

# Committee meeting for Connor Burbridge

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

Connor Burbridge

November 14, 2022

# Project Proposal Outline

- ▶ Background
- ▶ Research Problem
- ▶ Current Progress
- ▶ Timeline
- ▶ Deliverables

# 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
- ▶ Because of *Trichoderma*'s potential to produce materials which aid in resistance to bacteria and other fungi in soil, several *Trichoderma* strains have been well studied and are employed in the use of manufacturing antibiotics and other materials

## Background: Previous GIFS Work

Two strains have been sequenced in previous work within GIFS:

- ▶ DC1 and Tsth20
- ▶ Strains from the prairie regions of Canada, including Alberta and Saskatchewan
- ▶ One strain has been shown to provide resistance for plants growing in soils with high salt content, the other has potential for use as a bioremediation agent in soils contaminated with hydrocarbons
- ▶ How exactly do these processes work? Which genes are included in these processes?
- ▶ To lay groundwork for better understanding, we must identify potential genes in these strains
- ▶ 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, GenomeThreader, 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

# Evaluation of Gene Finding Tools

A methodology for evaluating and comparing the selected tools needs to be developed. Metrics for comparison will include:

- ▶ Gene finding features
- ▶ Efficiency of selected gene finding tools (i.e. runtimes, memory requirements)
- ▶ Requirements of selected gene finding tools and ease of installation
- ▶ Comparison ('validation') of called genes with existing RNASeq data if available
- ▶ Identification of small RNAs and genes in repetitive and AT-rich regions
- ▶ Distribution of lengths of called genes (particularly in the case of *ab initio* gene finders)

# Current Progress

Preliminary assemblies for both DC1 and Tsth20 are ready

- ▶ Two assemblies for each strain using both SPAdes and MaSuRCA
- ▶ Both assemblers use a hybrid assembly approach



# Assembly Metrics

## SPAdes

Strain	Total Contigs	Total Length	Largest Contig	GC%	N50	L50
Tsth20	611	41.88 Mb	2.44 Mb	47.28	1.17 Mb	14
DC1	181	38.60 Mb	1.85 Mb	47.95	807.44 Kb	17

## MaSuRCA

Strain	Total Contigs	Total Length	Largest Contig	GC%	N50	L50
Tsth20	8	41.52 Mb	9.97 Mb	47.36	4.96 Mb	3
DC1	13	38.60 Mb	7.36 Mb	47.96	4.06 Mb	4

## Existing Progress Cont.

Place images of genome alignments and GC content here

# Next Steps

- ▶ Finish assemblies of DC1 and Tsth20
- ▶ Identify repetitive regions of selected assemblies
- ▶ Identify non-coding RNAs in selected assemblies
- ▶ Apply gene finders to selected assemblies and evaluate

# Deliverables

- ▶ Assemblies of both Tsth20 and DC1
- ▶ Lists of potential genes for each *Trichoderma* assembly considered
- ▶ A consensus or 'core' genome for genes called by all gene finders
- ▶ Repetitive regions identified in all *Trichoderma* genomes considered
- ▶ Potential true positives supported by RNAseq evidence and existing annotations
- ▶ Final comparative tables including the evaluation metrics described previously

# Timeline

- ▶ Finishing assemblies of DC1 and Tsth20 assemblies (2 weeks)
- ▶ Collection of existing genome assemblies (1 week)
- ▶ Application of gene finding tools to selected genomes (1-2 months)
- ▶ Downstream analysis of gene finding results (1-2 months)

Questions and/or comments?