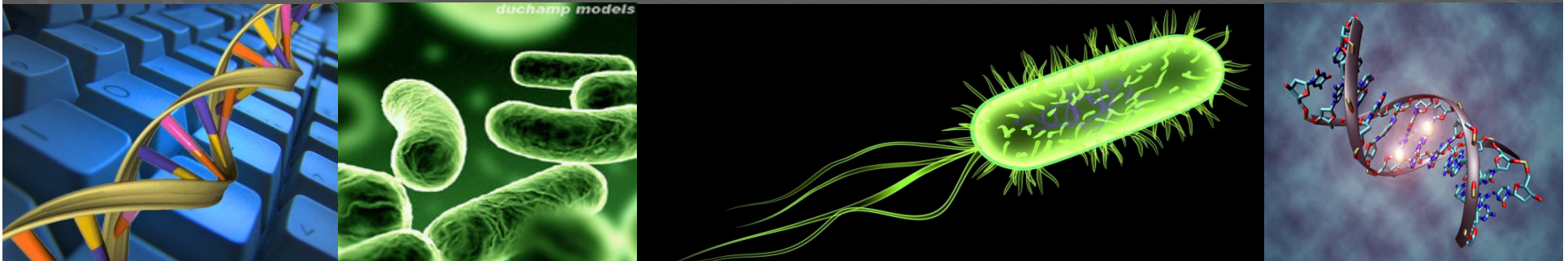


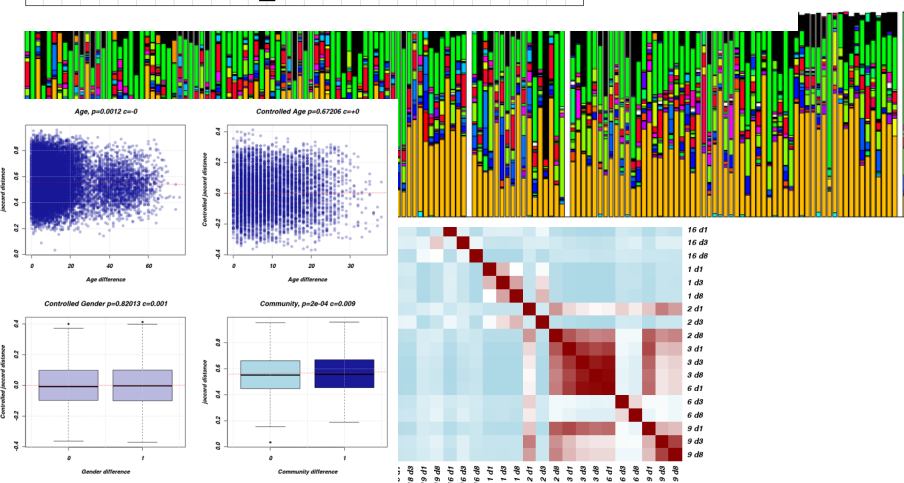
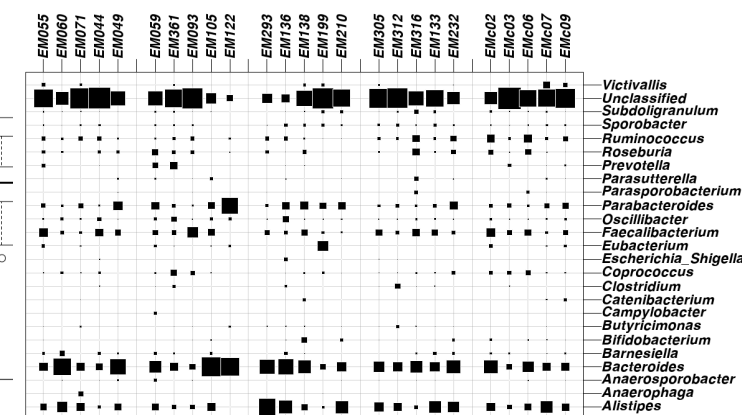
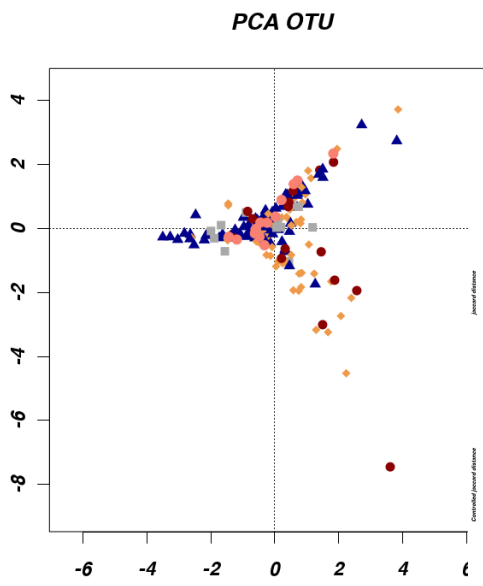
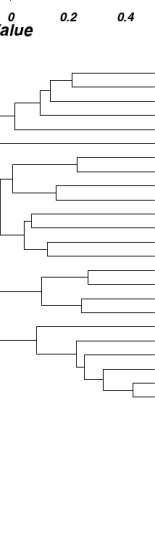
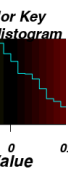


Queensland Institute of
Medical Research

Calypso – Mining, Visualizing and Comparing Multiple 16S rDNA Samples

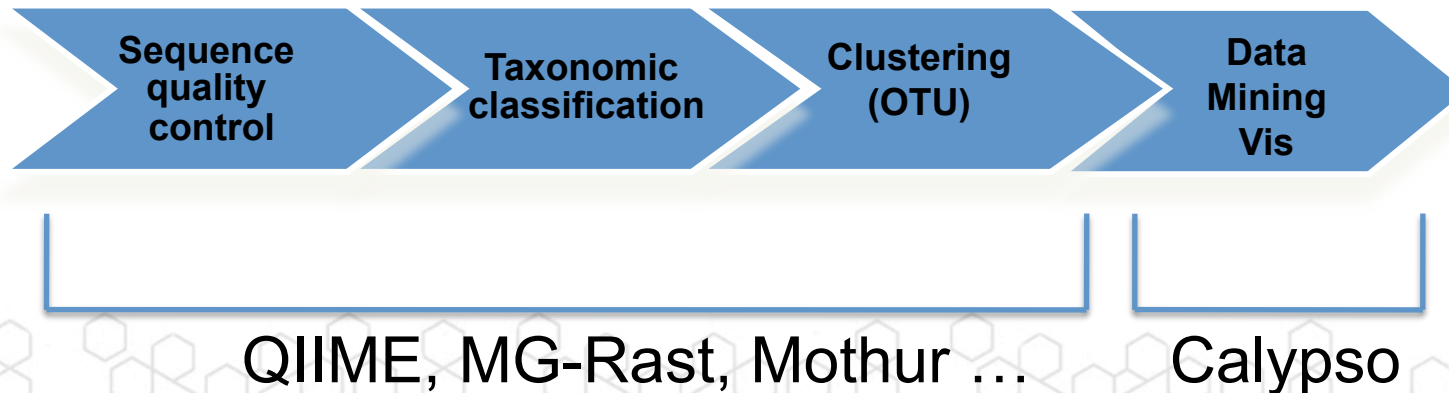
Lutz Krause





Calypso – User-Friendly Software for Analyzing Multiple and Large 16S rDNA Samples

- Easy to use web frontend, no installation required
- High-quality visualizations, figures can directly be used in publications
- Focus on data-mining, statistics and data visualization
- bioinfo.qimr.edu.au/calypso

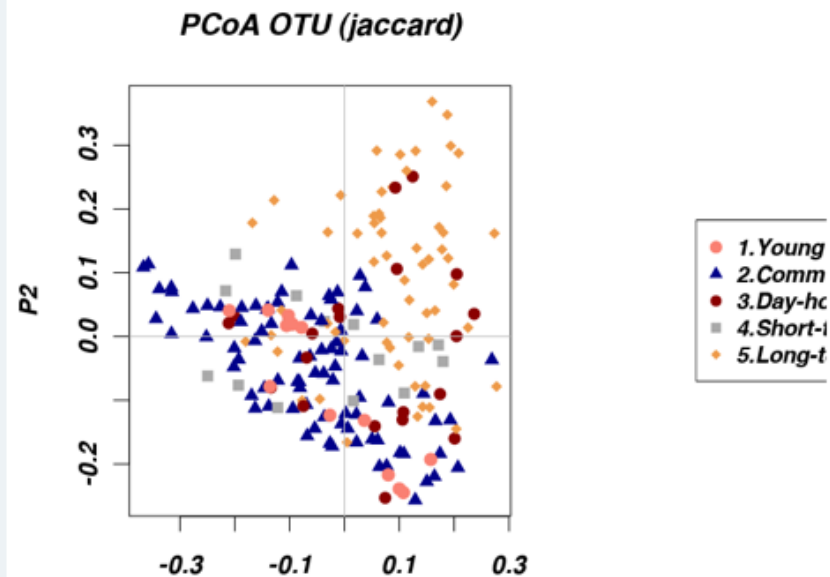


Calypso

| Home | Summary | Rarefaction | SamplePlots | GroupPlots | Stats | Multivariate | Diversity | Paired | Correlation | Regression |

Level Color Type: Color by:
Distance Method: TimePoint/Location: Components: Min proportion (0-100)
Relative Counts ☒ Scale: ☐ Legend: ☒ Loadings: ☐ Sample IDs: ☐ Resolution Width Height (mm)
Network settings: Edge Min Similarity Vertex Size

Multivariate analysis of microbial samples.



Example Data Set

Gut microbiota composition correlates with diet and health in the elderly

Marcus J. Claesson^{1,2*}, Ian B. Jeffery^{1,2*}, Susana Conde³, Susan E. Power¹, Eibhlís M. O'Connor^{1,2}, Siobhán Cusack¹, Hugh M. B. Harris¹, Mairead Coakley⁴, Bhuvaneswari Lakshminarayanan⁴, Orla O'Sullivan⁴, Gerald F. Fitzgerald^{1,2}, Jennifer Deane¹, Michael O'Connor^{5,6}, Norma Harnedy^{5,6}, Kieran O'Connor^{6,7,8}, Denis O'Mahony^{5,6,8}, Douwe van Sinderen^{1,2}, Martina Wallace⁹, Lorraine Brennan⁹, Catherine Stanton^{2,4}, Julian R. Marchesi¹⁰, Anthony P. Fitzgerald^{3,11}, Fergus Shanahan^{2,12}, Colin Hill^{1,2}, R. Paul Ross^{2,4} & Paul W. O'Toole^{1,2}

- Aims: Investigate correlation between intestinal microbiota of elderly people with environment, diet, and health
- Fecal samples from: 178 elderly (64 – 102y), non antibiotics treated, Irish Caucasian
- Recruited from two hospitals in Cork city
- 13 young adults (mean age 36)

Residence setting:

- 1) Community-dwelling
- 2) Out-patient day hospital
- 3) Short-term rehabilitation hospital care (<6 weeks),
- 4) Long-term institutionalized hospital care

Gut microbiota composition correlates with diet and health in the elderly

Dietary Information	Diet group 1-4	Complex and simple carbohydrates, meat, fish, fat, fruits, vegetables,...
	MNA	Mini nutritional assessment
Health Measurements	GDT	Geriatric depression test
	MMSE	Mini-mental state exam
	CCI	Charlson Comorbidity Index, predictor of survival
	CRP	Chronic Cardiac Failure
	Barthel	Index of activities of daily living
	FMI	Functional independence
Inflammation markers	IL6	
	IL8	
	TNF-alpha	

Input File: Read Counts

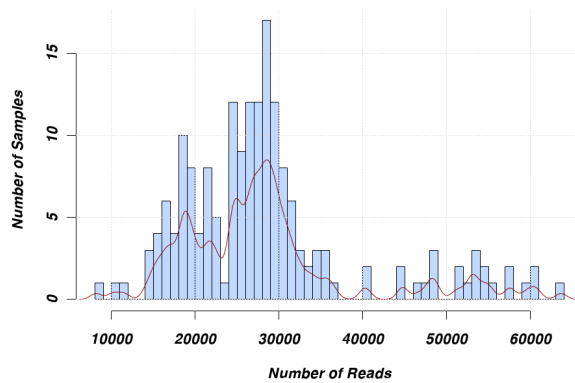
Header	Sample49	Sample54	Sample53	Sample60	Sample57	Sample56
Bifidobacterium	19	15	21	2	85	20
Escherichia/Shigella	0	0	1	0	0	2
Ochrobactrum	1	0	0	0	2	0
Streptococcus	2	2	2	3	105	52
Flavobacterium	0	0	0	0	1	0
Oribacterium	0	0	0	1	1	0
Vagococcus	0	0	0	0	3	0
Asaccharobacter	0	1	0	0	8	0
Acetivibrio	29	24	4	0	29	72
Dorea	0	0	0	0	0	1
Novosphingobium	0	0	0	0	0	0
Lachnobacterium	0	0	0	0	0	0
Pseudobutyrvibrio	0	0	0	0	0	0
Pseudolabrys	0	0	0	0	0	0
Hespellia	1	0	3	0	4	39
Brevundimonas	3	0	0	0	1	1
Veillonella	0	0	0	0	0	1
Salmonella	0	0	1	0	0	0
Anaeroplasma	0	0	0	2	0	36
Rothia	1	0	4	0	0	0

Input File: Sample Annotation

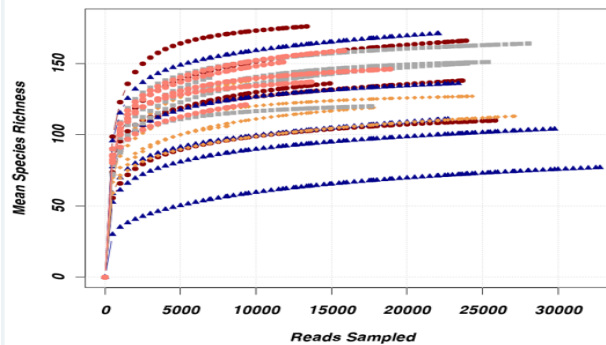
Sample	Group	Include	Age	Diet	BMI	IL-6	IL-8	TNFa
Sample1	Community	1	73	2	27	229.8	30	48.2
Sample2	Community	1	70	1	22.9	8.2	4.3	12.8
Sample3	Community	1	68	2	29	10.6	23.6	49.6
Sample4	Day-hospital	1	79	2	35.3	9.5	19.1	4.2
Sample5	Community	1	83	1	30.4	8.7	26.6	5.5
Sample6	Community	1	77	1	31.2	7.7	23.9	4.6
Sample7	Community	1	70	2	31.8	2.6	15.6	2.3
Sample8	Community	1	71	2	25.4	4.5	12.5	4.2
Sample9	Community	1	79	2	27.5	3	14.5	3.3
Sample10	Day-hospital	1	80	2	25.2	13.3	14	5.6
Sample11	Rehabilitation	1	84	2	32	23.2	26.3	8.6
Sample12	Community	1	78	1	27.4	4.9	11.1	4.3
Sample13	Long-term care	1	73	3	27.3	10.9	31.8	5.7
Sample14	Long-term care	1	78	3	28	26.3	26.1	12.8
Sample15	Rehabilitation	1	81	3	31.6	6.3	16.8	6.6
Sample16	Community	1	86	1	31.6	9.2	8.1	4.3
Sample17	Long-term care	1	66	3	34.4	4.2	13.4	3
Sample18	Long-term care	1	87	3	25.6	13.3	23.7	9.4
Sample19	Community	1	74	2	30	5.1	20.9	5.7
Sample20	Long-term care	1	87	4	26.6	13.1	27.1	7.8
Sample21	Long-term care	1	87	4	26.7	17.2	24.4	7
Sample22	Community	1	76	1	25.5	3.8	75.7	3.6
Sample23	Long-term care	1	82	3	13.7	7.9	34.9	6.6
Sample24	Long-term care	1	76	3	23.9	38.3	15.6	6.4
Sample25	Long-term care	1	88	3	31.6	8.7	26.7	7.8
Sample26	Long-term care	1	91	3	13.8	17	22.6	10.6

Basic Descriptive Statistics

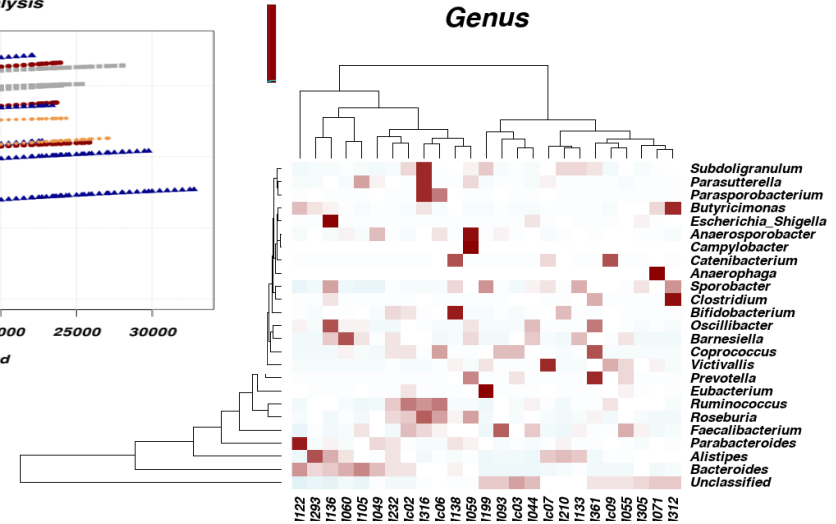
Reads per Sample, Min=8147 (EM232) Max=63675 (EM284)



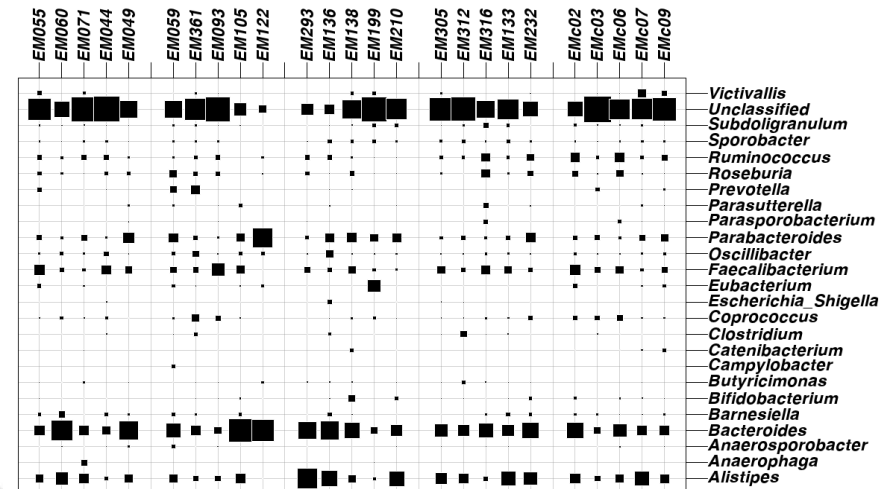
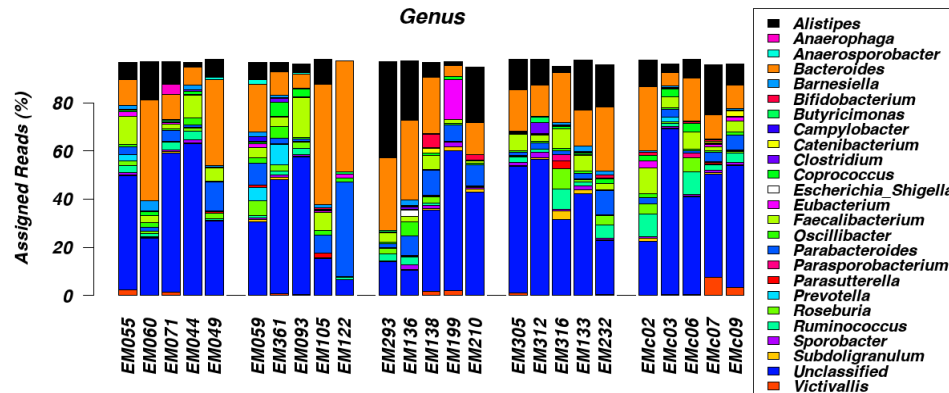
Rarefaction Analysis



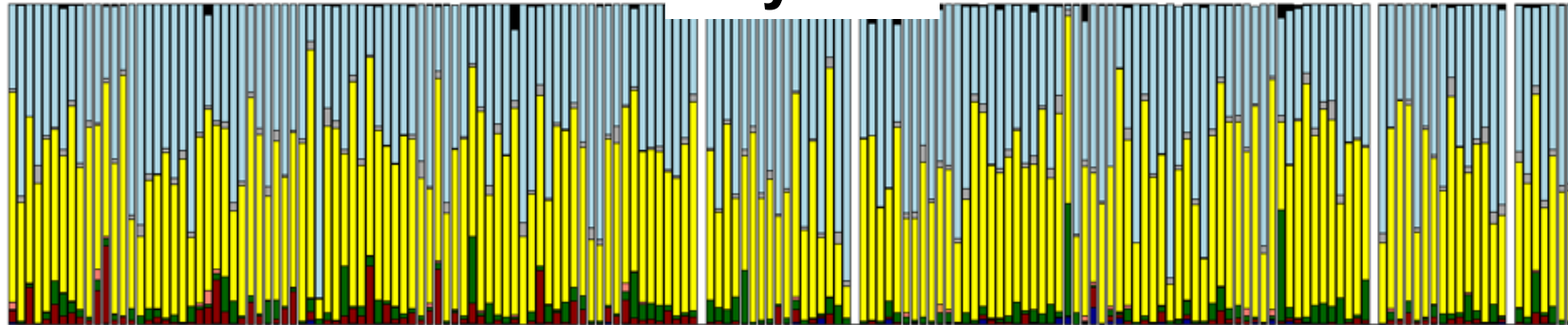
Genus



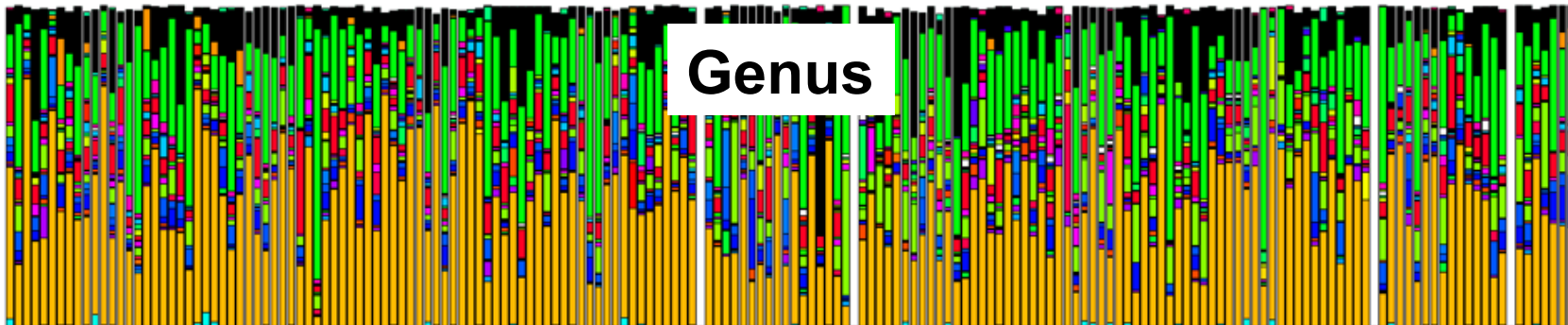
Genus



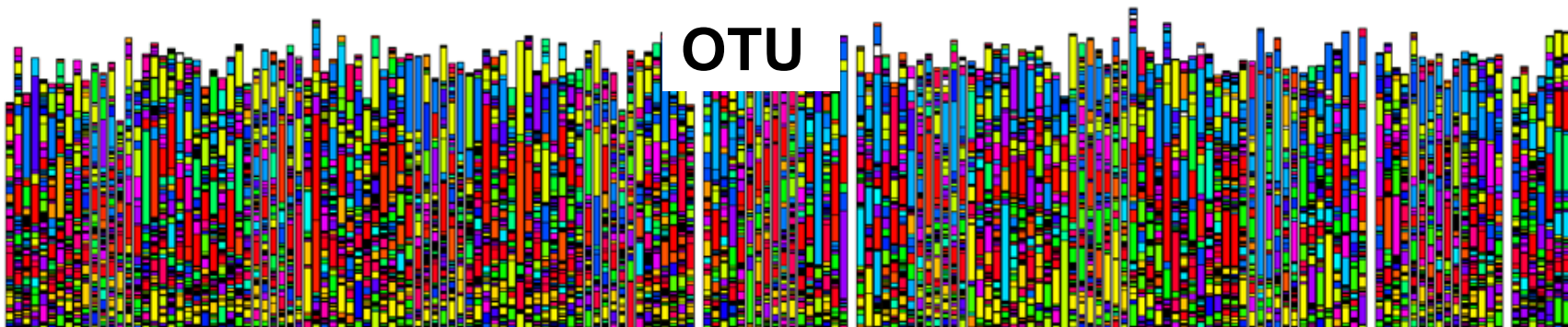
Phylum



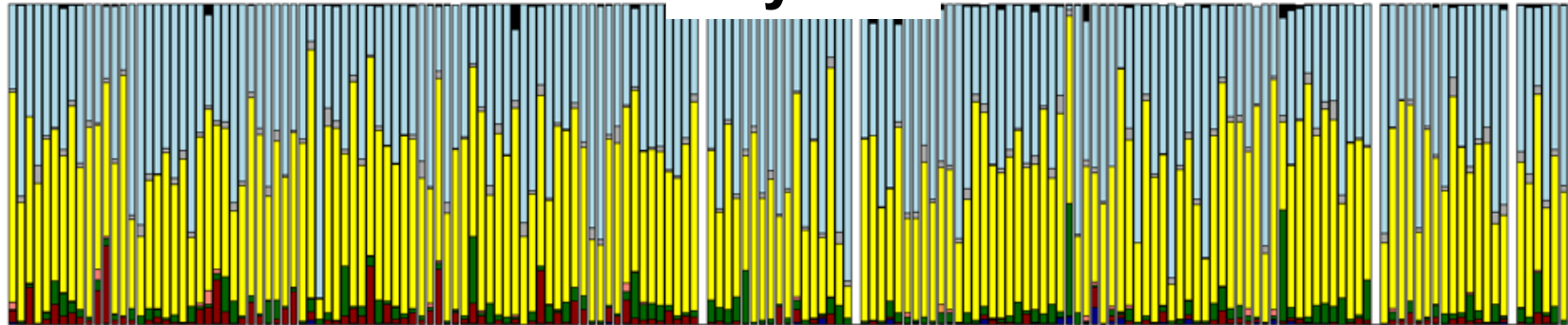
Genus



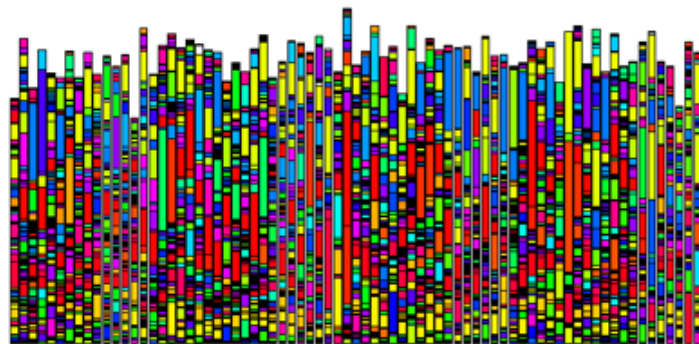
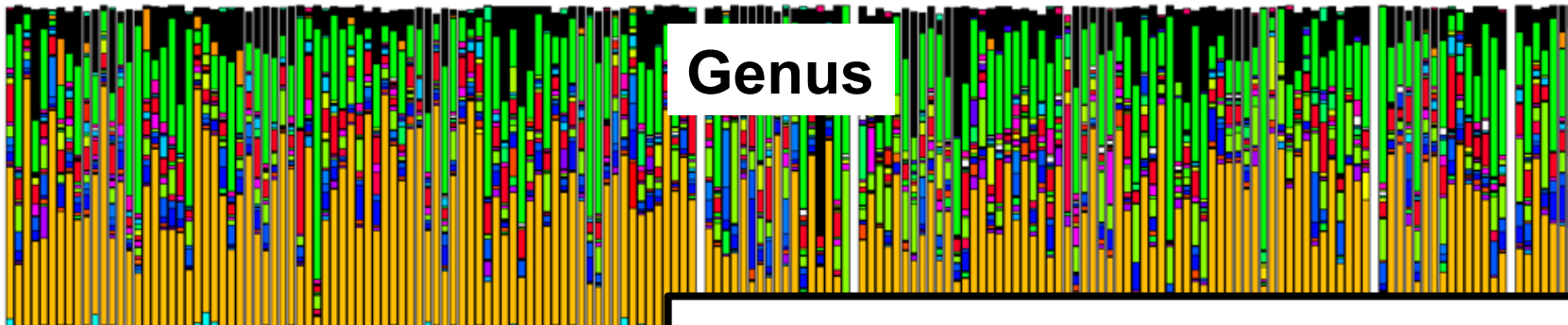
OTU



Phylum



Genus

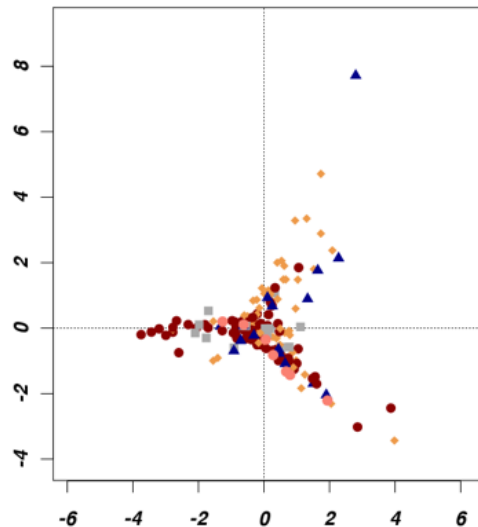


Human intestinal microbiota
similar on high taxonomic rank,
but highly diverse and individual
specific on low rank

Are Communities Different On Global, Whole-Community Level?

Type1 Unsupervised:

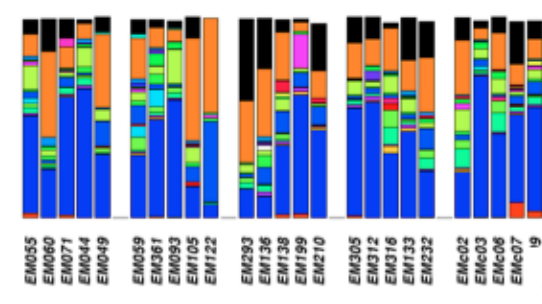
(PCA, DCA, NMDS, ...)



Type1 Supervised:

(LDA, RDA, ...)

Type 2 (Distance based):

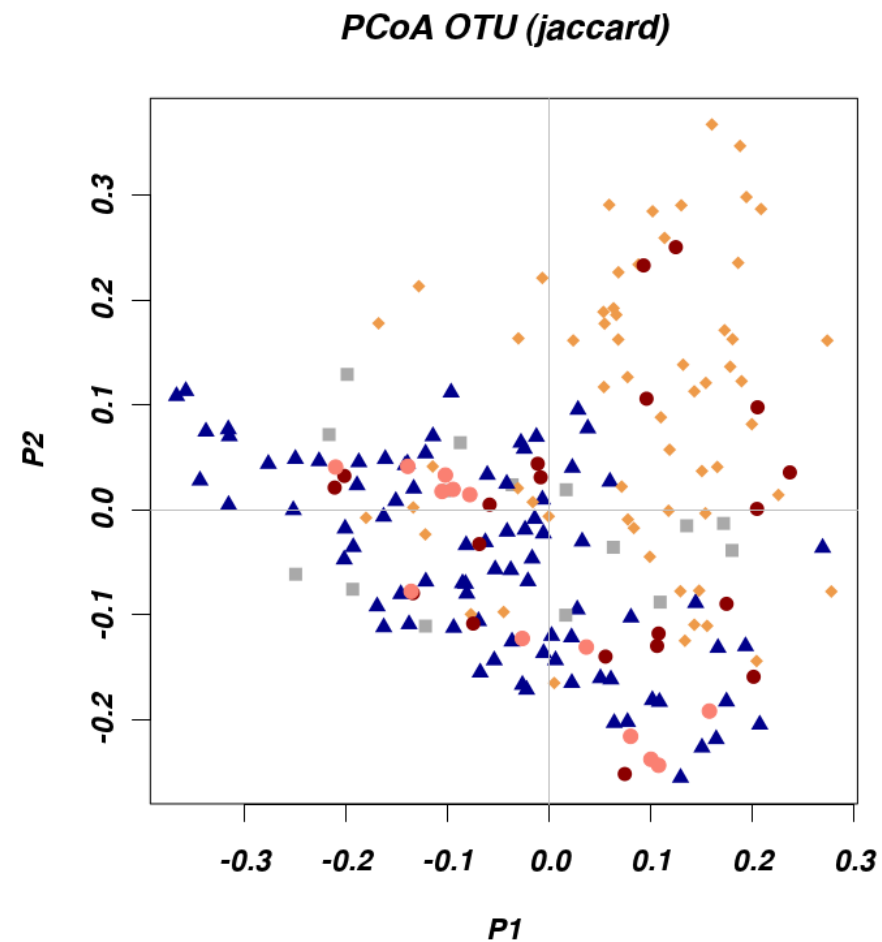
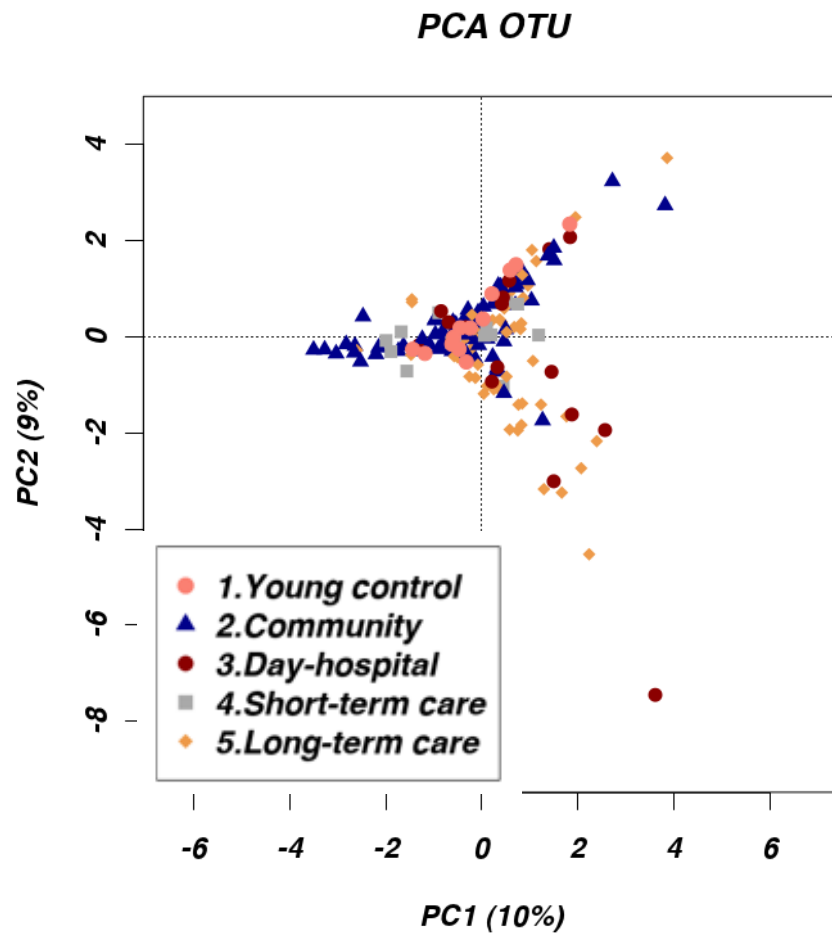


Distance Matrix

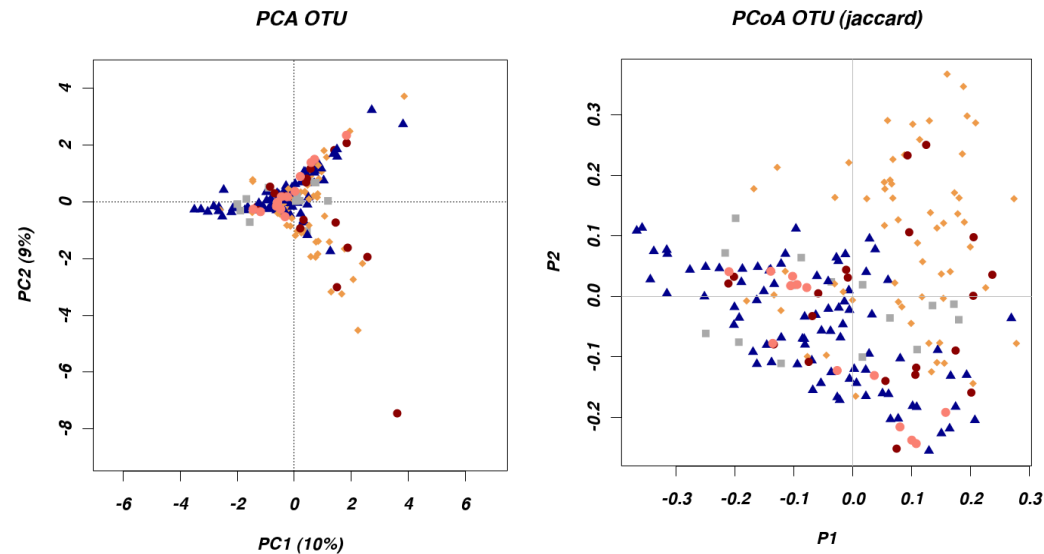


Hierarchical clustering, PCoA,
Heatmaps, ...

Does Residence Location Effect Community Structure on Global Level?



Does Residence Location Effect Community Structure on Global Level?



P-values for grouping:

	2.Community	3.Day-hospital	4.Short-term care	5.Long-term care
1.Young control	0.96741	0.19929	0.0398	0
2.Community		0	0.12997	0
3.Day-hospital			0.48583	0
4.Short-term care				0.02199

Detecting Differently Abundant Taxa in Calypso

Parametric tests

- Ttest, Anova

Non parametric tests

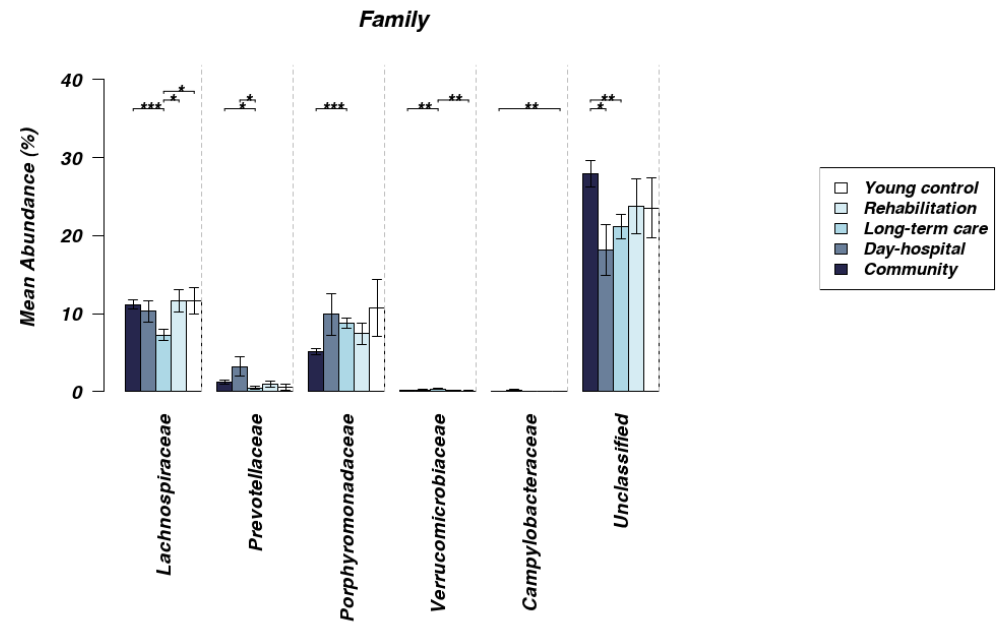
- Wilcoxon rank test
- Kruskal Wallis

Correction for multiple testing

- Bonferroni, FDR

Visualization

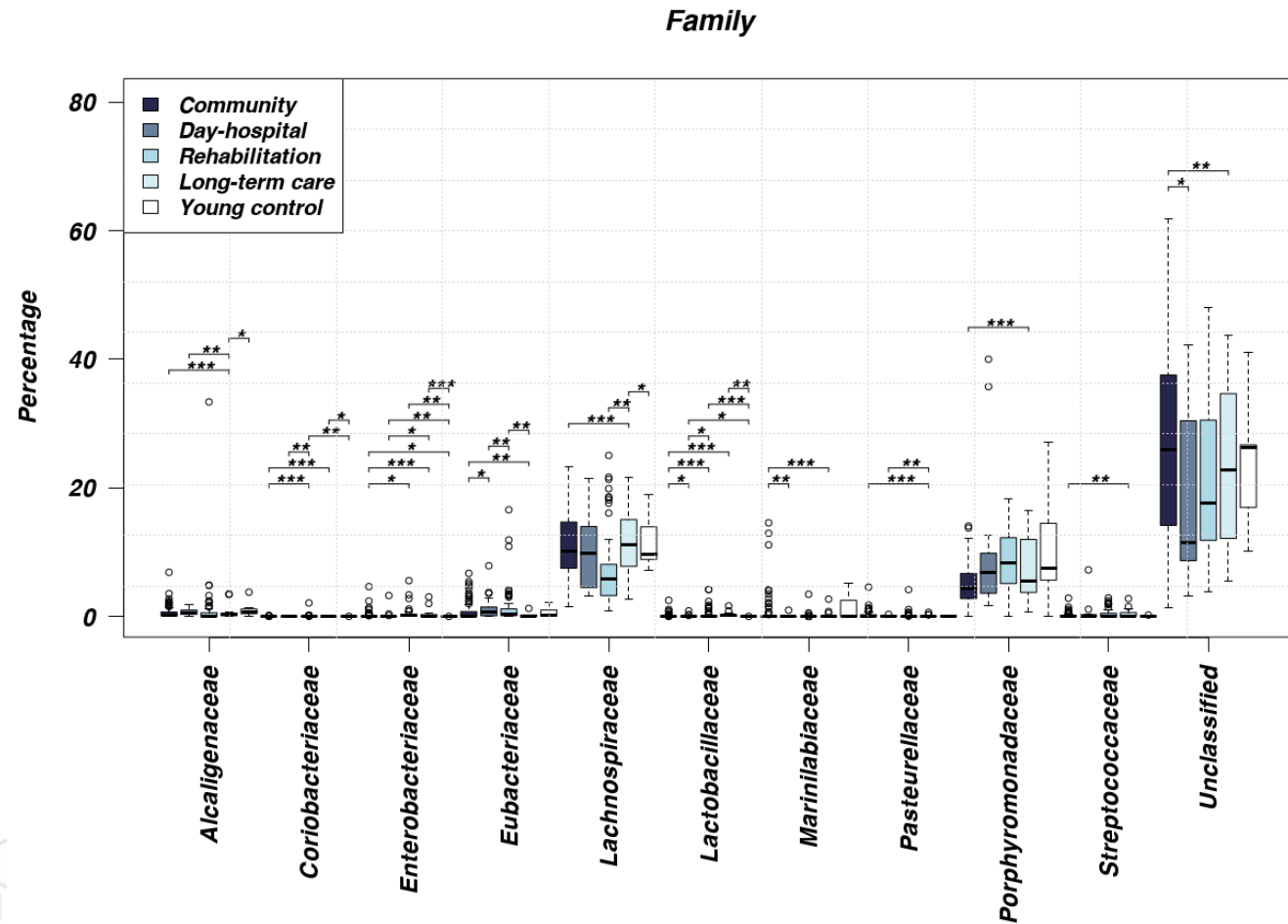
- Barcharts, boxplots,
- qq-plots, histograms



Differentially Abundant Bacterial Groups Elderly Study

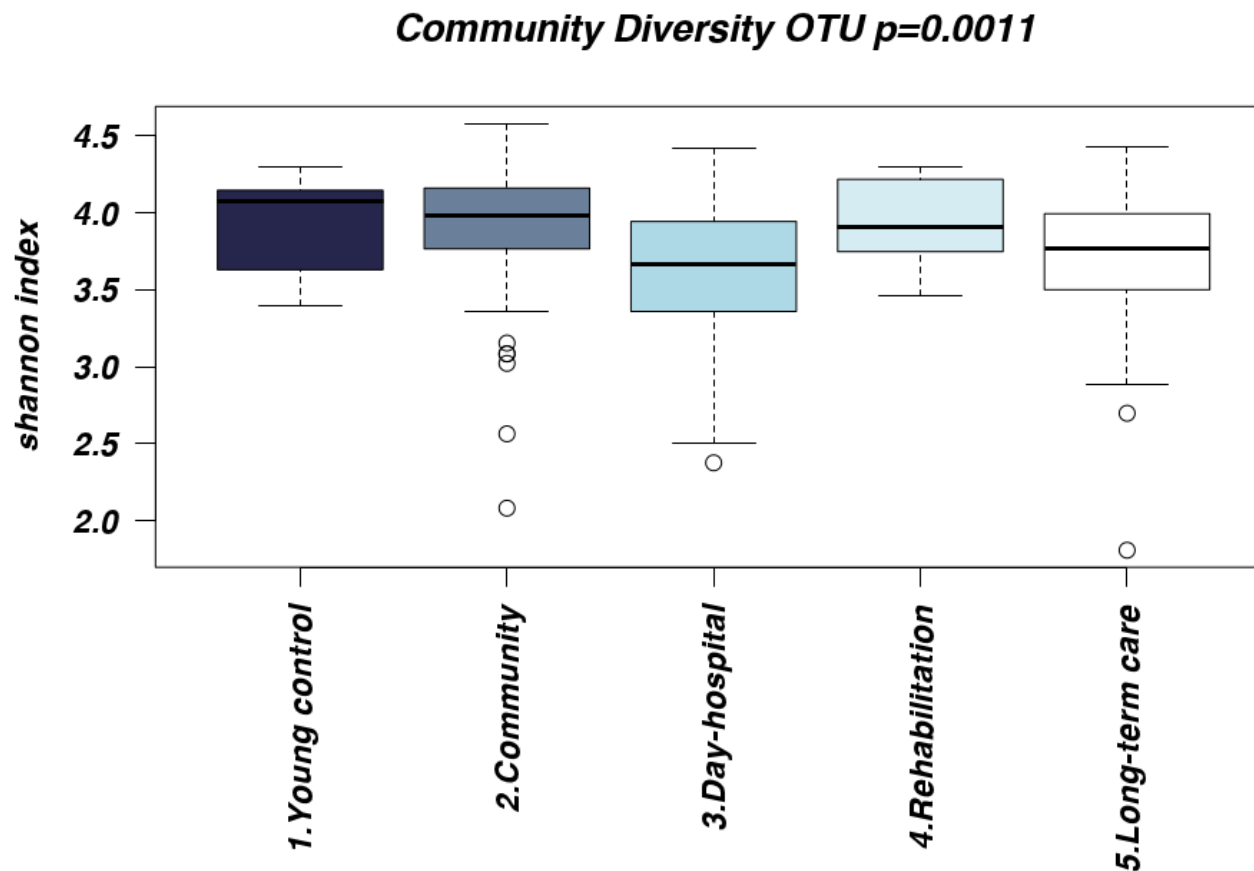
Differentially abundant groups with FDR < 0.05:

- 14 families
- 28 genera
- 220 OTUs

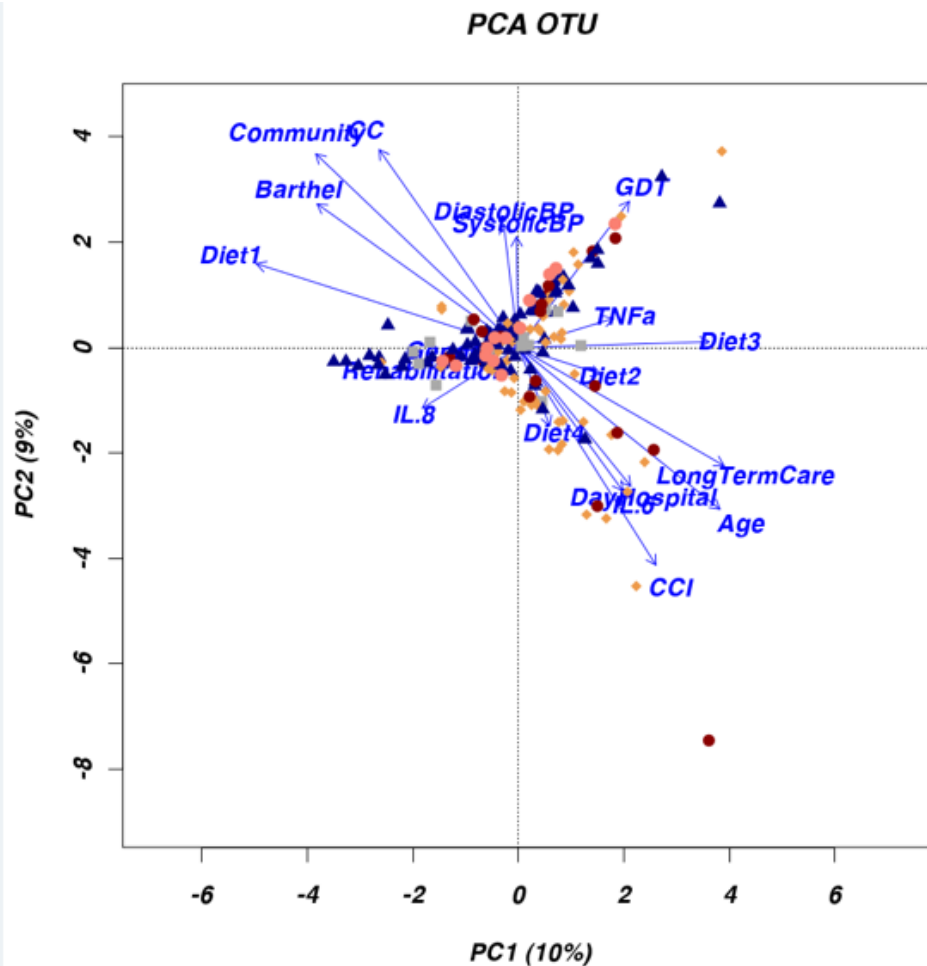


Overall Microbial Diversity (Shannon Index)

Available indexes: Shannon, richness, evenness, Simpson, Fisher's alpha, ...



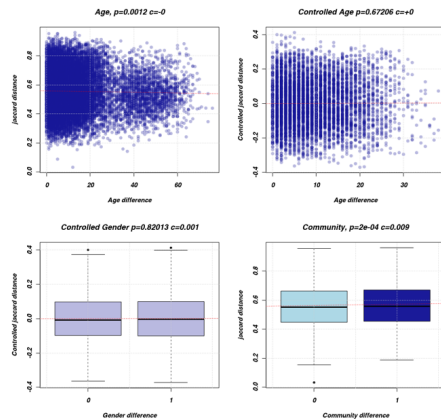
Associations Between Microbiota and Environmental Variables



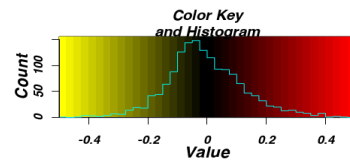
- Overlay on PCA
- Correlation
- Hierarchical clustering
- CCA, RDA
- Regression

Examples: age,
gender, BMI, diet; ph,
temperature, season,
...

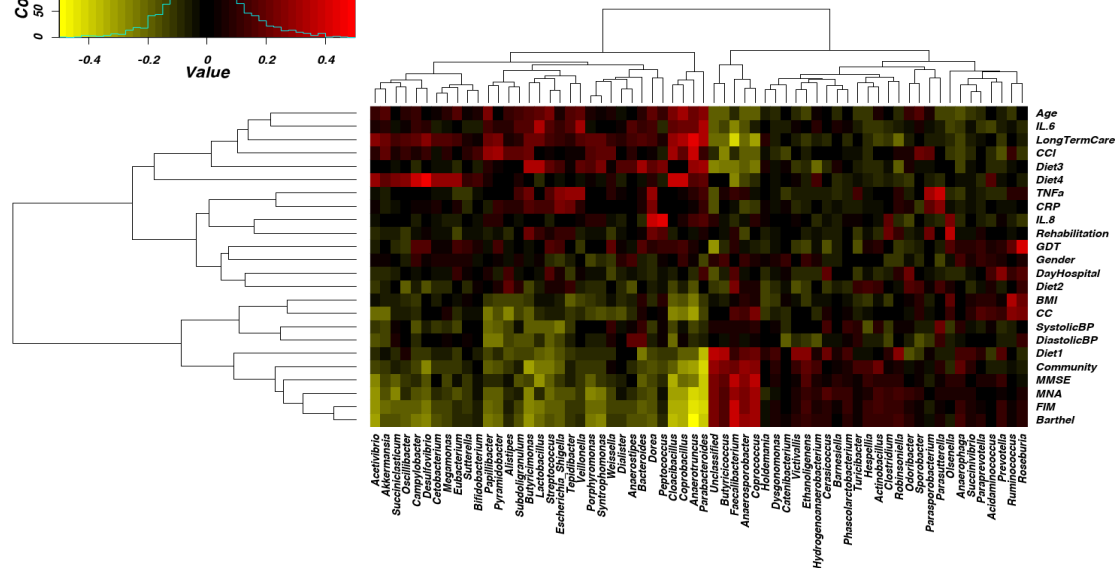
Associations Between Microbiota and Environmental Variables

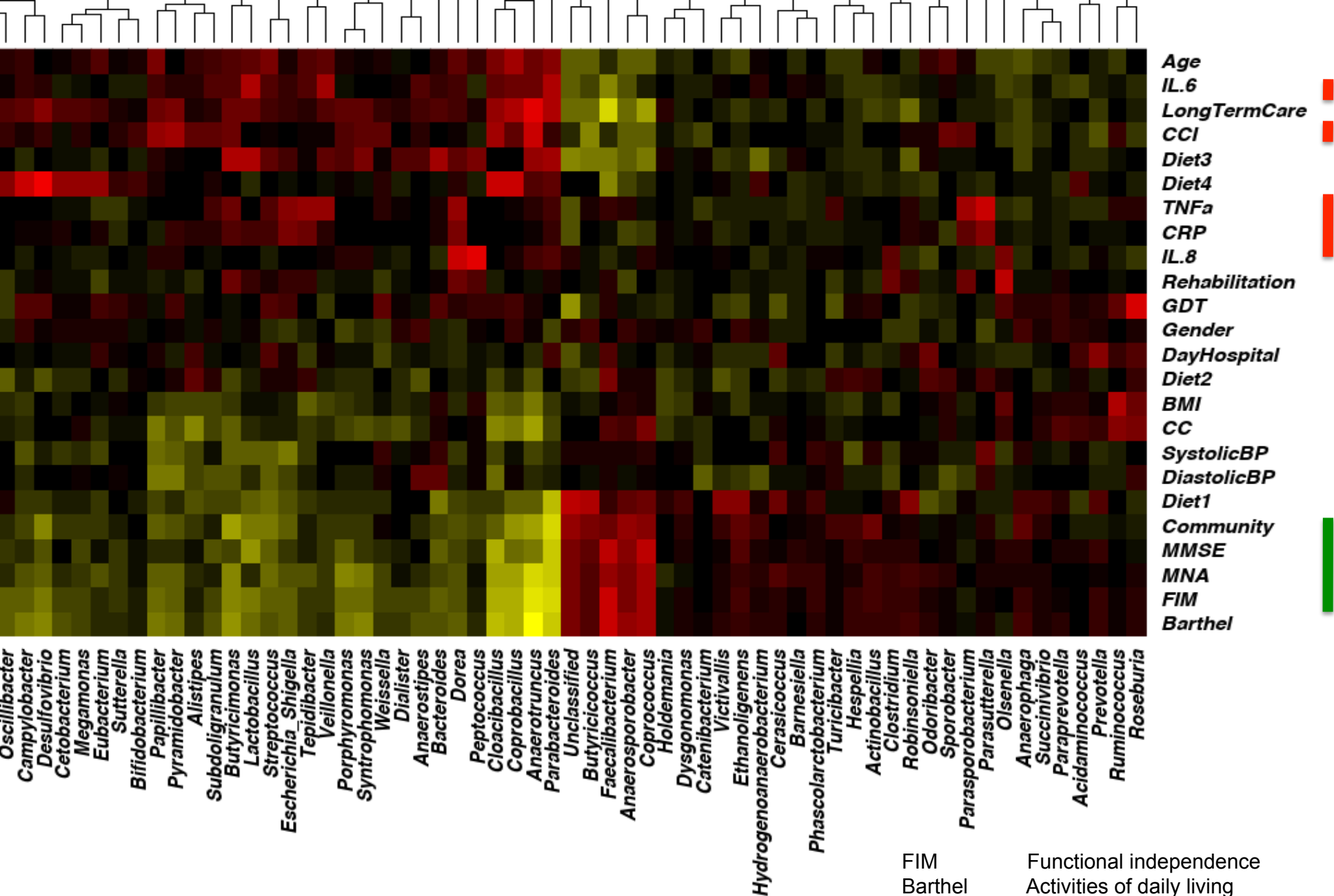


- Overlay on PCA
- **Correlation**
- Hierarchical clustering
- CCA, RDA
- Regression



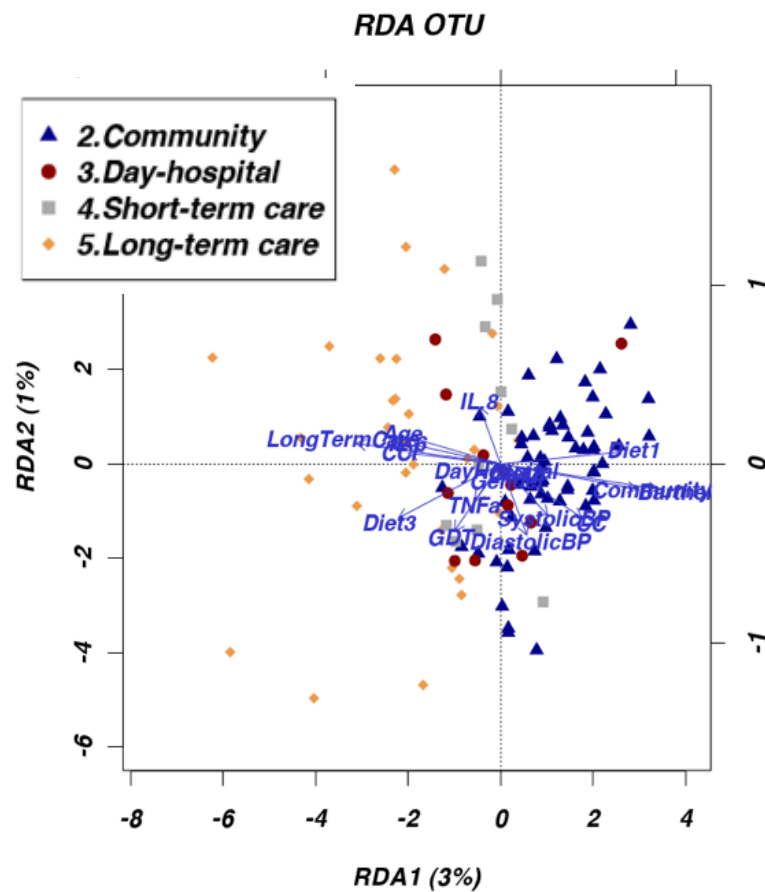
Pearson Correlation Genus





FIM Functional independence
 Barthel Activities of daily living
 MMSE Mini-mental state exam
 CCI Charlson Comorbidity
 CRP Chronic Cardiac Failure

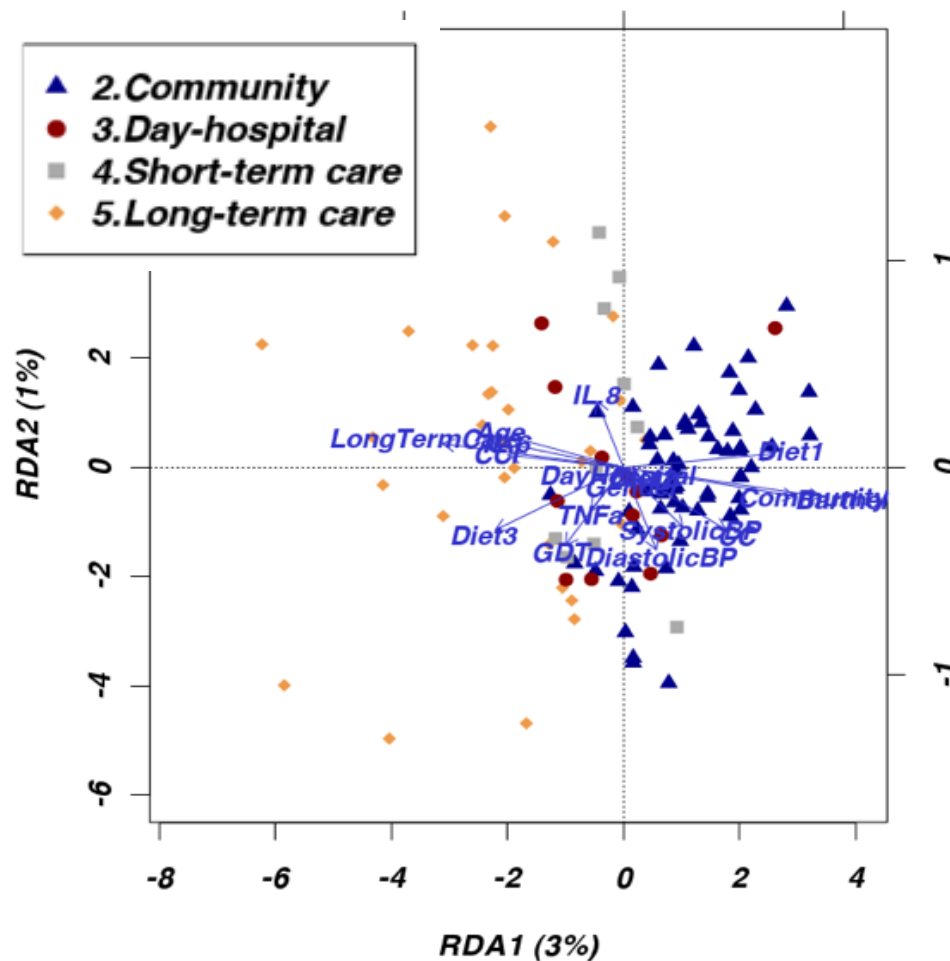
Associations Between Microbiota and Environmental Variables



- Overlay on PCA
- Correlation
- Hierarchical clustering
- **CCA, RDA**
- **Regression**

CCA and RDA

Aim: Identify clinical/environmental parameters explaining variance in abundance of bacterial groups



	<i>Df</i>	<i>N.Perm</i>	<i>P</i>	<i>Sign.</i>
Age	1	1999	0.000	***
Gender	1	1999	0.522	
Location	3	1999	0.001	***
Diet	3	1999	0.022	*
CCI	1	1999	0.042	*
FIM	1	1999	0.266	
Barthel	1	1999	0.242	
MMSE	1	1999	0.236	
BMI	1	1999	0.454	
CC	1	1999	0.378	
DiastolicBP	1	1999	0.058	
SystolicBP	1	1999	0.833	
GDT	1	1999	0.110	
MNA	1	1999	0.498	
CRP	1	1999	0.880	
IL.6	1	1999	0.060	
IL.8	1	1999	0.280	
TNFa	1	1999	0.711	

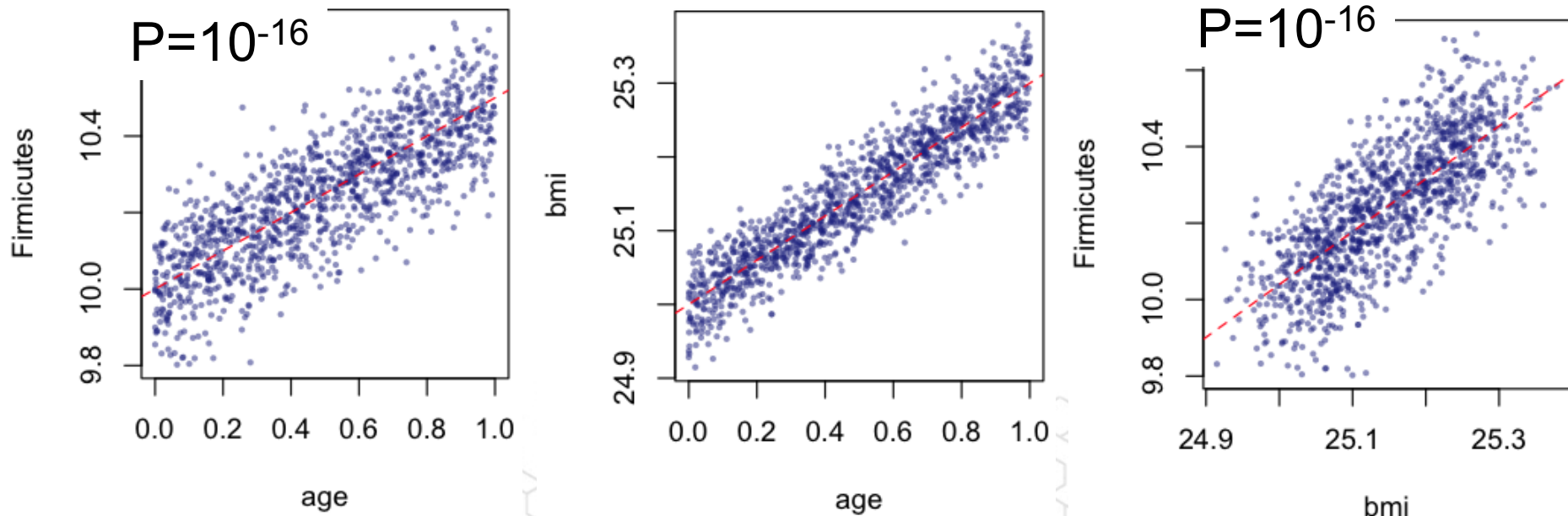
CCI: Charlson Comorbidity

Regression Analysis: Motivation

- **AIM:** Identify complex associations between microbiota and environmental parameters

Artificial example:

- Assume: age effects abundance of Firmicutes by: $\text{Firmicutes} = 10 + 0.5 * \text{age} + e$
- Assume BMI depends on age: $\text{BMI} = 25 + 0.3 * \text{age} + e$
- Simple statistics: Firmicutes sign. correlate with both BMI and age
- Fit regression model: $\text{Firmicutes} = c1 * \text{age} + c2 * \text{bmi} + c$
- Firmicutes sign. associated with age ($p=10^{-16}$) but not bmi ($p=0.98$)



Regression Analysis in Calypso

AIM: Identify complex associations between microbiota and clinical/environmental parameters

- 1) Identify parameters effecting global community composition:

Pairwise distance = age-diff + sex-diff + diet-diff + residential setting diff + ... + c

- 2) Identify single taxa/OTUs effected by environmental/clinical parameters:

OTU abundance = age + sex + diet + residential status + ... + c

- 3) Community diversity:

Diversity = age + sex + diet + residential status + ... + c

Regression Analysis: Global Composition

AIM: Identify complex associations between microbiota and clinical/environmental parameters

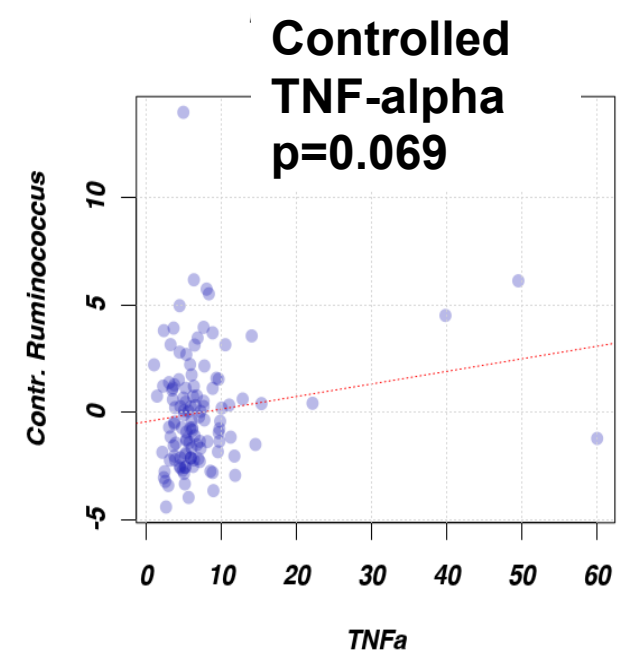
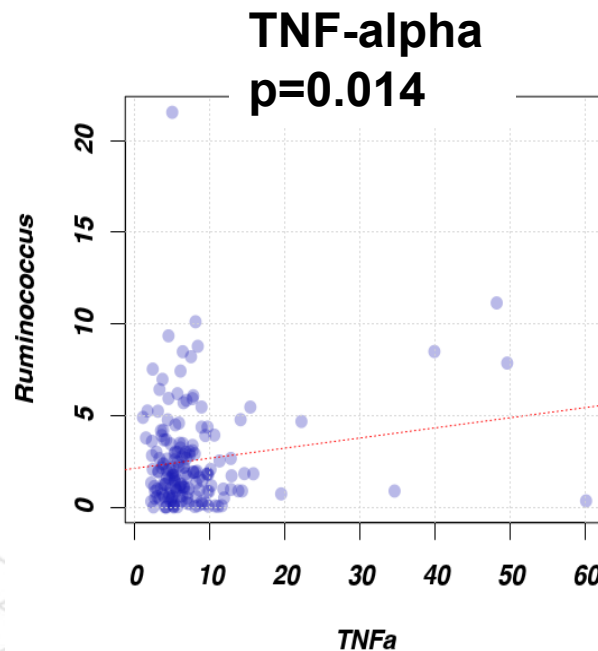
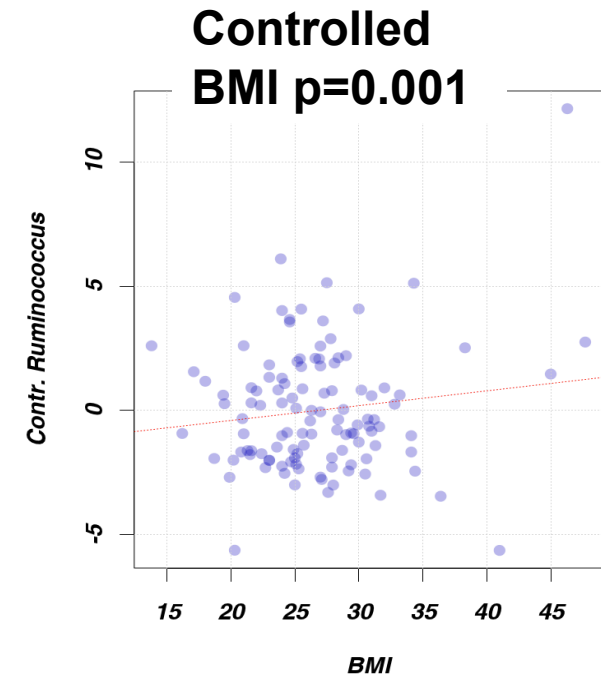
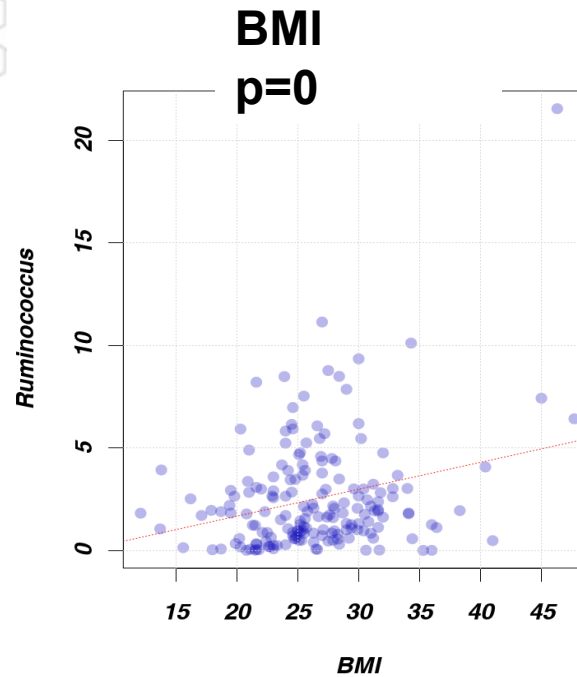
1) **Identify parameters effecting global community composition:**

Pairwise distance = age-diff + sex-diff + diet-diff + residential res. diff + ...

	<i>P</i>	<i>S</i>	
(Intercept)	0.000	***	
Age	0.000	***	
Gender	0.196		
Location	0.000	***	
Diet	0.000	***	
CCI	0.049	*	
FIM	0.000	***	General health
Barthel	0.000	***	
MMSE	0.168		
BMI	0.212		
CC	0.062		
DiastolicBP	0.000	***	
SystolicBP	0.026	*	
GDT	0.834		
MNA	0.000	***	Nutrition
CRP	0.785		
IL.6	0.000	***	Inflammation
IL.8	0.000	***	
TNFa	0.001	***	

FIM Functional independence
 Barthel Activities of daily living
 MMSE Mini-mental state exam
 CCI Charlson Comorbidity
 CRP Chronic Cardiac Failure

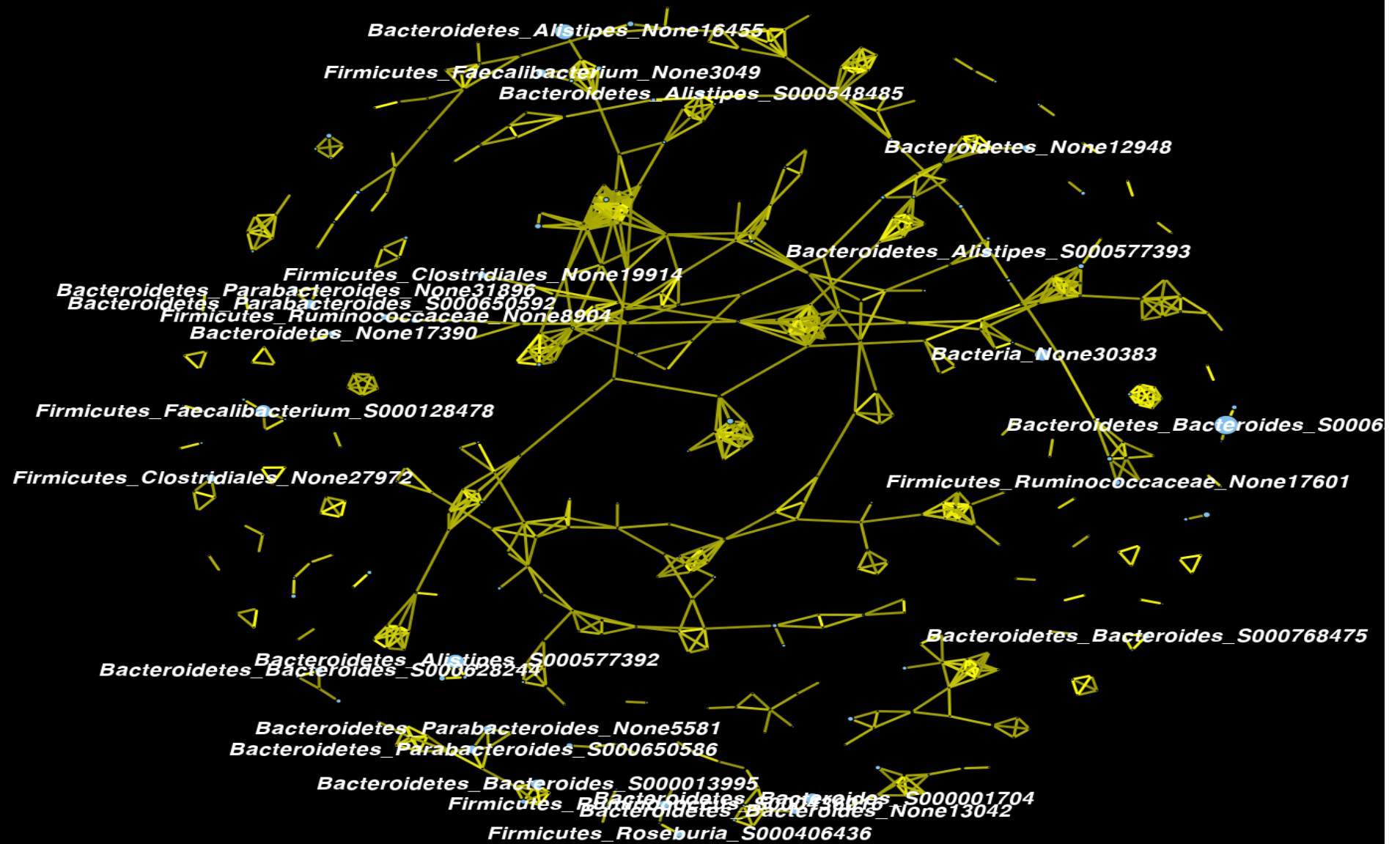
**Gut microbiota
& obesity:
Genus With
Strongest
Association
With BMI:
Ruminococcus
($p=0.001$)**

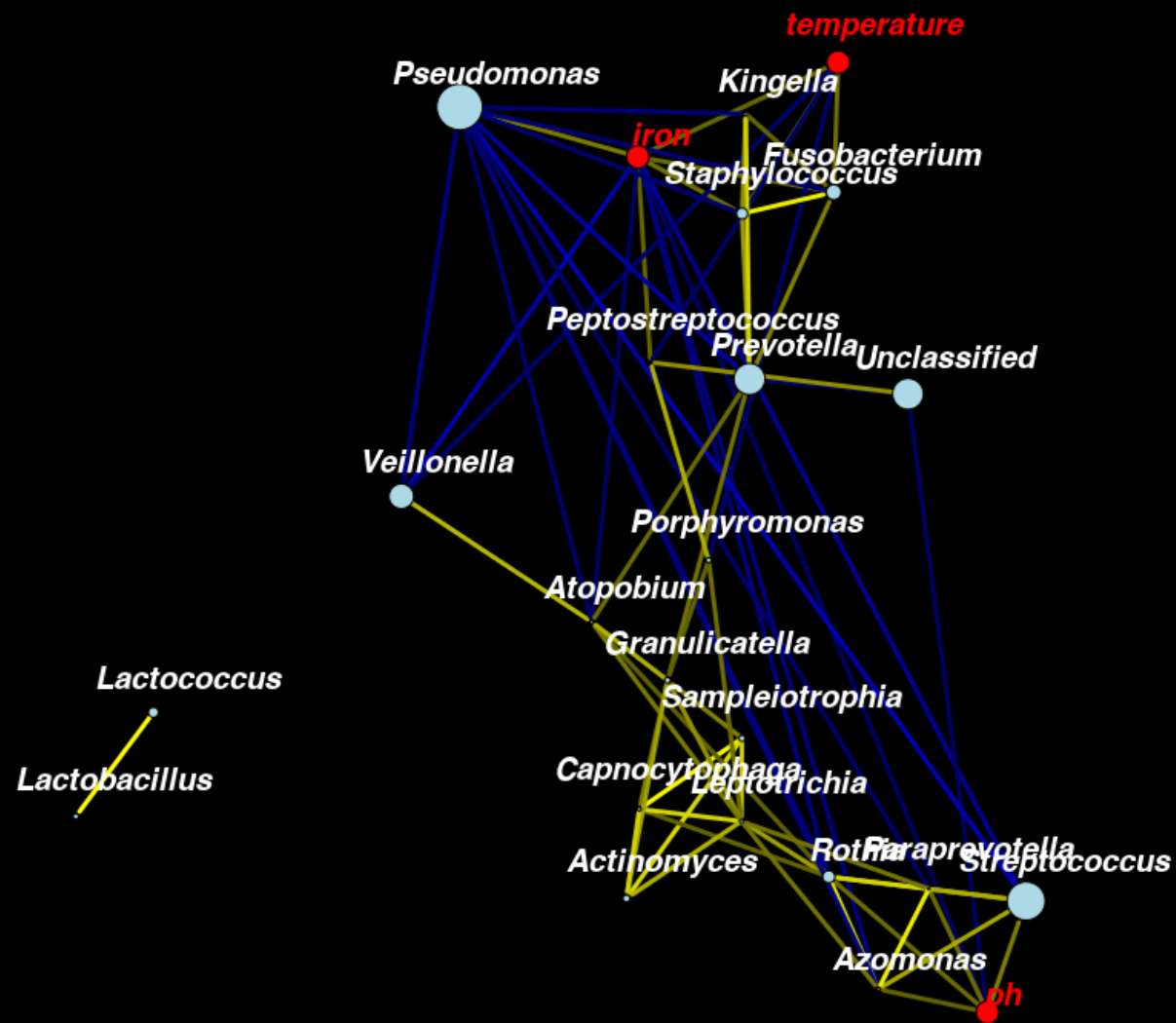


Analysis of Paired Samples

- Example: drug treatment, change of environmental parameters over time
- Tests: Paired t-test, paired Wilcoxon rank test, ...
- Regression analysis

OTU





Longitudinal Data

Linear mixed effect model:

Abundance of taxa = time point + pair,

Where time point is a fixed effect and pair is a random effect

