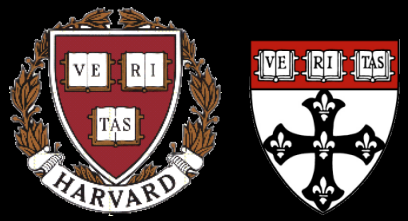
From microbes to microbiota and back:

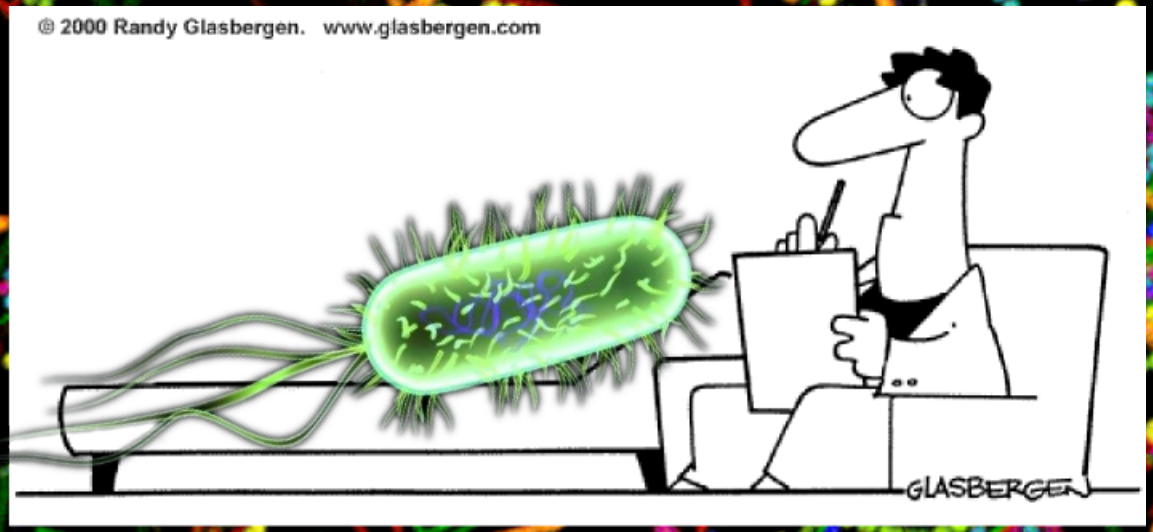
Using thousands of genomes to understand thousands of metagenomes

Curtis Huttenhower



Harvard School of Public Health
Department of Biostatistics

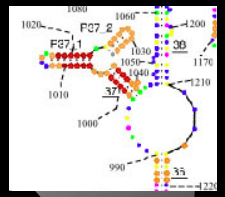
02-16-13



Valm et al, PNAS 2011

Sequencing as a tool for microbial community analysis

Lyse cells
Extract DNA (and/or RNA)

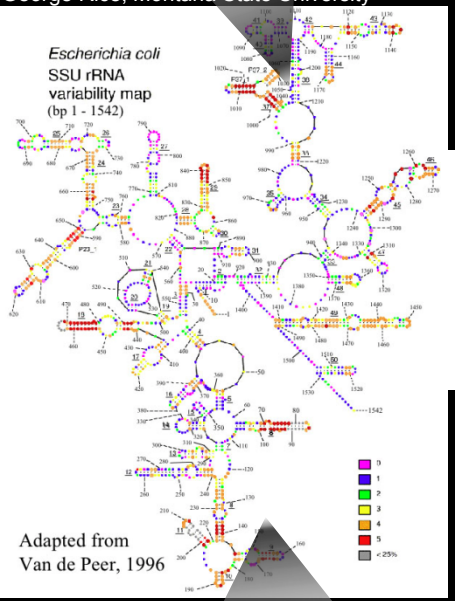


V6

16S amplicons

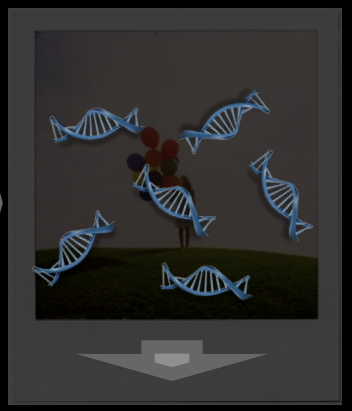
Meta'omic

George Rice, Montana State University



PCR to amplify the single 16S rRNA marker gene

Classify sequence
➔ microbe



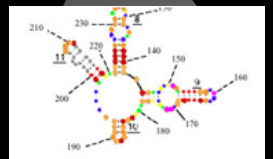
Samples

Microbes

Relative abundances

Genes,
Genomes,
Metabolic profiling,
Relative abundances,
Genetic variants...

V2



ChocoPhlAn: Cataloging core and unique marker gene sequences


Nicola Segata

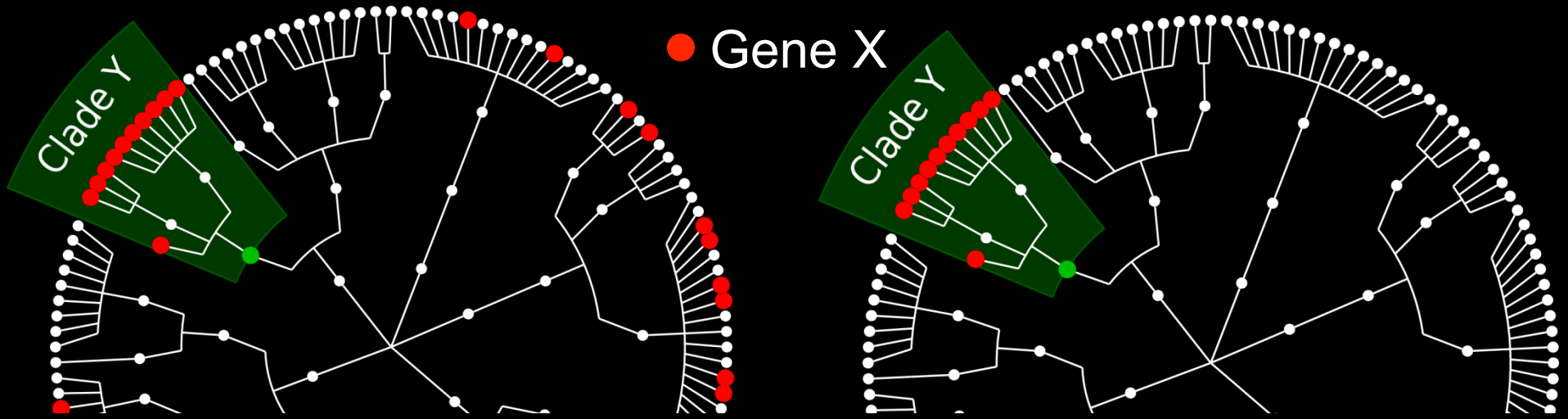
- IMG alone now contains ~3,100 bacterial genomes
  - Plus ~100 archaeal, ~100 eukaryotic, and a few thousand viruses
  - About half final and half draft
- These comprise 1,222 bacterial species
  - 652 genera, 278 families, 130 orders, 66 classes, 33 phyla
  - 2,383 total clades
- **And roughly 12M genes**

- These genes and genomes are a tremendous resource to:
  - *Identify **unique** markers that can be used to infer taxonomy*
  - *Identify **conserved** markers that can be used to infer phylogeny*
  - *Relate the microbial members of a community to their annotated metagenomic **functional potential***

MetaPhlAn: phylogenetically unique markers sequences for taxonomic profiling

X is a core gene for clade Y

X is a unique marker gene for clade Y



ChocoPhlAn (offline pipeline)

- Identify all **core genes** for all clades
- Screen core genes for unique **marker genes**
- Select most representative marker genes

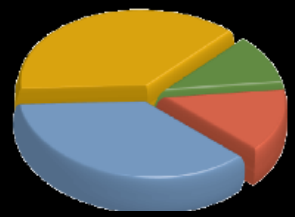
Unique marker genes DB

Available reference genomes

MetaPhlAn

- Blast reads agains the marker genes
- Assign, count, normalize reads

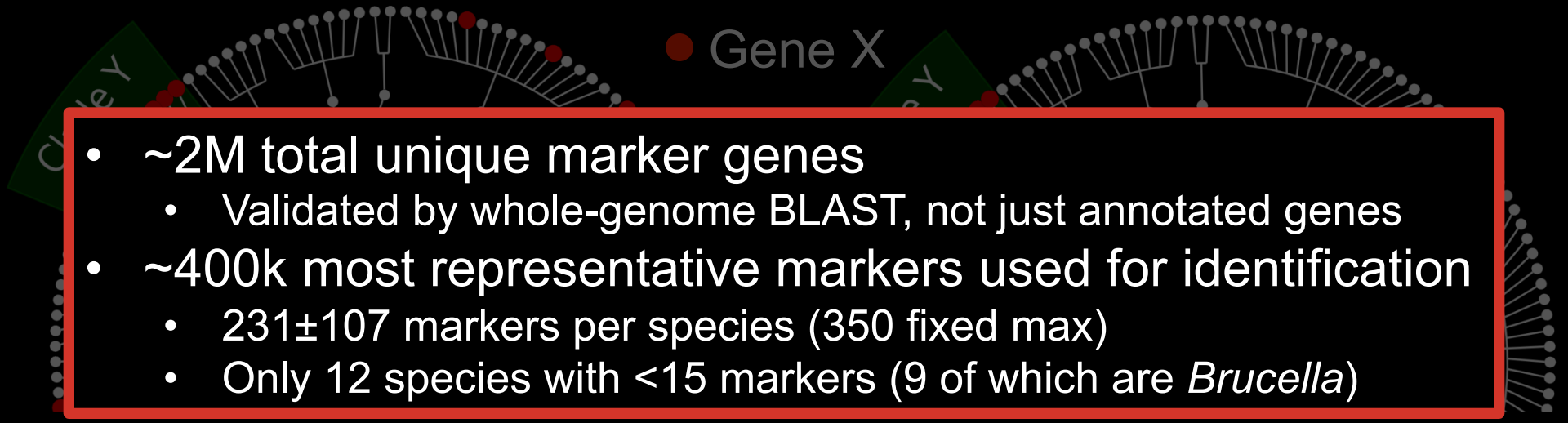
Metagenome



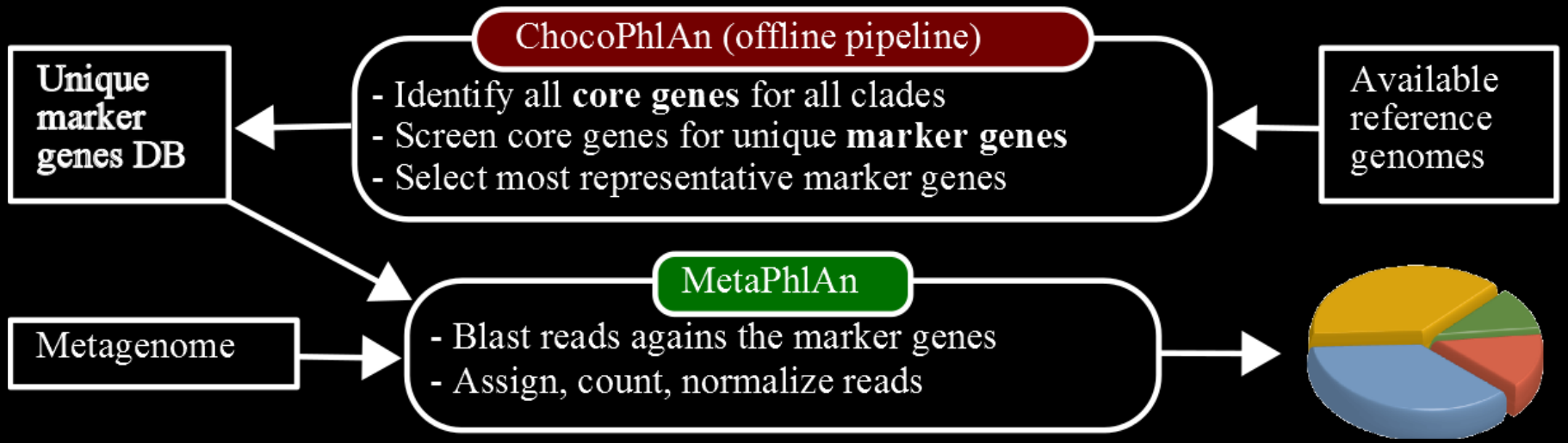
MetaPhlAn: phylogenetically unique markers sequences for taxonomic profiling

X is a core gene for clade Y

X is a unique marker gene for clade Y

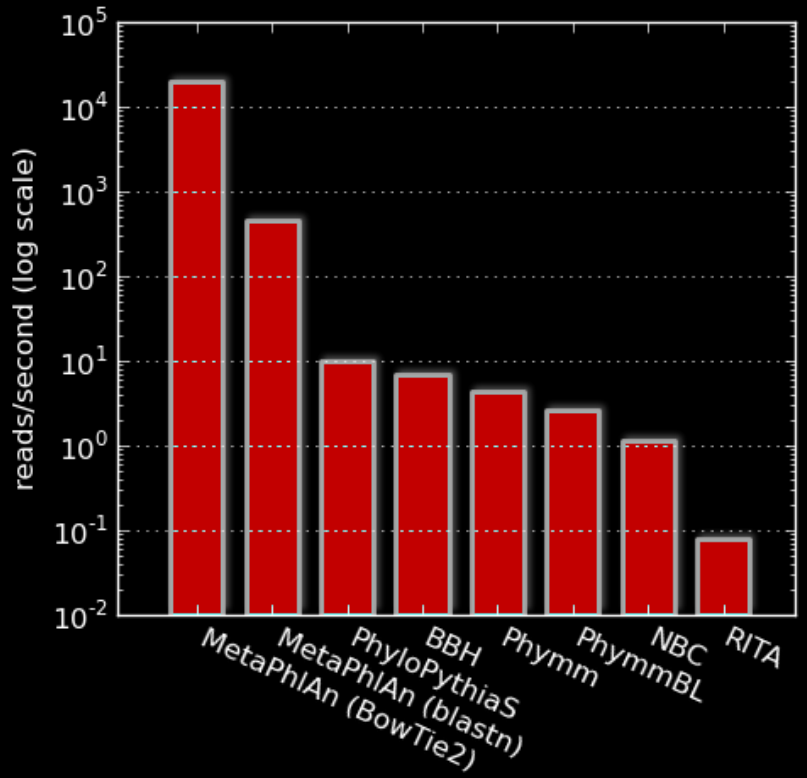


- ~2M total unique marker genes
  - Validated by whole-genome BLAST, not just annotated genes
- ~400k most representative markers used for identification
  - 231±107 markers per species (350 fixed max)
  - Only 12 species with <15 markers (9 of which are *Brucella*)



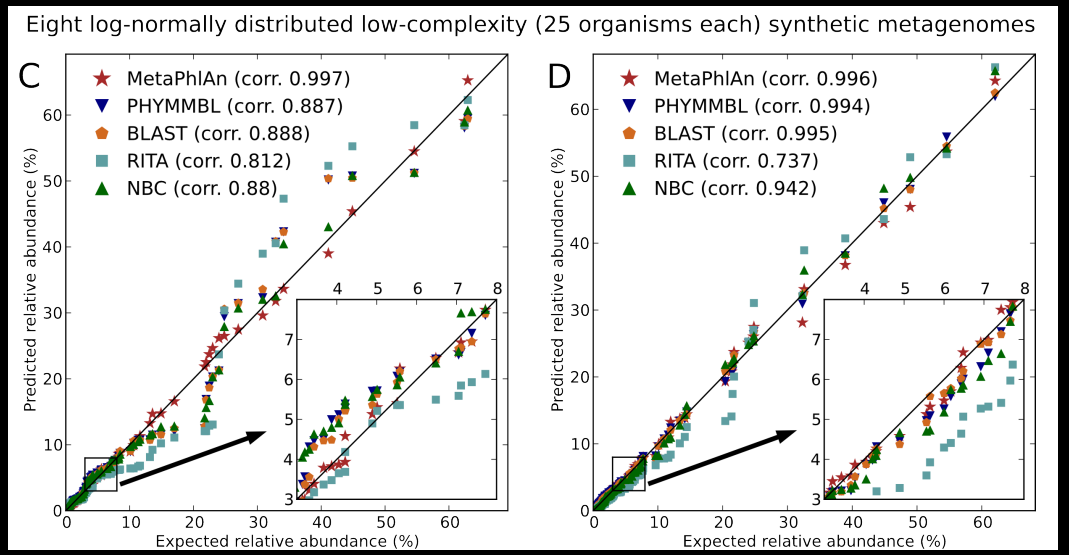
MetaPhlAn: inferring microbial abundances from metagenomic data using marker genes

- Map metagenomic reads to marker genes to infer microbial abundances
  - Normalizing for copy number, gene length, etc.



~1000x faster than previous approaches

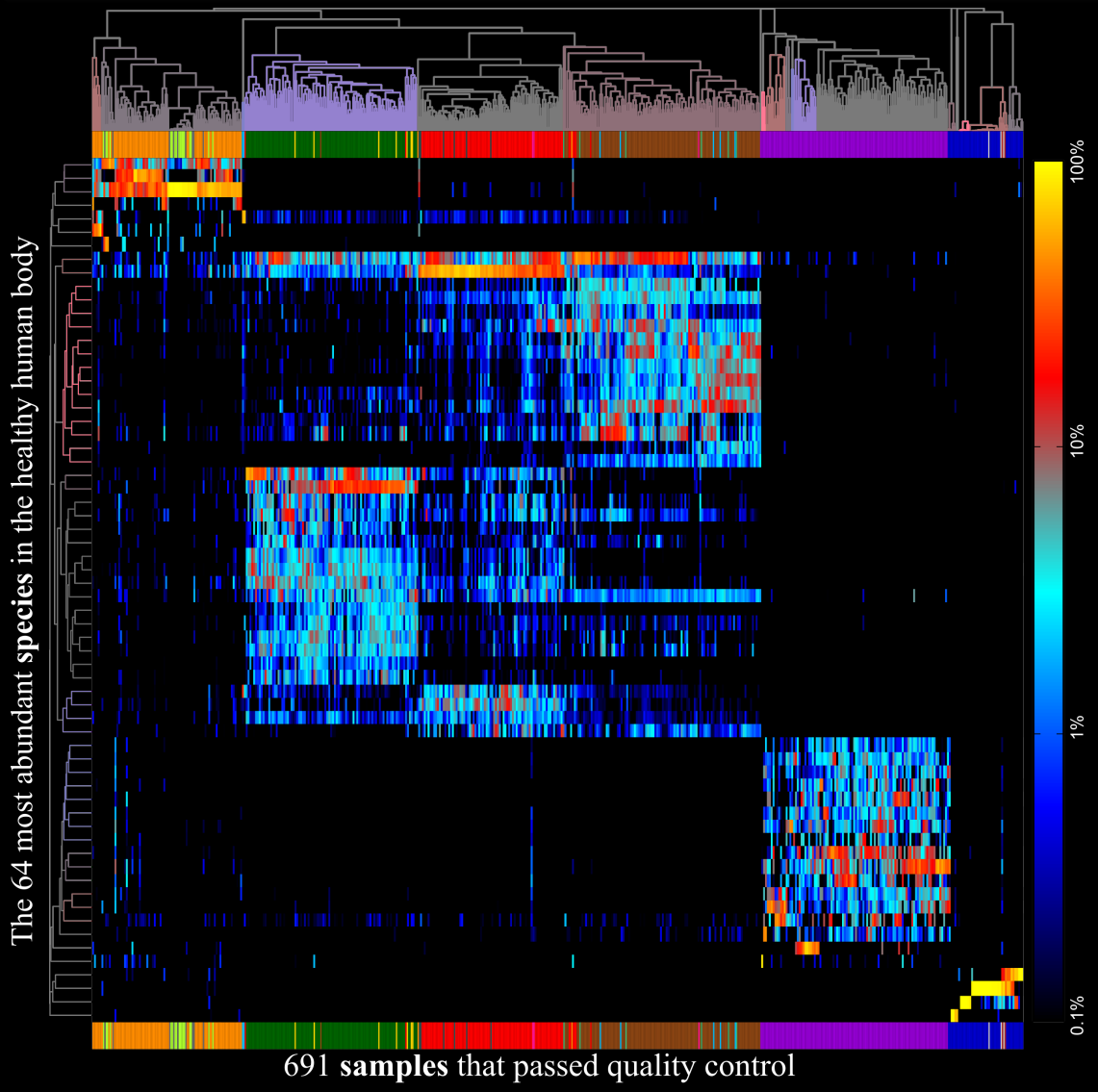
Hours instead of weeks for
Illumina samples with 100Gbs of sequence



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

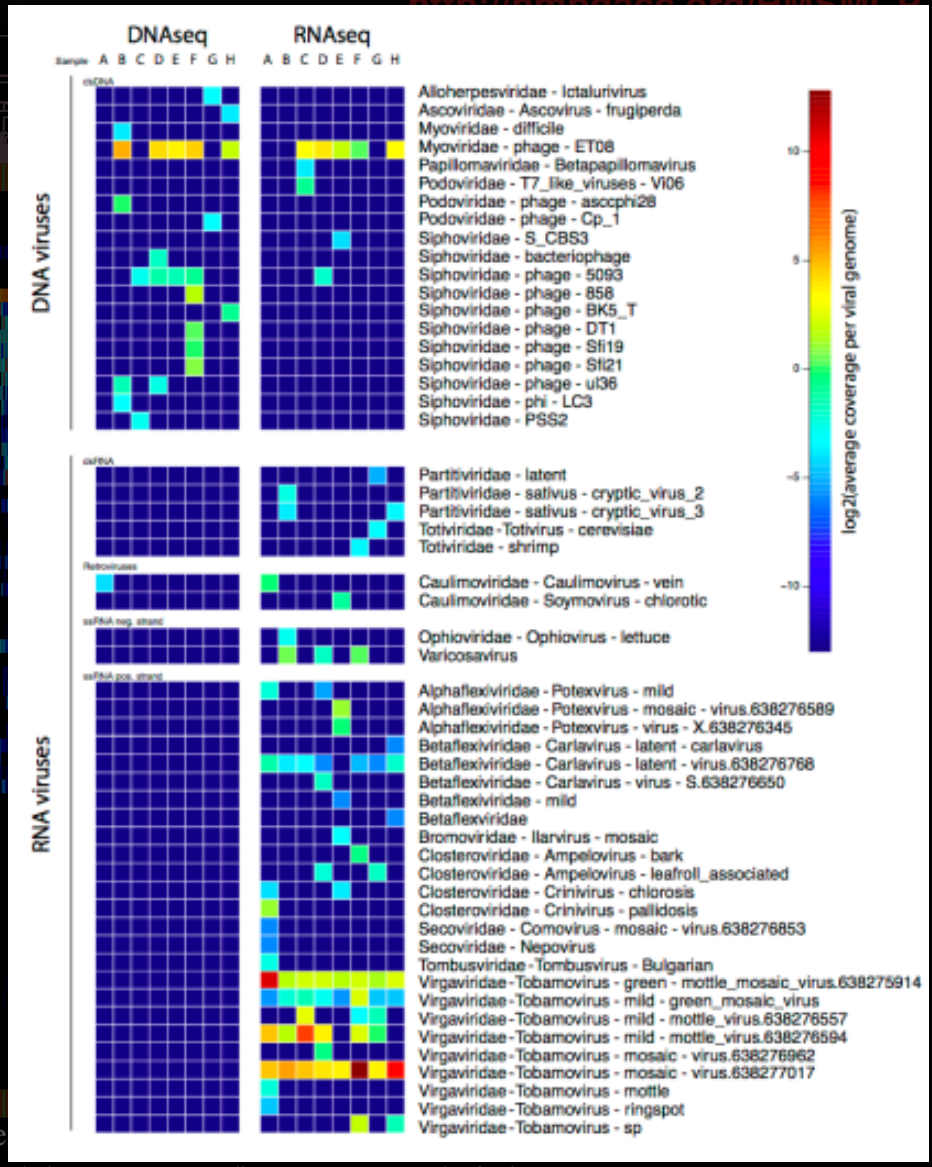
The HMP's human microbiome at species-level resolution

http://hmpdacc.org/HMSMCP



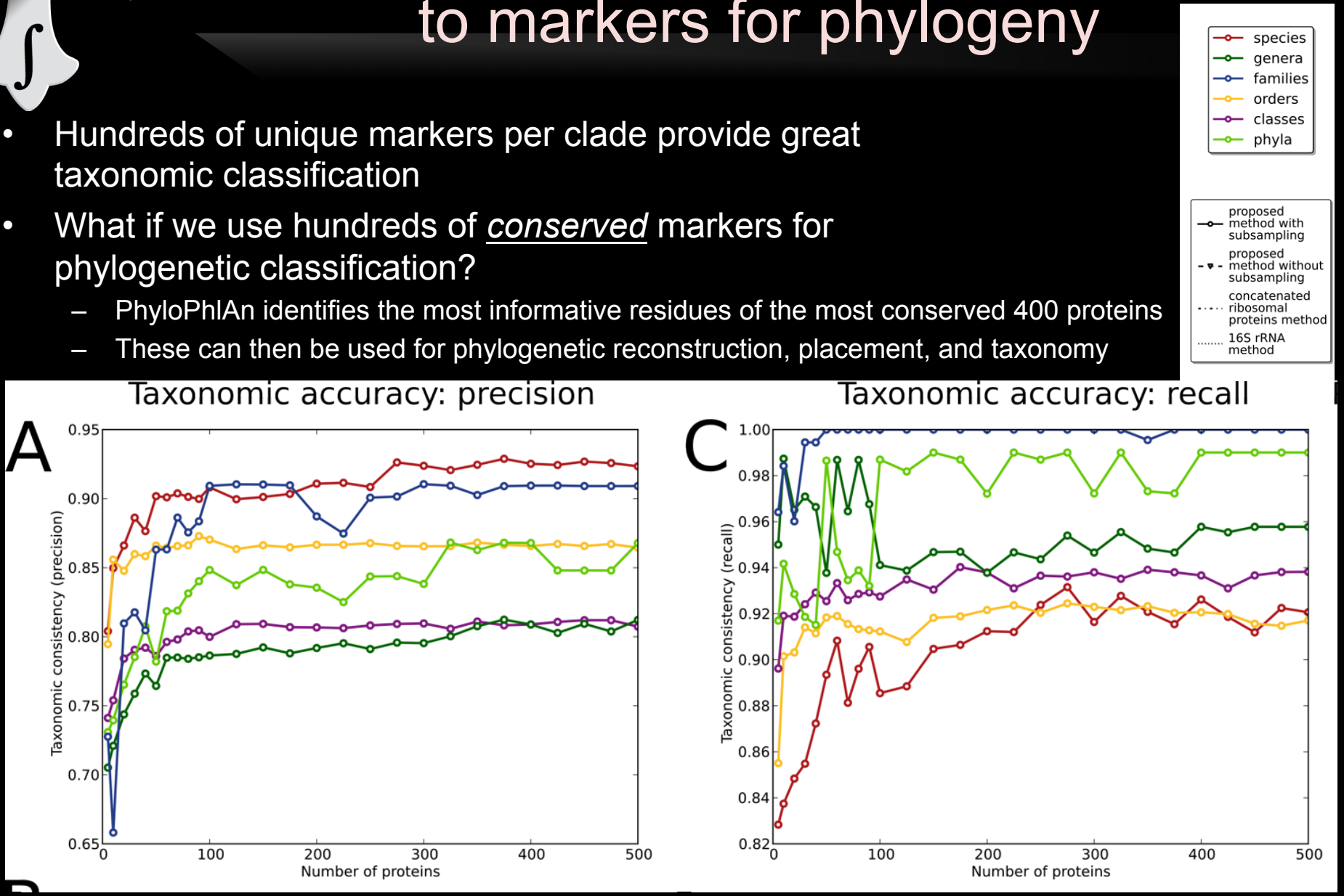
The HMP's human microbiome at species-level resolution

Watch for MetaPhlAn v2:
Additional archaea, DNA/RNA viruses, and eukaryotic microbes.



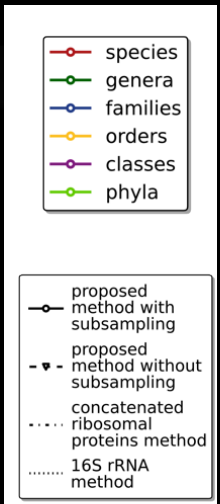
PhyloPhlAn: From markers for taxonomy to markers for phylogeny

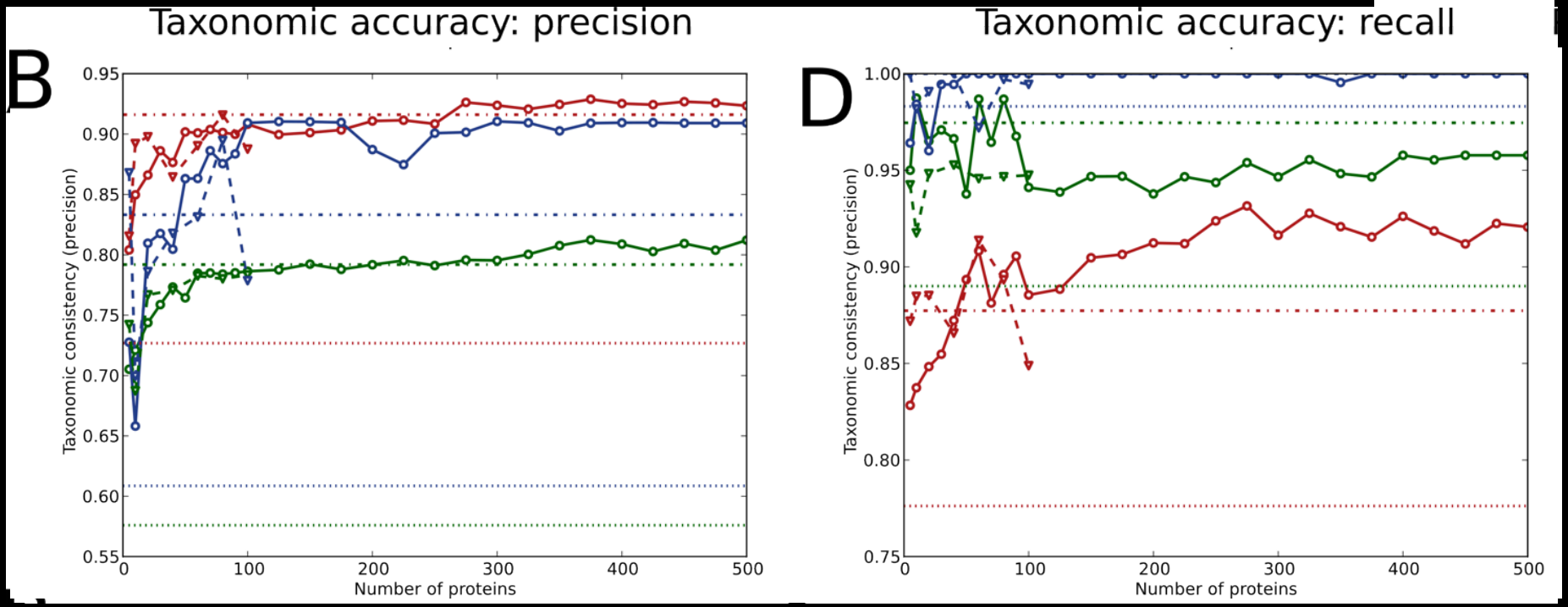
- Hundreds of unique markers per clade provide great taxonomic classification
- What if we use hundreds of *conserved* markers for phylogenetic classification?
  - PhyloPhlAn identifies the most informative residues of the most conserved 400 proteins
  - These can then be used for phylogenetic reconstruction, placement, and taxonomy



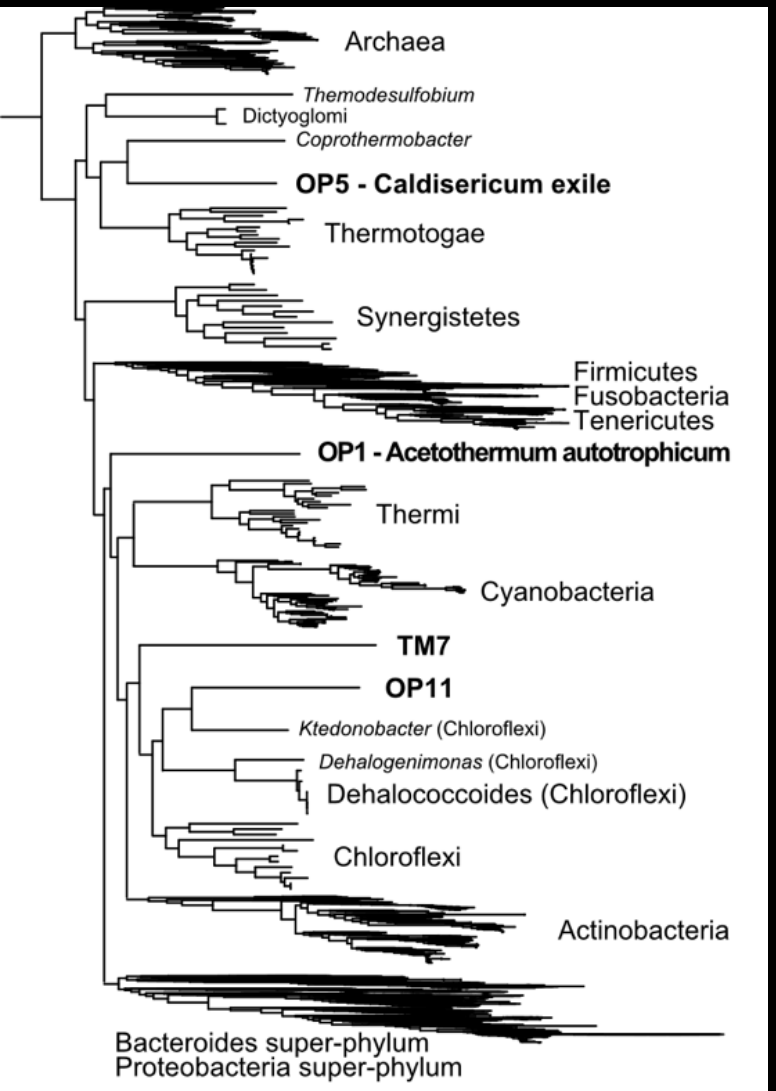
PhyloPhlAn: From markers for taxonomy to markers for phylogeny

- Hundreds of unique markers per clade provide great taxonomic classification
- What if we use hundreds of *conserved* markers for phylogenetic classification?
  - PhyloPhlAn identifies the most informative residues of the most conserved 400 proteins
  - These can then be used for phylogenetic reconstruction, placement, and taxonomy

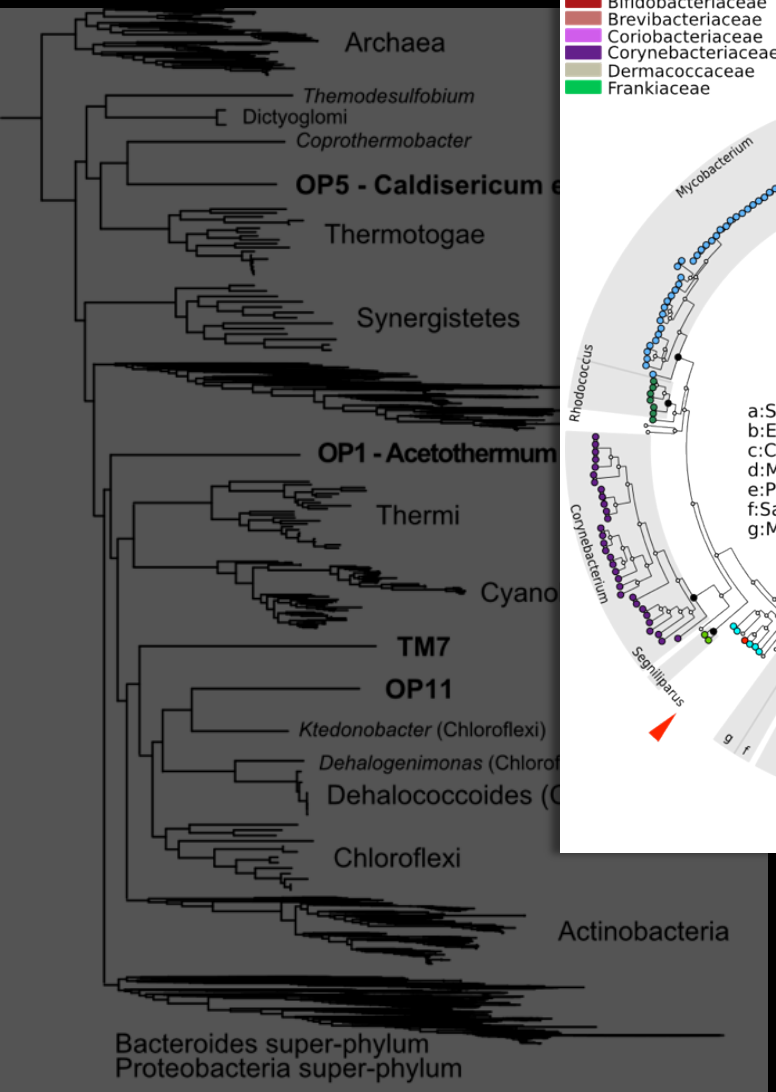


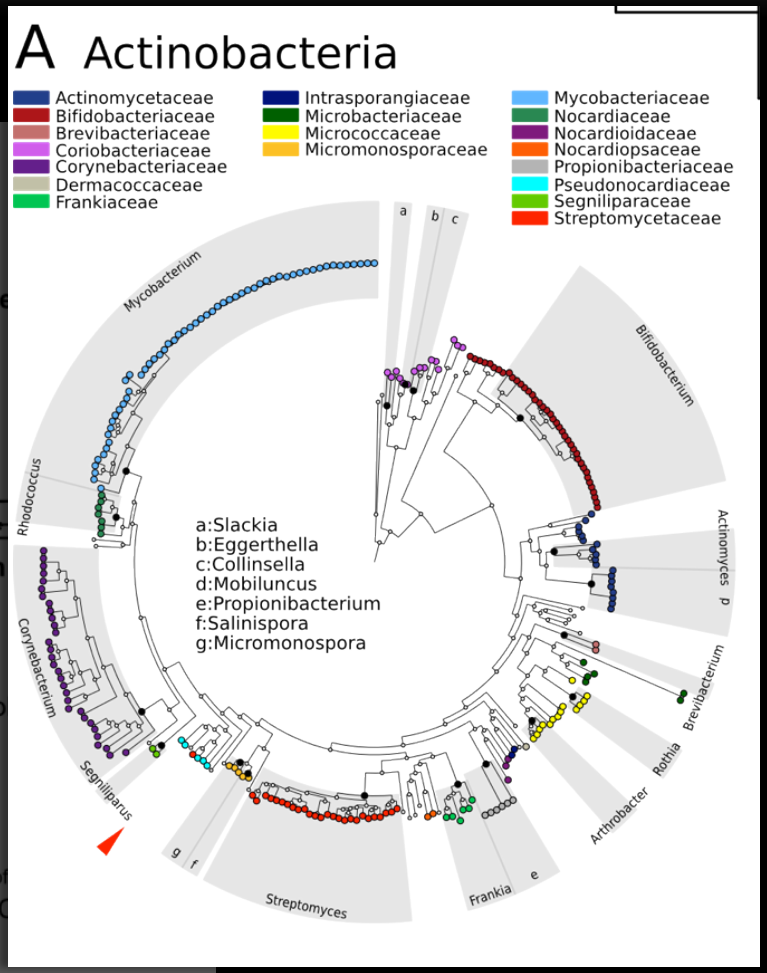


PhyloPhlAn: for phyla to subspecies

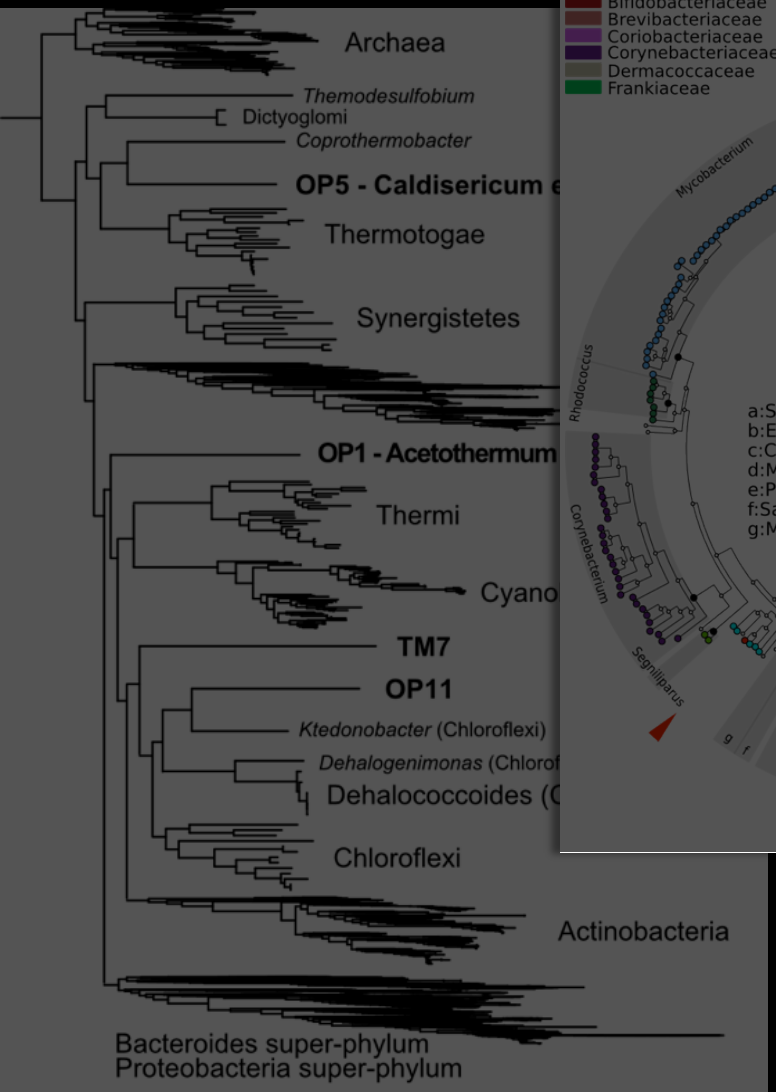


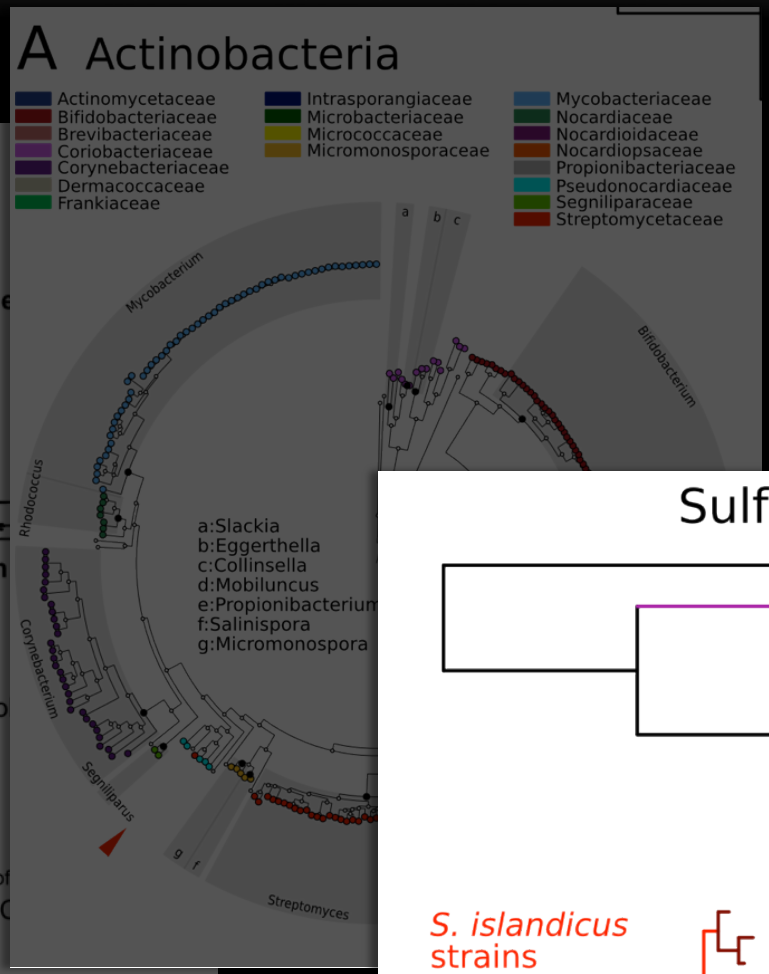
PhyloPhlAn: for phyla to subspecies

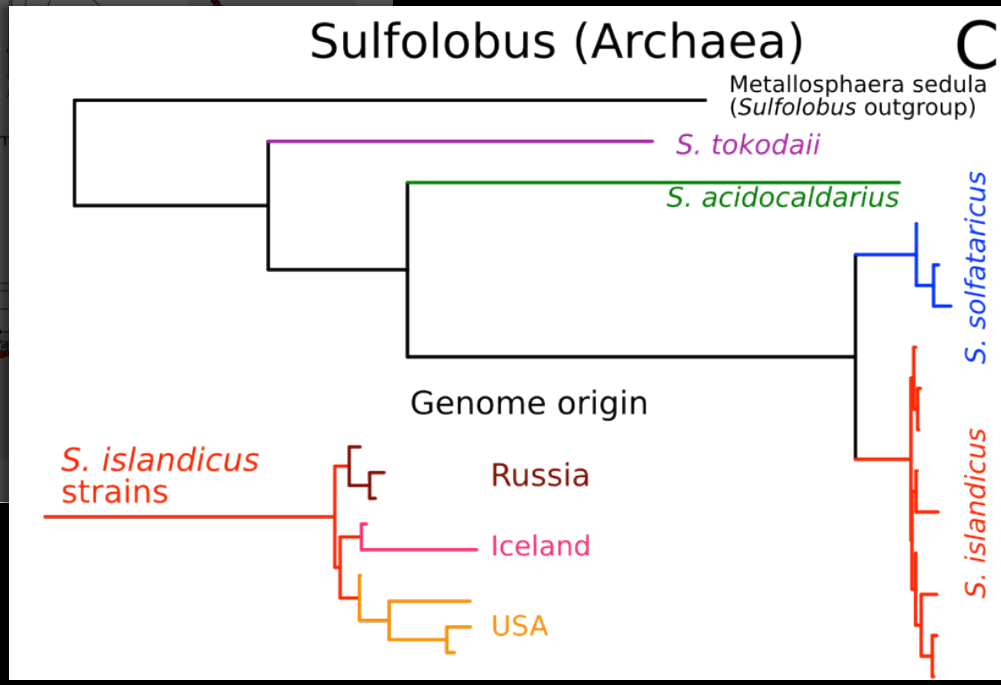


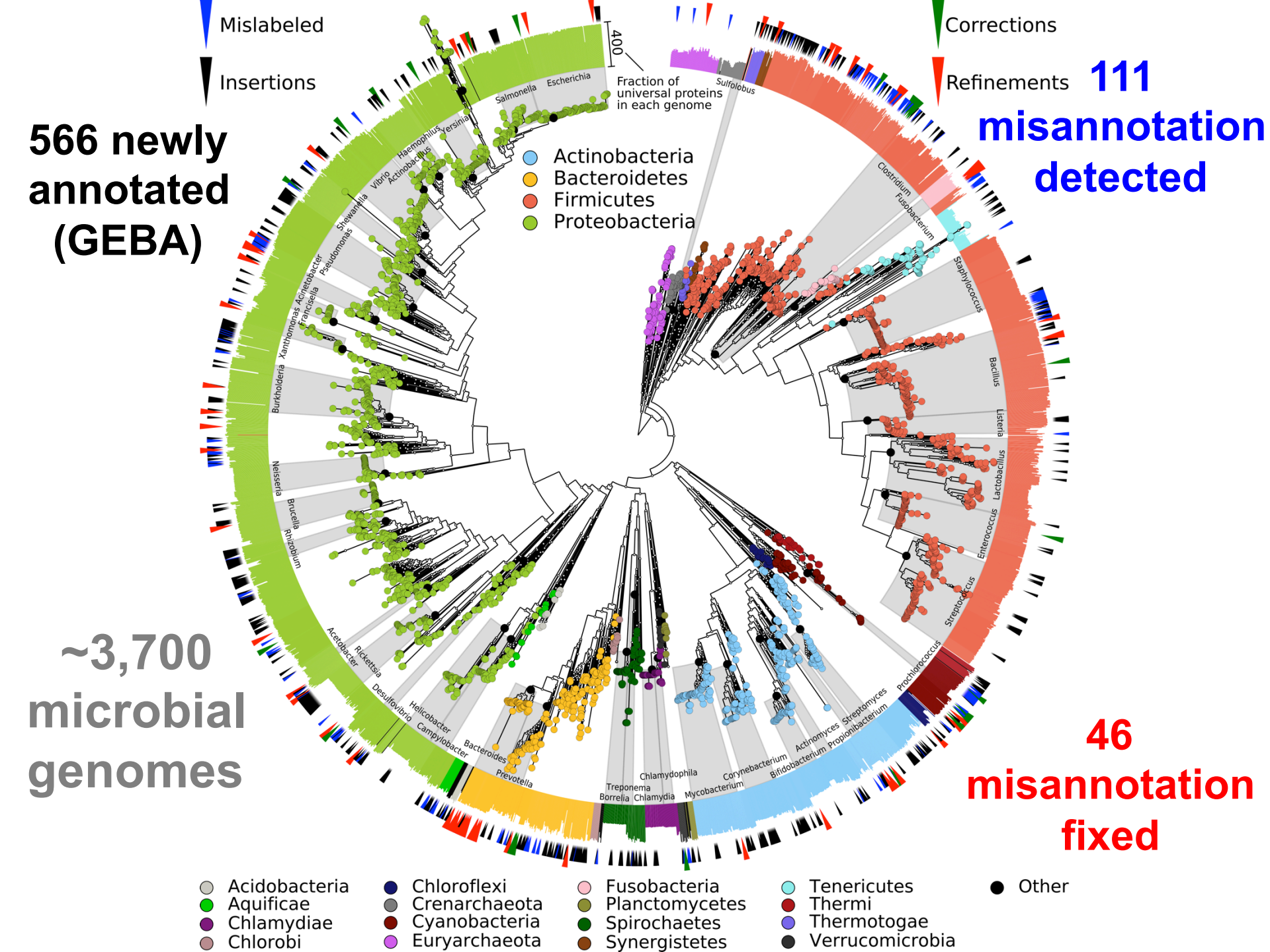


PhyloPhlAn: for phyla to subspecies

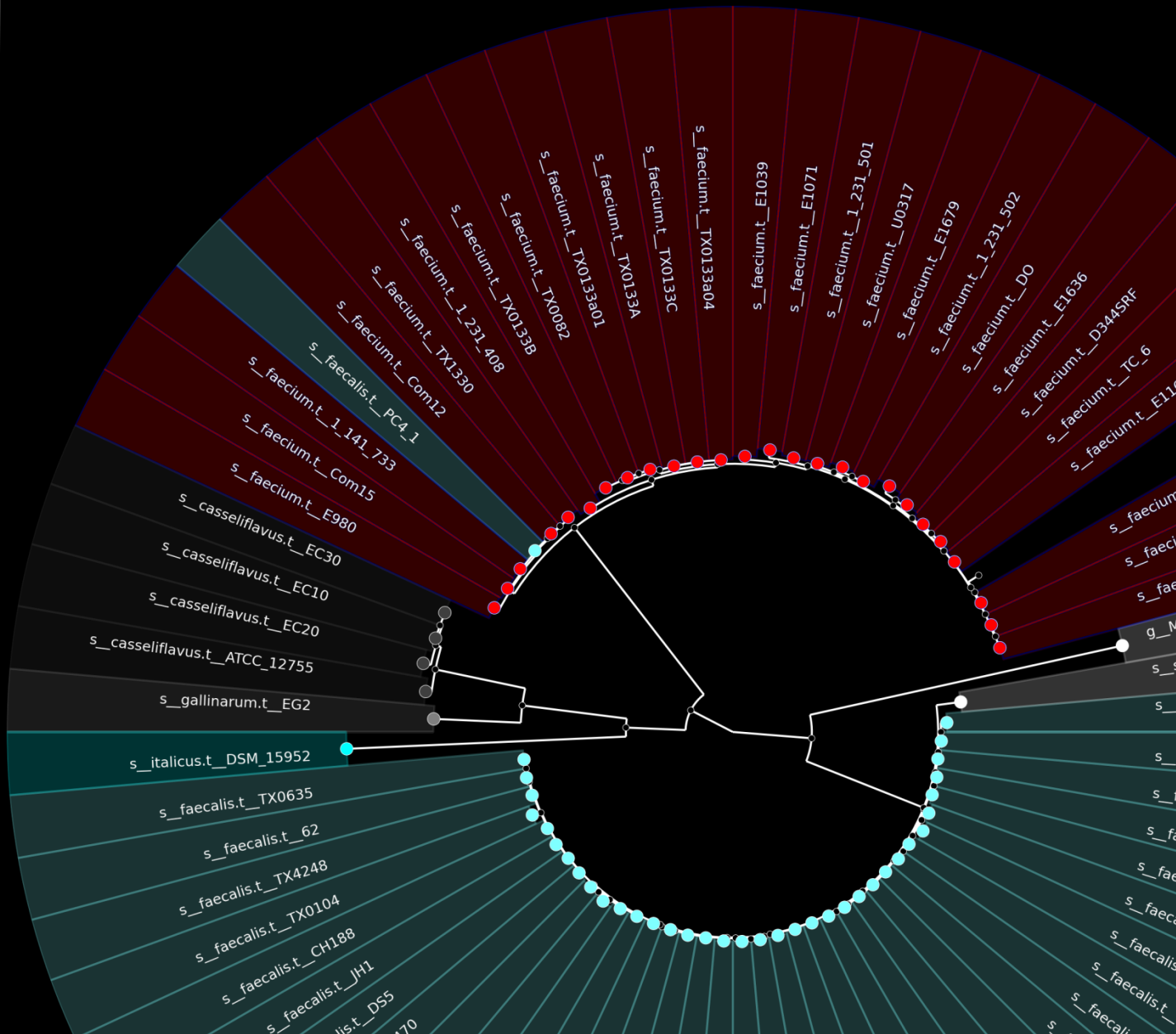






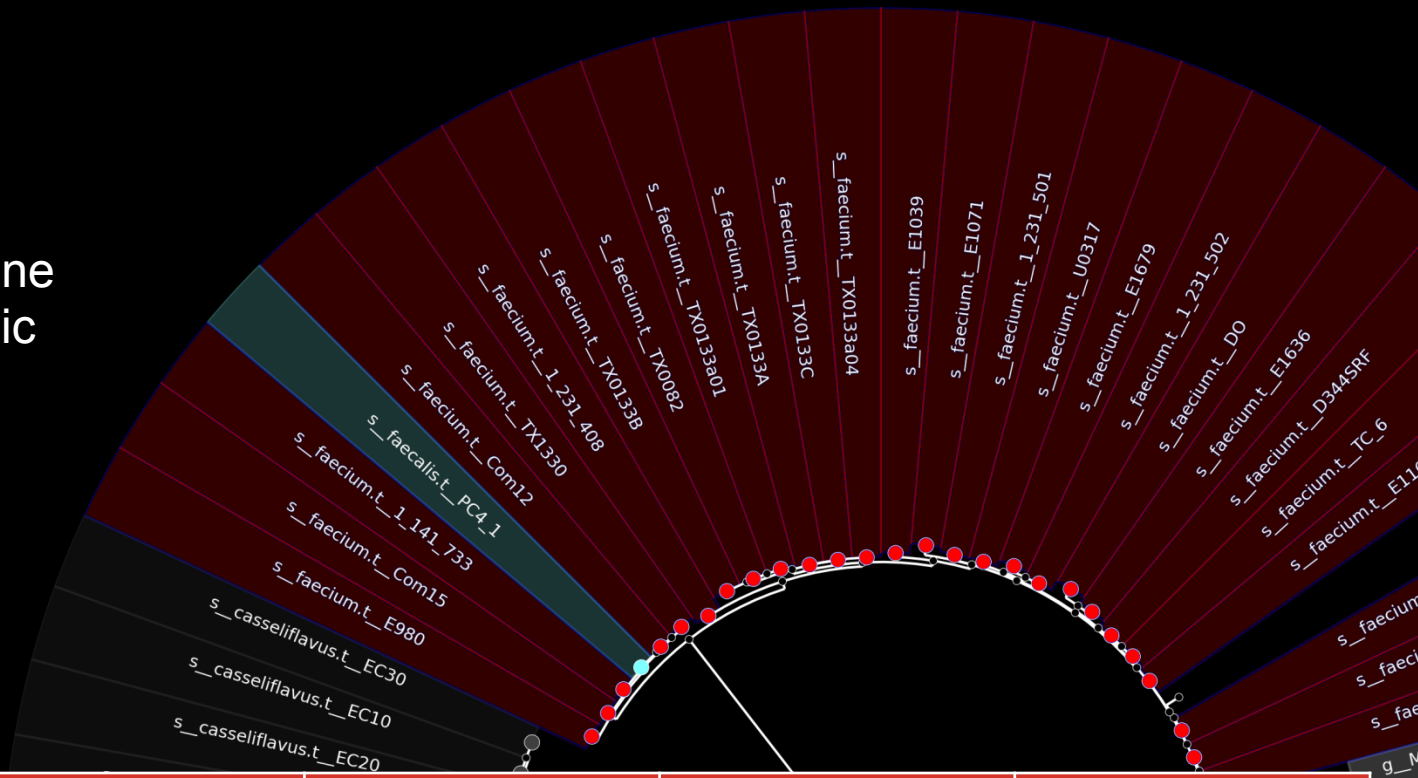

Mislabeled
Insertions
Corrections
Refinements
566 newly annotated (GEBA)
111 misannotation detected
400
Fraction of universal proteins in each genome
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Salmonella
Escherichia
Yersinia
Haemophilus
Actinobacillus
Vibrio
Shewanella
Pseudomonas
Acinetobacter
Xanthomonas
Francisella
Burkholderia
Neisseria
Brucella
Rhizobium
Acetobacter
Rickettsia
Desulfovibrio
Helicobacter
Campylobacter
Bacteroides
Prevotella
Treponema
Borrelia
Chlamydophila
Chlamydia
Mycobacterium
Corynebacterium
Bifidobacterium
Actinomyces
Streptomyces
Propionibacterium
Prochlorococcus
Streptococcus
Enterococcus
Lactobacillus
Listeria
Bacillus
Staphylococcus
Fusobacterium
Clostridium
Sulfolobus
~3,700 microbial genomes
46 misannotation fixed
Acidobacteria
Aquificae
Chlamydiae
Chlorobi
Chloroflexi
Crenarchaeota
Cyanobacteria
Euryarchaeota
Fusobacteria
Planctomycetes
Spirochaetes
Synergistetes
Tenericutes
Thermi
Thermotogae
Verrucomicrobia
Other

PhyloPhlAn: Taxonomic curation and reannotation



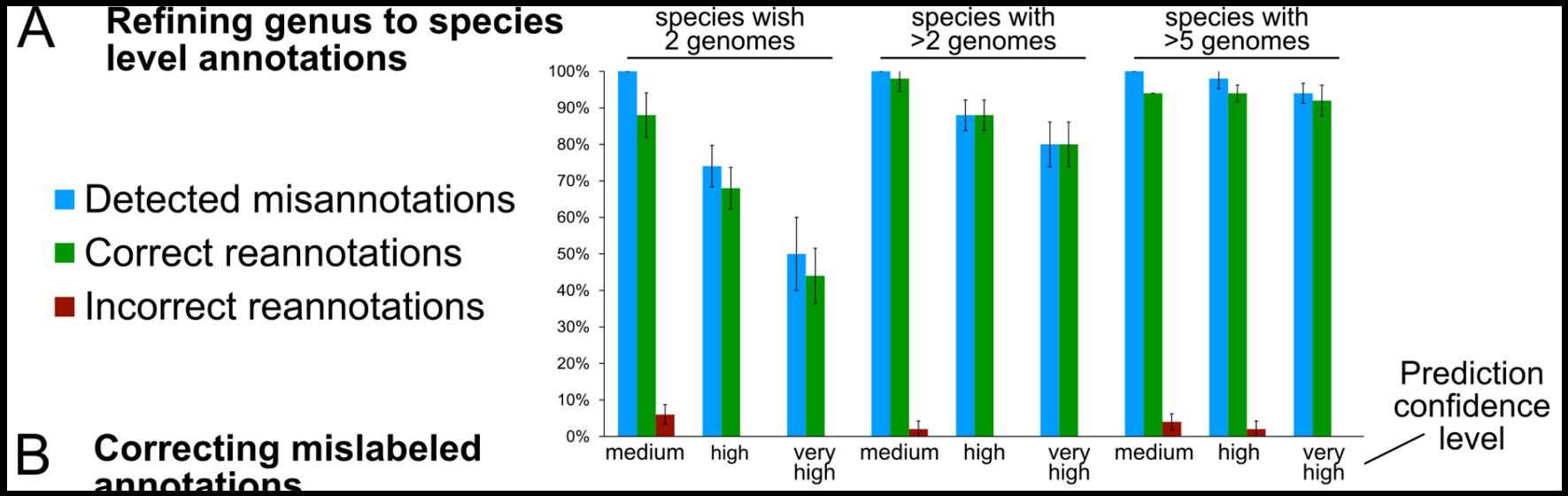
PhyloPhlAn: Taxonomic curation and reannotation

- Taxa with at least one ‘unknown’ taxonomic level: 445
- Additional taxa we detected as suspicious: 111

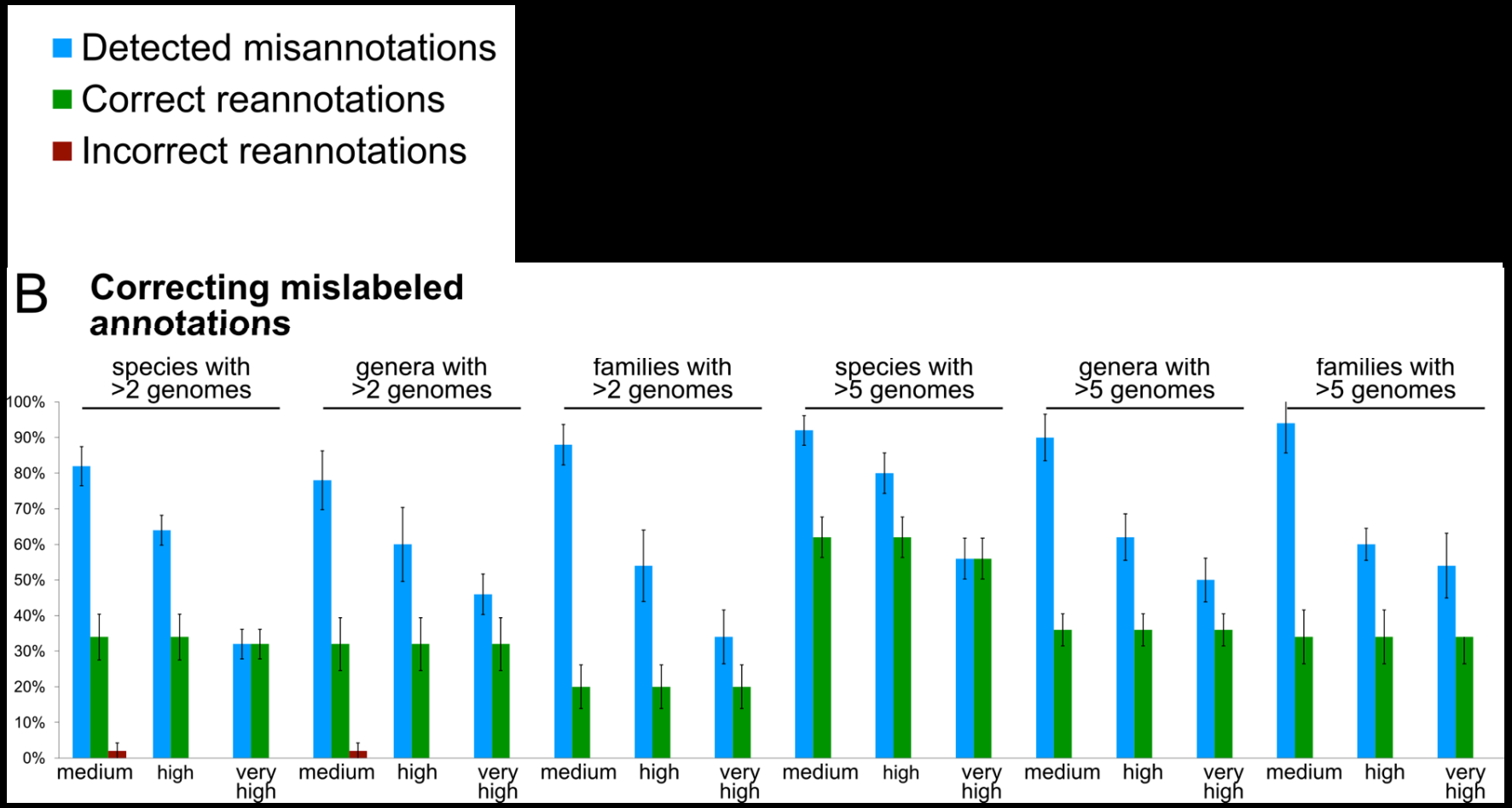


	Example	Very high confidence	High confidence	Medium confidence
Corrected	A|B|C→A|B|D	26	3	26
Refined	A|B|?→A|B|C	67	25	75
Removed	A|B|C→A|B|?	11	1	1
Incomplete	A|?|?→A|?|?	224	10	66

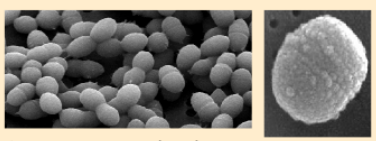
PhyloPhlAn: Automatically assigning precise taxonomy using precise phylogeny



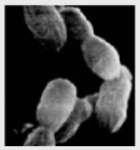
PhyloPhlAn: Automatically assigning precise taxonomy using precise phylogeny



A map of diversity in the human microbiome



Streptococcus dominates the oral cavity with *S. mitis* > 75% in the cheek



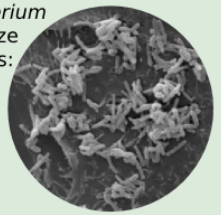
Propionibacterium acnes lives on the skin and nose of most people

Many *Corynebacterium* species characterize different body sites:

C. matruchoti the plaque

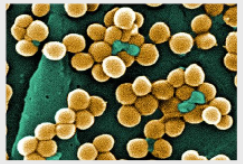
C. accolens the nose

C. croppenstedtii the skin

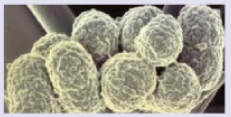


Lactobacillus species (*L. gasseri*, *L. jensenii*, *L. crispatus*, *L. iners*) are predominant but mutually exclusive in the vagina

Staphylococcus epidermidis colonizes external body sites

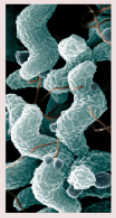


Several *Prevotella* species are present in the gastrointestinal tract. *P. copri* is present in 19% of the subjects and dominates the intestinal flora when present



Microscopy from http://bacmap.wishartlab.com

Bacteroides is the most abundant genus in the gut of almost all healthy subjects



Campylobacter includes opportunistic pathogens, but members live in the oral cavities of most healthy people in the cohort

E. coli is present in the gut of the majority of healthy subjects but at very low abundance

Commensal microbes

Potential pathogens

The four most abundant phyla

- Actinobacteria
- Bacteroidetes
- Firmicutes
- Proteobacteria

Low abundance phyla

- Chloroflexi
- Cyanobacteria
- Euryarchaeota
- Fusobacteria
- Lentisphaerae
- Spirochaetes
- Synergistetes
- Tenericutes
- Thermi
- Verrucomicrobia

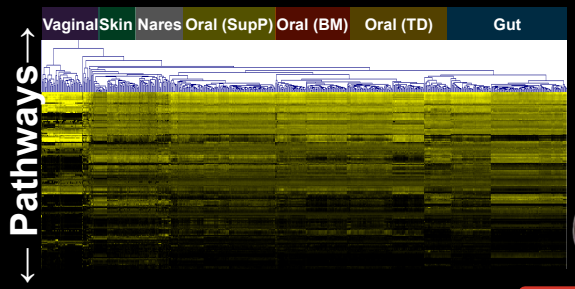
National Institutes of Health Human Microbiome Project

N. Segata & C. Huttenhower
http://huttenhower.sph.harvard.edu

(generated using CirClader and mTol-400 from MetaPhlAn analysis)

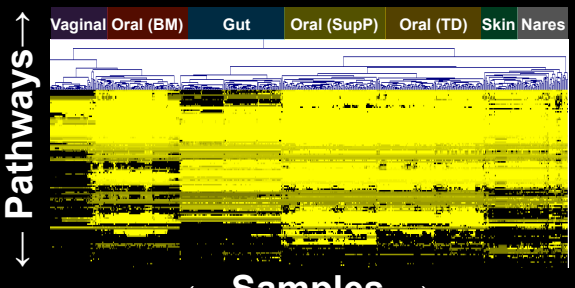
http://huttenhower.sph.harvard.edu/graphlan

HUMAnN: Metabolic profiling for microbial communities



Pathway abundance

Pathway coverage



100 subjects
1-3 visits/subject
~7 body sites/visit
10-200M reads/sample
100bp reads

BIOCYC Database Collection

BLAST

Functional seq.
KEGG + MetaCYC
CAZy, TCDB,
VFDB, MEROPS…

Metagenomic reads

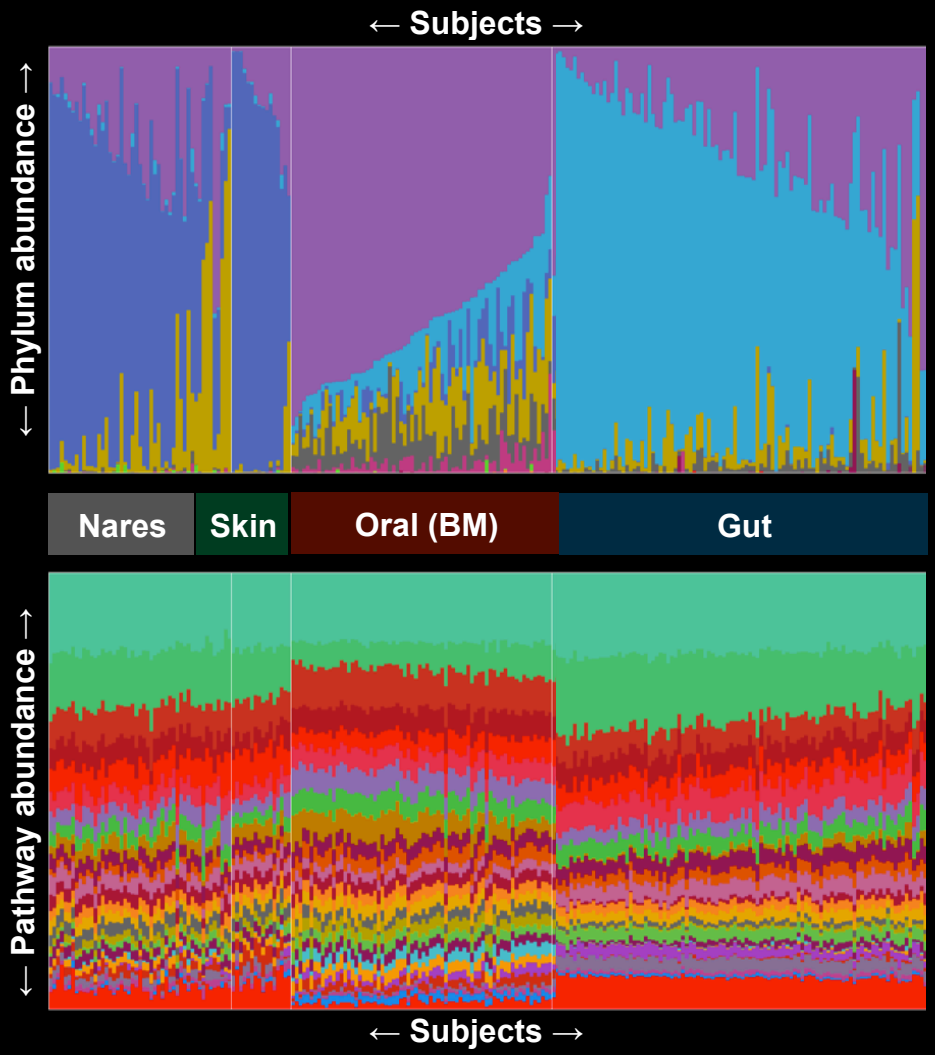
?

Enzymes and pathways

HUMAnN
HMP Unified Metabolic Analysis Network
http://huttenhower.sph.harvard.edu/humann

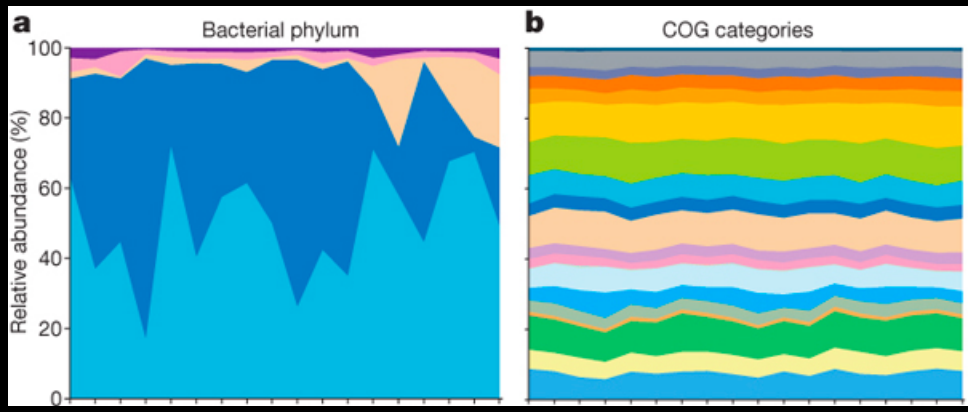
The “core” human microbiome consists of genes, not bugs.

http://hmpdacc.org/HMSMCP



http://hmpdacc.org/HMMRC

Turnbaugh 2009



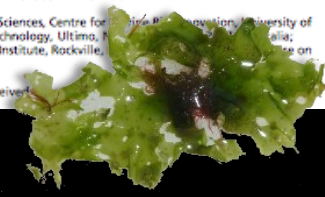
PNAS

Bacterial community assembly based on functional genes rather than species

Catherine Burke[a,b], Peter Steinberg[c,d], Doug Rusch[e], Staffan Kjelleberg[a,f], and Torsten Thomas[a,1]

[a]School of Biotechnology and Biomolecular Sciences, [b]School of Biological, Earth and Environmental Sciences, Centre for ... New South Wales, Sydney, New South Wales 2052, Australia; [c]The iThree Institute, University of Technology, Ultimo, ... [d]Sydney Institute of Marine Science, Mosman, New South Wales 2088, Australia; [e]The J. Craig Venter Institute, Rockville, ... Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved July 14, 2011 (receive...



PNAS

The convergence of carbohydrate active gene repertoires in human gut microbes

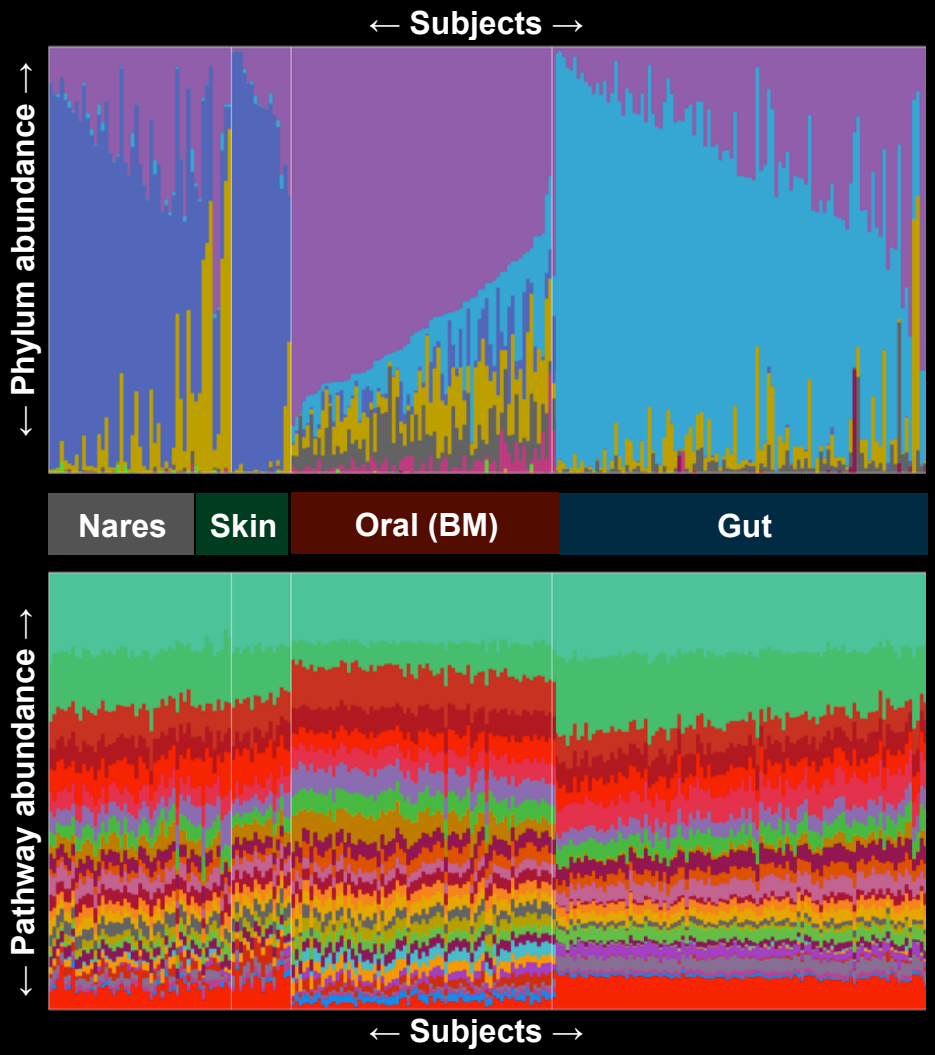
Catherine A. Lozupone*[†], Micah Hamady[‡], Brandi L. Cantarel[§¶], Pedro M. Coutinho[§¶], Bernard Henrissat[§¶], Jeffrey I. Gordon[†], and Rob Knight*[‖]

*Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309; [†]Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108; [‡]Department of Computer Science, University of Colorado, Boulder, CO 80309; and [§]Centre National de la Recherche Scientifique, Unite Mixte de Recherche 6098, [¶]Universités Aix-Marseille I and II, Marseille 13284, France

Contributed by Jeffrey I. Gordon, July 31, 2008 (sent for review June 13, 2008)

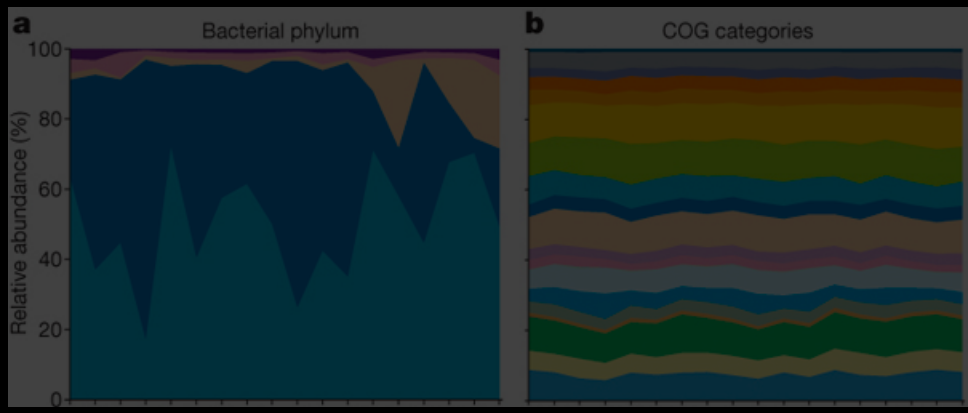
The “core” human microbiome consists of genes, not bugs.

http://hmpdacc.org/HMSMCP



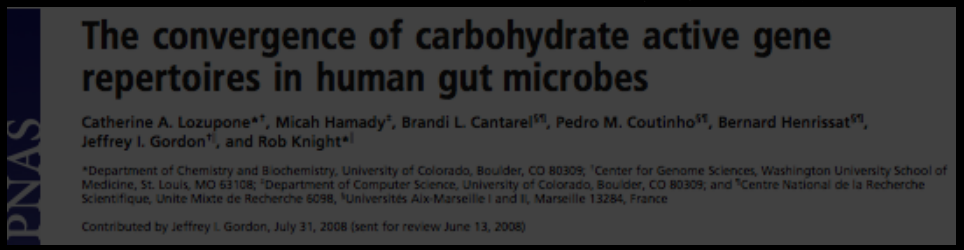
http://hmpdacc.org/HMMRC

Turnbaugh 2009



↙ This is the “core” human microbiome,
↖ Not this.

- Over 2/3 of its genes are uncharacterized, more than almost any single bacterial genome
- We don’t know how its “cell types” communicate
- We don’t know their physical structure or lineages



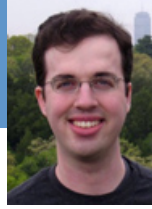
The convergence of carbohydrate active gene repertoires in human gut microbes

Catherine A. Lozupone*†, Micah Hamady‡, Brandi L. Cantarel§¶, Pedro M. Coutinho§¶, Bernard Henrissat§¶, Jeffrey I. Gordon†, and Rob Knight*‖

*Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309; †Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108; ‡Department of Computer Science, University of Colorado, Boulder, CO 80309; and §Centre National de la Recherche Scientifique, Unite Mixte de Recherche 6098, ¶Universités Aix-Marseille I and II, Marseille 13284, France

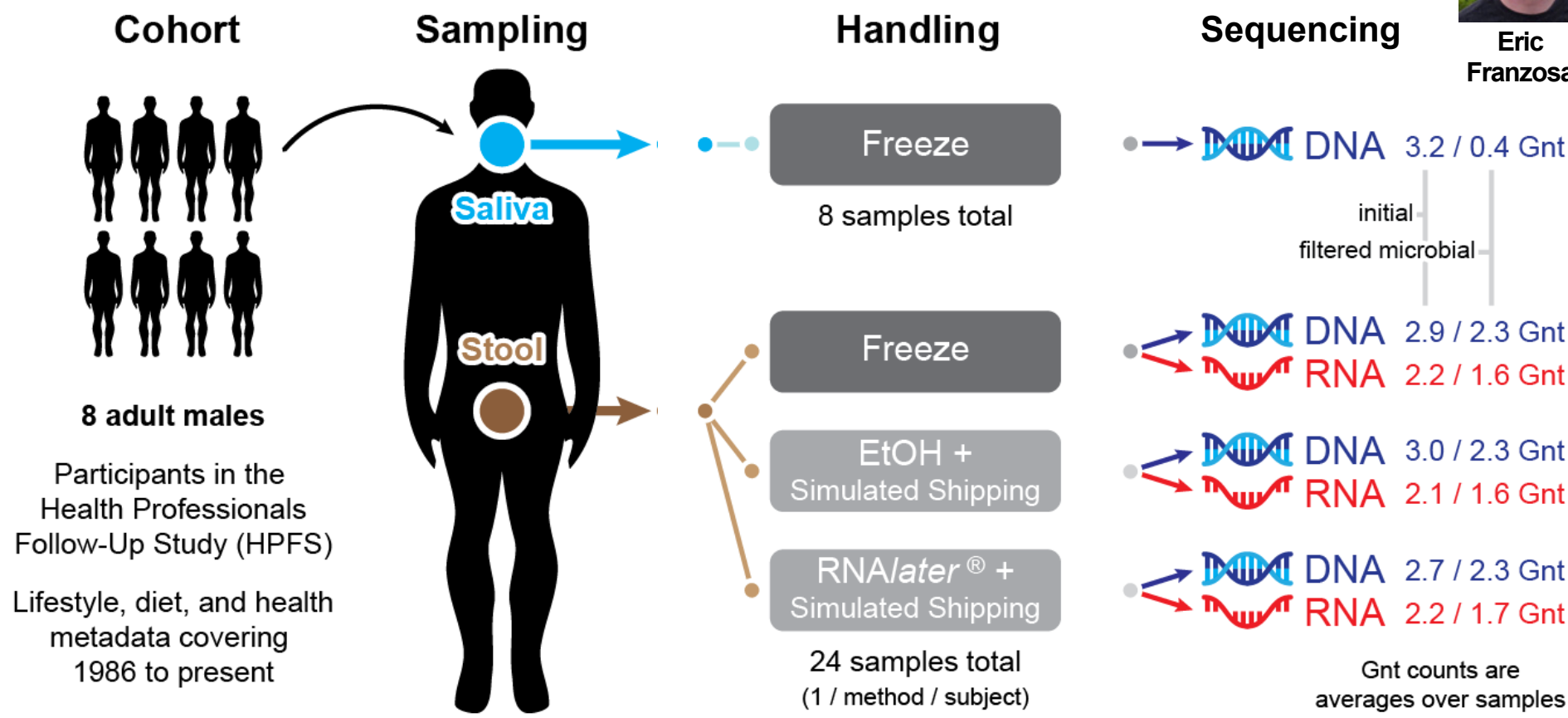
Contributed by Jeffrey I. Gordon, July 31, 2008 (sent for review June 13, 2008)

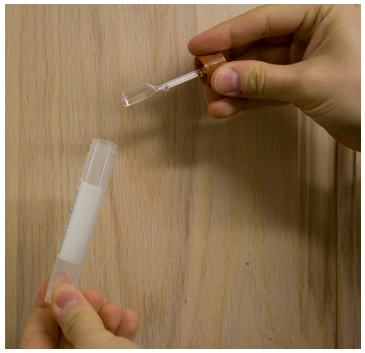
HPFS Pilot Project: Overview



Eric Franzosa

With Jacques Izard, Andy Chan, Wendy Garrett





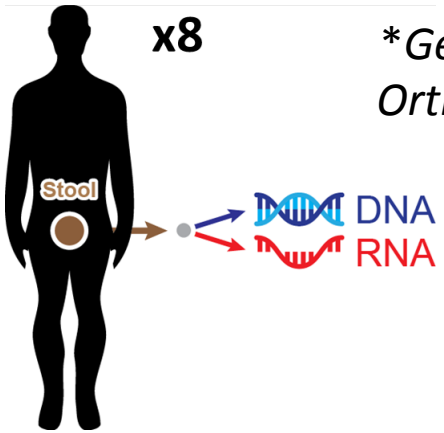
0) Investigate links between the mouth and gut microbiomes

1) Evaluate stability of meta'omic samples under subject-shipped conditions

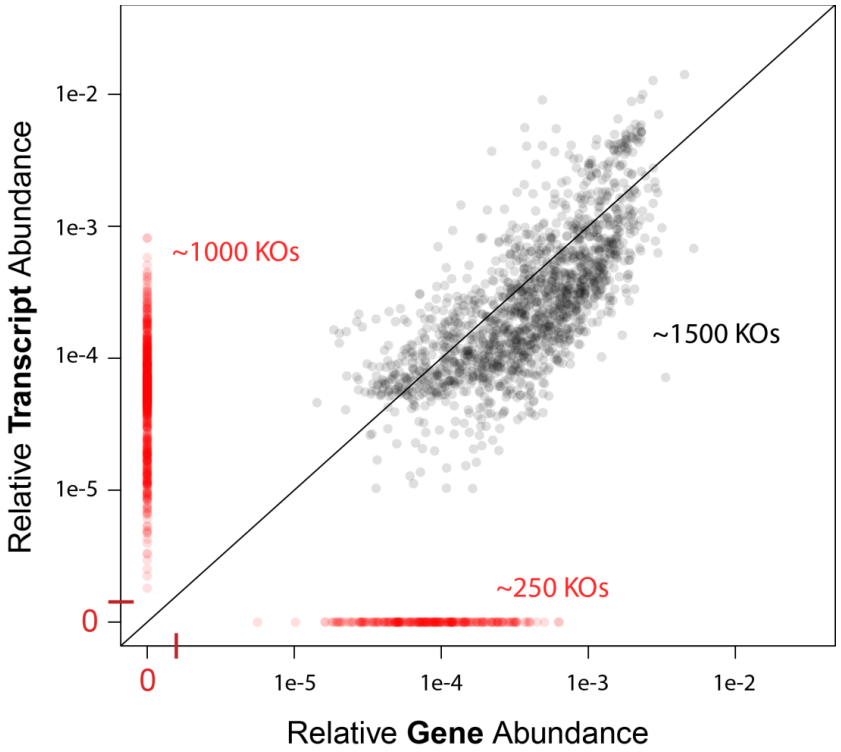
2) Relate the gut metagenome and metatranscriptome

2) Relating the gut metagenome and metatranscriptome

A large portion of genes (~50%) correlate well at the DNA and RNA levels

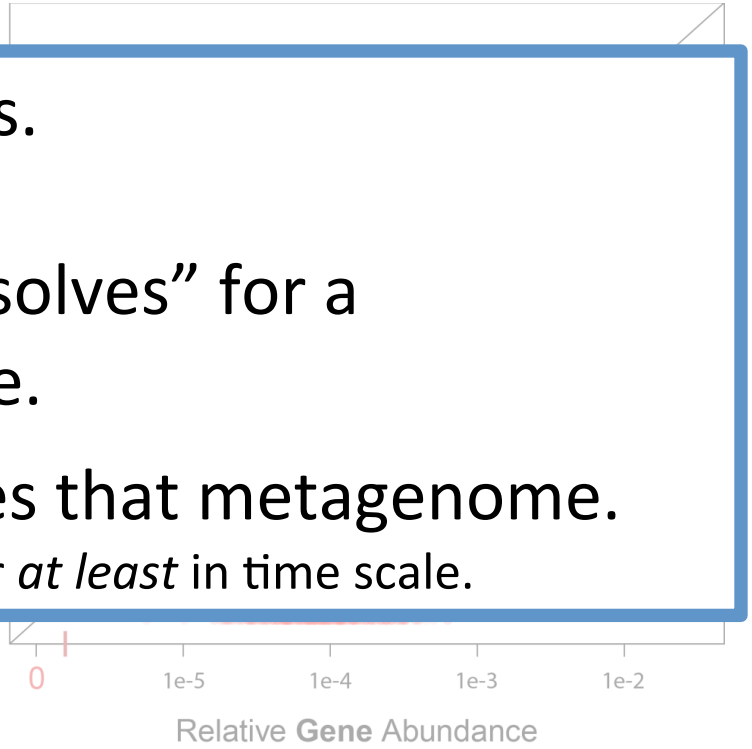


**Genes are KEGG Orthogroups, KOs*

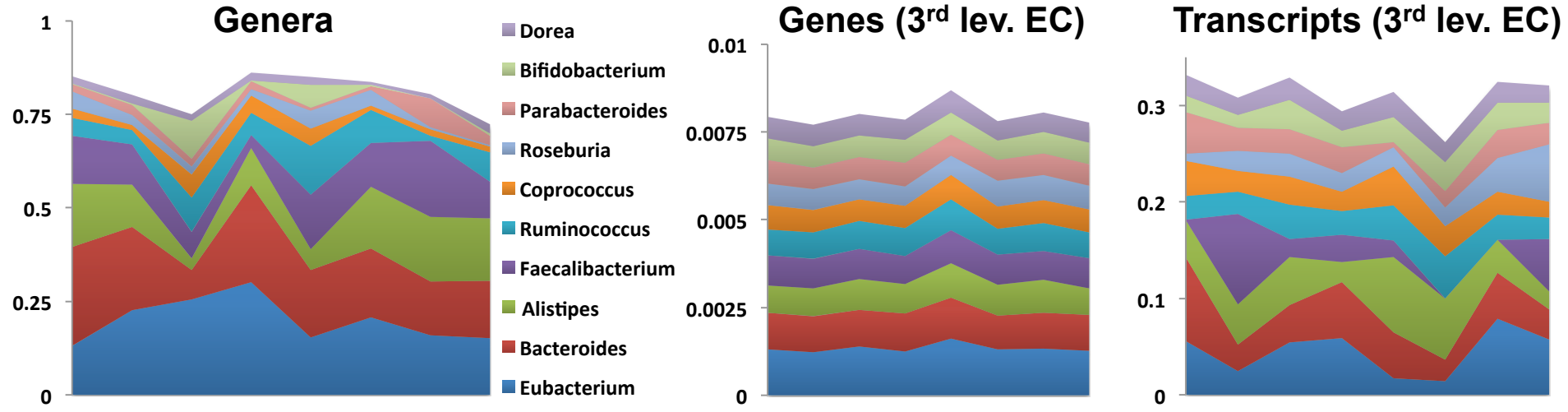


2) Relating the gut metagenome and metatranscriptome

- Microbial membership varies.
  - Early colonization? Genetics?
- Over time, the community "solves" for a habitat-specific metagenome.
- It then differentially regulates that metagenome.
  - These two types of regulation differ *at least* in time scale.

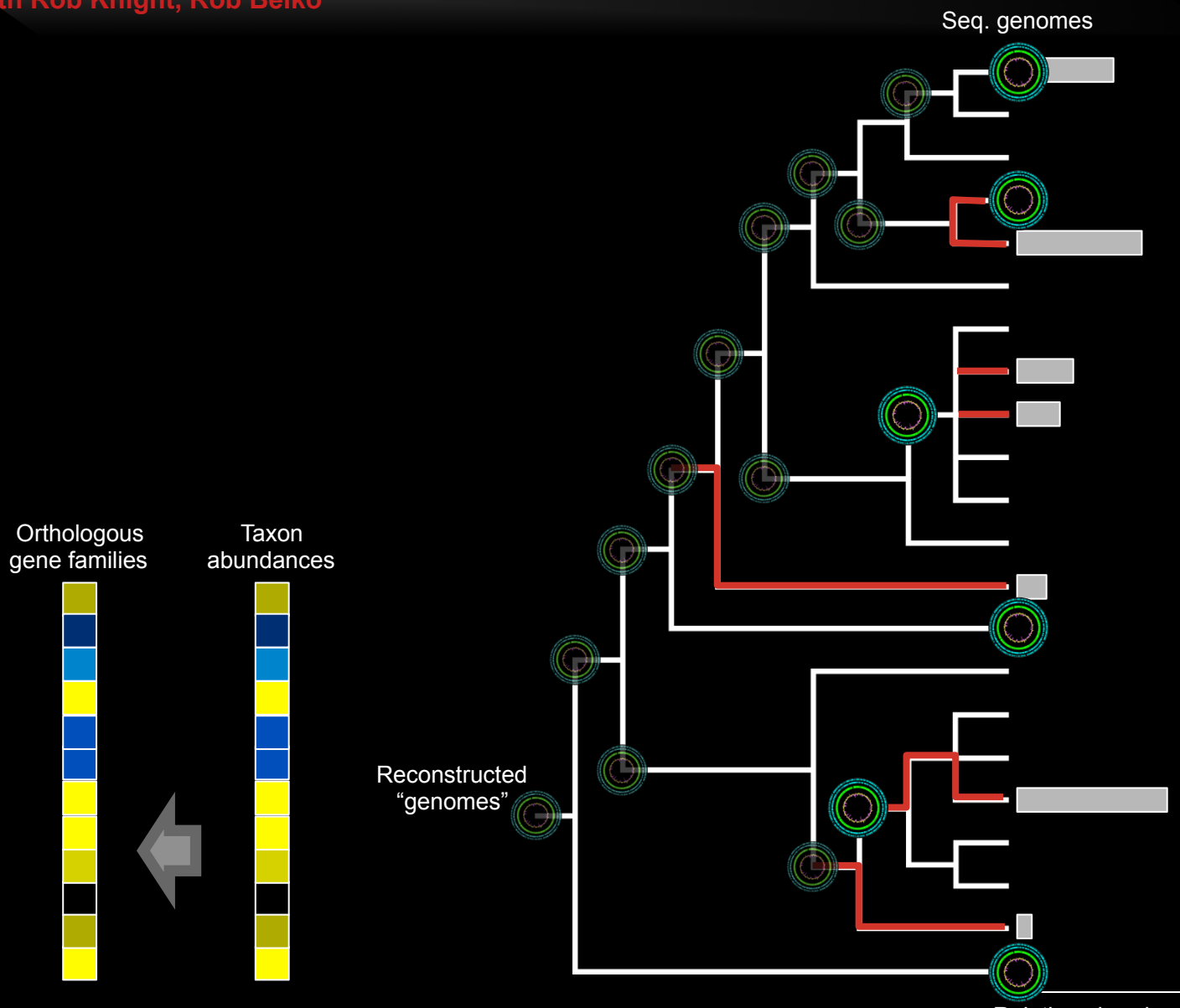


How consistent are the top ten...



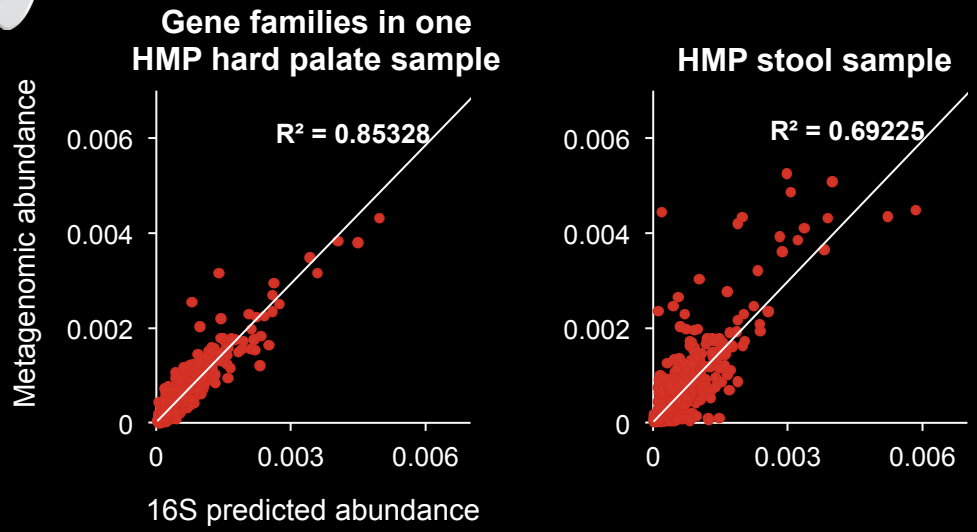
PICRUSt: Inferring community metagenomic potential from marker gene sequencing

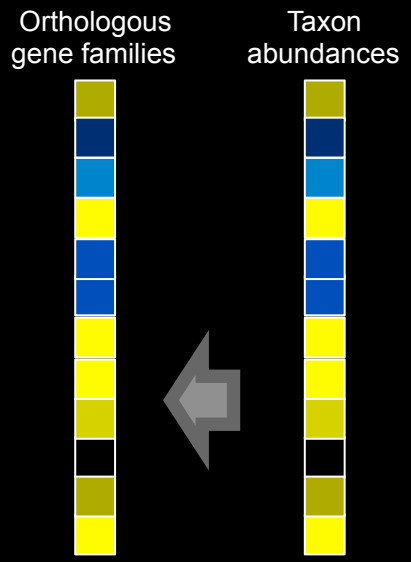
With Rob Knight, Rob Beiko

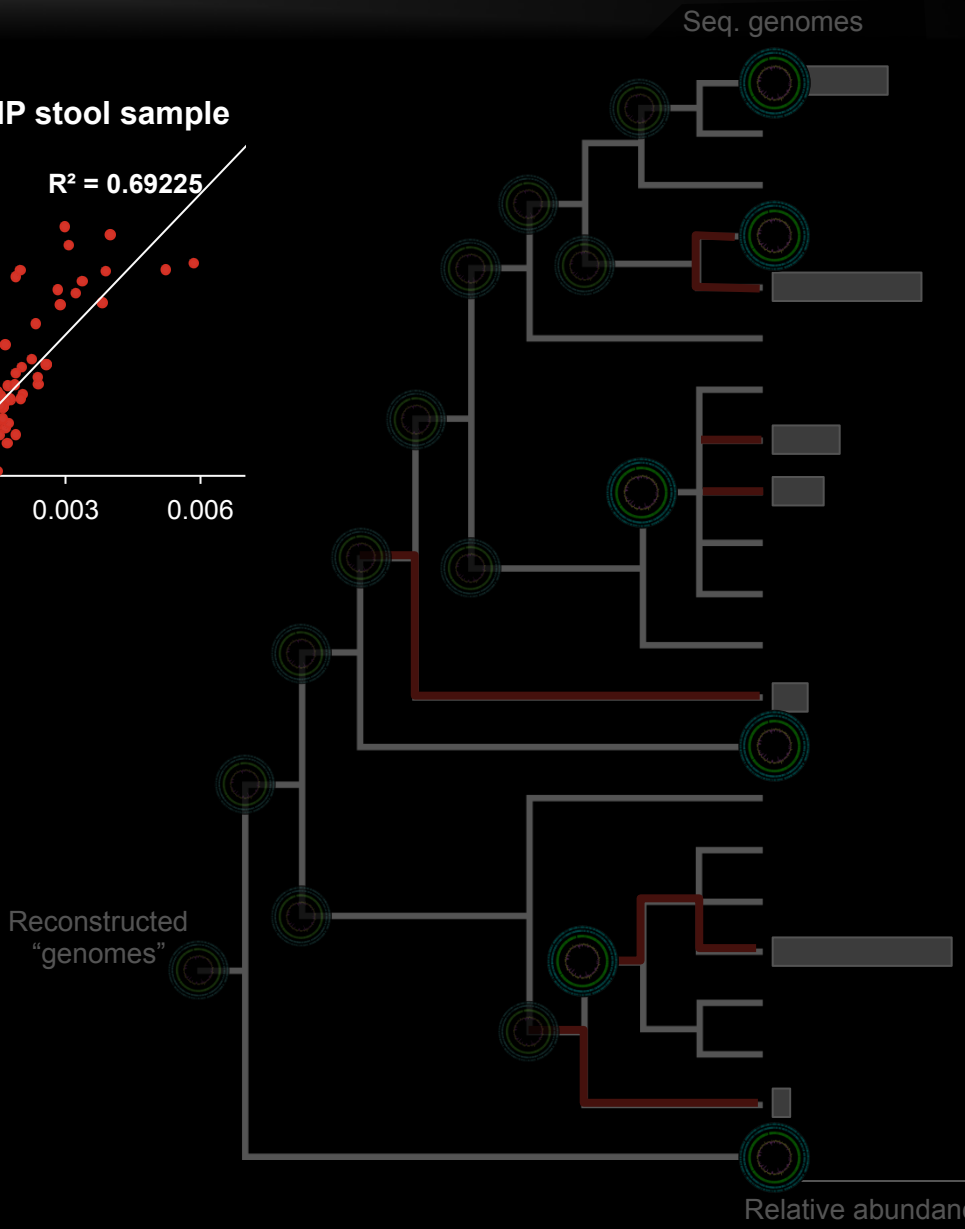


PICRUSt: Inferring community metagenomic potential from marker gene sequencing

With Rob Knight, Rob Beiko





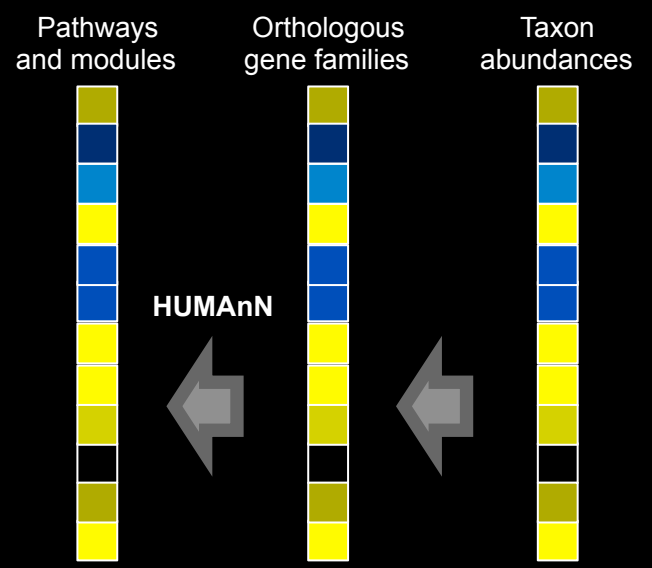


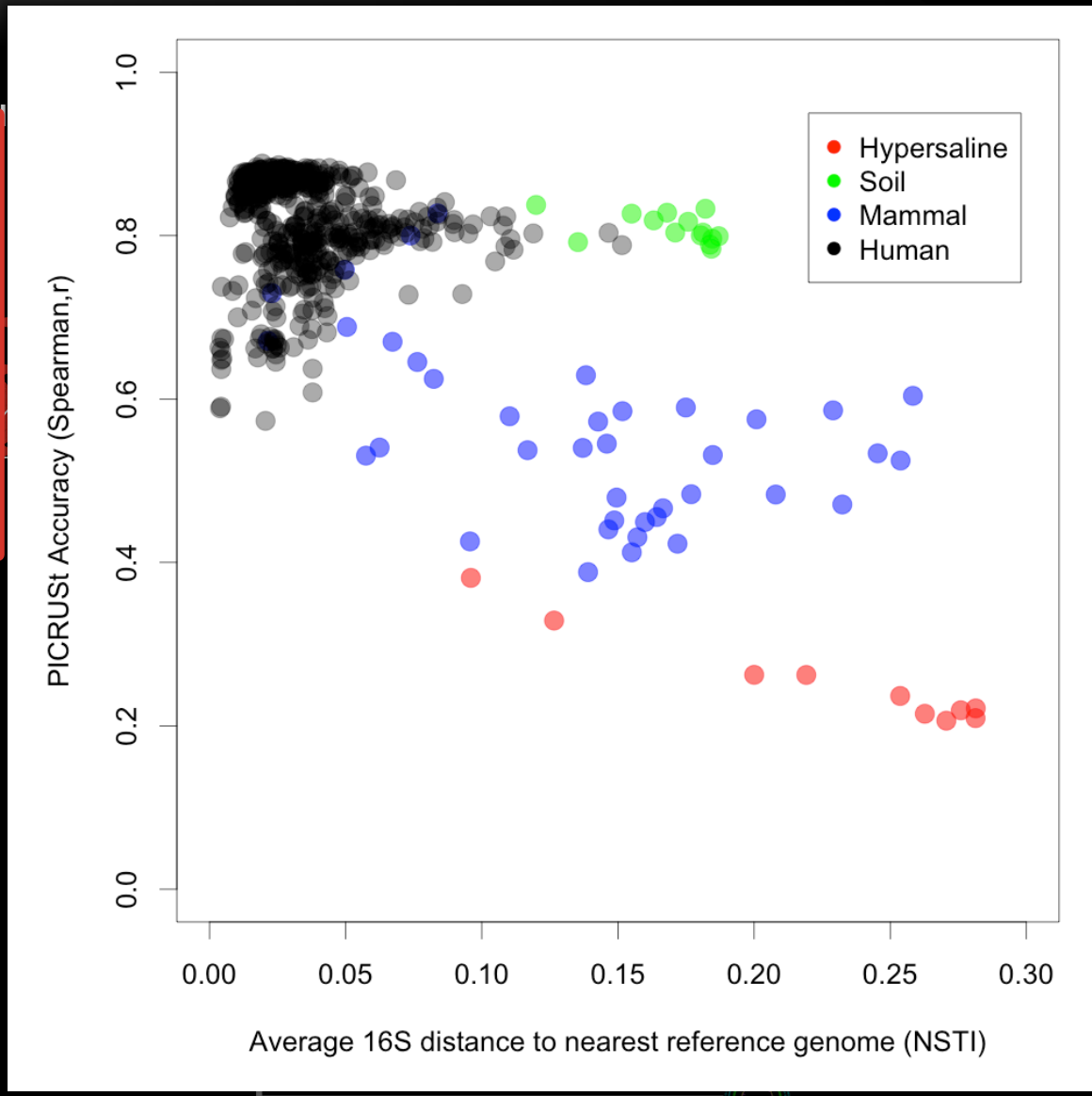
PICRUSt: Inferring community metagenomic potential from marker gene sequencing

With Rob Knight, Rob Beiko

One can recover ***general*** community function with reasonable accuracy from 16S profiles.

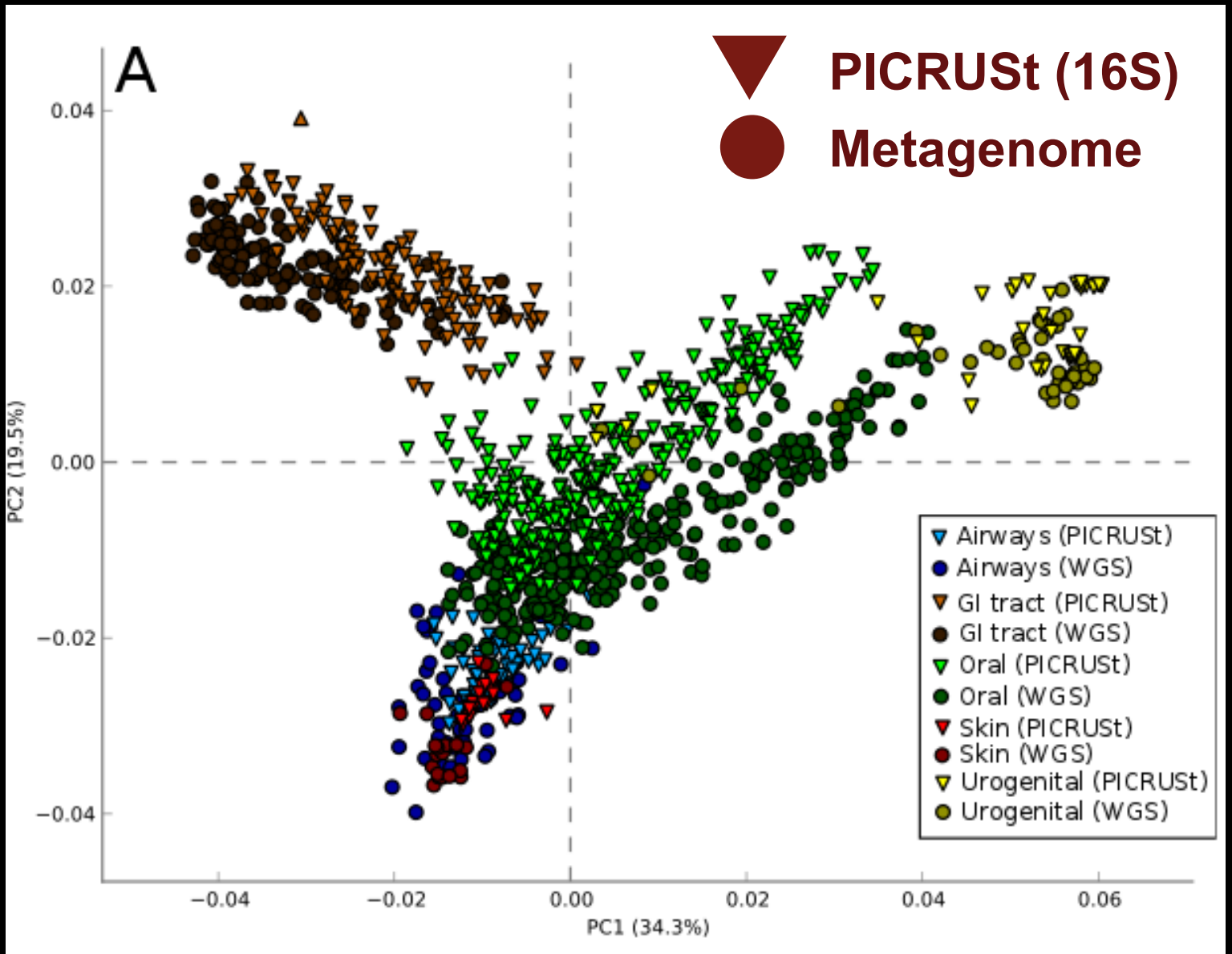
http://picrust.github.com





Relative abundance

It's not shotgun sequencing, but it's not too shabby, either



Ask *both* what you can do for your microbiome *and* what your microbiome can do for you

Thanks!


Nicola Segata

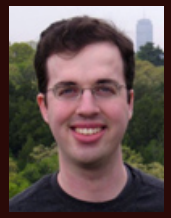

Levi Waldron


Xochi Morgan

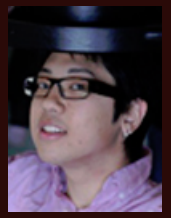

Tim Tickle


Vagheesh Narasimhan


Emma Schwager


Eric Franzosa


Daniela Boernigen


Joseph Moon


Jim Kaminski


Craig Bielski


Brian Palmer


Ren Lu


Hufeng Zhou


Dirk Gevers
Kat Huang


Ramnik Xavier
Harry Sokol
Dan Knights
Moran Yassour


Wendy Garrett
Michelle Rooks


Rob Beiko
Morgan Langille


Rob Knight
Greg Caporaso
Jesse Zaneveld


Mark Silverberg
Boyko Kabakchiev
Andrea Tyler

Ruth Ley
Omry Koren

Jacques Izard
Katherine Lemon

Bruce Sands

Human Microbiome Project

Owen White
Joe Petrosino
George Weinstock
Karen Nelson
Lita Proctor
Erica Sodergren
Anthony Fodor
Marty Blaser
Jacques Ravel
Pat Schloss
Sahar Abubucker
Brandi Cantarel
Alyx Schubert
Mathangi Thiagarajan
Beltran Rodriguez-Mueller
Makedonka Mitreva
Yuzhen Ye
Mihai Pop
Larry Forney
Barbara Methe

Bruce Birren Mark Daly
Doyle Ward Ashlee Earl

BROAD INSTITUTE

http://huttenhower.sph.harvard.edu