

# Microbial ecology principles: diversity, succession and biogeography

ANSC 516

January 12, 2023

# Today's topics?

1. The vastness of microbes and their complexity
2. Describing/comparing microbial communities
3. Microbial community succession

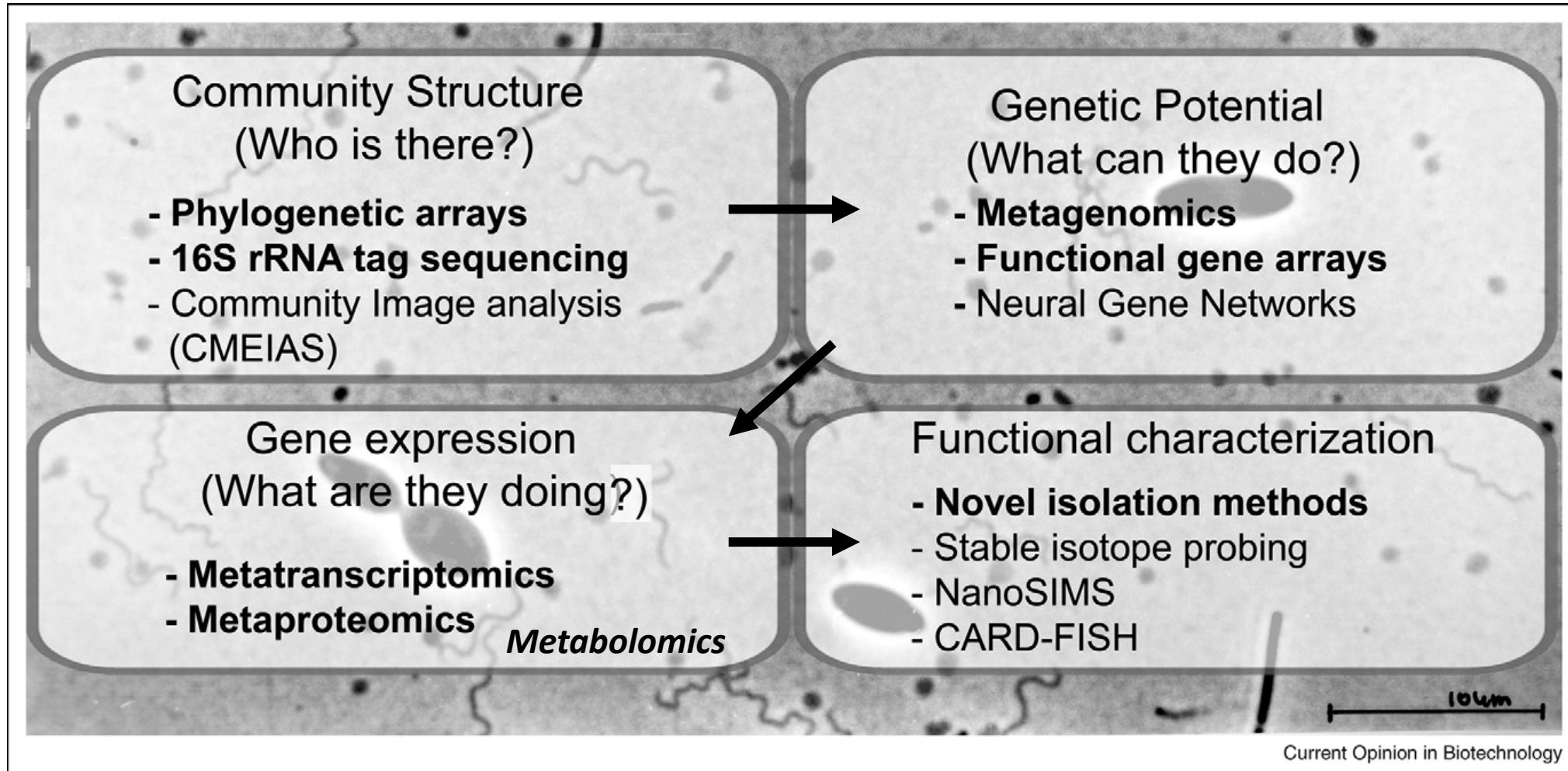
# Today's topics?

- 1. The vastness of microbes and their complexity**
2. Describing/comparing microbial communities
3. Microbial community succession

# What is microbial ecology?

- A branch of ecology dealing with microbes (bacteria, fungi and archaea)
- “The study of microbes in the environment and their interactions with each other.” - ISME website
- It’s the sociology (development, structure, and functioning) and psychology (mind and behavior) of microbes
- Despite their small size, they have a huge impact on us and on our environment.
  - Health and disease
  - Global nutrient cycling
  - Antimicrobials and antimicrobial resistance
  - Biodiversity
  - Metabolic potential
  - Biotechnology

# What Questions Does One Ask About a Community?



What do the differences mean?

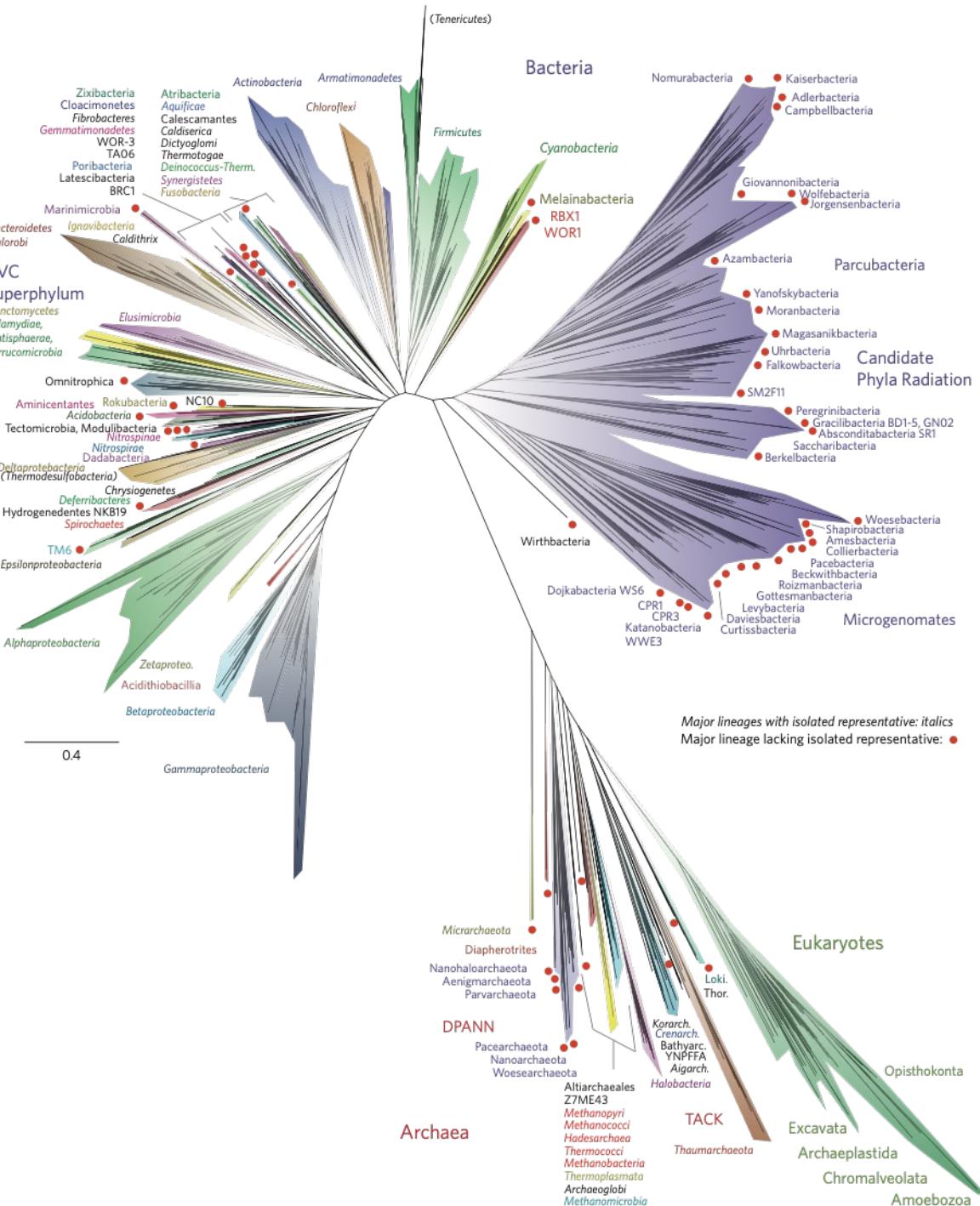
Can the community be managed?

# What makes microbial ecology exciting?

- Microbes rule the earth
  - Microbes are everywhere: air, water, soil, bottom of ocean, in acid, hot
  - Abundant: Soil:  $10^9/g$ , Intestinal tract:  $10^{12}/g$
  - Complex relationships: Several thousand species in a gram of soil.
  - Metabolic processes only they can do.
  - Control global nutrient cycles: carbon, nitrogen, phosphorus
  - Most types of microbes remain unknown.
- Archaea, Bacteria, Fungi, Protists, Viruses (tree of Life)

# Tree of Life:

relationship  
between all  
known phyla

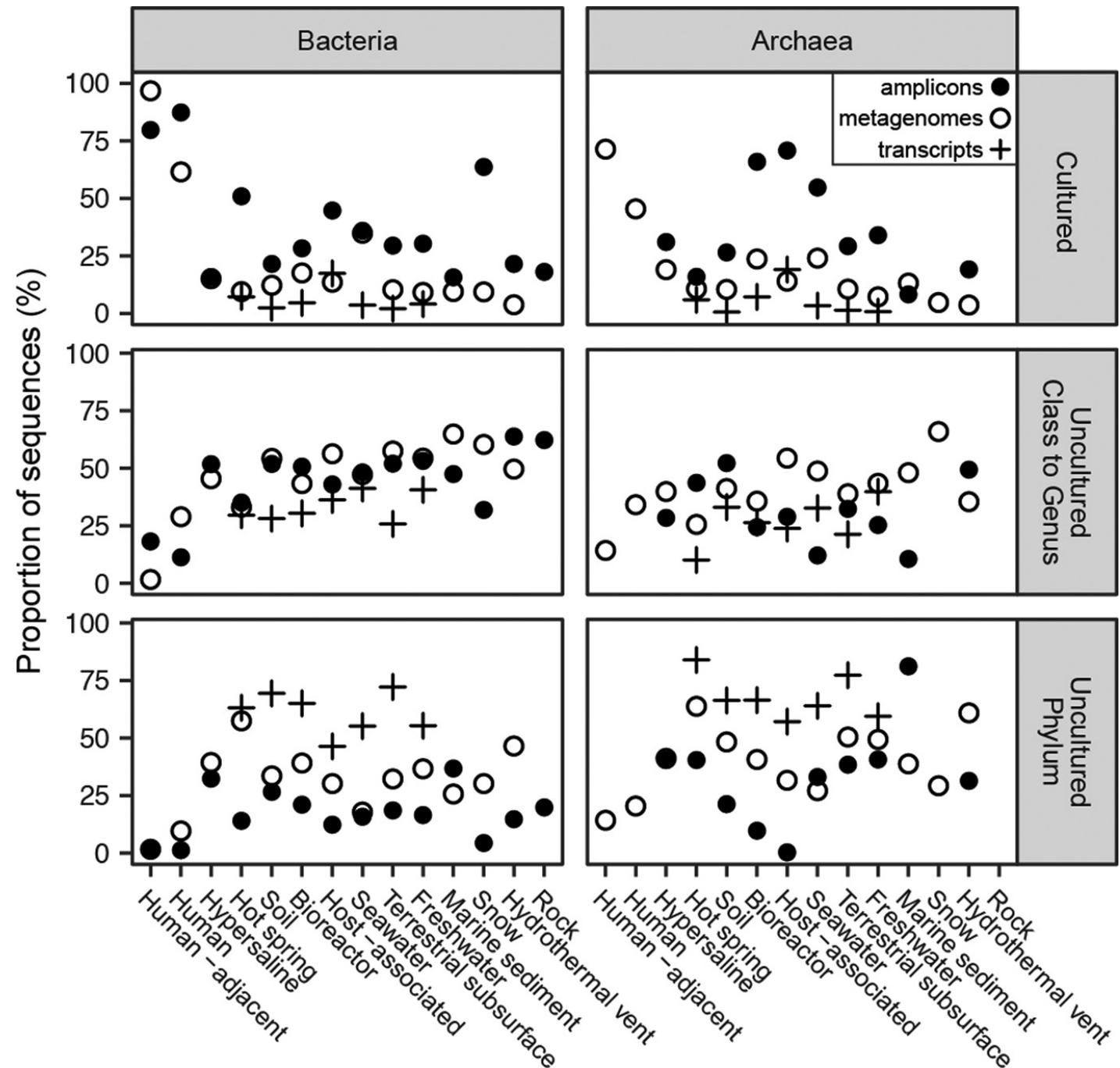


92 bacterial phyla  
26 archaeal phyla  
5 eukaryotic supergroups

Animals, fungi  
Flagellate protozoa  
Land plants, some algae  
Brown algae, diatoms  
Amoebas, slime molds

Hug, Nat Microbiol, 2016

Outside of human environments, most microbes remain uncultured



# Today's topics?

1. The vastness of microbes and their complexity
2. **Describing/comparing microbial communities**
3. Microbial community succession

## 2. Describing microbial communities

So we talk categorize, summarize, count and compare: DIVERSITY

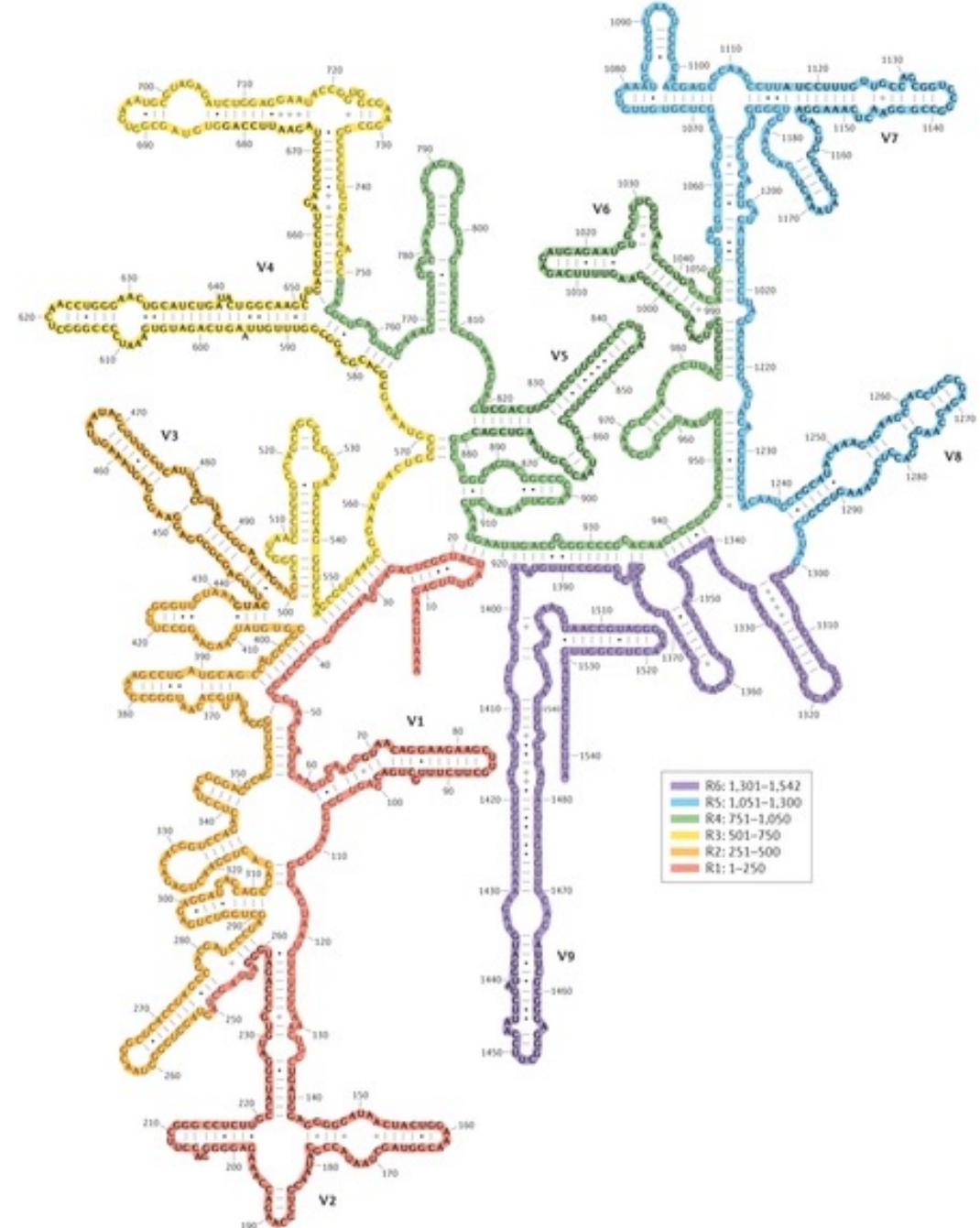
Describe interactions between microbial taxa

What can these organisms do metabolically?

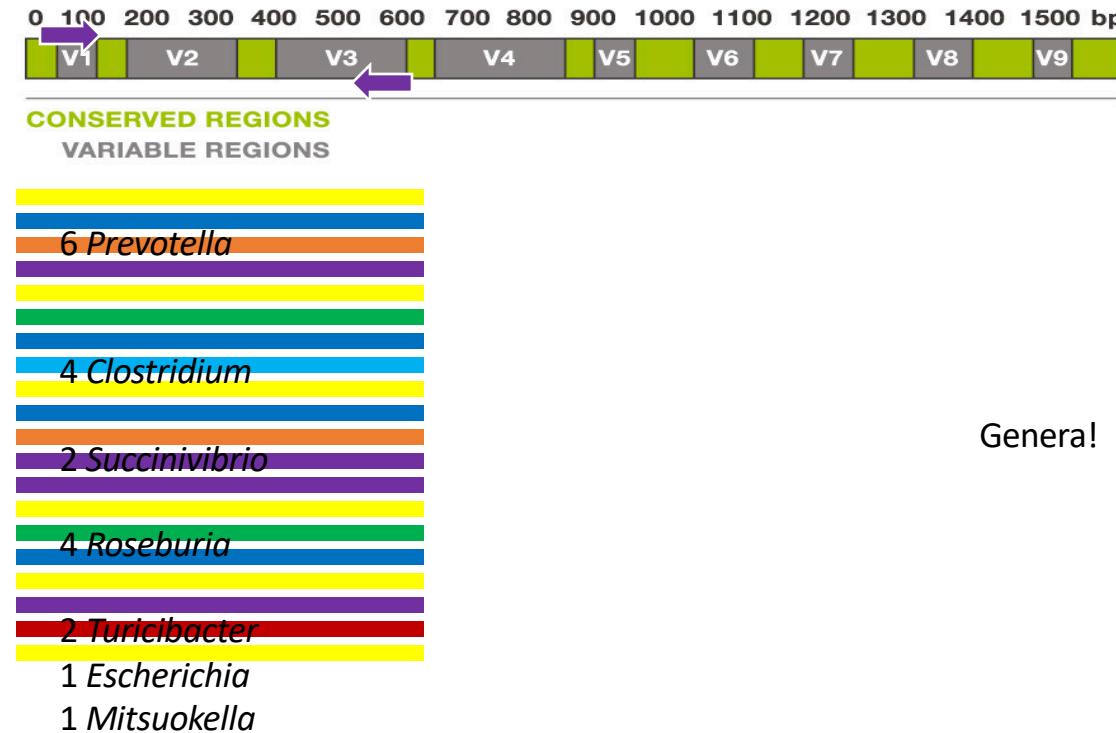
	16S amplicons	Genomes	metagenomics
<b>Alpha diversity</b>	Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?	If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.	Which microbiota are the most functionally diverse, with the most gene families?
<b>Beta diversity</b>	Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?	Beta diversity plots on the basis of taxonomic and functional composition are usually highly correlated. Therefore, functional genes have considerable phylogenetic signal.	Which factors, such as age, culture or disease state, correlate with differences in the overall composition of functional genes?
<b>Machine learning or classical statistics</b>	Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?	Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.	Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?
<b>Co-occurrence</b>	Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.	Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.	Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.
<b>Metabolic network modelling</b>	What metabolic activities are predicted for a collection of phylotypes?	Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.	What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.

# 16S rRNA gene sequence analysis

- Why the 16S gene and not some other gene?
  - It confers an essential function
  - It is present in every bacterial species
  - It is a good evolutionary clock, with both conserved regions and variable regions
- Potential pitfalls
  - Is it truly universal?
  - Subspecies resolution?
  - Species definition?

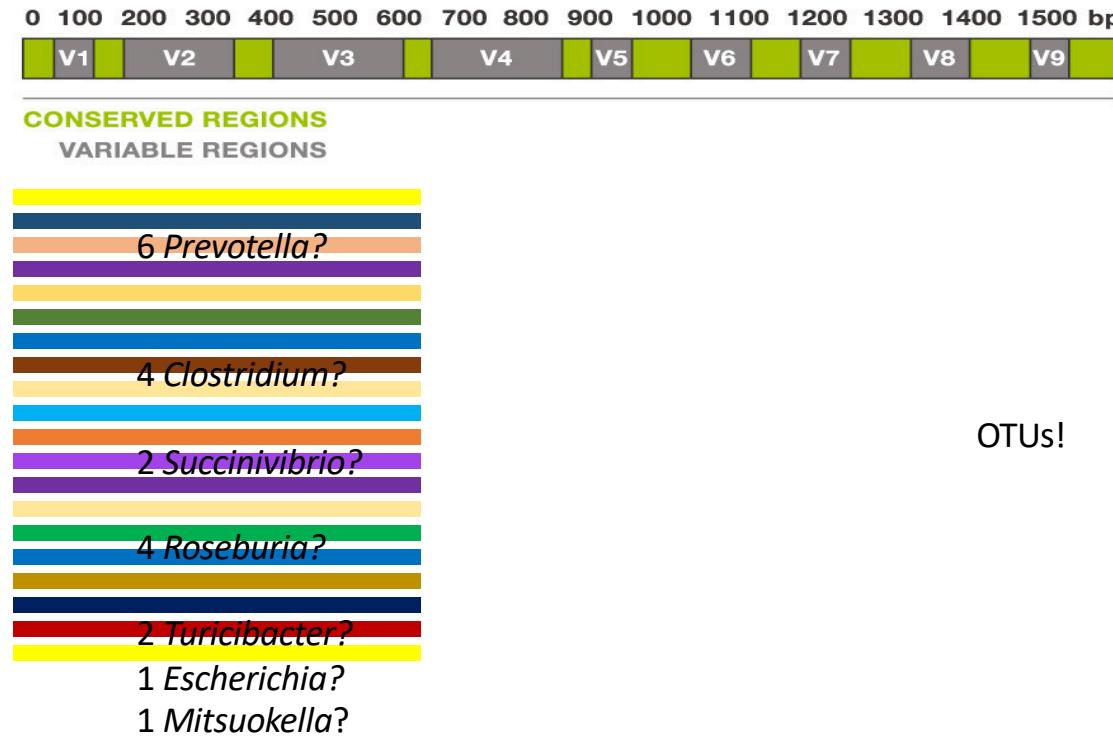


# 16S rRNA gene → microbiota analysis



Genera!

# 16S rRNA gene → microbiota analysis



# Community structure analyses

6 *Prevotella*

4 *Clostridium*

2 *Succinivibrio*

4 *Roseburia*

2 *Turicibacter*

1 *Escherichia*

1 *Mitsuokella*

- Any grouping of organisms is called a **taxon** (plural = taxa)
- Examples: species, genus, OTU
- **OTU** = operational taxonomic unit
- Choice of what criteria for grouping is important:
  - 97% similarity? Species-level
  - 95% similarity? Genus-level
  - Particularly important to mention in OTU-based analyses

## 16S amplicons

## Genomes

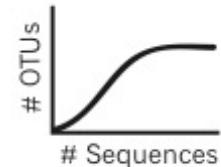
## metagenomics

### Alpha diversity

Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?

If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.

Which microbiota are the most functionally diverse, with the most gene families?

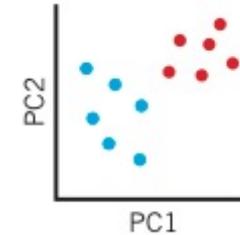


### Beta diversity

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?

Beta diversity plots on the basis of taxonomic and functional composition are usually highly correlated. Therefore, functional genes have considerable phylogenetic signal.

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of functional genes?



### Machine learning or classical statistics

Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?

Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.

Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?



### Co-occurrence

Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.

Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.

Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.

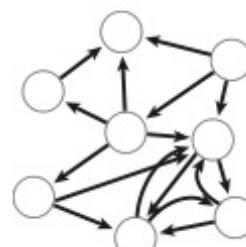


### Metabolic network modelling

What metabolic activities are predicted for a collection of phylotypes?

Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.

What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.



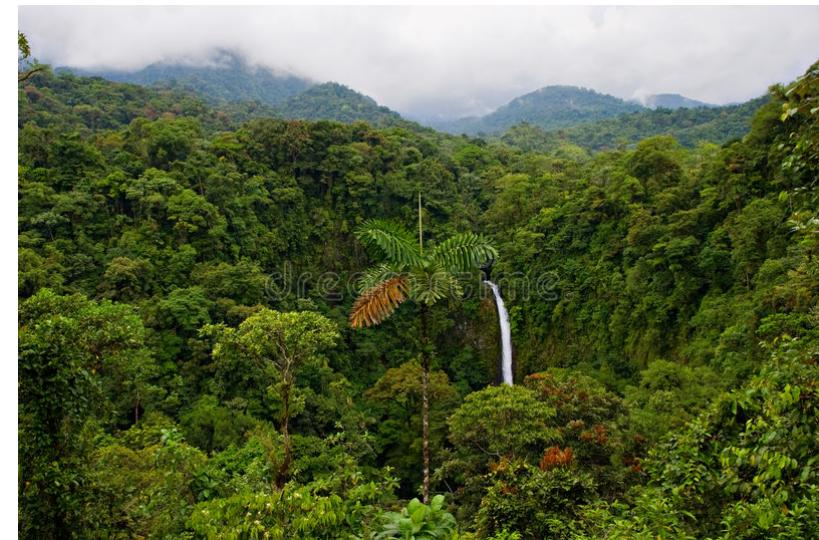
# Alpha Diversity

- Diversity in one location
  - Richness
  - Evenness
  - Phylogenetic diversity



# Richness

- The number of species



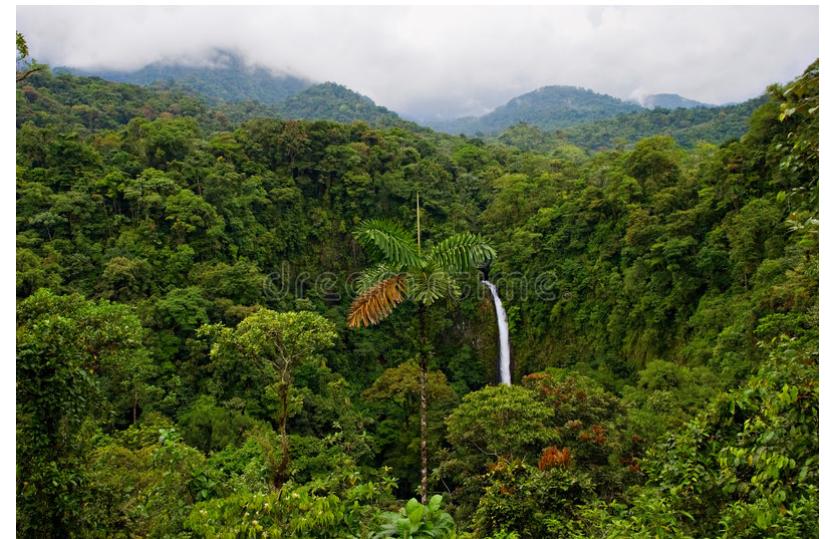
# Evenness

- Abundance distribution of species



# Phylogenetic diversity

- Breadth of dissimilarity between species that are present



# Alpha Diversity

- Number of different “species” in a location (richness)
- Number (abundance) of each of the “species” in that location
- BUT, in many ecosystems it is very difficult to identify and count every species present
- So an estimation must be made of the true number of species (richness). For example, Chao1

$$S_1 = S_{obs} + \frac{F_1^2}{2F_2}$$

where

$S_1$  estimated number of species,

$S_{obs}$  number of observed species,

$F_1$  number of singletons and

$F_2$  number of doubletons.

# e.g. Alpha Diversity Metrics

- Abundance-based Coverage Estimator (ACE) metric
- Berger-Parker Dominance Index
- Brillouin's index
- Chao1 confidence interval
- Chao1 index
- Dominance measure
- Effective Number of Species (ENS)/Probability of intra-or interspecific encounter (PIE) metric:  
Calculates Effective Number of Species (ENS)/Probability of intra-or interspecific encounter (PIE) metric
- Etsy confidence interval
- Faith's phylogenetic diversity
- Fisher's index
- Gini index
- Good's coverage of counts
- Heip's evenness measure
- Kempton-Taylor Q index
- Lladser's confidence interval
- Lladser's point estimate
- Margalef's richness index
- McIntosh dominance index D
- McIntosh evenness index E
- Menhinick's richness index
- Michaelis-Menten fit to rarefaction curve of observed OTUs:
- Number of distinct features: Calculates number of distinct OTUs
- Number of double occurrences: Calculates number of double occurrence OTUs (doubletons)
- Number of observed features, including singles and doubles: Calculates number of observed OTUs, singles, and doubles.
- Singles: Calculates number of single occurrence OTUs (singletons)
- Pielou's evenness
- Robbins' estimator
- Shannon's index
- Simpson evenness measure E
- Simpson's index
- Strong's dominance index (Dw)

# Equations

## Shannon

$$H' = - \sum_{i=1}^R p_i \ln p_i$$

where

$p_i$  species i relative abundance

R richness (observed species)

## Simpson

$$\lambda = \sum_{i=1}^R p_i^2$$

where

$p_i$  species i relative abundance

R richness (observed species)

## 16S amplicons

### Alpha diversity

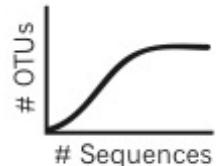
Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?

## Genomes

If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.

## metagenomics

Which microbiota are the most functionally diverse, with the most gene families?

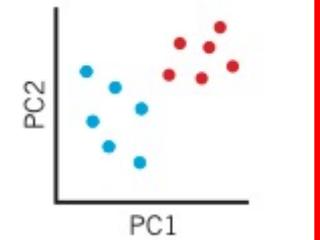


### Beta diversity

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?

Beta diversity plots on the basis of taxonomic and functional composition are usually highly correlated. Therefore, functional genes have considerable phylogenetic signal.

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of functional genes?



### Machine learning or classical statistics

Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?

Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.

Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?

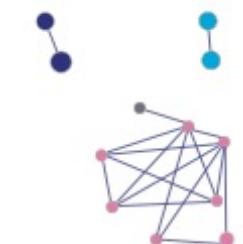


### Co-occurrence

Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.

Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.

Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.

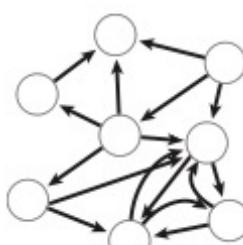


### Metabolic network modelling

What metabolic activities are predicted for a collection of phylotypes?

Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.

What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.



# Beta diversity: between habitat diversity

- 3 main methods of calculating  $\beta$  diversity
  1. “True  $\beta$  diversity” or species turn over. The relationship between  $\beta = \gamma / \alpha$ .
  2. Similarity or distance measures. Could include phylogenetic distance
  3. Species-area relationship metric related to species turn over by area
- From my experience, distance metrics are the only ones used in microbial ecology.
  - Presence/absence
  - Difference in relative abundance
  - Difference in phylogenetic overlap

# Beta diversity

- Qualitative
- Quantitative
- Phylogenetic diversity



# Beta diversity: between habitat diversity

- Bray-Curtis and Jaccard use presence/absence and abundance
- Sorenson is presence/absence only
- All are commonly used, but BC the most.

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$$

where

$S_i$  # of individuals in sample  $i$

$S_j$  # of individuals in sample  $j$

$C_{ij}$  is the sum of the lower value  
between each sample for all  
species

$$S_{ij} = 1 - \frac{2a}{2a + b + c}$$

where

$a$  # of species in both samples

$b$  # of species only in sample  $i$

$c$  # of species only in sample  $j$

Many other methods for beta diversity

## 16S amplicons

### Alpha diversity

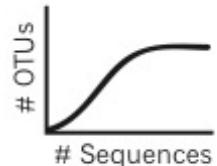
Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?

## Genomes

If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.

## metagenomics

Which microbiota are the most functionally diverse, with the most gene families?

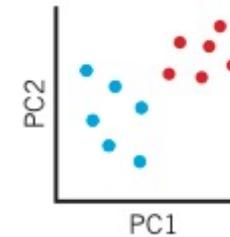


### Beta diversity

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?

Beta diversity plots on the basis of taxonomic and functional composition are usually highly correlated. Therefore, functional genes have considerable phylogenetic signal.

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of functional genes?



### Machine learning or classical statistics

Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?

Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.

Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?



### Co-occurrence

Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.

Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.

Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.



### Metabolic network modelling

What metabolic activities are predicted for a collection of phylotypes?

Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.

What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.



# Differential abundance

Many methods...

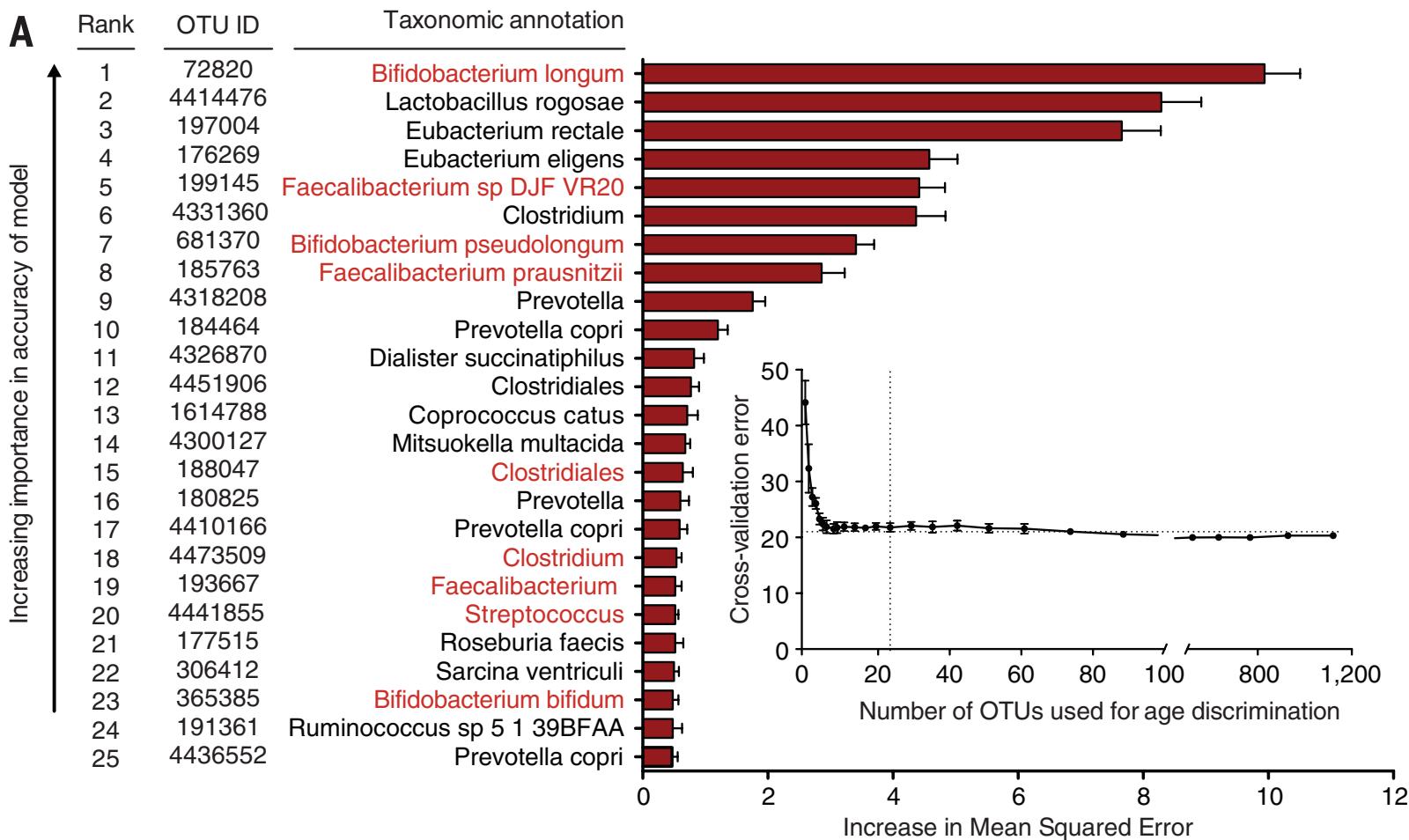
no settled best approach

1. Traditional stats
2. Metastats
3. DESeq2
4. Random Forests
5. LefSe
6. ANCOM
7. SIMPER

# Differential abundance

Many methods...  
no settled best approach

1. Traditional stats
2. Metastats
3. DESeq2
4. **Random Forests**
5. LefSe
6. ANCOM
7. SIMPER



## 16S amplicons

### Alpha diversity

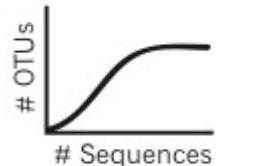
Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?

## Genomes

If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.

## metagenomics

Which microbiota are the most functionally diverse, with the most gene families?

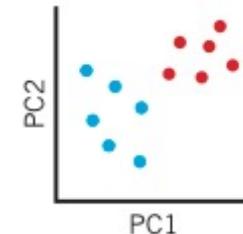


### Beta diversity

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?

Beta diversity plots on the basis of taxonomic and functional composition are usually highly correlated. Therefore, functional genes have considerable phylogenetic signal.

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of functional genes?



### Machine learning or classical statistics

Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?

Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.

Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?

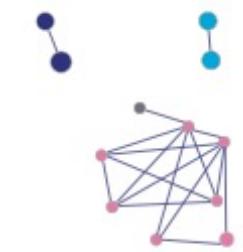


### Co-occurrence

Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.

Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.

Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.



### Metabolic network modelling

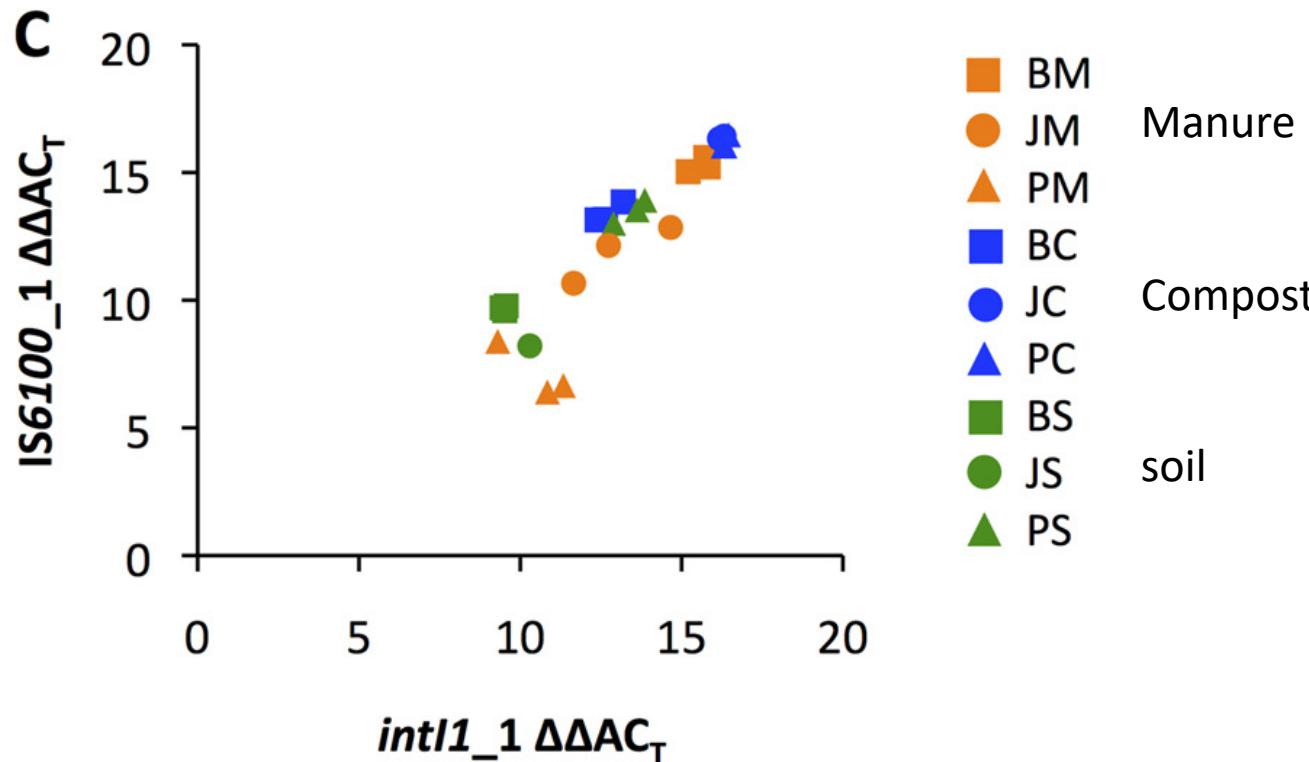
What metabolic activities are predicted for a collection of phylotypes?

Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.

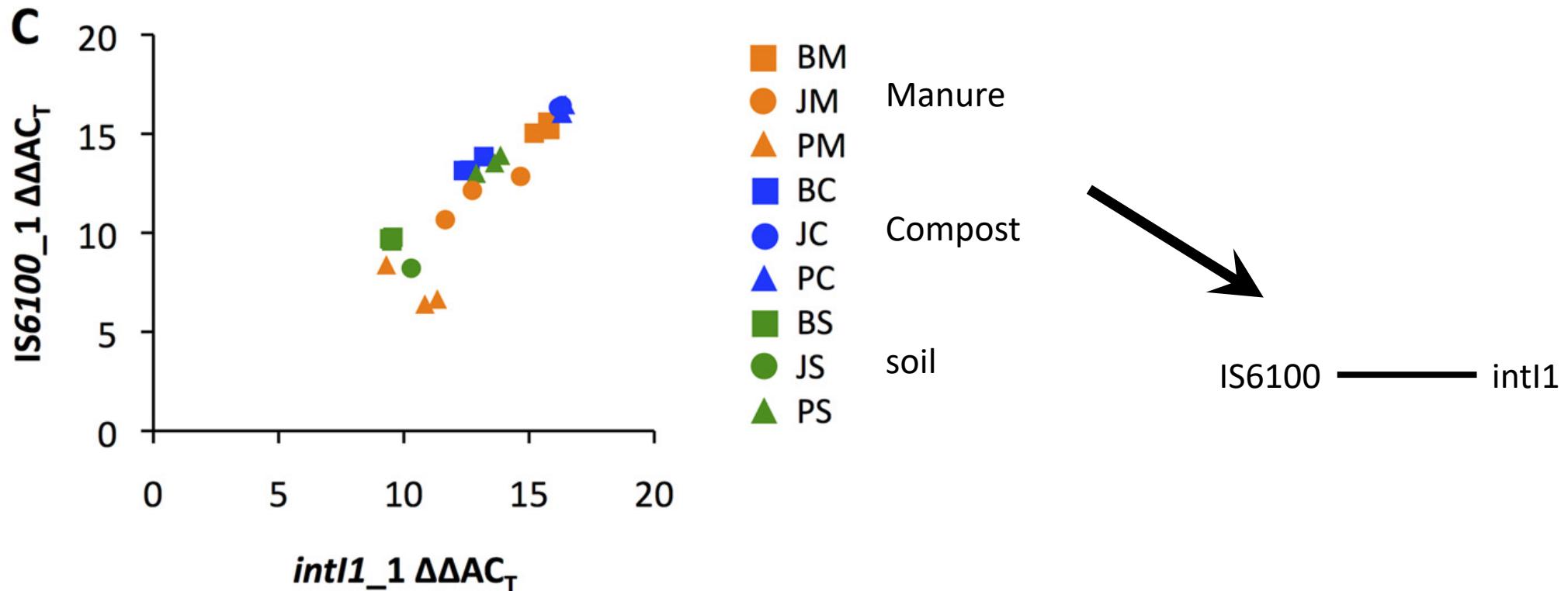
What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.



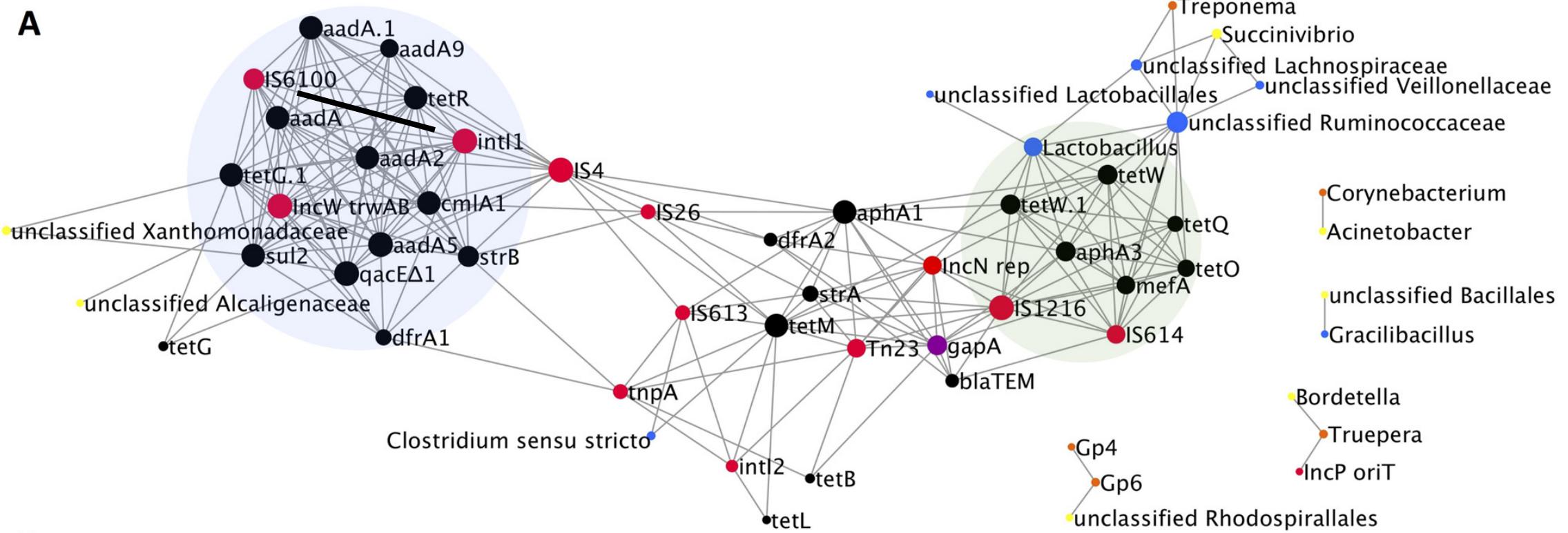
# Co-occurrence: who cohabitates?



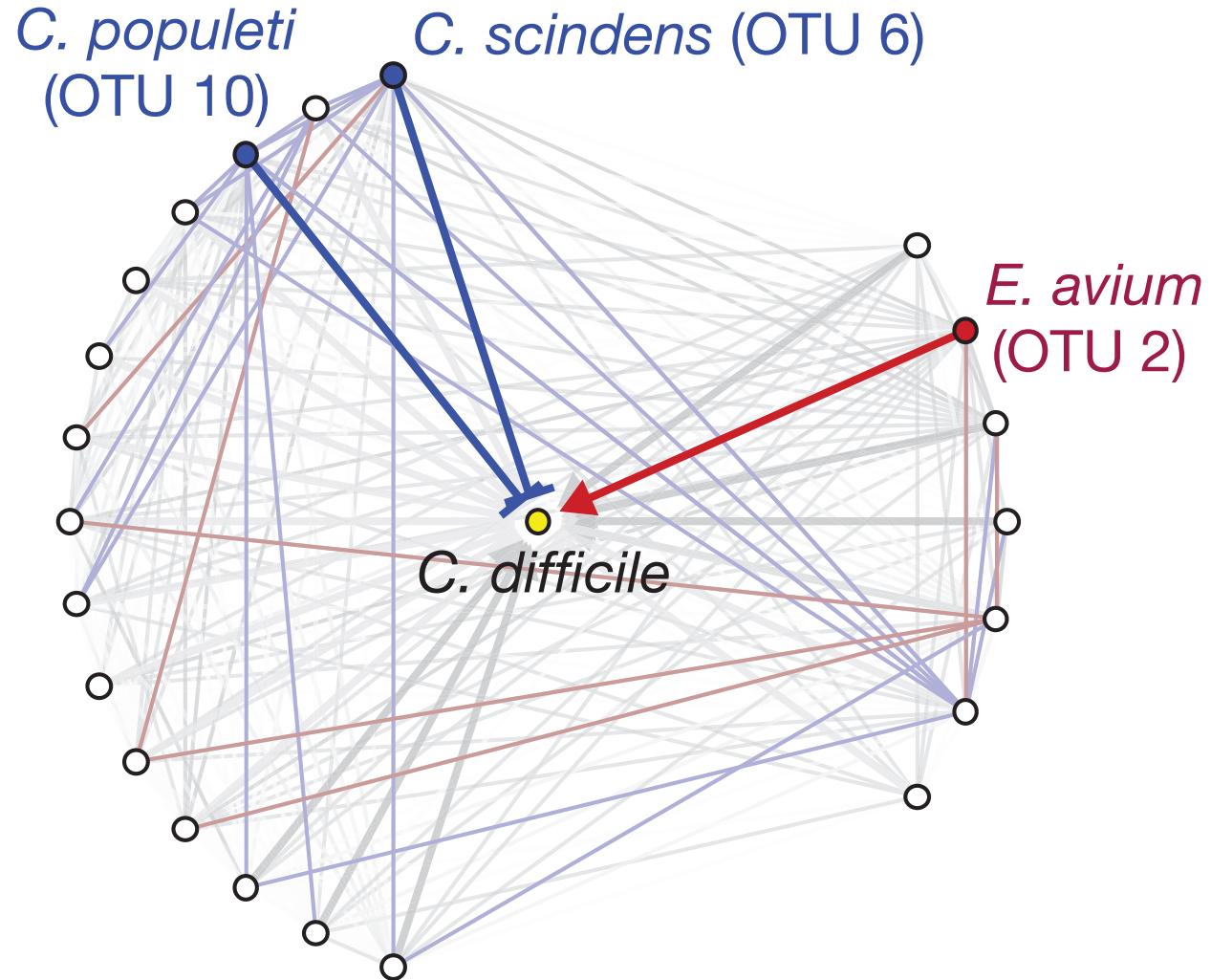
# Co-occurrence: who cohabitates?



# Co-occurrence



# Co-occurrence



Buffie, Nature, 2014

## 16S amplicons

### Alpha diversity

Which microbiota are the most compositionally diverse, with the most phylotypes or longest branch length?

## Genomes

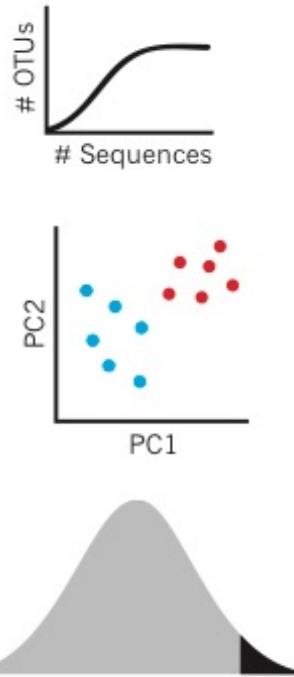
If additional taxa (compositional diversity) do not increase the number of functions (functional diversity), there is functional redundancy.

### Beta diversity

Which factors, such as age, culture or disease state, correlate with differences in the overall composition of phylotypes or phylogenetic lineages?

## metagenomics

Which microbiota are the most functionally diverse, with the most gene families?

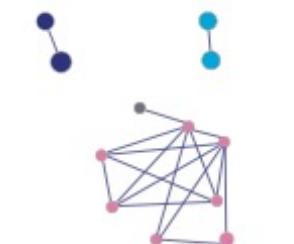


### Machine learning or classical statistics

Are there phylotypes or collections of phylotypes that discriminate or significantly differ between the microbiota of individuals in different groups – diets or disease states?

Functional differences are often reflective of taxonomic differences. Selected functions that cannot be explained by selected taxonomic groups indicate potential functional convergence.

Are there particular gene families or functions that discriminate or significantly differ between the microbiomes of individuals in different groups – diets or disease states?

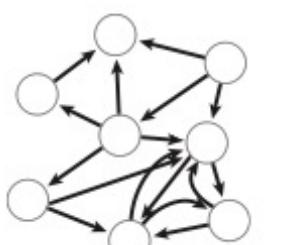


### Co-occurrence

Which phylotypes co-distribute across people? Positive correlations are driven by shared environmental preferences, symbiotic or syntrophic relationships. Negative correlations are driven by divergent environmental preferences or competitive relationships.

Performing co-occurrence analyses on species with complete genomes can allow for further exploration of the driving factors of positive or negative correlations. For example, genes or functions that are shared by phylogenetically unrelated bacteria that co-occur can indicate potential environmental driving factors.

Which genes co-distribute across people? Positive correlations will largely be driven by genes that co-occur in the same genome (phylotype). Genes with high rates of horizontal gene transfer and strong environmental selection may co-distribute independently.



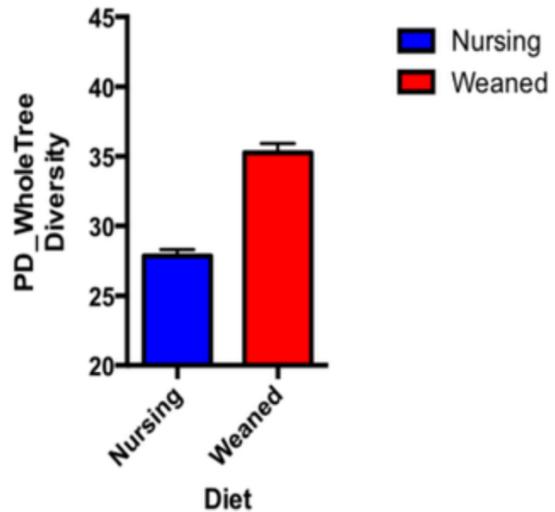
### Metabolic network modelling

What metabolic activities are predicted for a collection of phylotypes?

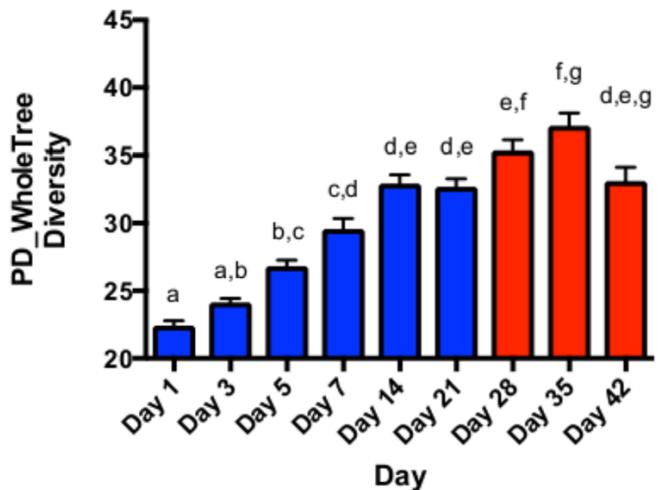
Metabolic predictions can, in principle, be made for collections of phylotypes if associated reference genome sequences are available.

What overall metabolic activities are predicted for microbiota? What spectrum of metabolites would microbiota produce from a defined diet? Promising for designing personalized therapeutic diets, but has challenges.

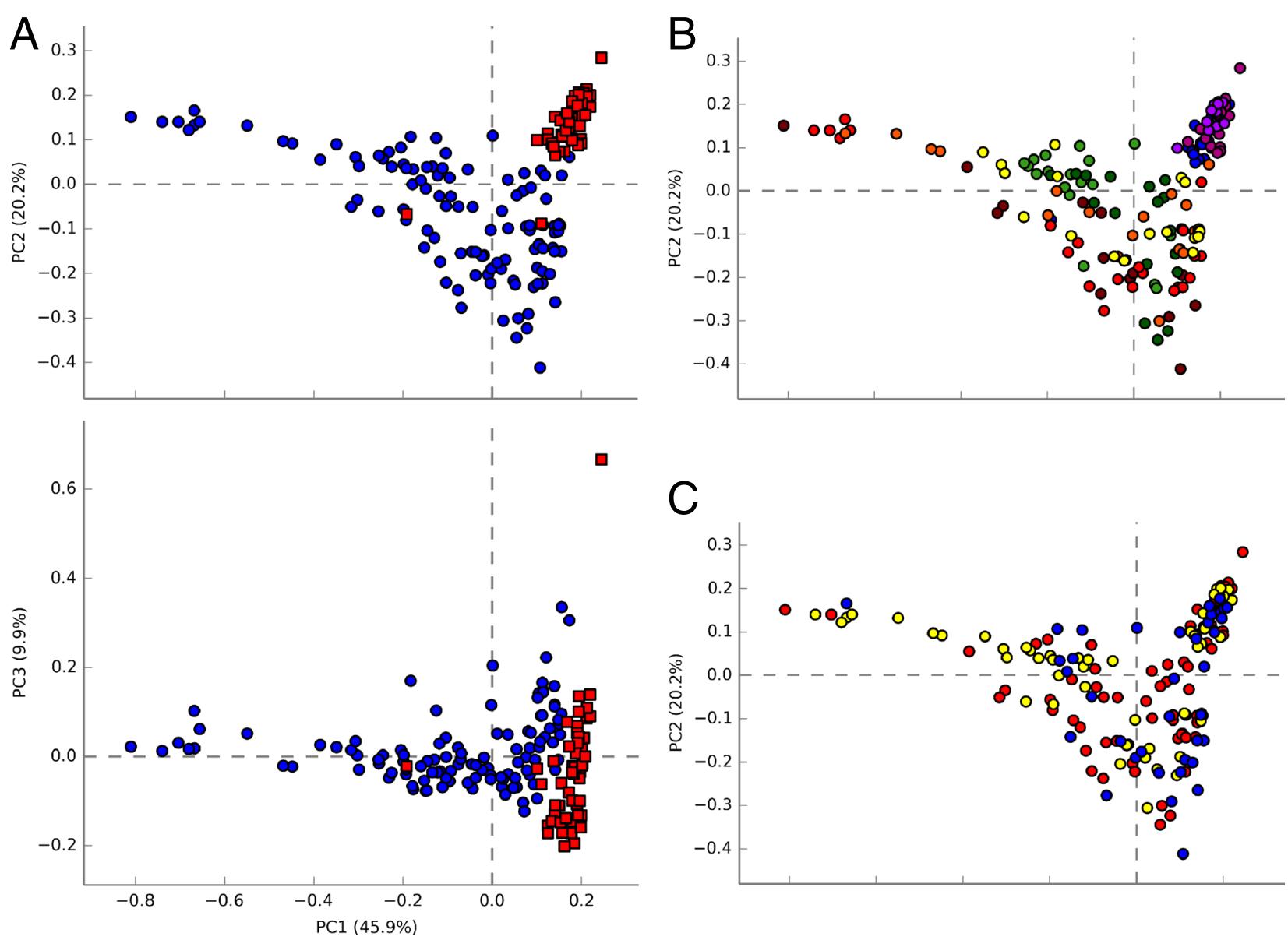
### A Alpha Diversity Compared Between Diets



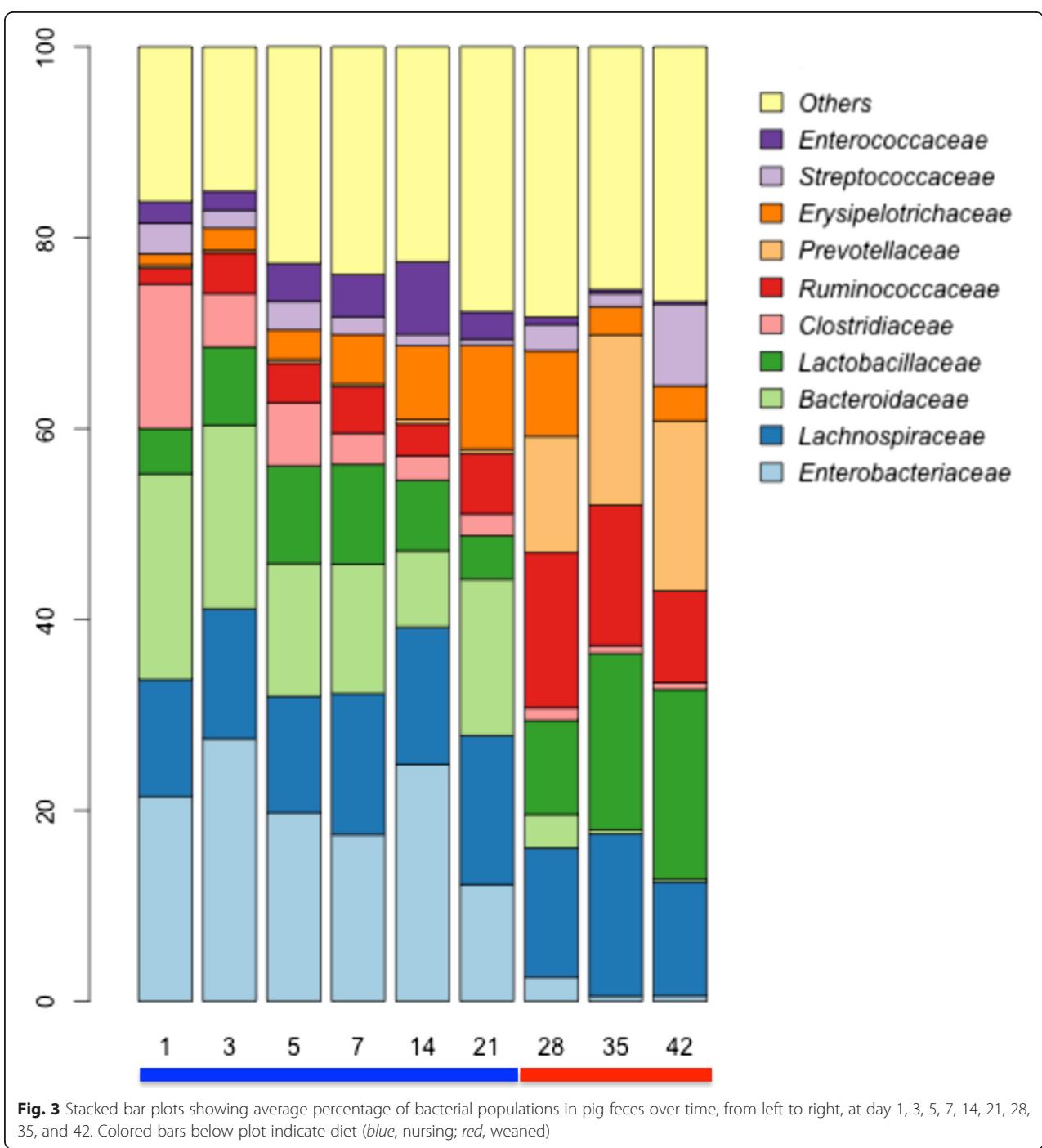
### B Alpha Diversity Compared by Sampling Day



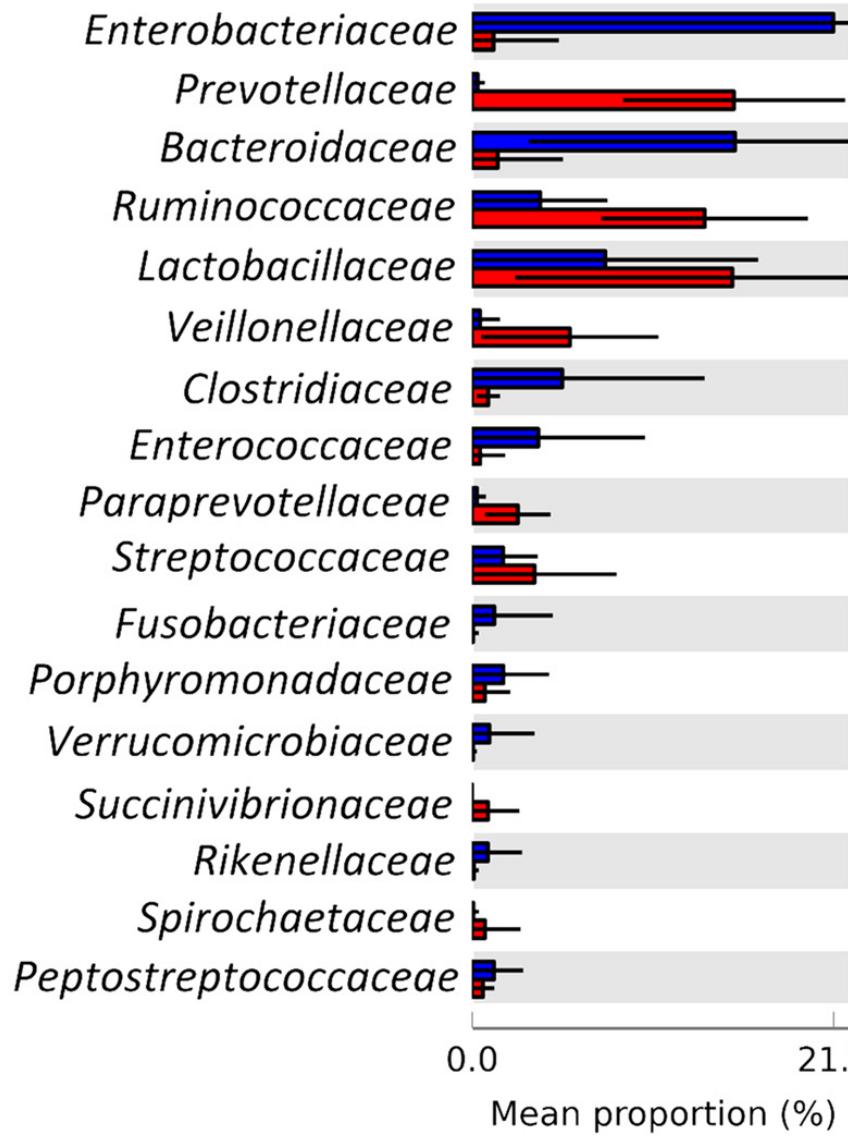
**Fig. 1** Pig fecal alpha diversity (phylogenetic distance, PD), measured in rarified samples, compared between diets (a) and over time (b). Bars are colored by diet, nursing (blue) or weaned (red). Bars are shown  $\pm$  SEM



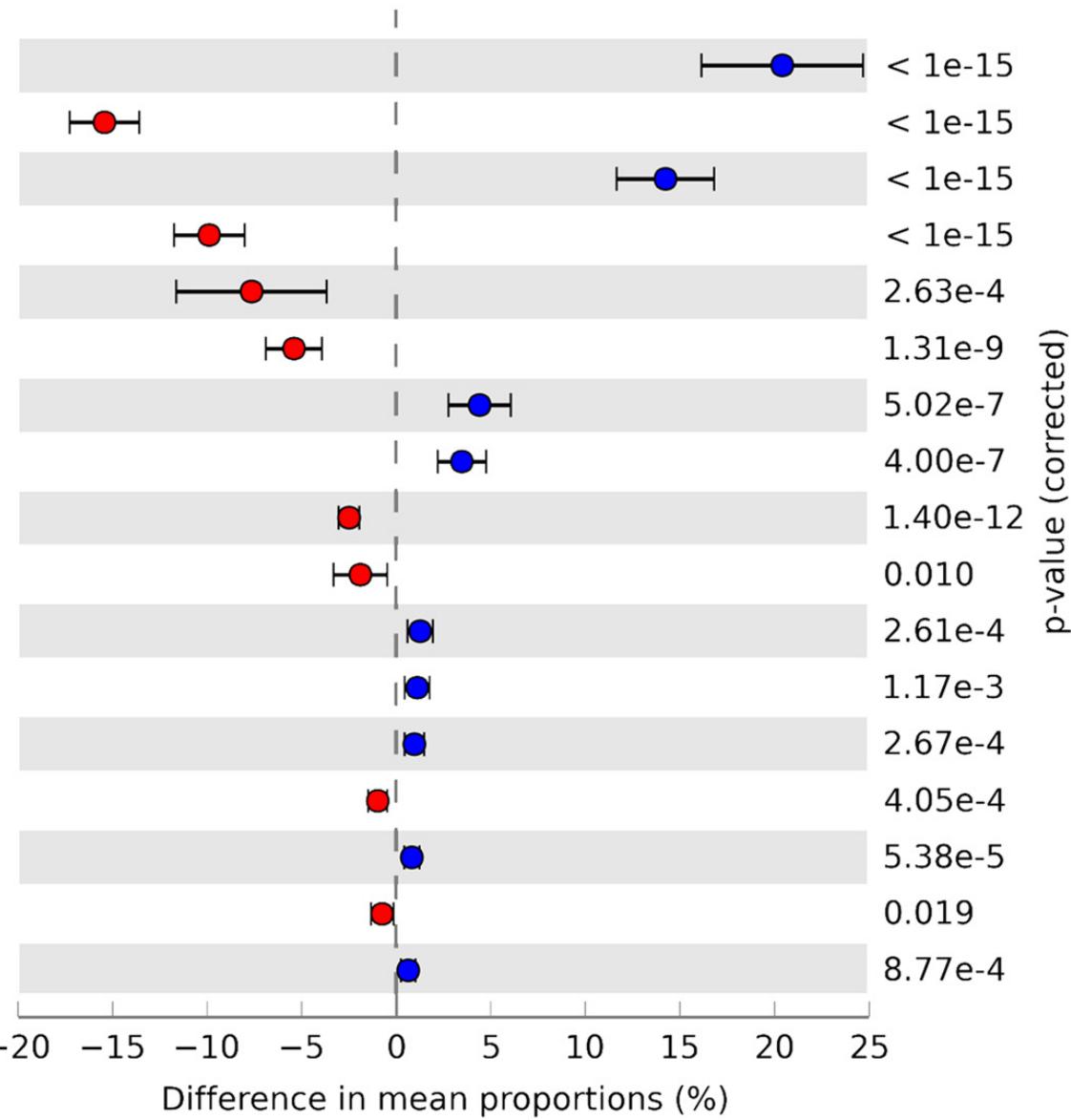
**Fig. 2** Principal component analysis plots of pig fecal microbiota colored by diet (**a** nursing, blue; weaned, red), day (**b** day 1, dark red; day 3, red; day 5, orange; day 7, yellow; day 14, green; day 21, dark green; day 28, blue; day 35, dark blue; day 42, purple), or litter (**c** litter 1, red; litter 2, yellow; litter 3, blue)



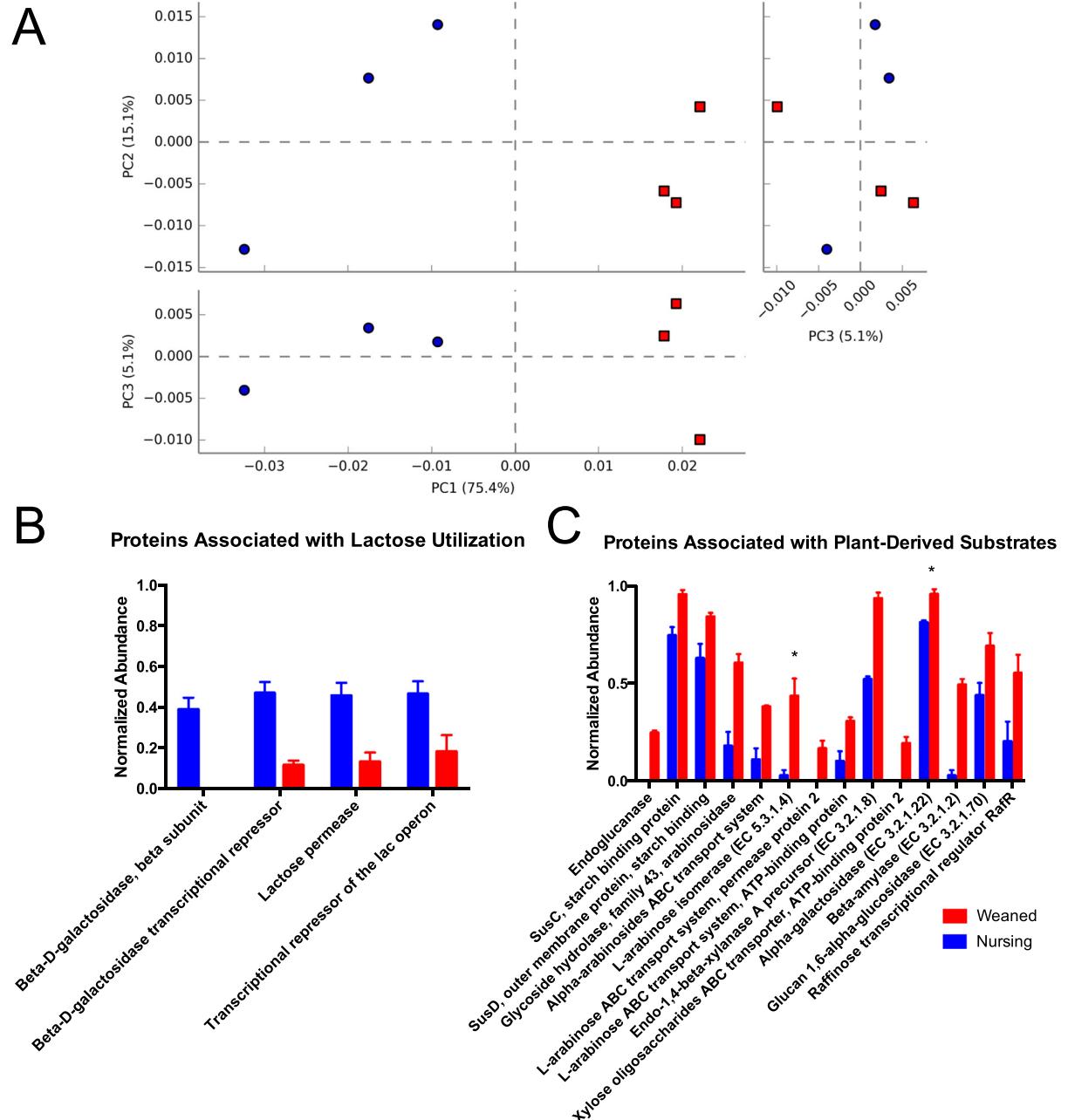
█ nursing    █ weaned



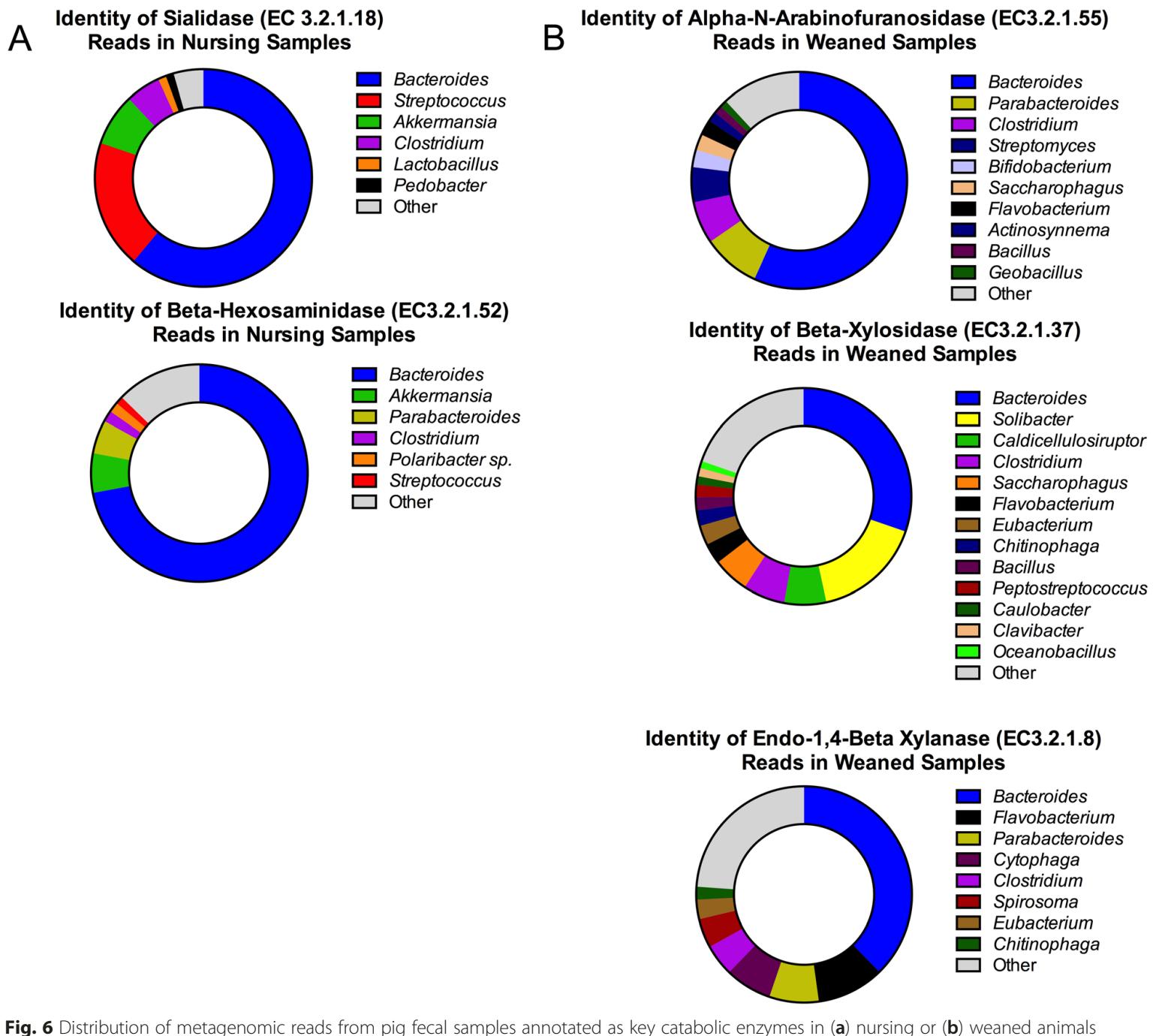
95% confidence intervals



**Fig. 4** Extended error bar plot identifying significant differences between mean proportions of bacterial taxa in nursing (blue) and weaned (red) samples. Corrected p values are shown at right



**Fig. 5** Metagenomic sequencing of pig fecal samples analyzed using PCA (a), and bar charts comparing normalized abundances of individual predicted genes involved in lactose utilization (b) and plant-derived glycans (c). All differences between diets shown (blue, nursing; red, weaned) were significant ( $p < 0.01$ ), unless indicated with an asterisk, where  $p < 0.05$



**Fig. 6** Distribution of metagenomic reads from pig fecal samples annotated as key catabolic enzymes in (a) nursing or (b) weaned animals

# Today's topics?

1. The vastness of microbes and their complexity
2. Describing microbial communities
- 3. Microbial community succession**

# Community Succession: is colonization random?



## Succession of microbial consortia in the developing infant gut microbiome

Jeremy E. Koenig<sup>a</sup>, Aymé Spor<sup>a</sup>, Nicholas Scalfone<sup>a</sup>, Ashwana D. Fricker<sup>a</sup>, Jesse Stombaugh<sup>b</sup>, Rob Knight<sup>b,c</sup>, Largus T. Angenent<sup>d</sup>, and Ruth E. Ley<sup>a,1</sup>

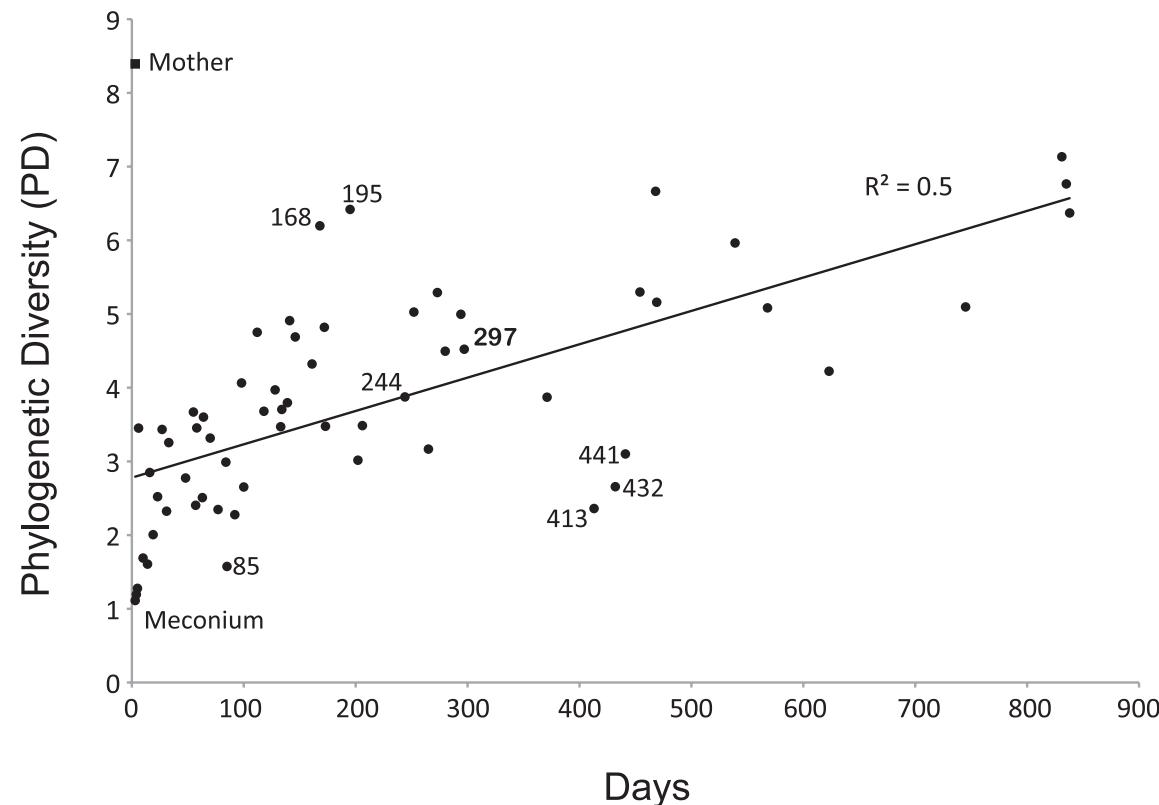
<sup>a</sup>Department of Microbiology, Cornell University, Ithaca, NY 14853; <sup>b</sup>Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309;

<sup>c</sup>Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309; and <sup>d</sup>Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY 14850

Edited by Todd R. Klaenhammer, North Carolina State University, Raleigh, NC, and approved June 24, 2010 (received for review March 2, 2010)

- 16S rRNA sequencing and shotgun sequencing
- Tracked a single child

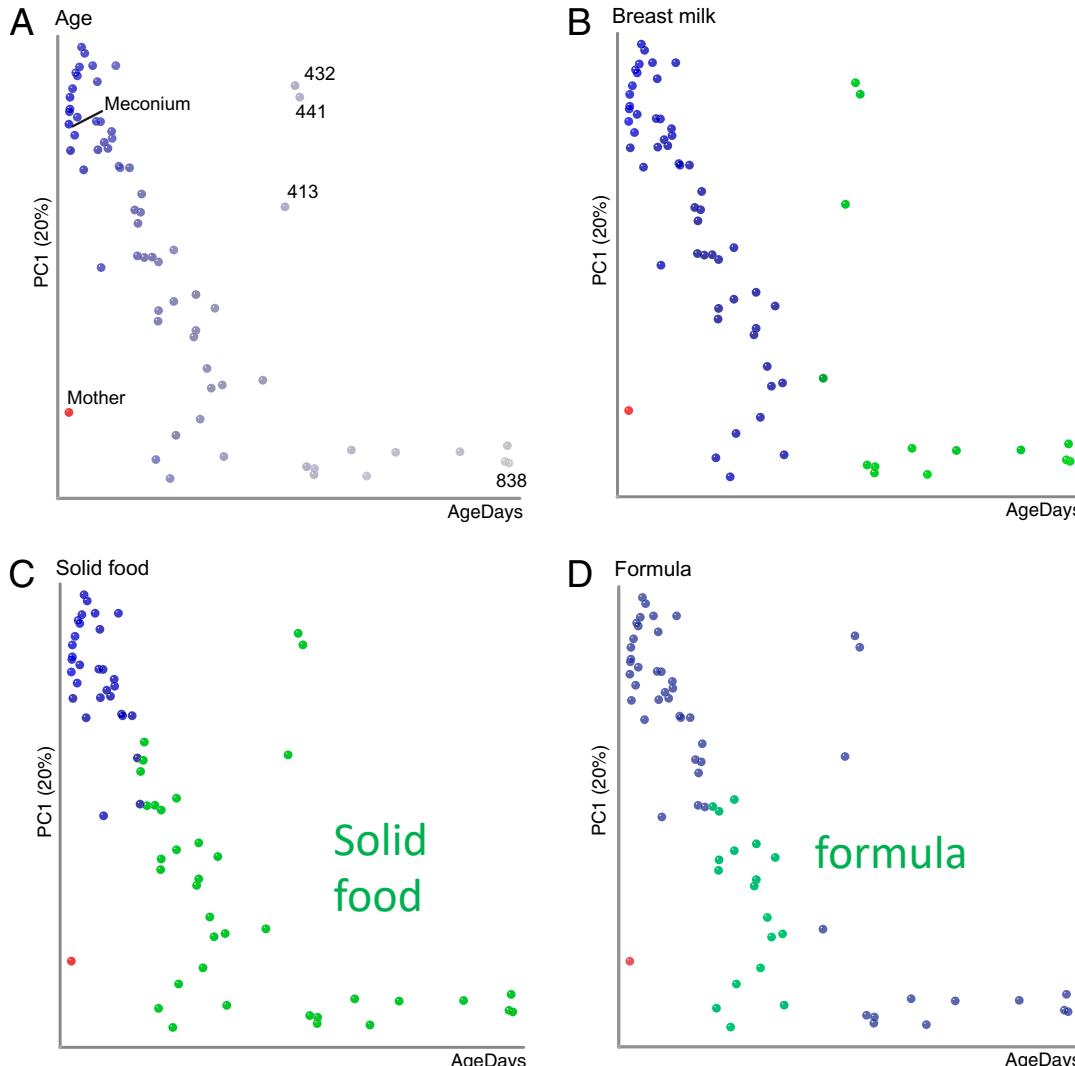
# Phylogenetic diversity increased with time



**Fig. 1.** Bacterial PD of the infant gut microbiota over time. PD provides a measure of the diversity within a community based on the extent of the 16S rRNA phylogenetic tree that is represented by that community. Symbols are fecal samples. The mother's fecal sample, collected at day 3, is denoted as a filled square.

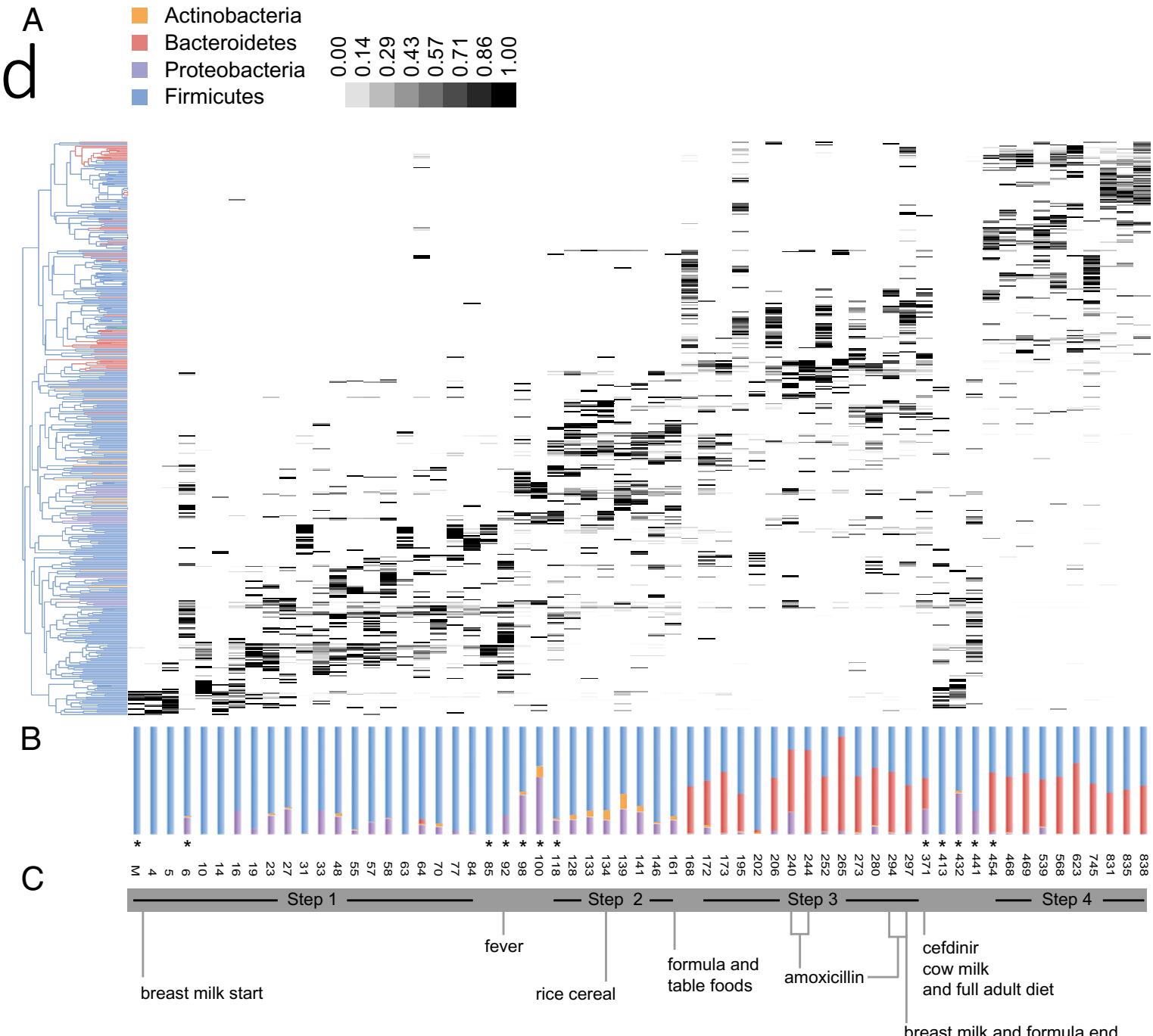
# Non-random progression through time

Only PC1 is shown  
Time is on x-axis

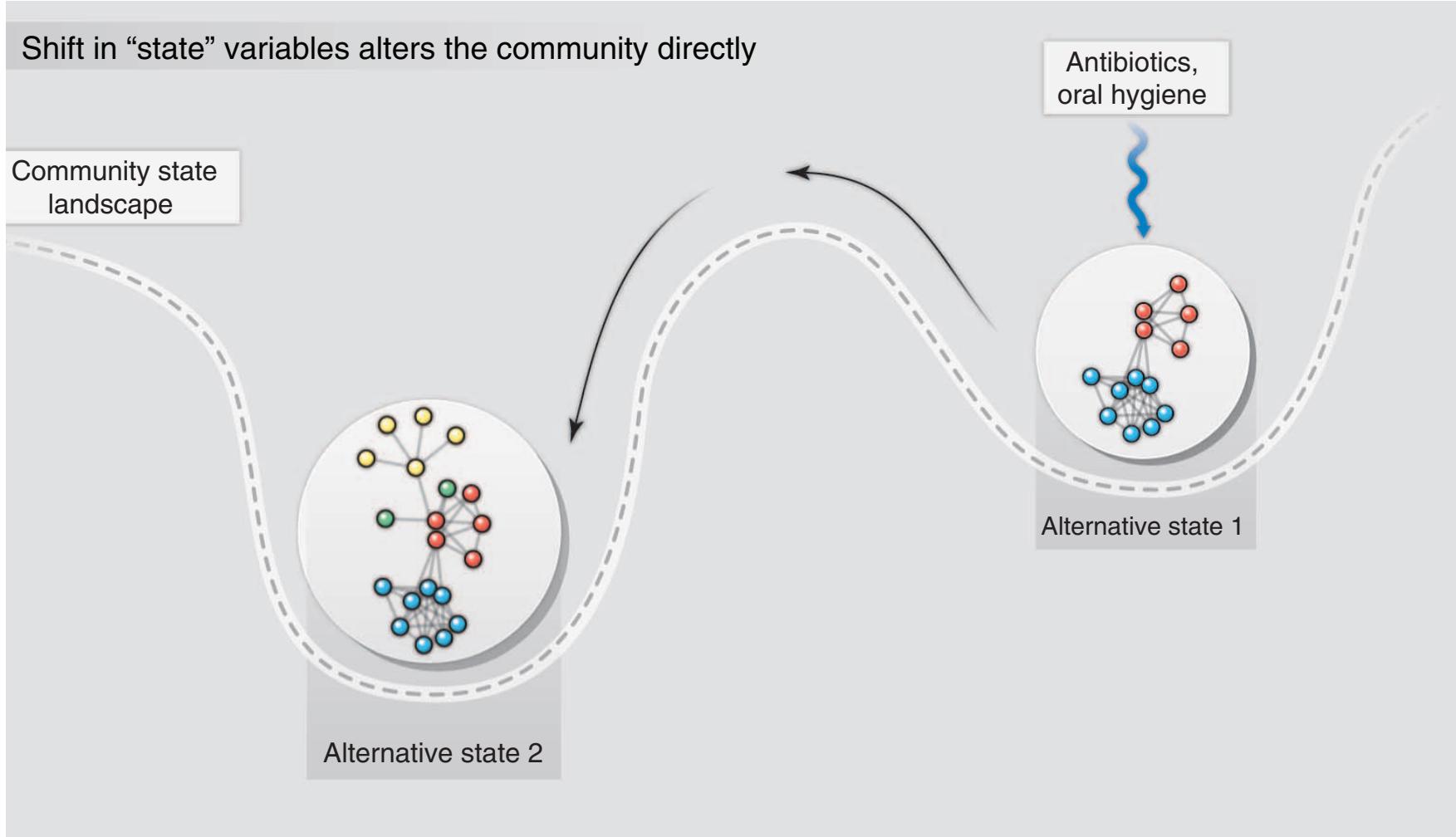


# Randomness explained by life events

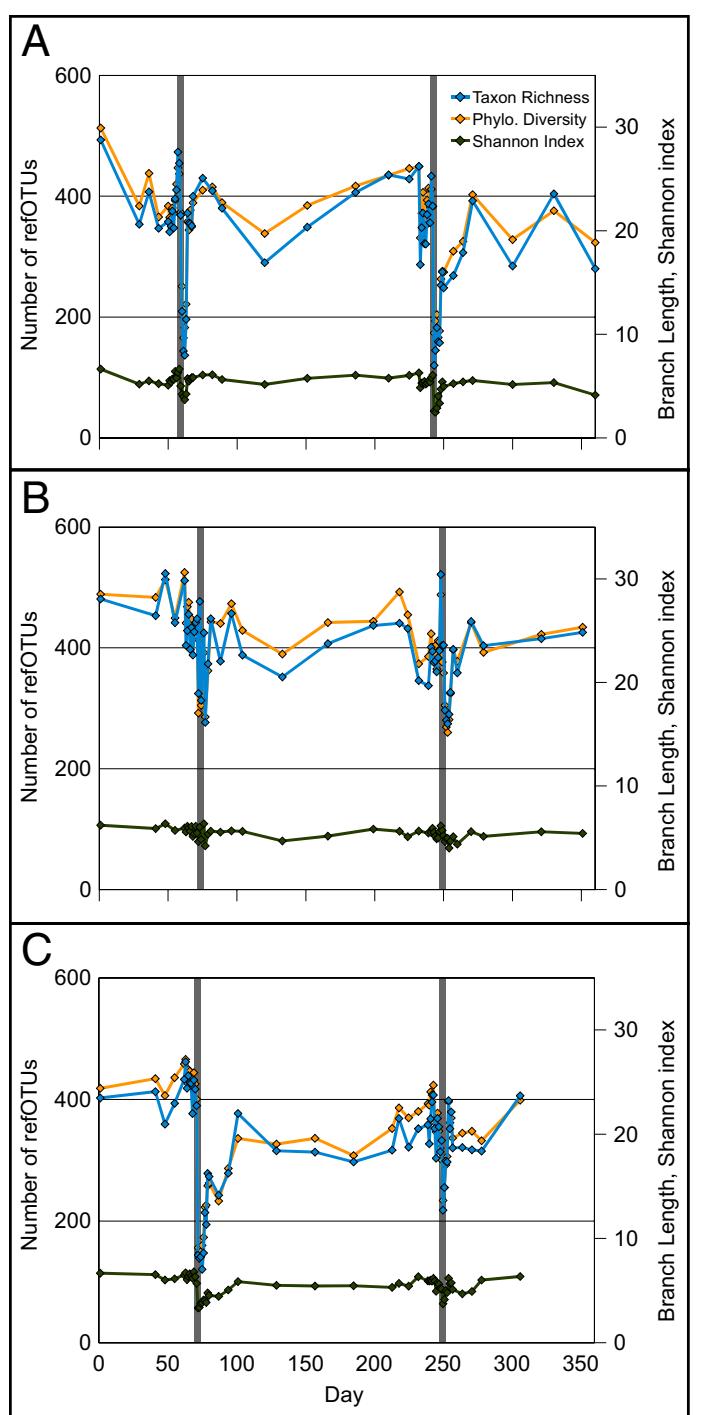
Antibiotics decrease PD in two of three treatments



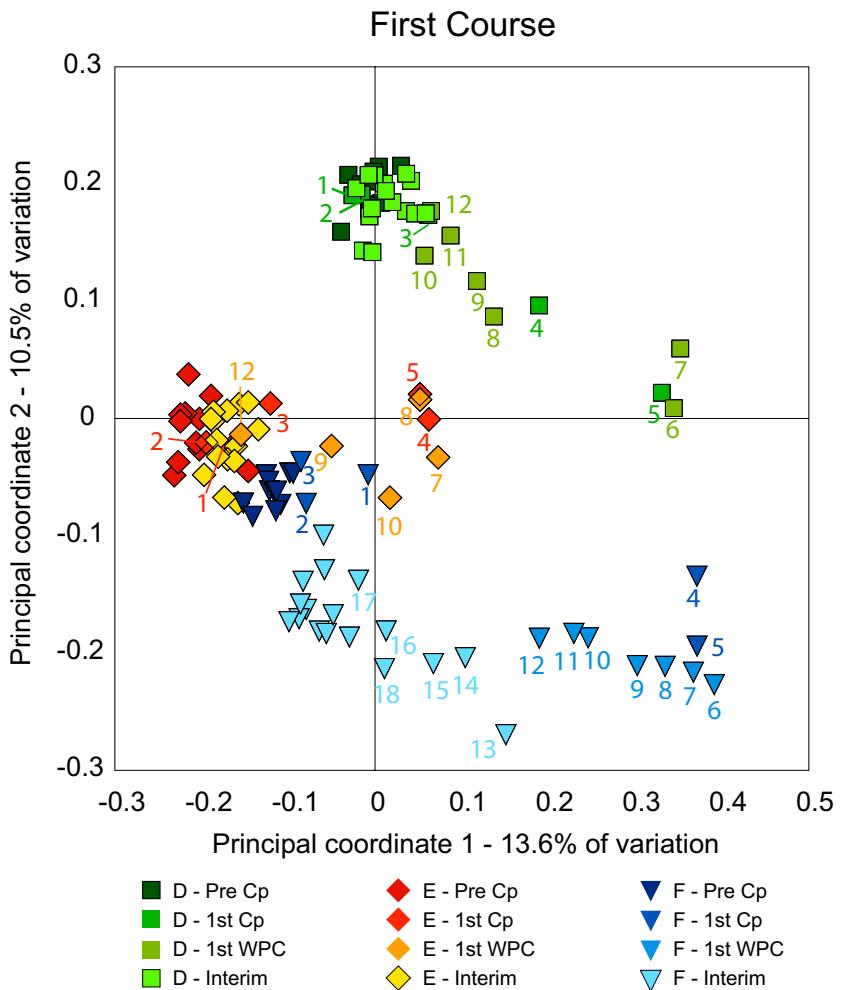
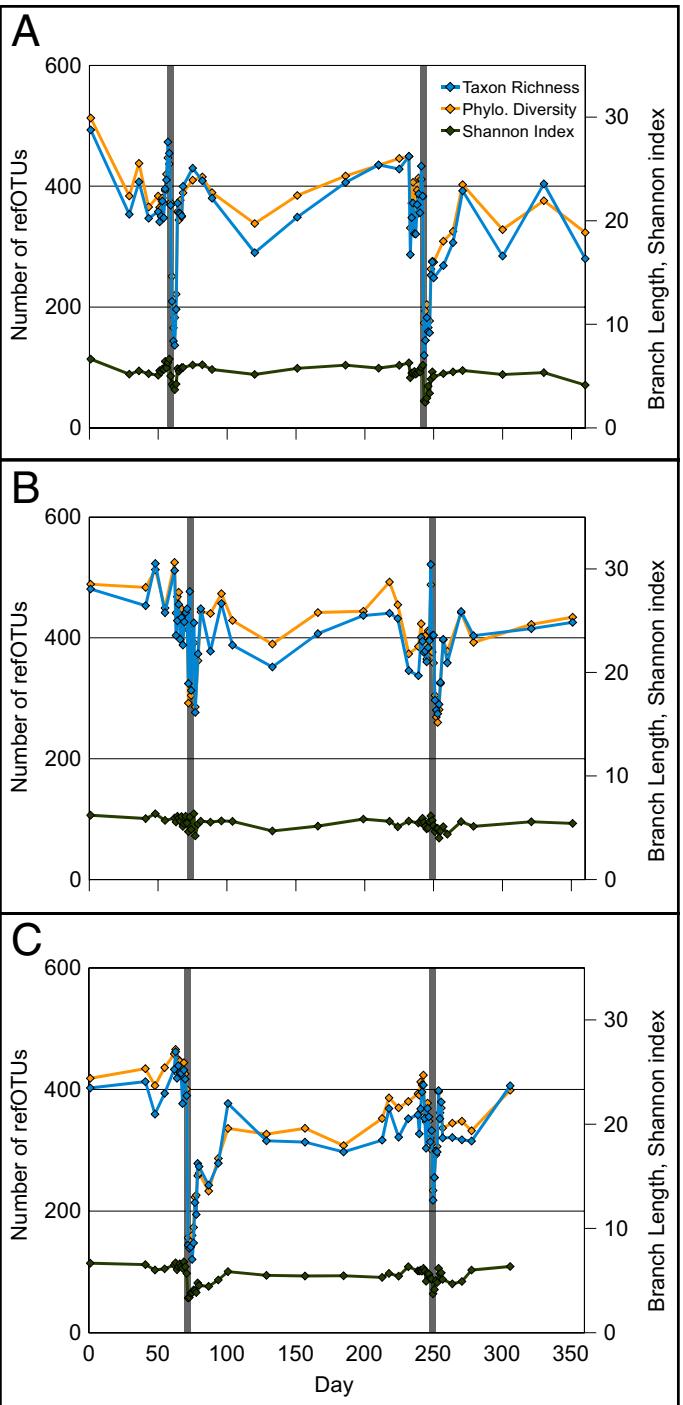
# Succession: Recolonization after a disturbance



# Recolonization after a disturbance

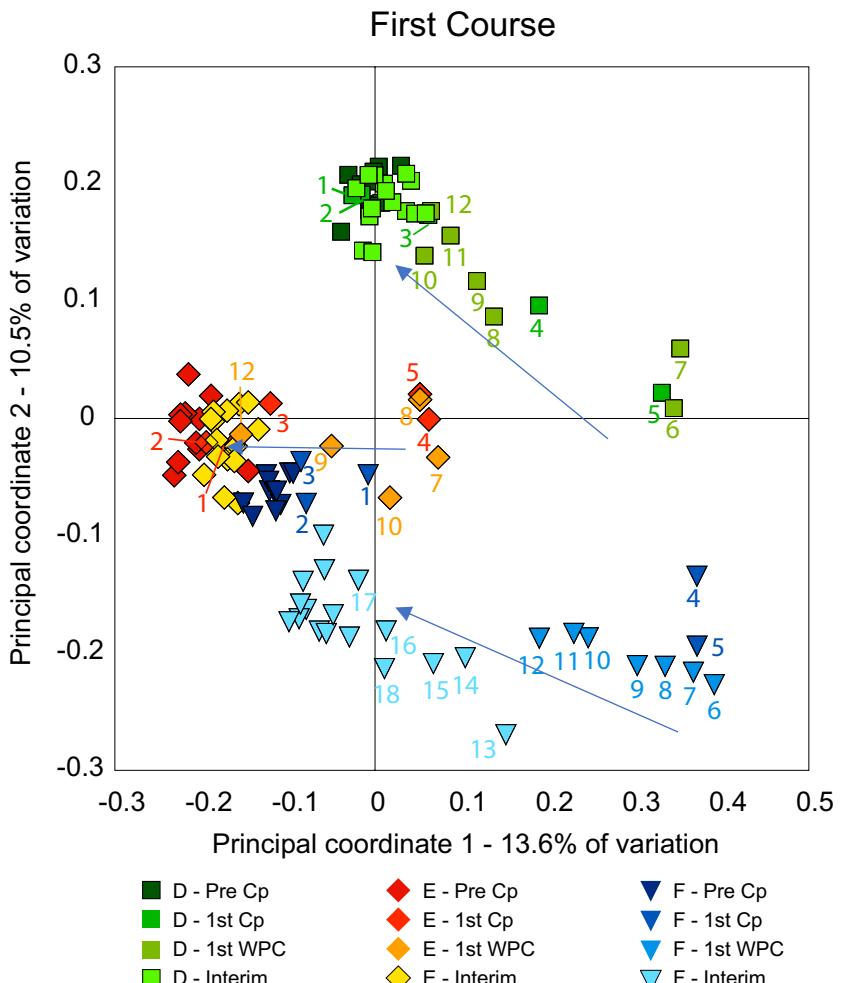
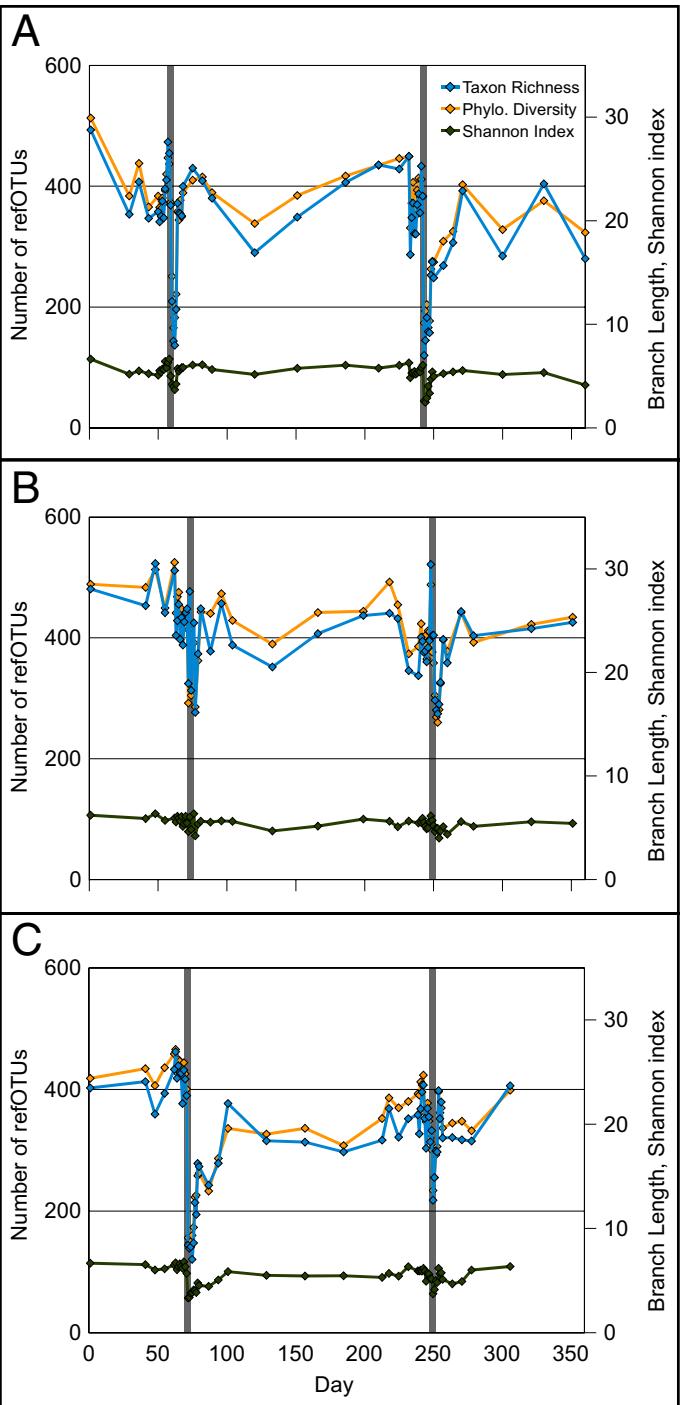


# Recolonization after a disturbance



Dethlefsen, PNAS, 2011

# Recolonization after a disturbance



Dethlefsen, PNAS, 2011

# What are your take away messages?

1. The vastness of microbes and their complexity
2. Describing microbial communities
3. Microbial community succession