



Eric
Franzosa

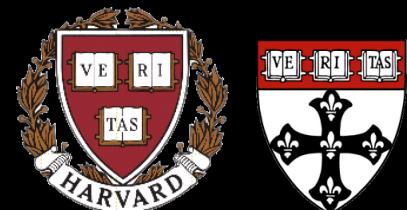


Jason
Lloyd-Price



Functional profiling with HUMAnN2

Curtis Huttenhower



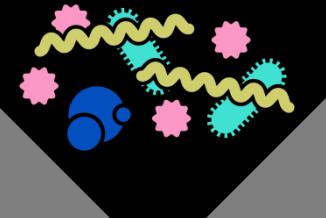
Harvard T.H. Chan School of Public Health
Department of Biostatistics

07-31-19

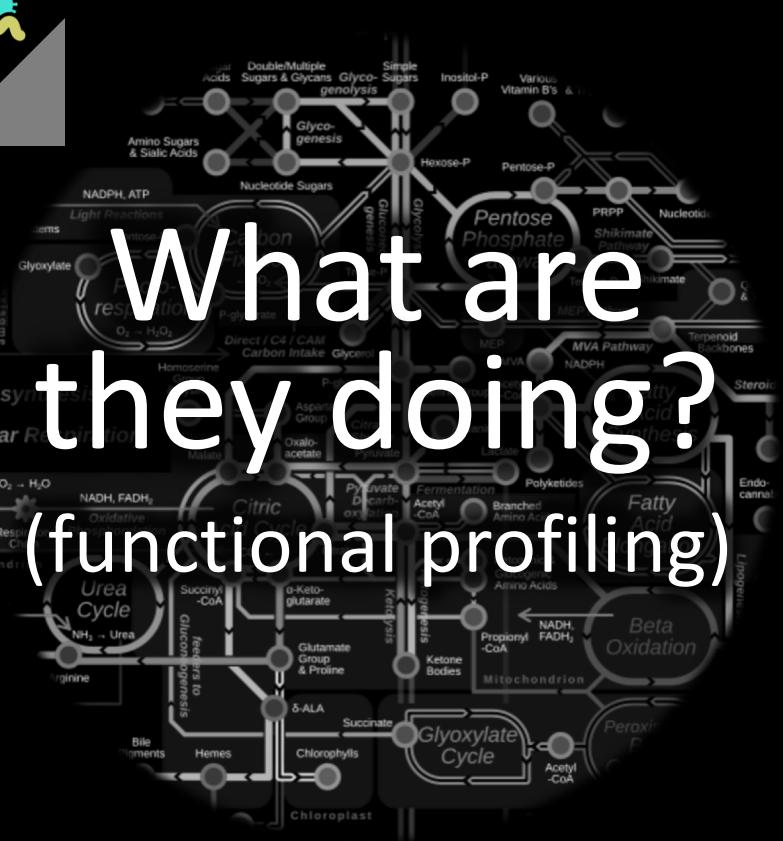


The two big questions of microbial community profiling:

Who is there?
(taxonomic profiling)



What are they doing?
(functional profiling)





HUMAnN2: stratified output

UniRef gene cluster	Gene name	Total gene abundance (RPK)
UniRef90_R6K3Z5: IMP dehydrogenase		600.95
UniRef90_R6K3Z5: IMP dehydrogenase <i>Bacteroides_caccae</i>		234.76
UniRef90_R6K3Z5: IMP dehydrogenase <i>Bacteroides_dorei</i>		107.38
UniRef90_R6K3Z5: IMP dehydrogenase <i>Bacteroides_ovatus</i>		92.18
UniRef90_R6K3Z5: IMP dehydrogenase <i>Bacteroides_stercoris</i>		83.95
UniRef90_R6K3Z5: IMP dehydrogenase <i>Bacteroides_vulgatus</i>		57.27
UniRef90_R6K3Z5: IMP dehydrogenase unclassified		25.41

~HUMAnN1

Per-species & unclassified stratifications

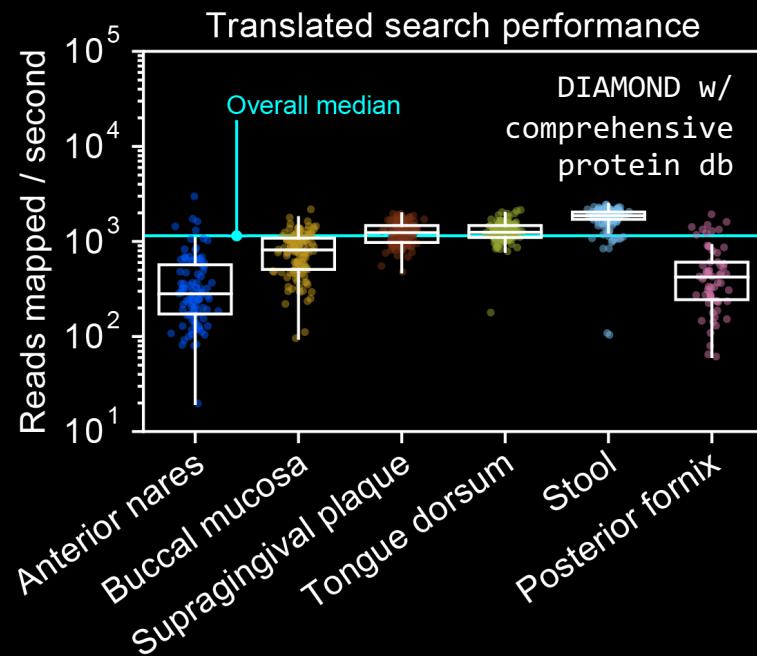
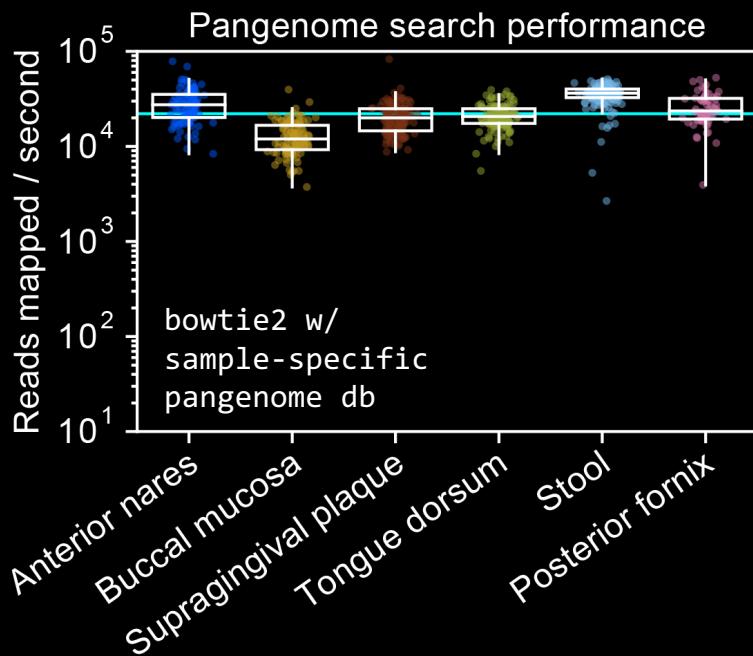
MetaCyc pathway	Pathway abundance	& coverage
PWY-7221: GTP biosynthesis	200.35	1
PWY-7221: GTP biosynthesis <i>Bacteroides_caccae</i>	120.23	1
PWY-7221: GTP biosynthesis <i>Bacteroides_dorei</i>	11.12	0



HUMAnN2 real-world performance

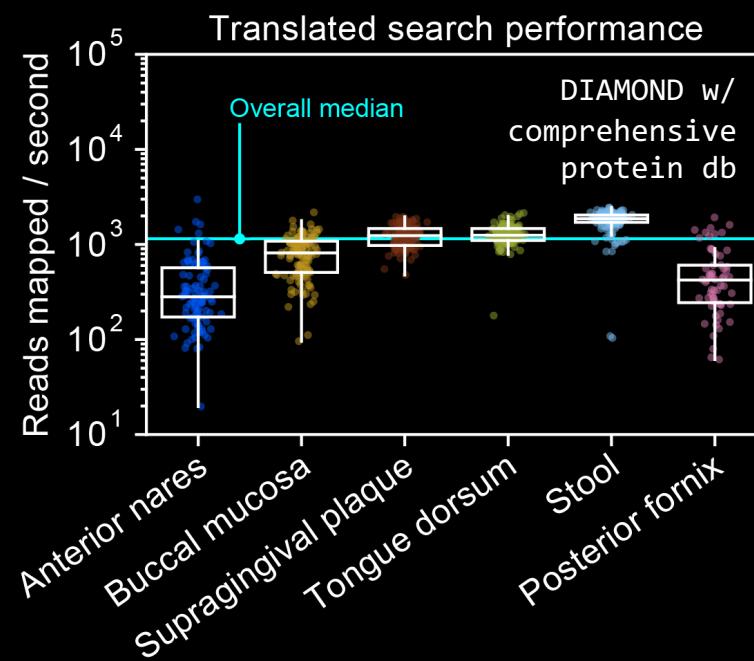
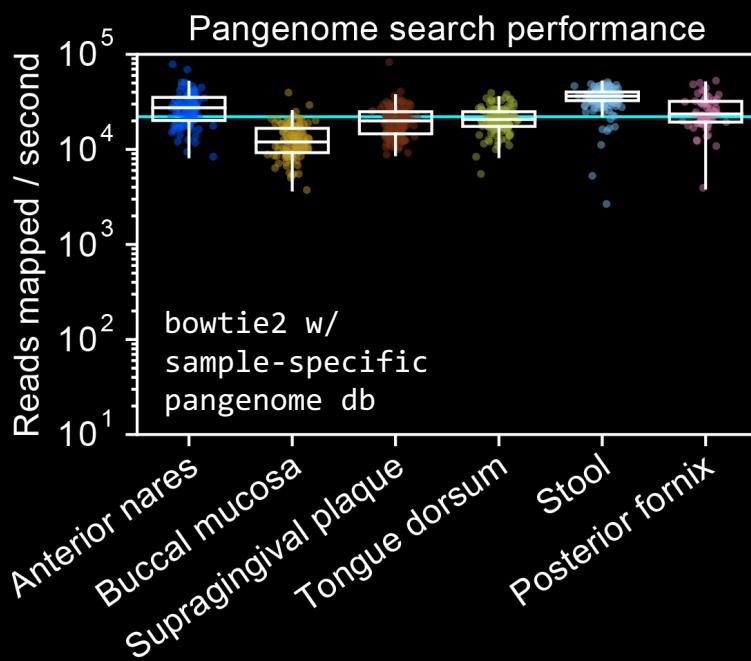
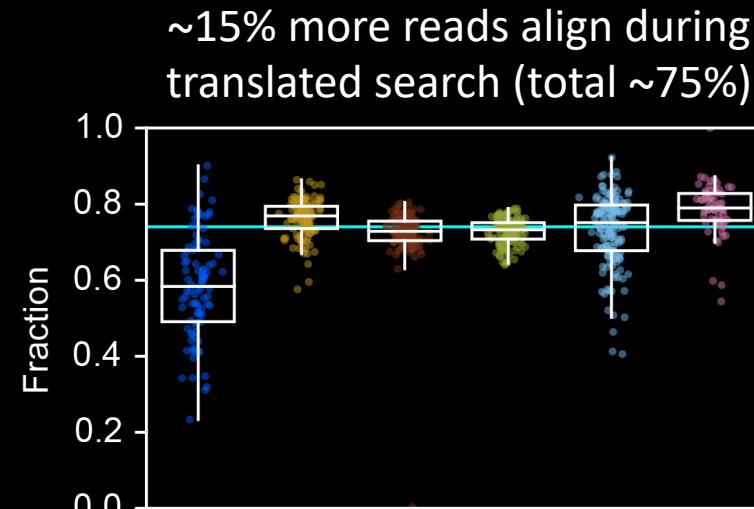
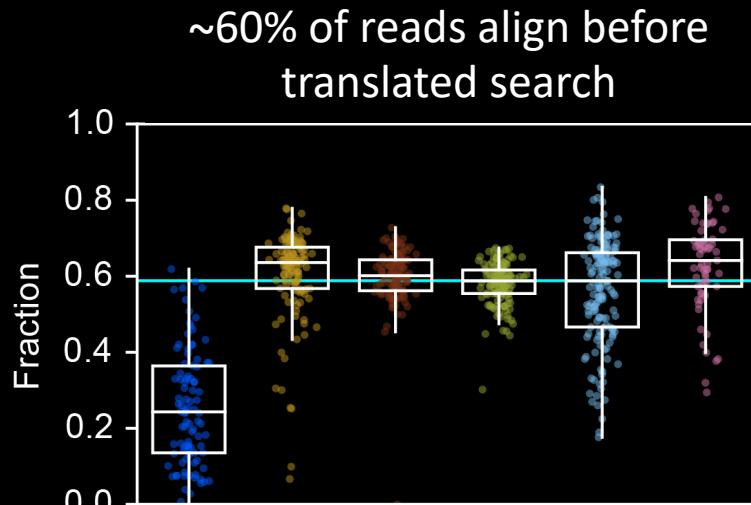
Applied HUMAnN2's tiered search to profile >2K human metagenomes (HMP1-II, six major body sites)

Pangenome search tier
1-2 orders of magnitude faster than comprehensive translated search





HUMAnN2 real-world performance

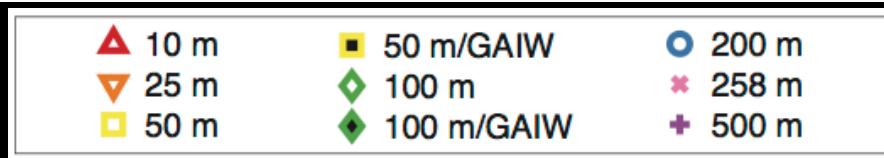
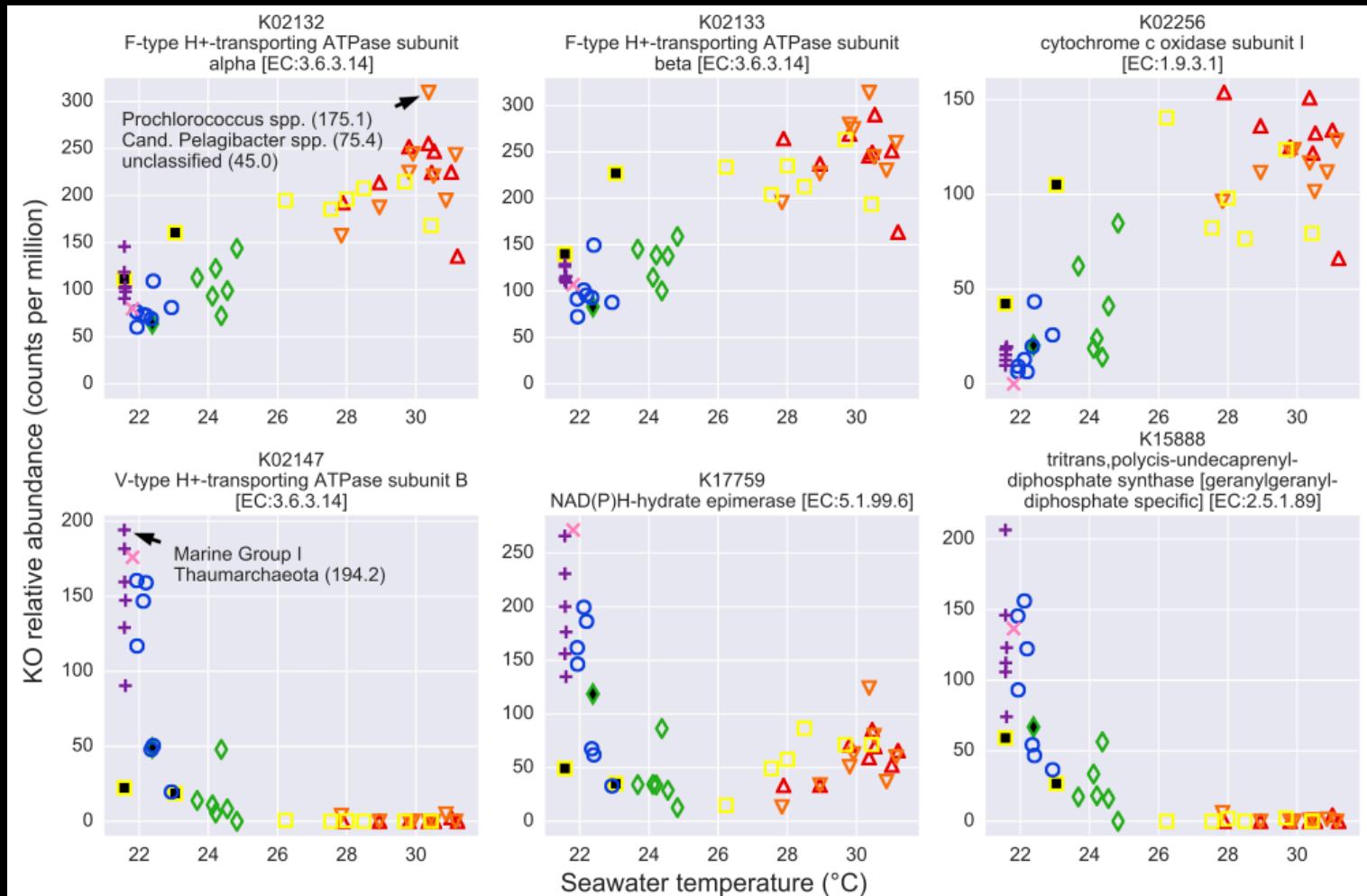




And it works on non-human meta'omes, too

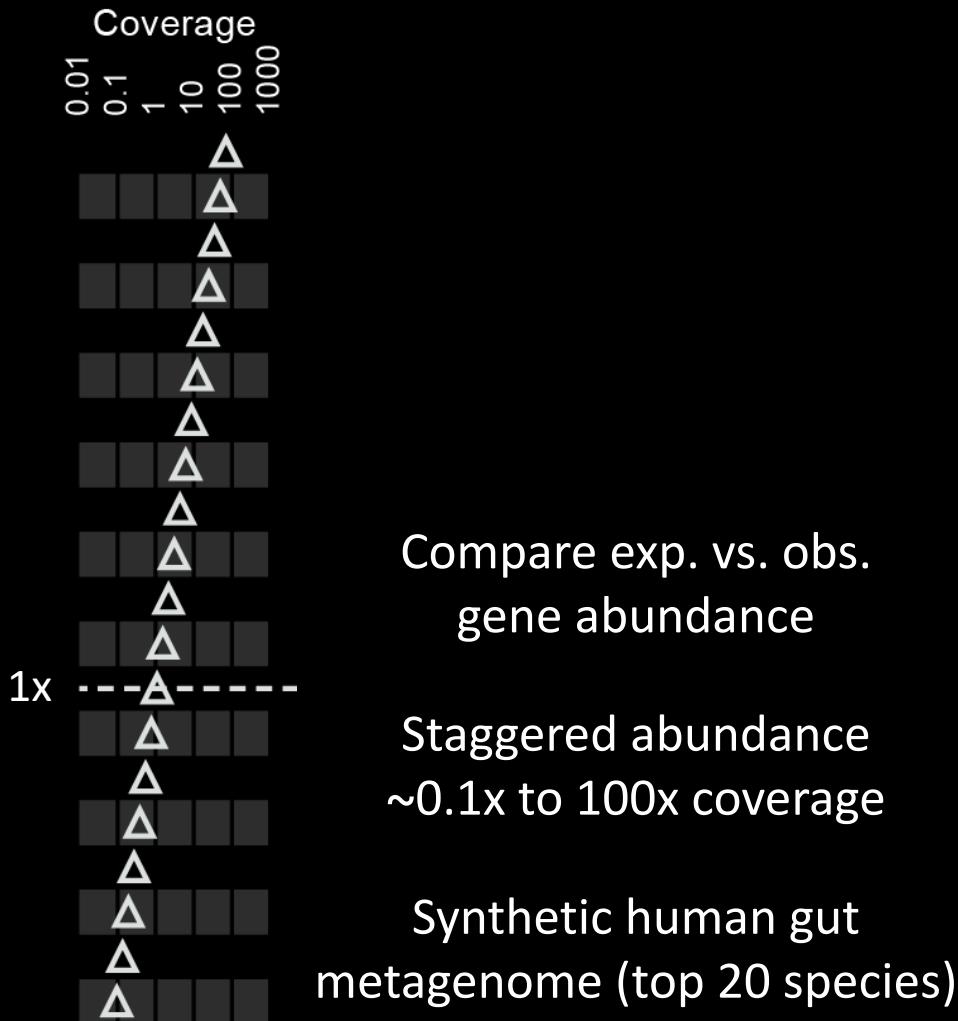


Luke
Thompson



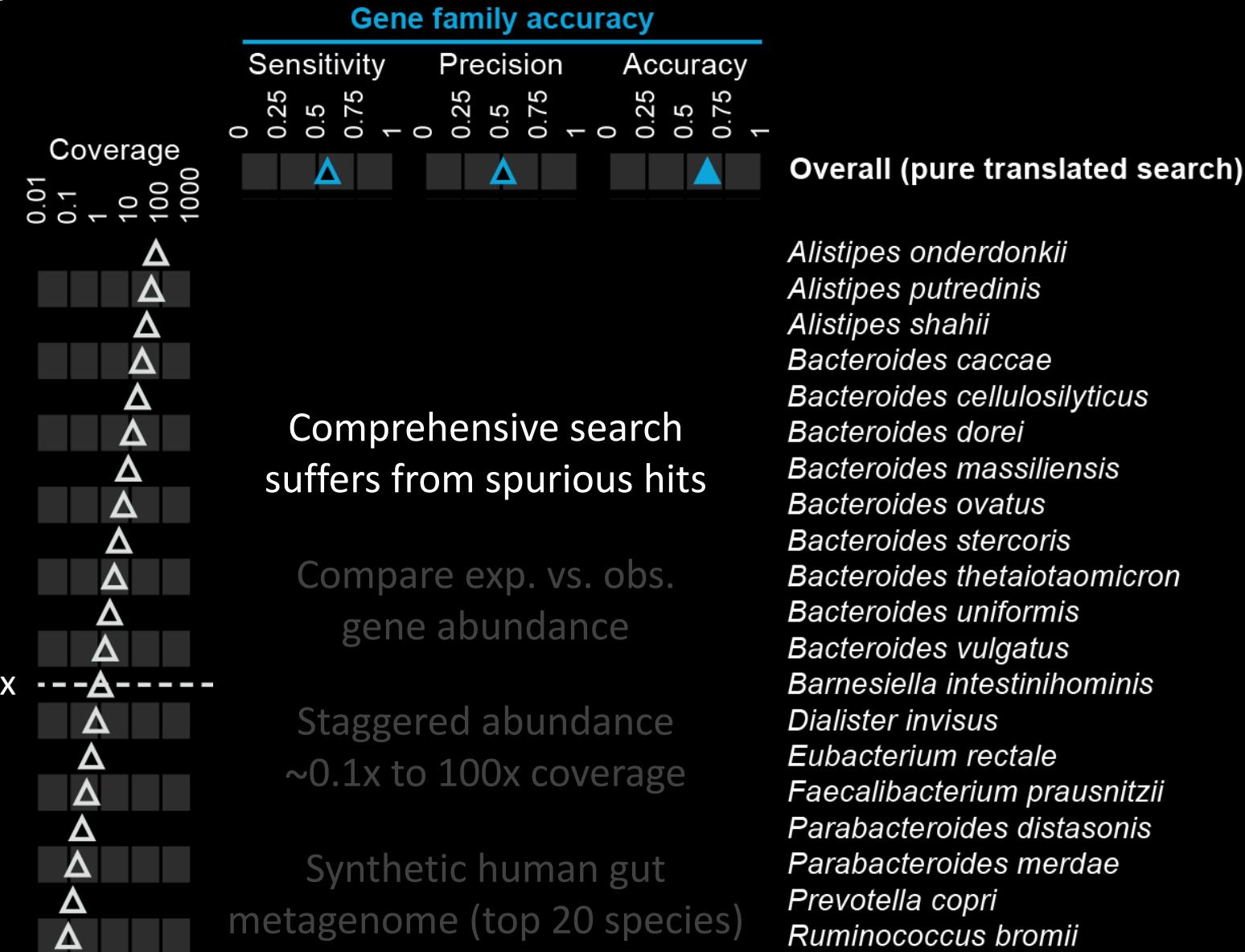


HUMAnN2 synthetic evaluation (genes)



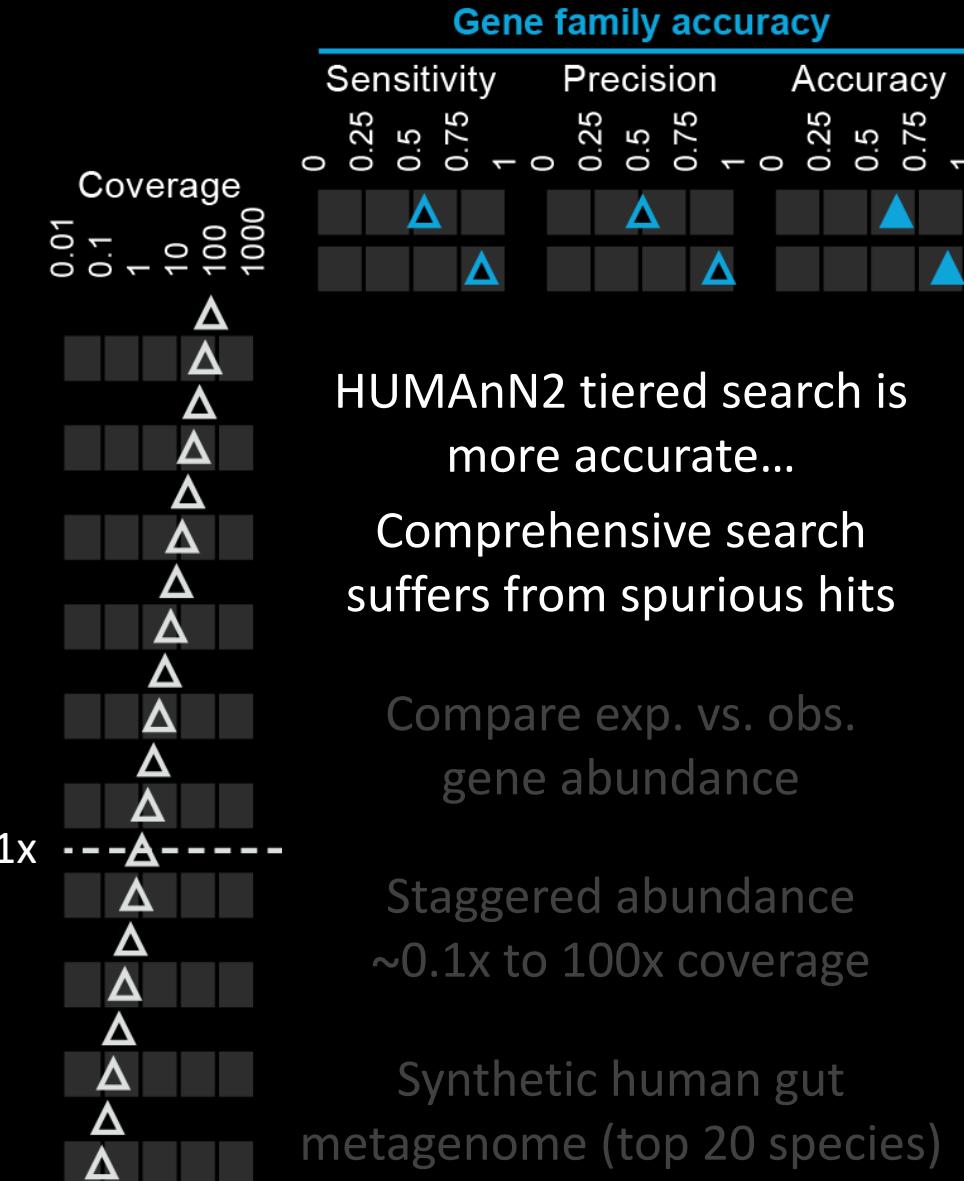


HUMAnN2 synthetic evaluation (genes)





HUMAnN2 synthetic evaluation (genes)



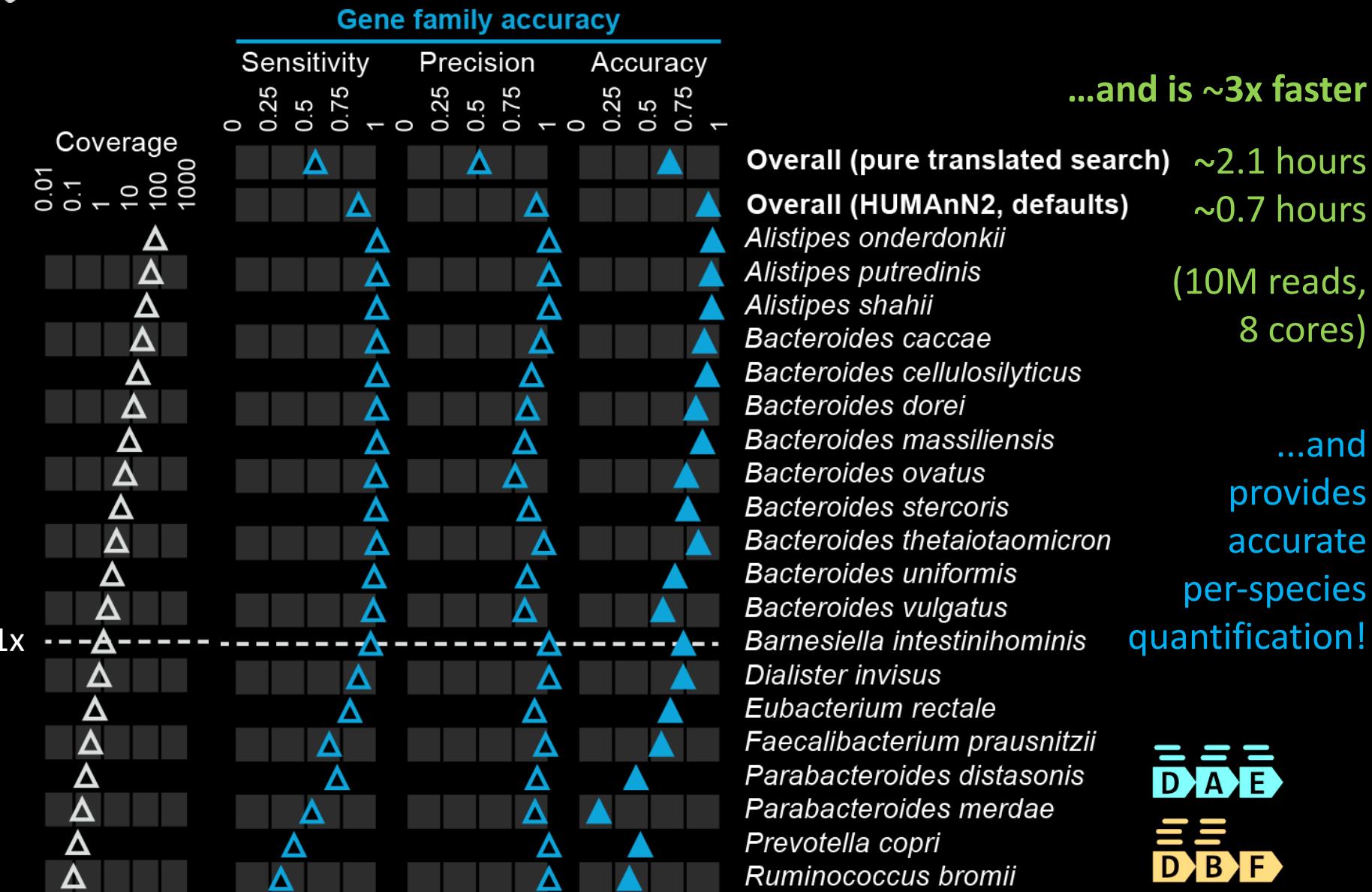
...and is ~3x faster

Overall (pure translated search)	~2.1 hours
Overall (HUMAnN2, defaults)	~0.7 hours
<i>Alistipes onderdonkii</i>	
<i>Alistipes putredinis</i>	
<i>Alistipes shahii</i>	
<i>Bacteroides caccae</i>	
<i>Bacteroides cellulosilyticus</i>	
<i>Bacteroides dorei</i>	
<i>Bacteroides massiliensis</i>	
<i>Bacteroides ovatus</i>	
<i>Bacteroides stercoris</i>	
<i>Bacteroides thetaiotaomicron</i>	
<i>Bacteroides uniformis</i>	
<i>Bacteroides vulgatus</i>	
<i>Barnesiella intestinihominis</i>	
<i>Dialister invisus</i>	
<i>Eubacterium rectale</i>	
<i>Faecalibacterium prausnitzii</i>	
<i>Parabacteroides distasonis</i>	
<i>Parabacteroides merdae</i>	
<i>Prevotella copri</i>	
<i>Ruminococcus bromii</i>	





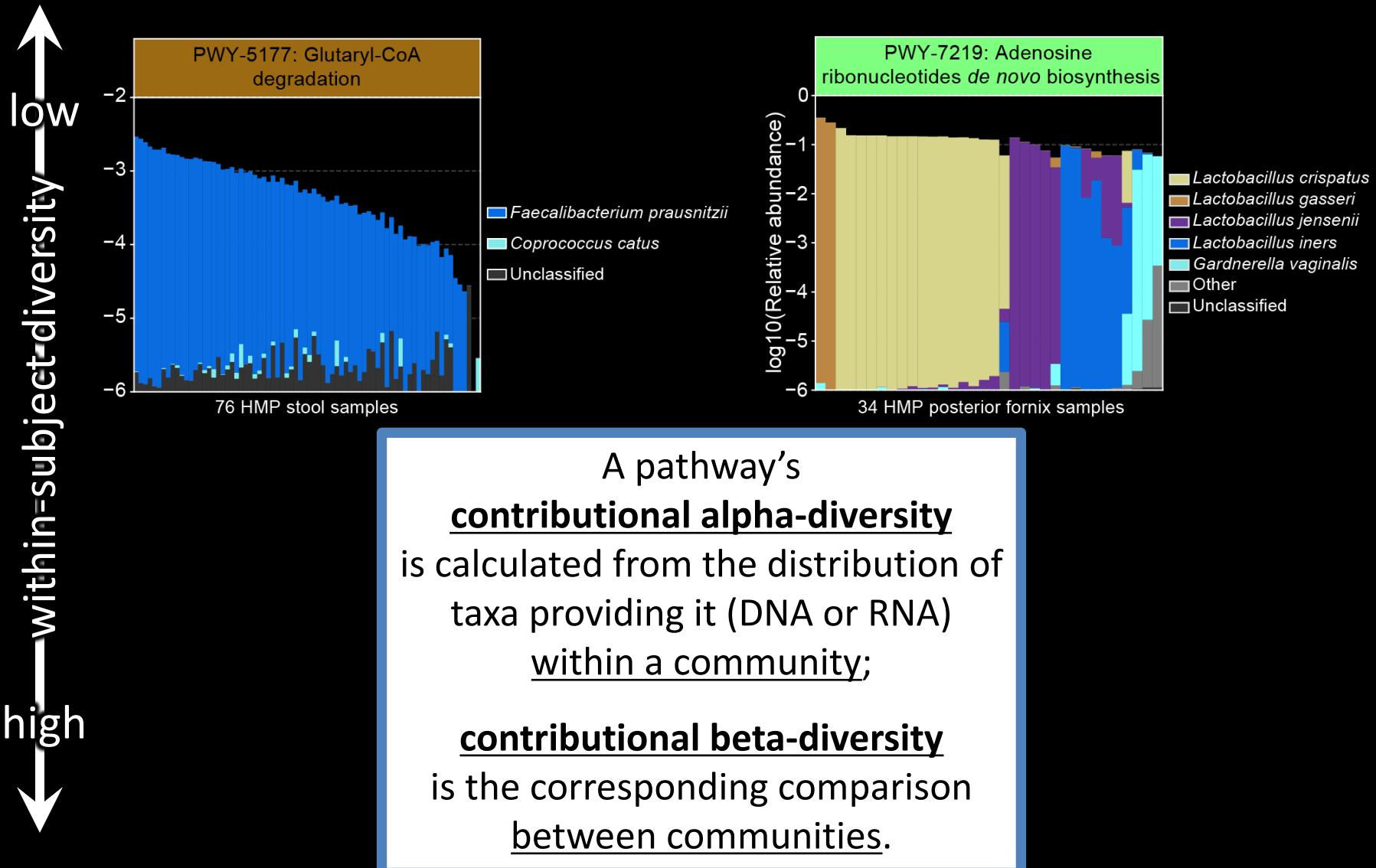
HUMAnN2 synthetic evaluation (genes)





Quantifying the diversity of species contributing a function within and across subjects

← low ————— between-subject-diversity ————— high →



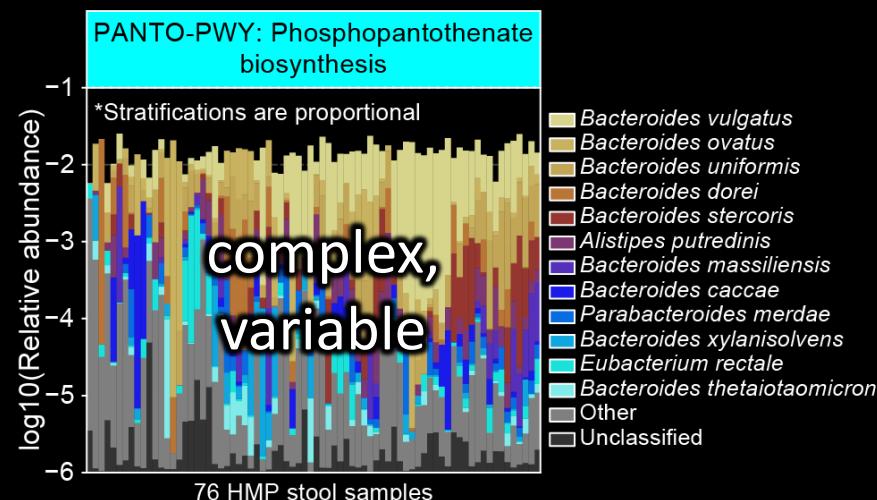
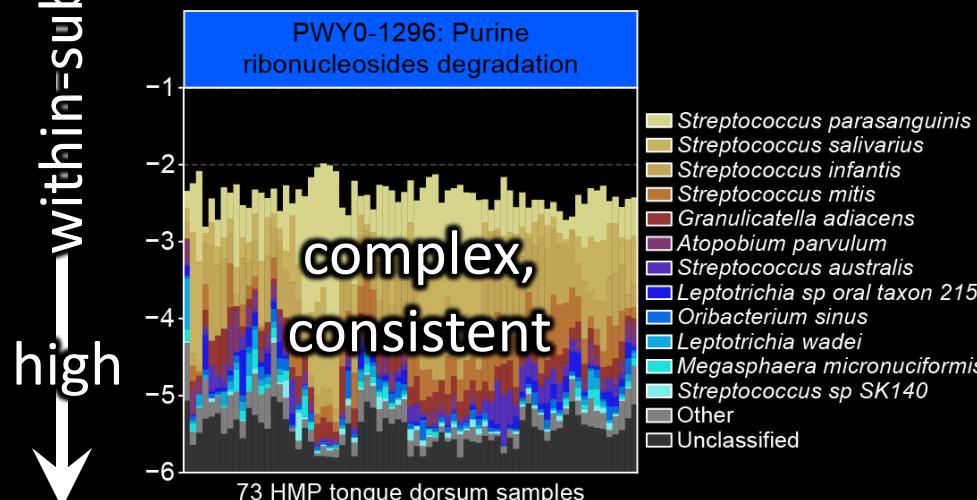
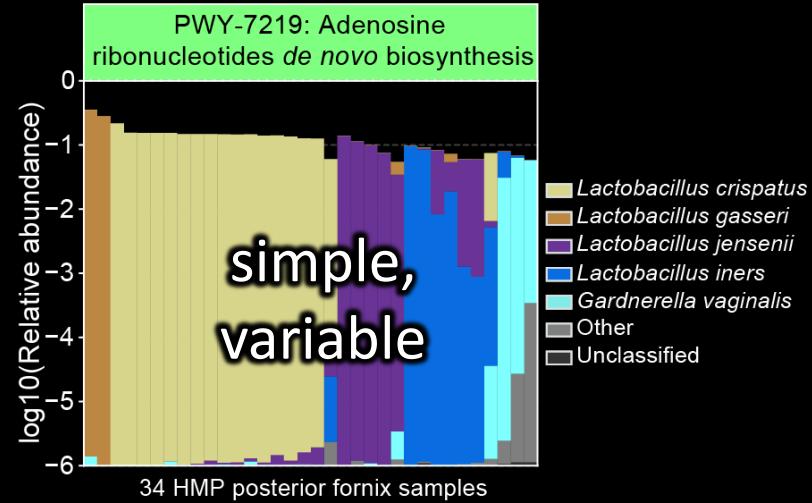
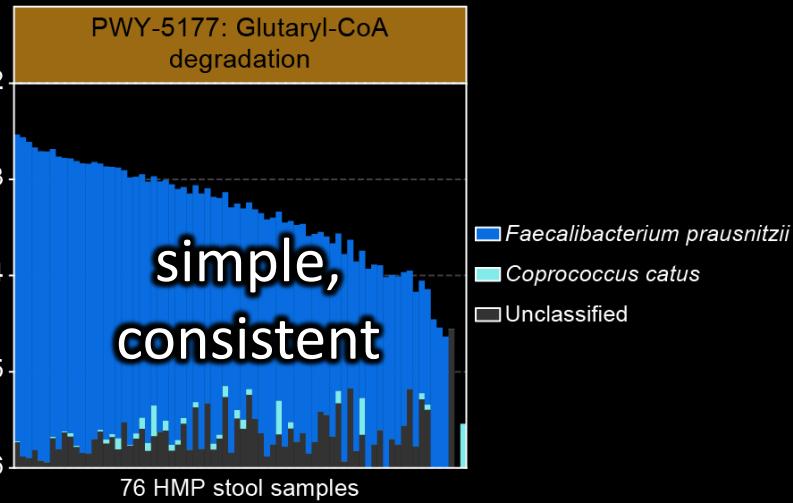


Quantifying the diversity of species contributing a function within and across subjects

12

← low → between-subject-diversity

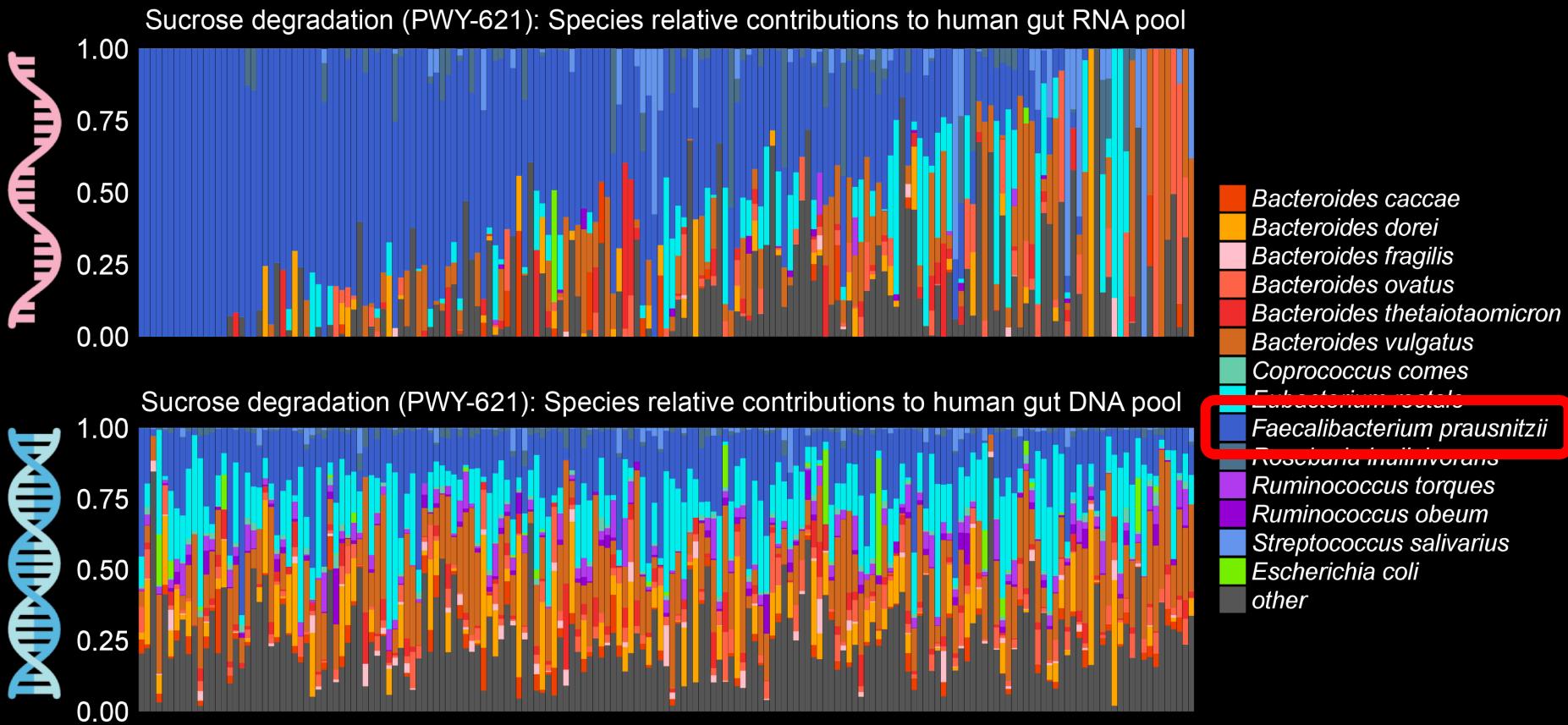
↑
low
within-subject-diversity
high
↓





HUMAnN2 reveals unusual “relative expression” in paired metatranscriptomes & metagenomes

13



Sucrose degradation follows a complex attribution pattern across ~200 human gut metagenomes...
...but its expression can be dominated by a single species in paired gut metatranscriptomes!



Galeb
Abu-Ali

In collaboration with
the STARR Consortium
& HPFS cohort



Considerations for paired metatranscriptomes & metagenomes

- \$ humann2_rna_dna_norm --input_dna <DNA genefamilies file> --input_rna <RNA genefamilies file> --output_basename <basename of the 3 output files>
 - Calculates RNA/DNA abundance ratios
 - Smooths the RNA and DNA abundances prior to taking the ratio
 - Also outputs smoothed RNA and DNA files

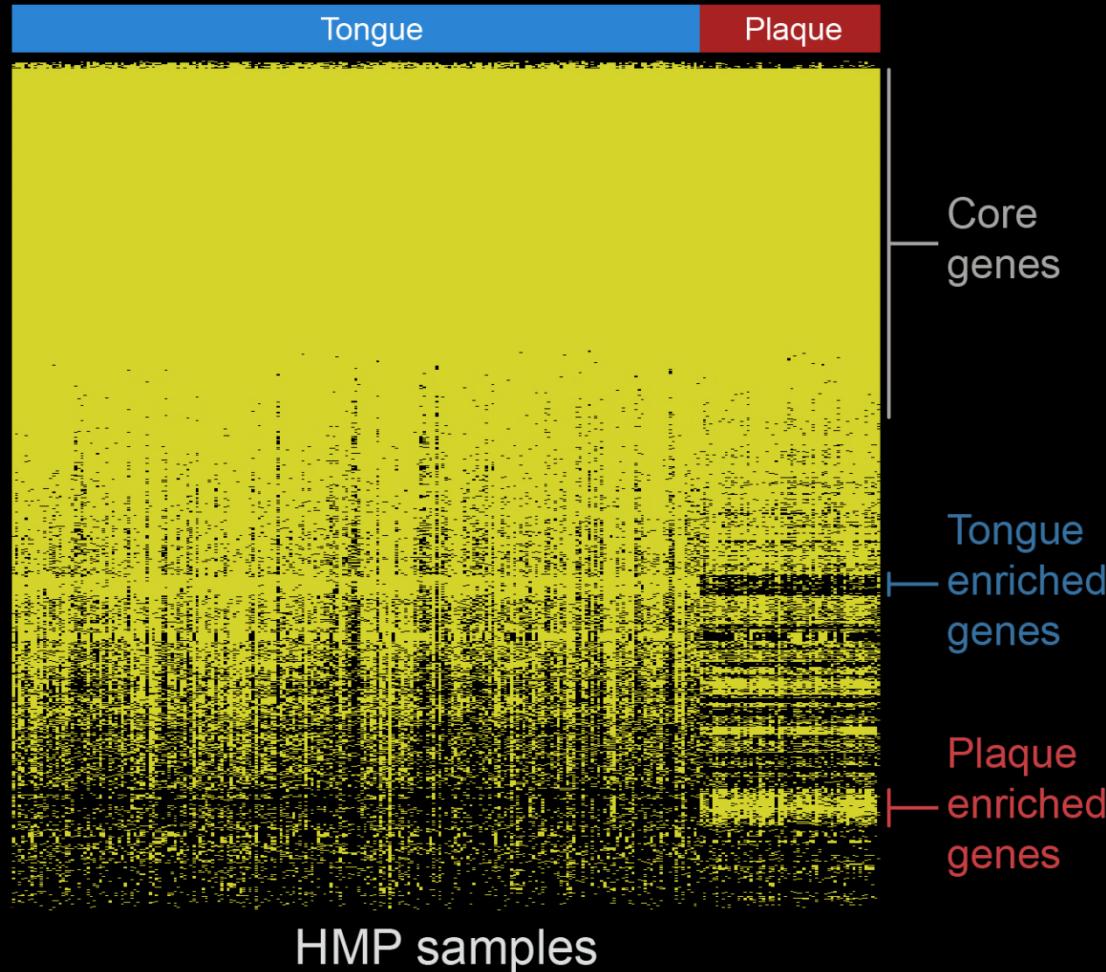
UniRef90_R6K3Z5: IMP dehydrogenase	2.02
UniRef90_R6K3Z5: IMP dehydrogenase Bacteroides_caccae	5.96
UniRef90_R6K3Z5: IMP dehydrogenase Bacteroides_dorei	3.82
UniRef90_R6K3Z5: IMP dehydrogenase Bacteroides_ovatus	1.80
UniRef90_R6K3Z5: IMP dehydrogenase Bacteroides_stercoris	0.87
UniRef90_R6K3Z5: IMP dehydrogenase Bacteroides_vulgatus	0.34
UniRef90_R6K3Z5: IMP dehydrogenase unclassified	1.96



HUMAnN2 gene family profiles identify niche-adapted subspecies

15

Pangenome profiles of *Campylobacter showae*

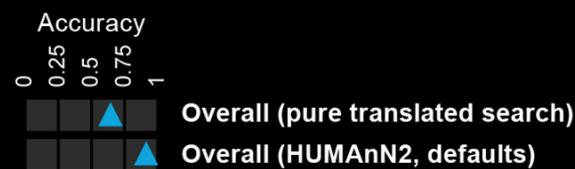
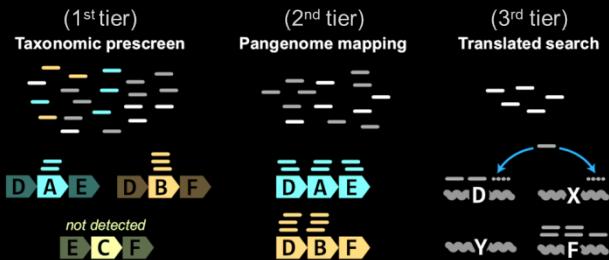


This example confirmed by SNP-level strain analysis

HUMAnN2 gene profiles can act as input for PanPhlAn strain tracking:
<http://segatalab.cibio.unitn.it/tools/panphlan>

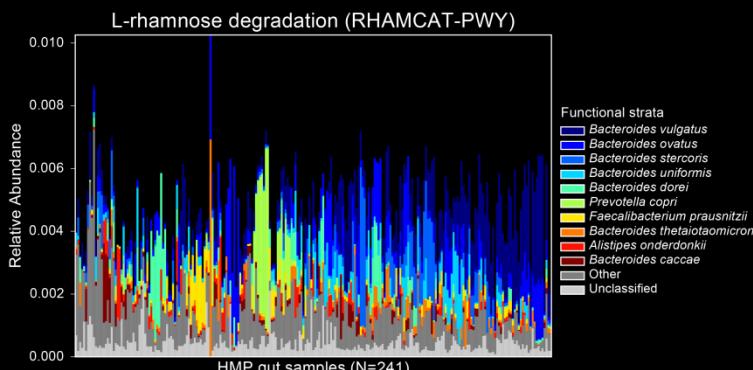


Take-home messages



HUMAnN2 implements a tiered approach to facilitate meta'omic functional profiling

This approach is more accurate & much faster than traditional comprehensive meta'omic search



Results stratify by species for free:
Answering “*who’s there?*” and
“*What are they doing?*” in tandem



HUMAnN2 tutorial

<https://bitbucket.org/biobakery/biobakery/wiki/humann2>

HUMAnN

Profile microbial genes
and pathways; stratify by
contributing species

Input

- Short reads

Output

- Gene abundance table
- Functional profiling
- Pathway reconstructions