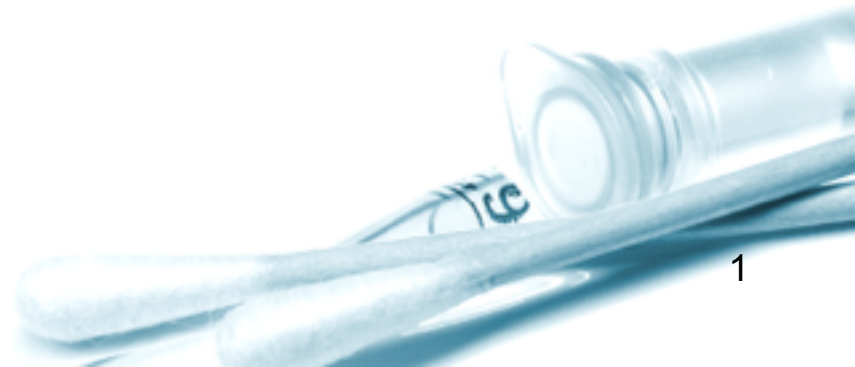


Host Microbiome and aging using 16s approach

Guillermo G. Torres

Ph.D. student

Genetics of Human Longevity Group



About the data

	<i>All</i>	<i>Males</i>	<i>Females</i>
<i>BSP/SPC</i>	133	68	65
<i>FOC</i>	101	49	52
<i>AGI</i>	9		
<i>Samples</i>	241	117+agi_M	117+agi_F

About the data

A whole genome sequencing (metagenes) was performed to 12 samples using our budget plus 9 from centenarians.

FOC_236, FOC_253, FOC_291, FOC_306, FOC_341, FOC_346, FOC_366, BSPSPC_168, BSPSPC_462, BSPSPC_468, BSPSPC_475, BSPSPC_489.

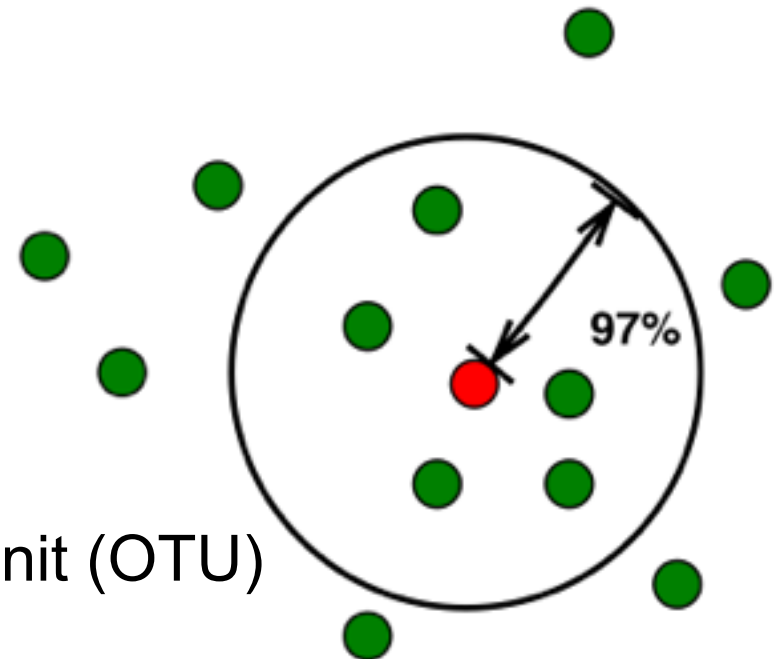
** Fiete, are These included in the whole batch of 54? or are extra.

We are going to use 54 more metagenomes from FOC/SPC batches for later analysis. From this 47 sequencing libraries got so far.

Working with data

- » Sequences are binned into OTU according sequence similarity.

	<i>No. Seqs.</i>
<i>Raw seqs.</i>	11674227
<i>Quality Control</i>	8573796
<i>OTUs</i>	203000



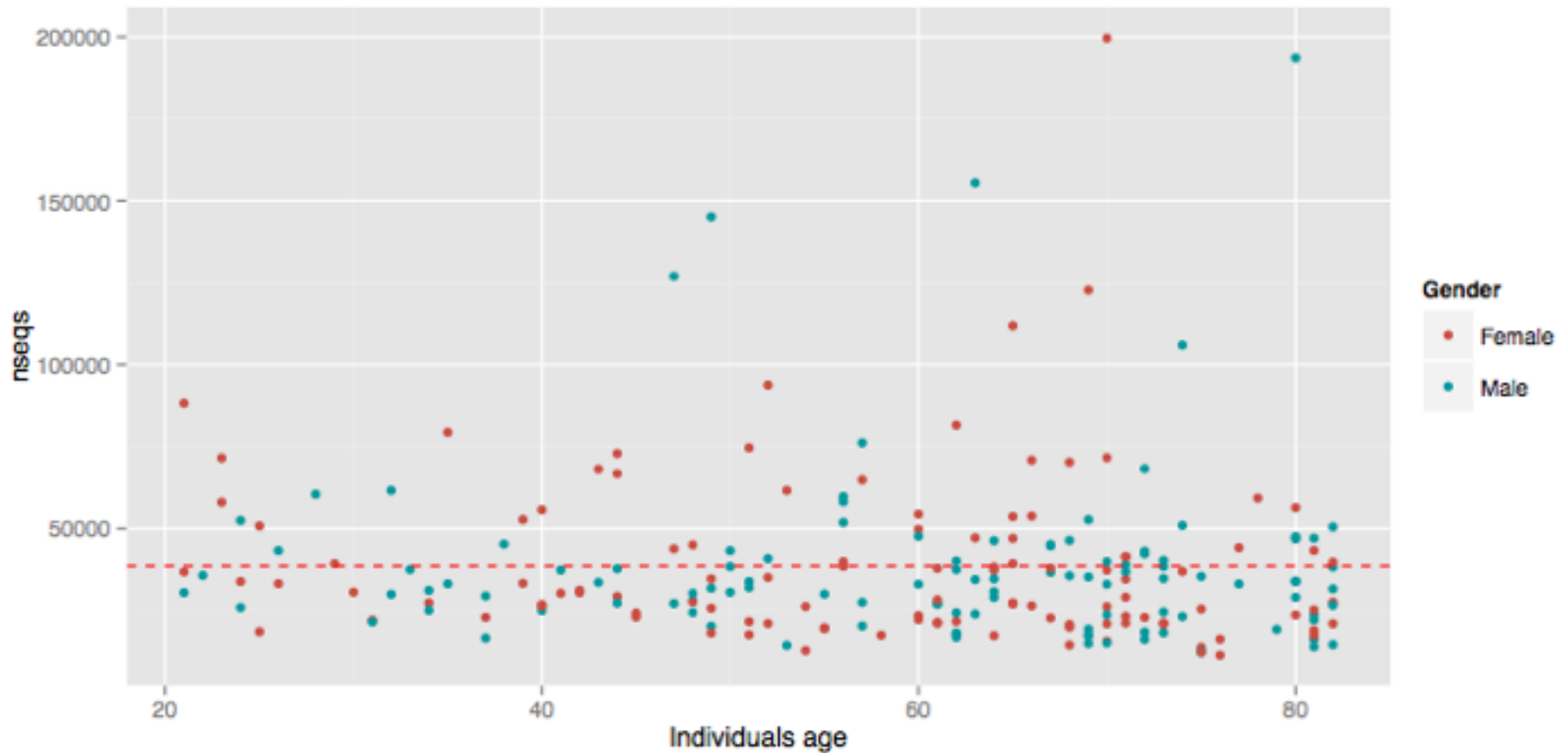
Operational Taxonomic Unit (OTU)



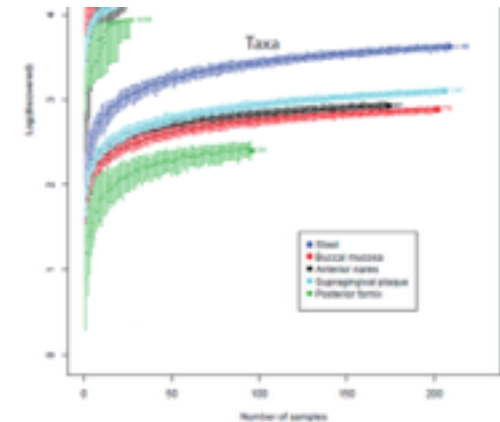
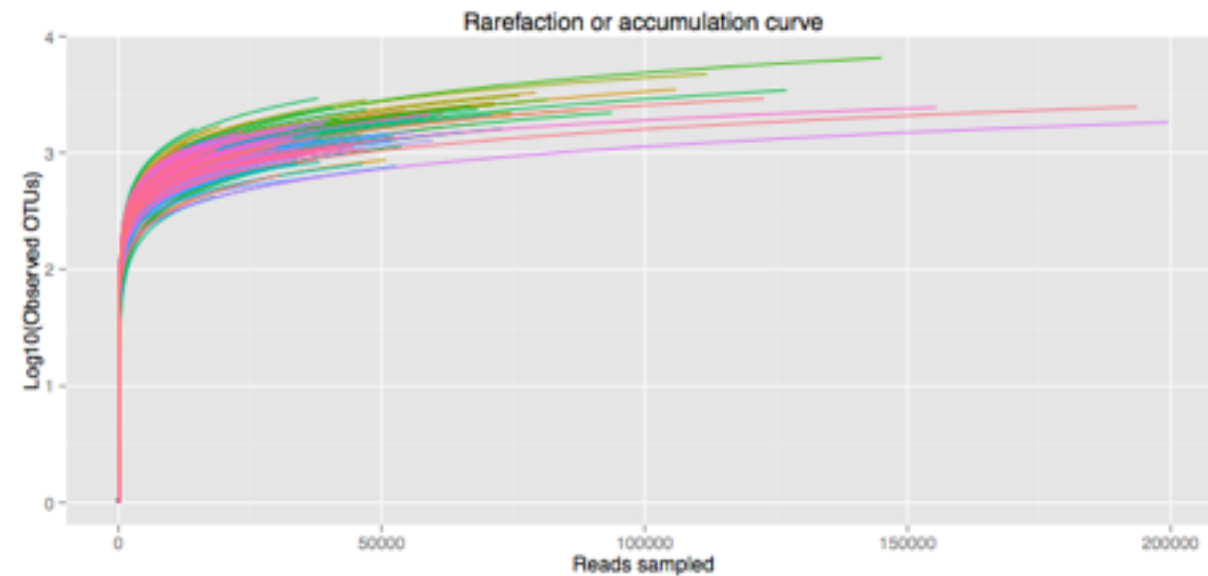
General overview

Alpha diversity

Reads per sample



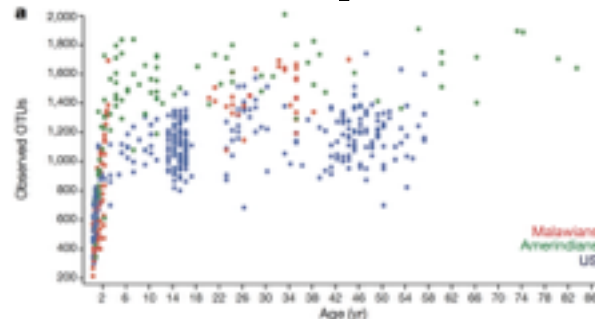
Rarefaction or cumulative curves



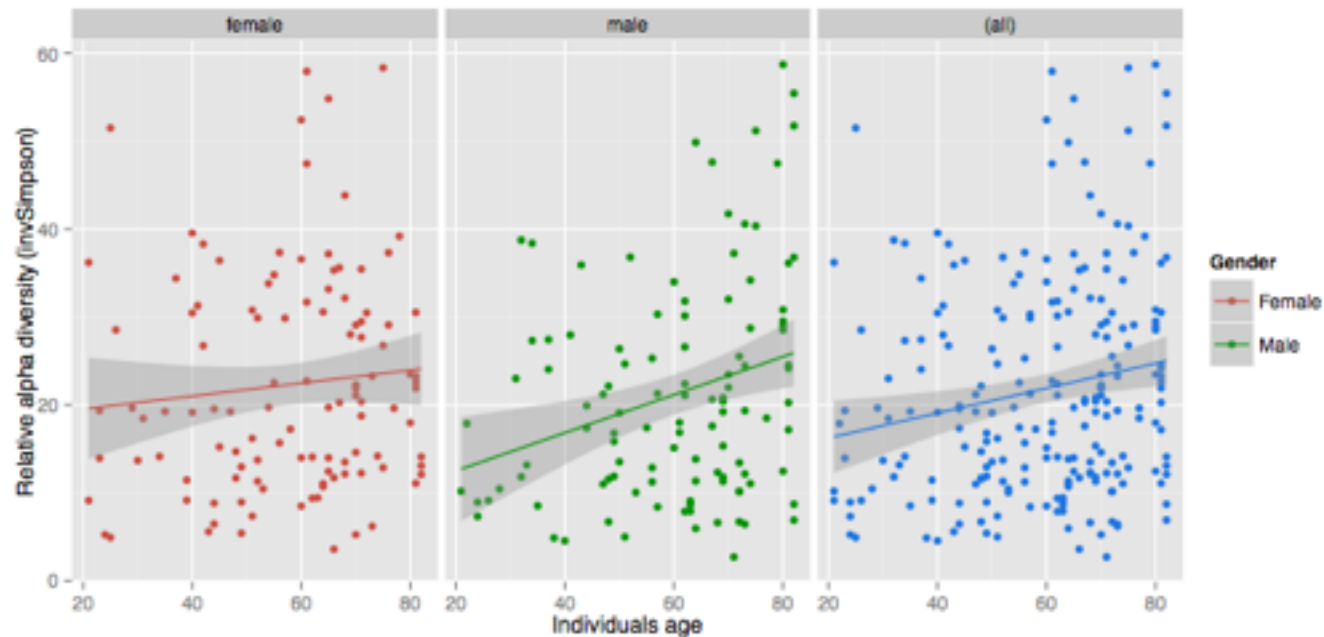
Supplementary Figure 9. Rates of gene and OTU discovery from five body sites using HMP taxonomic and metagenomic data. For improved ease of viewing, the accumulation curves for genes ("Genes" top) and OTUs ("Taxa" bottom) for five applicable body sites sampled by the HMP (and shown in Fig. 1a and 1b) are compared in this figure. The values given for each curve are final median bootstrap values.

The Human Microbiome Project Consortium 2012 b

OTUs per sample



Yatsunenکو *et al.* 2012



222 indiv.

110 Female

112 Male.

OTU classification

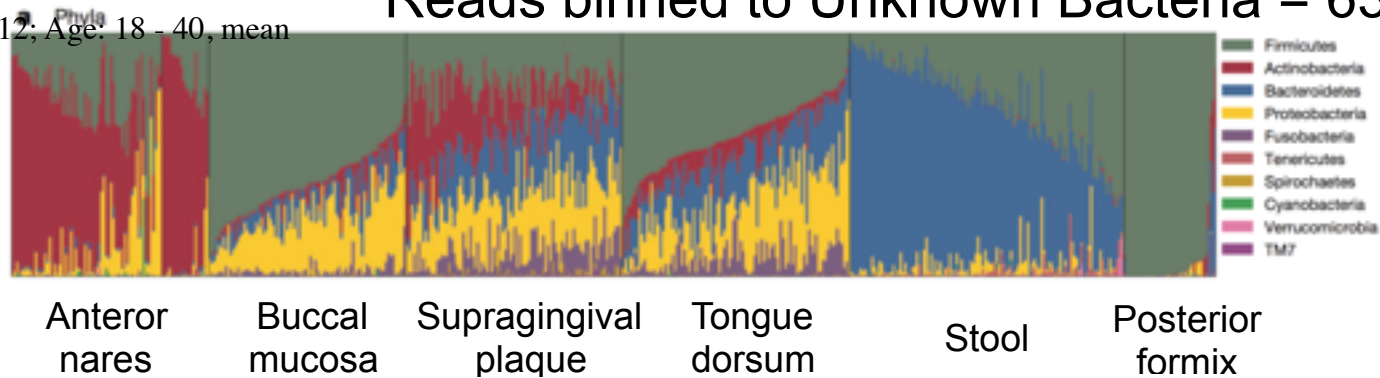
The Human Microbiome Project

Consortium 2012; Age: 18 - 40, mean

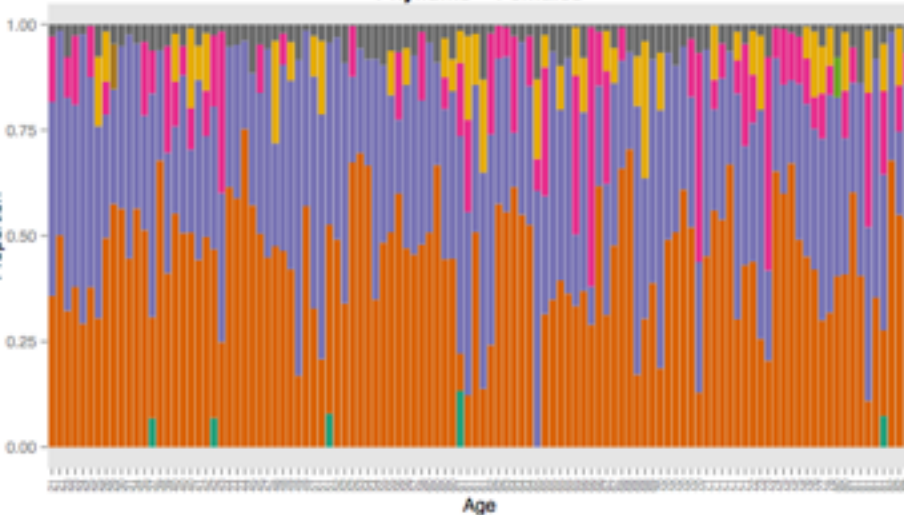
age of 27 sd. 5

All reads binned to certain OTU = 8,573,796

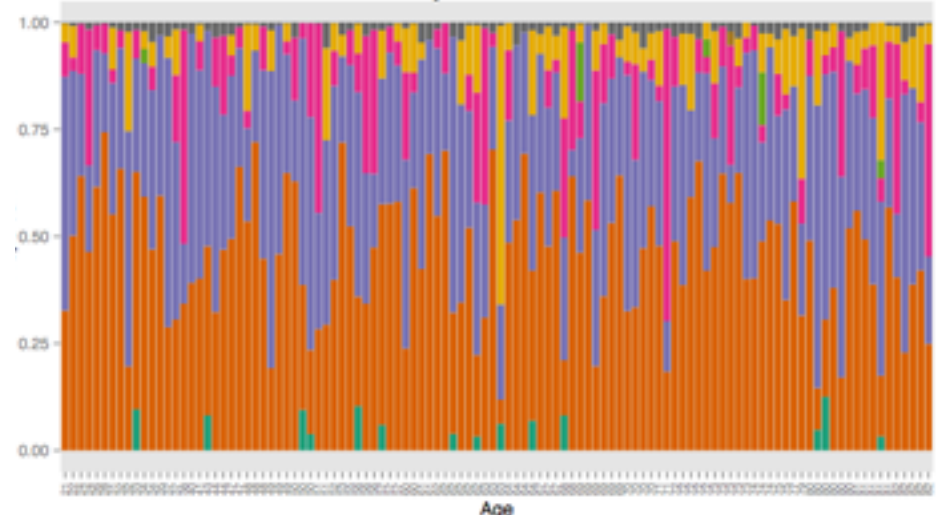
Reads binned to Unknown Bacteria = 638,980



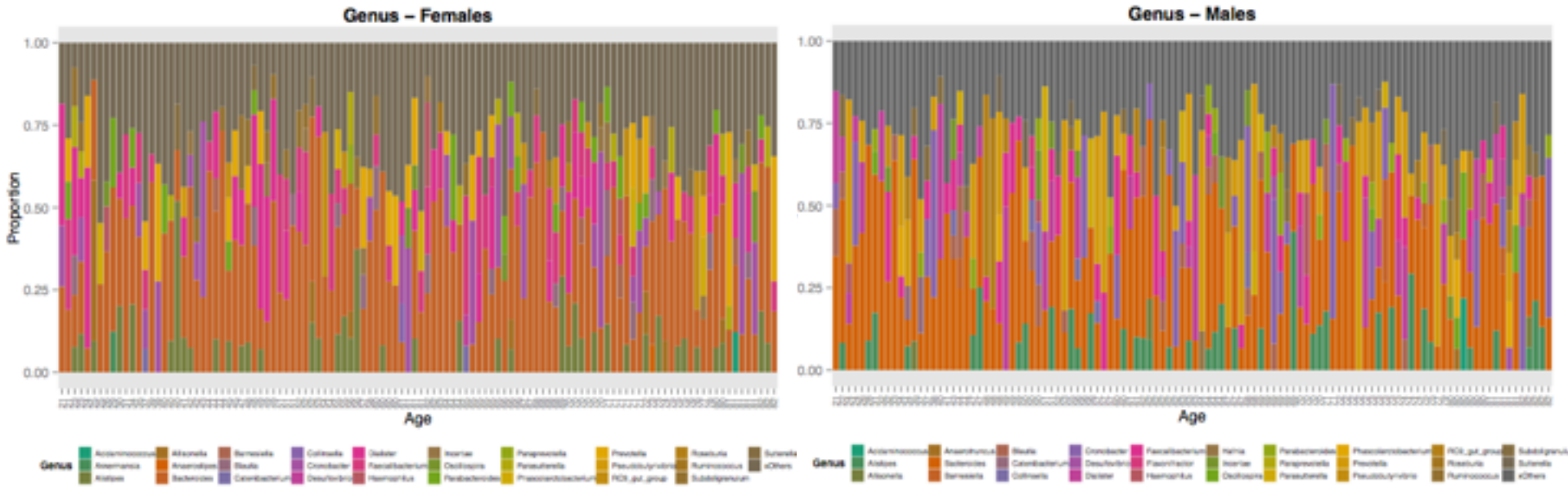
Phylums – Females



Phylums – Males



OTU classification



Bacteroides genus dominating.

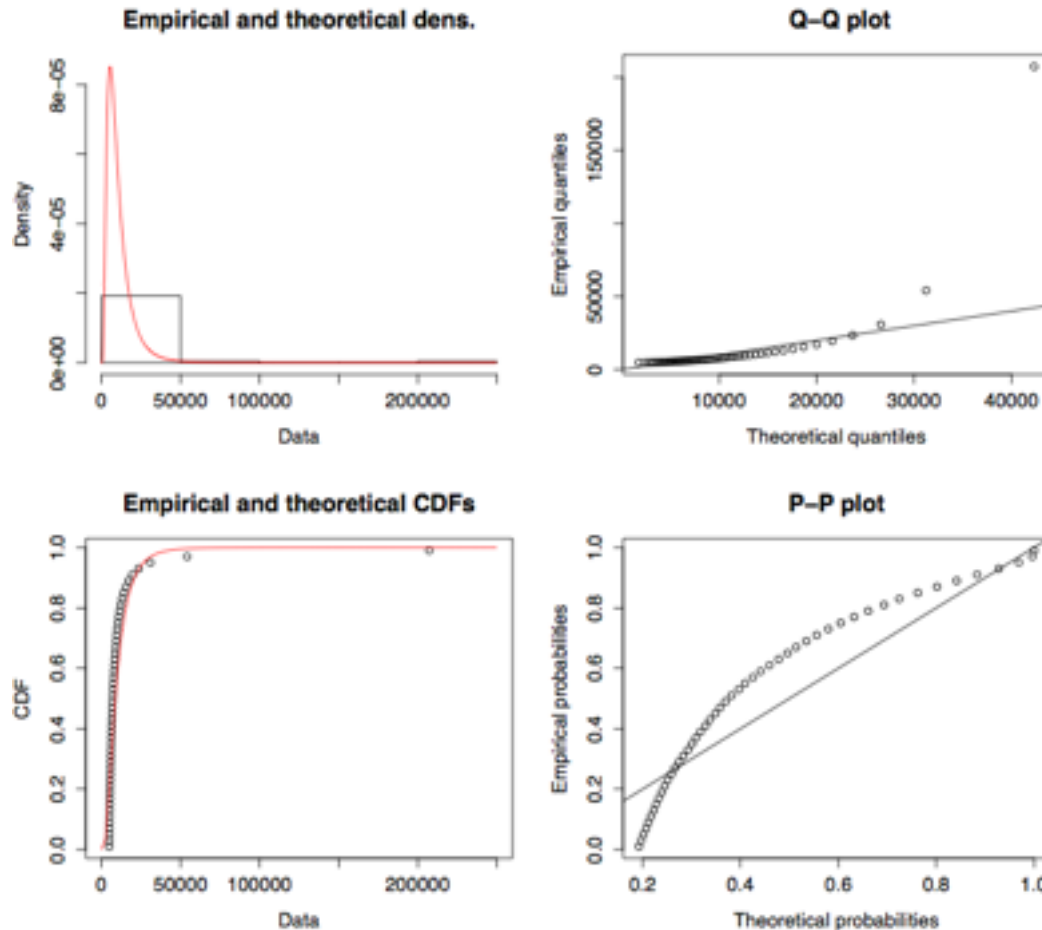


Networks approach

workflow

- 1) In ecology the competition between organisms could be seen from their presence abundanc.
 - » Patterns of co-occurrence due to ecological relationships (mutualism, comensalism, etc.) or just niche overlapping.
- 2) The interactions will be represented as a network (mathematical representation of real-world complex system).
- 3) Use the age variable to represent network serial time.
- 4) Calculate network global and local parameters to understand the dynamics of the organism interactions.

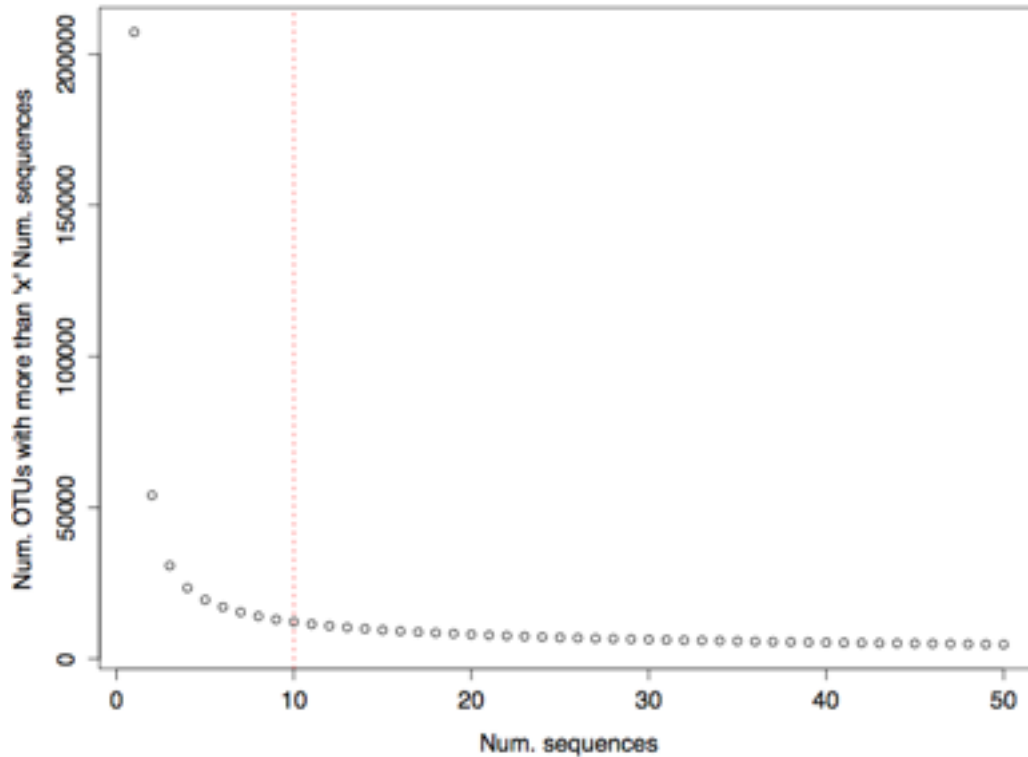
Data treatment



Raw data does not fit well to log-normal distribution

Data treatment

Distribution of OTU sizes



Many OTUs low covered

We estimated the population size of each sample based on rare OTUs using Chao and ACE estimators.

According to ACE, 85% of the samples estimate less than 12000 OTUs.

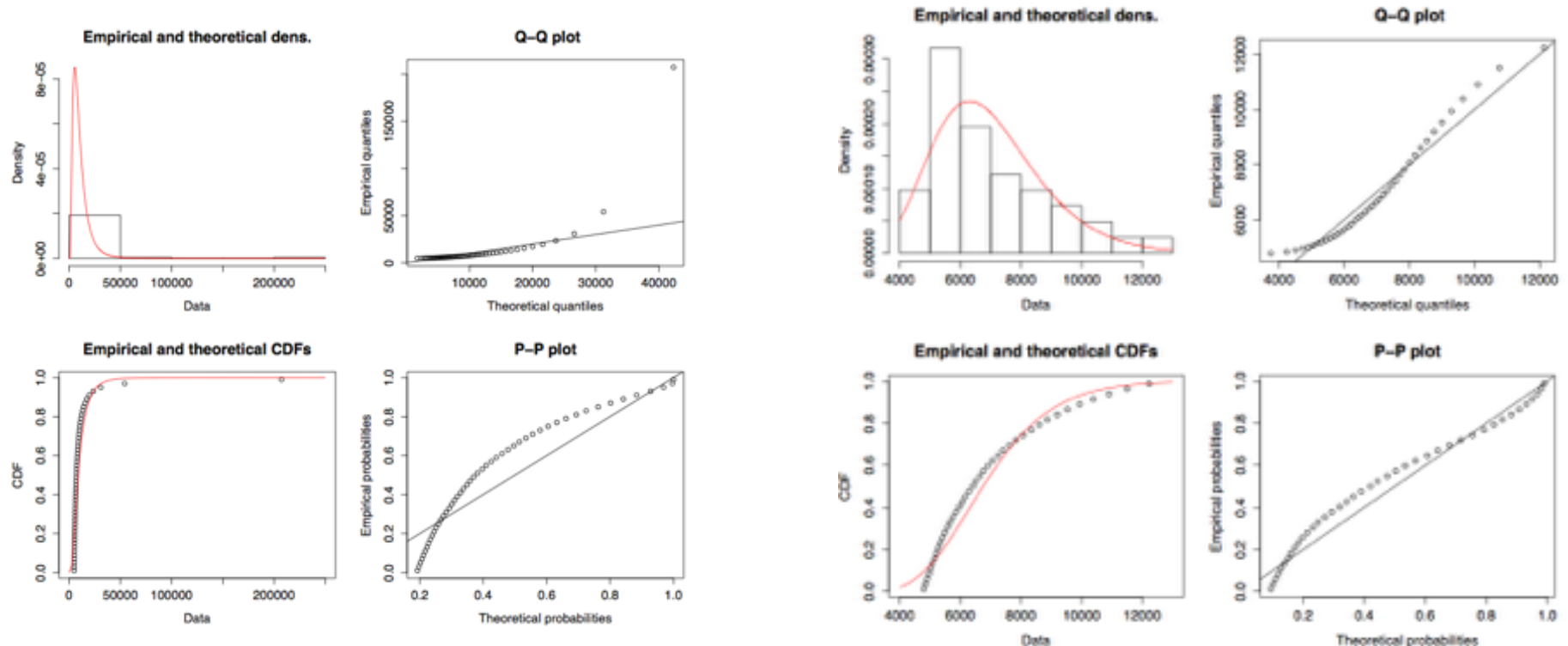
Congruent with rarefaction.

Around 12K OTUs with are formed with 10 read counts or more.
OTU size threshold = 10 read counts

Data treatment

Raw

Information lost = 3.3% - Trimmed



After trimmed OTUs with fewer read counts than 10 fits better to log-normal. Probably better distribution adjustment would be to free-scale

Data treatment

$$x_i = (c_{i1}/N_i, \dots, c_{ik}/N_i)$$

$$r_i = (r_{i1}, \dots, r_{ik})$$

$$r_{ij} = \begin{cases} t_{ij} \frac{s_i}{N_i + s_i} & \text{if } x_{ij} = 0, \\ x_{ij} \left(1 - \frac{s_i}{N_i + s_i} \sum_{k|x_{ik}=0} t_{ik} \right) & \text{if } x_{ij} > 0. \end{cases}$$

Most OTU appears in a few individuals with low abundance, and zero values

dominate most count values. Data structure falls into special Non-sufficient sampling effort. data class named as

compositional data (Aitchison

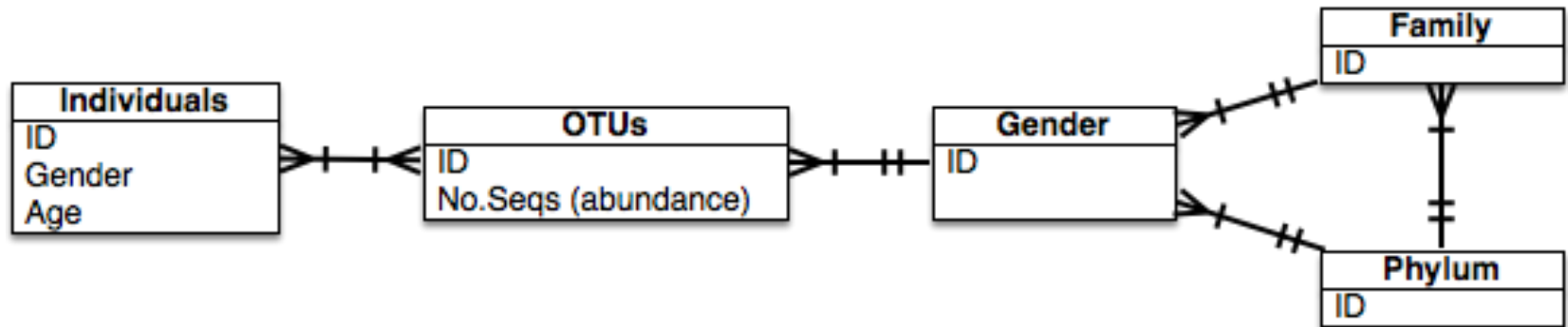
For replacing the zero counts by a proper value, 2003)

we managed the data using the Bayesian-

multiplicative approach (Pawlowsky-Glahn et. al.

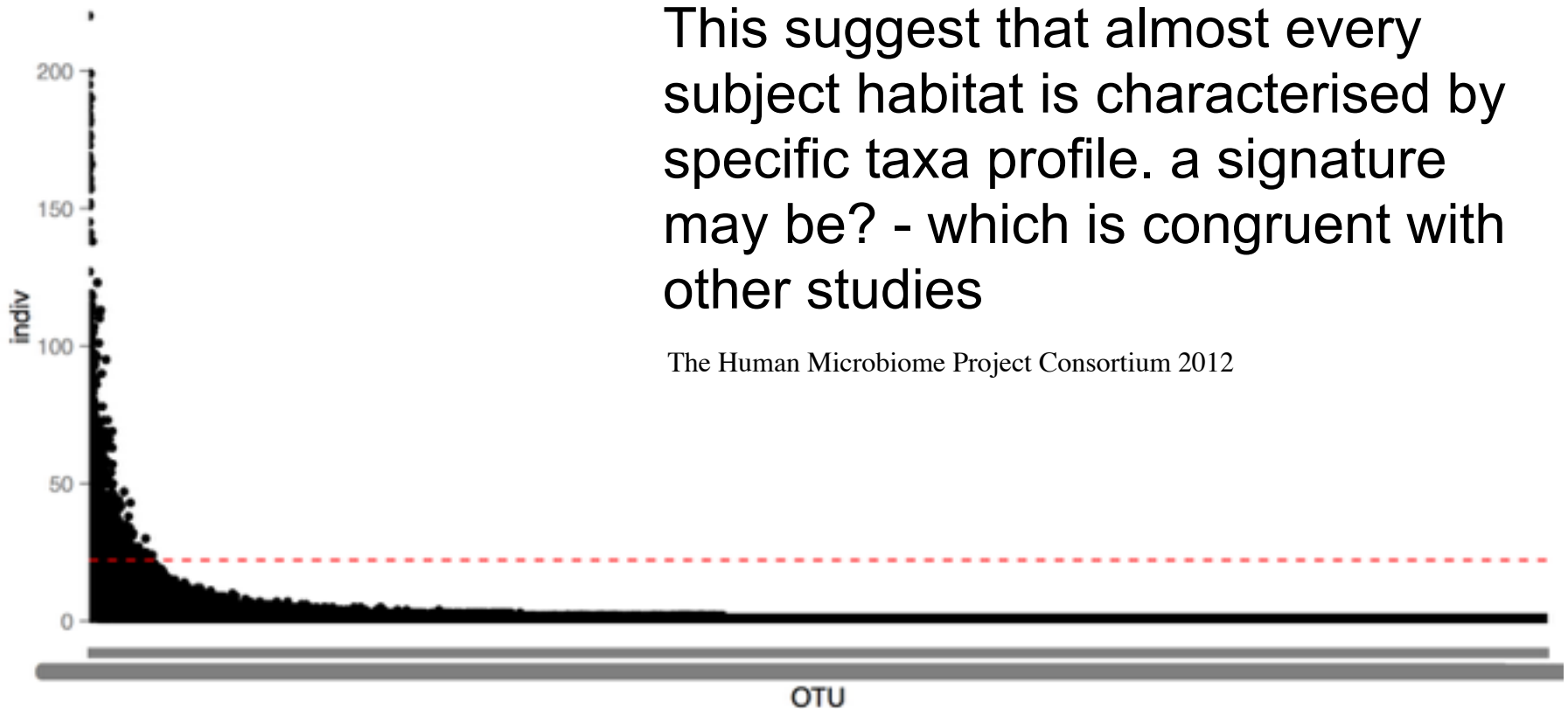
2011-chap4).

Data structure



- Construction of OTU correlation matrix. ~12K OTUs found.
- OTU filtration; assessed OTUs will be those that occurs in at least 15% of the samples. This reduction left 300 OTUs
- filtered OTUs may have accounted for **particularities** of each sample.

OTU in individuals

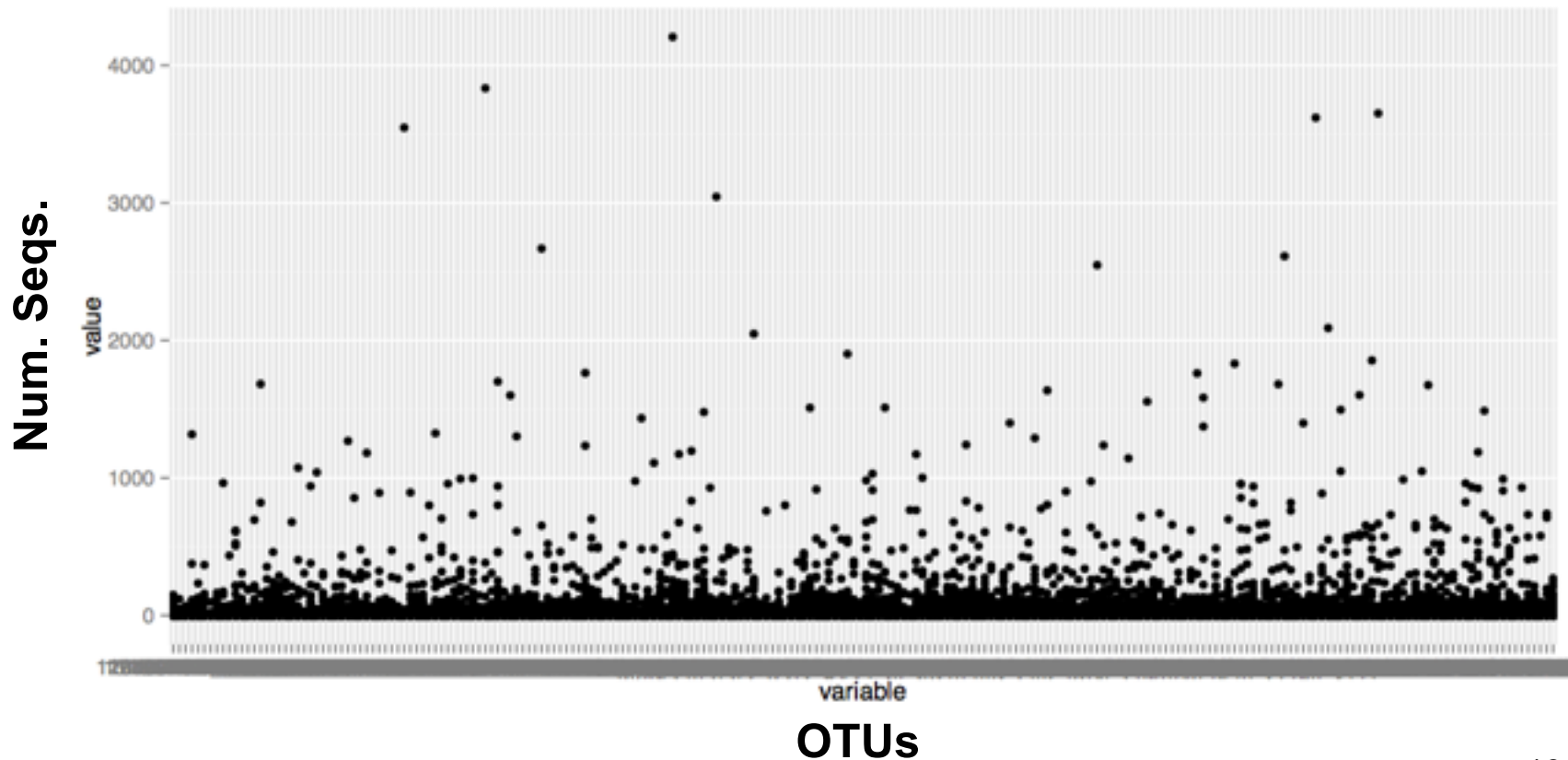


This suggest that almost every subject habitat is characterised by specific taxa profile. a signature may be? - which is congruent with other studies

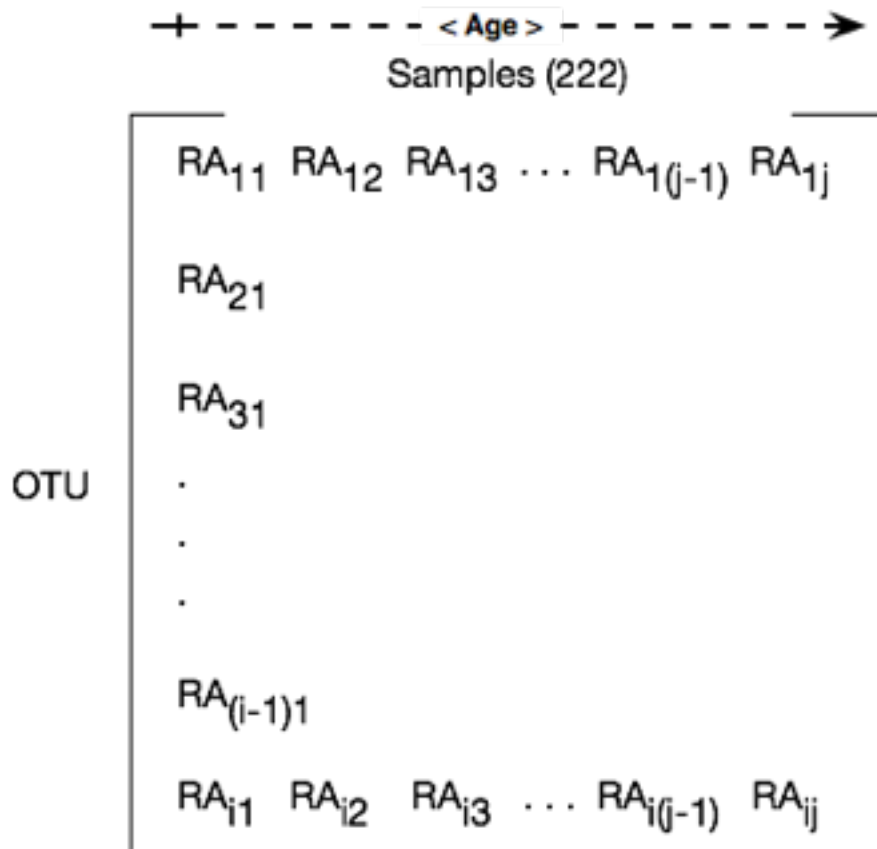
The Human Microbiome Project Consortium 2012

OTUs discarded

Get more deep into this later.



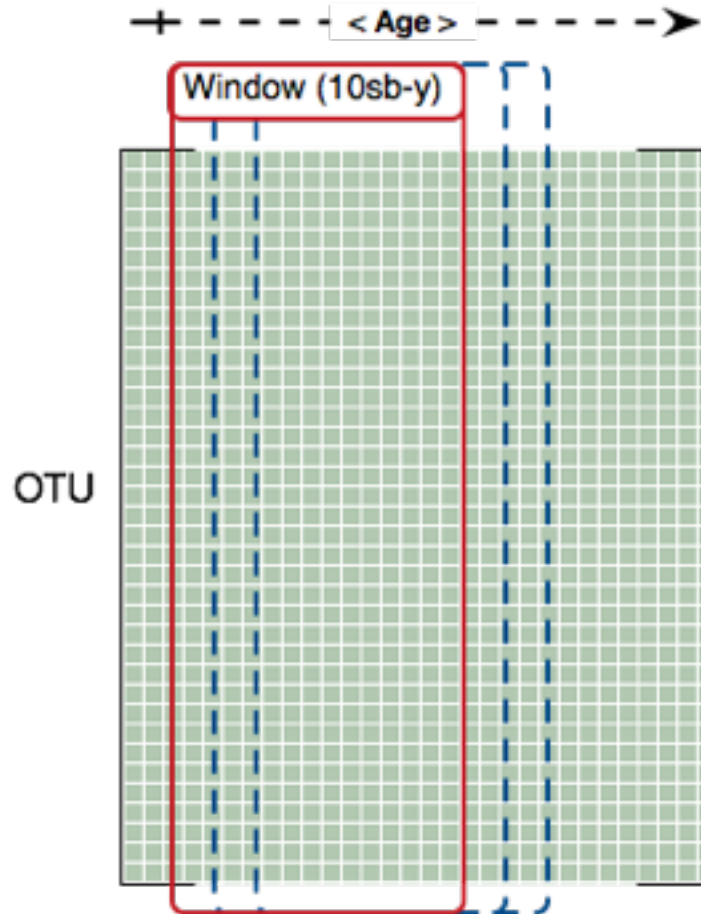
Data structure



Where **RA** means the transformed relative abundance of the OTU in an individual.

Due to data follows a compositional data statistics we use log-ratios for further analysis (Aitchison 2003).

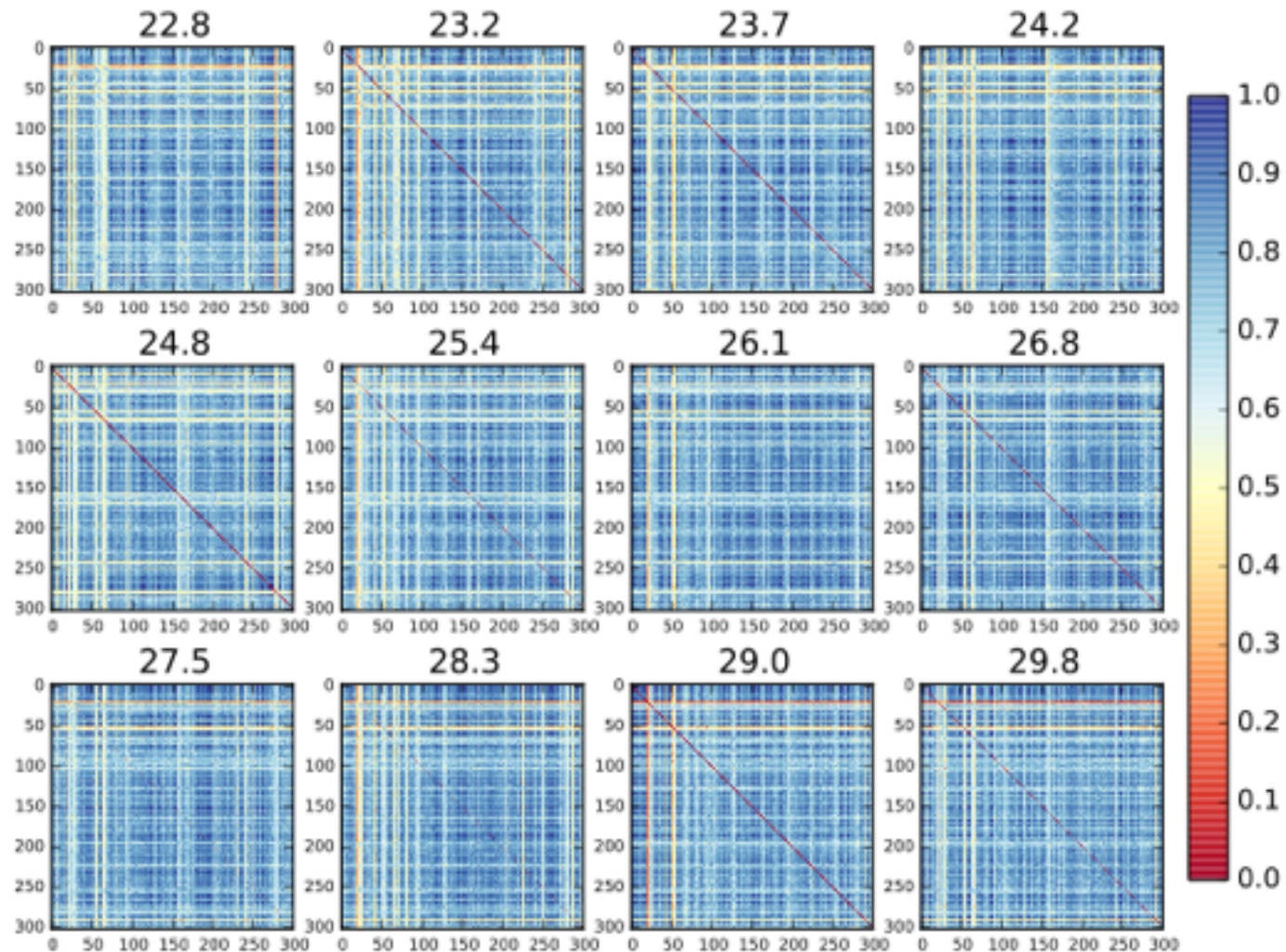
Correlation matrix



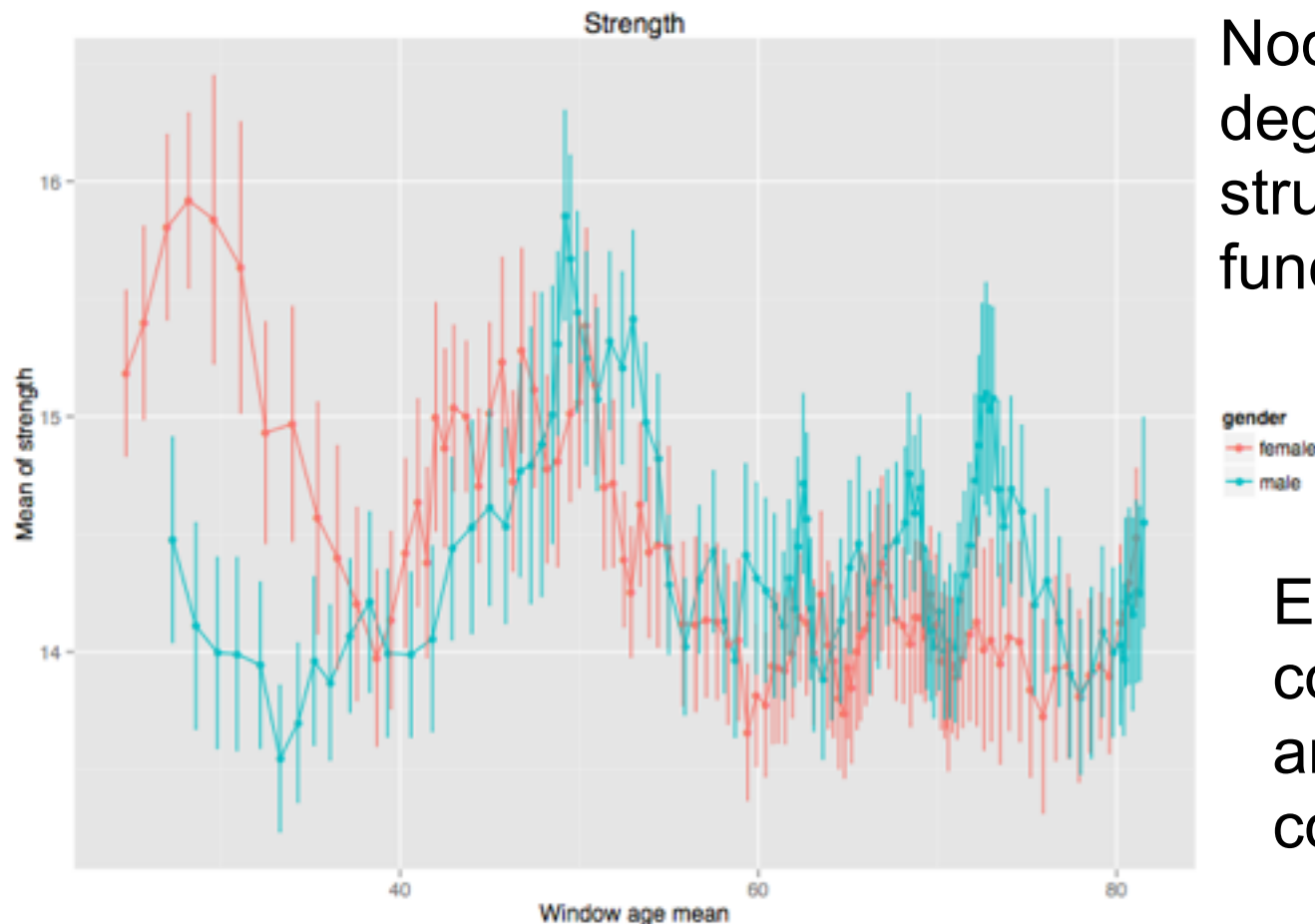
Using a sliding window we generate multiple correlation matrices.

Later evaluation as time series.

Correlation network matrices



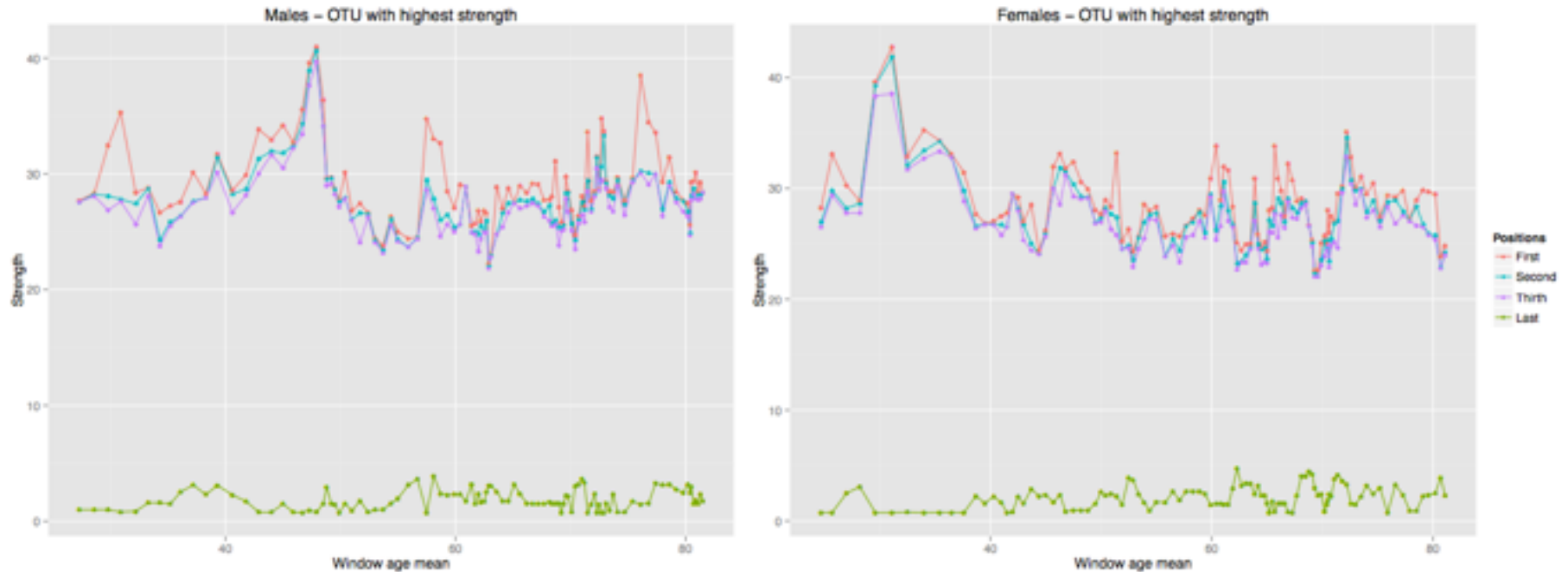
Network metics



Nodes with high degree are interacting, structurally or functionally.

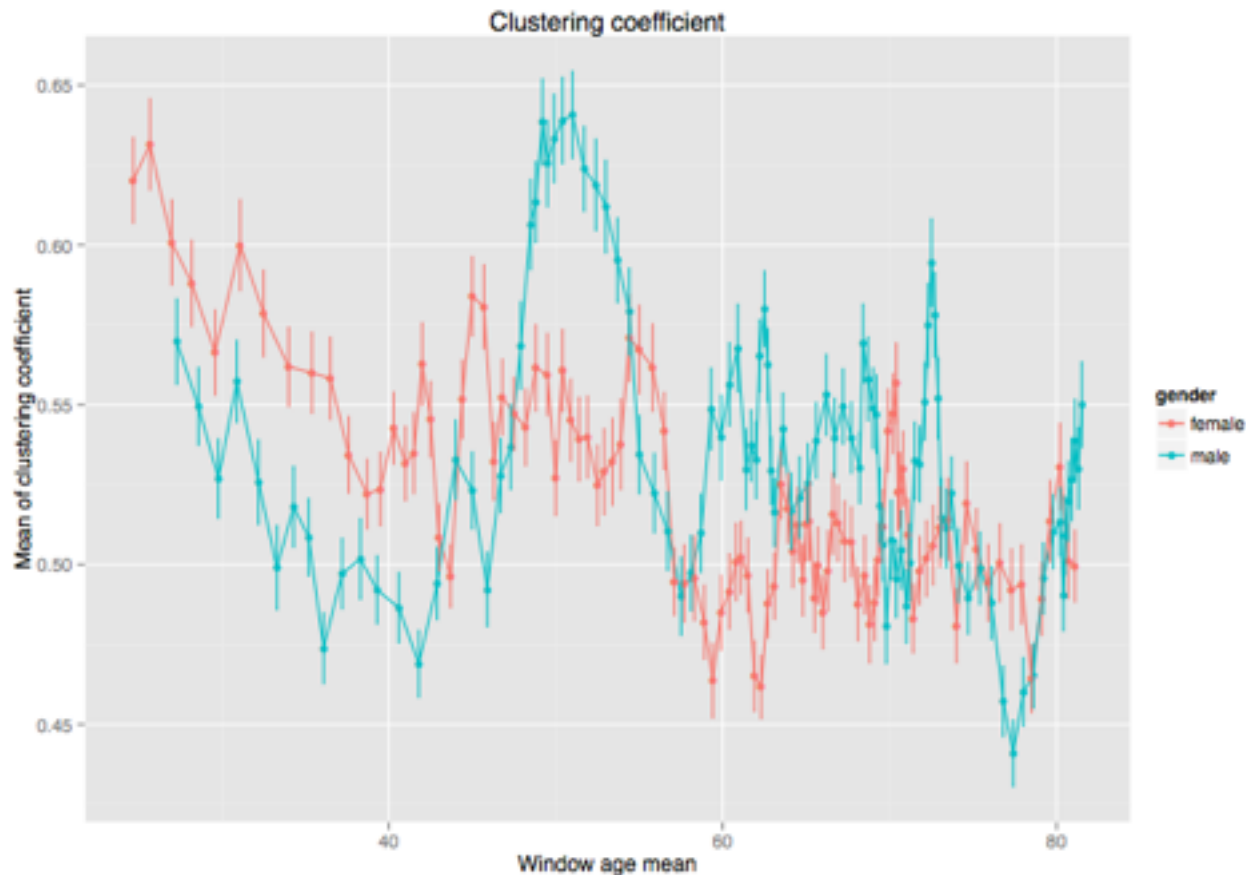
Elderly loose high connected nodes and gender convergence.

Network metics



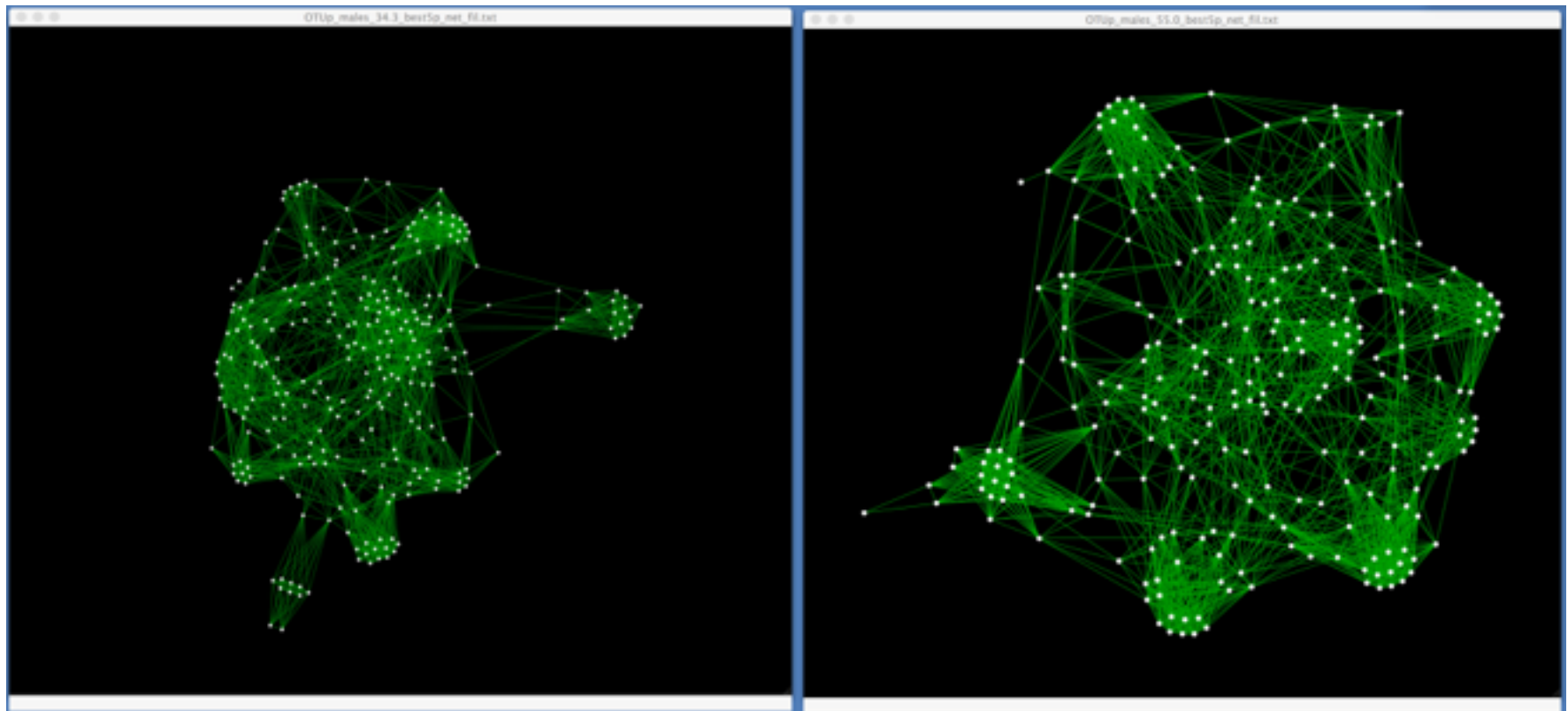
Hubs between males and females look a bit different.
Get in deep later.

Matrice metics



The level of clustering between male and female are different specially around 50's.

Topologies of the network for males with mean age of 34 (left) and 55 (right). Contrasting clustering and strength values.



Ongoing work...

Our first approach to generate the correlation matrixes using log-ratios and cosine distance give us many spurious values therefore we are exploring:

- 1) John Baines suggests to use Sparcc algorithm (2012) and relative ones (REBACCA - 2015) to generate the correlation matrix.
- 3) We are exploring to generate a Covariance matrix to assess OTUs relationships using SPIEC-EASI algorithm (2015). (more complex than Sparcc)

Next steps

- 1) Once the “relationship” matrices are done, collaborative partners from Spain (Dr. Javier Buldú, at the Center for Biomedical Engineering of the Technical University of Madrid) will assist me to analyze this time series framework to model the dynamics in microbial relationships through aging.
- 2) For most relevant microbial clusters we are going to generate its possible functional profile, and then with the assistance of Christoph Kaleta, we will perform a flux balance analysis to know whether the cluster has a functional dependency or not.

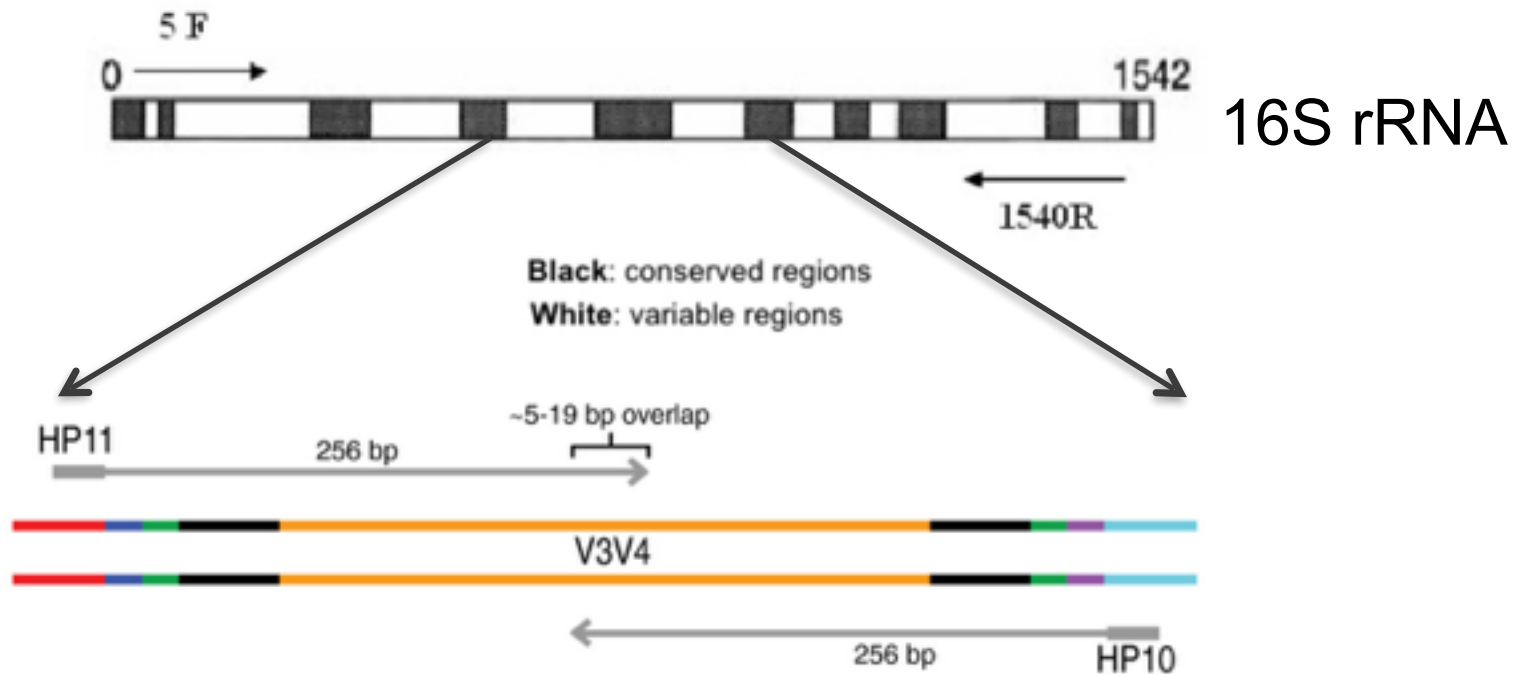




Extra...

Where data comes from?

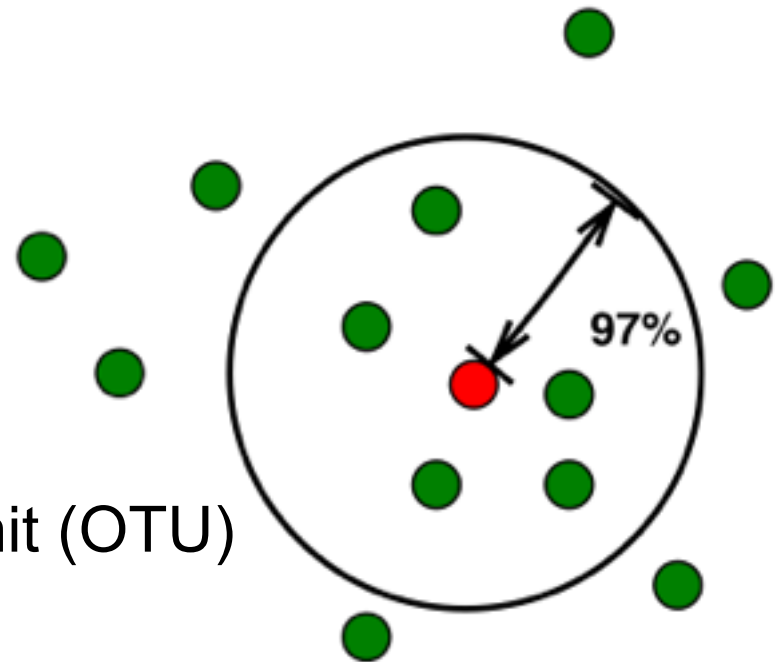
- » 16S gene - exists universally among bacteria
- » Highly conserved regions (V3-V4)



Working with data

» 16S variability rate has been determined experimentally:

- $\geq 99\%$ Species level
- $\geq 97\%$ Genus level
- $\geq 95\%$ Family level



Operational Taxonomic Unit (OTU)

Working with data

Abundance

Richness = No. OTUs

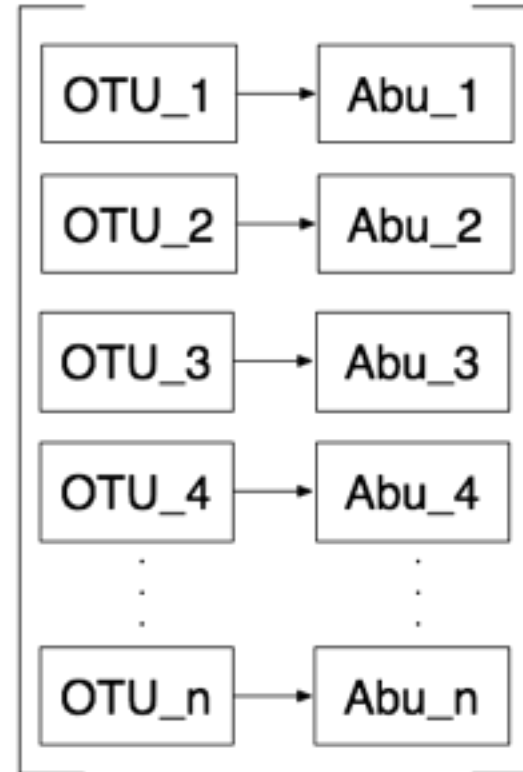
	<i>No. Seq.</i>
<i>OTU_1</i>	<i>50</i>
<i>OTU_2</i>	<i>20</i>
<i>OTU_3</i>	<i>300</i>
<i>OTU_4</i>	<i>100</i>

**Each sample has one
OTU profile**

Alpha Diversity

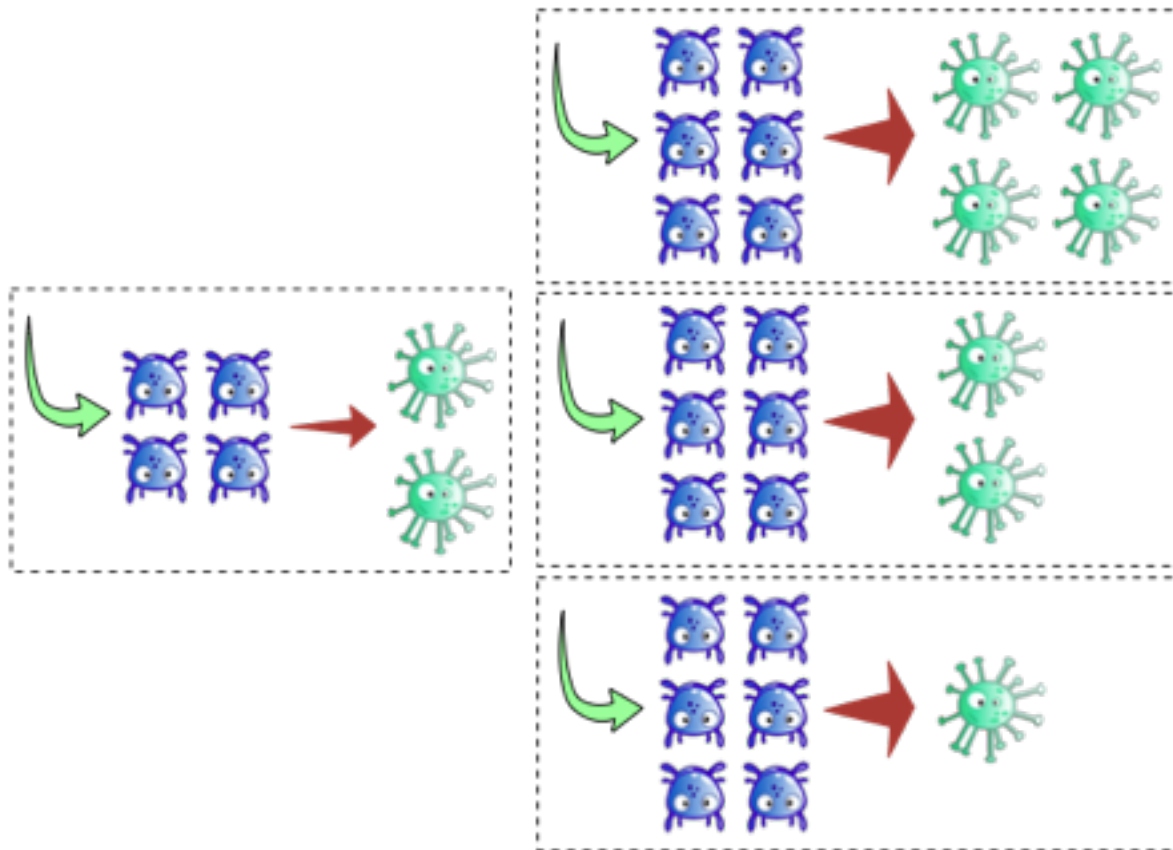
Identify Microbial composition with regard to age

- » Data combines quantitative and qualitative variables besides multiple environmental factors (OTUs).
- » Multivariate statistics is one alternative.
 - Classification
 - Placement of sample units into groups
 - Ordination
 - Arrangement or 'ordering' of sample units along gradients.

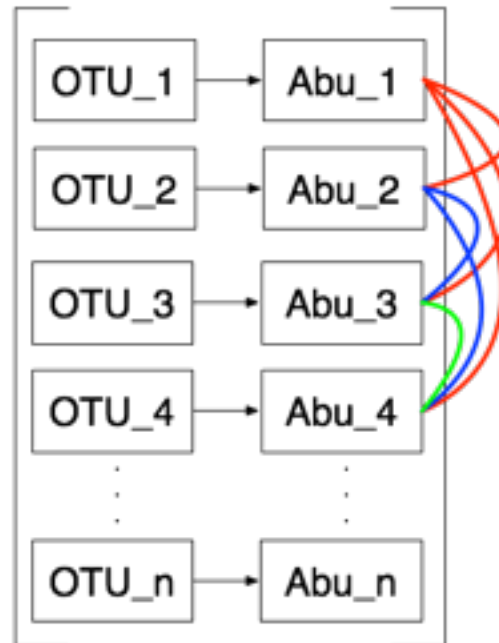
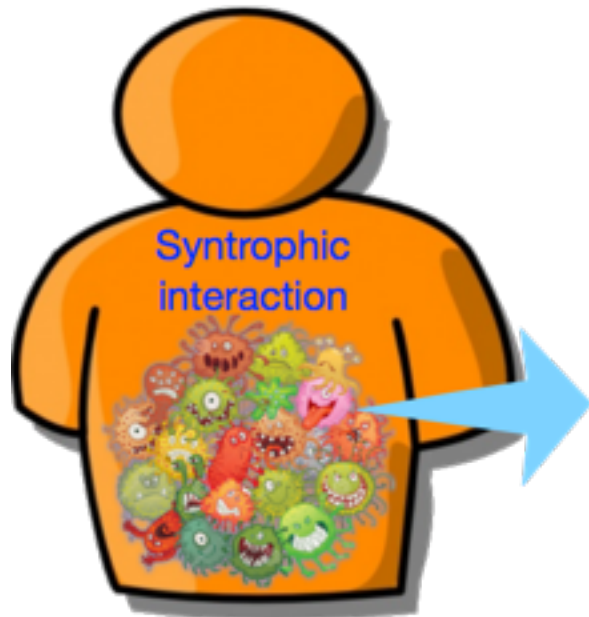


» OTU abundance change through age?
Is the richness changing through age?
Which taxa are changing?

Syntrophy = Cross-feeding, one species lives off the products of another species



Syntrophy = Cross-feeding, one species lives off the products of another species



The relationship will be measured in terms of the abundance.

Then following up through age.








Characterise Relationships among microbiota

- » Complex network analysis.
 - » Holistic approach which allows explore the interactants and their interaction:
 - Micro-scale (OTUs)
 - Meso-scale (Group of OTUs)
 - Macro-scale (Net)

Results

	<i>No. Seqs.</i>	<i>No.</i>
<i>Raw</i>	11674227	
<i>Quality Control</i>	10158262	
<i>Contigs</i>	5079131	
<i>Quimeras</i>	3047478.6	
<i>OTUs</i>	2031652.4	322000



Taxa	OTUs	Abundance
Taxon A	 OTU1	4000 reads
	 OTU2	200 reads
	 OTU3	10 reads
	 OTU4	1 reads
Taxon B	 OTU5	2000 reads
Taxon C	 OTU6	500 reads
Taxon D	 OTU7	10 reads

Data transformation

By column, RA values are transformed to lineal scale

1) linearisation:

$$m_{ij} = \log_2 RA_{ij}$$

$$SO = \begin{bmatrix} m_{a1} & m_{a2} & \dots & m_{aj} \\ m_{b1} & m_{b2} & \dots & m_{bj} \\ \dots & \dots & \dots & \dots \\ m_{i1} & m_{i2} & \dots & m_{ij} \end{bmatrix}$$

Data transformation

By column normalisation. m values are transformed into 0 - 1 scale:

2) Individual normalisation:

$$M_{ij} = \frac{m_{ij} - \min(SO_{.j})}{\max(SO_{.j}) - \min(SO_{.j})}$$

$$SO_{.j} = \begin{bmatrix} m_{aj} \\ m_{bj} \\ m_{cj} \\ \dots \\ m_{ij} \end{bmatrix} \rightarrow \begin{bmatrix} M_{aj} \\ M_{bj} \\ M_{cj} \\ \dots \\ M_{ij} \end{bmatrix} \mid \{M \in R: 0 \leq M \leq 1\}$$

Data transformation

$$SO_{sin} = \begin{bmatrix} M_{a1} & M_{a2} & \dots & M_{aj} \\ M_{b1} & M_{b2} & \dots & M_{bj} \\ \dots & \dots & \dots & \dots \\ M_{i1} & M_{i2} & \dots & M_{ij} \end{bmatrix}$$

$$\dim(SO_{sin}) = [samples, OTUs] = [222, 300]$$

OTU - OTU correlation

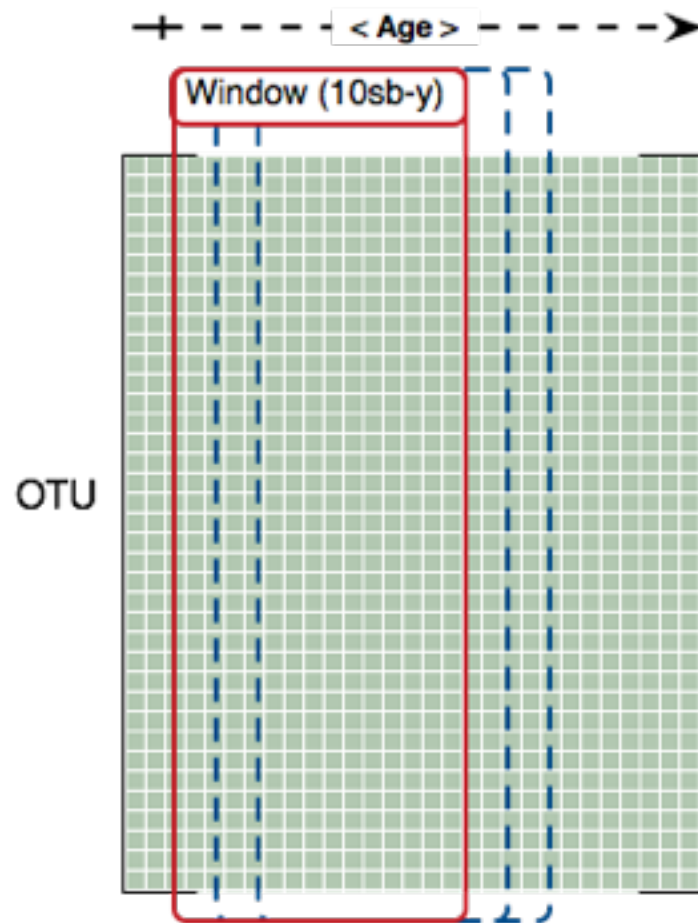
$$SO_{sin,j} = \begin{bmatrix} M_{aj} \\ M_{bj} \\ \dots \\ M_{pj} \\ \dots \\ M_{ij} \end{bmatrix}$$

" M_{aj} " represents the abundance of the OTU "a" in the organism "j" then:

- 1) Calculate all the OTU permutations
- 2) Calculate the distance "d" of each permutation

$$d_{ap} = M_{aj} - M_{pj} \implies Dis_j = \begin{bmatrix} d_{aa} \\ d_{ab} \\ \dots \\ d_{ai} \\ \dots \\ d_{ii} \end{bmatrix}$$

SO_{sin}



OTU - OTU correlation

$$D_w = \begin{bmatrix} d_{aa}^1 & d_{aa}^2 & \dots & d_{aa}^k \\ d_{ab}^1 & d_{ab}^2 & \dots & d_{ab}^k \\ \dots & \dots & \dots & \dots \\ d_{ai}^1 & d_{ai}^2 & \dots & d_{ai}^k \\ \dots & \dots & \dots & \dots \\ d_{ii}^1 & d_{ii}^2 & \dots & d_{ii}^k \end{bmatrix}$$

Dw matrix contains distance vectors ($Dis.j$) of individuals in slide window.

$$w_{pq} = 1 - sd(Win[pq,]) \implies Cor = \begin{bmatrix} w_{aa} \\ w_{ab} \\ \dots \\ w_{ai} \\ \dots \\ w_{ii} \end{bmatrix}$$

OTU - OTU correlation

$$\dim(Cor) = [OTUs, OTUs] = [300, 300]$$

$$Tensor = [Cor_1, Cor_2 \dots, Cor_k]$$

$$\dim(Tensor) = [(samples - Win_length) + 1] = [211]$$