The virome of the human gut: metagenomic analysis of changes associated with diet

James Lewis
Gary Wu
Frederic Bushman

Diet, Genetic Factors, and the Gut Microbiome in Crohn's Disease

University of Pennsylvania

Diet, Genetic Factors, and the Gut Microbiome in Crohn's Disease

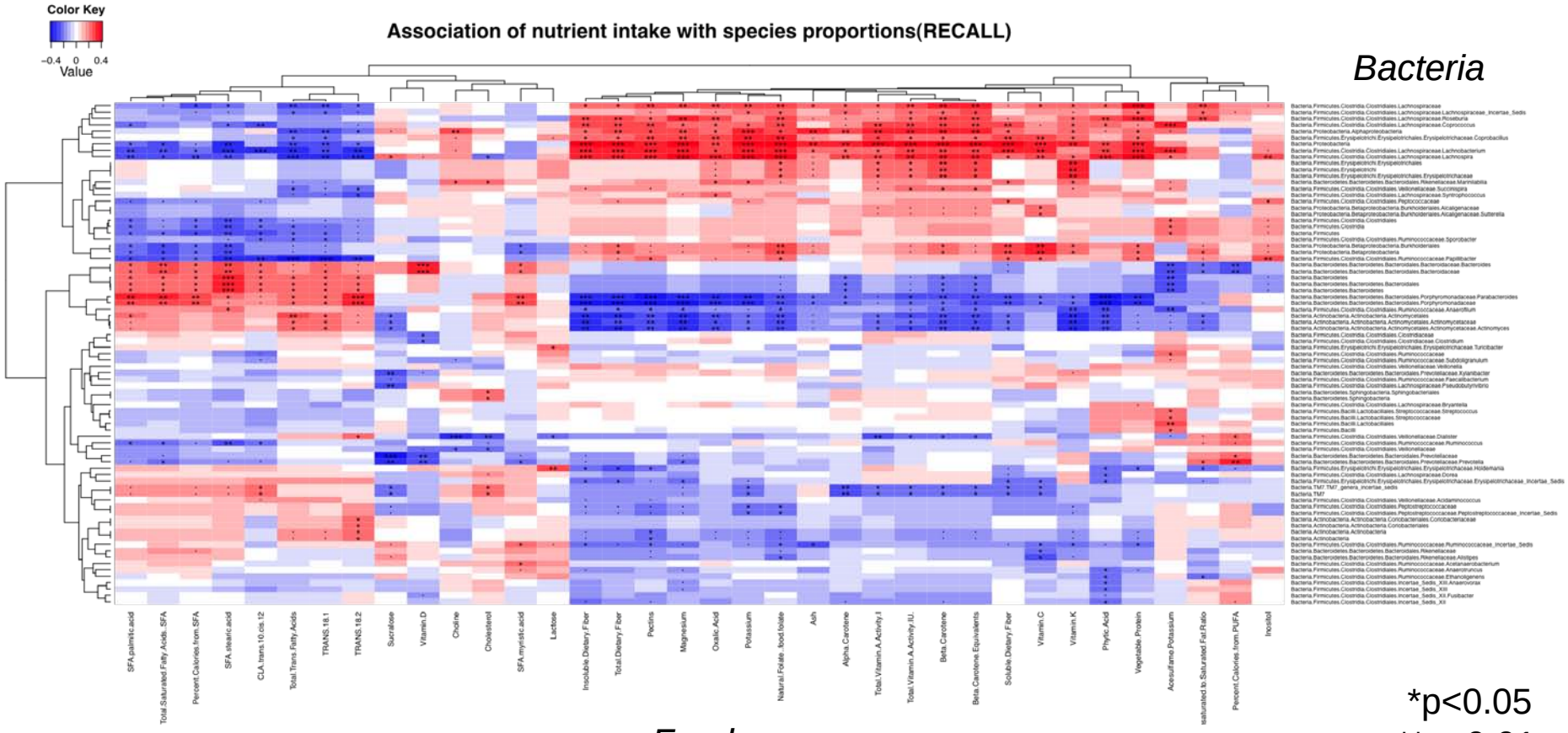
COMBO: Cross-Sectional Study of Diet and Stool Microbiome

CAFE: Controlled Feeding Experiment

Study virome under controlled feeding

COMBO

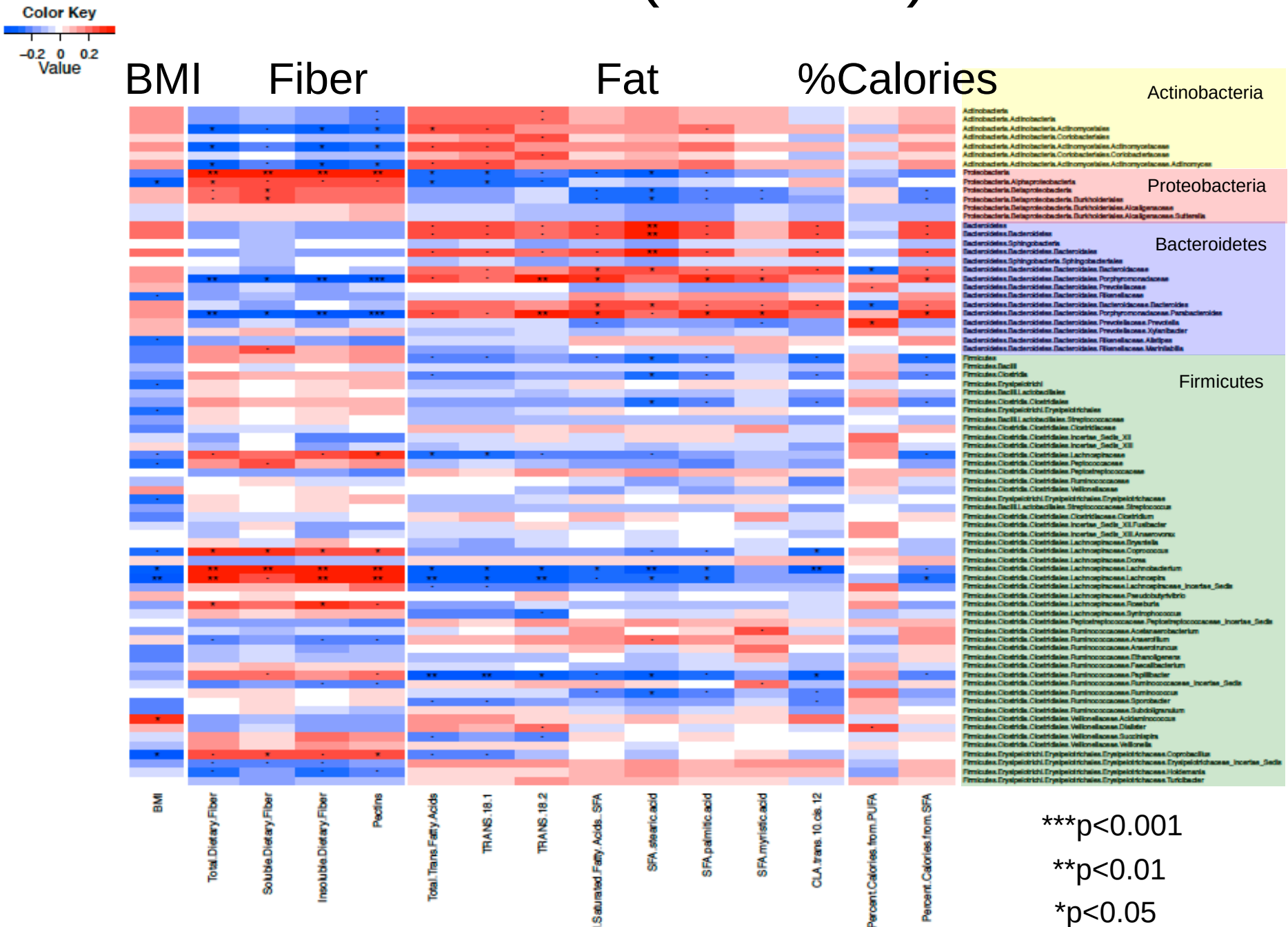
97 subjects, assess diet with food frequency questionnaire, 24 hr. recall questionnaire
454 tag sequencing of 16S regions



*p<0.05
**p<0.01
***p<0.001

diet components called at 25% FDR

COMBO (selected)



***p<0.001

**p<0.01

*p<0.05

Conclusions from COMBO

•At an FDR of 25%, ~30-40 nutrients correlated with changes in bacterial populations

•Correlations between food groups and Pyla detected, but considerable divergence among deeper taxa

•Correlated effects on microbiome among BMI, dietary fat, and percent calories from saturated fatty acids

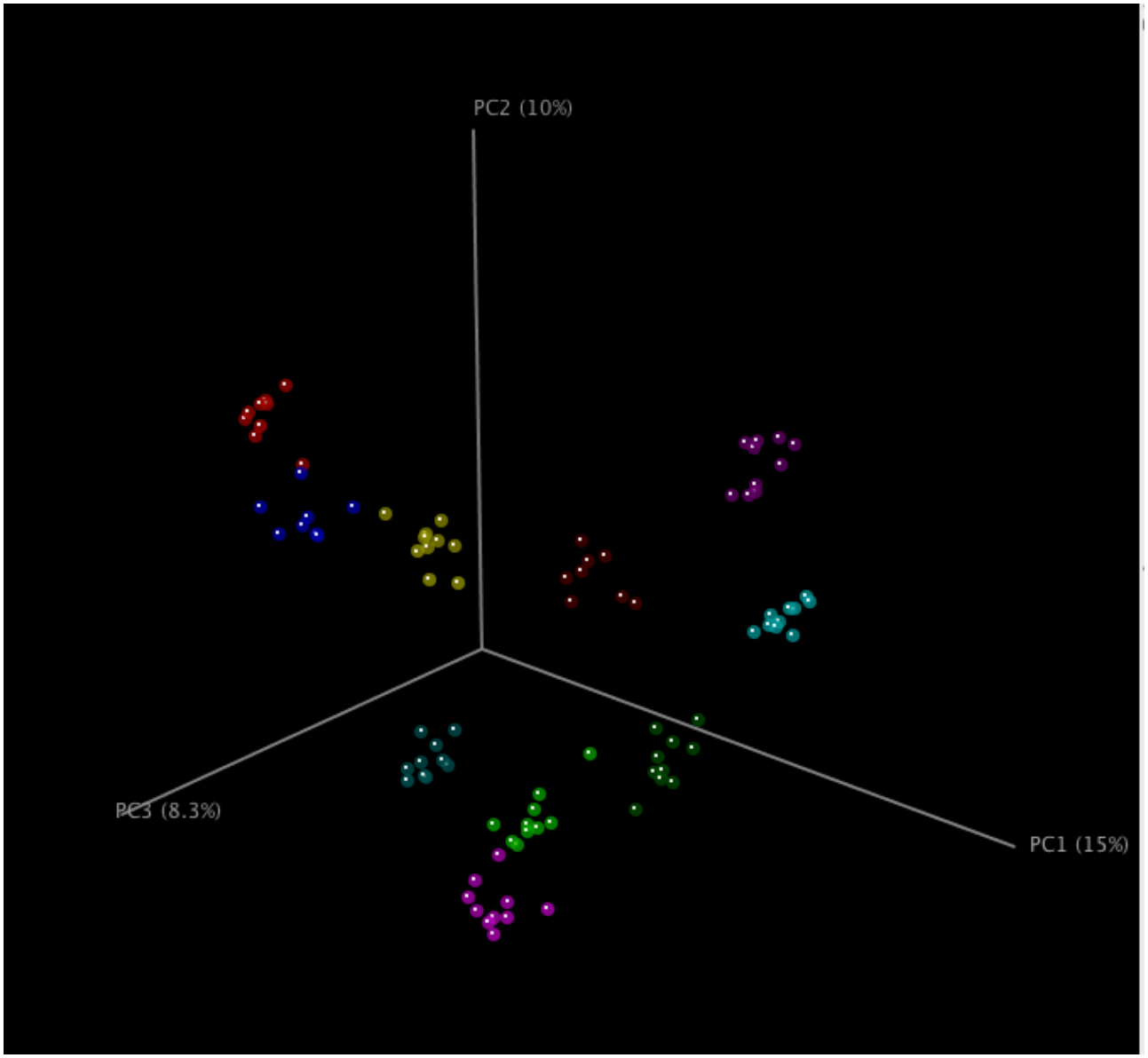
•Though effects were significant, measured dietary effects accounted for a small fraction of the total variation among subjects

Follow up in CAFÉ: 1) fat versus fiber, 2) total calories

CAFE1: Controlled Feeding Experiment

- 10 healthy volunteers
- Randomized to high fat vs. low fat diet
- 10 day inpatient stay with same meals each day
- Caloric intake adjusted to maintain current weight
- Daily stool sample collection
- Rectal biopsies on days 1 and 10
- Sitz marker study to assess transit time
- Sequencing: 16S tags, shot-gun metagenomic sequencing of total DNA and viral DNA, targeted analysis of Archaea and Eukarya

Longitudinal analysis of microbiome under controlled feeding



Unweighted Unifrac

Graphic from QIIME

Conclusions from CAFÉ 1 and 2

•Inter-individual variation predominates.

•Bacterial populations change within 24 hours of initiating controlled diet.

•High fat versus fiber has detectable effect (CAFE1), increased calories has detectable effect (CAFE2).

•Considerable longitudinal drift during stay in hospital in all groups. Specific vitamins? Hospital environment?

Approaches to virome analysis

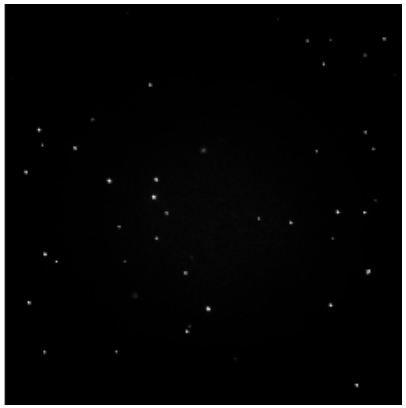
Multiple possible goals:

- Characterize overall viral communities
- Hunt for new viruses linked to disease
- Characterize populations of a specific virus

Viral Analysis on CAFE1 samples

- Purification of viral particles: filtration, CsCl gradient, DNAse digestion.
- Quantitative recovery of phage λ analyzed as a control
- Greatly reduced 16S rDNA
- Multiple displacement amplification
- Shot-gun sequencing, 454 Titanium, 992,309 reads, median length 380 nt

PID	Day 1	2	3	4	5	6	7	8	Diet
2011	X							X	Low Fat
2012		X					X	X	Low Fat
2016	X	X					X	X	High Fat
2020	X	X					X	X	Low Fat
2019	X						X		High Fat
1013	X								Ad lib



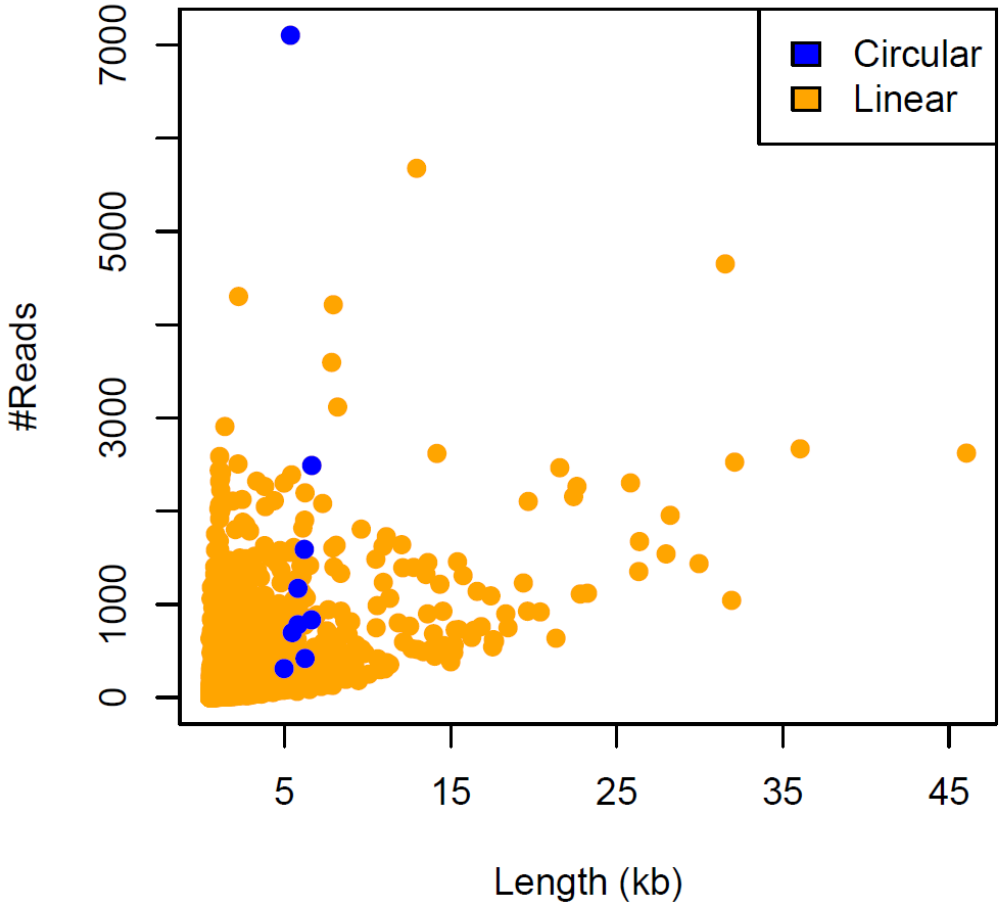
Sybr gold staining

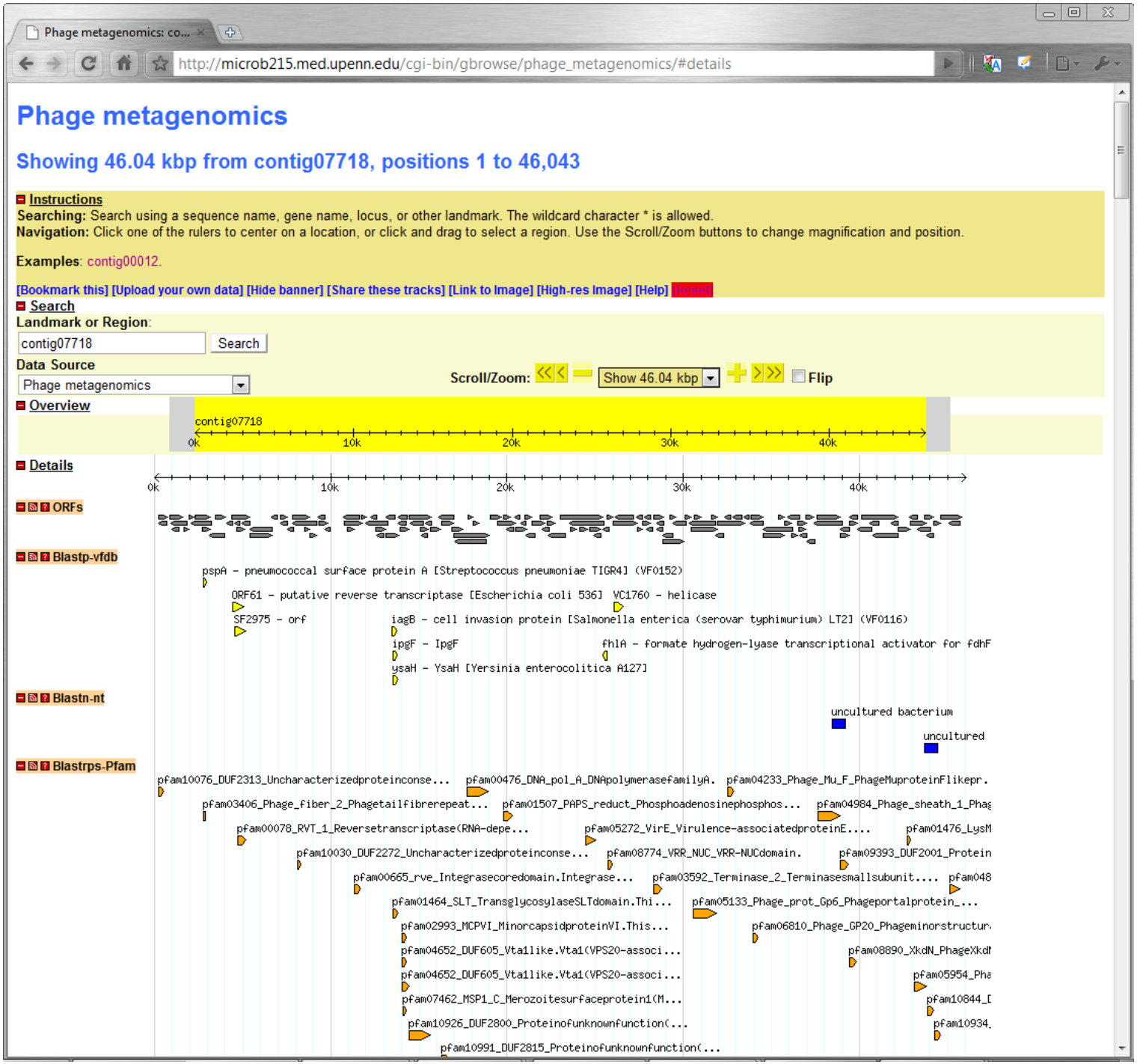
10^{10} phage per gram of stool

Sam Minot et al.

Assembly of viral sequence reads

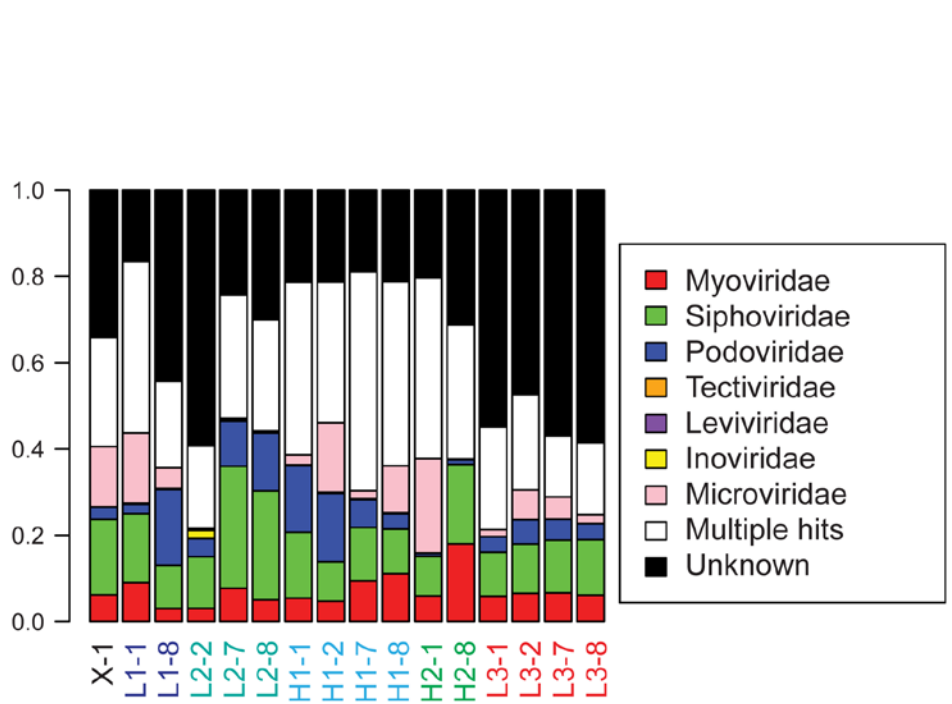
- Newbler assembler
  40bp overlap
  90% sequence identity
- 7,175 contigs >500 bp
- 86.6% of reads in contigs
- Custom code to allow circular assembly
- PHACCS: median species richness 44 (range=19-785)

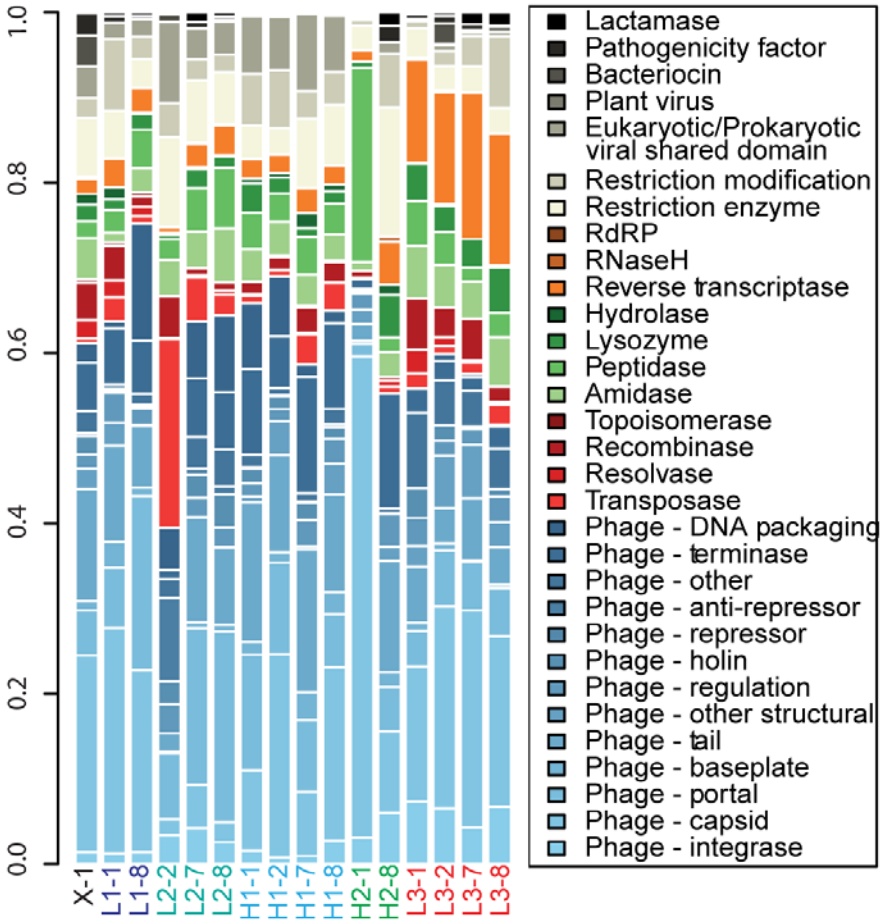



Phage metagenomics: co...
http://microb215.med.upenn.edu/cgi-bin/gbrowse/phage_metagenomics/#details
Phage metagenomics
Showing 46.04 kbp from contig07718, positions 1 to 46,043
Instructions
Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.
Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.
Examples: contig00012.
[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [High-res Image] [Help] Reset
Search
Landmark or Region:
contig07718
Search
Data Source
Phage metagenomics
Scroll/Zoom:
Show 46.04 kbp
Flip
Overview
contig07718
0k
10k
20k
30k
40k
Details
ORFs
Blastp-vfdb
pspA - pneumococcal surface protein A [Streptococcus pneumoniae TIGR4] (VF0152)
ORF61 - putative reverse transcriptase [Escherichia coli 536]
VC1760 - helicase
SF2975 - orf
iagB - cell invasion protein [Salmonella enterica (serovar typhimurium) LT2] (VF0116)
ipgF - IpgF
fhlA - formate hydrogen-lyase transcriptional activator for fdhF
ysaH - YsaH [Yersinia enterocolitica A127]
Blastn-nt
uncultured bacterium
uncultured
Blastrps-Pfam
pfam10076_DUF2313_Uncharacterizedproteinconse...
pfam00476_DNA_pol_A_DNApolymerasefamilyA.
pfam04233_Phage_Mu_F_PhageMuproteinFlikepr.
pfam03406_Phage_fiber_2_Phagetailfibrerepeat...
pfam01507_PAPS_reduct_Phosphoadenosinephosphos...
pfam04984_Phage_sheath_1_Phag
pfam00078_RVT_1_Reversetranscriptase(RNA-depe...
pfam05272_VirE_Virulence-associatedproteinE....
pfam01476_LysM
pfam10030_DUF2272_Uncharacterizedproteinconse...
pfam08774_VRR_NUC_VRR-NUCdomain.
pfam09393_DUF2001_Protein
pfam00665_rve_Integrasecoredomain.Integrase...
pfam03592_Terminase_2_Terminasesmallsubunit....
pfam048
pfam01464_SLT_TransglycosylaseSLTdomain.Thi...
pfam05133_Phage_prot_Gp6_Phageportalprotein_...
pfam02993_MCPVI_MinorcapsidproteinVI.This...
pfam06810_Phage_GP20_Phageminorstructur.
pfam04652_DUF605_Vta1like.Vta1(VPS20-associ...
pfam08890_XkdN_PhageXkdN
pfam04652_DUF605_Vta1like.Vta1(VPS20-associ...
pfam05954_Pha
pfam07462_MSP1_C_Merozoitesurfaceprotein1(M...
pfam10844_D
pfam10926_DUF2800_Proteinofunknownfunction(...
pfam10934_
pfam10991_DUF2815_Proteinofunknownfunction(...


Milk Duds
Milk Duds
CHARLESTON CHEWS
1.79
1.69
1.29
RAISINETS
RAISINETS
SOUR
Kids
1.79
1.02
Mike Ike
Mike Ike
Mike Ike
Mike Ike
1.79
1.02

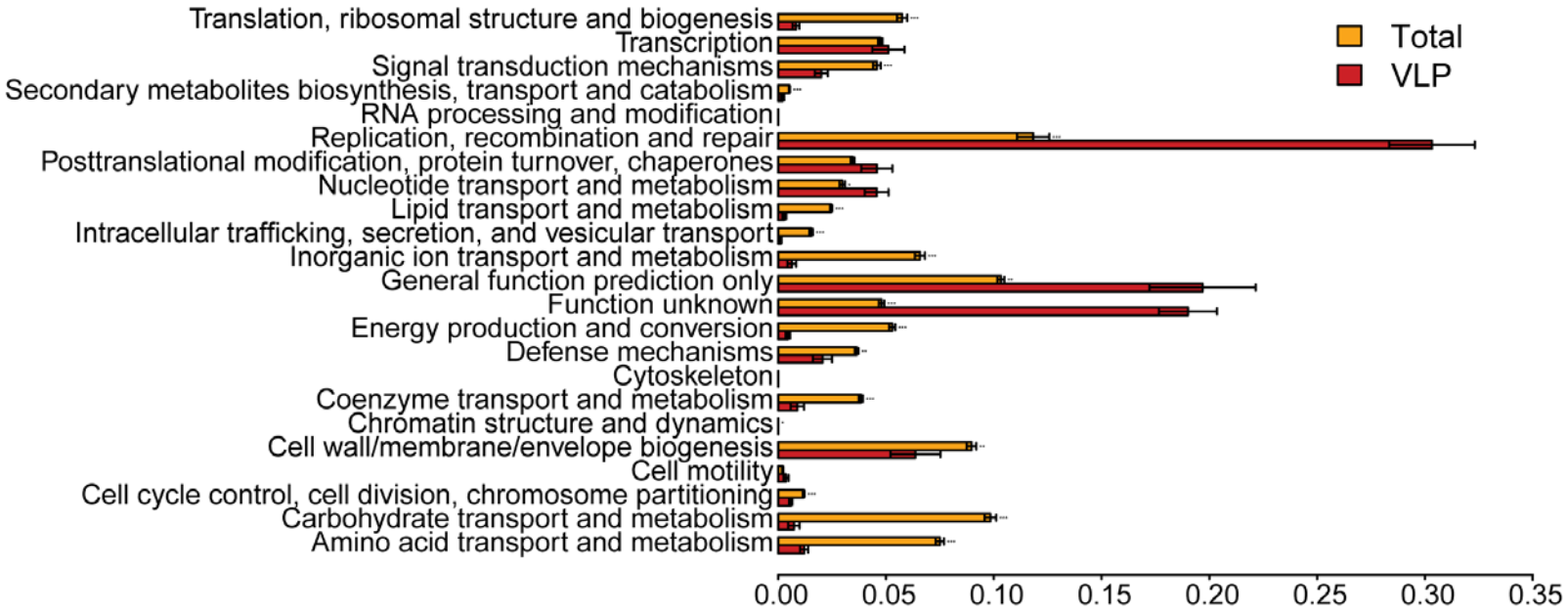
Gut virome characterization



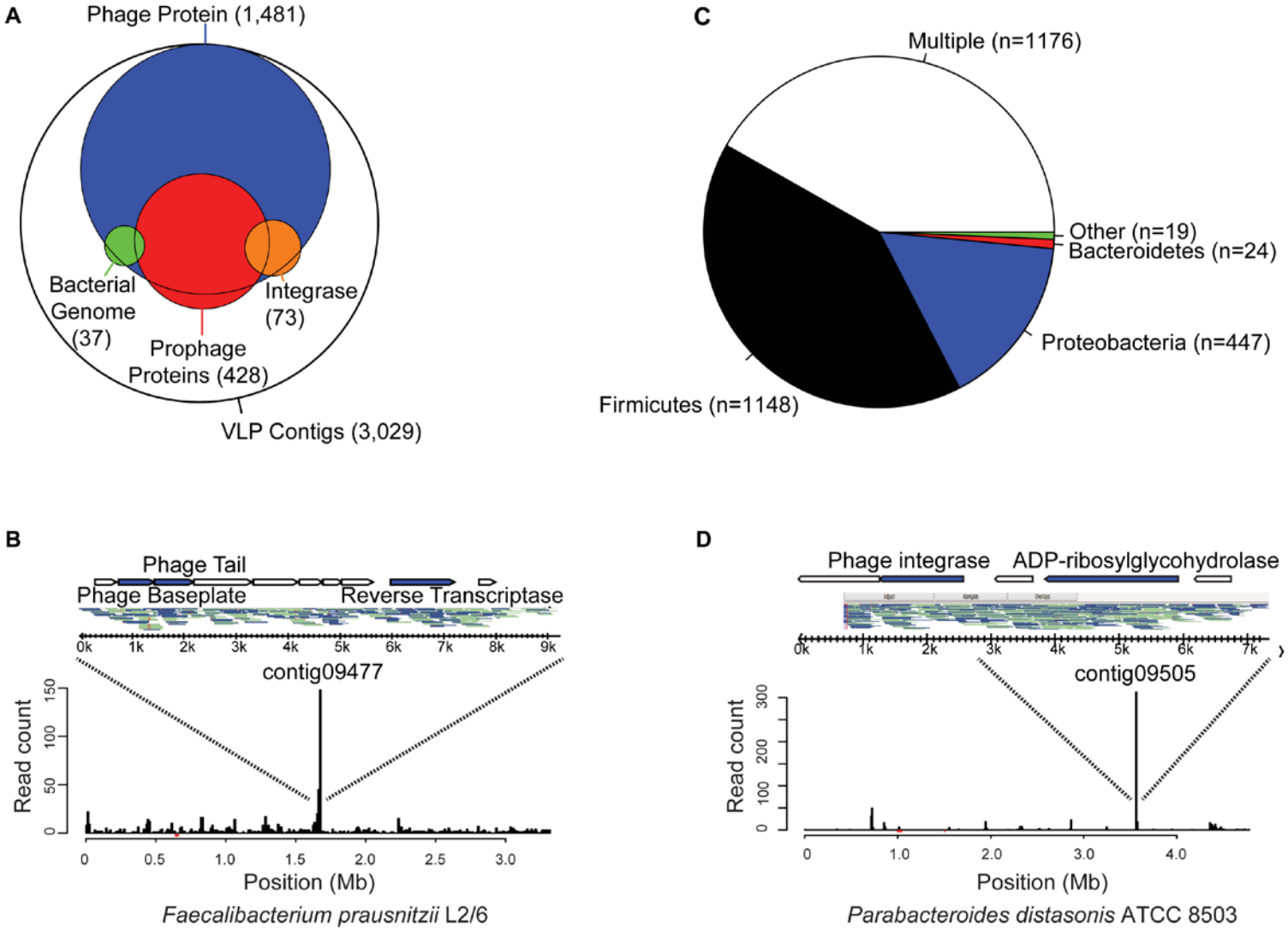


Also found 22 CRISPR arrays, one example of CRISPR spacer targeting another virus in the same individual

Comparative metagenomics: viruses are parasites



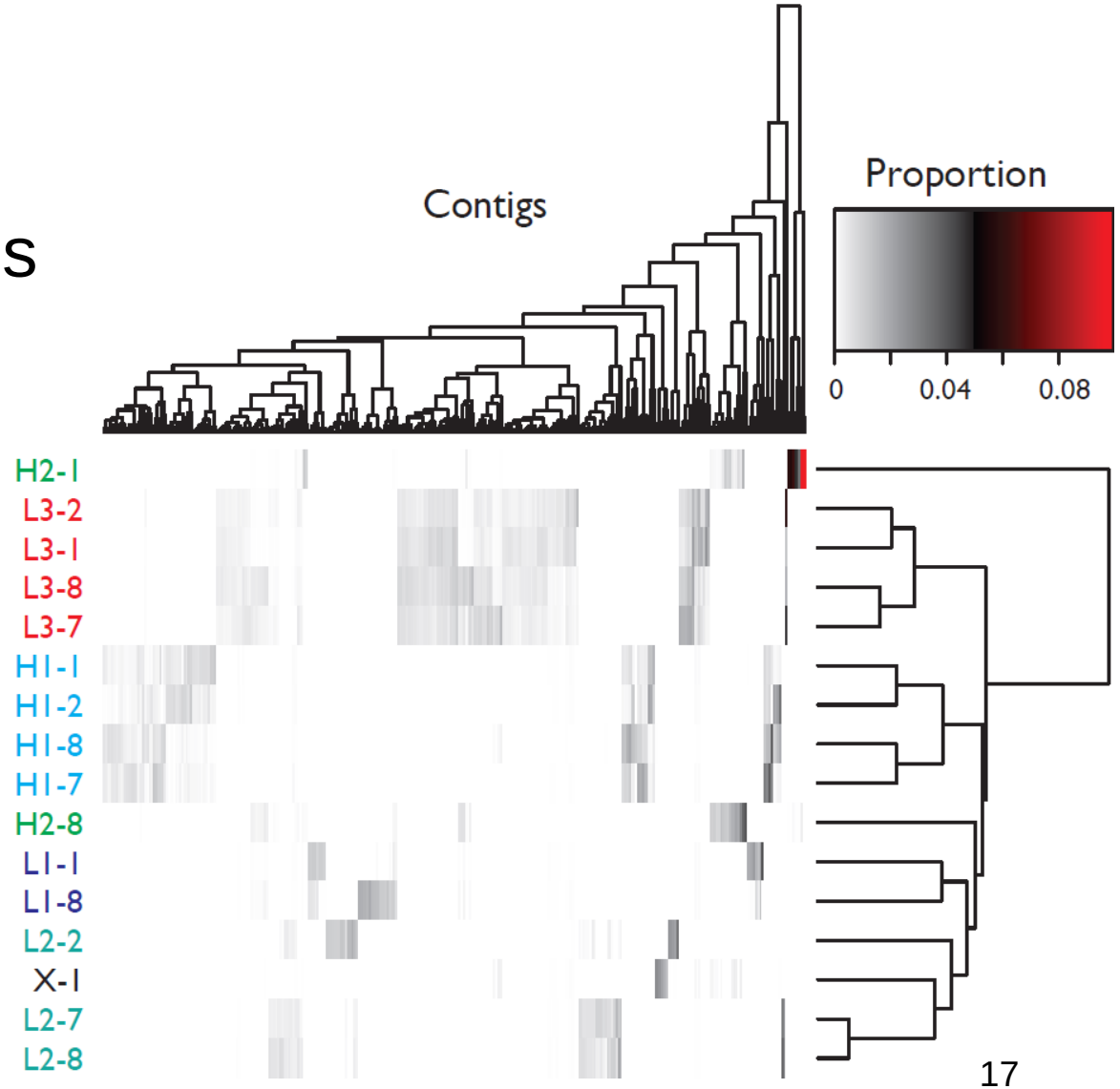
Lysogeny


A
Phage Protein (1,481)
Bacterial Genome (37)
Integrase (73)
Prophage Proteins (428)
VLP Contigs (3,029)
C
Multiple (n=1176)
Other (n=19)
Bacteroidetes (n=24)
Proteobacteria (n=447)
Firmicutes (n=1148)
B
Phage Tail
Phage Baseplate
Reverse Transcriptase
contig09477
Read count
Position (Mb)
Faecalibacterium prausnitzii L2/6
D
Phage integrase
ADP-ribosylglycohydrolase
contig09505
Read count
Position (Mb)
Parabacteroides distasonis ATCC 8503

Interpersonal variation

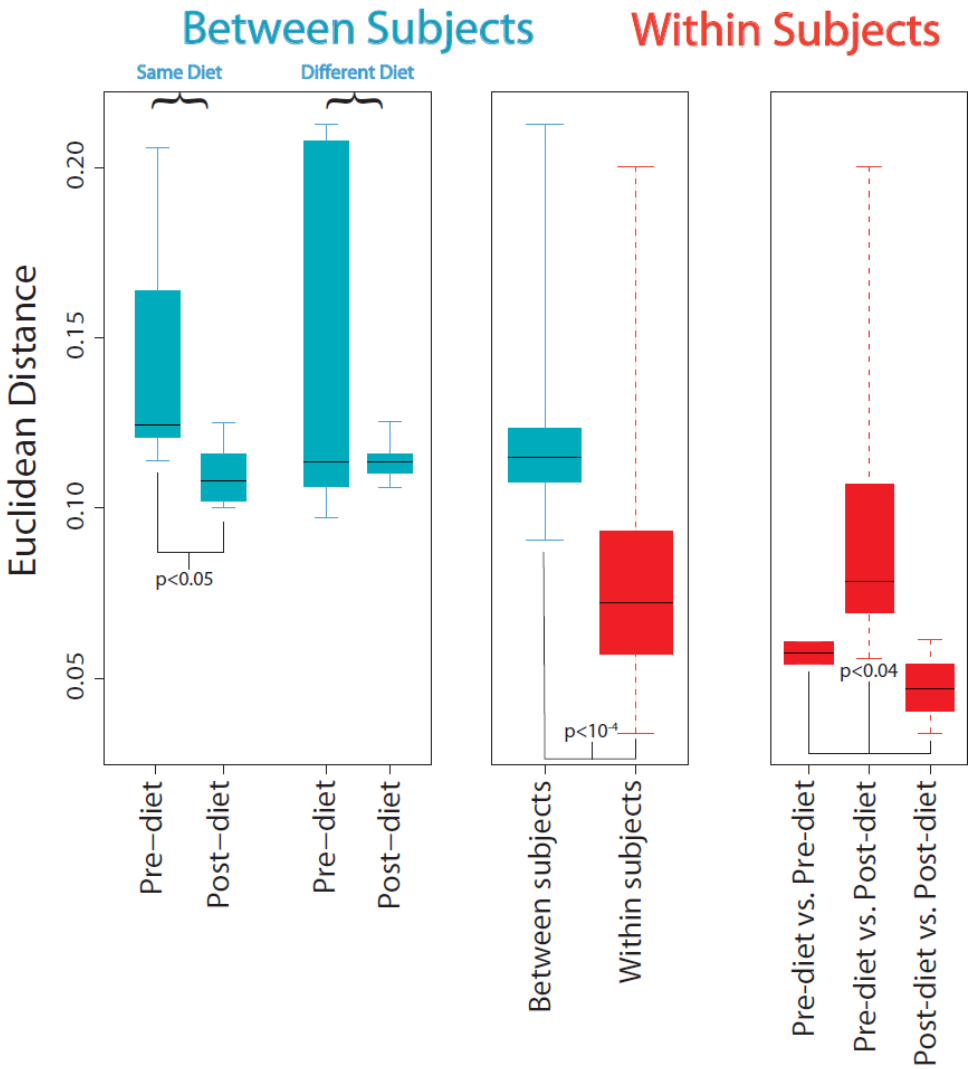
- Rows: samples
- Columns: contigs

- Clustering by individual



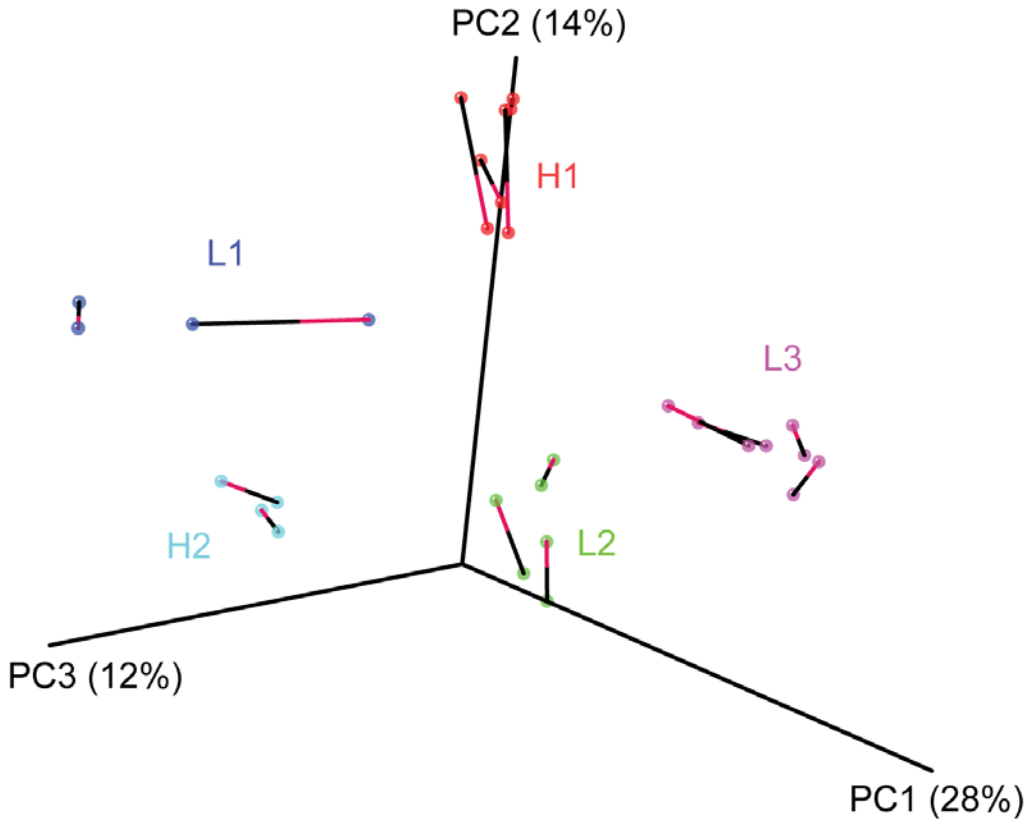
Variation associated with subject and diet

- People are more similar to themselves
- People on the same diet become more similar over time
- Specify the contigs involved



P values by label permutation

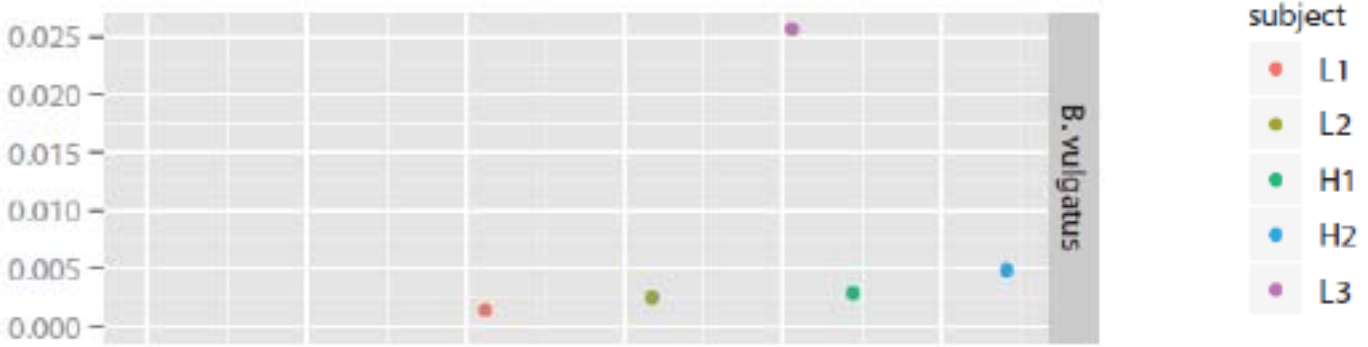
Procrustes analysis: covariation of host and viral communities



What determines phage abundance?

- Abundance of host?
- Lysogenic induction?

Phage abundance



Host abundance

In some cases the abundance of phage does not correlate with apparent abundance of host: possibly indicative of differential induction?

Summary

- Variation in both bacterial and viral communities dominated by interpersonal variation
- Dietary intervention associated with altered composition of both bacterial and viral communities
- Metagenomic analysis over six subjects yielded 7000 viral contigs; 19-785 types per sample (minimum estimate)
- Both expected and novel functionality, one example being viral CRISPRs targeting other viruses

Credits

UPenn/CHOP Group

Gary Wu*
James Lewis*
Frederic Bushman*
Sam Minot
Emily Charlson
Christian Hoffmann
Scott Sherrill-Mix
Serena Dollive
Kyle Bittinger
Rohini Sinha
Jennifer Hwang
Rebecca Custers-Allen
Stephanie Grunberg

Bob Baldassano
Sue Keilbaugh
Ying-Yu Chen
Lisa Nessel
Jun Chen
Hongzhi Li

** co-PIs*

Collaborators

Rob Knight
Catherine Lozupone
Dan Knights