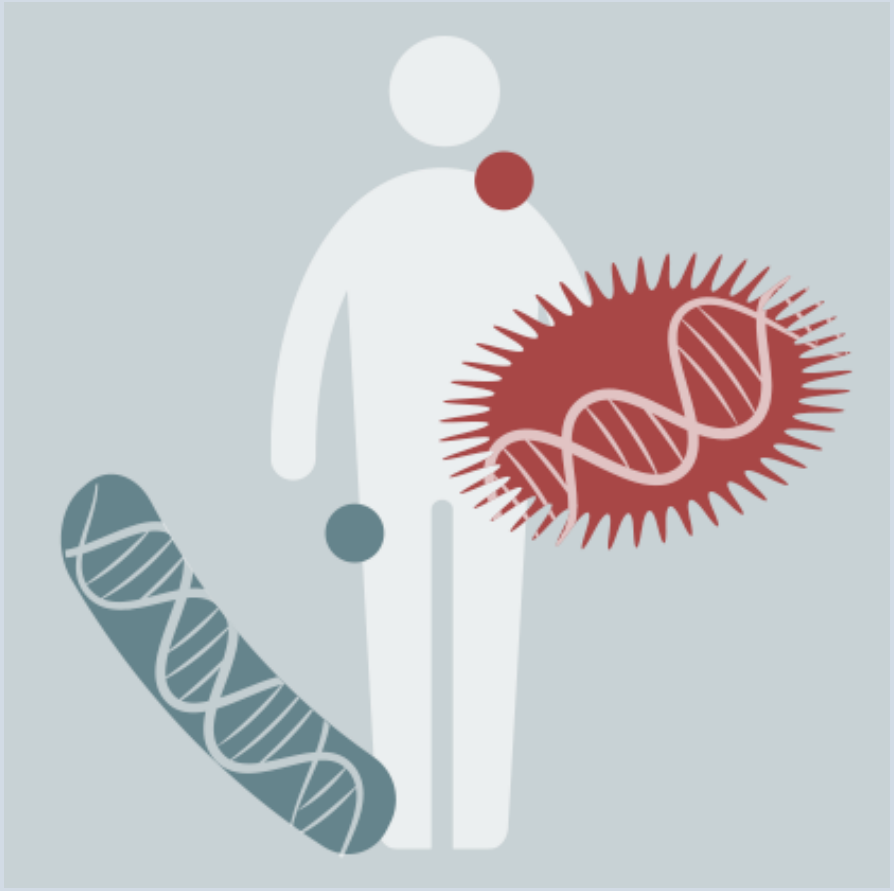
Module 6
Metatranscriptomics

John Parkinson

Microbiome Analysis

December 6th-7th, 2022



Parkinson Lab

SickKids
THE HOSPITAL FOR SICK CHILDREN

UNIVERSITY of TORONTO

Research in the Parkinson-Lab

Pathogen:host:microbiome interactions

Microbiome and malnutrition in pregnancy

Pediatric IBD – eukaryotic microbiome

Chickens!

Learning Objectives of Module

At the end of this module the student will have an appreciation of the opportunities and challenges of metatranscriptomics by:

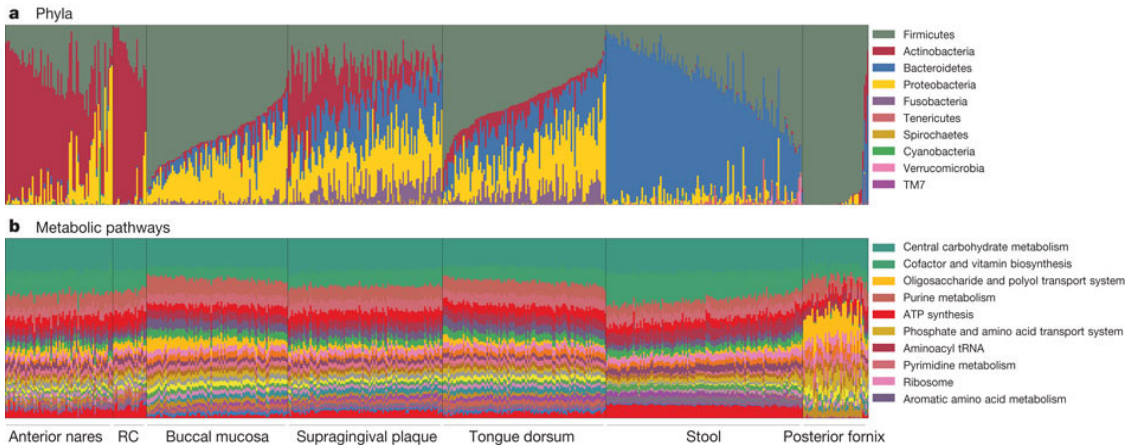
- Understanding the capabilities of metatranscriptomics
- Gaining an appreciation of sample collection and experimental design
- Learning important steps in data processing
- Processing a simple metatranscriptomic dataset

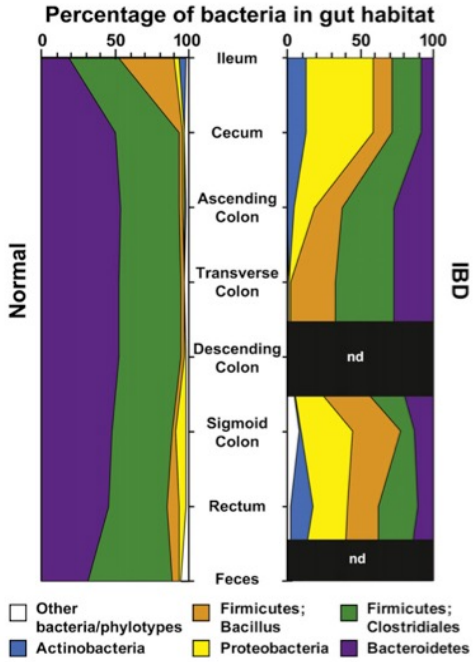
Overview

- Metatranscriptomics – what is it and why use it?
- Experimental design, sample collection and preparation
- Processing of reads
  - Filtering
  - Assembly
  - Functional annotation
  - Taxonomic annotation
- Statistical analysis
- Visualization and Interpretation

Metagenomics and metatranscriptomics reveal function

16S rRNA surveys ("Who is there?") have been widely applied but yield only limited mechanistic insights – cause or consequence?





Metagenomics ("What can they do?") have revealed dramatic differences in community **composition** but with conserved microbiome **functions**

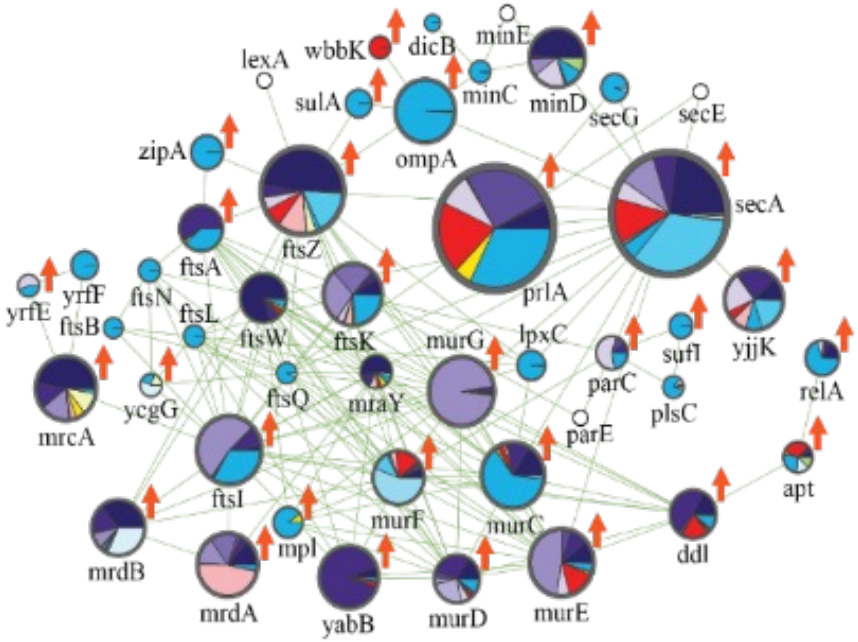
Metatranscriptomics ("Who is doing what?") examines microbiome **activity**

Frank et al Cell Host Microbe 2008 / The human microbiome consortium Nature 2012/ Xiong et al PLoS ONE 2012

Metatranscriptomics focuses on community activity

Metatranscriptomics exploits RNA-Seq to determine which genes and pathways are being actively expressed within a community

Genes involved in pathways associated with cell wall biogenesis from microbes in the chicken ceca



Genes upregulated in presence of antibiotics shown as red arrows

Metatranscriptomics can reveal active *functions*

It can also reveal which taxa are responsible for the active functions

Metatranscriptomics applied to a model of obesity

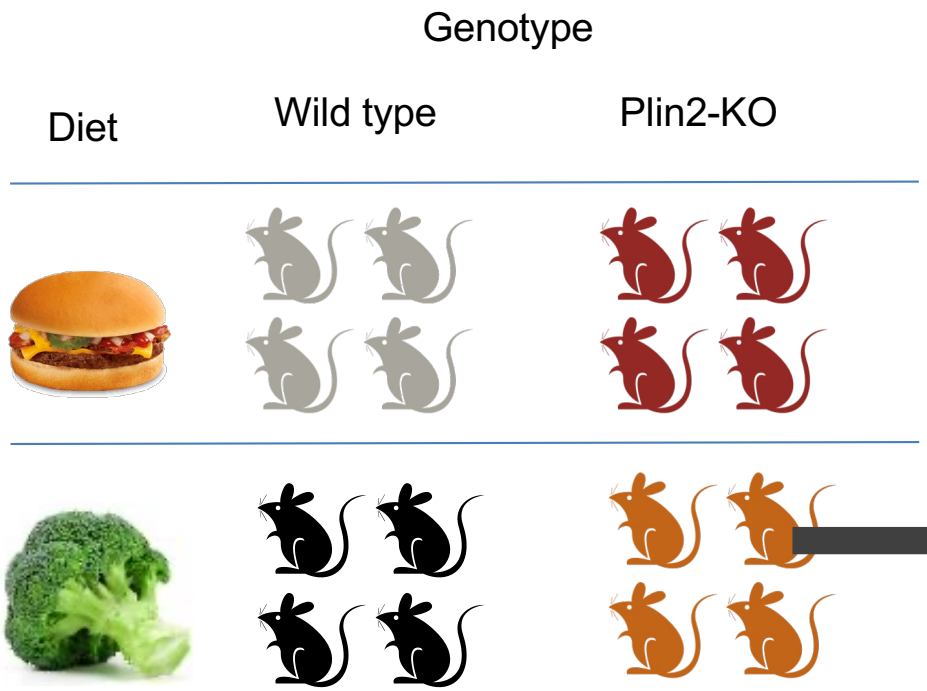
Xiong et al. Microbiome (2017) 5:117
DOI 10.1186/s40168-017-0327-x

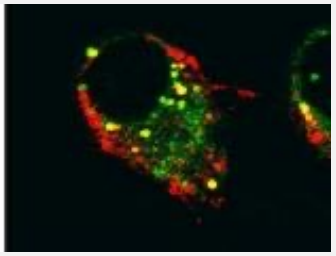
Microbiome

RESEARCH Open Access

Perilipin-2 modulates dietary fat-induced microbial global gene expression profiles in the mouse intestine

Xuejian Xiong[1], Elise S. Bales[2], Diana Ir[3], Charles E. Robertson[3,4], James L. McManaman[2,5], Daniel N. Frank[3,4*] and John Parkinson[1,6,7*]





Najt et al
Biochemistry
2014

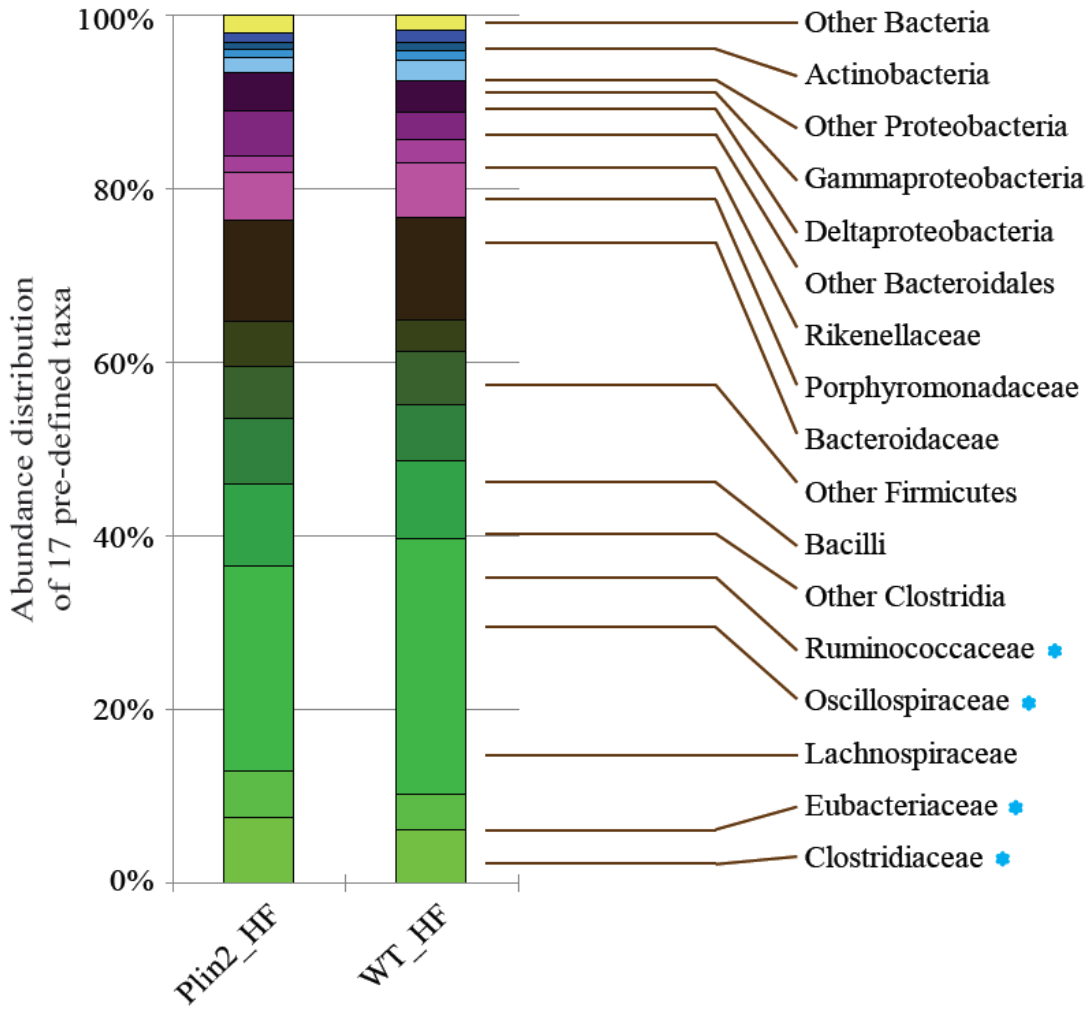
Perilipin2 (Plin2) interacts with lipid droplets and helps regulate lipid uptake

Deletion of Plin2 in mice largely abrogates deleterious effects of a high fat (HF) diet

What impact does Plin2-KO have on microbiome function?

Under a high fat diet metatranscriptomics reveals no differences in microbiome composition

Distribution of Taxa
(n=4; 20-30 million reads/sample)

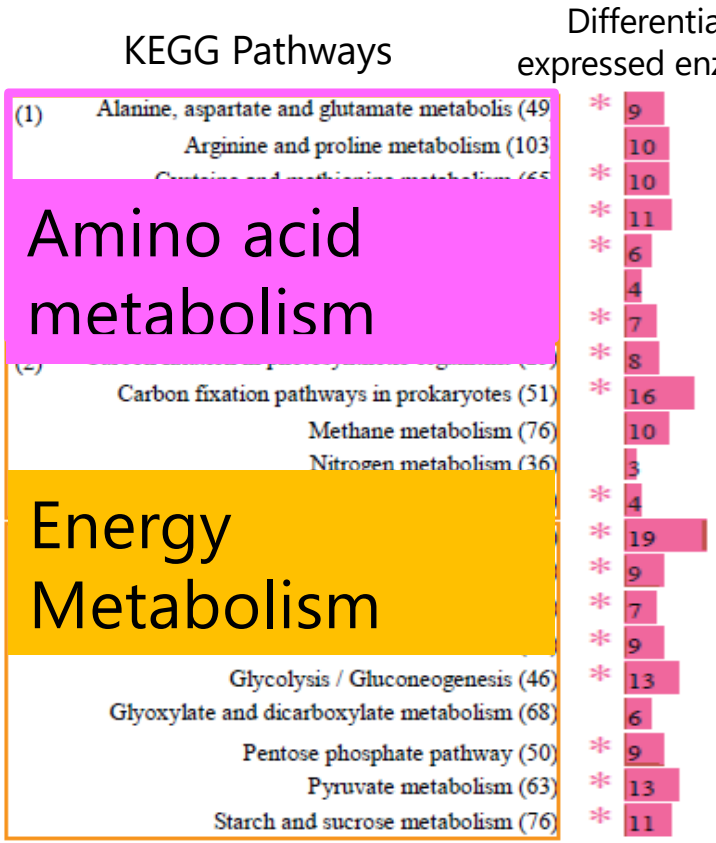


While changes in diet (low fat v high fat) result in significant shifts in taxa abundance, under a high fat diet, there were **no significant differences in composition** between the Plin2 Knockout mice and WT

Xiong et al Microbiome 2017

Similar microbiomes can express different functions

Under a high fat diet, Plin2 and WT mice exhibit genotype-specific differential expression of **over 1000** highly expressed microbial genes despite similar **taxonomic** composition!

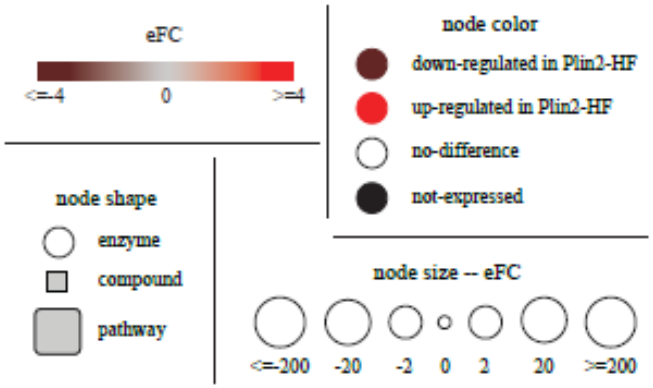


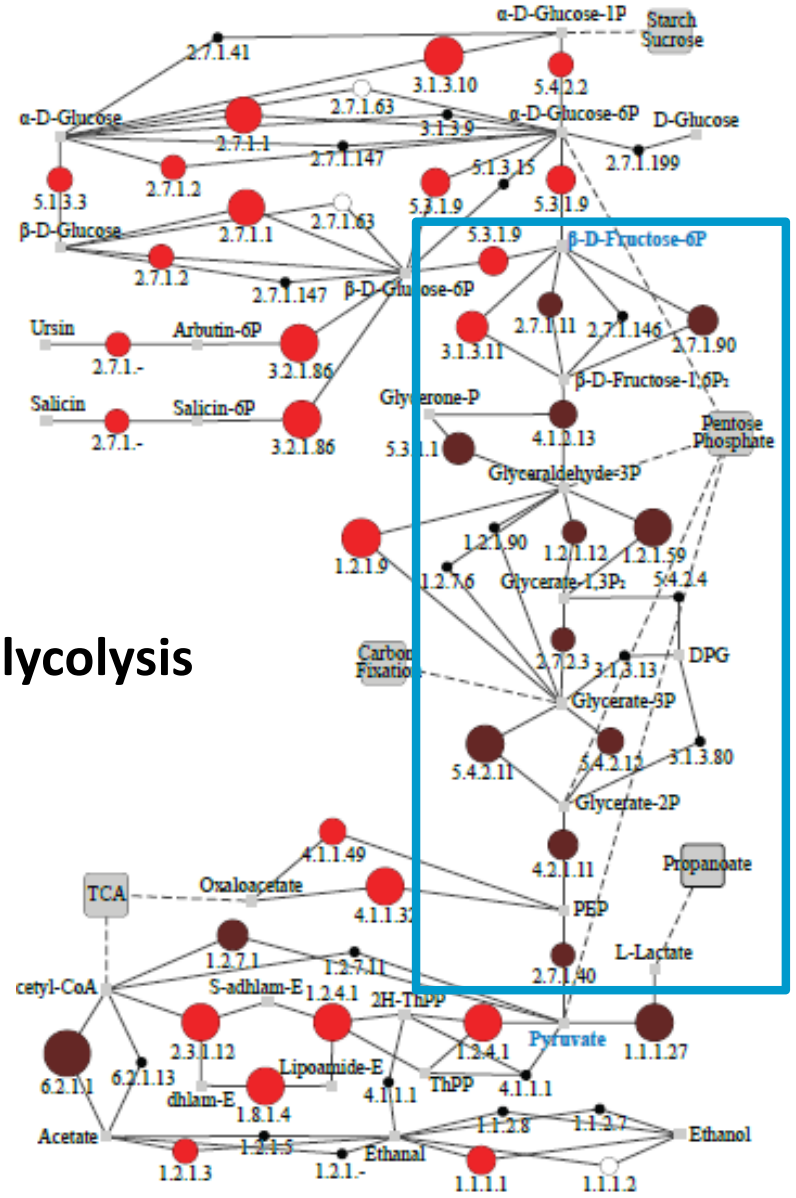
Many of these differentially expressed genes are associated with amino acid metabolism and energy metabolism

Xiong et al Microbiome 2017

Pathway analyses reveal potential impact of host genotype on microbial gene expression

Mapping of expression differences in a pathway context, reveal enzymes performing consecutive reactions in the production of pyruvate, exhibit **consistent down regulation** in Plin2 mice

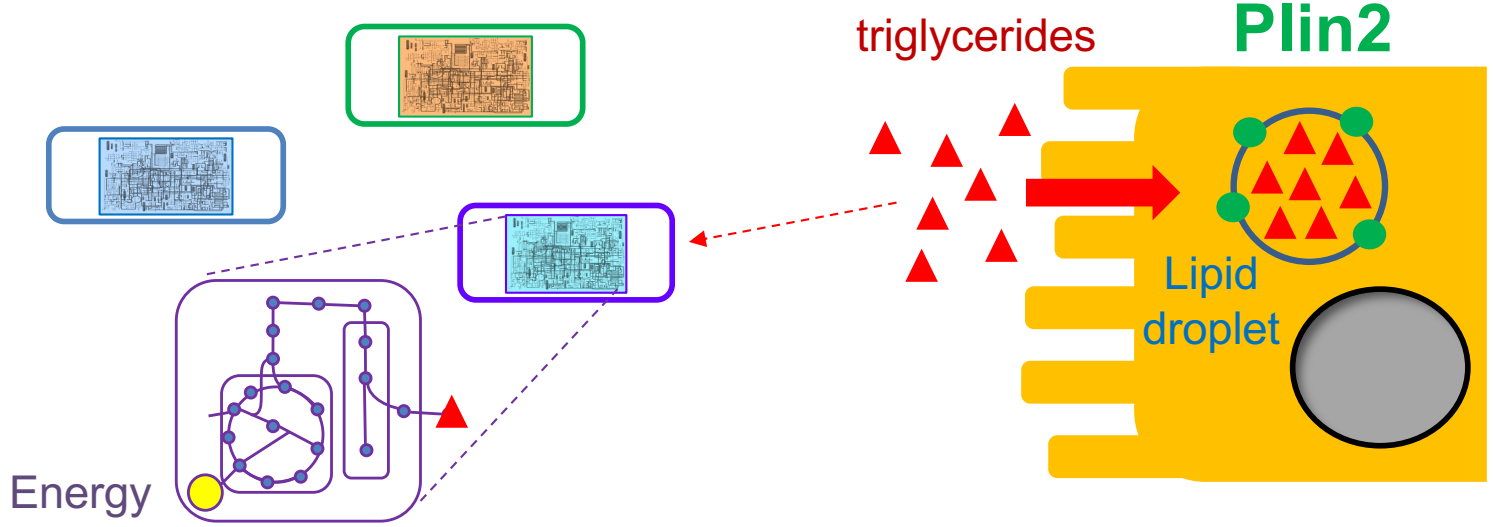




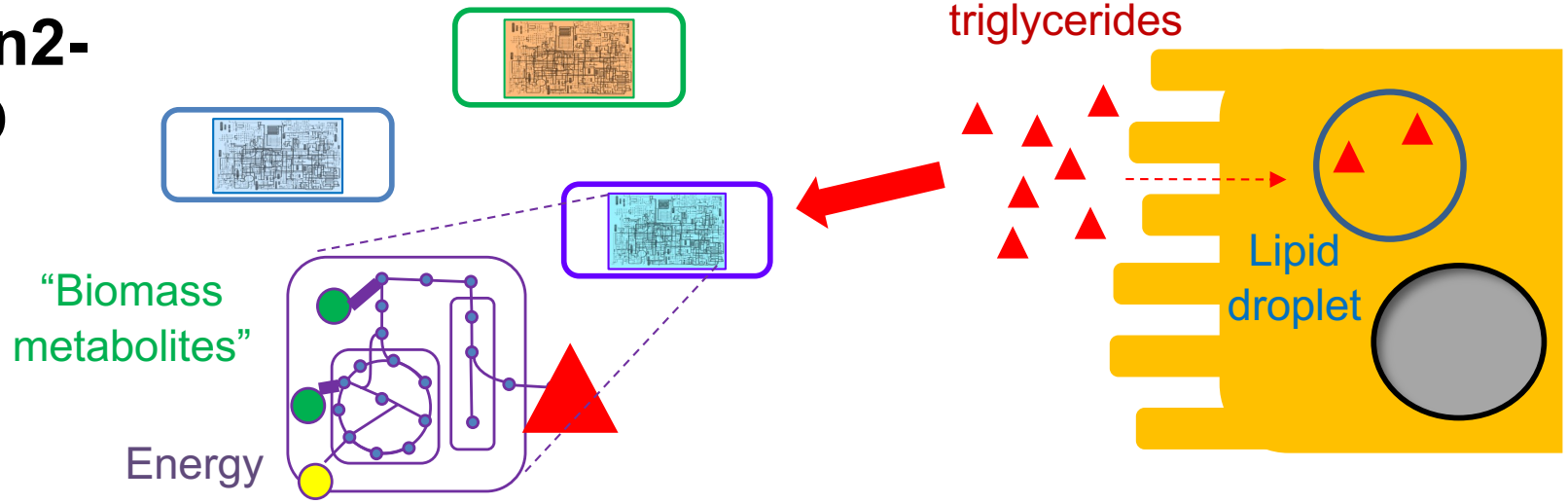
Glycolysis

Are energy producing pathways down-regulated due to accumulation of triglycerides under a high fat diet?

Wild type



Plin2-KO



Other Examples of Applications of Metatranscriptomics

ISME

www.nature.com/ismej

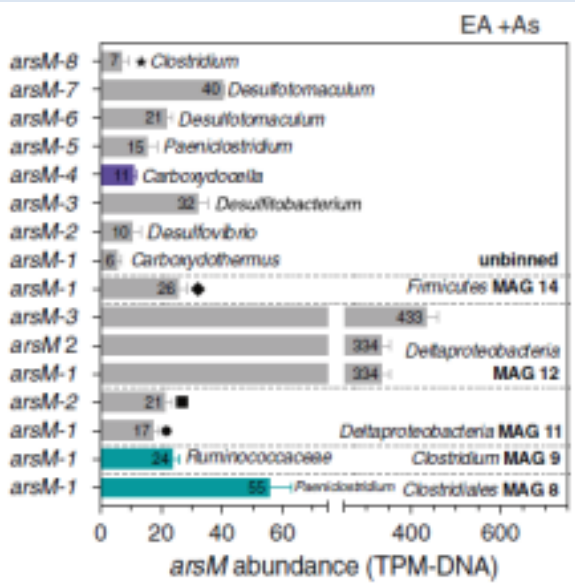
ARTICLE OPEN

Check for updates

Meta-omics-aided isolation of an elusive anaerobic arsenic-methylating soil bacterium

Karen Viacava[1,2], Jiangtao Qiao[1], Andrew Janowczyk[3], Suresh Poudel[4], Nicolas Jacquemin[1,5], Karin Lederballe Meibom[1], Him K. Shrestha[4,6], Matthew C. Reid[7], Robert L. Hettich[4] and Rizlan Bernier-Latmani[1]✉

Soil microbiomes were monitored by metagenomics, metatranscriptomics & proteomics, only taxa which expressed arsM genes were targeted for isolation (presence NOT sufficient)



nature microbiology

ARTICLES

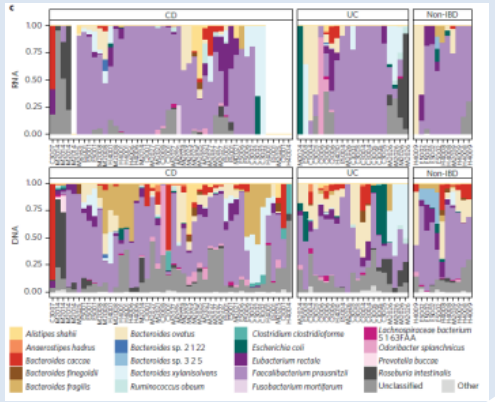
https://doi.org/10.1038/s41564-017-0089-z

Dynamics of metatranscription in the inflammatory bowel disease gut microbiome

Melanie Schirmer[1,2], Eric A. Franzosa[1,2], Jason Lloyd-Price[1,2], Lauren J. McIver[1,2], Randall Schwager[2], Tiffany W. Poon[1], Ashwin N. Ananthakrishnan[3], Elizabeth Andrews[3], Gildardo Barron[4], Kathleen Lake[5], Mahadev Prasad[6], Jenny Sauk[3,7], Betsy Stevens[3], Robin G. Wilson[3], Jonathan Braun[8], Lee A. Denson[5], Subra Kugathasan[6,9], Dermot P. B. McGovern[4], Hera Vlamakis[1], Ramnik J. Xavier[1,3,10,11]* and Curtis Huttenhower[1,2]*

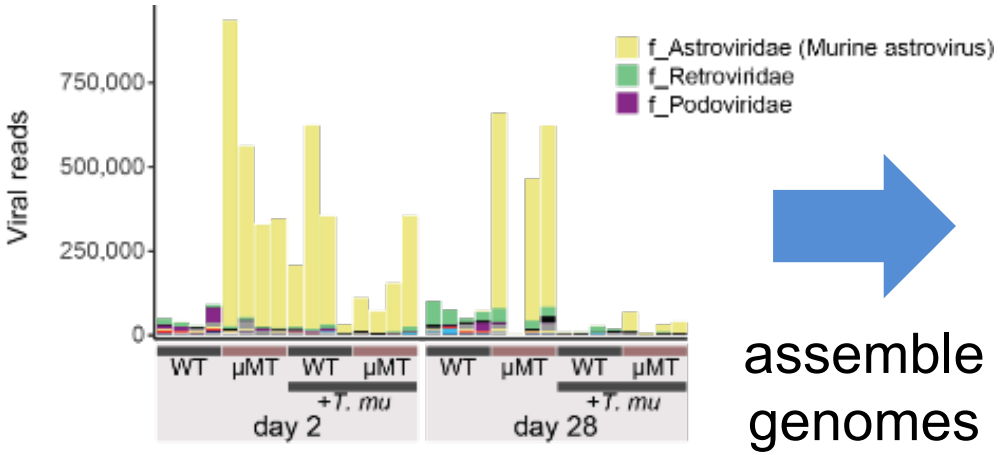
Applying metatranscriptomics to IBD samples found:

- Specific taxa providing unique pathway expression (glucoronate conversions by *F. prausnitzii*)
- Taxa that are abundant but not active (*Dialista invisus*)
- Important pathways can be contributed by different taxa

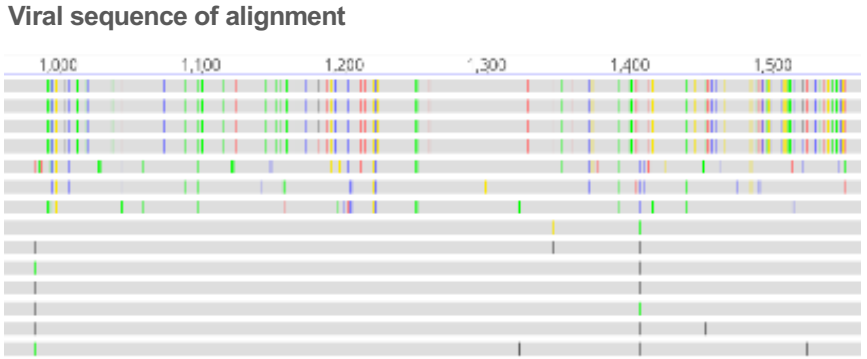


Recovery of RNA Viral Genomes

During assembly of metatranscriptomic datasets, assembly algorithms do a great job of recapitulating entire RNA viral genomes – here from mouse cecal samples



assemble genomes



Metatranscriptomics has been applied in this way to monitor RNA viral burdens in e.g. wastewater

Water Research 215 (2022) 118237

Contents lists available at ScienceDirect

Water Research

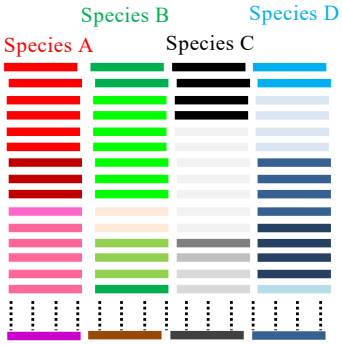
journal homepage: www.elsevier.com/locate/watres

A mixed blessing of viruses in wastewater treatment plants

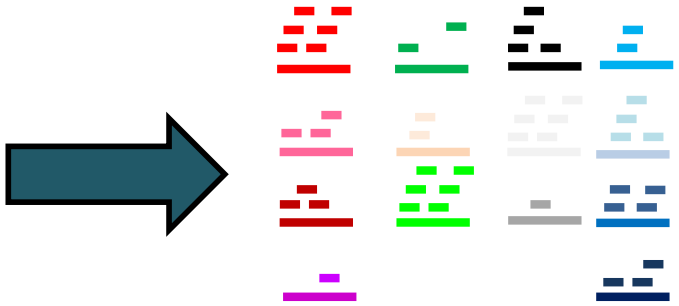
Ling-Dong Shi [a], Xiyang Dong [b,c], Zongbao Liu [d], Yuchun Yang [e], Jih-Gaw Lin [f], Meng Li [d,h], Ji-Dong Gu [g], Li-Zhong Zhu [a], He-Ping Zhao [a,*]

Metatranscriptomics through RNA-Seq

Extract RNA



Fragment and sequence



Align reads to known transcripts

Digital readout of bacterial gene expression

Metatranscriptomics is similar to single organism RNA-Seq

Typically applied to organisms with a reference (sequenced) genome, microbiome applications of RNASeq require specialized tools and approaches

Metatranscriptomics: Challenges

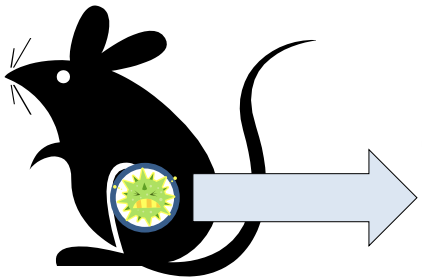
In a typical RNA-Seq experiment applied to a single eukaryotic organism, mRNA is isolated through polyA binding. After fragmentation and sequencing, reads are mapped to a reference genome using standard software to provide yield a readout on the relative abundance of the transcript

Microbiome samples face additional challenges

- compared to DNA, RNA is very unstable
- lack of polyA tails / host contamination
- complex datasets composed of hundreds of millions/billions of sequences
  - depth of sequencing
  - lack of reference sequences

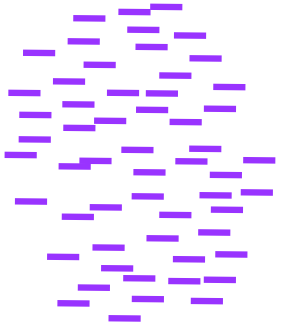
A typical metatranscriptomic analytical pipeline

1. Obtain RNA

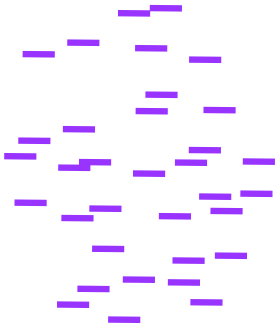


2. Prep for sequencing

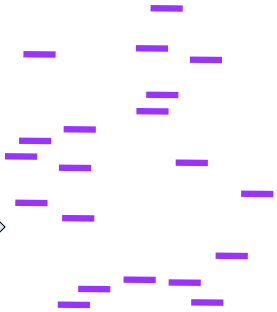
3. Generate Reads



4. Remove low quality



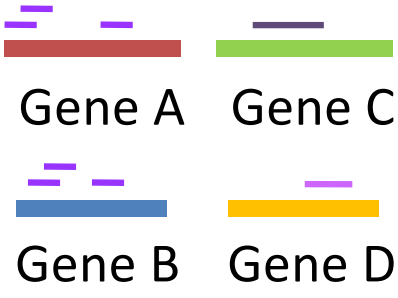
5. Remove rRNA reads



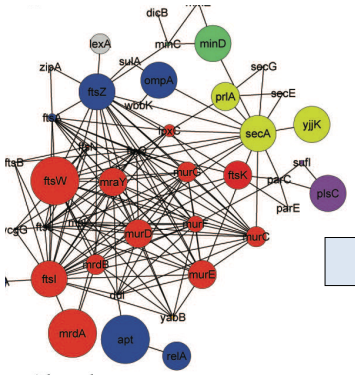
6. Remove host reads

7. Assemble

7. Identify bacterial transcript



8. Map to pathways



9. Sample comparisons

1. Sample collection and RNA extraction

Unlike DNA, RNA quality deteriorates rapidly – Method of storage and preparation can impact taxa recovered. While best practice is to process immediately to extract RNA then store at -80 or (next best) snap freeze in liquid nitrogen and store at -80, sometimes we can't do that…

Zymoresearch

DNA/RNA Shield
Fecal Collection Tube

Norgen

Stool Nucleic Acid
Collection and
Preservation Tubes

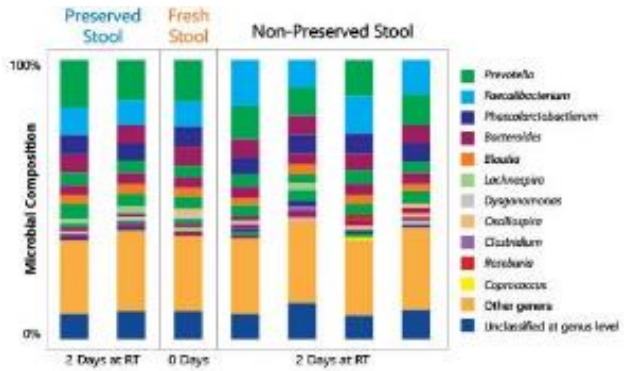
DNAGenoTek

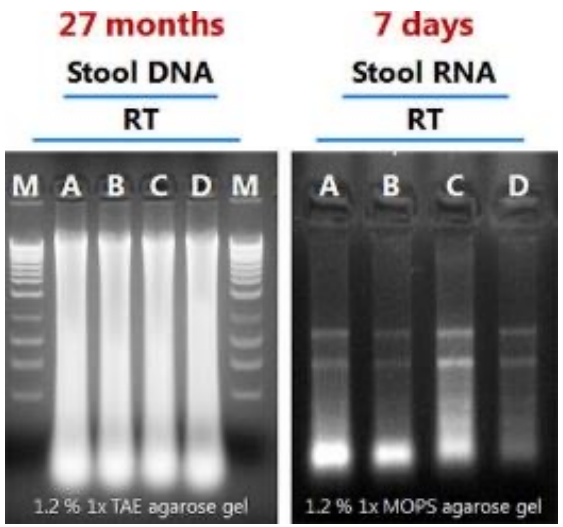
OMNIgene GUT
Released May 2022?

Avoid use of RNALater – it lyses some cells and can interfere with RNA extraction kits

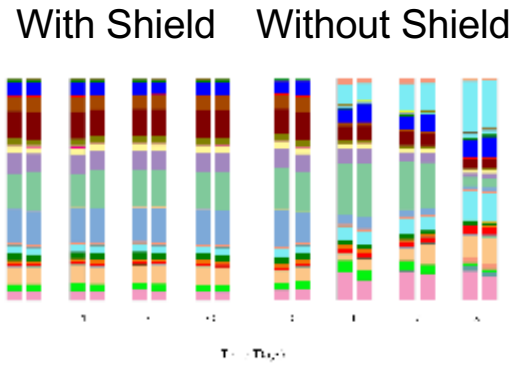
1. Sample collection and RNA extraction

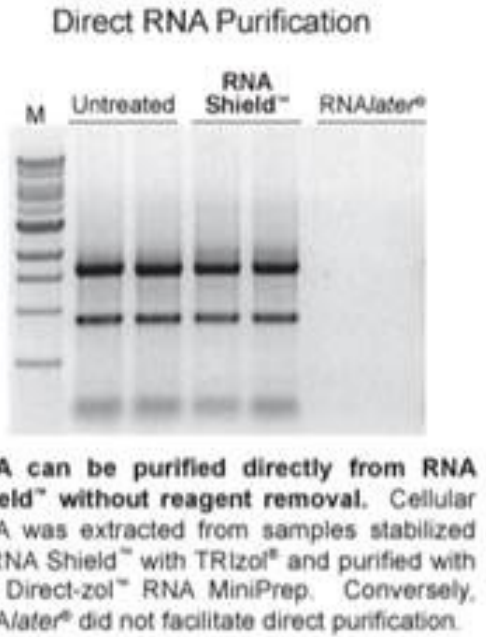
Norgen



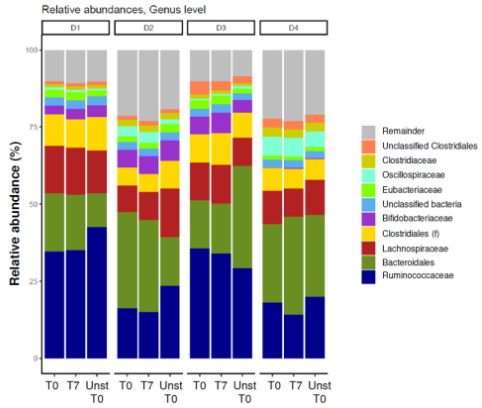


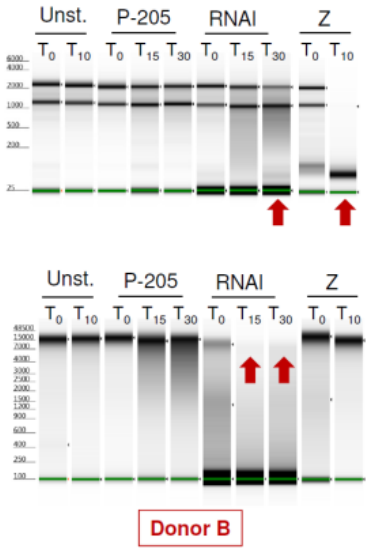
Zymoresearch





DNAGenoTek





1. Sample collection and RNA extraction

Metatranscriptomics is expensive mainly due to library preparation

Cost per sample (40 million reads)

~$300-$400

How many replicates?

"At least 4!" (depends on nature of samples)

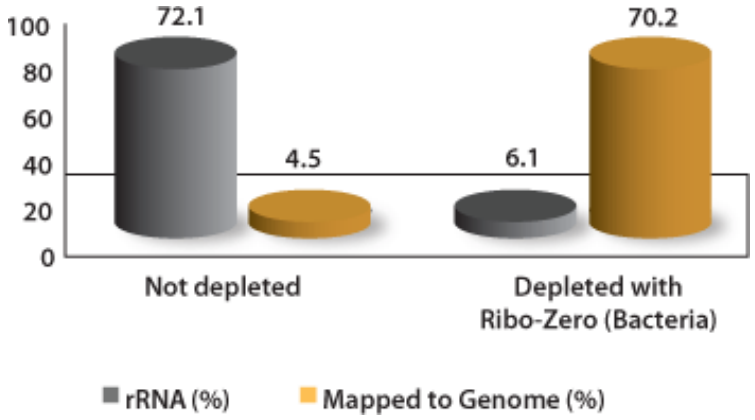
Lack of reference sets make power analyses challenging

2. Preparing sample for sequencing

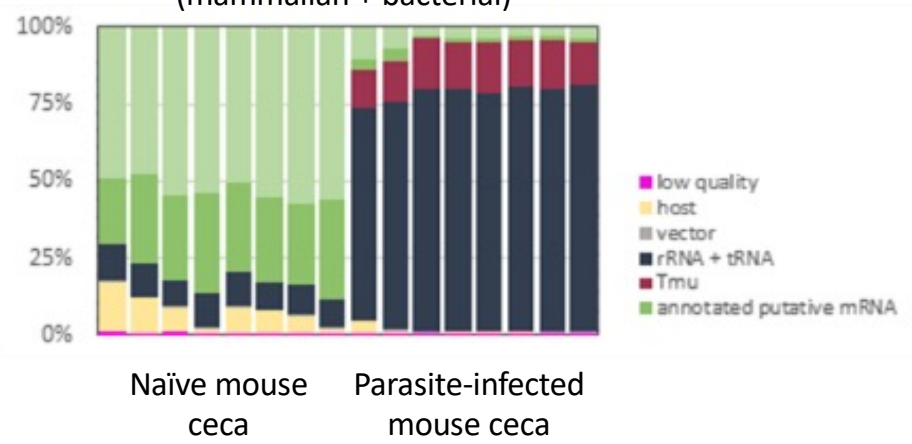
Bacterial mRNA's lack a polyA tail so how to remove abundant rRNA species?

Once RNA has been extracted, several kits are available to remove rRNA – need 500ng-2.5ug RNA/sample

Ribo-Zero (Illumina) provides reasonable success



NEB rRNA depletion kits (mammalian + bacterial)



Host mRNAs can also prove challenging – can also be informative!

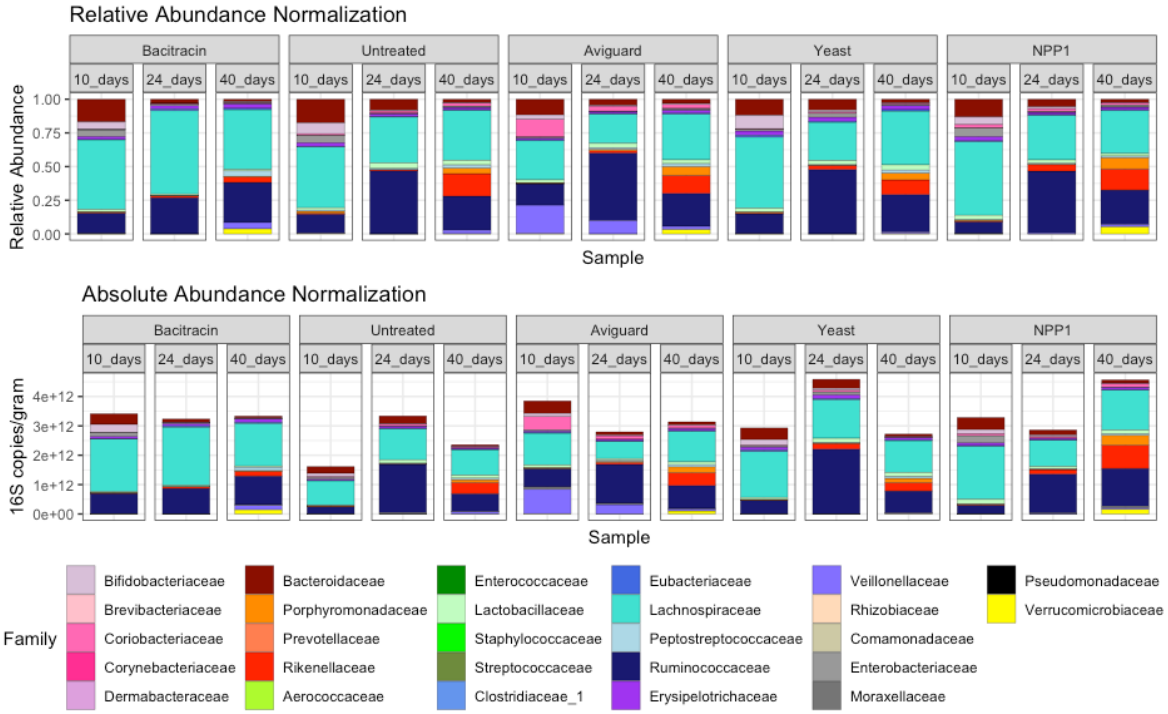
2. Spike in to quantify read abundance

Typical microbiome experiments yield only ‘relative abundance’ data which can yield misleading results – ‘absolute abundance’ requires quantification of bacterial cells in initial sample (e.g. CFU counts, Flow Cytometry, spike in’s)

ZYMO high microbial load spike-in

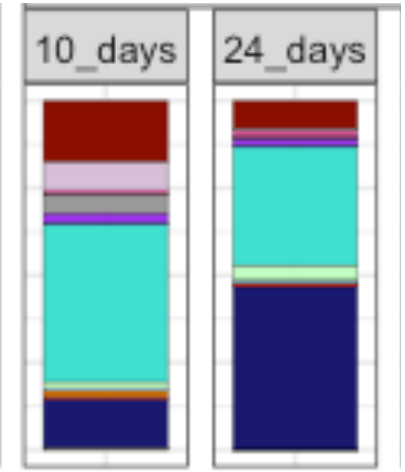
The spike-in consists of two species: *Imtechella halotolerans* and *Allobacillus halotolerans*, totaling 4 x 10^7 cells.

Tested on Cecal, Ileal and Jejunum samples from Chickens

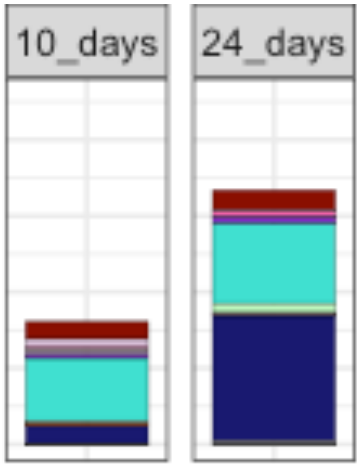


2. Absolute abundance analyses alter results

Relative abundance



Absolute abundance

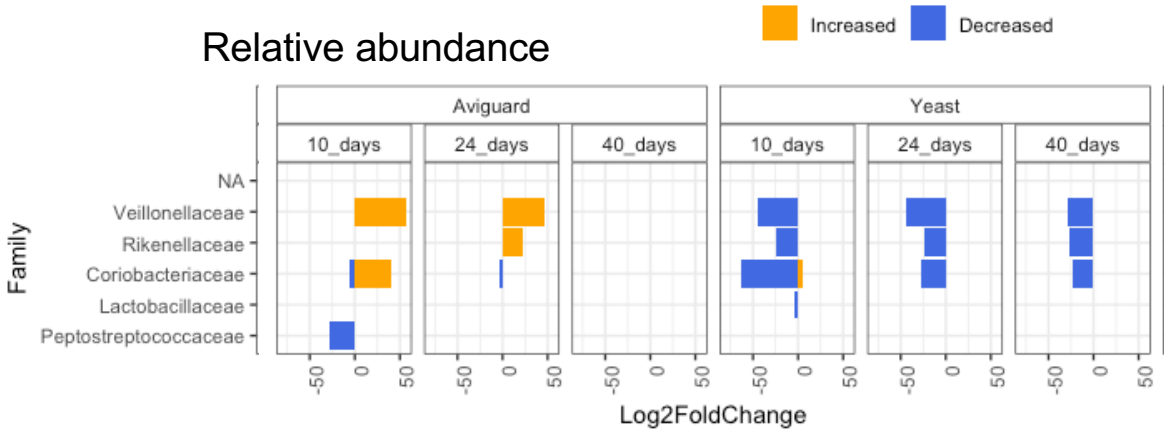


Decrease in Bacteroidaceae at 24 days post hatch?

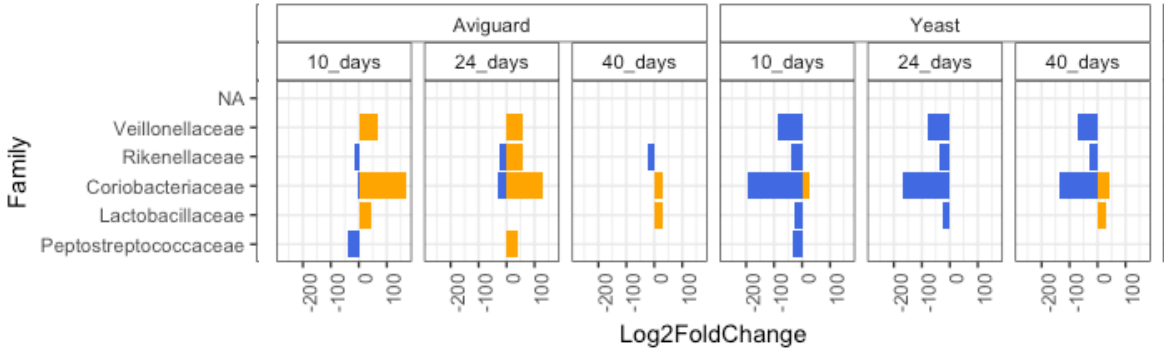
No change in Bacteroidaceae at 24 days post hatch

Accounting for absolute abundance alters taxa detected as significantly abundant – will impact significantly differentially expressed genes too

Relative abundance

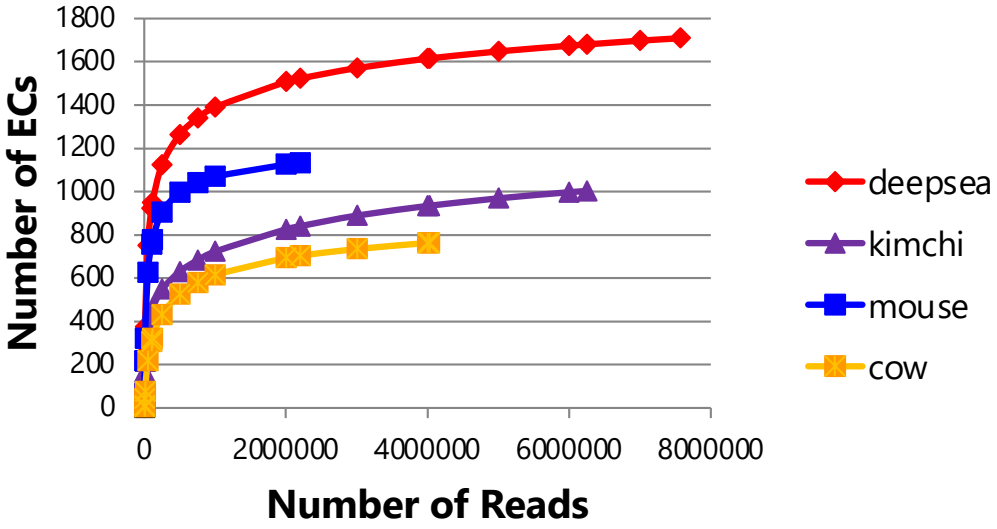


Absolute abundance



3. Generating reads

How many reads are “enough”?



While PacBio and MiSeq provide long reads: great for metagenomics, Metatranscriptomics requires large numbers of reads - Novoseq

~5 million mRNA reads provide 90-95% of enzymes (ECs) in a microbiome
With kits yielding mRNA read rates of ~25%, this suggests 20 million/sample mRNA

4. Analysing the data

Metatranscriptomics is a relatively new field requiring robust tools and pipelines to process and analyse – (SAMSA2 / HUMAnN3 and **MetaPro**)

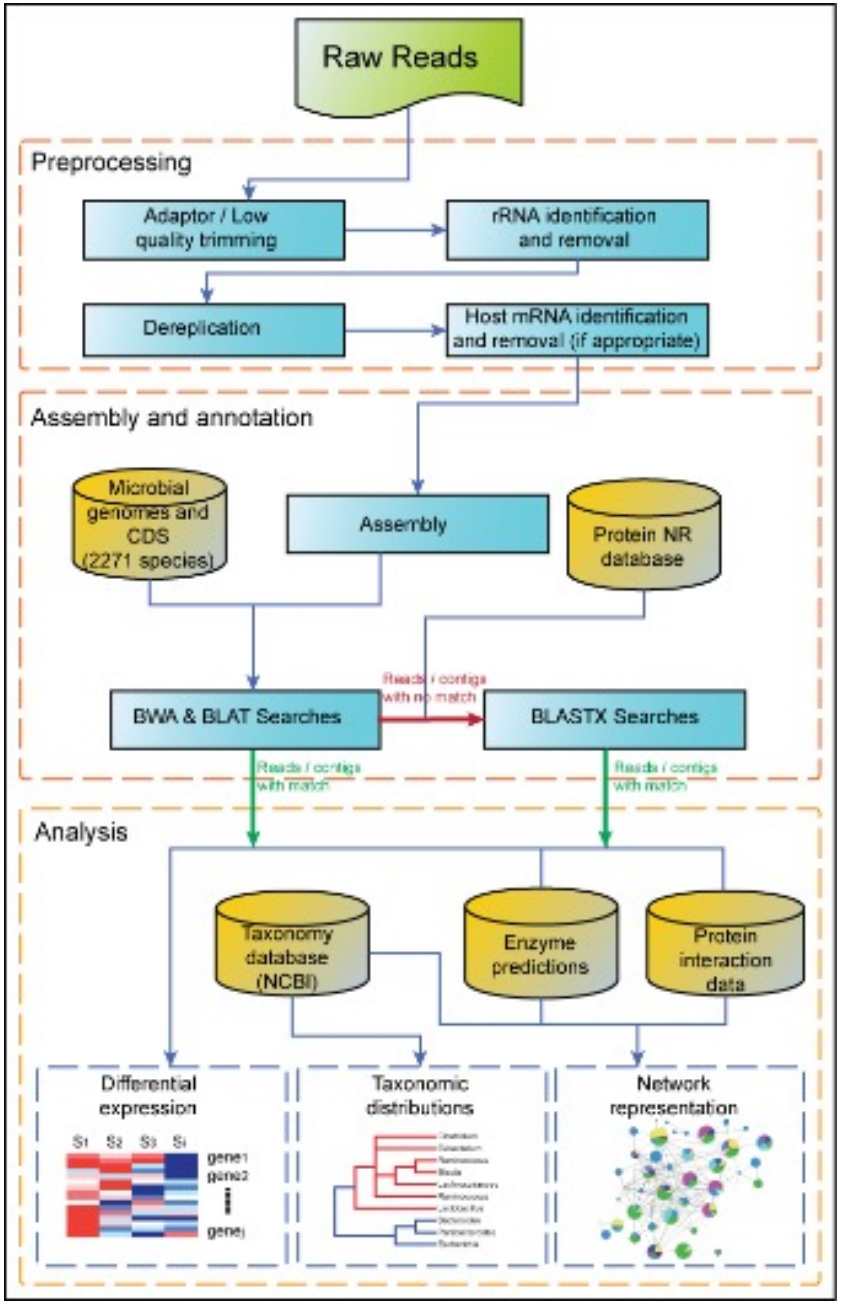
Due to their size (billions of sequence reads) – compute clusters are key

Read processing - filtering

To identify reads derived from mRNA bioinformatics pipelines need to be in place that remove contaminating reads:

Low quality - *Trimmomatic*
Adaptors – *Trimmomatic*
Host – *BWA / BLAT*
rRNA – *BLAT / Infernal*

Of these Infernal is the most time consuming but is considerably more sensitive than sortmeRNA

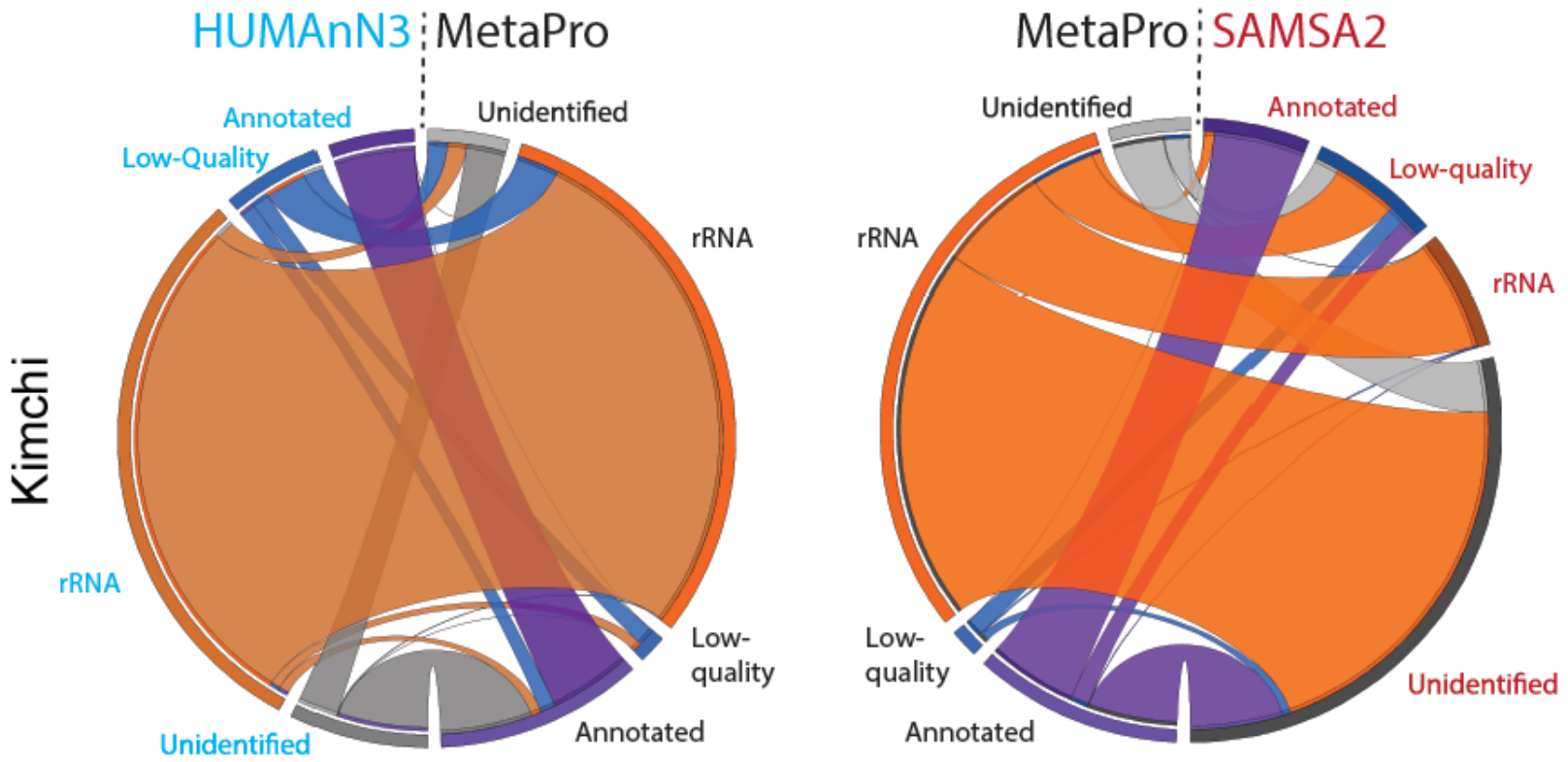


Jiang et al Microbiome 2016
Xiong et al Microbiome 2017

Comparing pipelines

Several pipelines are available for processing metatranscriptomic datasets
- e.g. MetaPro, HUMAnN3 and SAMSA2

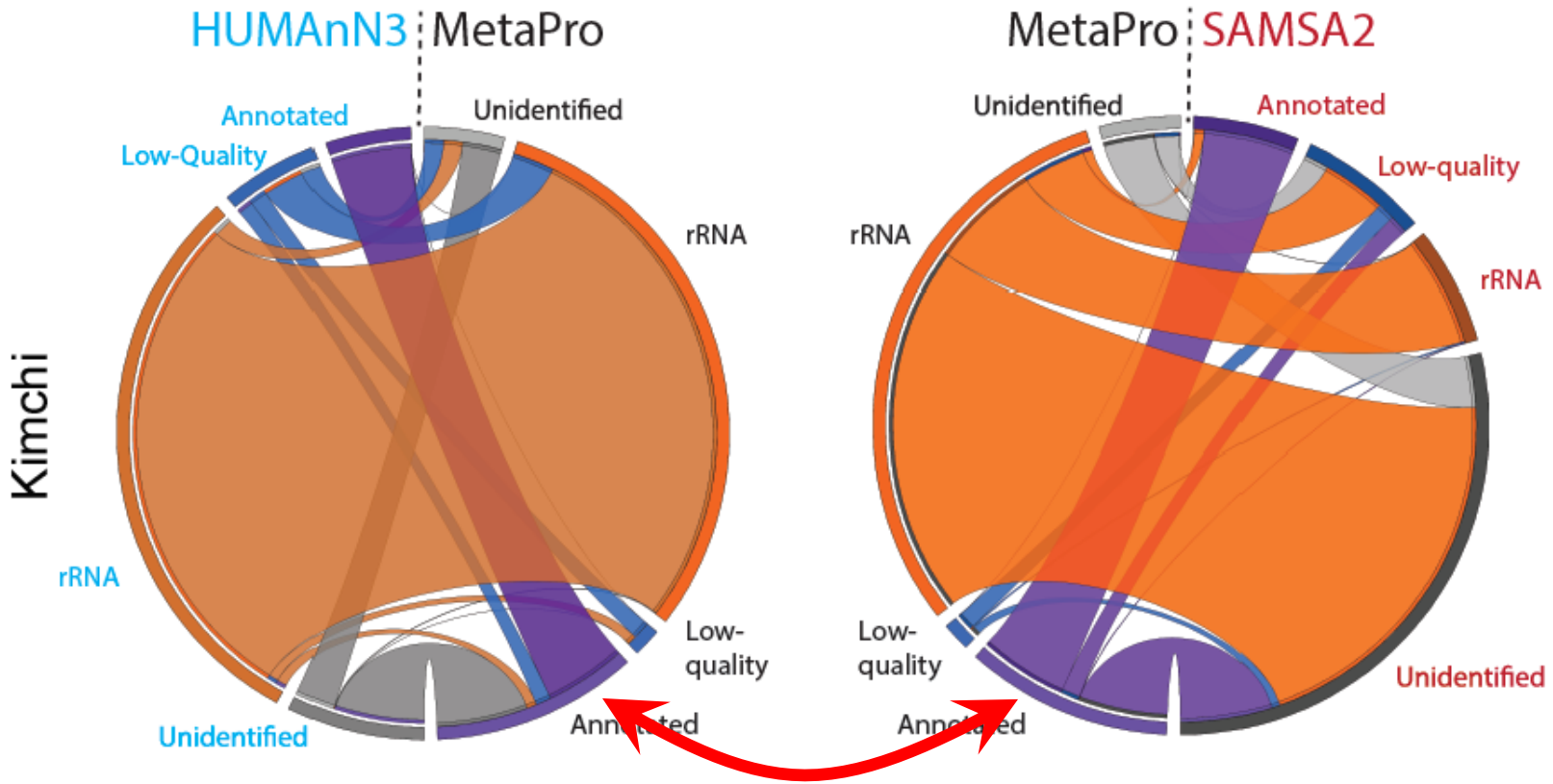
You could also build your own!



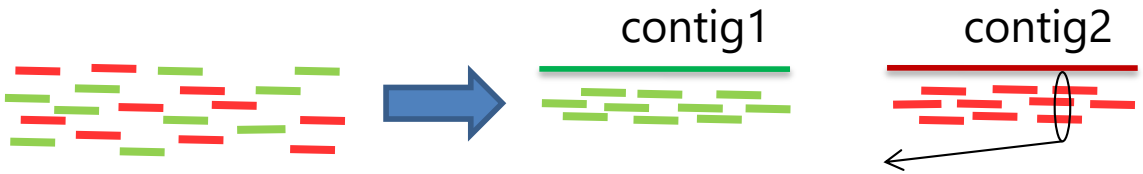
Comparing pipelines

Several pipelines are available for processing metatranscriptomic datasets
- e.g. MetaPro, HUMAnN3 and SAMSA2

You could also build your own!

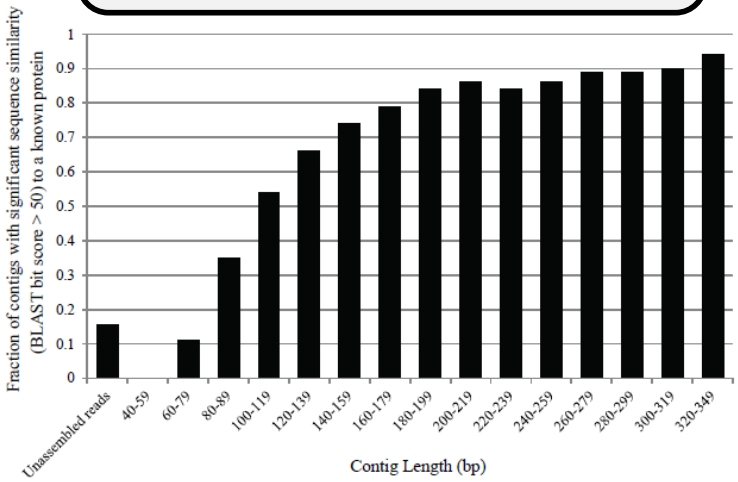


Read processing - Assembly



attagcggcgattttcggcgatcttatcttgatctgggcgcgtatcggtagcgtagcgattcgtagc
attagcggcgattttcggcgatcttatcttgatctgggcgcgtatcggtagcgtagcgattcgtagc
attagcggcgattttcggcgatcttatcttgatctgggcgcgtatcggtagcgtagcgattcgtagc

Assembly improves annotation accuracy



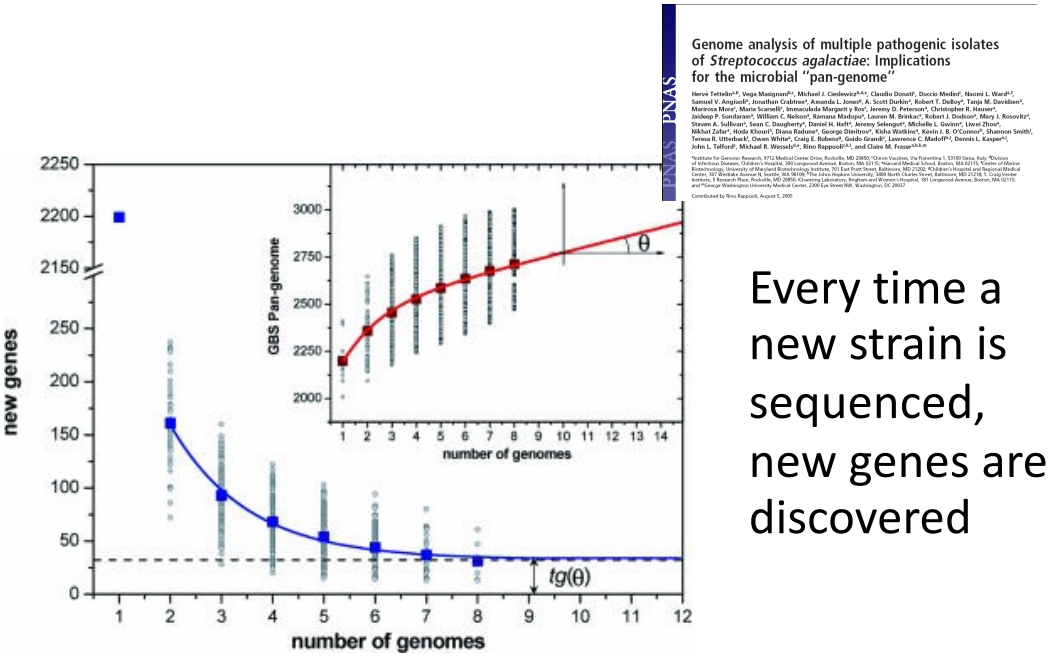
Here we use SPADes followed by MetaGeneMark to identify separate ORFs in transcripts that may represent operons

Chimera's, misassembled contigs, can become a problem due to reads derived from orthologs from different species

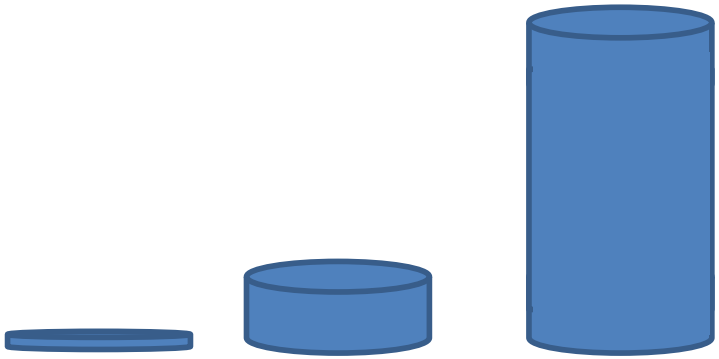
Read processing – Annotating to genes

Functional annotations rely on sequence similarity searches

BWA -> Fastest, strict
BLAT -> Fast, less strict
DIAMOND -> Slow, sensitive



Every time a new strain is sequenced, new genes are discovered



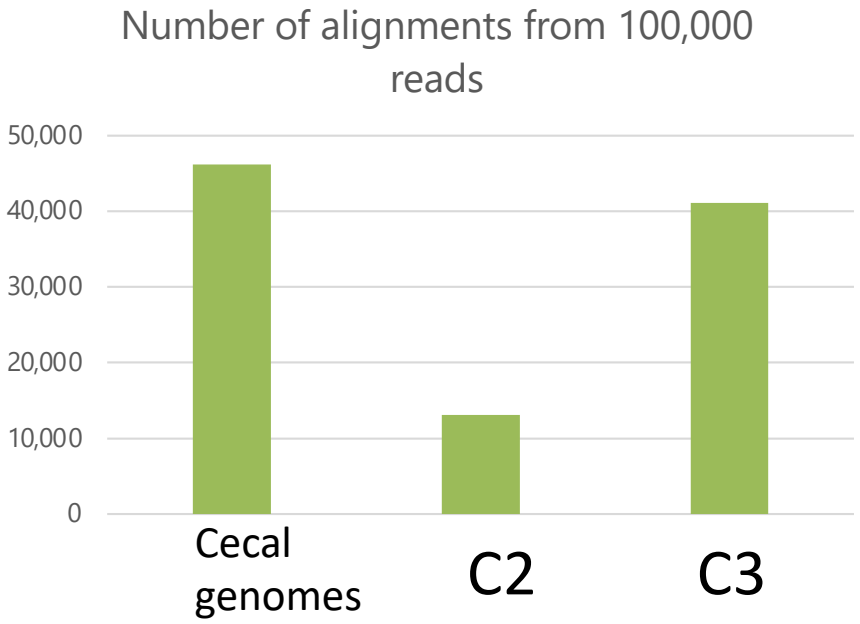
As the number of reference genomes increases, the amount of memory to perform BWA searches increases

Large memory compute clusters (100's Gb RAM)
Software solutions (splitting databases)

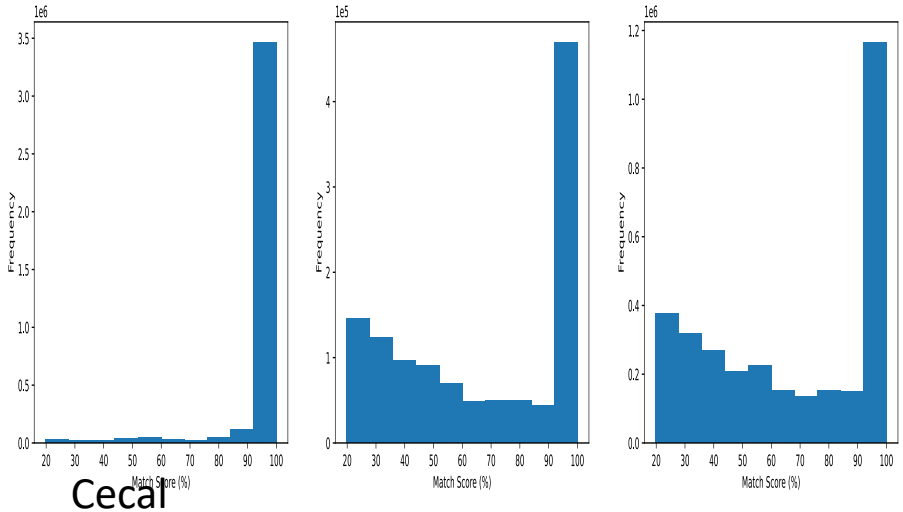
Custom databases can be faster and more accurate

Instead of searching against all genomes, a more appropriate subset can reduce search times

Here we compare the performance of a dataset of ~500 genomes assembled from the cecal microbiome of chickens (1.1 Gb) with two versions of the ChocoPhlan Database used by HUMAnN (19Gb & 66Gb). The smaller, focused. database runs 100x faster than C3.



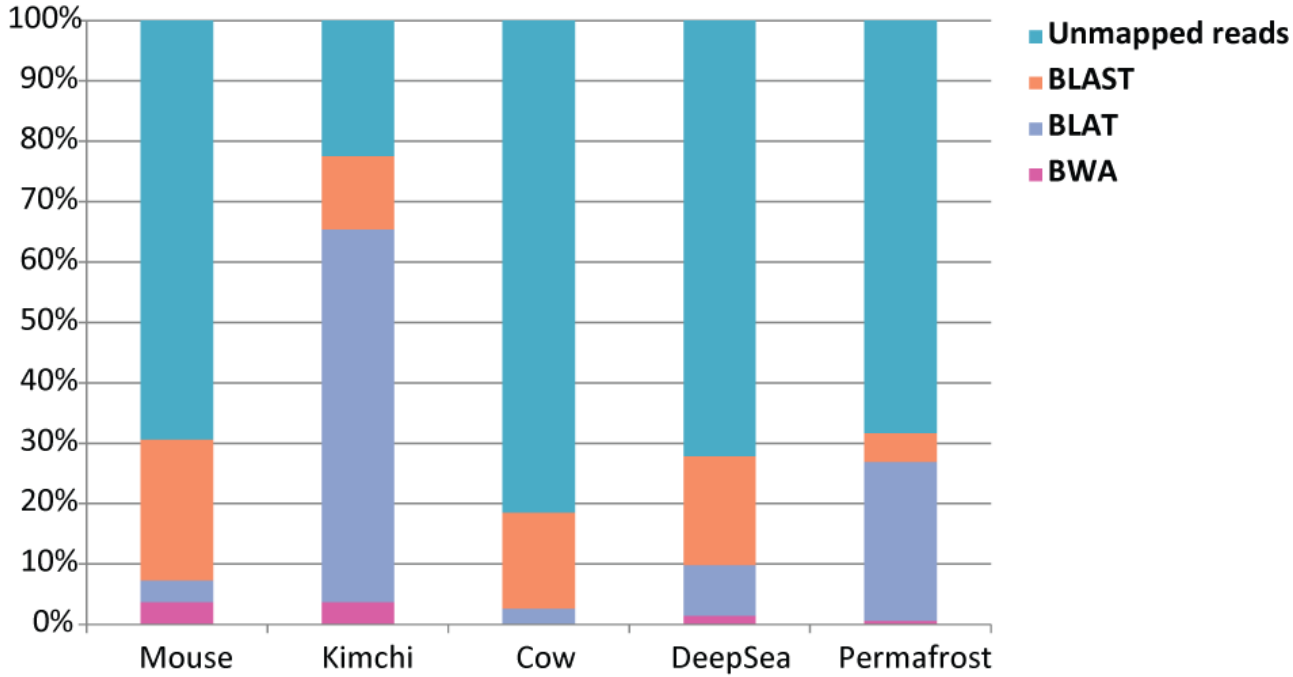
Quality of alignments



Read processing – annotating to genes

BWA works at the level of nucleotide sequences and requires precise matches – can be challenging for undersampled samples

DIAMOND/BLAST can work at the level of protein sequence, allowing flexibility at the nucleotide level



Level of annotation dependent on previous sampling of the niche explored

Even with DIAMOND/BLAST many reads remain unannotated

Quality of BLASTX matches

Typical match to a 71bp read

E-Value = 39 (NOT e^{-39})

```
>ref|ZP_06984083.1|  conserved hypothetical protein [Bacteroides sp. 3_1_19]
 gb|EFI10148.1|  conserved hypothetical protein [Bacteroides sp. 3_1_19]
Length=1166

 Score = 31.6 bits (70),  Expect =   39
 Identities = 16/23 (70%), Positives = 16/23 (70%), Gaps = 0/23 (0%)
 Frame = +3

Query  3     TYSVEAKESFRFFLTLDEAYSQS  71
             TYSVEA  SF F LTLD  Y QS
Sbjct  1032  TYSVEAGGSFSFSLTLDTDYDQS  1054
```

Species looks about right though

Summary for all matches

% of Read Length

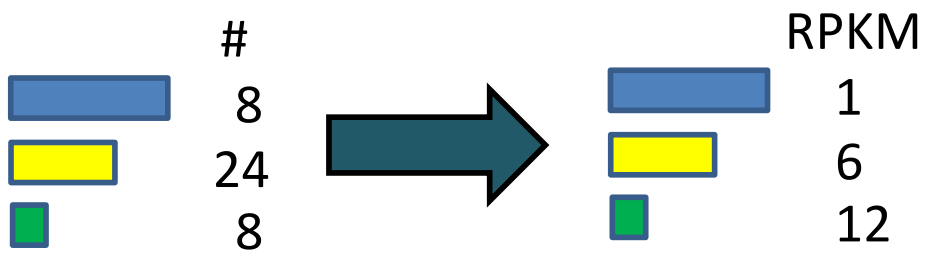
% ID of match

	55	60	65	70	75	80	85	90	95	100
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
45	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1
55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.4	0.3
60	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.5	0.7	1.6
65	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.9	1.3	1.5
70	0.0	0.0	0.0	0.1	0.1	0.3	0.3	2.2	1.6	1.7
75	0.0	0.0	0.1	0.2	0.3	0.4	0.5	1.6	3.8	2.0
80	0.0	0.0	0.0	0.3	0.5	0.5	0.5	1.7	2.4	4.9
85	0.0	0.0	0.1	0.5	0.9	1.1	0.8	1.8	2.7	3.3
90	0.0	0.0	0.2	0.9	0.6	0.6	1.7	1.9	3.3	4.0
95	0.1	0.1	0.2	0.8	0.6	0.6	0.7	1.9	3.7	4.8
100	0.1	0.2	0.3	0.6	0.7	0.7	1.3	2.9	8.6	12.1

Read processing – converting mappings to expression

To normalize expression levels to account for differences in gene length, read counts are converted to ***Reads per kilobase of transcript mapped (RPKM)***

Expression is biased for gene length (longer transcripts should have more reads) to normalize, reads are converted to Reads per Kilobase of transcript per million reads mapped



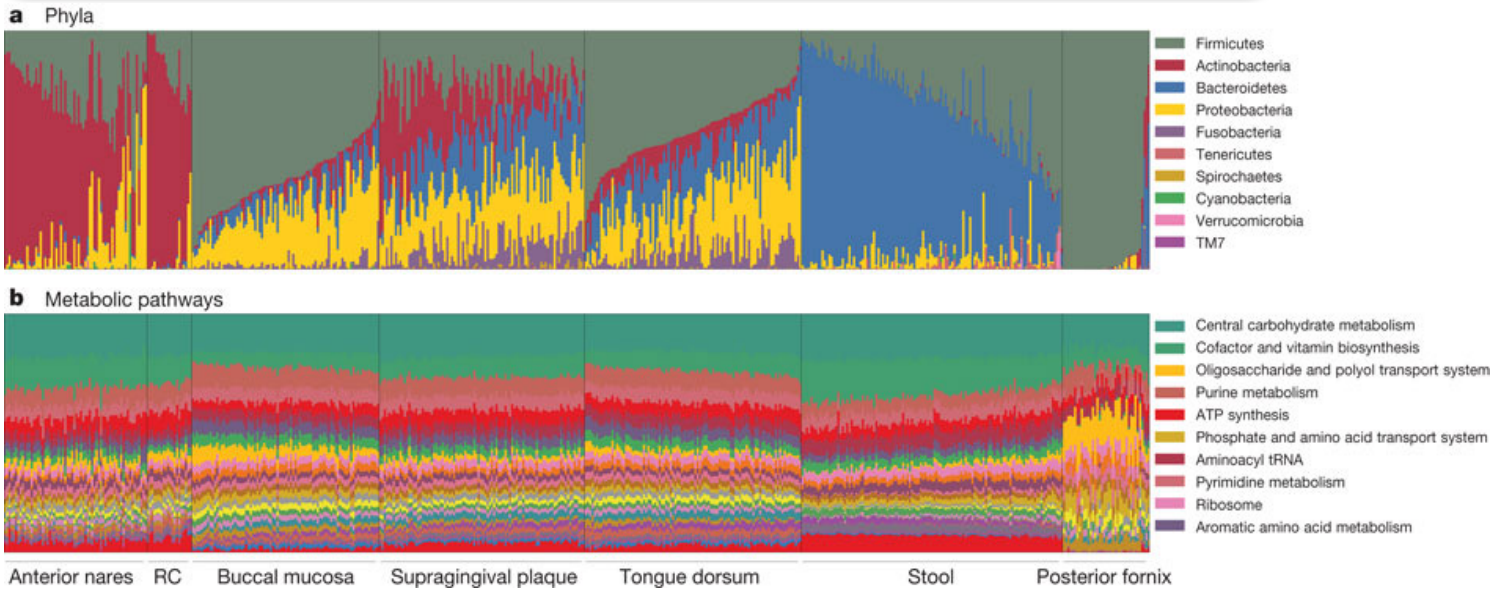
$$RPKM_{geneA} = 10^9 C_{geneA} / NL$$

C_{geneA} = number of reads mapped to geneA
N = total number of reads
L = length of transcript in units of Kb

Several software tools available to do mapping and calculate normalized expression (Bowtie and Cufflinks) or can be included as a simple calculation in your pipeline

Read processing – taxonomic annotation

Previous studies have shown that microbiomes can vary significantly in taxonomic contributions while yielding similar functionality



Knowing which species are providing which functions may not be that important - On the other hand, assigning RNA reads to taxa may reveal critical functions contributed by keystone taxa, can also help in binning for assembly

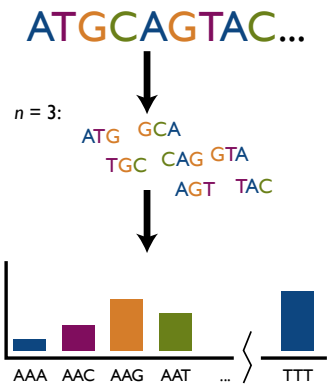
How do we assign taxonomic information?

The human microbiome consortium 2012

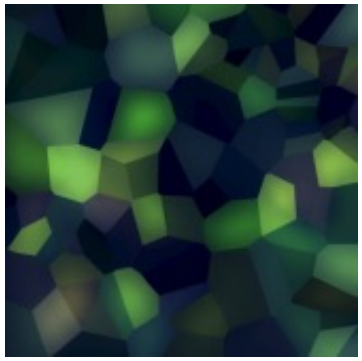
Read processing – taxonomic annotation

Alignment based methods such as BWA and DIAMOND can fail where we lack suitable reference genomes – particularly for short read datasets where assignments may be ambiguous

Compositional methods (e.g. nt frequency, codon bias) offer alternative strategies



Here a sequences is classified into frequencies of 3-mers



Nearest neighbours methods then try to assign a sequence to the genome with the closest distribution

Kaiju - 2016

ARTICLE

Received 26 Nov 2015 | Accepted 7 Mar 2016 | Published 13 Apr 2016

Fast and sensitive taxonomic classification for metagenomics with Kaiju

Peter Menzel[1], Kim Lee Ng[1] & Anders Krogh[1]

Published online 31 August 2012 Nucleic Acids Research, 2013, Vol. 41, No. 1 e3

Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms

Jiemeng Liu[1,2,3], Haifeng Wang[1,4], Hongxing Yang[1,4], Yizhe Zhang[5], Jinfeng Wang[6], Fangqing Zhao[6,*] and Ji Qi[1,4,*]

NBC - 2011

Method

Centrifuge: rapid and sensitive classification of metagenomic sequences

Daehwan Kim,[1,4] Li Song,[1,2,4] Florian P. Breitwieser,[1,4] and Steven L. Salzberg[1,2,3]

Centrifuge - 2016

Wood and Salzberg Genome Biology 2014, 15:R46

Genome Biology

METHOD Open Access

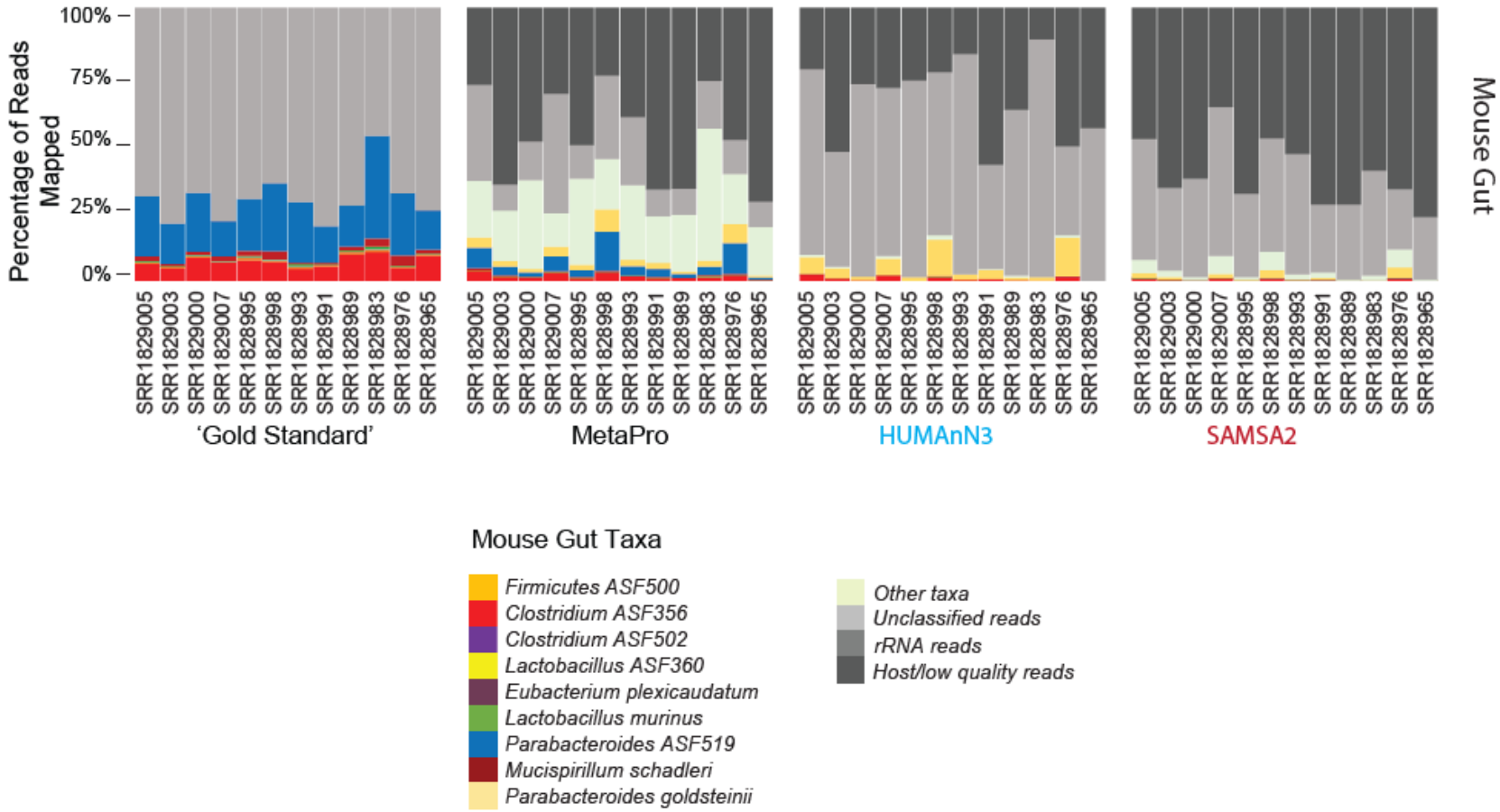
Kraken: ultrafast metagenomic sequence classification using exact alignments

Derrick E Wood[1,2*] and Steven L Salzberg[2,3]

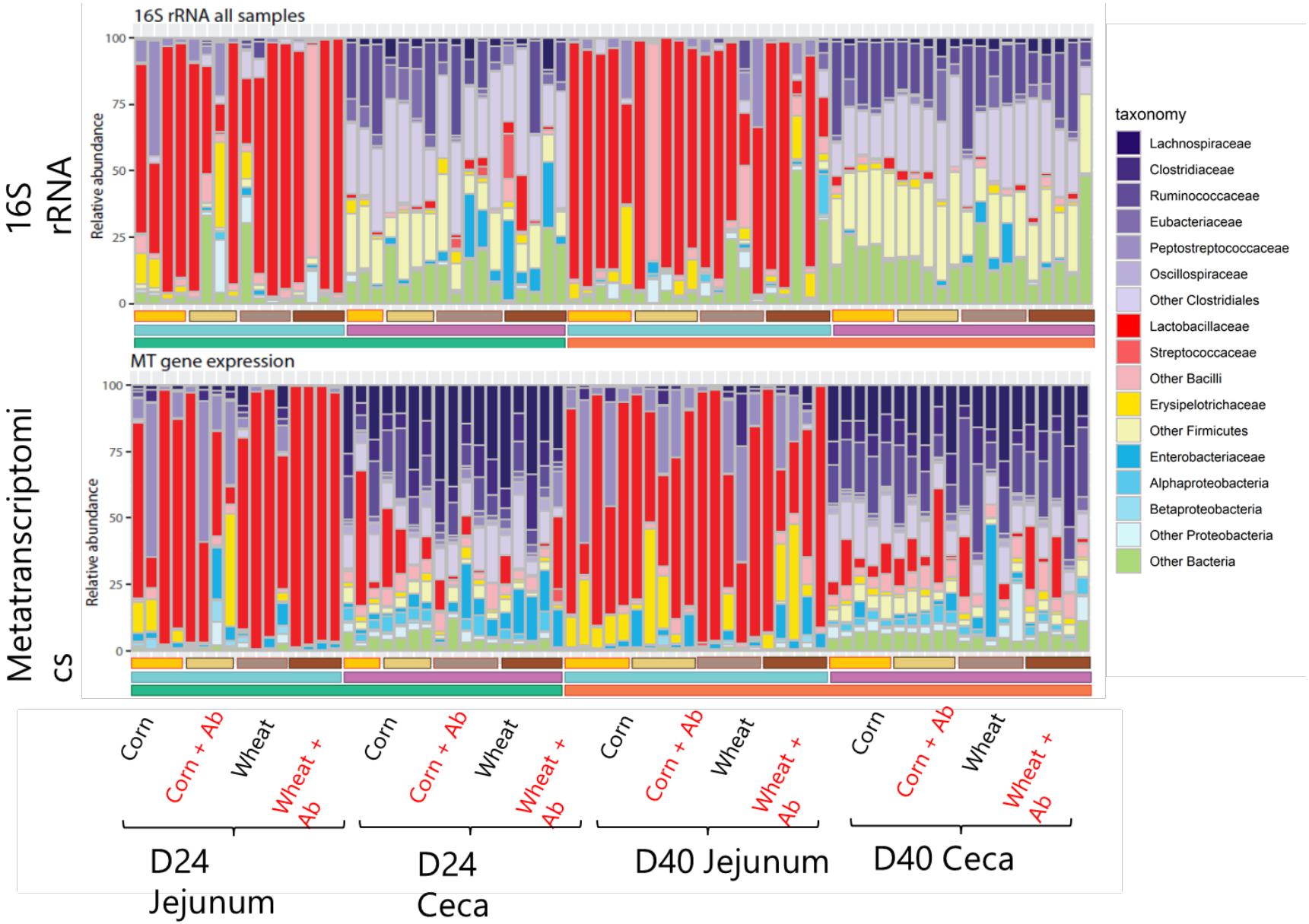
KRAKEN – 2014
KRAKEN2 - 2019

MetaPro employs “majority voting” for taxonomic classification

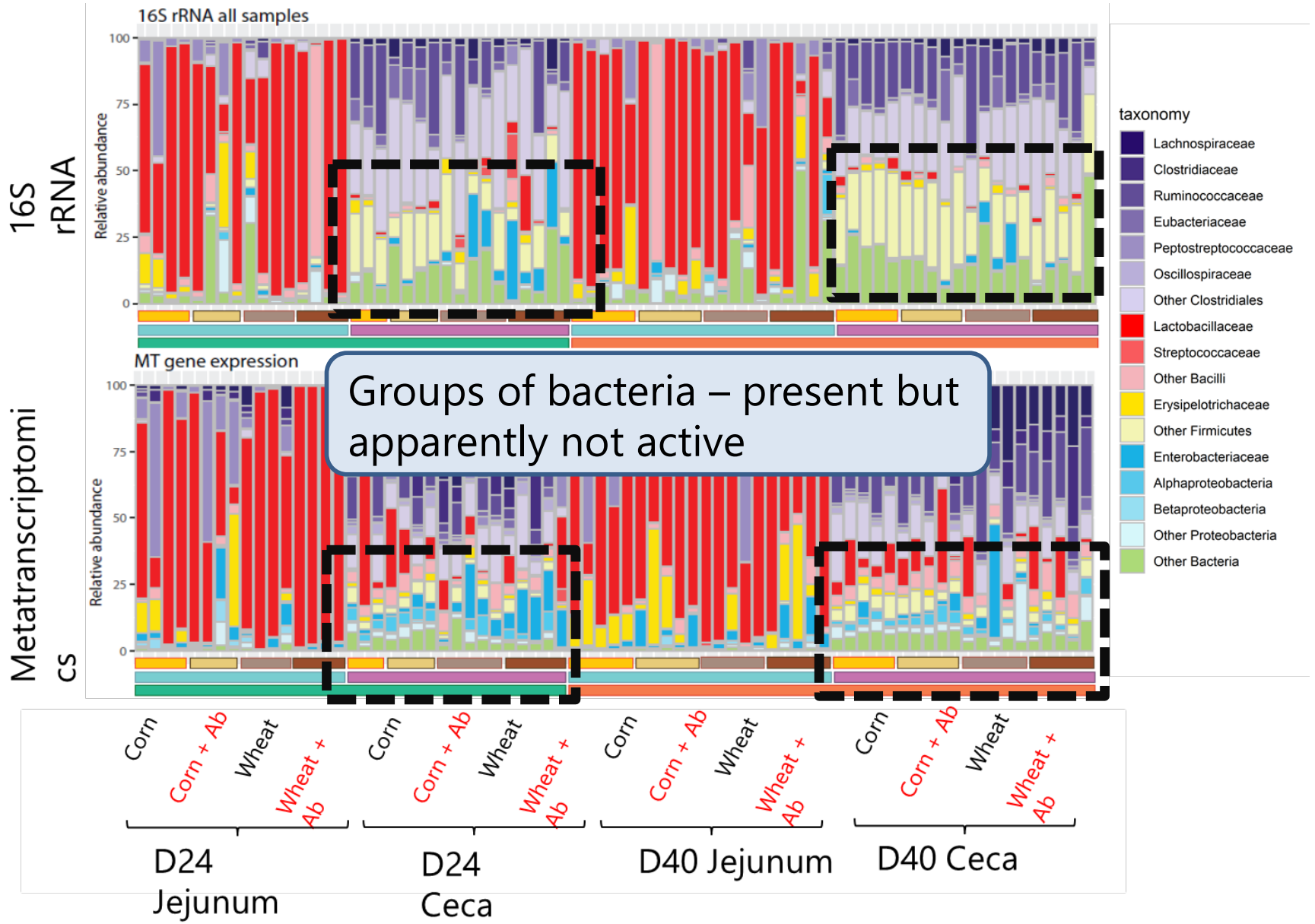
Different tools exhibit a range of sensitivity and recall. MetaPro combines results from three classifiers (DIAMOND searches, Kaiju and Centrifuge) to perform taxonomic assignments



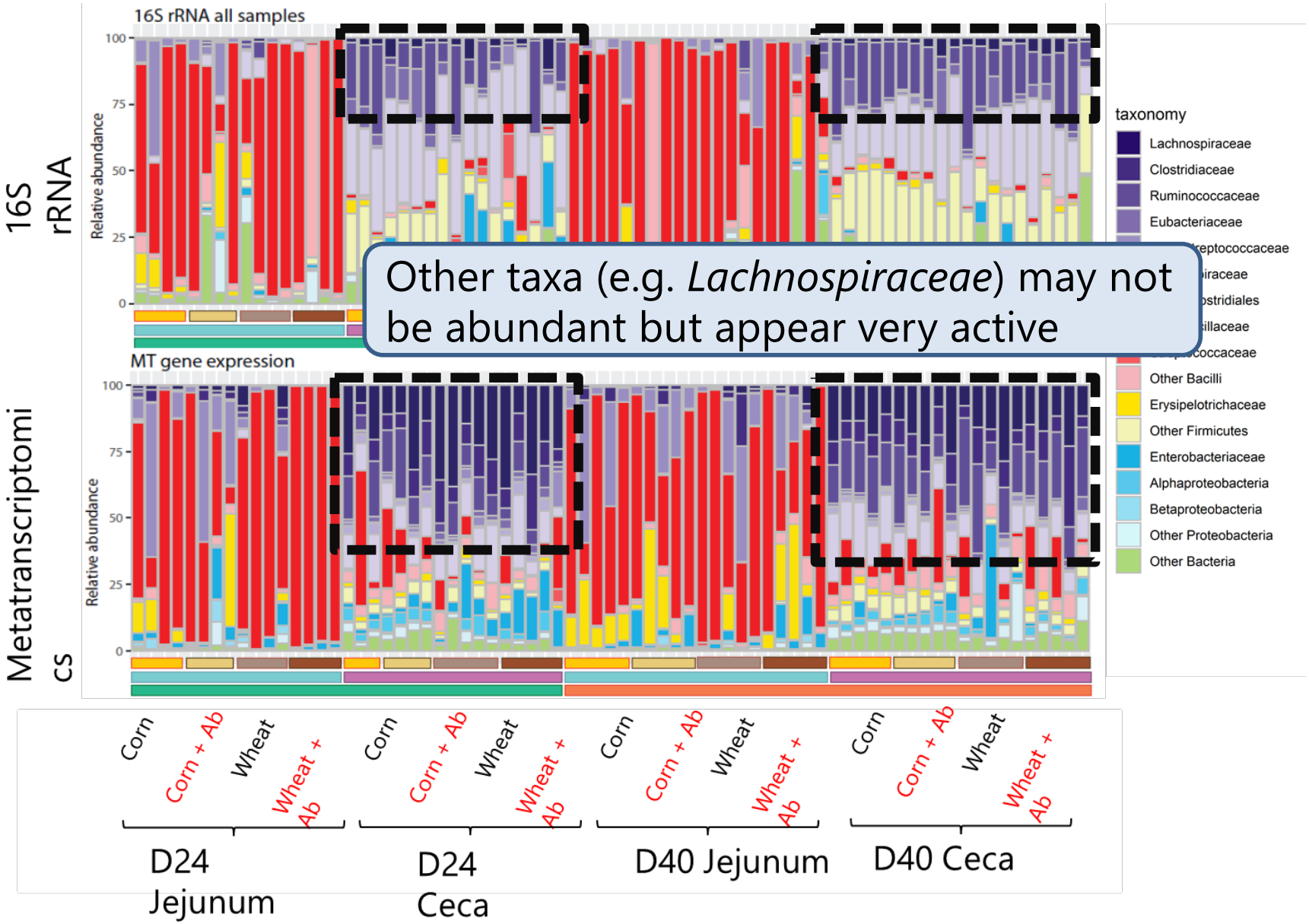
16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles



16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles



16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles



16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles

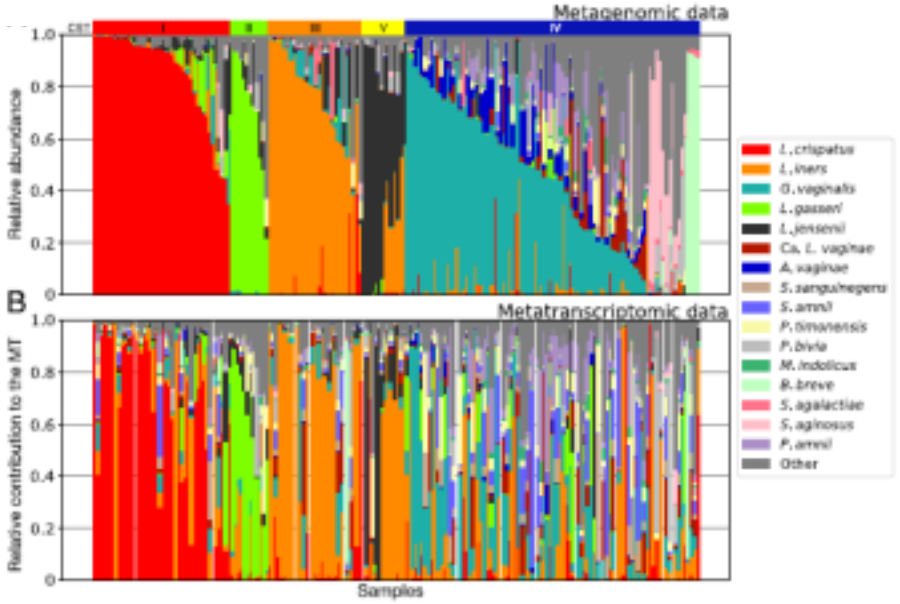
France et al. Genome Biology (2022) 23:66
https://doi.org/10.1186/s13059-022-02635-9

Genome Biology

RESEARCH Open Access

Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data

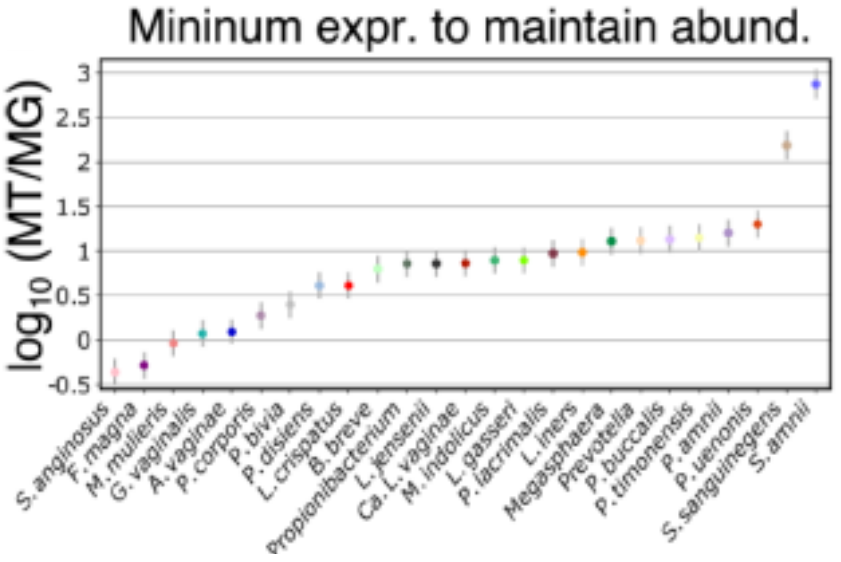
Michael T. France[1], Li Fu[1], Lindsay Rutt[1], Hongqiu Yang[1], Michael S. Humphrys[1], Shilpa Narina[1], Pawel M. Gajer[1], Bing Ma[1], Larry J. Forney[2] and Jacques Ravel[1*]



A recent study comparing metagenome and metatranscriptome data from vaginal swab samples also found differences between metagenomic and metatranscriptomic profiles

Metatranscriptomic data more reflective of future abundance

Given DNA is slow to degrade, does RNA provide a more accurate reflection of community abundance?



Functional Annotation

Once reads have been assigned to transcripts, transcripts may already be annotated with functions – otherwise annotations can be generated through mapping to annotated orthologs

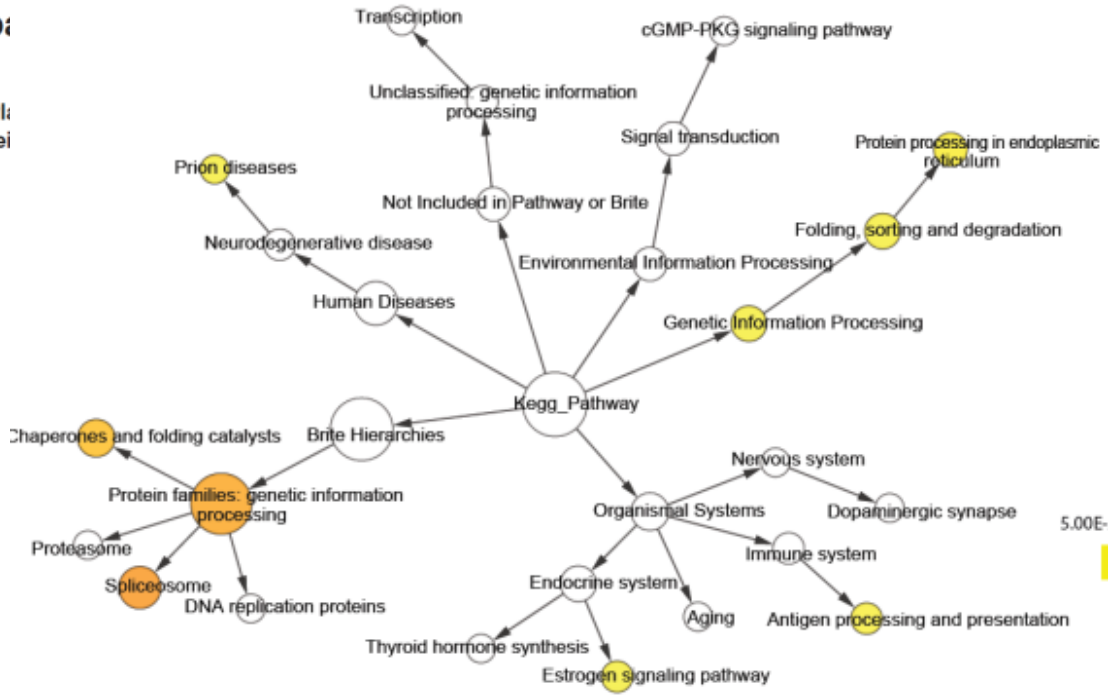
Published online 12 November 2018 *Nucleic Acids Research, 2019, Vol. 47, Database issue D309–D314*
doi: 10.1093/nar/gky1085

eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource ba on 5090 organisms and 2502 viruses

Jaime Huerta-Cepas[1,2,*,†], Damian Szklarczyk[3,†], Davide Heller[3], Ana Hernández-Pl
Sofia K. Forslund[1,4], Helen Cook[5], Daniel R. Mende[6], Ivica Letunic[7], Thomas Rattei
Lars J. Jensen[5], Christian von Mering[3] and Peer Bork[1,9,10,11,*]

EggNOG provides mappings to Gene Ontology (GO), KEGG enzymes, KEGG modules, CAZy

GO terms can be challenging to summarize…

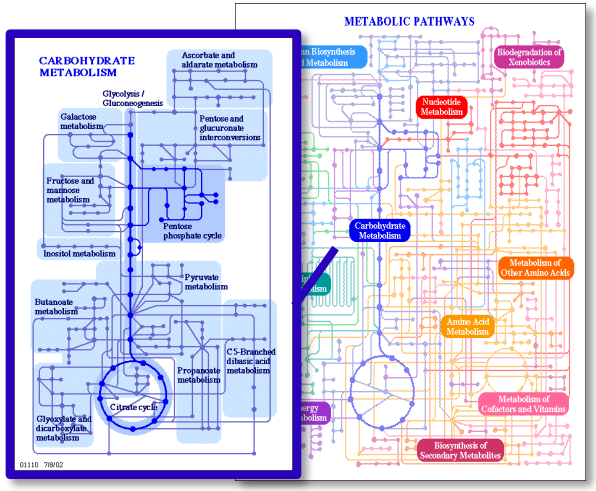


BinGO plug in for cytoscape can help interpretation of GO annotations

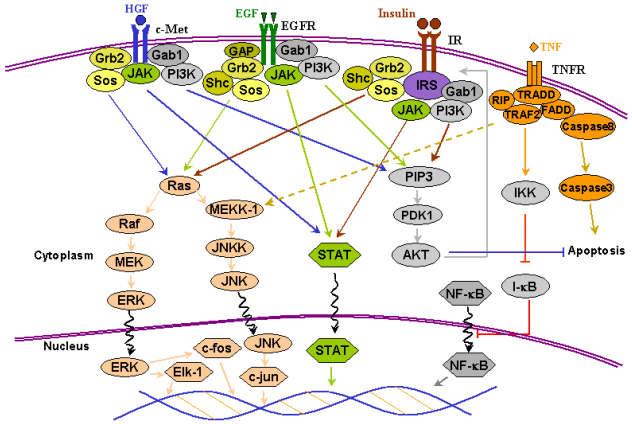
Functional Annotation

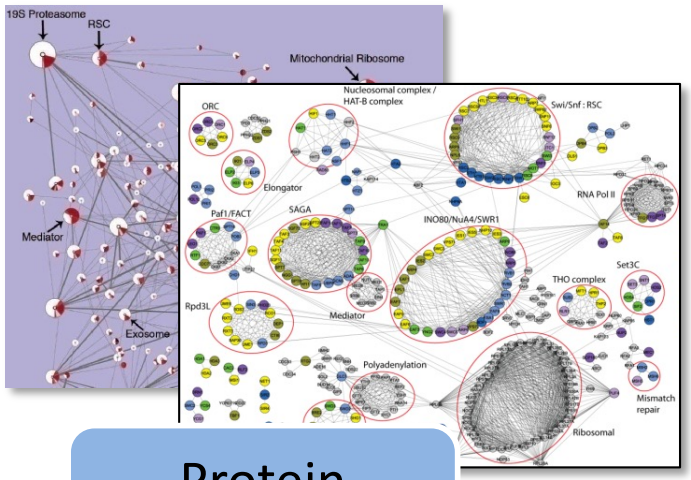
Beyond focusing on broad functional categories, we can also start to undertake systems based analyses

Genes and proteins do not operate in isolation but form parts of interconnected functional modules



Signalling networks





Protein complexes

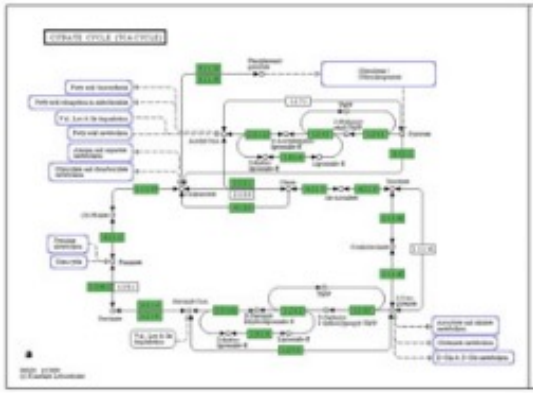
Metabolic pathways

By placing a bacterial transcripts within these functional contexts, we can understand how the microbiome functions at a molecular level

Functional Annotation – metabolic reconstructions

MG-RAST
metagenomics analysis server

MG-RAST uses the SEED framework, an alternative set of functional annotations; allows visualization of metabolic pathways for example



OLD!

NEW!

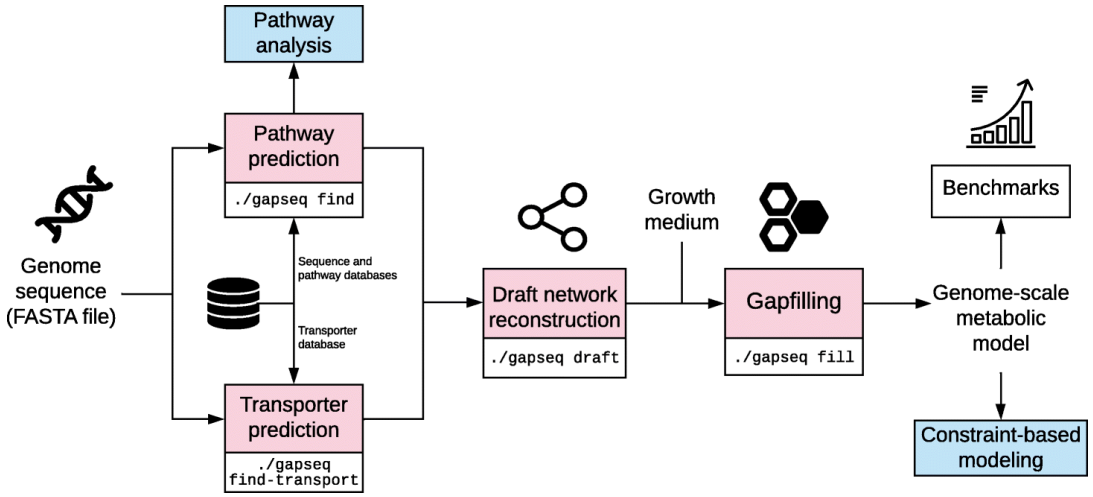
Zimmermann *et al. Genome Biology* (2021) 22:81
https://doi.org/10.1186/s13059-021-02295-1

Genome Biology

SOFTWARE Open Access

gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models

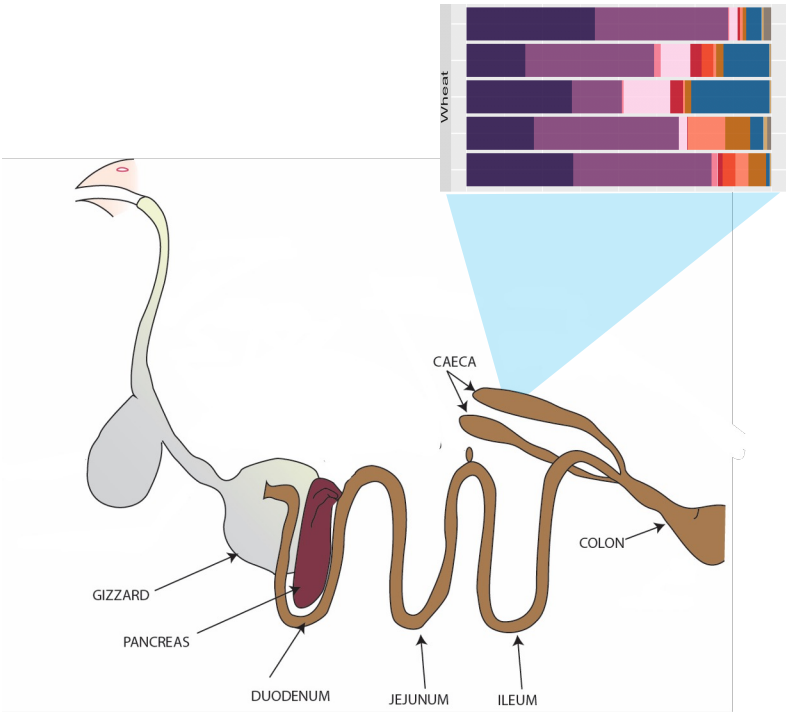
Johannes Zimmermann[1], Christoph Kaleta[1] and Silvio Waschina[1,2*]

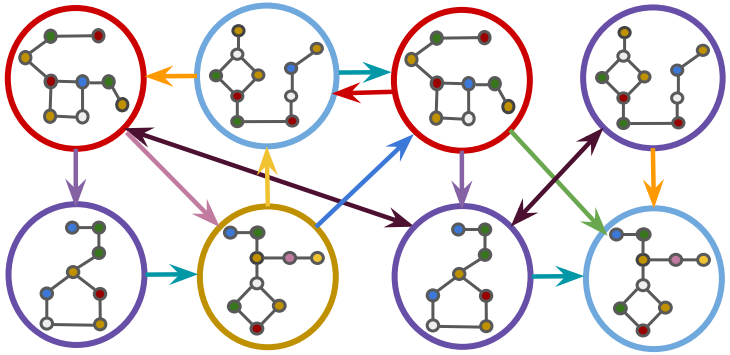


Gapseq identifies enzymes and builds complete metabolic reconstructions – can be used for metabolic modeling!

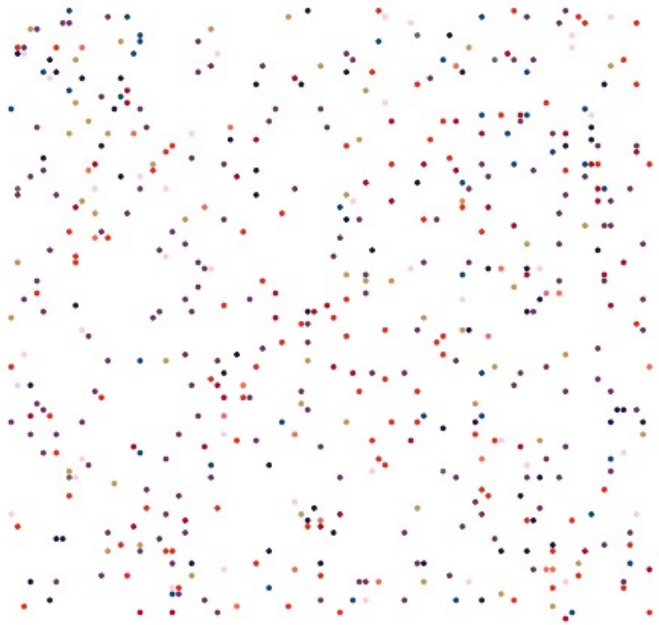
Functional Annotation – metabolic modeling

Metabolic modeling tools offer routes to understanding microbial community dynamics and predict production of key metabolites



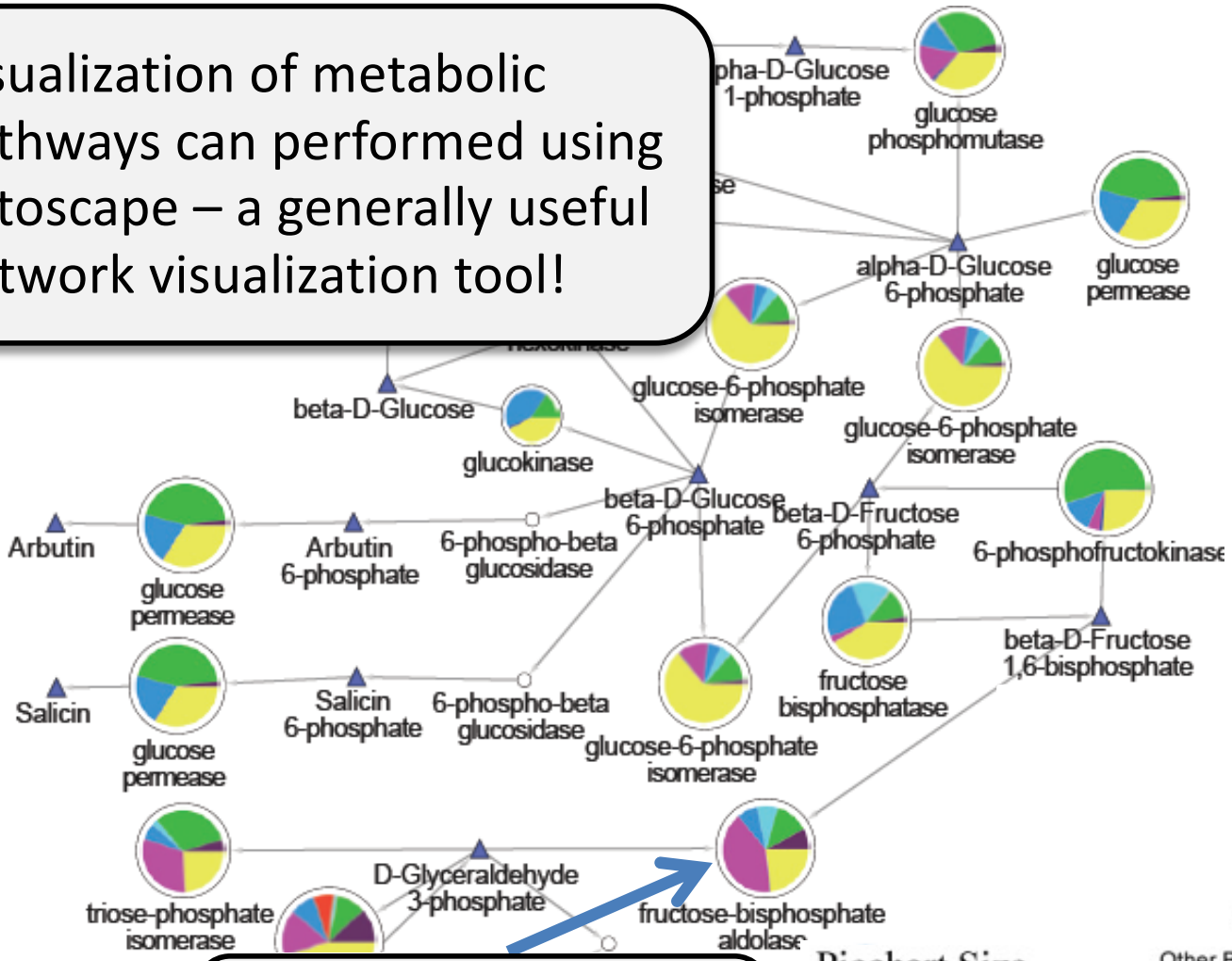


Population after 1 h

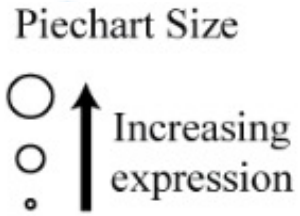


Visualizing results

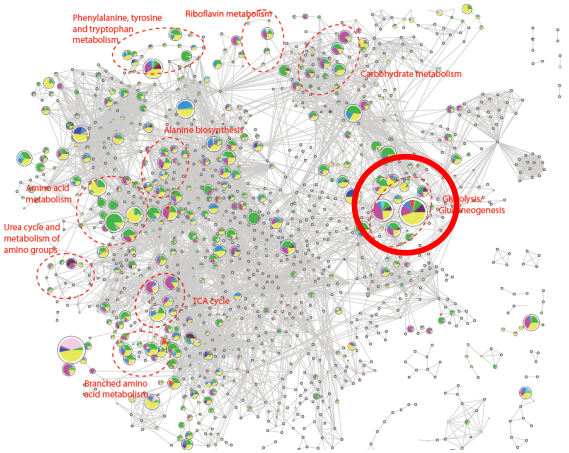
Visualization of metabolic pathways can performed using Cytoscape – a generally useful network visualization tool!



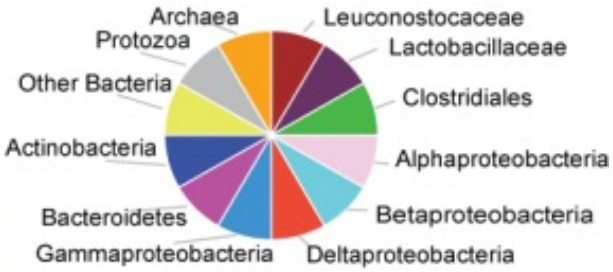
The interconversion of fructose-1,6 bisphosphate to glyceraldehyde 3-phosphate is largely mediated by bacteroides



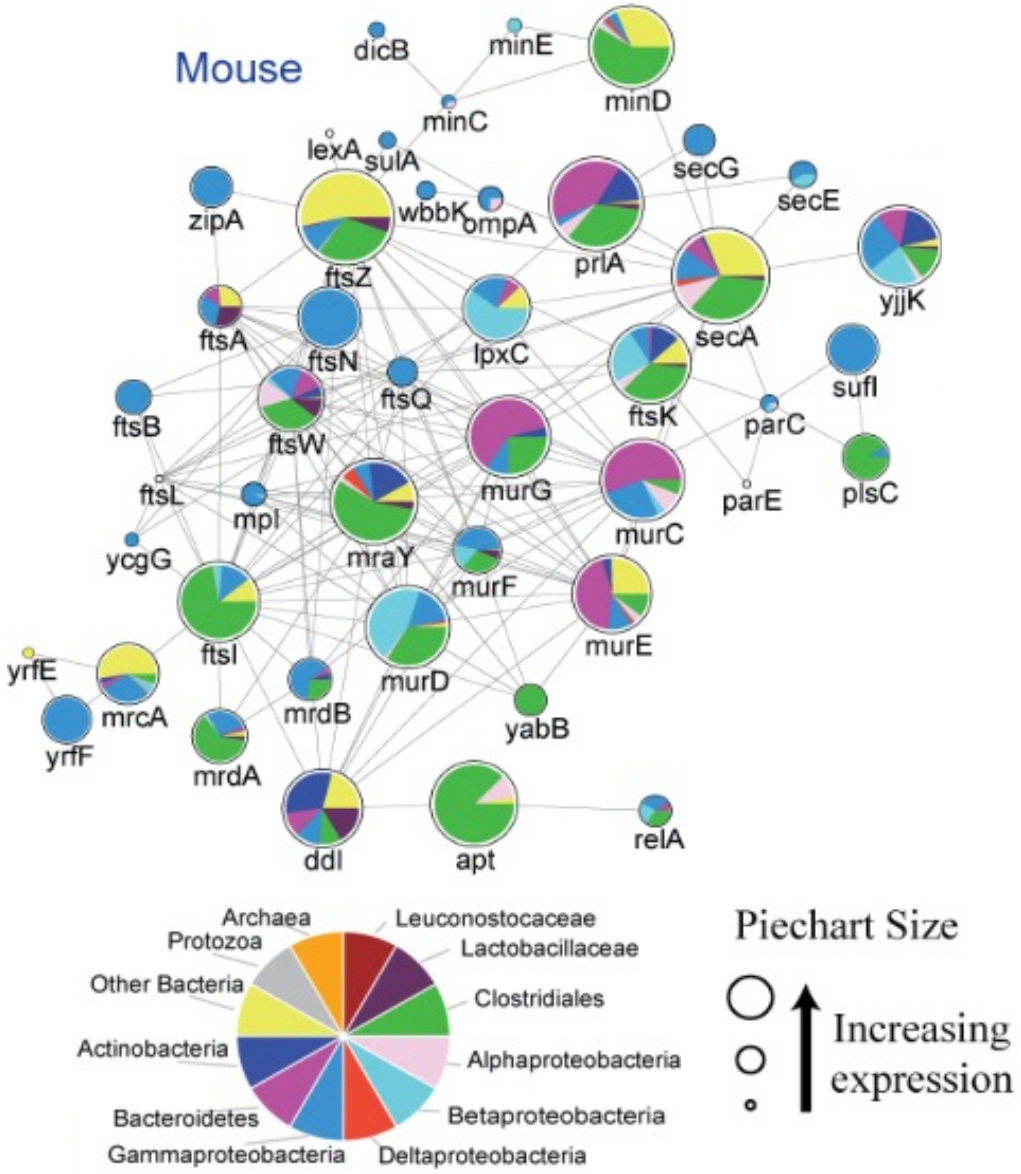
Metabolic network of a mouse gut microbiome



Piecharts indicate relative contribution of each taxon to an enzymatic activity



Functional annotation – Cell wall biogenesis

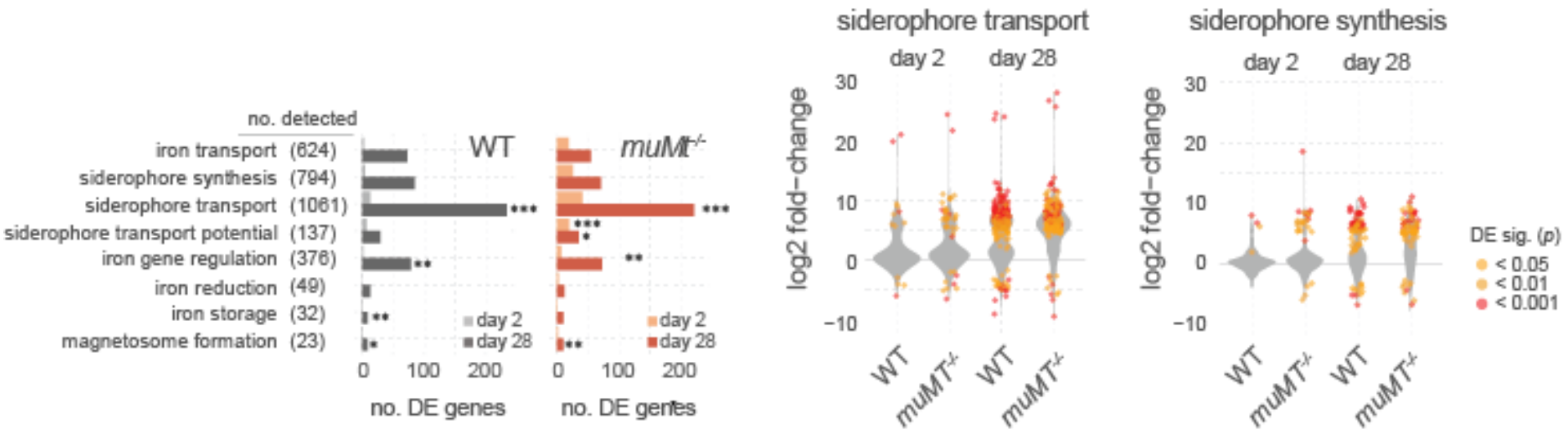


Beyond metabolism, other 'systems'-type datasets can be leveraged to help interpret datasets

Here we map metatranscriptomic reads onto a protein interaction network generated for genes involved in cell wall biogenesis in *E. coli*

However *E. coli* does not represent all bacteria – do we need maps for e.g. Gram +ve's or (better) a generic map of 'bacterial systems'?

Functional annotation – Iron uptake and storage



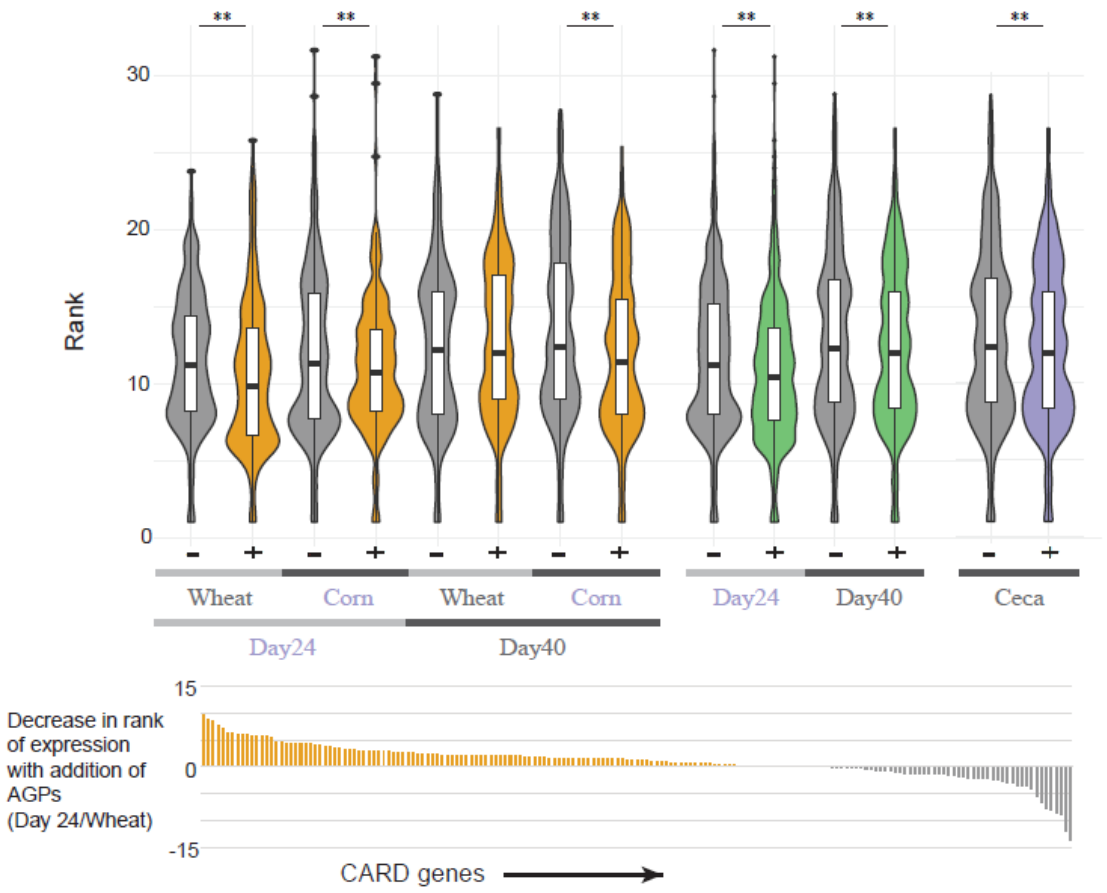
Previously we showed that iron supplements can promote colonization by parasites and bacterial dysbiosis in the gut

Applying metatranscriptomics to a mouse infection model, we used FeGenie to identify genes associated with iron uptake / storage systems

Bacteria in the gut of mice infected with the protozoan parasite – *Tritrichomonas musculis* – increased expression of iron uptake / iron storage systems

Is upregulation an attempt to maintain status quo? or an attempt by bacteria to compromise the parasite?

Functional annotation – AMR



In our chicken studies, we were interested in the impact of antibiotic growth promotants (AGPs) on AMR gene expression

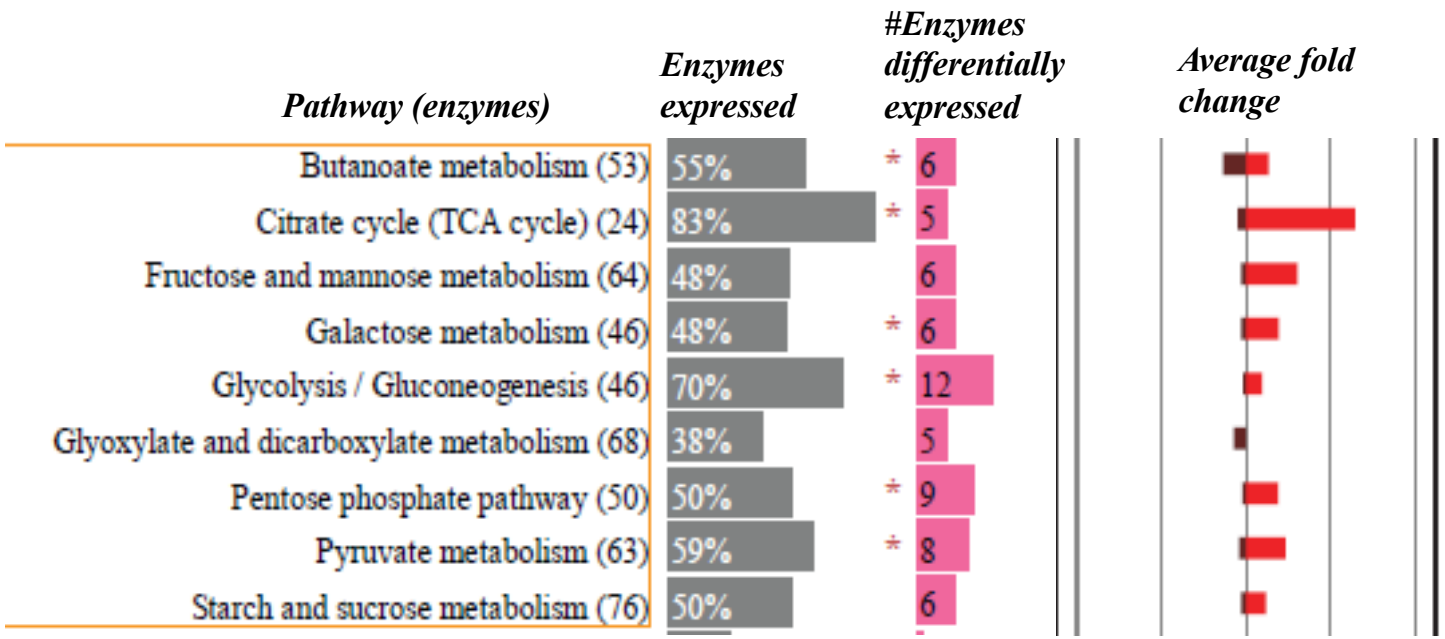
Using the CARD database (McMaster) we identified samples exposed to AGPs exhibited a significant upregulation in AMR gene expression

Metatranscriptomic datasets are rich in information that we are only scratching the surface of

Quorum sensing
Biofilm machinery
Secretion systems
….

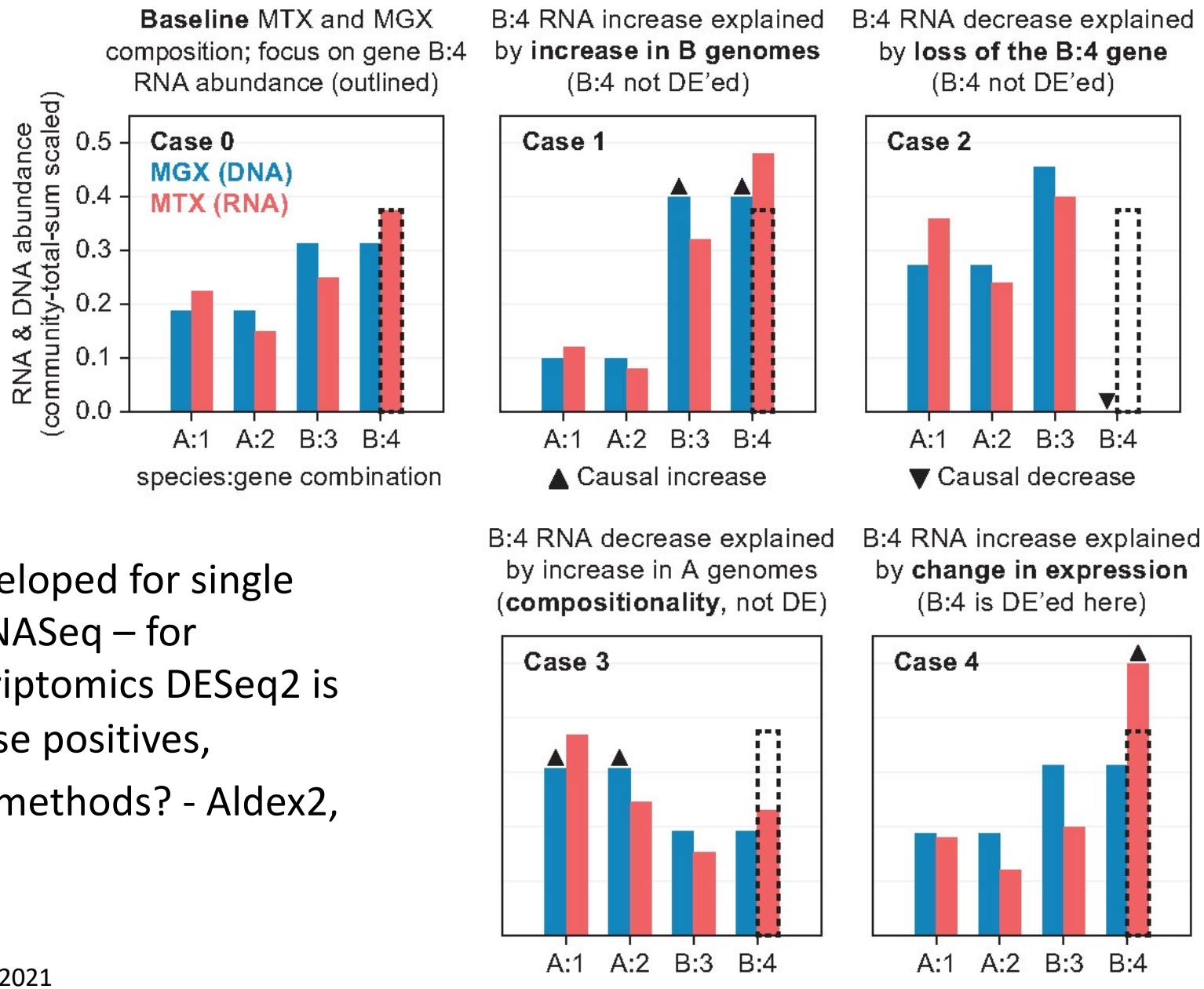
Analysing the data – differentially expressed genes

Established RNASeq Tools such as DESeq2 and EdgeR provide platforms to identify differentially expressed genes for subsequent gene set enrichment analyses



Hypergeometric tests of differentially expressed genes associated with KEGG pathways can identify metabolic pathways exhbiting differential expression (here comparing Plin2-KO and WT mice fed a high fat diet)

Normalizing for taxon/gene abundance



- DESeq2 developed for single organism RNASeq – for metatranscriptomics DESeq2 is prone to false positives,
- Alternative methods? - Aldex2, Ancom

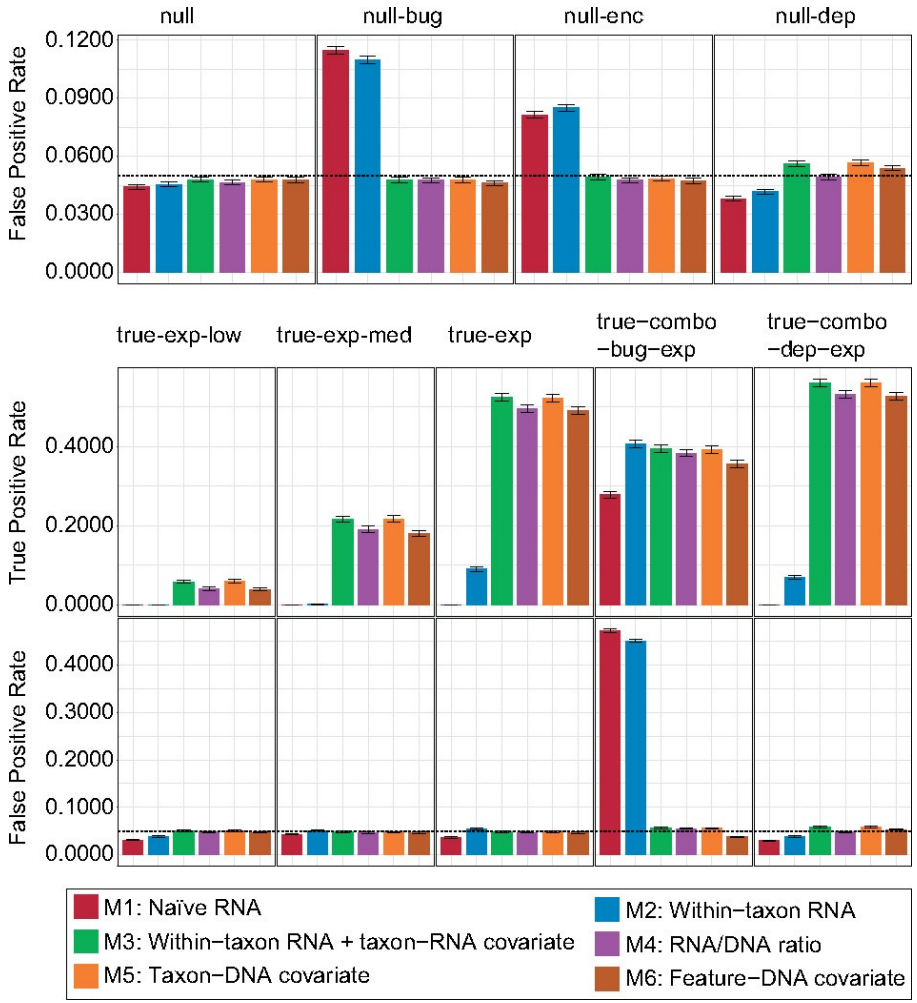
Zhang et al Bioinformatics 2021

Normalizing for taxon/gene abundance

‘Taxon-specific scaling’ normalization approximately transforms a MTX dataset into an aggregate of single-organism RNA-seq datasets

When samples are profiled with paired MTX and MGX sequencing, a gene’s RNA abundance can be normalized by its DNA abundance

Zhang et al recently performed a systematic analysis of six different linear models showing enhanced performance of abundance normalization relative to naïve RNA



Zhang et al Bioinformatics 2021

Analysing the data

PCA plots can be applied to identify clustering of samples

PERMANOVA identifies statistical differences across samples (here mice fed different diets with different genotypes have similar taxonomic distributions; mice fed a high fat diet exhibit genotype driven differences in enzyme expression)