Fine-tuning the association between pediatric arthritis and *Faecalibacterium prausnitzii*

Matthew Stoll MD

Wonderful World of Technology

March 23, 2017

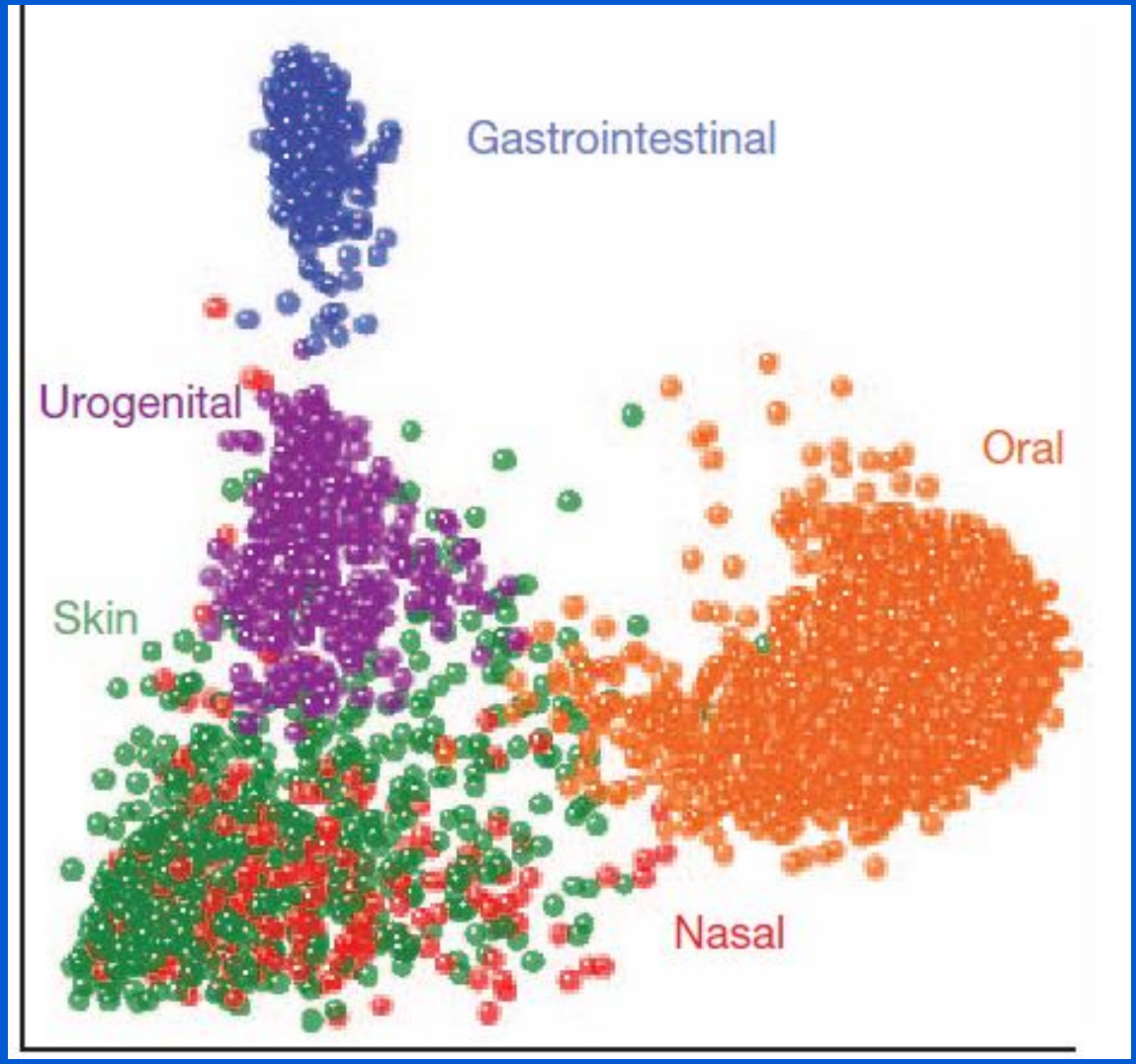
Microbiota and human diseases

- Inflammatory bowel disease
- Arthritis
  - Juvenile idiopathic arthritis
  - Rheumatoid arthritis
  - Spondyloarthritis
- Type I diabetes
- Obesity / metabolic syndrome

Outline

- **Where to sample**
- **Sequencing**
- **Data analysis**
- **Next steps**

Bacteria composition differs by site

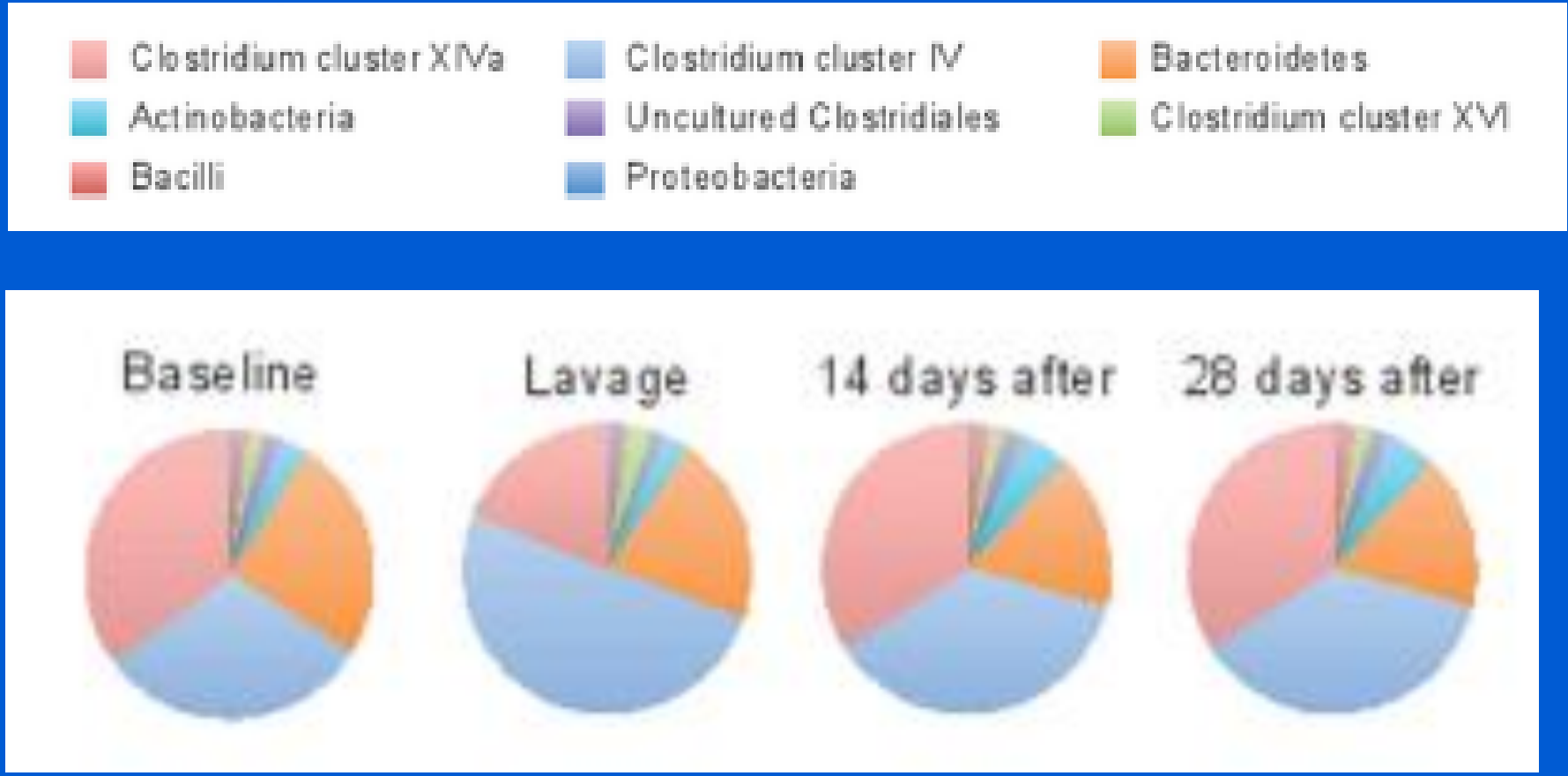


HMP Consortium, *Nature* 2012;486:207

Heterogeneity within habitats

- Gut
- Skin
- Mouth

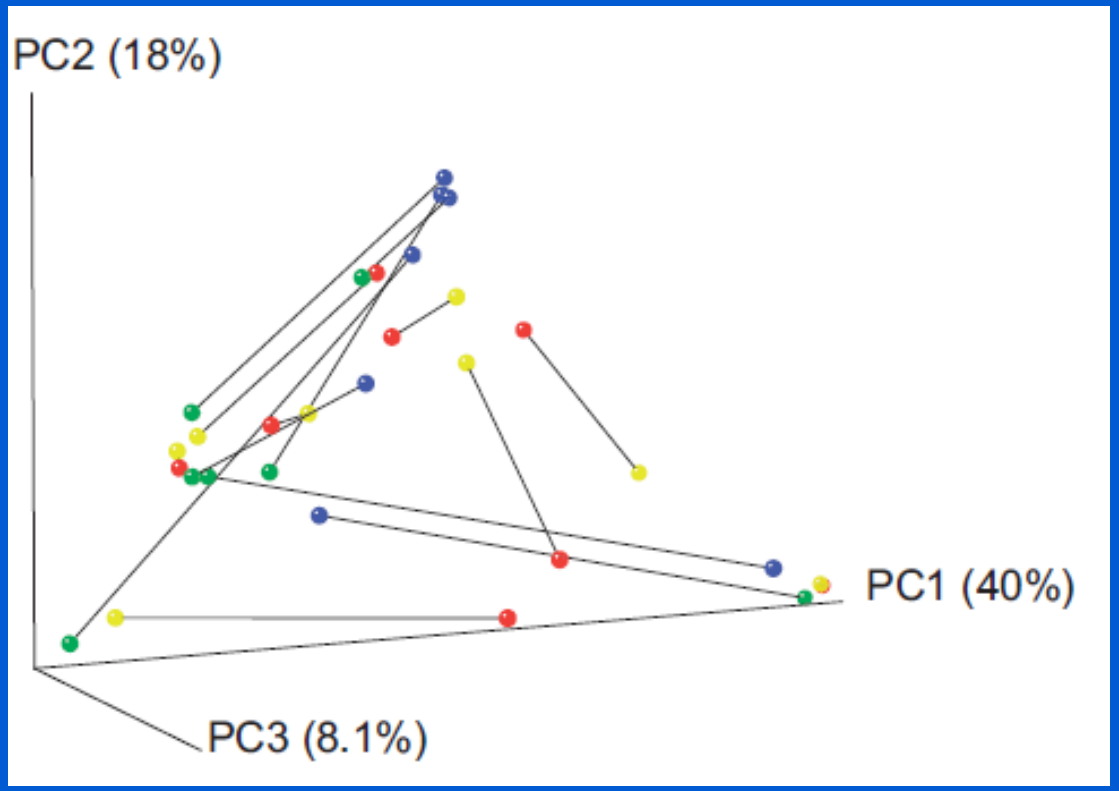
Bowel prep affects fecal microbiota



Jalanka, *Gut* 2015;64:1562

And the mucosal microbiota

Sigmoid biopsies before and after bowel preparation
Patients with IBD

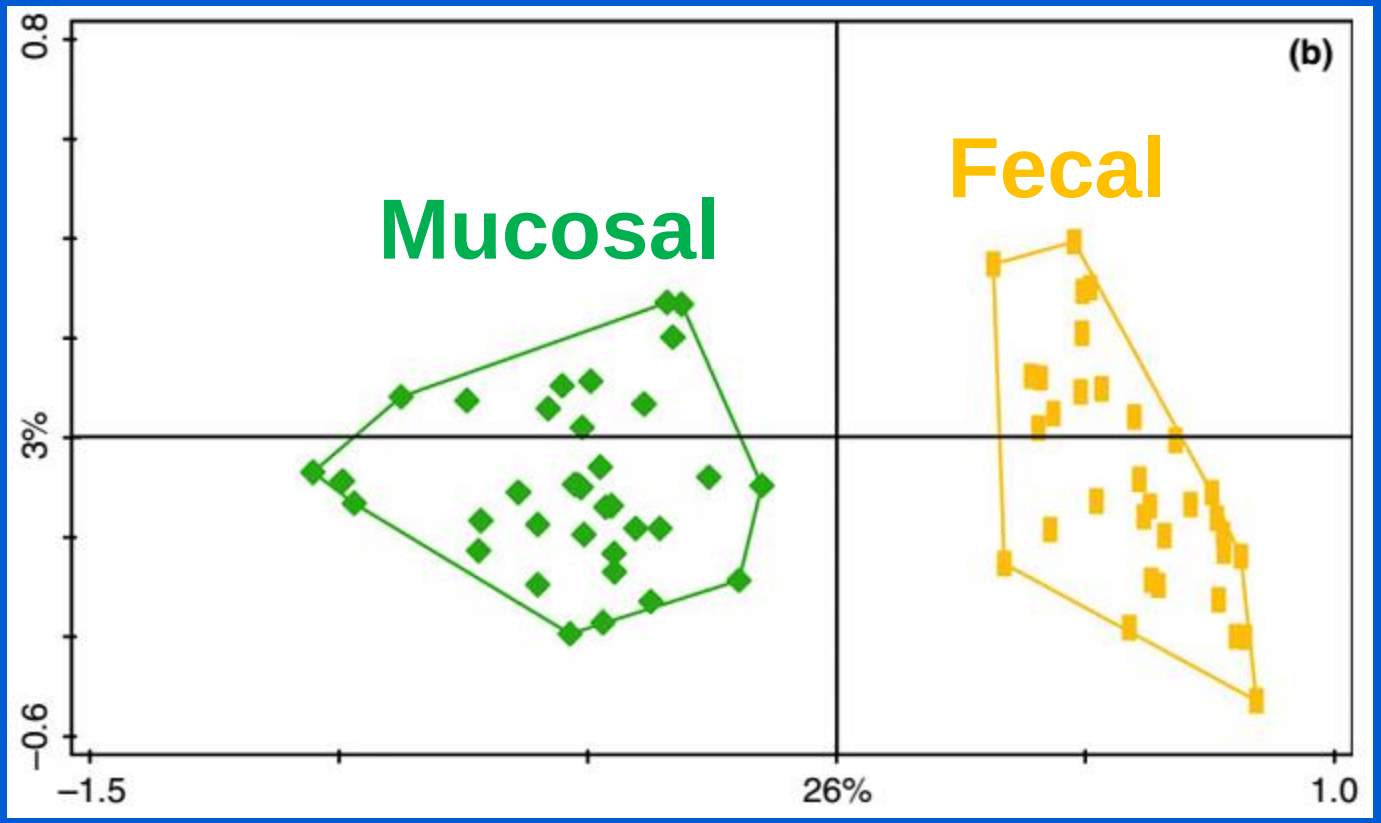


Red-yellow: Biopsy samples
Blue-green: Fecal samples

Shobar, *Clin and Transl Gastroenterol* 2016;7:e143

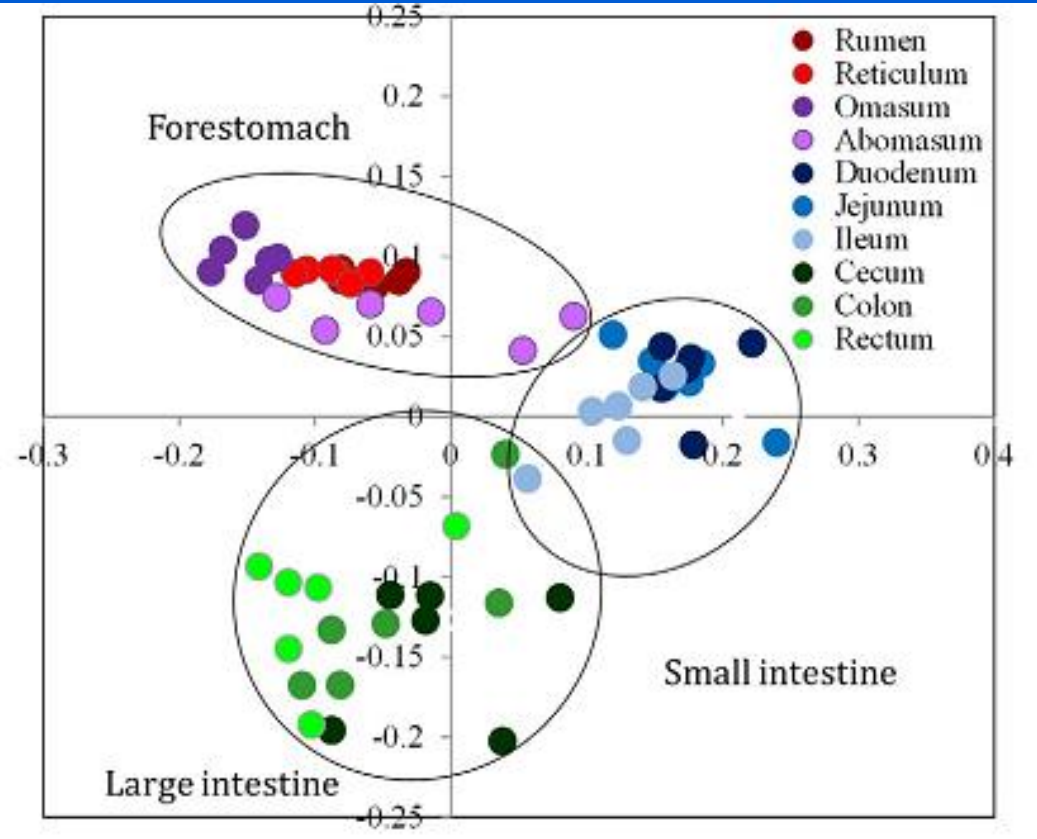
Fecal vs mucosal microbiota

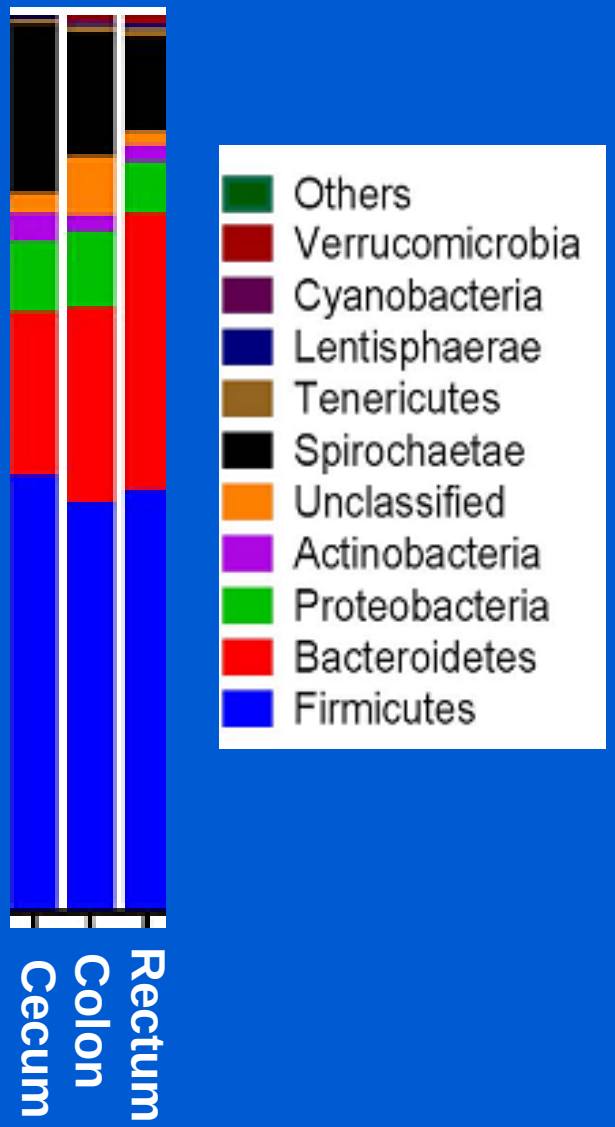
Sigmoid biopsies; no bowel prep



Rangel, *Aliment Pharmacol Ther* 2015;42:1211

Spatial heterogeneity in intestines





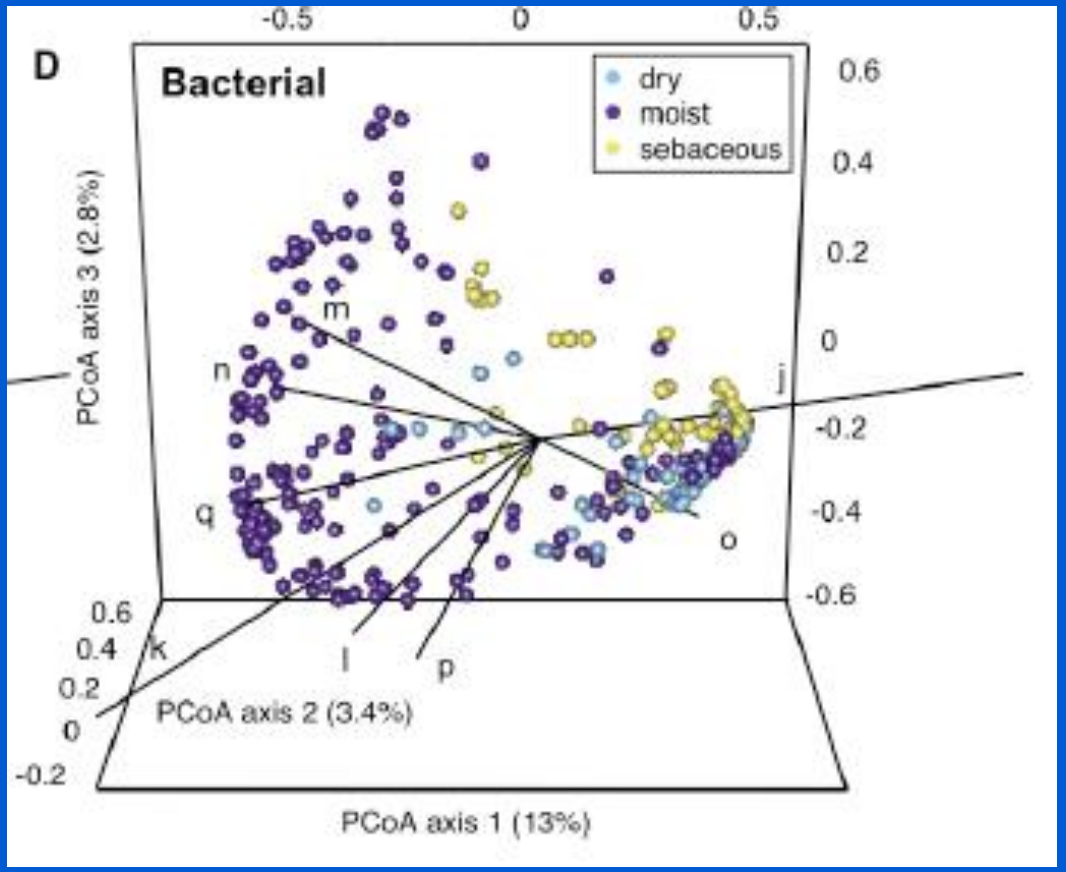
Mao, *Scientific Reports* 2015;5:16116

Skin has topographically distinct microbiota

Heel

Elbow

Back



Findley, *Nature* 2013;498:367

Mouth also has topographically distinct microbiota

Buccal mucosa
Keratinized gingiva
Hard palate

Throat
Tonsils
Tongue Dorsum
Saliva

Supragingival plaque
Subgingival plaque

Segata, *Genome Biol* 2012;13:R42

DNA preparation

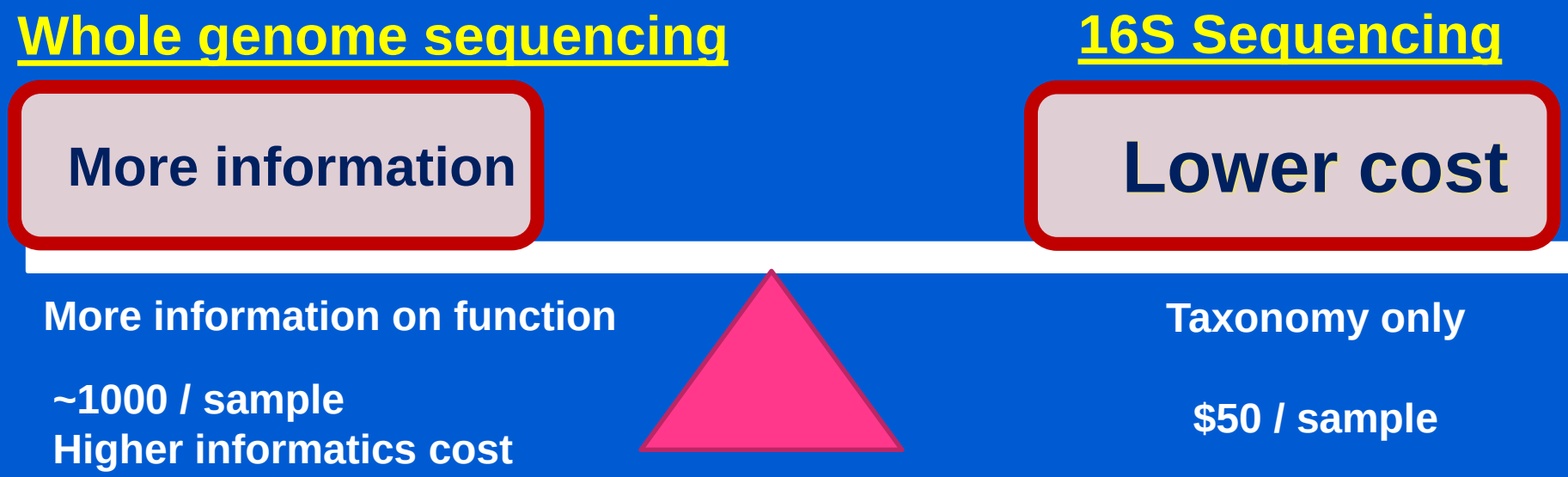
- Various kits used for DNA purification
  - MoBio preferred by HMP
  - We use Zymo for fecal collection
- Key is that you need conditions harsh enough to lyse the microbes

Outline

- Where to sample
- Sequencing
- Data analysis
- Next steps

Type of sequencing

- Amplification of 16S ribosomal DNA
- Whole genome sequencing



Sequencing prep at UAB

Starting with purified DNA

16S (Peter Eipers PhD)

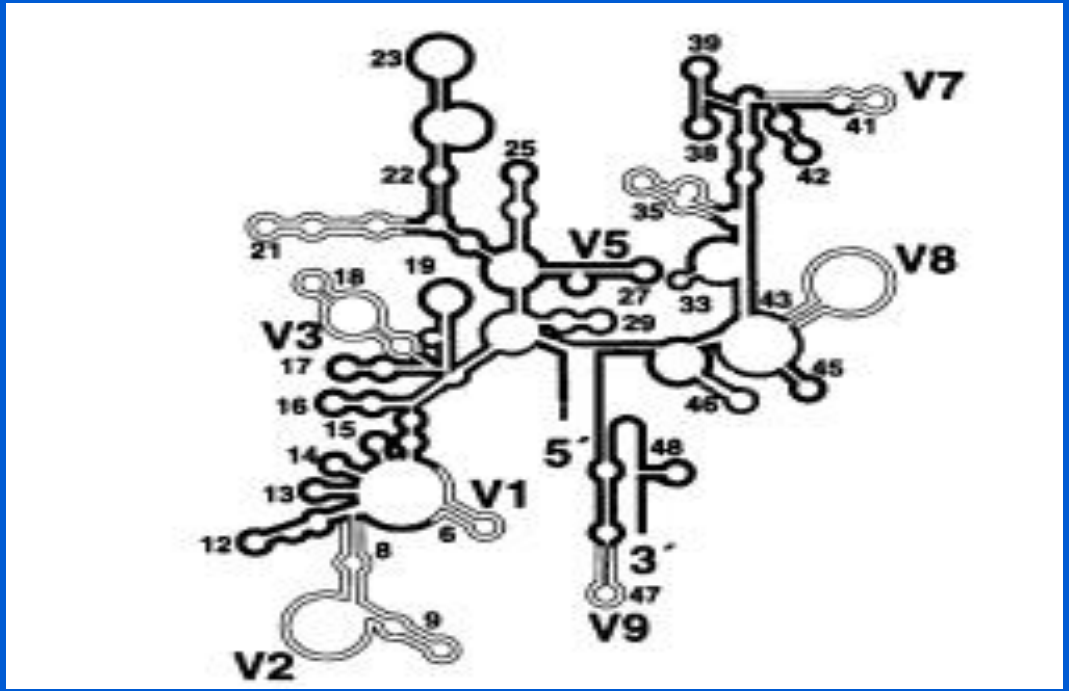
- PCR of V4 16S region
- Special primers
  - Barcode at one end
  - Adaptor at other

WGS (Mike Crowley PhD)

- Shear DNA
- Ligate adaptors to each end
  - Includes barcodes
- Short PCR
- Optional size selection

Key elements of 16S rDNA

- 9 conserved and 9 variable regions
  - Interspersed
- Primers bind to conserved regions
- Amplify variable sequences



Tortoli, *Clin Microbiol Rev* 2003;16:319

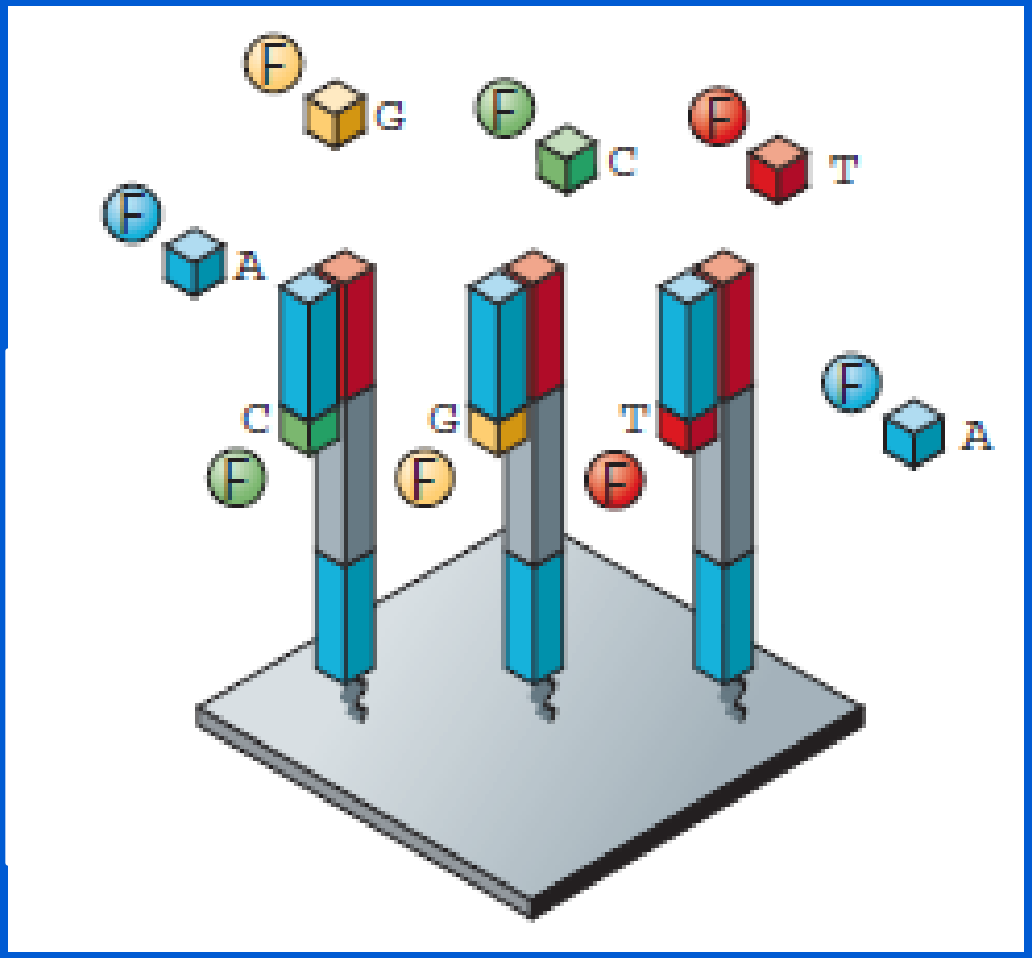
Foundation of 16S analysis

Position	Consensus composition	Purple bacteria				Gram-positive bacteria		Cyano-bacteria	Green sulfur bacteria	Spirochetes		Bacteroides		Planc-tomyces	Deino-coccus	Green non-sulfur bacteria
		α	β	γ	δ	Lo	High			Spiro-chetes	Lepto-spiras	Bacter-oides	Flavo-bacteria[a]			
47	C	●[b]	●	●	●	●	●	●	●	U[c]	U	●	●	G	●	●
48	Y	●	●	●	●	●	●	●	●	●	●	●	●	A	●	●
50	A	●	U	●	●	●	●	●	●	U	U	●	●	U	●	●
52	Y	●	●	●	●	●	●	●	●	A	A	●	●	G	G	●
53	A	●	●	●	●	●	●	●	●	G	●	●	●	G	●	G
353	A	●	●	●	●	●	●	●	●	●	●	●	●	U	●	●
570	G	●	●	●	●	●	●	●	●	●	●	U	U	U	●	●
698	G	U	●	●	Y	●	●	●	●	●	●	●	●	●	●	●
812	G	●	C	●c[d]	●	●	●	●	●	●	C	●	●c	●	C	●
906	G	Ag	Ag	Ag	Ag	●a	A	●	●	●	●	●	●	●	●	A
933	G	●	●	●	●	●	●	●	●a	●	●	●	●	A	●	●
955	U	●	Au	●	●	●	AC	●	●	●	●	●	●	C	●	●
976	G	●	A	●	●	●	●	●	●	●a	A	●	●	●	●	●
983:1	—[e]	—	—	—	—	—	—	—	—	—	—	—	—	U	—	—
995	C	●	●	●	●	●	●	●	(A)[f]	●	●	A	A	●	●	●

Carl Woese, *Microbiol Rev* 1987;51:221

Cyclic reversible termination

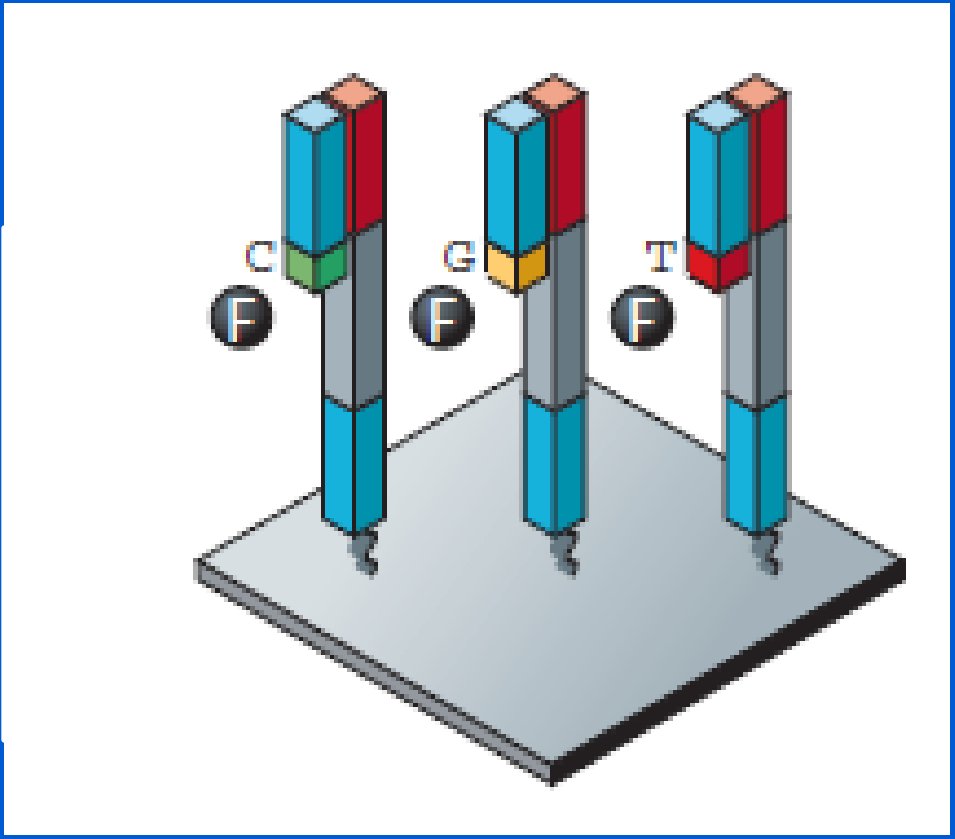
Metzker, *Nature Rev Genet* 2010;11:31



Incorporate all four nucleotides, each with different dye

Cyclic reversible termination

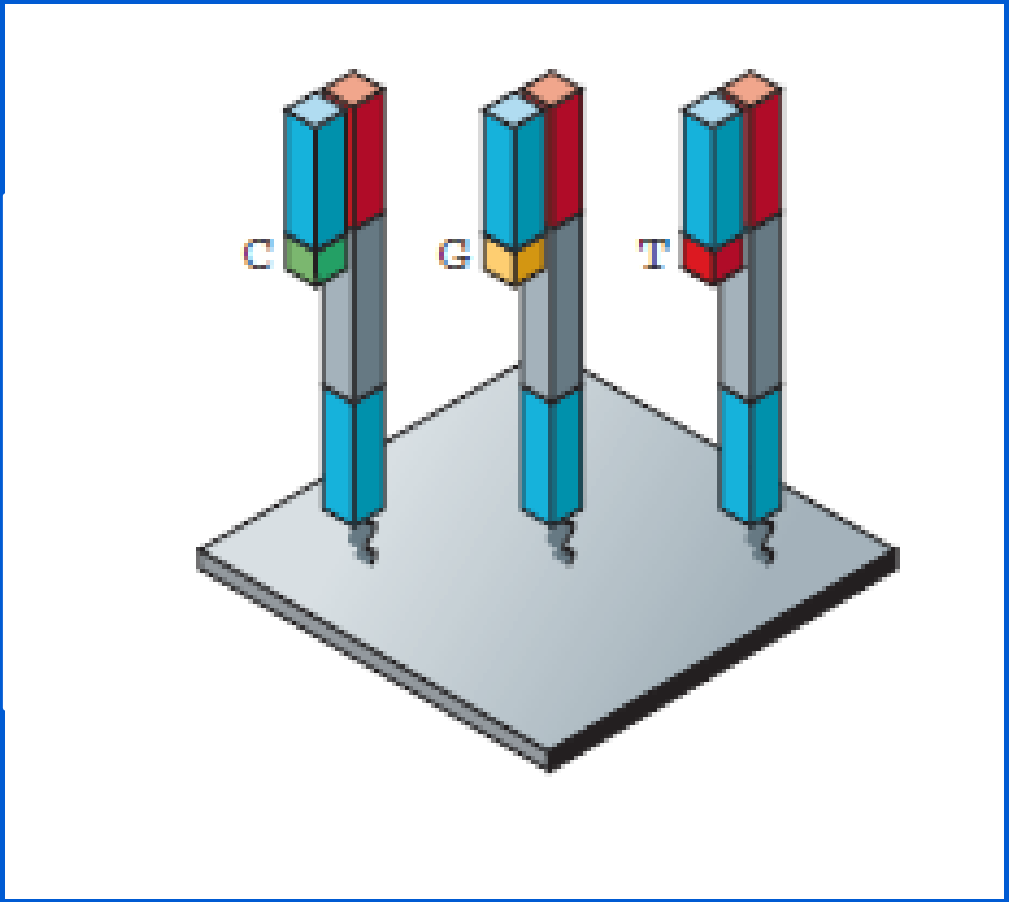
Metzker, *Nature Rev Genet* 2010;11:31



Wash out unused nucleotides; image

Cyclic reversible termination

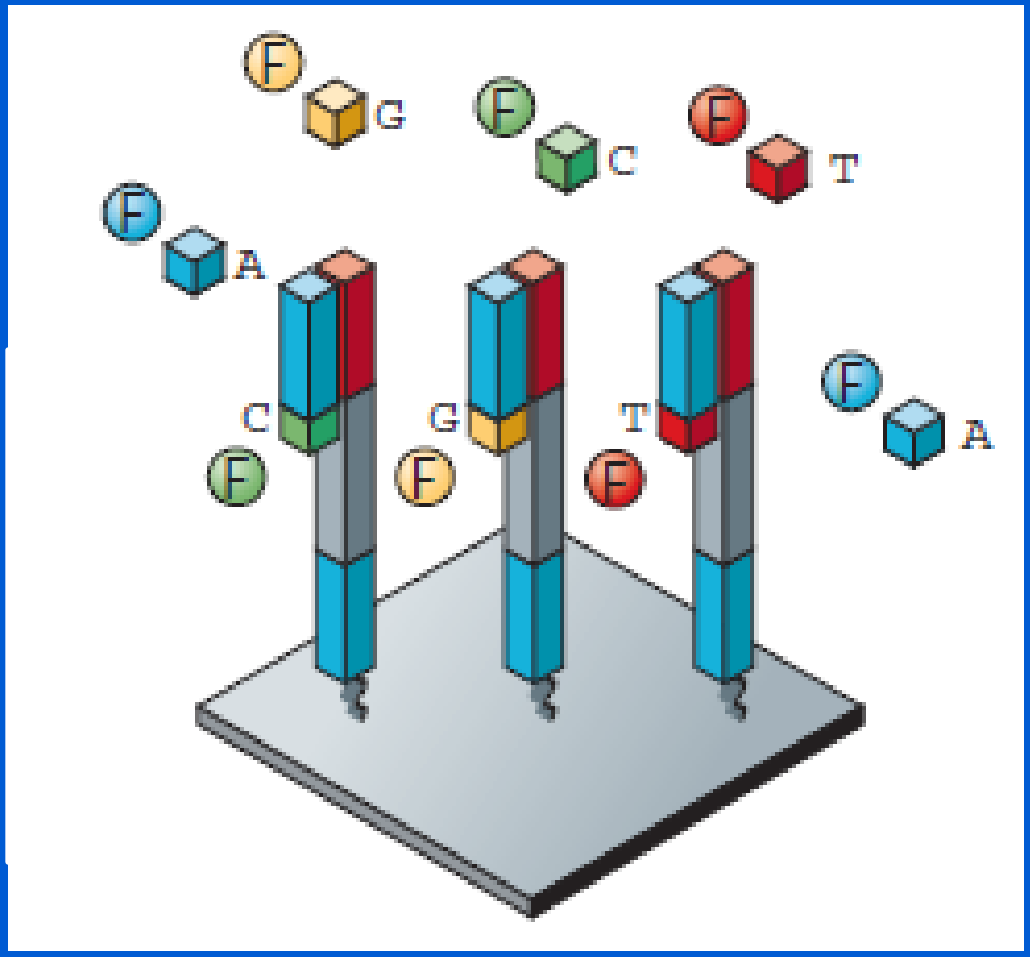
Metzker, *Nature Rev Genet* 2010;11:31



Cleave dye and terminating groups

Cyclic reversible termination

Metzker, *Nature Rev Genet* 2010;11:31



Repeat

Barcodes to sort out samples

Sample 1: AGGTTCCA
Sample 2: GGCAATTT
Sample 3: TTGGAAAC

Outline

- Where to sample
- Sequencing
- Data analysis
- Next steps

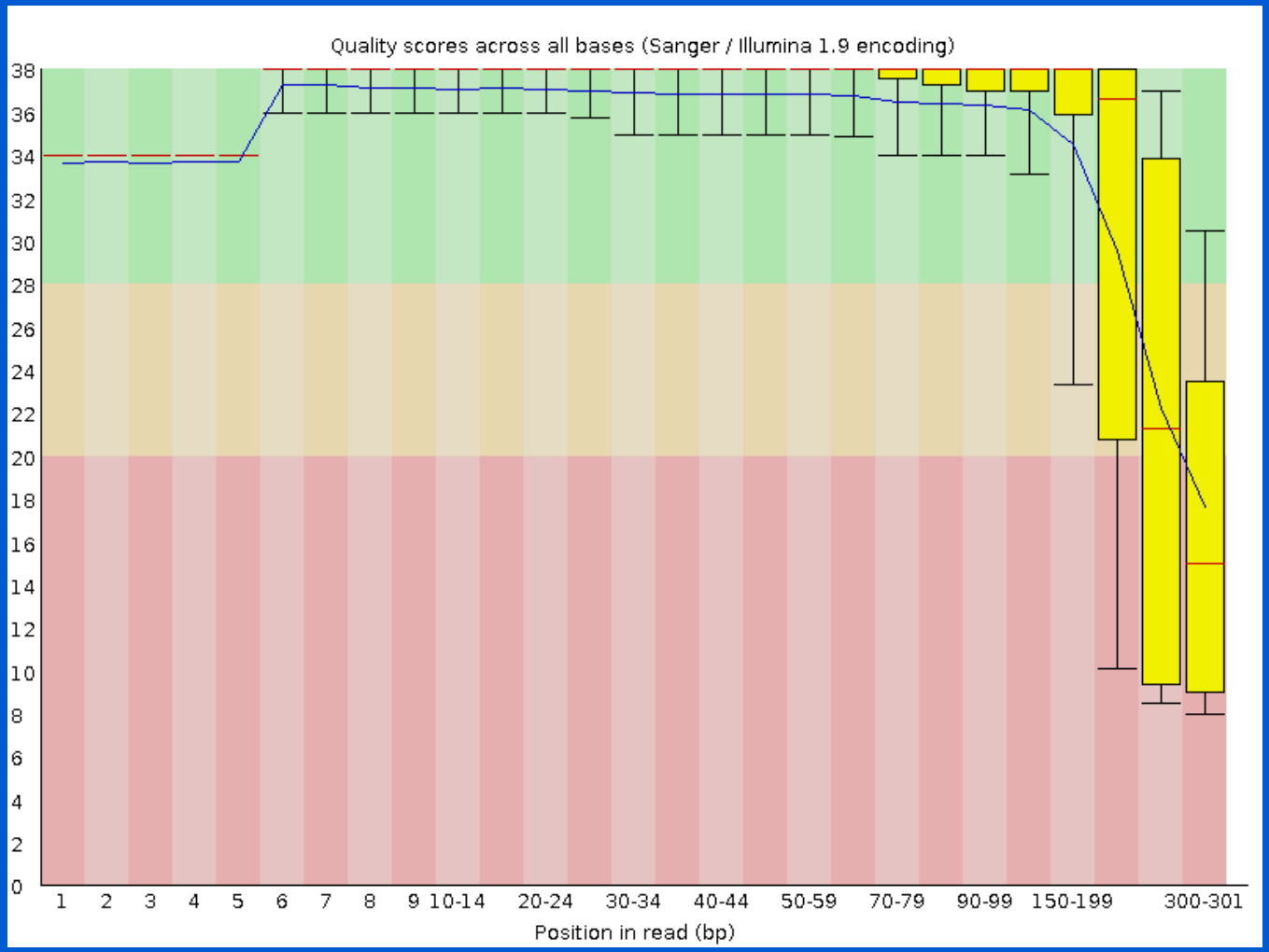
Output: fastq files

Sample Fastq output (two DNA strands)

Header: @M02079:147:000000000-AK0J5:1:1101:15736:1676 1:N:0:49
Sequence: TACAGAGGTCTCAAGCGTTGTTCGGAATCACTGGGCGTAA
Additional line: +
Quality: >//>>EEGGFFE/////</>-<0>DBF1<F<1.<<-<GD

Header: @M02079:147:000000000-AK0J5:1:1101:15989:1722 1:N:0:49
Sequence: TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAG
Additional line: +
Quality: B3EHFGGHF3FB43/E?EFGGFFGH3/B4?//B/FG?122FB

Assess quality of reads



Position in read (BP)

Quality filtering options

- **Trim the low-quality tails**
  - **Option: remove sequence if more than a set percentage of bases are trimmed**
- **Remove sequences with:**
  - **Ambiguous bases**
  - **Quality scores below a threshold**
  - **More than a certain number of predicted errors**

The bane of amplicon analysis is PCR errors

- PCR errors, especially if they occur early in the cycle, can look like unique species
- Additionally, sequencing errors may have occurred, especially if the quality is low
- A common approach to addressing this issue is clustering of similar sequences into operational taxonomic units (OTUs)

QIIME

- Open-source bioinformatics pipeline
- Designed for 16S sequence analysis
- Every step from fastq processing through data analysis

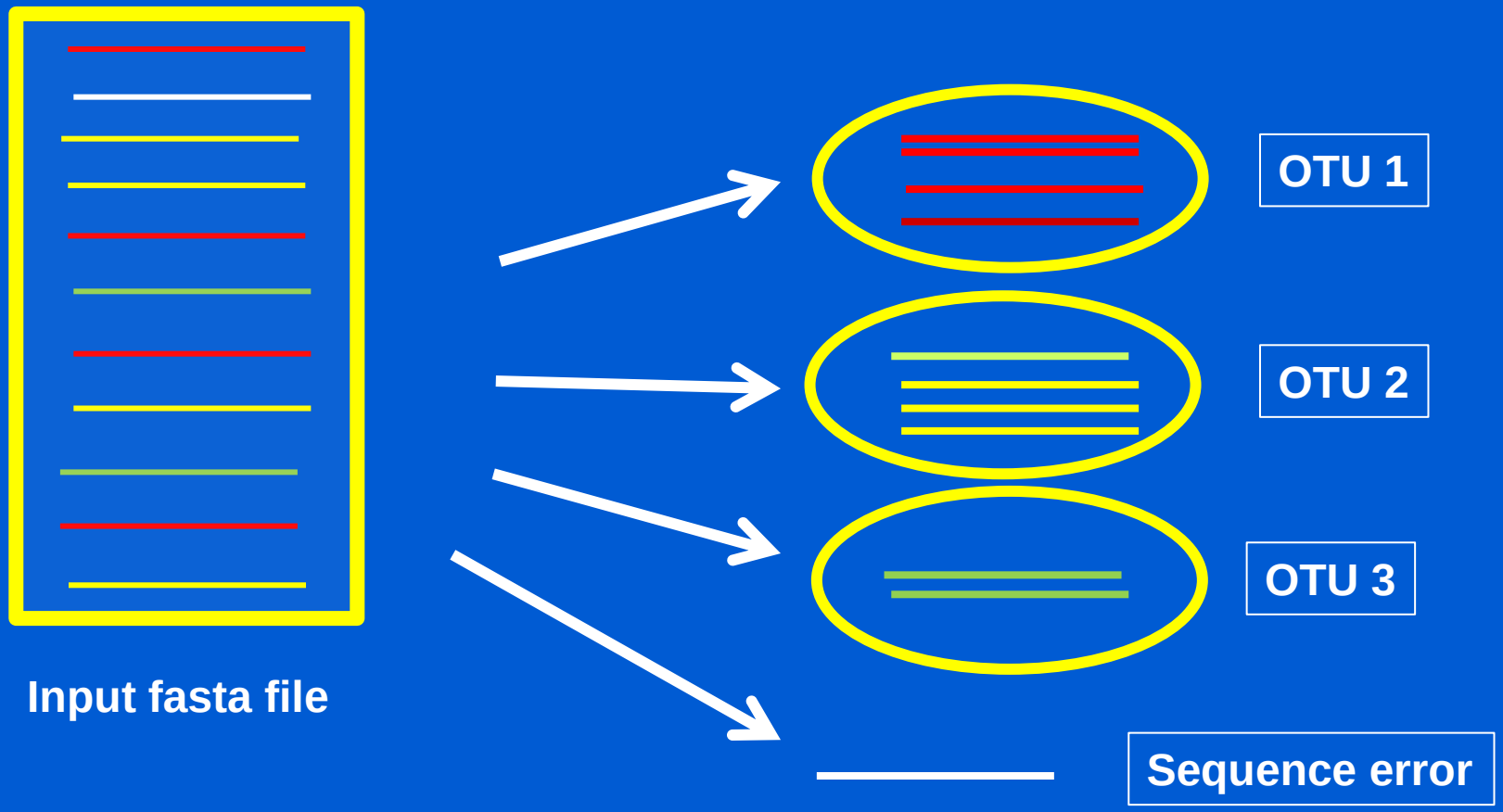
www.QIIME.org

Operational taxonomic unit picking

- Similar (typically 97%) sequences are clustered together
- Creates an "OTU table"
  - List of all OTUs
  - Frequency of each OTU in each sample
  - Taxonomic assignment of each OTU

OTUId	Sample1	Sample2	Sample3	Taxonomy
OTU_1	30	70	35	k__Bacteria;p__Firmicutes…
OTU_2	40	10	40	k__Bacteria;p__Bacteroidetes…
OTU_3	30	20	25	K__Bacteria;p__Firmicutes…

Basic idea of OTU picking



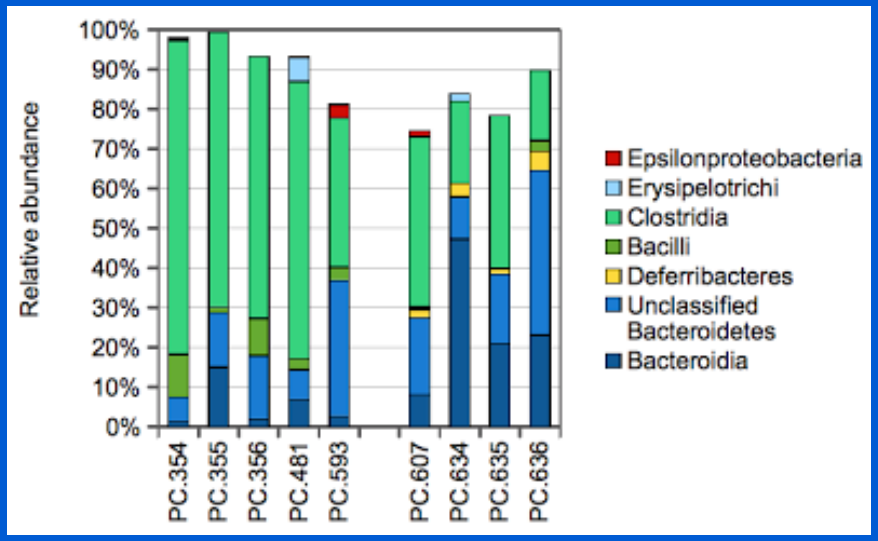
Limitations to OTU picking

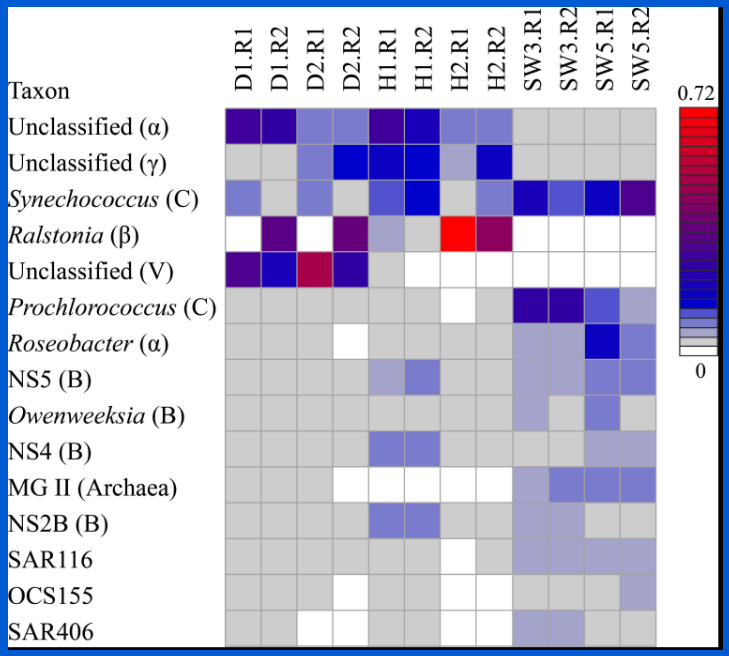
- OTUs are clustered based upon similarity
- Strain- and even species-level distinctions are often lost
- PCR amplification errors and low quality reads with high error rates can be mistaken as unique OTUs

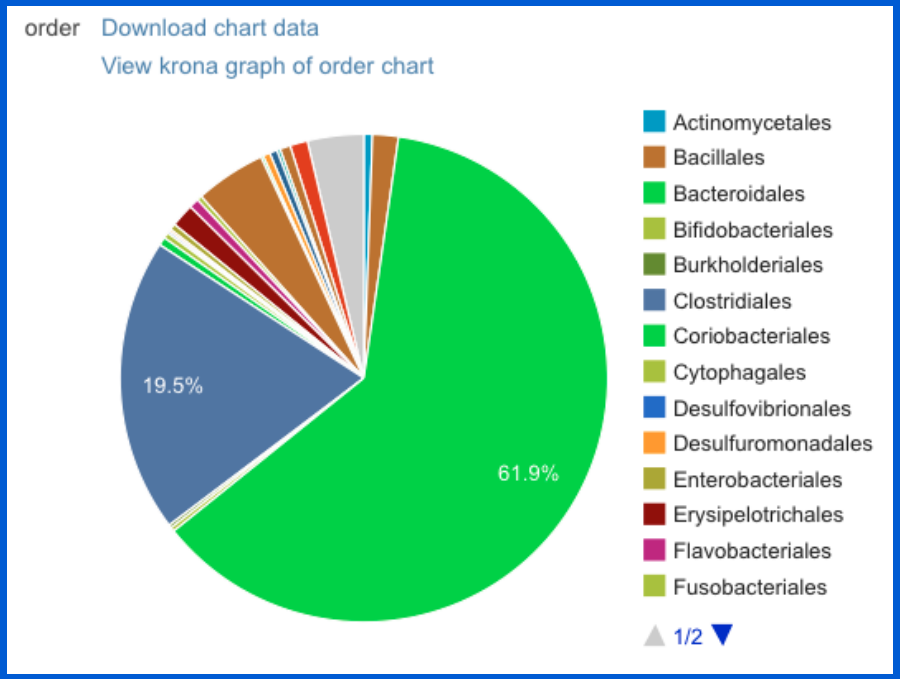
Downstream analyses with OTU tables

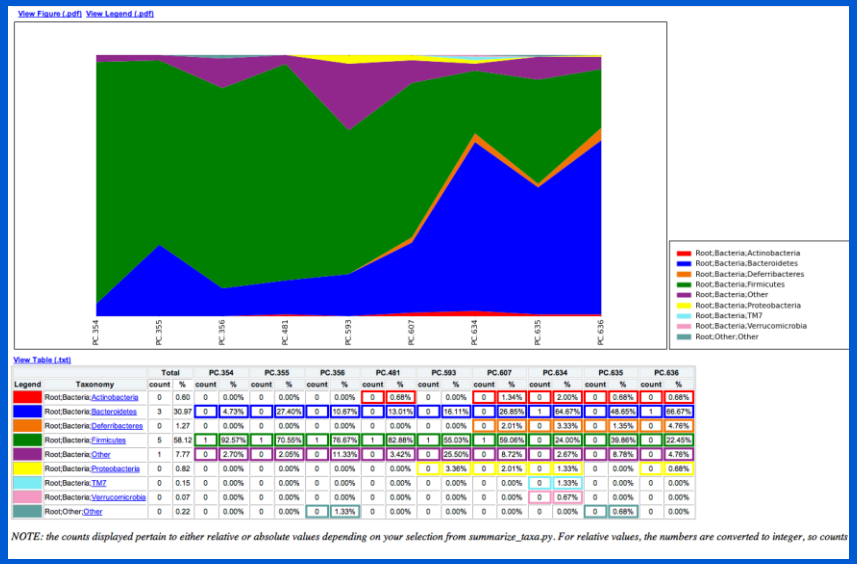
- Display taxonomy
- Alpha diversity (within group)
- Beta diversity (between groups)

Displaying taxonomy

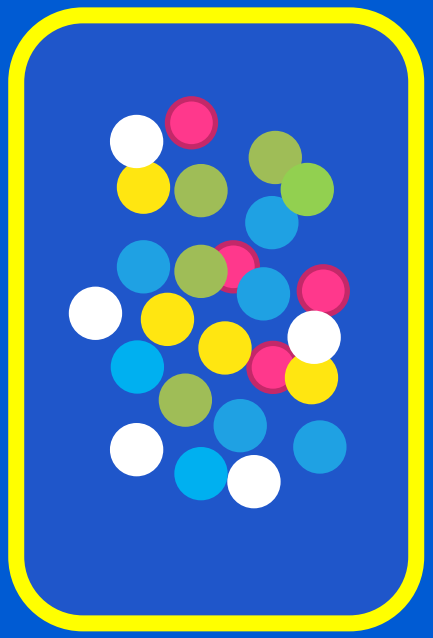




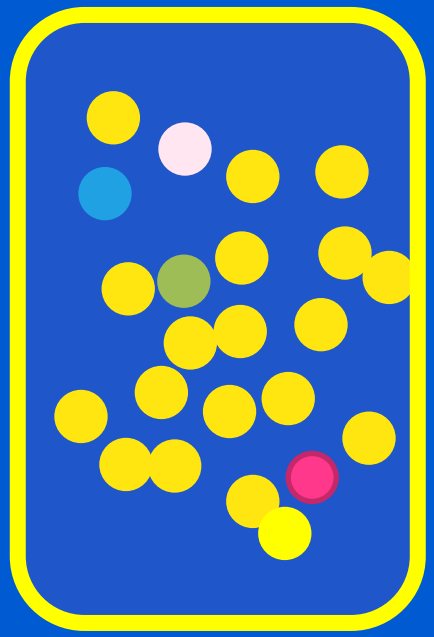




Alpha diversity



Rich and even



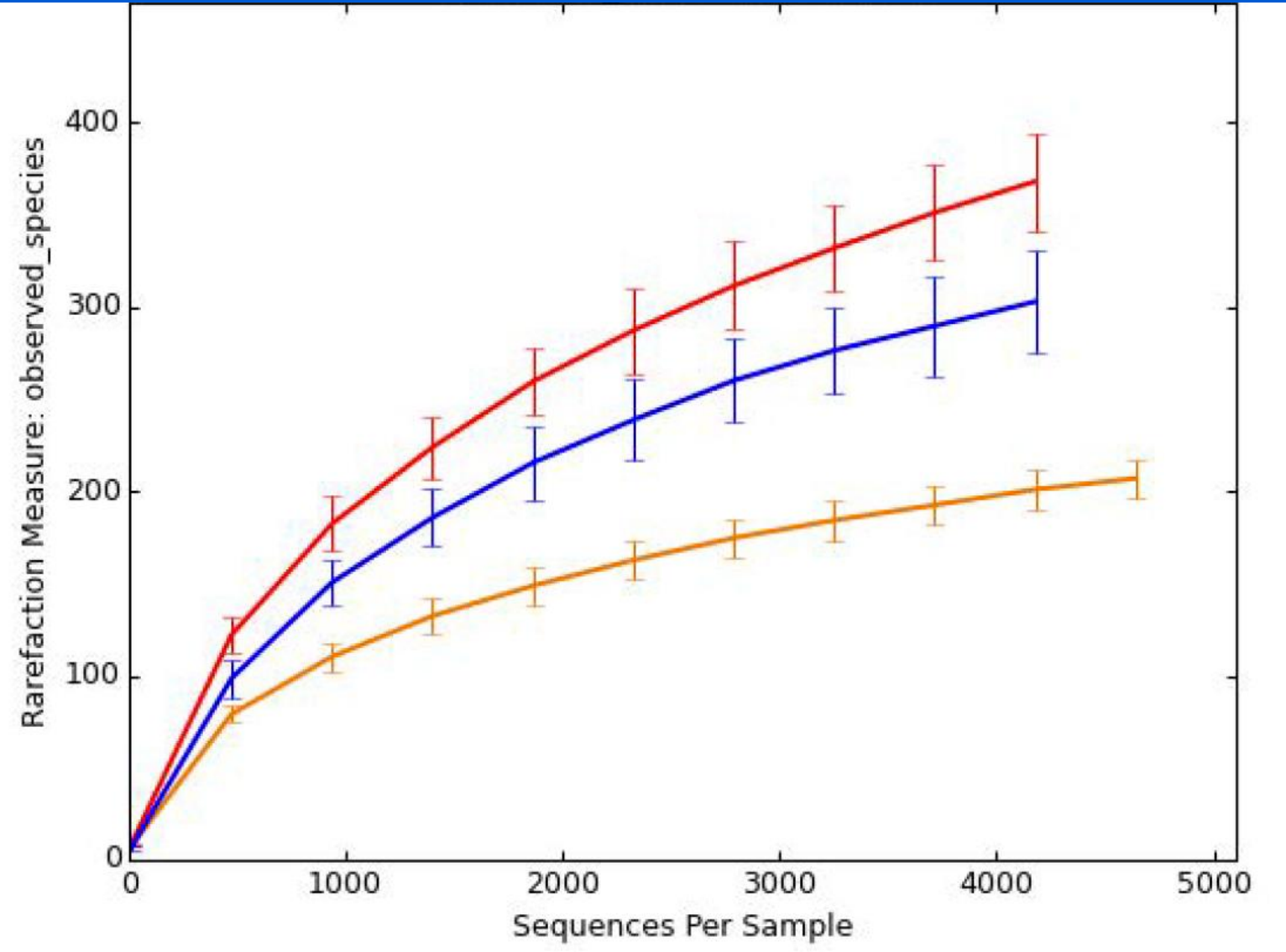
Rich, not even

Not rich or even

Cautionary note about measuring alpha diversity

- Must take into account sequencing depth
- Typically, at UAB, depth is 50K – 150K sequences per sample for 16S
- To a point, diversity increases with higher depth, as you pick up more rare species

Illustration of rarefaction curves



Beta diversity

- This speaks to the diversity between two different groups
- Typically depicted as PCOA plots
- Consider sequencing depth

Sample distance matrix

	C1	C2	C3	C4	T1	T2	T3
C1	0	0.35	0.31	0.39	0.88	0.79	0.91
C2	0.35	0	0.42	0.22	0.92	0.90	0.74
C3	0.31	0.42	0	0.35	0.74	0.79	0.91
C4	0.39	0.22	0.35	0	0.82	0.84	0.92
T1	0.88	0.92	0.74	0.82	0	0.29	0.21
T2	0.79	0.90	0.79	0.84	0.29	0	0.32
T3	0.91	0.74	0.91	0.92	0.21	0.32	0

Sample distance matrix

	C1	C2	C3	C4	T1	T2	T3
C1	0	0.35	0.31	0.39	0.88	0.79	0.91
C2		0	0.42	0.22	0.92	0.90	0.74
C3			0	0.35	0.74	0.79	0.91
C4				0	0.82	0.84	0.92
T1					0	0.29	0.21
T2						0	0.32
T3							0

Statistical tests compare distance between vs within groups

e.g. Permanova

Microbiota in juvenile spondyloarthritis

Spondyloarthritis

- Type of pediatric and adult arthritis
- Distinctive demographic & clinical features
- High prevalence of gut inflammation
  - Frank IBD: 5 – 10% (IBD-associated arthritis)
  - Subclinical: 67%
- In children, goes by the name enthesitis-related arthritis (ERA)

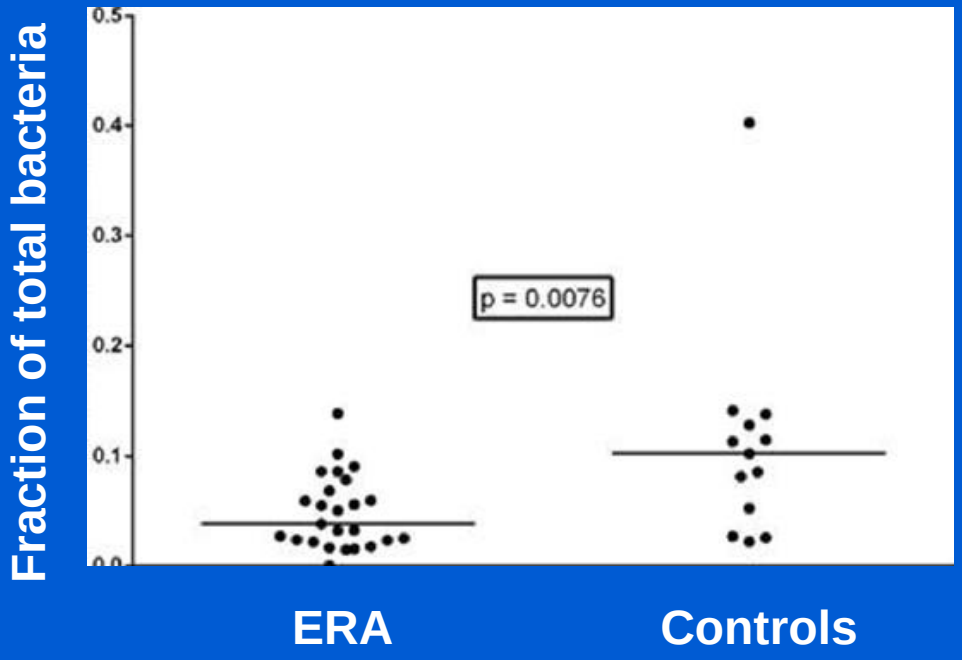
Initial foray into microbiota analysis

- Obtained fecal specimens from children with ERA, along with healthy controls
- Patients included those who were newly diagnosed as well as those with long-standing disease
- Performed sequencing of 16S rDNA
- Analyzed with QIIME platform

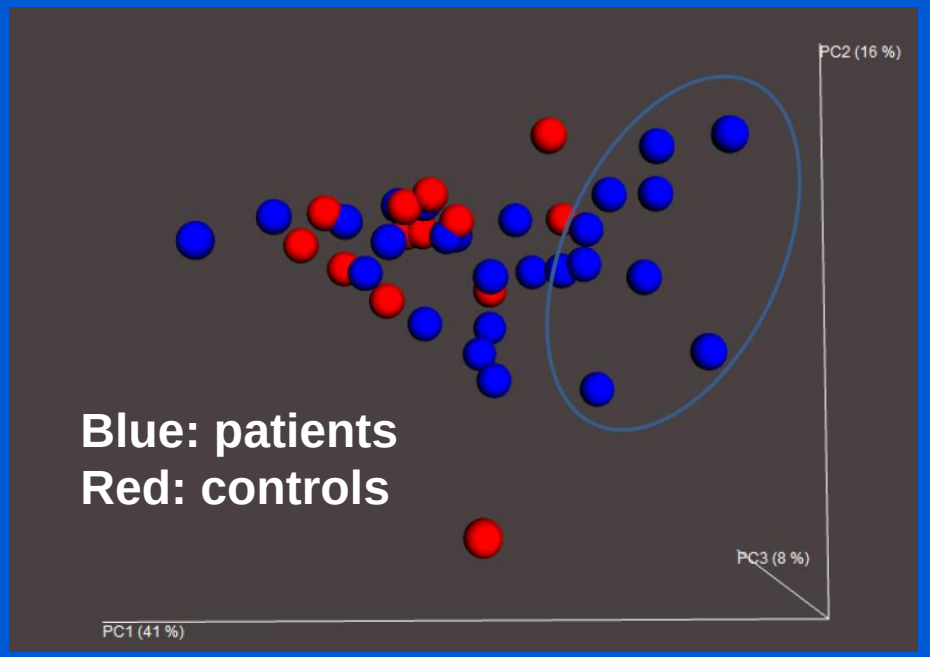
Microbiota in ERA

- Decreased *Faecalibacterium prausnitzii* in children with ERA
- Small patient cluster identified by PCoA

Fecal abundance of *F. prausnitzii*



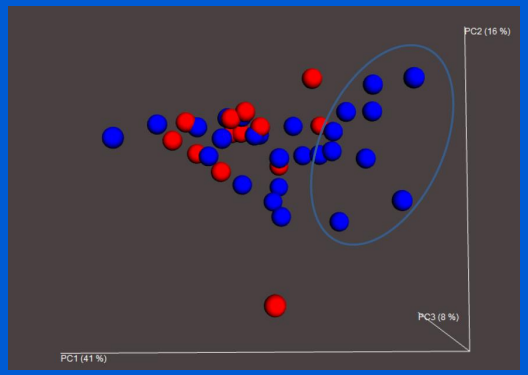
Principal coordinates analysis



Stoll, *Arth Res Ther* 2014;16:486

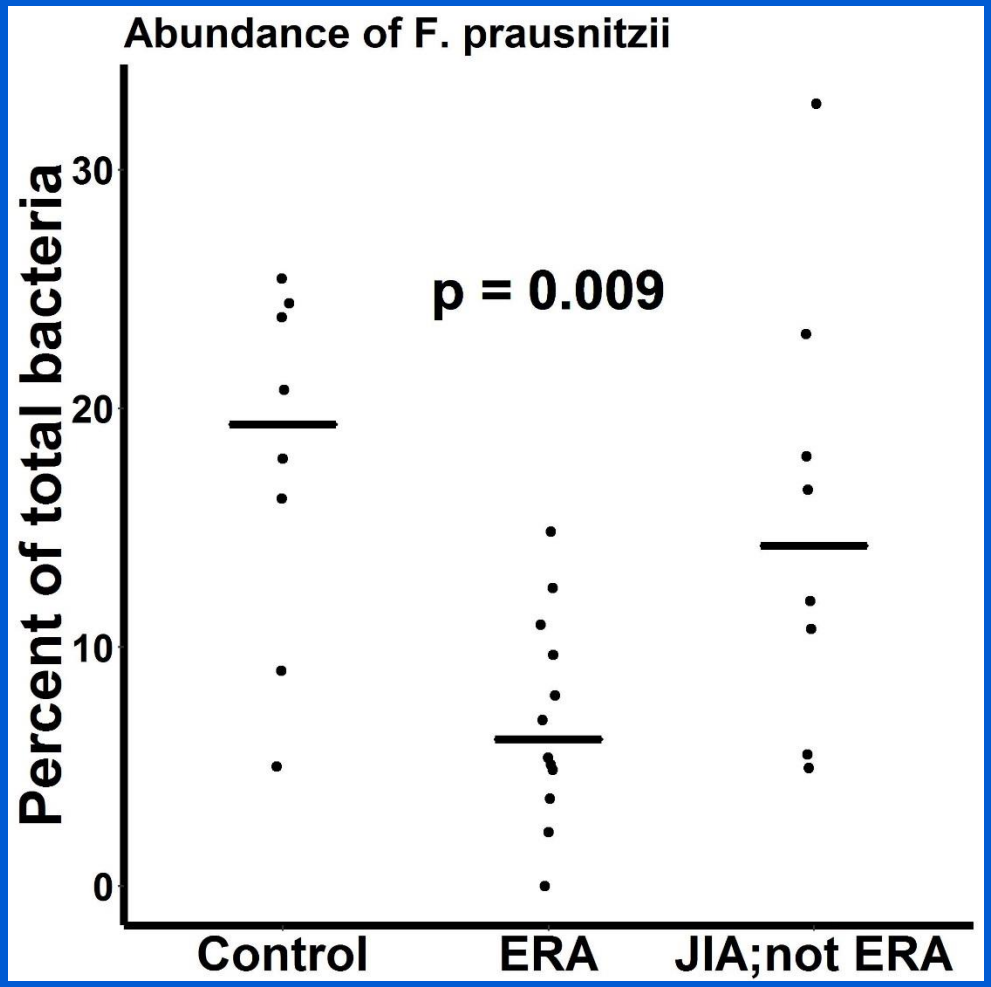
Clusters are defined by *Bacteroides* and *Akkermansia*

- **The two clusters had similar abundance of *F. Prausnitzii***
- **The subjects who clustered with the controls tended to have high abundance of *Akkermansia* (> 1% in 7/17 vs 0/8)**
- **The subjects forming their own cluster had high *Bacteroides* abundance (41% vs 13%)**



Stoll, *Arth Res Ther* 2014;16:486

Altered *F. prausnitzii* is largely limited to SpA subtype of pediatric arthritis



Stoll, *Clin Immunol* 2015;159:134

F. prausnitzii is an anti-inflammatory organism

- **Decreased fecal abundance in IBD[1]**
- **Increased IL-10 production by PBMCs[2,3]**
- **Reduces inflammation in colitis model[2]**
- **Major butyrate producer[4]**
  - **Beneficial effects on enterocytes[5]**
  - **Increases colonic regulatory T cells[6]**

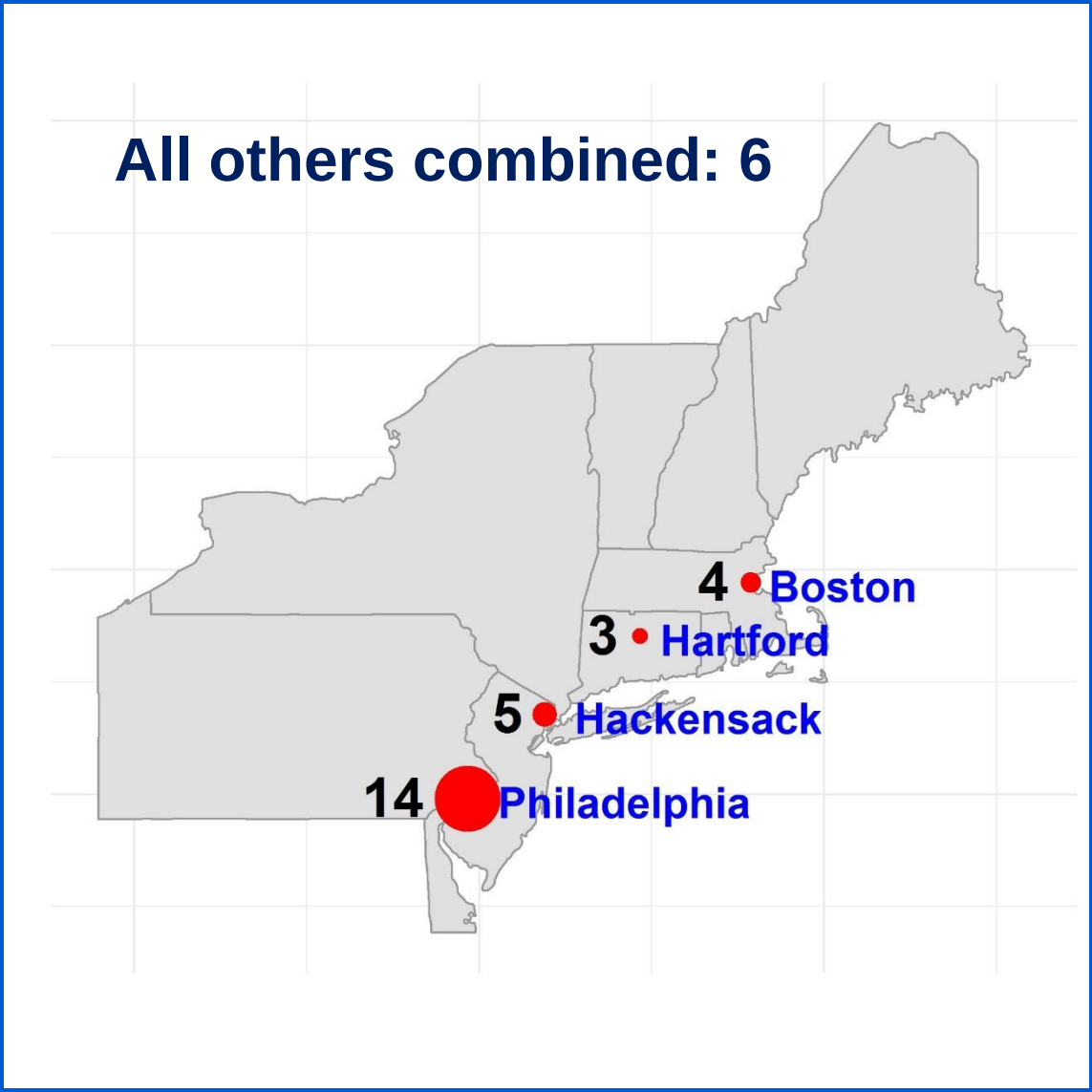
[1]Cao,*Gastroenterol Res Practice* 2014:872725 (review)
[2]Sokol, *PNAS* 2008;105:16731
[3]Rossi, *Sci Rep* 2016;6:18507
[4]Hold, *Appl Environ Microbiol* 2003;69:4320
[5]Hamer, *Aliment Pharmacol Ther 2008;27:104*
[6]Smith, *Science* 2013341:569

Focus on treatment-naïve subjects

- Recruited pediatric spondyloarthritis patients from centers around the country
- Attempted to recruit healthy controls from each site
- Analyzed with QIIME (initially)

Collaborating centers

Collaborators are not all created equal



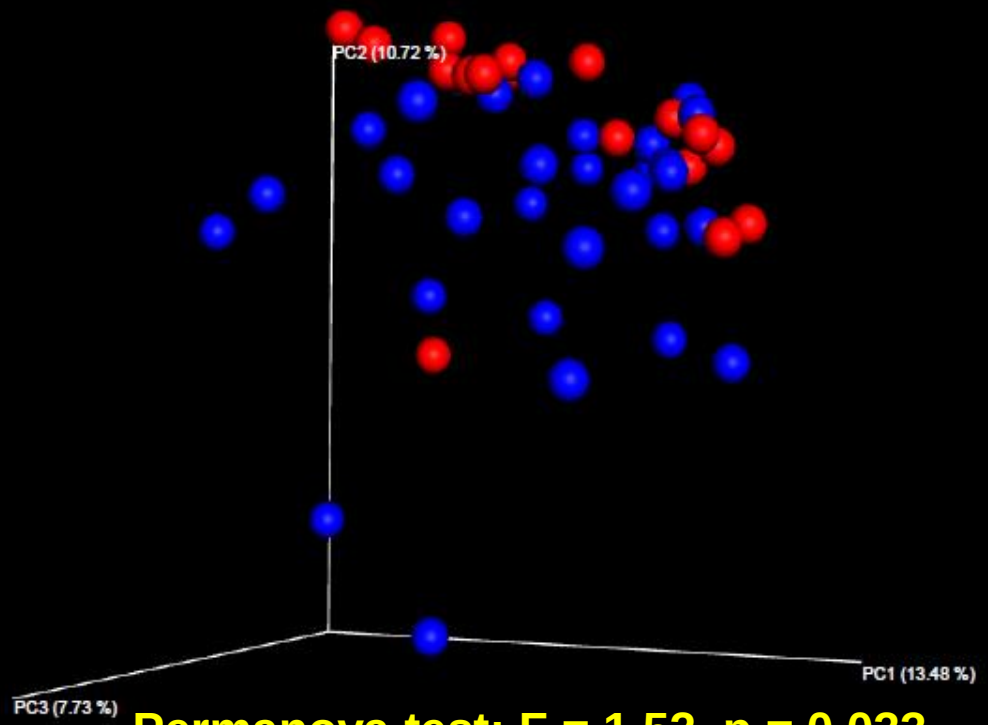
Subjects

Characteristic	ERA	Controls
n	30	19
Age (years; mean ± SD)	13.5 ± 3.0	13.6 ± 2.7
BMI (kg / m^2; mean ± SD)	20.7 ± 4.1	21.5 ± 6.1
Male	19, 63%	13, 68%
Caucasian	23, 77%	17, 89%

Clustering by diagnosis

Blue: patients
Red: controls

Bray Curtis

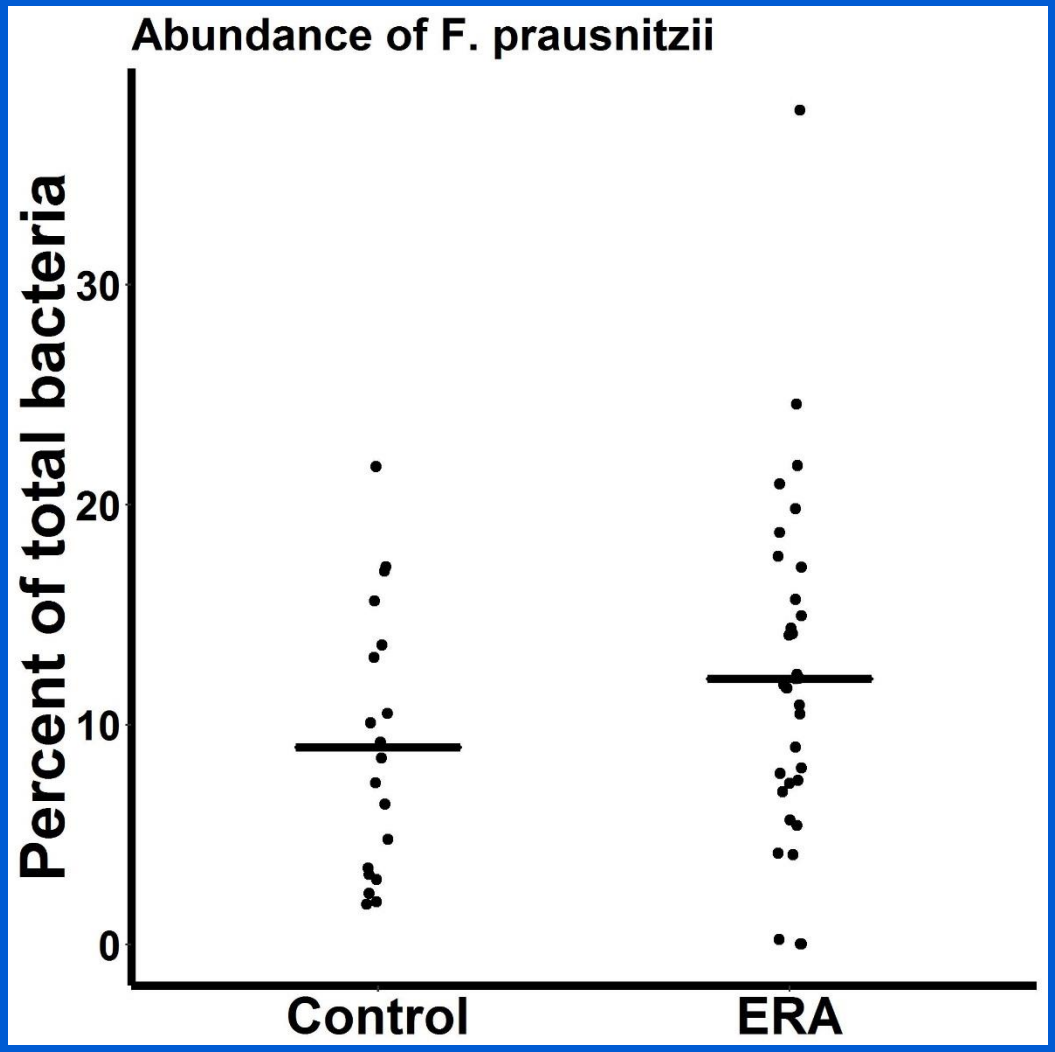


Permanova test: F = 1.52, p = 0.033

Taxonomic assessment

- **Pairwise comparisons revealed minimal differences between the groups**
  - **Especially after adjustment for multiple comparisons**
  - **At the phylum level, decreased Actinobacteria in ERA (3.3 vs 9.0%, $p < 0.002$)**
- **Focused on organism of interest based upon previous work**
  - ***Faecalibacterium prausnitzii***

Not what I expected



Limitations to OTU picking

- OTUs are clustered based upon similarity
- Strain- and even species-level distinctions are often lost
- PCR amplification errors and low quality reads with high error rates can be mistaken as unique OTUs

Alternative to OTU picking

Published in final edited form as:
Nat Methods. 2016 July ; 13(7): 581–583. doi:10.1038/nmeth.3869.

DADA2: High resolution sample inference from Illumina amplicon data

Benjamin J Callahan[1,*], **Paul J McMurdie**[2], **Michael J Rosen**[3], **Andrew W Han**[2], **Amy Jo A Johnson**[2], and **Susan P Holmes**[1]

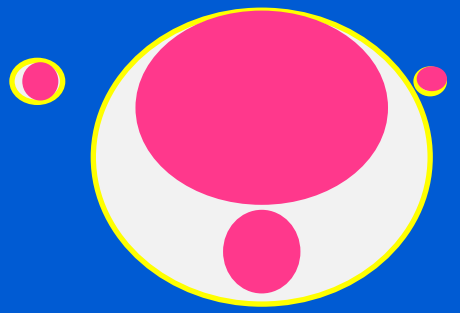
[1]Department of Statistics, Stanford University, Stanford, CA, USA

[2]Second Genome, South San Francisco, CA, USA

[3]Department of Applied Physics, Stanford University, Stanford, CA, USA

- **Use error modeling to distinguish sequence errors from sequence variants**
  - **Real OTUs versus fake OTUs**

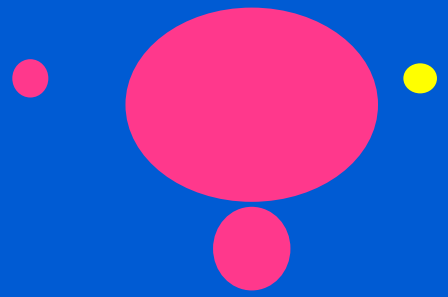
DADA2 offers a replacement to traditional OTU picking methods



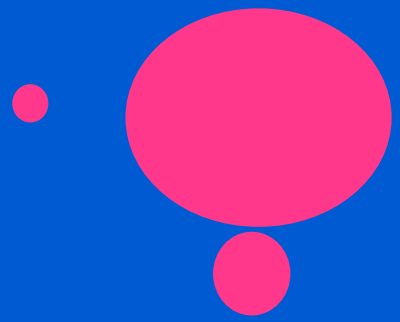
DADA2 – basic concept

- Start with the most abundant sequence (centroid)
- For each other sequence, test the null hypothesis that it is the same as the centroid
- Two factors are used to compare sequences with the centroids
  - Distance
    - Number of nucleotide differences
    - Quality of the sequences
    - Specific nucleotide transitions (e.g. T → A, T → C, etc)
  - Abundance
- If the lowest (smallest) p-value is below a cutoff, then the null hypothesis is rejected, and the sequence becomes a centroid
- Otherwise, the sequence is combined with the nearest centroid

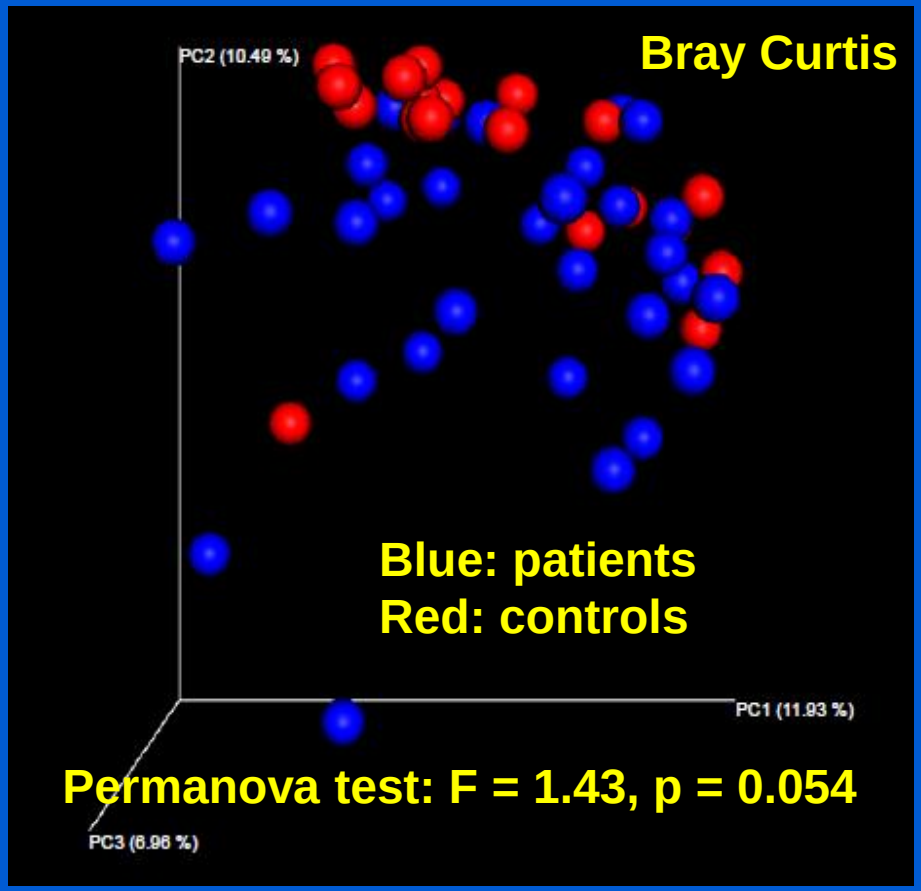
DADA2 offers a replacement to traditional OTU picking methods

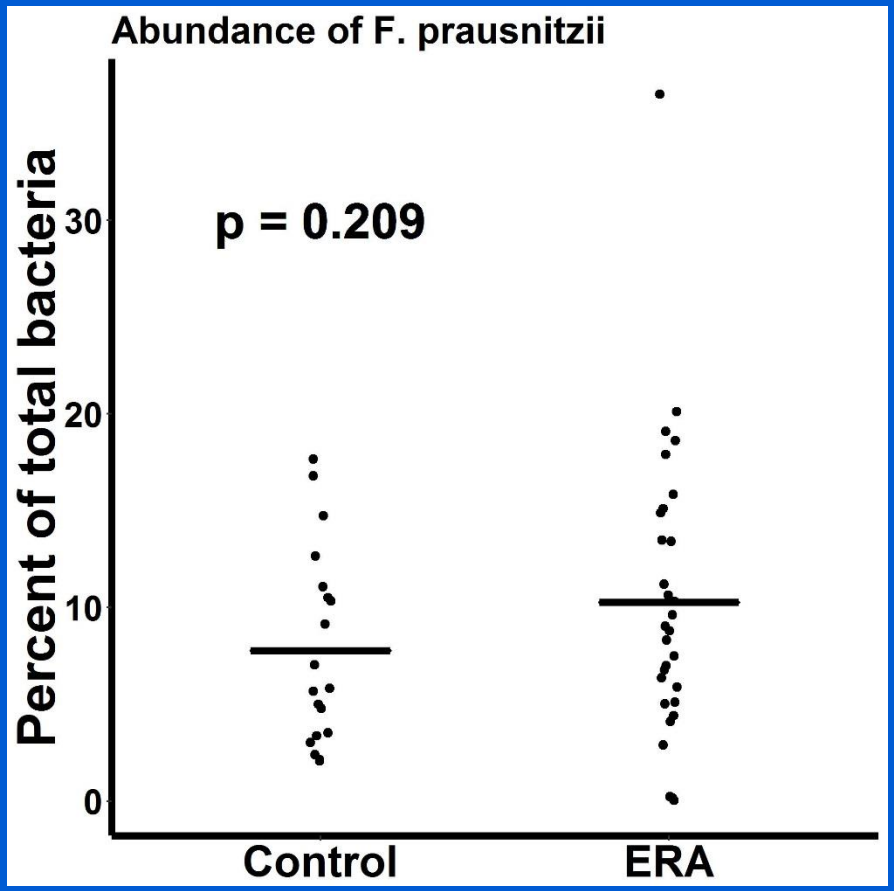


DADA2 offers a replacement to traditional OTU picking methods



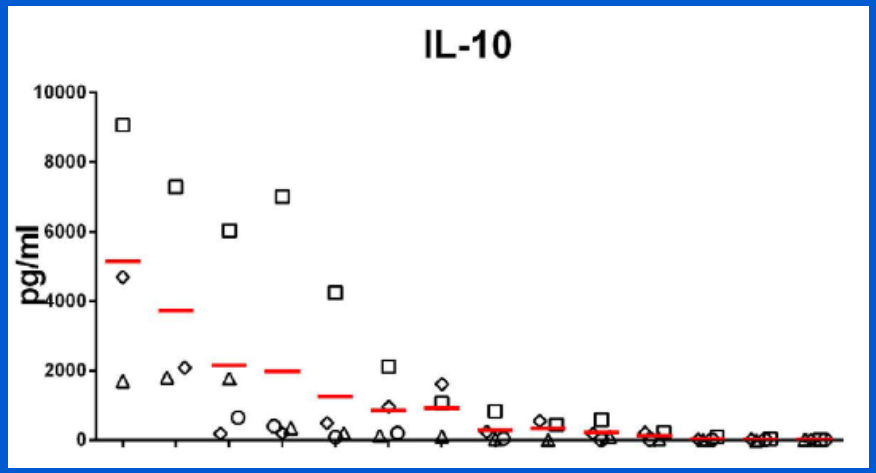
At first blush, results from the DADA2 analysis were similar





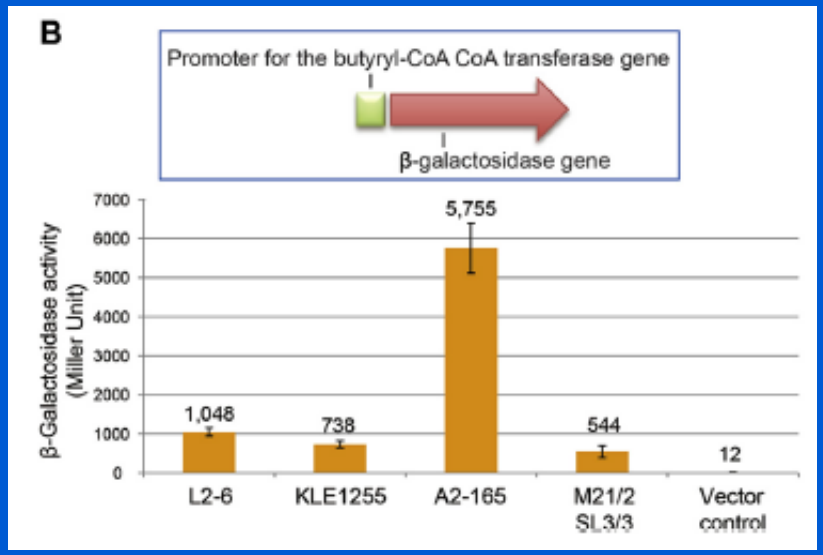
There are differences within strains of *F. praustnizii*

IL-10 production from hDC



Rossi, *Sci Rep* 2016;6:18507

Butyryl-CoA transferase gene promoter activity



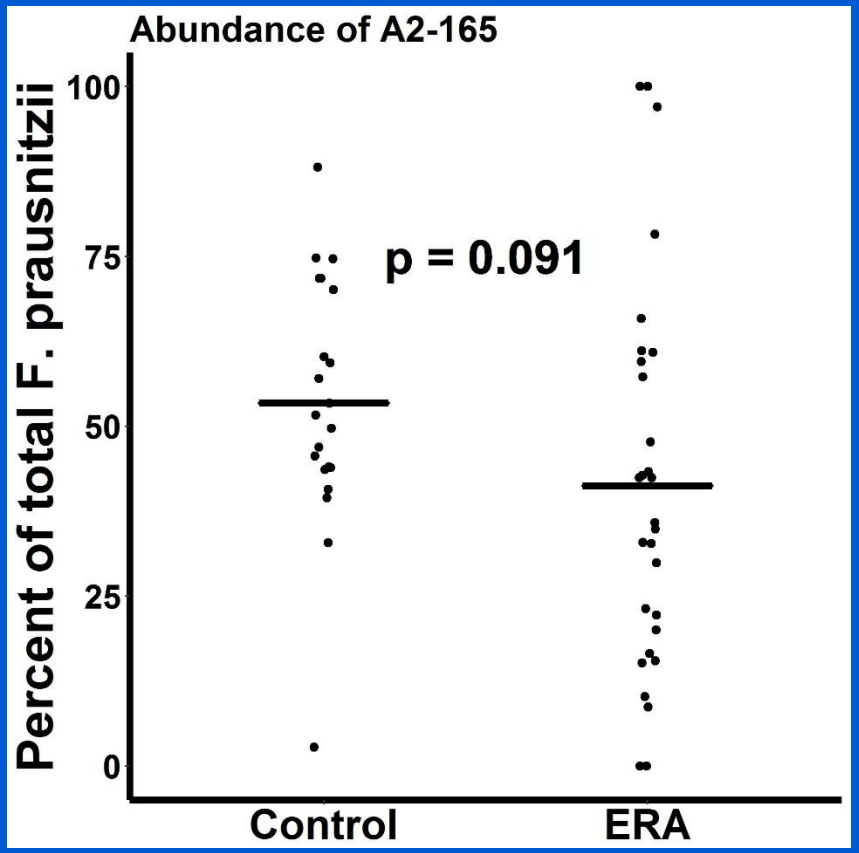
Song, *JACI* 2016;137:852

23 sequence variants matching *F. prausnitzii* were detected

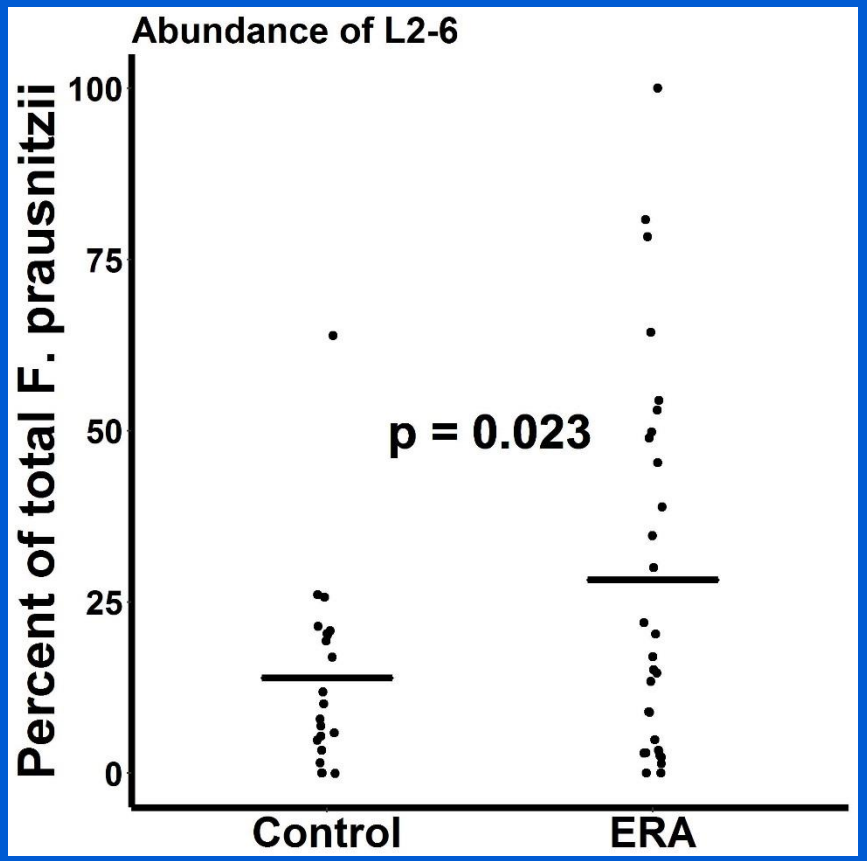
- I downloaded the 16S sequences for A2-165 and L2/6
- Ran BLAST of the 23 SVs against these two sequences
- One SV was a 100% match (253 / 253) to A2-165
- Another SV was a 99% match (252 / 253) to L2/6

Distribution of FP strains differs between patients and controls

A2-165 ("good")



L2/6 (less "good")



Whole genome sequencing

- Shotgun sequencing of all the DNA present in a sample
- May not include viral particles
- Will include human contaminant DNA

Removal of host DNA sequences

- **Not required with 16S analysis**
  - **Host DNA should not be amplified**
- **Contamination can occur with WGS**
  - **Variable with fecal microbes**
  - **High likelihood with other habitats**

Options for host DNA removal

- Reference database of host DNA
  - Filter out alignments
  - BLAST or Bowtie2 / BWA
- Reference database of microbes
  - Include sequences that align with dominant bacteria
  - Output will be limited to these bacteria

Assembly

- Most packages not designed for microbiota
  - Hundreds of species
- Unclear if even required

Options for taxonomic assessment

- **Metaphlan[1]**
  - **Assigns taxonomy based upon marker genes**
  - **Second version includes 17K reference genomes**
- **Alignment-based**
- **Composition-based: bacteria have unique sequence features**
  - **GC content**
  - **Nucleotide repeats**
  - **Codon usage**

[1]Truong *Nature Methods* 2015;10:902

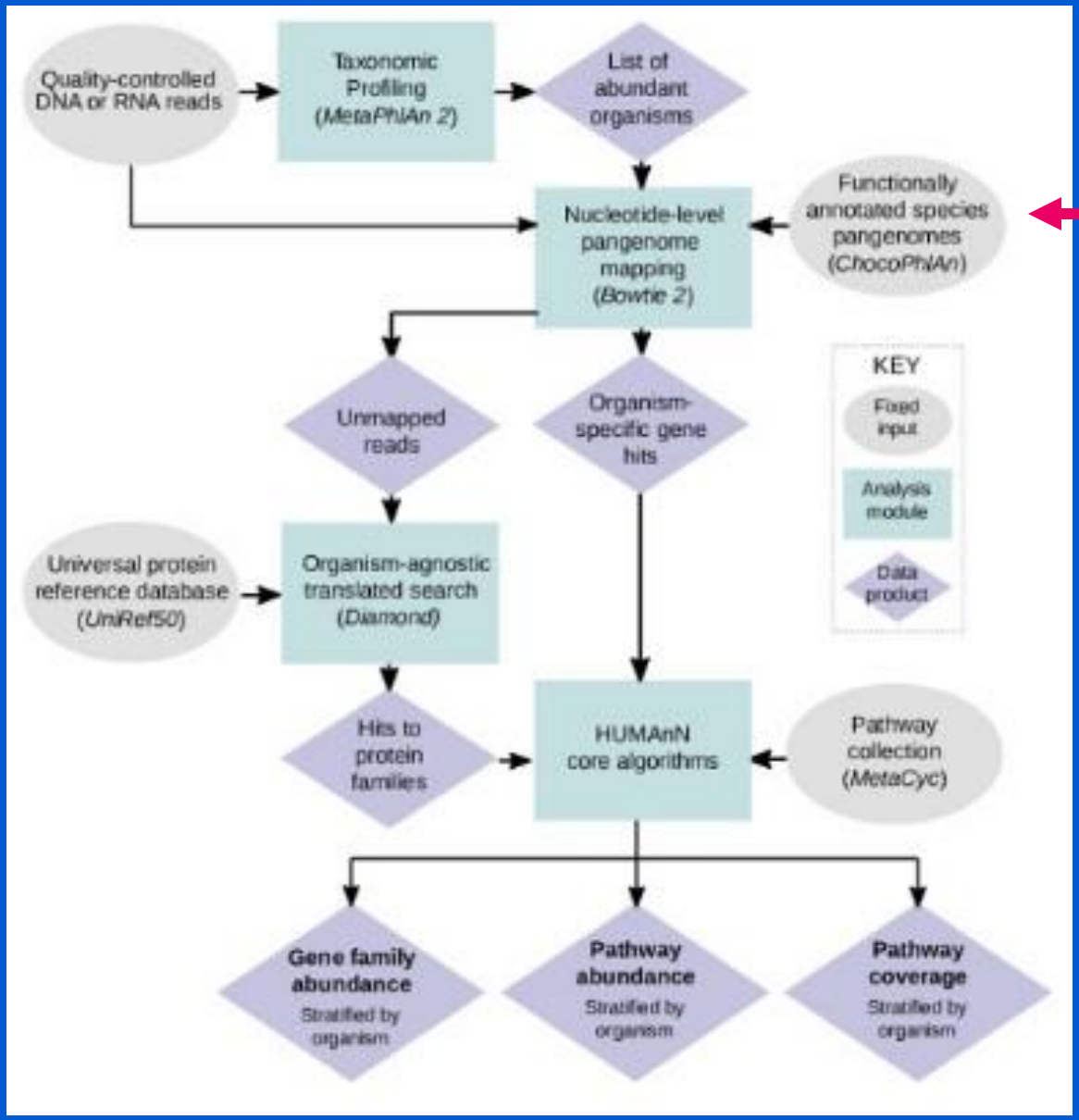
Functional annotation

- **Probably the most straightforward approach is with HUMAnN2**
- **“What are the microbes in my community of interest capable of doing?”**

[1]Abubucker *Plos Comput Biol* 2012;8:e1002358

http://huttenhower.sph.harvard.edu/humann2

HUMAnN2 workflow

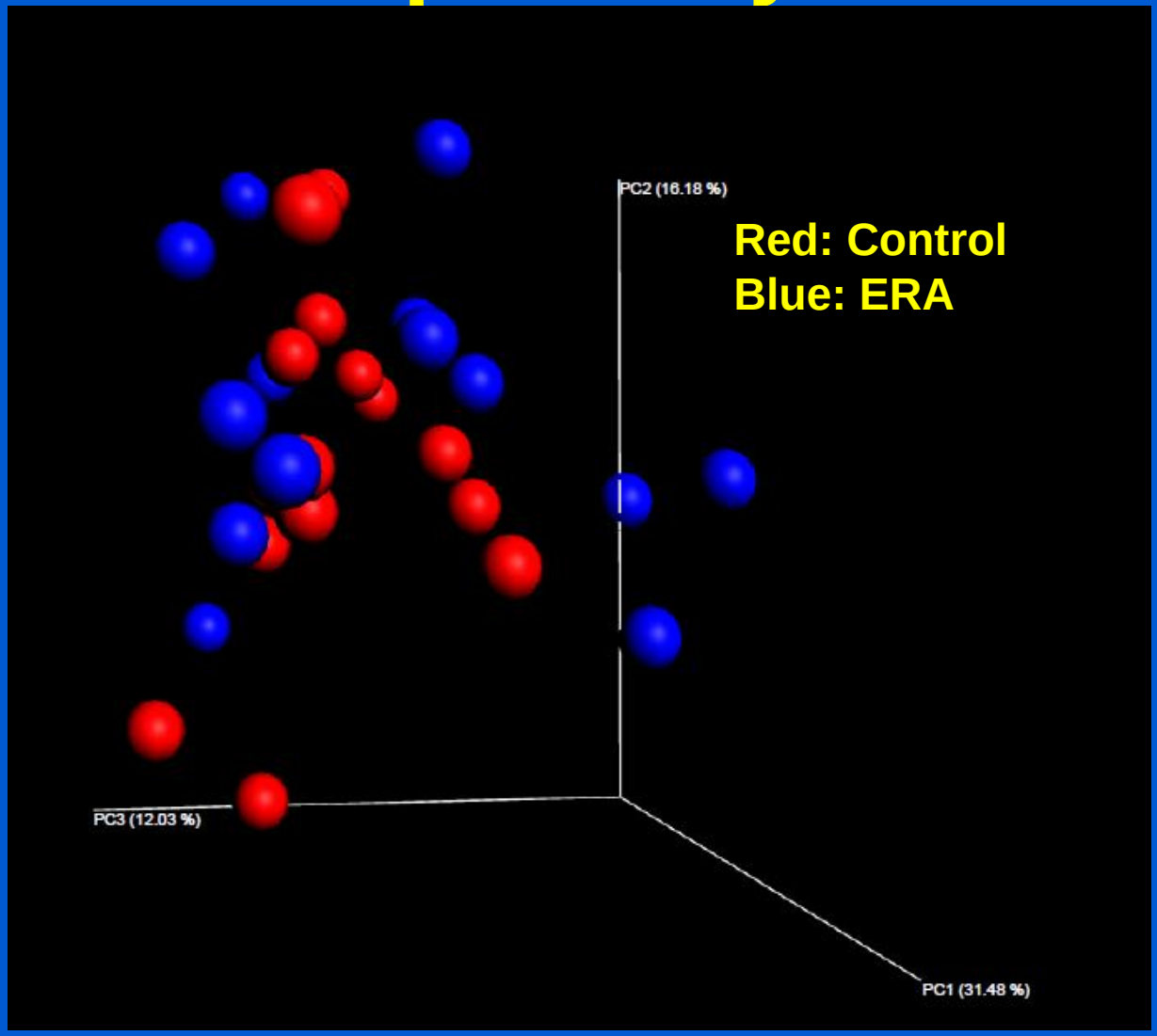


NCBI genomes

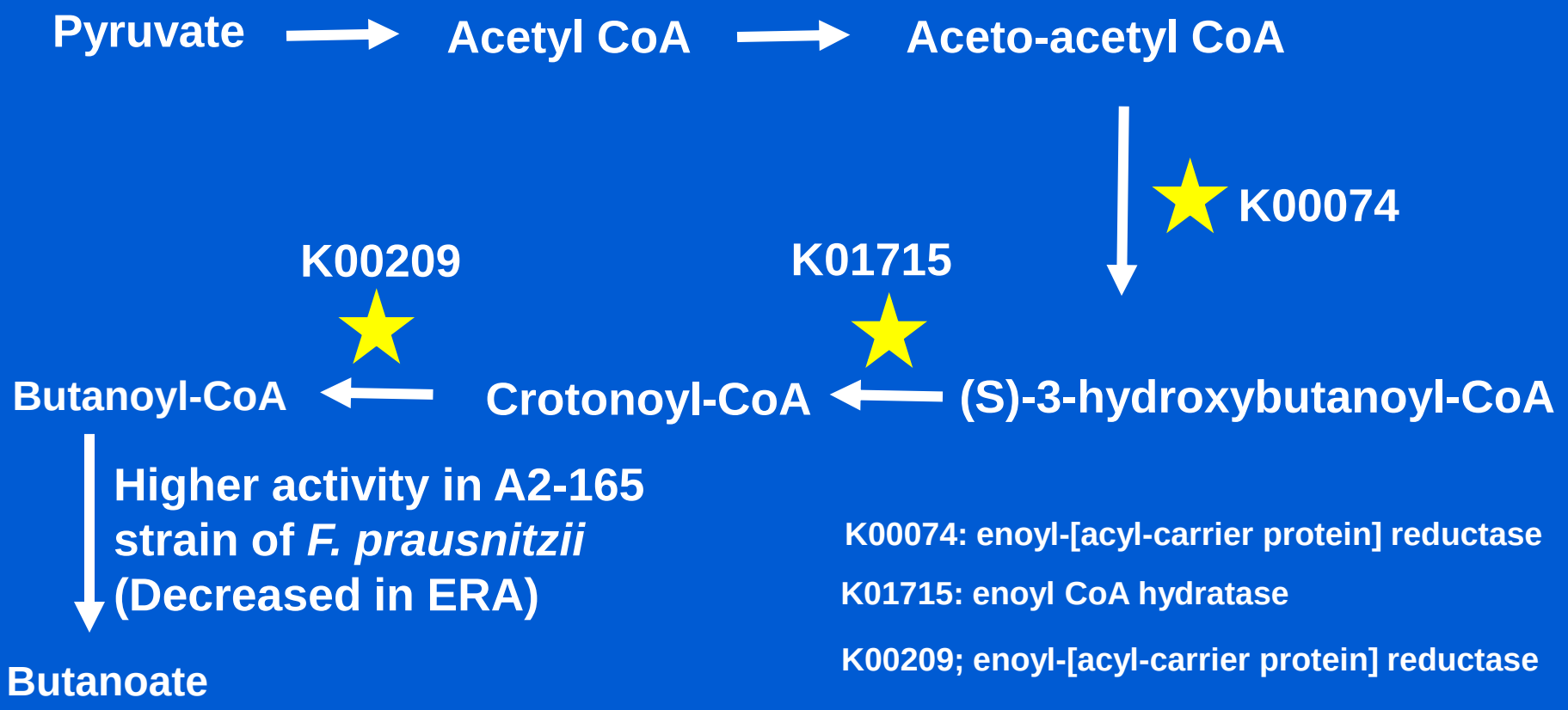
Performed whole genome sequencing of fecal microbiota from ERA patients and controls

- 14 ERA patients and 14 healthy controls
- Subset of newly diagnosed patients used for 16S analysis
- Custom quality control pipeline
- Analysis with HUMAnN2

No obvious global differences at the pathway level



Butanoate metabolism



K00074: enoyl-[acyl-carrier protein] reductase

K01715: enoyl CoA hydratase

K00209; enoyl-[acyl-carrier protein] reductase

Kanehisa *Nucleic Acids Res* 2016;44:D457

Summary

- **Study confirms association of F. *prausnitzii* in children with ERA**
  - **Findings are specific to the A2-165 strain**
- **Patients with ERA may have decreased ability to make butyrate**

Outline

- Where to sample
- Sequencing
- Data analysis
- Next steps

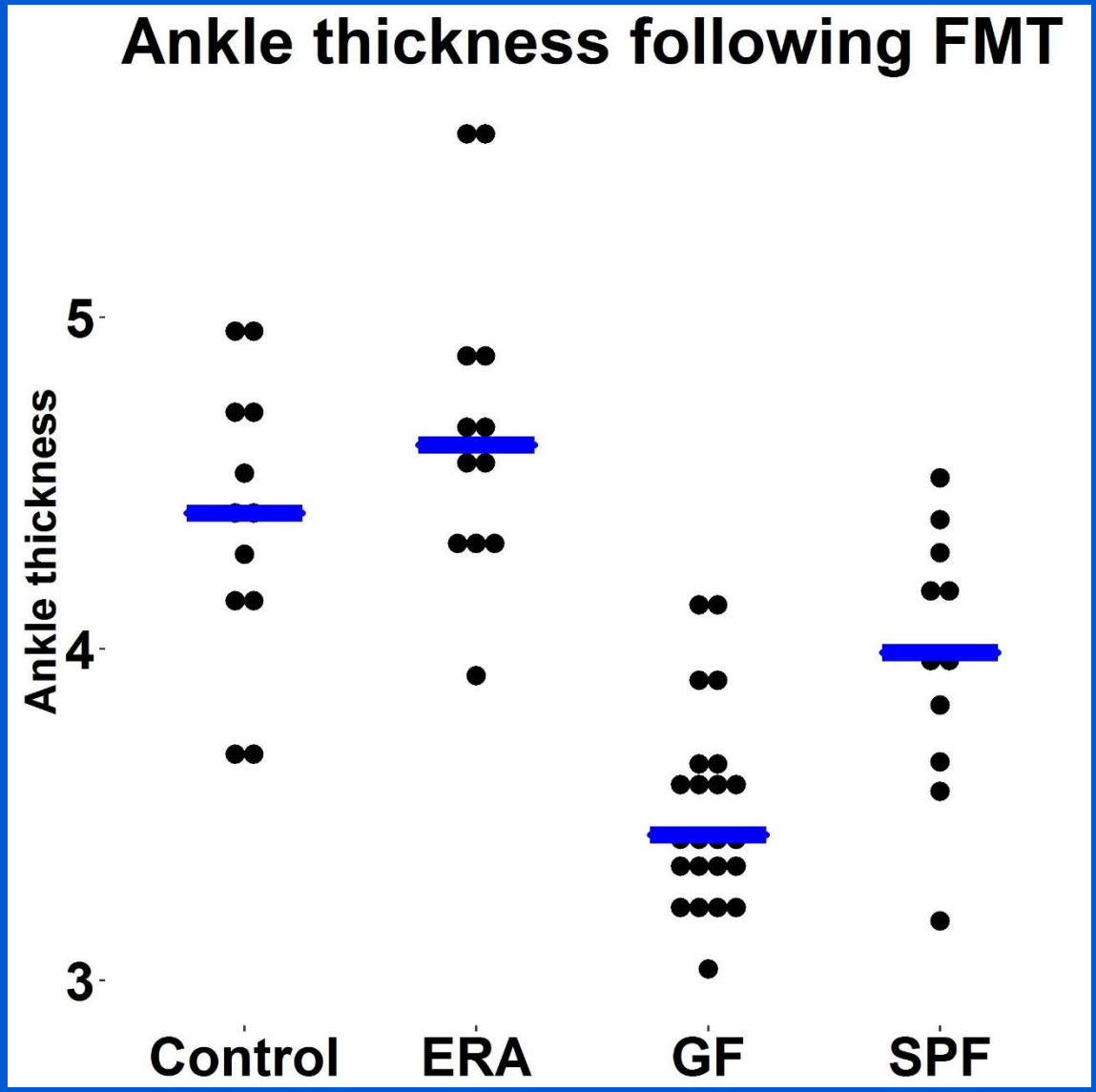
Assessing the role of the microbiota: FMT into mice

- Patient and control fecal specimens were collected in media that permits live recovery of organisms and frozen in glycerol
- Transferred to germ-free K/BxN mice
  - Arthritis model
  - Dependent on microbiota
  - Rapid onset
- Studied 3 – 4 weeks after transfer

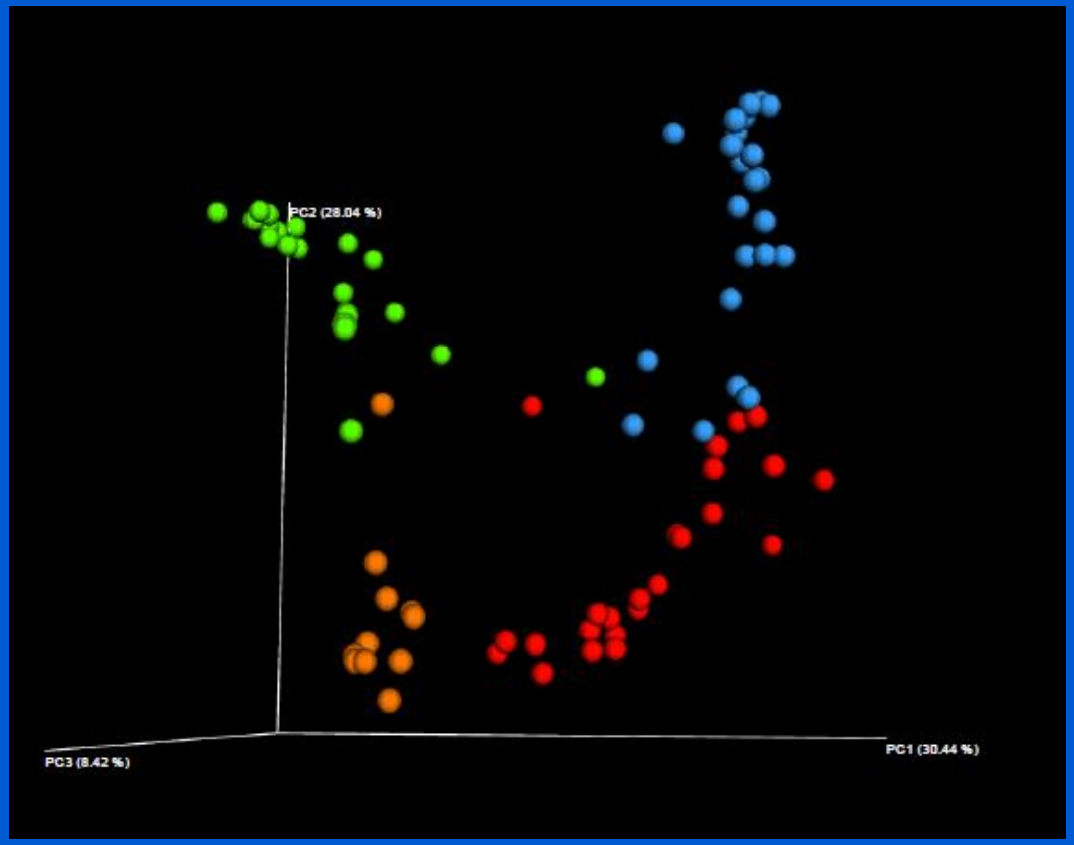
Groups

- Human donors
  - ERA, n = 12
  - Controls, n = 12
- Mouse recipients
  - ERA, n = 12
  - Controls, n = 12
- Germ-free (GF)
- Specific pathogen free (SPF)
  - Transferred from GF facility to SPF facility

Ankle swelling after FMT

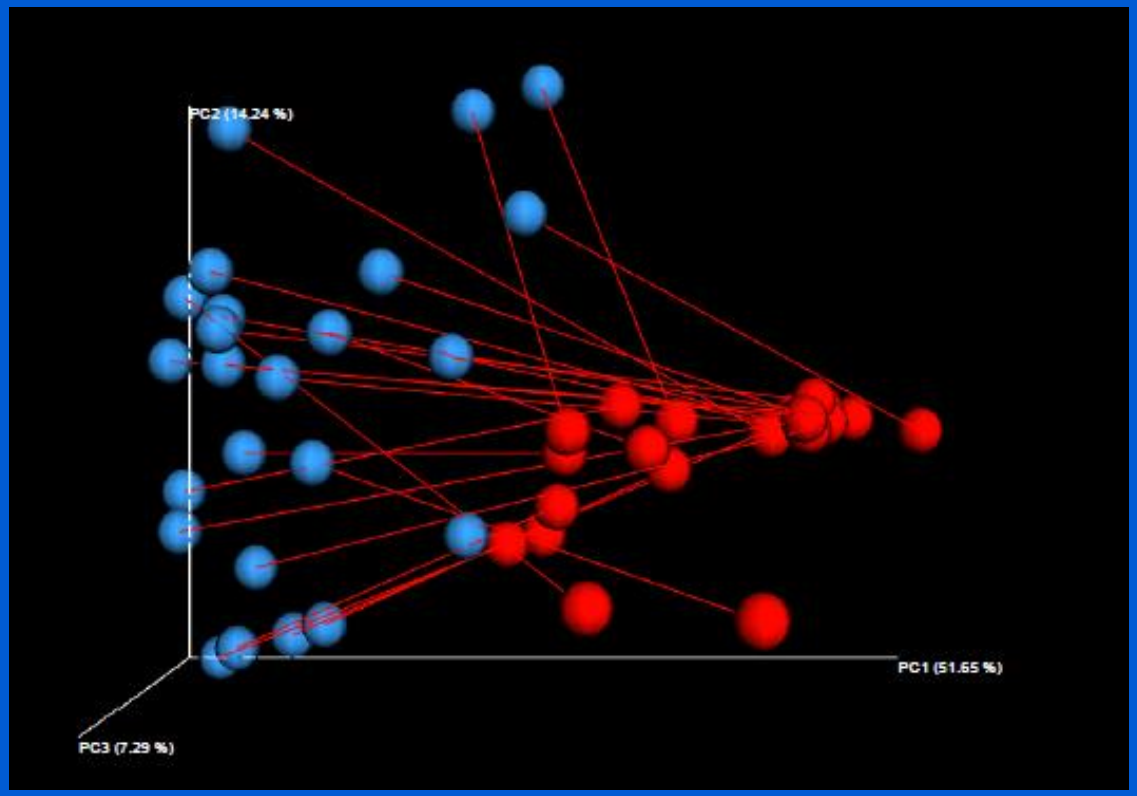


Incomplete humanization



Green: GF mice
Blue: Transplanted mice
Red: Humans
Orange: SPF mice

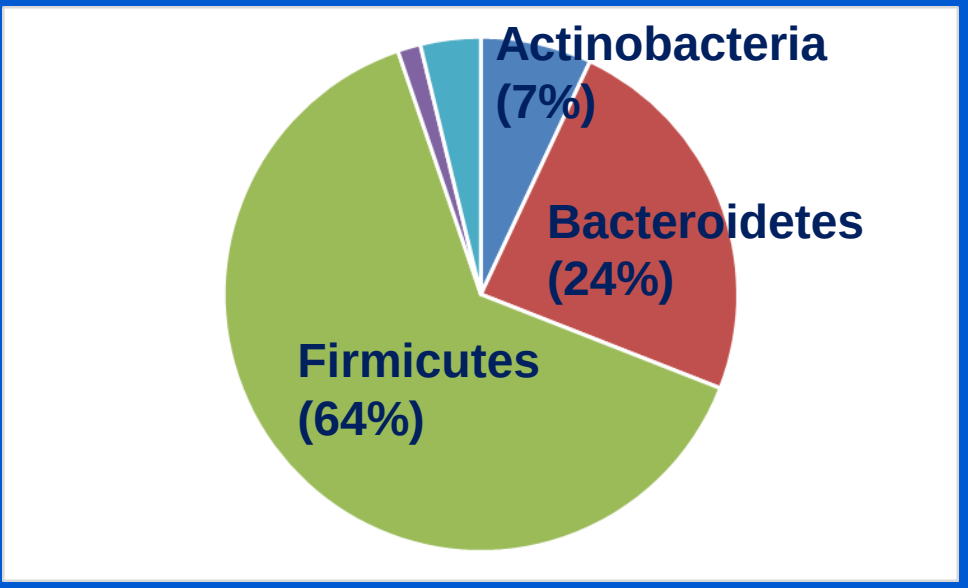
No clustering between individual patient/donor dyads



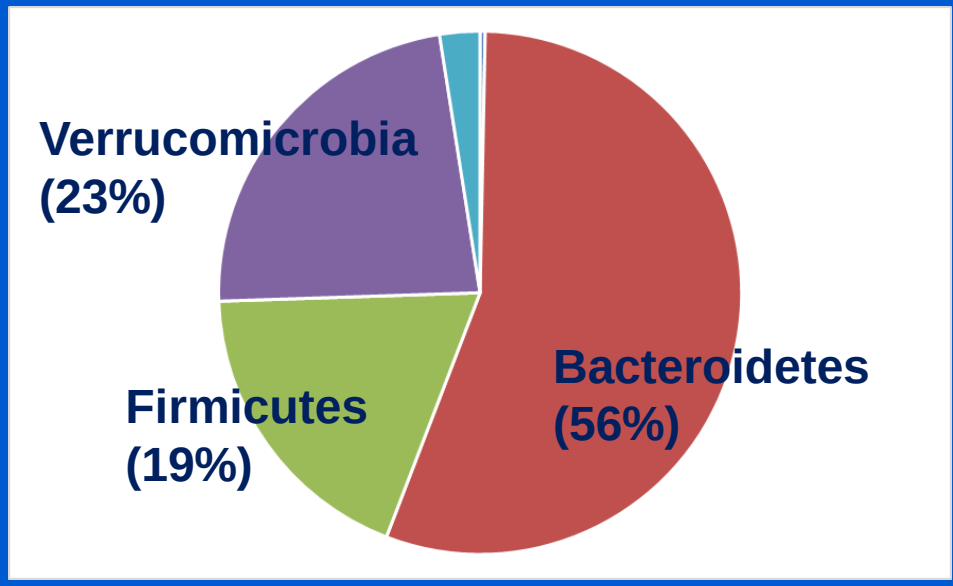
Red: Humans
Blue: Transplanted mice

Taxonomy: Phylum level

Humans

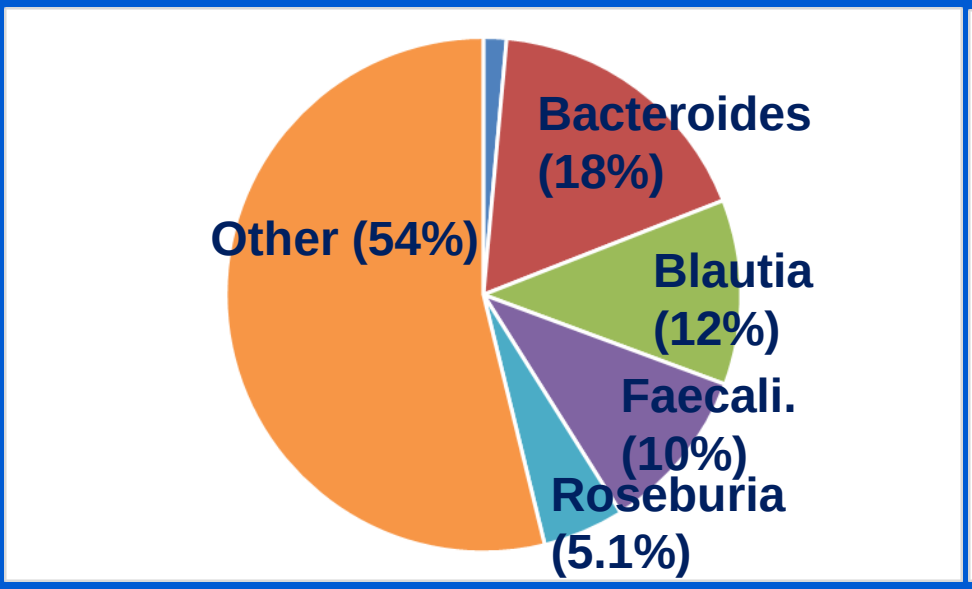


Mice

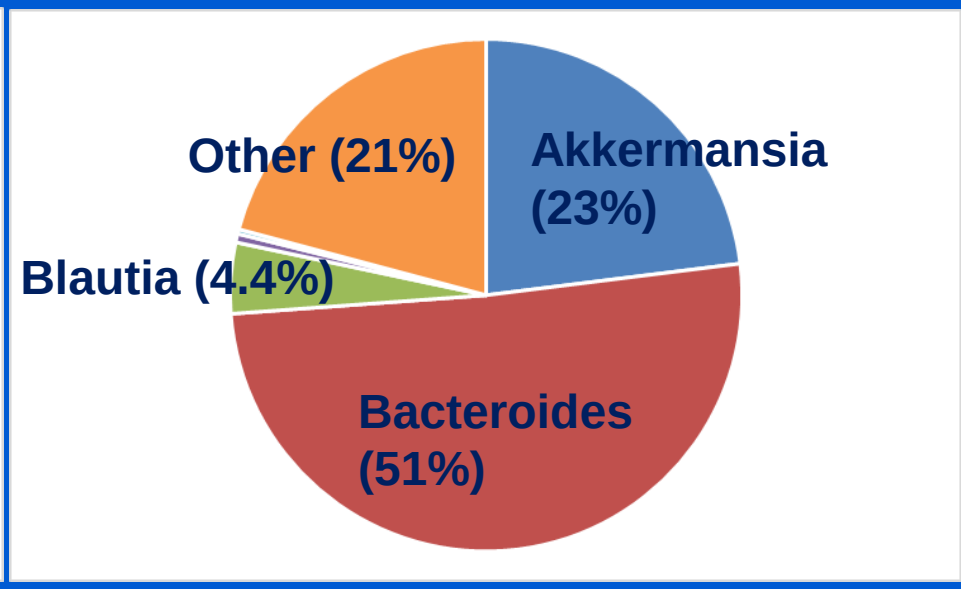


Taxonomy: Genus level

Humans



Mice



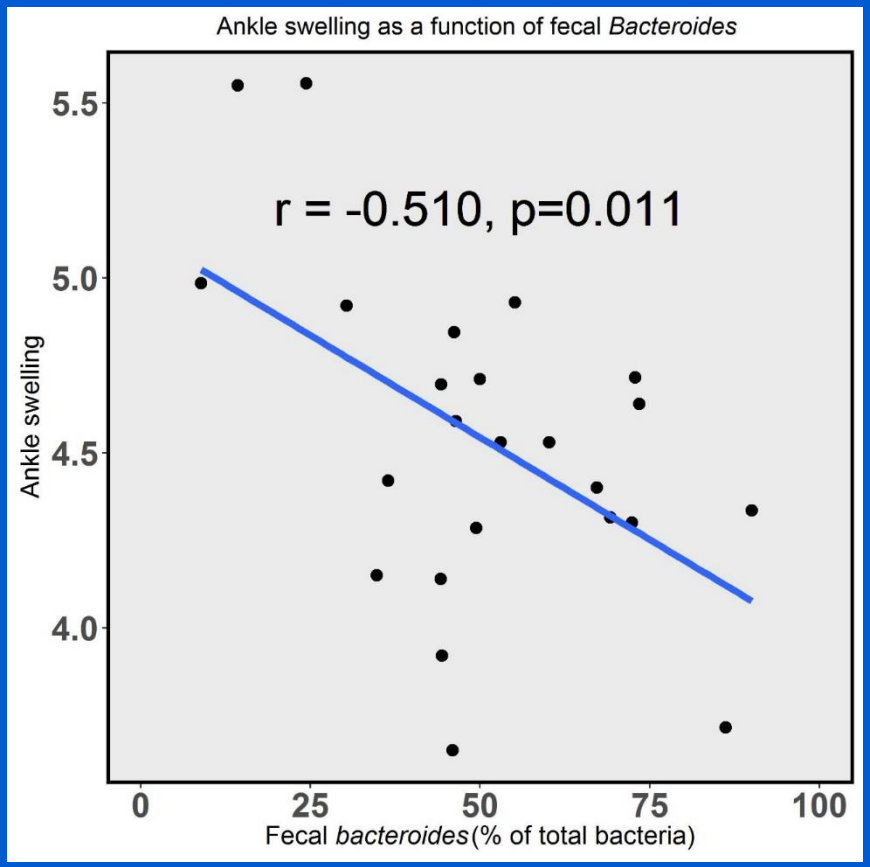
Genera present in ≥ 5% in either group

Relationship between microbiota and arthritis, among the transplanted mice

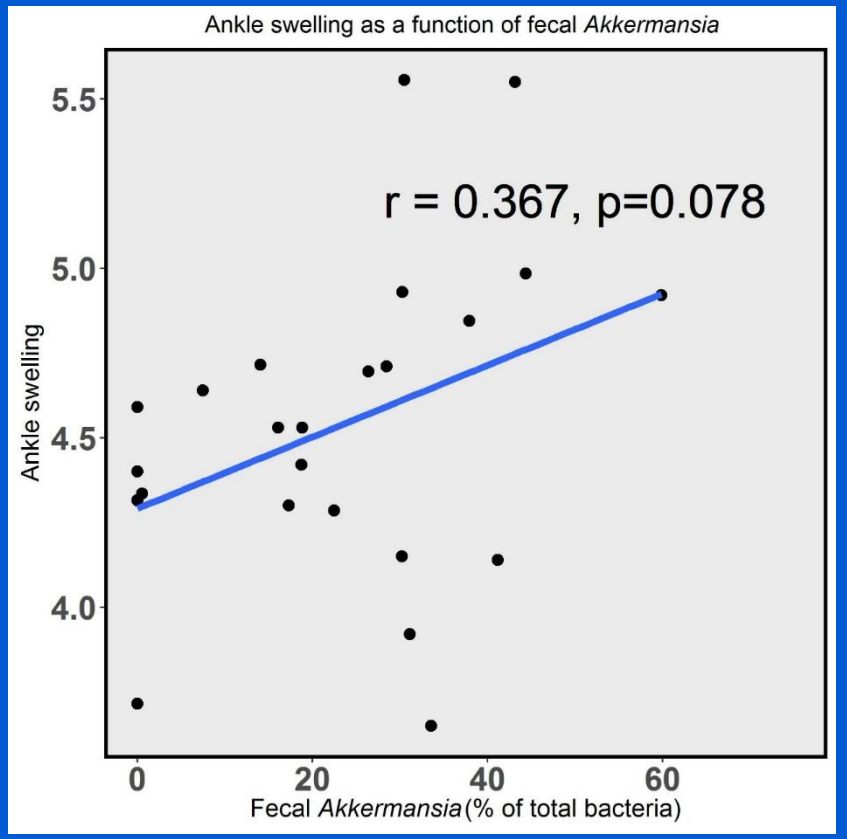
- Ran the Adonis test, which looks at whether a continuous predictor variable (in this case, ankle swelling) predicts the structure of the microbiota
- It did to some extent; $R2 = 0.185$, $p = 0.018$
- Which bacteria were driving this?

Association of common taxa with ankle swelling

Bacteroides



Akkermansia

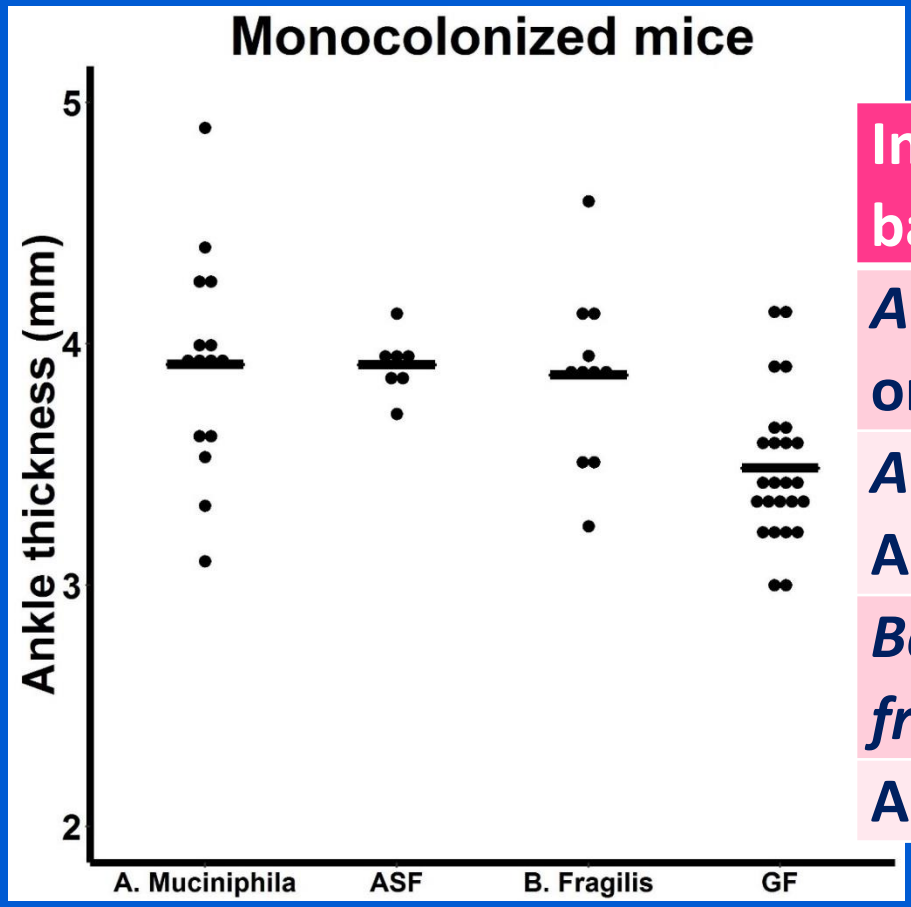


A. muciniphila may be pro- or anti-inflammatory in arthritis

- Given its name due to ability to degrade intestinal mucin
  - Potential for increased permeability
- Associated with disease in an animal model of spondyloarthritis[1]
- Decreased abundance in adults with psoriatic arthritis[2]

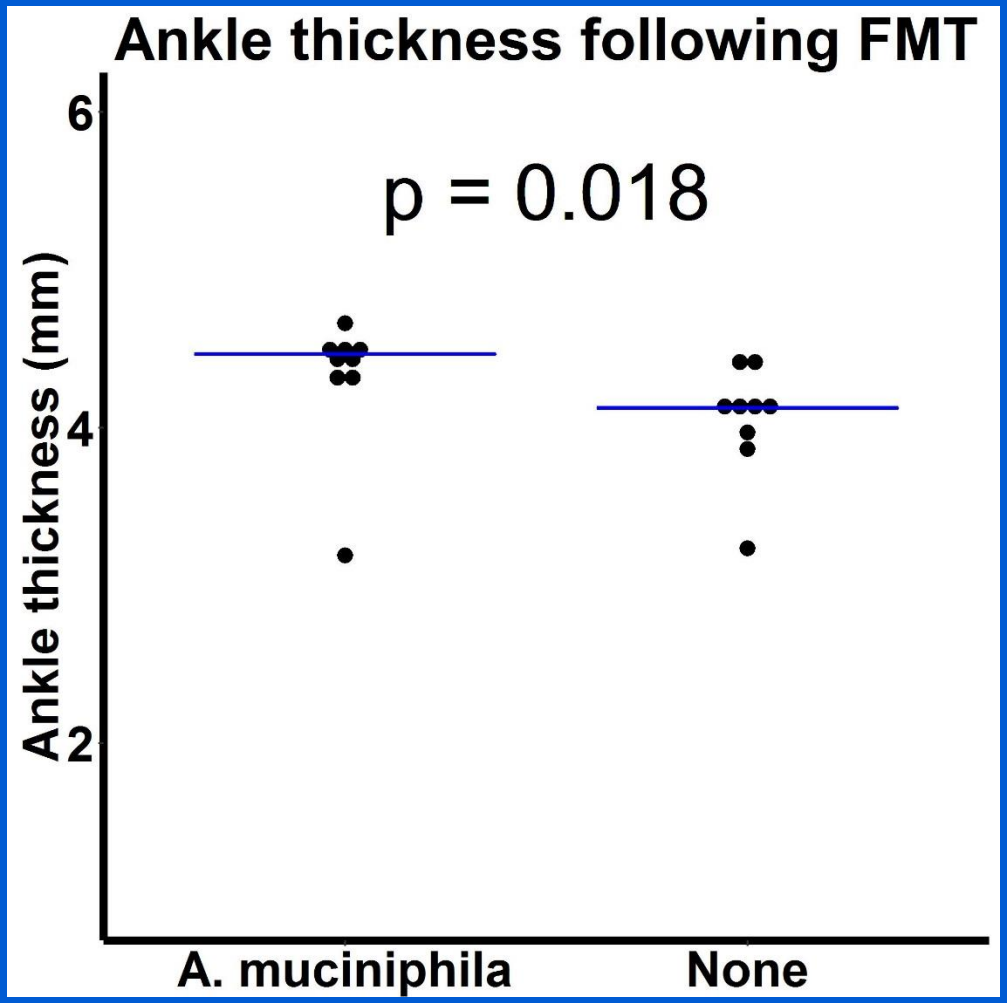
[1]Asquith, *Arth Rheum* 2016;68:2151
[2]Scher, *Arth Rheum* 2015;67:128

Monocolonization with *A. muciniphila* did not affect ankle swelling



Inoculated bacteria	Abundance of *Akkermansia* (%)
A. muciniphila only	88 ± 7.3
A. muciniphila + ASF	39 ± 14
Bacteroides fragilis + ASF	0.3 ± 0.5
ASF alone	0.3 ± 0.6

Adding *A. muciniphila* to mice receiving normal microbiota did



Summary

- ***Akkermansia*** **may facilitate arthritis, but does not appear to be directly pathogenic**
- **May relate to ability of this organism to thrive on intestinal mucin**
- **In progress: studies to evaluate for increased invasiveness of other organisms in mice receiving *Akkermansia***

Acknowledgements

UAB

Randy Cron
Wayne Duck
Peter Eipers
Charles Elson
Ranjit Kumar
Elliot Lefkowitz
Casey Morrow
Kathy Pierce
Trenton Schoeb

Other sites

Pamela Weiss
Jennifer Weiss
Barbara Edelheit
Peter Nigrovic
Charles Spencer
Lynn Punaro
Kenneth Schikler
Andreas Reiff

Funding

NIAMS
NIEHS
ACR
UAB
CoA
CARRA

Rheumatology Research Foundation

Advancing Treatment | Finding Cures

UAB MEDICINE
DEPARTMENT OF PEDIATRICS