

Causes and implications of compositionality on microbiome interpretation



Western

Greg Gloor

Biochemistry, U. Western Ontario

<https://github.com/ggloor/compositions/presentations>

ggloor@uwo.ca @gbgloor



Motivation

- Compositional nature of data recognized from start
 - but unknown how to deal with compositional problems
- Ecology/Microbiology
 - proportion/relative abundance
 - rarefaction
- SAGE/Transcriptome
 - count normalizations
 - RPKM (Mortazavi: 2008)
 - TPM
 - scaling normalizations
 - TMM (edgeR Robinson: 2010)
 - RLE (DESeq, DESeq2 Anders: 2010)
- ratio data
 - qPCR (Vandesompele 2002)
 - compositional approaches

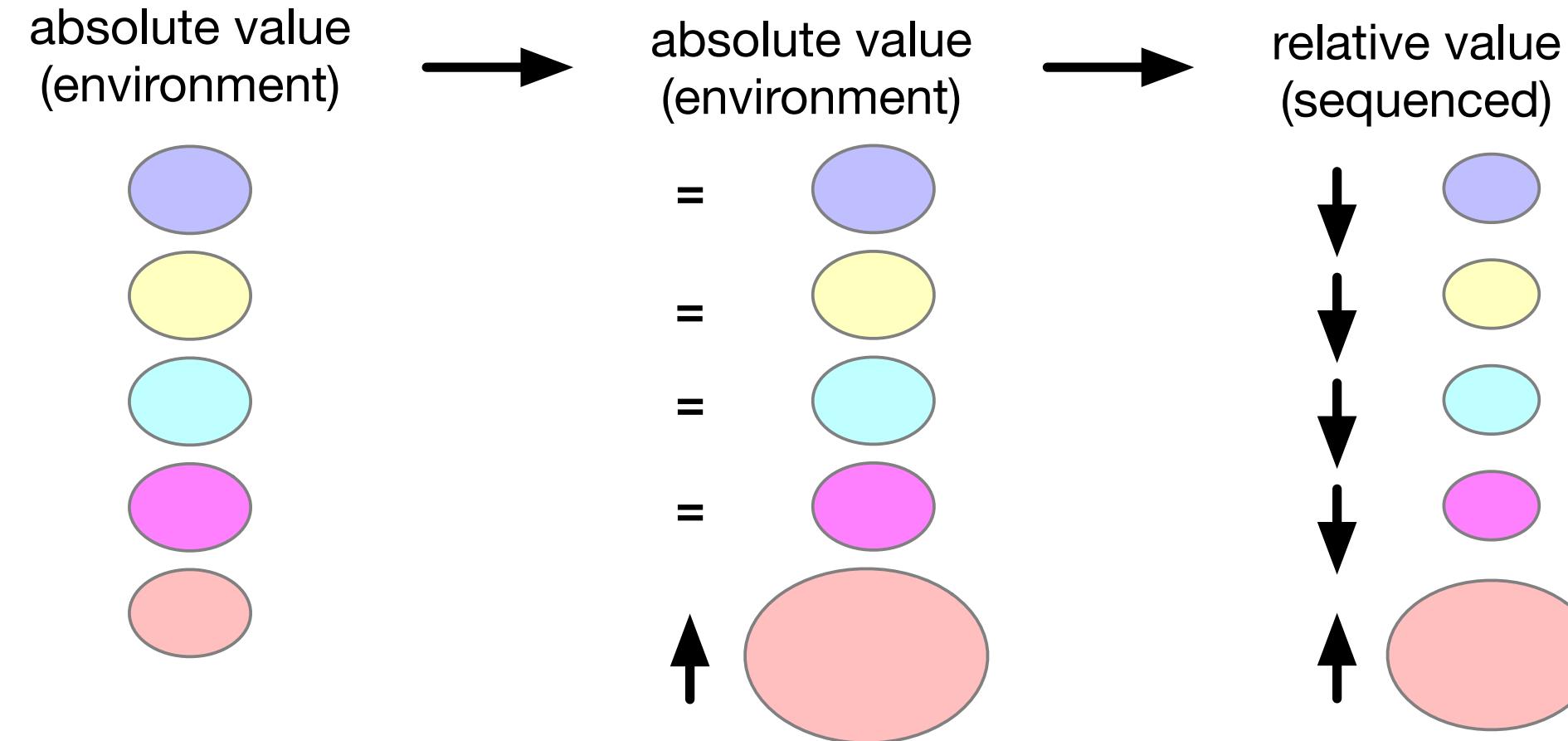


Objectives

- Demonstrate the use of multiple compositional approaches to examine microbiome datasets
- Show how to interpret the results
 - Exploratory PCA plots
 - effect sizes (features associated with the cohorts)
 - balances (best model to explain the cohorts)

Basic interpretation

- cannot ask "what has changed" (increased, decreased)
- can ask "what has changed relative to something" (what relationship has changed)



- single feature relative to a basket
 - clr, *lr
- single (or basket) relative to single (or basket)
 - balances, ilr

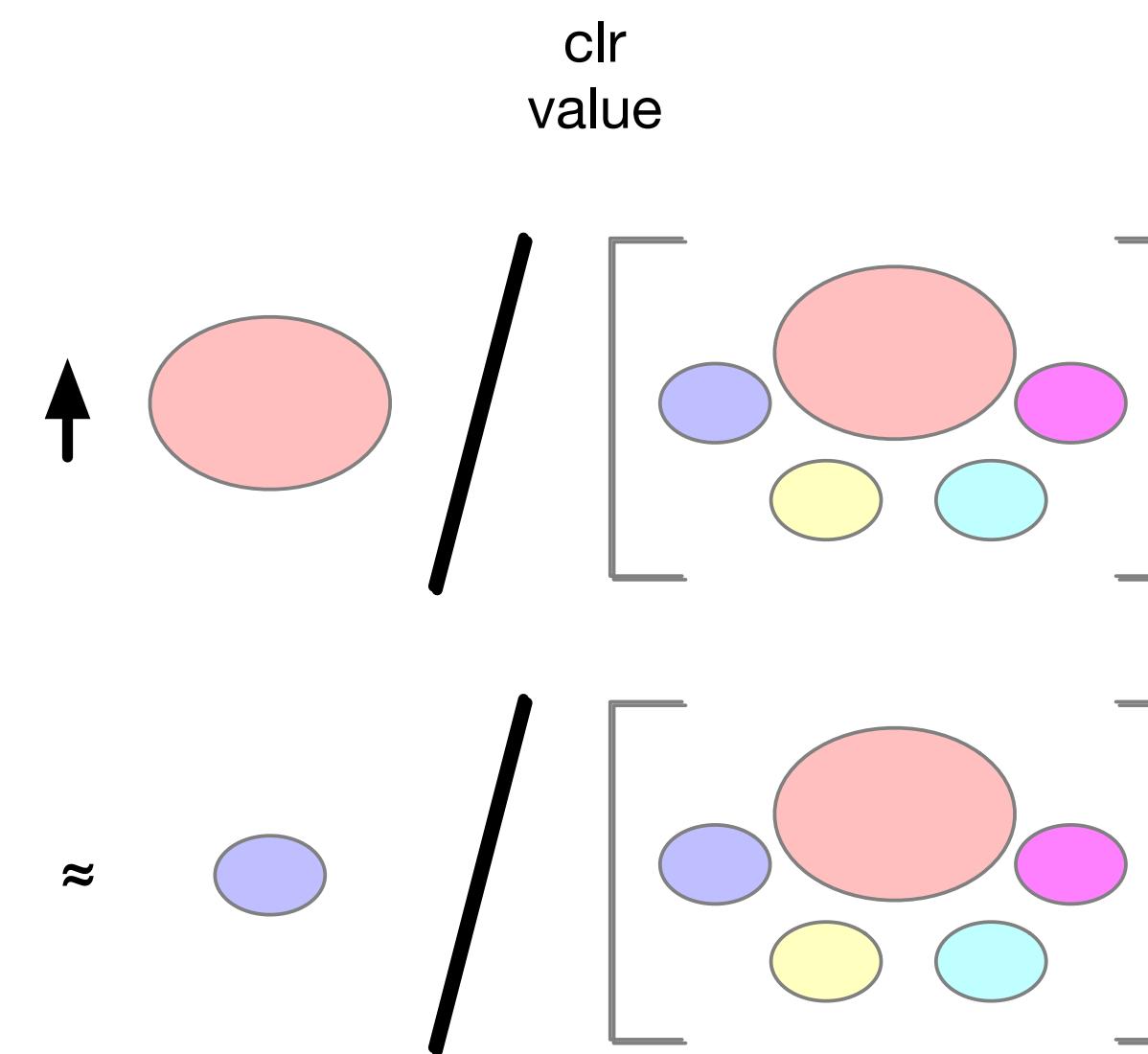
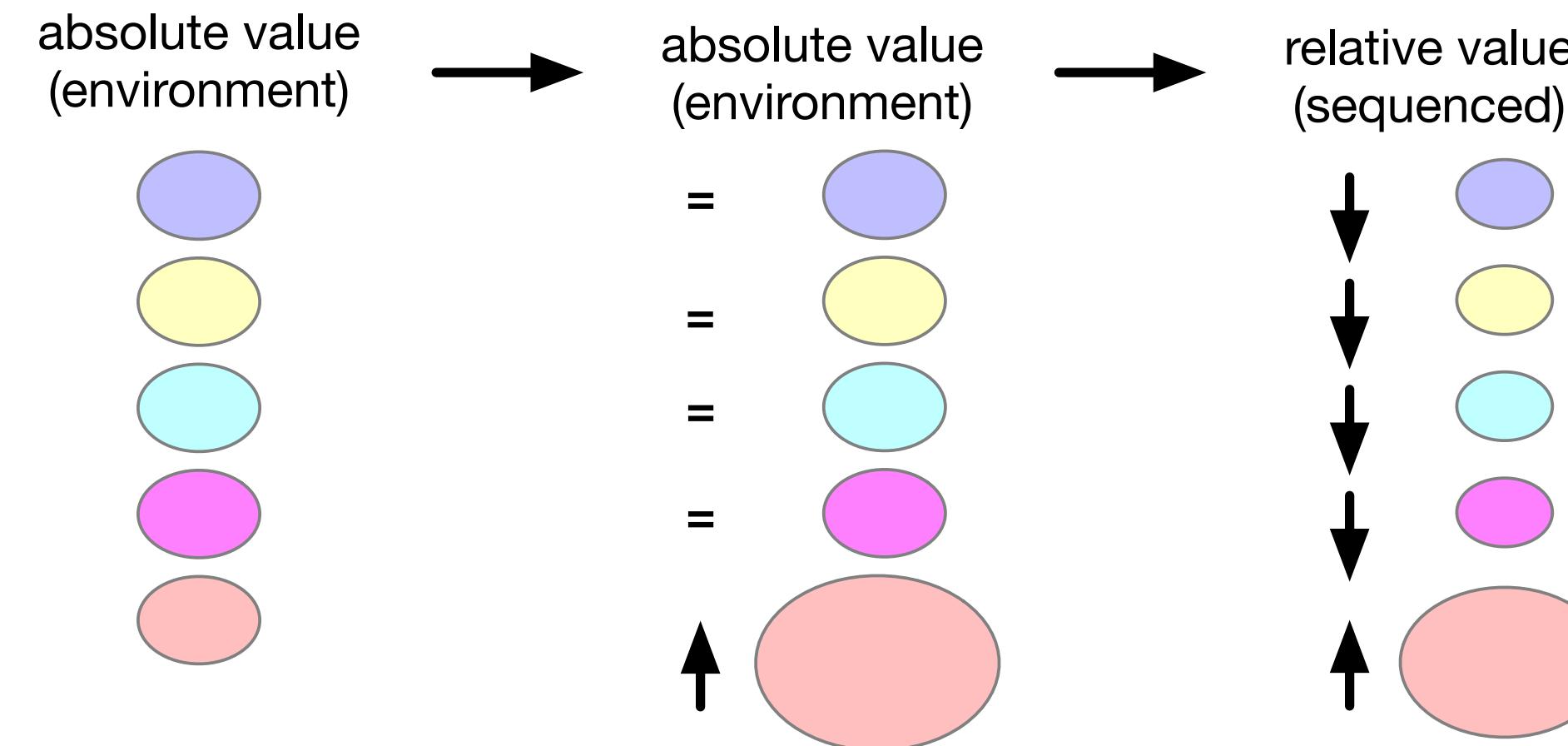
Basic interpretation - CLR

- cannot ask "what has changed" (increased, decreased)

- can ask "what has changed relative to something" (what relationship has changed)

- single feature relative to a basket

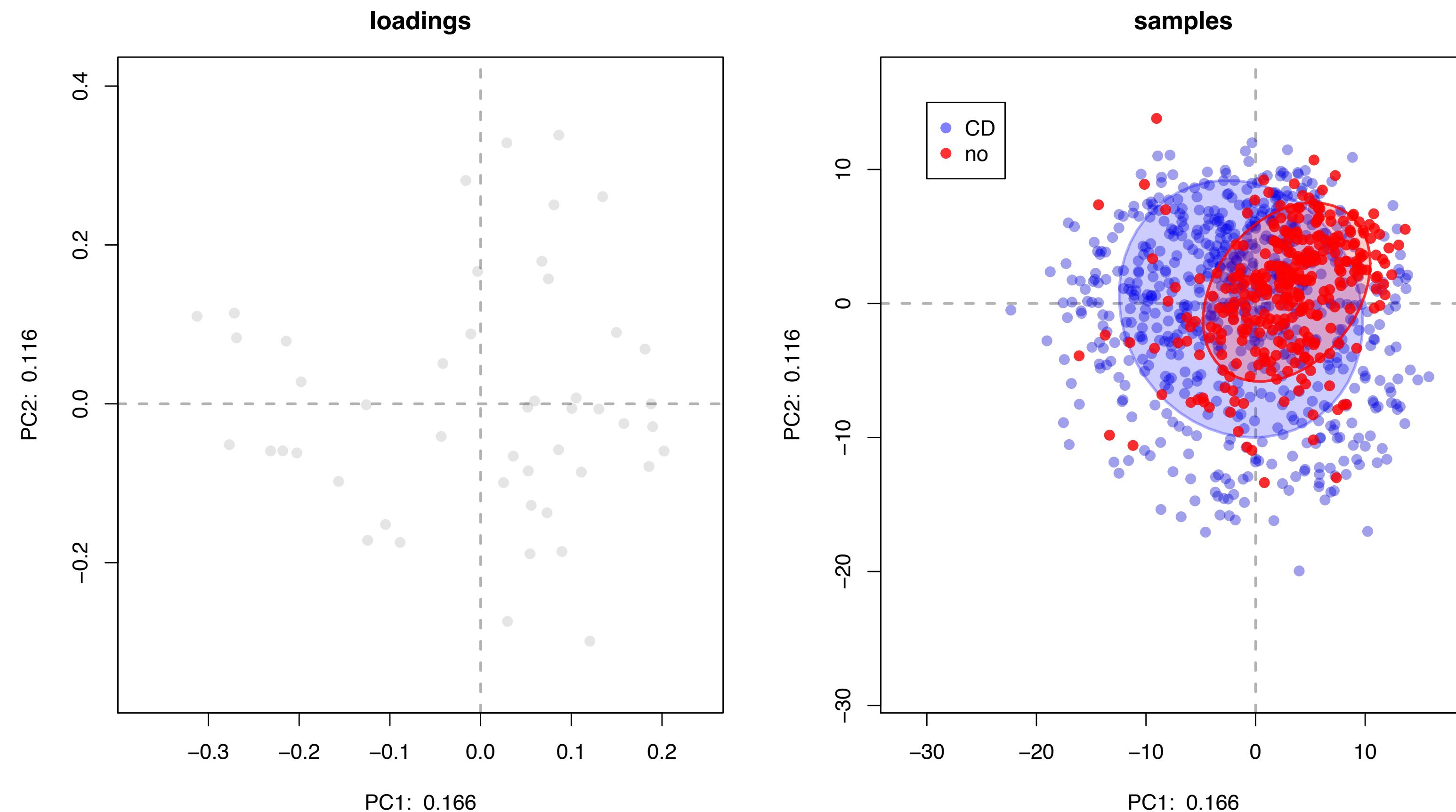
- clr, *lr
- single (or basket) relative to single (or basket)
- balances, ilr



Crohn's dataset

(Gevers CHM 2014)

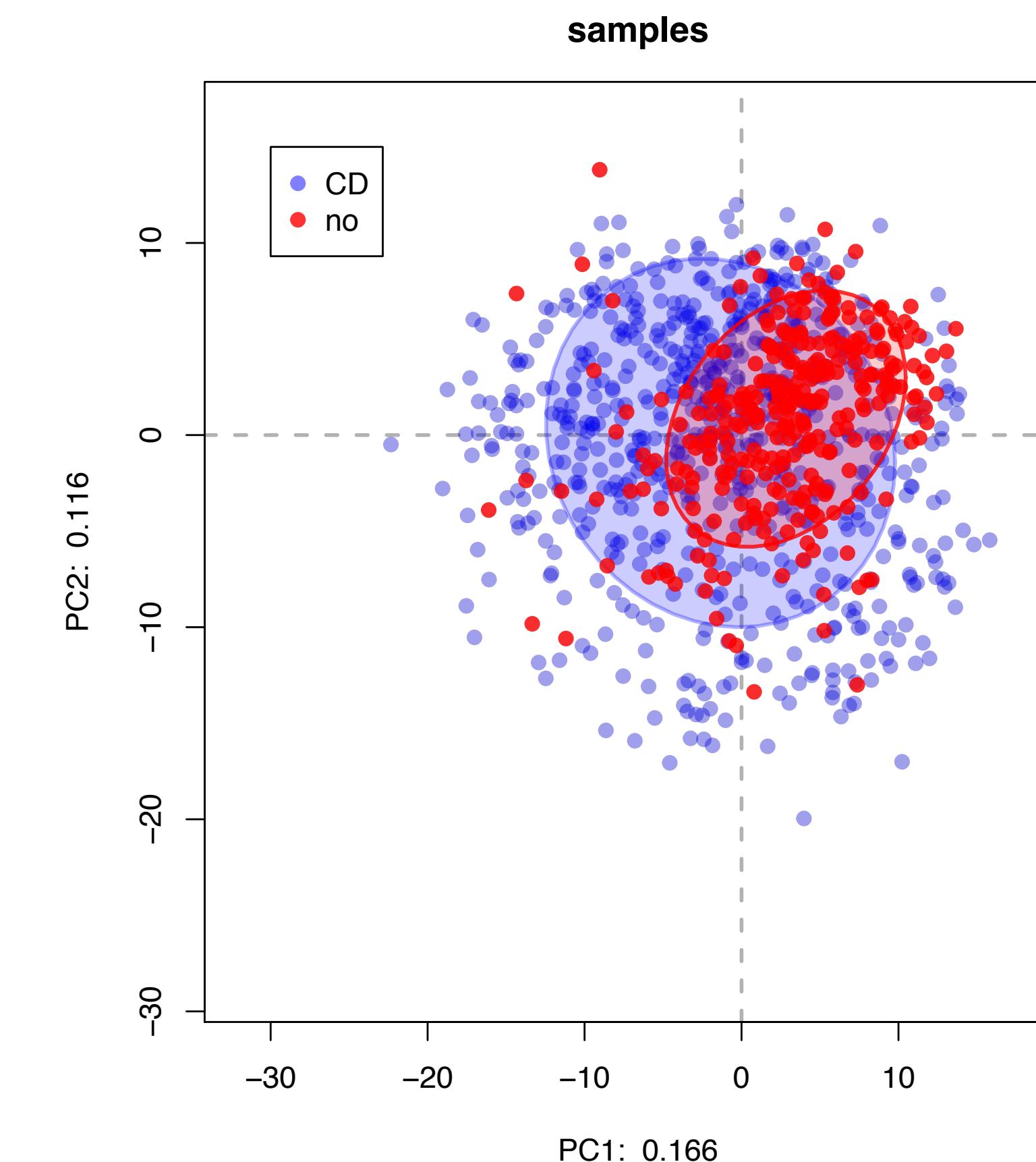
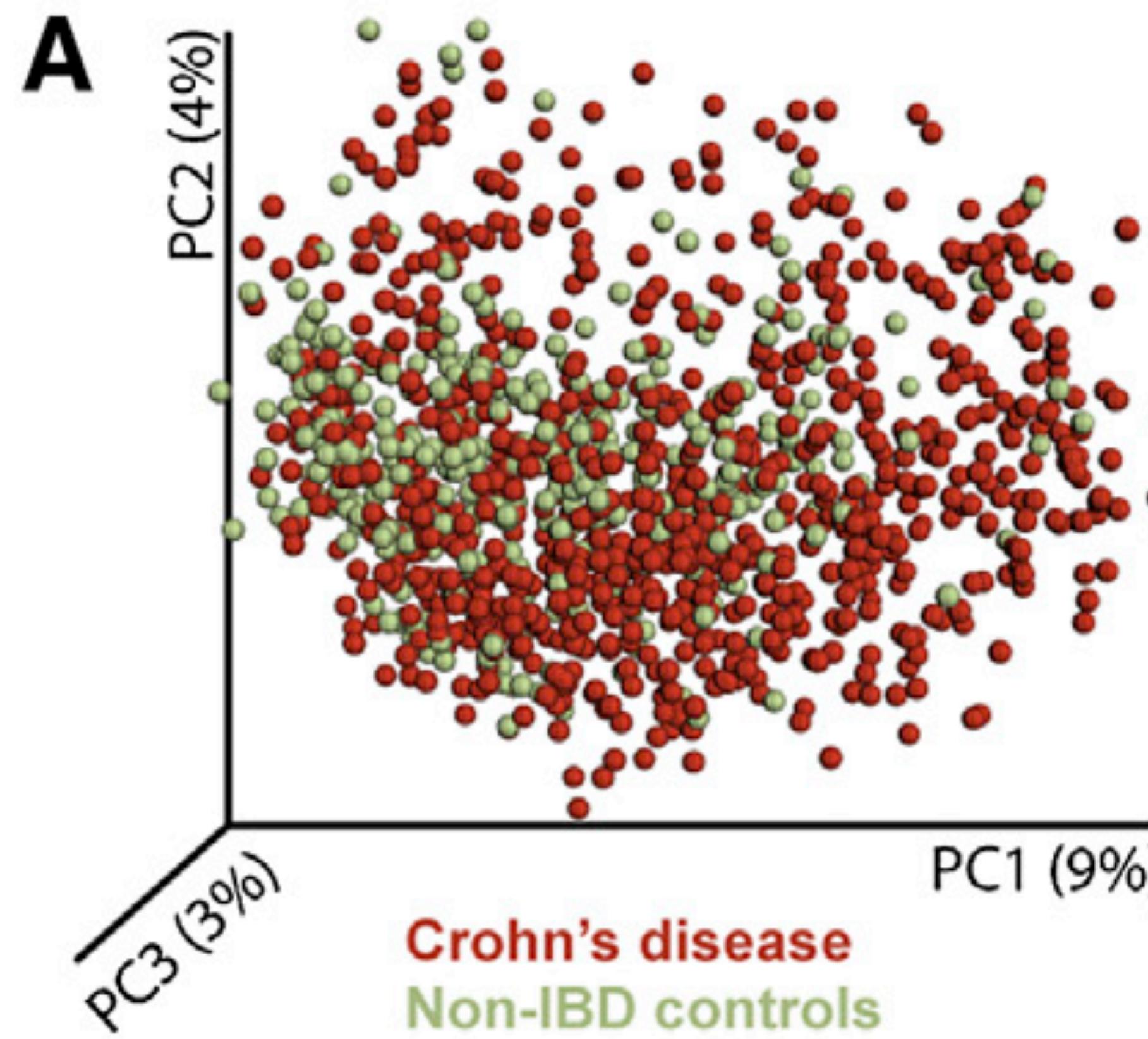
- Treatment naive CD (n=662) and controls (n=313). Analyzed at genus level (Rivera-Pinto mSystems 2018)
- Crohn's samples more variable and centre shifted
- Compositional PCA explains about 2X variance than does UniFrac in this dataset
- which taxa or group associate with each cohort?



Crohn's dataset

(Gevers CHM 2014)

- Treatment naive CD (n=662) and controls (n=313). Analyzed at genus level (Rivera-Pinto mSystems 2018)
- Crohn's samples more variable and centre shifted
- Compositional PCA explains about 2X variance than does UniFrac in this dataset
- which taxa or group associate with each cohort?



ALDEx2: what is different?

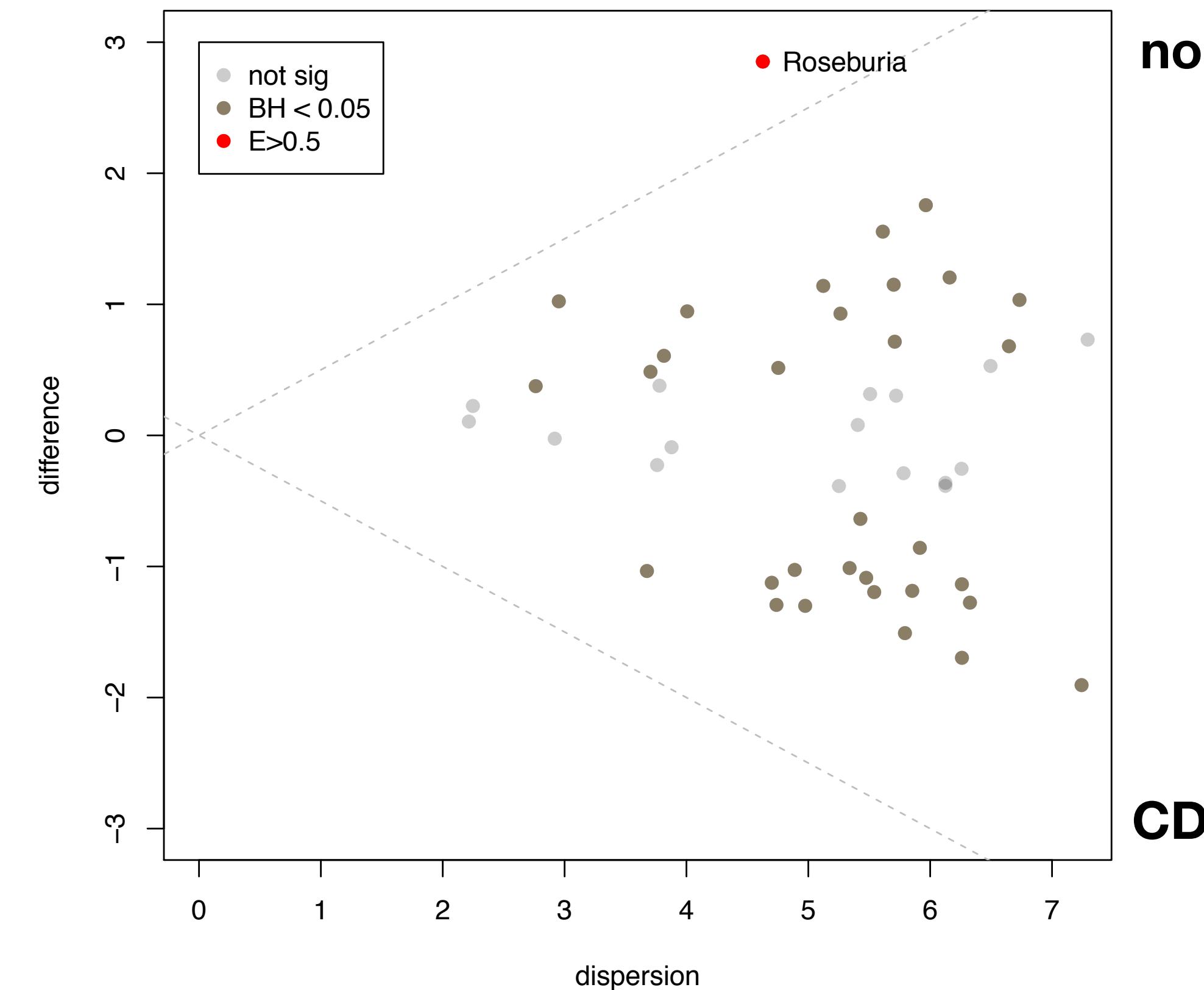
Fernandes PLoS ONE 2013

Uses CLR transform

- what has changed relative to everything else
- change is robust to sampling noise

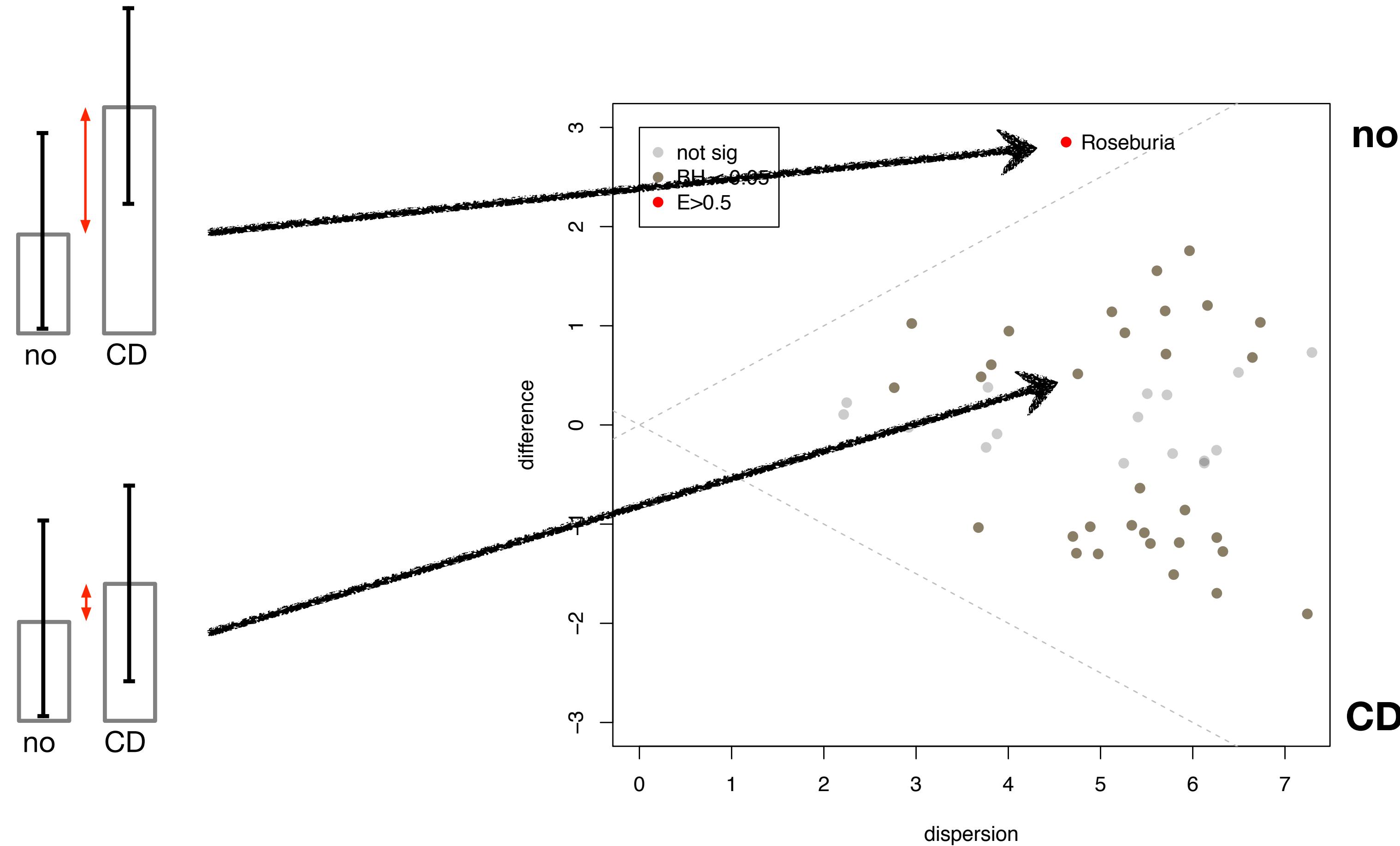
Only one genus with relatively large effect

- Roseburia is a relatively abundant butyrate producer
- Generally associated with a healthy gut microbiota



Effect plot

Gloor JCGS 2016



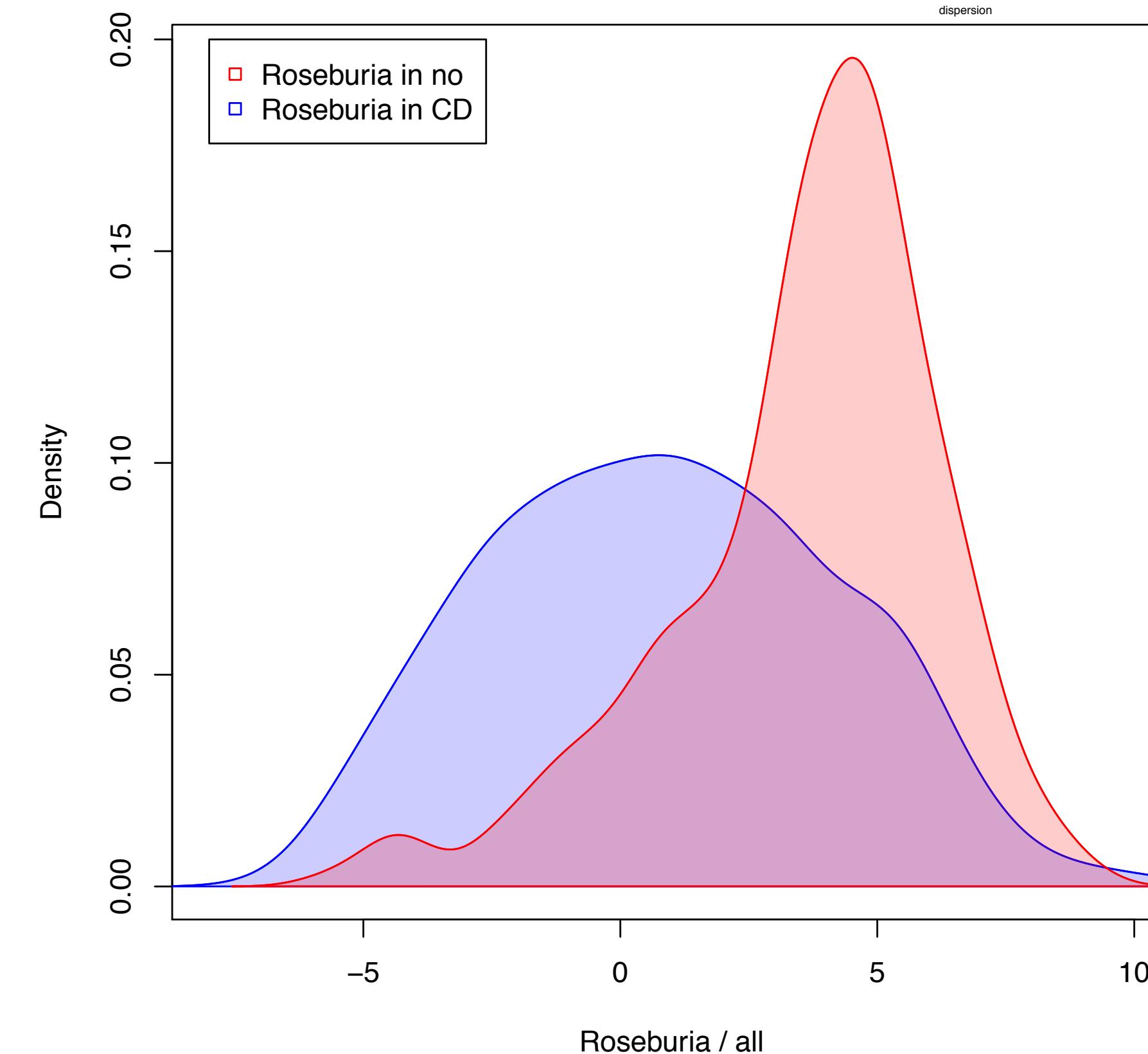
ALDEEx2: what is different

Uses CLR transform

- what has changed relative to everything else
- change is robust to sampling noise

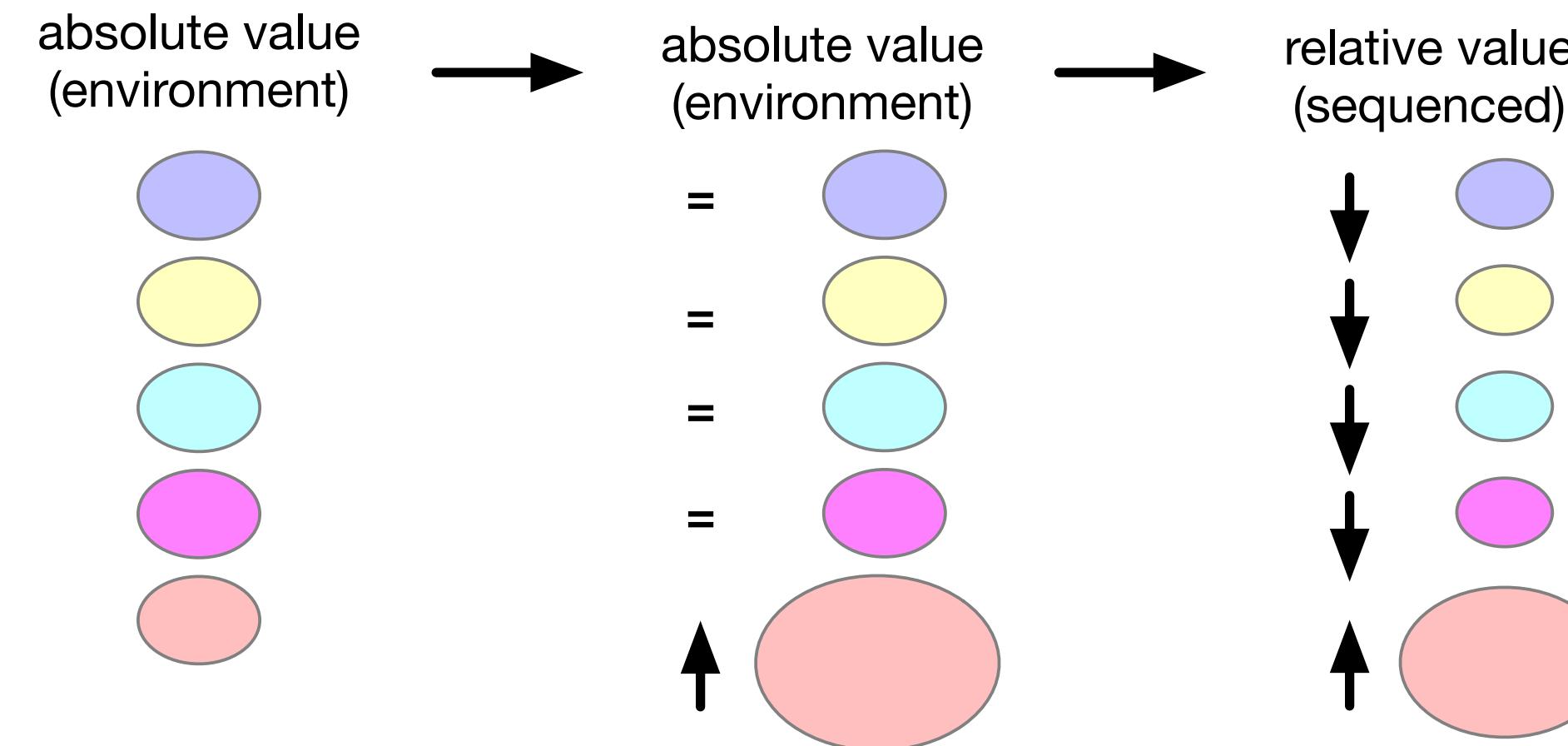
Only one genus with relatively large effect

- Roseburia is a relatively abundant butyrate producer
- Generally associated with a healthy gut microbiota

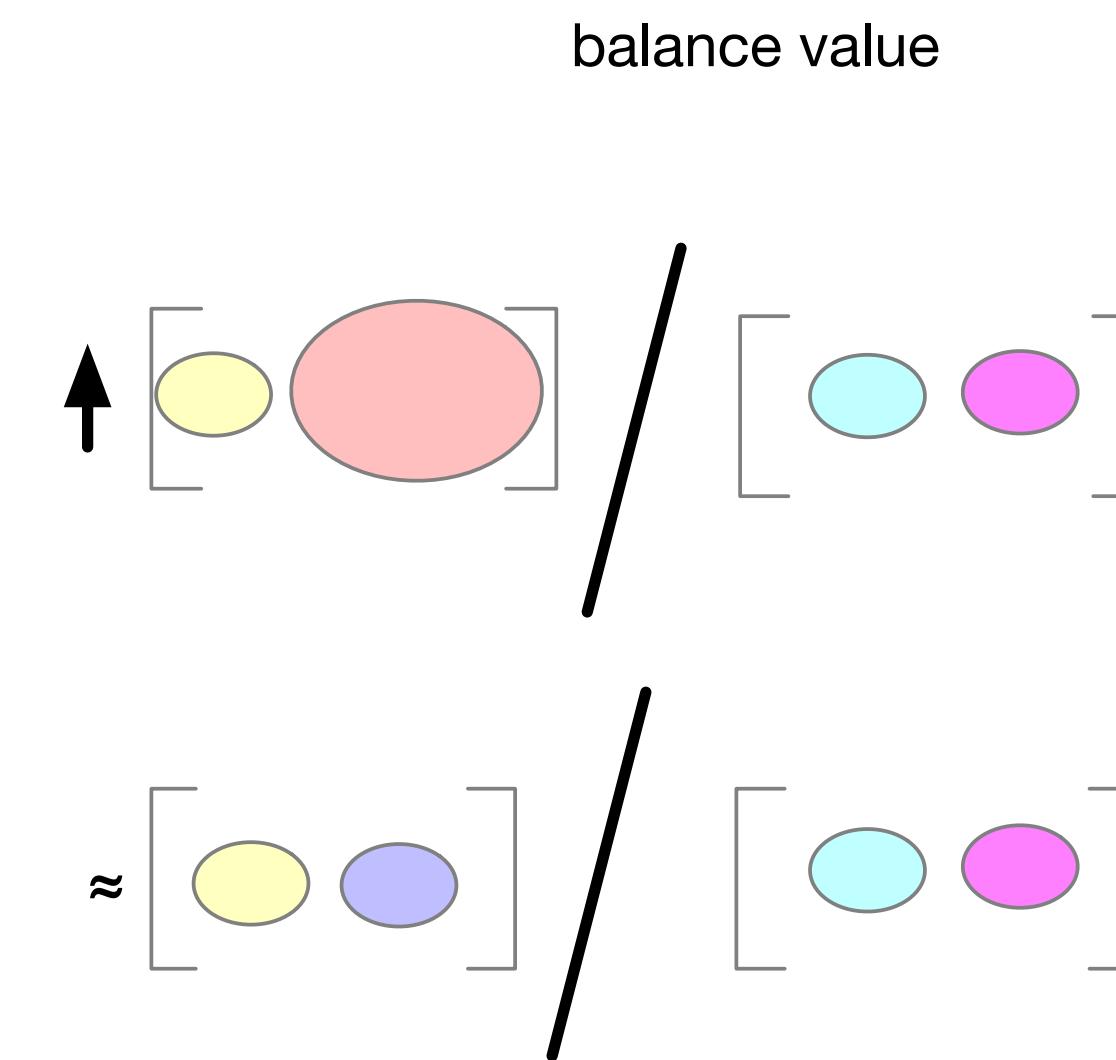


Basic interpretation: balances

- cannot ask "what has changed" (increased, decreased)
- can ask "what has changed relative to something" (what relationship has changed)



- single feature relative to a basket
 - clr, *lr
- single or basket, relative to single or basket
 - balances, ilr



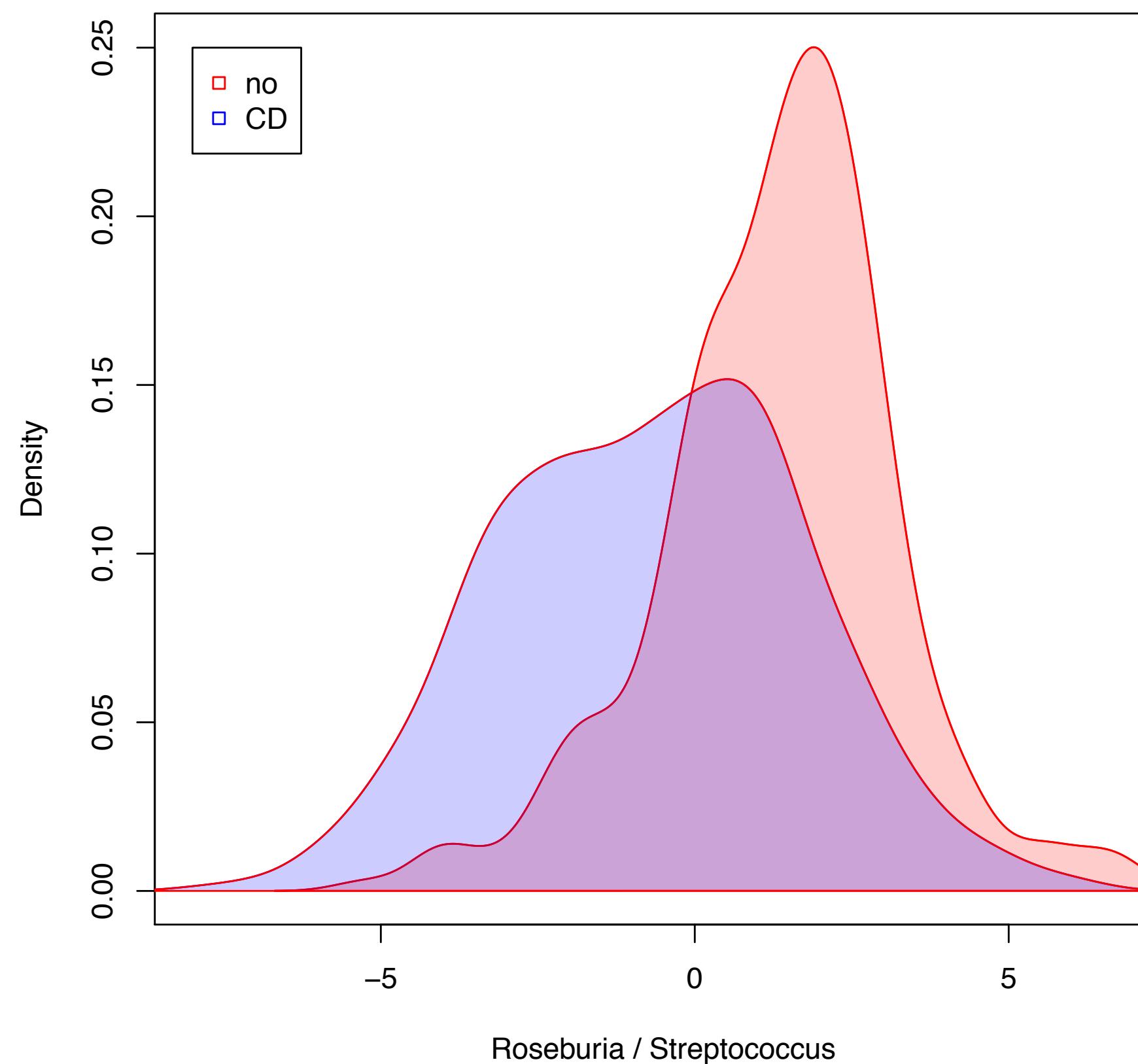
balance: pair(s) of features that are most explanatory (Erb & Quinn)

Uses balance transform (Quinn F1000 2018)

- what ratio between two (or three) taxa is different in the two groups

Only one pair of genera with a ratio of relatively large effect

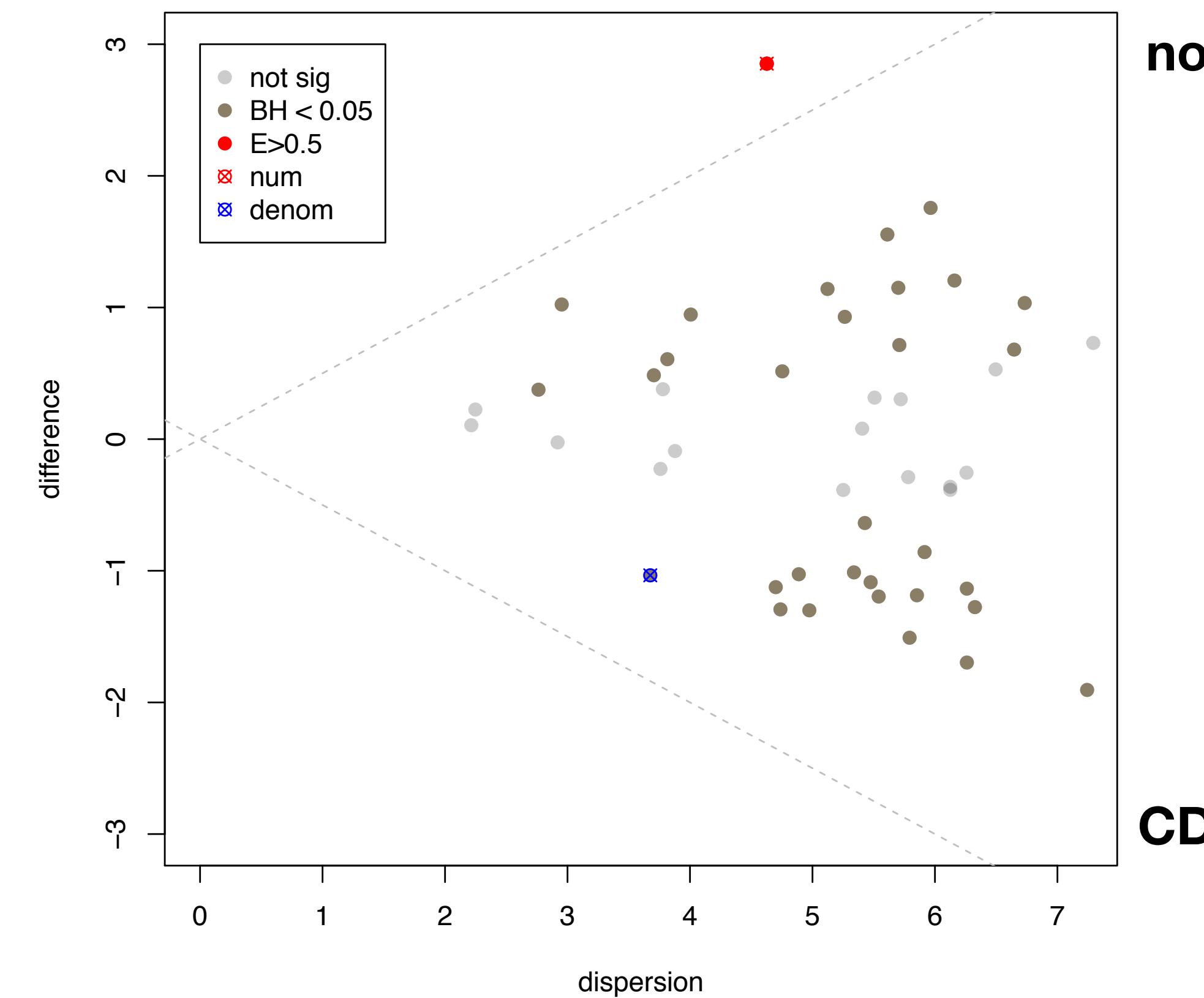
- **Roseburia is a relatively abundant butyrate producer**
- **Streptococcus is often associated with dysbiosis**



balance: pair of features that are most explanatory

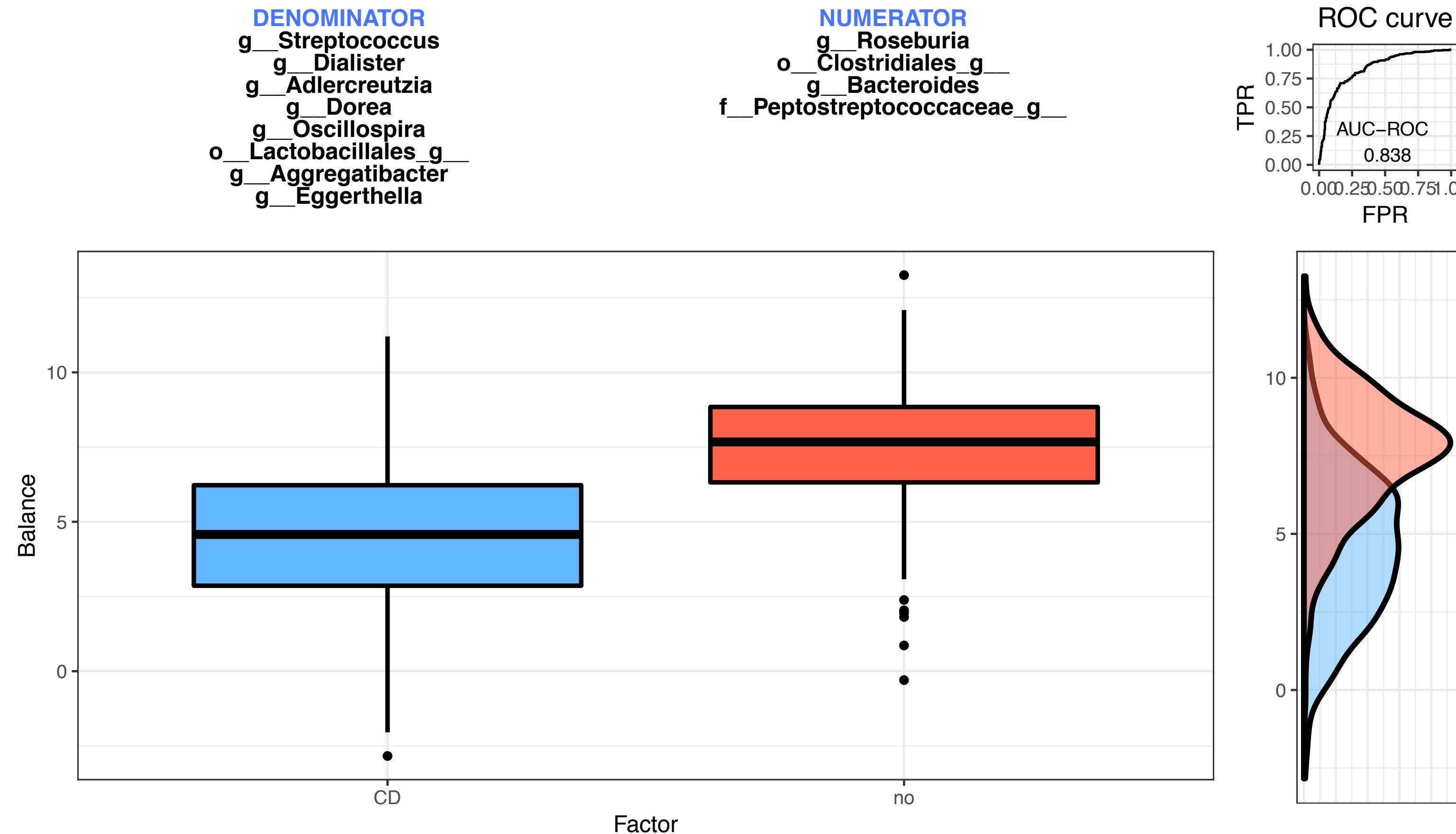
Balance result congruent with univariate CLR analysis

Roseburia and Streptococcus have the largest relative effects in the normal and CD cohorts

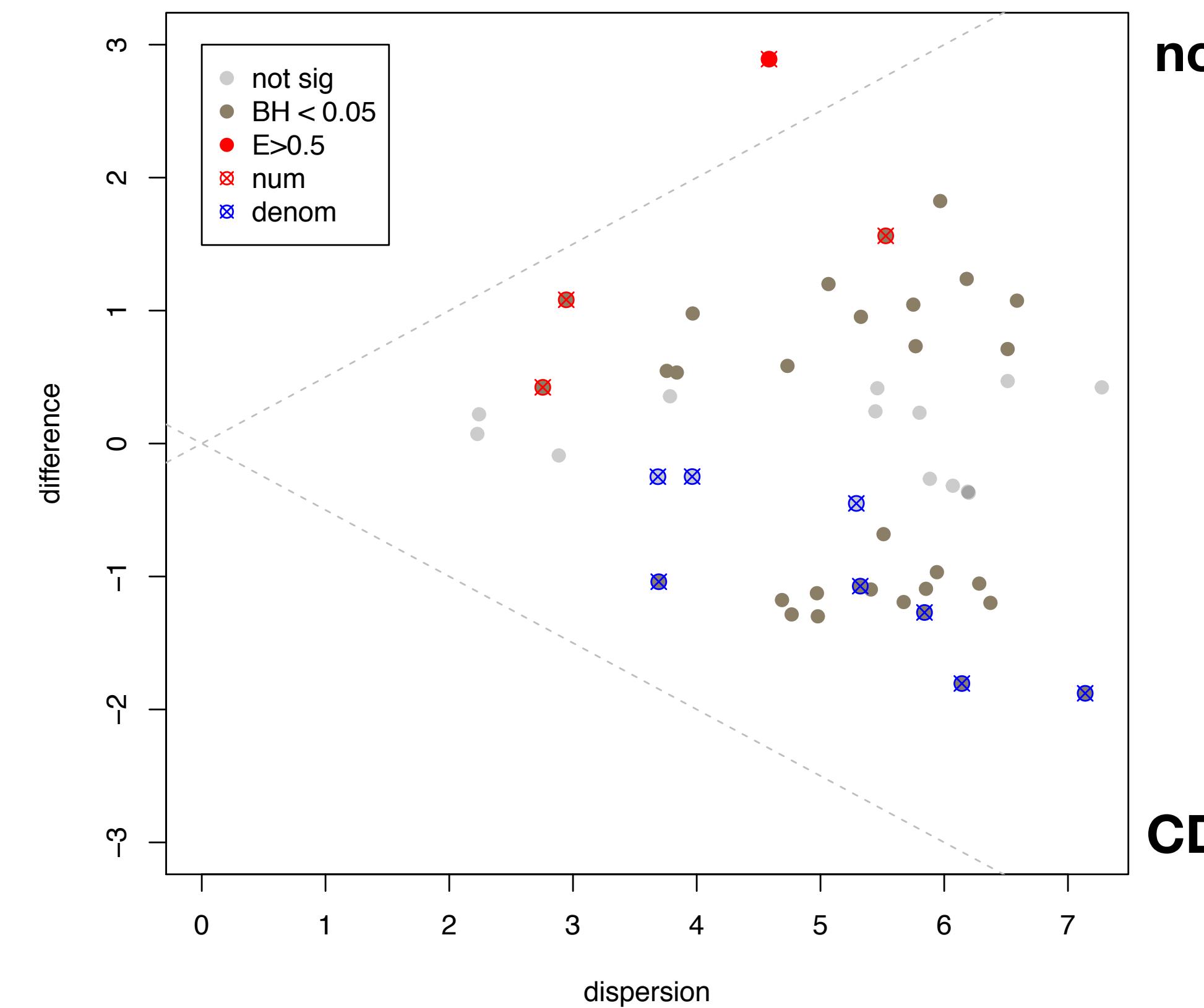
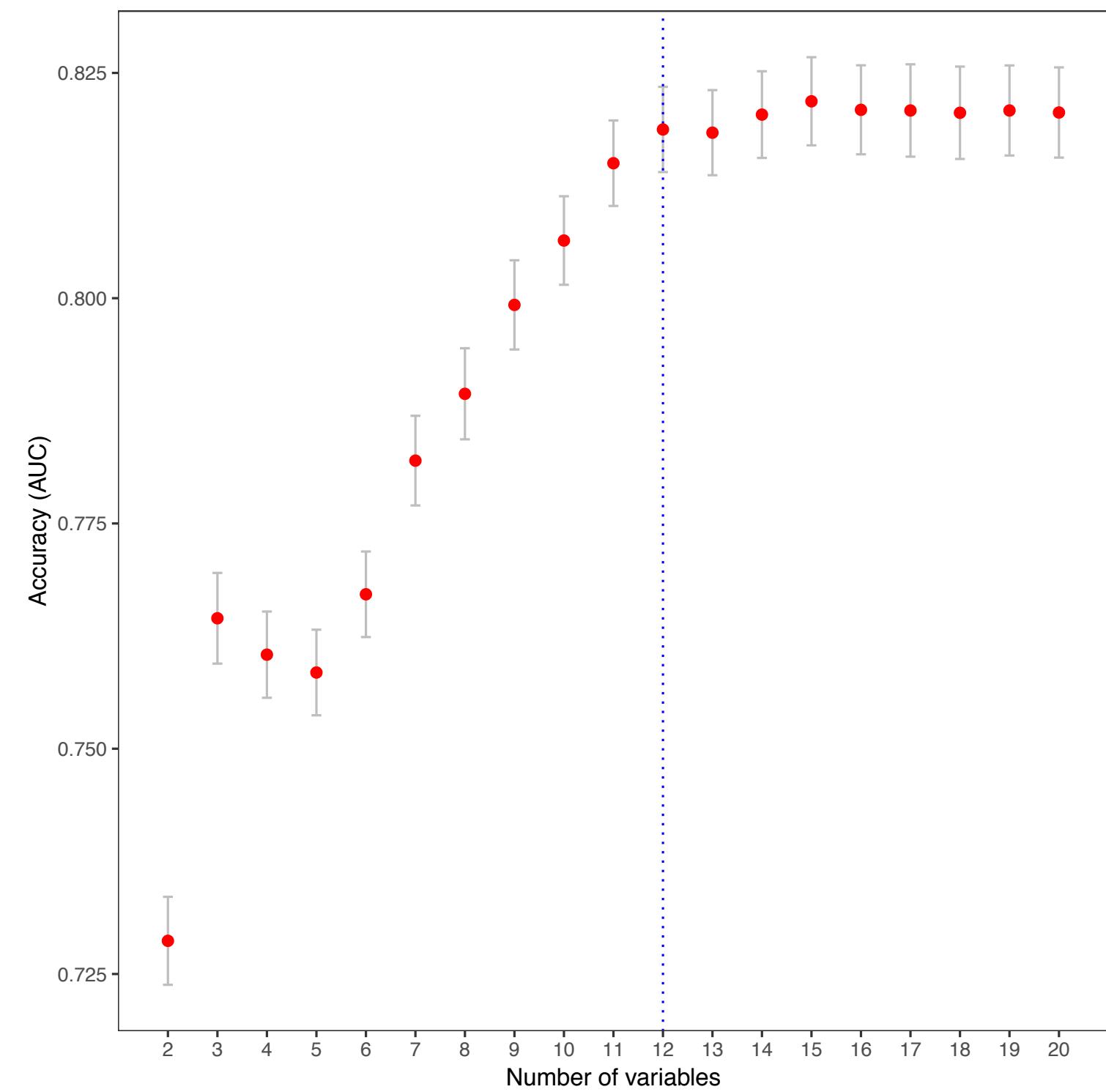


selbal: "basket of ratios" features that are most explanatory

Rivera-Pinto mSystems 2018



selbal



Intermediate summary

- analysis must be re-cast as "what has changed relative to a standard"
- multiple ways of choosing standard
 - one vs. many - clr, *lr, qPCR
 - one vs. one, many vs. many - balances, ilr
- full suite of tools available
 - multivariate, univariate, bivariate approaches
- results are internally consistent within and between approaches

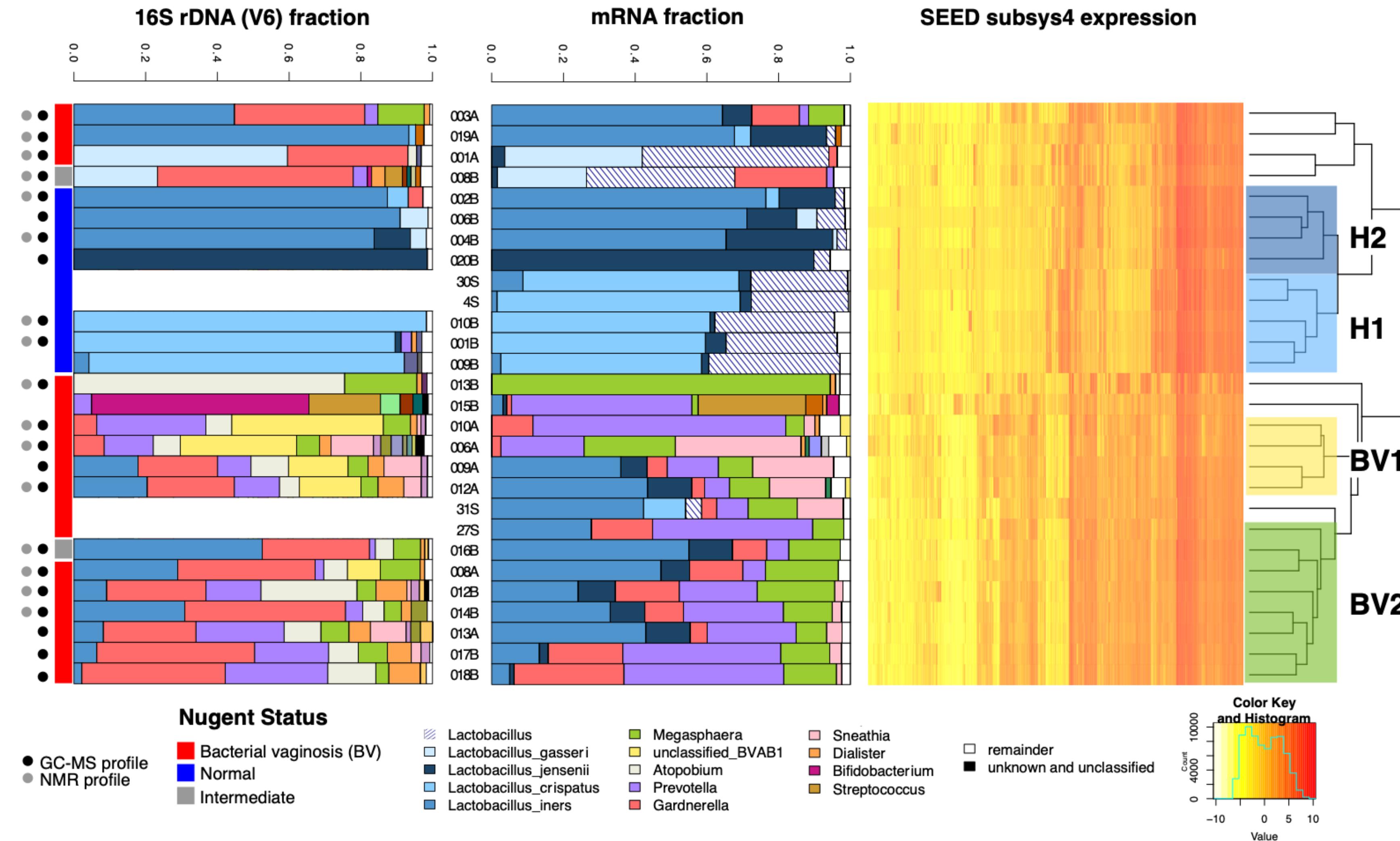
Meta-transcriptome of an ecosystem

- characterize mRNA of all species at once
- Information about what is being made now
- Unbalanced
 - Different conditions can have different taxonomic compositions
 - Both absolute and relative abundance of the taxa and their transcripts can change

Bacterial vaginosis

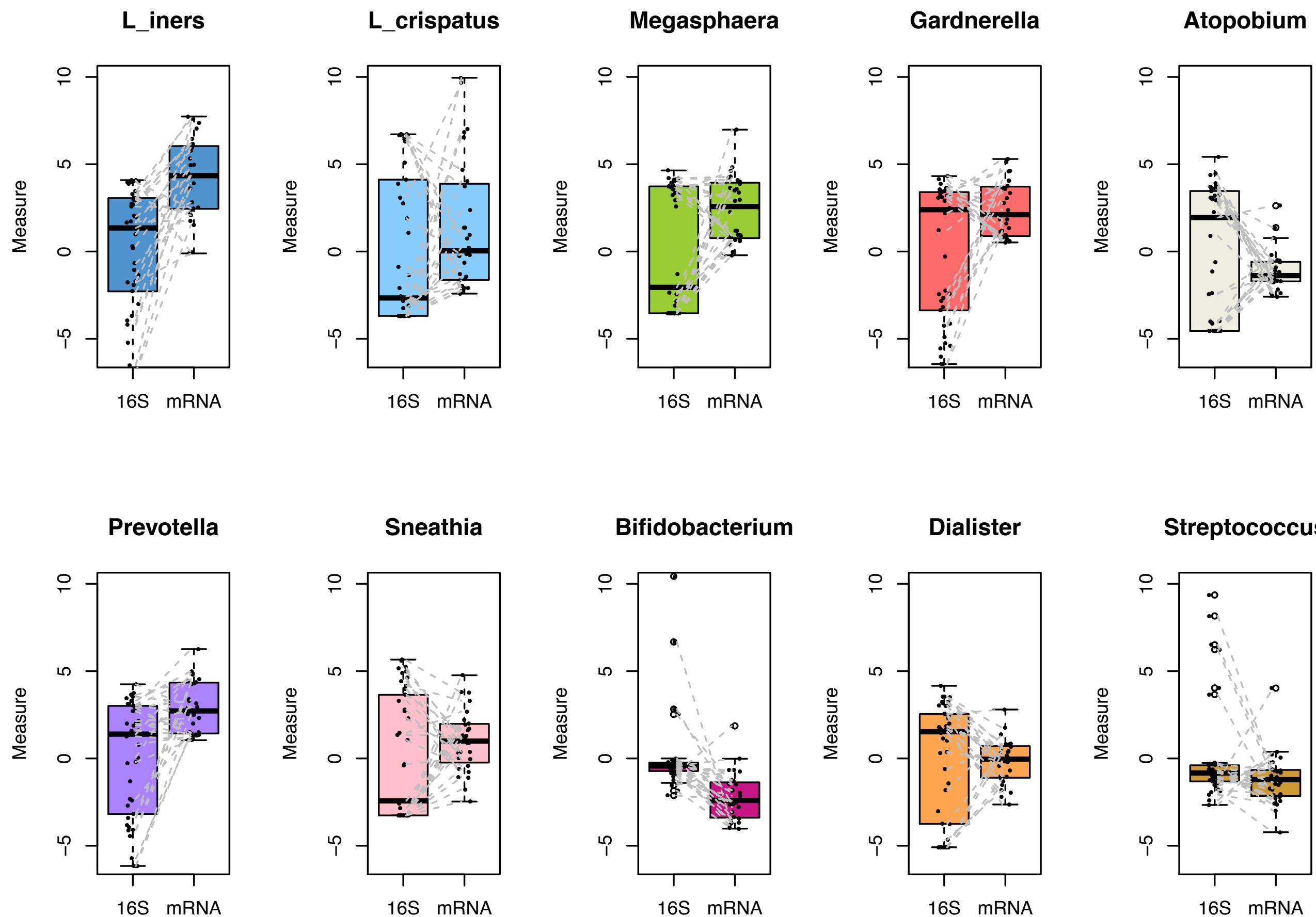
- Most common vaginal dysbiosis
 - H is predominantly *Lactobacillus* sp.
 - BV is mixed bag of anaerobes with *L. iners*
- Marked asymmetry in composition
 - Group genes to functional level (SEED, KEGG)
 - Sparse
- If everything is different, then nothing is important
 - We *must assume* something is invariant

Big Picture



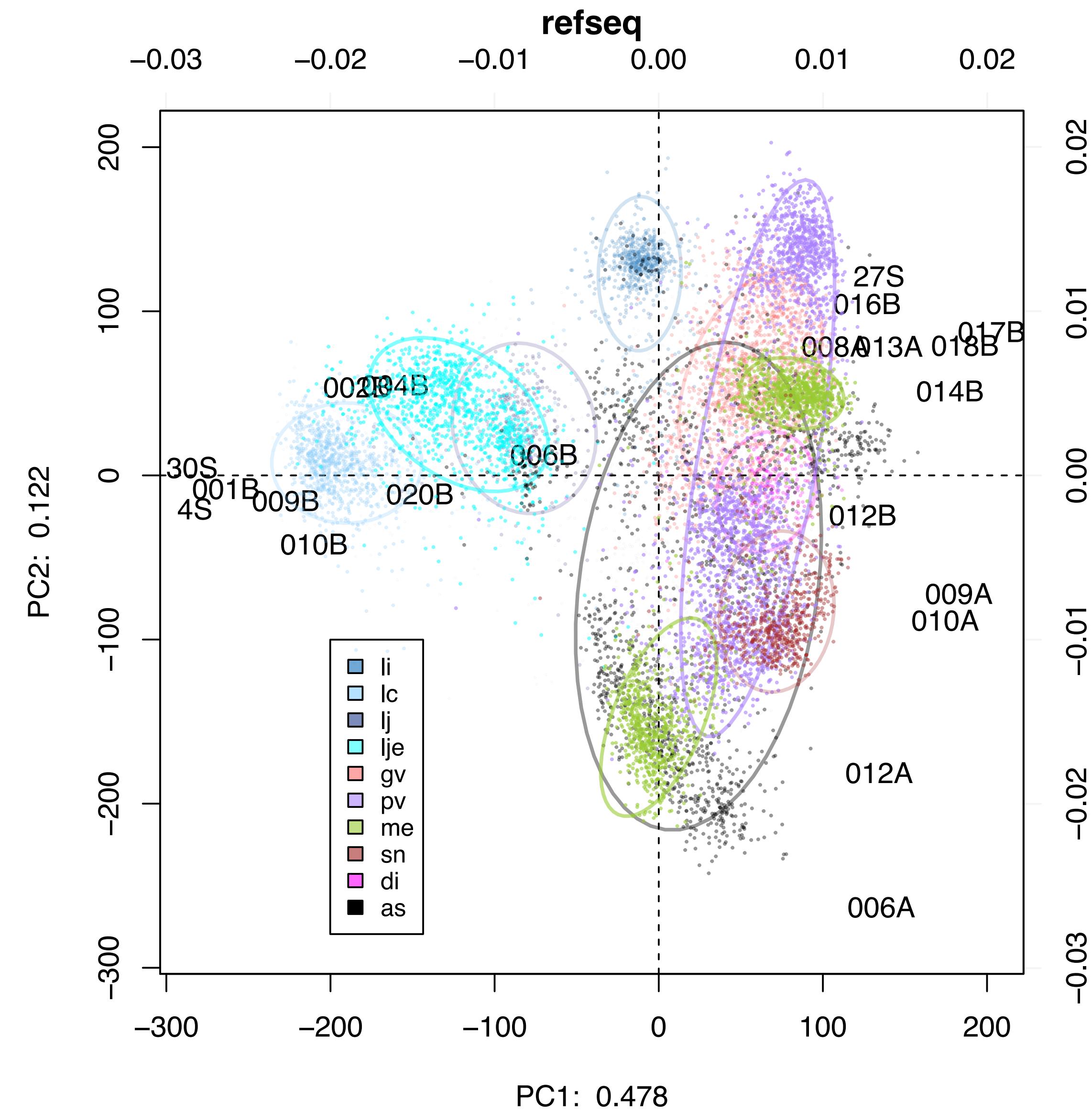
Macklaim, Macmillan unpublished
Macmillan Sic Rep 2015

16S vs. mRNA contribution

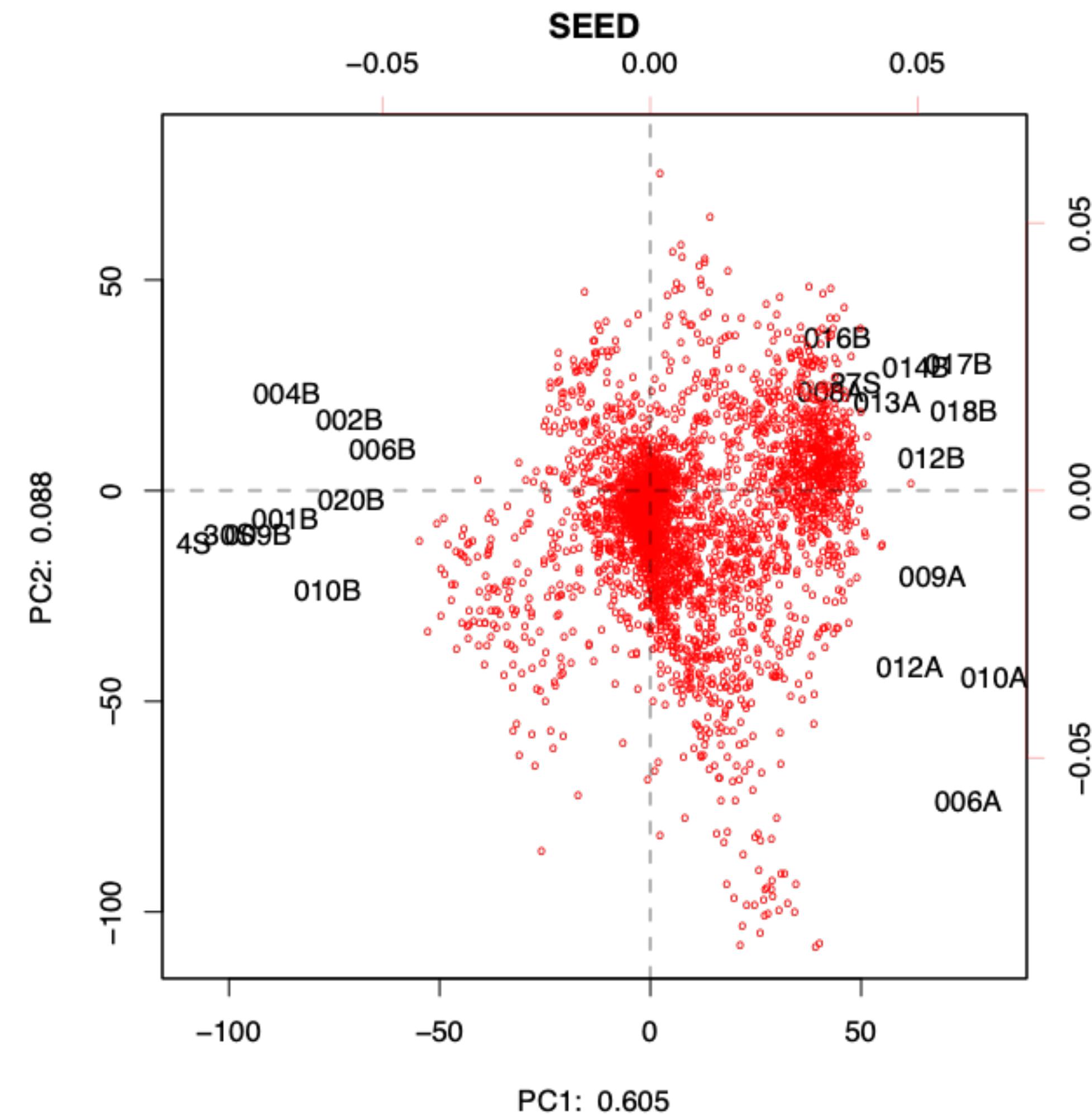


Macklaim, Macmillan unpublished
Macmillan Sic Rep 2015

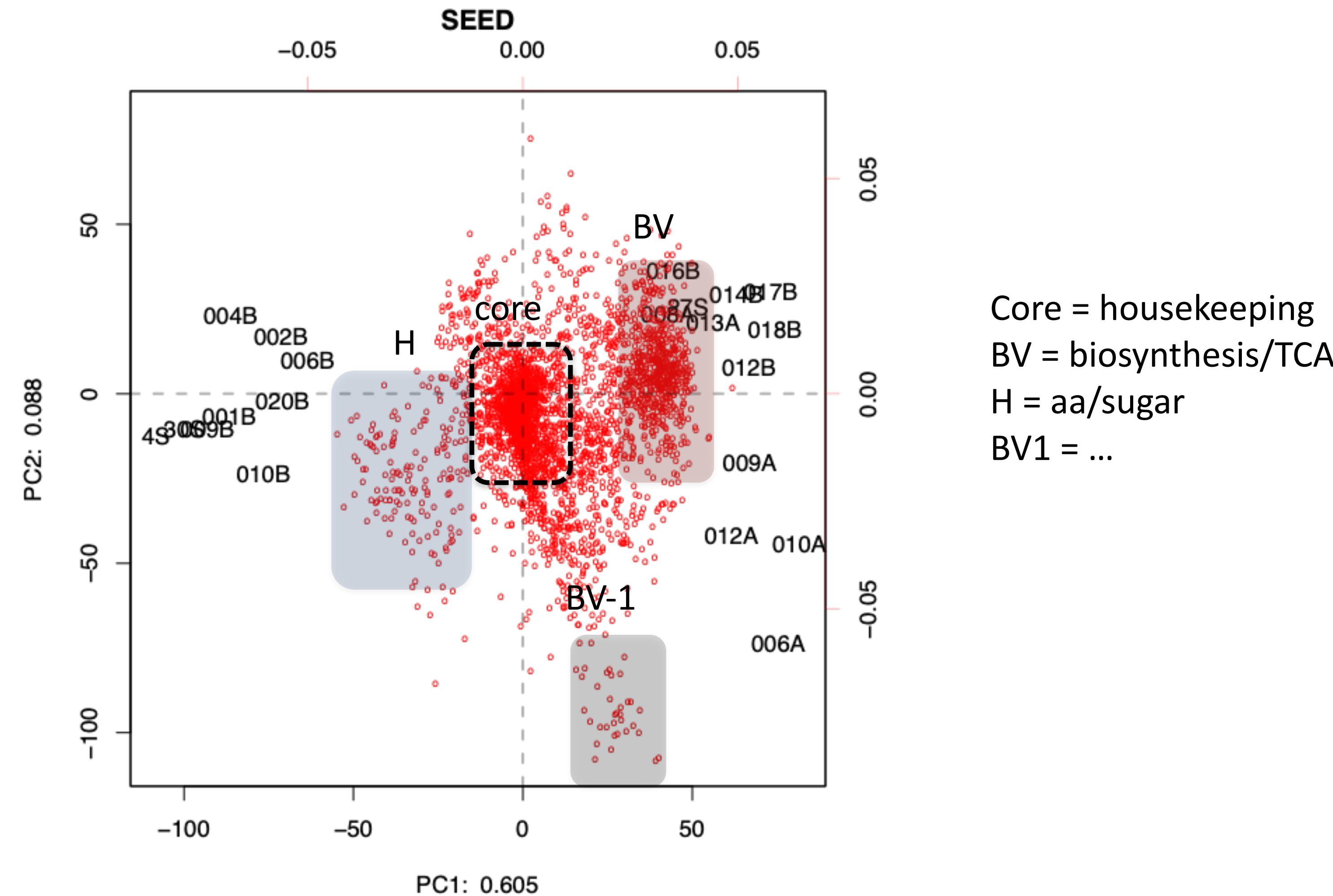
Reference sequence



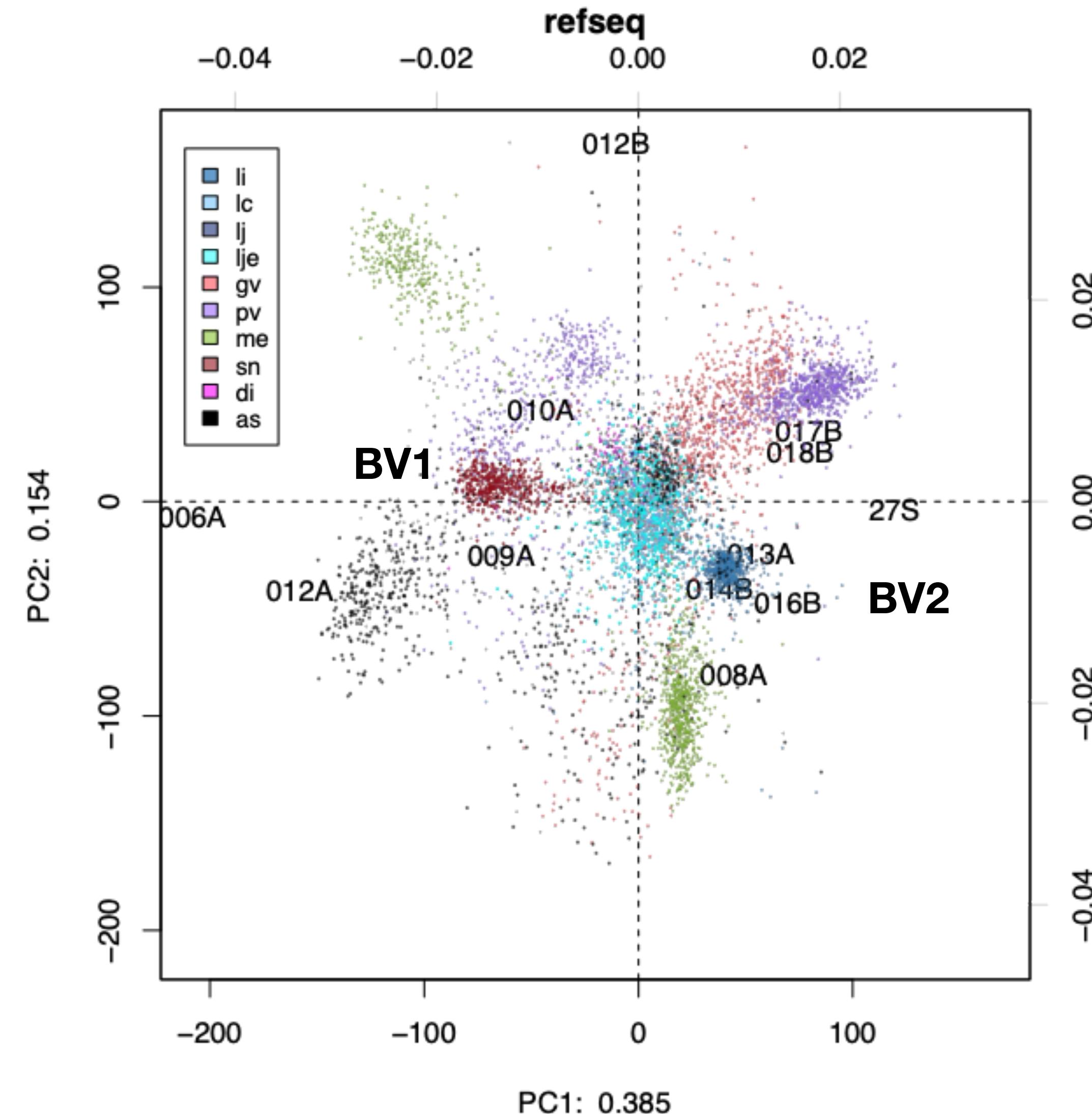
Aggregated by function



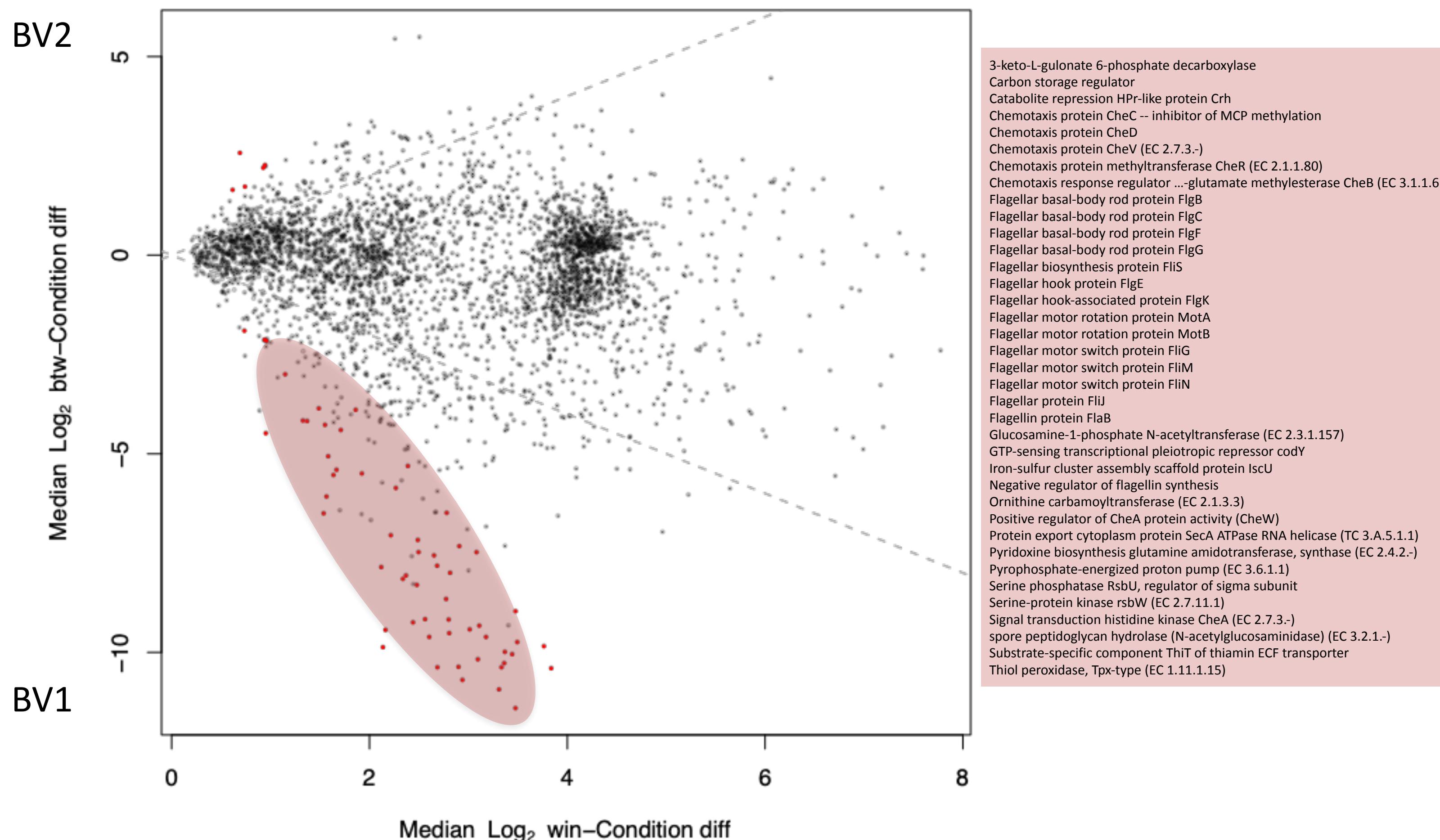
Four groups of functions



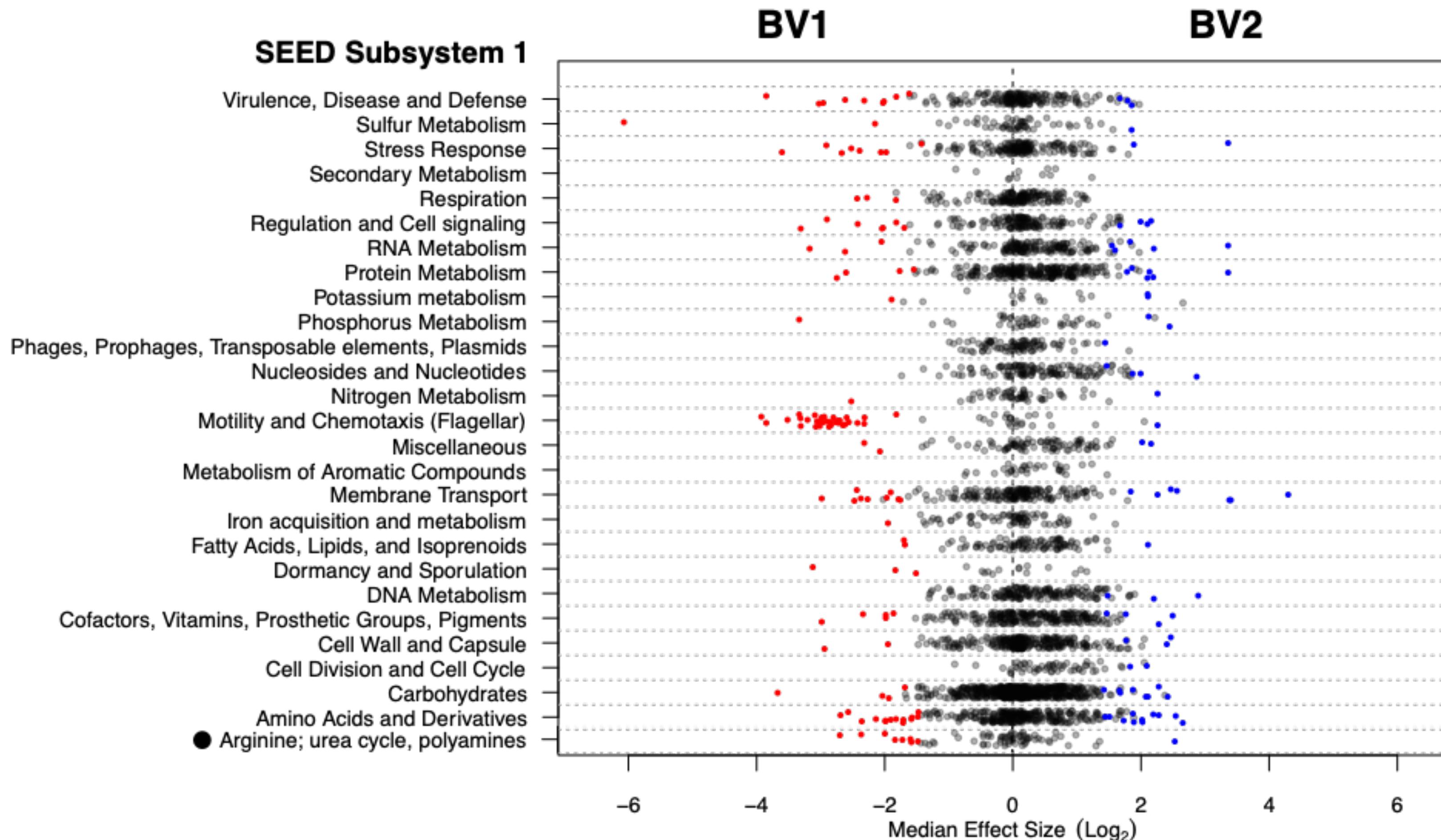
Reference sequence - BV



What is different between BV1 and BV2?



Distinguishing functions

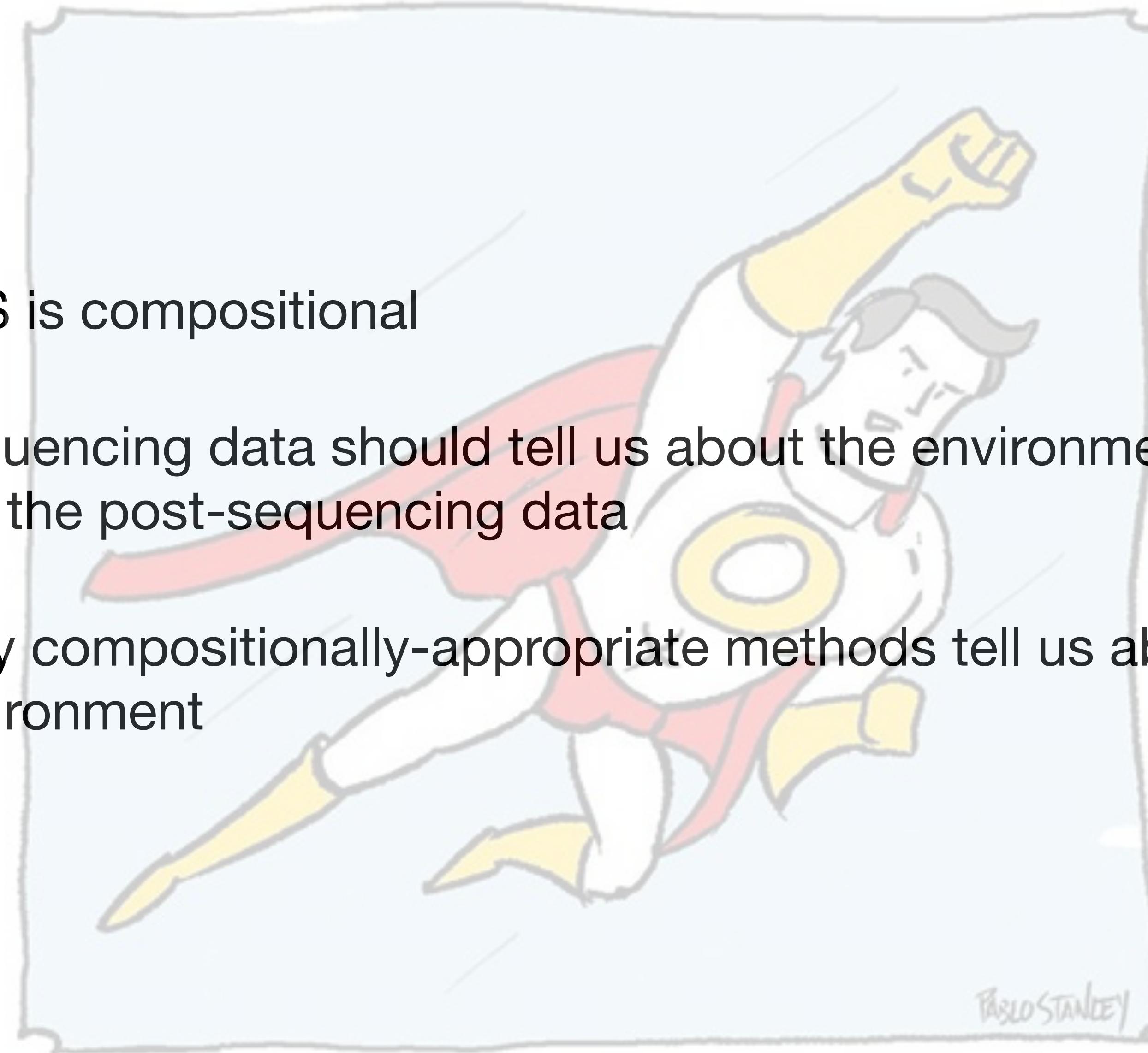
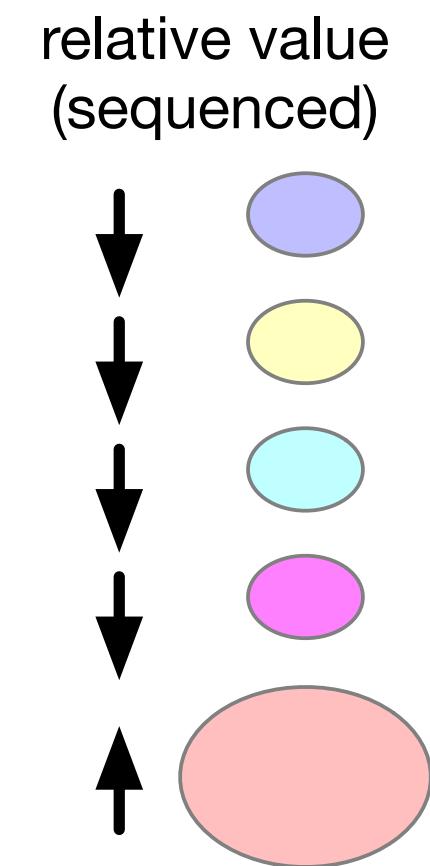


Summary

| Analysis | Standard (count) | CoDa/probability |
|------------------------|---|---|
| Beta Diversity | Driven by most abundant taxon or gene | Variance of ratios between taxa or genes |
| Clustering | Driven by most abundant taxon or gene | Variance of ratios between taxa or genes |
| Differential abundance | Usually rarest taxon or gene is most variable within and between groups | Variance of ratios between taxa or genes - most variable between groups |
| Correlation | Just wrong - many false positives | Pairs of taxa or genes that have common . |
| Interpretation | Seems simple but is not | Seems hard but is not |

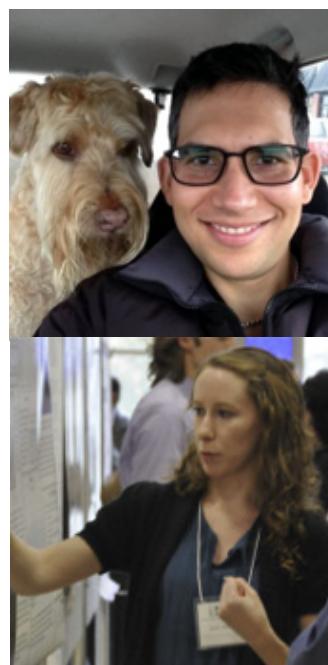
Summary

- HTS is compositional
- Sequencing data should tell us about the environment, not just the post-sequencing data
- Only compositionally-appropriate methods tell us about the environment

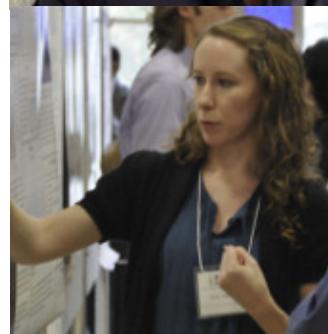


Acknowledgments

Canadian Centre for Human Microbiome
and Probiotic Research



Andrew
Fernandes



Jean
Macklaim

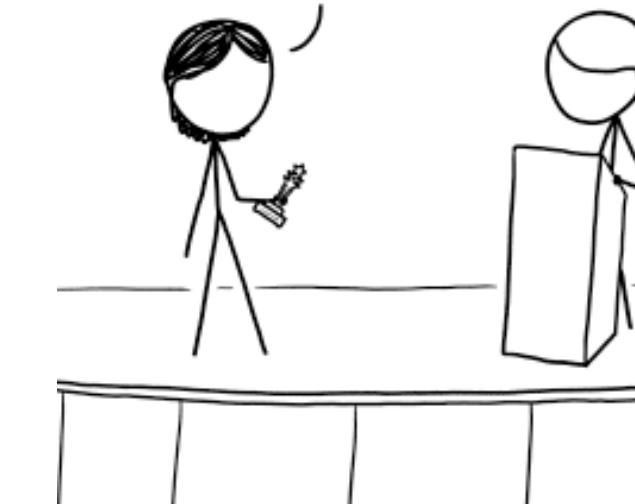
Gregor Reid
Jeremy Burton
Amy McMillan
Daniel Giguere
Jia Rong Wu



People. Discovery. Innovation.

I'D LIKE TO THANK MY DIRECTOR,
MY FRIENDS AND FAMILY, AND—
OF COURSE—THE WRETHER MASS
OF GUT BACTERIA INSIDE ME.

I MEAN, THERE'S LIKE ONE OR
TWO PINTS OF THEM IN HERE;
THEIR CELLS OUTNUMBER MINE!
ANYWAY, THIS WAS A
REAL TEAM EFFORT.



Coming soon:
Rob Finn et al.
Mgnify@EBI



Vera Pawlowsky-Glahn
Juan Jose Egozcue

Justin Silverman
philr
Tom Quinn, Ionas Erb
propr, balances



vaginal microbiome group initiative

Advancing Women's Health through
Microbiome Research



Canadian Institutes
of Health Research Instituts de recherche
en santé du Canada