HUMAnN & bioBakery

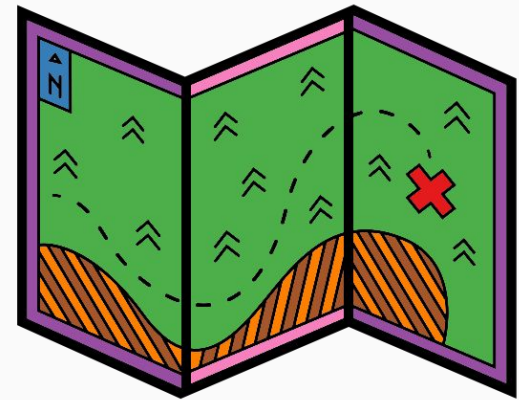# Plan

bioBakery

HUMAnN

Functional annotation

bioBakery

A meta'omic analysis environment

Microbial community profiling

Huttenhower lab

https://github.com/biobakery/biobakery

https://elifesciences.org/articles/65088

Microbial community profiling

HUMAnN
Profile microbial genes and pathways; stratify by contributing species

MetaPhlAn
Profile microbial clades and their abundances

PhyloPhlAn
Reconstruct microbial phylogenetic trees

PICRUSt
Predict metagenome functional content from marker genes

ShortBRED
Profile proteins of interest with very high specificity

PPANINI
Prioritize microbial genes based on their meta'omic properties

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

MelonnPan
Predict metabolome content from microbial enzyme abundances

WAAFLE
Identify novel LGT events in assembled metagenomes

MetaWIBELE
Identify potentially bioactive novel gene products in microbial communities

MACARRoN
Identify potentially bioactive small molecules in microbial communities

Downstream analysis and stats

HAllA
Perform well-powered comparisons of paired, high-dimensional datasets

ARepA
Extract and normalize 'omics data from online repositories

CCREPE
Assess the significance of similarity measures in compositional data

LEfSe
Associate up to two metadata with microbiome features

MaAsLin
Associate arbitrarily complex metadata with microbiome features

MMUPHin
Correct batch effects, meta-analyze microbes, genes, and pathways across multiple studies

MicroPITA
Select samples for follow-up analysis in two-stage tiered studies

SparseDOSSA
A hierarchical model of microbial ecological population structure

BAnOCC
Bayesian assessment of association in compositional data

Infrastructure and utilities

GraPhlAn
Generate cladograms and decorate with metadata

KneadData
Perform quality control of meta'omic reads, including host read removal

AnADAMA
Perform and document automated scientific workflows

Workflows
A collection of meta'omic data processing and visualization workflows

HUMAnN

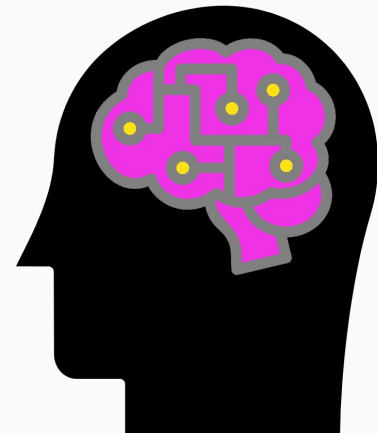HMP Unified Metabolic Analysis Network

Profile the abundance of microbial pathways & other microbial functions
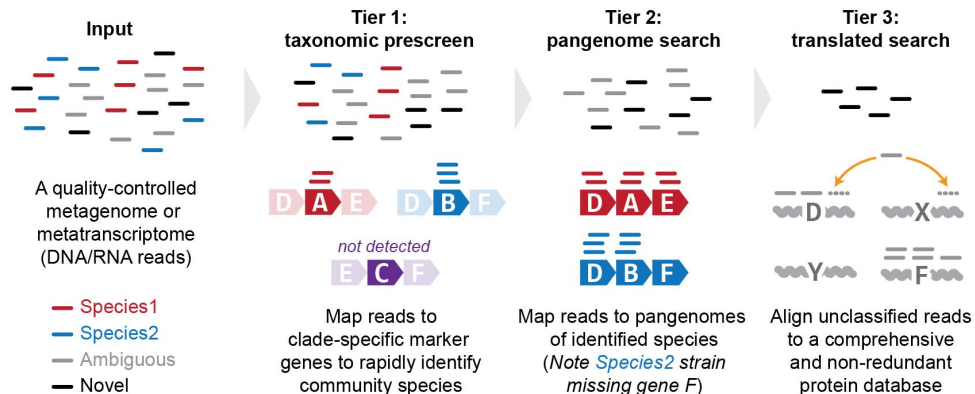
Metagenomic or metatranscriptomic data

Read data

On 3rd version

Co-created with MetaPhlAn 3.0

HUMAnN pipeline

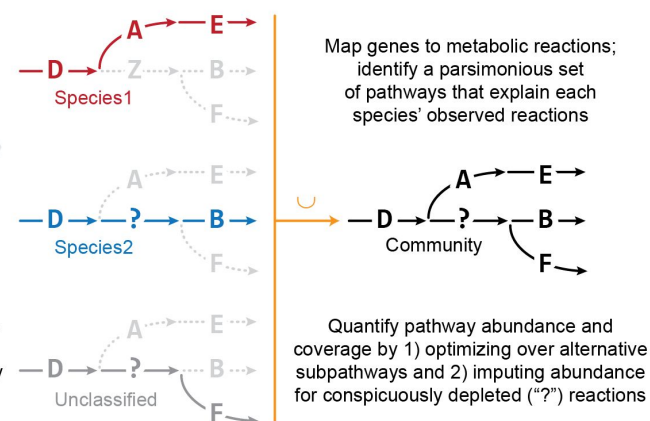**a** HUMAnN's tiered meta'omic search

Input | Tier 1: taxonomic prescreen | Tier 2: pangenome search | Tier 3: translated search

A quality-controlled metagenome or metatranscriptome (DNA/RNA reads)

— Species1
— Species2
— Ambiguous
— Novel

Map reads to clade-specific marker genes to rapidly identify community species

not detected

Map reads to pangenomes of identified species (*Note Species2 strain missing gene F*)

Align unclassified reads to a comprehensive and non-redundant protein database

b HUMAnN's gene family & pathway quantification

Gene abundance estimation

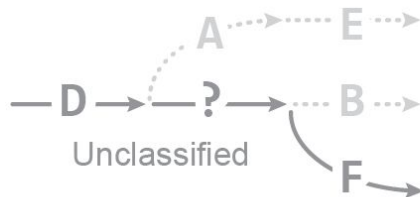| Feature	RPK
GeneA	2
GeneA|Species1	2
GeneB	3
GeneB|Species2	3
∑ GeneD	8
GeneD|Species1	2
GeneD|Species2	3
GeneD|unclassified	3
GeneE	2
GeneE|Species1	2
GeneF	5
GeneF|unclassified	5

Process mapping results to estimate per-species and community total gene family abundance, weighting by 1) alignment quality, 2) gene length, and 3) gene coverage

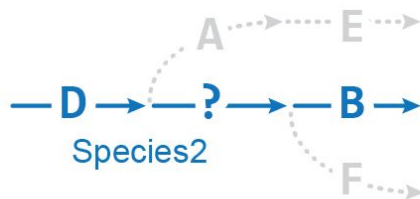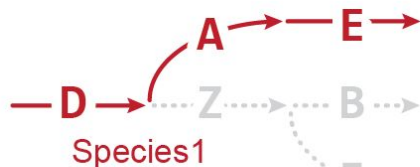**Per-species and community-level metabolic network reconstruction**

Species1

Species2

Unclassified

Community

Map genes to metabolic reactions; identify a parsimonious set of pathways that explain each species' observed reactions

Quantify pathway abundance and coverage by 1) optimizing over alternative subpathways and 2) imputing abundance for conspicuously depleted ("?") reactions

a HUMAnN's tiered meta'omic search

Input

A quality-controlled metagenome or metatranscriptome (DNA/RNA reads)

— Species1
— Species2
— Ambiguous
— Novel

Tier 1: taxonomic prescreen

not detected

Map reads to clade-specific marker genes to rapidly identify community species

Tier 2: pangenome search

Map reads to pangenomes of identified species (*Note Species2 strain missing gene F*)

Tier 3: translated search

Align unclassified reads to a comprehensive and non-redundant protein database

# b HUMAnN's gene family & pathway quantification

Gene abundance estimation

Feature	RPK
GeneA	2
GeneA|Species1	2
GeneB	3
GeneB|Species2	3
GeneD	8
GeneD|Species1	2
GeneD|Species2	3
GeneD|unclassified	3
GeneE	2
GeneE|Species1	2
GeneF	5
GeneF|unclassified	5

Process mapping results to estimate per-species and community total gene family abundance, weighting by 1) alignment quality, 2) gene length, and 3) gene coverage

Per-species and community-level metabolic network reconstruction

Map genes to metabolic reactions; identify a parsimonious set of pathways that explain each species' observed reactions
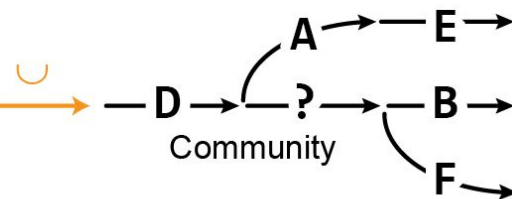
Quantify pathway abundance and coverage by 1) optimizing over alternative subpathways and 2) imputing abundance for conspicuously depleted ("?") reactions
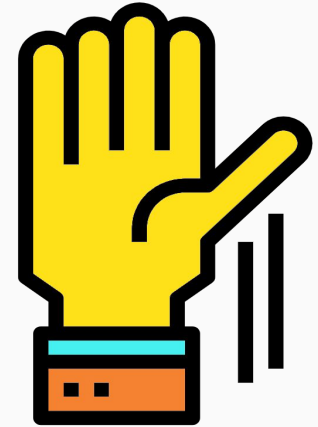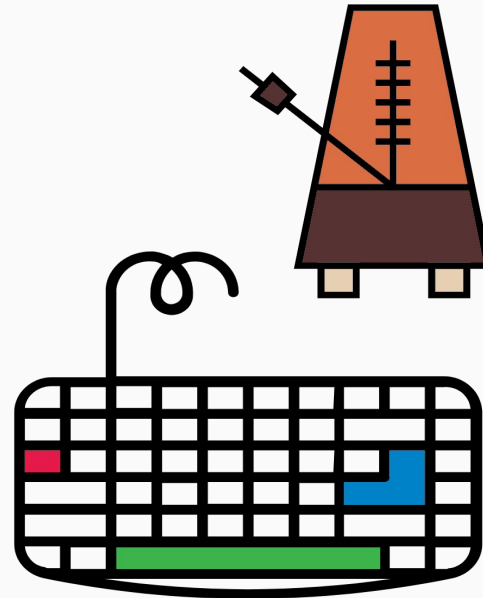
Reminders and Tips

Work at your own pace

Typos

Ask questions

Breaks are important

Tab, space, and enter

During sessions

Zoom - Ask via microphone if no question currently being asked/answered

Slack - Ask questions via the channel or ask to go into a zoom breakout room with one of us

WebVNC - We can connect to your webVNC to see and help with issues.
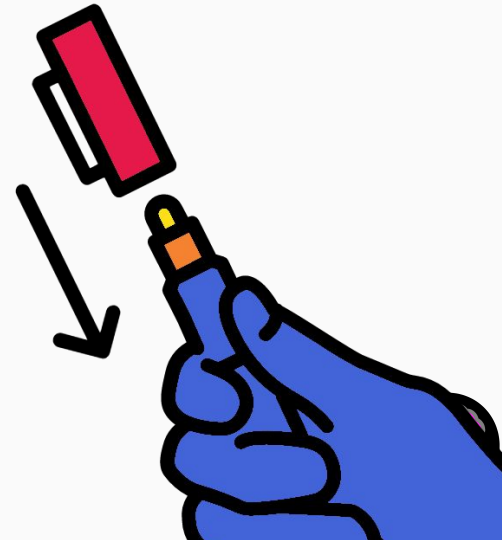
Breakout rooms upon request

BioBakery

HUMAnN

Profile the abundance of microbial pathways & other microbial functions

Thank you!

Questions?