

# Genome 540 Discussion

Conor Camplisson

February 23<sup>rd</sup>, 2023

# Outline

- Homework 8 overview
- Update on image processing Snakemake demo
- Homework 7 & 8 questions

# Outline

- Homework 8 overview
- Update on image processing Snakemake demo
- Homework 7 & 8 questions

# Homework 8 Overview

## *Pyrococcus horikoshii*

### Scientific classification

Domain: Archaea  
Kingdom: Euryarchaeota  
Phylum: Euryarchaeota  
Class: Thermococci  
Order: Thermococcales  
Family: Thermococcaceae  
Genus: *Pyrococcus*  
Species: *P. horikoshii*

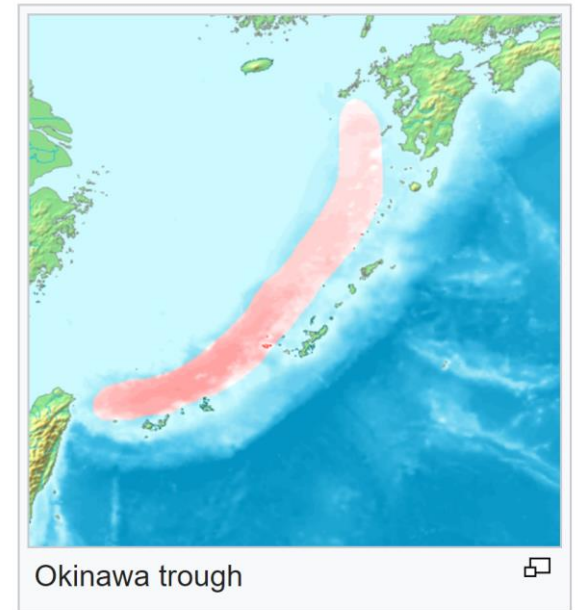
### Binomial name

*Pyrococcus horikoshii*

Erauso *et al.* 1993

## *Pyrococcus horikoshii*

- Hyperthermophile, 98 C !
- Anaerobic archaeon
- Isolated from Okinawa Trough
- Growth enhanced by Sulfur
- 32 min doubling time (growth rate)



package, 1996). The phylogenetic tree diagrams were generated by the PHYLIP suite of programs (Kuhner and Felsenstein 1994).

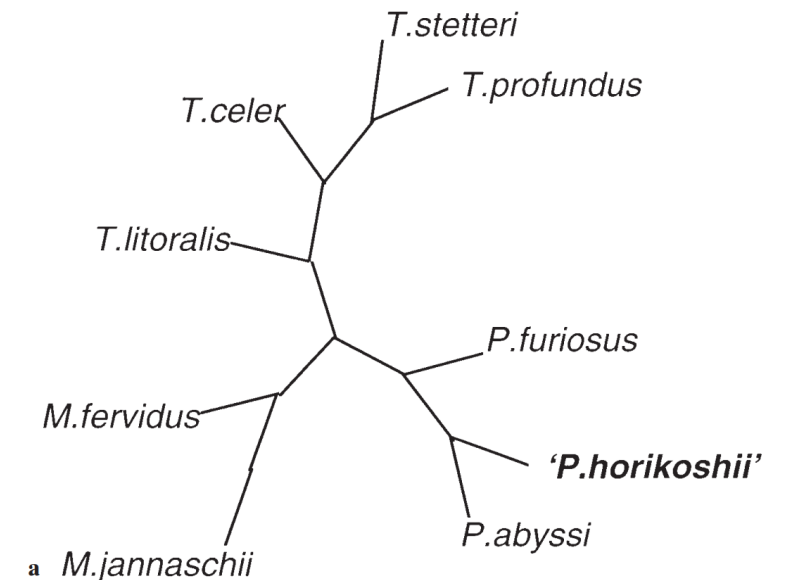
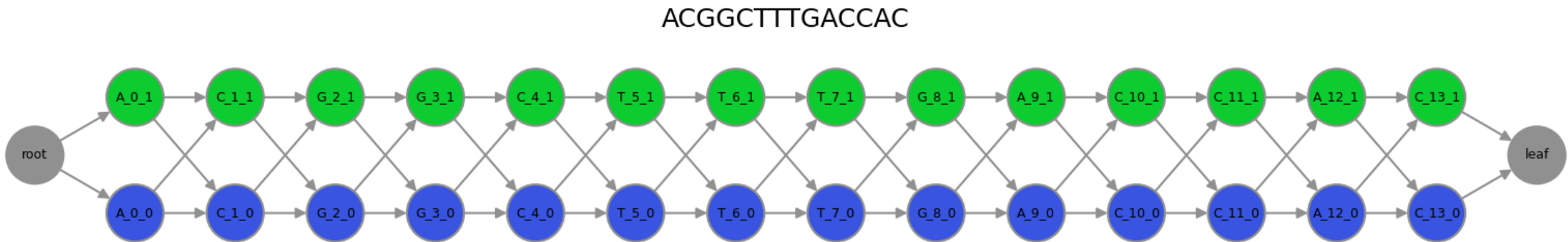


Fig. 3. Phylogenetic trees representing the relatedness of *P. horikoshii*

# Homework 8 Overview

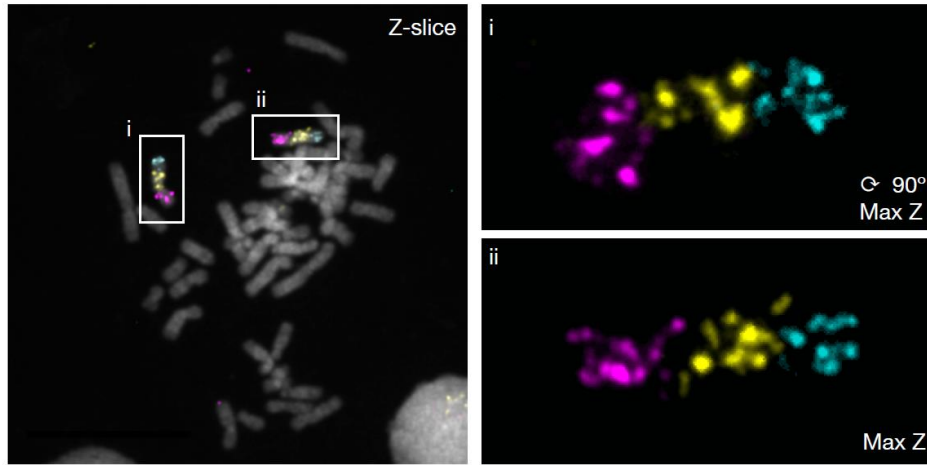
- 2-state HMM for detecting GC-rich regions in *Pyrococcus horikoshii* genome
  - state 1: AT-rich, state 2: GC-rich
  - starting parameters are given: initiation, transition, emission probabilities
- Use Baum-Welch training to find improved parameter estimates
  - each iteration: compute log-likelihood of the sequence, new probabilities
- Run until parameter estimates converge
  - stop when log-likelihood increase  $< 0.1$



# Outline

- Homework 8 overview
- Update on image processing Snakemake demo
- Homework 7 & 8 questions

# Snakemake Demo: Image Processing



## Pipeline Specification

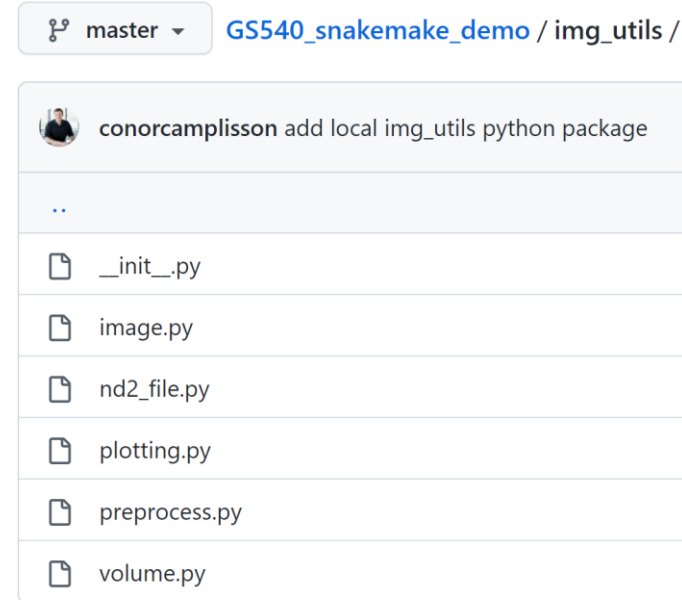
**Input:** .nd2 files (3D hyperstacks)

**Steps:** split channels, z-project, detect fluorescent objects (puncta), compute & plot stats

**Output:**

- plots of pixel intensity, spot size
- .csv file with stats per sample

Added img\_utils python package

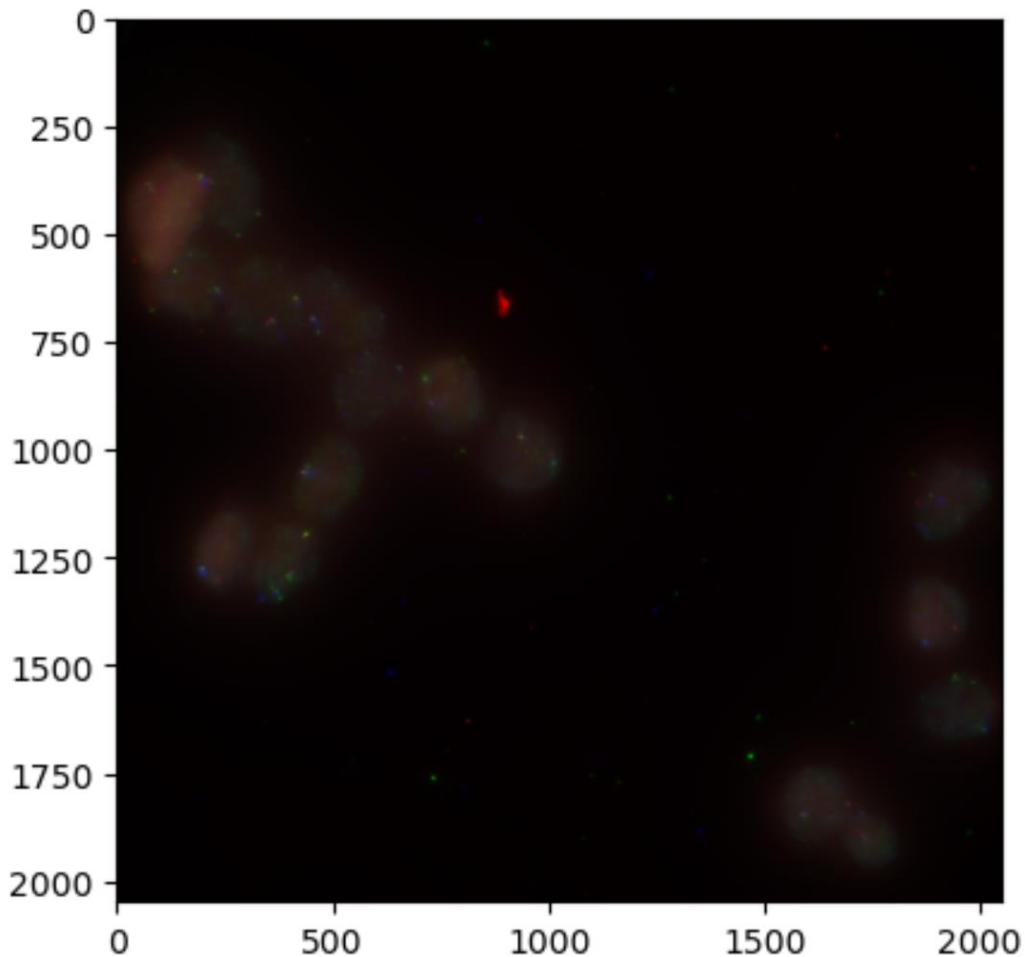


Access the demo pipeline repo:

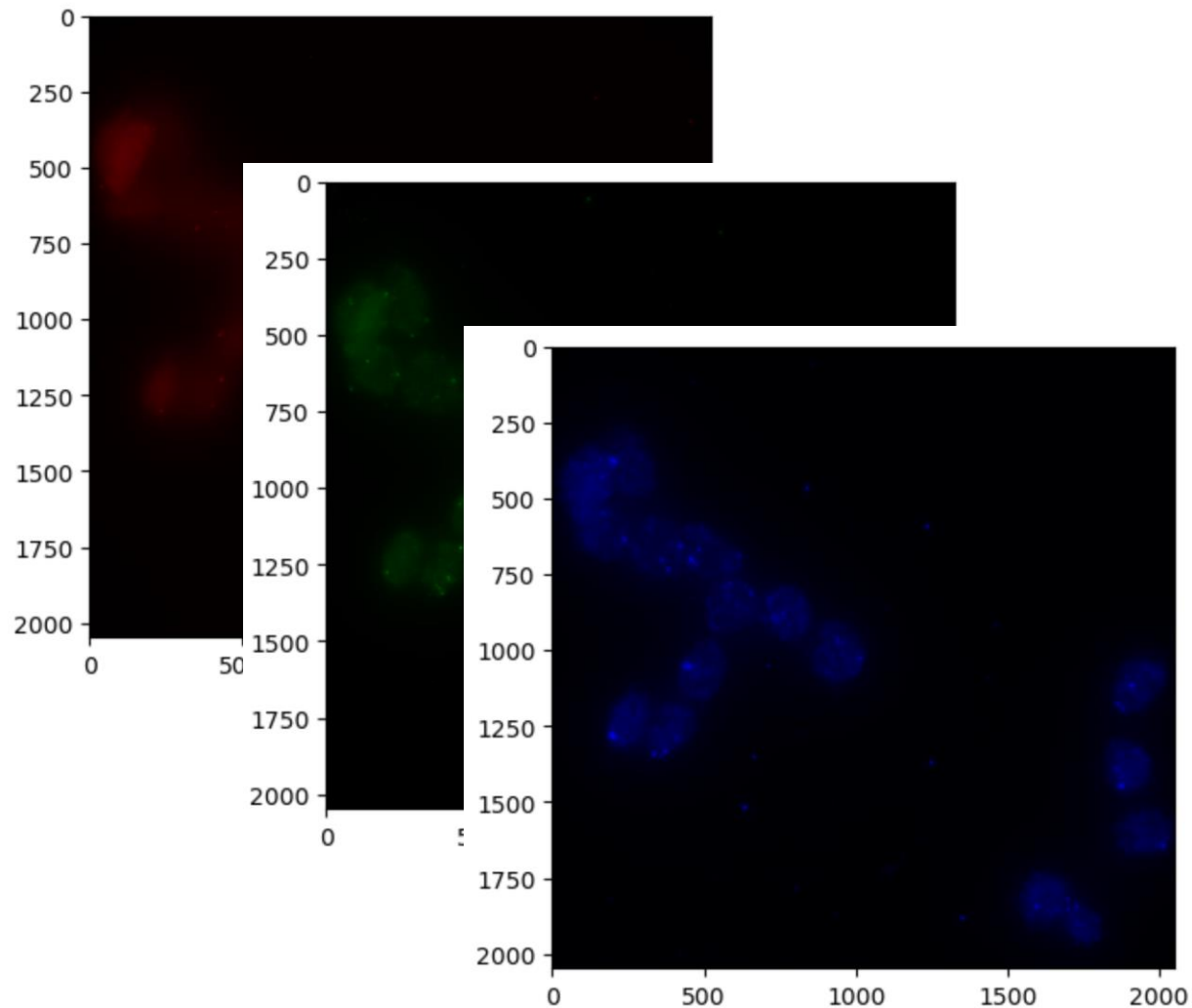
[https://github.com/conorcamlisson/GS540\\_snakemake\\_demo](https://github.com/conorcamlisson/GS540_snakemake_demo)

# Step 1: load .nd2, split fluor channels

3D multi-channel hyperstack



Individual channel 3D z-stacks

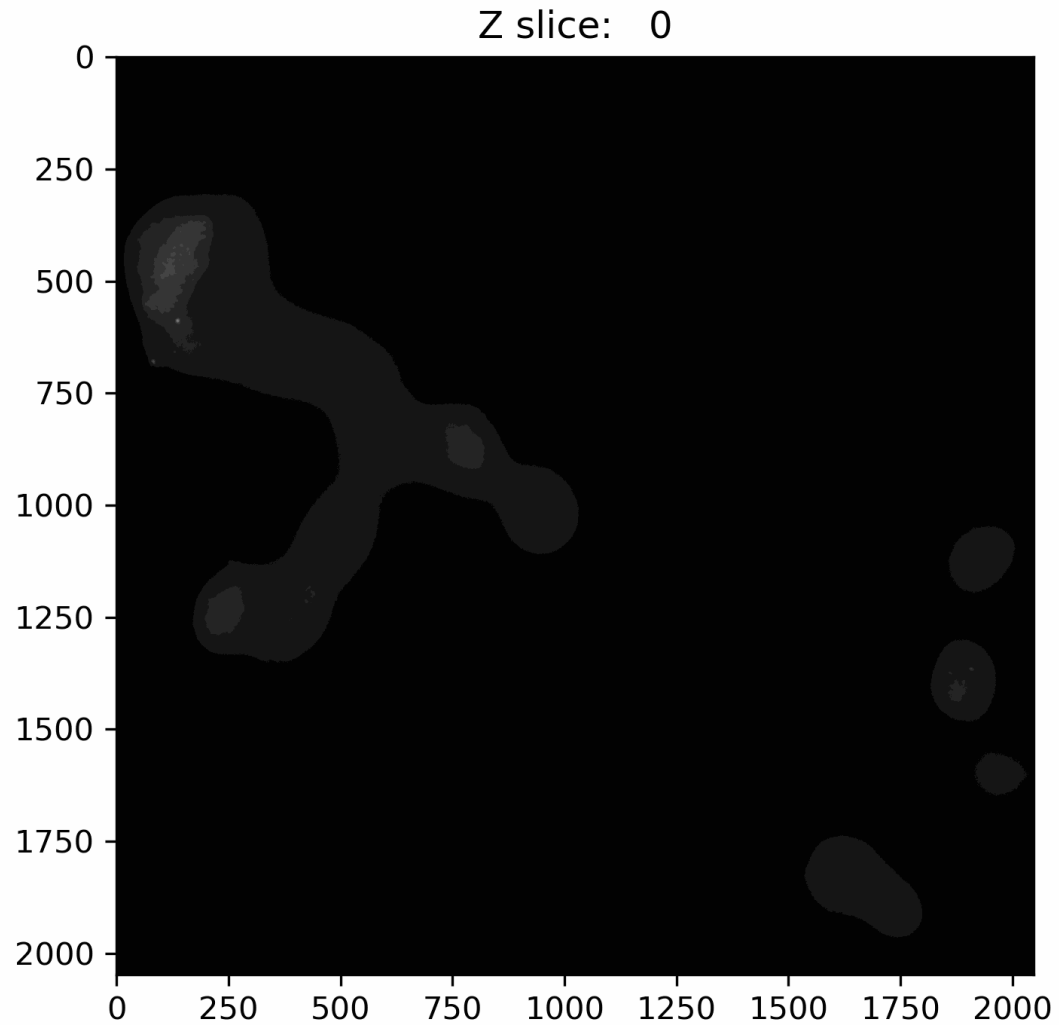


High level custom 'img\_utils' api uses nd2reader under the hood

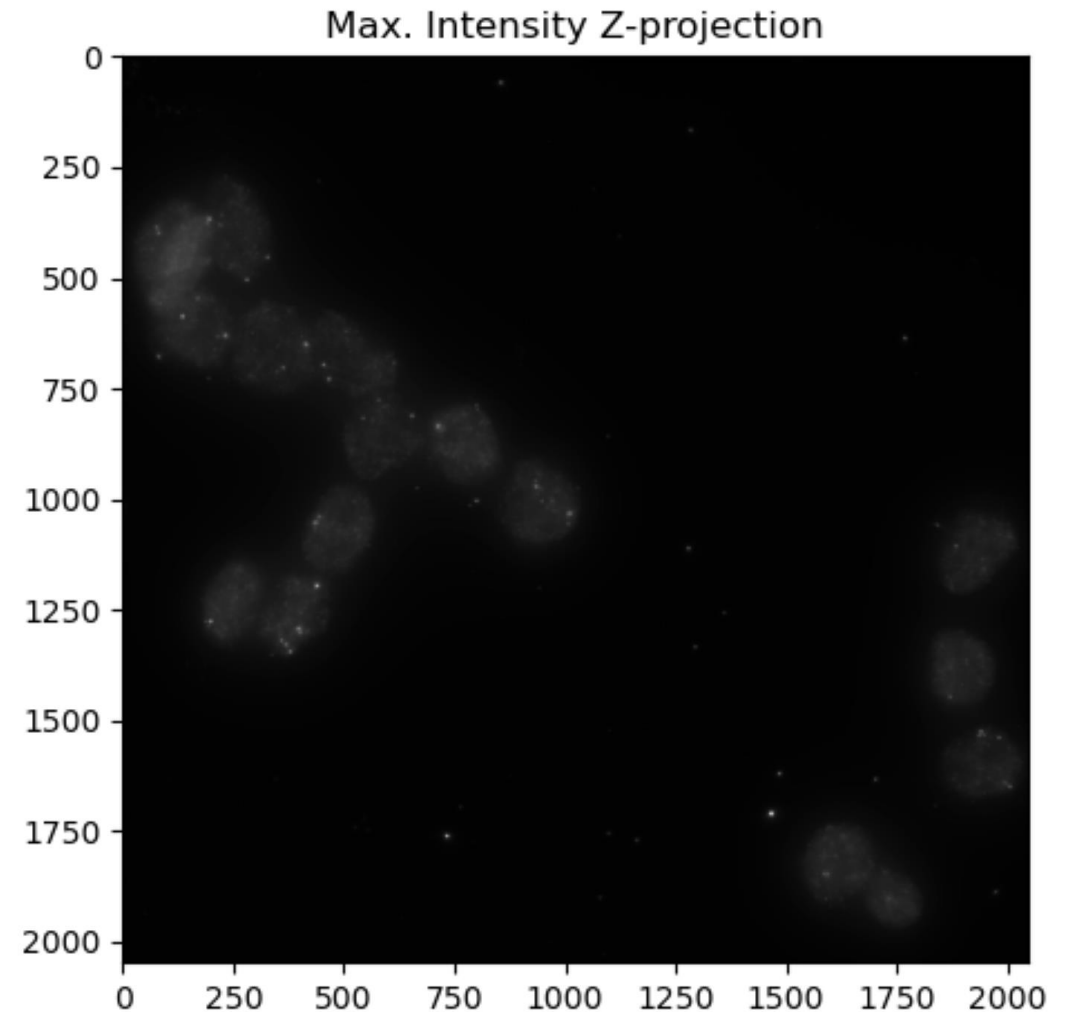


# Step 2: Max. z-project each channel

Individual channel 3D z-stacks

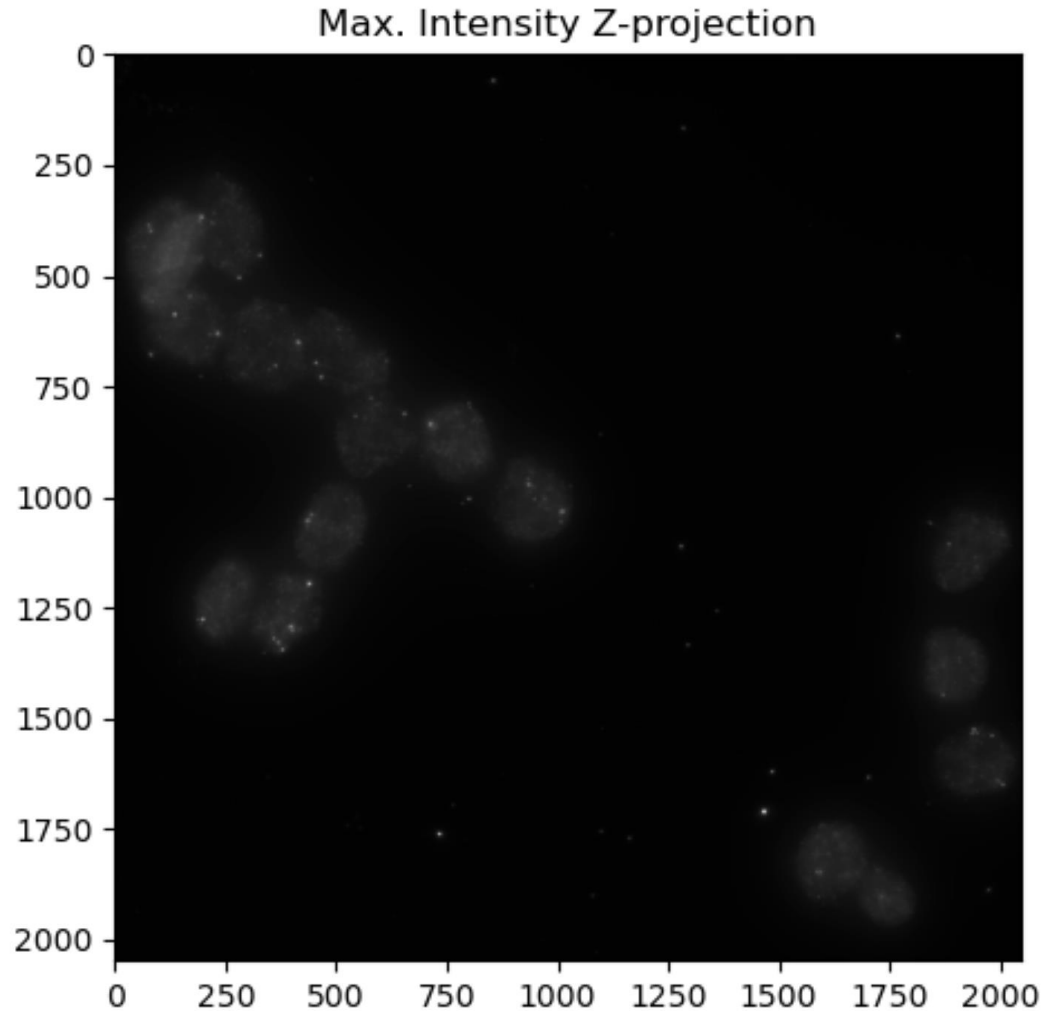


2D max z-projections

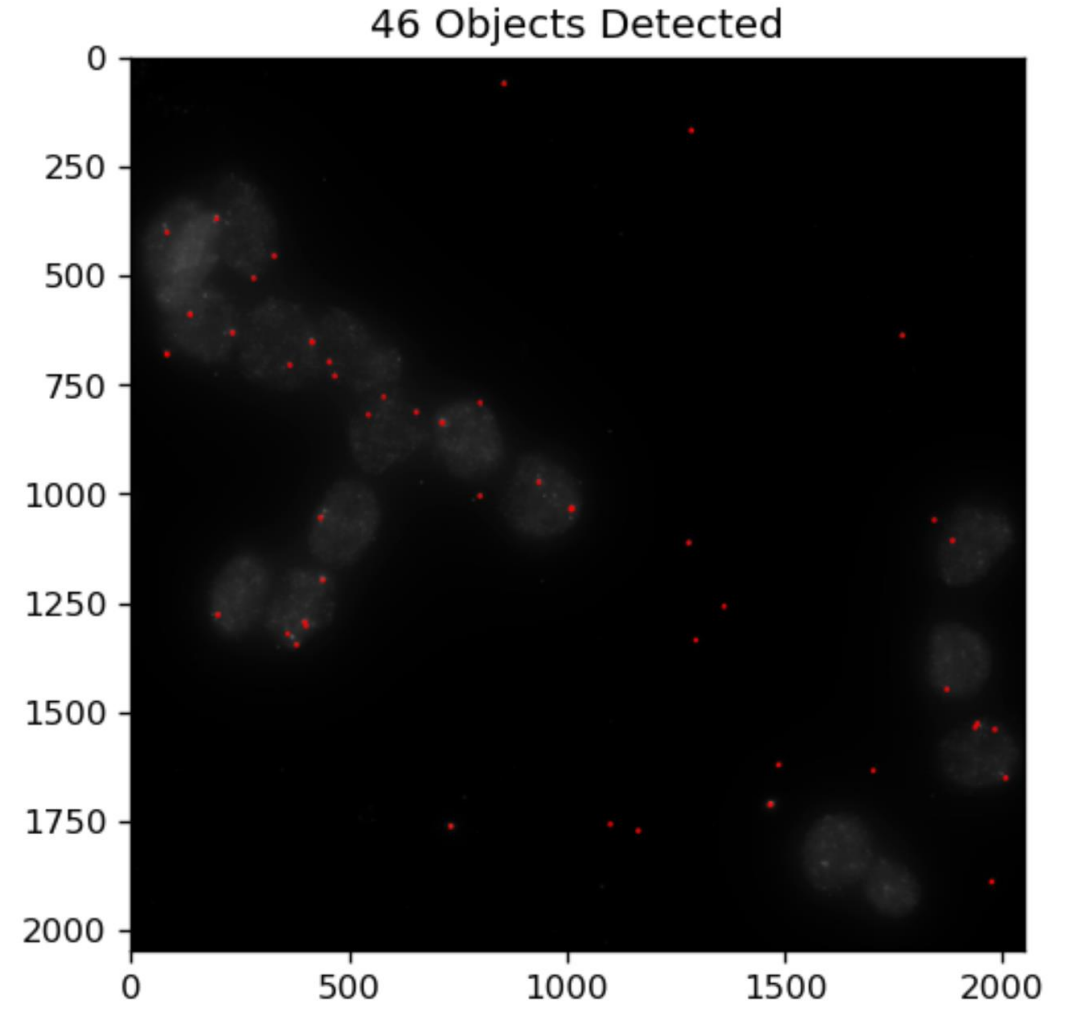


# Step 3: Find fluorescent objects

2D max z-projections



2D max z-projection



# Outline

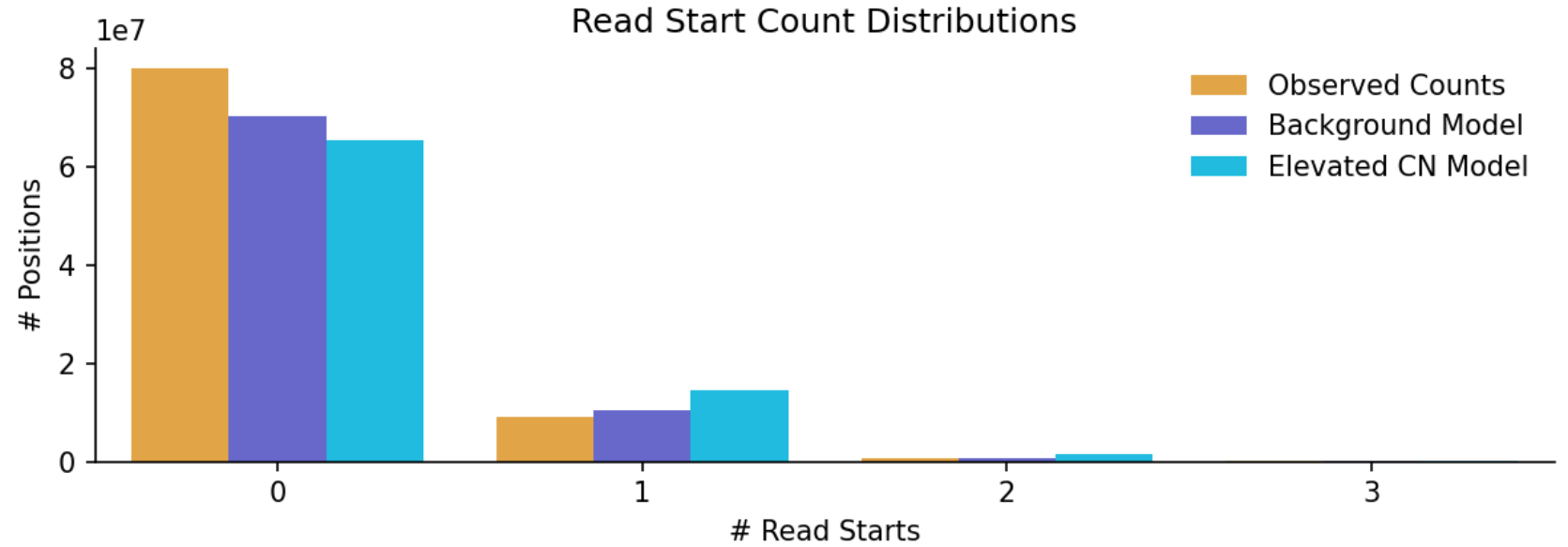
- Homework 8 overview
- Update on image processing Snakemake demo
- **Homework 7 & 8 questions**

# Homework 7 & 8 Questions ?

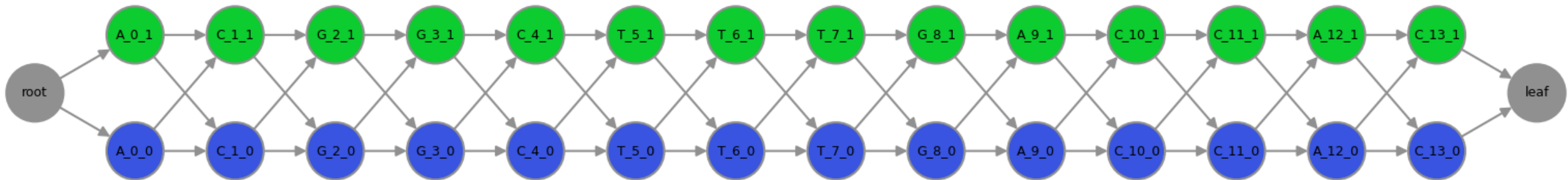
chm13.chr16.txt

```

16      1      0
16      2      0
16      3      0
16      4      0
16      5      0
[ ... ]
16     14793    0
16     14794    1
16     14795    3
16     14796    0
[ ... ]
    
```



ACGGCTTTGACCAC



# Reminders

- Homework 7 due this Sunday Feb. 26, 11:59 pm
- Homework 8 due this Sunday Mar. 5, 11:59 pm

