Reproducibility Analysis

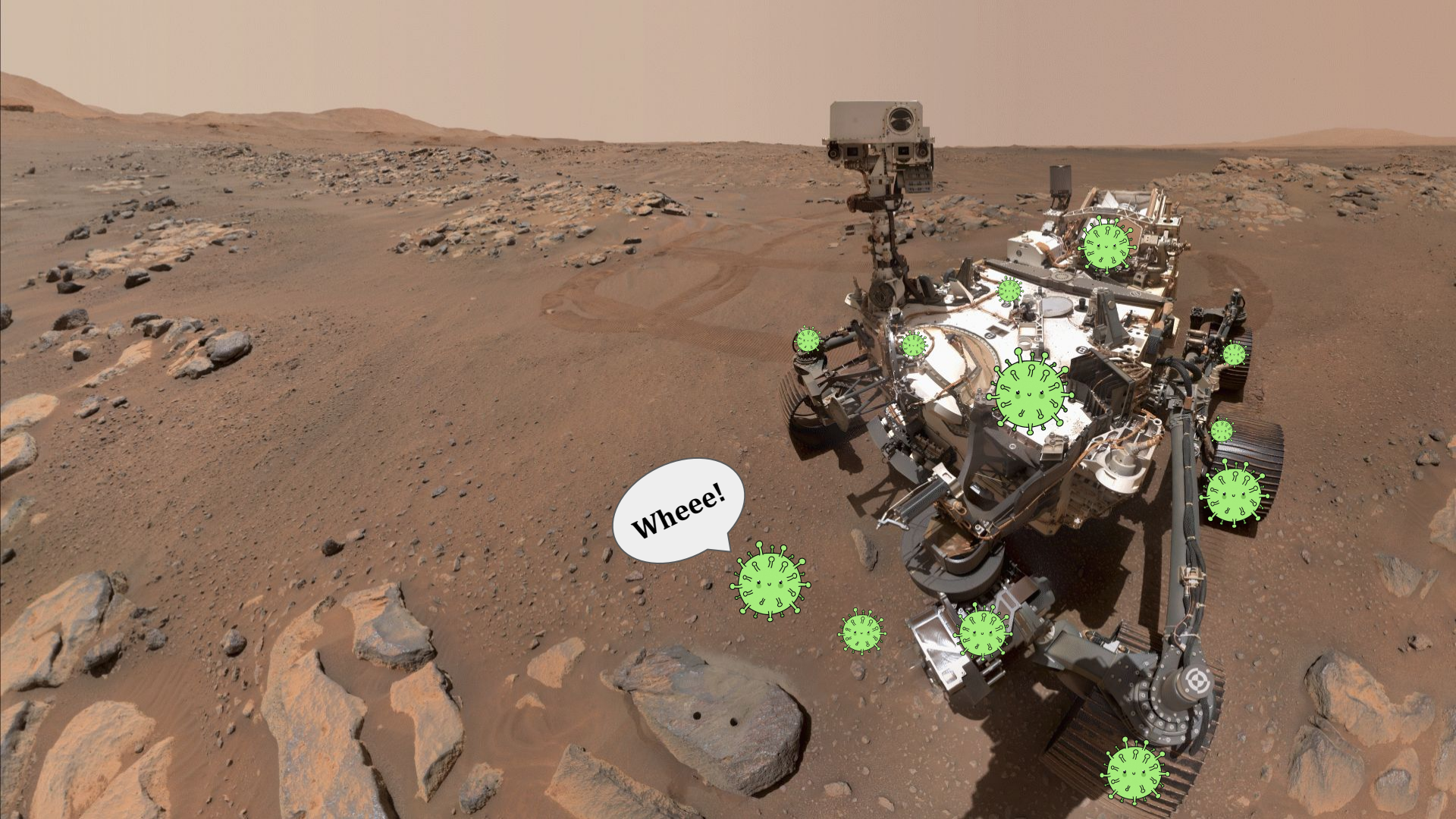
A comprehensive metagenomics framework to characterize organisms relevant for planetary protection

Annaliese Meyer and Matt Baldes
Environmental Bioinformatics
MIT-WHOI, 2021

RESEARCH Open Access

A comprehensive metagenomics framework to characterize organisms relevant for planetary protection

David C. Danko[1,2†], Maria A. Sierra[1,3†], James N. Benardini[4], Lisa Guan[4], Jason M. Wood[4], Nitin Singh[4], Arman Seuylemezian[4], Daniel J. Butler[2,3], Krista Ryon[1,3], Katerina Kuchin[2,3], Dmitry Meleshko[2,3], Chandrima Bhattacharya[2,3], Kasthuri J. Venkateswaran[4*] and Christopher E. Mason[1,3,5,6*]

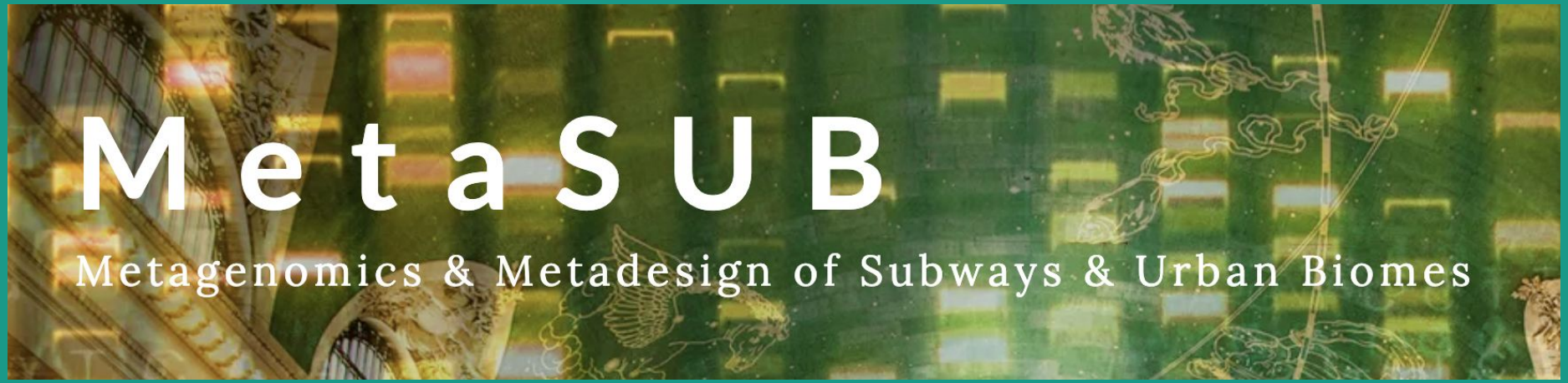

Wheee!

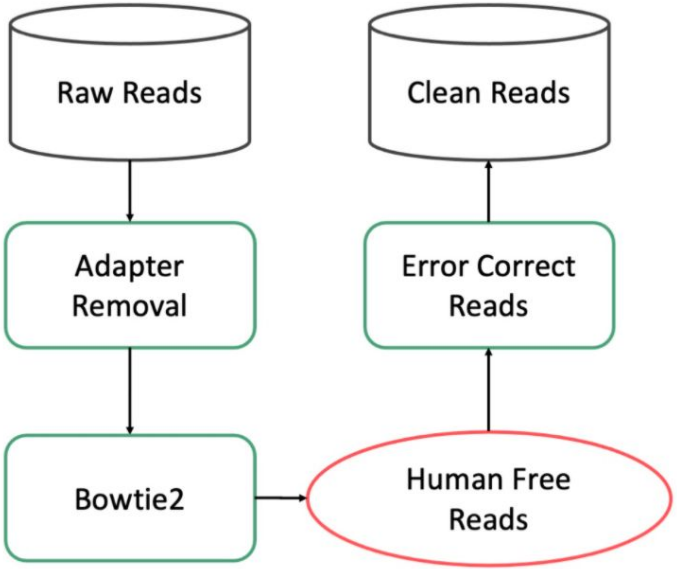
MetaSUB

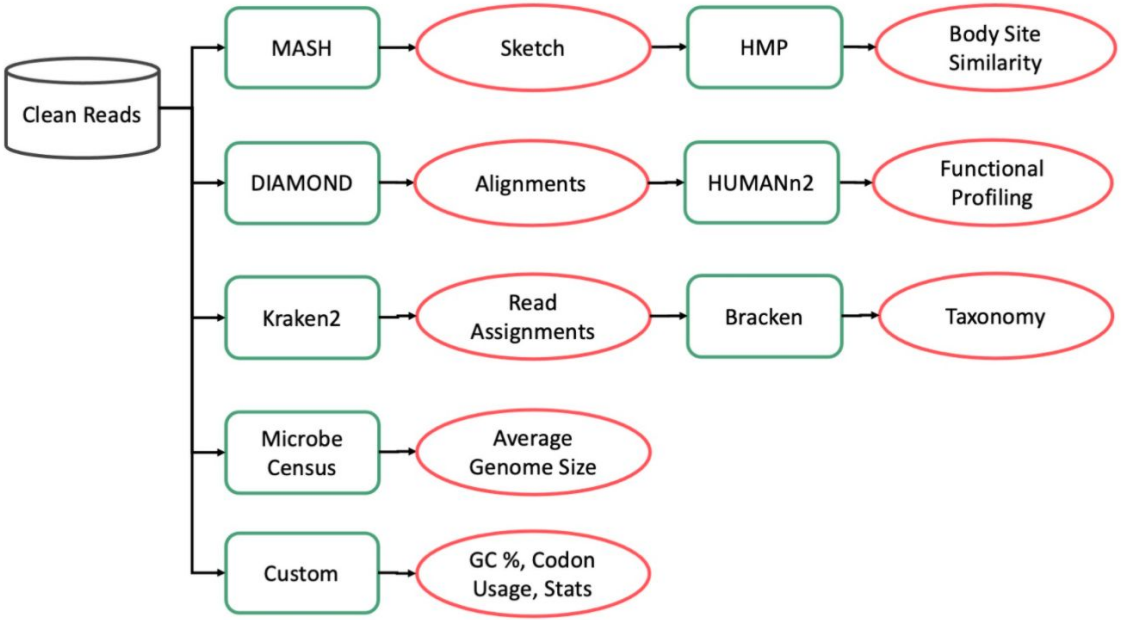
CAP2

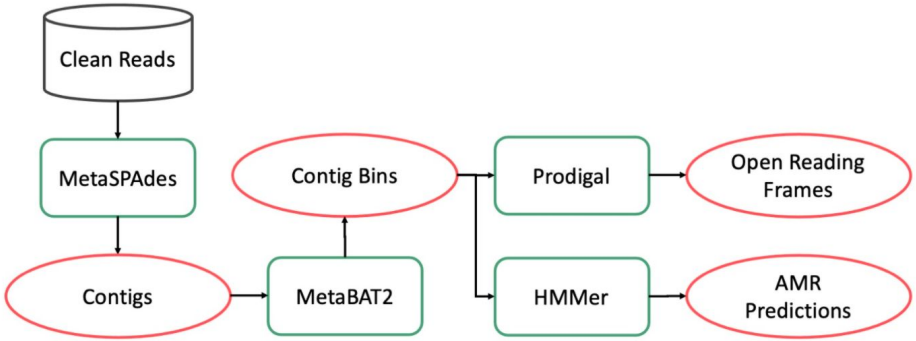
FAILED

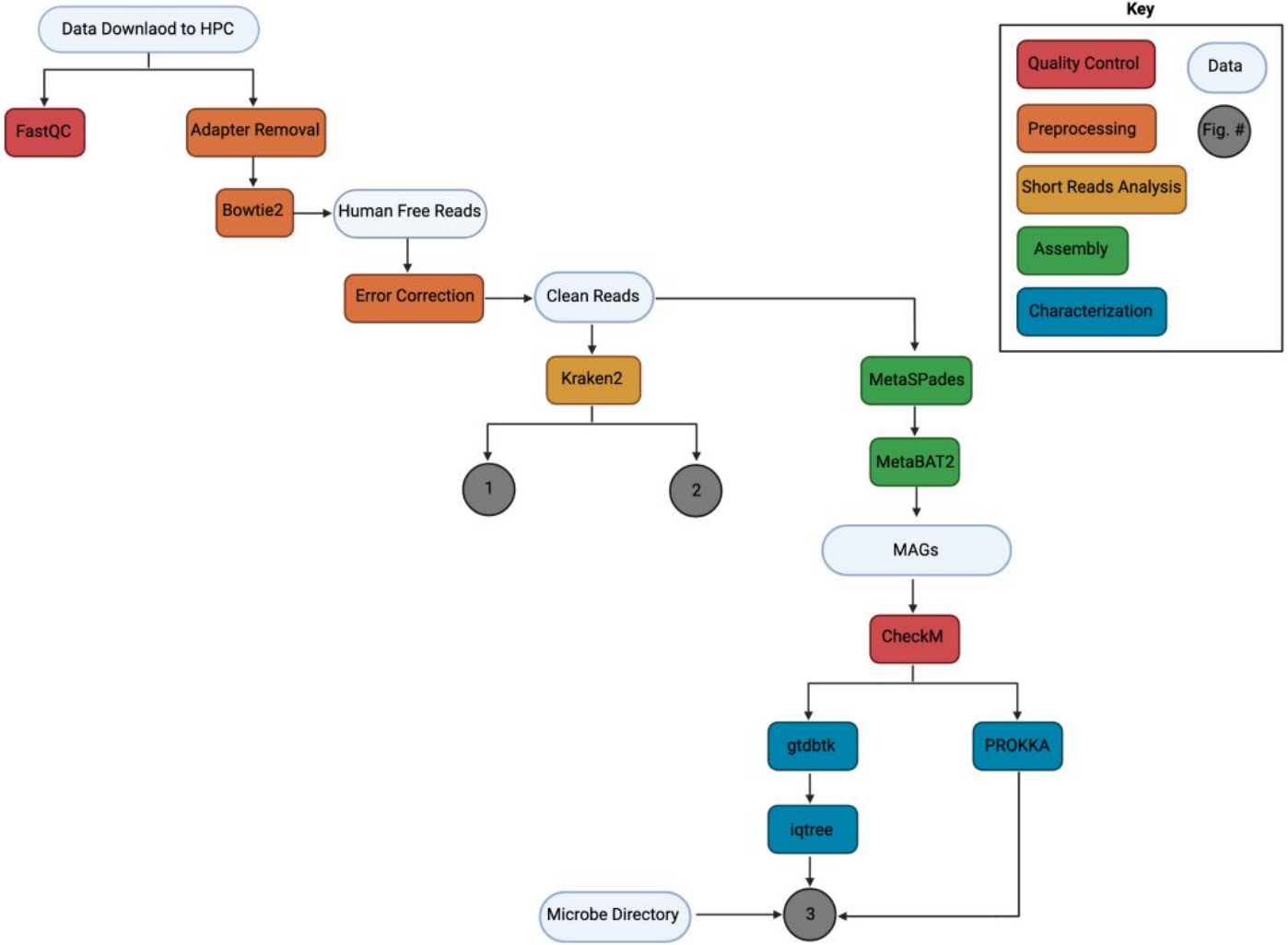
This repository contains code for the second version of the MetaSUB Core Analysis Pipeline. This pipeline is in Beta and under development.




Raw Reads
Adapter Removal
Bowtie2
Human Free Reads
Error Correct Reads
Clean Reads


Clean Reads
MASH
Sketch
HMP
Body Site Similarity
DIAMOND
Alignments
HUMANn2
Functional Profiling
Kraken2
Read Assignments
Bracken
Taxonomy
Microbe Census
Average Genome Size
Custom
GC %, Codon Usage, Stats

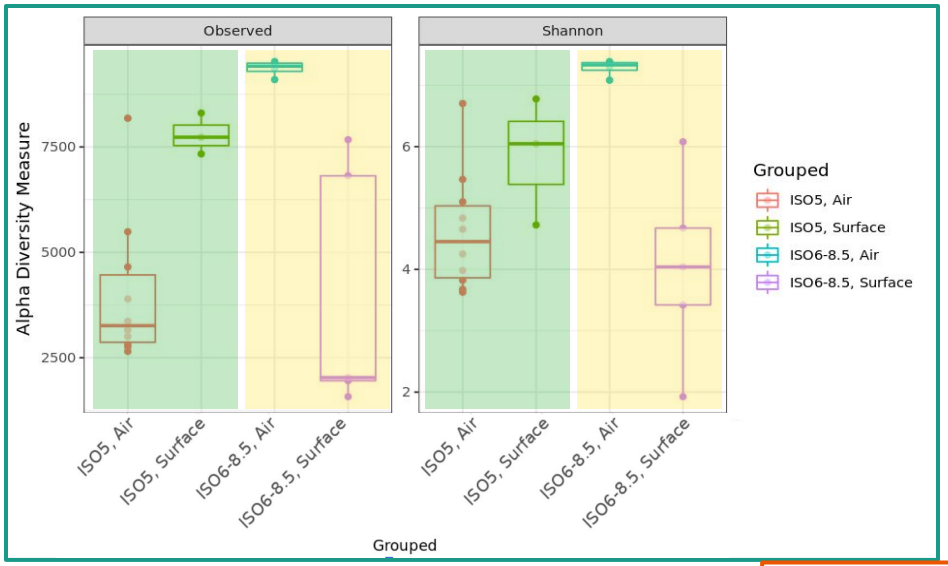

Clean Reads
MetaSPAdes
Contigs
MetaBAT2
Contig Bins
Prodigal
Open Reading Frames
HMMer
AMR Predictions


Data Downlaod to HPC
FastQC
Adapter Removal
Bowtie2
Human Free Reads
Error Correction
Clean Reads
Kraken2
1
2
MetaSPades
MetaBAT2
MAGs
CheckM
gtdbtk
PROKKA
iqtree
Microbe Directory
3
Key
Quality Control
Data
Preprocessing
Fig. #
Short Reads Analysis
Assembly
Characterization

ISO 5

ISO 6 ISO 7 ISO 8

Figures 1a and 1b



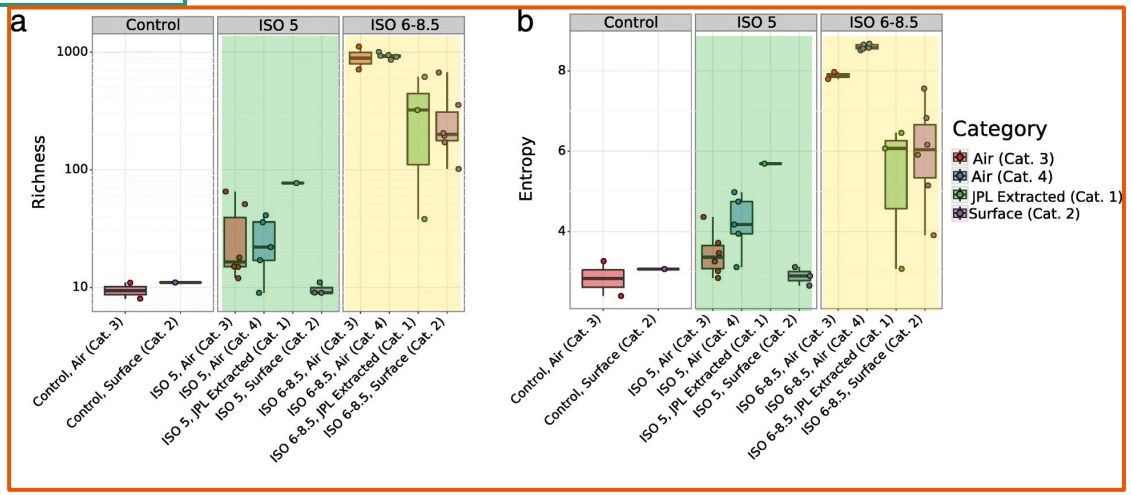
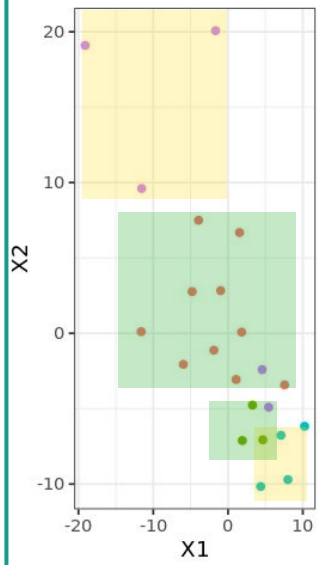
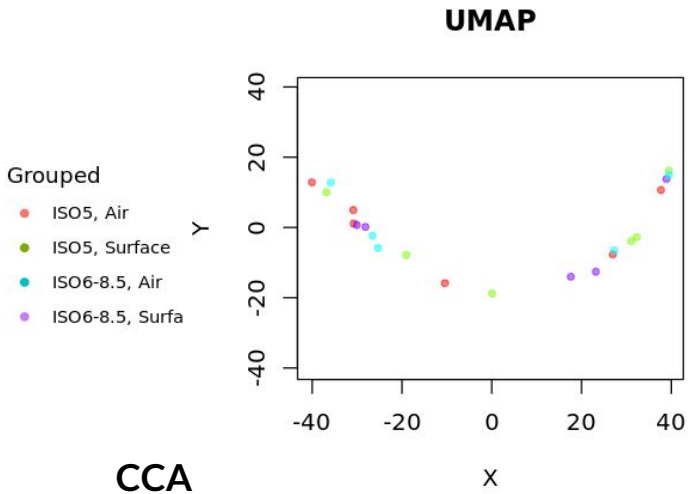
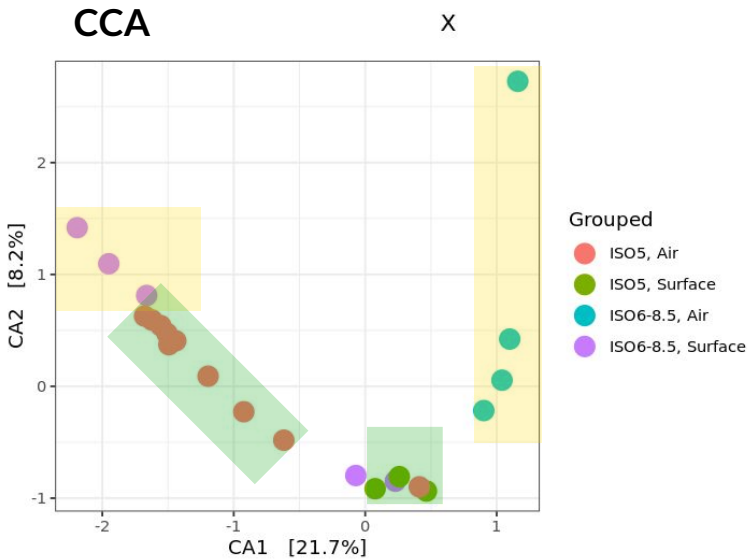


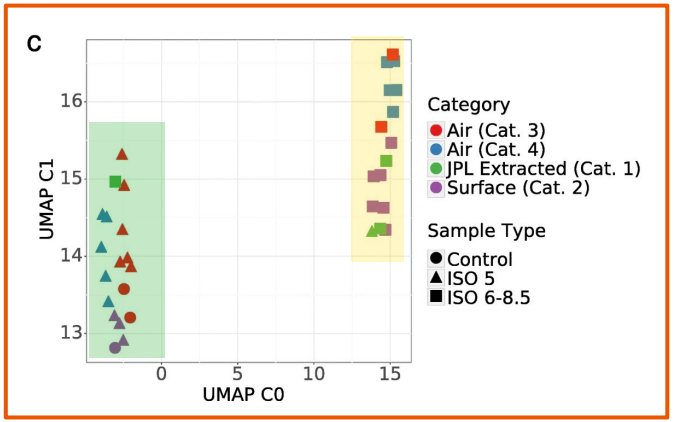
Figure 1c



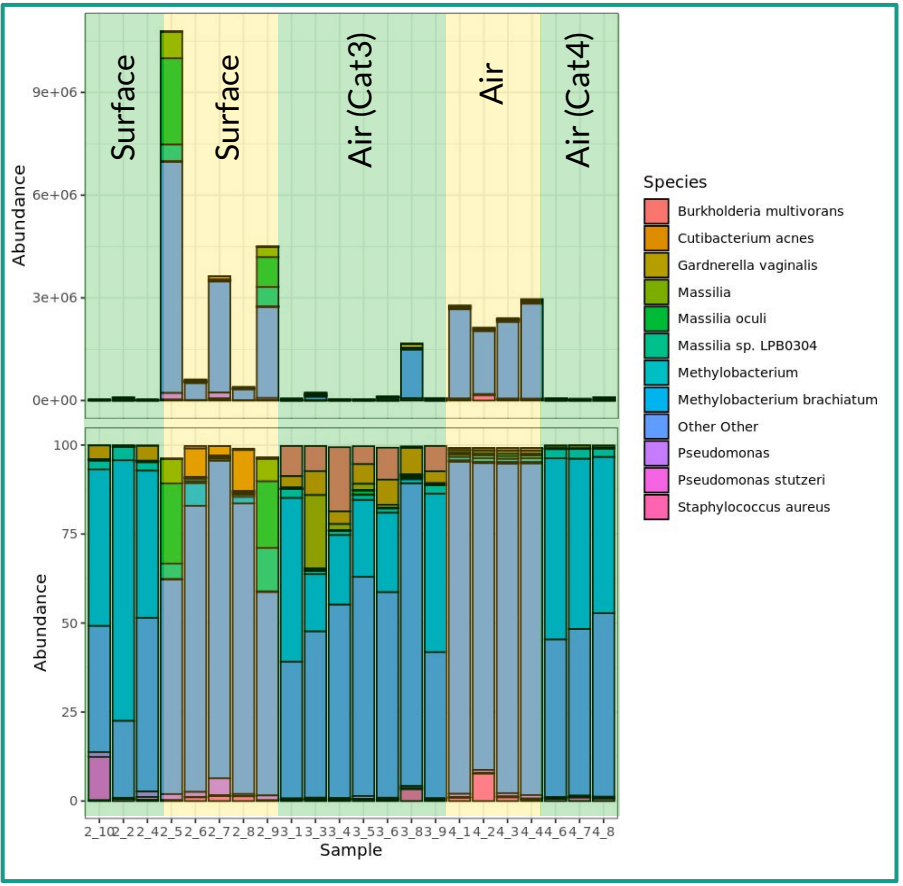
tSNE

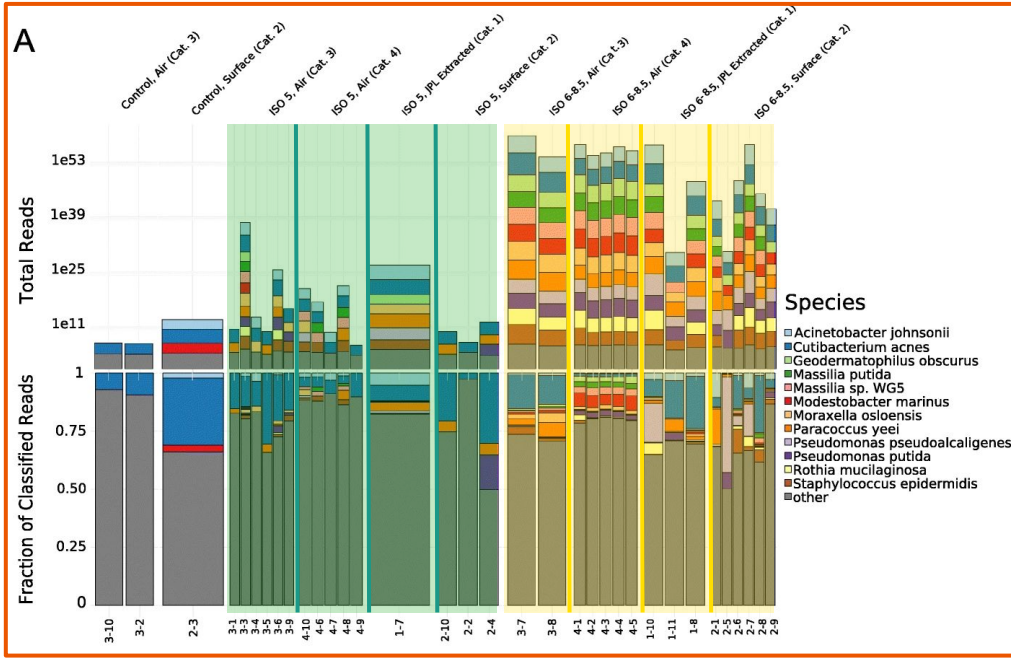


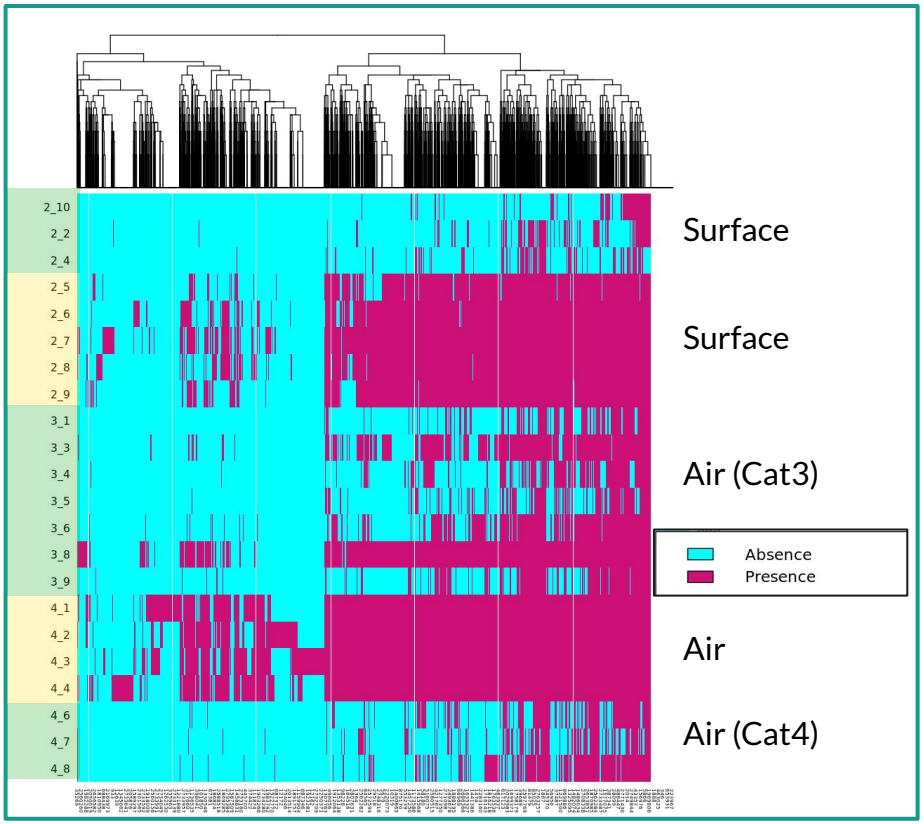


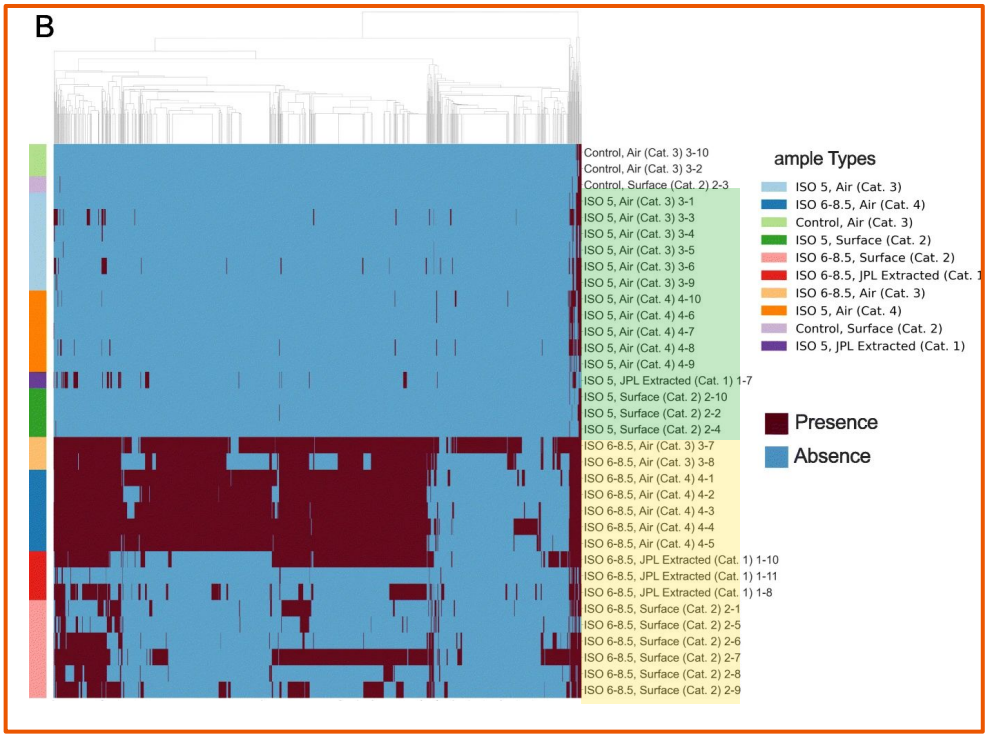


Note: Coloured boxes do not indicate statistically significant groupings


Surface
Surface
Air (Cat3)
Air
Air (Cat4)
Abundance
9e+06
6e+06
3e+06
0e+00
100
75
50
25
0
Sample
Species
Burkholderia multivorans
Cutibacterium acnes
Gardnerella vaginalis
Massilia
Massilia oculi
Massilia sp. LPB0304
Methylobacterium
Methylobacterium brachiatum
Other Other
Pseudomonas
Pseudomonas stutzeri
Staphylococcus aureus


A
Control, Air (Cat. 3)
Control, Surface (Cat. 2)
ISO 5, Air (Cat. 3)
ISO 5, Air (Cat. 4)
ISO 5, jPL Extracted (Cat. 1)
ISO 5, Surface (Cat. 2)
ISO 6-8.5, Air (Cat. 3)
ISO 6-8.5, Air (Cat. 4)
ISO 6-8.5, jPL Extracted (Cat. 1)
ISO 6-8.5, Surface (Cat. 2)
Total Reads
1e53
1e39
1e25
1e11
1
Fraction of Classified Reads
0.75
0.50
0.25
0
Species
Acinetobacter johnsonii
Cutibacterium acnes
Geodermatophilus obscurus
Massilia putida
Massilia sp. WG5
Modestobacter marinus
Moraxella osloensis
Paracoccus yeei
Pseudomonas pseudoalcaligenes
Pseudomonas putida
Rothia mucilaginosa
Staphylococcus epidermidis
other

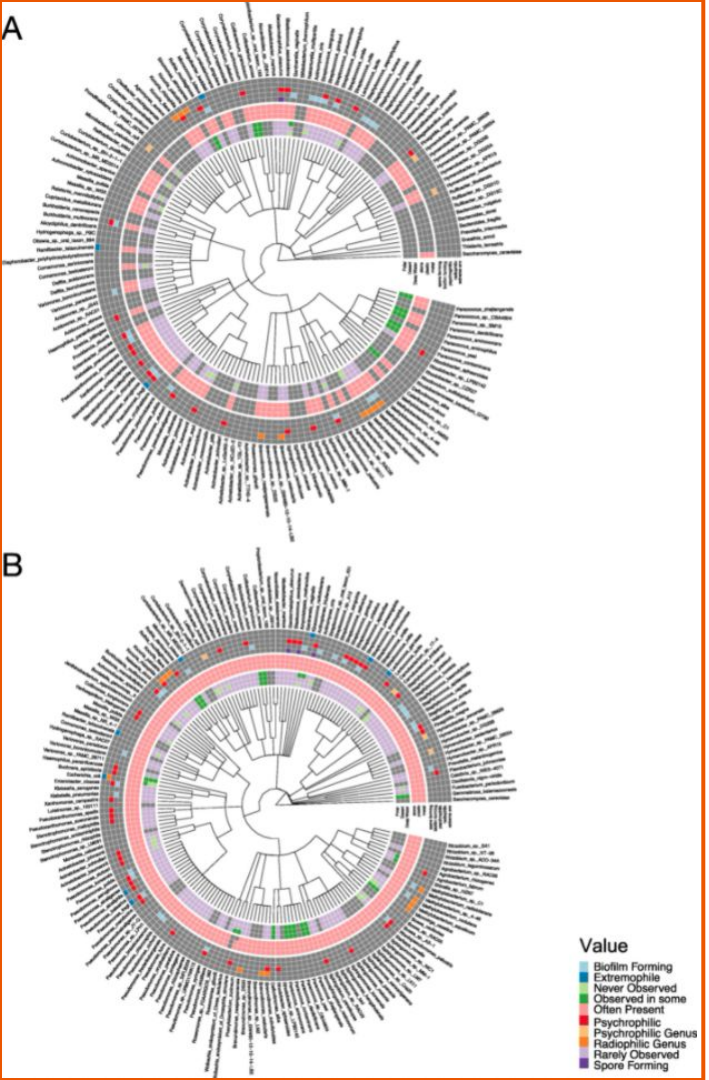

2_10
2_2
2_4
2_5
2_6
2_7
2_8
2_9
3_1
3_3
3_4
3_5
3_6
3_8
3_9
4_1
4_2
4_3
4_4
4_6
4_7
4_8
Surface
Surface
Air (Cat3)
Absence
Presence
Air
Air (Cat4)

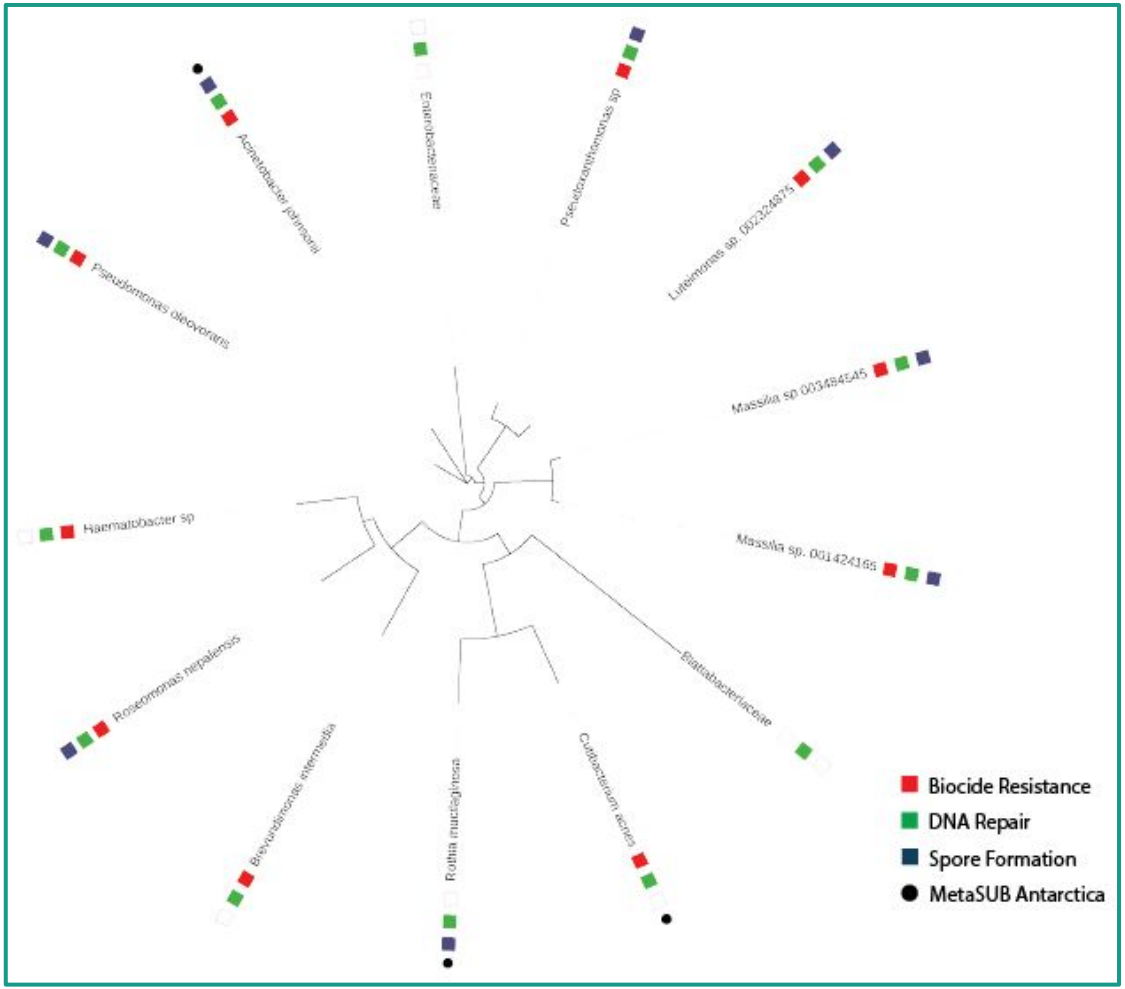

B
Control, Air (Cat. 3) 3-10
Control, Air (Cat. 3) 3-2
Control, Surface (Cat. 2) 2-3
ISO 5, Air (Cat. 3) 3-1
ISO 5, Air (Cat. 3) 3-3
ISO 5, Air (Cat. 3) 3-4
ISO 5, Air (Cat. 3) 3-5
ISO 5, Air (Cat. 3) 3-6
ISO 5, Air (Cat. 3) 3-9
ISO 5, Air (Cat. 4) 4-10
ISO 5, Air (Cat. 4) 4-6
ISO 5, Air (Cat. 4) 4-7
ISO 5, Air (Cat. 4) 4-8
ISO 5, Air (Cat. 4) 4-9
ISO 5, JPL Extracted (Cat. 1) 1-7
ISO 5, Surface (Cat. 2) 2-10
ISO 5, Surface (Cat. 2) 2-2
ISO 5, Surface (Cat. 2) 2-4
ISO 6-8.5, Air (Cat. 3) 3-7
ISO 6-8.5, Air (Cat. 3) 3-8
ISO 6-8.5, Air (Cat. 4) 4-1
ISO 6-8.5, Air (Cat. 4) 4-2
ISO 6-8.5, Air (Cat. 4) 4-3
ISO 6-8.5, Air (Cat. 4) 4-4
ISO 6-8.5, Air (Cat. 4) 4-5
ISO 6-8.5, JPL Extracted (Cat. 1) 1-10
ISO 6-8.5, JPL Extracted (Cat. 1) 1-11
ISO 6-8.5, JPL Extracted (Cat. 1) 1-8
ISO 6-8.5, Surface (Cat. 2) 2-1
ISO 6-8.5, Surface (Cat. 2) 2-5
ISO 6-8.5, Surface (Cat. 2) 2-6
ISO 6-8.5, Surface (Cat. 2) 2-7
ISO 6-8.5, Surface (Cat. 2) 2-8
ISO 6-8.5, Surface (Cat. 2) 2-9
ample Types
ISO 5, Air (Cat. 3)
ISO 6-8.5, Air (Cat. 4)
Control, Air (Cat. 3)
ISO 5, Surface (Cat. 2)
ISO 6-8.5, Surface (Cat. 2)
ISO 6-8.5, JPL Extracted (Cat. 1)
ISO 6-8.5, Air (Cat. 3)
ISO 5, Air (Cat. 4)
Control, Surface (Cat. 2)
ISO 5, JPL Extracted (Cat. 1)
Presence
Absence

The Microbe Directory: An annotated, searchable inventory of microbes' characteristics

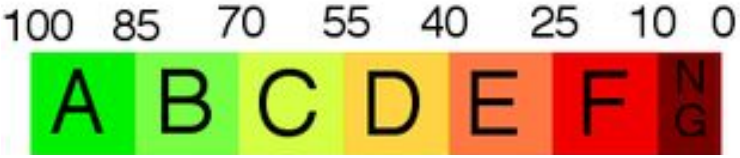
Heba Shaaban [#1,2,3], David A Westfall [#1,2,4], Rawhi Mohammad [1,2,5], David Danko [1,2], Daniela Bezdan [1,2], Ebrahim Afshinnekoo [1,2,6], Nicola Segata [7], Christopher E Mason [1,2,8]

Figure 3


A
B
Value
Biofilm Forming
Extremophile
Never Observed
Observed in some
Often Present
Psychrophilic
Psychrophilic Genus
Radiophilic Genus
Rarely Observed
Spore Forming


Acinetobacter johnsonii
Enterobacteriaceae
Pseudoxanthomonas sp
Luteimonas sp. 002324875
Massilia sp 003484545
Massilia sp. 001424165
Biatcabacteriaceae
Cutibacterium acnes
Rothia mucilaginosa
Brevundimonas intermedia
Roseomonas nepalensis
Haematobacter sp
Pseudomonas oleovorans
Biocide Resistance
DNA Repair
Spore Formation
MetaSUB Antarctica

Quantification of Reproducibility: Ease of use



Starting ‘grade’ of 100

Penalties:

- -1 per missing version number
- -2 per tool setting/flag not specified
- -5 per paid tool
- -8 per step/tool not outlined in manuscript or pipeline

Reproducibility Issues

- Non-standard repository
- Pipeline failed - both versions and many install parameters
    - Poorly designed custom workflow manager (v1)
- No version numbers specified outside of the source code
- No flags specified outside of the source code except in an isolated case
- Kmer length not specified
- Microbe Directory datasets not specified
- No figure documentation

27

(generously)