

# Analyzing Soil Microbial Community Sequencing Data with the CHTC

**Thea Whitman - Department of Soil Science - UW-Madison**

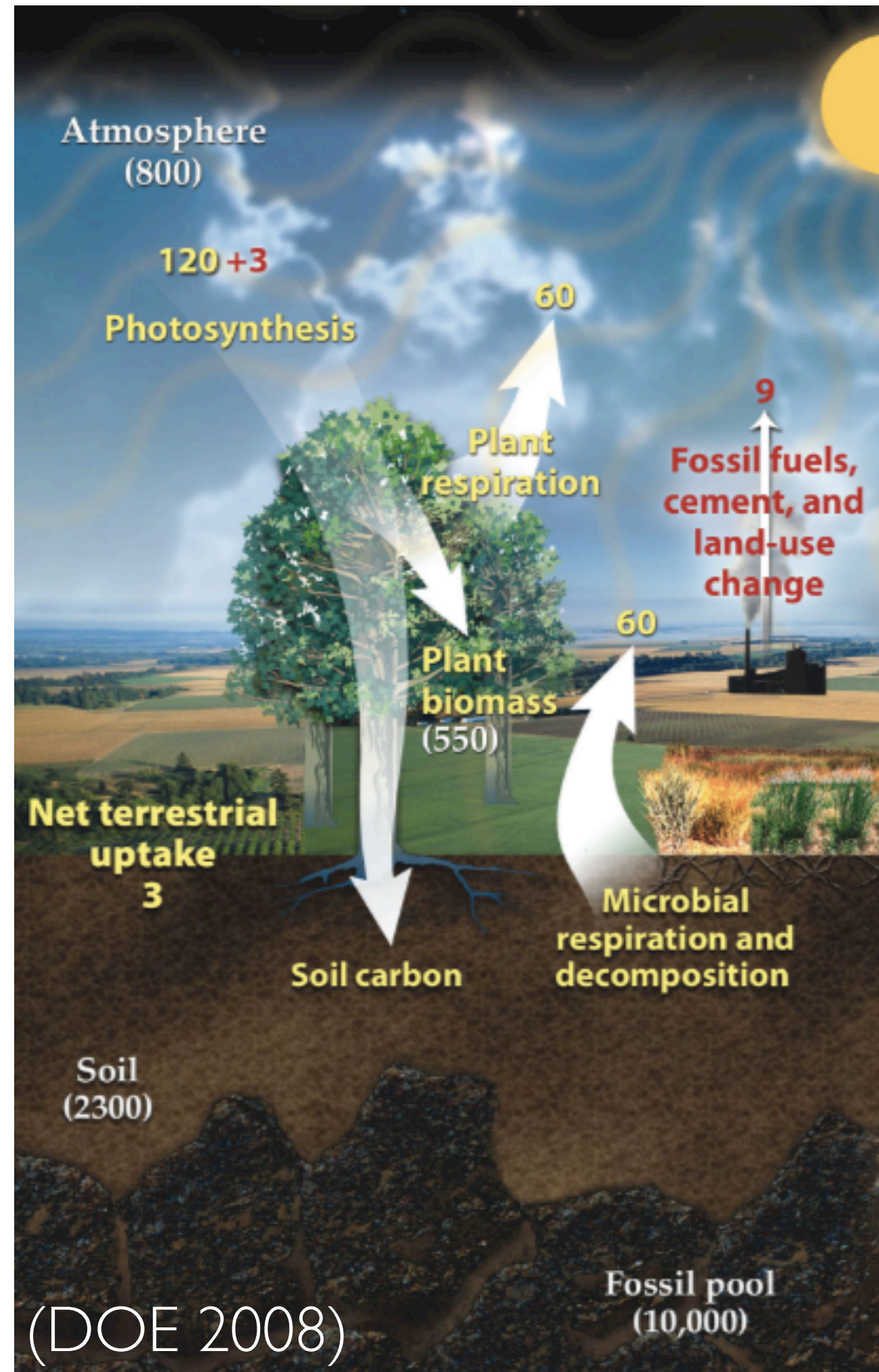
@theawhitman - [github.com/theawhitman](https://github.com/theawhitman) - [whitmanlab.soils.wisc.edu](http://whitmanlab.soils.wisc.edu)





# Key questions

- What drives organic matter (OM) cycling in soils?
- How can microbial ecology inform our understanding of OM cycling?
- How do forest fires impact the microbial community?



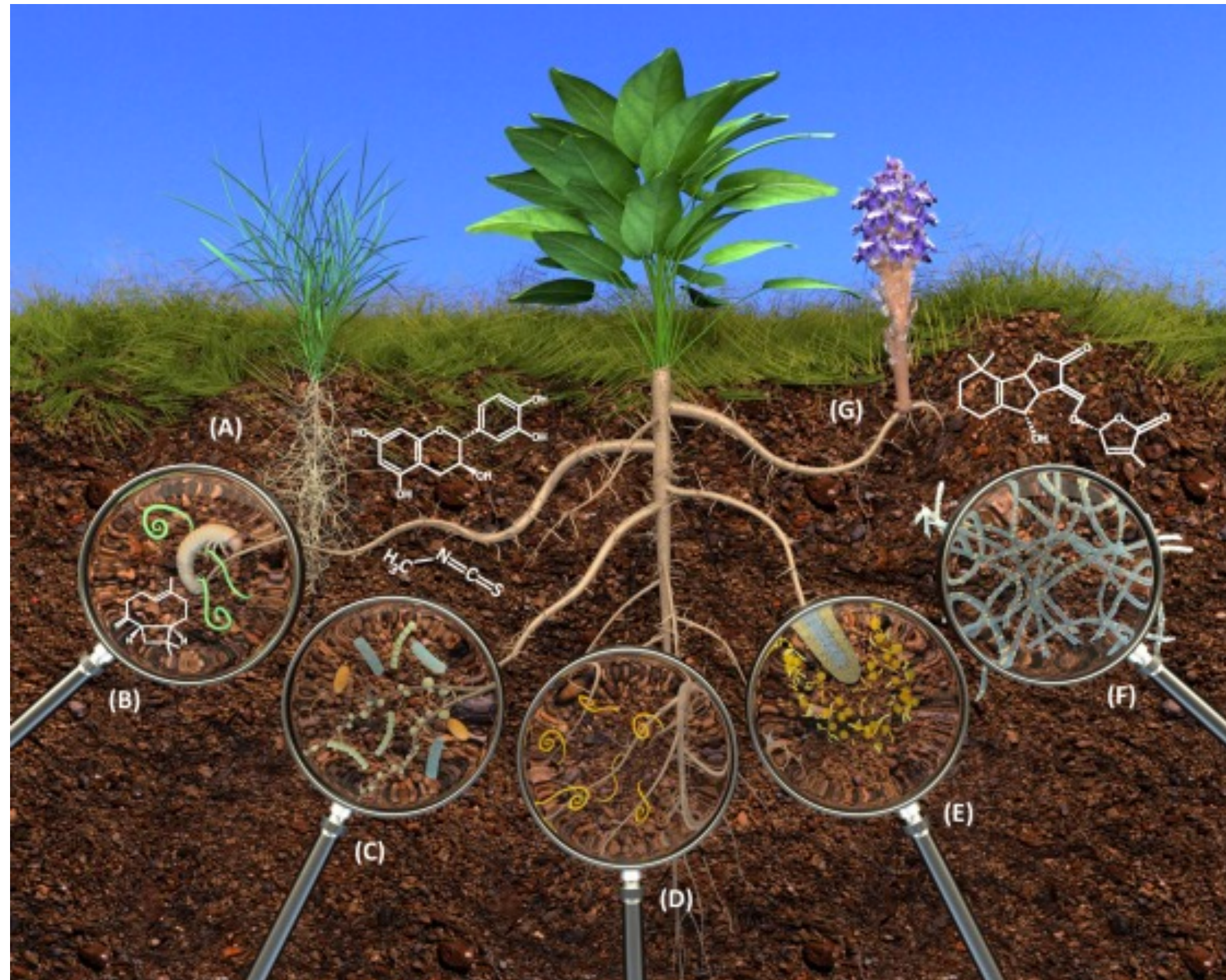


# Two approaches

- Amplicon Sequencing
- Environmental Shotgun Metagenomics

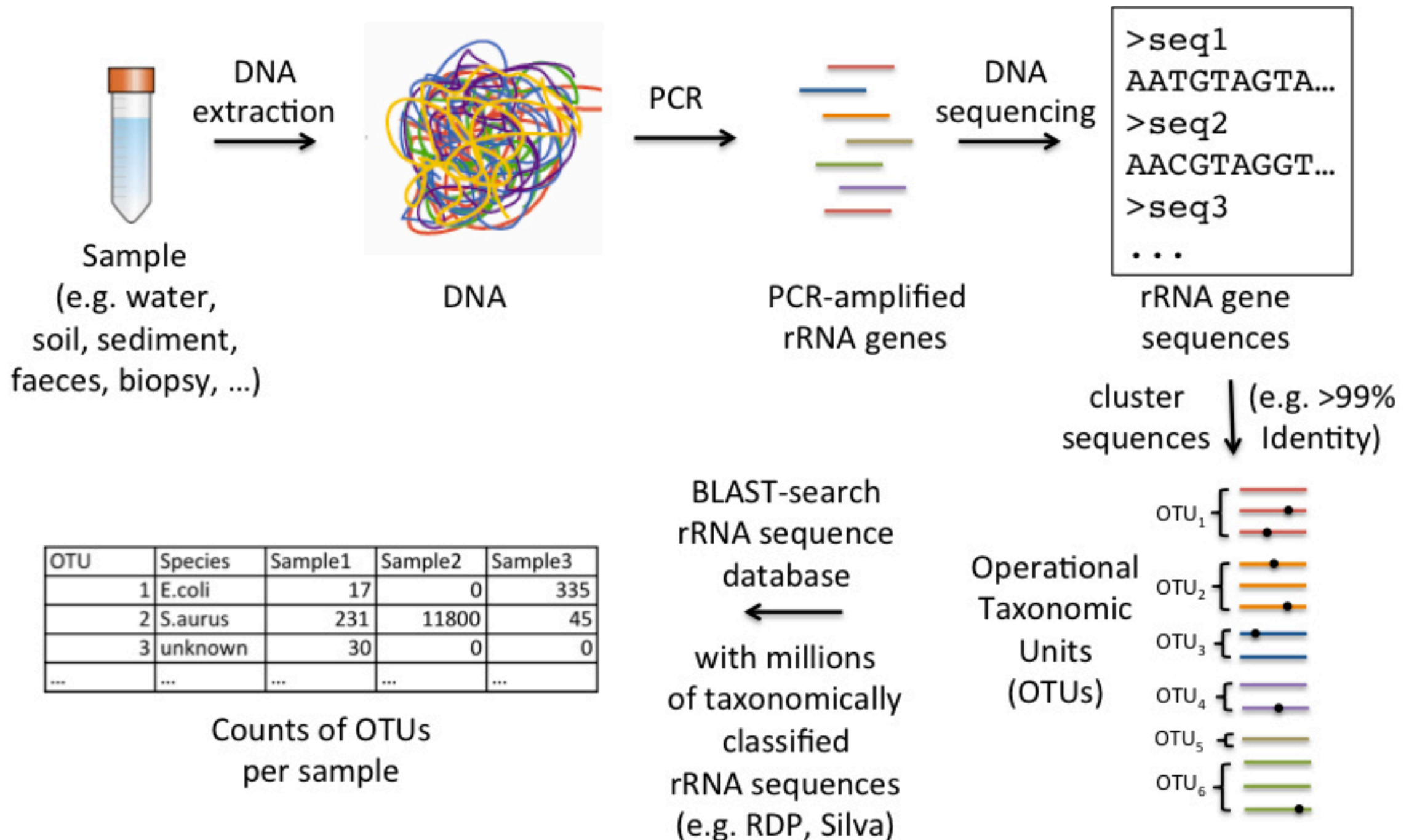


Dr. Timothy Berry



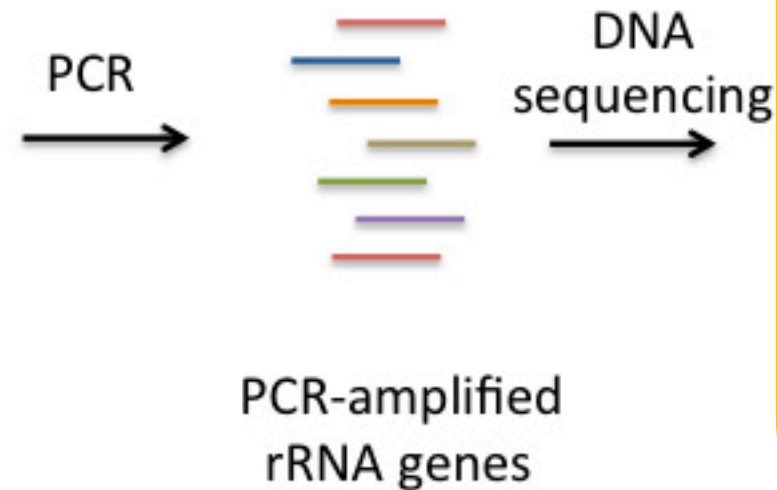
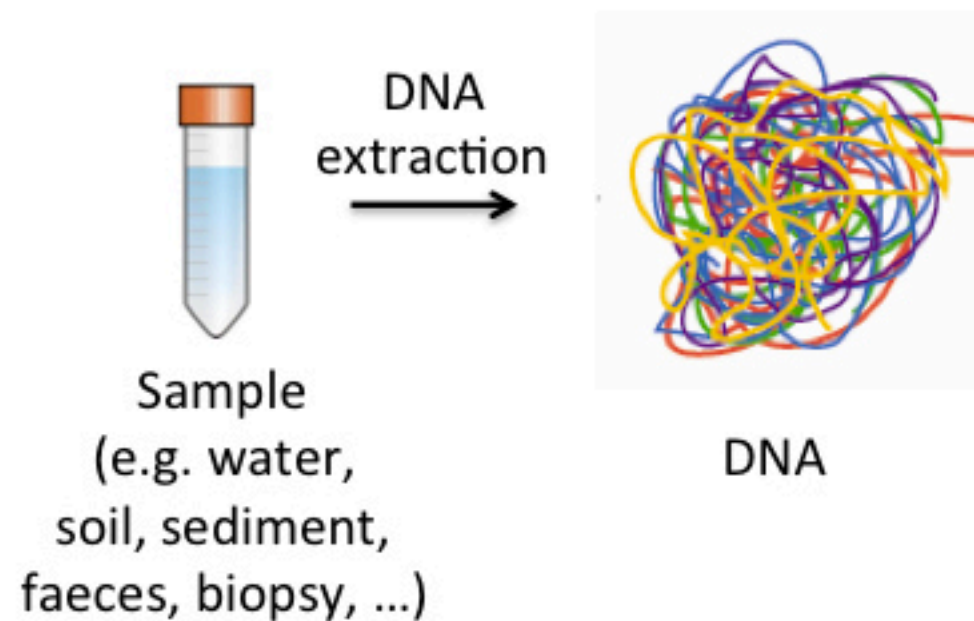


# Approach: Amplicon Sequencing





# Approach: Amplicon Sequencing

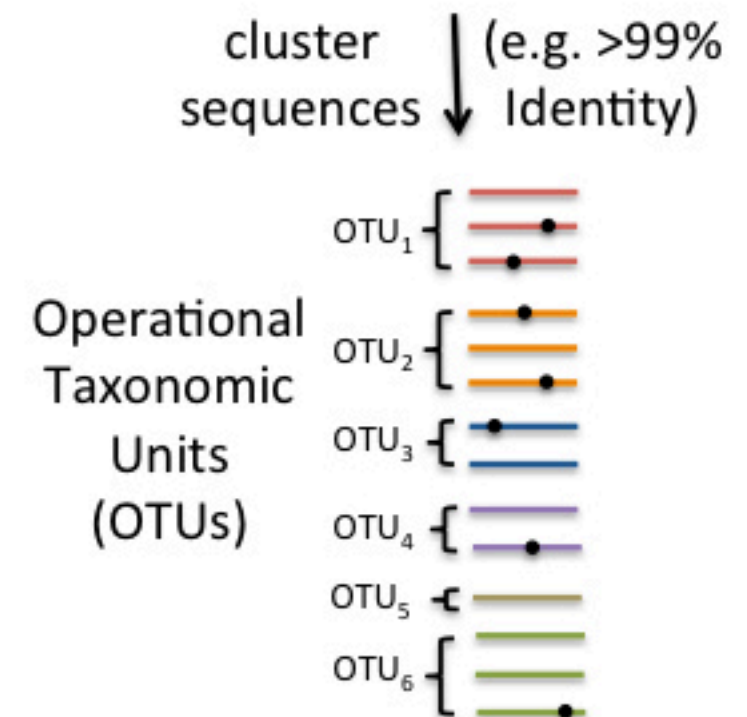


**Typical study:**  
~200 samples  
10k-100k seq/  
sample  
3-5 Gb total

OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

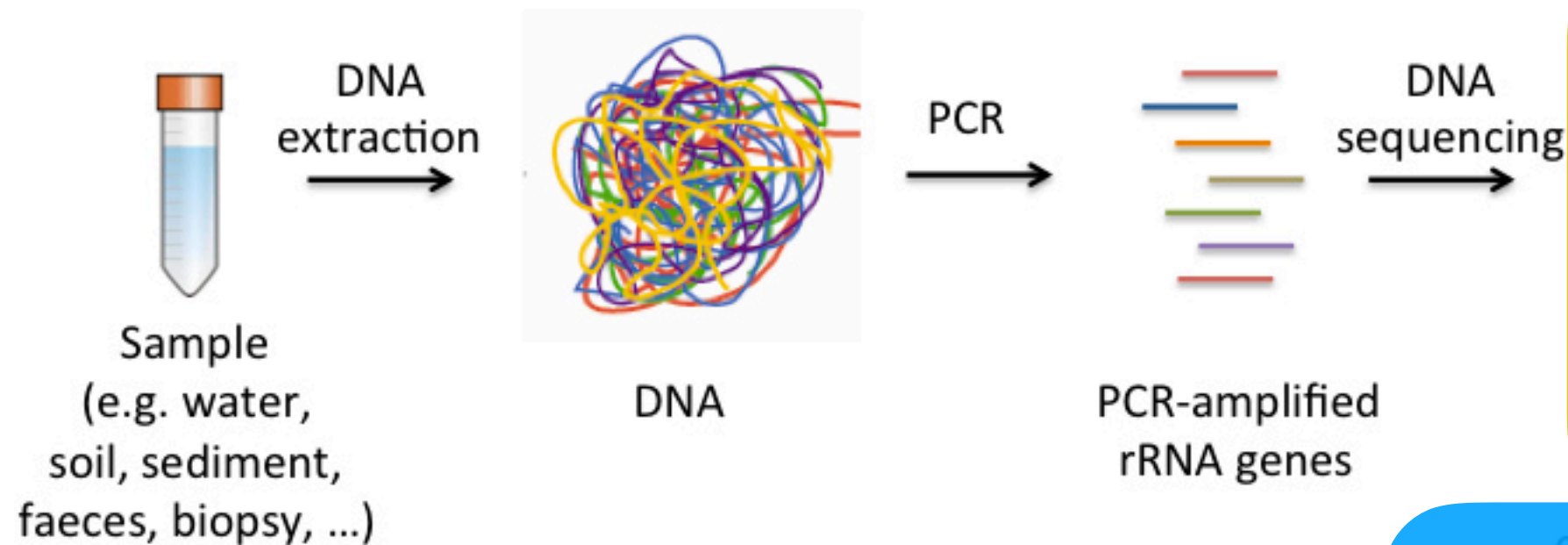
Counts of OTUs  
per sample

BLAST-search  
rRNA sequence  
database  
←  
with millions  
of taxonomically  
classified  
rRNA sequences  
(e.g. RDP, Silva)





# Approach: Amplicon Sequencing



**Typical study:**

**~200 samples**  
**10k-100k seq/  
sample**  
**3-5 Gb total**

OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

Counts of OTUs  
per sample

BLAST-search  
rRNA sequence  
database

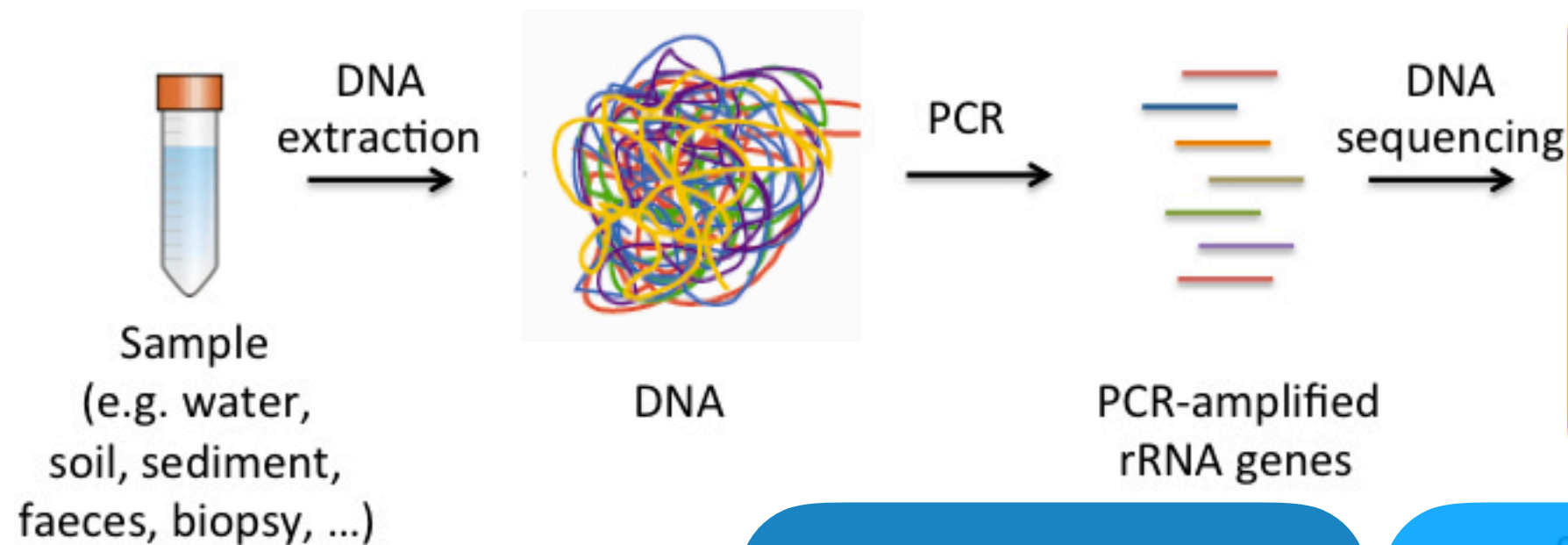
←  
with millions  
of taxonomically  
classified  
rRNA sequences  
(e.g. RDP, Silva)

**Approach:**

**Copy custom R  
installation with  
relevant packages  
(224Mb);  
Run R script**



# Approach: Amplicon Sequencing



## Typical study:

~200 samples  
10k-100k seq/  
sample  
3-5 Gb total

OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

Counts of OTUs  
per sample

## QC and data processing:

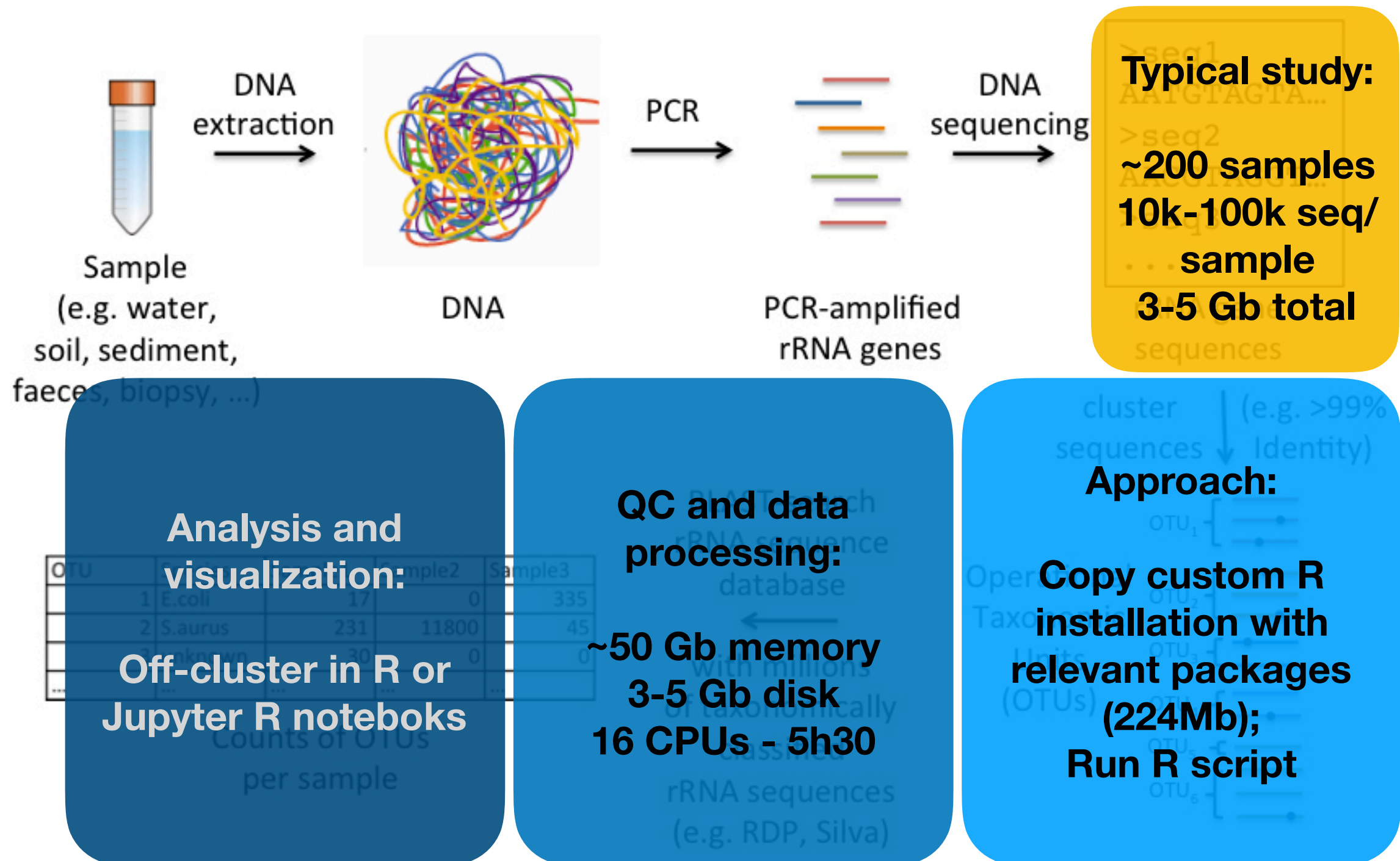
~50 Gb memory  
3-5 Gb disk  
16 CPUs - 5h30

## Approach:

Copy custom R  
installation with  
relevant packages  
(224Mb);  
Run R script



# Approach: Amplicon Sequencing





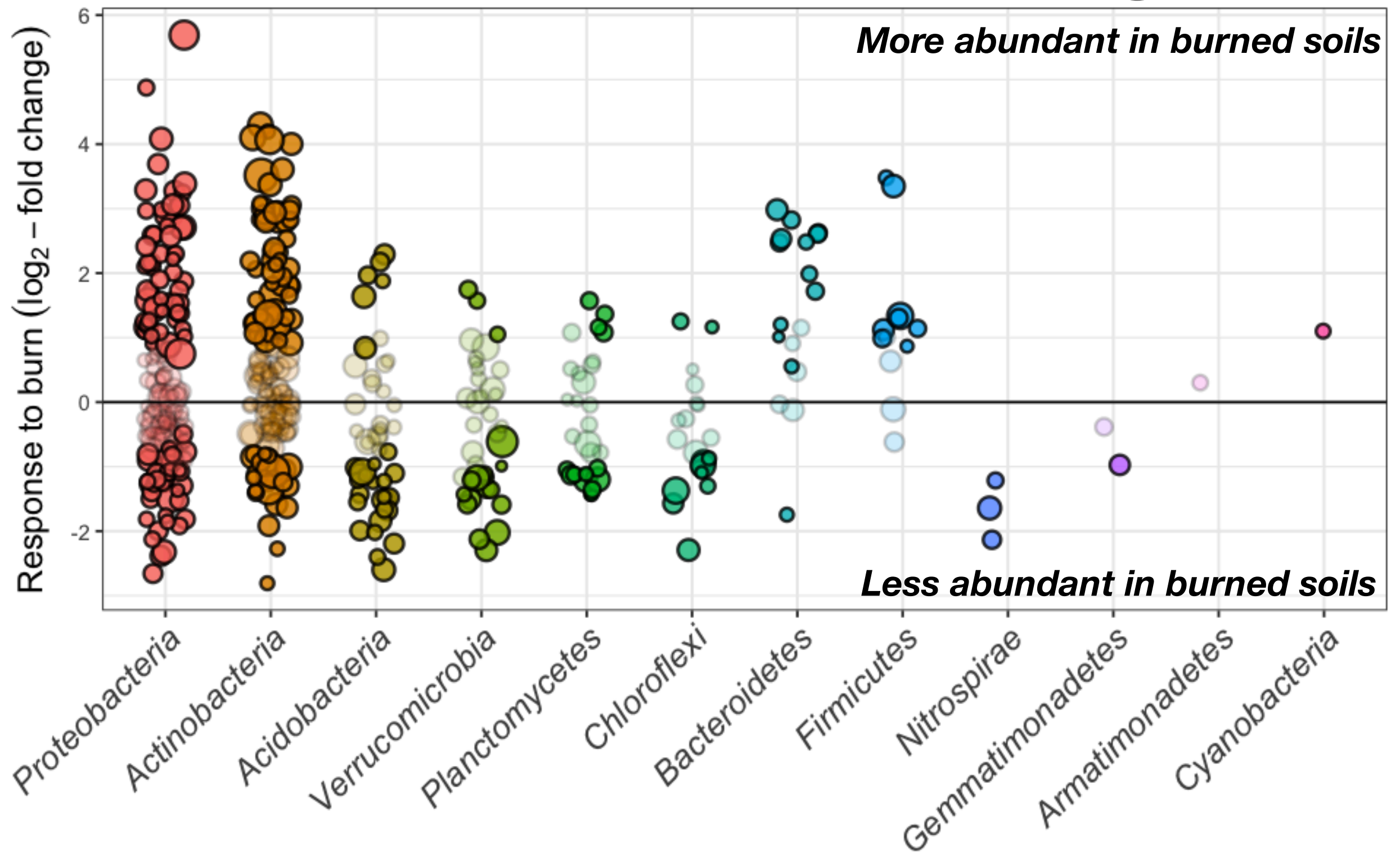
# Challenges: Amplicon Sequencing

- Almost ~tractable on desktop/laptop
- Working interactively vs. with scripts
- Rapidly evolving tools / best practices
- Bringing analysis into the classroom



# Results:

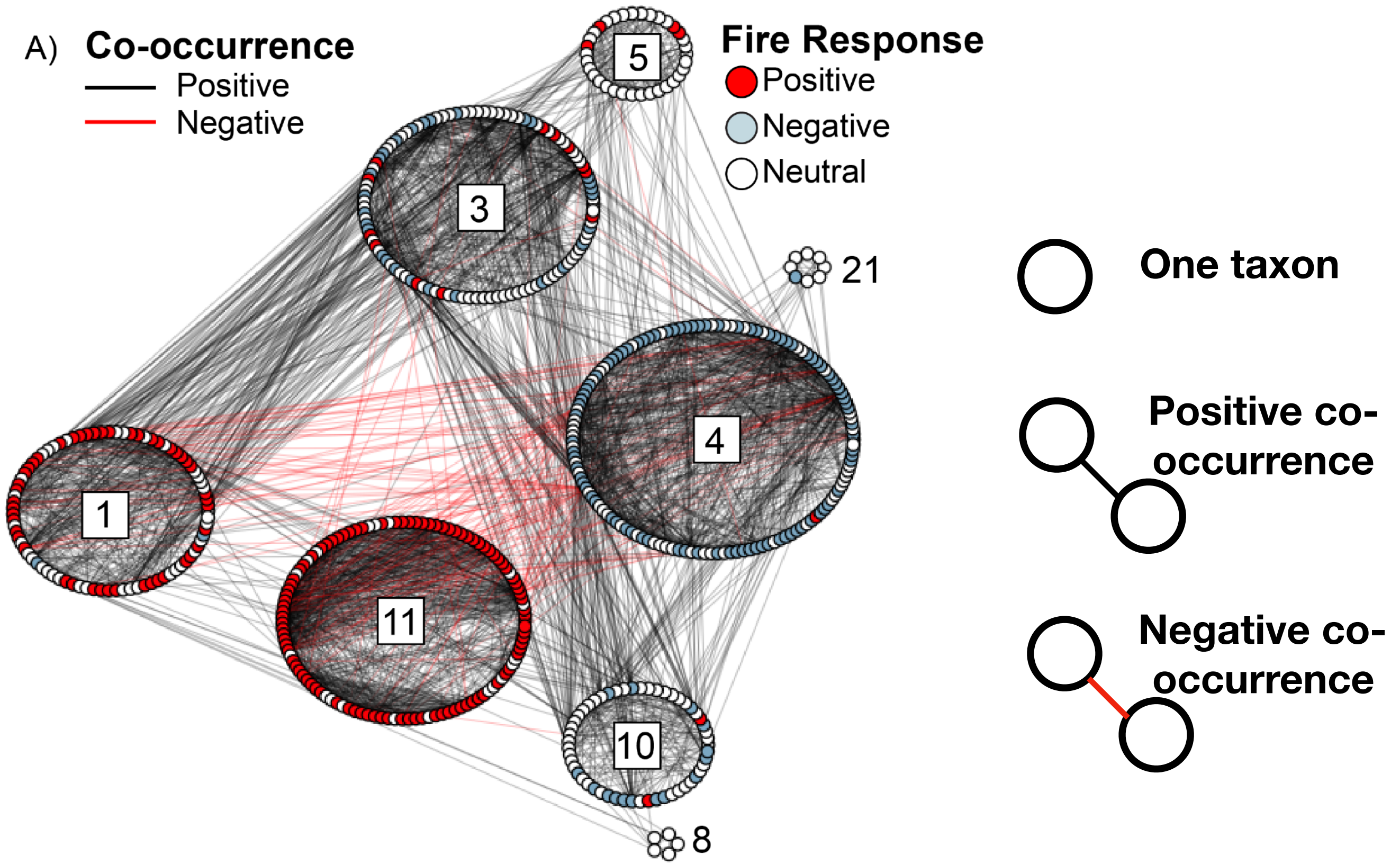
## Amplicon Sequencing



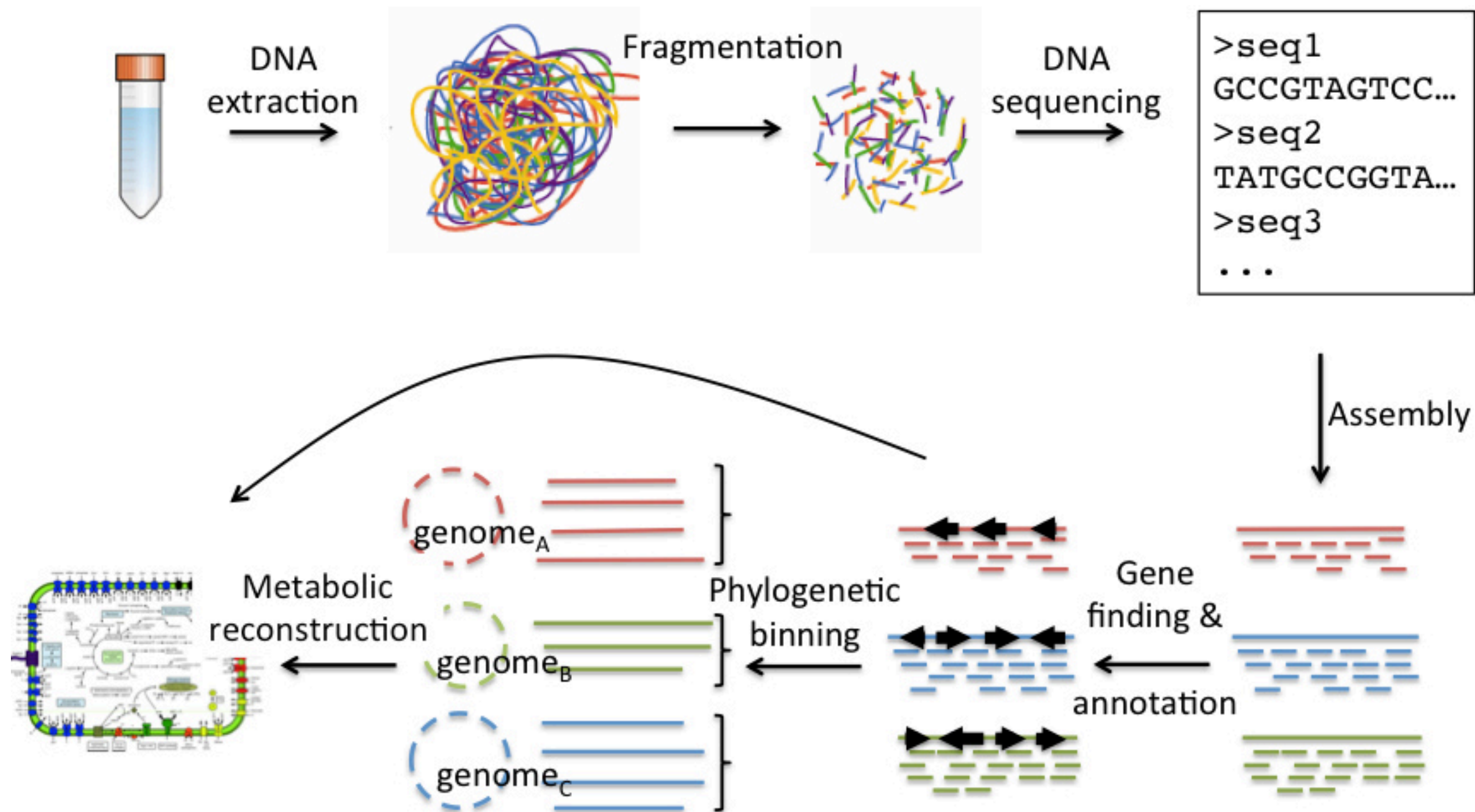
Log2-fold change with fire (controlling for pH, vegetation, and %C in soil) using *metagenomeSeq* (Paulson *et al.* 2016)  
Each point represents one OTU, sized by mean relative abundance. Faint OTUs were not significant ( $p_{adj} < 0.05$ )



# Results: Amplicon Sequencing

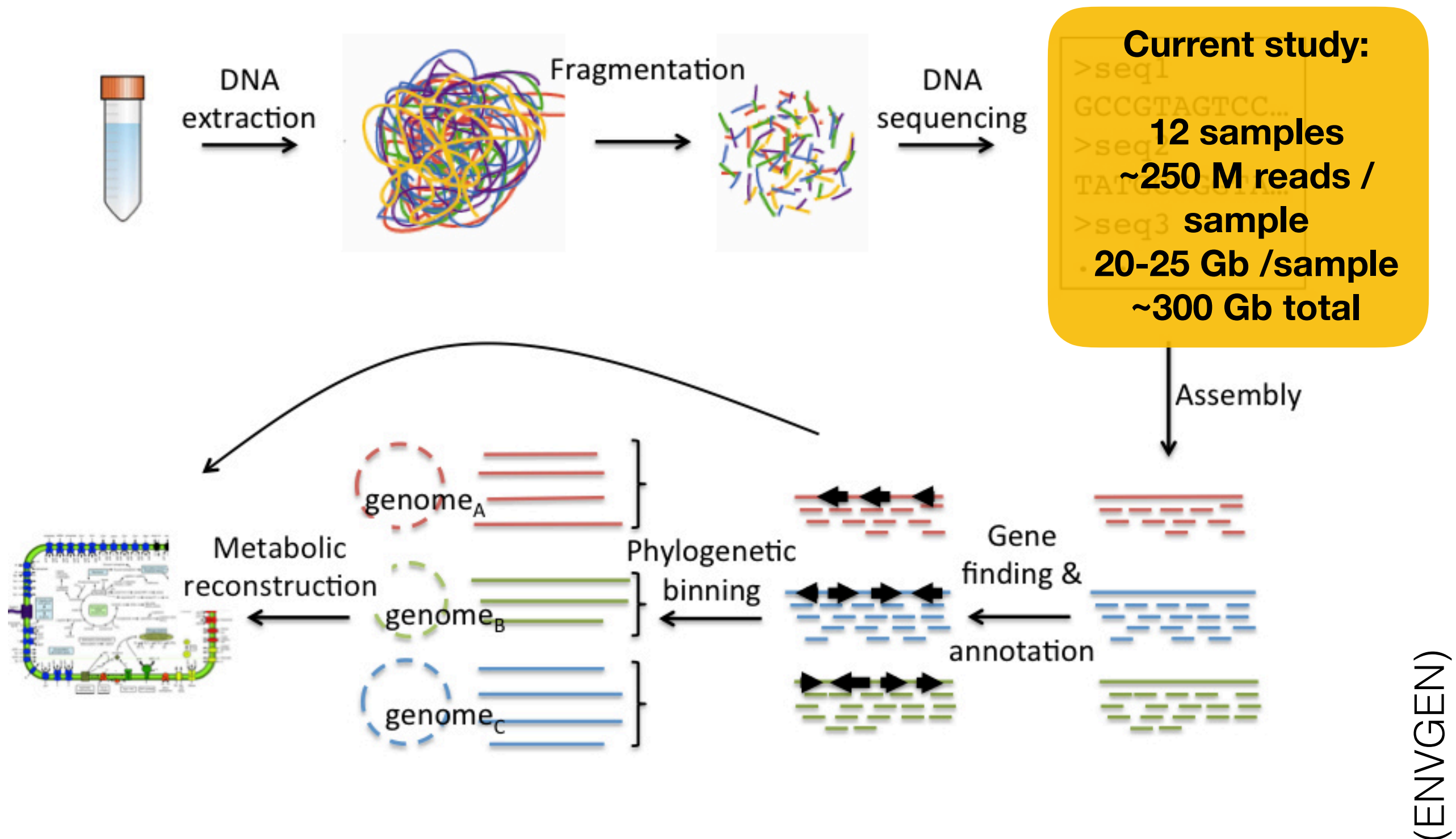


# Approach: Environmental Shotgun Metagenomics

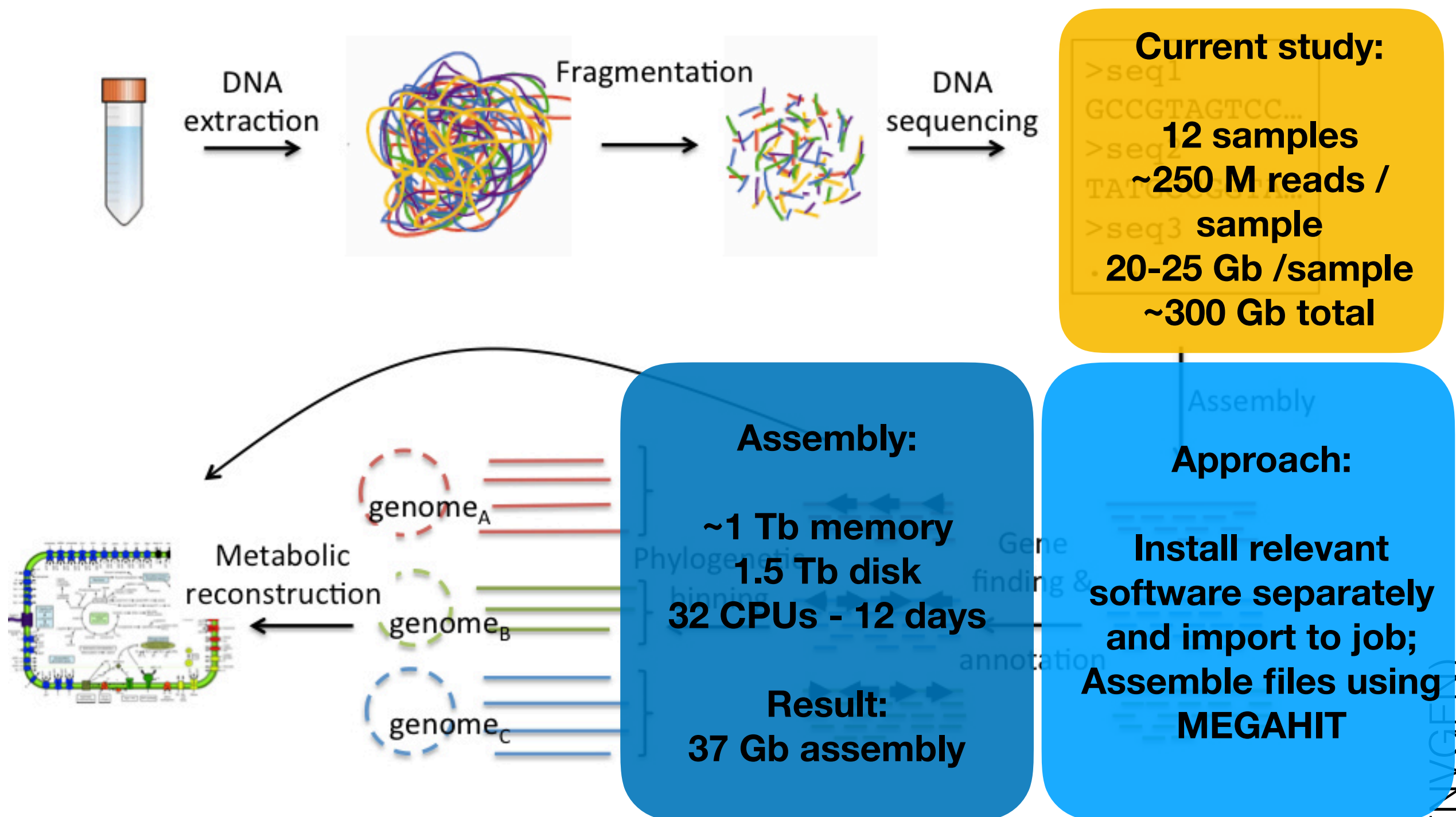




# Approach: Environmental Shotgun Metagenomics

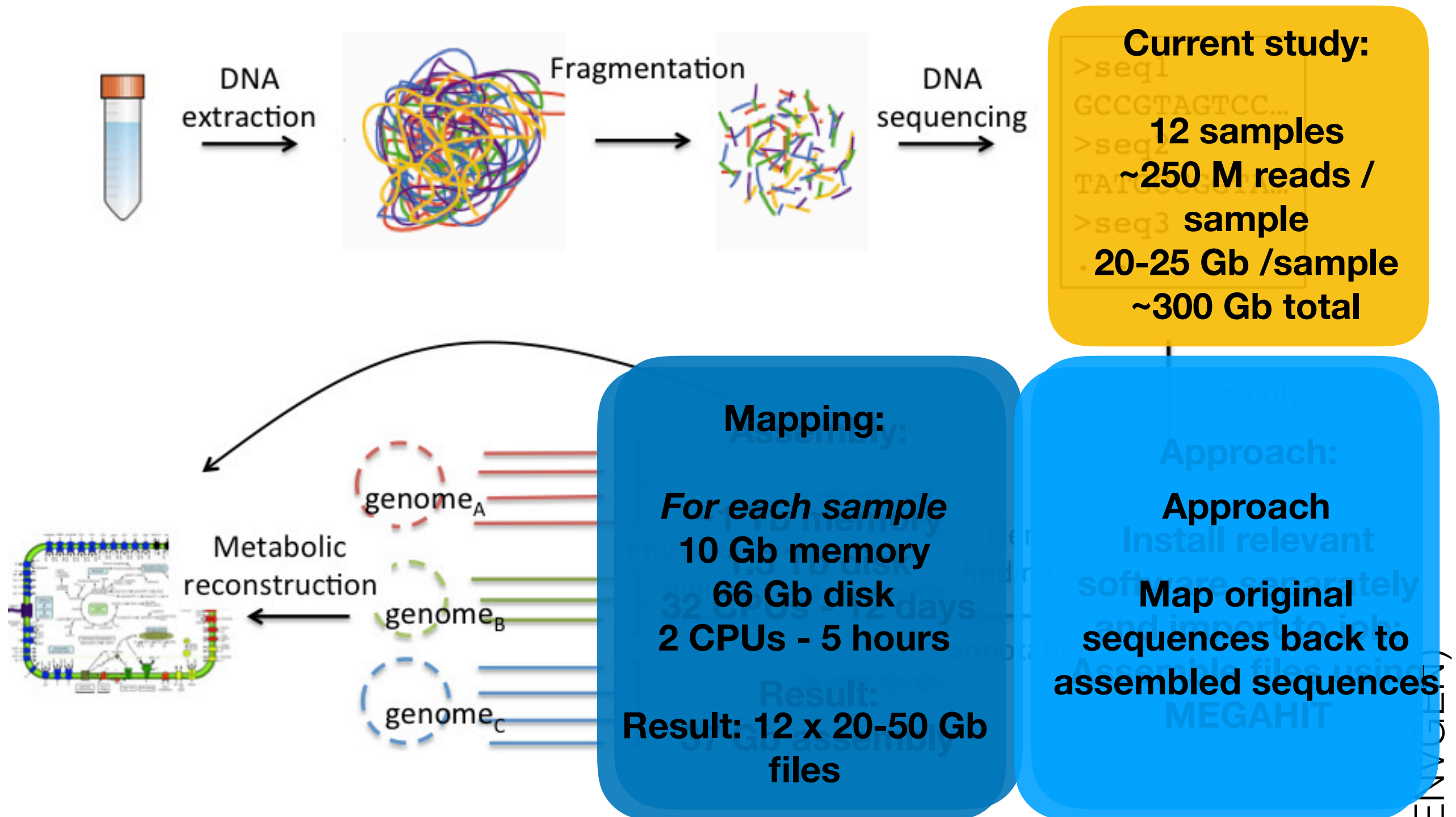


# Approach: Environmental Shotgun Metagenomics

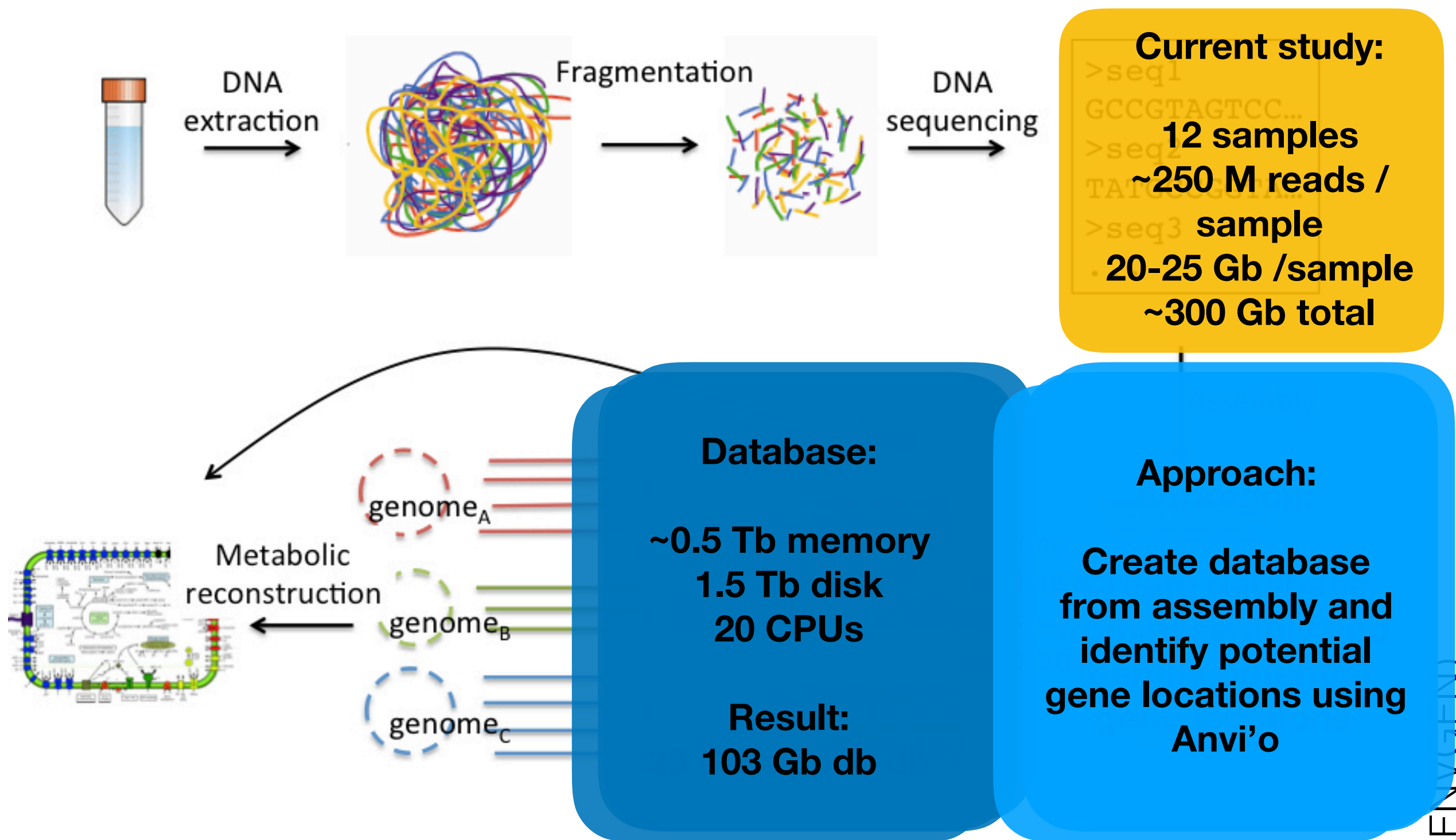




# Approach: Environmental Shotgun Metagenomics

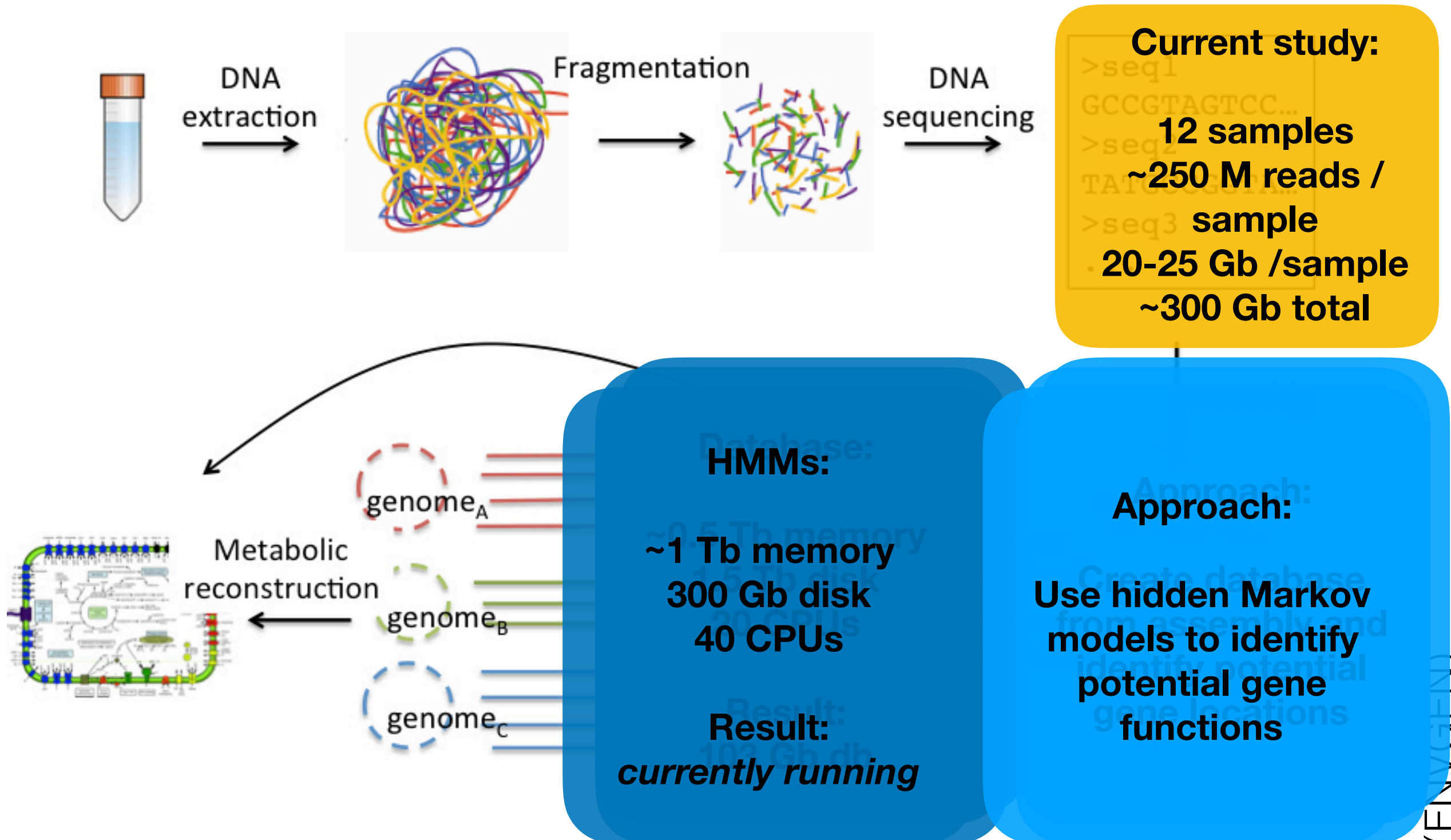


# Approach: Environmental Shotgun Metagenomics

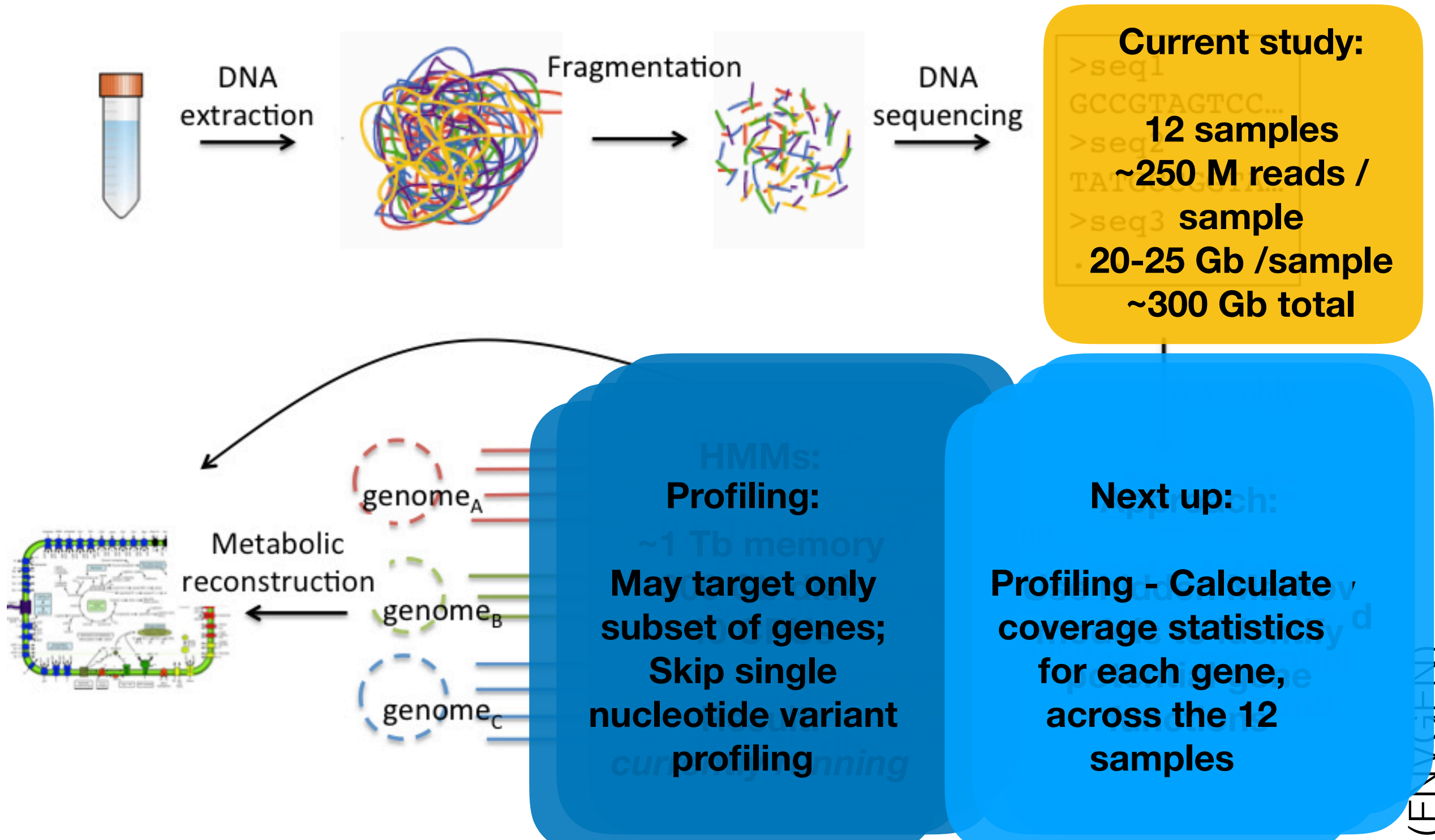




# Approach: Environmental Shotgun Metagenomics

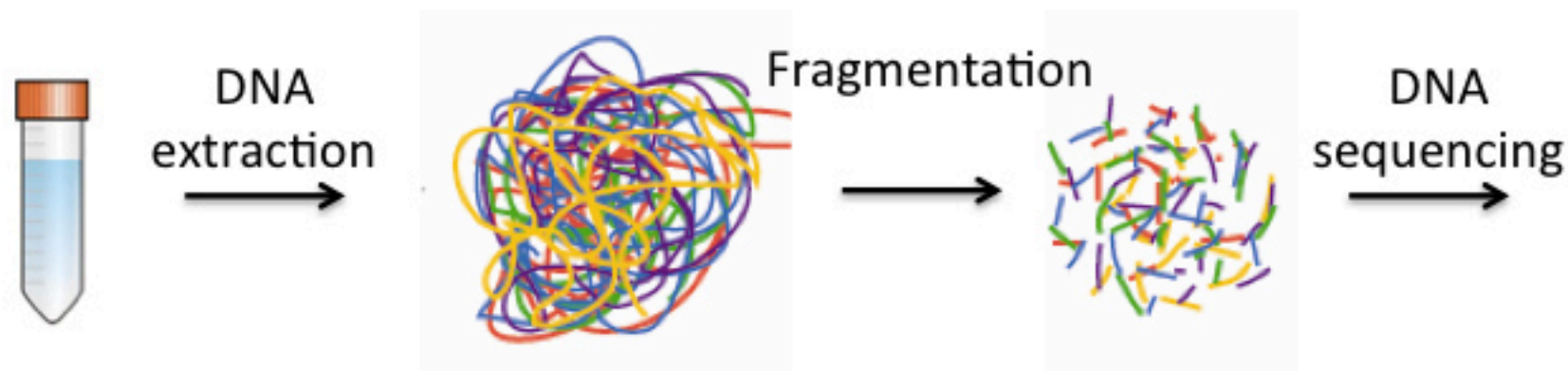


# Approach: Environmental Shotgun Metagenomics





# Approach: Environmental Shotgun Metagenomics



## Current study:

```
>seq1
GCCGTAGTCC...
>seq2
TATCGG...
>seq3
```

**12 samples**  
**~250 M reads / sample**  
**20-25 Gb /sample**  
**~300 Gb total**

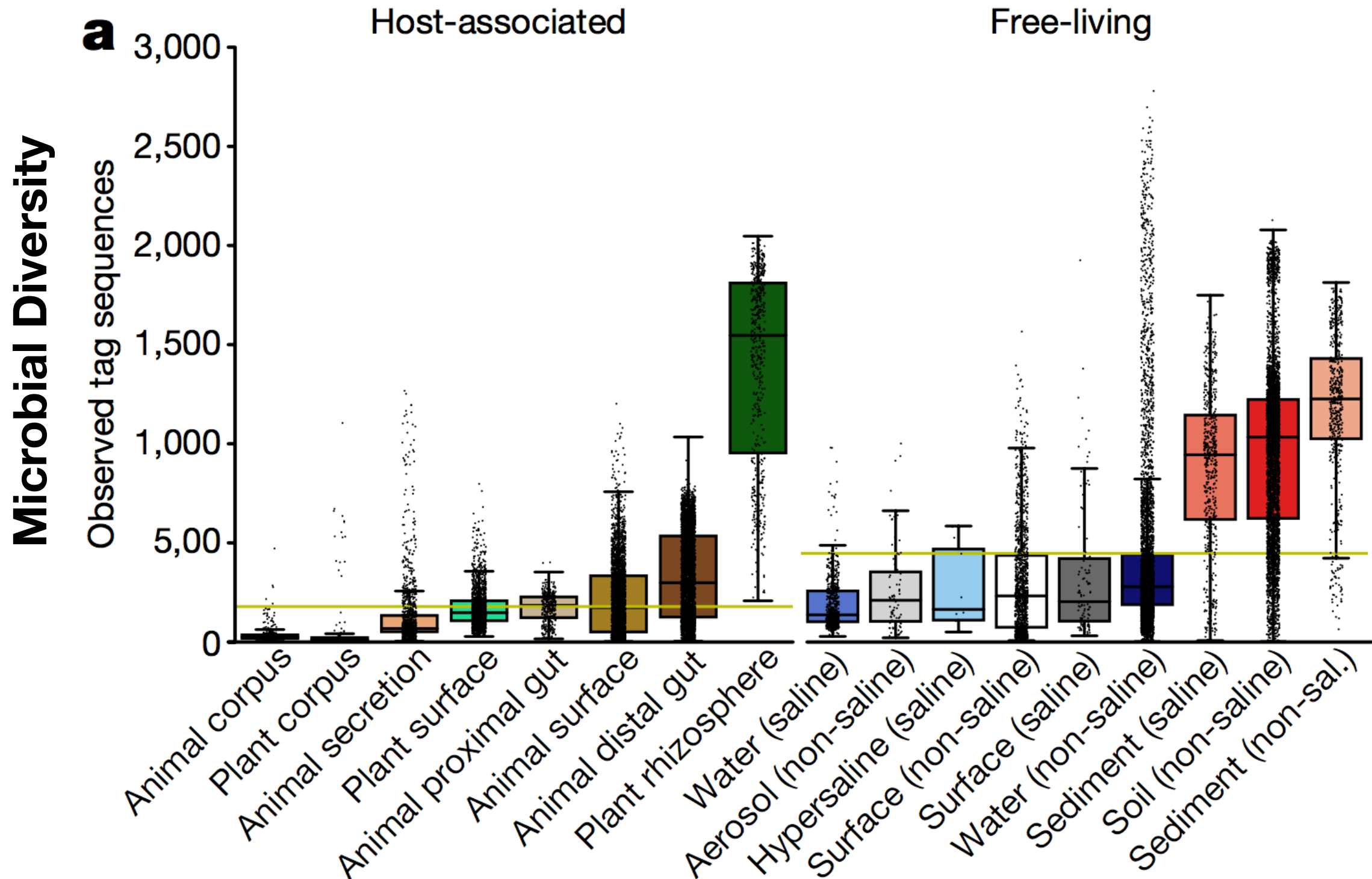
**Analysis and visualization:**  
**Interact with database within interactive job**

The visualization shows a circular genomic map with various colored segments and labels, including "genome\_A", "genome\_B", and "genome\_C", representing the reconstructed metagenome.

**HMMs:**  
**Profiling:**  
**~1 Tb memory**  
**May target only subset of genes;**  
**Skip single nucleotide variant profiling**

**Next up:**  
**Profiling - Calculate coverage statistics for each gene, across the 12 samples**

# Soil bacterial communities are among the most diverse



(Thompson *et al.*, 2017)

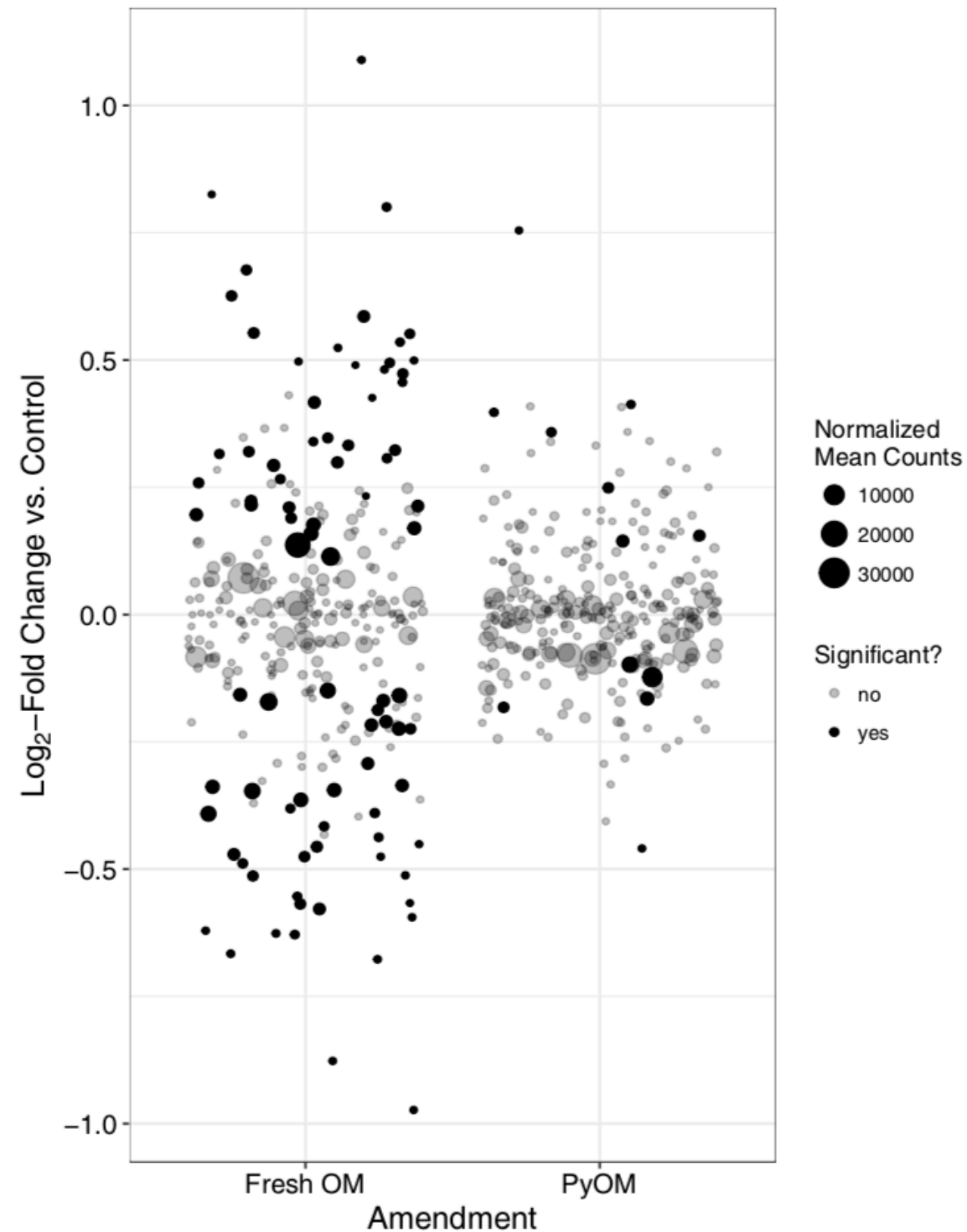


# Challenges: Environmental Shotgun Metagenomics

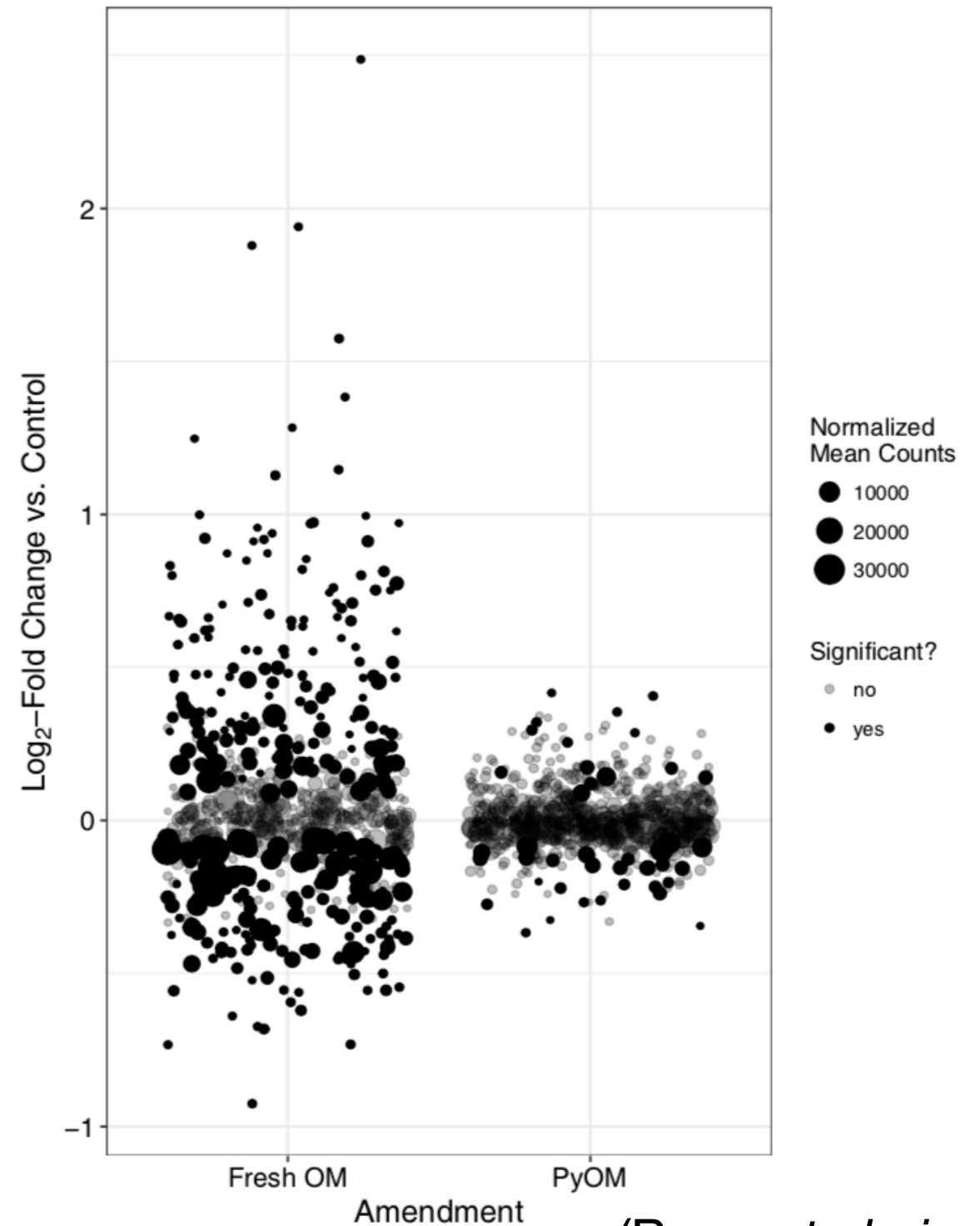
- Copying files (sometimes not zipping is faster)
- Soil microbial communities are particularly diverse
  - Comparatively large files
  - Tools may not be optimized for soil samples
  - Weeks-long running times
- User, not developer

# Results: Environmental Shotgun Metagenomics

## Xenobiotic Degradation KOs



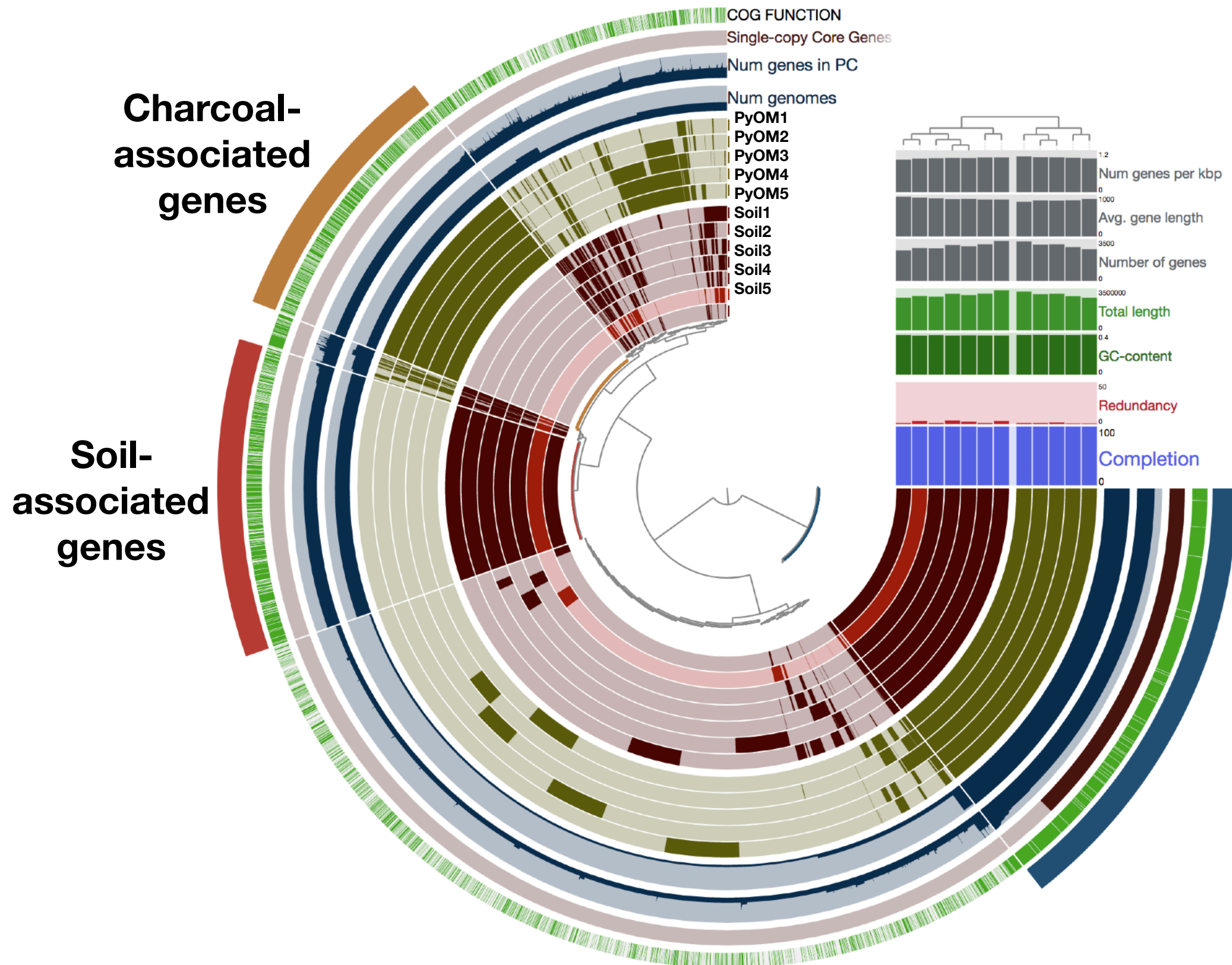
## Carbohydrate Metabolism KOs



(Berry *et al.*, *in prep*)



# *Future Results: Environmental Shotgun Metagenomics*





# Acknowledgements



**Co-authors: Ellen Whitman, Jamie Woolet, Marc-André Parisien, Daniel Thompson, Michael Flannigan, Timothy Berry**  
**Special thanks: Christina Koch and Lauren Michael**