



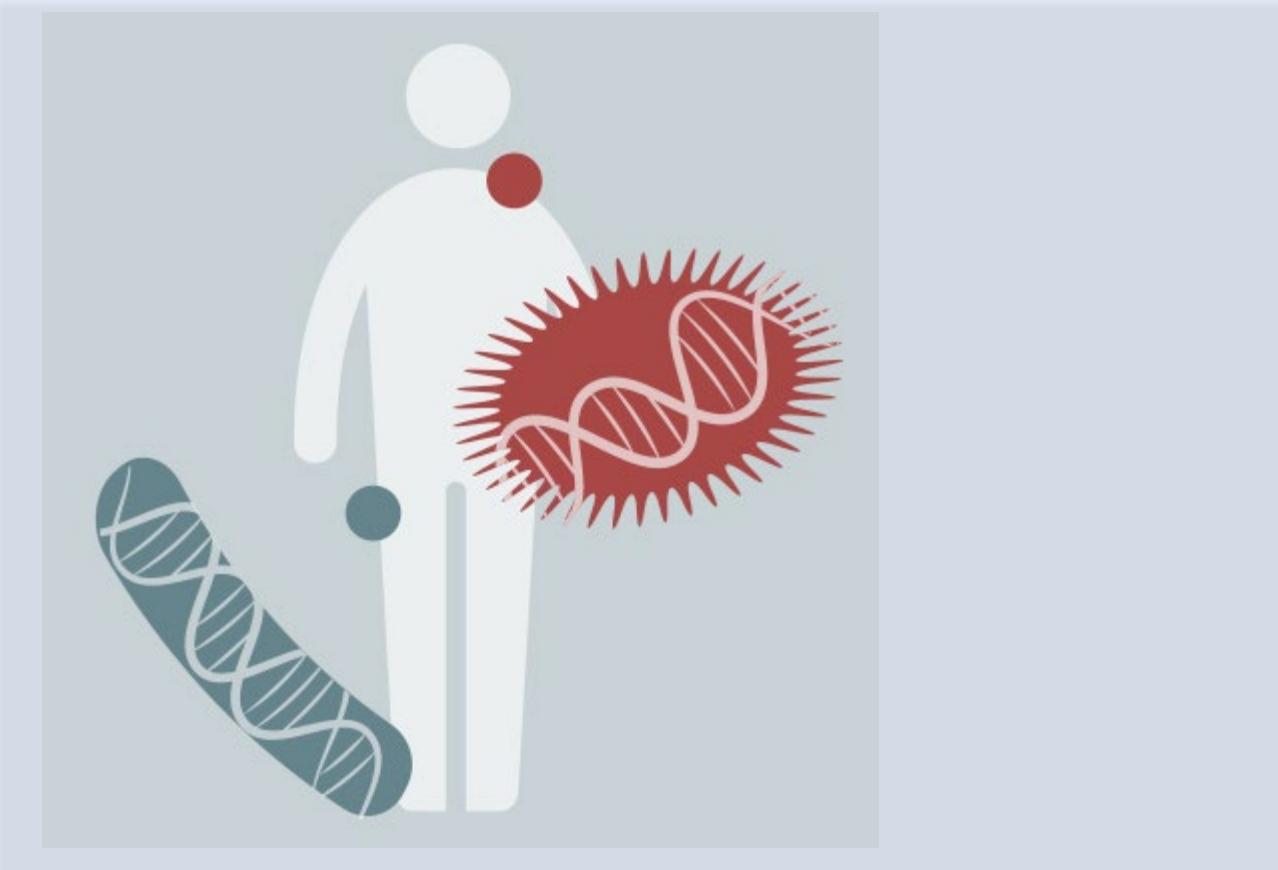
# Canadian Bioinformatics Workshops

[www.bioinformatics.ca](http://www.bioinformatics.ca)  
[bioinformaticsdotca.github.io](https://bioinformaticsdotca.github.io)

# Module 6

## Metatranscriptomics

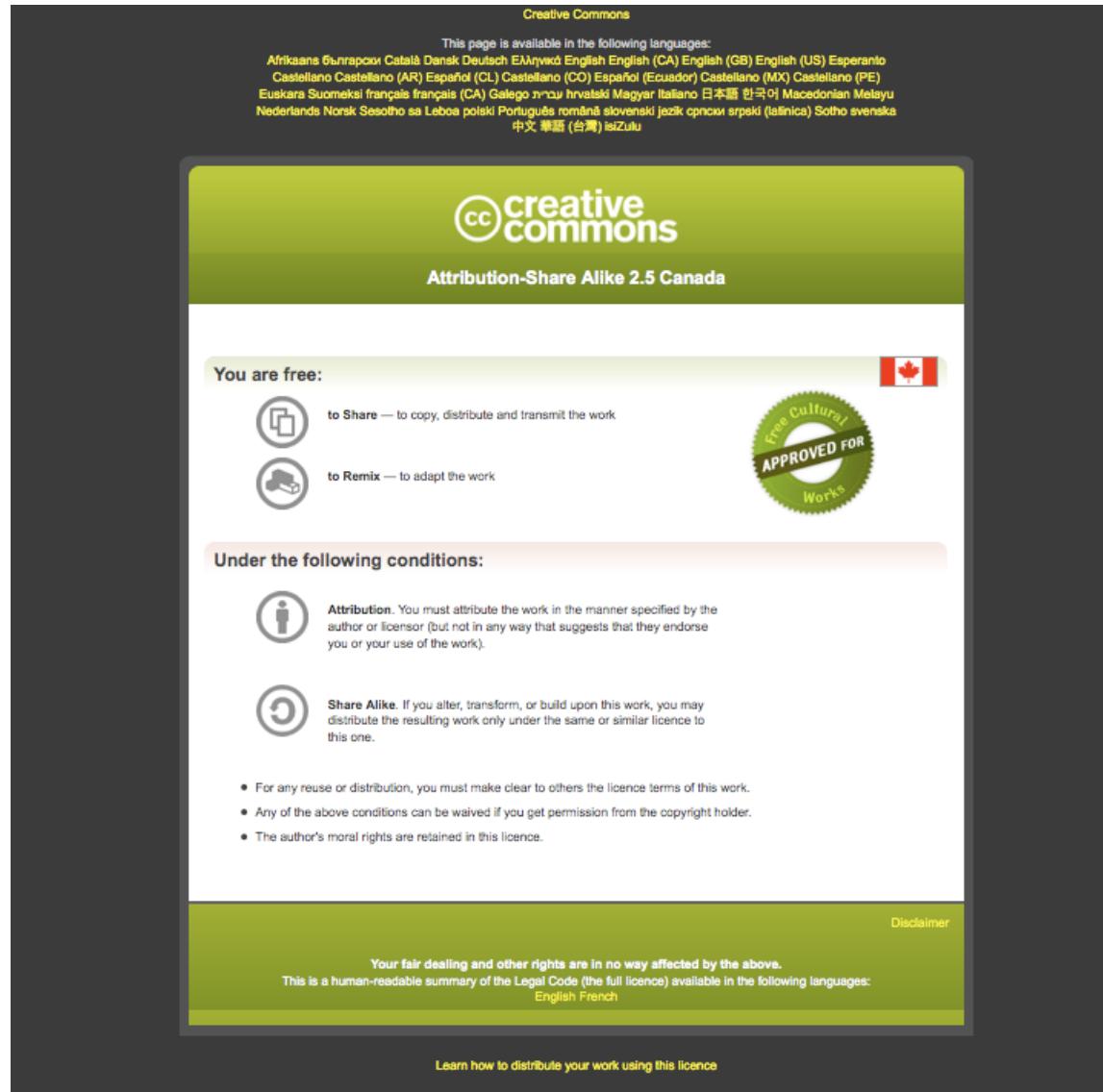
John Parkinson  
Microbiome Analysis  
December 6<sup>th</sup>-7<sup>th</sup>, 2022



Parkinson Lab

**SickKids**  
THE HOSPITAL FOR  
SICK CHILDREN





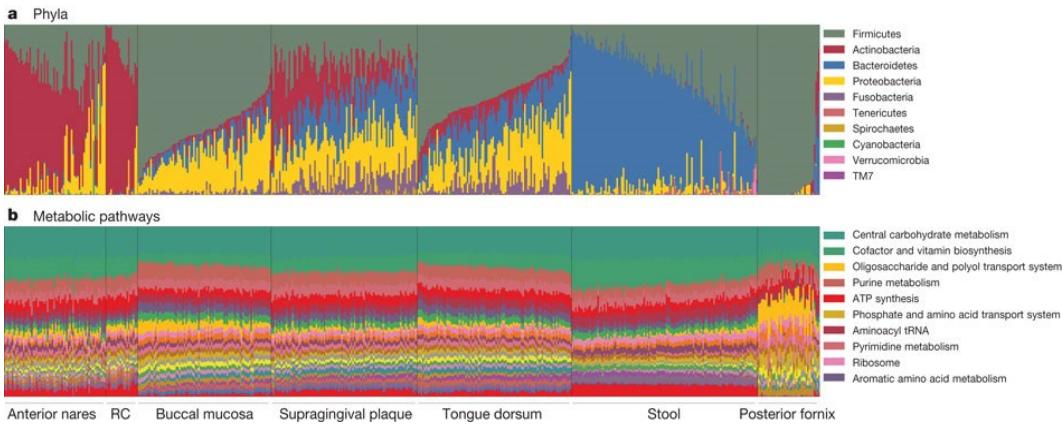
# Learning Objectives of Module

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

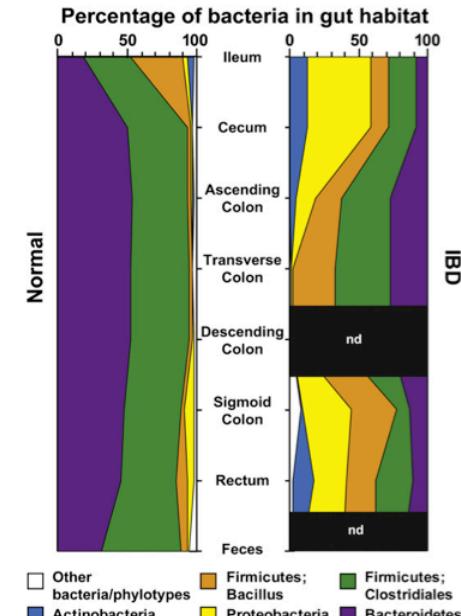
- Understanding the capabilities of metatranscriptomics
- Gaining an appreciation of sample collection and experimental design
- Learning important steps in data processing
- Processing a simple metatranscriptomic dataset

# 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

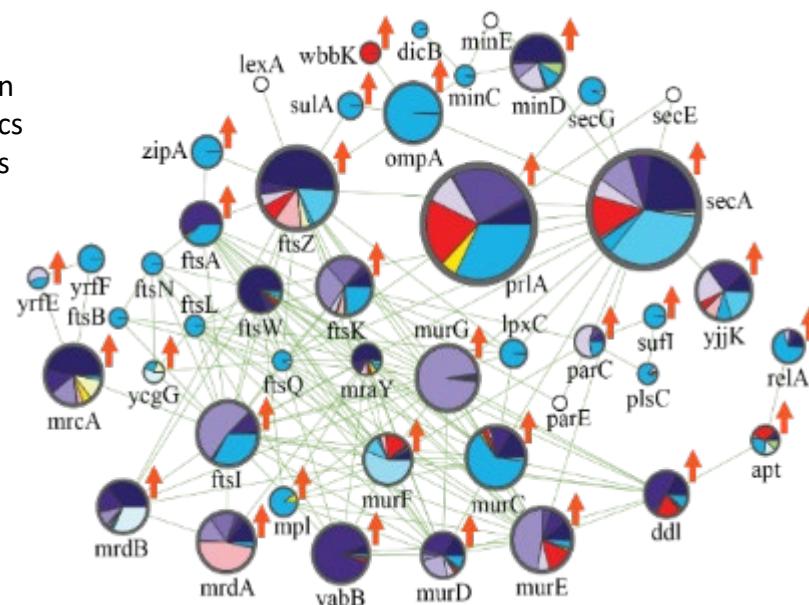


# Metatranscriptomics focuses on community activity

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

Zou et al Microbiome 2022

# Metatranscriptomics applied to a model of obesity

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

RESEARCH

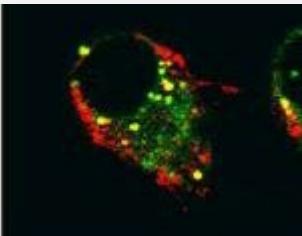
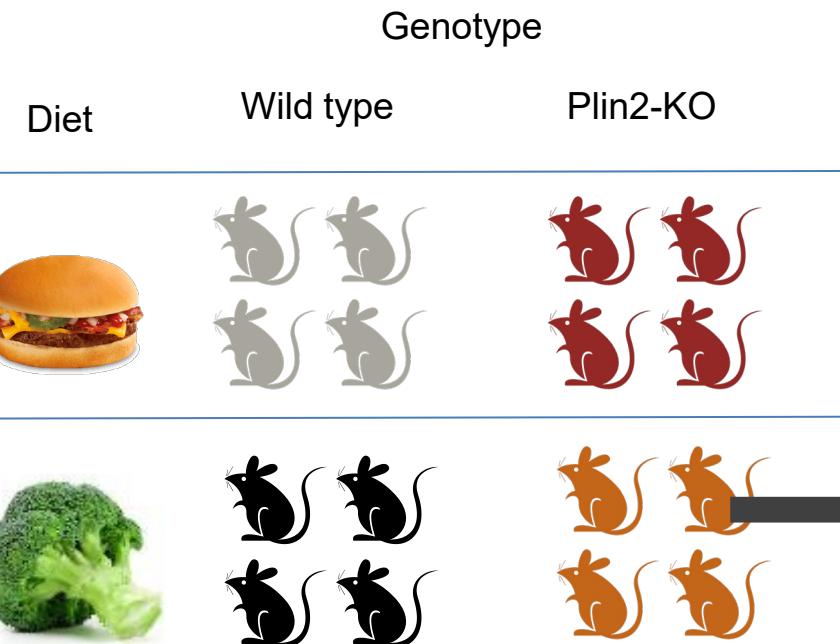
Microbiome

Open Access



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

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



Najt et al  
Biochemistry  
2014

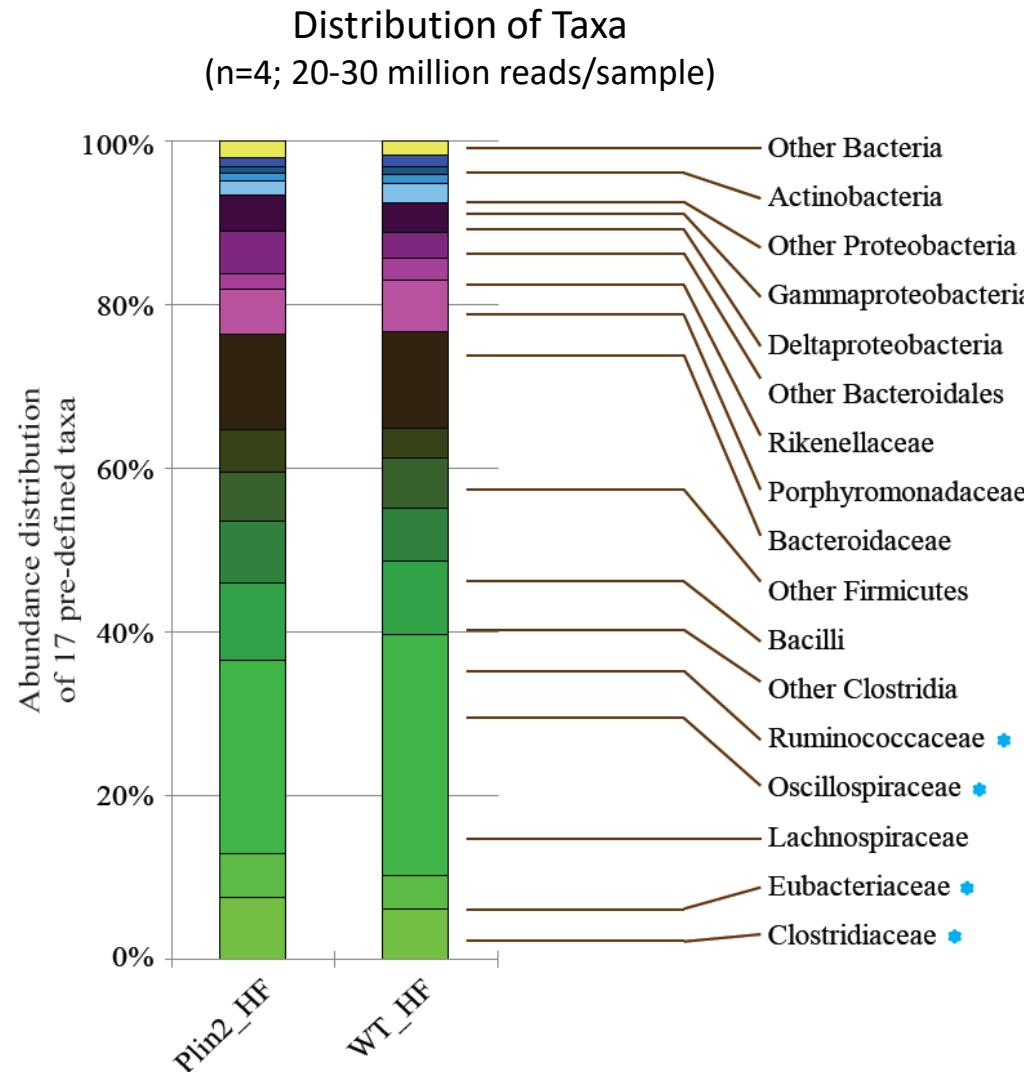
Perilipin2 (Plin2) interacts with lipid droplets and helps regulate lipid uptake

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

What impact does Plin2-KO have on microbiome function?

RNA-Seq  
20-30 million sequence reads/mouse

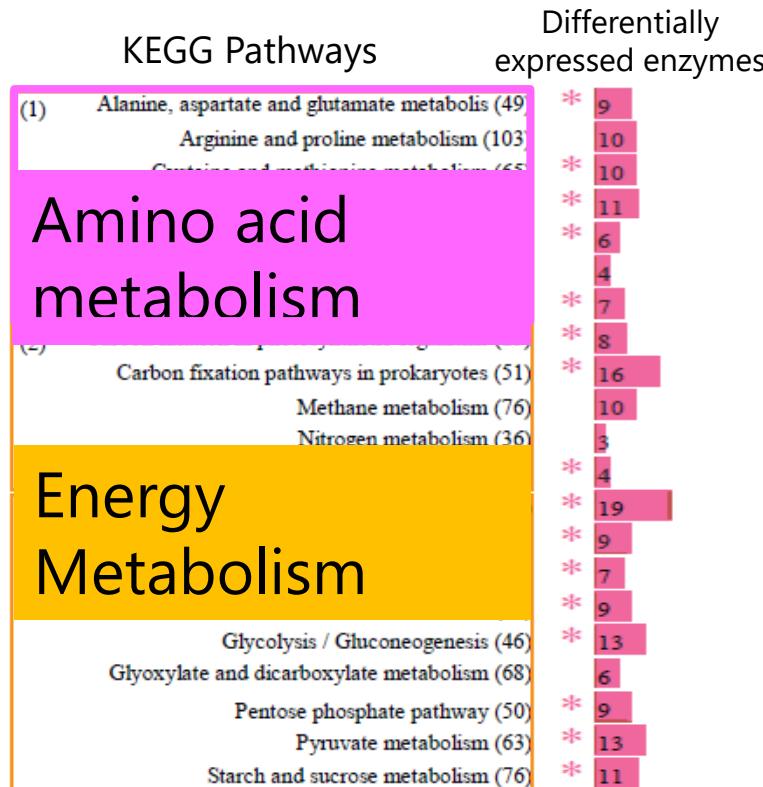
# Under a high fat diet metatranscriptomics reveals no differences in microbiome composition



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

# 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!

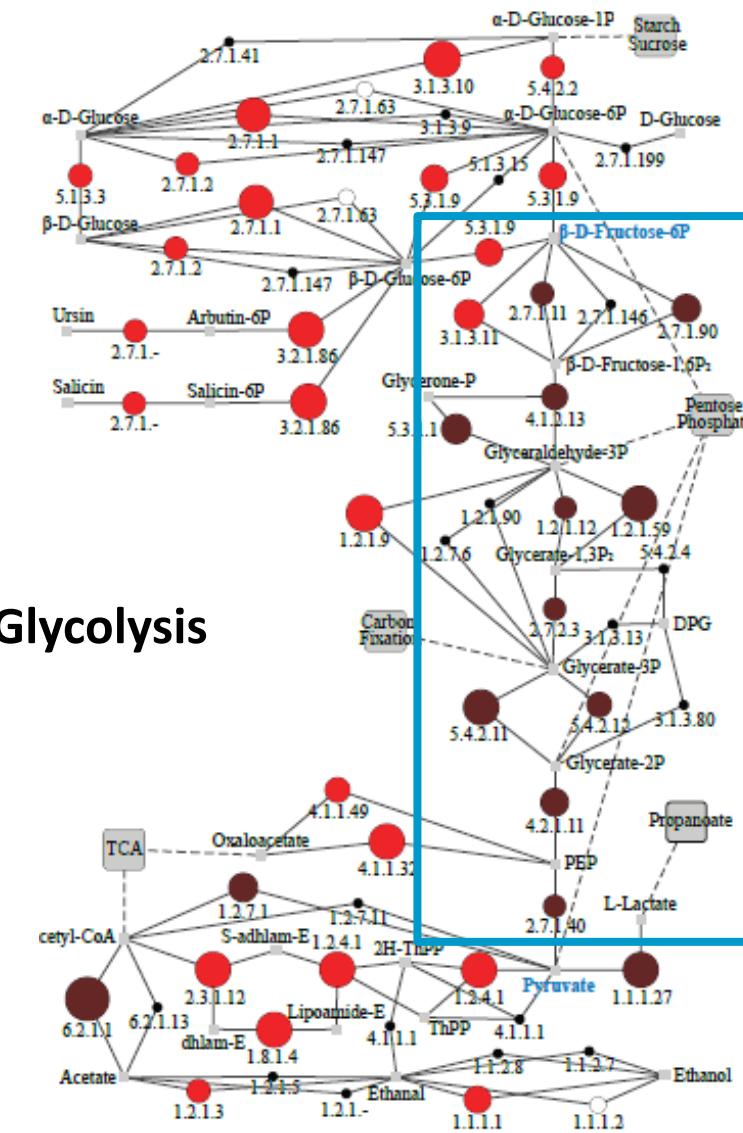
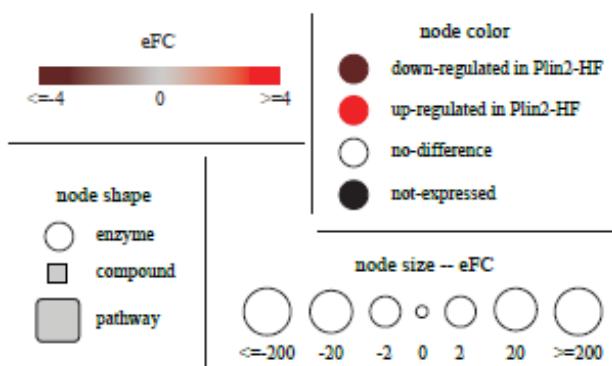


Many of these differentially expressed genes are associated with amino acid metabolism and energy metabolism

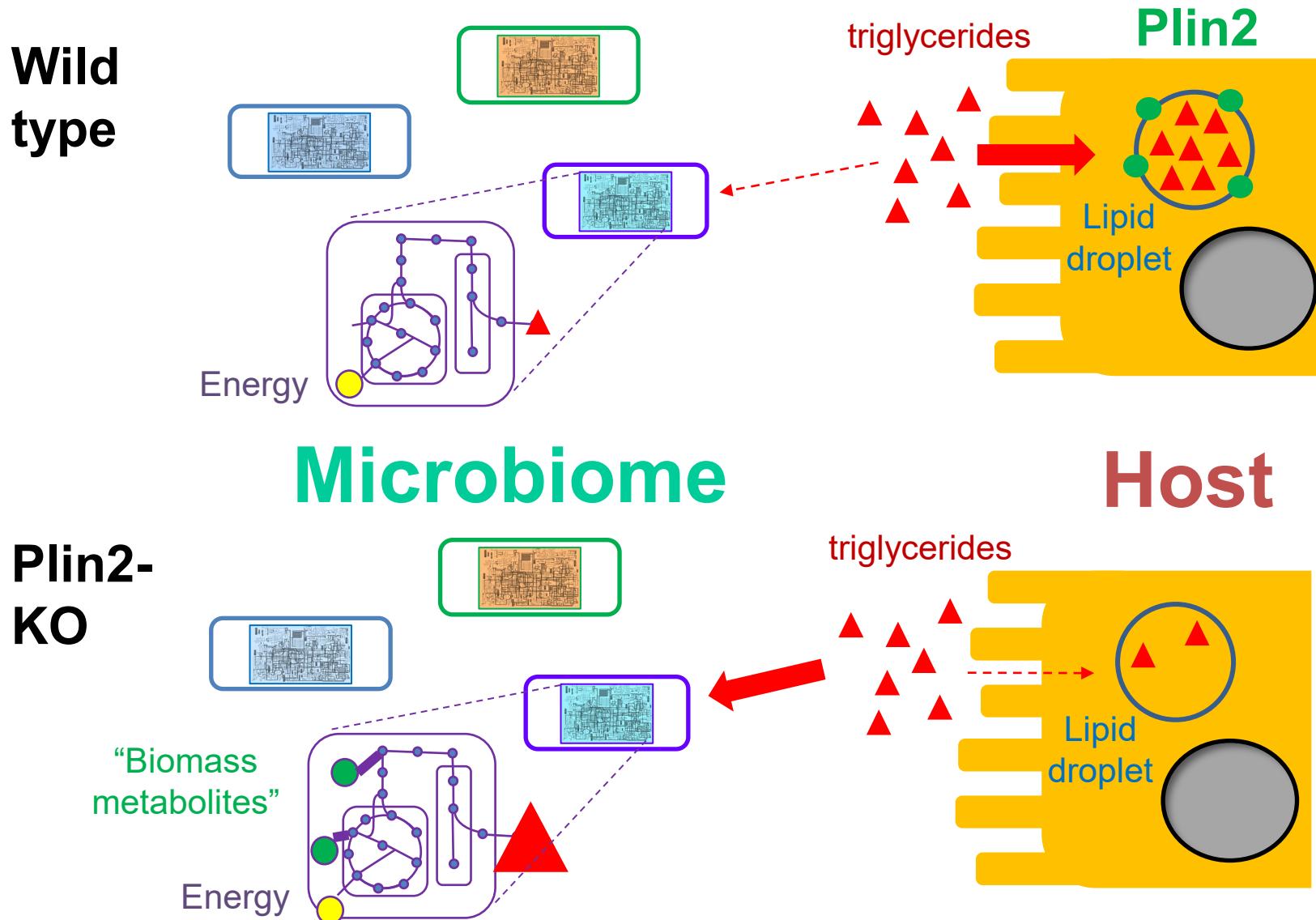
Xiong et al Microbiome 2017

# 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

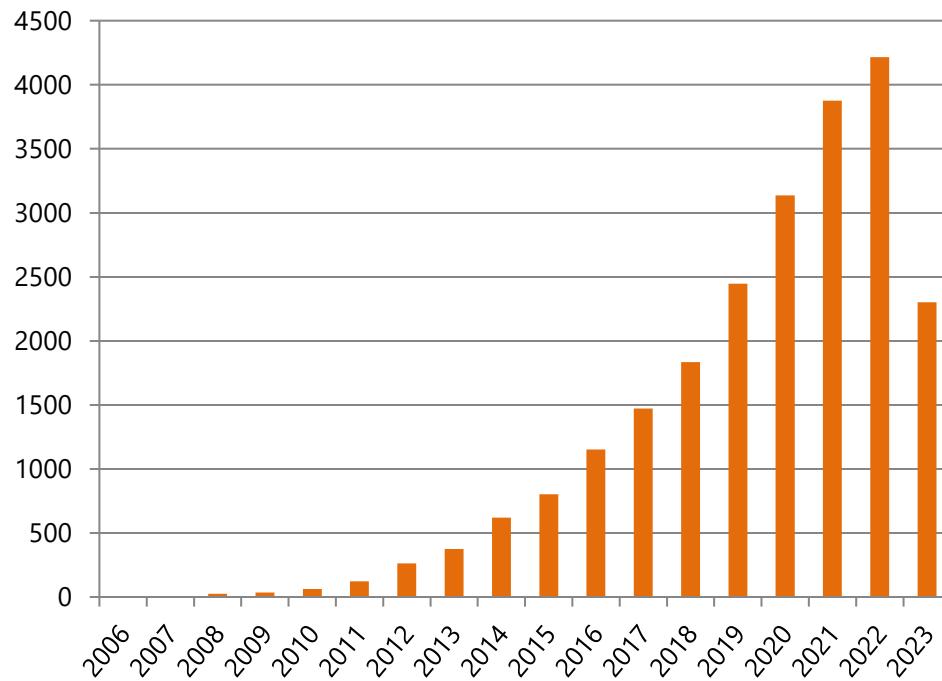


# Are energy producing pathways down-regulated due to accumulation of triglycerides under a high fat diet?

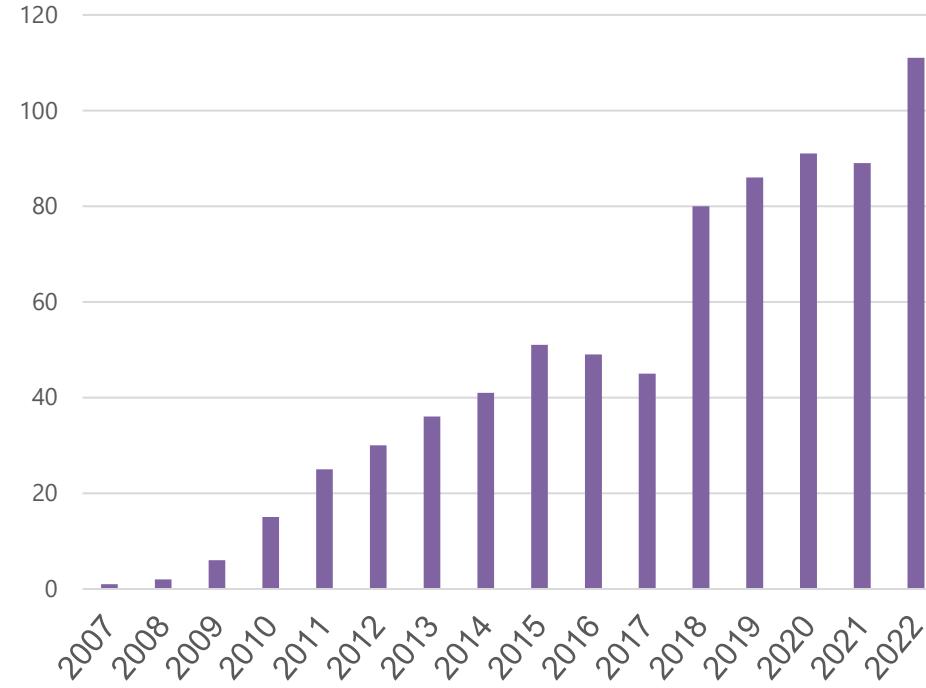


# Metatranscriptomics: starting to catch on?

**Number of publications  
with 'Microbiome' in Title**



**Number of publications with  
'Metatranscriptom\*' in Title**



# Other Examples of Applications of Metatranscriptomics



ARTICLE OPEN

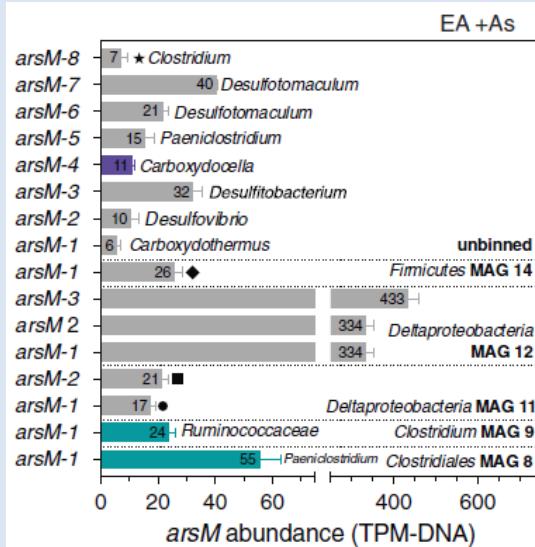
Meta-omics-aided isolation of an elusive anaerobic arsenic-methylating soil bacterium

Karen Vlachava<sup>1,2</sup>, Jiangtao Qiao<sup>1</sup>, Andrew Janowczyk<sup>3</sup>, Suresh Poudel<sup>4</sup>, Nicolas Jacquemin<sup>1,5</sup>, Karin Lederballe Meibom<sup>1</sup>, Him K. Shrestha<sup>1,6</sup>, Matthew C. Reid<sup>1,7</sup>, Robert L. Hettich<sup>1,8</sup> and Rizlan Bernier-Latmani<sup>1,9</sup>  
© The Author(s) 2022

[www.nature.com/ismej/](https://www.nature.com/ismej/)

[Check for updates](#)

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



nature  
microbiology

ARTICLES

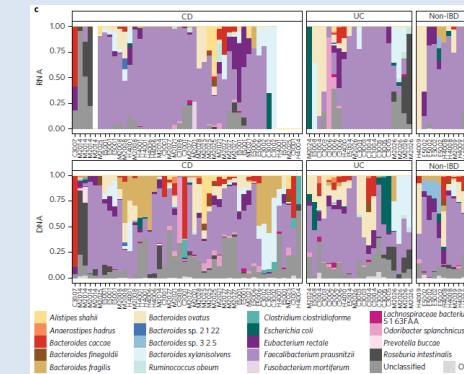
<https://doi.org/10.1038/s41564-017-0089-z>

Dynamics of metatranscription in the inflammatory bowel disease gut microbiome

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

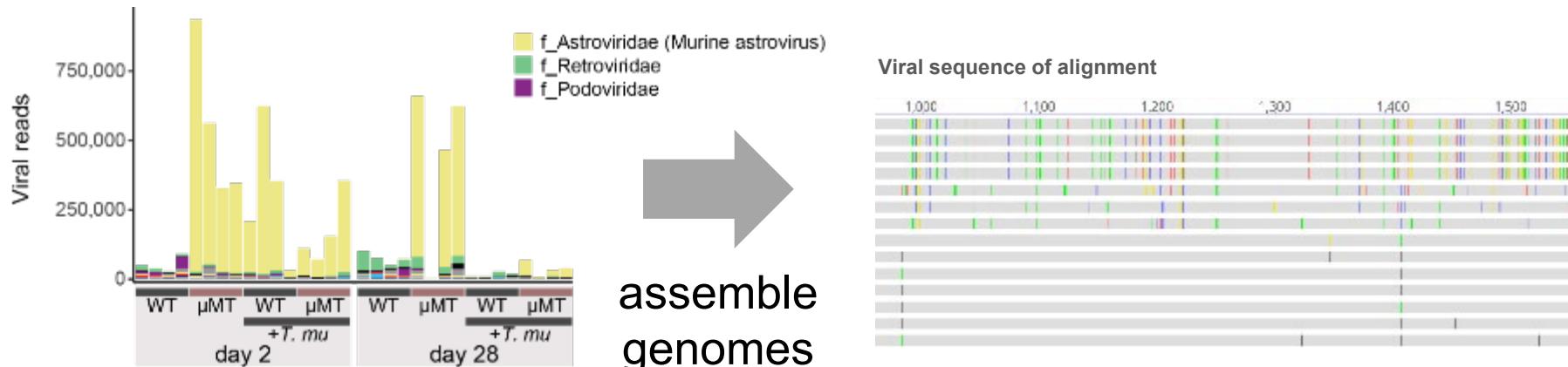
Applying metatranscriptomics to IBD samples found:

- 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



Metatranscriptomic datasets are being exploited to expand the universe of RNA viruses



RESEARCH ARTICLE  
May/June 2023 Volume 8 Issue 3 e01002-22  
<https://doi.org/10.1128/mSystems.01002-22>

Metatranscriptomic analysis uncovers prevalent viral ORFs compatible with mitochondrial translation

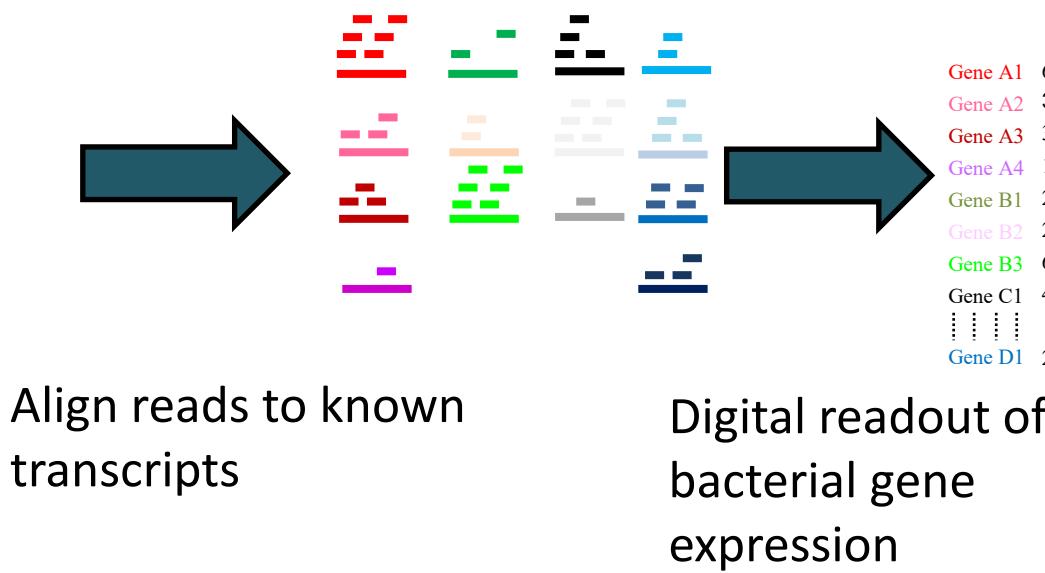
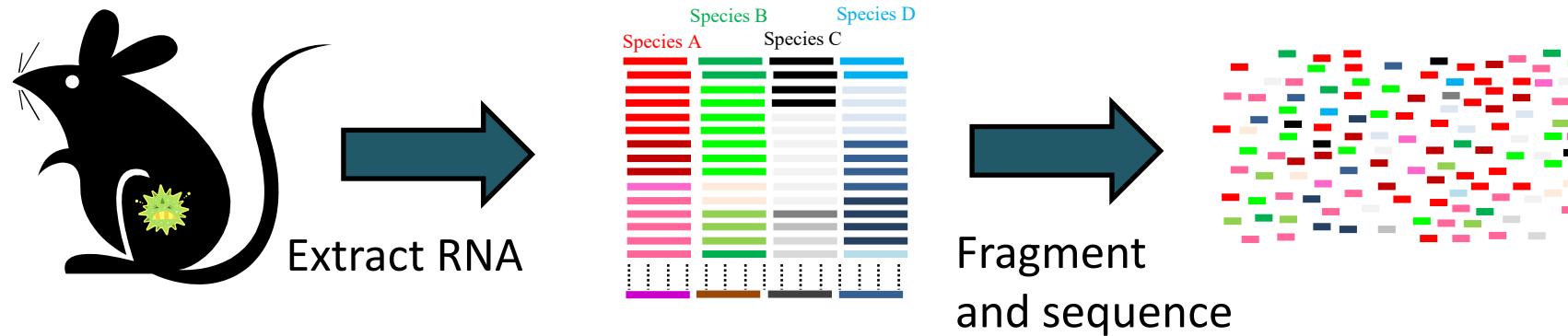
Adam Begeman<sup>1</sup>, Artem Babaian<sup>2,3</sup>, Samantha C. Lewis<sup>1</sup>

<sup>1</sup> Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

<sup>2</sup> Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

<sup>3</sup> Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada

# Metatranscriptomics through RNA-Seq



Metatranscriptomics is similar to single organism RNA-Seq but requires 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 from hundreds/thousands of taxa
  - depth of sequencing
  - lack of reference sequences

# 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



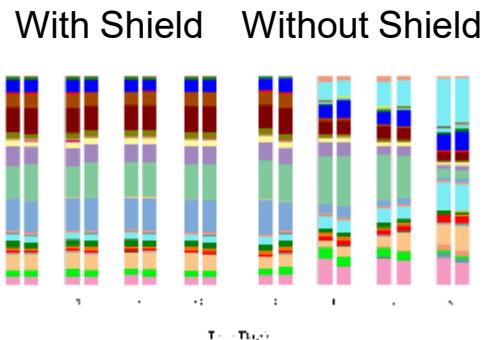
DNAGenoTek

OMNIgene GUT  
Released June 2022

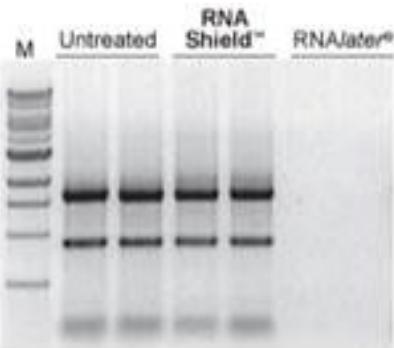


# 1. Sample collection and RNA extraction

Zymoresearch

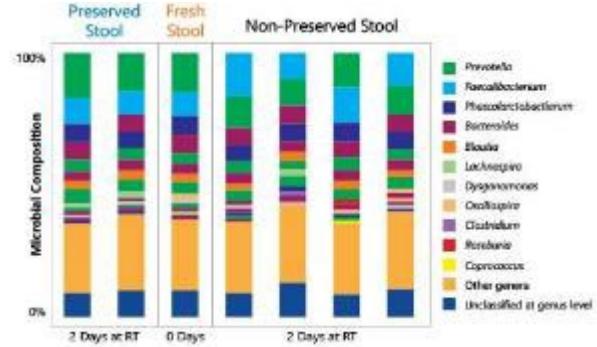


## Direct RNA Purification

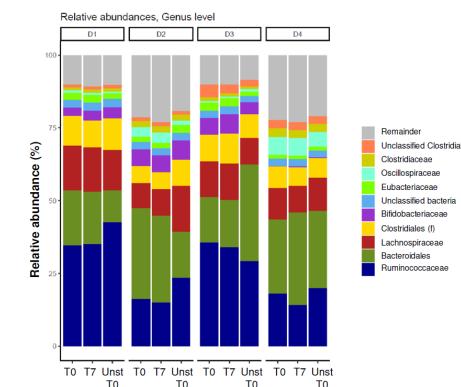


**RNA can be purified directly from RNA Shield™ without reagent removal.** Cellular RNA was extracted from samples stabilized in RNA Shield™ with TRIzol® and purified with the Direct-zol™ RNA MiniPrep. Conversely, RNaLater® did not facilitate direct purification.

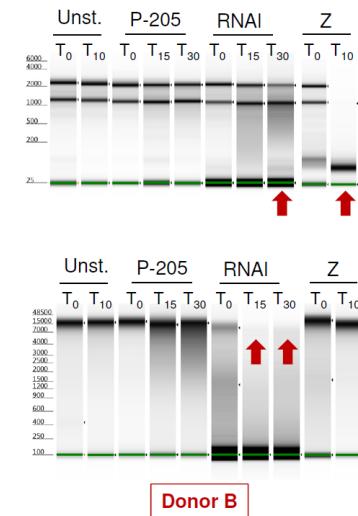
# Norgen



DNA GenoTek



The figure consists of two side-by-side gel electrophoresis images. The left panel, labeled "27 months", shows a 1.2% 1x TAE agarose gel with lanes labeled M, A, B, C, D, and M. Lanes A through D show distinct bands corresponding to different treatment groups. The right panel, labeled "7 days", shows a 1.2% 1x MOPS agarose gel with lanes labeled A, B, C, and D. Lanes A, B, and C show prominent bands, while lane D is mostly clear.



# 1. Sample collection and RNA extraction

Metatranscriptomics is expensive mainly due to library preparation



Cost per sample  
(60-80 million reads)

~\$300-\$400

How many replicates?



“At least 4!” (depends on nature of samples)



Lack of reference sets make power analyses challenging

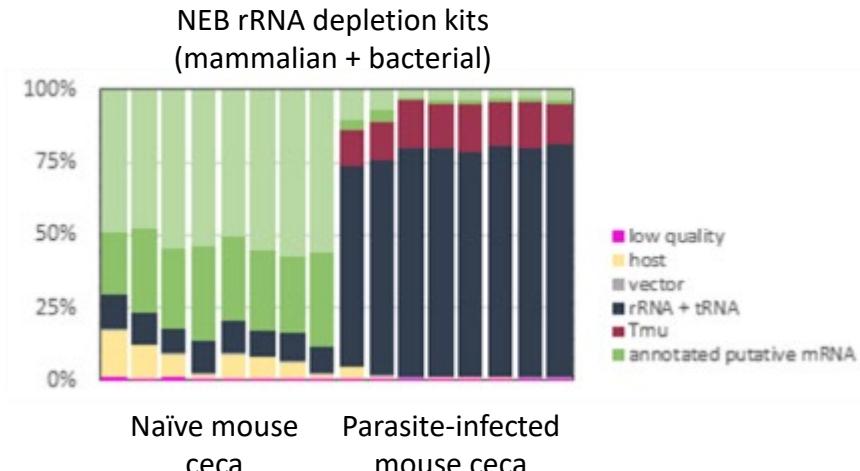
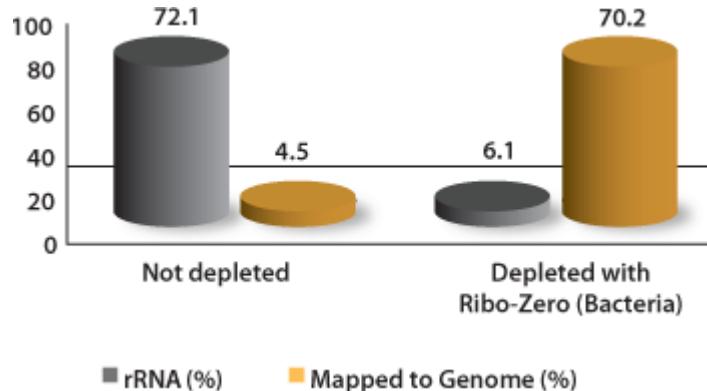


## 2. Preparing sample for sequencing

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 – aim for 500ng-2.5ug RNA/sample

Ribo-Zero (Illumina) provides reasonable success



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

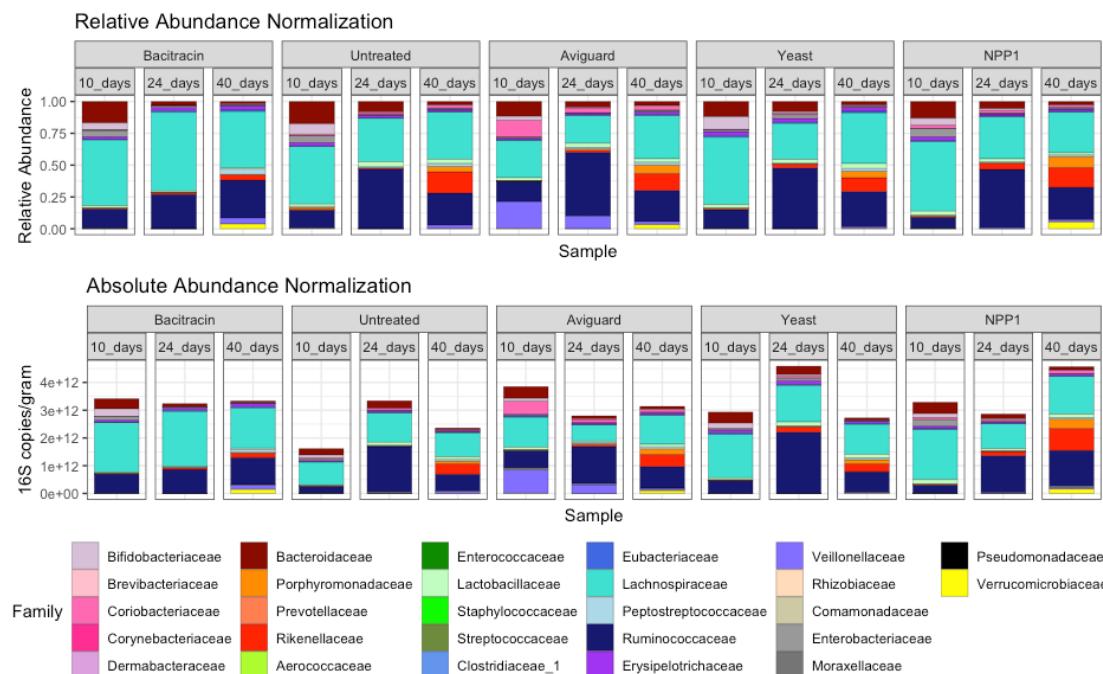
## 2. Spike in to quantify read abundance

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

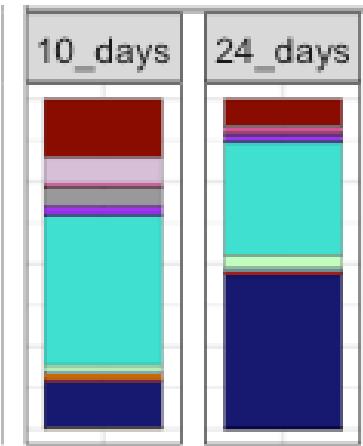
The spike-in consists of two species:  
*Imtechella halotolerans* and *Allobacillus halotolerans*, totaling  $4 \times 10^7$  cells.

Tested on Cecal, Ileal and Jejunum samples from Chickens



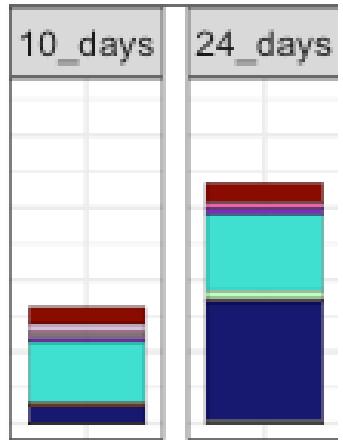
## 2. Absolute abundance analyses alter results

Relative abundance



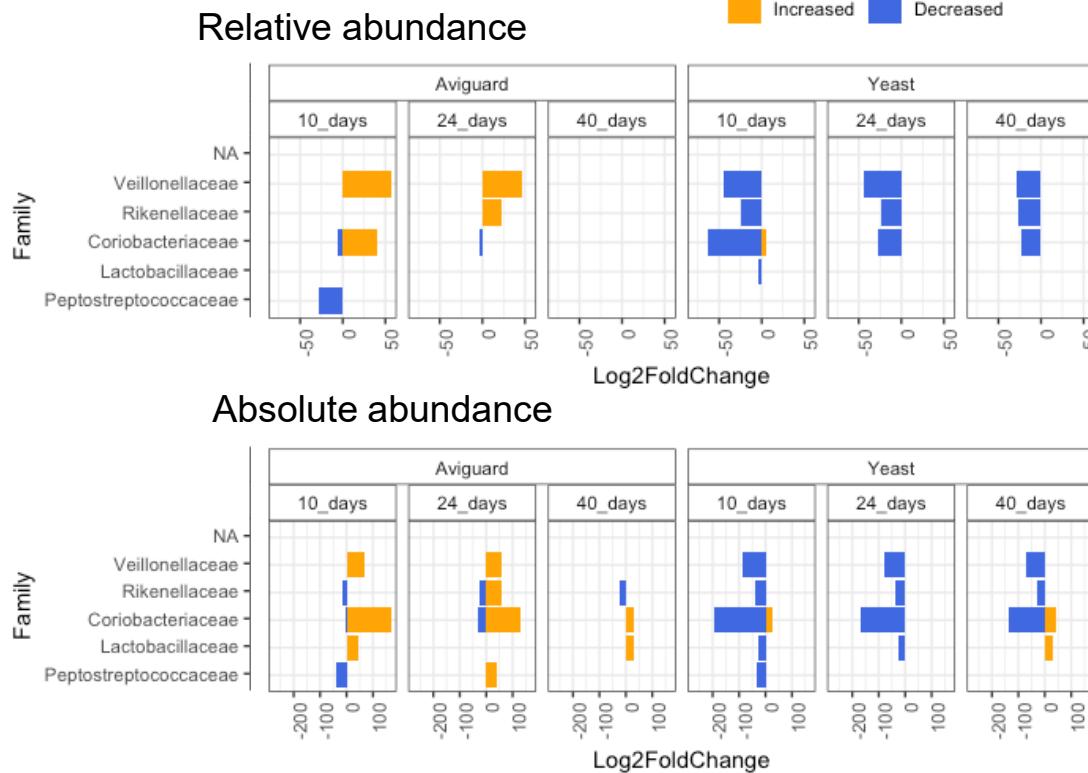
Decrease in  
Bacteroidaceae  
at 24 days post  
hatch?

Absolute abundance



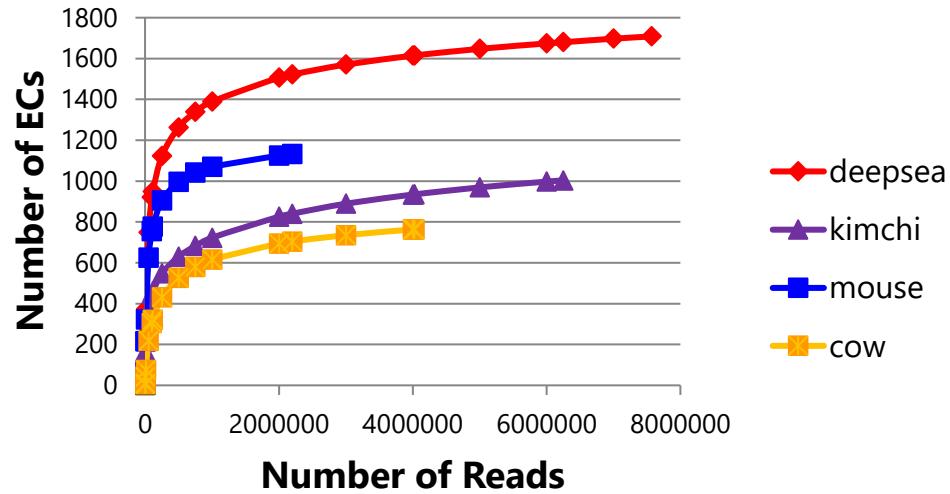
No change in  
Bacteroidaceae  
at 24 days post  
hatch

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



### 3. Generating reads

How many reads are “enough”?



~5 million mRNA reads provide 90-95% of enzymes (ECs) in a microbiome

Depending on complexity of sample may want to consider 40-80 million reads/sample



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

## 4. Analysing the data

Metatranscriptomics is a relatively new field requiring robust tools and pipelines to process and analyse  
Due to their size (billions of sequence reads) – compute clusters are key



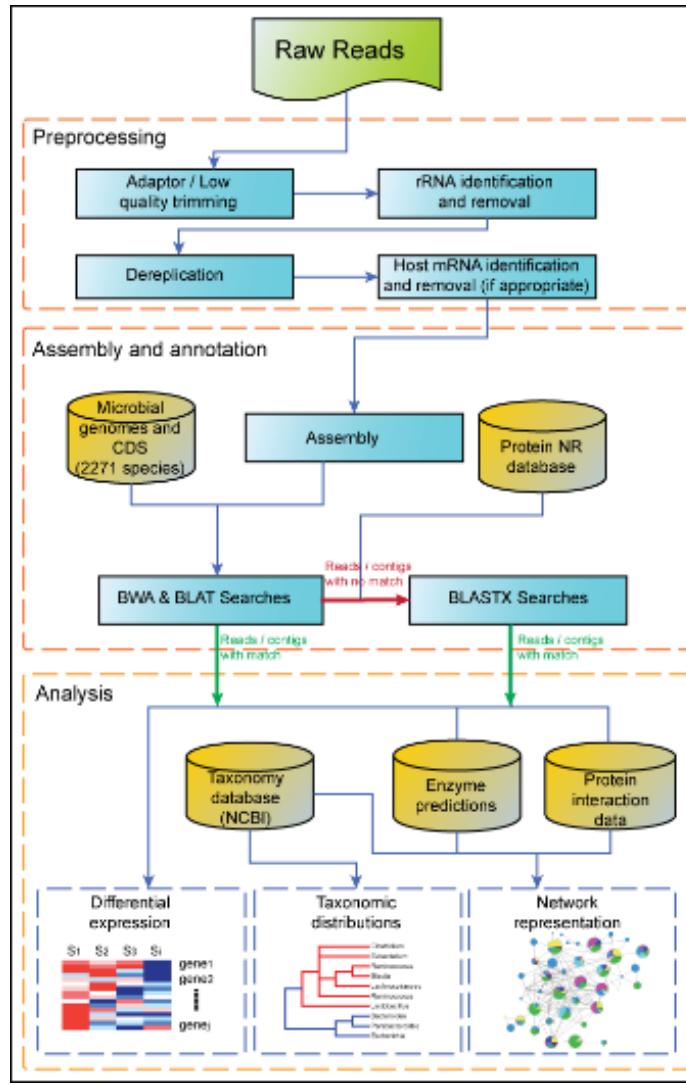
# Read processing - filtering

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

Taj et al Microbiome 2023



Taj et al. *Microbiome* (2023) 11:143  
<https://doi.org/10.1186/s40168-023-01562-6>

Microbiome

## SOFTWARE

## Open Access



MetaPro: a scalable and reproducible data processing and analysis pipeline for metatranscriptomic investigation of microbial communities

Billy Taj<sup>1†</sup>, Mobolaji Adeolu<sup>1†</sup>, Xuejian Xiong<sup>1</sup>, Jordan Ang<sup>2</sup>, Nirvana Nursimulu<sup>1,3</sup> and John Parkinson<sup>1,4,5\*</sup>

## Abstract

**Background** Whole microbiome RNASeq (metatranscriptomics) has emerged as a powerful technology to functionally interrogate microbial communities. A key challenge is how best to process, analyze, and interpret these complex datasets. In a typical application, a single metatranscriptomic dataset may comprise from tens to hundreds of millions of sequence reads. These reads must first be processed and filtered for low quality and potential contaminants, before being annotated with taxonomic and functional labels and subsequently collated to generate global bacterial gene expression profiles.

**Results** Here, we present MetaPro, a flexible, massively scalable metatranscriptomic data analysis pipeline that is cross-platform compatible through its implementation within a Docker framework. MetaPro starts with raw sequence read input (single-end or paired-end reads) and processes them through a tiered series of filtering, assembly, and annotation steps. In addition to yielding a final list of bacterial genes and their relative expression, MetaPro delivers a taxonomic breakdown based on the consensus of complementary prediction algorithms, together with a focused breakdown of enzymes, readily visualized through the Cytoscape network visualization tool. We benchmark the performance of MetaPro against two current state-of-the-art pipelines and demonstrate improved performance and functionality.

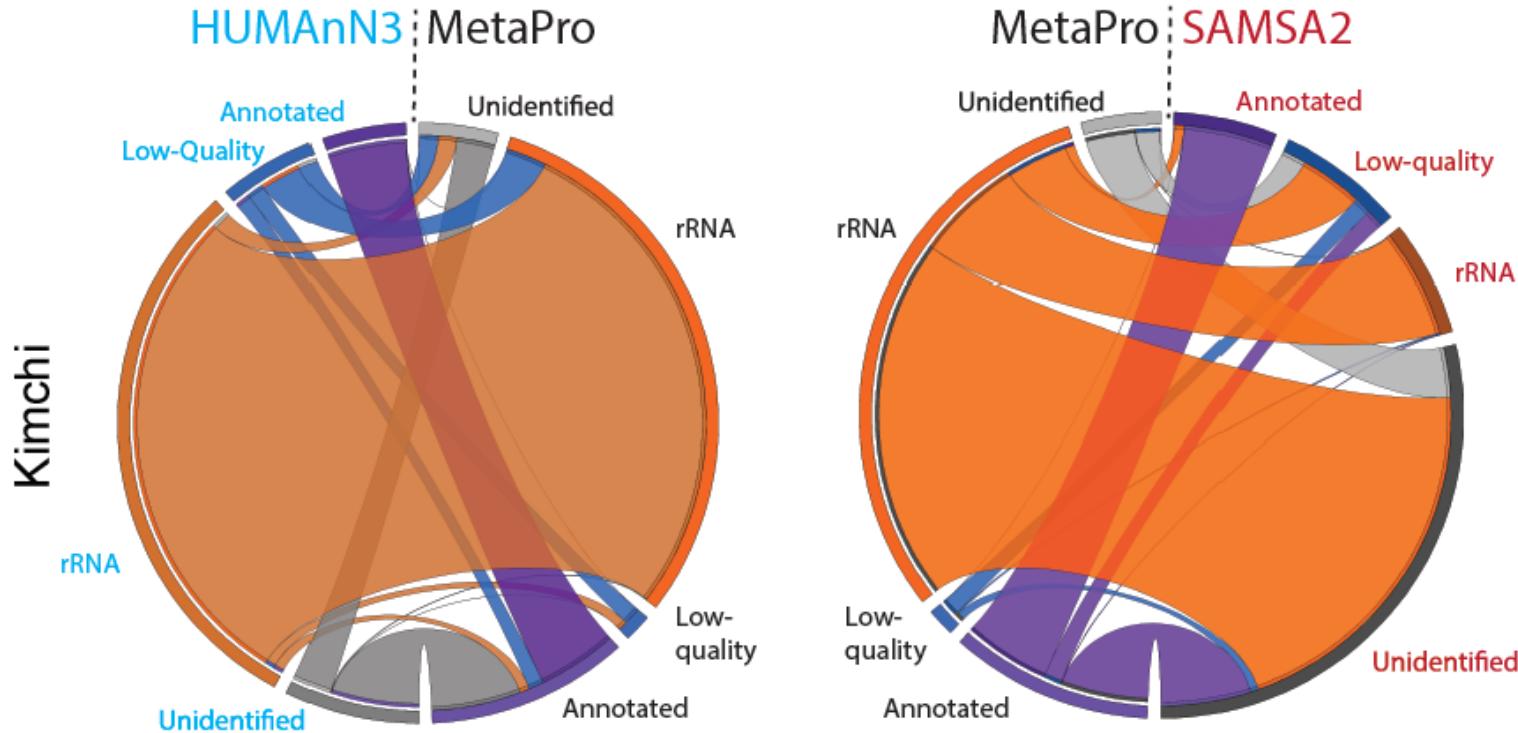
**Conclusions** MetaPro represents an effective integrated solution for the processing and analysis of metatranscriptomic datasets. Its modular architecture allows new algorithms to be deployed as they are developed, ensuring its longevity. To aid user uptake of the pipeline, MetaPro, together with an established tutorial that has been developed for educational purposes, is made freely available at <https://github.com/ParkinsonLab/MetaPro>. The software is freely available under the GNU general public license v3.

**Keywords** Metatranscriptomics, Microbiome function, Taxonomic annotation, Sequence analysis pipeline

# Comparing pipelines

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

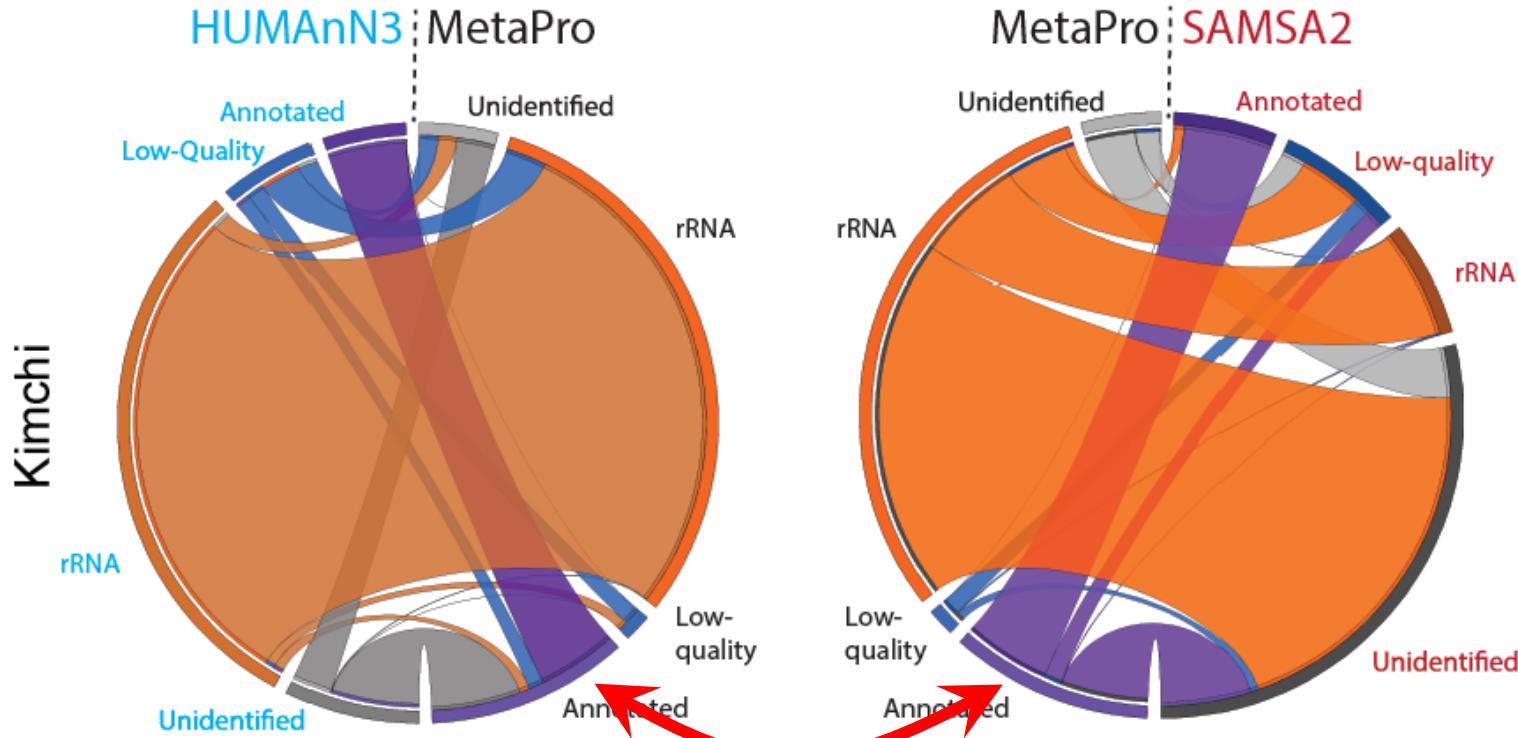
You could also build your own!



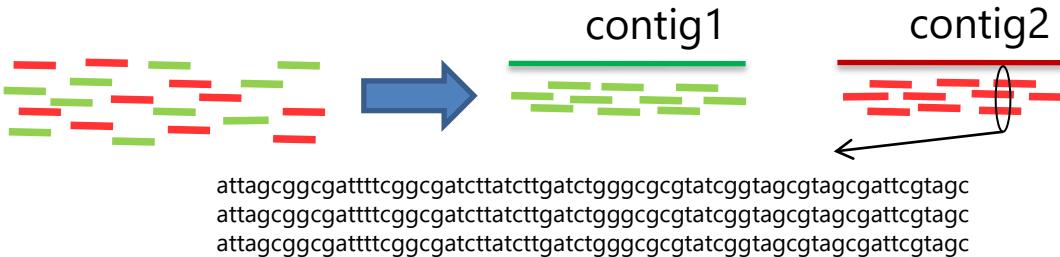
# Comparing pipelines

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

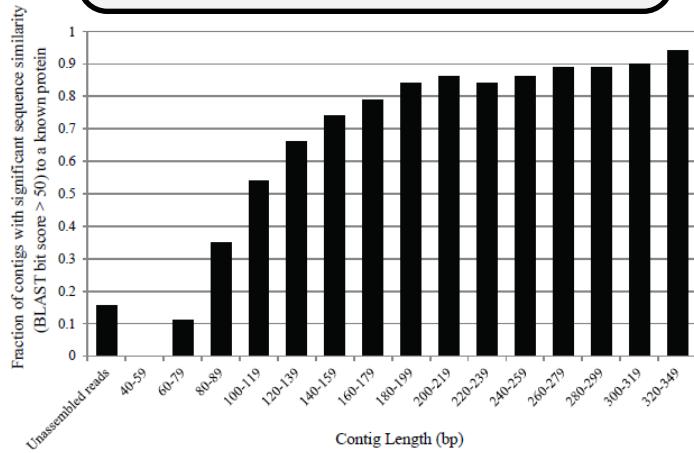
You could also build your own!



# Read processing - Assembly



Assembly improves annotation accuracy



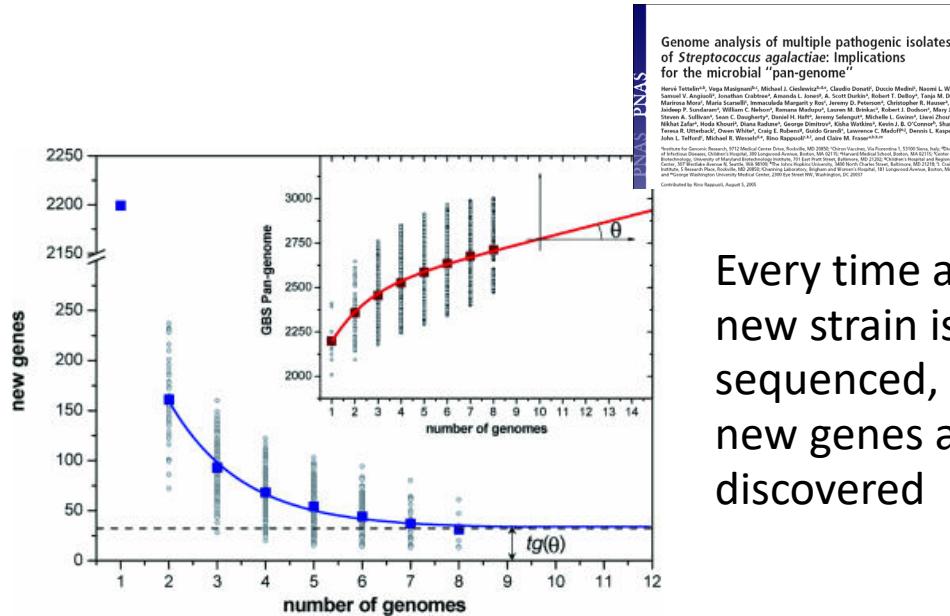
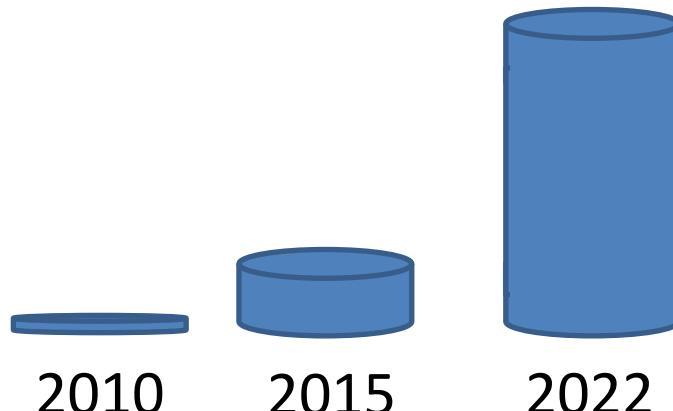
Here we use SPADes followed by MetaGeneMark to identify separate ORFs in transcripts that may represent operons

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

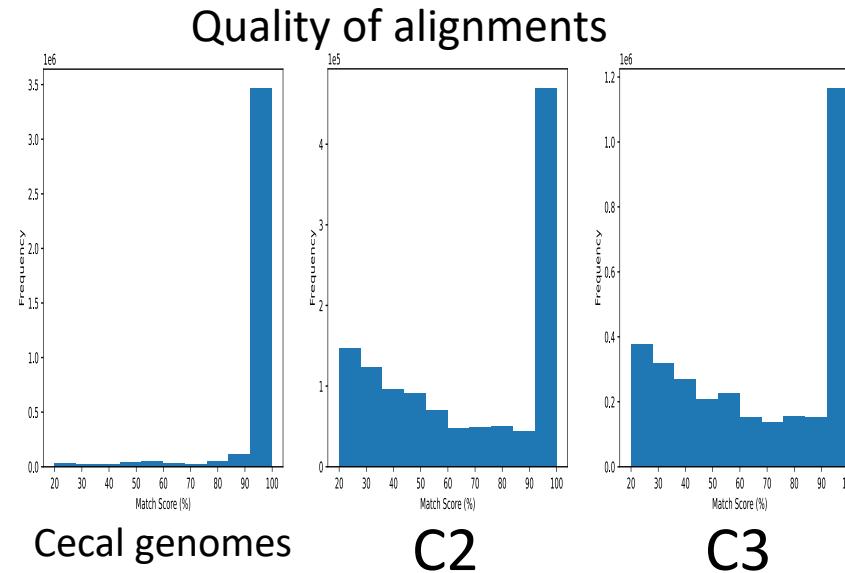
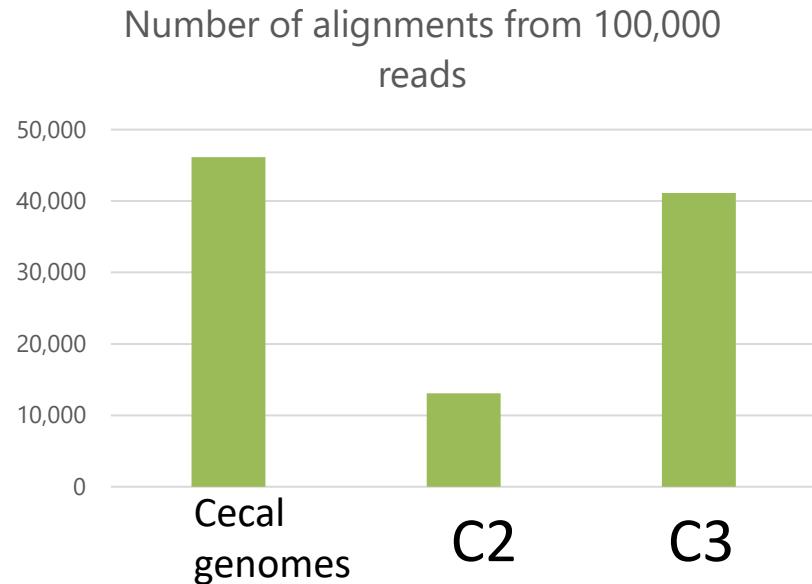
As the number of reference genomes increases, the amount of memory to perform BWA searches increases

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 HUMAnN2 & 3 (19Gb & 66Gb). The smaller, focused database runs 100x faster than C3.



# MGnify database for niche specific MAGs

EMBL-EBI | MGnify

# MGnify

Submit, analyse, discover and compare microbiome data

Search MGnify 

Example searches: Tara oceans, MGYS00000410, Human Gut

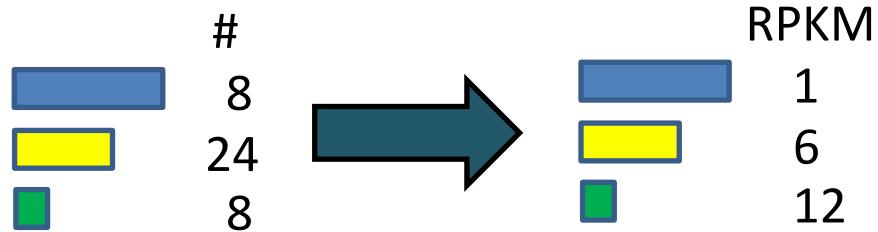
Overview Submit data Text search  Sequence search Browse data  API About Help Login

MGnify team are compiling niche-specific collections of MAGs to address the issue of database creep

# 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



$$\text{RPKM}_{\text{geneA}} = 10^9 \frac{C_{\text{geneA}}}{NL}$$

$C_{\text{geneA}}$  = number of reads mapped to geneA

N = total number of reads

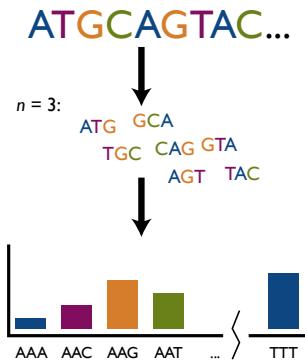
L = length of transcript in units of Kb

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

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



Published online 31 August 2012

Nucleic Acids Research, 2013, Vol. 41, No. 1 e3  
doi:10.1093/nar/gk220

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

Jiemeng Liu<sup>1,2,3</sup>, Haifeng Wang<sup>1,4</sup>, Hongxing Yang<sup>1,4</sup>, Yizhe Zhang<sup>5</sup>, Jinfeng Wang<sup>6</sup>, Fangqing Zhao<sup>6,\*</sup> and Ji Qian<sup>1,\*</sup>

<sup>1</sup>State Key Laboratory of Genetic Engineering, <sup>2</sup>State Key Laboratory of Surface Physics, <sup>3</sup>The T-Life Research Center Institute of Plant Biology, School of Life Sciences, Fudan University, <sup>4</sup>School of Life Sciences, Shanghai Jiaotong University, Shanghai, China, <sup>5</sup>School of Life Sciences, Beijing 100101, People's Republic of China, <sup>6</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

Received March 20, 2012; Revised July 1, 2012

Accepted

Published 13 Apr 2013

DOI: 10.1038/nature12287

Fast and sensitive taxonomic classification for metagenomics with Kaiju

Peter Menzel<sup>1</sup>, Kim Lee Ng<sup>1</sup> & Anders Krogh<sup>1</sup>

Method

Centrifuge: rapid and sensitive classification of metagenomic sequences

Wood and Salzberg Genome Biology 2014, 15:R46

http://genomebiology.com/2014/15/5/R46



Open Access

METHOD  
Kraken: ultrafast metagenomic sequence classification using exact alignments

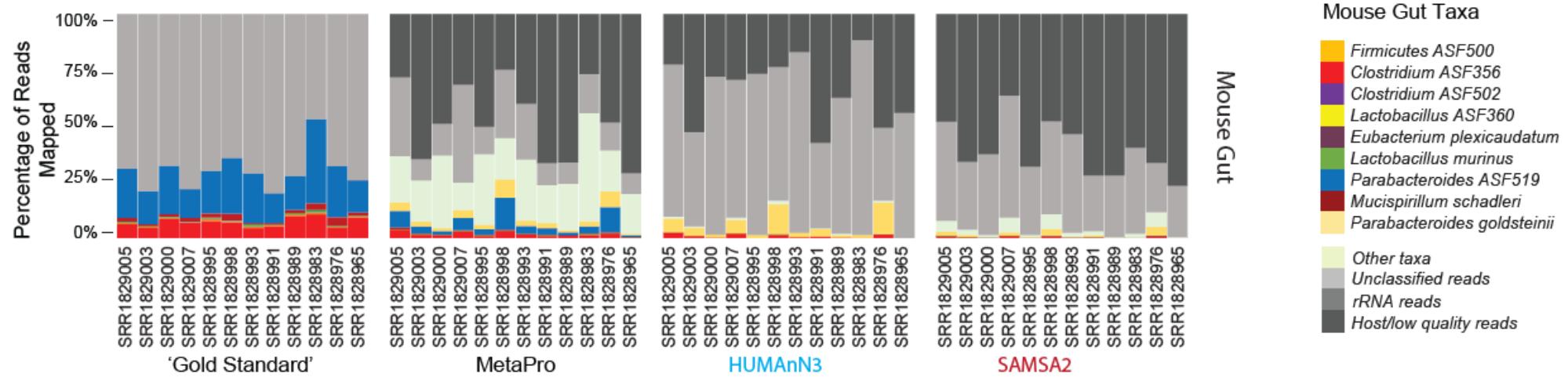
Derrick E Wood<sup>1,2\*</sup> and Steven L Salzberg<sup>2,3</sup>

Centrifuge - 2016

KRAKEN – 2014  
KRAKEN2 – 2019

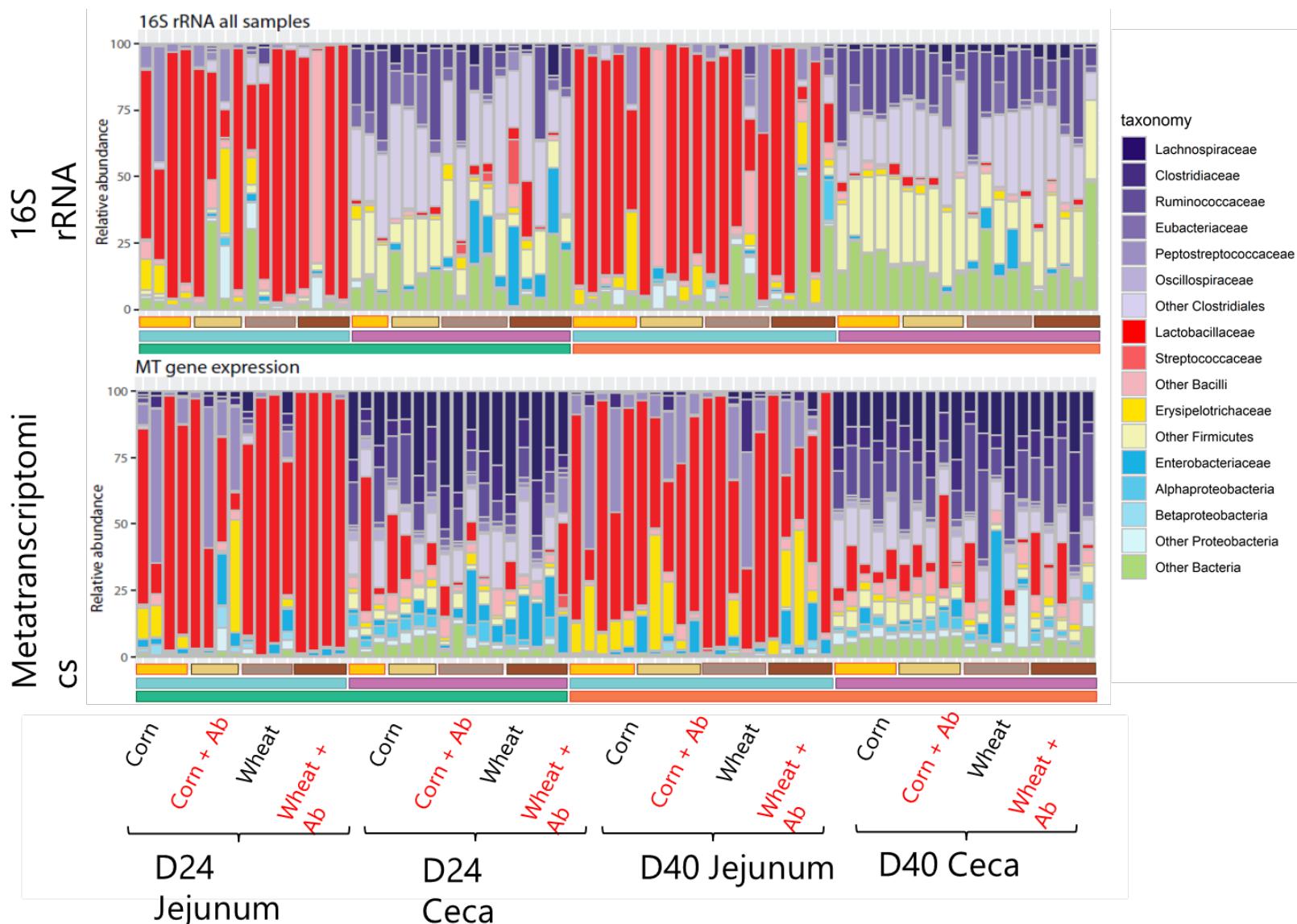
# MetaPro employs “majority voting” for taxonomic classification

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

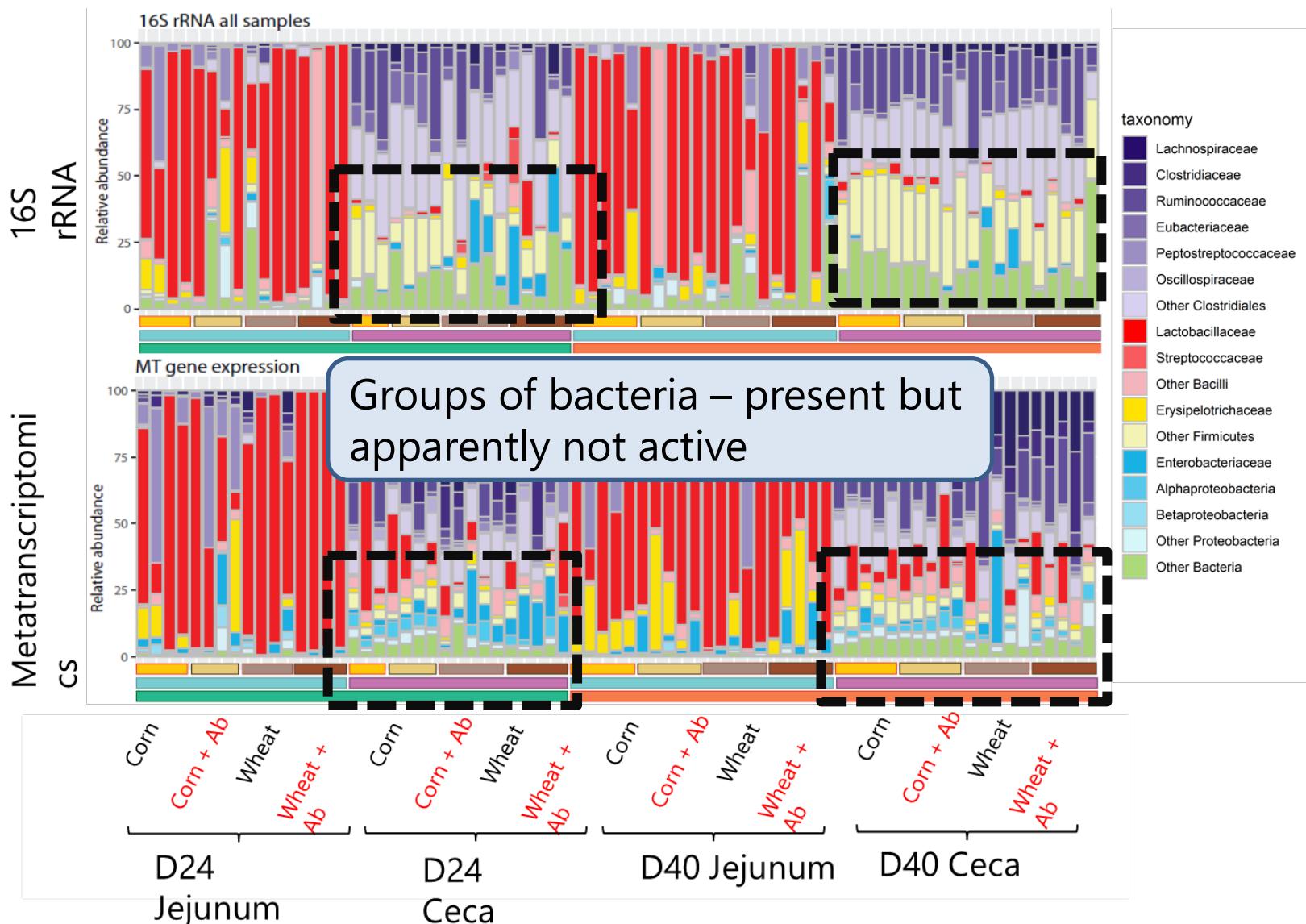




# 16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles

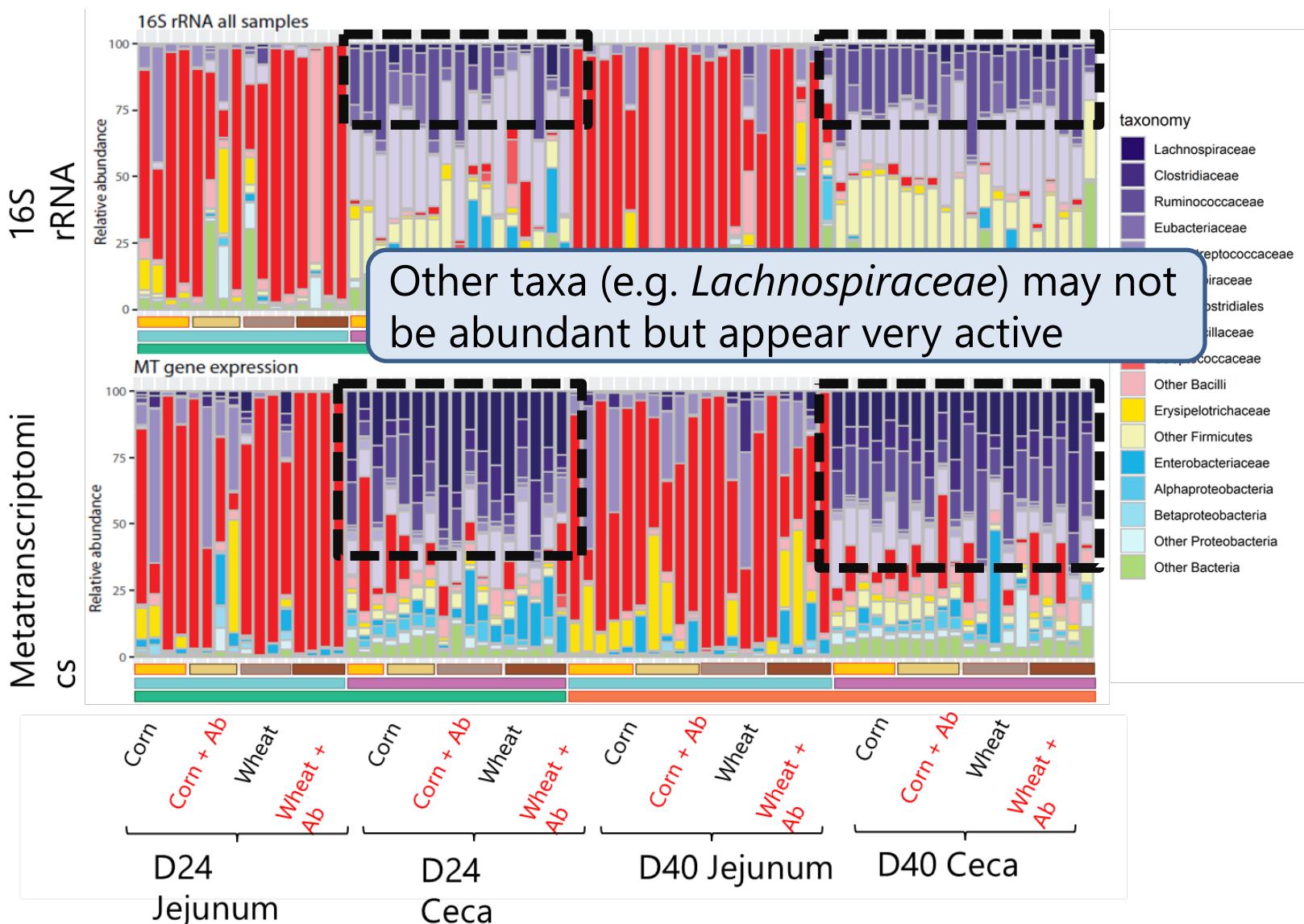


# 16S rRNA survey data and Metatranscriptomic data share similar, but not identical, taxonomic profiles





# 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



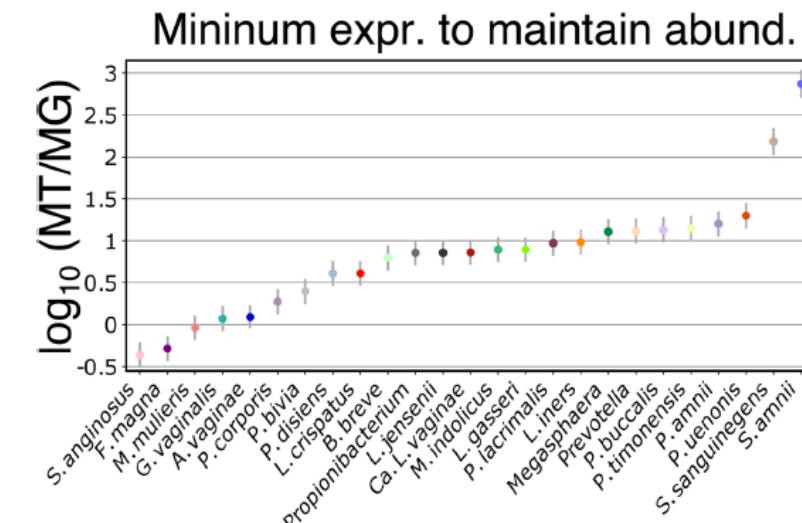
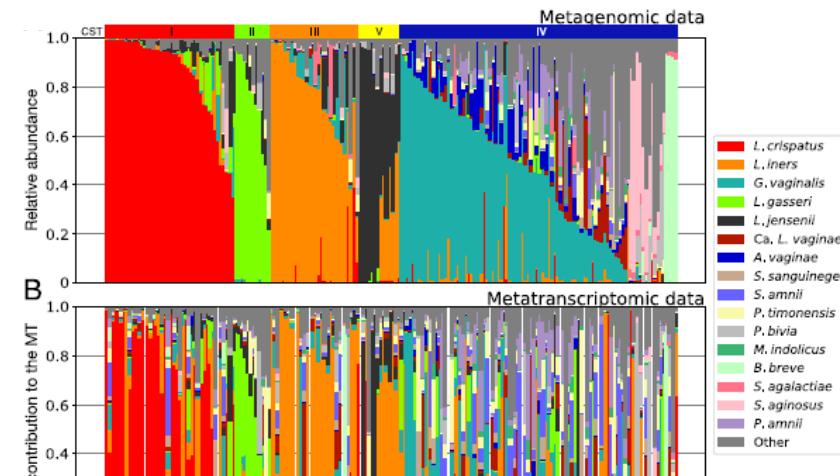
Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data

Michael T. France<sup>1</sup>, Li Fu<sup>1</sup>, Lindsay Rutt<sup>1</sup>, Hongqiu Yang<sup>1</sup>, Michael S. Humphrys<sup>1</sup>, Shilpa Narina<sup>1</sup>, Paweł M. Gajer<sup>1</sup>, Bing Ma<sup>1</sup>, Larry J. Forney<sup>2</sup> and Jacques Ravel<sup>1,3,4,5</sup>

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

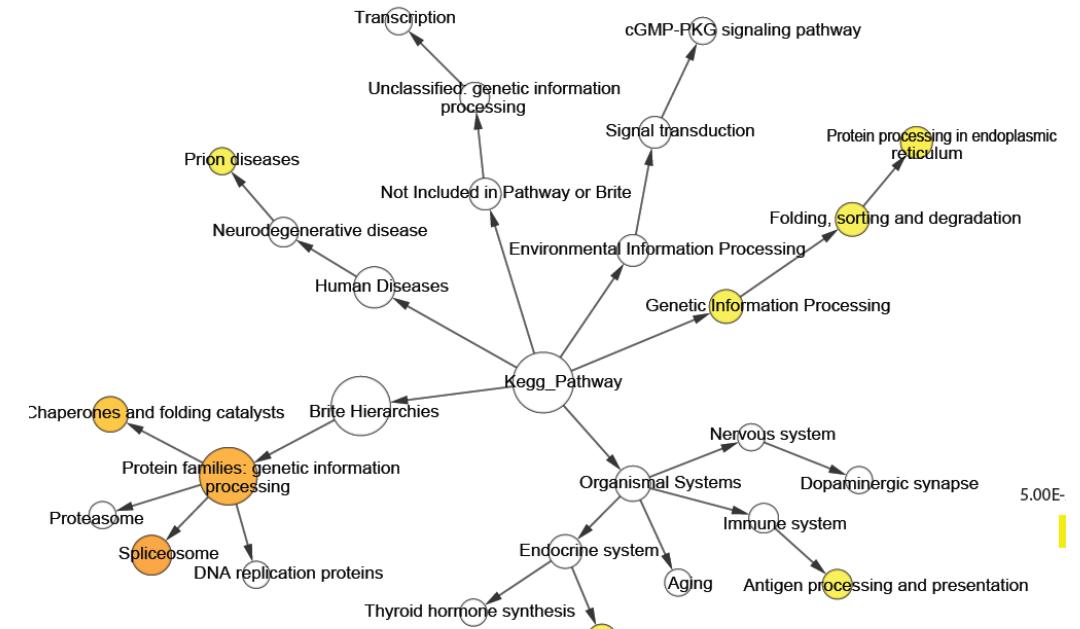
Nucleic Acids Research, 2019, Vol. 47, Database issue D309–D314  
doi: 10.1093/nar/gky1085

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

Jaime Huerta-Cepas<sup>1,2,\*†</sup>, Damian Szkłarczyk<sup>3,†</sup>, Davide Heller<sup>3</sup>, Ana Hernández-Plaza<sup>2</sup>, Sofia K. Forslund<sup>1,4</sup>, Helen Cook<sup>5</sup>, Daniel R. Mende<sup>6</sup>, Ivica Letunic<sup>7</sup>, Thomas Rattei<sup>8</sup>, Lars J. Jensen<sup>5</sup>, Christian von Mering<sup>5,3</sup> and Peer Bork<sup>1,9,10,11,\*</sup>

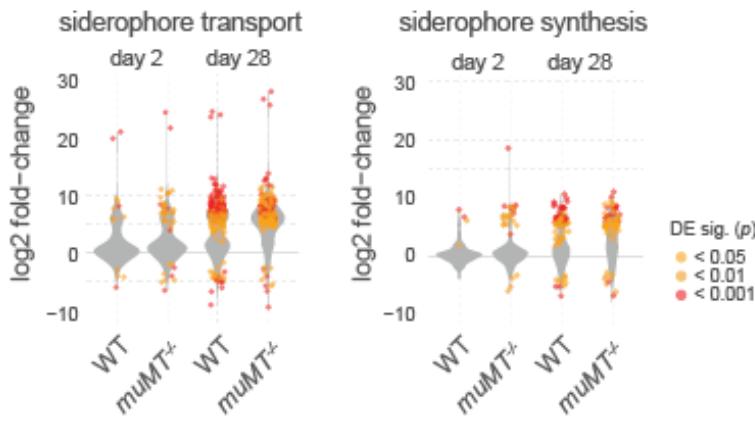
In addition to UniProt - EggNOG provides mappings to Gene Ontology (GO), KEGG enzymes, KEGG modules, CAZy

GO terms can be challenging to summarize...

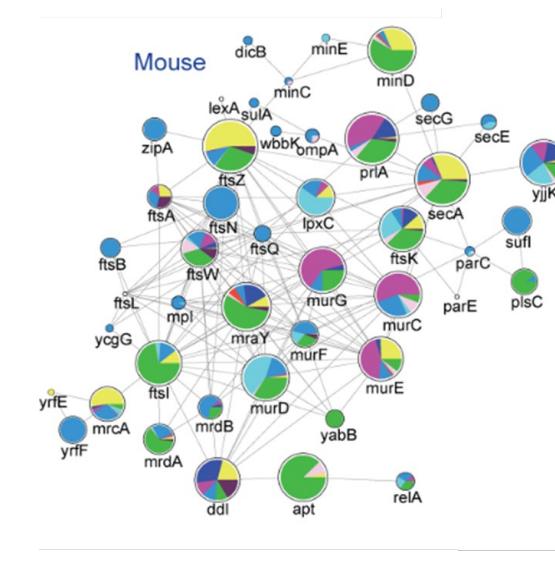
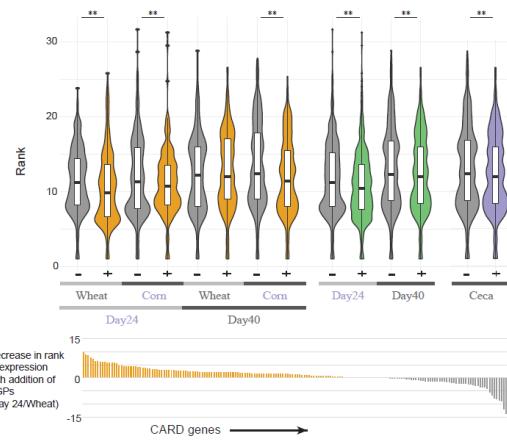


# Functional Annotation

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



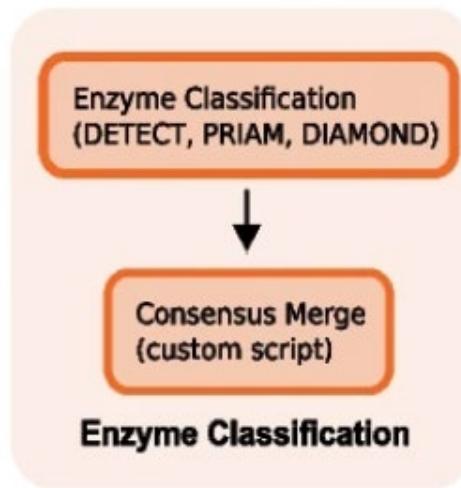
## AMR genes with CARD



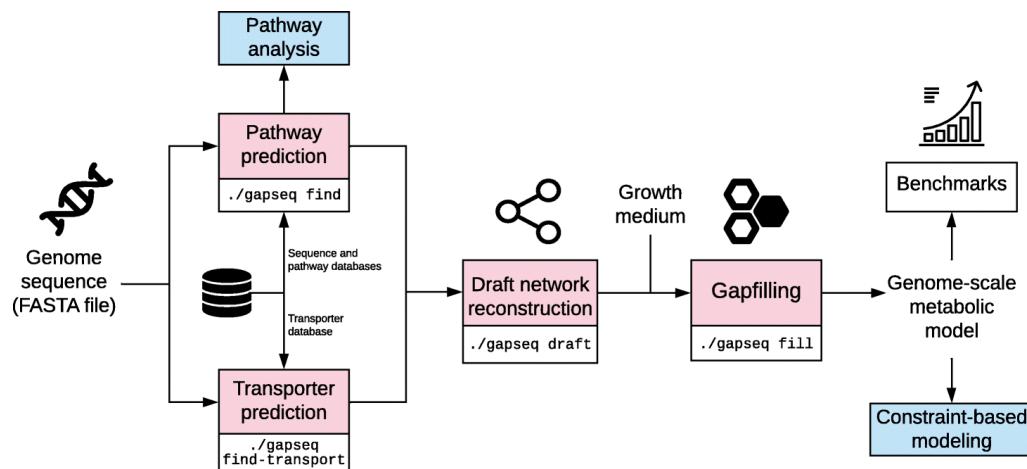
## Iron capture / storage with FeGenie

## Cell wall biogenesis genes from protein interaction datasets

# Functional Annotation – metabolic reconstructions



MetaPro relies on an in-house algorithm (DETECT) supplemented with PRIAM predictions to perform high quality annotation of enzymes  
These can be used for metabolic reconstructions



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

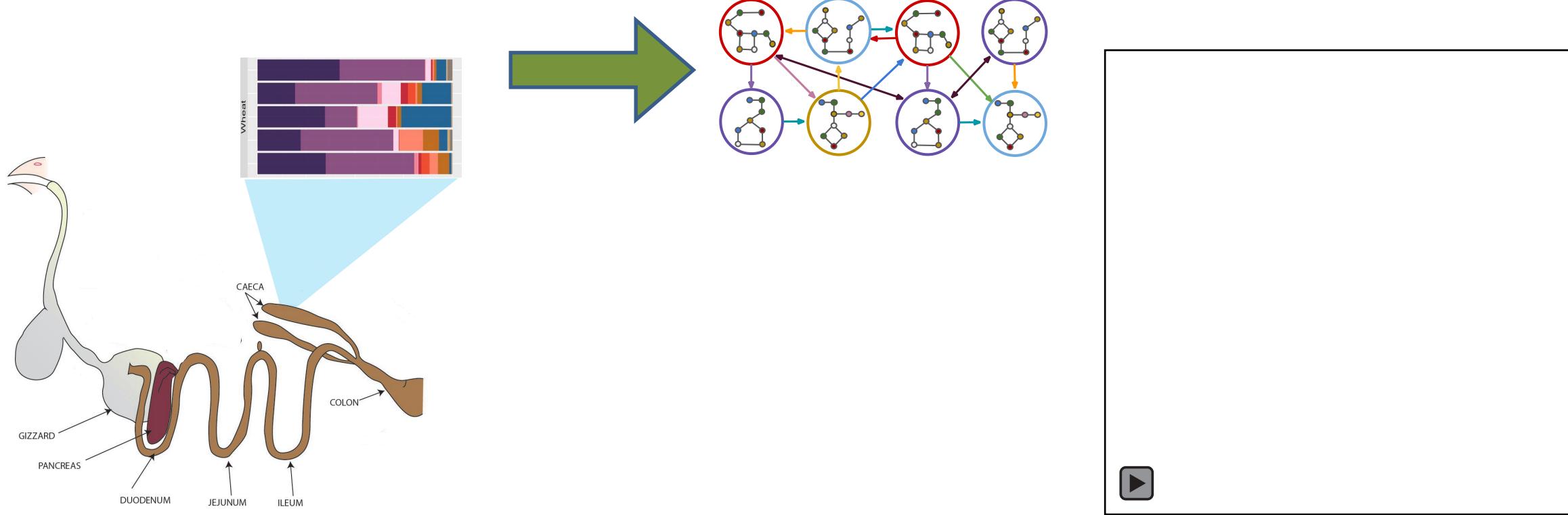


Johannes Zimmermann<sup>1</sup>, Christoph Kaleta<sup>1</sup> and Silvio Waschini<sup>1,2\*</sup>

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

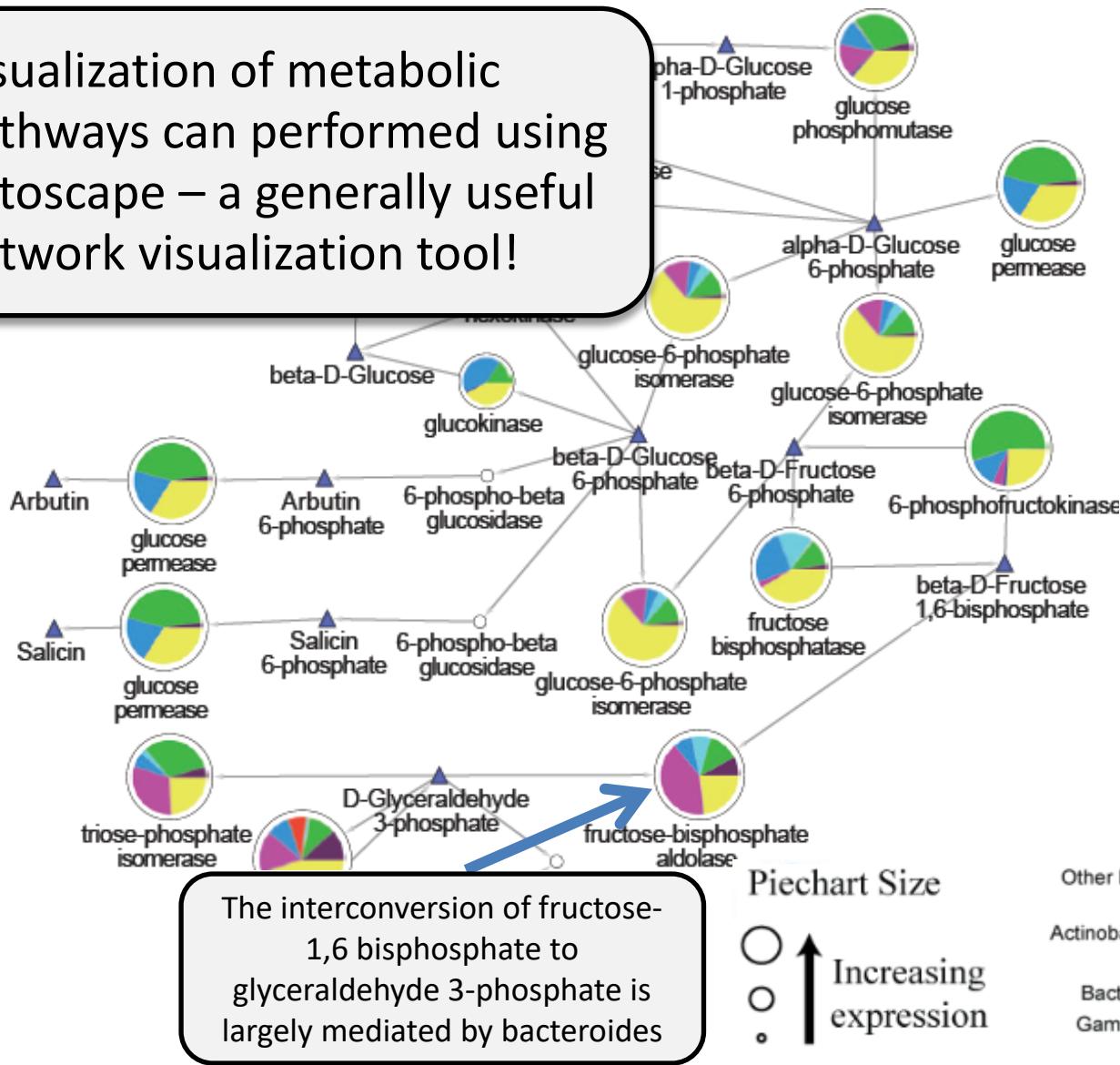
# Functional Annotation – metabolic modeling

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

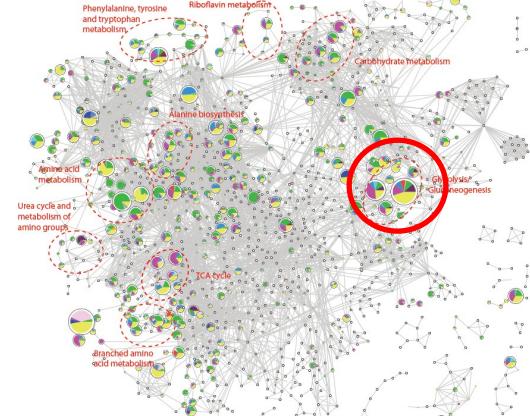


# Visualizing results

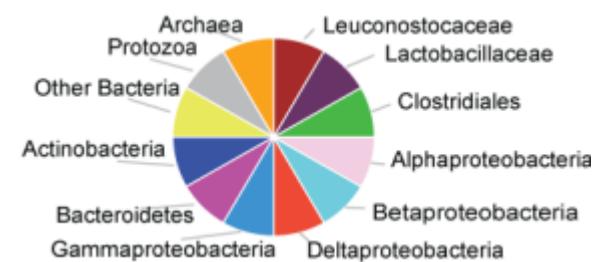
Visualization of metabolic pathways can be performed using Cytoscape – a generally useful network visualization tool!



## Metabolic network of a mouse gut microbiome



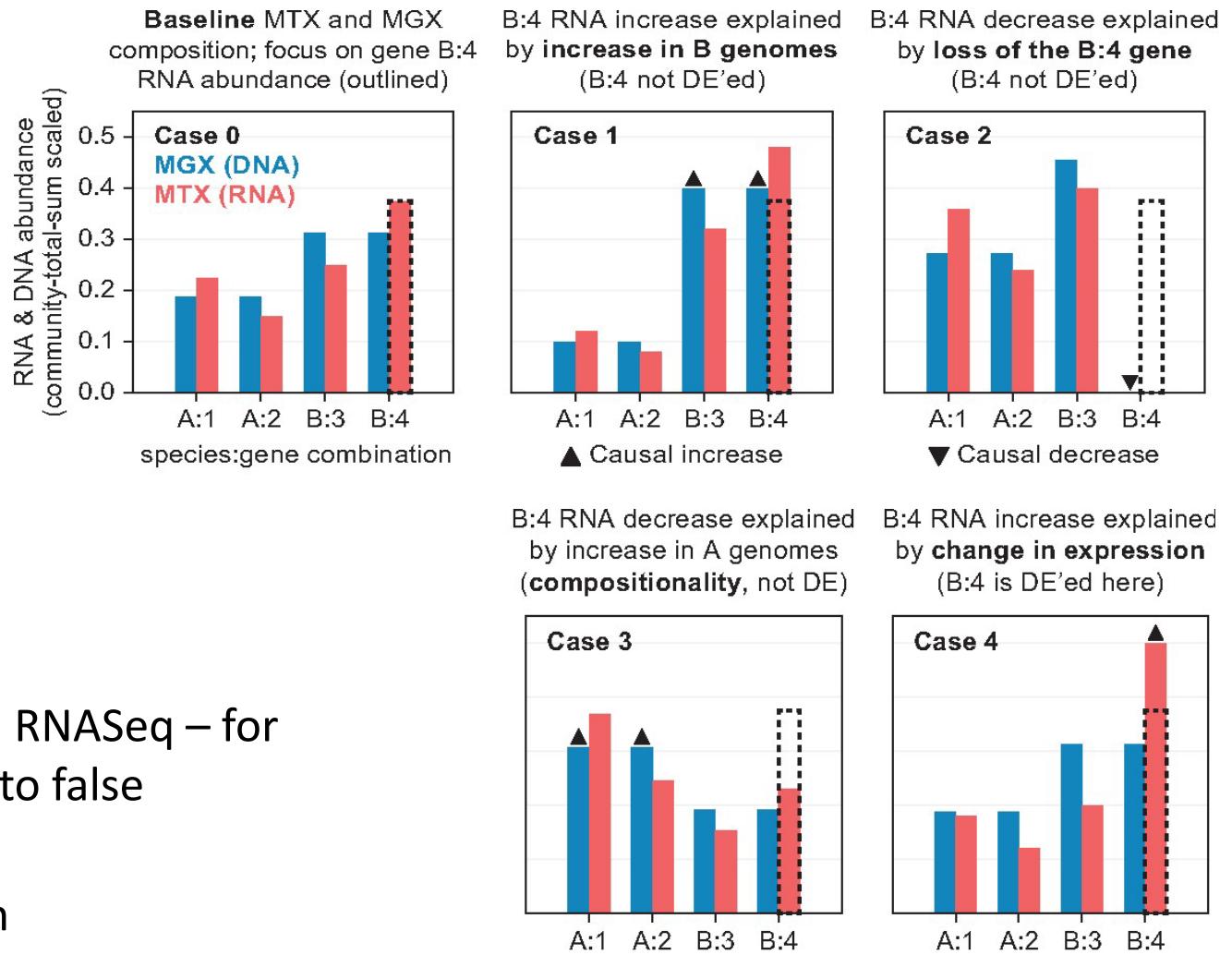
Piecharts indicate relative contribution of each taxon to an enzymatic activity



# Detecting differentially expressed transcripts

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

- DESeq2 developed for single organism RNASeq – for metatranscriptomics DESeq2 is prone to false positives,
- Alternative methods? - Aldex2, Ancom
- **Need to normalize for taxon/gene abundance**



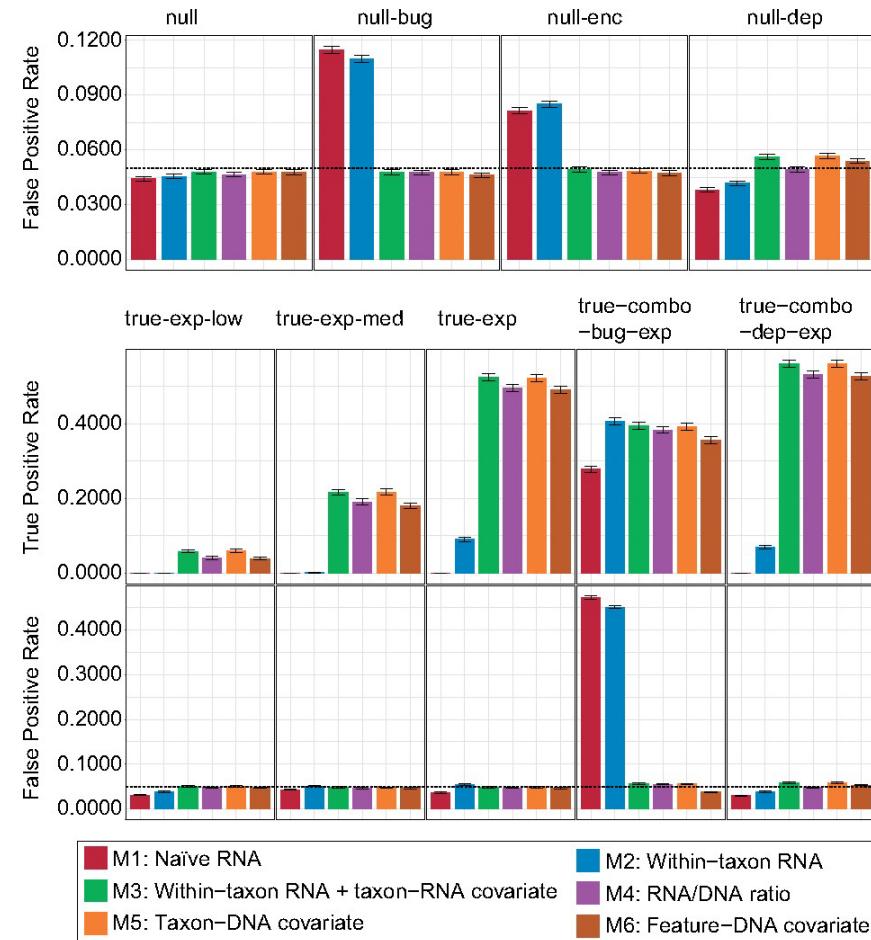
Zhang et al Bioinformatics 2021

# Normalizing for taxon/gene abundance

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

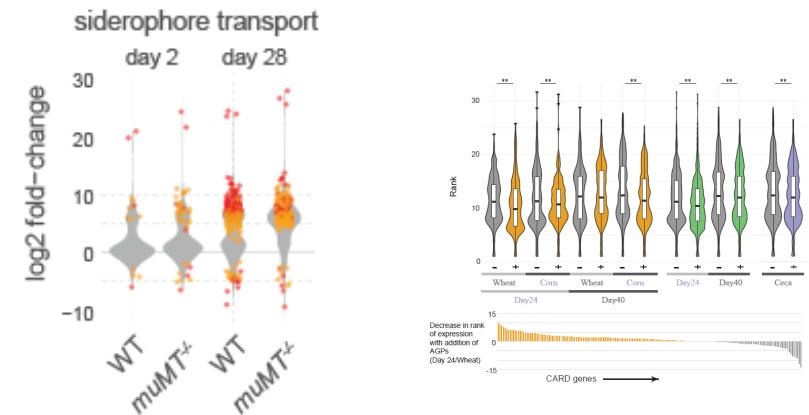
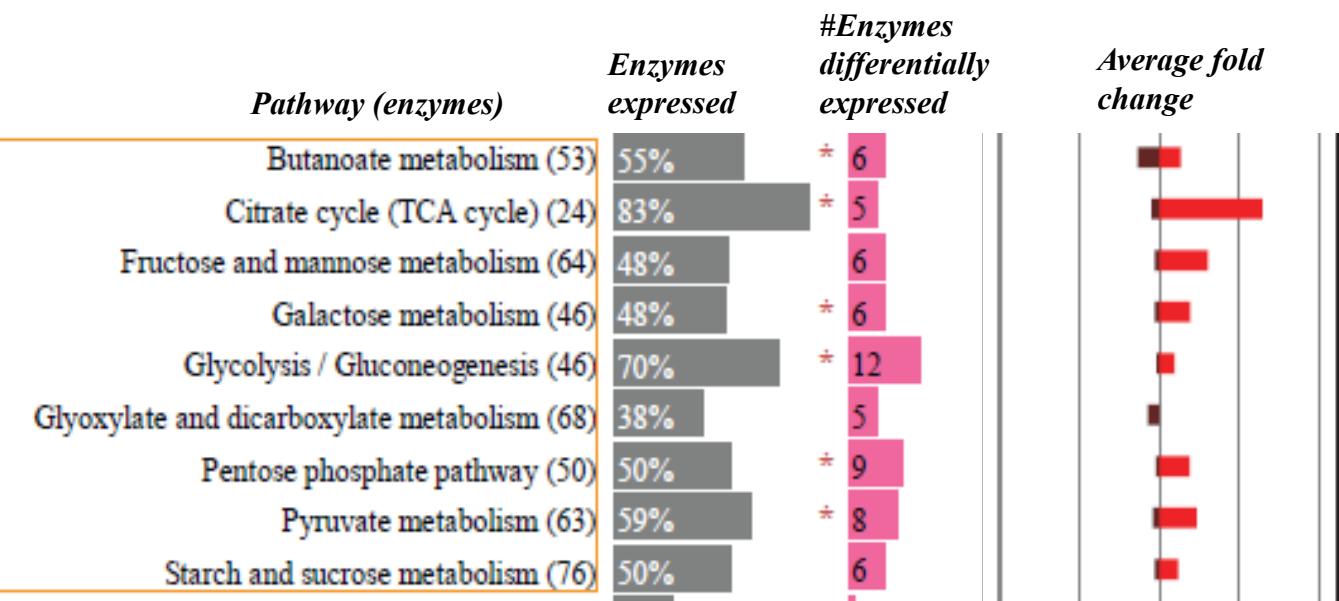
When samples are profiled with paired MTX and MGX sequencing, a gene's RNA abundance can be normalized by its DNA abundance

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



# Gene set enrichment analyses can help identify enriched functions

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



Can be applied to other functional groups (e.g. iron transport / AMR genes)