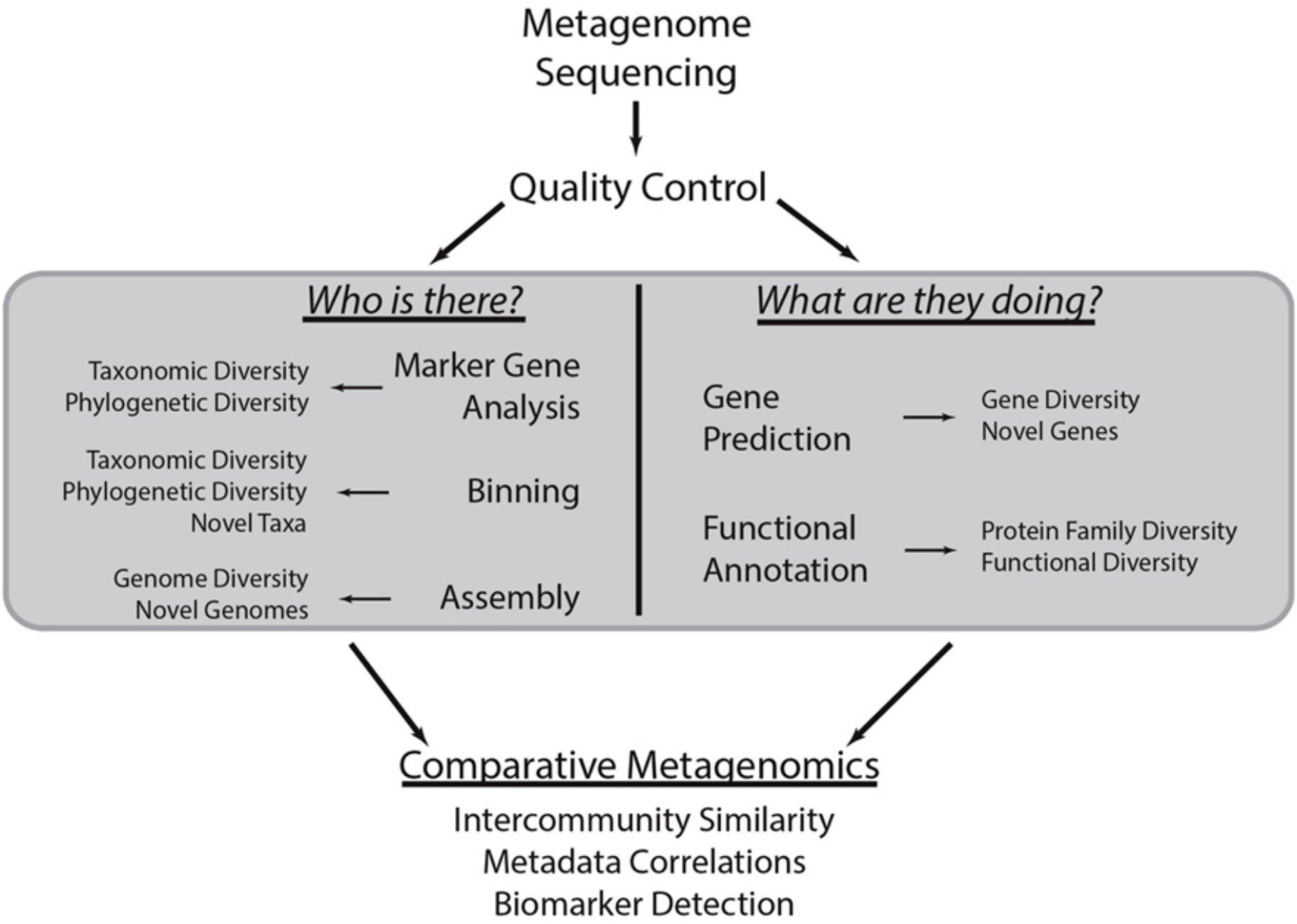


UNTARGETED METAGENOMICS: “SHOTGUN METAGENOMES”

Shotgun Metagenomics

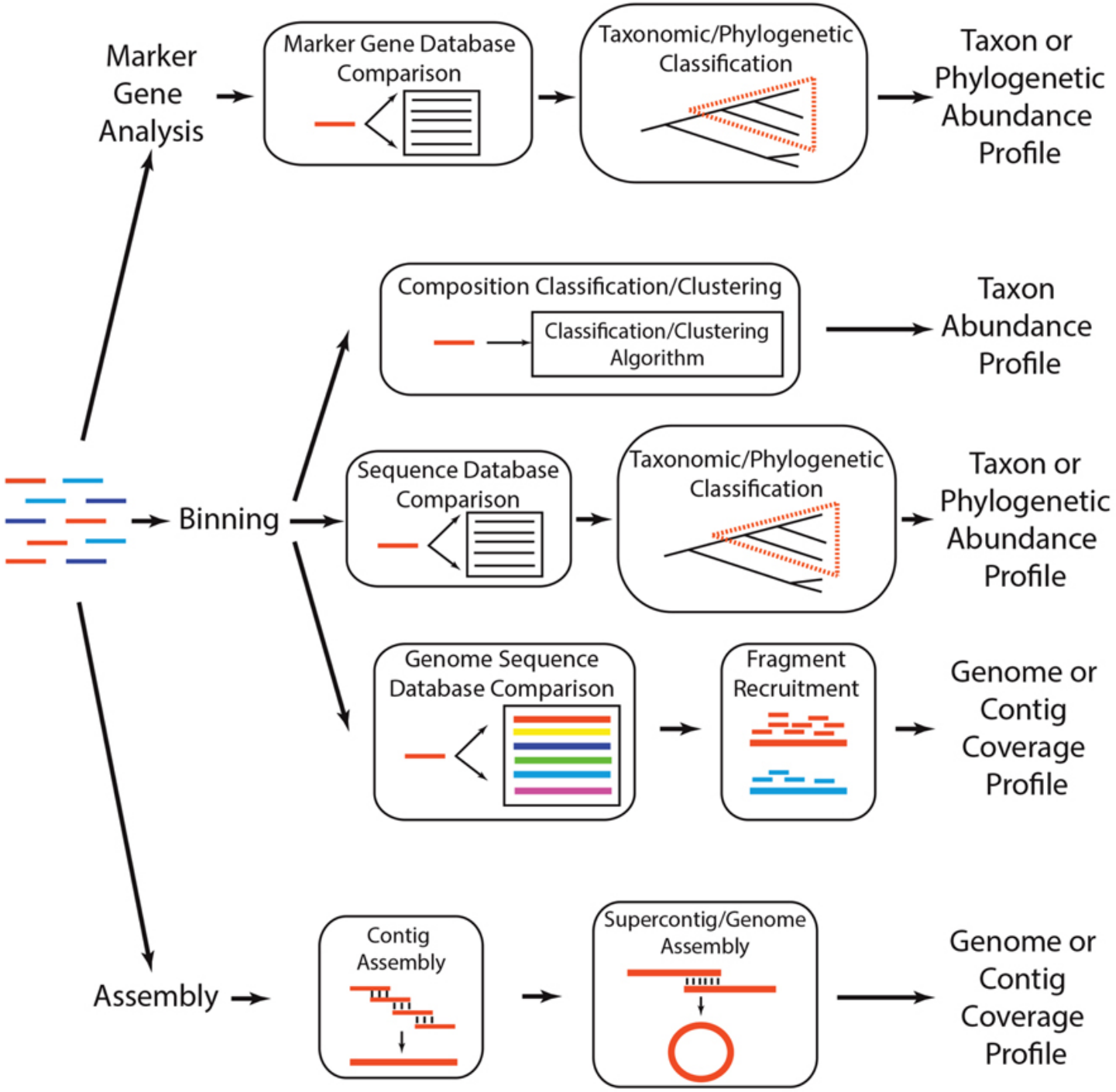
- Opportunities
  - No PCR step! This mean no bias, and errors
  - Not limited to one taxonomic category (e.g. will detect viruses, prokaryotes, eukaryotes)
  - Provides both taxonomic and functional (potential) information
  - Can better identify novel or highly divergent things

- Challenges
  - Expensive
    - e.g. soil targeted metagenome: $70/sample
    - Soil shotgun metagenome: $2000/sample
  - Requires a higher concentration of DNA/RNA; can be a challenge for low biomass samples
  - Computationally intensive
  - Contamination of host DNA



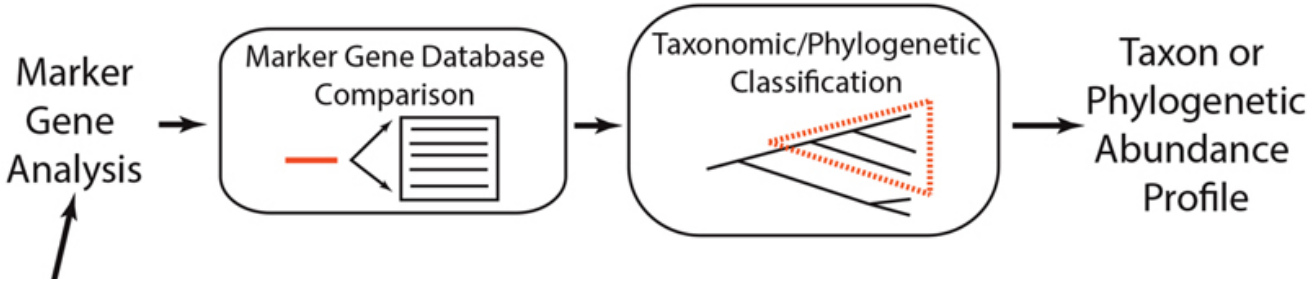
Sharpton (2014)

Who is there?



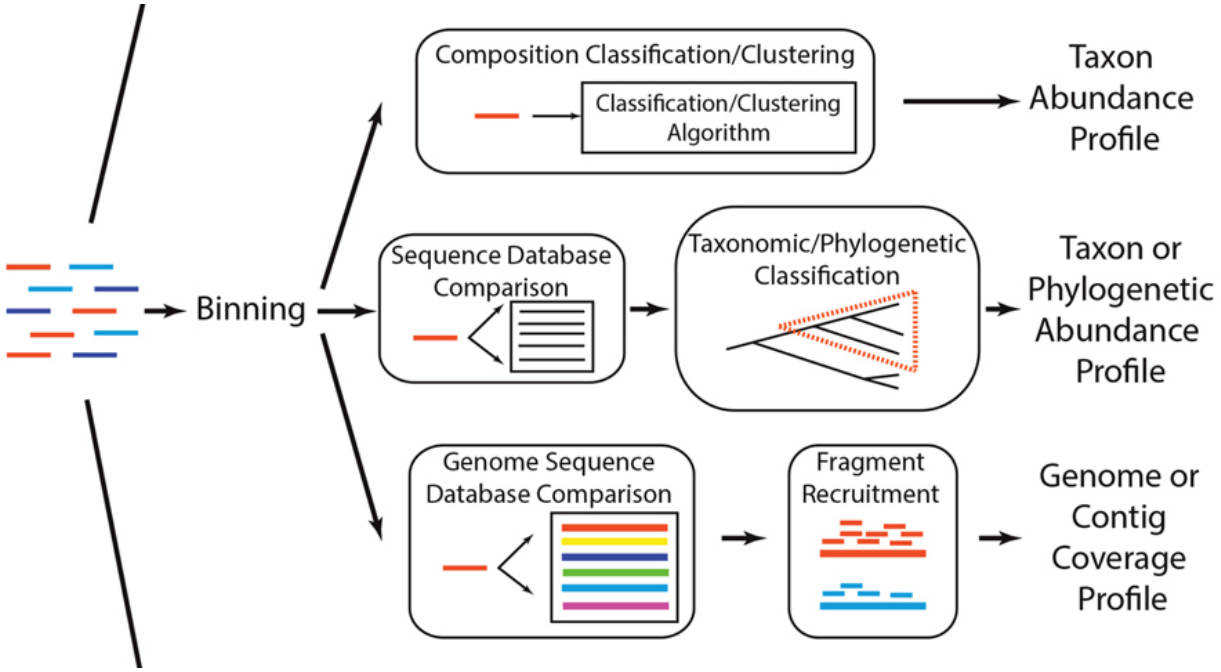
Sharpton (2014)

Who is there?



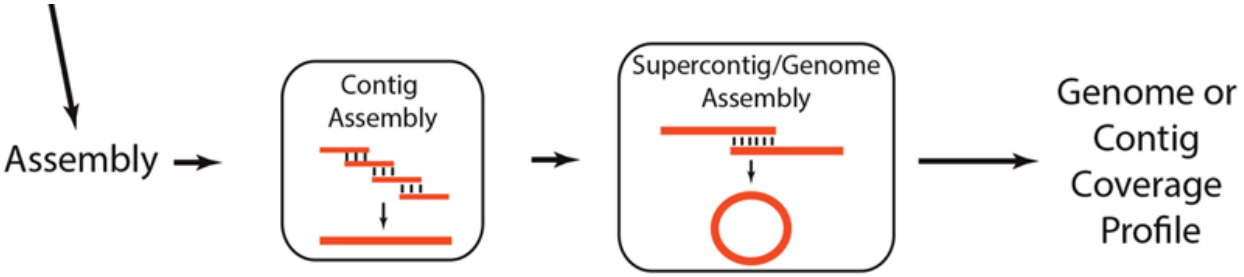
- Marker Gene Methods
  - Sequence similarity to known marker genes
    - E.g. MetaPhyler (Liu et al. 2011), MetaPhlAn (Segata et al. 2012)
  - Phylogenetic reconstruction
    - E.g. AMPHORA (Wu and Eisen 2008), PhyloSift (Darling et al. 2014), PhylOTU (Sharpton et al. 2011)
- Caveats:
  - Only works for taxa that have the marker
  - Requires reference database, so limited to what's in the database

Who is there?



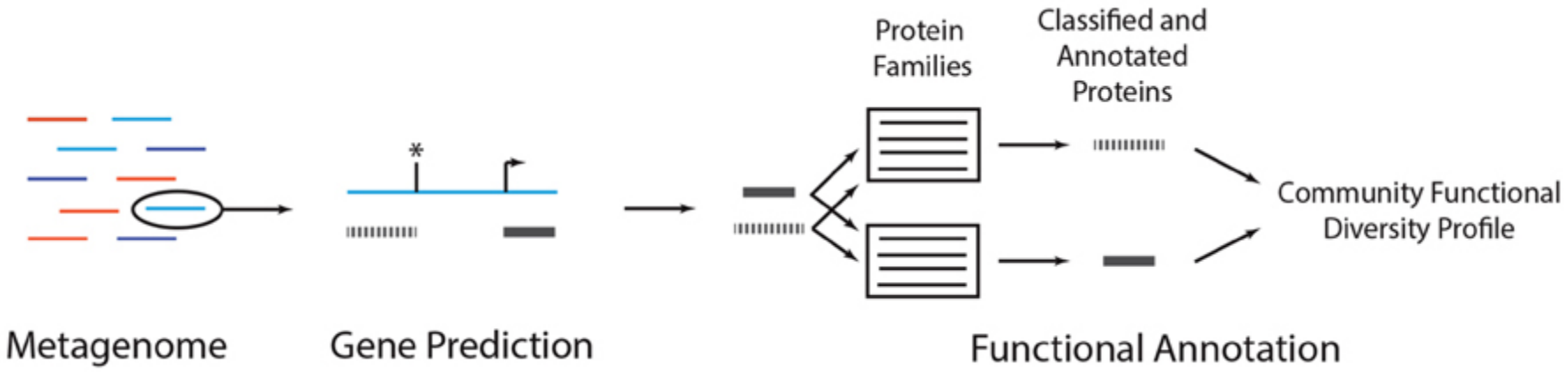
- Binning: attempts to assign every read to a taxonomic group
- Methods: Classification based on
  - Compositional Characteristics (e.g. GC content, tetramer frequency)
    - E.g. PhyloPithia (McHardy et al. 2007)
  - Similarity to reference
    - E.g. MEGAN (Huson et al. 2011), MG-RAST (Meyer et al. 2008)
  - Fragment recruitment
    - E.g. MOSAIK (Lee et al. 2013)
- Caveats:
  - Accuracy/specificity of annotation can vary widely across reads
  - Convergent evolution (horizontal gene transfer) reduces binning accuracty
  - Difficult to validate for novel organisms

Who is there?



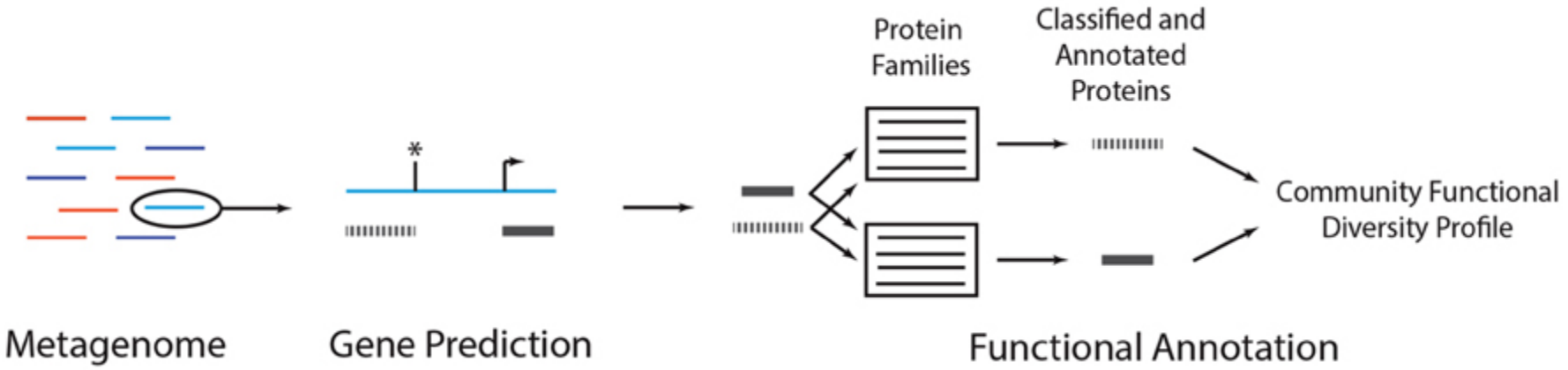
- Assembly: merges collinear reads from the same genome into longer contigs
- Opportunities:
  - More information about functional genes
  - Linking function to phylogeny (i.e. taxonomic markers like 16S)
  - Reconstructing whole genomes
  - Can do *de novo* assembly without reference genome
- Assemblers:
  - Many based on kmer searching and de Bruijn graphs e.g. Meta-IDBA (Peng et al. 2011), MetaVelvet (Namiki et al. 2012), MEGAHIT (Li et al. 2015)
- Challenges and caveats:
  - Short reads
  - High diversity (especially soils!)
  - Expensive to get the required coverage
  - Usually limited to most abundant in the community (not good for rare organisms)
  - Can generate chimeras

What are they doing?



- Types of functions and their relative abundance
- Compare communities for metabolic similarities
- Reveal functions associated with specific environment/treatment
- May reveal novel genes
- Insight on ecological conditions related to genes of unknown function

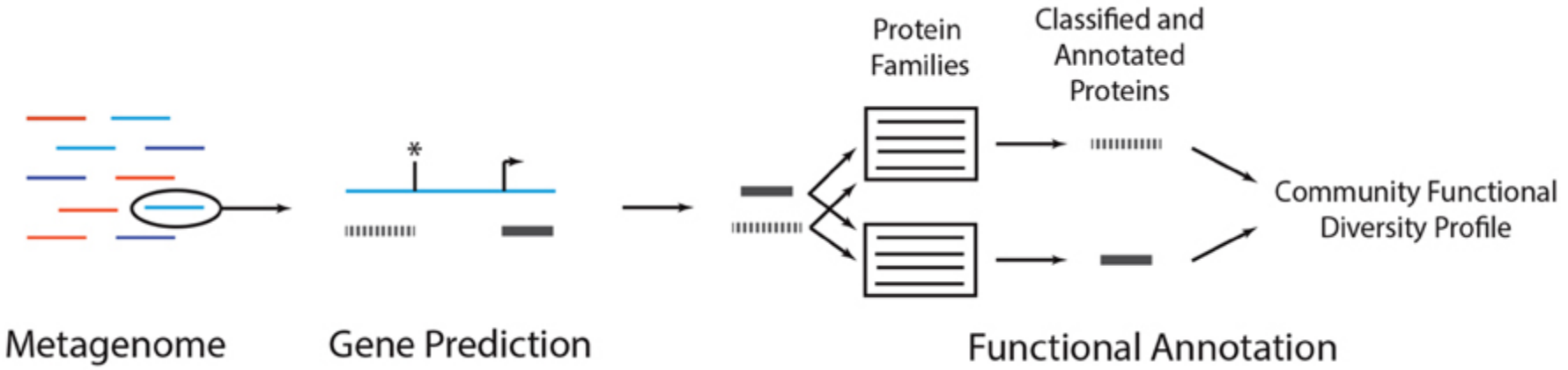
Sharpton (2014)



Gene prediction

- Methods:
  - Database: Fragment recruitment; aligns translated reads to homologs in database
    - E.g. USEARCH (Edgar 2010), VSEARCH (Rognes et al. 2016)
  - *De novo*: based on properties of microbial genes (e.g. length, codon use, GC bias)
    - E.g. MetaGene (Noguchi et al. 2006), Glimmer-MG (Kelley et al. 2012)

Sharpton (2014)



Functional Annotation

- Classify into protein families by comparing to database
- Common databases:
  - SEED: links family level functions to higher order subsystems
  - KEGG orthology groups: map to KEGG metabolic pathways
- Web servers:
  - MG-RAST, CAMERA, IMG/M
- Comparing communities:
  - ShotgunFunctionalizeR (Kristiansson et al. 2009), LEfSe (Segata et al. 2012), STAMP
- Caveats and Limitations:
  - Presence of gene doesn't mean it's active
  - Many protein families still have no known function
  - Database dependency can lead to phylogenetic bias, misses novel proteins/taxa
  - Assumes function is evolutionary static

Sharpton (2014)

Further Reading

Mothur:

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology. 2009 Dec 1;75(23):7537-41.

Westcott SL, Schloss PD. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ. 2015 Dec 8;3:e1487.

Metagenomics:

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5:209

Gilbert JA, CL Dupont (2011). Microbial Metagenomics: Beyond the Genome. Annual Review of Marine Science 3: 347-371

Zhou J, Z He, Y Yang, Y Deng, S Tringe, L Alvarez-Cohen (2015). High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. mBio 6: e02288-14

Howe, Adina, and Patrick SG Chain (2015). Challenges and opportunities in understanding microbial communities with metagenome assembly (accompanied by IPython Notebook tutorial). Frontiers in microbiology 6.

Sharon I, Banfield JF (2013). Genomes from metagenomics. Science. 342(6162):1057-8.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015 Jul 9;523(7559):208-11.

Informatics Reviews

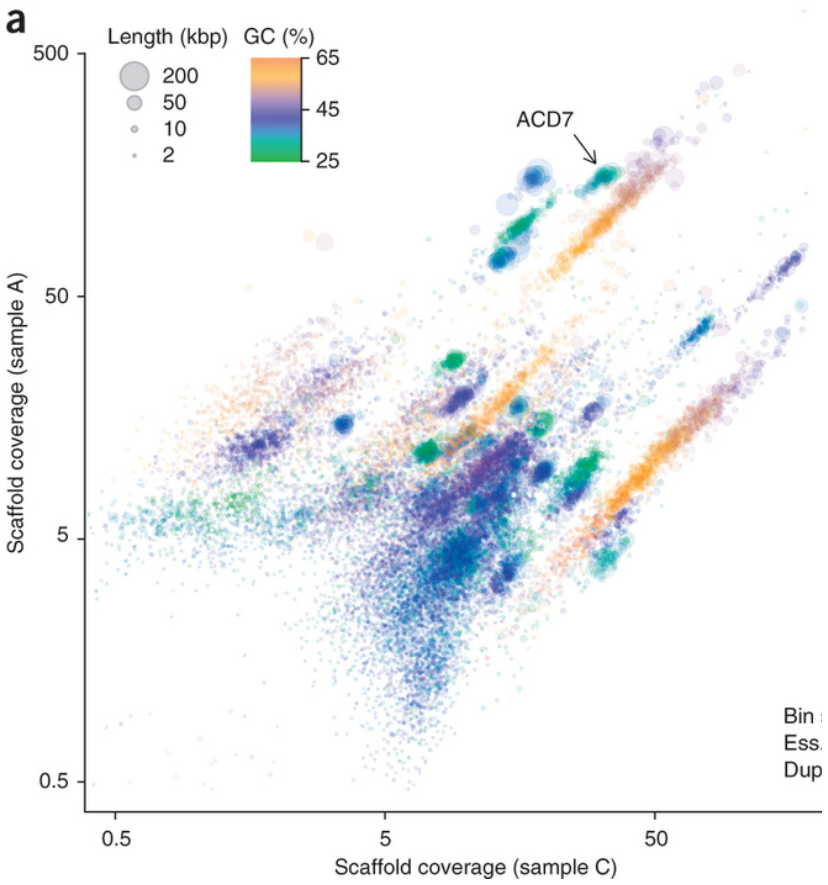
Mendoza ML, Sicheritz-Pontén T, Gilbert MT. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Briefings in bioinformatics. 2015 Feb 11:bbv001.

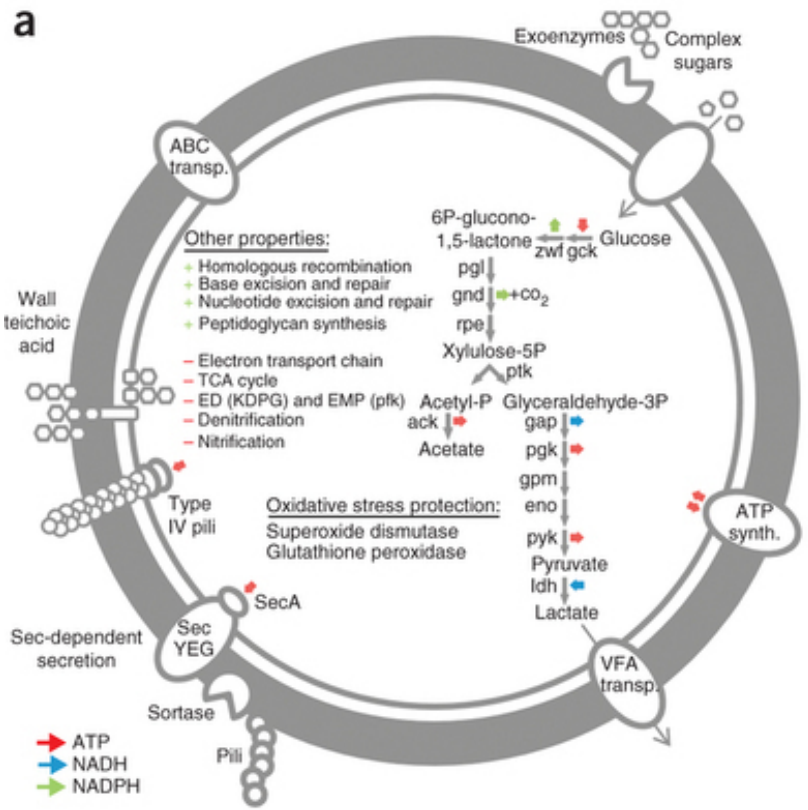
Dudhagara P, Bhavsar S, Bhagat C, Ghelani A, Bhatt S, Patel R. Web resources for metagenomics studies. Genomics, proteomics & bioinformatics. 2015 Oct 31;13(5):296-303.

Escobar-Zepeda A, de León AV, Sanchez-Flores A. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Frontiers in genetics. 2015;6.

Example: Deep sequencing + compositional binning (database-independent) to recover rare taxa

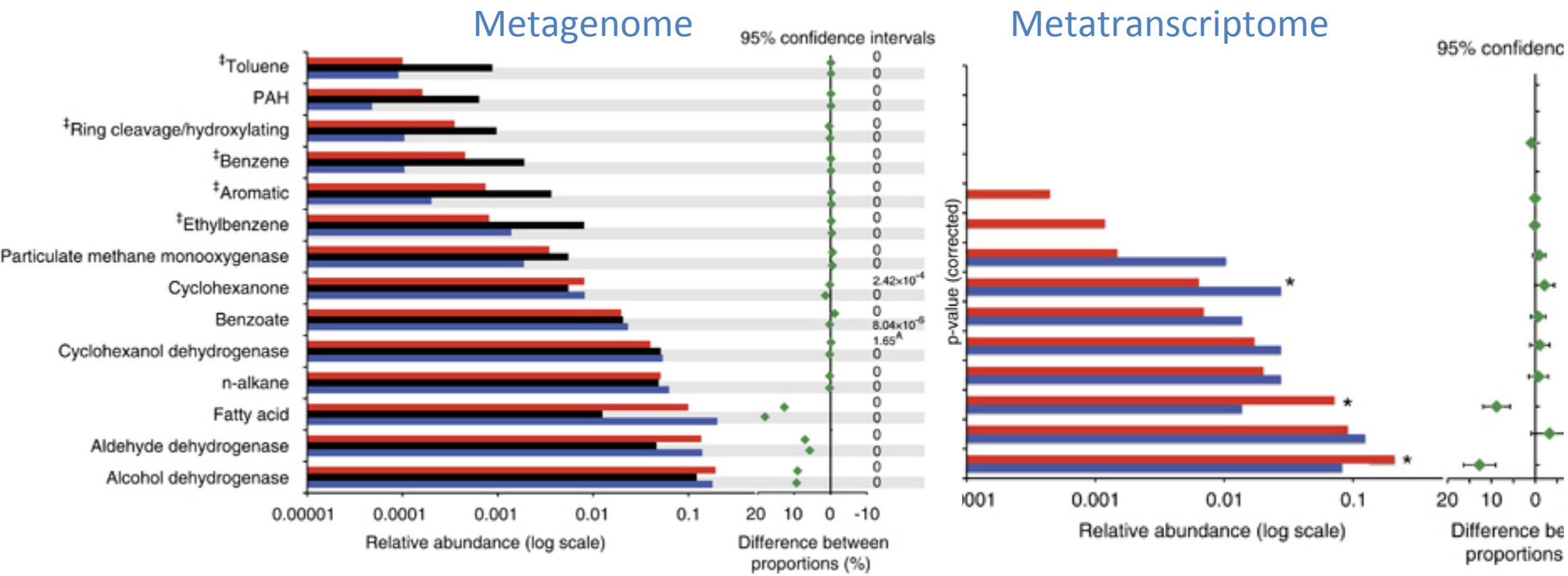
Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013) **Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes**. Nat Biotech 31:533-538





Example: Marine bacteria following Deepwater Horizon Oil Spill

Mason et al. (2012) **Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill.** Isme J. 6:1715-1727



Example: Binning + Functional Annotation in tundra soil metagenomes

ER Johnston, Luis M. Rodriguez, Chengwei Luo, Mengting M. Yuan, Liyou Wu, Zhili He, Edward A. G. Schuur, Yiqi Luo, James M. Tiedje, Jizhong Zhou and KT Konstantinidis. **Metagenomics Reveals Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem** Front. Microbiol., 25 April 2016