

Dr. Richard Allen White III
Washington State University
FABI Workshop - Oct 25, 2018
Day 2

Learning Objectives

1. Review omics, data formats and assembly terms
2. Awk tutorial for text wrangling
3. Core and Pan genome analysis
4. Mercat DIPA analysis
5. Decontaminate reads via mapping
6. Trim reads via Trimmomatic
7. Assemble good reads with Spades

Omics terms Review



1. Metagenome?
2. Meta-metabolomics or community metabolomics?
3. Genomics?
4. Metaproteomics?
5. Multiomics?
6. Special bonus -> Metaepigenomics?

Data format - Review

*.fna, .fasta, .fa ?

*.faa ?

*.gff ?

*.gbk ?

*.fq or fastq ?

*.gtf ?

*.sam ?

*.bam ?

How obtain a genome?

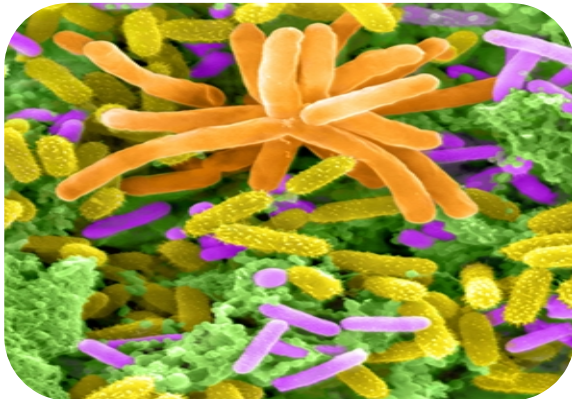


Culturing $< 1\%$ can be easily cultured



Single cell genomics

Only a few labs in the world, very incomplete genomes (30% avg, 10-90%)



Metagenomic - population genome binning

Why do we need complete genomes in microbial ecology?



- Functional studies demand an error-free genome sequence as a starting point¹
- Availability of data on genome organization provides biological insights¹
- Comparative genomics is meaningful only in terms of complete genome sequences¹
- Microbial forensics requires at least one complete reference genome sequence¹
- A complete genome sequence is a permanent, valuable scientific resource¹

Assembly term - Review

- Read vs. sequence?
- Contig?
- kmer?
- Scaffold?
- Coverage?
- Mapping?

Assembly term - Review



k-mer (4-mer)

ATTACCGG

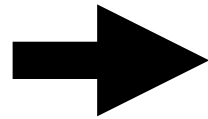
TTTTCCGG

GGGCCGG

TTTAATTA

Reads

***De novo*
assembly**



TTTTCCGGGGGGCCGG

TTTAATTACCGG

Contigs

TTTTCCGGGGGGCCGG
TTTTCCGG

**Mapping
(Coverage)**



TTTTCCGGGGGGCCGG
NNNNTTTAATTACCGG

Scaffolds

Command line review



grep ?

sed?

awk?

cat?

echo?

Printf?

For loop?

While loop?

Unix/Bash - check point review

- Make a folder with your name
- Make two files in as .txt (labeled 1 and 2)
- Write your name in file 1
- Write your name in file 2 with _a at the end
- Combine those two files to a new file
- Copy the whole folder of your name with a new name

ATLAS (Automatic Tool for Local Assembly Structures) architecture



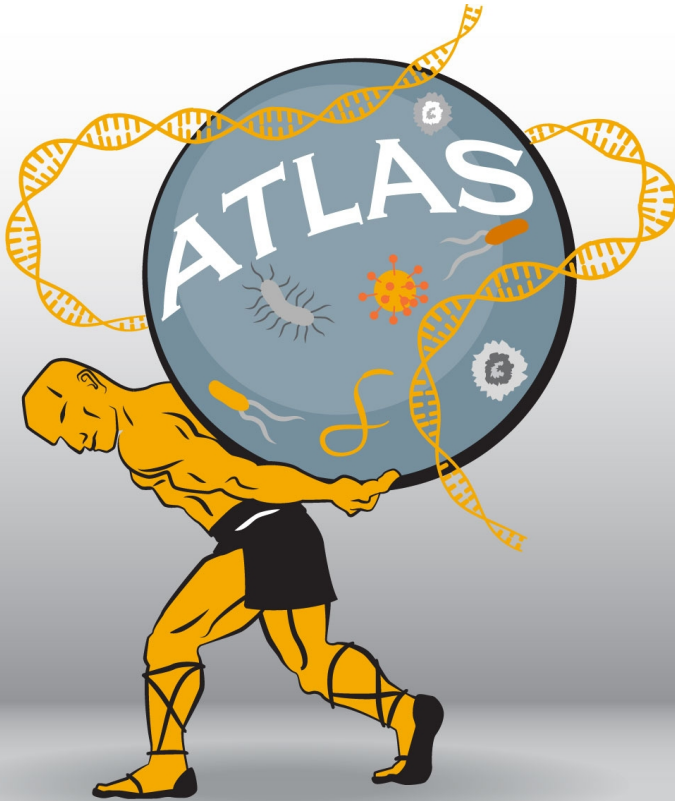
White III et al., 2017 PeerJ

github.com/raw937/atlas

- Open sourced - freely on github
- Easy to install with python3/Bioconda
- Little computational experience needed (great tutorials)
- High scaling (>100 gbp)
- Highly parallel (>1000 samples)
- Many tools integrated/custom workflows (snakemake)
- Standardization/reproducibility (docker)
- Long term updates/sustainability (ongoing)
- Free customer support/install
- ~50% faster then IMP (96 Gbp soil meta)

ATLAS steps and workflow

(Step 1 - 5)



Illumina paired end data

- Step 1 – Quality control
- Step 2 – Assembly
- Step 3 – Annotation
- Step 4 – Genomic Binning
- Step 5 – Quantification

Awk tutorial

https://github.com/raw937/FBI_workshop/blob/master/bin/day4/awk_tutorial.md

Pan and Core genome tutorial

https://github.com/raw937/FABl_workshop/blob/master/bin/day4/Core-pangenome_analysis.md

Mercat tutorial

https://github.com/raw937/FABl_workshop/blob/master/bin/day4/mercato_dipa.md

Decon tutorial

https://github.com/raw937/FABl_workshop/blob/master/bin/day4/decon.md

Trim tutorial

https://github.com/raw937/FABl_workshop/blob/master/bin/day4/trim.md

Assembly and Annotation tutorial

https://github.com/raw937/FABl_workshop/blob/master/bin/day4/assembly-annotation.md

Git tutorial

Make an account

<https://github.com/>

Follow tutorial on

https://github.com/raw937/FABI_workshop/edit/master/bin/git%20tutorial.md