

# End-to-End Metagenomics: decision making

CINEMA Workshop

July 2023

# Is there a “standard” pipeline?

## SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline

 Javier Tamames\* and  Fernando Puente-Sánchez

Software | [Open Access](#) | Published: 15 September 2018

### MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis

[Gherman V. Uritskiy](#), [Jocelyne DiRuggiero](#) & [James Taylor](#)

[Microbiome](#) 6, Article number: 158 (2018) | [Cite this article](#)

Article | [Open Access](#) | Published: 07 June 2022

### Metagenomic pipeline for identifying infections among distinct SARS-CoV-2 sequences of concern: study cases from

[Jose Arturo Molina-Mora](#), [Estela Cordero-Laurent](#), [Chacón-Ramírez](#) & [Francisco Duarte-Martínez](#)

[Scientific Reports](#) 12, Article number: 9377 (2022)

### Struo: a pipeline for metagenome profilers

[Jacobo de la Cuesta-Zuluaga](#), [Ruth E Ley](#), [Nicholas D Youngblut](#)

*Bioinformatics*, Volume 36, Issue 7, April 2020, Pages 2314–2315,

## Microseek: A Protein-Based Metagenomic Pipeline for Virus Diagnostic and Discovery

by  Philippe Pérot<sup>1,†</sup>,  Thomas Bigot<sup>2,†</sup>,  Sarah Temmam<sup>1</sup>,  Béatrice Regnault<sup>1</sup> and  Marc Eloit<sup>1,3,\*</sup>

## MetAMOS: a modular and open source metagenomic assembly and analysis pipeline

[Todd J Treangen](#), [Sergey Koren](#), [Daniel D Sommer](#), [Bo Liu](#), [Irina Astrovskaya](#), [Brian Ondov](#), [Aaron E Darling](#), [Adam M Phillippy](#) & [Michael D Benson](#)

[Genome Biology](#)

## Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics

Methodology | [Open Access](#) | [Léa Siegwald](#), [Hélène Touzet](#), [Yves Lemoine](#), [David Hot](#), [Christophe Audebert](#), [Ségolène Caboche](#)

BioMaS: a Review | Published: January 4, 2017 • <https://doi.org/10.1371/journal.pone.0169563>

Review | [Published: 18 October 2021](#)

## Music of metagenomics—a review of its applications, analysis pipeline, and associated tools

[Bilal Wajid](#), [Faria Anwar](#), [Imran Wajid](#), [Haseeb Nisar](#), [Sharoze Meraj](#), [Ali Zafar](#), [Mustafa Kamal Al-Shawaqfeh](#), [Ali Riza Ekti](#), [Asia Khatoon](#) & [Jan S. Suchodolski](#)

[Functional & Integrative Genomics](#) 22, 3–26 (2022) | [Cite this article](#)

[Sergei Kosakovsky Pond](#)<sup>1,2,6,9</sup>, [Samir Wadhawan](#)<sup>3,6,7</sup>, [Francesca Chiaromonte](#)<sup>4</sup>, [Guruprasad Ananda](#)<sup>1,3</sup>, [Wen-Yu Chung](#)<sup>1,3,8</sup>, [James Taylor](#)<sup>1,5,9</sup>, [Anton Nekrutenko](#)<sup>1,3,9</sup> and The Galaxy Team<sup>1</sup>

## m: an extensible pipeline for analyzing metagenomic sequencing experiments

[Louis J. Taylor](#), [Chunyu Zhao](#), [Andrew Connell](#), [Jung-Jin Lee](#), [Bryton Fett](#), [Frederic D. Bushman](#) & [Kyle Bittinger](#)

[Microbiome](#) 7, Article number: 46 (2019) | [Cite this article](#)

Review  
Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline

[Matthew B. Sullivan](#)<sup>b</sup>

with the  
le

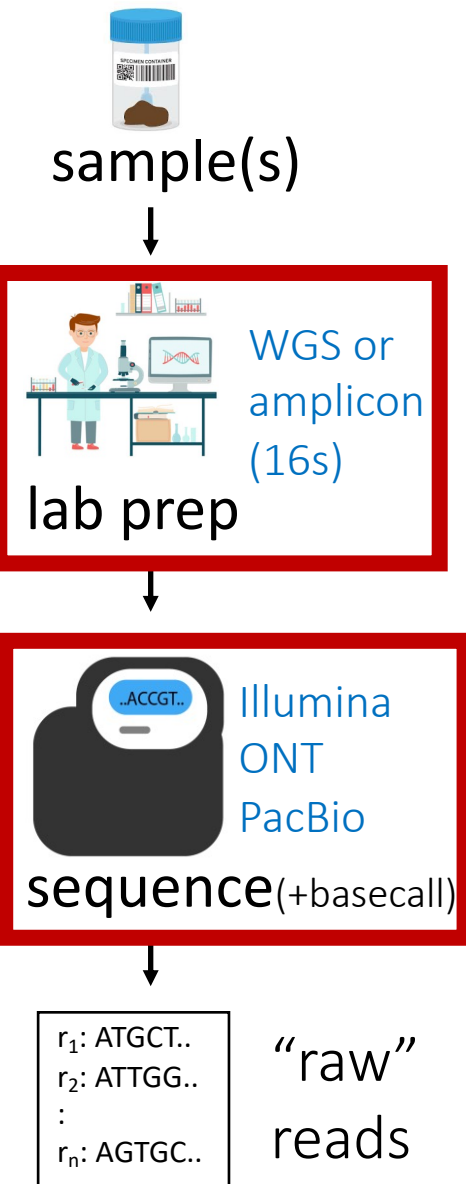
# Which questions are you looking to answer?

---

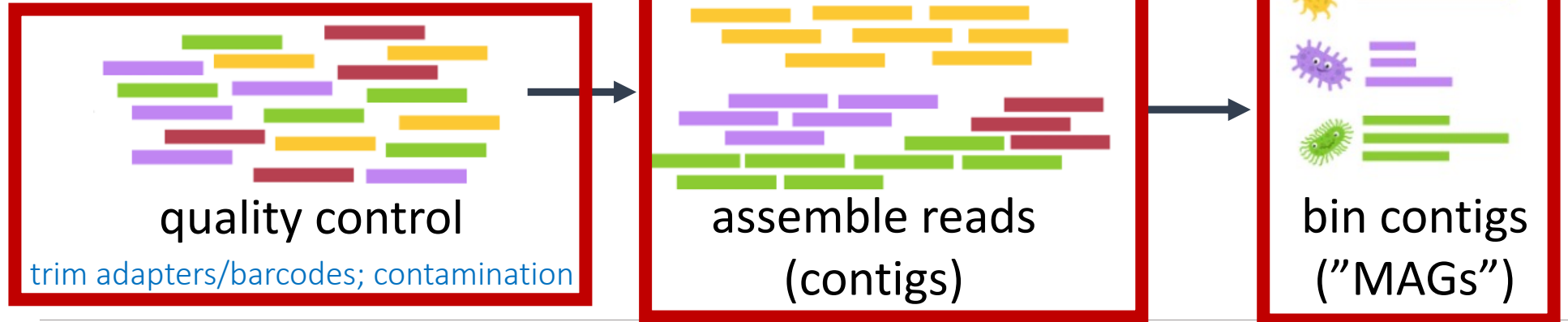
- |   |                                 |
|---|---------------------------------|
| - Detecting known genes in your community                     | <u>library prep? sequencer?</u> |
| - linked known genes with taxonomy                            | -                               |
| - taxonomic profiles (cheap and fast :)                       | -                               |
| - ..strain specific   | -                               |
| - searching for a known strain in communities                 | -                               |
| GROUP A   | -                               |
| - MGE detection – gene/pangenome detection and classification | -                               |
|   | -                               |
| - ZYMO ONT data   | -                               |
| GROUP C   | -                               |
| - assemble + binning  | -                               |
| GROUP B   |                                 |

# Metagenomics pipeline

## 1. generate reads



## 2. clean and sort

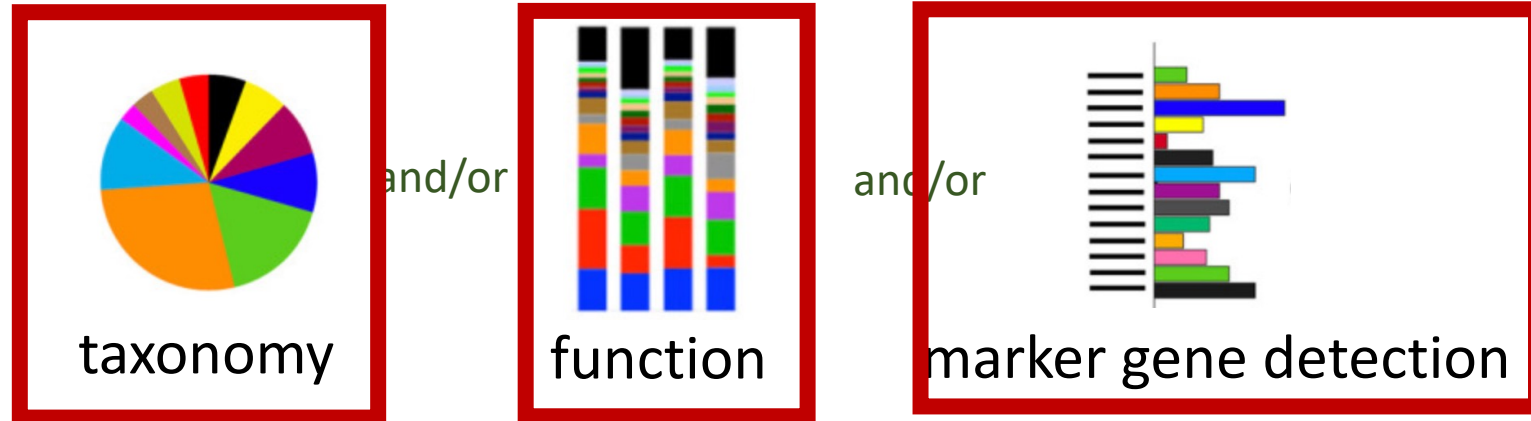


## 3. classify

(non-exhaustive list)

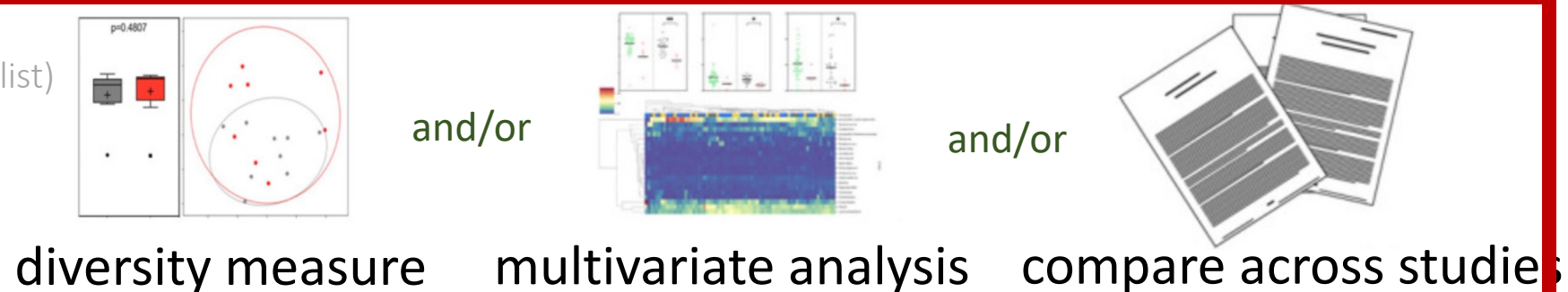
input:

reads, contigs, or MAGs



## 4. stats

(non-exhaustive list)





# Long or short reads?

---



# WGS or targeted amplicon (16s)?

---

# Utilizing Rice Virtual Machines

---

1. Obtain a login from one of your TAs

1. Login from your command line

\$ssh [studentX@cinema1.rice.edu](mailto:studentX@cinema1.rice.edu)

2. Paths to datasets

Zymo Mock (Tuesday): /home/kdc10/datasets/zymo\_mock

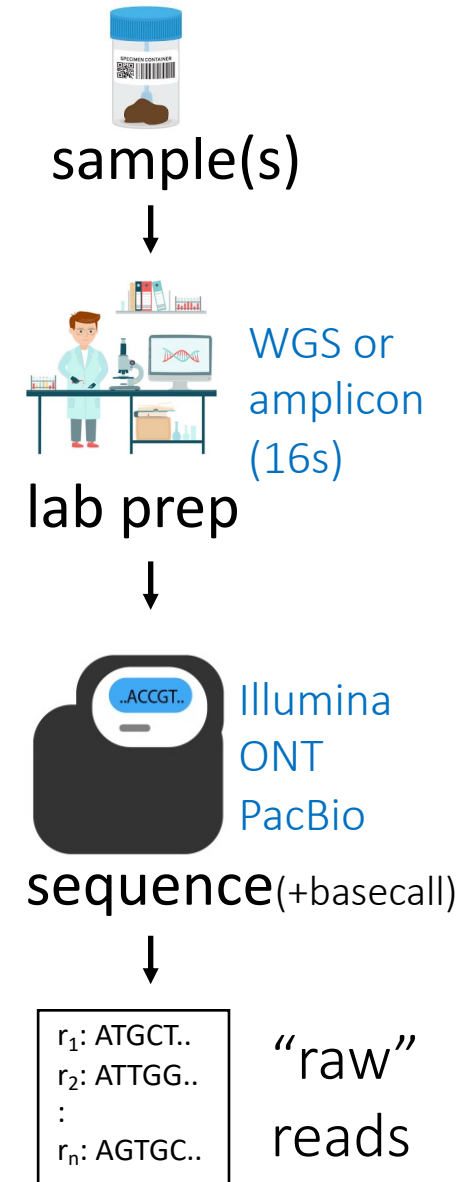
(also on Box: zymo\_hmw.tgz)

bioreactor contigs (Wednesday): /home/kdc10/datasets/bioreactor

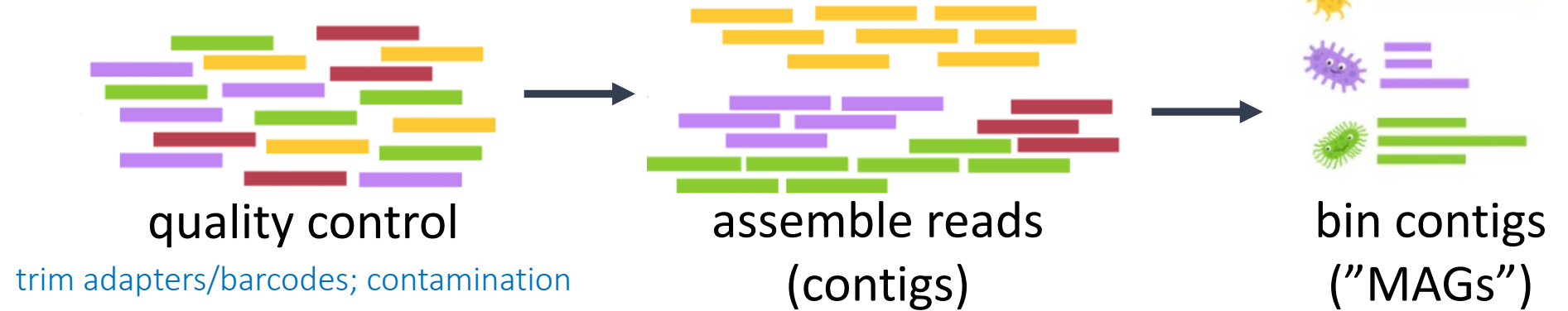
Vaginal 16S ONT (Friday): /home/kdc10/datasets/vaginal\_16S\_ONT

# Metagenomics pipeline

## 1. generate reads



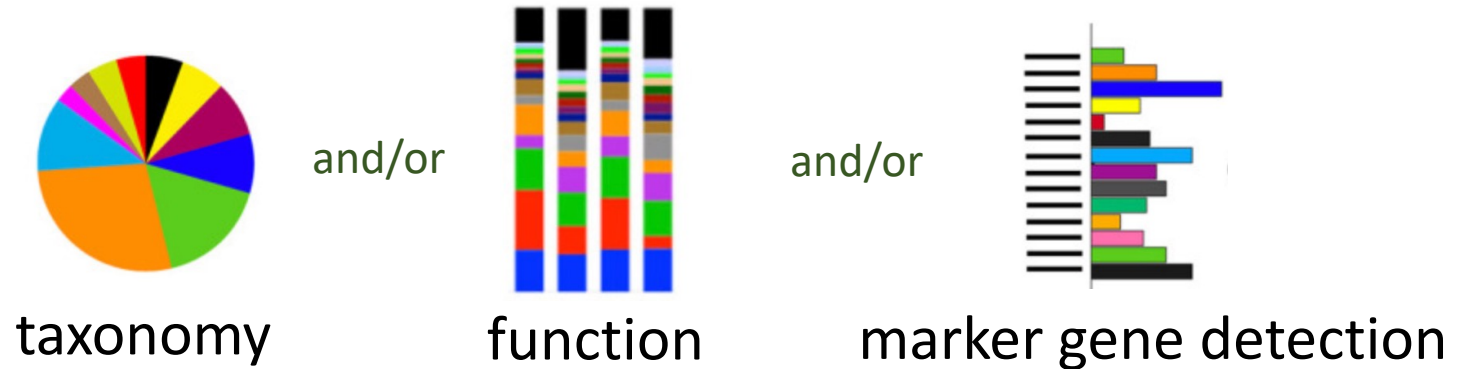
## 2. clean and sort



## 3. classify

(non-exhaustive list)

input:  
reads, contigs, or MAGs



## 4. stats

(non-exhaustive list)

