Comparison of Gene Finding Tools in the Context of *Trichoderma* Genomes

Committee members: Dave Schneider, Tony Kusalik, Matthew Links, Leon Kochian

Connor Burbridge

May 29, 2023

Background: *Trichoderma*

What is *Trichoderma*?

▶ *Trichoderma* is an opportunistic symbiotic fungi, which can colonize the roots of plants

▶ *Trichoderma* strains have been shown to provide several benefits to the host plant it colonizes, those generally being:
  ▶ Increased resistance to abiotic and biotic stressors
  ▶ Facilitating nutrient uptake
  ▶ Increased germination rates

▶ These benefits have resulted in *Trichoderma* being used in manufacturing processes for antibiotics and other materials

Background: Previous GIFS Work

Two strains have been sequenced in previous work within GIFS:

- ▶ These strains have been named DC1 and Tsth20
- ▶ Strains from the prairie regions of Canada, including Alberta and Saskatchewan
- ▶ How exactly do these processes work? Which genes are included in these processes?
- ▶ To answer these questions, both strains were sequenced with Illumina and Nanopore technologies

Research Problem

These sequenced strains offer an opportunity to assemble and annotate them:

- ▶ Genome assembly is 'relatively' straight-forward
- ▶ **However, the choice of a tool for gene finding or annotation is uncertain**
- ▶ There has been relatively little comparative analysis for gene finding tools in fungi, and even fewer for *Trichoderma*
- ▶ **This raises questions. How do different gene finding tools perform in fungi and Trichoderma in particular?**

Project Goal

This project aims to evaluate several different gene finding tools in the context of Trichoderma genomes

- ▶ Gene finding tools currently selected are GeneMark-ES and Braker2
- ▶ These tools include a mix of *ab initio*, evidence-based and hybrid gene finding methods
- ▶ This list is not final and may include more tools if desired or necessary

Methodology
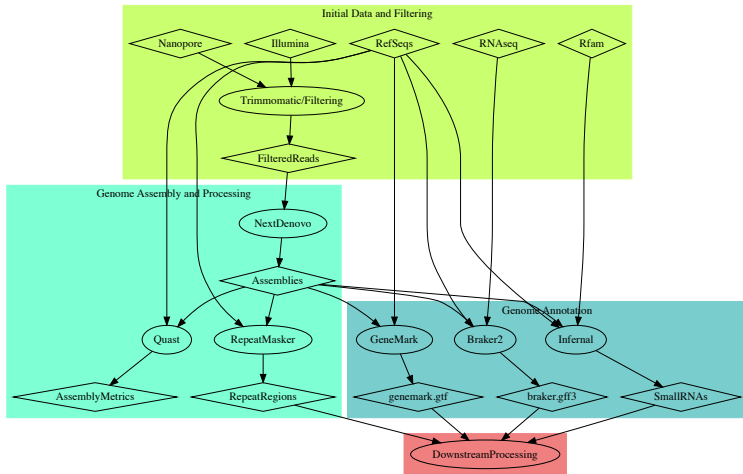
Evaluation of Gene Finding Tools

- ▶ A plan for evaluating and comparing the selected tools needs to be developed
- ▶ Metrics for comparison will include both quantitative and qualitative observations

Quantitative Metrics

- Total genes predicted
- Total transcripts predicted
- Genes predicted in repetitive regions
- Genes overlapping smallRNAs
- Length of gene models predicted
- Comparison to genes predicted in other fungal species (Yeast)
- Run times and memory usage

Qualitative Metrics

- Features of the gene finding tools
- Ease of software installation and their dependencies
- Ease of use
- Popularity among other research