



Tutorial 8 : Metagenome De Novo Assembly and Binning using Nextflow nf-core/mag

Zoe Sucato, Alix Foster, Lane Lerner



Overview

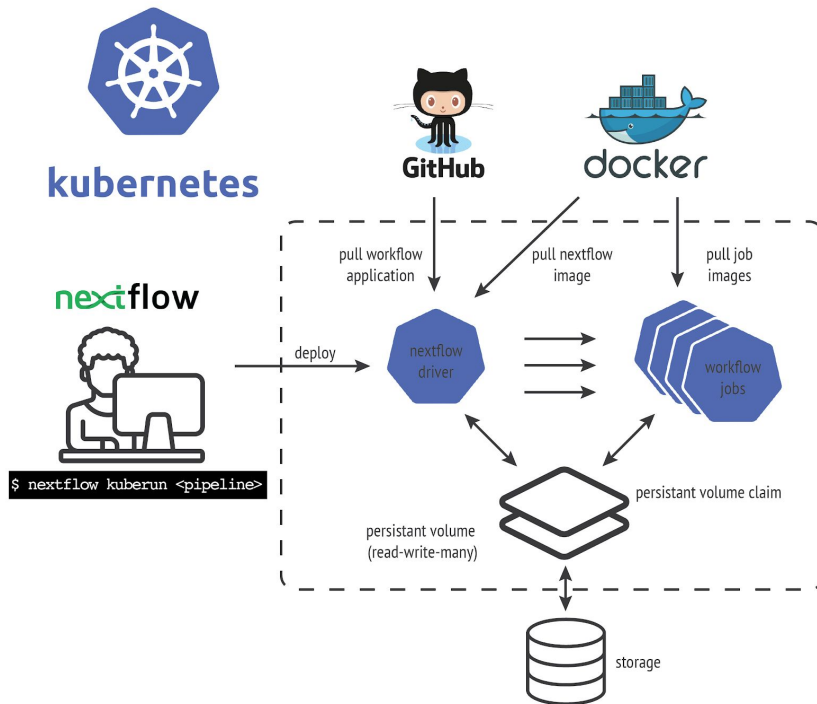
1. What is Nextflow and nf-core/mag?
2. How do we run the pipeline?
3. What can we learn from the results?



**What is Nextflow
and nf-core/mag?**

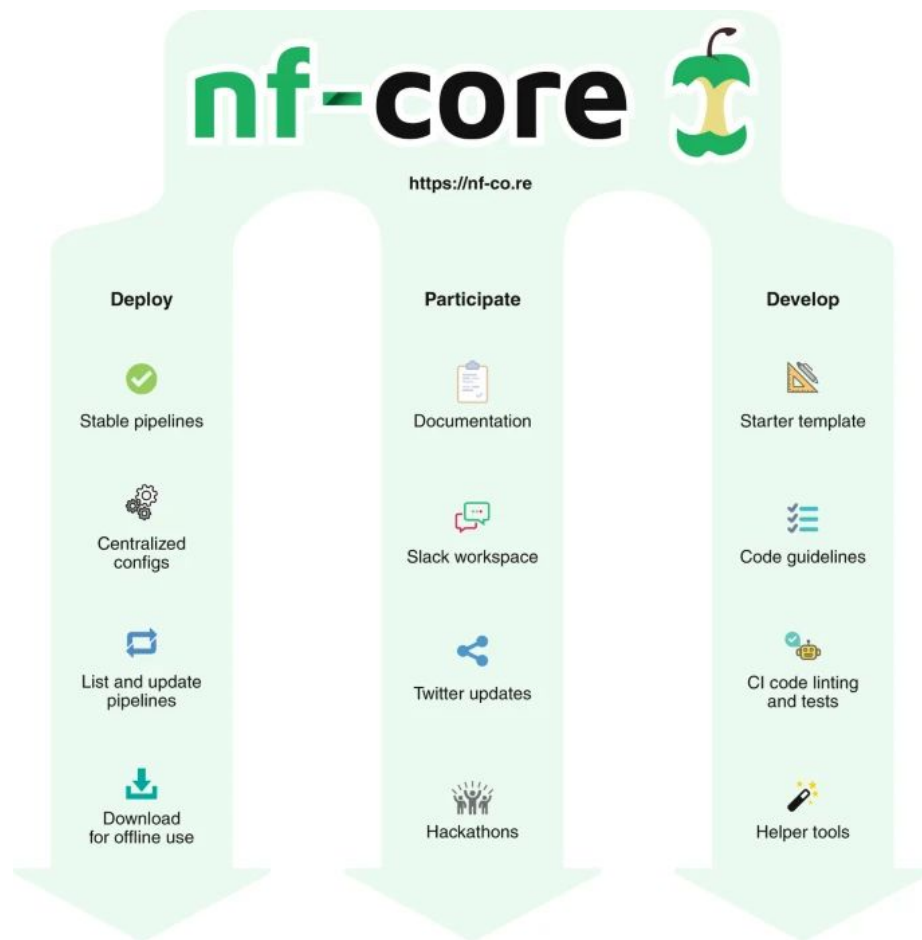
NextFlow Background

- Fast prototyping
- Reproducibility
- Portable
- Unified parallelism
- Continuous checkpoints
- Stream oriented



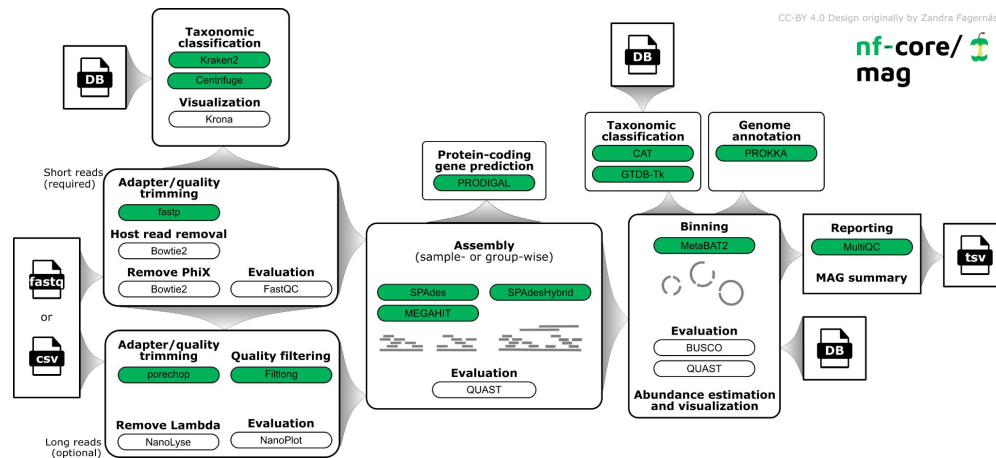
Nf-core Background

- Built using Nextflow
- Community supported
 - Curation and best-practice analysis pipelines
- Goal is to provide analysis pipelines that are fully reproducible and interoperable across many systems and institutions despite differences in hardware, operating systems, and software



Nf-core/mag Background

- Assigns Taxonomy
 - Centrifuge
 - Kraken2
- Assembly
 - MEGAHIT
 - SPAdes
 - Quast - quality check
- Protein-coding gene prediction
 - Prodigal
- Metagenome binning
 - MetaBAT2
 - Busco - quality check
- Assign Taxonomy to bins
 - GTDB-Tk
 - CAT





**How do we run
nf-core/mag?**


nf-core/mag requirements



- Pre-installed softwares
 - Nextflow
 - Singularity
- Working directory
 - 750+ GB
- RAM requirements
 - 500+ GB
- CPU requirements
 - 40+ CPUs
- Time
 - 38+ hours

Run in the `/BeeGFS/scratch` folder, using a big memory node partition, executed well in advance.

To run

- 
- Verify pre-installed softwares are available
 - Prepare data location
 - `/ifs/groups/eces450650Grp/ECES450650_SP22/data/CAMISIM_hybrid/`
 - Create directory in BeeGFS
 - `/BeeGFS/scratch/zas37; chmod 700`
 - Set up folder to store Singularity images
 - Create job script and custom configuration file

Job script



```
#!/bin/bash
#
#SBATCH --mail-user=zas37@drexel.edu
#SBATCH --nodes=1
#SBATCH --cpus-per-task=48
#SBATCH --time=48:00:00
#SBATCH --mem=1000GB
#SBATCH --partition=bm

module load nextflow
export NXF_SINGULARITY_CACHEDIR=/ifs/groups/eces450650Grp/ECES450650_SP22/zas37/singularity/cache
nextflow run nf-core/mag -r 2.1.0 -profile singularity
-c /beegfs/scratch/zas37/tutorial18/custom.config
--input /ifs/groups/eces450650Grp/ECES450650_SP22/data/CAMISIM_hybrid/samplesheet.CAMISIM_hybrid.csv
--outdir /ifs/groups/eces450650Grp/ECES450650_SP22/zas37/results --coassemble_group
--binning_map_mode all --max_cpus 48 --max_memory 1000.GB --spades_fix_cpus 40
--spadeshybrid_fix_cpus 40 --skip_megahit
```

Config file

```
process {

    withName: SPADES {

        memory = 500.GB

    }

    withName: SPADESHYBRID {

        memory = 500.GB

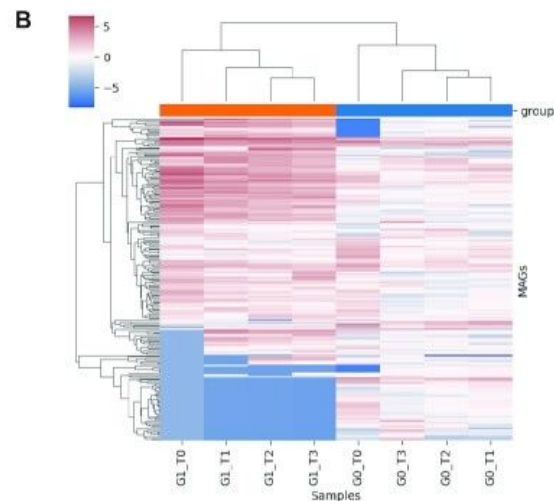
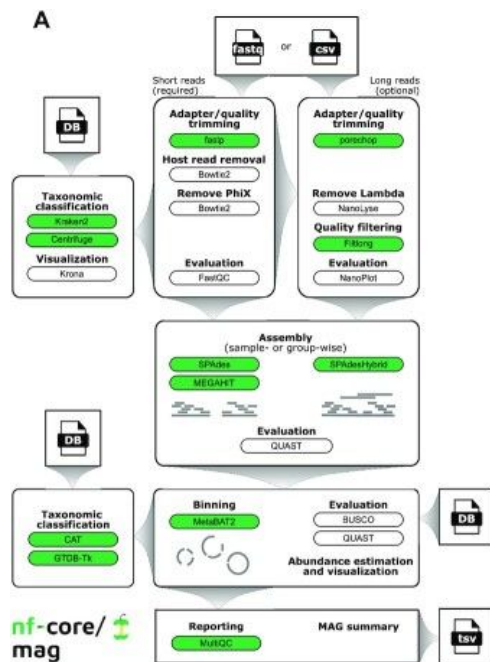
    }

}
```



**What can we learn
from the results?**

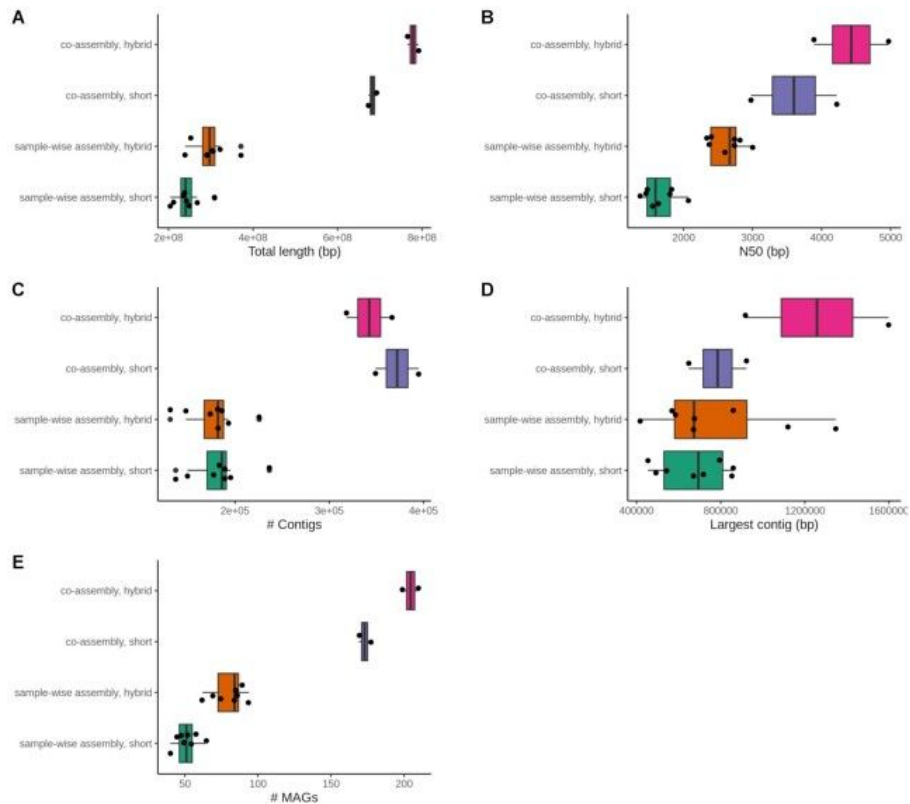
Results - Pipeline Overview



C

	Abundances	BUSCO results		QUAST results		GTDB-TK results	
bin	Depth S1 ...	%Complete	%Missing ...	# contigs	Total length ...	classification	
SPAdes-S1.1.fa	7.80	67.7	31.5	511	1411805	d__Bacteria;...;s__Lachnospira eligens_A	...
SPAdes-S1.2.fa	1.47	96.8	0.8	79	2350285	d__Bacteria;...;s__Sutterella wadsworthensis	
SPAdes-S1.3.fa	0.21	2.1	97.4	435	1120996	NA	
...							

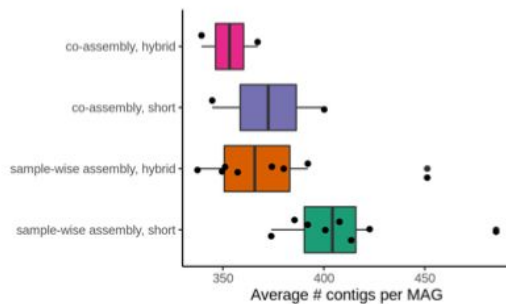
Results



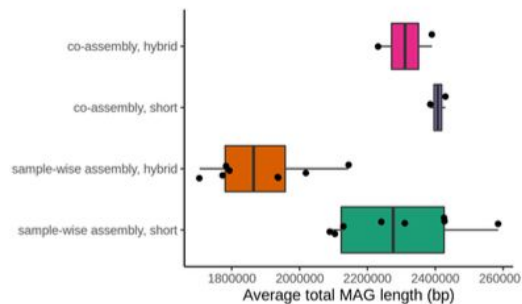
Results



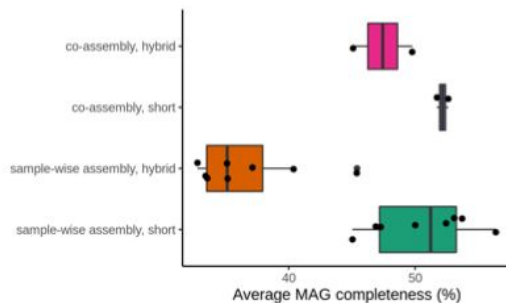
a)



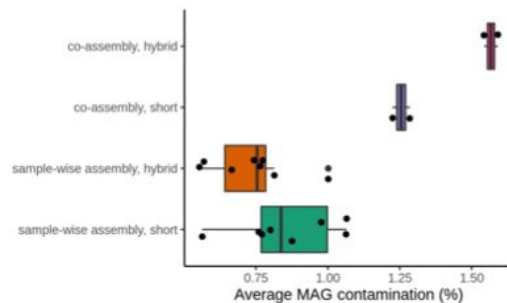
b)



c)



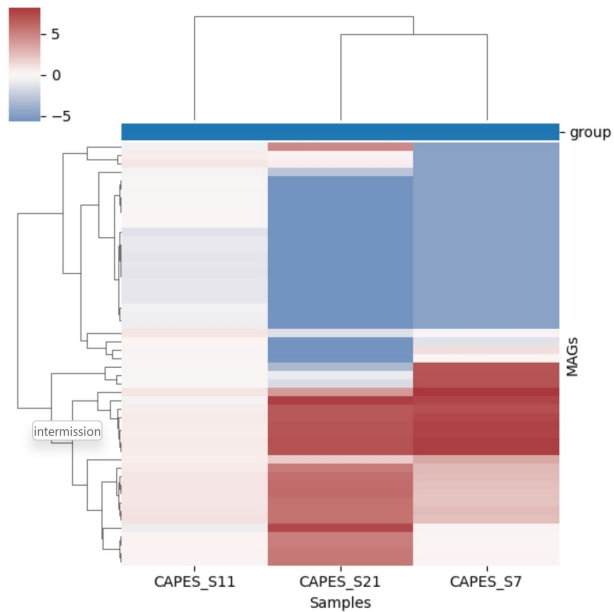
d)



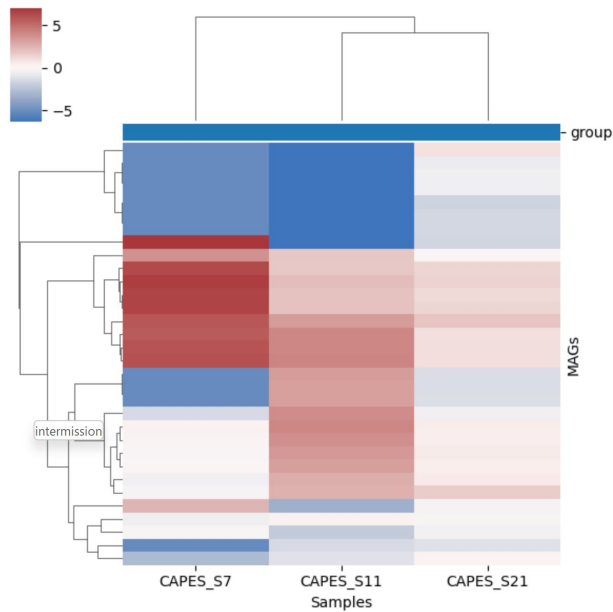
SPAdes Assembly Bin Heatmap



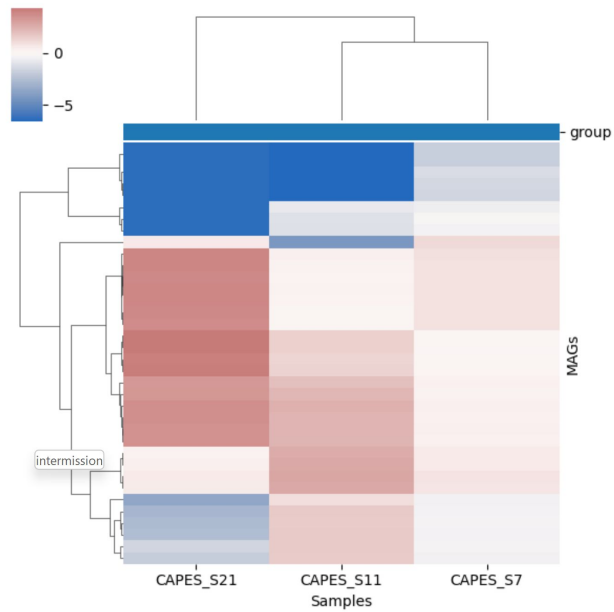
S11



S21



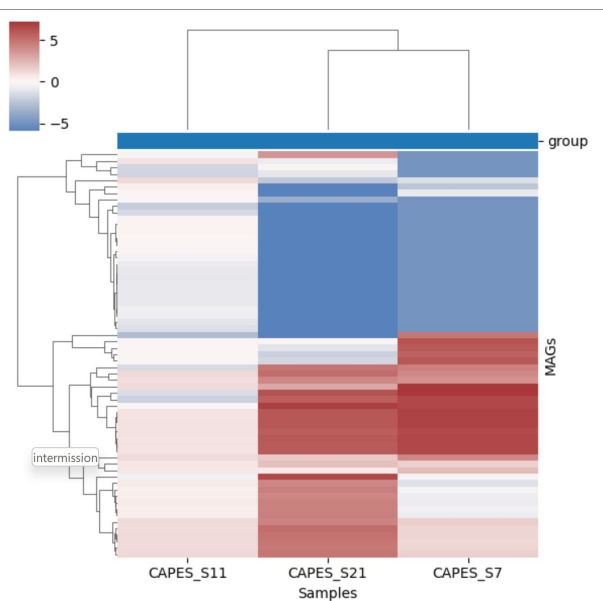
S7



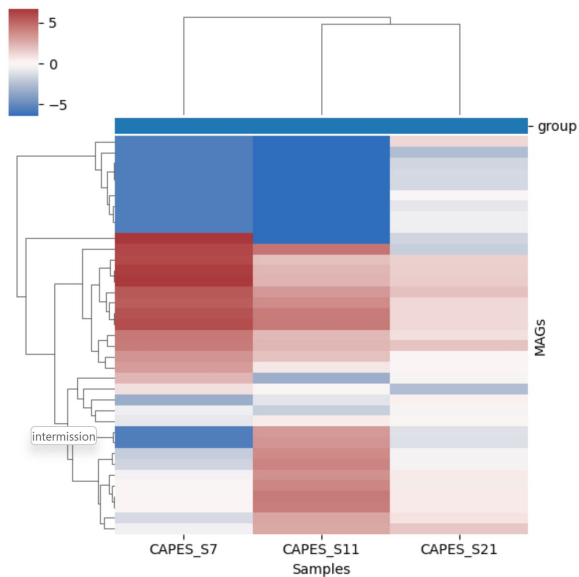
SPAdesHybrid Assembly Bin Heatmap



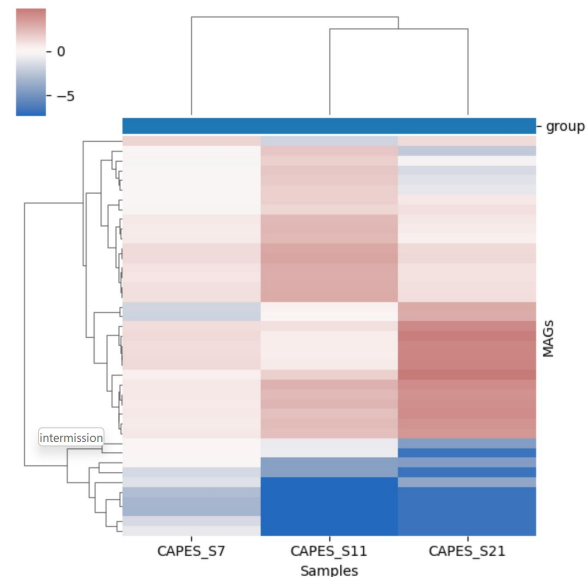
S11



S21



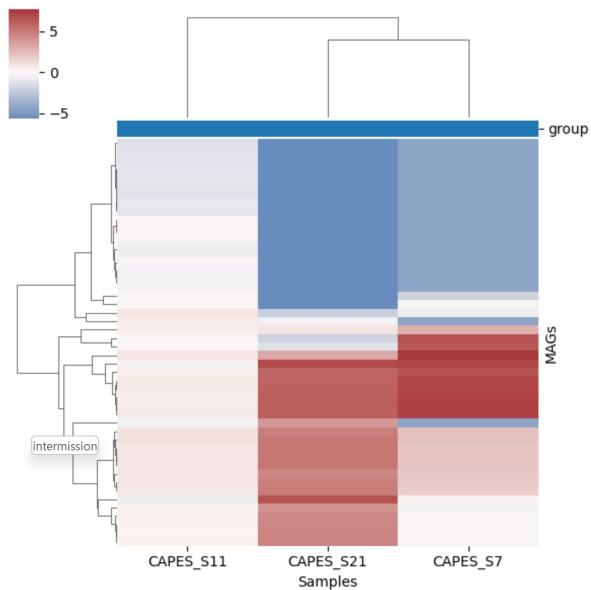
S7



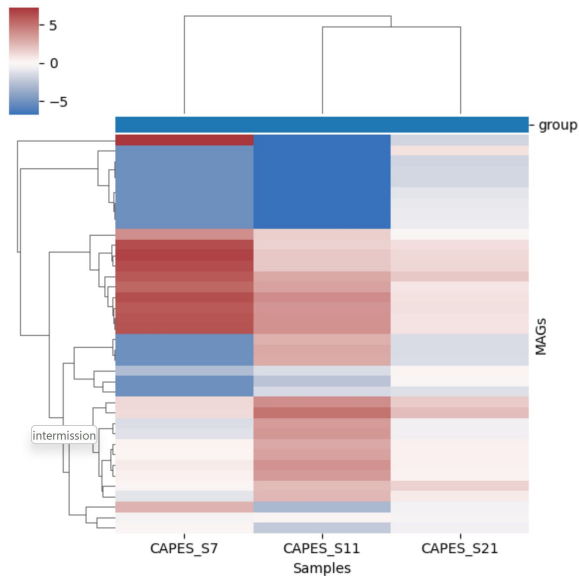
MEGAHIT Assembly Bin Heatmap



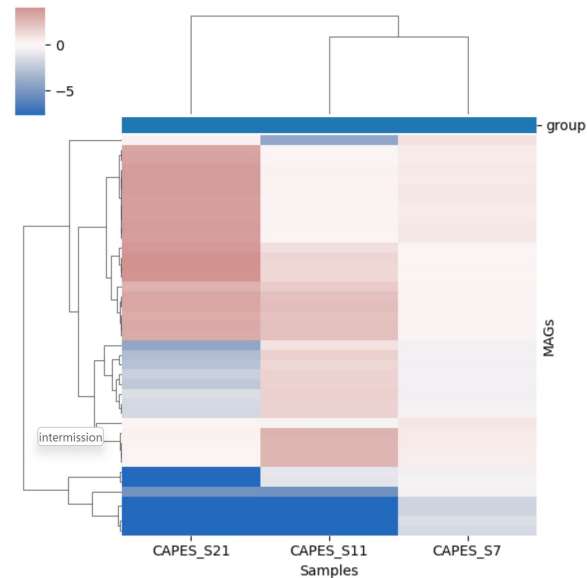
S11



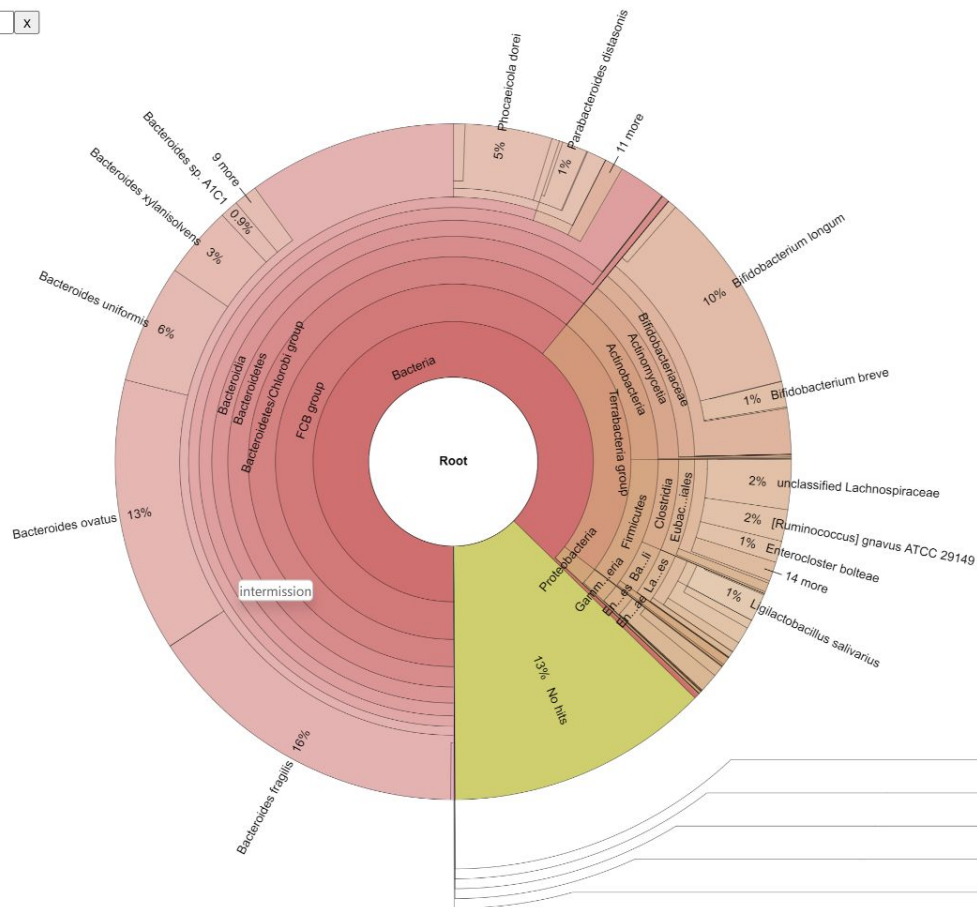
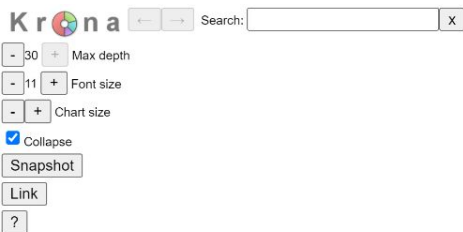
S21



S7



Taxonomy Krona - Kraken S21



Root 

Count: 21288326

Unassigned: 13921



Taxonomy Krona - Centrifuge S21

Krona Search: X

- 30 + Max depth

- 11 + Font size

- + Chart size

☒ Collapse

Snapshot

Link

?

Root

Count: 22589520

Unassigned: 11440



X

Homo 0.08%

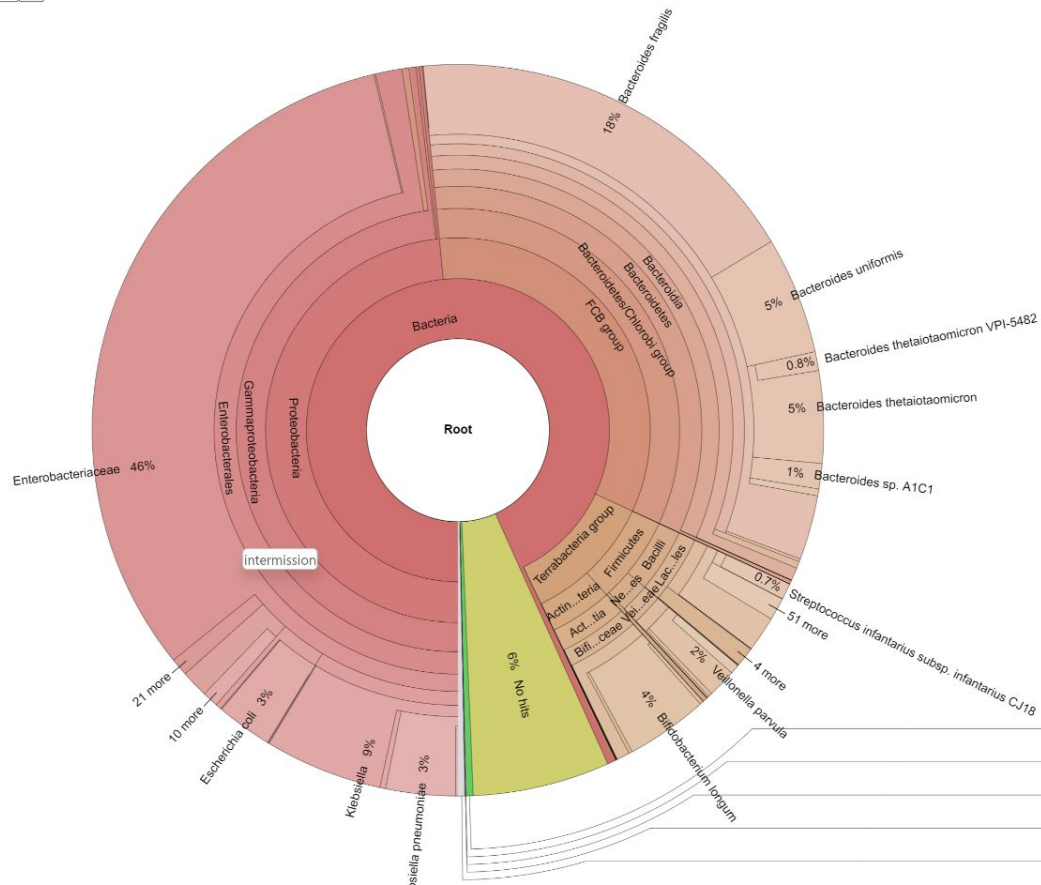
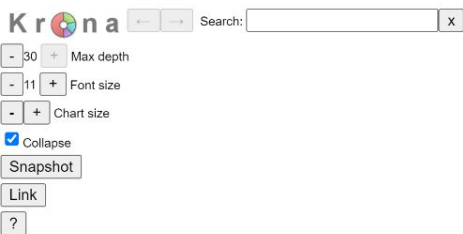
Viruses 0.04%

Archaea 0.03%

synthetic construct 0.02%

[other Root] 0.05%

Taxonomy Krona - Kraken2 S7



Root 

Count: 16965109

Unassigned: 56710



Taxonomy Krona - Centrifuge S7

Krona X

30 Max depth
 11 Font size
 Chart size

☒ Collapse

Root

Count: 21917603
Unassigned: 36638



Taxonomy Krona - Kraken2 S11

Krona Search: X

- 30 + Max depth

- 11 + Font size

- + Chart size

☒ Collapse

Snapshot

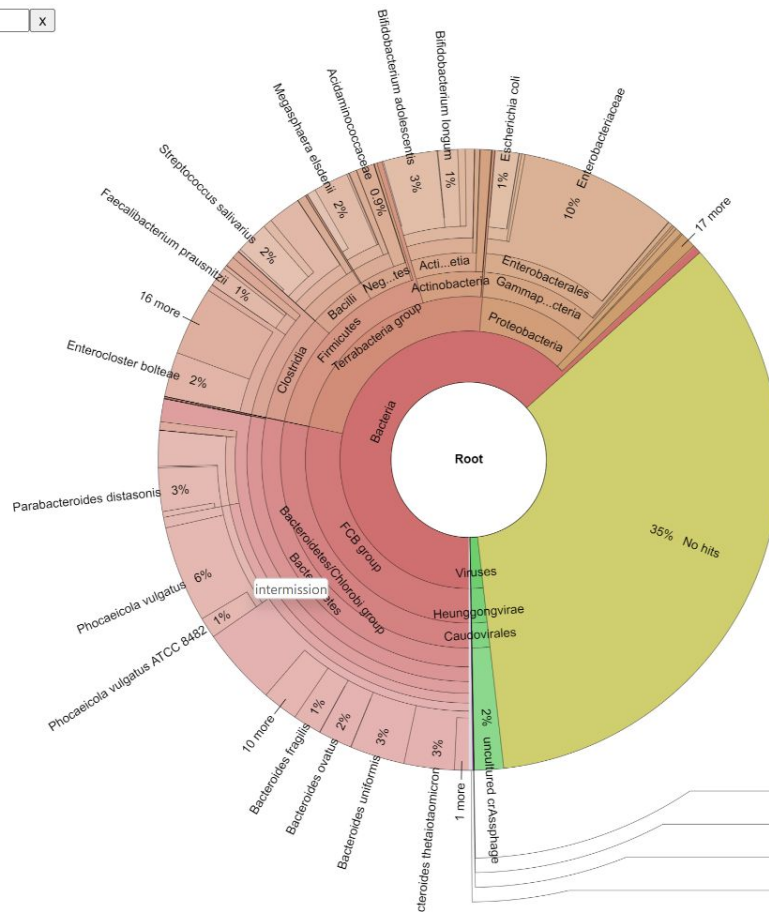
Link

?

Root

Count: 14098038

Unassigned: 34564



Homo sapiens 0.01%

Archaea 0.01%

other sequences 0.00006%

[other Root] 0.2%

Taxonomy Krona - Centrifuge S11

Krona Search: X

30 Max depth

11 Font size

Chart size

☒ Collapse

Snapshot

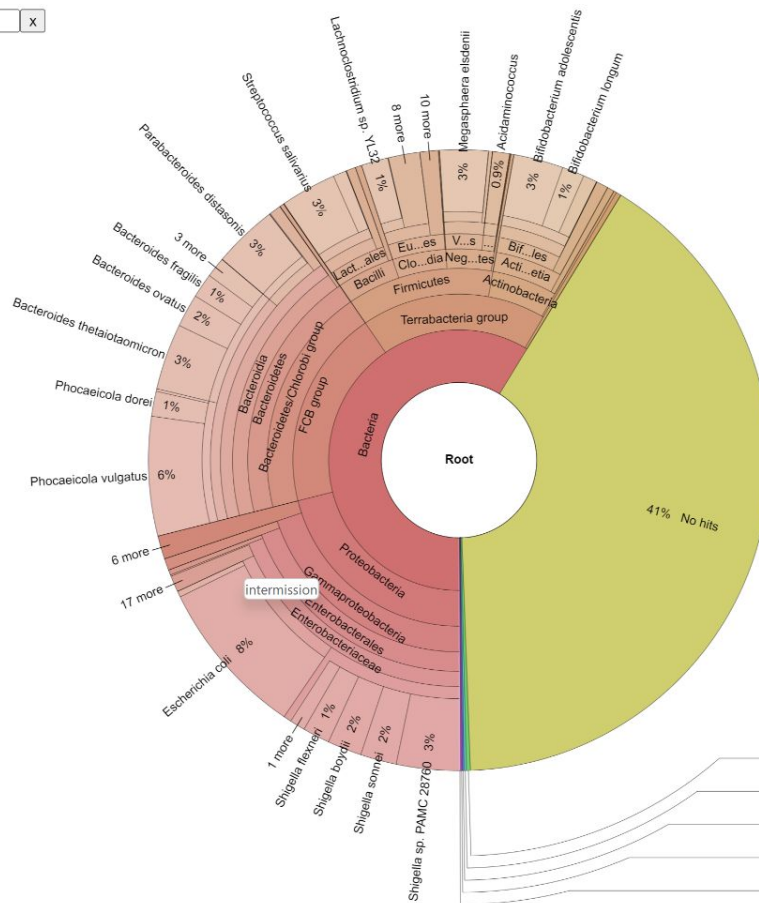
Link

?

Root

Count: 16348494

Unassigned: 16730





Work Cited

1. Krakau, S., Straub, D., Gourel, H., Gabernet, G., & Nahnsen, S. (2022). nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. *NAR genomics and bioinformatics*, 4(1), lqac007.
<https://doi.org/10.1093/nargab/lqac007>
2. Ewels, Phil. "NF-Core/MAG." Mag " Nf-Core, <https://nf-co.re/mag>.
3. Ewels, Phil. "Pipeline Configuration." Pipeline Configuration " Nf-Core, <https://nf-co.re/usage/configuration>.
4. Ewels, Philip A., et al. "The NF-Core Framework for Community-Curated Bioinformatics Pipelines." *Nature News*, Nature Publishing Group, 13 Feb. 2020, <https://www.nature.com/articles/s41587-020-0439-x>.
5. Nextflow - URCFwiki - Proteusmaster. [urcf.drexel.edu](https://proteusmaster.urcf.drexel.edu).
<https://proteusmaster.urcf.drexel.edu/urcfwiki/index.php/Nextflow>.
6. URCFwiki - Proteusmaster. [urcf.drexel.edu](https://proteusmaster.urcf.drexel.edu/urcfwiki/index.php/Main_Page). https://proteusmaster.urcf.drexel.edu/urcfwiki/index.php/Main_Page.
7. zas37. "Zas37/tutorial8-NF-Core-Mag: Tutorial 8 for Eces 450/650 with Dr. Gail Rosen." GitHub, <https://github.com/zas37/tutorial8-nf-core-mag>.