

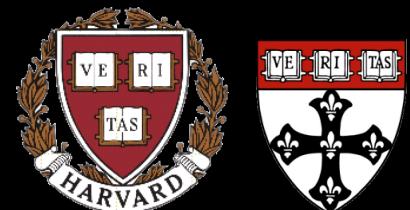


Nicola
Segata



Strain profiling with StrainPhIAn and PanPhIAn

Curtis Huttenhower



Harvard T.H. Chan School of Public Health
Department of Biostatistics

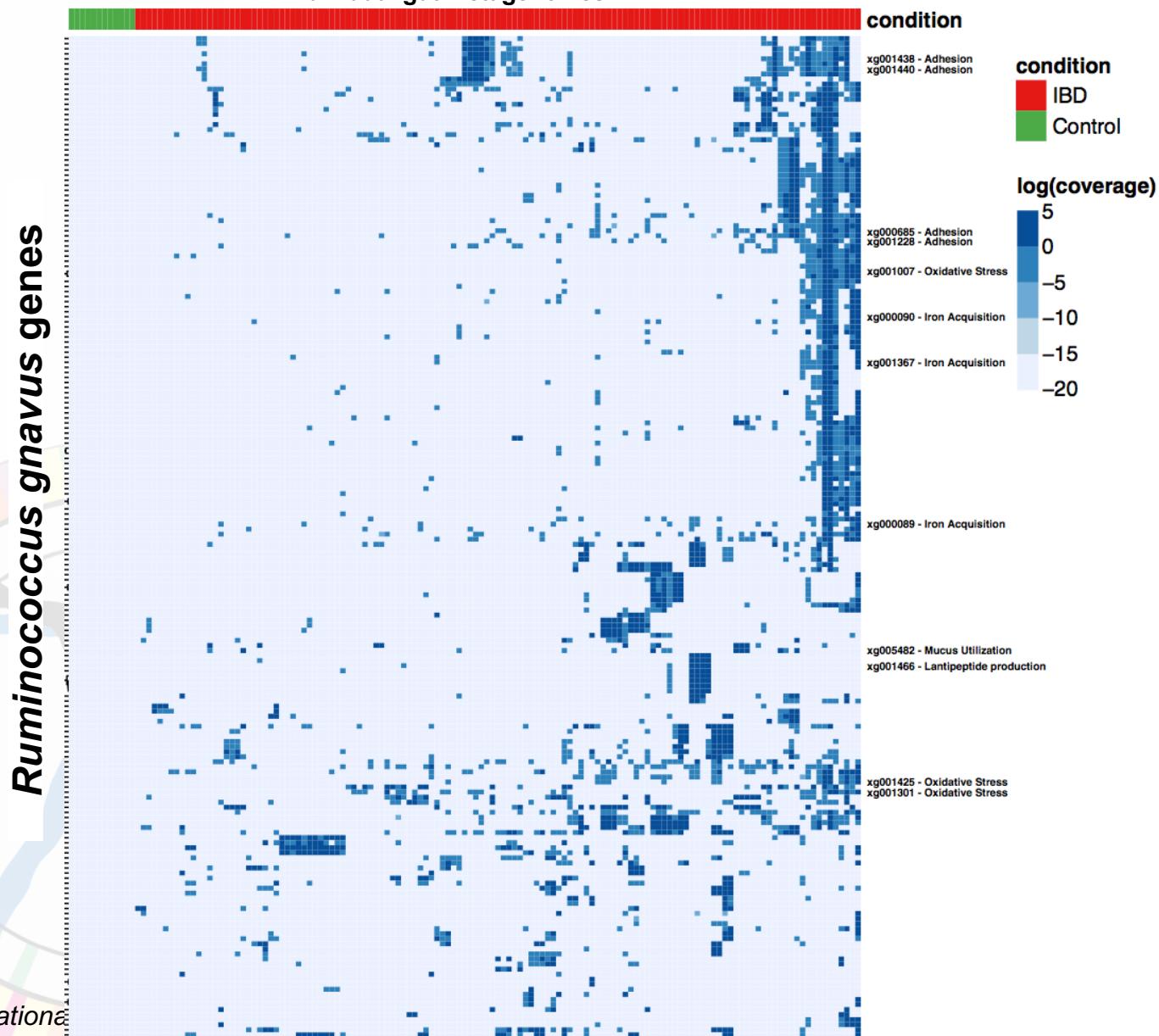
07-31-19



Linking microbial function to strain-specific phenotypes in IBD

Individual gut metagenomes

<http://segatalab.cibio.unitn.it/tools/panphlan>



Funded by National



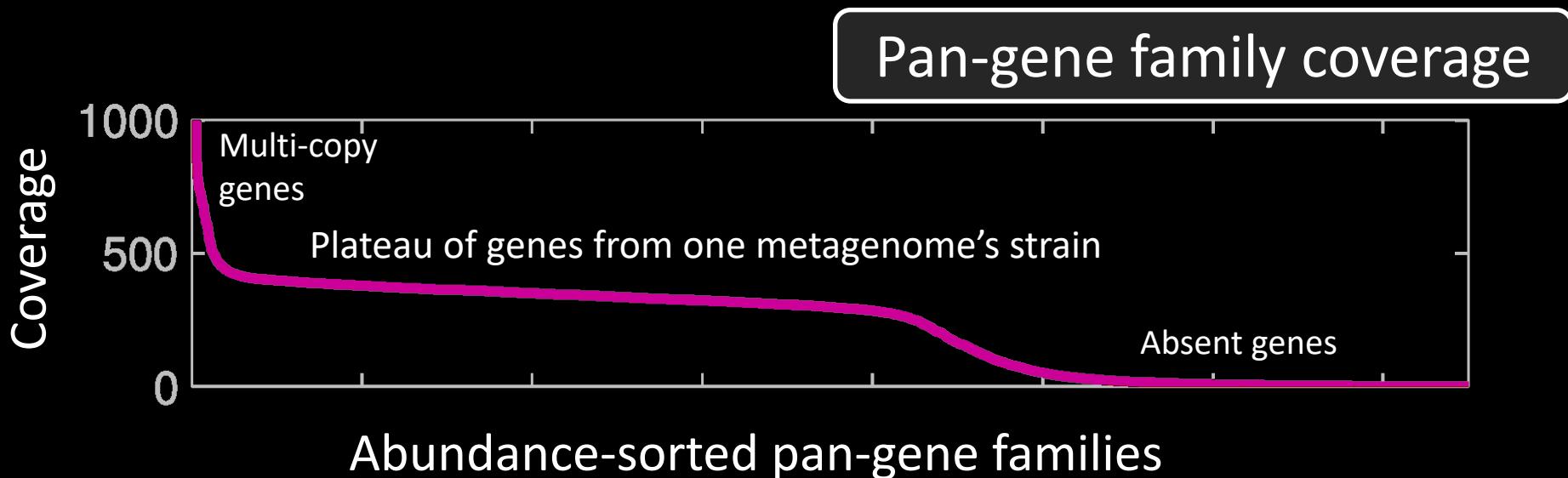
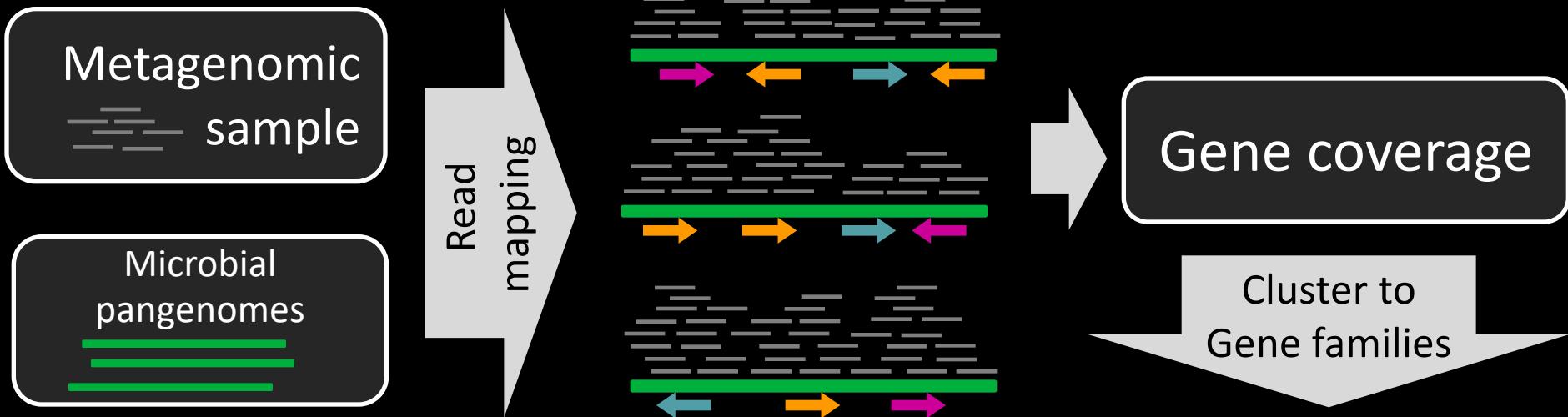
Brantley Hall Moran Yassour



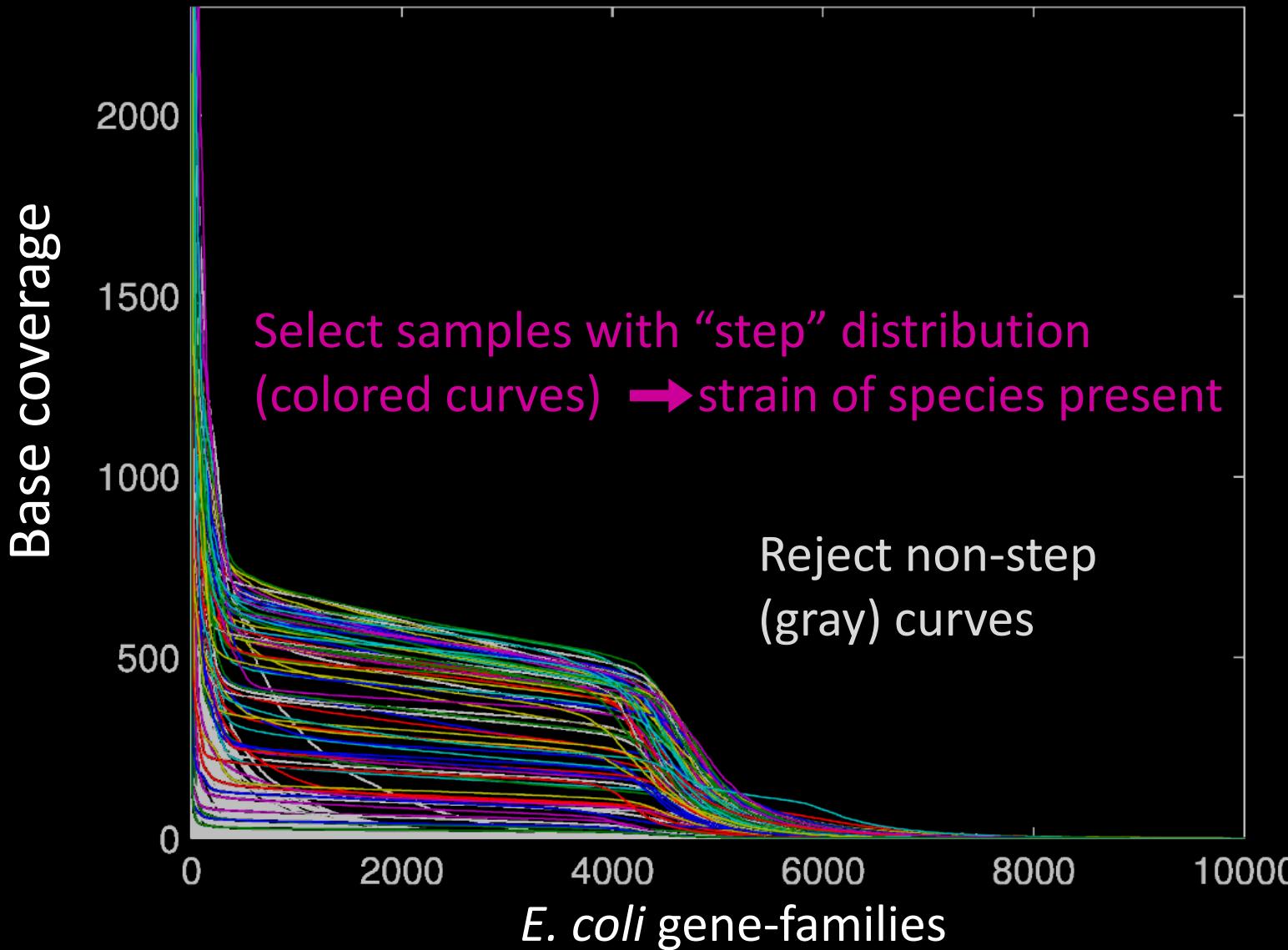


PanPhlAn: the approach

<http://bitbucket.org/CibioCM/panphlan>



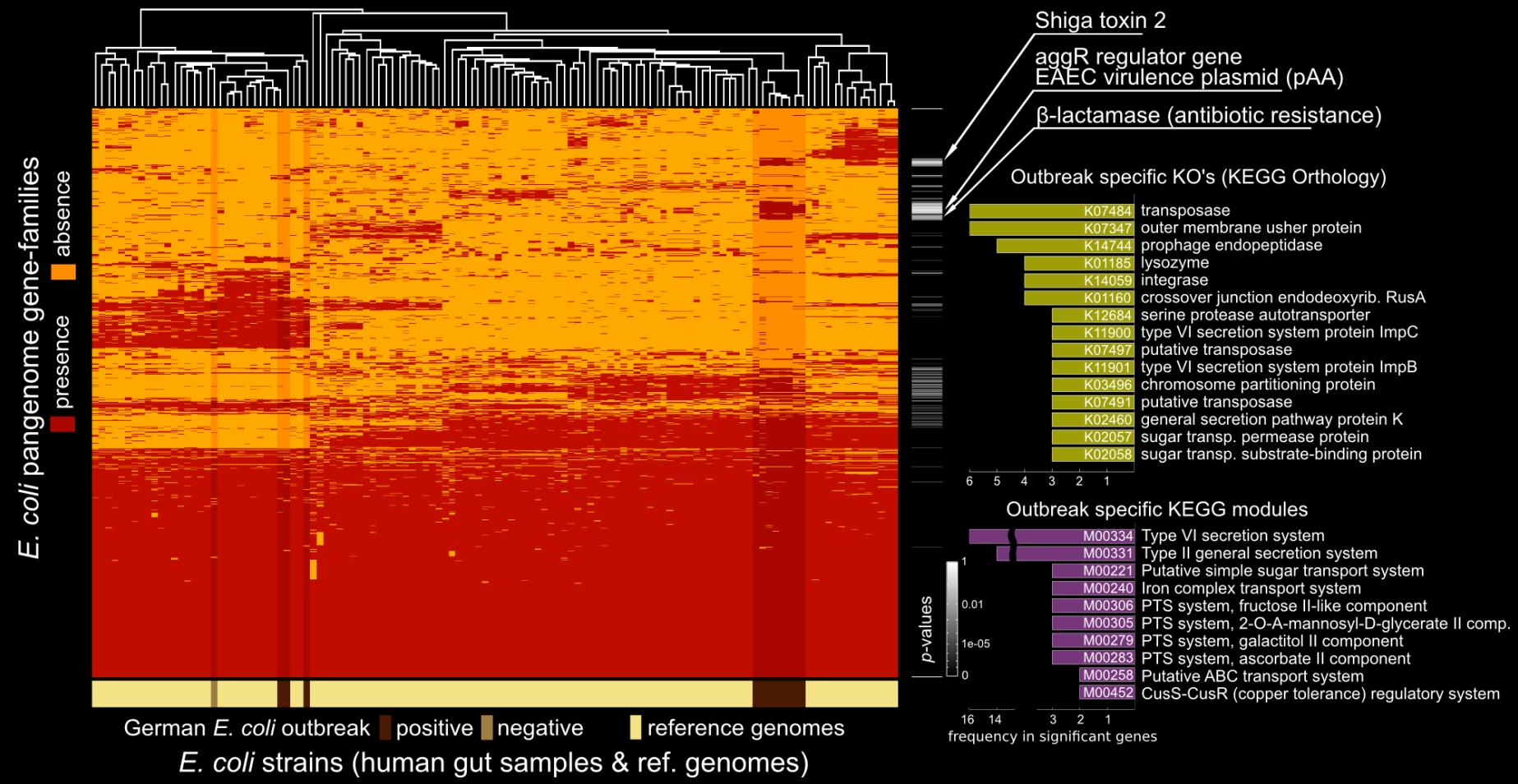
Gene-family distribution curves



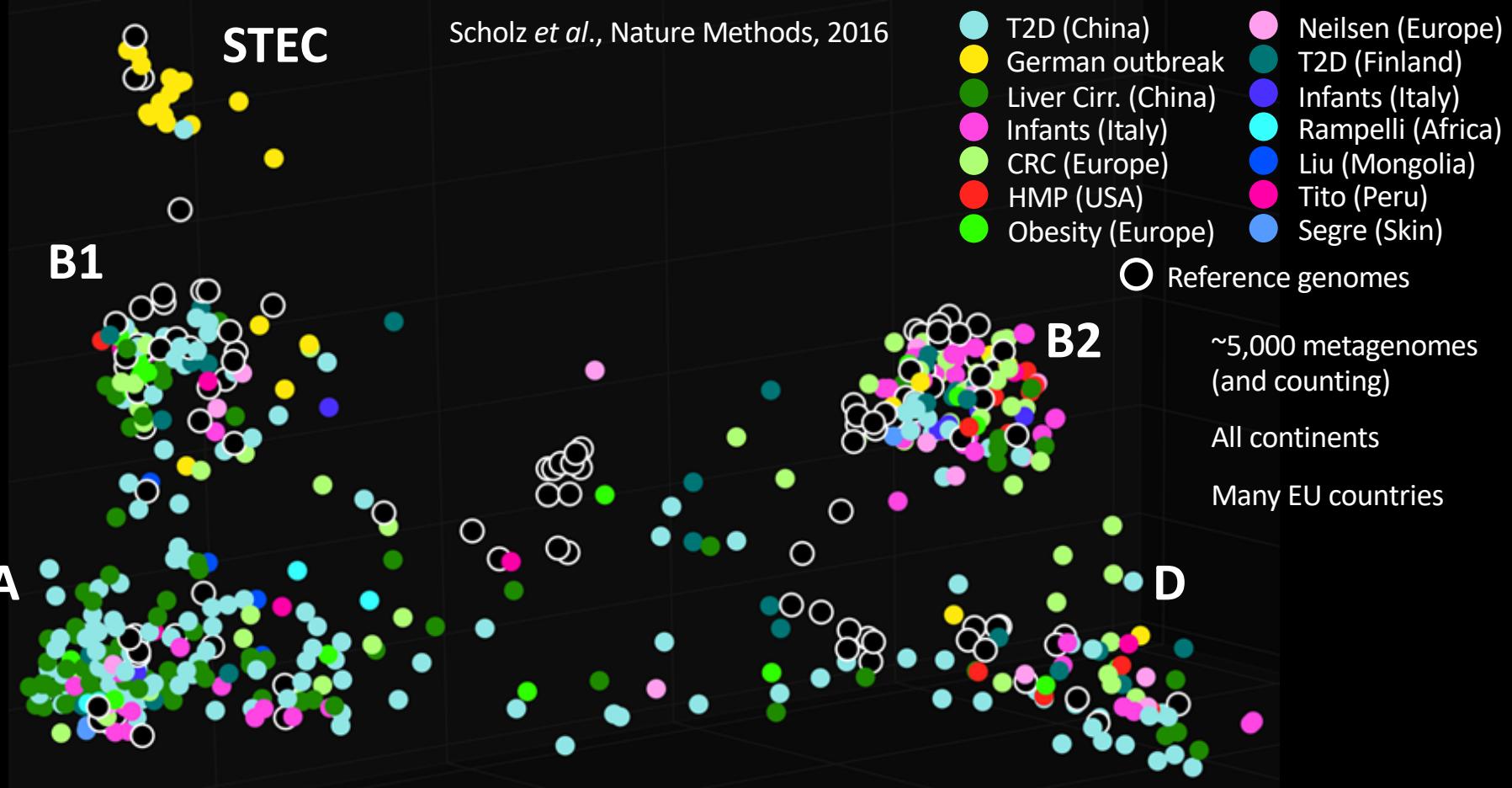


PanPhlAn for “meta-epidemiology”

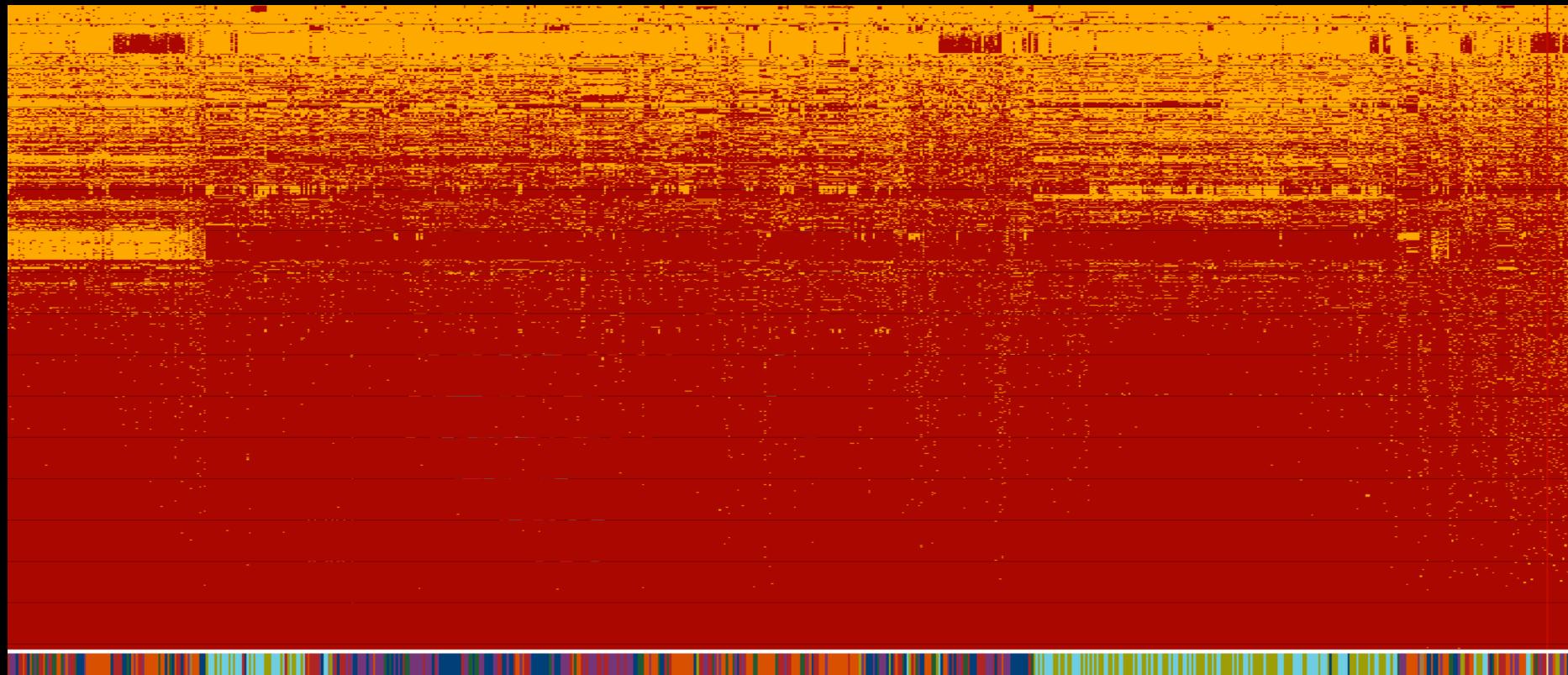
<http://bitbucket.org/CibioCM/panphlan>



Strain-level epidemiology of human-associated *E. coli* with PanPhlAn



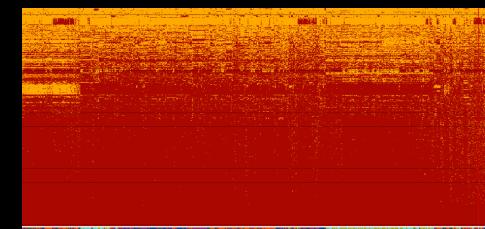
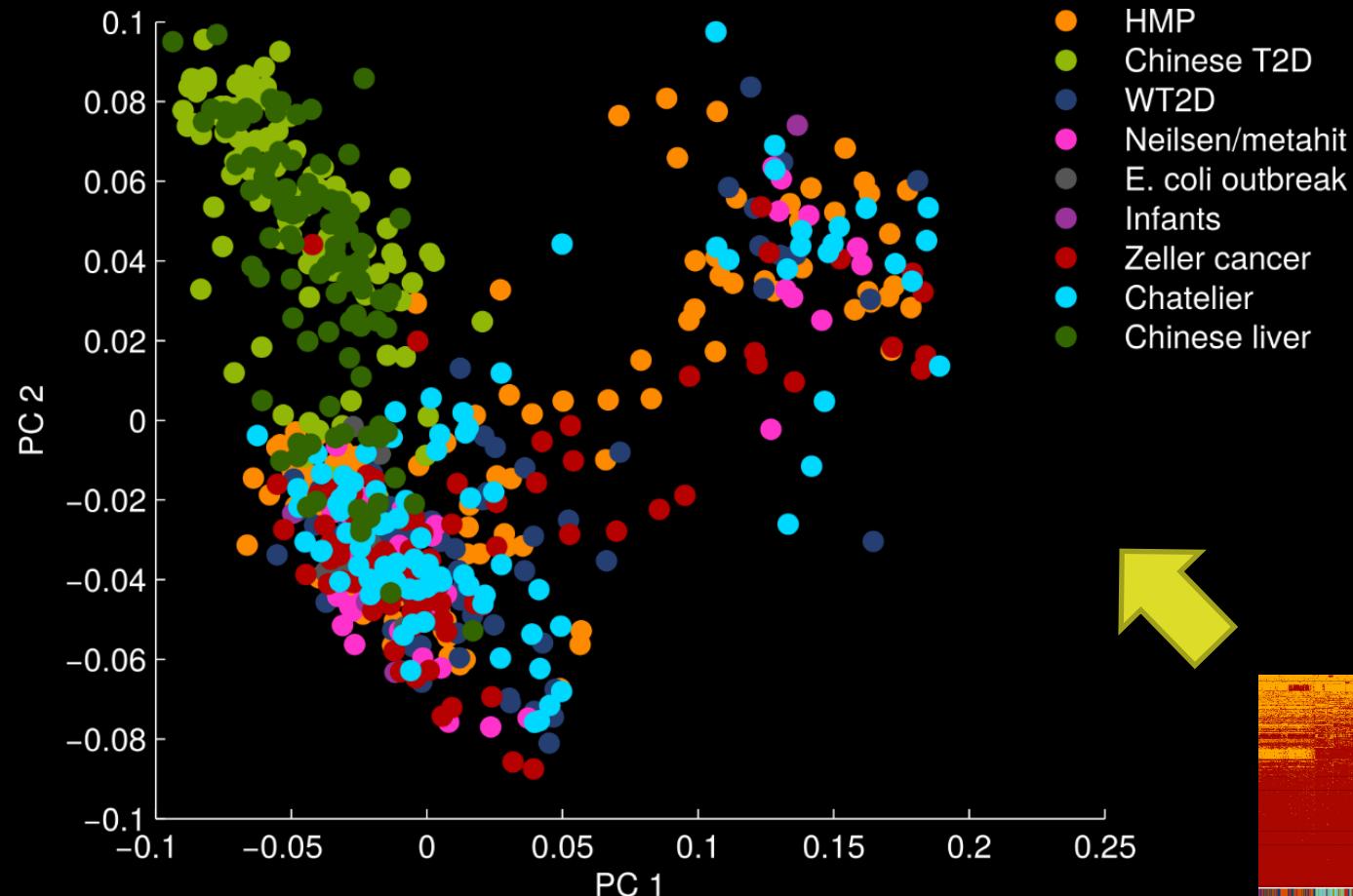
PanPhlAn on *Eubacterium rectale*



- HMP
- Chinese T2D
- WT2D
- Nielsen/metahit
- E. coli outbreak
- Infants
- Zeller cancer
- Chatelier
- Chinese liver

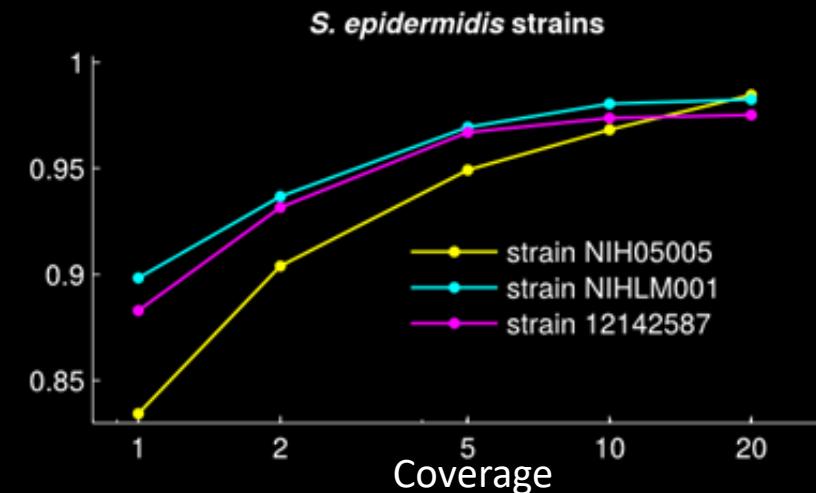
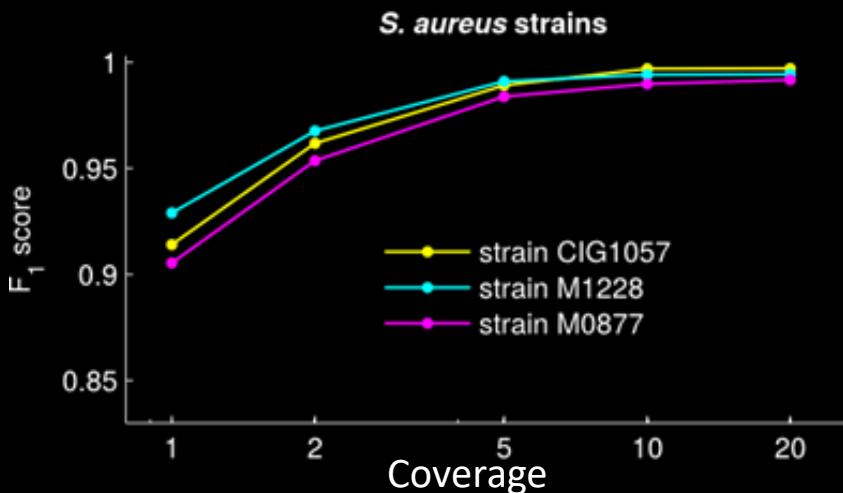
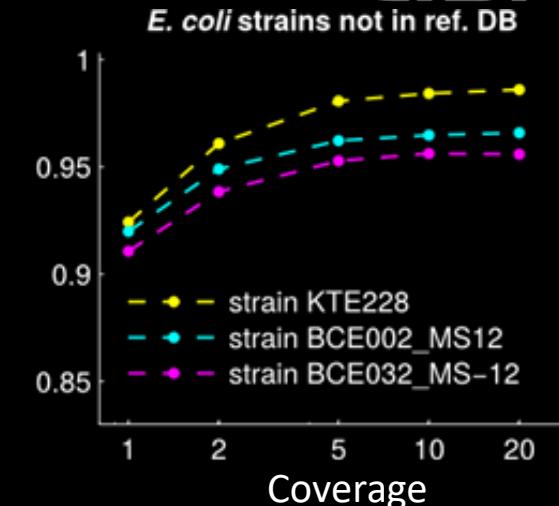
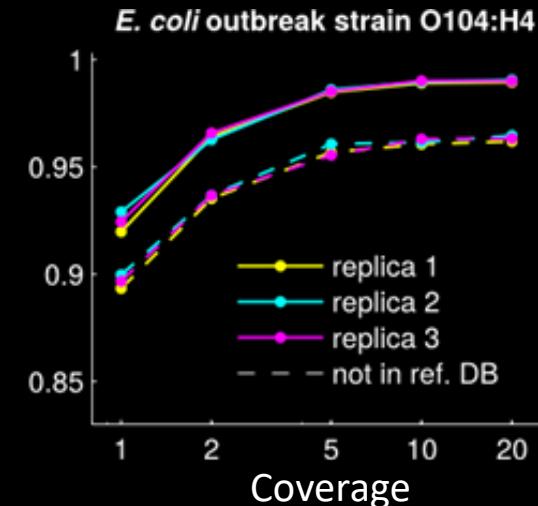
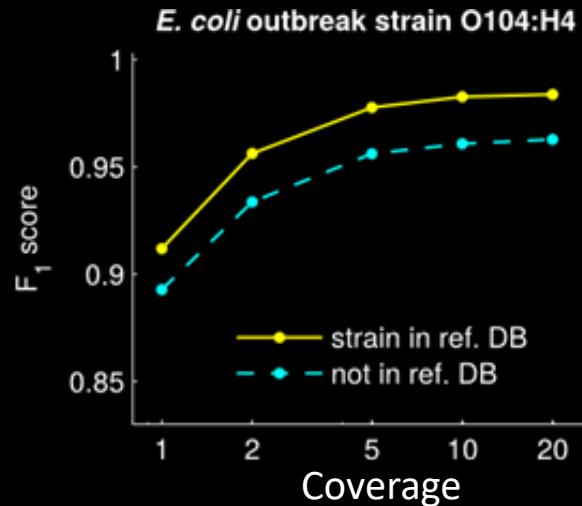
Only one *Eubacterium rectale* genome used here

PanPhlAn on *Eubacterium rectale*

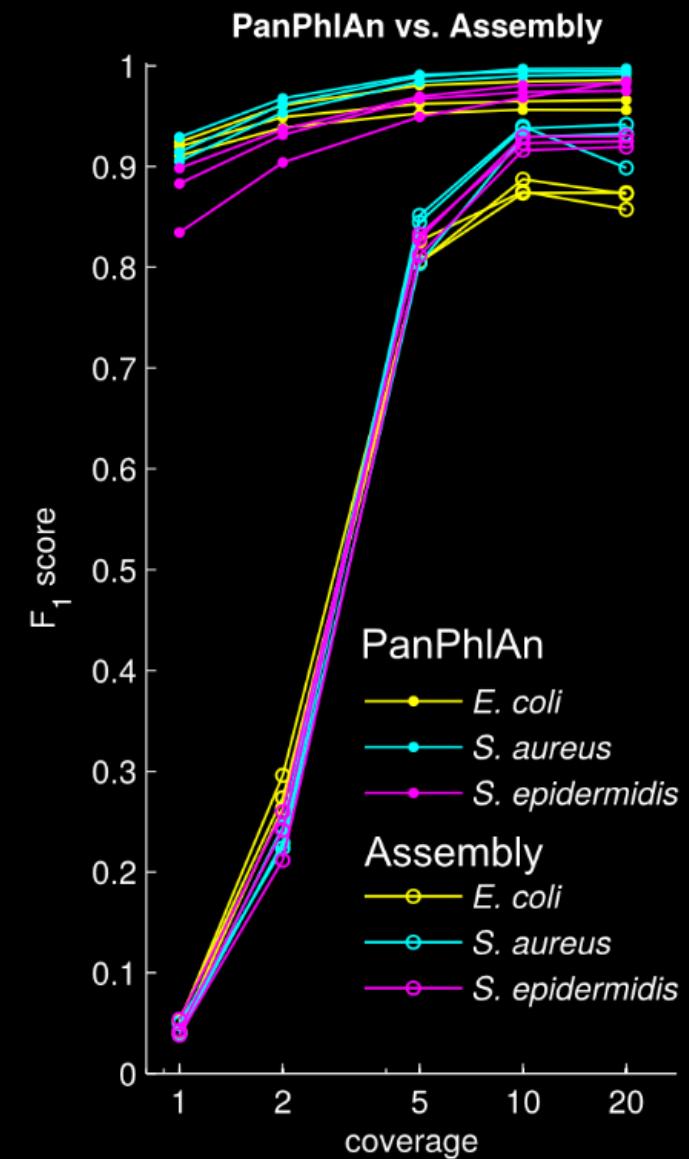
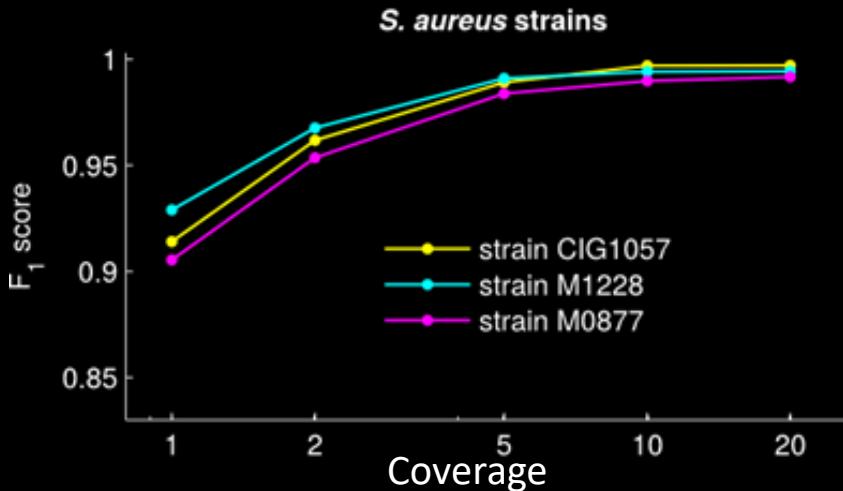
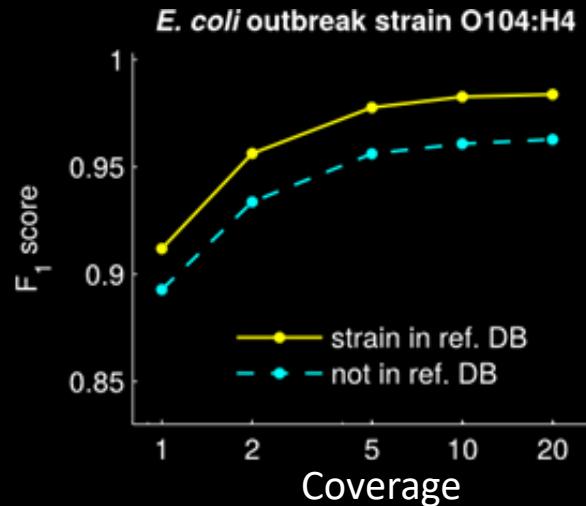




Synthetic and semi-synthetic validation



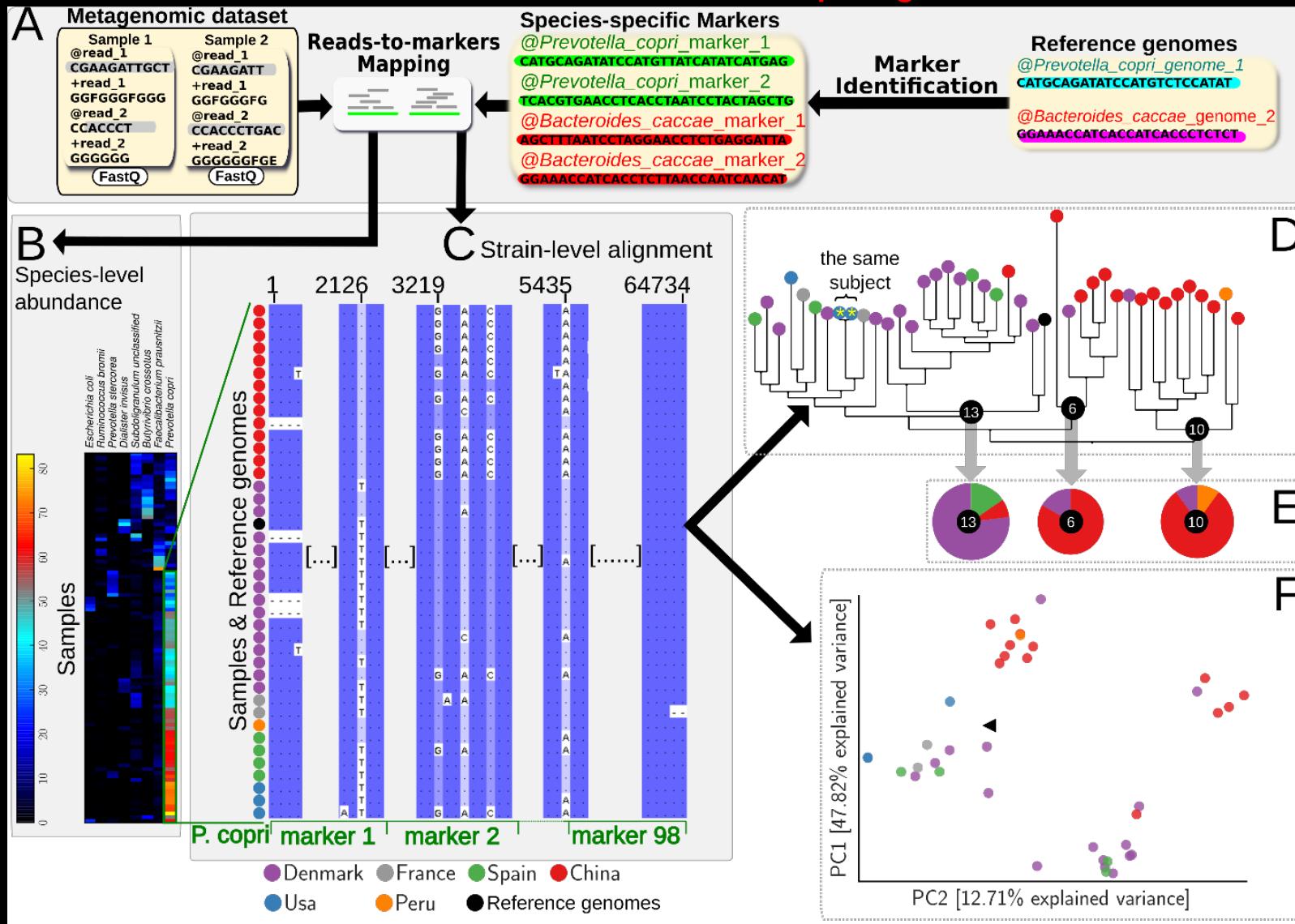
Synthetic and semi-synthetic validation





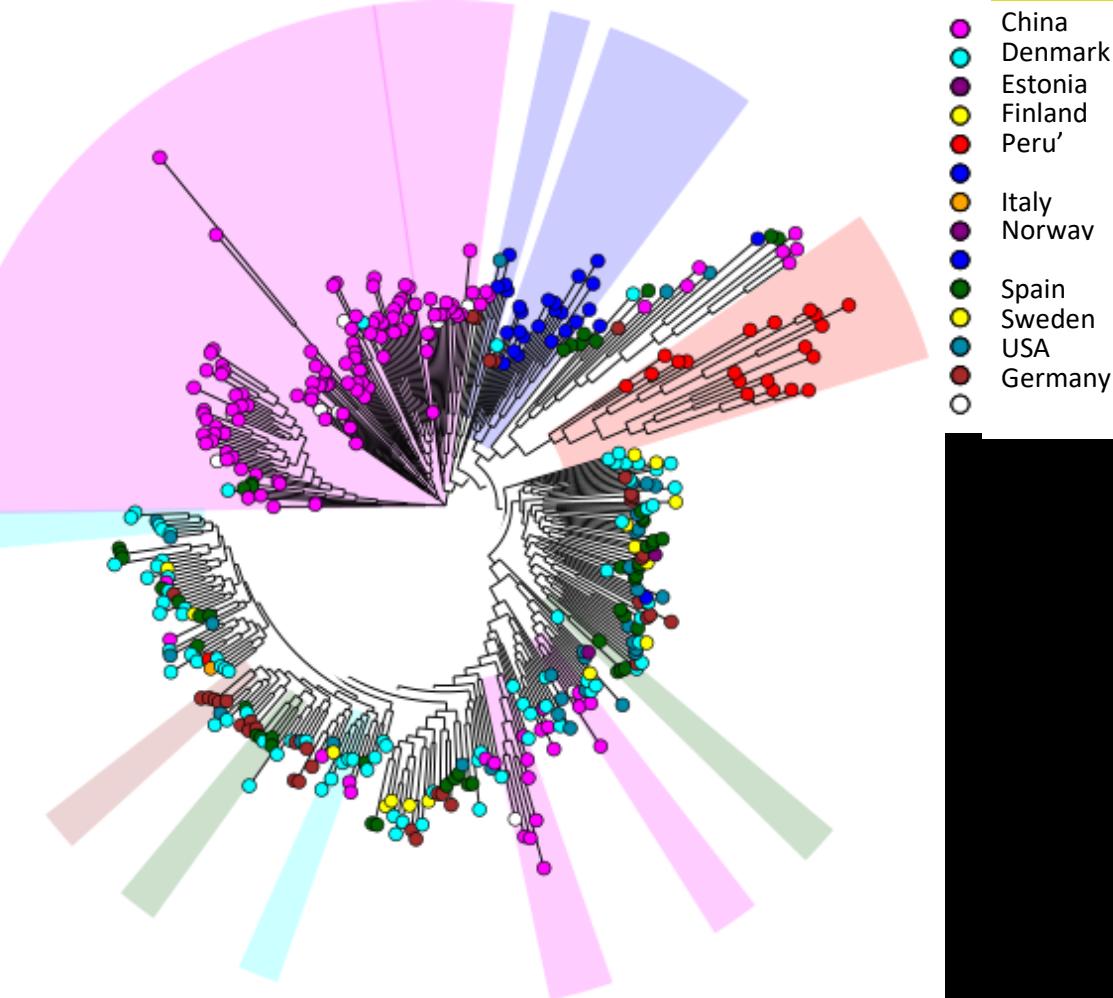
StrainPhIAn: metagenomic strain identification and tracking

<http://segatalab.cibio.unitn.it/tools/strainphian>





A tool for strain level population genomics

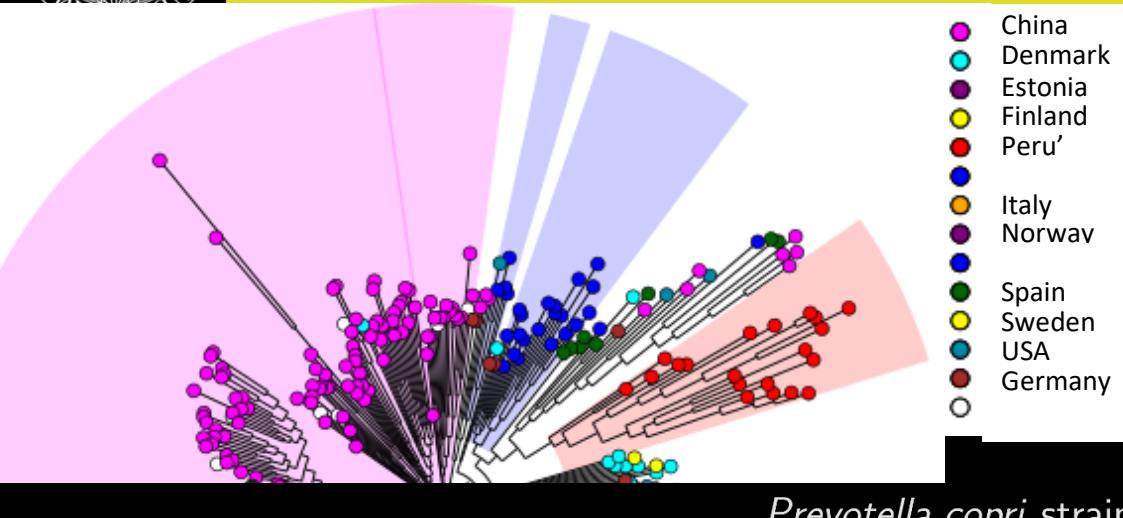


P. copri as an example species
Alignment length: 66k nt
Median SNPs: 830 [3.6%]
pos. samples: 123

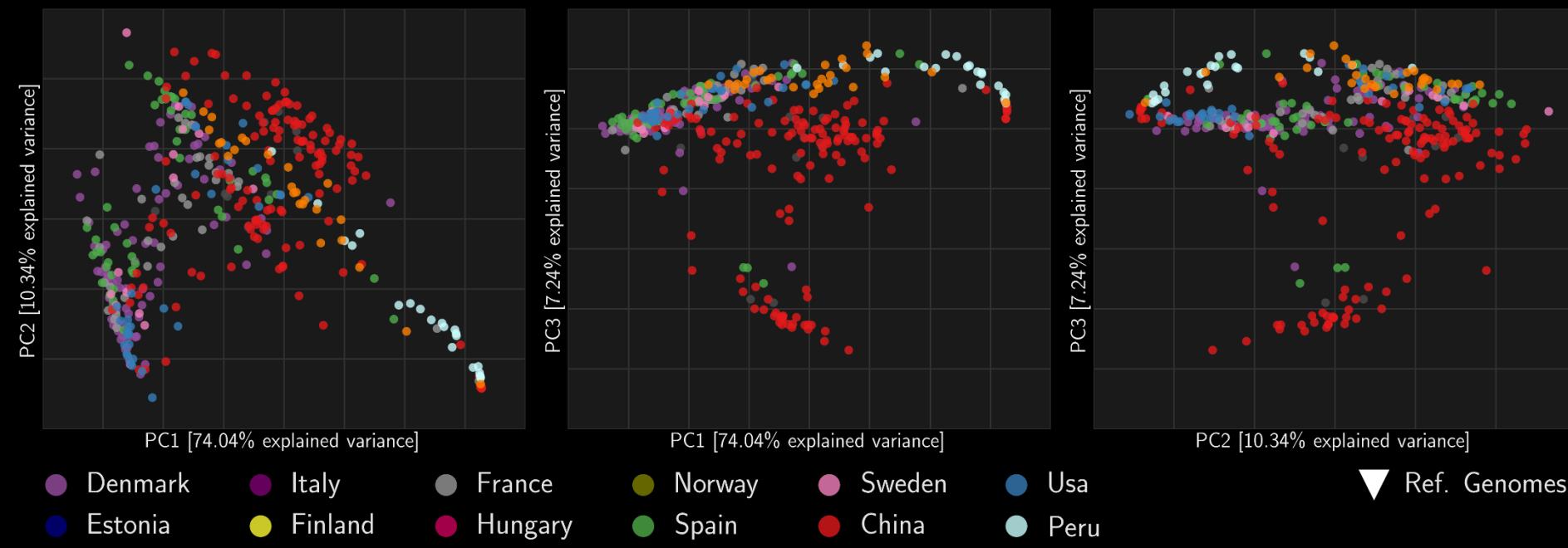




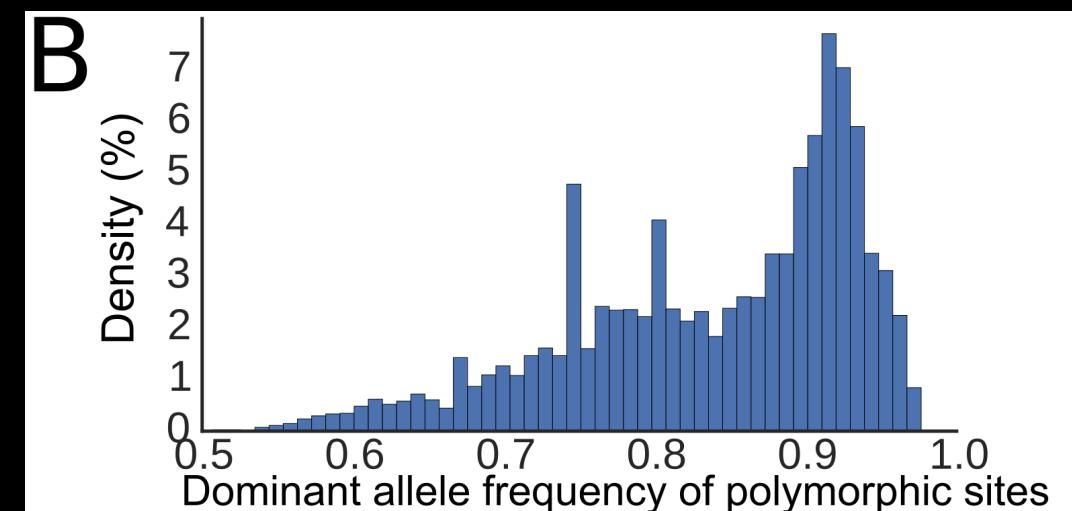
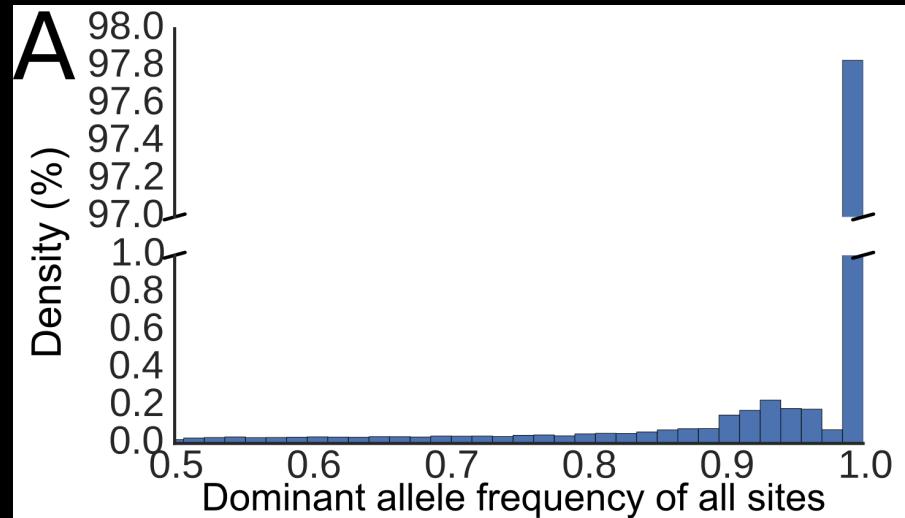
A tool for strain level population genomics



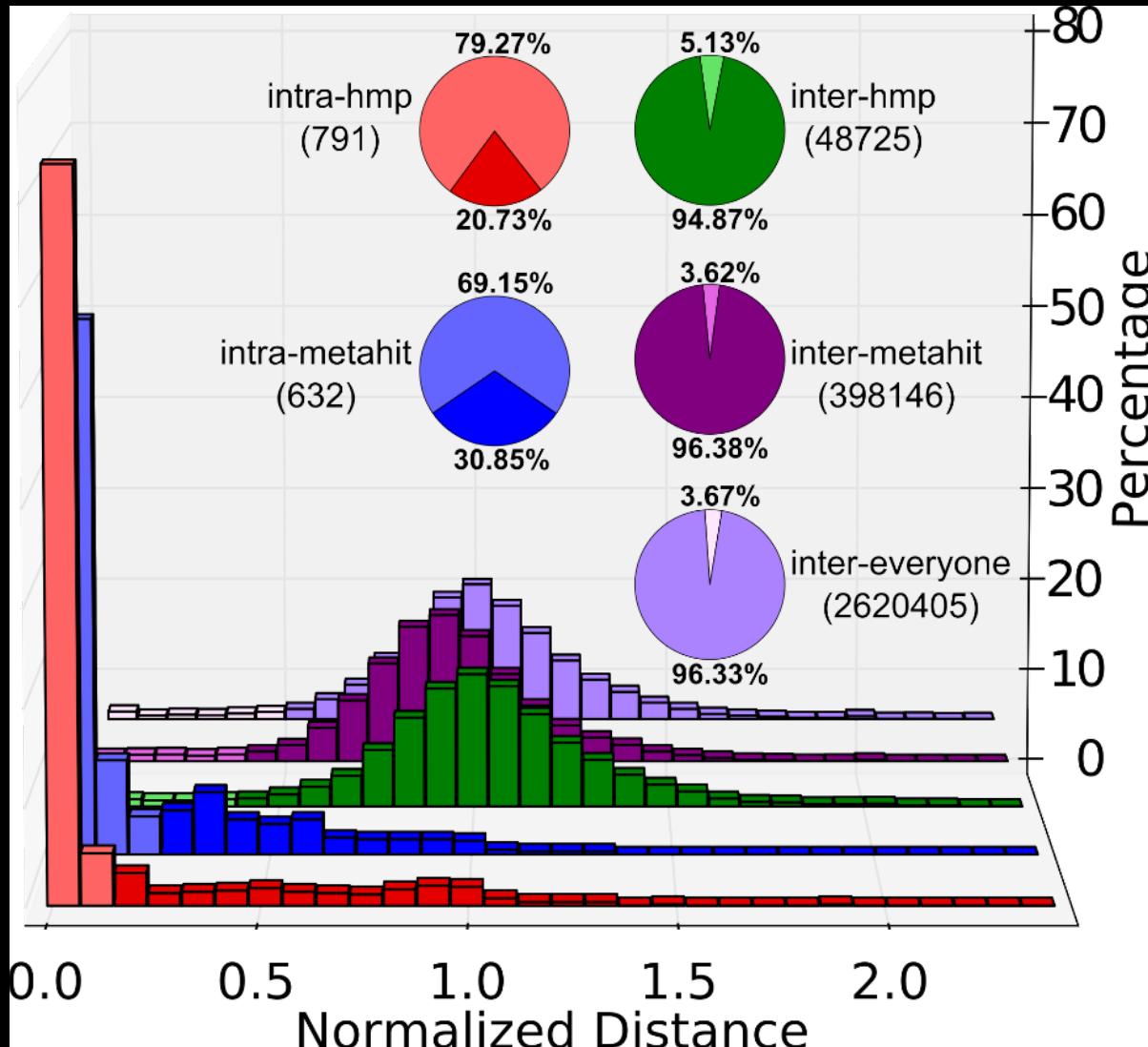
P. copri as an example species
Alignment length: 66k nt
Median SNPs: 830 [3.6%]
pos. samples: 123



Most bugs (in the gut) are dominated by one stable strain

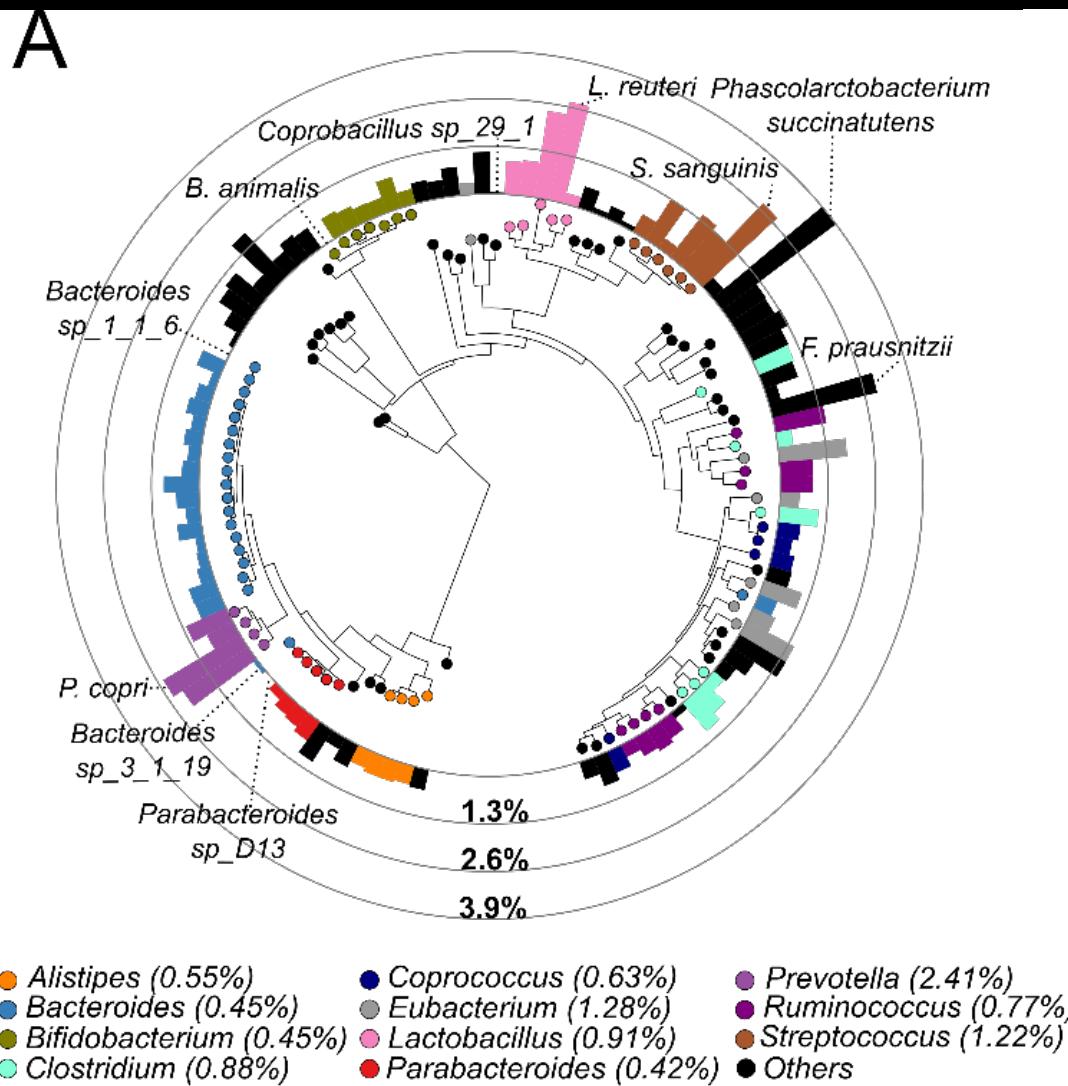


Most bugs (in the gut) are dominated by one stable strain





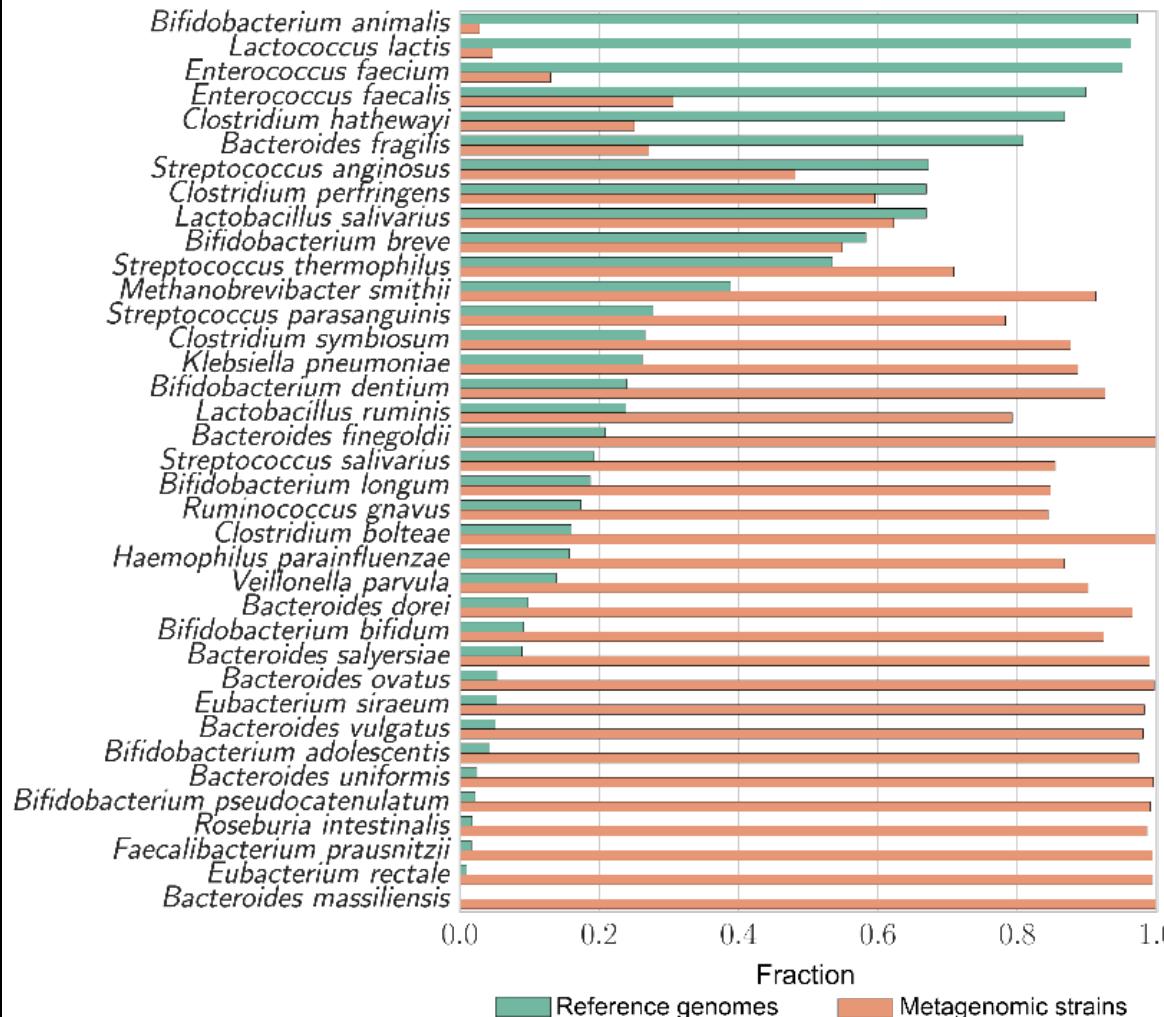
There's a lot of strain-level variation left to discover





There's a lot of strain-level variation left to discover

B



Phylogenetic branch % spanned by reference vs. “wild” bugs

Multiple options for strain tracking in metagenomes

- StrainPhlAn:
 - Map reads to core markers and call SNPs.
 - Requires ~10x coverage, ~0.1% error rate.
- PanPhlAn:
 - Map reads to pan-genomes and identify absent genes.
 - Requires ~1x coverage, ~1% error rate.
- Both work uniquely well for meta-analysis.
 - Not sensitive to typical batch effects.



StrainPhlAn tutorial

<https://bitbucket.org/biobakery/biobakery/wiki/strainphlan>

StrainPhlAn

Identify and track
microbial strains using
SNPs in conserved,
unique marker genes

Input

- Shotgun metagenome sequencing results (e.g fasta)

Output

- Phylogenetic resolution of subspecies OTUs and multiple sequence alignment

Visualization

- Phylogenetic trees and ordination plots