

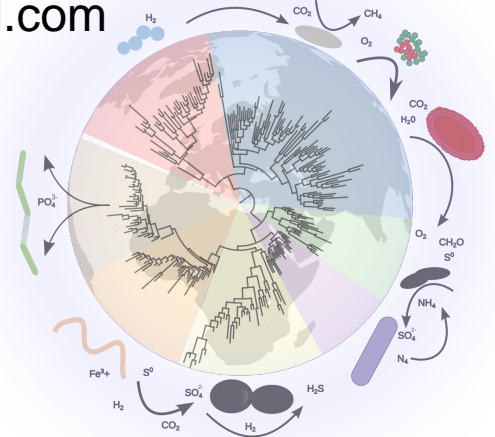
MICB 425: Project 2

Who is reducing my Nitrogen?

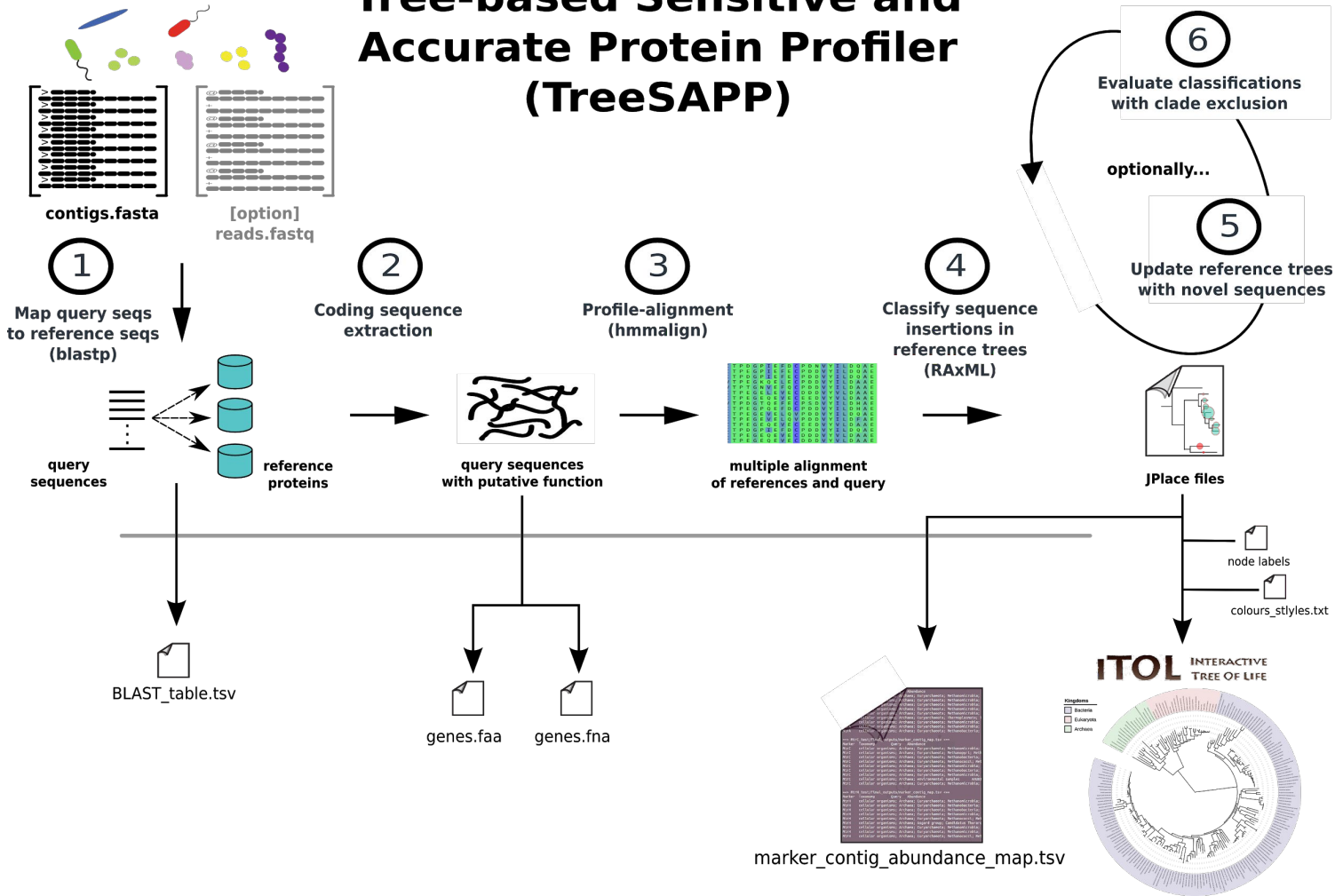
Connor Morgan-Lang: c.morganlang@gmail.com



GitHub: hallamlab/TreeSAPP

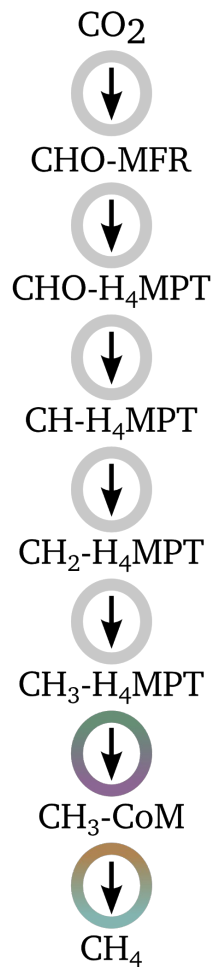


Tree-based Sensitive and Accurate Protein Profiler (TreeSAPP)

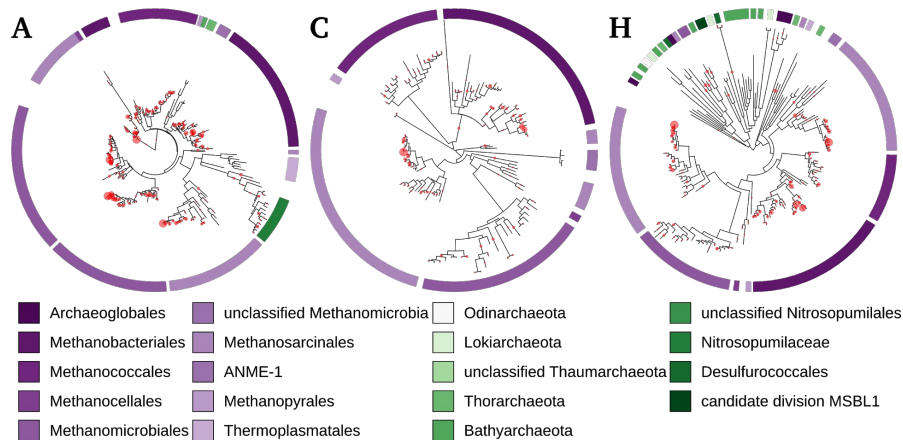


- Queries all sequence information
- Automated classifications
- Phylogenetically informed classifications
- Not slow

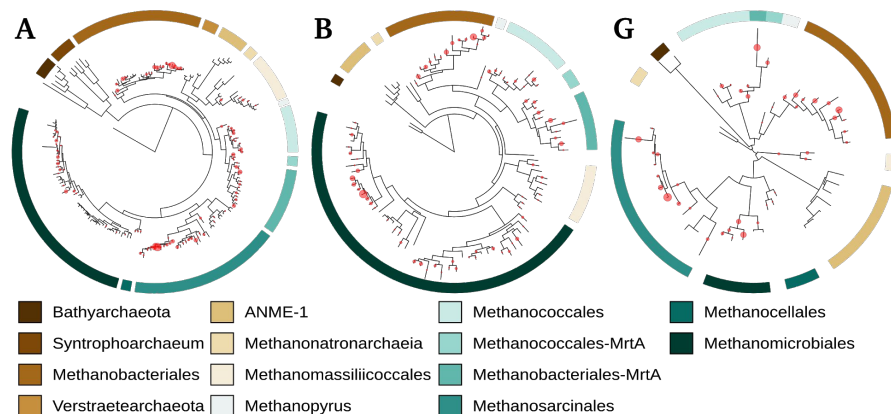
Hydrogenotrophic methanogenesis



Methyl-H₄MPT-coenzyme M methyltransferase subunits



Methyl-coenzyme M reductase subunits



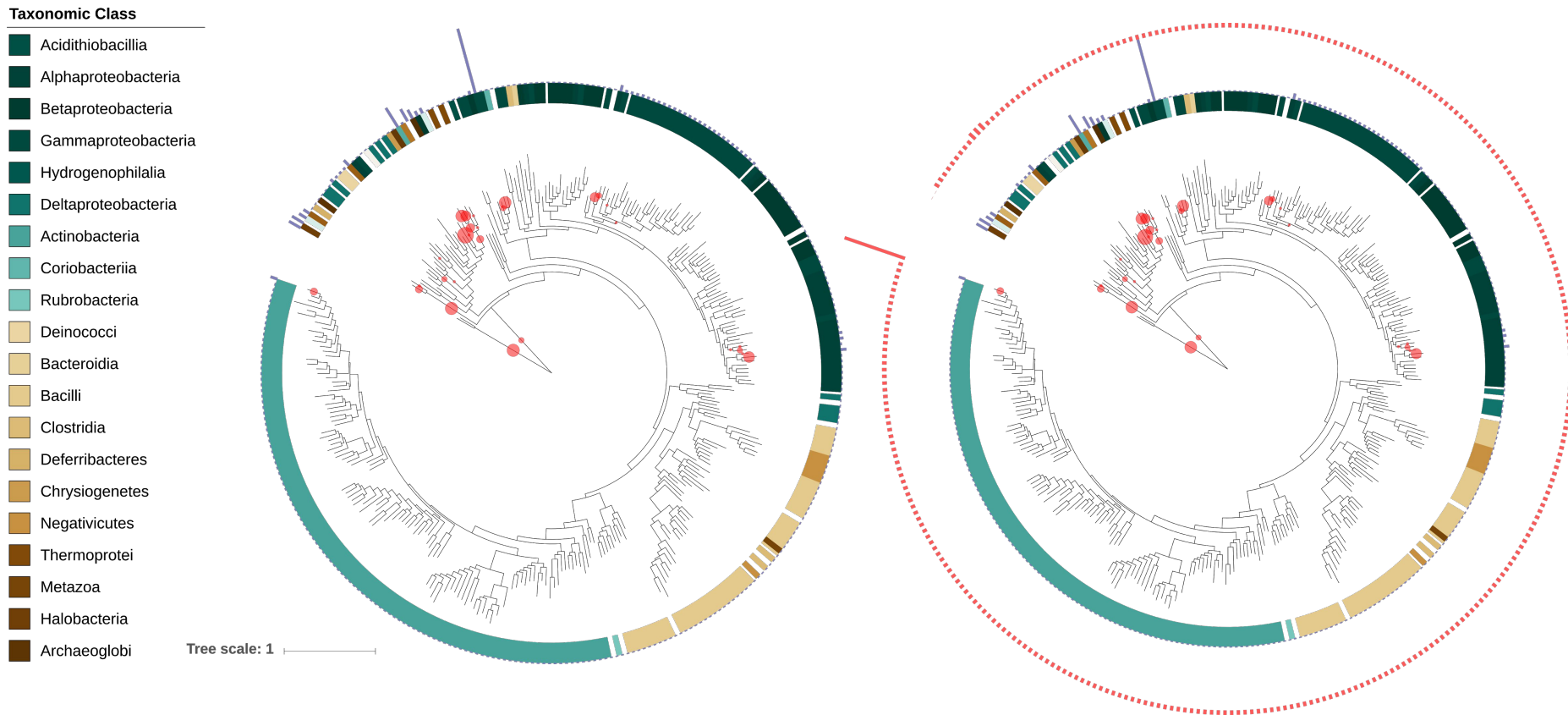
Mapping phylogeny of a pathway

All MAGs from IMG/M were queried against *mcr* and *mtr* TreeSAPP references

Takeaways:

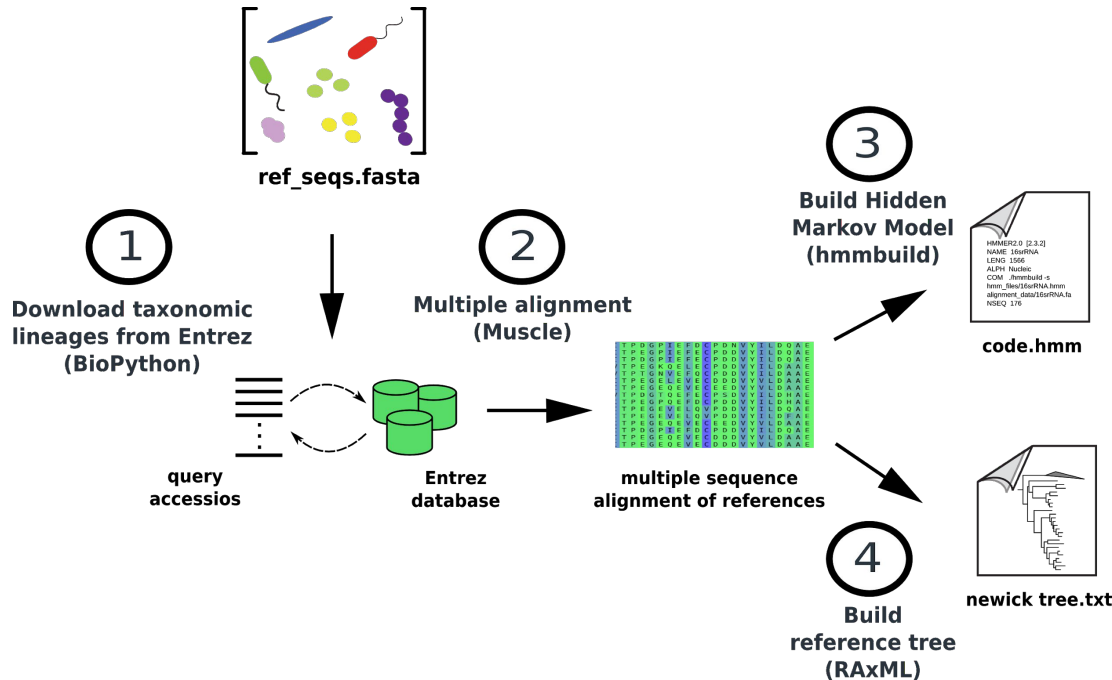
- *Mcr* is confined to only the known methanogenic lineages
- *Mtr* is found in non-methanogenic lineages such as the Asgard Archaeota

Layering trees



Creating reference packages for your *favourite* gene

```
$ ./create_treesapp_ref_data.py --fasta_input seqs.faa --name name
```



Tips:

1. Cluster first
2. Filter thoroughly
3. Sample comprehensively

Tips for running TreeSAPP:

- Use the flags: `--delete`, `--verbose` to delete useless (to you) files and know what's happening during runtime
- Use `-T 8` to take advantage of all cores
- REMOVE SAM FILES AFTER EVERY RUN
- Select your trees using `-t code_name` where `code_name` corresponds to the gene:

| | |
|-------------|--------------|
| narG | D0101 |
| napA | D0201 |
| nirK | D0301 |
| nirS | D0302 |
| norB | D0501 |
| norC | D0502 |
| nosZ | D0601 |

- Write a bash script to loop over all 14 TreeSAPP runs. REMOVE SAM FILES.



Why the cloud?

A: Scalable, consistent, on-demand resource

Why not my laptop?

A: RAM and CPU requirements are beyond your laptop. Also, TreeSAPP cannot be installed on all operating systems.

What do I need to get started?

A: Nothing! TreeSAPP has been installed on your group's instance.

How do I access it?

A: gsutil



gsutil:

Installation is easy: https://cloud.google.com/storage/docs/gsutil_install

Connecting should also be easy:

```
`gcloud compute ssh group8`
```


LIVE DEMO

Example command:

```
time treesapp.py -T 8 --verbose --delete \  
-t M0701 -i assembly.fasta --rpkm -r reads.fastq -o treesapp_out_dir
```

iTOL web address:

<http://itol.embl.de/>