

# Using metagenome assembled genomes (MAGs) to investigate thermophiles in the soils overlying the Centralia, PA coal fire

---

Jane Lee

27 July 2017

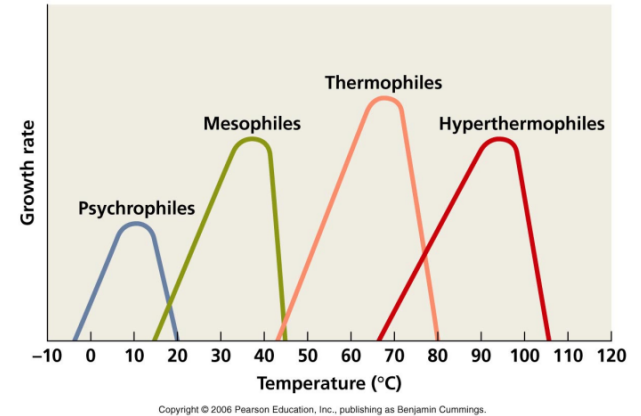
# Background

- Thermophiles

- Models for early life
- Source of biologically active enzymes and compounds
- Ubiquitous despite need for higher temperatures

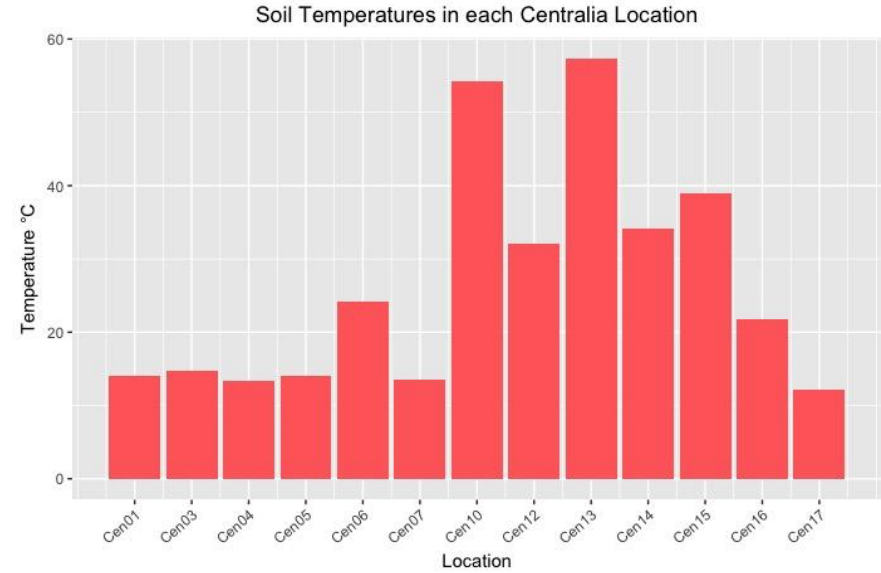
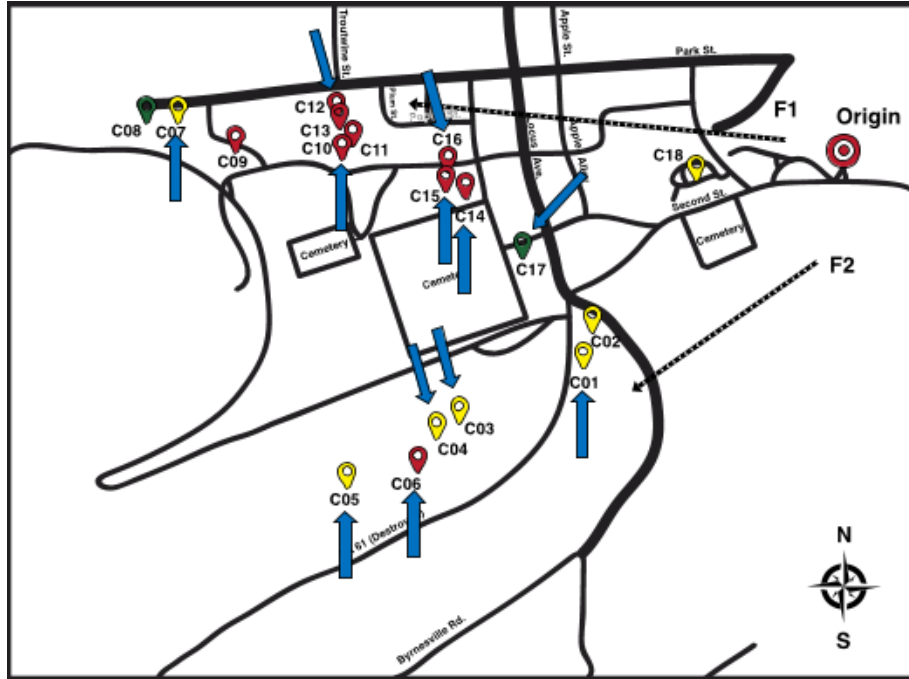
- Centralia

- Coal seam fire burning since 1962
- Interesting location to study thermophiles
  - Unusual environment



- Terrestrial, normally temperate

# Centralia Samples



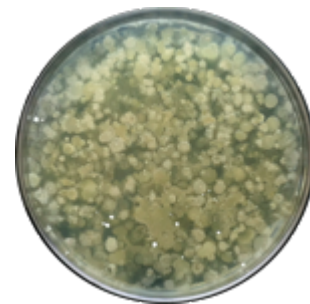
How homogeneous are thermophiles' genomes across temperature gradients and space in soil?

Are there any instances of gene loss and or gain in the thermophiles present in the soils overlying the Centralia coal mine fire?

# My Dataset



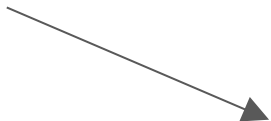
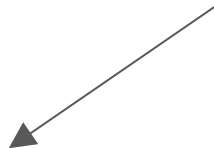
Cultured at 60°C



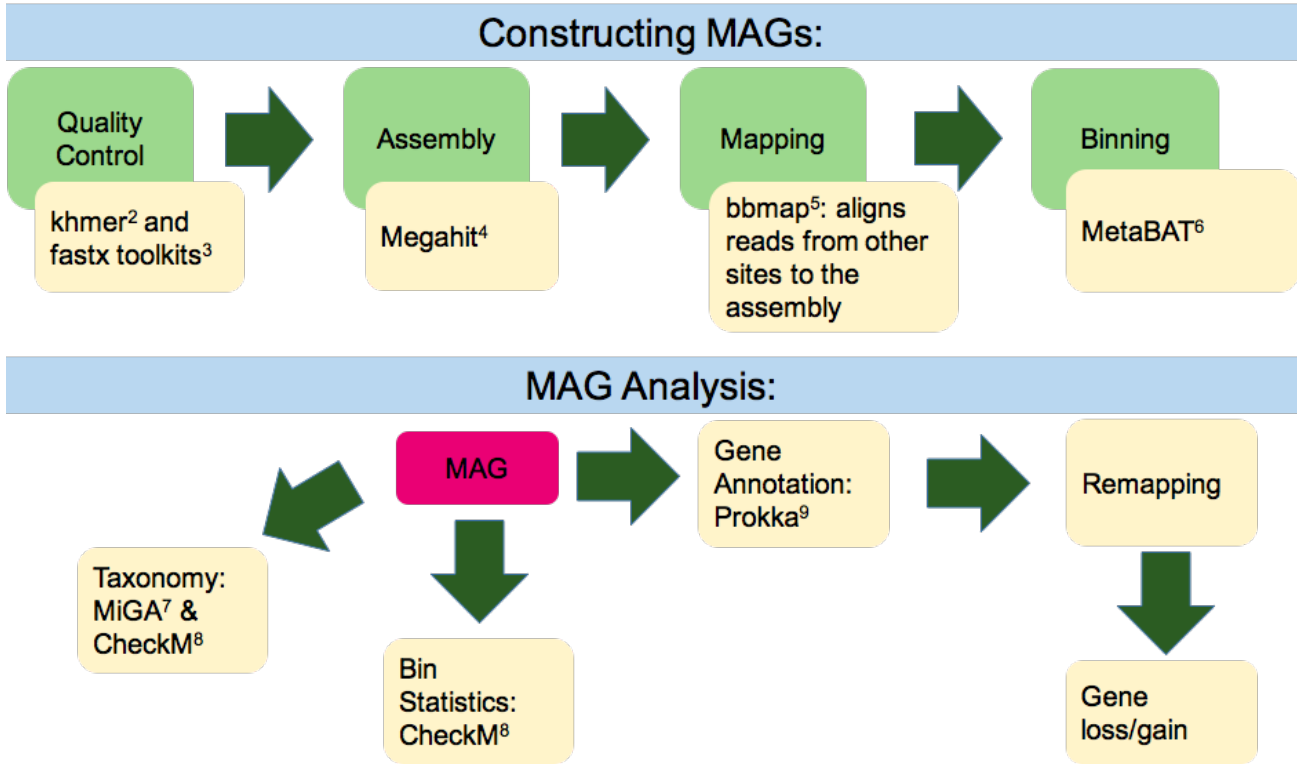
Uncultured DNA  
Extraction

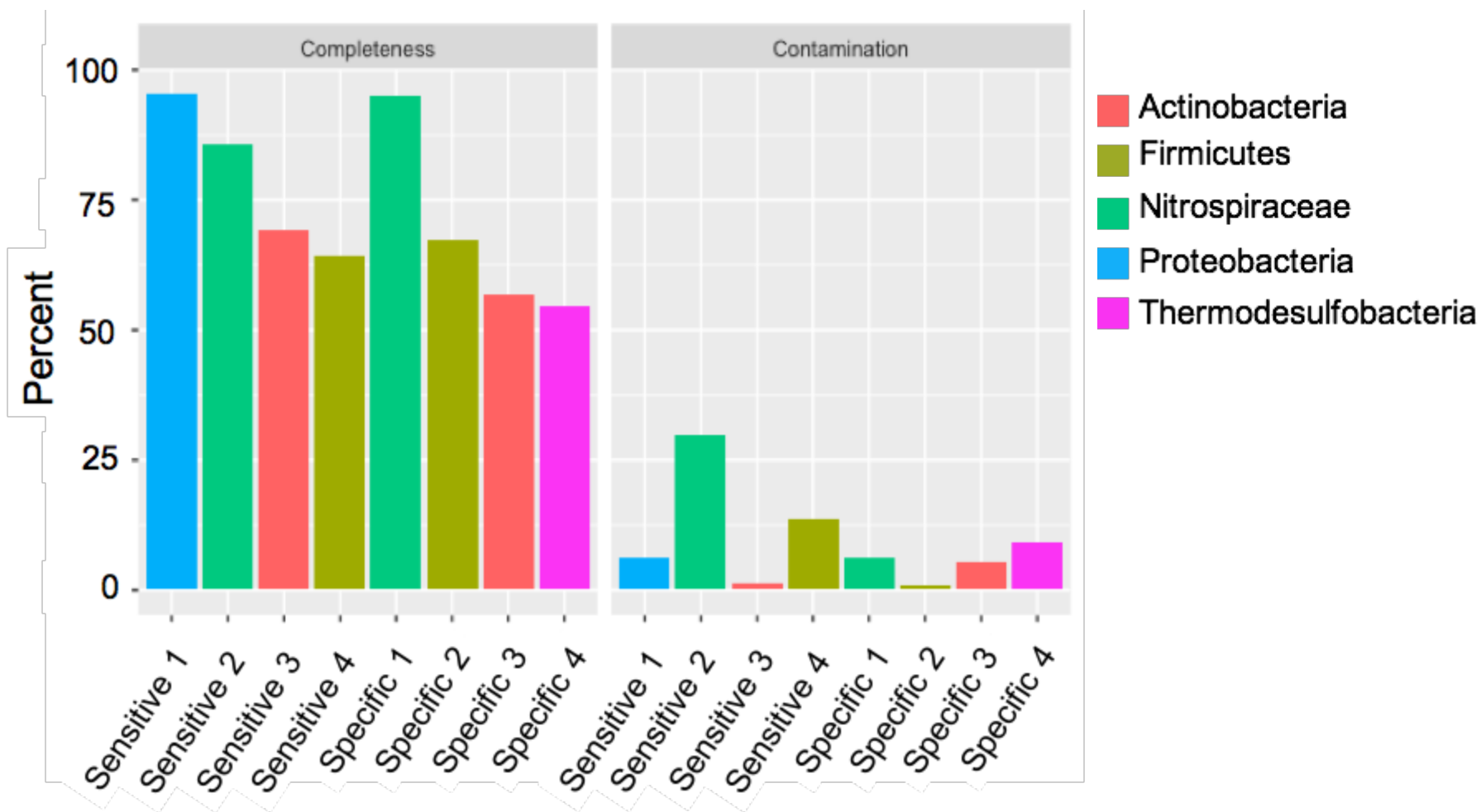
Cultured DNA  
Extraction

Sequencing

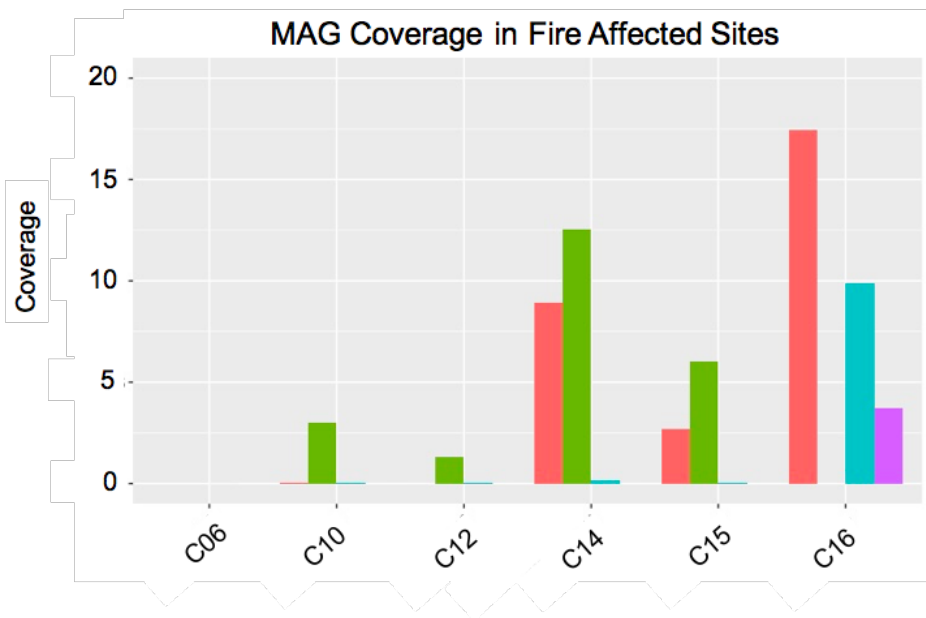


# Workflow



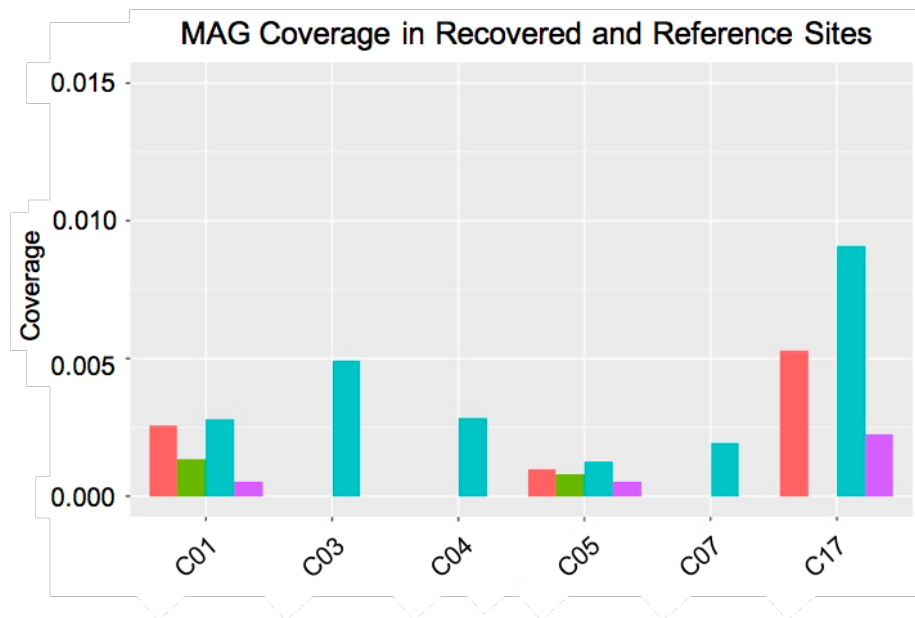


MAG Coverage in Fire Affected Sites

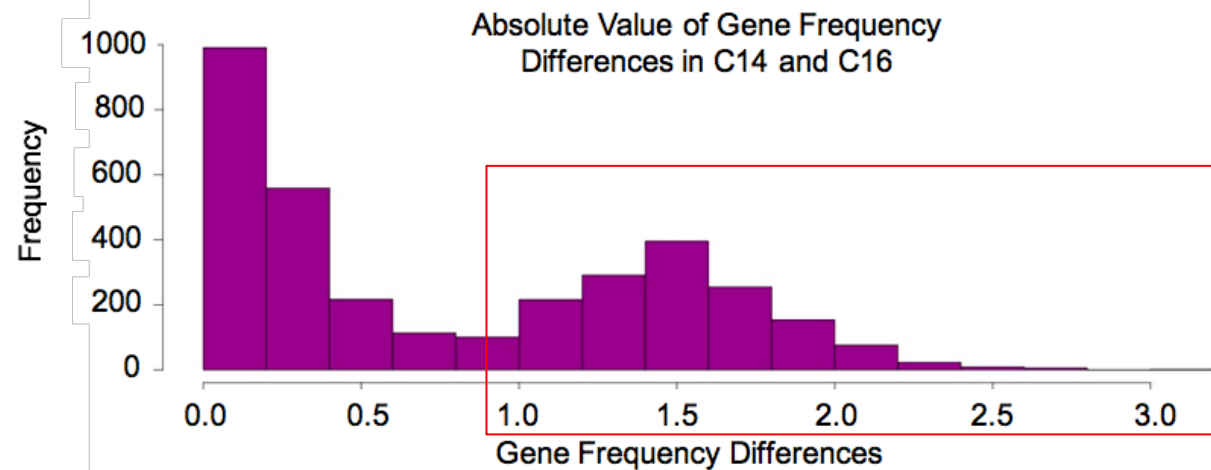
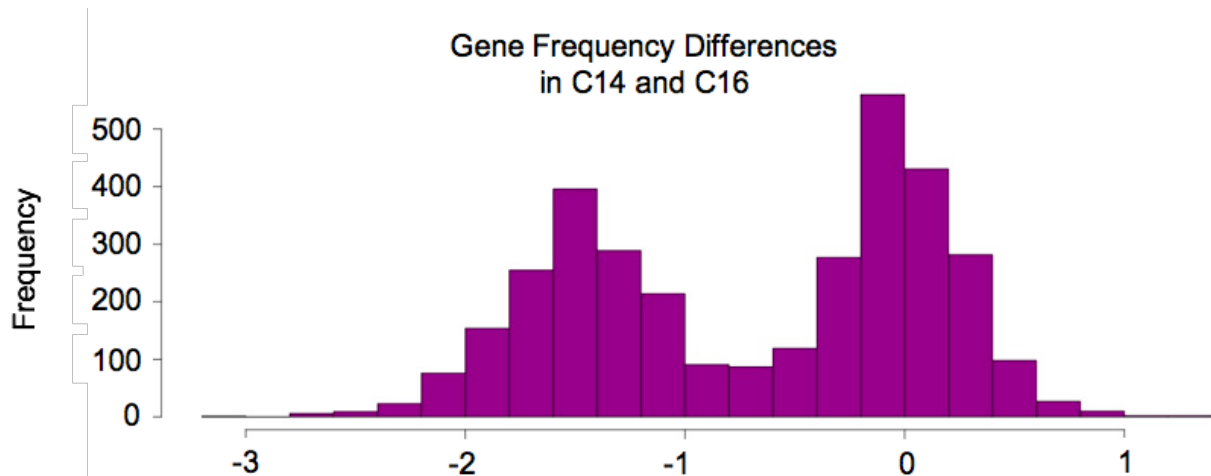


- Specific 1
- Specific 2
- Specific 3
- Specific 4

MAG Coverage in Recovered and Reference Sites







Gene frequency = 
$$\frac{\text{Gene coverage}}{\text{Median gene coverage of metagenome}}$$

# Conclusion

Within a MAG population, genes have different abundances at different sites suggesting loss or gain

Nearly complete and minimally contaminated MAGs can be binned from complex soil metagenomes

# Acknowledgements/References

Shade Lab

Dr. Ashley Shade

Jackson Sorensen

ACRES

Camille Archer

Brian O'Shea

Institute for Cyber-Enabled Research

Pat Bills

NSF

khmer and fastx toolkits

Megahit

bbmap

MetaBAT

MiGA

CheckM

Prokka