

COMPARATIVE ANALYSIS OF GENE FINDING TOOLS WHEN
APPLIED TO *Trichoderma* GENOMES

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Science
in the Applied Computing of Computer Science
University of Saskatchewan
Saskatoon

By
Connor Burbridge, Dave Schneider, Tony Kusalik

©Connor Burbridge, Dave Schneider, Tony Kusalik, All rights reserved.
Unless otherwise noted, copyright of the material in this thesis belongs to
the author.

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Disclaimer

Reference in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building, 110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

placeholder

Acknowledgements

I would like to acknowledge both Dr. Kusalik and Dave Schneider for their excellent support and mentorship during both COVID and my MSc. project. I would also like to thank Brendan Ashby and Dr. Leon Kochian for providing both data and additional financial support.

Contents

Permission to Use	i
Abstract	ii
Acknowledgements	iii
Contents	iv
List of Tables	v
List of Figures	vi
List of Abbreviations	vii
1 Background	1
1.1 <i>Trichoderma</i> Species and their Evolution	1
1.2 Novel <i>Trichoderma</i> Genomes	2
1.3 Gene Predictions and Complementing Similarity Searches	2
1.4 Secondary Metabolites	4
1.5 Genome Assembly	4
1.6 Identification of Anomalous Genomic regions	5
2 Research Questions	6
2.1 Research Questions	6
2.2 <i>Trichoderma</i> Assembly Results	6
2.3 Profiling of Gene Finding Tools	6
2.4 Number of Features Predicted	7
2.5 Lengths of Predicted Genes	7
2.6 Identifying Regions of Agreement and Disagreement	7
2.7 Validation of Predicted Genes via InterProScan	7
2.8 Similarity searches of Predicted Genes with tblastn	8
2.9 Performance in Regions of Anomalous Sequence Content	8
2.10 BUSCO Completeness	8
2.11 Selection of a Gene Finding Tool	8
References	9

List of Tables

List of Figures

1.1	Example of a eukaryotic gene structure, showing a promoter region (blue), the 5' and 3' UTRs (purple), start codon (green), exons (orange), and introns (grey). The directed arrow indicates the direction of transcription.	3
-----	--	---

List of Abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
CDS	Coding sequence
Mb	Megabases
Kb	Kilobase
WGS	Whole Genome Shotgun (sequencing)
NGS	Next generation sequencing
PCR	Polymerase chain reaction
GMO	Genetically modified organism
CPU	Central processing unit
GC	Guanine cytosine
HMM	Hidden Markov model
UTR	Untranslated region
GFF	General feature format
BUSCO	Benchmarking Universal Single-Copy Orthologs
RSMI	Root, Soil and Microbial Interactions
MITE	Minature inverted-repeat transposable element
TIR	Terminal inverted repeat
TDR	Terminal direct repeat
maybe	Should I include tools like BLAST and resources like NCBI?

1 Background

Chapter one provides a brief background on *Trichoderma* and the context in which is being studied. Section 1.1 discusses the evolutionary mechanisms that may have led to the increased production of secondary metabolites in *Trichoderma* species, such as horizontal gene transfer, gene duplication, and transposable elements.

1.1 *Trichoderma* Species and their Evolution

Trichoderma are a genus of ascomycete fungi found in a wide variety of soils, and are well-known for their use in biomanufacturing of cellulases and hemicellulases as well as their roles as non-toxic, avirulent opportunistic plant symbionts [17] [9]. Many of the features of *Trichoderma* can be attributed to the large number of secondary metabolites produced by these fungi, which are used to interact with plants and other organisms in the soil [12]. How some *Trichoderma* species evolved to produce more secondary metabolites than others, is a subject of frequent study.

There are many mechanisms organisms may leverage that result in gain or loss of function, and ultimately evolution of a species. One well-studied mechanism is that of horizontal gene transfer (HGT), which is the process of acquiring genetic material from another organism, rather than through inheritance from a parent organism [6]. HGT is a common mechanism in bacteria, but it has also been observed in fungi, including *Trichoderma* species [6]. Another mechanism of interest is the duplication of genes, which can result in the production of more than one copy of a gene, leading to increased expression of that gene [6]. This is particularly relevant in the case of secondary metabolite production, as many *Trichoderma* species have been shown to have multiple copies of genes involved in secondary metabolite production [12]. Lastly, transposable elements (TEs) are another mechanism of interest, as they can insert themselves into the genome and disrupt or enhance the expression of genes [6]. TEs have been shown to play a role in the evolution of *Trichoderma* species, particularly in the case of secondary metabolite production [6].

Interestingly, both HGT and TEs tend to present themselves in regions where GC content differs from the rest of the genome, [6]. Abnormal GC content is also associated with other genomic features, such as centromeres [14] and repetitive regions [16], both of which may have an effect on the production of secondary metabolites in *Trichoderma* species. As a result, it is important to consider the GC content of a genome when studying *Trichoderma* species, which forms the basis of the Research Question 2.9. Other studies have shown that *Trichoderma* species contain very few if any transposable elements, which is unusual for fungi [8].

It is hypothesized that TEs were lost in *Trichoderma* due to repeat induced point mutations (RIP), which is a process that occurs in fungi where TEs are silenced and eventually lost from the genome [8]. Genes responsible for RIP mutations also tend to target CA dinucleotides, converting them to TA nucleotides, which explains the decreased GC content observed in some regions of *Trichoderma* genomes [6].

Given the proximity of *Trichoderma* species to other fungi, bacteria, and plants in soils, it is possible that *Trichoderma* species acquired genes which are involved in response to antagonistic and sympathetic intercellular interactions from other organisms. While difficult to prove, studies have shown that HGT events have occurred in a number of fungal species [5]. It is also important to note that some *Trichoderma* species have higher numbers of genes involved in secondary metabolite production than others, indicating an evolutionary divergence, making comparative analysis of *Trichoderma* species and strains a useful tool for understanding the mechanisms behind secondary metabolite production [12].

1.2 Novel *Trichoderma* Genomes

Recently, two strains of *Trichoderma* have been identified in the prairie regions of Alberta and Saskatchewan. These two strains, named Tsth20 and DC1, have been found to have beneficial properties when used as an inoculant for plants in the soils mentioned before. In addition to these beneficial properties, the two strains mentioned previously provide even further protection for plants in dry, salty soils and one strain also has potential for use as a bioremediation tool in soils contaminated with hydrocarbon content. Bioremediation and resistance to drought tolerance has also been investigated in other strains of *Trichoderma* as well [15]. However, little is known about the mechanisms at work in these strains, so DC1 and Tsth20 were sequenced by the Global Institute for Food Security (no publication yet) in an initial attempt to better understand the details of these genomes. While this research does not explicitly identify genomic elements related the beneficial properties of these genomes, it may serve as a foundation for future research of *Trichoderma*. The assembly of these genomes come as a result of Research Question 2.2.

1.3 Gene Predictions and Complementing Similarity Searches

Gene prediction methods are generally based on the idea that genes can be identified by searching for patterns in a sequence that match an expected gene structure or model. These gene structures begin from a 5' start codon, continue through a series of exons and introns, and end with a 3' stop codon [10]. Flanking the gene structure are promoter regions, which are typically found upstream of the start codon, and untranslated regions (UTRs), which are found both upstream and downstream of the start and stop codons. Promoter regions are important for the regulation of gene expression, while UTRs are important for the stability and translation of the mRNA [10]. Sequences are translated from the 5' start codon through to the 3' stop codon, splicing together exons and ignoring introns. An example of a eukaryotic gene structure is shown in Figure 1.1.

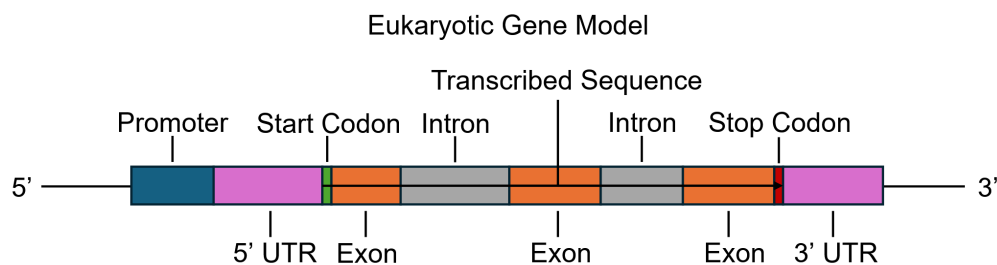


Figure 1.1: Example of a eukaryotic gene structure, showing a promoter region (blue), the 5' and 3' UTRs (purple), start codon (green), exons (orange), and introns (grey). The directed arrow indicates the direction of transcription.

Gene prediction methods vary in their approach, but they generally fall into two categories: *ab initio* methods and evidence-based methods [4]. *Ab initio* methods rely on the identification of patterns in the sequence that match an expected gene structure, while evidence-based methods use prior information such as RNAseq data, expressed sequence tags (ESTs), and expressed protein sequences to identify genes within a new genome [4]. Gene prediction methods can also vary in their ability to identify different components of a gene, such as the promoter region, UTRs, and introns. Some methods may only identify the coding sequence (CDS) of a gene, while others may also identify the promoter region and UTRs [4]. Making things more complicated, gene finders may predict differing combinations of exons and introns, leading to different predicted transcripts for the same gene. Different transcripts may produce different proteins, or the same protein with different post-translational modifications, which can significantly impact the function of the resulting protein [4]. Given the already complex nature of secondary metabolites, it is important to consider the impact of variability in gene prediction methods, especially in the case of *Trichoderma* species and their secondary metabolite production. This consideration is one of the main motivators behind this research, and forms the basis of Research Question 2.6.

While the basic structure of a gene may be present in a sequence, that does not mean that the gene codes for a functional protein, or in the case of secondary metabolites, a component that helps produce them. As a result, gene prediction methods are often used in conjunction with similarity searches, which compare the predicted genes to known genes in other organisms to identify potential functions, and serve as a form of validation for the predicted genes [10]. Similarity searches can come in several different forms, such as blast searches, which compare the predicted genes to a database of known proteins, or InterProScan, which compares the predicted genes to a database of known protein domains and binding motifs [10]. These similarity searches can provide additional information about the predicted genes, such as potential functions, and can also help to validate the predicted genes by comparing them to known genes in other organisms. Similarity searches can also be used to evaluate the completeness of the predicted genes, by comparing the predicted genes to a set of conserved single-copy orthologs expected in fungal genomes, such as those provided by BUSCO [11]. Together, these methods can provide a more complete picture of the predicted genes and their potential functions, and can help to identify potential targets for further research. This forms the basis

of Research Questions 2.7, 2.8, and 2.10.

1.4 Secondary Metabolites

While cellular products essential to an organism’s viability are of great interest, there are a vast number of cellular products that while not essential, still provide great benefit to the organism and may be necessary for survival in some situations [3] [12]. These other cellular products are known as secondary metabolites, and they are involved in several roles ranging from cellular signalling to antibiotic activity, making them a frequent subject of study in pharmaceutical research. Enzymes that produce secondary metabolites are comprised of non-ribosomal peptide synthetases (NRPS) and polyketide synthetases (PKS), which synthesize amino acids and other basic enzymatic building blocks into proteins [7]. Genes encoding the NRPSs and PKSs responsible for production of secondary metabolites are often found in clusters within a genome, but their products remain unknown [12].

The study of these products is difficult, as the genes encoding the modules that make up NRPSs and PKSs are not expressed under normal laboratory conditions [12]. The NRPSs and PKSs that have been studied have been shown to be large enzymatic structures containing several functional modules, each responsible for a specific step in the synthesis of a protein [12]. This concept of completeness of gene predictions is explored in Research Question 2.10. In addition, gene predictions can be processed further with tools such as antiSMASH, which can identify secondary metabolite biosynthetic gene clusters (BGCs) in a genome [1]. Comparison of predicted genes with known secondary metabolite BGCs can provide insight into completeness of the predicted genes with a specific focus on secondary metabolite production. This work is not covered in this thesis, but is worth noting for context as it is a common approach in the field of secondary metabolite research.

Another interesting feature of NRPSs and PKSs is the length of the genes encoding them, which can be quite long, often exceeding 10,000 base pairs in length [7]. Examining the outputs of gene prediction tools can provide insight into the lengths of predicted genes, and whether or not the tools are able to capture the full length of these genes. This topic serves as the basis for Research Question 2.5.

1.5 Genome Assembly

Sequence assembly has been a long-standing problem in the field of bioinformatics [13]. Determining the correct order and combination of smaller subsequences into an accurate complete sequence assembly is computationally difficult in terms of compute resources such as memory, CPU cycles and storage required for input sequences [13]. In addition to these difficulties, there can be other issues encountered during assembly due to the nature of the data or genomes themselves, such as low quality base calls for long read data, which is not necessarily the case today, or the inherent content of genomes themselves using repetitive regions as an example. Insufficient data may result in short, fragmented assemblies, depending on the size of the genomes,

while sequence data that is not long enough can fail to fully capture repetitive regions in an assembly. A wide range of assembly tools have been developed with their own unique approaches to the genome assembly problem, so it is important to use an appropriate assembler for the task at hand, and also important to evaluate the assembly thoroughly.

Genome assembly tools generally approach the assembly problem using a graph-based approach. The most common graph-based approach is the de Bruijn graph assembly [2]. A graph in this context, is set of nodes (k -mers from sequences) connected by edges (overlaps between k -mers). Traversing through this graph results in longer subsequences that ultimately result in a set of sequences referred to as an assembly. In the early years of long read sequence data, sequencing platforms encountered difficulties producing consistently high scores for base calls when sequencing. To combat this, some assembly workflows may also include a polishing or correction step once the initial assembly is completed in which high quality short read sequences are used as supplemental information to correct low quality regions in the assembly. These low quality base calls are typically not present in modern long read sequencing approaches as the methodology and quality of calls have improved drastically. While the polishing step is arguably unnecessary in modern assemblies, the polishing programs remain available should researchers be interested in applying additional reads for polishing.

One approach to aid in the previously mentioned issue of assembly correctness is to use a combination of long and short reads in what is known as a hybrid assembly. Combining both highly accurate short reads with deep coverage along with less accurate but much longer reads can produce high quality genome assemblies that capture long repetitive regions. Hybrid assembly approaches have been shown to produce high quality assemblies in a wide variety of organisms as they combine long read data with short data to produce assemblies that properly represent long repetitive regions with additionally high quality Illumina sequences for correction. Once assembled, the sequences must also be evaluated with measures such as N50, L50, coverage, average contig length and total assembled length to ensure that the genomes are well assembled, at least based on these metrics [13]. Following appropriate assembly protocols is essential to the further success of a project as downstream processing such as annotation depends on a high-quality assembly.

1.6 Identification of Anomalous Genomic regions

One important aspect of interest when assembling any form of sequence is GC content or percent GC of the assembled sequence. Large regions of anomalous GC content may be of interest to researchers as they may contain repetitive regions and unique features responsible for traits specific to the organism in question.

2 Research Questions

2.1 Research Questions

With an ever-increasing number of gene prediction tools available to users, it is important to assess and understand their behaviour and performance in the context, particularly in the context of new genome assemblies of lesser studied organisms, where a reference prediction set may not be available. The main purpose of this research is to evaluate and compare gene finding tools in the context of *Trichoderma* assemblies where a gold standard set of gene predictions does not exist. To assess behaviour and performance in these contexts, we have defined X problems to profile the selected gene finding tools. In addition to applying selected gene finding tools to novel *Trichoderma* isolates, Tsht20 and DC1, we also applied selected gene finding tools to existing *Trichoderma* assemblies from the National Center for Biotechnology Information (NCBI).

2.2 *Trichoderma* Assembly Results

Since gene finding tools operate on an assembled genomic sequence, it must follow that the results will be influenced by the supplied assembly. Before applying gene finders to the new assemblies, we should first investigate the new assemblies by generating general assembly metrics for the new assemblies to contrast and compare with existing assemblies. We ask: **how do assemblies of DC1 and Tsth20 compare to existing *Trichoderma* assemblies?** With these isolates being from the *Trichoderma* family, we expect assembly metrics to be similar in nature to existing assemblies from NCBI, but do they?

2.3 Profiling of Gene Finding Tools

Different gene finding tools may predict different types of features associated with gene structures. The question arises: **which (if any) gene finders predict additional features outside of the standard gene model?** Additional features in this case include promoter sequences, transcription binding sites, activating sequences and other upstream or downstream sequences. In addition, different gene finding tools employ differing programming languages and algorithms which raises several questions. **How are these gene finding tools implemented? Is the software straightforward to install? Are the tools user-friendly? What is the processing time and memory consumption of different gene finding**

tools in the context of *Trichoderma*?

2.4 Number of Features Predicted

Do gene finders predict similar numbers of features in the context of *Trichoderma* genomes?

One common method for evaluating gene finding tools is by looking at the number of features predicted by each tool. These features make an obvious point of comparison for selected gene finding tools. The term ‘feature’ here is somewhat ambiguous, referring to many possible categories of genomic feature. For each gene finding tool, we compare the counts for predicted genes, transcripts, and coding sequences.

2.5 Lengths of Predicted Genes

Genome assemblies can contain a wide range of gene lengths. For some users, genes of a specific length may be a key point of interest, so the ability of a gene prediction tool to capture the broad range of possible gene lengths is another important metric for comparison. Thus we ask the question: **do different gene finders predict genes of similar lengths in *Trichoderma*?**

2.6 Identifying Regions of Agreement and Disagreement

With predicted genes from several tools available, the question we would like to ask is whether or not the gene finders agree with one another for any given prediction. To answer this question, we will identify ‘regions’ of overlapping predictions. A region can be defined as a start and stop position of a set of individual or overlapping features from one or more gene finding tools and external sources. With regions identified, we can determine agreement, or more importantly, disagreement in predictions between gene finding tools from which we can ask: **do gene finders agree on their predictions? If no, to what extent do they disagree? Are there genomic regions where agreement or disagreement are more prevalent?**

2.7 Validation of Predicted Genes via InterProScan

In an effort to validate, or at least provide supporting evidence for any given gene prediction we will apply InterProScan to coding sequences predicted by each of the gene finders to identify features associated with protein function. Genes will be considered as ‘valid’ if the gene’s protein sequence contains binding sites, motifs, or other functional characteristics of proteins. Using the results from InterProScan, we ask the question: **in the context of *Trichoderma*, do proteins predicted by gene finders contain functional signatures?**

2.8 Similarity searches of Predicted Genes with tblastn

2.9 Performance in Regions of Anomalous Sequence Content

One of the inspirations for this research is the unique composition of genomic sequence in *Trichoderma*. Results from the assembly process show that GC content in *Trichoderma* strains is abnormal throughout most assemblies. These regions of assemblies present an interesting opportunity to assess gene finding performance in regions of anomalous GC content. The question follows: **do gene finders behave differently in regions of anomalous sequence content?**

2.10 BUSCO Completeness

The use of existing benchmarks for gene finding performance is useful when assessing performance of gene finding tools, particularly in the case of genes that should be evolutionarily conserved. **Do gene finders predict conserved single-copy orthologs expected in fungal genomes?**

2.11 Selection of a Gene Finding Tool

With all the results generated, we can provide insight to the question: **which gene finding tool should one choose?**

References

- [1] Kai Blin, Simon Shaw, Hannah E Augustijn, Zachary L Reitz, Friederike Biermann, Mohammad Alanjary, Artem Fetter, Barbara R Terlouw, William W Metcalf, Eric J N Helfrich, Gilles P van Wezel, Marnix H Medema, and Tilmann Weber. antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51(W1):W46–W50, July 2023.
- [2] Phillip E. C. Compeau, Pavel A. Pevzner, and Glenn Tesler. How to apply de Bruijn graphs to genome assembly. *Nature biotechnology*, 29(11):987–991, November 2011.
- [3] Arryn Craney, Salman Ahmed, and Justin Nodwell. Towards a new science of secondary metabolism. *The Journal of Antibiotics*, 66(7):387–400, 2013.
- [4] Girum Fitihamlak Ejigu and Jaehee Jung. Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9(9), September 2020.
- [5] David A. Fitzpatrick. Horizontal gene transfer in fungi. *FEMS Microbiology Letters*, 329(1):1–8, 2012.
- [6] Carla Gonçalves, Chris Todd Hittinger, and Antonis Rokas. Horizontal Gene Transfer in Fungi and Its Ecological Importance. In Yen-Ping Hsueh and Meredith Blackwell, editors, *Fungal Associations*, pages 59–81. Springer International Publishing, Cham, 2024.
- [7] Hisayuki Komaki and Tomohiko Tamura. Polyketide Synthase and Nonribosomal Peptide Synthetase Gene Clusters in Type Strains of the Genus Phytohabitans. *Life*, 10(11):257, October 2020.
- [8] Christian P. Kubicek, Alfredo Herrera-Estrella, Verena Seidl-Seiboth, Diego A. Martinez, Irina S. Druzhinina, Michael Thon, Susanne Zeilinger, Sergio Casas-Flores, Benjamin A. Horwitz, Prasun K. Mukherjee, Mala Mukherjee, László Kredics, Luis D. Alcaraz, Andrea Aerts, Zsuzsanna Antal, Lea Atanasova, Mayte G. Cervantes-Badillo, Jean Challacombe, Olga Chertkov, Kevin McCluskey, Fanny Culpier, Nandan Deshpande, Hans von Döhren, Daniel J. Ebbole, Edgardo U. Esquivel-Naranjo, Erzsébet Fekete, Michel Flippin, Fabian Glaser, Elida Y. Gómez-Rodríguez, Sabine Gruber, Cliff Han, Bernard Henrissat, Rosa Hermosa, Miguel Hernández-Oñate, Levente Karaffa, Idit Kost, Stéphane Le Crom, Erika Lindquist, Susan Lucas, Mette Lübeck, Peter S. Lübeck, Antoine Margeot, Benjamin Metz, Monica Misra, Helena Nevalainen, Markus Omann, Nicole Packer, Giancarlo Perrone, Edith E. Uresti-Rivera, Asaf Salamov, Monika Schmoll, Bernhard Seiboth, Harris Shapiro, Serenella Sukno, Juan Antonio Tamayo-Ramos, Doris Tisch, Aric Wiest, Heather H. Wilkinson, Michael Zhang, Pedro M. Coutinho, Charles M. Kenerley, Enrique Monte, Scott E. Baker, and Igor V. Grigoriev. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma. *Genome Biology*, 12(4):R40, April 2011.
- [9] Christian P. Kubicek, Andrei S. Steindorff, Komal Chenthamara, Gelsomina Manganiello, Bernard Henrissat, Jian Zhang, Feng Cai, Alexey G. Kopchinskiy, Eva M. Kubicek, Alan Kuo, Riccardo Baroncelli, Sabrina Sarrocco, Eliane Ferreira Noronha, Giovanni Vannacci, Qirong Shen, Igor V. Grigoriev, and Irina S. Druzhinina. Evolution and comparative genomics of the most common Trichoderma species. *BMC Genomics*, 20(1):485, 2019.
- [10] Brendan Loftus. 4 - Genome Sequencing, Assembly and Gene Prediction in Fungi. In Dilip K. Arora and George G. Khachatourians, editors, *Fungal Genomics*, volume 3, pages 65–81. Elsevier, 2003.
- [11] Mosè Manni, Matthew R. Berkeley, Mathieu Seppey, and Evgeny M. Zdobnov. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, 1(12):e323, 2021.

- [12] Prasun K. Mukherjee, Benjamin A. Horwitz, and Charles M. Kenerley. Secondary metabolism in *Trichoderma* – a genomic perspective. *Microbiology*, 158(1):35–45, 2012.
- [13] Niranjana Nagarajan and Mihai Pop. Sequence assembly demystified. *Nature Reviews Genetics*, 14(3):157–167, 2013.
- [14] Miroslav Plohl, Nevenka Meštrović, and Brankica Mravinac. Centromere identity from the DNA point of view. *Chromosoma*, 123(4):313–325, August 2014.
- [15] Biancamaria Senizza, Fabrizio Araniti, Simon Lewin, Sonja Wende, Steffen Kolb, and Luigi Lucini. *Trichoderma* spp.-mediated mitigation of heat, drought, and their combination on the *Arabidopsis thaliana* holobiont: A metabolomics and metabarcoding approach. *Frontiers in Plant Science*, 14, 2023.
- [16] David J. Winter, Austen R. D. Ganley, Carolyn A. Young, Ivan Liachko, Christopher L. Schardl, Pierre-Yves Dupont, Daniel Berry, Arvina Ram, Barry Scott, and Murray P. Cox. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLOS Genetics*, 14(10):1–29, 2018.
- [17] Sheridan L. Woo, Rosa Hermosa, Matteo Lorito, and Enrique Monte. *Trichoderma*: A multipurpose, plant-beneficial microorganism for eco-sustainable agriculture. *Nature Reviews Microbiology*, 21(5):312–326, 2023.