

# Evaluating assembly quality

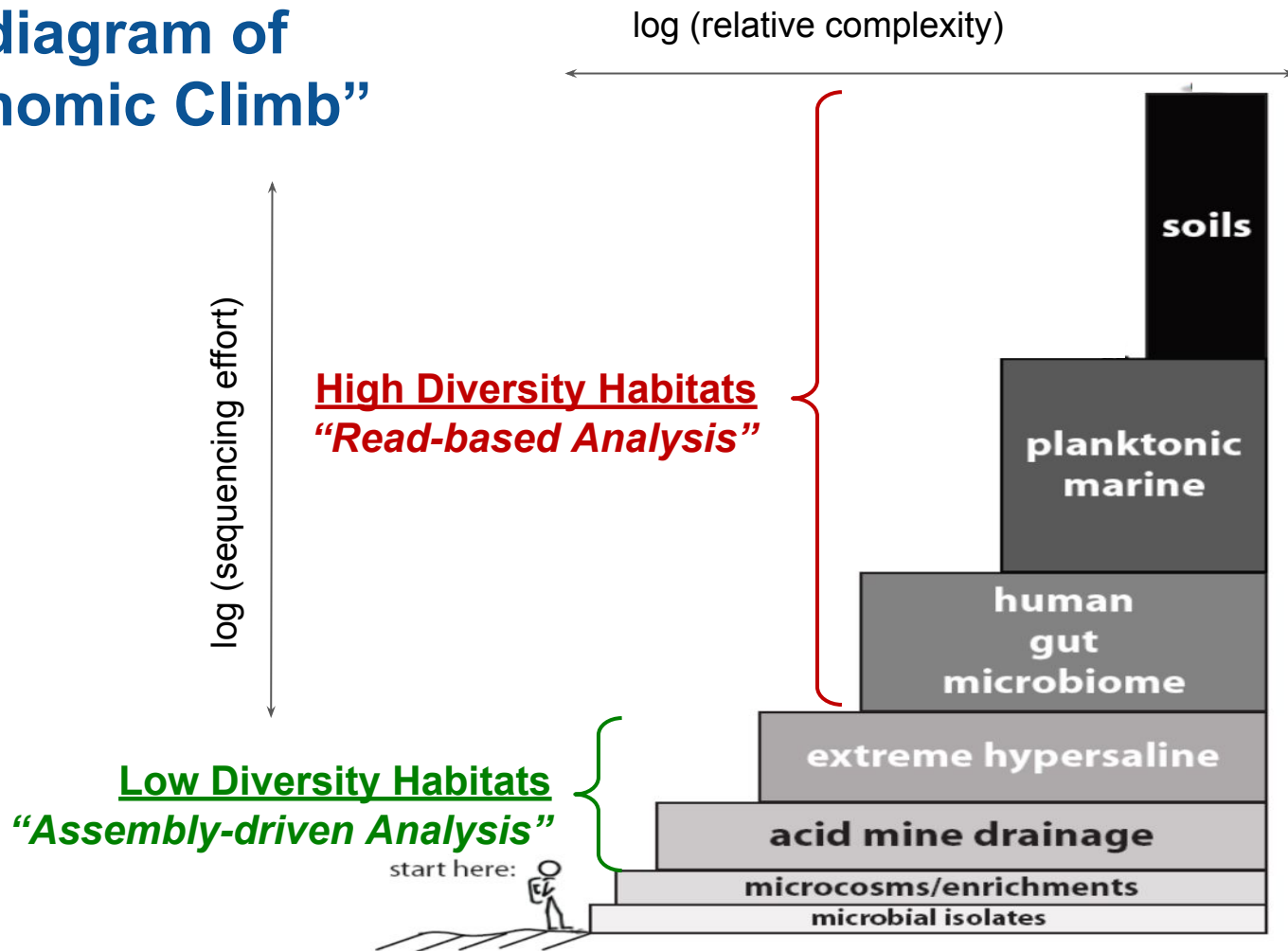
# How good is my assembly?

1. Is the assembler “correct” ?

2. Will the assembly answer your question?

- Functional Profiling - want high assembly rate
- Novel taxa, common or rare - need accurate genome binning
- Biosynthetic gene cluster finding - requires long sequences

# Eric Allen's diagram of the "Metagenomic Climb"



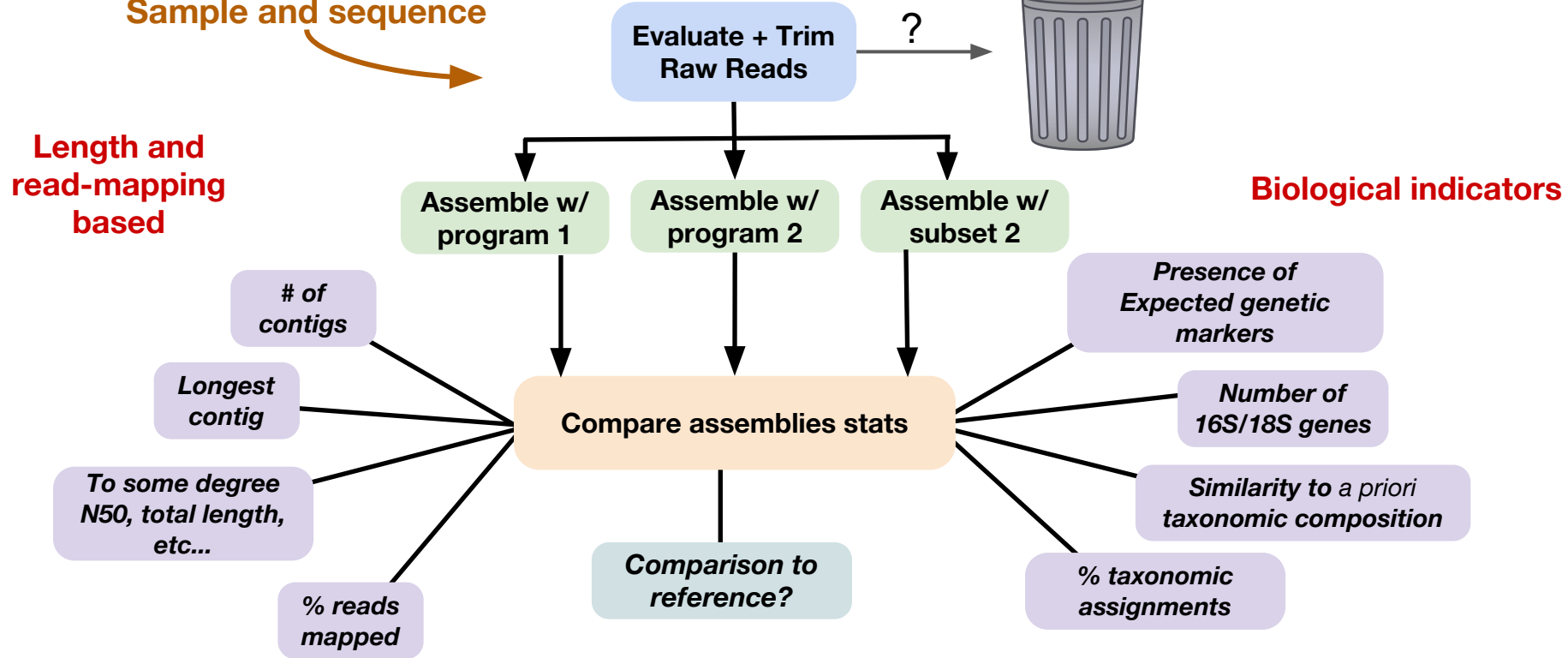
# How good is my assembly?

**-Or-**

**How well does the assembly represent the reads dataset that when into it?**

# Many different statistics to compare assemblies

Starting with biological question in hand...  
Sample and sequence



# Many metrics are “relative”

*Some may only tell you if one assembly is different (maybe better/worse) than another.*

# Some metrics are more important to improve than others

*Improving % reads is possibly the most important*

*But luckily, sequencing tech will only get better and cheaper!*

# Some great recent references:

## 1. **Critical Assessment of Metagenome Interpretation (CAMI) challenge**

- a. Comparison using mock communities from isolate

<http://www.biorxiv.org/content/biorxiv/early/2017/01/09/099127.full.pdf>

## 2. **Vollmer et al** “Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist’s Perspective - Not Only Size Matters!”

- a. Comparison using experimental (real) metagenomic data

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169662>

## 3. **Awad et al** “Evaluating Metagenome Assembly on a Simple Defined Community with Many Strain Variants”

<http://www.biorxiv.org/content/early/2017/06/25/155358>