Thesis Defence: Evaluation of Gene Finding Tools When Applied to *Trichoderma* Genomes

January, 2026

Connor Burbridge

UNIVERSITY OF SASKATCHEWAN

Outline

- *Trichoderma* fungi.
- Novel *Trichoderma* genomes.
- Motivation.
- Research objectives.
- Workflow overview.
- Results.
- Conclusions and recommendations.
- Questions.

What is *Trichoderma*?

- *Trichoderma* is a genus of filamentous fungi that is ubiquitous in soil and plays a significant role in nutrient cycling.
- They are also used in biocontrol and as biofertilizers.
- Known for their production of plant cell wall degrading enzymes, and **secondary metabolites**.
- Further genomic studies can help in understanding their biology and potential applications.
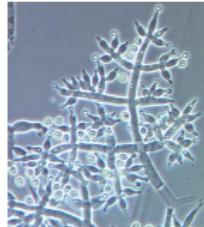
Figure: *T. harzianum*
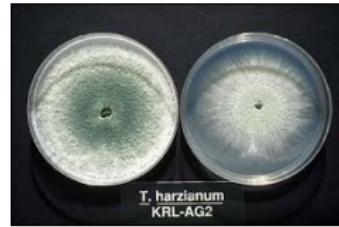
Figure: *Trichoderma* colony

Novel *Trichoderma* Genomes

- *Trichoderma* species are diverse, with many species not yet fully characterized.

- DC1 and Tsth20, shown to improve drought and salt tolerance when applied to crops, and have been shown to breakdown hydrocarbons in soils.

- Recent advances in sequencing technology have made it possible to generate high-quality genomes for these species.

Motivation
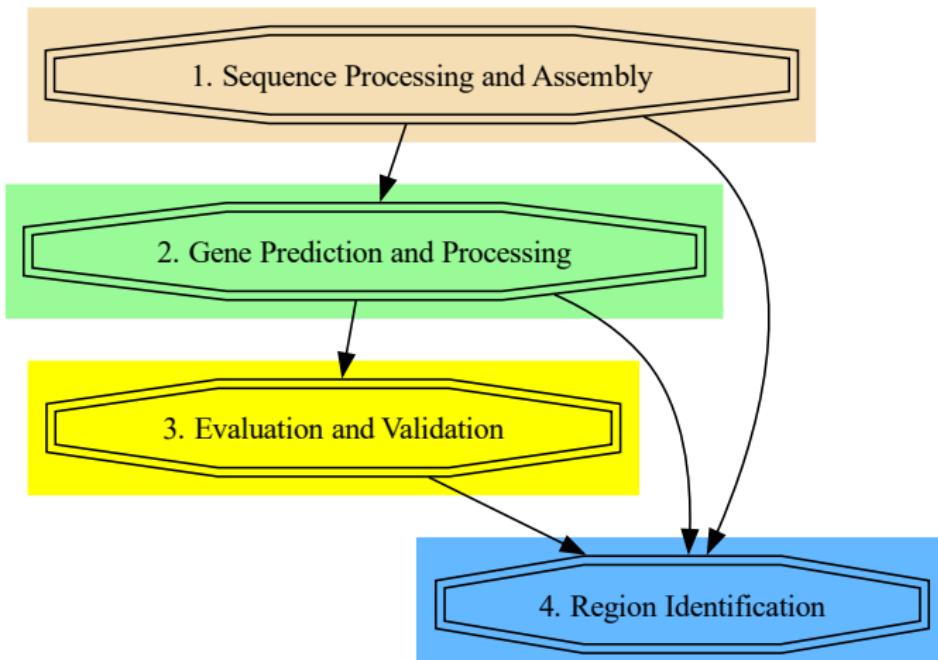
- To better understand the biological mechanisms at work in *Trichoderma* spp., we first need to understand their genomes by identifying genes within them.

- Various tools exist for gene prediction, but their implementations and performance can vary significantly.

- **Few studies have compared these tools in fungi, and even fewer in *Trichoderma*.**

- In addition, increased accessibility to high-quality sequencing services has led to the generation of many new *Trichoderma* genomes, including DC1 and Tsth20.

- These genomes provide an opportunity to evaluate gene finding tools in a comparative context, and contribute to the understanding of *Trichoderma* biology.

Research Objectives

- Assemble and evaluate novel assemblies of DC1 and Tsth20.
- Apply gene finders Braker2 and GeneMark in DC1, Tsth20, and three other RefSeq assemblies - *T. reesei, T.harzianum* and *T. virens*.
- Compare gene finding tools based on relevant criteria, including:
  - Proportions of gene lengths predicted.
  - Presence of functional domains and closely related protein sequences.
  - Presence of genes in AT-rich sequence.
  - Agreement of gene finders on start and stop positions of a gene.

Workflow Overview

1. Sequence Processing and Assembly

2. Gene Prediction and Processing

3. Evaluation and Validation

4. Region Identification

Acknowledgements

- My supervisors for their committed support through COVID and other challenges.
- The Global Institute for Food Security for providing the data for this project as well as a portion of my funding.
- My committee members for their feedback and support.

Datasets and Tools

- **Datasets:**
  - Novel *Trichoderma* genomes: DC1 and Tsth20.
  - Reference genomes and annotations: *T. reesei, T. harzianum, T.virens.*
  - RNAseq training data from *T. reesei.*
  - Benchmarking Universal Single-Copy Orthologs (BUSCO) fungal database.
  - Protein sequence queries for tblastn from *T. atroviride, Fusarium graminearum, and Saccharomyces cerevisiae.*
- **Tools:**
  - Sequence processing: FastQC, Trimmomatic, Hisat2.
  - Genome assembly: NextDenovo and NextPolish.
  - Gene finding tools: Braker2 and GeneMark-ES.
  - Evaluation tools: BUSCO, tblastn, InterProScan, and custom scripts.

Category	Braker2	GeneMark	RefSeq
Availability	3	3	0
Ease of install	1	2	0
Ease of use	3	3	0
# of genes predicted	0	3	3
# of transcripts predicted	3	0	2
Predicts shortest genes	2	1	0
Predicts more shorter genes	1	0	3
BUSCO Performance	2	1	3
Performance in AT-rich sequence	2	1	3
Predictions with InterProScan support	3	3	3
Final Score (Publicly Available)	20	17	N/A
Final Score (Ignoring Availability)	13	9	17

Image Credits

- *T. harzianum* image:
  `https://en.wikipedia.org/wiki/Trichoderma`
- *Trichoderma colony* image:`https://biocontrol.`
  `entomology.cornell.edu/pathogens/trichoderma.php`