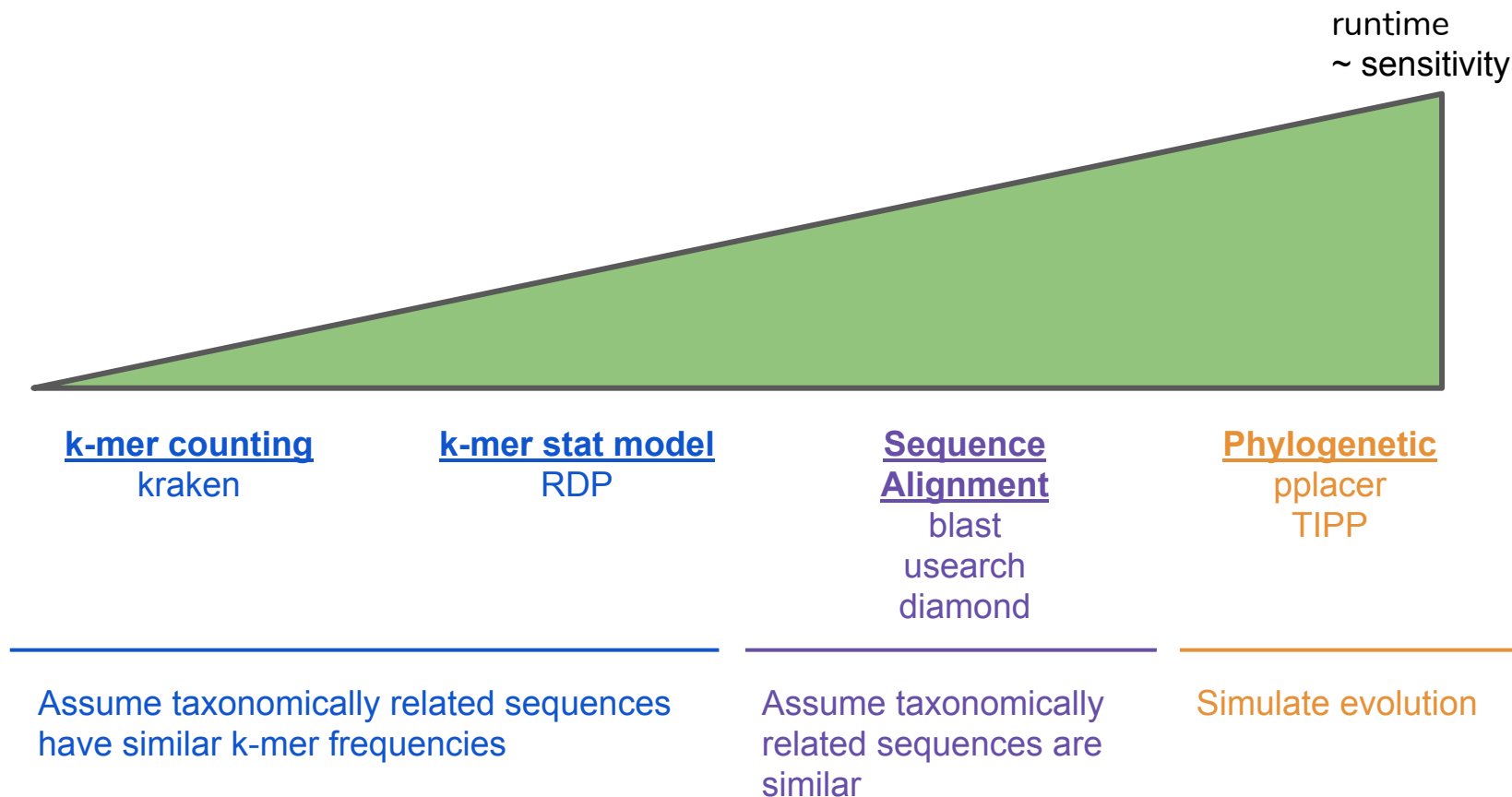


Comparing Taxonomic Tools

M³ Workshop
January 10, 2019

Landscape of Taxonomic Approaches



What are the similarities and differences of the taxonomic assignment methods discussed today?

How do you decide what is the right database for you?

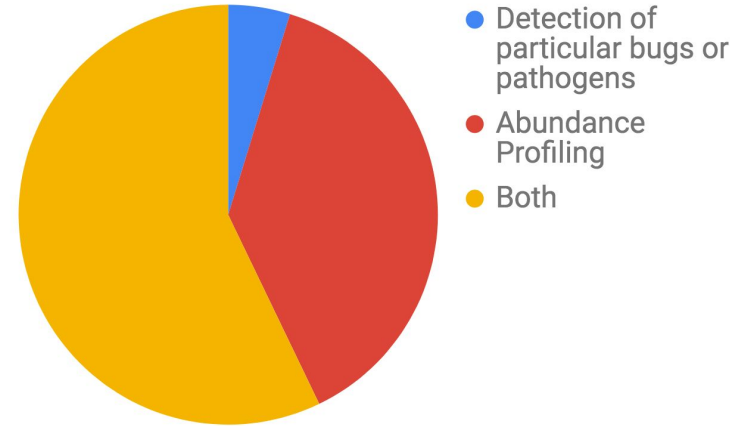
How do you best assess confidence in your methods?

What are the similarities and differences of the taxonomic assignment methods discussed today?

How do you decide what is the right database for you?

How do you best assess confidence in your methods?

WHAT IS THE GOAL OF YOUR STUDY?



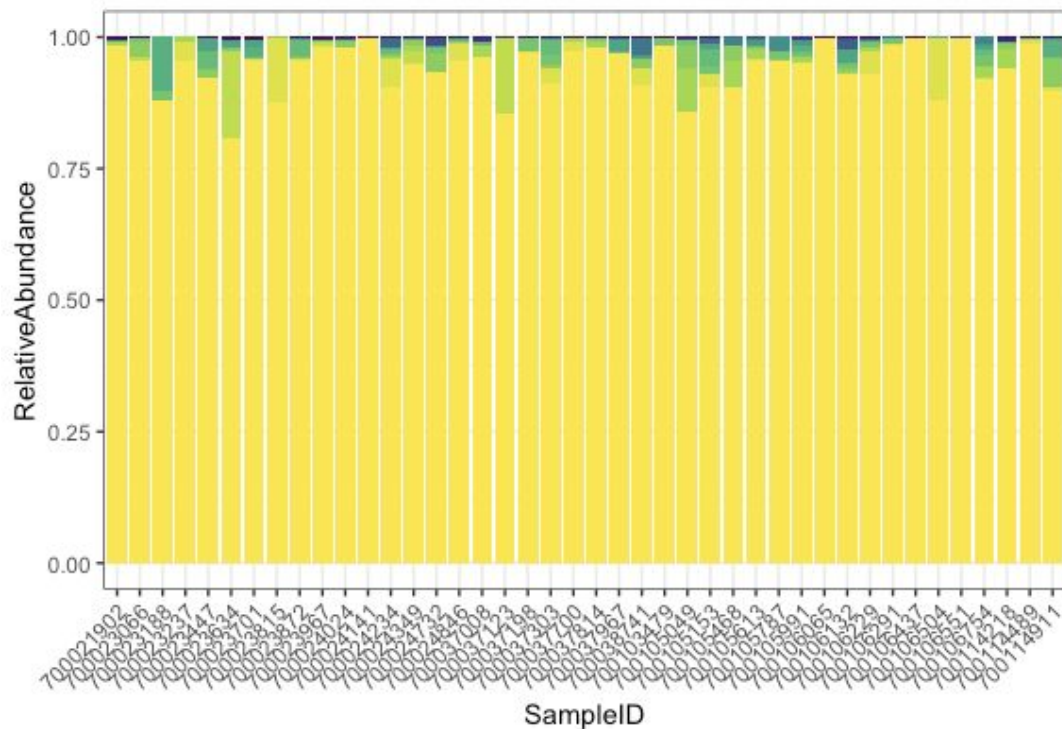
Overall Recommendations

- Retain OTU-level information (even if you aggregate)
- Use a domain-specific database if available
- Most tools are OK for initial broad classification
- If you really care about taxonomic label – dig deep
 - search multiple databases
 - look for related type strains
 - use phylogenetic approaches

Discussion Question 1

What are the similarities and differences of the taxonomic assignment methods discussed today?

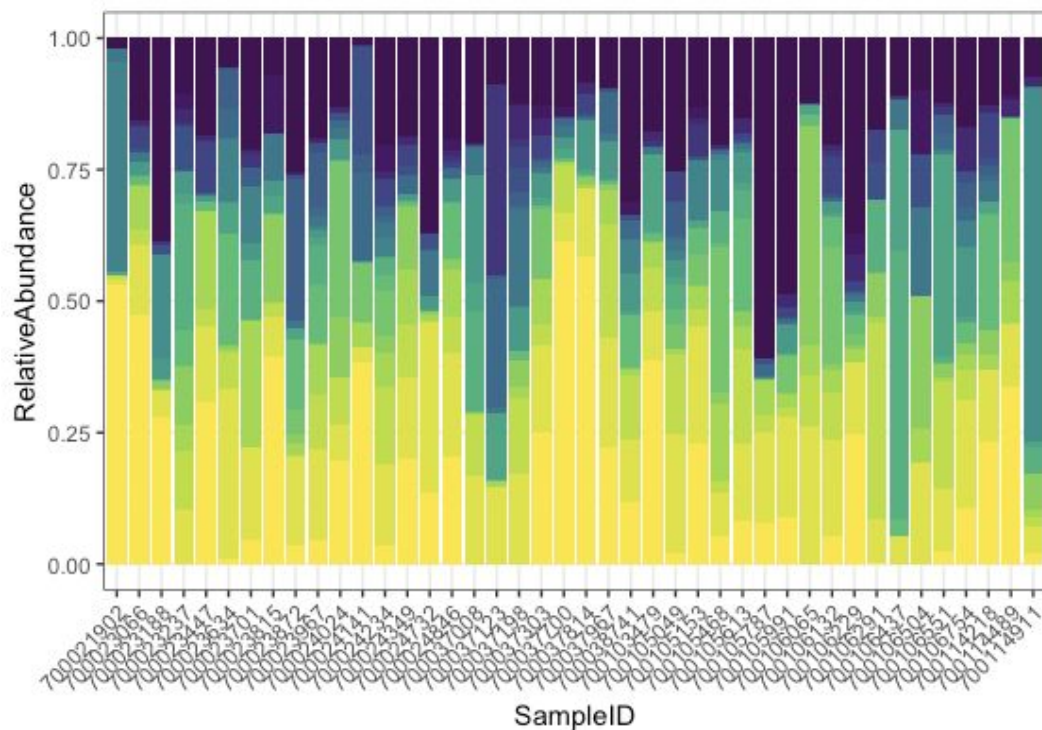
Blast Partitions-Species



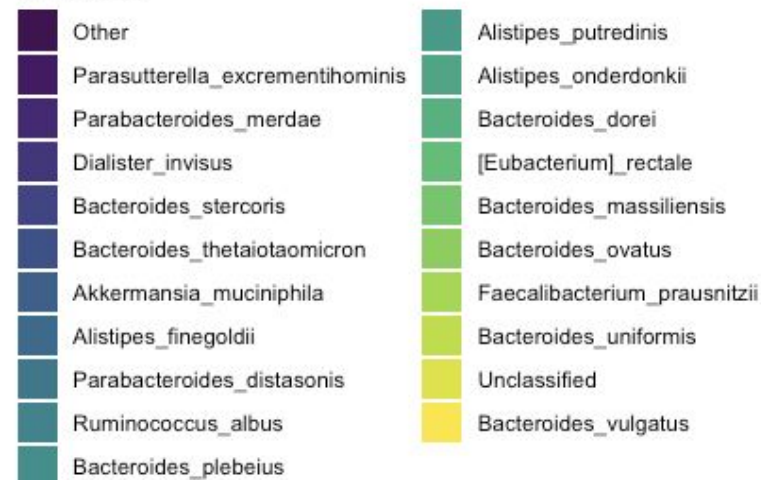
Taxonomy

Other	Alistipes indistinctus
Prevotella bivia	Ruminococcus bicirculans
Bacteroides nordii	Paraprevotella clara
Oxalobacter formigenes	Coprococcus eutactus
Barnesiella intestinihominis	Odoribacter splanchnicus
Victivallis vadensis	Acidaminococcus intestini
Bacteroides eggerthii	Bacteroides caccae
Bacteroides salyersiae	Breznakia pachnodae
Odoribacter laneus	Parasutterella excrementihominis
Butyrivibrio crossotus	Unclassified
[Clostridium] leptum	

TIPP-Species



Taxonomy



Discussion Question 2

How do you decide what is the right database for you? How much granularity do you need? When adding to databases, how does adding multiple strains of the same species confuse classifiers?

Discussion Question 3

Do you use different databases if you have very diverse data versus data from mostly a single genus but multiple species--how will this affect your results and your confidence in these results?

Discussion Question 4

How do you assess confidence in all of these methods when you don't know the ground truth? How do we know when the sequence is novel versus a close relative to something in our database? How do we improve our databases to improve overall classification?

Interactive Exercise

Try running one of the pipelines on a new dataset and/or with a new database. What are some challenges looking at a different environment?

Downloading the Data You Generated

```
scp -r
```

```
USER@openclass.umiacs.umd.edu:/classhomes/USER/m3-taxonomy-workshop/  
./m3_taxa_workshop_output/
```

Downloading the Test Datasets

```
scp -r
```

```
USER@openclass.umiacs.umd.edu:/fs/m3taxworkshop/data/1-datasets/hmp/  
./m3_taxa_workshop_hmp_test_datasets/
```

Downloading the Singularity Containers

```
scp -r USER@openclass.umiacs.umd.edu:/fs/m3taxworkshop/images/  
./m3_taxa_singularity_images/
```

Thank you!!!



<https://m3.groups.io/g/main>

JHU

Jocelyne
DiRuggiero
Sarah Preheim
Kristin Staicos
Cindy Holstein



JOHNS HOPKINS
ENVIRONMENTAL HEALTH
& ENGINEERING



JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

UMD

Mihai Pop
Barbara Lewis
Derek Yarnell



JOHNS HOPKINS
DEPARTMENT OF BIOLOGY



JOHNS HOPKINS
KRIEGER SCHOOL
of ARTS & SCIENCES

M³ Sponsors

Please don't hesitate to contact us with any additional thoughts or questions!

Jennifer Lu - JHU - jlu26@jhmi.edu

Jackie Meisel - UMD - meiselj@umiacs.umd.edu

Dan Nasko - UMD - dnasko@umiacs.umd.edu

Nidhi Shah - UMD - nidhi@cs.umd.edu

Please fill out the post-workshop survey!

<https://goo.gl/forms/oT2wDIMA8wMJbI552>