Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.

http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.

http://www.ub.tuwien.ac.at/eng

DIPLOMARBEIT

CONVERGENT EVOLUTION OF PROTEINS WITH ANKYRIN DOMAINS THE MAIN GENOMIC HALLMARK OF AN INDUSTRIALLY RELEVANT FUNGUS *TRICHODERMA*

Ausgeführt am Institut für
Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften
der Technischen Universität Wien

unter der Anleitung von
Assoc. Prof. Dr.rer.nat. Irina Druzhinina
und
DI Komal Chenthamara

durch

Vladimir Gojic BSc
1225096
Grünbergstraße 27/10
1120, Wien

Datum

(Unterschrift Student)

(Unterschrift Betreuer)

Acknowledgments

First, I would like to thank Prof. Dr. Irina Druzhinina for giving me the opportunity to conduct my master thesis in her research group Microbiology and Applied Genomics and for her feedback in every step of this project.

Second, I would like to thank DI Komal Chenthamara for her constant support and encouragement, especially for giving me directions and help in crucial moments of my master thesis.

I would also like to thank all members of the Research Group Microbiology and Applied Genomics for a pleasant work atmosphere and time spent together.

Last but not least I would like to thank my friends and family for their unconditional love and support.

Abstract

The comparative genomics study of the industrially relevant *Trichoderma* spp. (Hypocreales, Ascomycota) revealed a considerable expansion of ankyrin-domain-containing- (ANKdc-) proteins when compared to other ecologically similar Pezizomycotina fungi. Ankyrin-(ANK-) domains are found in all domains of life as the only identifiable structural motif or as associates of other PFAM (host) domains that may be involved in a diversity of functions such as signal transduction, transport, transcription regulation, inflammatory response and other essential cell processes. The role of ANKdc-proteins in fungi remains poorly understood. In this thesis, a comprehensive *in silico* analysis of *Trichoderma* ANKyrome was performed to uncover the role of expansion of ANKdc-proteins in the ecology of *Trichoderma* spp. and provide the first hypotheses into their role in *Trichoderma* interactomes. A comparative genomics approach was used for genome mining, annotation, enumeration and classification of proteins with ANK-repeats in 10 *Trichoderma* species. In total, eighteen Hypocreales genomes were mined for ANKdc-proteins and 2406 of these proteins were retrieved. Our study revealed that the expansion of ANKdc-proteins is not exclusive to *Trichoderma*, but also evident in closely related fungi such as phytopathogenic Nectriaceae and Bionectriaceae. Indication that expansion of ANKdc is characteristic of phytopathogens as well, compels to look at more transcriptomic data of *Trichoderma* when interacting with plants or growing in soil to understand the role of ANKdc-proteins. By orthology inference, 1172 homologous ANKdc-proteins were distributed between 183 orthogroups of which 18 orthologous ANKdc-proteins were identified as core in *Trichoderma* spp. and 11 as core in all of 18 considered Hypocreales fungi. Thus, the results indicate that the larger share of ANKdc-proteins is unique to most species (orphans). All ANK-repeats in *Trichoderma* ANKyrome were annotated by InterProScan within Blast2GO using three databases, Pfam, SMART, and Prosite. Additionally, all host-domains were annotated within *Trichoderma* ANKyrome using the same approach but with all 14 databases available in InterProScan. This study revealed that two most frequent host-domains in *Trichoderma* ANKyrome are P-loop containing nucleoside triphosphate hydrolase and nucleoside phosphorylase, but also that majority of these ANKdc-proteins are orphan proteins. As nucleoside phosphorylases are known to play a role in immune responses, these results allow us to hypothesize that abundance of ANK-repeats in *Trichoderma* might be contributing to general fitness of this genus. Furthermore, SignalP and TMHMM analyses revealed that only a small share of *Trichoderma* ANKyrome consists of proteins involved in cell signaling or

transmembrane proteins, respectively. Evolutionary analyses of the core *Trichoderma* ANKyrome were performed using vertical evolution and purifying selection acting across each of 18 core ANKdc-proteins from *Trichoderma* spp.

Kurzfassung

Die komparative genomische Studie der industriell relevanten Pilze *Trichoderma* spp. (Hypocreales, Ascomycota) zeigte eine beträchtliche Expansion von Ankyrin-Domäne-enthaltenden- (ANKdc-) Proteinen im Vergleich zu anderen ökologisch ähnlichen Pezizomycotina-Pilzen. Ankyrin- (ANK-) Domäne finden sich in allen Domänen des Lebens als das einzige identifizierbare Strukturmotif oder assoziiert mit anderen PFAM (host) Domänen, die an einer Vielfalt von Funktionen wie Signaltransduktion, Transport, Transkriptionsregulation, Entzündungsantwort und anderen essentiellen Zellprozessen beteiligt sein können. Die Rolle von ANKdc-Proteinen in Pilzen ist noch weitgehend ungeklärt. In dieser Arbeit wurde eine umfassende *in silico* Analyse der *Trichoderma* ANKyrome durchgeführt, um die Rolle der Expansion von ANKdc-Proteinen in der Ökologie von *Trichoderma* spp. zu erforschen und die ersten Hypothesen über ihre Rolle in *Trichoderma* Interaktomen zu stellen. Ein komparativer genomischer Ansatz wurde für Genom-Mining, Annotierung, Verzeichnis und Klassifizierung von Proteinen mit ANK-Repeats in zehn *Trichoderma*-Arten verwendet. Insgesamt wurden 18 Hypocreales-Genome nach ANKdc-Proteinen durchsucht und 2406 dieser Proteine gefunden. Unsere Studie zeigte, dass die Expansion von ANKdc-Proteinen nicht ausschließlich bei *Trichoderma* auftritt, sondern auch bei nahe verwandten Pilzen wie phytopathogenen Nectriaceae und Bionectriaceae.

Der Hinweis, dass die Expansion von ANKdc auch für Phytopathogene charakteristisch ist, zwingt dazu, mehr transkriptomische Daten von *Trichoderma* zu betrachten, wenn sie mit Pflanzen interagieren oder im Boden wachsen, um die Rolle von ANKdc-Proteinen zu verstehen. Durch Orthologie-Inferenz wurden 1172 homologe AKNdc-Proteine auf 183 Orthogruppen verteilt, von denen 18 orthologe ANKdc-Proteine als Core-Proteine in *Trichoderma* spp identifiziert wurden und 11 als Core-Proteine in allen der 18 betrachteten Hypocreales-Pilzen.

Somit zeigen die Ergebnisse, dass der größere Anteil an ANKdc-Proteinen für die meisten Arten einzigartig ist (Orphan). Alle ANK-Repeats in *Trichoderma* ANKyromen wurden annotiert durch InterProScan in Blast2GO mit drei Datenbanken, Pfam, SMART und Prosite. Darüber hinaus wurden alle Host-Domänen in *Trichoderma* ANKyromen mit dem gleichen Ansatz, aber mit allen 14 Datenbanken in InterProScan annotiert. Diese Studie ergab, dass die beiden häufigsten Host-Domänen in *Trichoderma* ANKyromen P-Loop-haltige Nucleosidtriphosphat-Hydrolase und Nucleosidphosphorylase sind, aber auch, dass die Mehrheit dieser ANKdc-Proteine

Orphan-Proteine sind. Da bekannt ist, dass Nucleosidphosphorylasen eine Rolle bei Immunantworten spielen, erlauben diese Ergebnisse die Hypothese, dass die Häufigkeit von ANK-Repeats in *Trichoderma* zur allgemeinen Fitness dieser Gattung beitragen könnte. SignalP- und TMHMM-Analysen zeigten, dass nur ein kleiner Teil vom *Trichoderma* ANKyrom aus Proteinen die an der Zellsignalisierung beteiligt sind bzw. Transmembranproteinen besteht. Evolutionäre Analysen des Core-ANKyrom von *Trichoderma* wurden durchgeführt unter der Verwendung von vertikaler Evolution und reinigender Selektion, die in jedem der 18 Core-Proteine in *Trichoderma* spp. wirkt.

Abbreviations

AICc - Akaike Information Criterion corrected

ANK - Ankyrin

ANKdc - Ankyrin domain containing

BIC - Bayesian Inference Criterion

BLAST - Basic Local Alignment Search Tool

BUSTED - Branch-Site Unrestricted Statistical Test for Episodic Diversification

DGE - Differential Gene Expression

FUBAR - Fast Unconstrained Bayesian AppRoximation

GARD - Genetic Algorithm for Recombination Detection

GTR - Generalized Time Reversible (model)

hdc-ANKdc - host-domain containing ANKdc

HET - Heterokaryon Incompatibility

JGI - Joint Genome Institute

JTT - Jones, Taylor, and Thornton model

LG - Le and Gascuel model

LRT - Likelihood Ratio Test

Mbp - Million base pairs

MCL - Markov Cluster Algorithm

MEME - Mixed Effects Model of Evolution

ML - Maximum Likelihood

MSA - Multiple Sequence Analysis

NTHGs - Non-*Trichoderma* Hypocrealean Genomes

NP - Nucleoside Phosphorylase

NTPase - Nucleoside Triphosphatase

OG - Orthogroup

PK - Protein Kinase

PSRF - Potential Scale Reduction Factor

PLNTPH - P-loop containing Nucleoside Triphosphate Hydrolase

RPKM - Reads Per Kilobase Million

Contents

Acknowledgments 3

Abstract 5

Kurzfassung 7

Abbreviations 9

1. Introduction 13

  1.1 *Trichoderma* and their neighbors 13

  1.2 Ankyrin-domain-containing proteins 17

2 Aims of the study and hypothesis 21

3 Materials and Methods 23

  3.1 Preliminary investigation 23

    3.1.1 Whole Genome and ANKdc-protein statistics among Hypocreales 23

    3.1.2 Retrieval of putative ANKdc-protein sequences 23

  3.2 Re-annotations of Hyporcrealean ANKyrome 24

  3.3 Orthology Analysis of Hypocrealean ANKyrome 25

    3.3.1 UpSet analysis to reveal shared ANKyrome and core orthologous genes 27

  3.4 Evolutionary Analysis for core orthologous genes 27

    3.4.1 Phylogenetic Analysis 27

    3.4.2 Selection Pressure Analysis 28

  3.5 TMHMM and SignalP analyses 30

  3.6 Transcriptomic data analysis 30

4 Results and discussions 31

  4.1 Inventory of the *Trichoderma* ANKyrome 31

    4.1.1 Genome and ANKdc-protein statistics comparison 31

    4.1.2 Retrieval of putative ANKdc-proteins from selected Hypocreales 33

    4.1.3 Re-annotated Hypocrealean ANKyrome 35

    4.1.4 Orthology Inference within the Hypocrealean ANKyrome 47

    4.1.5 UpSet visualization of ANKyrome intersections 49

  4.2 Evolutionary Analysis 50

    4.2.1 Phylogenetic Analysis 50

    4.2.2 Selection Pressure Analysis 70

  4.3 TMHMM and SignalP analysis 72

  4.4 Transcriptomic data analysis 74

    4.4.1 Differential Gene Expression results from *T. guizhouense* NJAU 4742 74

    4.4.2 Differential Gene Expression results from *T. harzianum* TUCIM 916 77

5 Concluding remarks 79
6 Outlook 81
7 Curriculum Vitae 83
8 References 85
Supplementary Materials 91
  Supplementary material 1 91
  Supplementary material 2 92
  Supplementary material 3 93
  Supplementary material 4 95
  Supplementary material 5 96
  Supplementary material 6 97
  Supplementary material 7 101

1. Introduction

1.1 *Trichoderma* and their neighbors

Species of *Trichoderma* are filamentous Ascomycota found all over the world occupying vastly different ecological habitats, which is why many of them are often referred to as cosmopolitan and opportunistic organisms (Druzhinina *et al.* 2011; Gupta *et al.* 2014).

Taxonomically, *Trichoderma* belongs to the order of Hypocreales and the class of Sordaryomycetes. *Trichoderma* spp. are haploid organisms, but most also possess a functional heterothallic mode of sexual reproduction system, that is possible only between two different mating type individuals, namely *mat1-1* and *mat1-2* mating locus types (bipolar heterothallism) (Seidl *et al.* 2009).

Some members of *Trichoderma* genus secrete large amounts of certain CAZymes and find their application mainly in biotechnology as cellulase and hemicellulase producers for the food and biofuel industries. The most studied species for this application and in general is the industrial cellulase producer *T. reesei* (see Druzhinina *et al.* 2016 for the references; Stricker, Mach, and De Graaff 2008; Seiboth, Ivanova, and Seidl-Seiboth 2011). For this reason, *T. reesei* was sequenced in 2008, becoming the first genome from this genus to be studied (Martinez *et al.* 2008).

Druzhinina et al. (2018) described the expansion of nutritional strategies adapted by *Trichoderma*. These species have developed multiple nutritional modes which enable them to occupy vastly different ecological niches like other fungi, plants, dead wood, soil etc. Their ability to parasitize and kill other fungi, called mycoparasitism, as well as their production of antifungal secondary metabolites is applied as a biocontrol against plant pathogenic fungi (Harman *et al.* 2004). Furthermore, some *Trichoderma* spp. are used for stimulation of plant growth as associates or endophytes of plants, contributing to resistance to plant stresses and diseases (Harman 2011). Although there are many advantages of *Trichoderma* spp. applications, there are studies showing certain *Trichoderma* spp. causing green mold disease on mushroom farms (Komoń-Zelazowska *et al.* 2007). More importantly, some *Trichoderma* spp. act as pathogens of immunocompromised patients, especially organ transplant recipients (Chouaki *et al.* 2002). Due to above reasons, the interest to study their genotypes resulted in availability of 15 different genomes of *Trichoderma* in 2018. However, only 10 *Trichoderma* genomes were publicly available at the time of beginning of this master thesis.

In 2011, Kubicek *et al.* sequenced strongly opportunistic and cosmopolitan *T. atroviride* and *T. virens* and performed the first comparative genomics study with already published *T. reesei* genome (Kubicek *et al*. 2011). Inspired by this study, a more extensive Markov cluster algorithm (MCL) analysis of 44 Pezizomycotina genomes in total, including the three previously mentioned *Trichoderma* spp. was done (Irina Druzhinina, personal communication). The MCL analysis revealed that *Trichoderma* spp. contained an increased number of ankyrin-domain-containing (ANKdc) proteins compared to the average number of these genes in considered Pezizomycotina species (Figure 1). Furthermore, within *Trichoderma* genus, in strongly opportunistic and cosmopolitan *T. atroviride* and *T. virens,* ANKdc-proteins were even more expanded than in *T. reesei*. With this thesis, we aim to set foundation towards explaining the expansion and roles of ANKdc genes. Therefore, we selected genomes of *Trichoderma* and other closely related species for a comparative genomics study of their ANKdc-protein.

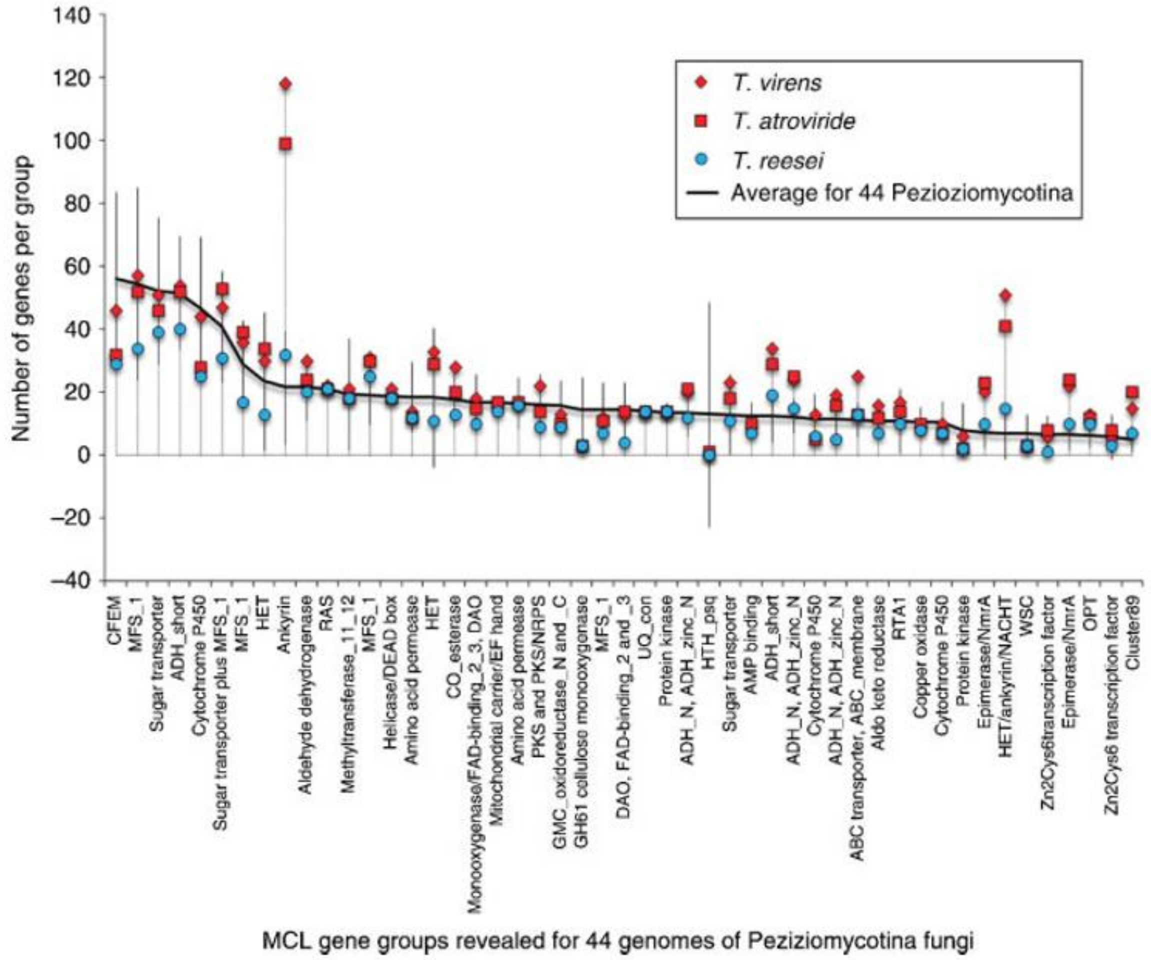


Figure 1: MCL analysis of gene families in 44 Pezizomycotina genomes. Irina Druzhinina, unpublished

Druzhinina *et al.* (2018) performed a multilocus phylogenetic analysis with 100 housekeeping genes from 21 Hypocreales genomes and two species from Sordariales order (Figure 2; adapted from Druzhinina *et al.* 2018), revealing the close neighbors of *Trichoderma*. This phylogenetic analysis confirmed the assumed classification of *Trichoderma* spp. into taxonomical sections *Longibrachiatum*, *Harzianum*, *Virens* and *Trichoderma*. It is evident from the phylogram that species of the *Longibrachiatum* section are evolutionarily most derived, while the Trichoderma section is closest to the last common ancestor of all *Trichoderma* spp. Furthermore, *E. weberi* (de Man *et al.* 2016) shares the last common ancestor with *Trichoderma* genus and it is a specialized mycoparasite, reflecting the ancestral state of *Trichoderma* spp. Moreover, it was revealed that *Trichoderma* spp. are closer to entomopathogens (carnivory) than their plant-associated hosts (herbivory). Availability of these eight other Hypocrealean genomes and their well-studied ecology (Smith 2007; Roy *et al.* 2010; Xiao *et al.* 2012) make them interesting candidates for comparison with the genomes of *Trichoderma,* in order to reveal clues about ANKdc-proteins expansion in *Trichoderma*.

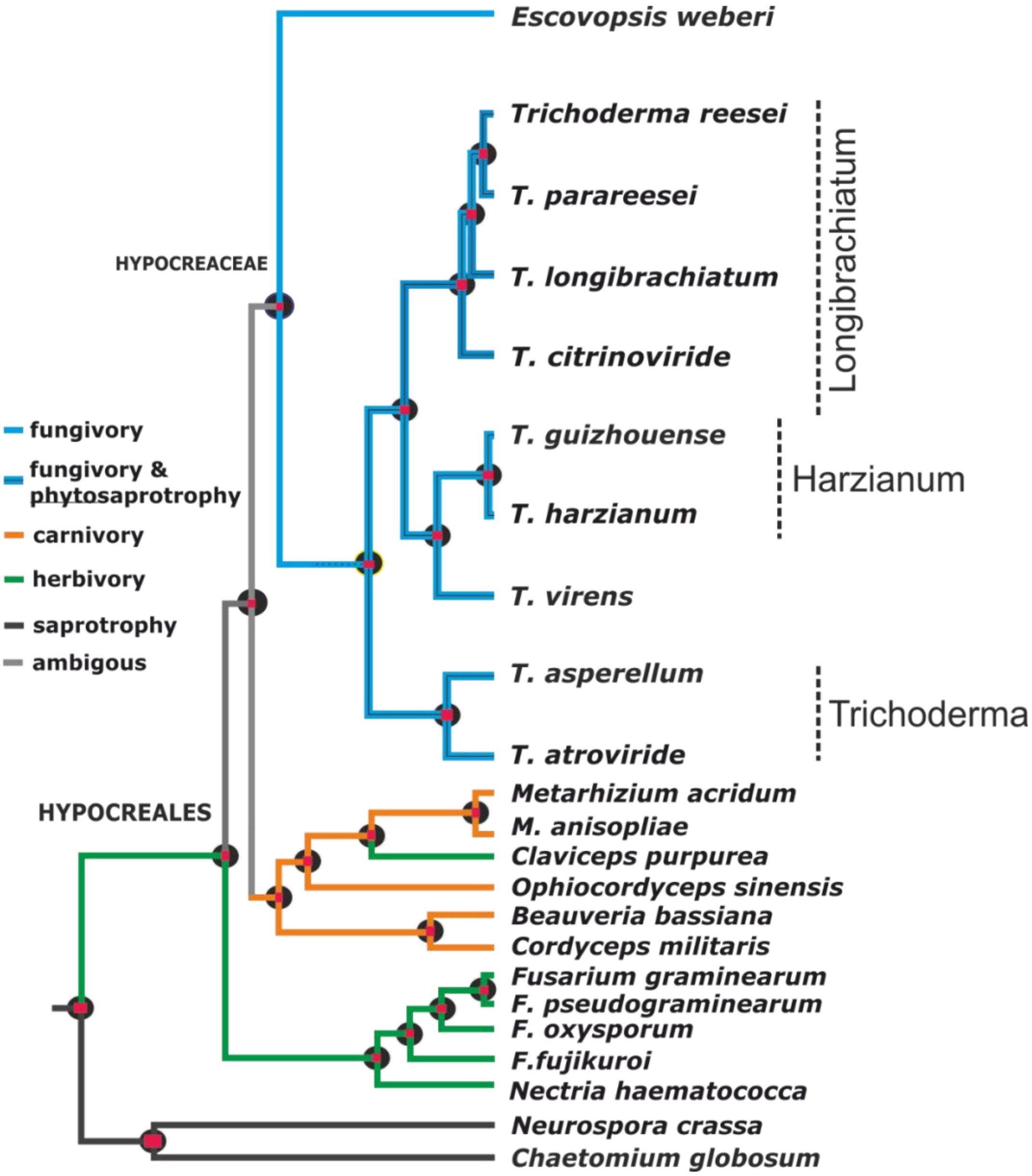


Figure 2: Phylogram of Hyocreales spp. based on 100 orhtologous proteins (~50,000 resdiues). Modified from Druzhinina et al. 2018

1.2 Ankyrin-domain-containing proteins

The ankyrin- (ANK-) repeat is a protein-protein interaction motif found in proteins in all living organisms (Figure 3, Chenthamara *et al*. unpublished). It was named after a cytoskeletal protein ankyrin which contains 22 copies of the ANK-repeat (Lux, John, and Bennett 1990).

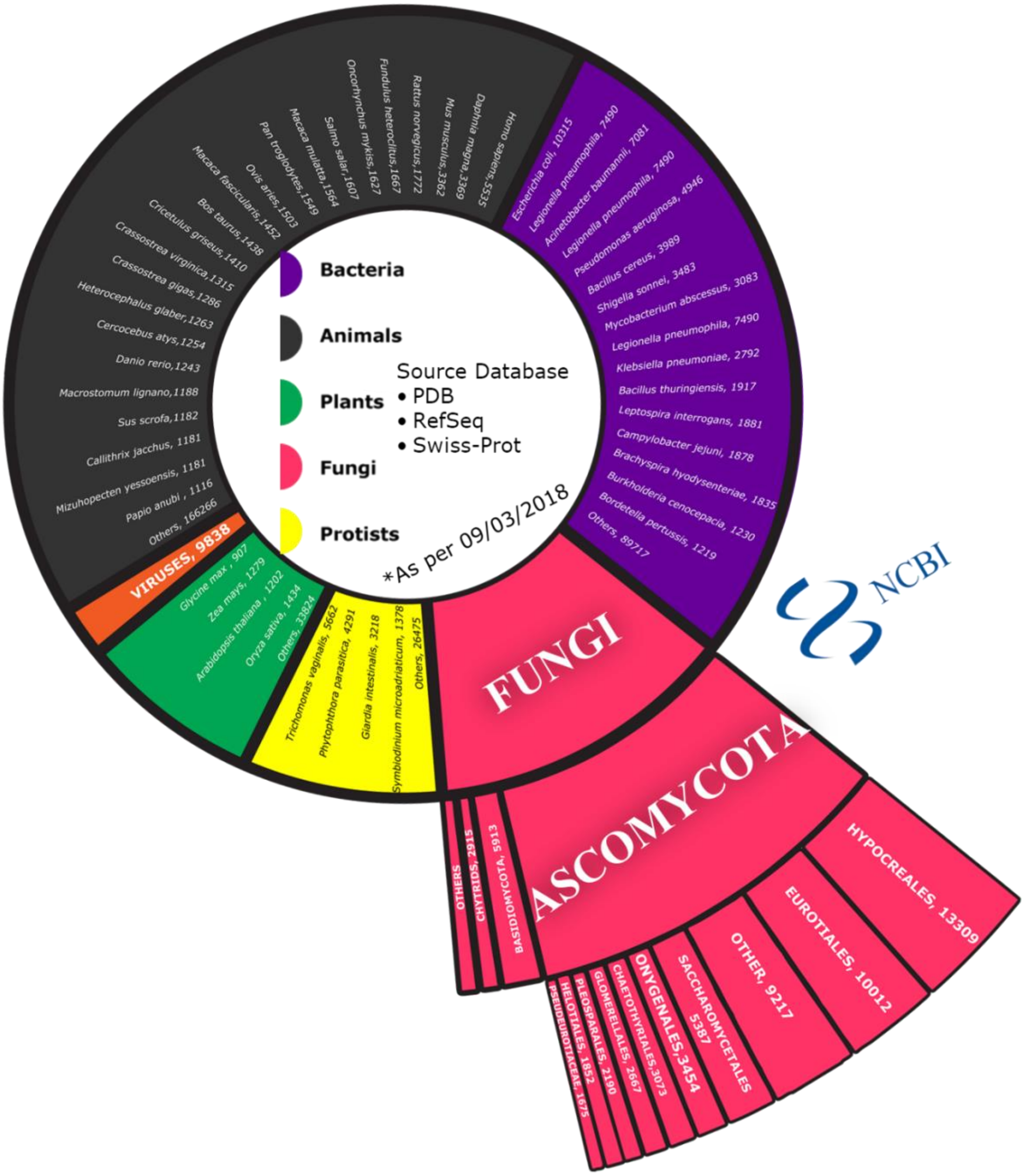


Figure 3: A sunburst diagram showing distribution of ANKdc-proteins across all domains of life. The number next to species name is the respective number of ANKdc-proteins reported in the respective species. Chenthamara et al; unpublished

A single ANK-repeat consists of 33-amino acid residues that form a motif of a helix-turn-helix-beta-hairpin/loop fold, which is shown in Figure 4A. The two helices are arranged in antiparallel fashion followed by a loop region that points outward at an approximately 90° angle, resembling a structure similar to letter L (Gorina and Pavletich 1996). Specific residues, often referred to as signature residues, are relatively conserved to keep the structural integrity of the motif. The positions 4-7 are usually occupied by T-P-L-H (Threonine-Proline-Leucine-Histidine) amino acids. Proline at 5th position is responsible for the L-shaped structure by forming the tight turn and initiating the first helix. Hydrogen bonding between the hydroxyl group of threonine and the imidazole ring of histidine contribute to the stability of ANK-domain. In positions 17-22, amino acids V/I-V-X-L/V-LL (Valin/Isoleucine-Valin-XX(hydrophilic)-Leucine/Valine-Leucine-Leucine) form the central part of the second alpha-helix and inter- and intra-hydrophobic networks to stabilize the whole ANK-domain (Mosavi, Minor, and Peng 2002; Li, Mahajan, and Tsai 2006). Solvent accessibility affects the amino acid composition, so those residues that do not come into contact with solvent are usually hydrophobic and constitute the hydrophobic core of the ANK-domain. Some of these residues are replaced with hydrophilic residues in terminal ANK-repeats of an ANK-domain-containing (ANKdc)-protein because the solvent has access to them.

In ANK-domains, helices of one ANK-repeat pack against the helices of the adjacent ANK-repeat while the beta-hairpin/loop region, in some cases, forms a continuous beta-sheet as shown in Figure 4B. The tips of beta-hairpins/loop region and the surface of packed helices facing them are responsible for binding of the target molecule. The residues constituting these regions of ANK-domain are generally variable to ensure functional and binding specificity (Sedgwick and Smerdon 1999). On average there are 4-7 ANK-repeats per ANKdc-protein, and proteins with more ANK-repeats generally have more compact and concave structure which reflects their modular nature and enables their wide functional diversity (Mosavi *et al*. 2004).

The ANKdc-proteins can be loner-proteins, containing exclusively ANK-domain(s), or they can be multidomain proteins where ANK-domain(s) are associated with a functional (host) domain. These functional domains are involved in transcription regulation, signal transduction, cell-cell signaling, cell-cycle regulation, inflammatory response, cytoskeleton integrity, toxin-encoding, transport phenomena etc. (Bork 1993; Mosavi *et al.* 2004). The function of ANKdc-proteins in *Trichoderma* and fungi, in general, is not well understood.

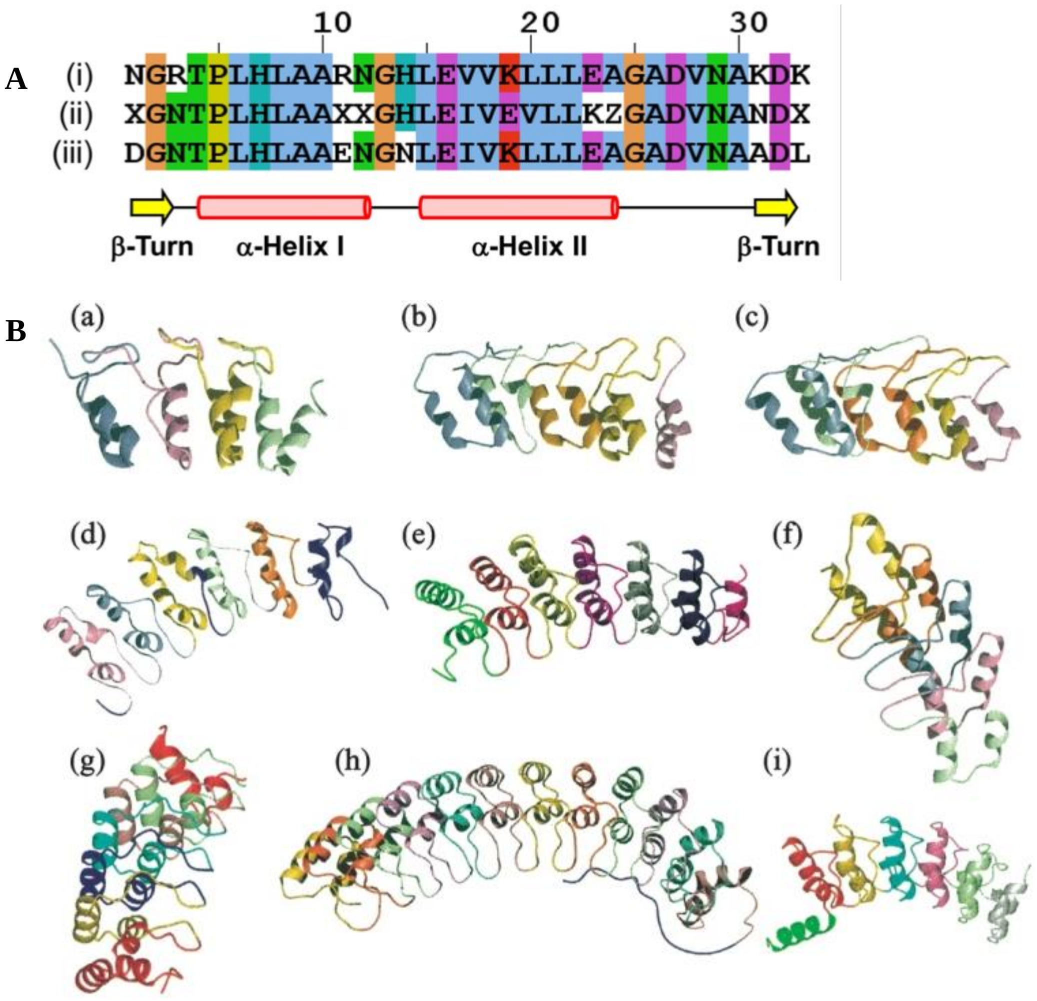

A
10
20
30
(i) NGRTPLHLAARNGHLEVVKLLLEAGADVNAKDK
(ii) XGNTPLHLAAXXGHLEIVEVLLKZGADVNANDX
(iii) DGNTPLHLAAENGNLEIVKLLLEAGADVNAADL
β-Turn
α-Helix I
α-Helix II
β-Turn
B
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)

Figure 4A and 4B: Conserved structural features of the ANK-repeat and ANKdc-proteins with high-resolution structures in the PDB.Al-Khodor & Souhaila, et al., 2010

2 Aims of the study and hypothesis

The hypothesis of our study is that the overrepresentation of ANKdc-proteins in *Trichoderma* spp. is linked to the fitness and thereby to the environmental success of this genus. Therefore, the aim of this thesis is a comprehensive *in silico* inventory and evolutionary analysis of *Trichoderma* ANKyrome. To achieve this aim the following tasks should be completed:

- Building an inventory of all ANKdc-proteins from 10 *Trichoderma* spp. and eight closely related non-*Trichoderma* Hypocrealean genomes (NTHGs), referred to as Hypocrealean ANKyrome
- Re-annotation of ANK-repeats and other associated host-domains (where present) in Hypocrealean ANKyrome using InterProScan
- Domain architecture analysis
- Comparison of *Trichoderma* ANKyrome to other closely related Hypocreales.
- Orthology inferences within the Hypocrealean ANKyrome
- Phylogenetic and selection pressure analysis of core ANKdc-proteins from considered *Trichoderma* spp.

3 Materials and Methods

3.1 Preliminary investigation

3.1.1 Whole Genome and ANKdc-protein statistics among Hypocreales

JGI (Joint Genome Institute) Mycocosm website (https://genome.jgi.doe.gov/programs/fungi/index.jsf) enables filtering of species at different taxonomical levels by using an interactive phylogram (Grigoriev *et al*. 2013). Genome statistics of all species from a selected subdivision, order or genus were downloaded and transferred into excel for data analysis. This data was used to create box plots of genome sizes and N° of gene models at different taxonomic levels from Ascomycota (division) to *Trichoderma* (species).

All species belonging to Hypocreales taxonomical level were filtered and by searching for "ankyrin" keyword in JGI search total number of ANKdc-proteins per species was obtained. All hits were transferred into excel and counted to obtain the number of ANKdc-proteins per species.

3.1.2 Retrieval of putative ANKdc-protein sequences

All ANKdc-proteins from *Trichoderma guizhouense* NJAU 4742 (Druzhinina *et al.* 2018), *T. parareesei* TUCIM 717 (Yang *et al.* 2015) and *T. afroharzianum* LTR were obtained from a non-public local database based on already available annotations by InterProScan (Jones *et al.* 2014). The ANKdc-proteins from other 15 Hypocreales species were downloaded from JGI's Mycocosm website. The keyword "ankyrin" was used in JGI website´s search function to find all proteins that contained at least one ANK-domain, based on JGI's annotation. The ANKdc-protein sequences were downloaded in amino acid fasta format from genomes listed in Table 1.

Hypocrealean ANKyrome is the inventory of all ANKdc-proteins from 10 *Trichoderma* spp. and eight closely related non-*Trichoderma* Hypocrealean Genomes (NTHGs).

Note: This study is based on the assumption that protein domain assignations of the genomes carried out by JGI using their custom annotation pipelines are correct.

Genome	Reference	NCBI Accession
Trichoderma reesei Qm6a	Martinez *et al.* 2008b	PRJNA225530
T. parareesei TUCIM 717	Yang *et al.* 2015	LFMI00000000
T. longibrachiatum ATCC 18648	Druzhinina *et al.* 2018	MBDJ00000000
T. citrinoviride TUCIM 6016		MBDI00000000
T. guizhouense NJAU 4742		LVVK00000000
T. harzianum CBS 226.95		MBGI00000000
T. afroharzianum LTR		
T. virens Gv29-8		PRJNA264113
T. asperellum CBS 433.97	Kubicek *et al.* 2011	MBGH00000000
T. atroviride IMI 206040		PRJNA264112
Metarhizium acridum CQMa102	Gao *et al.* 2011	PRJNA38715
M. robertsii ARSEF 23		PRJNA245140
Beauveria bassiana ARSEF 2860	Xiao *et al.* 2012	PRJNA225503
Cordyceps militaris CM01	Zheng *et al.* 2012	PRJNA225510
Fusarium graminearum PH-1	Cuomo *et al.* 2007	AACM00000000
F. oxysporum f. sp. *lycopersici* strain 4287	Ma *et al.* 2010	PRJNA342688
F. fujikuroi IMI 58289	Wiemann *et al.* 2013	PRJEB185
Nectria haematococca MPVI-77-13-4	Coleman *et al.* 2009	PRJNA51499

Table 1: Hypocreales species from which ANKdc-proteins were retrieved

3.2 Re-annotations of Hyporcrealean ANKyrome

ANK-repeat annotations

The re-annotation of domains present in Hypocrealean ANKyrome was performed in batch by using Blast2GO software (Conesa *et al*. 2005, 2) with inbuilt InterProScan tool (Jones *et al*. 2014).

Three main databases for ANK repeat domain annotation implemented in InterProScan were Pfam (Bateman *et al*. 2004), Prosite (Hofmann *et al*. 1999) and SMART (Letunic, Doerks, and Bork 2011). The assigned IPR signature in InterProScan for ANK-repeat is IPR002110. ANK-repeats were filtered and to avoid duplicates specific database signature IDs for ANK-repeat were used. SMART database hits were filtered first, and the corresponding annotated proteins removed from the list. Then, the remaining proteins were filtered by Prosite and Pfam annotations, respectively.

Host-domain annotations

The annotations of various host-domains associated with ANK-domains within the Hypocrealean ANKyrome were obtained from 14 different databases. Therefore, duplicate annotations had to be removed after a consensus for each host-domain was determined. The cured annotations were used to find those host-domains that were present in higher abundance in *Trichoderma* spp. or in NTHGs. Then, these domains were compared at species level for each of 18 considered Hypocreales genomes. Furthermore, the most common host-domains in orphan proteins were investigated and compared to host-domain annotations in orthogroup proteins in *Trichoderma* spp.

Determination of positional preference of ANK-repeats in T. virens ANKyrome

After reannotation the position of each ANK-domain in ANKyrome from *T. virens* was visually determined and classified into 3 groups, N-terminal, C-terminal or Mid-section. *T. virens* was selected because it had the largest number of ANKdc-proteins in its ANKyrome.

3.3 Orthology Analysis of Hypocrealean ANKyrome

OrthoFinder v1.1.8 (Emms and Kelly 2015) was used to identify orthologs within the Hypocrealean ANKyrome obtained in 3.1.2 and then to group them into independent orthogroups. A complete OrthoFinder analysis was performed consisting of BLAST all-vs-all and MCL clustering, as described in the following paragraph. A multiple sequence alignment (MSA) of each orthogroup was performed as part of the OrthoFinder algorithm by using MAFFT auto strategy that selects an alignment method based on data size (Katoh *et al.* 2002; Katoh and Standley 2013).

Curing and filtering of putative orthogroups

The inferred MSAs of putative orthogroups were visually inspected and realigned in AliView (Larsson 2014) with MUSCLE (Edgar 2004) to detect misaligned regions more precisely. Certain putative orthogroups were split into separate orthogroups and some were rejected as orthogroups. ANKdc-proteins that were not assigned to any orthogroup will be referred to as orphan proteins in this thesis.

To confirm their orphan status, orphan ANKdc-proteins were subjected to second OrthoFinder analysis. The resulting orthogroups had to be filtered and cured as previously explained. Based on visual inspection of each orthogroup's alignment using Geneious (Kearse *et al.* 2012), a 23 % pairwise identity was accepted as threshold for orthogroups.

OrthoFinder algorithm

An overview of steps in OrthoFinder algorithm is described in Figure 5 . The first step of the algorithm is a BLAST all-versus-all (Altschul *et al*. 1990) with a higher than recommended E-value threshold of 10^{-3} to prevent exclusion of very short sequences (Figure 5b).

In the second step BLAST bit scores are normalized to ensure that the best hit is assigned score independent of protein length or phylogenetic distance Bit scores are used because the lowest E-value is e^{-180} and no resolution of sequences with high similarities is achieved (Figure 5c).

In the next step, RBNHs (Reciprocal Best Length-Normalized hit) is calculated for each protein to set the lower limit for acceptance of putative orthologs (Figure 5d).

In the last two steps, putative orthologs are connected in the orthogroup graph based on normalized BLAST bit scores and then clustered into orthogroups using MCL (Figure 5f).

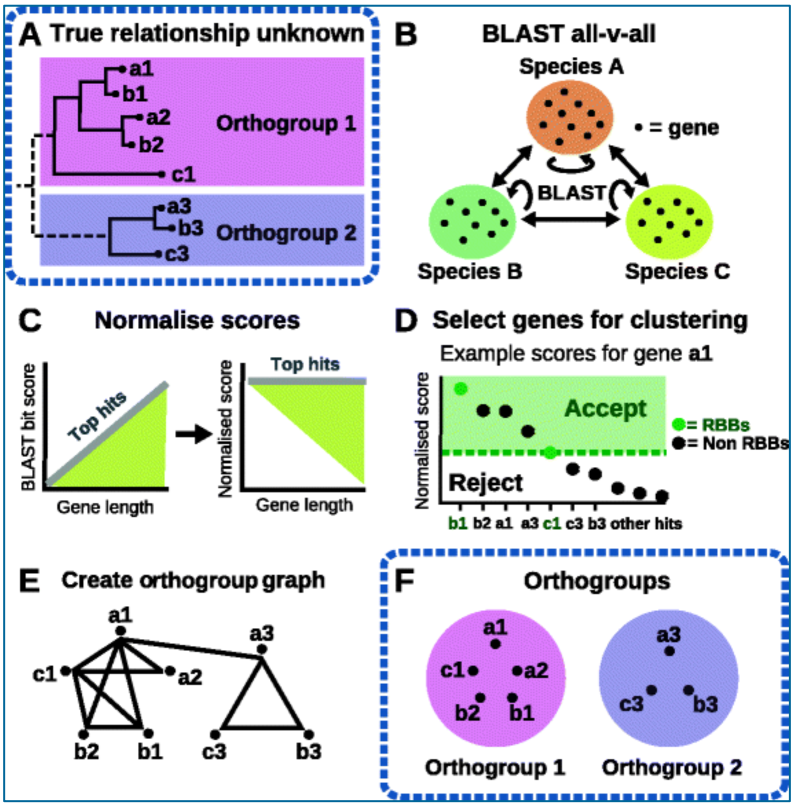


Figure 5: Overview of the steps in the OrthoFinder algorithm. Emms and Kelly 2015

3.3.1 UpSet analysis to reveal shared ANKyrome and core orthologous genes

Universal ANKdc-proteins shared by species of different taxonomical sections were determined by analyzing inferred orthogroups. The results were summarized and visualized using an R-package (Team 2013) UpSet (Lex *et al*. 2014). This tool visualizes set intersections quantitatively in a matrix layout. A set is defined as a collection of distinct elements that describes a common characteristic. ANKdc-proteins (distinct elements) inventory, or ANKyrome, from one species (common characteristic) was considered a single set. If at least one ortholog from each of two different species is present in same orthogroup, this orthogroup is considered an intersection between ANKyromes of these two species. By counting all orthogroups where at least one representative from each of these two species is present, the number of intersections between them is obtained. Accordingly, core *Trichoderma* orthogroups were considered to be those orthogroups that contained at least one representative of each of the 10 considered *Trichoderma* spp.

3.4 Evolutionary Analysis for core orthologous genes

3.4.1 Phylogenetic Analysis

BLAST with representatives from core orthogroups

Only four sequences, one from each taxonomical section of *Trichoderma* genus, *T. harzianum*, *T. virens*, *T. atroviride* and *T. reesei,* from each core orthogroup were used for similarity search by Geneious 10.2.2. batch BLAST (Kearse *et al*. 2012). A batch BLAST was performed by using an E-value cutoff of e^{-30}. The BLAST results were merged with the original orthogroups before curing. After preliminary alignment using Muscle, duplicates were removed, and only unique sequences were retained. Furthermore, accession numbers without species names were searched in NCBI Protein (Coordinators 2016) to find their corresponding source. Each orthogroup was visually inspected and realigned with MUSCLE. Non-homologs were removed in multiple iterative steps.

Multiple sequence alignments

The sequences were realigned using MAFFT accurate algorithm. Next, Gapstreeze tool (https://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html) was used to remove flanking regions by assigning a threshold of 84-90% for gaps in aligned columns. Cured MSAs were used in following phylogenetic analysis.

Maximum Likelihood phylogeny inference

Cured protein alignments were subjected to maximum likelihood phylogenetic estimation by uploading them to PhyML 3.0 public server (http://www.atgc-montpellier.fr/phyml/) (Stephane Guindon *et al*. 2005; Stéphane Guindon *et al.* 2010). An automatic substitution model selection was performed by SMS (Lefort, Longueville, and Gascuel 2017) by using the Bayesian Information Criterion. For the branch support of maximum likelihood phylogeny, 1000 bootstrap replicates were performed.

Bayesian phylogeny inference

Two simultaneous, independent analyses starting from different random trees were run in MrBayes (Ronquist *et al*. 2012), each using three heated chains and one "cold" chain. Once the analyses were completed, 7500 trees were summarized after discarding the first 25% of the obtained 10,000 trees, resulting in a consensus tree.

The output was assessed based on a convergence diagnostic PSRF (Potential scale reduction factor) which compares variance within and between runs and had to be close to 1.0, otherwise the analysis had to be ran longer. If PSRF value was close to 1.0 the trees were summarized by removing 25 % of samples. The output contained a phylogram with mean branch lengths. The same substitution models selected for each alignment by SMS previously were also used in the Bayesian analyses.

3.4.2 Selection Pressure Analysis

Preparation of codon alignments with Pal2Nal

Corresponding cDNA sequences for core protein ANKyrome were retrieved from JGI. PAL2NAL (Suyama, Torrents, and Bork 2006) online tool (http://www.bork.embl.de/pal2nal/) uses a protein MSA and its corresponding cDNA sequences to prepare a cDNA-MSA. Since cDNA sequences of proteins from *T. guizhouense* NJAU 4742, *T. parareesei* TUCIM 717 and *T. afroharzianum* LTR were not available, these genomes were excluded from subsequent analyses.

Preparation of Bayesian trees for Selection Pressure analyses

The protein MSA alignments with reduced MOTUs that correspond to MOTUs in cDNA MSA were subjected to Bayesian phylogenetic analyses according to the procedure described in 3.4.1. Inferred trees were combined with codon alignments to obtain a nexus file necessary for selection pressure analyses.

Selection Pressure Analysis

It has been shown that recombinant sequences lead to the false positive detection of positive selection in selection pressure analyses (Shriner *et al*. 2003). This can be prevented if the MSA is split into respective non-recombinant parts. For this purpose, Genetic Algorithm for Recombination Detection (GARD) for identification of non-recombinant fragments in an MSA was developed by Kosakovsky Pond *et al.* 2006. The nexus files containing both MSAs and trees were subjected to GARD. If a recombination breakpoint is detected in an MSA, the MSA is split into non-recombinant parts, which can be independently subjected to selection pressure analyses.

The first selection pressure analysis was performed by using BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification), a protein-wide positive selection pressure analysis. This tool indicates if at least one site or one branch in the protein has evolved under positive selection. The significant result is not conclusive, so further analyses at site-level are necessary for a complete hypothesis testing (Murrell *et al*. 2015).

In the next step, MEME (Mixed Effects Model of Evolution) tool was used. MEME is an algorithm for detection of episodic (on a subset of branches) positive selection pressure analysis at individual sites that allows the levels of positive selection to vary from branch to branch (Murrell *et al*. 2012). For the visualization of results β - α (β is the relative rate of nonsynonymous substitutions, α is the relative rate of synonymous substitutions) is used as a measure of the intensity of selection, because for small α values the ratio $\beta/\alpha = \omega$ is misleading ($\omega = 1$ represents neutral evolution, $\omega < 1$ negative selection and $\omega > 1$ positive selection). Both MEME and BUSTED were used as online implementations on datamonkey.org server (Delport *et al*. 2010).

In the last step of selection pressure analysis, FUBAR (Fast, Unconstrained Bayesian AppRoximation) tool was used. This tool is used for detection of pervasive (across whole phylogeny) positive or purifying selection at individual sites based on a Bayesian approach (Murrell *et al*. 2013). The selection pressure for each site is constant along the whole phylogeny and the inferred rates are supported by their corresponding posterior probabilities. FUBAR was used as part of the offline HyPhy tool package (Pond and Muse 2005).

Inferences of selection pressure analysis were made based on results from all of the above-mentioned tools.

3.5 TMHMM and SignalP analyses

The Hypocrealean ANKyrome was subjected to signal peptide cleavage site prediction on the online server http://www.cbs.dtu.dk/services/SignalP/ (Petersen *et al*. 2011). Furthermore, these proteins were also subjected to the prediction of transmembrane helices in proteins on the public server http://www.cbs.dtu.dk/services/TMHMM/ (Sonnhammer, Von Heijne, and Krogh 1998). These analyses were used to detect trends in the share distribution of signal peptides and transmembrane helices in Hypocrealean ANKyrome and compare them on the species level.

3.6 Transcriptomic data analysis

Each protein from *Trichoderma* ANKyrome was screened against differentially expressed genes (RNA deepseq) obtained during dual confrontation assays of *Trichoderma guizhouense* NJAU 4742 and *Trichoderma harzianum* TUCIM 916 *(*=CBS 226.95) against *Fusarium oxysporum* f. sp. *cubense* 4 strain (Foc4). Transcriptional response was investigated in a confrontation of *T. guizhouense* against itself, against *T. harzianum* and finally against Foc4. The same approach was employed for *T. harzianum* (Zhang, Miao, Rahimi, Shen, Druzhinina, unpublished).

4 Results and discussions

4.1 Inventory of the *Trichoderma* ANKyrome

4.1.1 Genome and ANKdc-protein statistics comparison

Genomes from the taxonomical order Hypocreales, including those that were considered in this study, that were selected for statistical comparison are presented in Table 2. The listed *Trichoderma* spp. belong to four different taxonomical sections also shown in the phylogram from Figure 2, the *Longibrachiatum*, *Harzianum*, *Trichoderma* and *Virens*, respectively. Within the *Trichoderma* genus, average genome size is 36.3 Mbp, while the average number of gene models is 11618. *T. harzianum* CBS 226.95 is the largest genome with highest number of gene-models among all *Trichoderma* spp., whereas *T. parareesei* TUCIM 717 is the smallest genome. The *Longibrachiatum* section's genomes are generally smaller, when compared to the rest of considered *Trichoderma* spp. and this is also the youngest section among *Trichoderma* spp.

Genome	Genome size [Mbp]	N° of gene models	TAXONOMICAL SECTIONS
Trichoderma reesei Qm6a	34.1	9129	***LONGIBRACHIATUM***
T. parareesei TUCIM 717	31.13	9318	
T. longibrachiatum ATCC 18648	32.2	10938	
T. citrinoviride TUCIM 6016	33.22	9737	
T. guizhouense NJAU 4742	38.29	11297	***HARZIANUM***
T. harzianum CBS 226.95	**40.98**	14095	
T. afroharzianum LTR	-	-	
T. virens Gv29-8	39	12518	***VIRENS***
T. asperellum CBS 433.97	37.5	12586	***HARZIANUM***
T. atroviride IMI 206040	36.1	11863	
T. gamsii T6085	37.97	10944	
Escovopsis weberi CC031208-10	29.45	6870	
Metarhizium acridum CQMa102	39.42	9849	
M. robertsii ARSEF 23	39.15	10583	
Calviceps purpurea 20.1	32.09	8979	
Ophiocordyceps sinensis CO18	**78.52**	6972	
Beauveria bassiana ARSEF 2860	33.69	10364	
Cordyceps militaris CM01	32.27	9651	
Fusarium graminearum PH-1	36.45	13322	
F. pseudograminearum CS3096	36.93	12448	
F. oxysporum f. sp. *lycopersici* strain 4287	61.36	17708	
F. fujikuroi IMI 58289	43.83	14813	
Nectria haematococca MPVI-77-13-4	51.49	15707	

Table 2: Hypocreales genomes that were used for characteristics comparison

Comparison of Genome sizes between Hypocrealean and Non Hypocrealean spp

Data from JGI was used to compare all available species from Ascomycota level to Hypocreales level. At both Ascomycota and Pezizomycotina levels the largest genome is almost 180 Mbp, specifically the *Cenoccum geophilum* 1.58 from Dothideomycetes group. The second largest genome among all Ascomycota is *Blumeria graminis* f. sp. *tritici* 96224. The results are visualized in a box plot in Figure 6a.

In the second box plot in Figure 6b number of gene models (or JGI predicted gene) per species is compared. The largest number of genes are found in *Fusarium oxysporum* f. sp*. lycopersici strain* 4287.

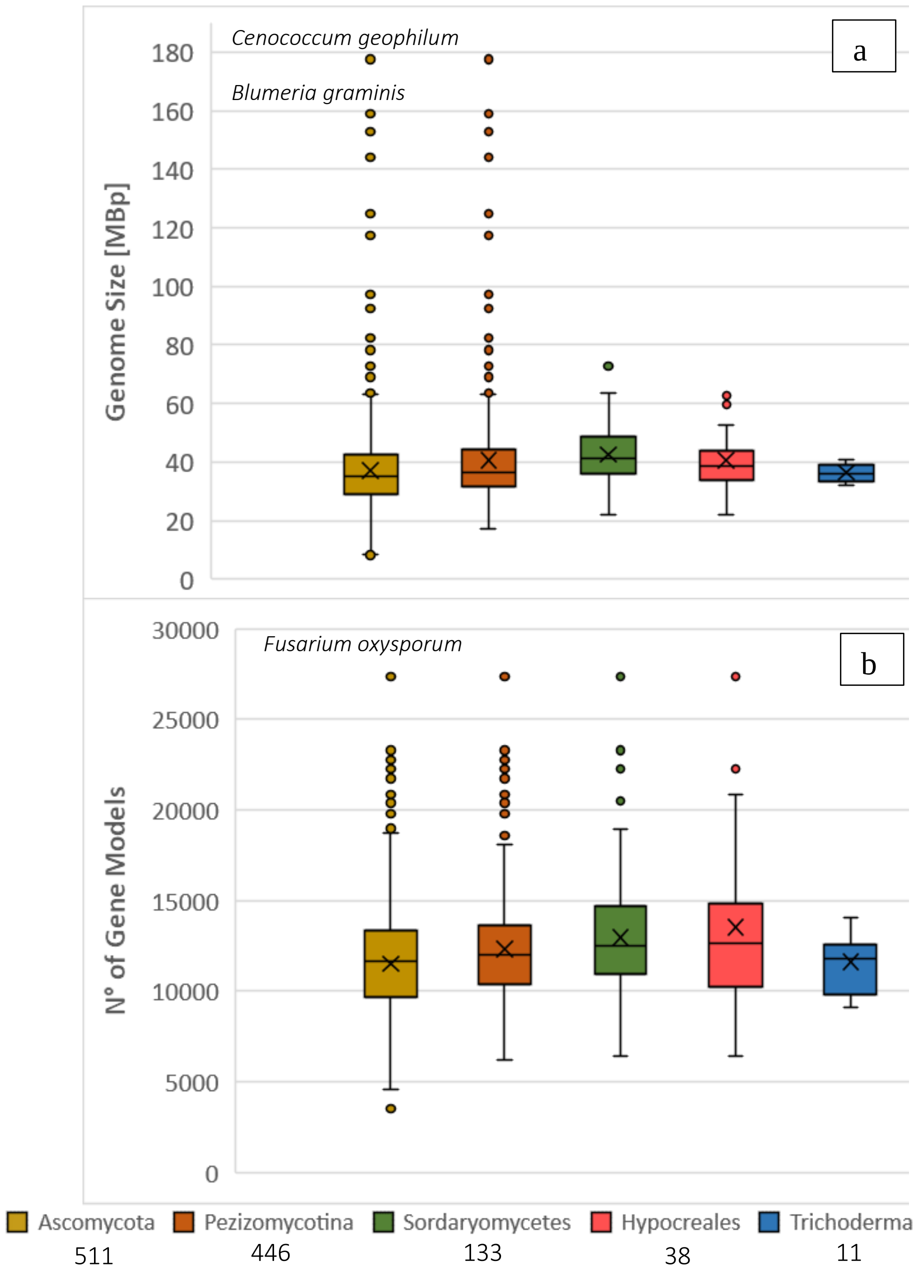


Figure 6a and 6b: Box plots showing range of genome sizes and number of total proteins at different taxonomical levels

ANKdc-protein expansion in species of higher taxonomical orders

The results of ANKdc-proteins abundance comparison in Hypocreales genomes are summarized in Table 3. In *T. virens* ANKdc-proteins account for 1.56 % of all gene models which is only superseded by *Ilyonectria robusta* PMI 751 inside the Hypocreales order. Some of the highest shares of ANKdc-proteins among considered Hypocreales species are found in *T. atroviride* with 1.3 %, *T. harzianum* with 1.27 %, *T. gamsii* with 1.22 % and *T. guizhouense* with 1.22 % of all genes. Below 1 % share of ANKdc-proteins are found in *Trichoderma* spp. from *Longibrachiatum* section. These findings are in concordance with results from the Druzhinina *et al.* unpublished work and confirm that the expansion of ANKdc-proteins in most aggressive *Trichoderma* spp. is present. But, according to these updated results the expansion is not exclusive to *Trichoderma* spp. even if only the species from Hypocreales order are taken into account. Two *Ilyonectria* spp. had the highest share of ANKdc-genes and also the highest total number of ANKdc-proteins. These genomes were sequenced and published in 2016 and 2017 so the data was not available at the time of preliminary investigation by Kubicek *et al.* in 2011. The species from *Ilyonectria* genus are common soil fungi that cause root rot as opportunistic phytopathogen but can also be found as endophytes of healthy plants (dos Santos *et al.* 2014).

4.1.2 Retrieval of putative ANKdc-proteins from selected Hypocreales

Total number of mined ANKdc-proteins from the respective source species, along with minimum and maximum protein lengths are summarized in Table 4. The highest number of ANKdc-proteins are found in *Fusarium oxysporum,* which also hast the largest number of proteins in general. Among *Trichoderma* spp. the most ANKdc-proteins were obtained from *T. virens*, while the least were present in *T. reesei.*

Organism Name	Genome Size [Mbp]	N° of Gene Models	N° of ANKdc-proteins	Share of ANKdc-proteins [%]
Ilyonectria robusta PMI 751	59.65	20499	345	1.68
I. europaea CBS 129078	62.83	20870	326	1.56
Trichoderma virens Gv29-8	39.02	12423	194	1.56
Clonostachys rosea CBS125111	52.44	18639	267	1.43
T. harzianum TR274	40.87	13932	182	1.31
T. atroviride IMI 206040	36.14	11828	154	1.30
T. harzianum CBS 226.95	40.98	14095	179	1.27
T. gamsii T6085	37.97	10944	134	1.22
T. guizhouense NJAU 4742	38.29	11297	138	1.22
Pochonia chlamydosporia 170	44.22	14204	172	1.21
Metarhizium robertsii ARSEF 23	41.66	11688	135	1.16
Stachybotrys elegans LAHC-LSPK-M15	43.47	14925	169	1.13
Fusarium fujikuroi IMI 58289	43.83	14813	163	1.10
F. redolens A4	52.56	17051	185	1.08
Nectria haematococca	51.29	15707	170	1.08
T. asperellum TR356	35.39	12320	133	1.08
T. asperellum CBS 433.97	37.46	12586	135	1.07
T. parareesei TUCIM 717	31.13	9318	95	1.02
Myrothecium inundatum CBS 120646	39.21	13553	134	0.99
T. citrinoviride TUCIM 6016	33.22	9737	96	0.99
Neonectria ditissima R09/05	45.72	12685	123	0.97
F. oxysporum f. sp. lycopersici 4287	61.36	27347	264	0.97
Beauveria bassiana ARSEF 2860	33.69	10364	98	0.95
M. acridum CQMa 102	39.42	9849	92	0.93
F. pseudograminearum CS3096	36.33	12395	115	0.93
F. verticillioides 7600	41.78	20553	188	0.91
T. reesei Qm6a	33.45	9143	83	0.91
Mariannaea sp. PMI_226	42.25	12638	113	0.89
T. longibrachiatum ATCC 18648	32.24	10938	96	0.88
F. graminearum v1.0	36.45	13322	114	0.86
Niesslia exilis CBS 358.70 v1.0	35.38	13499	99	0.73
Tolypocladium inflatum NRRL 8044	30.35	9998	67	0.67
Purpureocillium sp. UdeA0106 v1.0	36.08	13642	90	0.66
Acremonium strictum DS1bioAY4a v1.0	35.79	13158	84	0.64
Cordyceps militaris CM01	32.27	9651	61	0.63
A. chrysogenum ATCC 11550	28.56	8899	51	0.57
Ustilaginoidea virens	33.57	6451	29	0.45
Valetoniellopsis laxa CBS 191.97 v1.0	22.13	8026	36	0.45

Table 3: N° of ANKdc-proteins in different Hypocreales fungi publicly available in JGI, 6th Nov, 2017

Genome	N° of proteins downloaded	Average N° of proteins per taxonomical section	Max Sequence Length [AA]	Min Sequence Length [AA]
Trichoderma parareesei TUCIM 717	95	92.5	2125	97
T. reesei Qm6a	83		2123	88
T. longibrachiatum ATCC 18648	96		2120	72
T. citrinoviride TUCIM 6016	96		1924	63
T. guizhouense NJAU 4742	138	152	2563	154
T. harzianum CBS 226.95	179		2113	56
T. afroharzianum LTR	139		2645	154
T. virens Gv29-8	194	194	2147	57
T. atroviride IMI 206040	154	144.5	2105	68
T. asperellum CBS 433.97	135		2114	53
Beauveria bassiana ARSEF 2860	98	96.5	2259	118
Cordyceps militaris CM01	61		2106	181
Metarhizium robertsii ARSEF 23	135		2251	91
M. acridium CQMa102	92		1849	87
Fusarium oxysporum f. sp. lycopersici strain 4287	264	177.5	2209	107
F. graminearum PH-1	114		2087	65
F. fujikuroi IMI 58289	163		2533	157
Nectria haematococca MPVI-77-13-4	170		2242	73

Table 4: Genome statistics in Hypocreales

4.1.3 Re-annotated Hypocrealean ANKyrome

Annotation of ANK-repeats

From 2406 proteins from Hypocrealean ANKyrome 2036 had at least one annotated ANK repeat from at least one of the selected databases. All filtered hits were sorted by their database signature ID and the total number of hits per database are shown in Table 5.

	Database Signature ID	N° of annotations
Pfam	PF00023	373
	PF13606	105
Prosite	PS50088	6882
SMART	SM00248	13498
TOTAL	-	20858

Table 5: Annotations of ANK repeats per database

The results from Table 5 indicated that SMART database was superset of other two databases, which was additionally confirmed by randomly analyzing some proteins in online InterProScan and visual inspection of graphical outputs. For this reason, SMART annotations were filtered first, which resulted in annotation of 13502 ANK-repeats from 1961 out of 2406 query proteins. The annotations in remaining 445 proteins, were filtered first by Prosite and then by Pfam results. Finally, additional 75 ANK-repeats were obtained, for the total of 13581 annotated ANK-repeat sequences. The results are summarized in Table 6.

	Signature ID	N° of annotations
SMART	SM00248	13502
Prosite	PS50088	74
Pfam	PF00023	5
TOTAL		**13581**

Table 6: Source databases of final ANK-repeat annotations

N° of ANK-repeats per ANKdc-proteins on species level were analyzed and the results are shown in Figure 7 and the corresponding Supplementary material 1. On average there are 6.5 ANK-repeats per ANKdc-protein within the *Trichoderma* ANKyrome and within the NTHGs ANKyrome there are 6.9 ANK-repeats per ANKdc-protein on average. The most ANK-repeats per ANKdc-protein within the Hypocrealean ANKyrome were present in *T. guizhouense* with 9.4 followed by 8.5 ANK-repeats per ANKdc-protein in *T. afroharzianum*.

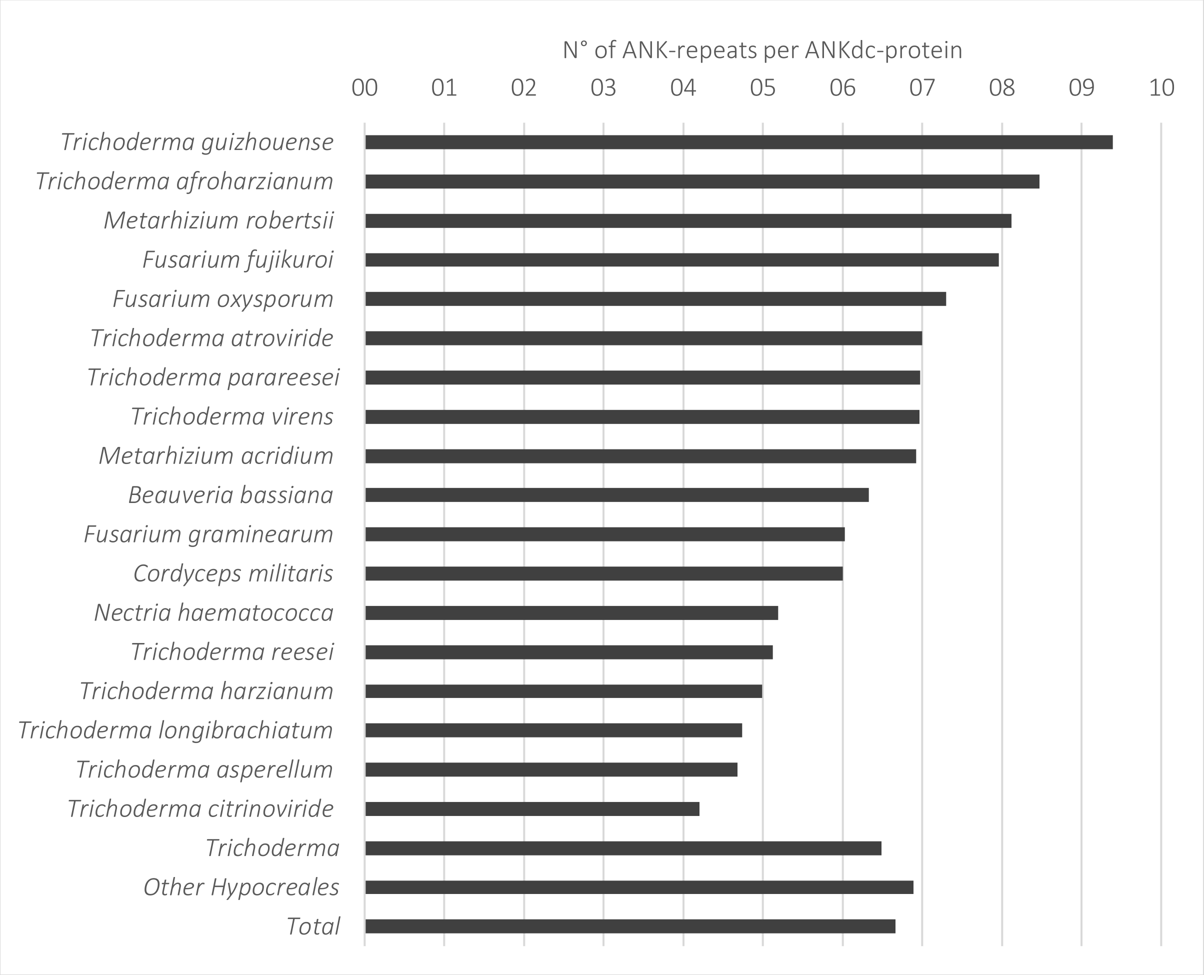


Figure 7: Comparison of number of annotated repeats per species

Host-domain annotation in ANKdc-proteins

According to InterProScan analysis results, 48.7% of ANKdc-proteins in the Hypocrealean ANKyrome contained ANK-domain as the only identified domain (loner-ANKdc-proteins). The ANKdc-proteins that contain other associated domains will be referred to as host-domain containing ANKdc-proteins (hdc-ANKdc-proteins) in this thesis.

In the *Trichoderma* ANKyrome of 1309 ANKdc-proteins, 49% or 642 proteins are hdc-ANKdc-proteins. The *T. guizhouense* and *T. afroharzianum* ANKyromes contained the highest share of hdc-ANKdc-proteins with 63% and 59%, respectively. *T. virens* ANKyrome has the lowest share of hdc-ANKdc-proteins with 37.1%, implying that most of the ANKdc proteins in *T. virens* are loner-ANKdc proteins. The complete list of hdc-ANKdc-proteins in individual species are shown in Figure 8 and Supplementary material 2

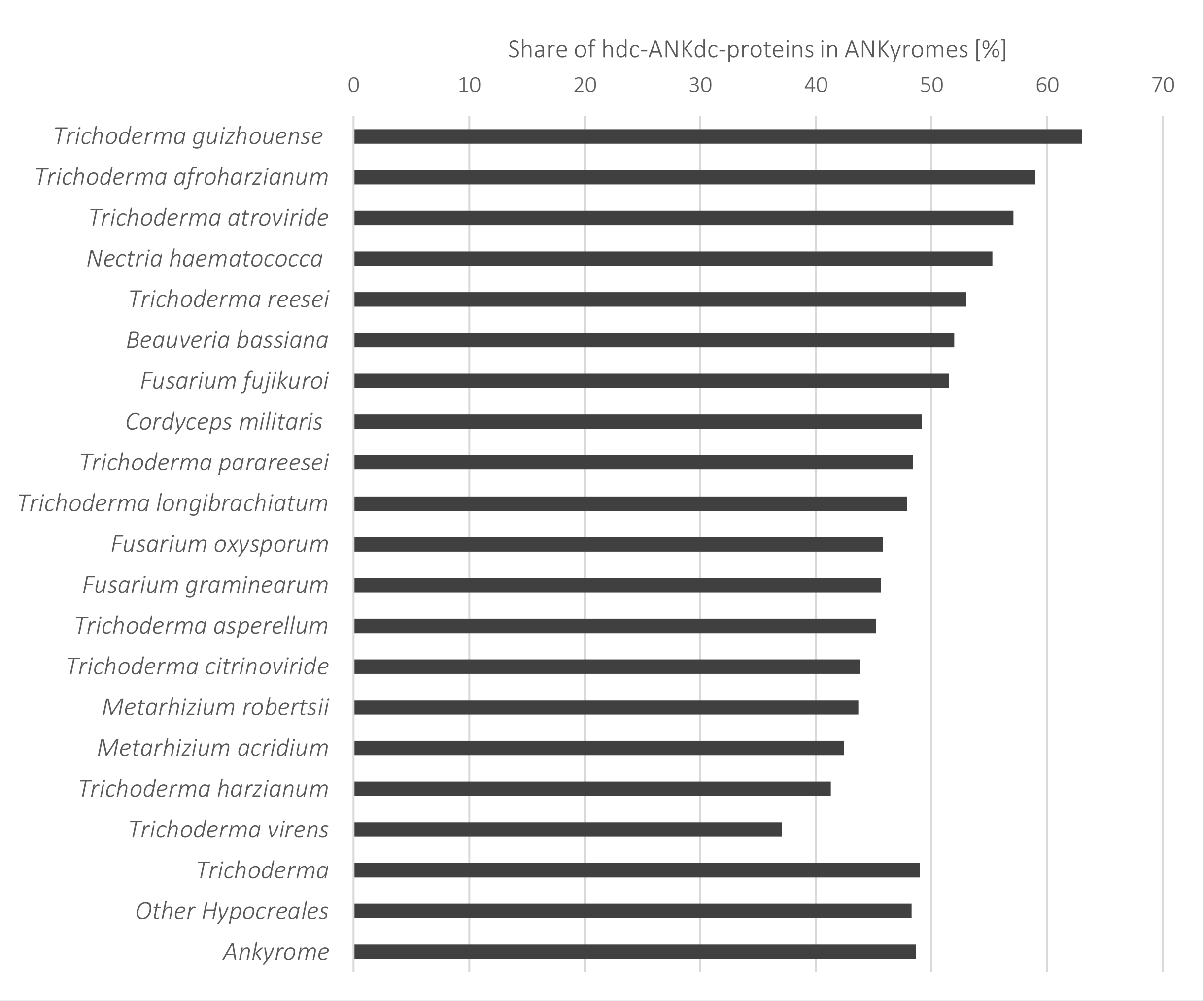


Figure 8: Share distribution of hdc-ANKdc-proteins in Trichoderma *spp. and NTHGs*

The share distribution of twelve most common associated domains found in hdc-ANKdc-proteins from *Trichoderma* spp. is shown in Figure 9, while the full list of annotations is given in Supplementary material 3.

The three most common domains that are associated with ANK-domain in hdc-ANKdc-proteins from *Trichoderma* ANKyrome are P-loop containing nucleoside triphosphate hydrolase (PLNTH), nucleoside phosphorylase (NP) and NACHT domain (named after proteins it is present in, NAIP: NLP family apoptosis inhibitor protein, CIITA: C2TA or MHC class II transcription activator, HET-E: incompatibility locus protein from *Podospora anserina* and TEP1: TP1 or telomerase-associated protein). PLNTHs are present in 24.5 % of hdc-ANKdc-proteins from *Trichoderma* spp. and in 26 % of hdc-ANKdc-proteins from NTHGs. The PLNTHs catalyze the hydrolysis of NTPs (specific for ATP/GTP) beta-gamma bond and the released energy is used for conformational changes in other molecules (Leipe, Koonin, and Aravind 2004).

NPs were found in almost 13 % of hdc-ANKdc-proteins from *Trichoderma* spp. and 19 % of hdc-ANKdc-proteins from NTHGs. Function of this domain depends on which kind of phosphorylase it encodes, eg PurineNP, UridineNP etc. It catalyzes the phosphorolytic breakdown of the N-glycosidic bond to the respective nucleoside and sugar-1-phosphate molecules, which are either used as carbon and energy sources or for nucleotide synthesis (Takehara *et al*. 1995).

The next most common host-domain in *Trichoderma* ANKyrome is NACHT. This domain consists of ATP/GTPase specific P-loop domain, the Mg^{2+}-binding site (Walker A and B motifs, respectively) and five more specific motifs. It is essentially a modified nucleoside phosphatase, but its function remains unknown (Koonin and Aravind 2000).

Further most common domain in *Trichoderma* ANKyrome is protein kinase (PK), F-box and heterokaryon incompatibility motifs (HET). Each of these domains was present in around 5 % of hdc-ANKdc-proteins from *Trichoderma* spp.

PKs catalyze the transfer of gamma phosphate groups from NTPs on proteins. Their function in cells is associated with division, proliferation, apoptosis, and differentiation (Hanks, Quinn, and Hunter 1988).

F-box domain is a part of the E3 ubiquitin ligase complex which beside F-box, contains Skp and Cullin domains. This complex catalyzes the ubiquitation of proteins for their degradation by proteases. The function of F-box is to connect the target protein with Skp (OmpH Chaperone protein) and thereby bring it closer to the E2 enzyme that is loaded with ubiquitin (Kipreos and Pagano 2000).

Filamentous fungi use individual specific HET domains as self-/non-self-discrimination during vegetative fusion. Heterokaryotic cells produced by vegetative fusion of genetically different cells trigger the vegetative incompatibility associated programmed cell death reaction (Paoletti and Clave 2007).

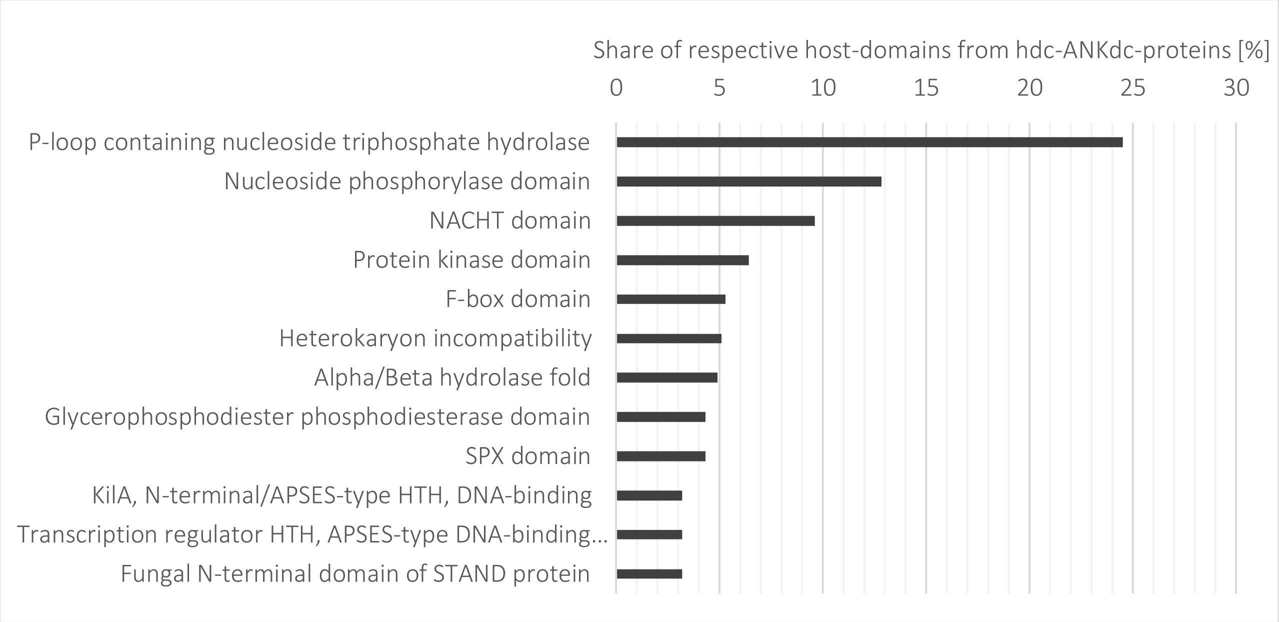


Figure 9: Share distribution of the most common host-domains found in hdc-ANKdc-proteins from Trichoderma *spp.*

The distribution of host-domains was compared between orphan and orthogroup ANKdc-proteins from *Trichoderma* ANKyrome and the results are shown in Figure 10. The two most common host-domains in *Trichoderma* ANKyrome, NP and PLNTH, are predominantly found in orphan ANKdc-proteins. The annotations were sorted based on the largest absolute differences between orphan and orthogroup ANKdc-proteins. There are 123 ANKdc-proteins containing NP domain of which 82% are orphan proteins. Almost 70% of ANKdc proteins containing PLNTH domain are orphan proteins. Furthermore, 85 % of Peptidase S8/S53 domain annotations were found in orphan proteins of hdc-ANKdc proteins from *Trichoderma* spp.

NWD-NACHT NTPase, Sigma domain on NACHT NTPase, BTB/POZ, Kila, HTH, PK, SPX, F-box, Allantoicase, Phosphodiesterase domain were predominantly found in ANKdc-proteins that belong to orthogroups.

In general, 41 % of all 642 *Trichoderma* hdc-ANKdc-proteins are orphan.

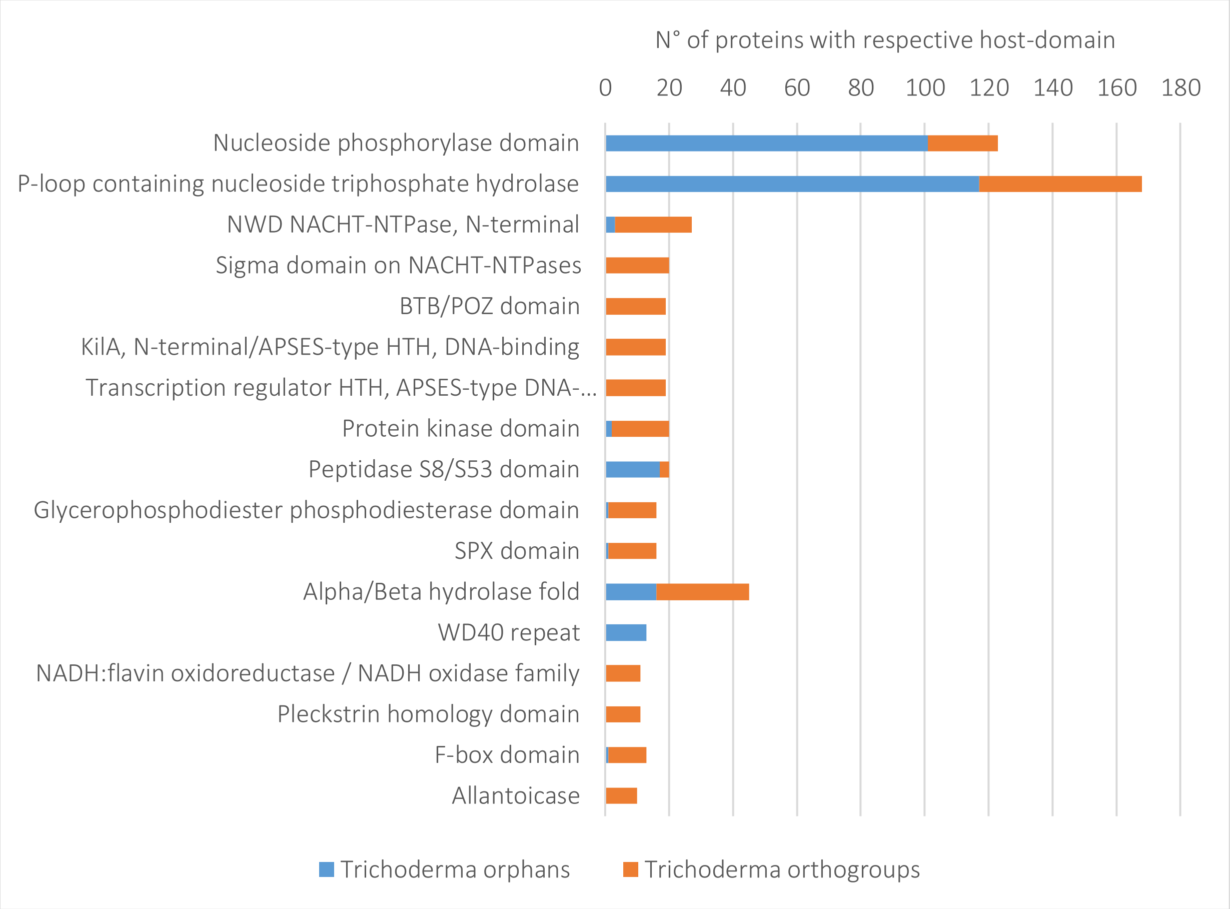


Figure 10: Host-domain annotation distribution comparison between Trichoderma *orhpan ANKdc-proteins and orthogroup ANKdc-proteins*

A direct comparison of share distributions of annotated domains in ANKdc-proteins between *Trichoderma* spp. and NTHGs. was performed to identify group specific host-domains. 20 annotated domains with largest difference in their share abundance among hdc-ANKdc-proteins between *Trichoderma* spp. and NTHGs are shown in Figure 11.

Largest difference was found in NP domain abundance, which was present in 19 % of hdc-ANKdc-proteins from *Trichoderma* spp. and 13 % of hdc-ANKdc-proteins from NTHGs. The most significant difference is the absence of Eisosome protein SEG1/Sle1 in ANKyromes from NTHGs, which was found in almost 4 % of hdc-ANKdc-proteins from *Trichoderma* spp. The function of this protein is unknown.

Other important differences are increased share of ANKdc-proteins containing PK and F-box domains in Hypocreales, but also an increased share of NWD-NACHT NTPase, Sigma domain on NACHT NTPase, NACHT and WD40 repeat domain containing ANKdc-proteins in *Trichoderma* ANKyrome.

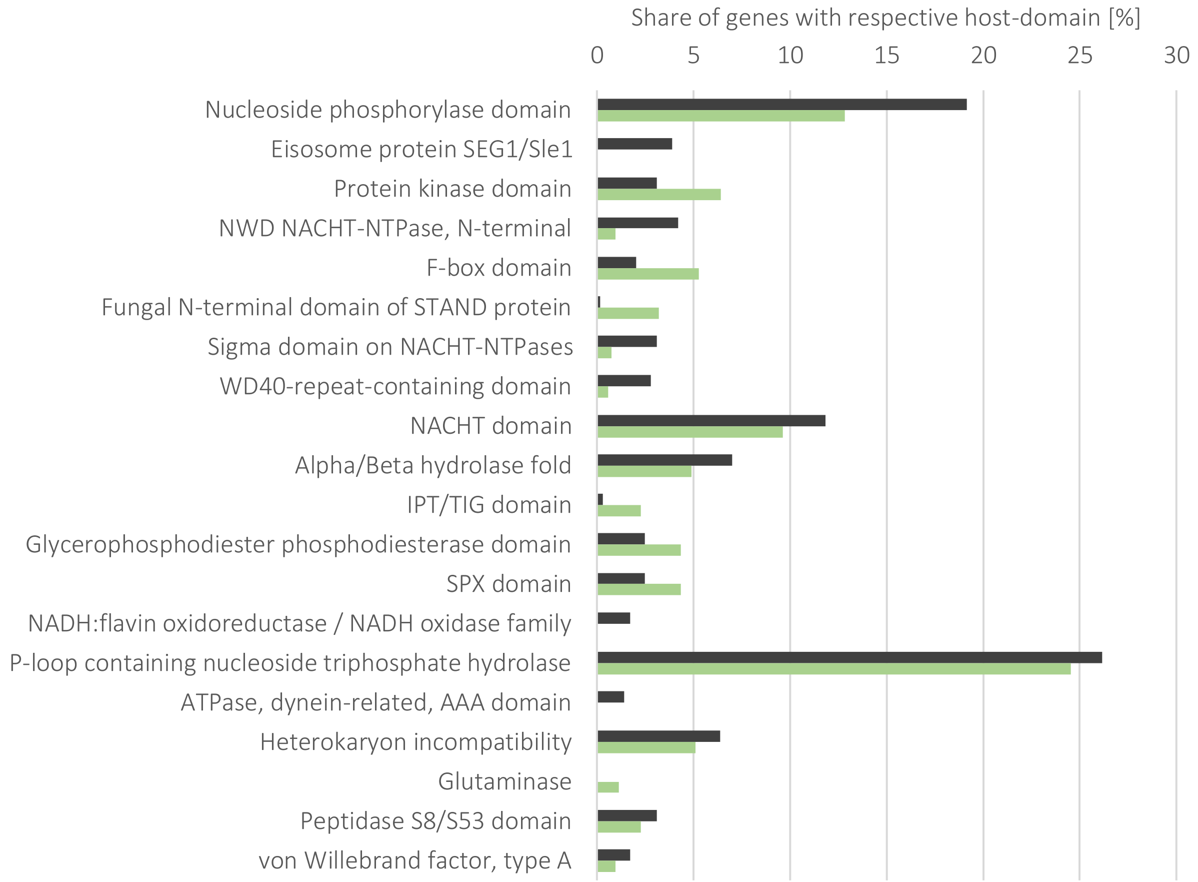


Figure 11: Host-domains annotations with largest differences in their share abundance among hdc-ANKdc-proteins between Trichoderma *spp. and NTHGs*

Comparison of host-domains at the species level

To get a deeper insight into the distribution of host-domains with largest differences in abundance between ANKdc-proteins from *Trichoderma* spp. and ANKdc-proteins from NTHGs, they were investigated at the level of individual species. The results are presented in Figure 12 and Figure 13 and corresponding Supplementary material 3.

Nucleoside phosphorylase domain containing ANKdc-proteins made up the largest share of hdc-ANKdc-proteins with 27 % or 24 of all hdc-ANKdc-proteins from *T. guizhouense*, while in *T. virens* and *T. harzianum* it was found in 25 % and 24 % of all hdc-ANKdc-proteins, respectively (18 proteins each). This type of ANKdc-proteins made up the lowest share of all hdc-ANKdc-proteins in *C. militaris*, accounting only for 3.3 % of all hdc-ANKdc-proteins.

Eisosome protein SEG1/Sle1 is a host-domain specific for ANKdc-proteins from *Trichoderma* spp. as it was not present in hdc-ANKdc-proteins from NTGHs. It is found in 8 % (6 ANKdc-proteins) of *T. harzianum* and 7 % (5 ANKdc-proteins) of hdc-ANKdc-proteins from *T. virens*. But, it is not present in any of hdc-ANKdc-proteins from *T. parareesei*.

Protein kinase domain made up 12 % (15 proteins) of hdc-ANKdc-proteins from *F. oxysporum*, which was by far the largest share % of all analyzed species and the largest absolute number of proteins.

NWD NACHT-NTPase, N-terminal domain is found in 6.9 % in *T. virens* and 6.5 % in *T. asperellum*. It is mostly found in *Harzianum* and *Trichoderma* sections with five to four proteins, while other species either have only one or no proteins containing this domain.

F-box domain is present in 9% of of hdc-ANKdc-proteins from *B. bassiana* and 8 % from *F. oxysporum*. It is important to mention, that the absolute number is highest in *F. oxysporum* with 10 proteins in total, while *B. bassiana* has five ANKdc-proteins containing this domain. Each considered *Trichoderma* sp. has one F-box domain containing ANKdc-protein except for species from *Trichoderma* section and *T. longibrachiatum* which have two ANKdc-proteins of this kind.

Fungal N-terminal domain of STAND protein was almost exclusively found in Phytopathogens except for *T. parareesei* which had only one ANKdc-protein with this host-domain. *F. oxysporum* had the most with eight (6.6 %) and *F. fujikuroi* with five proteins (6 %) with this host-domain. The function of this domain is unknown.

Sigma domain on NACHT-NTPases was present in three proteins in each of the species of *Harzianum* section. The most ANKdc-proteins were found in *T. virens* and *T. atroviride* as each have four ANKdc-proteins with this domain. With the exception of *T. longibrachiatum*, which had one ANKdc-protein containing this domain, this domain was almost absent from the *Longibrachiatum* section.

WD40-repeat-containing domain is most common in *Harzianum* section, where 6.1 % of all hdc-ANKdc-proteins in *T. afroharzianum*, 4.6 % in *T. guizhouense* and 4.1 % in *T. harzianum* contain this domain. In the *Longibrachiatum* section *T. reesei* and *T. parareesei* each have two ANKdc-proteins (4%) containing this domain.

In *Trichoderma* spp., **NACHT domain** is predominantly found in ANKdc-proteins of *Harzainum* section. In *T. guizhouense* 21.8 % of all hdc-ANKdc-proteins or 19 ANKdc-proteins in total contain this domain, while *T. afroharzianum* and *T. harzianum* have 10 and nine of these ANKdc-proteins, respectively. In Hypocreales, *F. oxysporum* has 12, *F. fujikuroi* 11 and *M. robertsii* nine NACHT domain containing ANKdc-proteins in total.

IPT/TIG domain is only present in ANKdc-proteins from *Trichoderma* section and NTHGs. Each species contains only one ANKdc-protein with this domain, whereas only *F. oxysporum* has five proteins of this kind.

Glycerophosphodiester phosphodiesterase domain and **SPX domain** are only absent in hdc-ANKdc-proteins from *T. reesei*, *T. citrinoviride* and *T. longibrachiatum*.

NADH:flavin oxidoreductase / NADH oxidase domain family is specific to hdc-ANKdc-proteins from *Trichoderma* spp. *T. parareesei* has two proteins with this host-domain, while other *Trichoderma* spp. have only 1 ANKdc-protein each.

Heterokaryon incompatibility domain was present in eight (13.5 %) hdc-ANKdc-proteins from *F. fujikuroi,* seven from *T. afroharzianum* and six from *T. guizhouense* and *T. parareesei* each.

Glutaminase was not present in any of hdc-ANKdc-proteins from *Trichoderma* spp. It is found in one ANKdc-protein in each of the Phytopathogens section species and in *B. bassiana* and *C. militaris*.

Concanavalin A-like lectin/glucanase domain superfamily and **SPRY domain** were predominantly found in hdc-ANKdc-proteins from *T. guizhouense*, where each host-domain was present in 8.5 % (5 ANKdc-proteins) of hdc-ANKdc-proteins.

ATPase, dynein-related, AAA domain was exclusively present in species of *Trichoderma* genus. Each species contained one hdc-ANKdc-protein of this type, but it was absent in *T. parareesei*.

P-loop containing nucleoside triphosphate hydrolase (PLNTH) containing ANKdc-proteins are present in highest numbers in *F. oxysporum* with 33 and *F. fujikuroi* with 29 ANKdc-proteins. The highest share of hdc-ANKdc-proteins of this type was found in *T. virens* with 36.1 % (26 ANKdc-proteins), *F. fujikuroi* with 34.5 % and *T. citrinoviride* with 33%, but it was also expanded in species of *Harzianum* section where each species had more than 20 hdc-ANKdc-protein with this domain. Furthermore, in *T. atroviride* there were 26 ANKdc-proteins containing this domain, while in *T. asperellum* only 12 ANKdc-proteins of this type were present.

Leucine-rich repeat domain superfamily and **GAR domain profile** containing ANKdc-proteins are specific to *Longibrachiatum* taxonomical section except for *T. parareesei.*

The majority of host-domains from Hypocrealean ANKyrome are involved in essential cellular functions. As an example, it can be assumed that NTPases are associated with ANK-domain because it can recognize and specifically direct the enzyme to the place where produced energy is needed. The same can be applied to other above-mentioned host-domains.

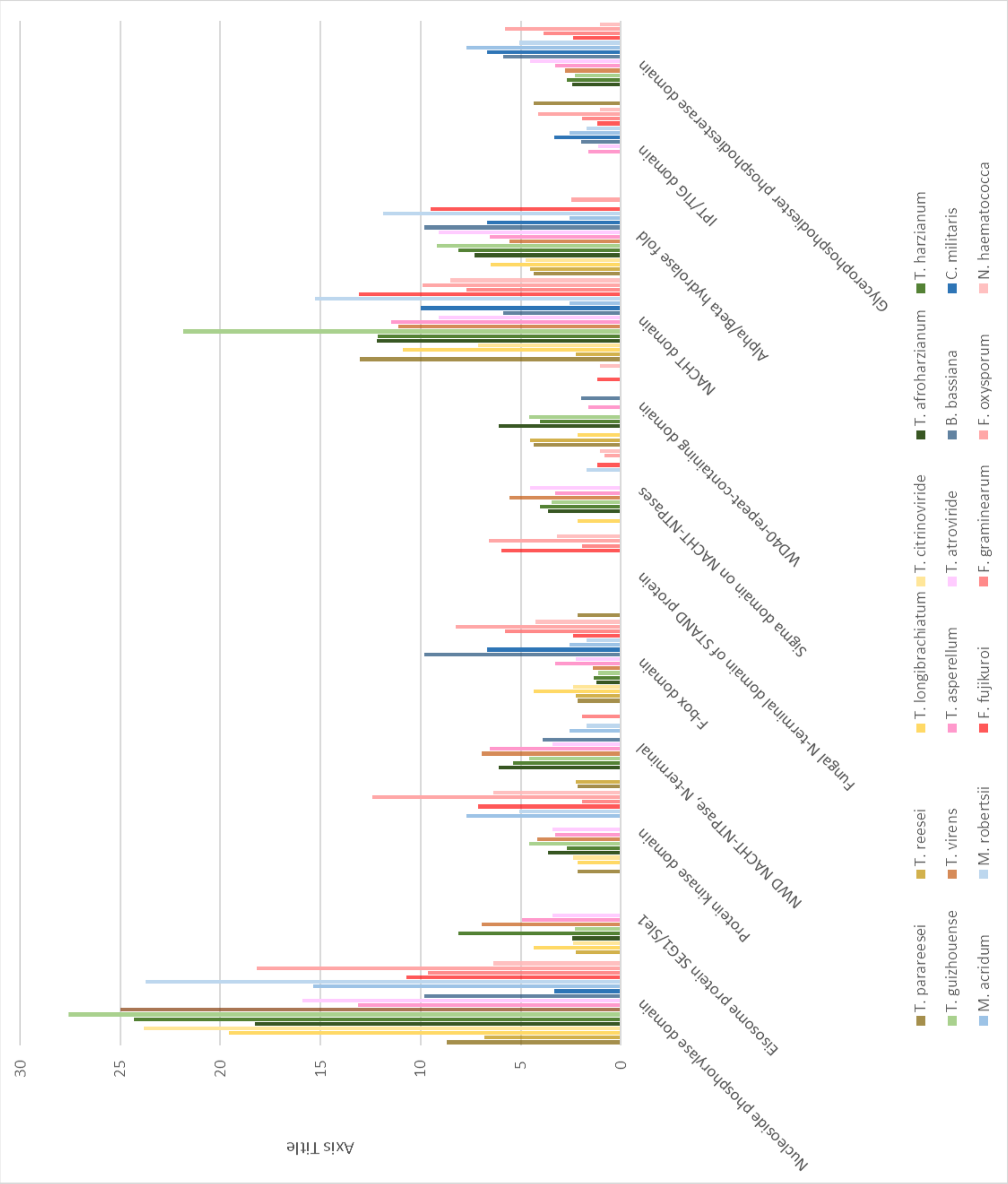


Figure 12: Share distribution of host-domains among hdc-ANKdc-proteins at the species level (1st part)

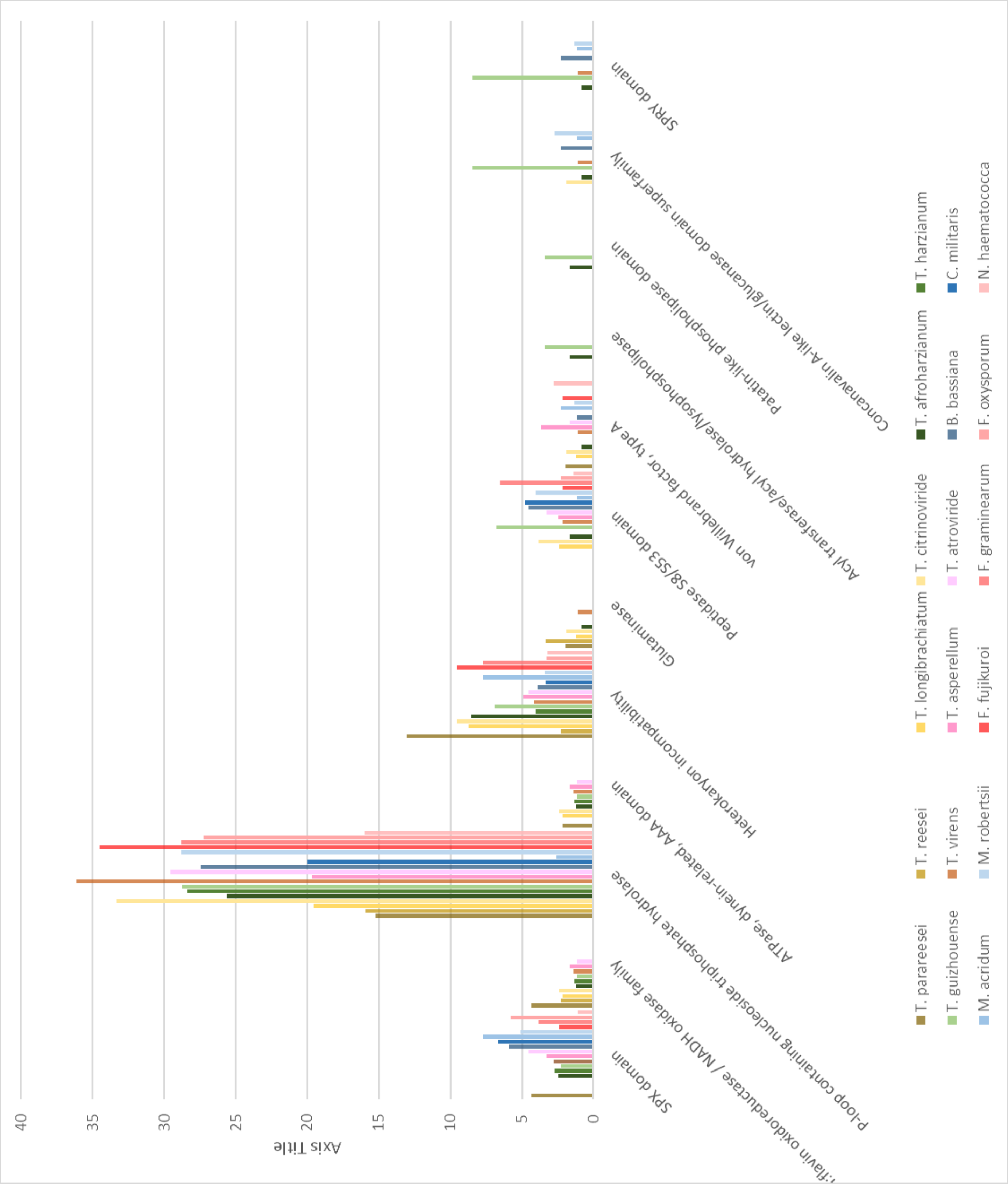


Figure 13: Share distribution of host-domains in hdc-ANKdc-proteins at the species level (2nd part)

Annotation of core orthogroups

11 out of 18 *Trichoderma* spp. core orthogroups consisted of loner-ANKdc-protein. Proteins belonging to these orthogroups contained exclusively ANK-domains. The other 7 orthogroups, which consisted of hdc-ANKdc-proteins, were also core orthogroups for the Hypocrealean ANKyrome. The annotations of ANKdc-proteins from these orthogroups and descriptions of their putative functions are shown in Table 7.

Orthogroup	Associated Domains	Function
16	VPS9	Vacuolar protein sorting- (fungal) transport of proteins from biosynthetic and endocytic pathways into the vacuole
	Phox homologous	Cell signaling, vesicular trafficking, protein sorting, lipid modification and protein-protein interaction
17	Allantoicase	Allantoate amidinohydrolase- catalyzes a step in uricolytic pathway: allantoate + H_2O = (S)-ureidoglycolate + urea
	Dilute	Associates myosins with different organelles, membrane vesicles mRNA; Myosins are molecular motors
20	Palmitoyltransferase, DHHC	Post-translational modification; transfers palmitoyl group from palmitoyl-CoA to the thiol group of Cys residues
26	BTB/POZ	Transcriptional regulators
28	PH-Pleckstrin	Recruiting proteins to different membranes
	Oxysterol-binding protein	signaling, vesicular trafficking, lipid metabolism, and non-vesicular sterol transport.
30	HTH, APSES-type DNA-binding	Transcription regulator, *N. crassa:* role in spore maturation. *A. nidulans*: transformation of undifferentiated hyphal elements into a complex multicellular structure *Candida albicans* (Yeast): enhanced filamentous growth protein
31	L-asparaginase N and C terminal	Catalyzes deamination of asparagine to aspartic acid

Table 7: Summarized InterProScan results for core hdc-ANKdc-proteins

4.1.4 Orthology Inference within the Hypocrealean ANKyrome

OrthoFinder analysis with total 2406 ANK-dc proteins resulted in 200 putative orthogroups that had to be filtered by visual inspection based on homology between MOTUs. After this step, 140 orthogroups containing 1053 proteins, as well as 1353 proteins as putative orphans (these proteins were not assigned to any orthogroup) were obtained. To verify their orphan status, a new OrthoFinder analysis was performed with only putative orphans from the previous step.

After visual inspection, further 43 orthogroups, containing 119 proteins, were identified. The main reason why further orthogroups were identified is the calculation of RBNHs (Reciprocal Best Length-Normalized hit) which led to less strict lower limit for acceptance of putative orthologs as explained in 3.3.

In conclusion, out of 2406 proteins 1172 were assigned to 183 separate orthogroups, while the remaining 1234 proteins were classified as orphan proteins.

A closer look into the number of orphan ANKdc-protein per species is shown in Figure 14 and the corresponding Supplementary material 4. The results reveal that the largest share of *Nectria haematococca* ANKyrome are orphans with 71% of all ANKdc-proteins. Within the *Trichoderma* genus, 61 % of *T. virens* ANKyrome are orphans, followed by *T. asperellum* ANKyrome which has 60.7 % orphans. In the ANKyromes from *T. parareesei* and *T afroharzianum* orphans account for 33.7 % and 36.7 %, respectively.

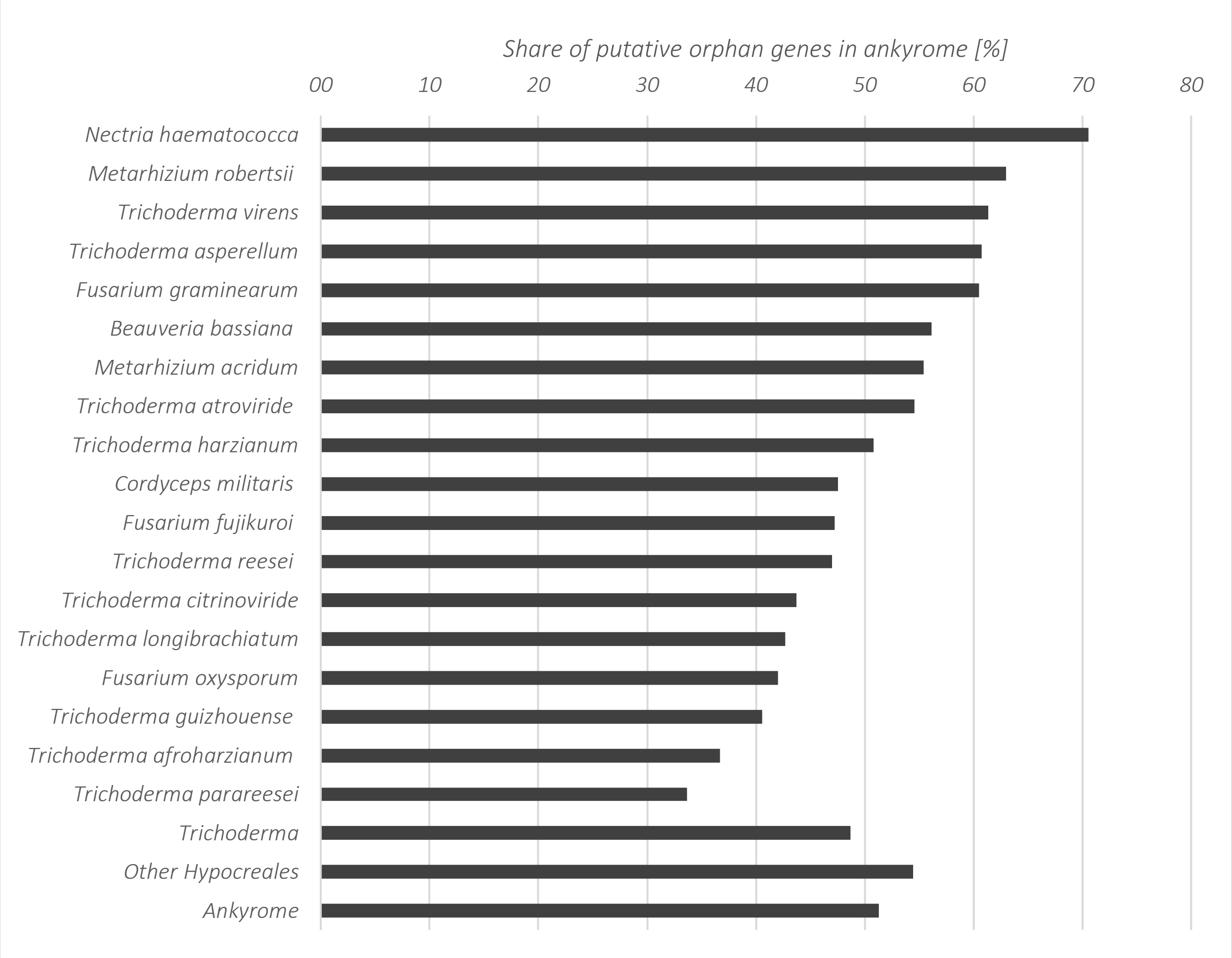


Figure 14: Share of orphan ANKdc-proteins per species

4.1.5 UpSet visualization of ANKyrome intersections

ANKdc-proteins shared between different taxonomical sections are shown in Figure 15. The connected dots represent which sections are being considered, while the corresponding column with the number on top corresponds to the total number of orthogroups in which respective species are present.

Out of 183 orthogroups only 18 orthogroups contained at least one representative from each of 10 considered *Trichoderma* species. This implies only 18 ANKdc-proteins are present in all of the 10 considered *Trichoderma* genomes. These 18 orthogroups represent the core ANKdc-proteins for *Trichoderma* spp.

Highest number of shared ANKdc-proteins were found in *Harzianum* section, to which *T. harzianum*, *T. afroharzianum* and *T. guizhouense* belong. Furthermore, only 11 core ANKdc-proteins among the 18 considered Hypocreales spp. were detected.

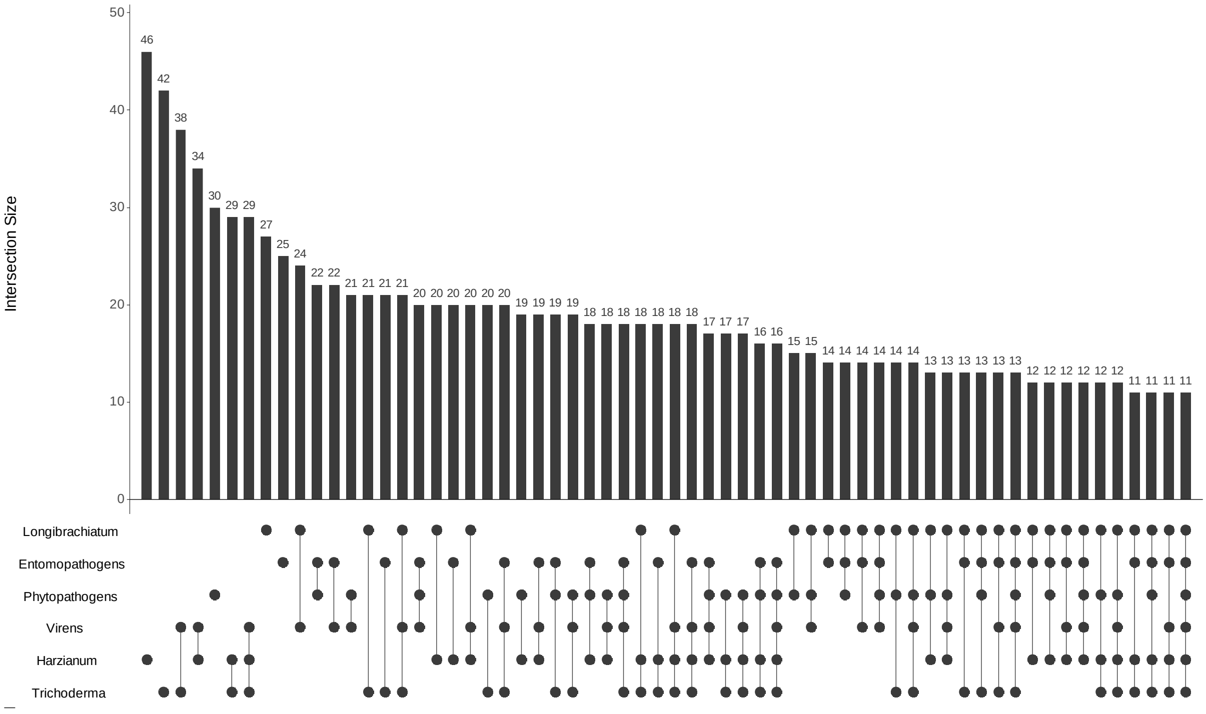


Figure 15: UpSet visualization of ANKyrome orthogroup intersections by taxonomical sections

***Phytopathogens** are plant pathogenic fungi* Nectria haematococca, Fusarium oxysporum, F. graminearum*, and* F. fujikuroi.
***Entomopathogens** are insect pathogenic fungi* Beauveria bassiana, Cordyceps militaris*,* Metarhizium acridum *and* M. robertsii.

4.2 Evolutionary Analysis

4.2.1 Phylogenetic Analysis

The best substitution model determined by SMS for each MSA as well as the N° of taxa in each MSA are listed in Supplementary material 5.

The inferred phylogenetic trees will be interpreted individually but there are some general characteristics which should be mentioned before. The topologies of trees inferred by Maximum likelihood and Bayesian method were similar and congruent with the multilocus gene genealogy of Hypocreales from Figure 2. The taxonomical sections *Longibrachiatum*, *Harzianum*, *Virens* and *Trichoderma* were consistently identifiable in each tree topology. The node support values in Maximum likelihood trees are shown explicitly, where well supported nodes had bootstrap values greater than 50%. The Bayesian inferred trees were manually processed by adding a grey circle to nodes that had a posterior probability ≥ 95 %. The nodes of *Trichoderma* spp*.* clades were predominantly well supported in both phylogenetic analyses. Only Bayesian trees are included since there was no significant difference in comparison to Maximum likelihood trees.

The trees were grouped for interpretation based on species they contained as shown in Table 8.

Class	Sordariomycetes										Leotiomycetes		Microbotryomycetes	
Subclass	Hypocreomycetidae			Sordariomycetidae						Xylariomycetida	-	-	-	-
Order	Hypocreales	Microascales	Glomerellales	Sordariales	Conio	Diaporthales	Ophiostomatales	Magnaporthales	Togniniales	Xylariales	Erysiphales	Helotiales	Leucosporidiales	Microbotryales
43	+	+	+	-	-	+	-	-	-	+	-	+	+	+
21	+	+	+	+	+	+	+	+	+	+	+	+	-	-
28	+	-	+	+	-	+	+	+	+	+	+	+	-	-
20	+	+	+	+	+	+	+	+	+	+	-	-	-	-
26	+	+	+	+	+	+	+	+	+	+	-	-	-	-
30	+	+	+	+	+	+	+	+	+	+	-	-	-	-
16	+	-	+	-	+	-	+	+	-	+	-	-	-	-
17	+	+	+	+	+	-	-	+	+	-	-	-	-	-
25	+	+	+	-	-	-	-	-	-	-	-	-	-	-
66	+	-	+	-	-	-	-	-	-	-	-	-	-	-
64	+	-	-	-	-	-	-	-	-	-	-	-	-	-
22	+	-	-	-	-	-	-	-	-	-	-	-	-	-
23	+	-	-	-	-	-	-	-	-	-	-	-	-	-
31	+	-	-	-	-	-	-	-	-	-	-	-	-	-
50	+	-	-	-	-	-	-	-	-	-	-	-	-	-
58	+	-	-	-	-	-	-	-	-	-	-	-	-	-
39	+	-	-	-	-	-	-	-	-	-	-	-	-	-
48	+	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 8: Summary of taxonomical range of MOTUs present in phylogenetic trees

Phylogenetic trees 39, 48, 58

The MSAs 39, 48 contained exclusively *Trichoderma* spp. homologs including proteins from *T. gamsii*. The MSA 58, in addition to all *Trichoderma* spp. proteins, had only one homologous protein outside *Trichoderma* genus that originated from *Oidiodendron maius* (Ascomycota, Leotiomycetes). *Oidiodendron* spp. are symbionts of roots of Ericaceae plants so they belong to a group called Ericoid mycorrhizal fungi. These species improve nutrient exchange and protect the plant from heavy metal toxicity. A comparative genomics study from 2018 found that Ericoid mycorrhizal fungi are more similar to saprotrophs and pathogens than to ectomycorrhyzal fungi based on their genetic machinery and secondary metabolism but also that they can be both saprotrophs and biotrophs (Wei *et al*. 2016; Martino *et al.* 2018). The corresponding trees are shown in Figure 16, Figure 17 and Figure 18.

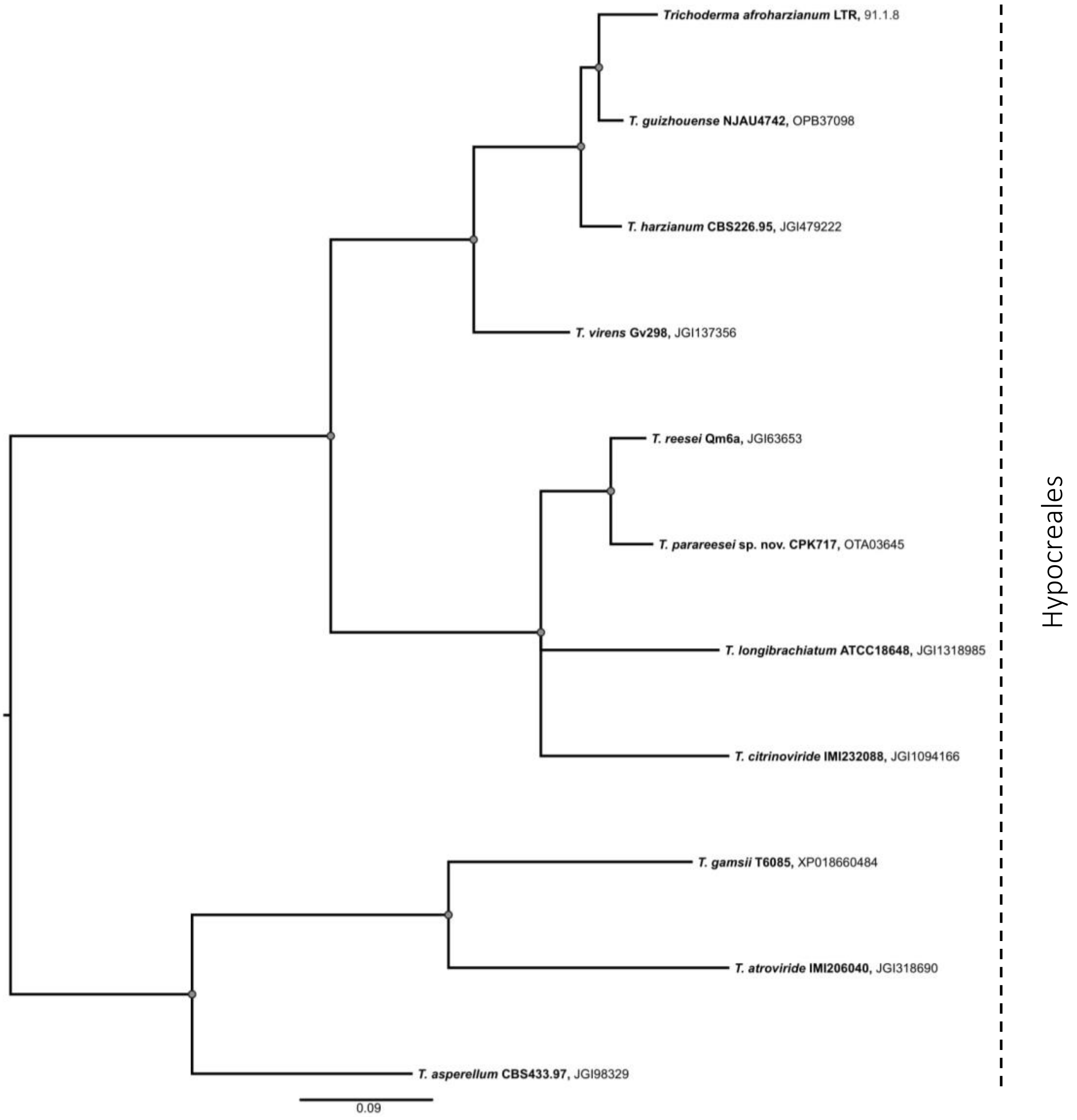

Trichoderma afroharzianum LTR, 91.1.8
T. guizhouense NJAU4742, OPB37098
T. harzianum CBS226.95, JGI479222
T. virens Gv298, JGI137356
T. reesei Qm6a, JGI63653
T. parareesei sp. nov. CPK717, OTA03645
T. longibrachiatum ATCC18648, JGI1318985
T. citrinoviride IMI232088, JGI1094166
T. gamsii T6085, XP018660484
T. atroviride IMI206040, JGI318690
T. asperellum CBS433.97, JGI98329
0.09
Hypocreales

Figure 16: Bayesian tree of OG 39

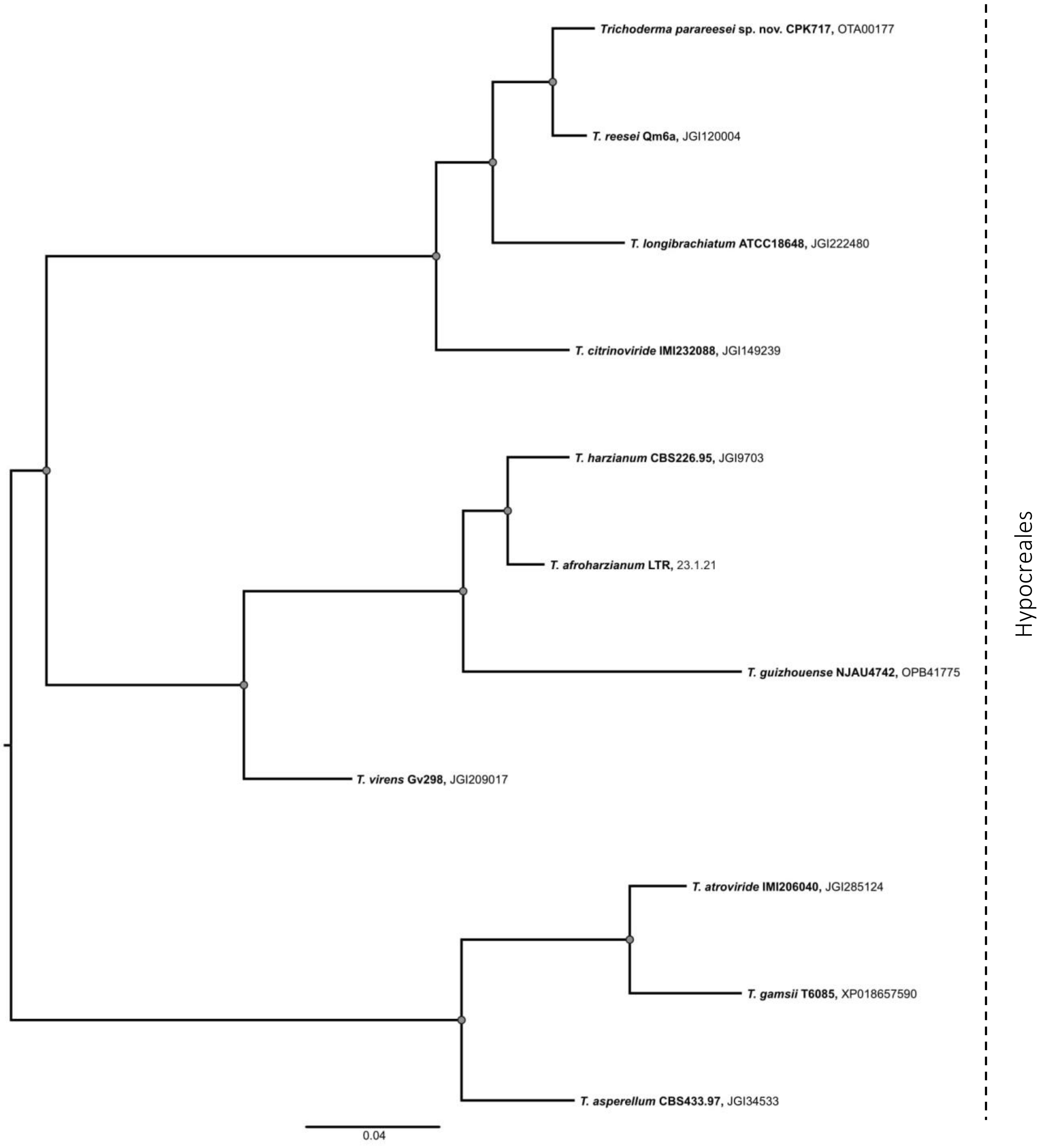

Trichoderma parareesei sp. nov. CPK717, OTA00177
T. reesei Qm6a, JGI120004
T. longibrachiatum ATCC18648, JGI222480
T. citrinoviride IMI232088, JGI149239
T. harzianum CBS226.95, JGI9703
T. afroharzianum LTR, 23.1.21
T. guizhouense NJAU4742, OPB41775
T. virens Gv298, JGI209017
T. atroviride IMI206040, JGI285124
T. gamsii T6085, XP018657590
T. asperellum CBS433.97, JGI34533
0.04
Hypocreales

Figure 17: Bayesian tree of OG 48

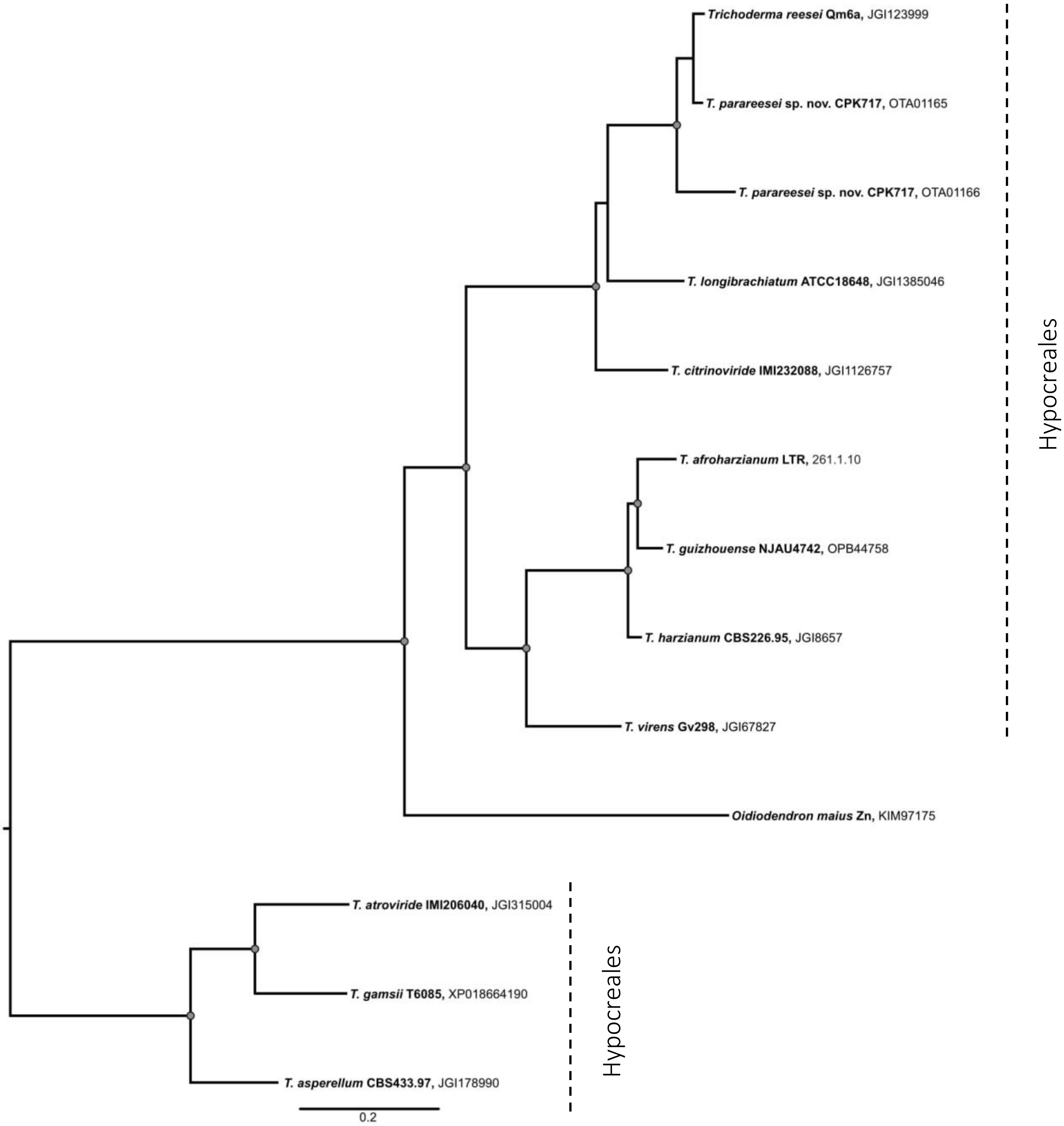

Trichoderma reesei Qm6a, JGI123999
T. parareesei sp. nov. CPK717, OTA01165
T. parareesei sp. nov. CPK717, OTA01166
T. longibrachiatum ATCC18648, JGI1385046
T. citrinoviride IMI232088, JGI1126757
T. afroharzianum LTR, 261.1.10
T. guizhouense NJAU4742, OPB44758
T. harzianum CBS226.95, JGI8657
T. virens Gv298, JGI67827
Oidiodendron maius Zn, KIM97175
T. atroviride IMI206040, JGI315004
T. gamsii T6085, XP018664190
T. asperellum CBS433.97, JGI178990
0.2
Hypocreales
Hypocreales

Figure 18: Bayesian tree of OG 58

Phylogenetic trees 22, 23, 31, 50

All of these trees contained exclusively homologs from species belonging to the Hypocreales order. The corresponding trees are shown in Figure 19, Figure 20, Figure 21 and Figure 22.

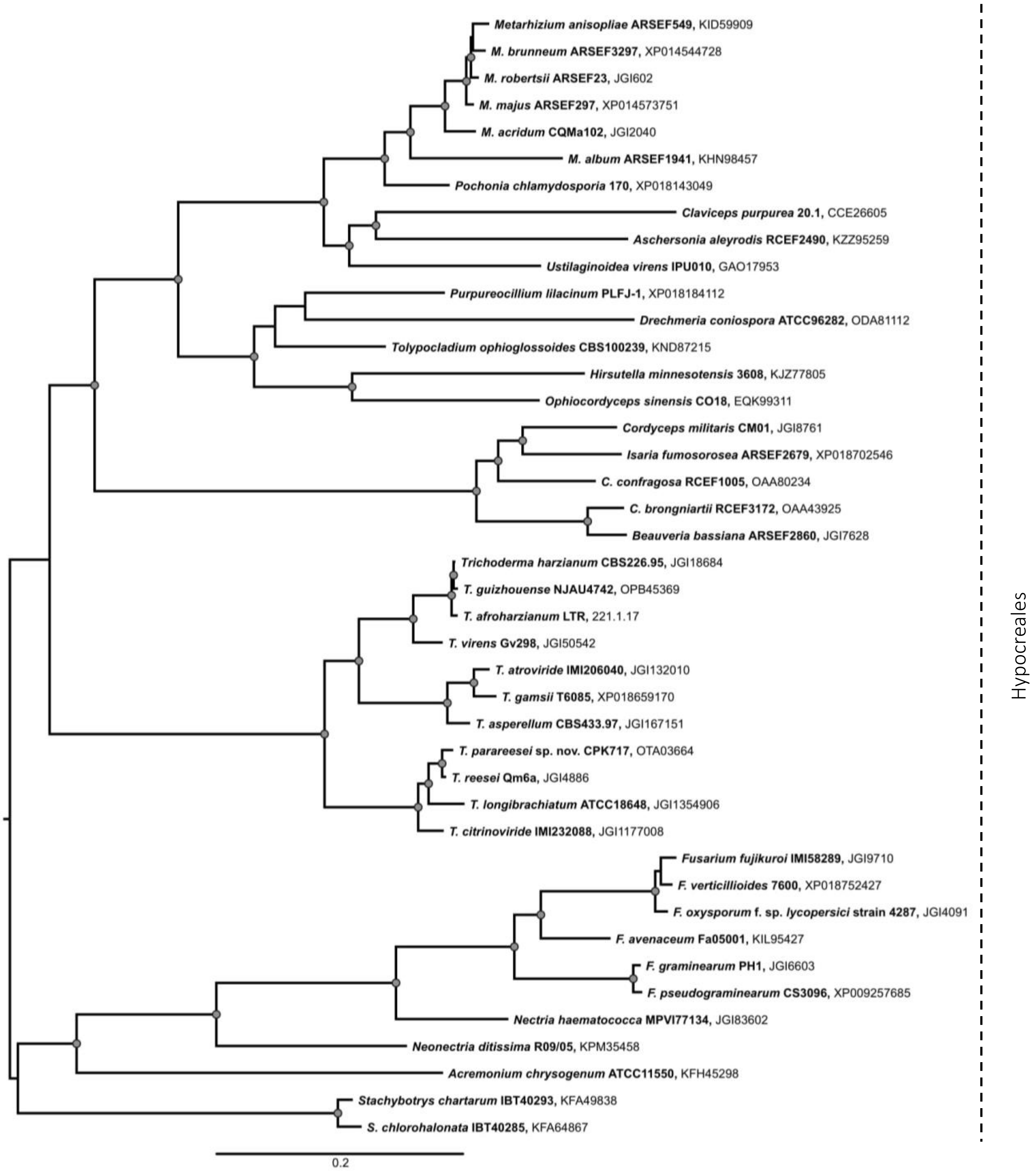

Metarhizium anisopliae ARSEF549, KID59909
M. brunneum ARSEF3297, XP014544728
M. robertsii ARSEF23, JGI602
M. majus ARSEF297, XP014573751
M. acridum CQMa102, JGI2040
M. album ARSEF1941, KHN98457
Pochonia chlamydosporia 170, XP018143049
Claviceps purpurea 20.1, CCE26605
Aschersonia aleyrodis RCEF2490, KZZ95259
Ustilaginoidea virens IPU010, GAO17953
Purpureocillium lilacinum PLFJ-1, XP018184112
Drechmeria coniospora ATCC96282, ODA81112
Tolypocladium ophioglossoides CBS100239, KND87215
Hirsutella minnesotensis 3608, KJZ77805
Ophiocordyceps sinensis CO18, EQK99311
Cordyceps militaris CM01, JGI8761
Isaria fumosorosea ARSEF2679, XP018702546
C. confragosa RCEF1005, OAA80234
C. brongniartii RCEF3172, OAA43925
Beauveria bassiana ARSEF2860, JGI7628
Trichoderma harzianum CBS226.95, JGI18684
T. guizhouense NJAU4742, OPB45369
T. afroharzianum LTR, 221.1.17
T. virens Gv298, JGI50542
T. atroviride IMI206040, JGI132010
T. gamsii T6085, XP018659170
T. asperellum CBS433.97, JGI167151
T. parareesei sp. nov. CPK717, OTA03664
T. reesei Qm6a, JGI4886
T. longibrachiatum ATCC18648, JGI1354906
T. citrinoviride IMI232088, JGI1177008
Fusarium fujikuroi IMI58289, JGI9710
F. verticillioides 7600, XP018752427
F. oxysporum f. sp. lycopersici strain 4287, JGI4091
F. avenaceum Fa05001, KIL95427
F. graminearum PH1, JGI6603
F. pseudograminearum CS3096, XP009257685
Nectria haematococca MPVI77134, JGI83602
Neonectria ditissima R09/05, KPM35458
Acremonium chrysogenum ATCC11550, KFH45298
Stachybotrys chartarum IBT40293, KFA49838
S. chlorohalonata IBT40285, KFA64867
0.2
Hypocreales

Figure 19: Bayesian tree of OG 22

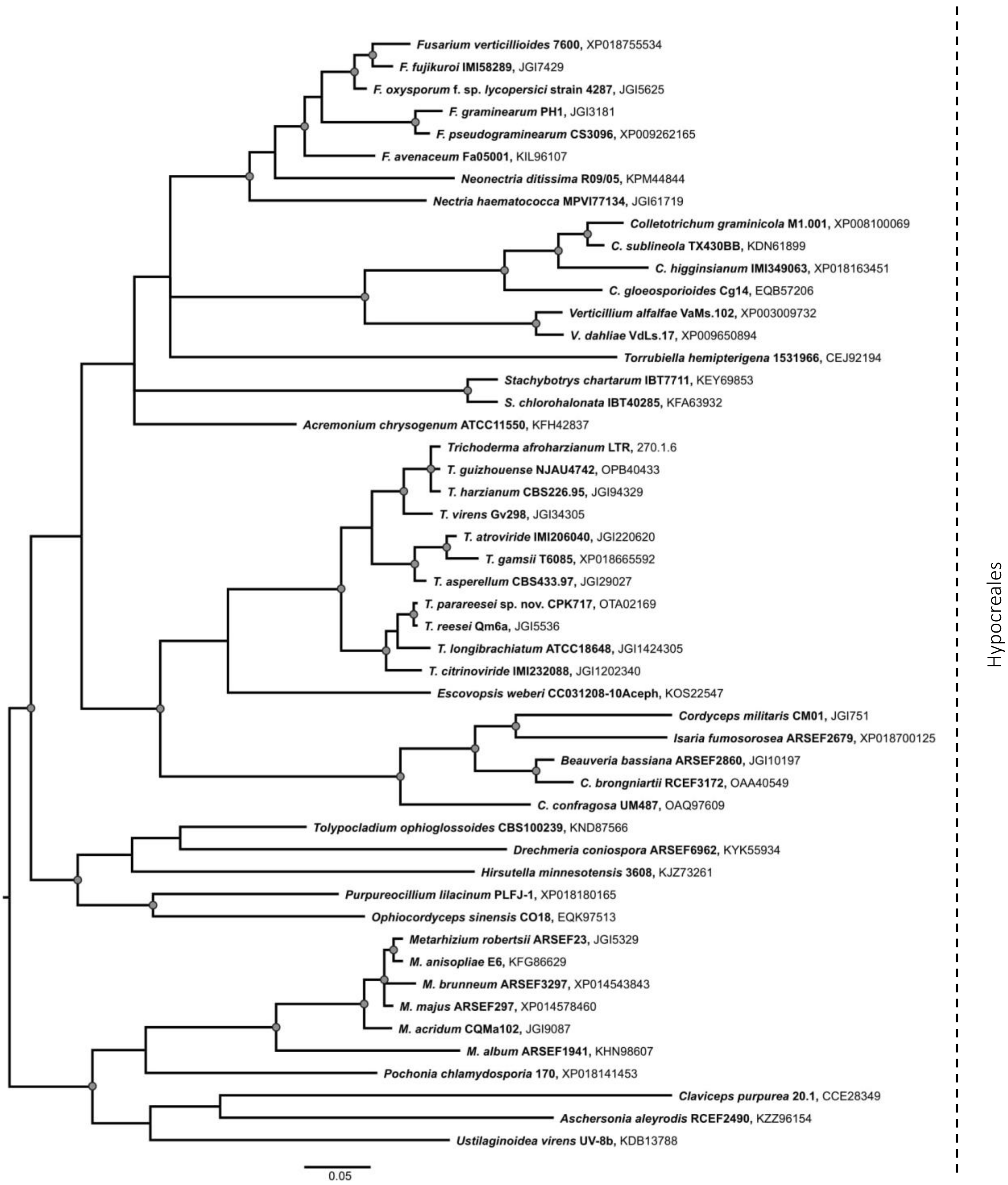

Fusarium verticillioides 7600, XP018755534
F. fujikuroi IMI58289, JGI7429
F. oxysporum f. sp. lycopersici strain 4287, JGI5625
F. graminearum PH1, JGI3181
F. pseudograminearum CS3096, XP009262165
F. avenaceum Fa05001, KIL96107
Neonectria ditissima R09/05, KPM44844
Nectria haematococca MPVI77134, JGI61719
Colletotrichum graminicola M1.001, XP008100069
C. sublineola TX430BB, KDN61899
C. higginsianum IMI349063, XP018163451
C. gloeosporioides Cg14, EQB57206
Verticillium alfalfae VaMs.102, XP003009732
V. dahliae VdLs.17, XP009650894
Torrubiella hemipterigena 1531966, CEJ92194
Stachybotrys chartarum IBT7711, KEY69853
S. chlorohalonata IBT40285, KFA63932
Acremonium chrysogenum ATCC11550, KFH42837
Trichoderma afroharzianum LTR, 270.1.6
T. guizhouense NJAU4742, OPB40433
T. harzianum CBS226.95, JGI94329
T. virens Gv298, JGI34305
T. atroviride IMI206040, JGI220620
T. gamsii T6085, XP018665592
T. asperellum CBS433.97, JGI29027
T. parareesei sp. nov. CPK717, OTA02169
T. reesei Qm6a, JGI5536
T. longibrachiatum ATCC18648, JGI1424305
T. citrinoviride IMI232088, JGI1202340
Escovopsis weberi CC031208-10Aceph, KOS22547
Cordyceps militaris CM01, JGI751
Isaria fumosorosea ARSEF2679, XP018700125
Beauveria bassiana ARSEF2860, JGI10197
C. brongniartii RCEF3172, OAA40549
C. confragosa UM487, OAQ97609
Tolypocladium ophioglossoides CBS100239, KND87566
Drechmeria coniospora ARSEF6962, KYK55934
Hirsutella minnesotensis 3608, KJZ73261
Purpureocillium lilacinum PLFJ-1, XP018180165
Ophiocordyceps sinensis CO18, EQK97513
Metarhizium robertsii ARSEF23, JGI5329
M. anisopliae E6, KFG86629
M. brunneum ARSEF3297, XP014543843
M. majus ARSEF297, XP014578460
M. acridum CQMa102, JGI9087
M. album ARSEF1941, KHN98607
Pochonia chlamydosporia 170, XP018141453
Claviceps purpurea 20.1, CCE28349
Aschersonia aleyrodis RCEF2490, KZZ96154
Ustilaginoidea virens UV-8b, KDB13788
0.05
Hypocreales

Figure 20: Bayesian tree of OG 23

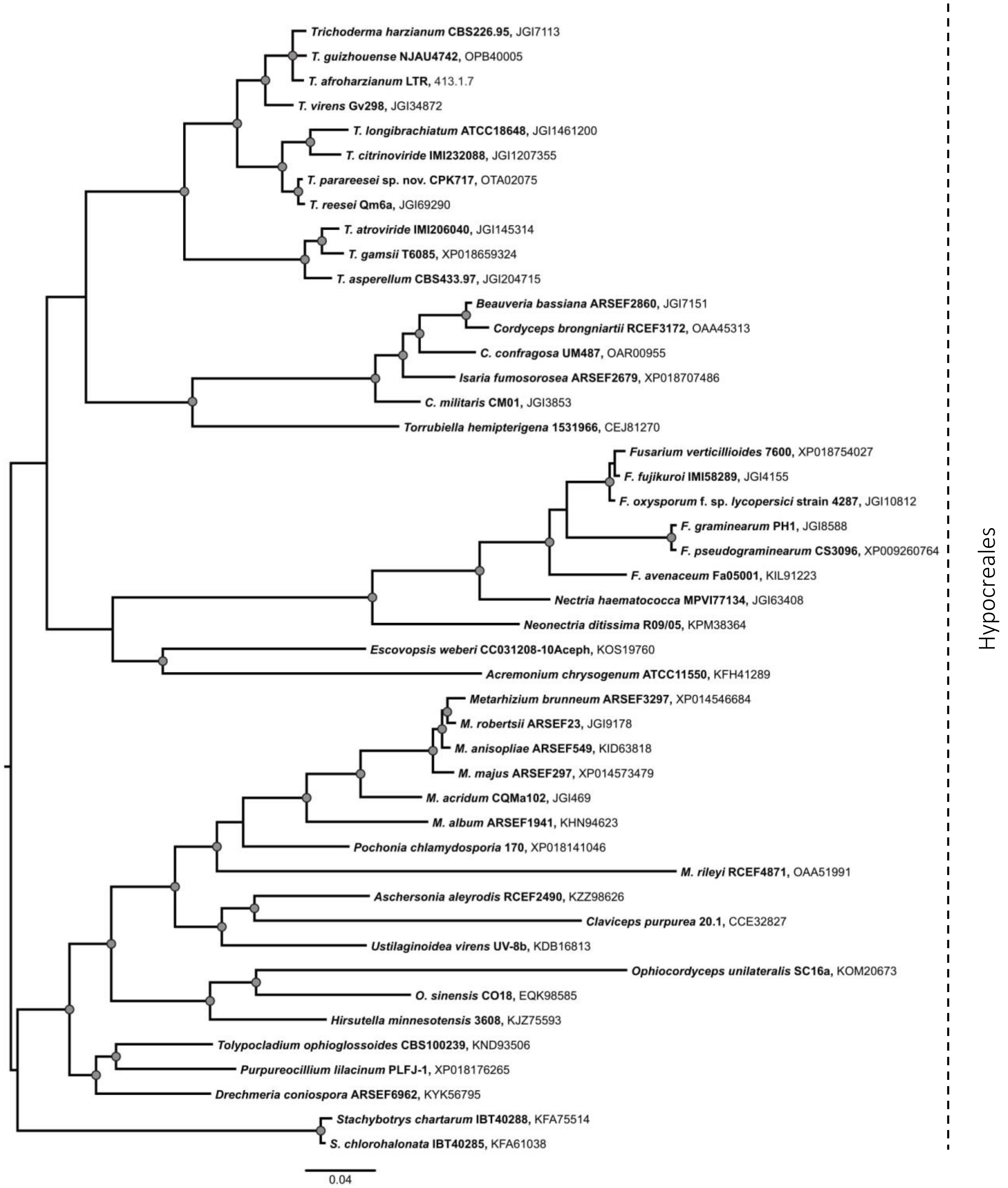

Trichoderma harzianum CBS226.95, JGI7113
T. guizhouense NJAU4742, OPB40005
T. afroharzianum LTR, 413.1.7
T. virens Gv298, JGI34872
T. longibrachiatum ATCC18648, JGI1461200
T. citrinoviride IMI232088, JGI1207355
T. parareesei sp. nov. CPK717, OTA02075
T. reesei Qm6a, JGI69290
T. atroviride IMI206040, JGI145314
T. gamsii T6085, XP018659324
T. asperellum CBS433.97, JGI204715
Beauveria bassiana ARSEF2860, JGI7151
Cordyceps brongniartii RCEF3172, OAA45313
C. confragosa UM487, OAR00955
Isaria fumosorosea ARSEF2679, XP018707486
C. militaris CM01, JGI3853
Torrubiella hemipterigena 1531966, CEJ81270
Fusarium verticillioides 7600, XP018754027
F. fujikuroi IMI58289, JGI4155
F. oxysporum f. sp. lycopersici strain 4287, JGI10812
F. graminearum PH1, JGI8588
F. pseudograminearum CS3096, XP009260764
F. avenaceum Fa05001, KIL91223
Nectria haematococca MPVI77134, JGI63408
Neonectria ditissima R09/05, KPM38364
Escovopsis weberi CC031208-10Aceph, KOS19760
Acremonium chrysogenum ATCC11550, KFH41289
Metarhizium brunneum ARSEF3297, XP014546684
M. robertsii ARSEF23, JGI9178
M. anisopliae ARSEF549, KID63818
M. majus ARSEF297, XP014573479
M. acridum CQMa102, JGI469
M. album ARSEF1941, KHN94623
Pochonia chlamydosporia 170, XP018141046
M. rileyi RCEF4871, OAA51991
Aschersonia aleyrodis RCEF2490, KZZ98626
Claviceps purpurea 20.1, CCE32827
Ustilaginoidea virens UV-8b, KDB16813
Ophiocordyceps unilateralis SC16a, KOM20673
O. sinensis CO18, EQK98585
Hirsutella minnesotensis 3608, KJZ75593
Tolypocladium ophioglossoides CBS100239, KND93506
Purpureocillium lilacinum PLFJ-1, XP018176265
Drechmeria coniospora ARSEF6962, KYK56795
Stachybotrys chartarum IBT40288, KFA75514
S. chlorohalonata IBT40285, KFA61038
0.04
Hypocreales

Figure 21: Bayesian tree of OG 31

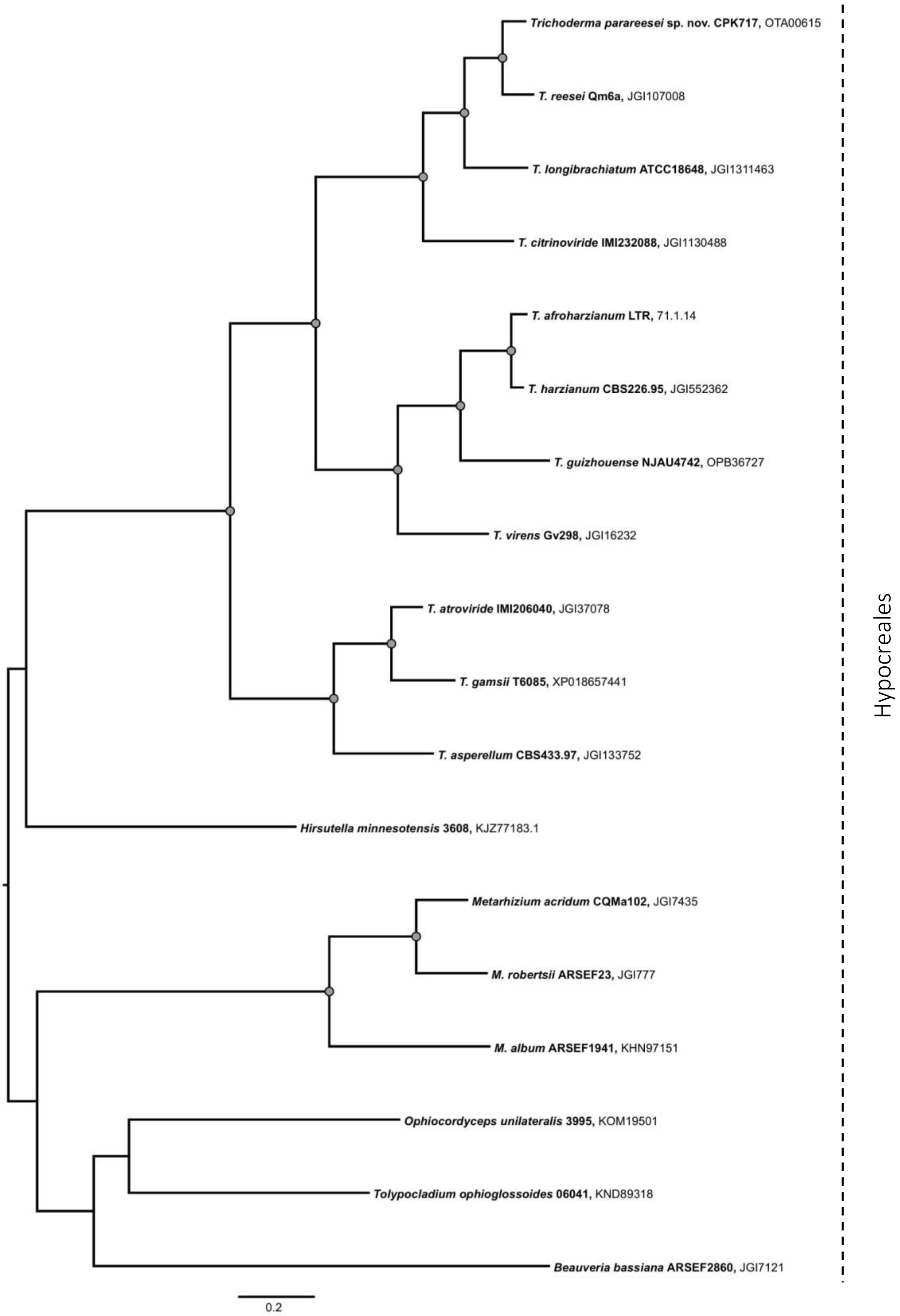

Trichoderma parareesei sp. nov. CPK717, OTA00615
T. reesei Qm6a, JGI107008
T. longibrachiatum ATCC18648, JGI1311463
T. citrinoviride IMI232088, JGI1130488
T. afroharzianum LTR, 71.1.14
T. harzianum CBS226.95, JGI552362
T. guizhouense NJAU4742, OPB36727
T. virens Gv298, JGI16232
T. atroviride IMI206040, JGI37078
T. gamsii T6085, XP018657441
T. asperellum CBS433.97, JGI133752
Hirsutella minnesotensis 3608, KJZ77183.1
Metarhizium acridum CQMa102, JGI7435
M. robertsii ARSEF23, JGI777
M. album ARSEF1941, KHN97151
Ophiocordyceps unilateralis 3995, KOM19501
Tolypocladium ophioglossoides 06041, KND89318
Beauveria bassiana ARSEF2860, JGI7121
Hypocreales
0.2

Figure 22: Bayesian tree of OG 50

Phylogenetic tree 64

The MSA 64 and its corresponding tree contained only proteins from *Trichoderma* spp. but no *T. gamsii* protein. The homologs from *T. gamsii* were found by homology search and were not included in the initial orthology inference because the genome was not available. The orthology inference can be repeated after retrieval of ANKyrome from published *T. gamsii* genome to eventually reject this orthogroup. The corresponding tree is shown in Figure 23.

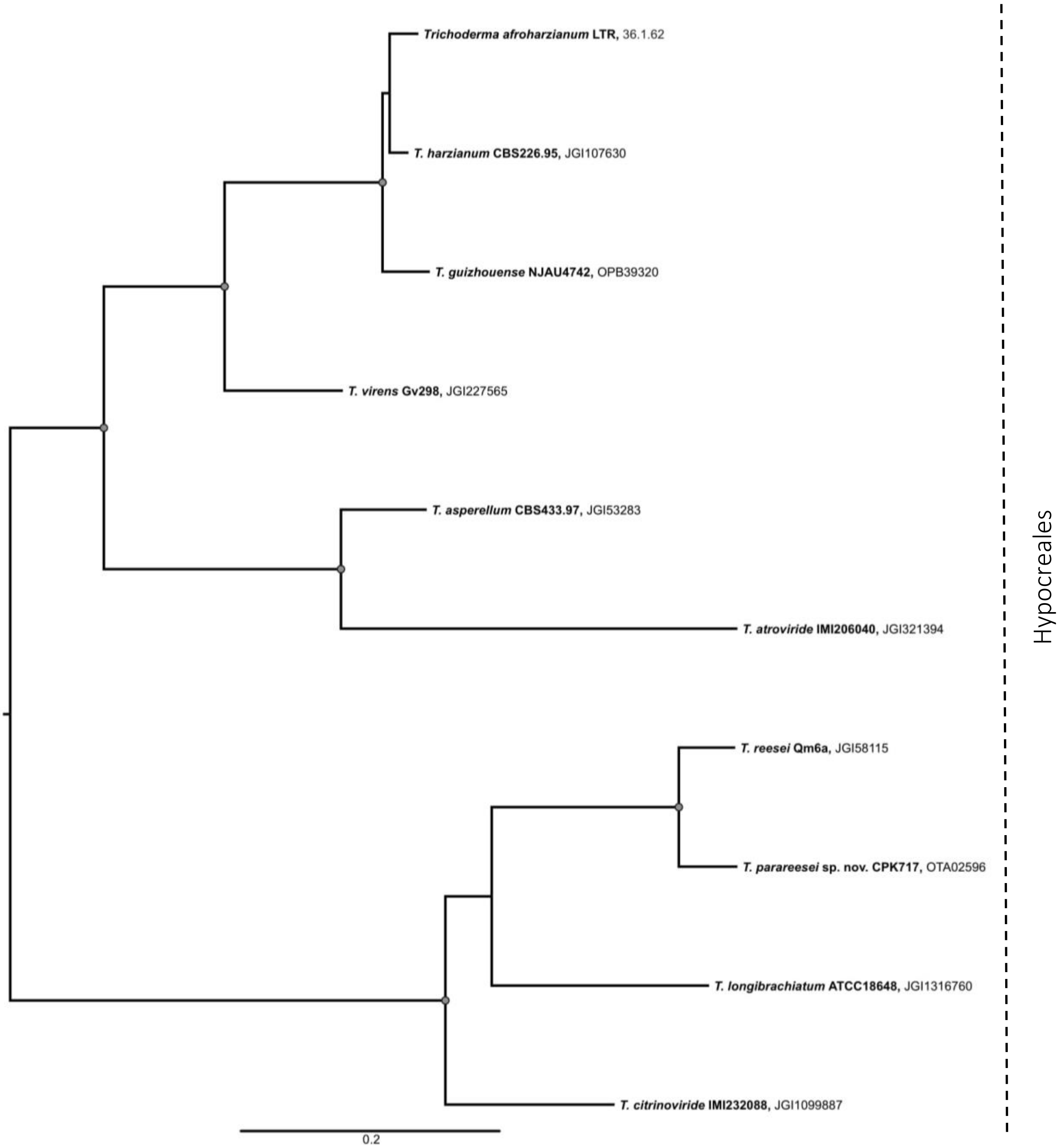

Trichoderma afroharzianum LTR, 36.1.62
T. harzianum CBS226.95, JGI107630
T. guizhouense NJAU4742, OPB39320
T. virens Gv298, JGI227565
T. asperellum CBS433.97, JGI53283
T. atroviride IMI206040, JGI321394
T. reesei Qm6a, JGI58115
T. parareesei sp. nov. CPK717, OTA02596
T. longibrachiatum ATCC18648, JGI1316760
T. citrinoviride IMI232088, JGI1099887
0.2
Hypocreales

Figure 23: Bayesian tree of OG 64

Phylogenetic trees 25, 66

The MSA 66 contained in addition to species from Hypocreales order homologs from Glomerellales order, while the phylogenetic tree 25 contained also species from Microascales order. The corresponding trees are shown in Figure 24 and Figure 25.

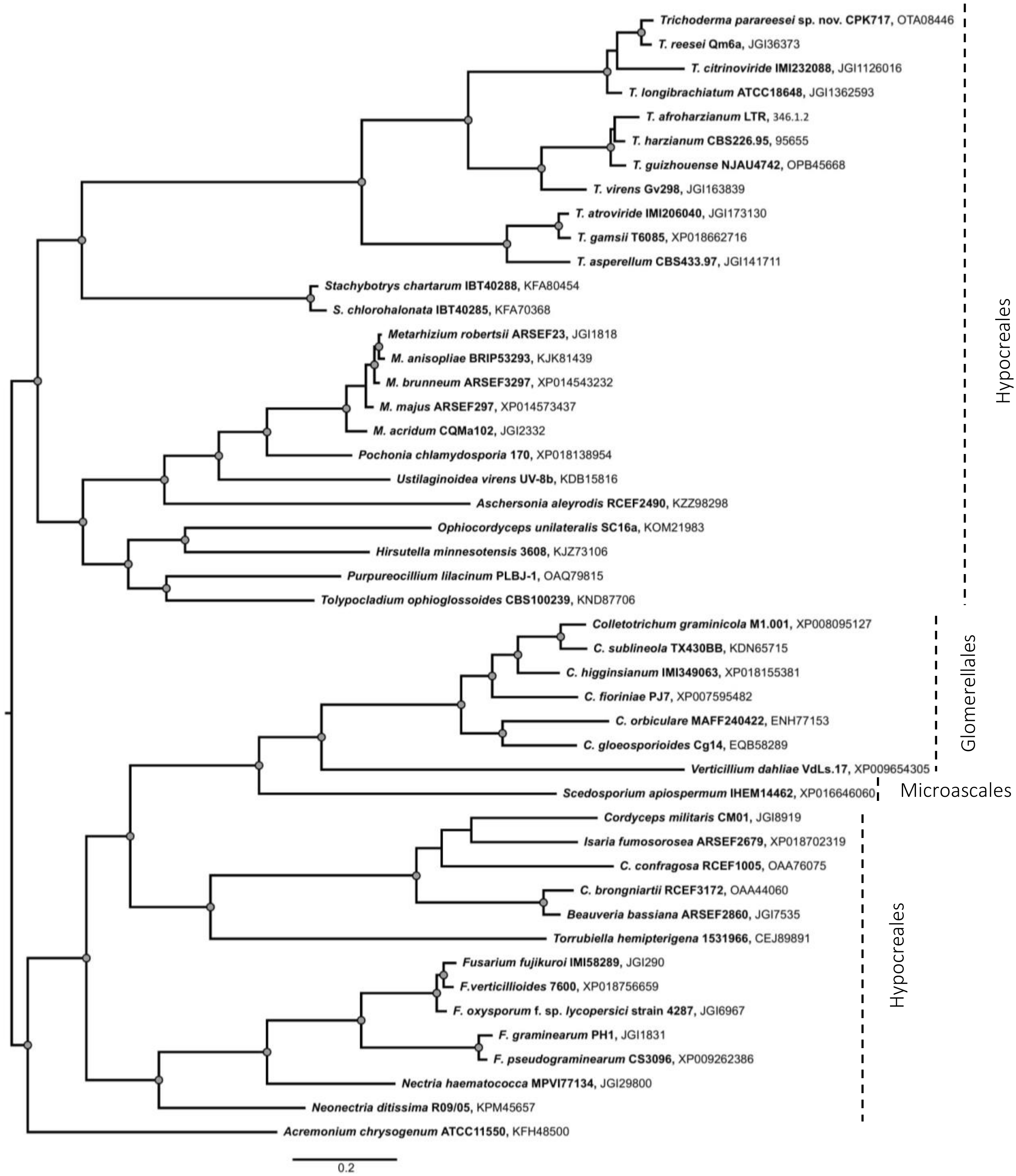

Trichoderma parareesei sp. nov. CPK717, OTA08446
T. reesei Qm6a, JGI36373
T. citrinoviride IMI232088, JGI1126016
T. longibrachiatum ATCC18648, JGI1362593
T. afroharzianum LTR, 346.1.2
T. harzianum CBS226.95, 95655
T. guizhouense NJAU4742, OPB45668
T. virens Gv298, JGI163839
T. atroviride IMI206040, JGI173130
T. gamsii T6085, XP018662716
T. asperellum CBS433.97, JGI141711
Stachybotrys chartarum IBT40288, KFA80454
S. chlorohalonata IBT40285, KFA70368
Metarhizium robertsii ARSEF23, JGI1818
M. anisopliae BRIP53293, KJK81439
M. brunneum ARSEF3297, XP014543232
M. majus ARSEF297, XP014573437
M. acridum CQMa102, JGI2332
Pochonia chlamydosporia 170, XP018138954
Ustilaginoidea virens UV-8b, KDB15816
Aschersonia aleyrodis RCEF2490, KZZ98298
Ophiocordyceps unilateralis SC16a, KOM21983
Hirsutella minnesotensis 3608, KJZ73106
Purpureocillium lilacinum PLBJ-1, OAQ79815
Tolypocladium ophioglossoides CBS100239, KND87706
Hypocreales
Colletotrichum graminicola M1.001, XP008095127
C. sublineola TX430BB, KDN65715
C. higginsianum IMI349063, XP018155381
C. fioriniae PJ7, XP007595482
C. orbiculare MAFF240422, ENH77153
C. gloeosporioides Cg14, EQB58289
Verticillium dahliae VdLs.17, XP009654305
Glomerellales
Scedosporium apiospermum IHEM14462, XP016646060
Microascales
Cordyceps militaris CM01, JGI8919
Isaria fumosorosea ARSEF2679, XP018702319
C. confragosa RCEF1005, OAA76075
C. brongniartii RCEF3172, OAA44060
Beauveria bassiana ARSEF2860, JGI7535
Torrubiella hemipterigena 1531966, CEJ89891
Fusarium fujikuroi IMI58289, JGI290
F.verticillioides 7600, XP018756659
F. oxysporum f. sp. *lycopersici* strain 4287, JGI6967
F. graminearum PH1, JGI1831
F. pseudograminearum CS3096, XP009262386
Nectria haematococca MPVI77134, JGI29800
Neonectria ditissima R09/05, KPM45657
Acremonium chrysogenum ATCC11550, KFH48500
Hypocreales
0.2

Figure 24: Bayesian tree of OG 25

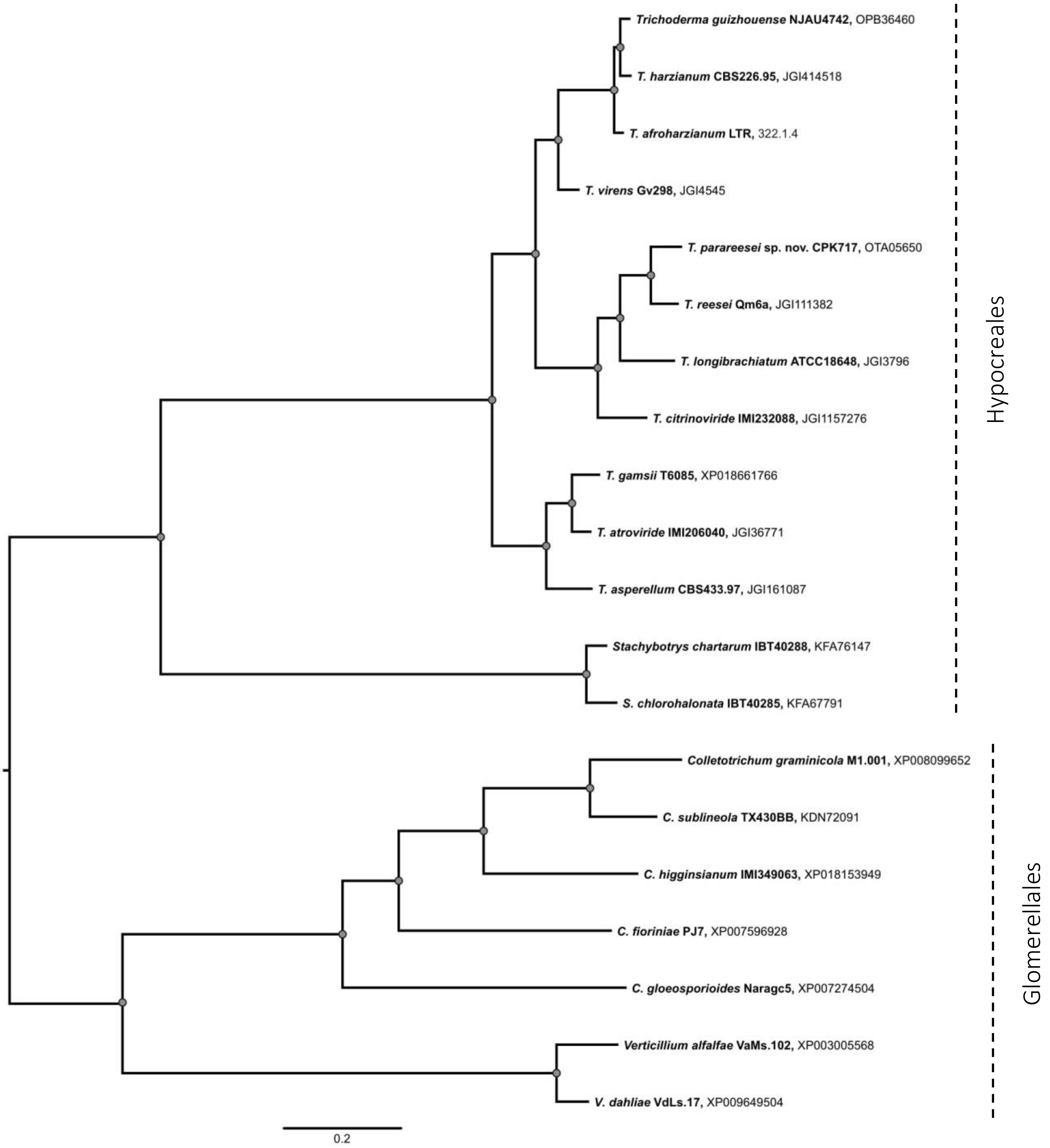

Trichoderma guizhouense **NJAU4742,** OPB36460
T. harzianum **CBS226.95,** JGI414518
T. afroharzianum **LTR,** 322.1.4
T. virens **Gv298,** JGI4545
T. parareesei **sp. nov. CPK717,** OTA05650
T. reesei **Qm6a,** JGI111382
T. longibrachiatum **ATCC18648,** JGI3796
T. citrinoviride **IMI232088,** JGI1157276
T. gamsii **T6085,** XP018661766
T. atroviride **IMI206040,** JGI36771
T. asperellum **CBS433.97,** JGI161087
Stachybotrys chartarum **IBT40288,** KFA76147
S. chlorohalonata **IBT40285,** KFA67791
Colletotrichum graminicola **M1.001,** XP008099652
C. sublineola **TX430BB,** KDN72091
C. higginsianum **IMI349063,** XP018153949
C. fioriniae **PJ7,** XP007596928
C. gloeosporioides **Naragc5,** XP007274504
Verticillium alfalfae **VaMs.102,** XP003005568
V. dahliae **VdLs.17,** XP009649504
0.2
Hypocreales
Glomerellales

Figure 25: Bayesian tree of OG 66

Phylogenetic trees 20, 26 and 30

In these phylogenetic trees species from both Hypocreomycetidae and Sordariomycetidae subclasses are found. Also, in each of the trees there are species from Xylariales order. The corresponding trees are shown in Figure 26, Figure 27 and Figure 28.

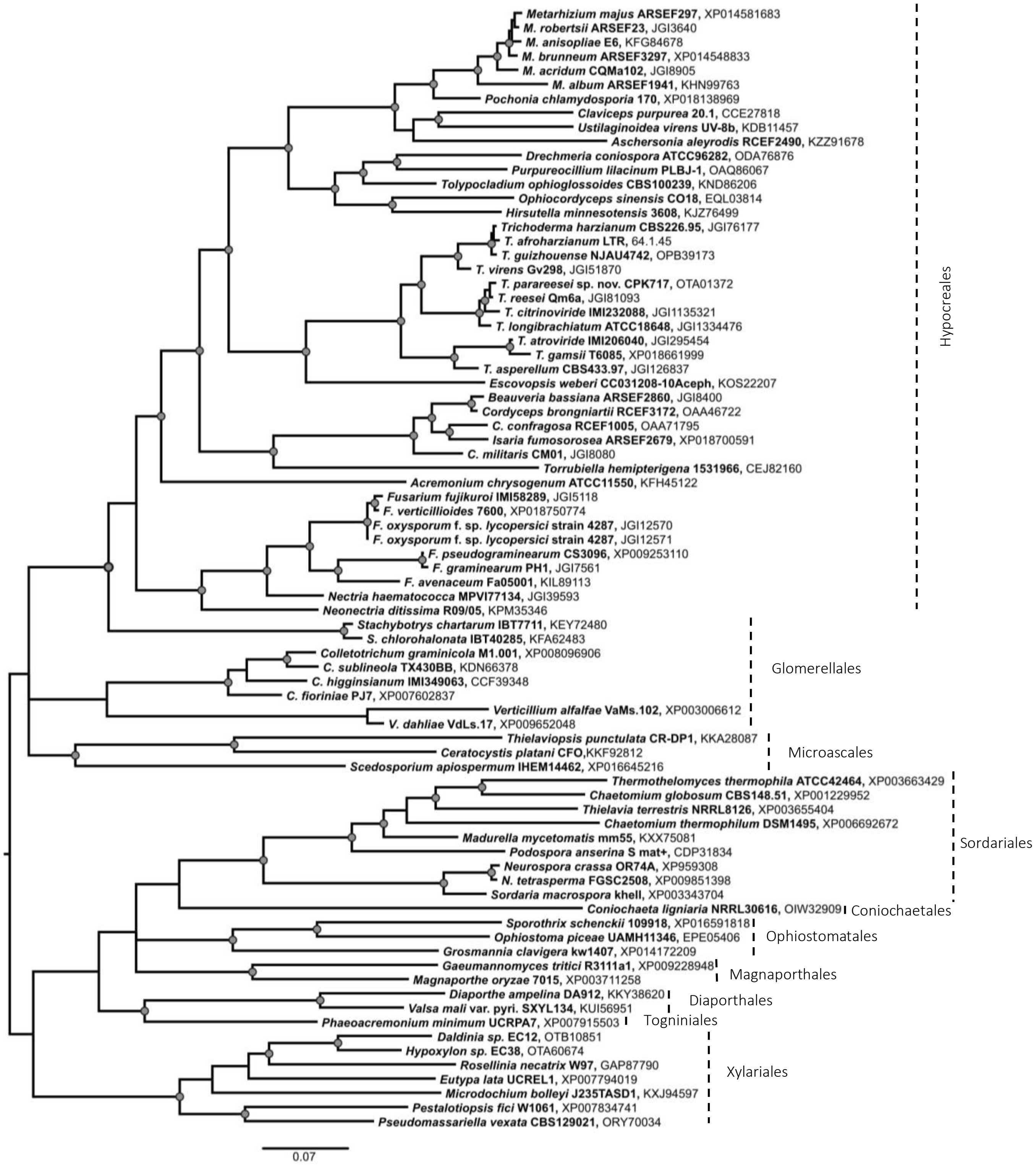

Metarhizium majus ARSEF297, XP014581683
M. robertsii ARSEF23, JGI3640
M. anisopliae E6, KFG84678
M. brunneum ARSEF3297, XP014548833
M. acridum CQMa102, JGI8905
M. album ARSEF1941, KHN99763
Pochonia chlamydosporia 170, XP018138969
Claviceps purpurea 20.1, CCE27818
Ustilaginoidea virens UV-8b, KDB11457
Aschersonia aleyrodis RCEF2490, KZZ91678
Drechmeria coniospora ATCC96282, ODA76876
Purpureocillium lilacinum PLBJ-1, OAQ86067
Tolypocladium ophioglossoides CBS100239, KND86206
Ophiocordyceps sinensis CO18, EQL03814
Hirsutella minnesotensis 3608, KJZ76499
Trichoderma harzianum CBS226.95, JGI76177
T. afroharzianum LTR, 64.1.45
T. guizhouense NJAU4742, OPB39173
T. virens Gv298, JGI51870
T. parareesei sp. nov. CPK717, OTA01372
T. reesei Qm6a, JGI81093
T. citrinoviride IMI232088, JGI1135321
T. longibrachiatum ATCC18648, JGI1334476
T. atroviride IMI206040, JGI295454
T. gamsii T6085, XP018661999
T. asperellum CBS433.97, JGI126837
Escovopsis weberi CC031208-10Aceph, KOS22207
Beauveria bassiana ARSEF2860, JGI8400
Cordyceps brongniartii RCEF3172, OAA46722
C. confragosa RCEF1005, OAA71795
Isaria fumosorosea ARSEF2679, XP018700591
C. militaris CM01, JGI8080
Torrubiella hemipterigena 1531966, CEJ82160
Acremonium chrysogenum ATCC11550, KFH45122
Fusarium fujikuroi IMI58289, JGI5118
F. verticillioides 7600, XP018750774
F. oxysporum f. sp. lycopersici strain 4287, JGI12570
F. oxysporum f. sp. lycopersici strain 4287, JGI12571
F. pseudograminearum CS3096, XP009253110
F. graminearum PH1, JGI7561
F. avenaceum Fa05001, KIL89113
Nectria haematococca MPVI77134, JGI39593
Neonectria ditissima R09/05, KPM35346
Hypocreales
Stachybotrys chartarum IBT7711, KEY72480
S. chlorohalonata IBT40285, KFA62483
Colletotrichum graminicola M1.001, XP008096906
C. sublineola TX430BB, KDN66378
C. higginsianum IMI349063, CCF39348
C. fioriniae PJ7, XP007602837
Verticillium alfalfae VaMs.102, XP003006612
V. dahliae VdLs.17, XP009652048
Glomerellales
Thielaviopsis punctulata CR-DP1, KKA28087
Ceratocystis platani CFO,KKF92812
Scedosporium apiospermum IHEM14462, XP016645216
Microascales
Thermothelomyces thermophila ATCC42464, XP003663429
Chaetomium globosum CBS148.51, XP001229952
Thielavia terrestris NRRL8126, XP003655404
Chaetomium thermophilum DSM1495, XP006692672
Madurella mycetomatis mm55, KXX75081
Podospora anserina S mat+, CDP31834
Neurospora crassa OR74A, XP959308
N. tetrasperma FGSC2508, XP009851398
Sordaria macrospora khell, XP003343704
Sordariales
Coniochaeta ligniaria NRRL30616, OIW32909
Coniochaetales
Sporothrix schenckii 109918, XP016591818
Ophiostoma piceae UAMH11346, EPE05406
Grosmannia clavigera kw1407, XP014172209
Ophiostomatales
Gaeumannomyces tritici R3111a1, XP009228948
Magnaporthe oryzae 7015, XP003711258
Magnaporthales
Diaporthe ampelina DA912, KKY38620
Valsa mali var. pyri. SXYL134, KUI56951
Diaporthales
Phaeoacremonium minimum UCRPA7, XP007915503
Togniniales
Daldinia sp. EC12, OTB10851
Hypoxylon sp. EC38, OTA60674
Rosellinia necatrix W97, GAP87790
Eutypa lata UCREL1, XP007794019
Microdochium bolleyi J235TASD1, KXJ94597
Pestalotiopsis fici W1061, XP007834741
Pseudomassariella vexata CBS129021, ORY70034
Xylariales
0.07

Figure 26: Bayesian tree of OG 20

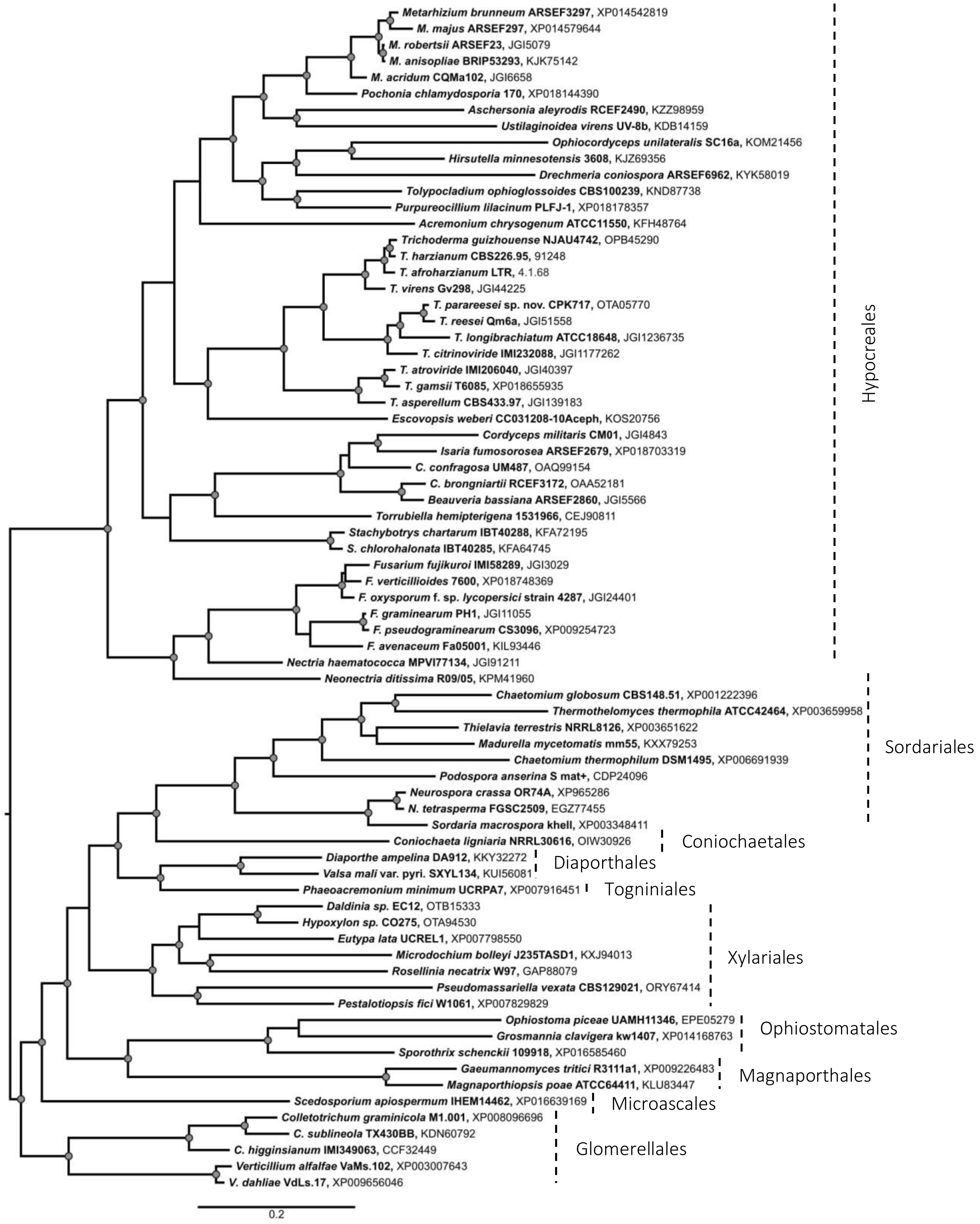

Metarhizium brunneum ARSEF3297, XP014542819
M. majus ARSEF297, XP014579644
M. robertsii ARSEF23, JGI5079
M. anisopliae BRIP53293, KJK75142
M. acridum CQMa102, JGI6658
Pochonia chlamydosporia 170, XP018144390
Aschersonia aleyrodis RCEF2490, KZZ98959
Ustilaginoidea virens UV-8b, KDB14159
Ophiocordyceps unilateralis SC16a, KOM21456
Hirsutella minnesotensis 3608, KJZ69356
Drechmeria coniospora ARSEF6962, KYK58019
Tolypocladium ophioglossoides CBS100239, KND87738
Purpureocillium lilacinum PLFJ-1, XP018178357
Acremonium chrysogenum ATCC11550, KFH48764
Trichoderma guizhouense NJAU4742, OPB45290
T. harzianum CBS226.95, 91248
T. afroharzianum LTR, 4.1.68
T. virens Gv298, JGI44225
T. parareesei sp. nov. CPK717, OTA05770
T. reesei Qm6a, JGI51558
T. longibrachiatum ATCC18648, JGI1236735
T. citrinoviride IMI232088, JGI1177262
T. atroviride IMI206040, JGI40397
T. gamsii T6085, XP018655935
T. asperellum CBS433.97, JGI139183
Escovopsis weberi CC031208-10Aceph, KOS20756
Cordyceps militaris CM01, JGI4843
Isaria fumosorosea ARSEF2679, XP018703319
C. confragosa UM487, OAQ99154
C. brongniartii RCEF3172, OAA52181
Beauveria bassiana ARSEF2860, JGI5566
Torrubiella hemipterigena 1531966, CEJ90811
Stachybotrys chartarum IBT40288, KFA72195
S. chlorohalonata IBT40285, KFA64745
Fusarium fujikuroi IMI58289, JGI3029
F. verticillioides 7600, XP018748369
F. oxysporum f. sp. lycopersici strain 4287, JGI24401
F. graminearum PH1, JGI11055
F. pseudograminearum CS3096, XP009254723
F. avenaceum Fa05001, KIL93446
Nectria haematococca MPVI77134, JGI91211
Neonectria ditissima R09/05, KPM41960
Hypocreales
Chaetomium globosum CBS148.51, XP001222396
Thermothelomyces thermophila ATCC42464, XP003659958
Thielavia terrestris NRRL8126, XP003651622
Madurella mycetomatis mm55, KXX79253
Chaetomium thermophilum DSM1495, XP006691939
Podospora anserina S mat+, CDP24096
Neurospora crassa OR74A, XP965286
N. tetrasperma FGSC2509, EGZ77455
Sordaria macrospora khell, XP003348411
Sordariales
Coniochaeta ligniaria NRRL30616, OIW30926
Coniochaetales
Diaporthe ampelina DA912, KKY32272
Valsa mali var. pyri. SXYL134, KUI56081
Diaporthales
Phaeoacremonium minimum UCRPA7, XP007916451
Togniniales
Daldinia sp. EC12, OTB15333
Hypoxylon sp. CO275, OTA94530
Eutypa lata UCREL1, XP007798550
Microdochium bolleyi J235TASD1, KXJ94013
Rosellinia necatrix W97, GAP88079
Pseudomassariella vexata CBS129021, ORY67414
Pestalotiopsis fici W1061, XP007829829
Xylariales
Ophiostoma piceae UAMH11346, EPE05279
Grosmannia clavigera kw1407, XP014168763
Sporothrix schenckii 109918, XP016585460
Ophiostomatales
Gaeumannomyces tritici R3111a1, XP009226483
Magnaporthiopsis poae ATCC64411, KLU83447
Magnaporthales
Scedosporium apiospermum IHEM14462, XP016639169
Microascales
Colletotrichum graminicola M1.001, XP008096696
C. sublineola TX430BB, KDN60792
C. higginsianum IMI349063, CCF32449
Verticillium alfalfae VaMs.102, XP003007643
V. dahliae VdLs.17, XP009656046
Glomerellales
0.2

Figure 27: Bayesian tree of OG 26

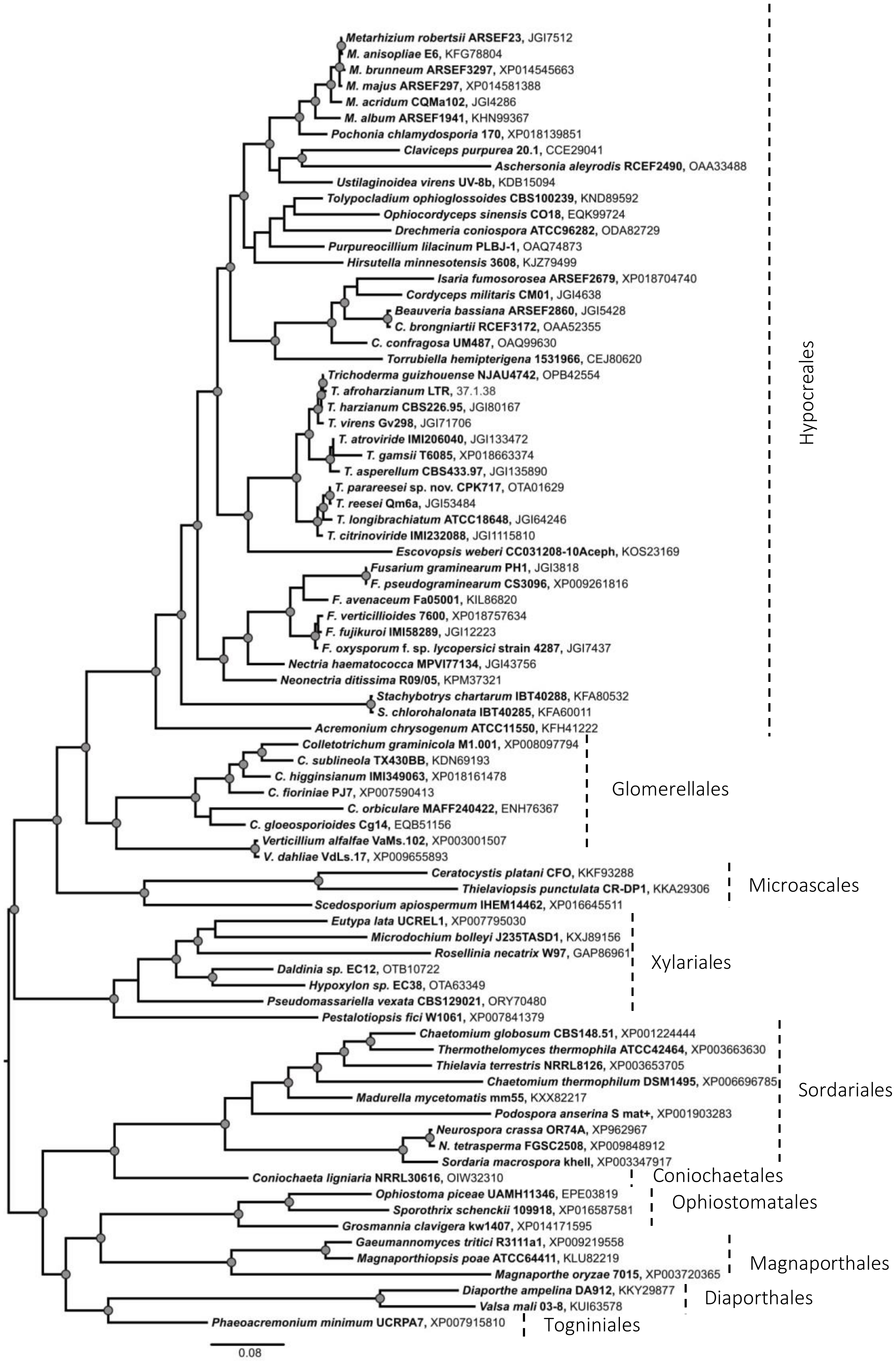

Metarhizium robertsii ARSEF23, JGI7512
M. anisopliae E6, KFG78804
M. brunneum ARSEF3297, XP014545663
M. majus ARSEF297, XP014581388
M. acridum CQMa102, JGI4286
M. album ARSEF1941, KHN99367
Pochonia chlamydosporia 170, XP018139851
Claviceps purpurea 20.1, CCE29041
Aschersonia aleyrodis RCEF2490, OAA33488
Ustilaginoidea virens UV-8b, KDB15094
Tolypocladium ophioglossoides CBS100239, KND89592
Ophiocordyceps sinensis CO18, EQK99724
Drechmeria coniospora ATCC96282, ODA82729
Purpureocillium lilacinum PLBJ-1, OAQ74873
Hirsutella minnesotensis 3608, KJZ79499
Isaria fumosorosea ARSEF2679, XP018704740
Cordyceps militaris CM01, JGI4638
Beauveria bassiana ARSEF2860, JGI5428
C. brongniartii RCEF3172, OAA52355
C. confragosa UM487, OAQ99630
Torrubiella hemipterigena 1531966, CEJ80620
Trichoderma guizhouense NJAU4742, OPB42554
T. afroharzianum LTR, 37.1.38
T. harzianum CBS226.95, JGI80167
T. virens Gv298, JGI71706
T. atroviride IMI206040, JGI133472
T. gamsii T6085, XP018663374
T. asperellum CBS433.97, JGI135890
T. parareesei sp. nov. CPK717, OTA01629
T. reesei Qm6a, JGI53484
T. longibrachiatum ATCC18648, JGI64246
T. citrinoviride IMI232088, JGI1115810
Escovopsis weberi CC031208-10Aceph, KOS23169
Fusarium graminearum PH1, JGI3818
F. pseudograminearum CS3096, XP009261816
F. avenaceum Fa05001, KIL86820
F. verticillioides 7600, XP018757634
F. fujikuroi IMI58289, JGI12223
F. oxysporum f. sp. lycopersici strain 4287, JGI7437
Nectria haematococca MPVI77134, JGI43756
Neonectria ditissima R09/05, KPM37321
Stachybotrys chartarum IBT40288, KFA80532
S. chlorohalonata IBT40285, KFA60011
Acremonium chrysogenum ATCC11550, KFH41222
Hypocreales
Colletotrichum graminicola M1.001, XP008097794
C. sublineola TX430BB, KDN69193
C. higginsianum IMI349063, XP018161478
C. fioriniae PJ7, XP007590413
C. orbiculare MAFF240422, ENH76367
C. gloeosporioides Cg14, EQB51156
Verticillium alfalfae VaMs.102, XP003001507
V. dahliae VdLs.17, XP009655893
Glomerellales
Ceratocystis platani CFO, KKF93288
Thielaviopsis punctulata CR-DP1, KKA29306
Scedosporium apiospermum IHEM14462, XP016645511
Microascales
Eutypa lata UCREL1, XP007795030
Microdochium bolleyi J235TASD1, KXJ89156
Rosellinia necatrix W97, GAP86961
Daldinia sp. EC12, OTB10722
Hypoxylon sp. EC38, OTA63349
Pseudomassariella vexata CBS129021, ORY70480
Pestalotiopsis fici W1061, XP007841379
Xylariales
Chaetomium globosum CBS148.51, XP001224444
Thermothelomyces thermophila ATCC42464, XP003663630
Thielavia terrestris NRRL8126, XP003653705
Chaetomium thermophilum DSM1495, XP006696785
Madurella mycetomatis mm55, KXX82217
Podospora anserina S mat+, XP001903283
Neurospora crassa OR74A, XP962967
N. tetrasperma FGSC2508, XP009848912
Sordaria macrospora khell, XP003347917
Sordariales
Coniochaeta ligniaria NRRL30616, OIW32310
Coniochaetales
Ophiostoma piceae UAMH11346, EPE03819
Sporothrix schenckii 109918, XP016587581
Grosmannia clavigera kw1407, XP014171595
Ophiostomatales
Gaeumannomyces tritici R3111a1, XP009219558
Magnaporthiopsis poae ATCC64411, KLU82219
Magnaporthe oryzae 7015, XP003720365
Magnaporthales
Diaporthe ampelina DA912, KKY29877
Valsa mali 03-8, KUI63578
Diaporthales
Phaeoacremonium minimum UCRPA7, XP007915810
Togniniales
0.08

Figure 28: Bayesian tree of OG 30

Phylogenetic trees 16, 17

Here all MOTUs are from Sordaryomycetes class but in contrast to the phylogenetic trees 20,26 and 30 species from certain orders are missing. In both phylogenetic trees 16 and 17 homologs from species of Diaporthales order are missing. Furthermore, in the MSA 16 there are no representatives of Microascales, Sordariales and Togniniales orders and only one species from Xylariales order is present. In the MSA 17 species from Ophiostomatales and Xylariales orders are not present. The corresponding trees are shown in Figure 29 and Figure 30.

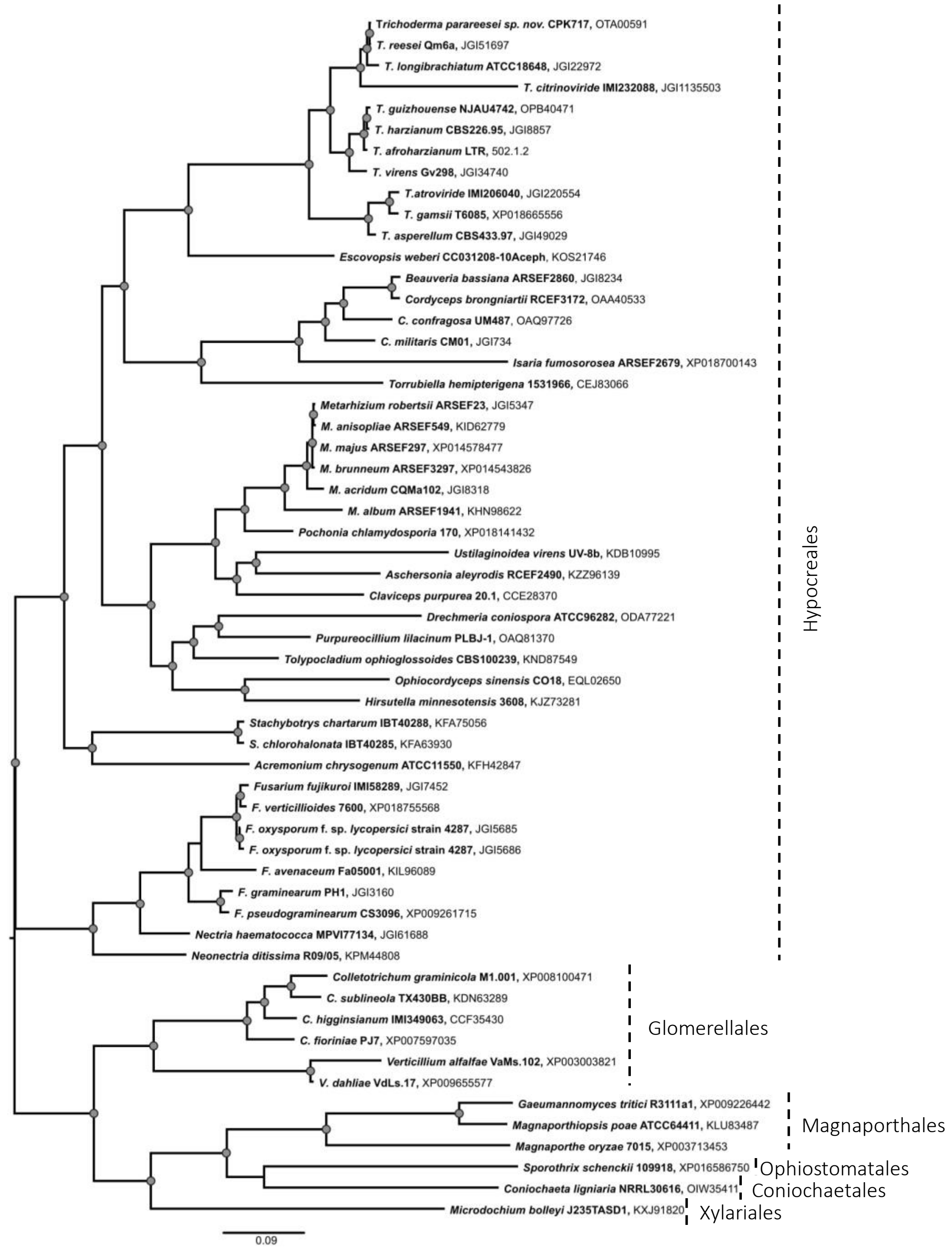

Trichoderma parareesei sp. nov. CPK717, OTA00591
T. reesei Qm6a, JGI51697
T. longibrachiatum ATCC18648, JGI22972
T. citrinoviride IMI232088, JGI1135503
T. guizhouense NJAU4742, OPB40471
T. harzianum CBS226.95, JGI8857
T. afroharzianum LTR, 502.1.2
T. virens Gv298, JGI34740
T.atroviride IMI206040, JGI220554
T. gamsii T6085, XP018665556
T. asperellum CBS433.97, JGI49029
Escovopsis weberi CC031208-10Aceph, KOS21746
Beauveria bassiana ARSEF2860, JGI8234
Cordyceps brongniartii RCEF3172, OAA40533
C. confragosa UM487, OAQ97726
C. militaris CM01, JGI734
Isaria fumosorosea ARSEF2679, XP018700143
Torrubiella hemipterigena 1531966, CEJ83066
Metarhizium robertsii ARSEF23, JGI5347
M. anisopliae ARSEF549, KID62779
M. majus ARSEF297, XP014578477
M. brunneum ARSEF3297, XP014543826
M. acridum CQMa102, JGI8318
M. album ARSEF1941, KHN98622
Pochonia chlamydosporia 170, XP018141432
Ustilaginoidea virens UV-8b, KDB10995
Aschersonia aleyrodis RCEF2490, KZZ96139
Claviceps purpurea 20.1, CCE28370
Drechmeria coniospora ATCC96282, ODA77221
Purpureocillium lilacinum PLBJ-1, OAQ81370
Tolypocladium ophioglossoides CBS100239, KND87549
Ophiocordyceps sinensis CO18, EQL02650
Hirsutella minnesotensis 3608, KJZ73281
Stachybotrys chartarum IBT40288, KFA75056
S. chlorohalonata IBT40285, KFA63930
Acremonium chrysogenum ATCC11550, KFH42847
Fusarium fujikuroi IMI58289, JGI7452
F. verticillioides 7600, XP018755568
F. oxysporum f. sp. lycopersici strain 4287, JGI5685
F. oxysporum f. sp. lycopersici strain 4287, JGI5686
F. avenaceum Fa05001, KIL96089
F. graminearum PH1, JGI3160
F. pseudograminearum CS3096, XP009261715
Nectria haematococca MPVI77134, JGI61688
Neonectria ditissima R09/05, KPM44808
Colletotrichum graminicola M1.001, XP008100471
C. sublineola TX430BB, KDN63289
C. higginsianum IMI349063, CCF35430
C. fioriniae PJ7, XP007597035
Verticillium alfalfae VaMs.102, XP003003821
V. dahliae VdLs.17, XP009655577
Gaeumannomyces tritici R3111a1, XP009226442
Magnaporthiopsis poae ATCC64411, KLU83487
Magnaporthe oryzae 7015, XP003713453
Sporothrix schenckii 109918, XP016586750
Coniochaeta ligniaria NRRL30616, OIW35411
Microdochium bolleyi J235TASD1, KXJ91820
Hypocreales
Glomerellales
Magnaporthales
Ophiostomatales
Coniochaetales
Xylariales
0.09

Figure 29: Bayesian tree of OG 16

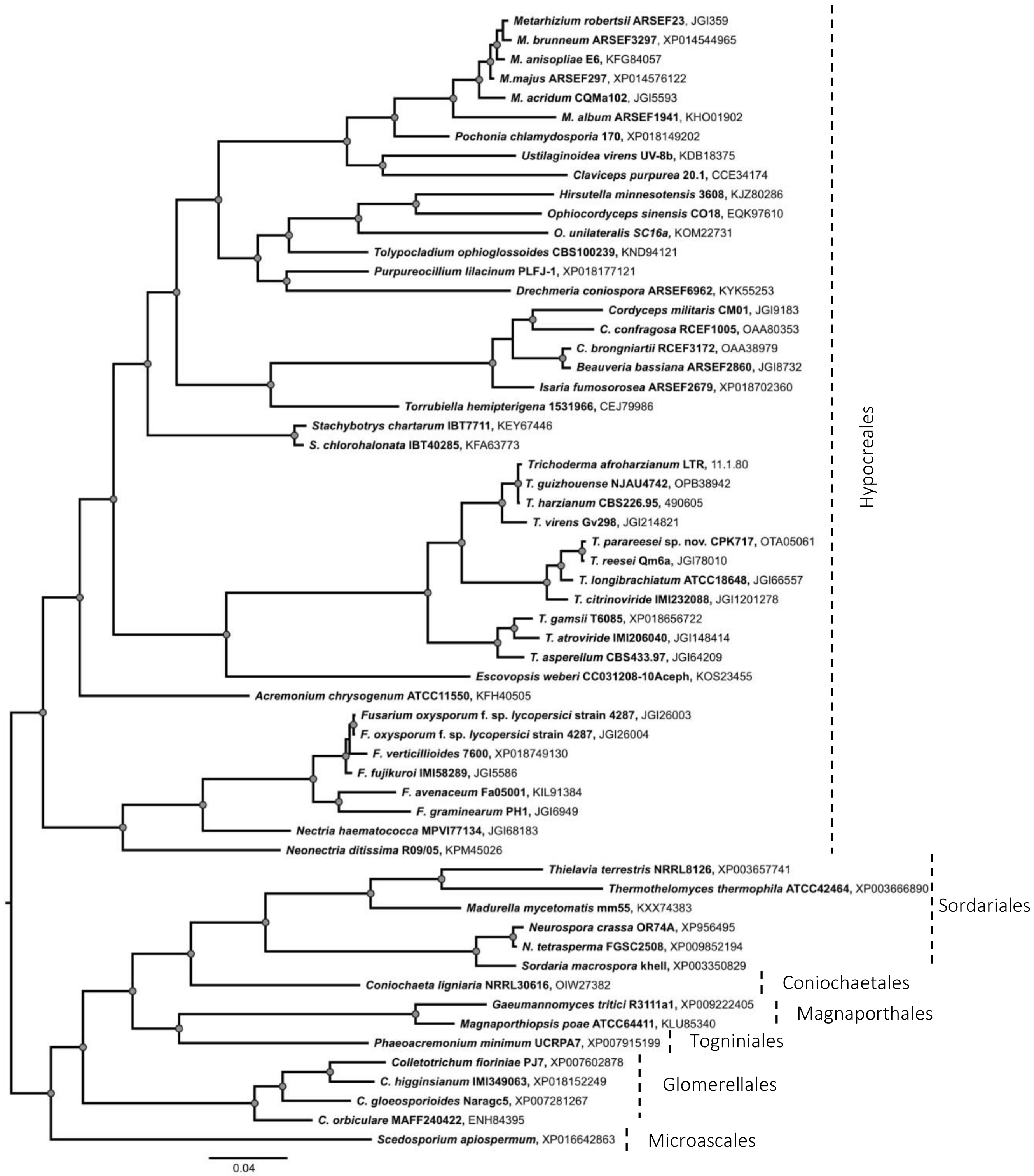

Metarhizium robertsii ARSEF23, JGI359
M. brunneum ARSEF3297, XP014544965
M. anisopliae E6, KFG84057
M.majus ARSEF297, XP014576122
M. acridum CQMa102, JGI5593
M. album ARSEF1941, KHO01902
Pochonia chlamydosporia 170, XP018149202
Ustilaginoidea virens UV-8b, KDB18375
Claviceps purpurea 20.1, CCE34174
Hirsutella minnesotensis 3608, KJZ80286
Ophiocordyceps sinensis CO18, EQK97610
O. unilateralis SC16a, KOM22731
Tolypocladium ophioglossoides CBS100239, KND94121
Purpureocillium lilacinum PLFJ-1, XP018177121
Drechmeria coniospora ARSEF6962, KYK55253
Cordyceps militaris CM01, JGI9183
C. confragosa RCEF1005, OAA80353
C. brongniartii RCEF3172, OAA38979
Beauveria bassiana ARSEF2860, JGI8732
Isaria fumosorosea ARSEF2679, XP018702360
Torrubiella hemipterigena 1531966, CEJ79986
Stachybotrys chartarum IBT7711, KEY67446
S. chlorohalonata IBT40285, KFA63773
Trichoderma afroharzianum LTR, 11.1.80
T. guizhouense NJAU4742, OPB38942
T. harzianum CBS226.95, 490605
T. virens Gv298, JGI214821
T. parareesei sp. nov. CPK717, OTA05061
T. reesei Qm6a, JGI78010
T. longibrachiatum ATCC18648, JGI66557
T. citrinoviride IMI232088, JGI1201278
T. gamsii T6085, XP018656722
T. atroviride IMI206040, JGI148414
T. asperellum CBS433.97, JGI64209
Escovopsis weberi CC031208-10Aceph, KOS23455
Acremonium chrysogenum ATCC11550, KFH40505
Fusarium oxysporum f. sp. lycopersici strain 4287, JGI26003
F. oxysporum f. sp. lycopersici strain 4287, JGI26004
F. verticillioides 7600, XP018749130
F. fujikuroi IMI58289, JGI5586
F. avenaceum Fa05001, KIL91384
F. graminearum PH1, JGI6949
Nectria haematococca MPVI77134, JGI68183
Neonectria ditissima R09/05, KPM45026
Hypocreales
Thielavia terrestris NRRL8126, XP003657741
Thermothelomyces thermophila ATCC42464, XP003666890
Madurella mycetomatis mm55, KXX74383
Neurospora crassa OR74A, XP956495
N. tetrasperma FGSC2508, XP009852194
Sordaria macrospora khell, XP003350829
Sordariales
Coniochaeta ligniaria NRRL30616, OIW27382
Coniochaetales
Gaeumannomyces tritici R3111a1, XP009222405
Magnaporthiopsis poae ATCC64411, KLU85340
Magnaporthales
Phaeoacremonium minimum UCRPA7, XP007915199
Togniniales
Colletotrichum fioriniae PJ7, XP007602878
C. higginsianum IMI349063, XP018152249
C. gloeosporioides Naragc5, XP007281267
C. orbiculare MAFF240422, ENH84395
Glomerellales
Scedosporium apiospermum, XP016642863
Microascales
0.04

Figure 30: Bayesian tree of OG 17

Phylogenetic trees 21, 28

In the phylogenetic tree 21 species from all of the above-mentioned orders from Hypocreales to Xylariales but also two additional homologs from Erysiphales and Helotiales orders (Leotiomycetes) have been found. In the MSA 28 the same orders are represented with the exception of Sordariales and Coniochaetales orders. The corresponding trees are shown in Figure 31 and Figure 32.

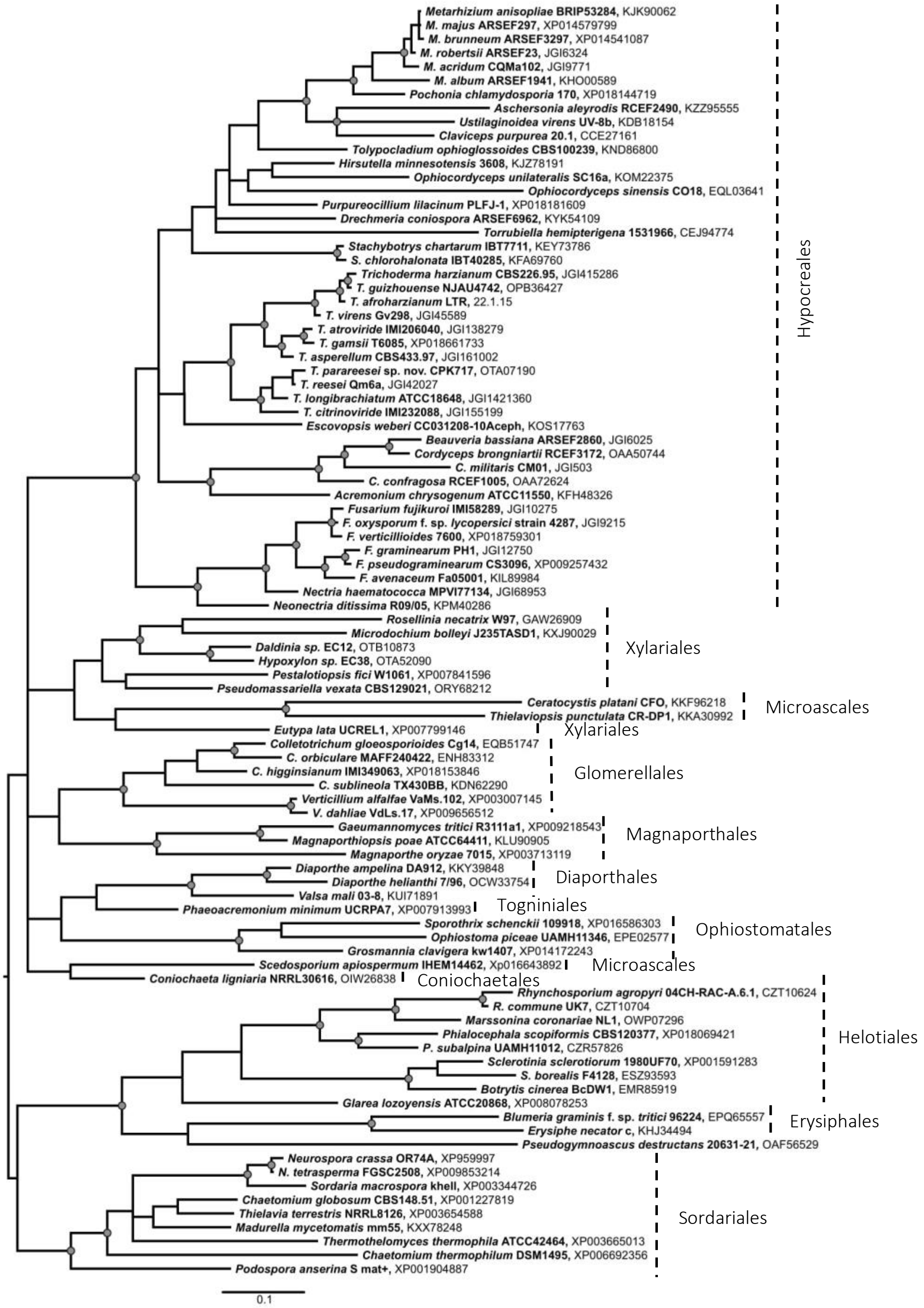

Metarhizium anisopliae BRIP53284, KJK90062
M. majus ARSEF297, XP014579799
M. brunneum ARSEF3297, XP014541087
M. robertsii ARSEF23, JGI6324
M. acridum CQMa102, JGI9771
M. album ARSEF1941, KHO00589
Pochonia chlamydosporia 170, XP018144719
Aschersonia aleyrodis RCEF2490, KZZ95555
Ustilaginoidea virens UV-8b, KDB18154
Claviceps purpurea 20.1, CCE27161
Tolypocladium ophioglossoides CBS100239, KND86800
Hirsutella minnesotensis 3608, KJZ78191
Ophiocordyceps unilateralis SC16a, KOM22375
Ophiocordyceps sinensis CO18, EQL03641
Purpureocillium lilacinum PLFJ-1, XP018181609
Drechmeria coniospora ARSEF6962, KYK54109
Torrubiella hemipterigena 1531966, CEJ94774
Stachybotrys chartarum IBT7711, KEY73786
S. chlorohalonata IBT40285, KFA69760
Trichoderma harzianum CBS226.95, JGI415286
T. guizhouense NJAU4742, OPB36427
T. afroharzianum LTR, 22.1.15
T. virens Gv298, JGI45589
T. atroviride IMI206040, JGI138279
T. gamsii T6085, XP018661733
T. asperellum CBS433.97, JGI161002
T. parareesei sp. nov. CPK717, OTA07190
T. reesei Qm6a, JGI42027
T. longibrachiatum ATCC18648, JGI1421360
T. citrinoviride IMI232088, JGI155199
Escovopsis weberi CC031208-10Aceph, KOS17763
Beauveria bassiana ARSEF2860, JGI6025
Cordyceps brongniartii RCEF3172, OAA50744
C. militaris CM01, JGI503
C. confragosa RCEF1005, OAA72624
Acremonium chrysogenum ATCC11550, KFH48326
Fusarium fujikuroi IMI58289, JGI10275
F. oxysporum f. sp. lycopersici strain 4287, JGI9215
F. verticillioides 7600, XP018759301
F. graminearum PH1, JGI12750
F. pseudograminearum CS3096, XP009257432
F. avenaceum Fa05001, KIL89984
Nectria haematococca MPVI77134, JGI68953
Neonectria ditissima R09/05, KPM40286
Hypocreales
Rosellinia necatrix W97, GAW26909
Microdochium bolleyi J235TASD1, KXJ90029
Daldinia sp. EC12, OTB10873
Hypoxylon sp. EC38, OTA52090
Pestalotiopsis fici W1061, XP007841596
Pseudomassariella vexata CBS129021, ORY68212
Xylariales
Ceratocystis platani CFO, KKF96218
Thielaviopsis punctulata CR-DP1, KKA30992
Microascales
Eutypa lata UCREL1, XP007799146
Xylariales
Colletotrichum gloeosporioides Cg14, EQB51747
C. orbiculare MAFF240422, ENH83312
C. higginsianum IMI349063, XP018153846
C. sublineola TX430BB, KDN62290
Verticillium alfalfae VaMs.102, XP003007145
V. dahliae VdLs.17, XP009656512
Glomerellales
Gaeumannomyces tritici R3111a1, XP009218543
Magnaporthiopsis poae ATCC64411, KLU90905
Magnaporthe oryzae 7015, XP003713119
Magnaporthales
Diaporthe ampelina DA912, KKY39848
Diaporthe helianthi 7/96, OCW33754
Valsa mali 03-8, KUI71891
Diaporthales
Phaeoacremonium minimum UCRPA7, XP007913993
Togniniales
Sporothrix schenckii 109918, XP016586303
Ophiostoma piceae UAMH11346, EPE02577
Grosmannia clavigera kw1407, XP014172243
Ophiostomatales
Scedosporium apiospermum IHEM14462, Xp016643892
Microascales
Coniochaeta ligniaria NRRL30616, OIW26838
Coniochaetales
Rhynchosporium agropyri 04CH-RAC-A.6.1, CZT10624
R. commune UK7, CZT10704
Marssonina coronariae NL1, OWP07296
Phialocephala scopiformis CBS120377, XP018069421
P. subalpina UAMH11012, CZR57826
Sclerotinia sclerotiorum 1980UF70, XP001591283
S. borealis F4128, ESZ93593
Botrytis cinerea BcDW1, EMR85919
Glarea lozoyensis ATCC20868, XP008078253
Helotiales
Blumeria graminis f. sp. tritici 96224, EPQ65557
Erysiphe necator c, KHJ34494
Erysiphales
Pseudogymnoascus destructans 20631-21, OAF56529
Neurospora crassa OR74A, XP959997
N. tetrasperma FGSC2508, XP009853214
Sordaria macrospora khell, XP003344726
Chaetomium globosum CBS148.51, XP001227819
Thielavia terrestris NRRL8126, XP003654588
Madurella mycetomatis mm55, KXX78248
Thermothelomyces thermophila ATCC42464, XP003665013
Chaetomium thermophilum DSM1495, XP006692356
Podospora anserina S mat+, XP001904887
Sordariales
0.1

Figure 31: Bayesian tree of OG 21

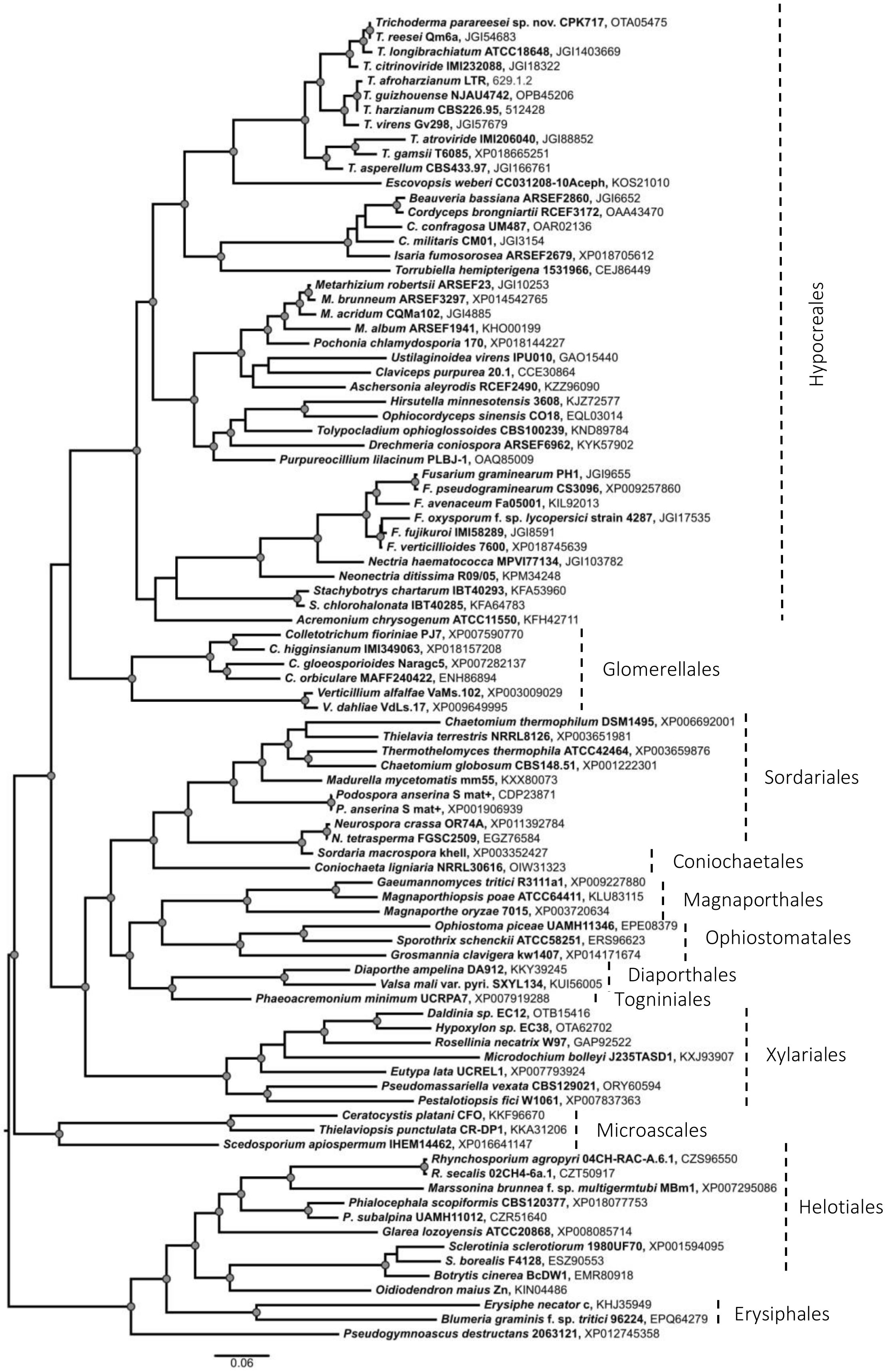

Trichoderma parareesei sp. nov. CPK717, OTA05475
T. reesei Qm6a, JGI54683
T. longibrachiatum ATCC18648, JGI1403669
T. citrinoviride IMI232088, JGI18322
T. afroharzianum LTR, 629.1.2
T. guizhouense NJAU4742, OPB45206
T. harzianum CBS226.95, 512428
T. virens Gv298, JGI57679
T. atroviride IMI206040, JGI88852
T. gamsii T6085, XP018665251
T. asperellum CBS433.97, JGI166761
Escovopsis weberi CC031208-10Aceph, KOS21010
Beauveria bassiana ARSEF2860, JGI6652
Cordyceps brongniartii RCEF3172, OAA43470
C. confragosa UM487, OAR02136
C. militaris CM01, JGI3154
Isaria fumosorosea ARSEF2679, XP018705612
Torrubiella hemipterigena 1531966, CEJ86449
Metarhizium robertsii ARSEF23, JGI10253
M. brunneum ARSEF3297, XP014542765
M. acridum CQMa102, JGI4885
M. album ARSEF1941, KHO00199
Pochonia chlamydosporia 170, XP018144227
Ustilaginoidea virens IPU010, GAO15440
Claviceps purpurea 20.1, CCE30864
Aschersonia aleyrodis RCEF2490, KZZ96090
Hirsutella minnesotensis 3608, KJZ72577
Ophiocordyceps sinensis CO18, EQL03014
Tolypocladium ophioglossoides CBS100239, KND89784
Drechmeria coniospora ARSEF6962, KYK57902
Purpureocillium lilacinum PLBJ-1, OAQ85009
Fusarium graminearum PH1, JGI9655
F. pseudograminearum CS3096, XP009257860
F. avenaceum Fa05001, KIL92013
F. oxysporum f. sp. lycopersici strain 4287, JGI17535
F. fujikuroi IMI58289, JGI8591
F. verticillioides 7600, XP018745639
Nectria haematococca MPVI77134, JGI103782
Neonectria ditissima R09/05, KPM34248
Stachybotrys chartarum IBT40293, KFA53960
S. chlorohalonata IBT40285, KFA64783
Acremonium chrysogenum ATCC11550, KFH42711
Hypocreales
Colletotrichum fioriniae PJ7, XP007590770
C. higginsianum IMI349063, XP018157208
C. gloeosporioides Naragc5, XP007282137
C. orbiculare MAFF240422, ENH86894
Verticillium alfalfae VaMs.102, XP003009029
V. dahliae VdLs.17, XP009649995
Glomerellales
Chaetomium thermophilum DSM1495, XP006692001
Thielavia terrestris NRRL8126, XP003651981
Thermothelomyces thermophila ATCC42464, XP003659876
Chaetomium globosum CBS148.51, XP001222301
Madurella mycetomatis mm55, KXX80073
Podospora anserina S mat+, CDP23871
P. anserina S mat+, XP001906939
Neurospora crassa OR74A, XP011392784
N. tetrasperma FGSC2509, EGZ76584
Sordaria macrospora khell, XP003352427
Sordariales
Coniochaeta ligniaria NRRL30616, OIW31323
Coniochaetales
Gaeumannomyces tritici R3111a1, XP009227880
Magnaporthiopsis poae ATCC64411, KLU83115
Magnaporthe oryzae 7015, XP003720634
Magnaporthales
Ophiostoma piceae UAMH11346, EPE08379
Sporothrix schenckii ATCC58251, ERS96623
Grosmannia clavigera kw1407, XP014171674
Ophiostomatales
Diaporthe ampelina DA912, KKY39245
Valsa mali var. pyri. SXYL134, KUI56005
Diaporthales
Phaeoacremonium minimum UCRPA7, XP007919288
Togniniales
Daldinia sp. EC12, OTB15416
Hypoxylon sp. EC38, OTA62702
Rosellinia necatrix W97, GAP92522
Microdochium bolleyi J235TASD1, KXJ93907
Eutypa lata UCREL1, XP007793924
Pseudomassariella vexata CBS129021, ORY60594
Pestalotiopsis fici W1061, XP007837363
Xylariales
Ceratocystis platani CFO, KKF96670
Thielaviopsis punctulata CR-DP1, KKA31206
Scedosporium apiospermum IHEM14462, XP016641147
Microascales
Rhynchosporium agropyri 04CH-RAC-A.6.1, CZS96550
R. secalis 02CH4-6a.1, CZT50917
Marssonina brunnea f. sp. multigermtubi MBm1, XP007295086
Phialocephala scopiformis CBS120377, XP018077753
P. subalpina UAMH11012, CZR51640
Glarea lozoyensis ATCC20868, XP008085714
Sclerotinia sclerotiorum 1980UF70, XP001594095
S. borealis F4128, ESZ90553
Botrytis cinerea BcDW1, EMR80918
Helotiales
Oidiodendron maius Zn, KIN04486
Erysiphe necator c, KHJ35949
Blumeria graminis f. sp. tritici 96224, EPQ64279
Erysiphales
Pseudogymnoascus destructans 2063121, XP012745358
0.06

Figure 32: Bayesian tree of OG 28

Phylogenetic tree 43

The phylogenetic tree 43 contains homologs from all above-mentioned orders of Hypocreomycetidae subclass of all orders from Sordariomycetidae subclass only Diaporthales order is represented. Furthermore, homologs from Xylariales and Helotiales orders were found as well as homologs from species belonging to Leucosporidiales and Microbotryales orders from Basidiomycota.

Surprisingly, none of the phylogenetic trees contained MOTUs from Eurotiomycetes and Dothideomycetes classes. The corresponding tree is shown in Figure 33.

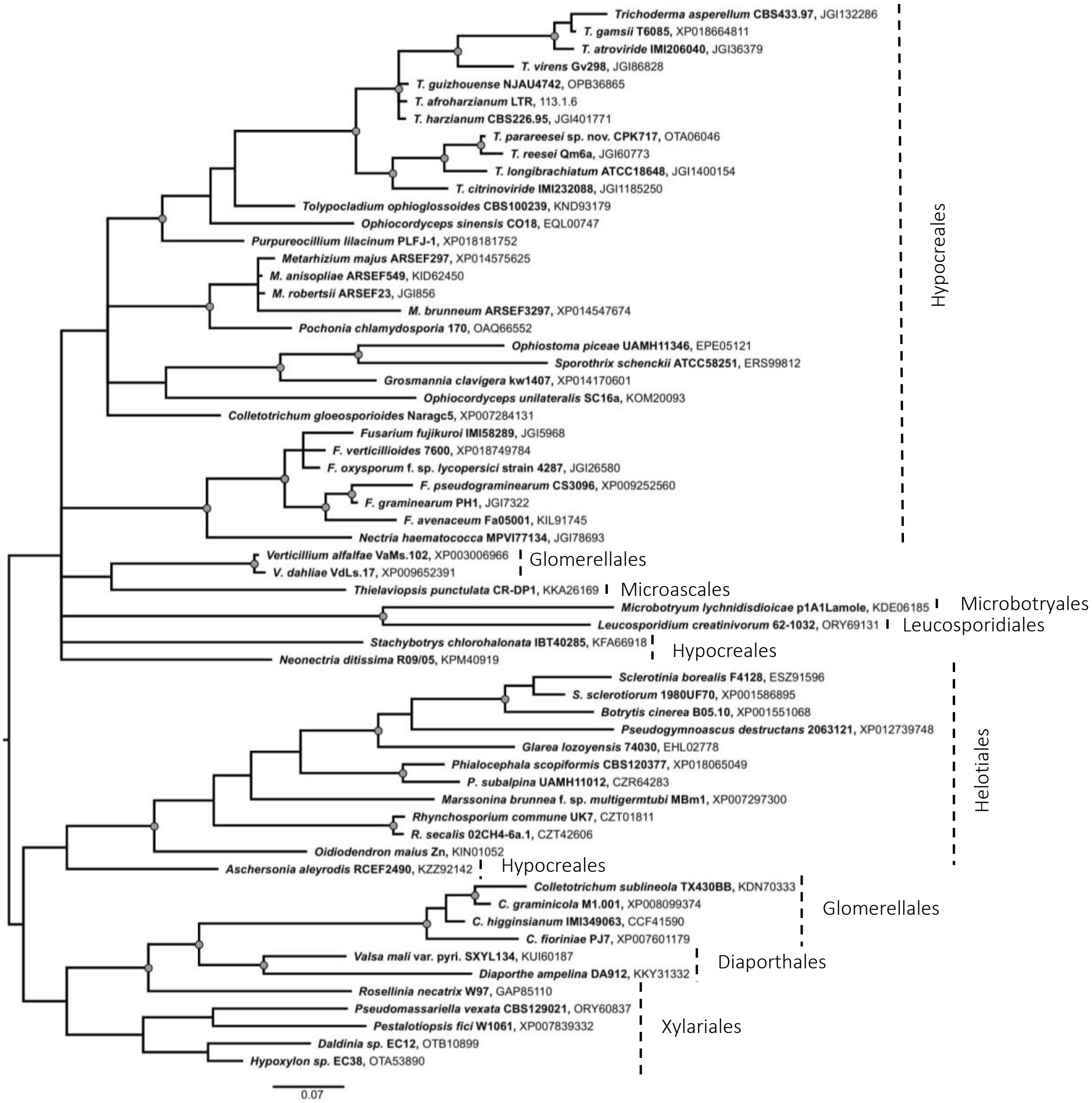

Trichoderma asperellum CBS433.97, JGI132286
T. gamsii T6085, XP018664811
T. atroviride IMI206040, JGI36379
T. virens Gv298, JGI86828
T. guizhouense NJAU4742, OPB36865
T. afroharzianum LTR, 113.1.6
T. harzianum CBS226.95, JGI401771
T. parareesei sp. nov. CPK717, OTA06046
T. reesei Qm6a, JGI60773
T. longibrachiatum ATCC18648, JGI1400154
T. citrinoviride IMI232088, JGI1185250
Tolypocladium ophioglossoides CBS100239, KND93179
Ophiocordyceps sinensis CO18, EQL00747
Purpureocillium lilacinum PLFJ-1, XP018181752
Metarhizium majus ARSEF297, XP014575625
M. anisopliae ARSEF549, KID62450
M. robertsii ARSEF23, JGI856
M. brunneum ARSEF3297, XP014547674
Pochonia chlamydosporia 170, OAQ66552
Ophiostoma piceae UAMH11346, EPE05121
Sporothrix schenckii ATCC58251, ERS99812
Grosmannia clavigera kw1407, XP014170601
Ophiocordyceps unilateralis SC16a, KOM20093
Colletotrichum gloeosporioides Naragc5, XP007284131
Fusarium fujikuroi IMI58289, JGI5968
F. verticillioides 7600, XP018749784
F. oxysporum f. sp. lycopersici strain 4287, JGI26580
F. pseudograminearum CS3096, XP009252560
F. graminearum PH1, JGI7322
F. avenaceum Fa05001, KIL91745
Nectria haematococca MPVI77134, JGI78693
Hypocreales
Verticillium alfalfae VaMs.102, XP003006966
V. dahliae VdLs.17, XP009652391
Glomerellales
Thielaviopsis punctulata CR-DP1, KKA26169
Microascales
Microbotryum lychnidisdioicae p1A1Lamole, KDE06185
Microbotryales
Leucosporidium creatinivorum 62-1032, ORY69131
Leucosporidiales
Stachybotrys chlorohalonata IBT40285, KFA66918
Neonectria ditissima R09/05, KPM40919
Hypocreales
Sclerotinia borealis F4128, ESZ91596
S. sclerotiorum 1980UF70, XP001586895
Botrytis cinerea B05.10, XP001551068
Pseudogymnoascus destructans 2063121, XP012739748
Glarea lozoyensis 74030, EHL02778
Phialocephala scopiformis CBS120377, XP018065049
P. subalpina UAMH11012, CZR64283
Marssonina brunnea f. sp. multigermtubi MBm1, XP007297300
Rhynchosporium commune UK7, CZT01811
R. secalis 02CH4-6a.1, CZT42606
Oidiodendron maius Zn, KIN01052
Helotiales
Aschersonia aleyrodis RCEF2490, KZZ92142
Hypocreales
Colletotrichum sublineola TX430BB, KDN70333
C. graminicola M1.001, XP008099374
C. higginsianum IMI349063, CCF41590
C. fioriniae PJ7, XP007601179
Glomerellales
Valsa mali var. pyri. SXYL134, KUI60187
Diaporthe ampelina DA912, KKY31332
Diaporthales
Rosellinia necatrix W97, GAP85110
Pseudomassariella vexata CBS129021, ORY60837
Pestalotiopsis fici W1061, XP007839332
Daldinia sp. EC12, OTB10899
Hypoxylon sp. EC38, OTA53890
Xylariales
0.07

Figure 33: Bayesian tree of OG 43

4.2.2 Selection Pressure Analysis

At p = 0.01 GARD found no evidence of recombination breakpoints in any of the 18 analyzed MSAs. Therefore, selection pressure analysis was performed without splitting of MSAs.

BUSTED was used by specifying forward branches in each MSA. In six out of 18 core ANKdc-proteins BUSTED found no evidence of positive diversifying selection in any of sites or along branches. It detected evidence of positive diversifying selection in a very low percentage < 1 % of sites in three proteins. In another six ANKdc-proteins it found evidence of positive diversifying selection in a low percentage of sites ≥ 1% and < 3 %. The remaining three proteins had more than 15 % of sites that were under positive selection, whereas MSA 50 contained almost one third of sites that were under positive selection. Model selection for each MSA is shown in Supplementary material 6.

MEME also found no sites under positive selection in MSAs 20 21 and 23, which is in congruence with the findings from BUSTED. In the remaining MSAs where BUSTED found no evidence of positive selection, MEME found 2 sites under positive selection in MSAs 16 and 43, which translates into 0.1 % and 1.5 % of all sites, respectively, while it detected eight sites, or 0.4 % of sites, under positive selection in MSA 22.

In most of the remaining MSAs a very low percentage (< 1 %) of sites was identified under positive selection by MEME, which is mostly in concordance with results from BUSTED. The largest deviations were present in MSAs that according to BUSTED contained more than 15 % of sites under positive selection. In the MSA 25 MEME found evidence of positive selection in 0.6 % (5) of sites, in MSA 39 only 0.9 % (5) of sites and in MSA 50 only 1.1 % (eight sites) of all sites. Model selection for each MSA is shown in Supplementary material 6.

FUBAR found evidence of pervasive positive selection at ≥ 0.95 posterior probability only in 3 proteins. In most MSAs 60 % to 90 % of sites were under purifying (negative) selection based on the FUBAR analysis at ≥ 0.95 posterior probability. In MSA 66 49 % of sites and in MSA 48 41 % of sites were under negative selection pressure. Negative selection was detected in least number of sites in MSAs 58, 64, 39 and 50 ranging from 33 % to 6 % of sites.

Model selection for each MSA is shown in Supplementary material 6.

The results from all analyses are summarized in

Table 9.

	BUSTED	FUBAR		MEME	
Treshold	*p = 0,05*	*posterior probability 0.95*		*p = 0,05*	
Orthogroup	Evidence of diversifying selection	N° of sites diversifying selection	N° of sites purifying selection	N° of sites diversifying selection	N° of sites in MSA
16	no	0	1176	2	1398
17	yes	0	949	3	1274
20	no	0	600	0	757
21	no	0	204	0	272
22	no	0	1680	8	1871
23	no	0	152	0	220
25	yes	0	492	5	820
26	yes	1	477	5	705
28	yes	0	1101	6	1377
30	yes	0	727	1	849
31	yes	0	409	2	578
39	yes	0	58	5	586
43	no	0	102	2	134
48	yes	0	364	5	879
50	yes	2	42	8	735
58	yes	0	438	15	1326
64	yes	0	72	4	522
66	yes	1	965	8	1953

Table 9: Summary of selection pressure analyses BUSTED, FUBAR, MEME

4.3 TMHMM and SignalP analysis

The largest share of ANKdc-proteins with at least one transmembrane helix (TMH) were found in *C. militaris*. Among *Trichoderma* spp., *T. virens, T.afroharzianum* and *T. guizhouense* had the largest share of ANKdc-proteins containing at least one TMH (Figure 34).

In conclusion, the share of proteins with transmembrane helices in Hypocrealean ANKyrome is on average very low, so the function of ANKdc-proteins in *Trichoderma* spp. is not exclusively associated with cell membranes.

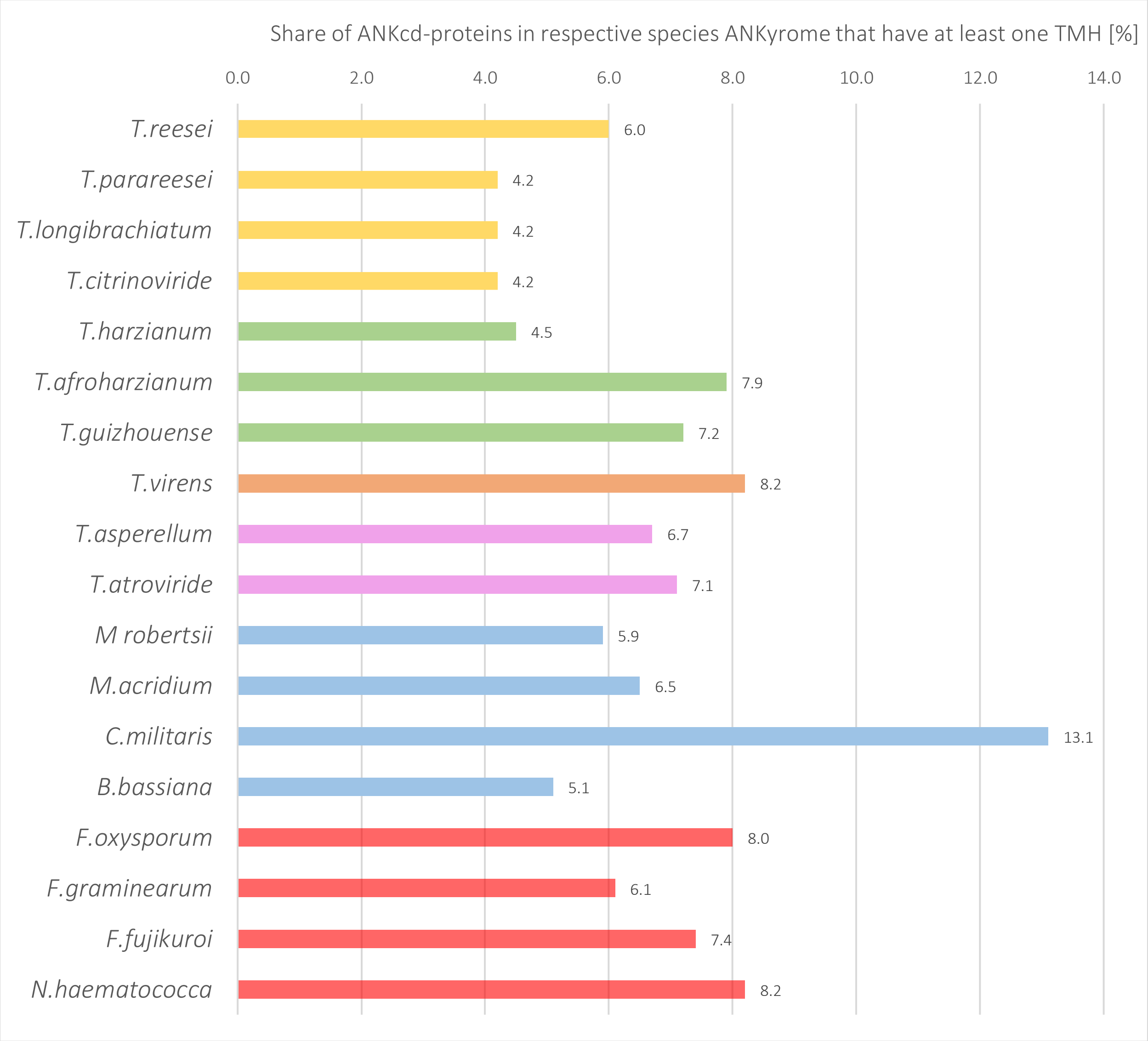


Figure 34: Normalized TMHMM results from Hypocrealean ANKyrome

The largest share of ANKdc-proteins with at least one signal peptide cleavage site were found in *T. reesei* followed by *T. virens*. If absolute numbers of signal peptide containing ANKdc-proteins is considered, the highest number was found in *T. virens* with 11 ANKdc-proteins containing signal peptide cleavage sites (Figure 35).

In conclusion, the share of ANKdc-proteins with signal peptides is on average is also very low, so the function of ANKdc-proteins is not limited to cell signaling pathways in *Trichoderma* spp.

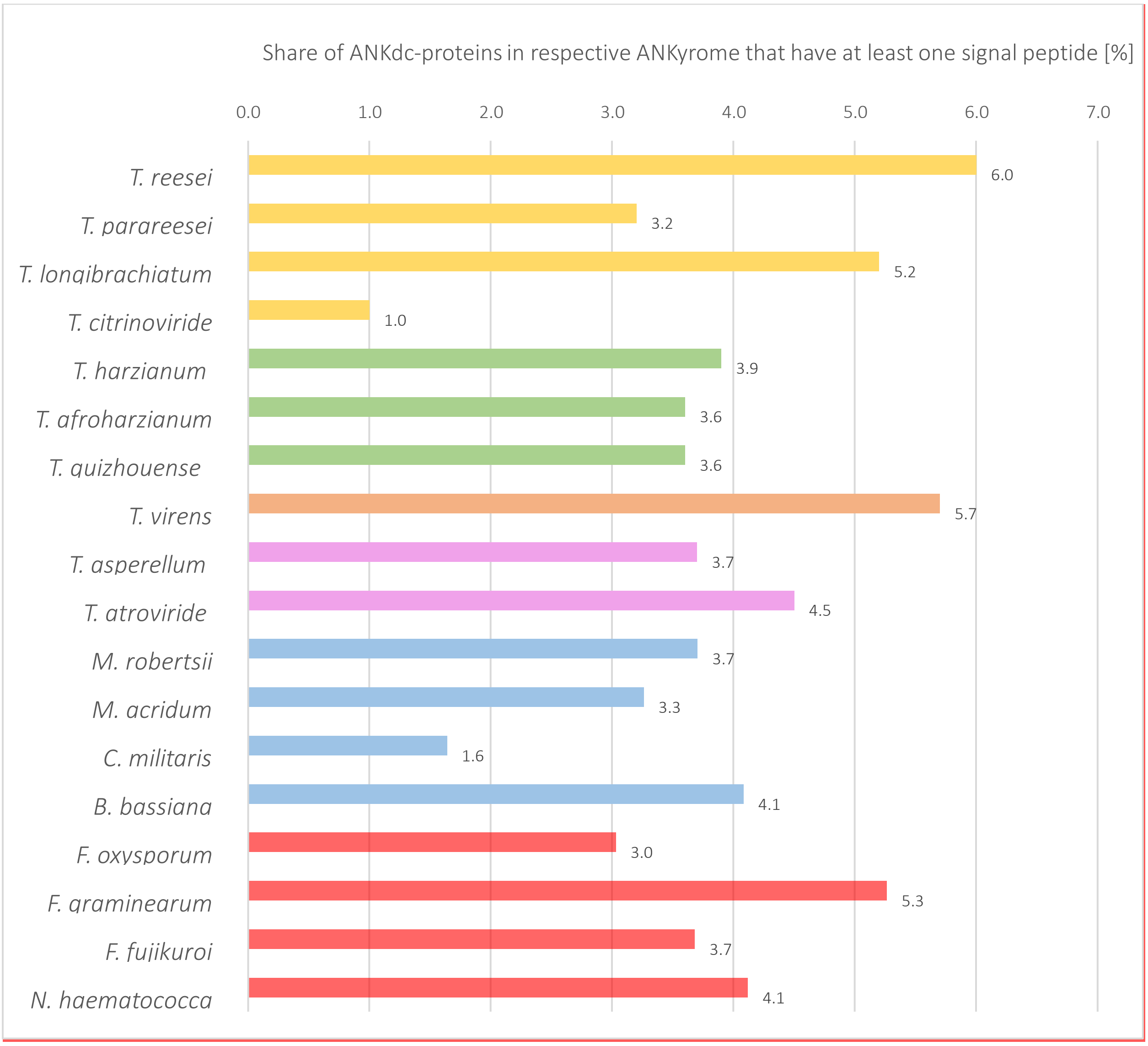


Figure 35: Normalized SignalP results from Hypocrealean ANKyrome

4.4 Transcriptomic data analysis

4.4.1 Differential Gene Expression results from *T. guizhouense* NJAU 4742

The search for ANKdc-proteins in DGE results from *T. guizhouense* transcriptomic analysis are summarized in Table 10. The search revealed a Nucleoside phosphorylase (NP) domain containing ANKdc-protein in *T. guizhouense* (NJAU 4742) which was significantly downregulated after contact with itself, with *T. harzianum* (TUCIM 916 = CBS 226.95) and with *F. oxysporum* (Foc4). Furthermore, two ANKdc-proteins; Protein-kinase (PK) domain containing ANKdc-protein, as well as a PLNTH and Alpha/beta hydrolase fold domain containing ANKdc-protein were significantly upregulated in each of the three confrontation assays.

The expression of seven ANKdc-proteins was upregulated in *T. guizhouense* both after contact with *T. harzianum* and after contact with *F. oxysporum*. Among these proteins, the most significantly upregulated protein is a von Willebrand factor, type A domain containing ANKdc-protein.

The expression of one loner-ANKdc-protein was significantly upregulated and one P-loop containing nucleoside triphosphate hydrolase (PLNTH) domain containing ANKdc-protein was significantly downregulated in *T. guizhouense* exclusively after contact with itself and with *T. harzianum*. The PLNTH containing ANKdc-protein was eight times more downregulated in *T. guizhouense* after contact with itself, compared to confrontation with *T. harzianum*.

Eight ANKdc-proteins were upregulated in *T. guizhouense* both after contact with itself and with *T. harzianum.* The most strongly upregulated expression was observed in a loner-ANKdc-protein OPB40539.

The expression of six ANKdc-proteins was downregulated in *T. guizhouense*, while the expression of one Sigma domain on NACHT NTPases containing ANKdc-protein was upregulated in *T. guizhouense* exclusively after contact with itself.

There are 6 ANKdc-proteins, that are upregulated in *T. guizhouense* exclusively after contact with *T. harzianum*. Among these, the most significantly upregulated ANKdc-protein contains a Protein kinase host-domain.

ProteinID *T. guizhouense*	Host-Domain Annotations	self	916	Foc4
OPB44870	Sigma domain on NACHT-NTPases	2,06	0	0
OPB42534	ANK-domain	-2	0	0
OPB39582	ANK-domain	-2,22	0	0
OPB44758	NADH:flavin oxidoreductase / NADH oxidase family	-2,31	0	0
OPB35975	NACHT domain Nucleoside phosphorylase domain	-2,45	0	0
OPB42857	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain	-2,64	0	0
OPB44958	ANK-domain	-2,79	0	0
OPB41799	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain	-2,88	0	0
OPB40217	Protein kinase domain	0	4,19	0
OPB40882	SPRY domain P-loop containing nucleoside triphosphate hydrolase Concanavalin A-like lectin/glucanase domain Alpha/Beta hydrolase fold	0	3,88	0
OPB39071	NACHT domain,	0	2,8	0
OPB46030	ANK-domain	0	2,19	0
OPB40809	ANK-domain	0	2,15	0
OPB42080	NWD NACHT-NTPase, N-terminal,	0	2,14	0
OPB46886	ATPase, dynein-related, AAA domain,	0	0	4,58
OPB40228	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain,	0	0	3,71
OPB40005	Asparaginase, N-terminal	0	0	2,04
OPB40466	NACHT domain Nucleoside phosphorylase domain	2,24	2,08	0
OPB40539	ANK-domain	3,82	2,8	0

OPB44326	P-loop containing nucleoside triphosphate hydrolase Protein kinase domain	2,45	2,81	0
OPB47160	NACHT domain	2,26	2,48	0
OPB45668	ANK-domain	2,48	2,18	0
OPB42886	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain	-6,16	-3,07	0
OPB37153	Peptidase S8/S53 domain	2,09	2,54	0
OPB37151	Zinc finger, ZZ-type Alpha/Beta hydrolase fold	2,13	2,65	0
OPB46923	ANK-domain	5,19	0	4,68
OPB44812	Protein kinase domain	2,23	0	3,62
OPB42070	ANK-domain	0	2,33	2,31
OPB42501	NACHT domain P-loop containing nucleoside triphosphate hydrolase	0	2,59	2,01
OPB46656	NACHT domain, Nucleoside phosphorylase domain	0	2,42	2,26
OPB45571	ANK-domain	0	4,97	2,62
OPB38898	P-loop containing nucleoside triphosphate hydrolase Putative serine esterase (DUF676) Alpha/Beta hydrolase fold	0	3,97	2,05
OPB36873	P-loop containing nucleoside triphosphate hydrolase	0	2,86	2,49
OPB37316	von Willebrand factor, type A	0	4,41	3,1
OPB38680	ANK-domain	3,36	3,34	2,46
OPB36460	P-loop containing nucleoside triphosphate hydrolase Alpha/Beta hydrolase fold	2,67	3,25	2,55
OPB44602	Nucleoside phosphorylase domain	-4,81	-4,22	-4,22

Table 10: RNA deepseq results from confrontation assays of T. guizhouense *4742. Confrontation with itself,* T. harzianum *916 and* Fusarium oxysporum *Foc4. Protein IDs that are marked red belong to the core* Trichoderma *ANKyrome.*

4.4.2 Differential Gene Expression results from *T. harzianum* TUCIM 916

The search for ANKdc-proteins in DGE results from *T. harzianum* TUCIM 916 transcriptomic analysis are summarized in Table 11. All three confrontation assays led to common transcriptional response in *T. harzianum* by upregulation of expression of eight ANKdc-proteins in total. Five of these ANKdc-proteins were loner-ANKdc-proteins, of which jgi479222 was most significantly upregulated, especially in confrontations with self and with *T. guizhouense*. The rest of the ANkdc-proteins that were upregulated in *T. harzianum* in each confrontation assay were Sigma domain on NACHT-NTPases containing ANKdc-protein, PLNTH- and NP-domain containing ANKdc-protein, as well as the most strongly upregulated von Willebrand factor, type A domain containing ANKdc-protein.

Three ANKdc-proteins were upregulated in *T. harzianum* during the confrontation assay with itself and with *T. guizhouense*. Only two ANKdc-proteins, one loner- and one NP-domain containing ANKdc-proteins were upregulated only after contact with *T. guizhouense* and *F. oxysporum*.

Three ANKdc-proteins were upregulated exclusively after contact with *T. guizhouense*, one loner-ANKdc-protein, one NACHT domain containing ANKdc-protein and one NWD NACHT-NTPase domain containing ANKdc-protein. A PLNTH containing ANKdc-protein was upregulated exclusively in confrontation with *F. oxysporum*. In *T. harzianum* during confrontation with itself, the expression of a WD40- repeat and NP-domain containing ANKdc-protein was upregulated. Also, during the confrontation with itself three different loner-ANKdc-proteins were downregulated in *T. harzianum*.

ProteinID *T. harzianum*	Host-Domain Annotations	self	4742	Foc4
jgi488896	NWD NACHT-NTPase, N-terminal	0	2,12	0
jgi427104	NACHT domain	0	2,11	0
jgi39922	ANK-domain	0	2,1	0
jgi459023	WD40-repeat-containing domain Nucleoside phosphorylase domain	2,38	0	0
jgi98872	ANK-domain	-2,12	0	0
jgi505734	ANK-domain	-2,3	0	0
jgi479284	ANK-domain	-2,8	0	0
jgi63203	P-loop containing nucleoside triphosphate hydrolase	0	0	2,89
jgi44122	ANK-domain	2,48	3,05	0
jgi515791	ANK-domain	3,95	2,55	0
jgi67504	ANK-domain	3,44	2,9	0
jgi508483	Nucleoside phosphorylase domain	0	2,58	2,15
jgi78458	ANK-domain	0	2,04	2,6
jgi145036	ANK-domain	4,27	4,8	3,15
jgi270473	Sigma domain on NACHT-NTPases	2,72	2,63	2,1
jgi335626	P-loop containing nucleoside triphosphate hydrolase NACHT domain,	2,1	2,09	2,27
jgi476900	ANK-domain	4,15	3,97	2,09
jgi479222	ANK-domain	6,96	6,14	4,59
jgi70585	ANK-domain	2,06	2,26	2,38
jgi551129	ANK-domain	2,65	2,57	3,25
jgi505476	von Willebrand factor, type A	5,29	4,61	3,62

Table 11: RNA deepseq results from confrontation assays of T. harzianum *916. Confrontation with self,* T. guizhouense *4742 and* Fusarium oxysporum *Foc4. Protein IDs that are marked red belong to the core* Trichoderma *ANKyrome.*

According to the DGE results, ANKdc-proteins predominantly have a function that is not linked to the interactions with other *Trichoderma* spp. and *Fusarium oxysporum* based on data obtained from Zhang, Miao, Rahimi, Shen, Druzhinina, unpublished.

5 Concluding remarks

The ANKdc-gene expansion was confirmed in strongly opportunistic *Trichoderma* spp. It has been also shown that this expansion is not exclusive to *Trichoderma* spp. In total 2406 ANKdc-proteins were mined from 10 *Trichoderma* spp., and eight genomes of closely related NTHGs. *T. virens* had the largest ANKyrome with 194 ANKdc-proteins, while *T. reesei* with 83 ANKdc-proteins had the smallest ANKyrome. ANK-repeats were re-annotated by InterProScan within Blast2GO software in *Trichoderma* ANKyrome using three databases, Pfam, SMART and Prosite. No specific pattern was observed in the location of ANK-domains in ANKdc-proteins or where they occur in genes (N-terminal, C-terminal or Mid-section) in association with other domains. Additionally, all associated host-domains were annotated by the same approach but using 14 available databases in InterProScan. *T. virens* had the smallest share of hdc-ANKdc-proteins, while the largest share of hdc-ANKdc-proteins was present in *T. guizhouense*. This study also revealed that two most frequent host-domains in *Trichoderma* ANKyrome were P-loop containing nucleoside triphosphate hydrolase (PLNTH) and nucleoside phosphorylase (NP), but also that majority of ANKdc-proteins with these host-domains were orphans.

OrthoFinder orthology inference found 1172 homologous proteins distributed between 183 orthogroups of which 18 were identified as core for *Trichoderma* spp. and 11 as core to all considered fungi from the taxonomical order Hypocreales. In most Hypocreales species the larger share of their respective ANKyrome consisted of orphan ANKdc-proteins. Phylogenetic analysis of 18 *Trichoderma* spp. core ANKdc-proteins resulted in tree topologies that were congruent with the multilocus tree without any deviations. Selection pressure analyses of the same core ANKdc-proteins using HyPhy methods BUSTED, MEME and FUBAR confirmed a significant ($p < 0.05$ and $P > 0.95$) purifying selection acting across all of 18 core ANKdc-proteins.

The trends demonstrated in the *Trichoderma* ANKyrome cannot be explained, until the purpose of these proteins is understood in more detail. Potential transformation candidates from transcriptomic data analysis were identified but generally most of ANKdc-proteins were not differentially expressed in confrontation assays with other fungi, so an alternative function must be proposed and tested.

The high abundance of orphans in *Trichoderma* ANKyrome suggest occurrence of evolutionary mechanisms such as segmental duplications and lateral gene transfer (LGT) events by which these genes may have originated. Horizontal gene transfer between Prokaryotes and Eukaryotes as one of the gene expansion mechanisms has been suggested first by (Bork 1993), but later questioned by (Al-Khodor *et al*. 2010)., who suggested that expansion of ANKdc-genes probably occurred by gene duplication and convergent evolution. Their hypothesis is based on the fact that genes containing repeats are generally more prone to such mechanisms (Lynch and Conery 2000), but also that ANK-repeats interact with universal proteins in nature.

Recently, Druzhinina *et al.* described LGT of plant cell wall degradation enzymes from a wide range of phytopathogenic Ascomycota to *Trichoderma* which parasitizes on them. In this study it was suggested that LGT could have been assisted by adelphoparasitism of *Trichoderma*, meaning that *Trichoderma* spp*.* can prey on *Fusarium* spp*.* and even other *Trichoderma* (Druzhinina *et al*. 2018).Based on the findings from this thesis it cannot be concluded whether the orphan ANKdc-genes were obtained by means of convergent evolution or HGT. The most probable hypothesis is that these genes originated by a combination of multiple mechanisms. Further research is necessary to investigate these possibilities.

Mercer, Fleming, and Ueda found in 2005 that majority of ANKdc-proteins in poxvirus contain an F-box like domain. They suggested that ANK-domains selectively bind a certain protein and thereby directs the ubiquitation of this protein driven by F-box-like domains. Furthermore, since a wide range of ANKdc-proteins containing an F-box-like domain were found in poxvirus they proposed that their purpose is to degrade different host-proteins during viral infection. In *Trichoderma* only one protein of this kind was identified but it would be interesting to observe the effect of a knock-out of this protein on species parasitism.

In general, a similar assumption as with F-box containing ANKdc-proteins can be made with PLNTHs. These domains release energy from NTPs that could be used by a protein bound by ANK-domain. This could explain the presence of a high number of these proteins in *Trichoderma* spp*.* The transcriptomic data indicated an important role of PLNTHs containing ANKdc-proteins in mycoparasitism, but this can be confirmed only by specifically designed experiments.

Finally, the selection pressure analysis of core ANKdc-proteins indicate that they have a key function in *Trichoderma* interactomes that is maintained by purifying selection.

6 Outlook

In this thesis, only the core *Trichoderma* ANKyrome was used for selection pressure analysis, while selection pressure analysis of other ANKdc-proteins was outside the scope of this study. A complete phylogenetic analysis as well as selection pressure would give a deeper insight into the evolution of these proteins, especially the orphan ANKdc-proteins. Since the majority of *Trichoderma* ANKyrome was probably obtained by a mechanism different from vertical evolution, a horizontal gene transfer (HGT) analysis such as NOTUNG (which also points out duplication events) of orphan ANKdc-proteins could be included in a complete evolutionary analysis.

The most common host-domains in ANKdc-proteins in *Trichoderma* spp*.,* the NP and PLNTH were predominantly found in orphan ANKdc-proteins. Proteins containing these domains are mostly involved in energy release from NTPs and they could be a good starting point for the evolutionary analysis of orphan ANKdc-proteins because of their disproportionate abundance. Transcriptomics data revealed that the majority of the *Trichoderma* ANKyrome is not differentially expressed during confrontation studies with other fungi, but a few potential knock-out candidates in *T. guizhouense and T. harzianum* both from their core and non-core ANKyromes have been identified. A deeper investigation by knock-out of these ANKdc-proteins, could provide us with more concrete information on the role of ANKdc-proteins in *Trichoderma*.

Furthermore, confrontation studies of *Trichoderma* spp. with certain mycoviruses, bacteria, and plants followed by transcriptomic analysis could potentially reveal a different ANKyrome expression profile. A more recent study identified a protein Caiap (CARD and ANK domains) with a 16 C-terminal ANK-domain, which acts as an adaptor protein for inflammasome-dependent resistance to *Salmonella enterica* infection of zebra fish (Tyrkalska *et al*. 2017). Even though this study was performed on zebra fish, it supports the idea of performing confrontation studies with bacteria to investigate the transcriptional response in *Trichoderma* spp. when in contact with certain bacterial species.

7 Curriculum Vitae

Vladimir Gojic

📍 Vienna, Austria
✉ vladimir.gojic@gmail.com
📅 21.08.1992, Teslic (Bosnia and Herz.)

EDUCATION

Oct 2016 – Jul 2018 **Master's degree - Technical Chemistry- Biotechnology, TU Wien**

Master thesis: "Convergent evolution of proteins with ankyrin domains the main genomic hallmark of an industrially relevant fungus *Trichoderma",* Prof. I. Druzhinina research group"

Jul - Aug 2017 Optional Internship - Bioorganic synthesis

Co-Author - Publication ChemCatChem, Prof. M.D. Mihovilovic research group

(https://onlinelibrary.wiley.com/doi/full/10.1002/cctc.201800272)

"Chemoenzymatic cascade reaction for the synthesis of enantiopure alcohols"

Aug - Sep 2017 Optional Internship - Optimization of industrial antibody production in CHO cells

Prof. Christoph Herwig research group

"Two-compartment scale-down model of fed-batch CHO cell culture"

Oct 2012 – Sep 2016 **Bachelor`s degree - Technical Chemistry, TU Wien - BSc degree**

Bachelor thesis: "Chemoenzymatic one-pot cascade reaction for the production of enantiopure alcohols", Prof. M.D. Mihovilovic research group

Sep 2009 – Jun 2012 Secondary school BRG Rosasgasse, Vienna
Sep 2007 – Jun 2009 Secondary school BRG Jovan Ducic, Teslic, Bosnia and Herz.

WORK EXPERIENCE

Jul - Sep 2014 Baxter (today Shire), Vienna
Internship Quality Control

ADDITIONAL

Languages	Serbo-croatian (native), german (fluent) und english (fluent)
PC-skills	Python Basics, AutoCAD Basics, MS Office (VBA Basics)

8 References

Al-Khodor, Souhaila, Christopher T. Price, Awdhesh Kalia, and Yousef Abu Kwaik. 2010. ‘Functional Diversity of Ankyrin Repeats in Microbial Proteins’. *Trends in Microbiology* 18 (3): 132–39.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. ‘Basic Local Alignment Search Tool’. *Journal of Molecular Biology* 215 (3): 403–10.

Bateman, Alex, Lachlan Coin, Richard Durbin, Robert D. Finn, Volker Hollich, Sam Griffiths-Jones, Ajay Khanna, Mhairi Marshall, Simon Moxon, and Erik LL Sonnhammer. 2004. ‘The Pfam Protein Families Database’. *Nucleic Acids Research* 32 (suppl_1): D138–41.

Bork, Peer. 1993. ‘Hundreds of Ankyrin-like Repeats in Functionally Diverse Proteins: Mobile Modules That Cross Phyla Horizontally?’ *Proteins: Structure, Function, and Bioinformatics* 17 (4): 363–74.

Chouaki, T., V. Lavarde, L. Lachaud, C. P. Raccurt, and C. Hennequin. 2002. ‘Invasive Infections Due to *Trichoderma* Species: Report of 2 Cases, Findings of in Vitro Susceptibility Testing, and Review of the Literature’. *Clinical Infectious Diseases* 35 (11): 1360–67.

Coleman, Jeffrey J., Steve D. Rounsley, Marianela Rodriguez-Carres, Alan Kuo, Catherine C. Wasmann, Jane Grimwood, Jeremy Schmutz, Masatoki Taga, Gerard J. White, and Shiguo Zhou. 2009. ‘The Genome of *Nectria Haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion’. *PLoS Genetics* 5 (8): e1000618.

Conesa, Ana, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. 2005. ‘Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research’. *Bioinformatics* 21 (18): 3674–76.

Coordinators, NCBI Resource. 2016. ‘Database Resources of the National Center for Biotechnology Information’. *Nucleic Acids Research* 44 (Database issue): D7.

Cuomo, Christina A., Ulrich Güldener, Jin-Rong Xu, Frances Trail, B. Gillian Turgeon, Antonio Di Pietro, Jonathan D. Walton, Li-Jun Ma, Scott E. Baker, and Martijn Rep. 2007. ‘The *Fusarium Graminearum* Genome Reveals a Link between Localized Polymorphism and Pathogen Specialization’. *Science* 317 (5843): 1400–1402.

Delport, Wayne, Art FY Poon, Simon DW Frost, and Sergei L. Kosakovsky Pond. 2010. ‘Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology’. *Bioinformatics* 26 (19): 2455–57.

Druzhinina, Irina S., Komal Chenthamara, Jian Zhang, Lea Atanasova, Dongqing Yang, Youzhi Miao, Mohammad J. Rahimi, Marica Grujic, Feng Cai, and Shadi Pourmehdi. 2018a. ‘Massive Lateral Transfer of Genes Encoding Plant Cell Wall-Degrading Enzymes to the Mycoparasitic Fungus *Trichoderma* from Its Plant-Associated Hosts’. *PLoS Genetics* 14 (4): e1007322.

———. 2018b. ‘Massive Lateral Transfer of Genes Encoding Plant Cell Wall-Degrading Enzymes to the Mycoparasitic Fungus Trichoderma from Its Plant-Associated Hosts’. *PLoS Genetics* 14 (4): e1007322.

Druzhinina, Irina S., Alexey G. Kopchinskiy, Eva M. Kubicek, and Christian P. Kubicek. 2016. ‘A Complete Annotation of the Chromosomes of the Cellulase Producer *Trichoderma Reesei* Provides Insights in Gene Clusters, Their Expression and Reveals Genes Required for Fitness’. *Biotechnology for Biofuels* 9 (1): 75.

Druzhinina, Irina S., Verena Seidl-Seiboth, Alfredo Herrera-Estrella, Benjamin A. Horwitz, Charles M. Kenerley, Enrique Monte, Prasun K. Mukherjee, Susanne Zeilinger, Igor V.

Grigoriev, and Christian P. Kubicek. 2011. '*Trichoderma:* The Genomics of Opportunistic Success'. *Nature Reviews Microbiology* 9 (10): 749.

Edgar, Robert C. 2004. 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput'. *Nucleic Acids Research* 32 (5): 1792–97.

Emms, David M., and Steven Kelly. 2015. 'OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy'. *Genome Biology* 16 (1): 157.

Gao, Qiang, Kai Jin, Sheng-Hua Ying, Yongjun Zhang, Guohua Xiao, Yanfang Shang, Zhibing Duan, Xiao Hu, Xue-Qin Xie, and Gang Zhou. 2011. 'Genome Sequencing and Comparative Transcriptomics of the Model Entomopathogenic Fungi *Metarhizium Anisopliae* and *M. Acridum*'. *PLoS Genetics* 7 (1): e1001264.

Gorina, Svetlana, and Nikola P. Pavletich. 1996. 'Structure of the P53 Tumor Suppressor Bound to the Ankyrin and SH3 Domains of 53BP2'. *Science* 274 (5289): 1001–5.

Grigoriev, Igor V., Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, and Frank Korzeniewski. 2013. 'MycoCosm Portal: Gearing up for 1000 Fungal Genomes'. *Nucleic Acids Research* 42 (D1): D699–704.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0'. *Systematic Biology* 59 (3): 307–21.

Guindon, Stephane, Franck Lethiec, Patrice Duroux, and Olivier Gascuel. 2005. 'PHYML Online—a Web Server for Fast Maximum Likelihood-Based Phylogenetic Inference'. *Nucleic Acids Research* 33 (suppl_2): W557–59.

Gupta, Vijai G., Monika Schmoll, Alfredo Herrera-Estrella, R. S. Upadhyay, Irina Druzhinina, and Maria Tuohy. 2014. *Biotechnology and Biology of Trichoderma*. Newnes.

Hanks, Steven K., Anne Marie Quinn, and Tony Hunter. 1988. 'The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains'. *Science* 241 (4861): 42–52.

Harman, Gary E. 2011. 'Multifunctional Fungal Plant Symbionts: New Tools to Enhance Plant Growth and Productivity'. *New Phytologist* 189 (3): 647–49.

Harman, Gary E., Charles R. Howell, Ada Viterbo, Ilan Chet, and Matteo Lorito. 2004. '*Trichoderma* Species—Opportunistic, Avirulent Plant Symbionts'. *Nature Reviews Microbiology* 2 (1): 43.

Hofmann, Kay, Philipp Bucher, Laurent Falquet, and Amos Bairoch. 1999. 'The PROSITE Database, Its Status in 1999'. *Nucleic Acids Research* 27 (1): 215–19.

Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, and Gift Nuka. 2014. 'InterProScan 5: Genome-Scale Protein Function Classification'. *Bioinformatics* 30 (9): 1236–40.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. 'MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform'. *Nucleic Acids Research* 30 (14): 3059–66.

Katoh, Kazutaka, and Daron M. Standley. 2013. 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability'. *Molecular Biology and Evolution* 30 (4): 772–80.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, and Chris Duran. 2012.

‘Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data’. *Bioinformatics* 28 (12): 1647–49.

Kipreos, Edward T., and Michele Pagano. 2000. ‘The F-Box Protein Family’. *Genome Biology* 1 (5): reviews3002. 1.

Komoń-Zelazowska, Monika, John Bissett, Doustmorad Zafari, Lóránt Hatvani, László Manczinger, Sheri Woo, Matteo Lorito, László Kredics, Christian P. Kubicek, and Irina S. Druzhinina. 2007. ‘Genetically Closely Related but Phenotypically Divergent *Trichoderma* Species Cause Green Mold Disease in Oyster Mushroom Farms Worldwide’. *Applied and Environmental Microbiology* 73 (22): 7415–26.

Koonin, Eugene V., and L. Aravind. 2000. ‘The NACHT Family–a New Group of Predicted NTPases Implicated in Apoptosis and MHC Transcription Activation’. *Trends in Biochemical Sciences* 25 (5): 223–24.

Kosakovsky Pond, Sergei L., David Posada, Michael B. Gravenor, Christopher H. Woelk, and Simon DW Frost. 2006. ‘GARD: A Genetic Algorithm for Recombination Detection’. *Bioinformatics* 22 (24): 3096–98.

Kubicek, Christian P., Alfredo Herrera-Estrella, Verena Seidl-Seiboth, Diego A. Martinez, Irina S. Druzhinina, Michael Thon, Susanne Zeilinger, Sergio Casas-Flores, Benjamin A. Horwitz, and Prasun K. Mukherjee. 2011. ‘Comparative Genome Sequence Analysis Underscores Mycoparasitism as the Ancestral Life Style of *Trichoderma*’. *Genome Biology* 12 (4): R40.

Larsson, Anders. 2014. ‘AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets’. *Bioinformatics* 30 (22): 3276–78.

Lefort, Vincent, Jean-Emmanuel Longueville, and Olivier Gascuel. 2017. ‘SMS: Smart Model Selection in PhyML’. *Molecular Biology and Evolution* 34 (9): 2422–24.

Leipe, Detlef D., Eugene V. Koonin, and L. Aravind. 2004. ‘STAND, a Class of P-Loop NTPases Including Animal and Plant Regulators of Programmed Cell Death: Multiple, Complex Domain Architectures, Unusual Phyletic Patterns, and Evolution by Horizontal Gene Transfer’. *Journal of Molecular Biology* 343 (1): 1–28.

Letunic, Ivica, Tobias Doerks, and Peer Bork. 2011. ‘SMART 7: Recent Updates to the Protein Domain Annotation Resource’. *Nucleic Acids Research* 40 (D1): D302–5.

Lex, Alexander, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. ‘UpSet: Visualization of Intersecting Sets’. *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 1983–92.

Li, Junan, Anjali Mahajan, and Ming-Daw Tsai. 2006. ‘Ankyrin Repeat: A Unique Motif Mediating Protein− Protein Interactions’. *Biochemistry* 45 (51): 15168–78.

Lux, Samuel E., Kathryn M. John, and Vann Bennett. 1990. ‘Analysis of CDNA for Human Erythrocyte Ankyrin Indicates a Repeated Structure with Homology to Tissue-Differentiation and Cell-Cycle Control Proteins’. *Nature* 344 (6261): 36.

Lynch, Michael, and John S. Conery. 2000. ‘The Evolutionary Fate and Consequences of Duplicate Genes’. *Science* 290 (5494): 1151–55. http://science.sciencemag.org/content/sci/290/5494/1151.full.pdf.

Ma, Li-Jun, H. Charlotte Van Der Does, Katherine A. Borkovich, Jeffrey J. Coleman, Marie-Josée Daboussi, Antonio Di Pietro, Marie Dufresne, Michael Freitag, Manfred Grabherr, and Bernard Henrissat. 2010. ‘Comparative Genomics Reveals Mobile Pathogenicity Chromosomes in *Fusarium*’. *Nature* 464 (7287): 367.

Man, Tom JB de, Jason E. Stajich, Christian P. Kubicek, Clotilde Teiling, Komal Chenthamara, Lea Atanasova, Irina S. Druzhinina, Natasha Levenkova, Stephanie SL Birnbaum, and

Seth M. Barribeau. 2016. ‘Small Genome of the Fungus *Escovopsis Weberi*, a Specialized Disease Agent of Ant Agriculture’. *Proceedings of the National Academy of Sciences* 113 (13): 3567–72.

Martinez, Diego, Randy M. Berka, Bernard Henrissat, Markku Saloheimo, Mikko Arvas, Scott E. Baker, Jarod Chapman, Olga Chertkov, Pedro M. Coutinho, and Dan Cullen. 2008a. ‘Genome Sequencing and Analysis of the Biomass-Degrading Fungus *Trichoderma Reesei* (Syn. *Hypocrea Jecorina*)’. *Nature Biotechnology* 26 (5): 553.

———. 2008b. ‘Genome Sequencing and Analysis of the Biomass-Degrading Fungus *Trichoderma Reesei* (Syn. *Hypocrea Jecorina*)’. *Nature Biotechnology* 26 (5): 553.

Martino, Elena, Emmanuelle Morin, Gwen-Aëlle Grelet, Alan Kuo, Annegret Kohler, Stefania Daghino, Kerrie W. Barry, Nicolas Cichocki, Alicia Clum, and Rhyan B. Dockter. 2018. ‘Comparative Genomics and Transcriptomics Depict Ericoid Mycorrhizal Fungi as Versatile Saprotrophs and Plant Mutualists’. *New Phytologist* 217 (3): 1213–29.

Mercer, Andrew A., Stephen B. Fleming, and Norihito Ueda. 2005. ‘F-Box-like Domains Are Present in Most Poxvirus Ankyrin Repeat Proteins’. *Virus Genes* 31 (2): 127–33.

Mosavi, Leila K., Tobin J. Cammett, Daniel C. Desrosiers, and Zheng-yu Peng. 2004. ‘The Ankyrin Repeat as Molecular Architecture for Protein Recognition’. *Protein Science* 13 (6): 1435–48.

Mosavi, Leila K., Daniel L. Minor, and Zheng-yu Peng. 2002. ‘Consensus-Derived Structural Determinants of the Ankyrin Repeat Motif’. *Proceedings of the National Academy of Sciences* 99 (25): 16029–34.

Murrell, Ben, Sasha Moola, Amandla Mabona, Thomas Weighill, Daniel Sheward, Sergei L. Kosakovsky Pond, and Konrad Scheffler. 2013. ‘FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection’. *Molecular Biology and Evolution* 30 (5): 1196–1205.

Murrell, Ben, Steven Weaver, Martin D. Smith, Joel O. Wertheim, Sasha Murrell, Anthony Aylward, Kemal Eren, Tristan Pollner, Darren P. Martin, and Davey M. Smith. 2015. ‘Gene-Wide Identification of Episodic Selection’. *Molecular Biology and Evolution* 32 (5): 1365–71.

Murrell, Ben, Joel O. Wertheim, Sasha Moola, Thomas Weighill, Konrad Scheffler, and Sergei L. Kosakovsky Pond. 2012. ‘Detecting Individual Sites Subject to Episodic Diversifying Selection’. *PLoS Genetics* 8 (7): e1002764.

Paoletti, M., and C. Clave. 2007. ‘The Fungus-Specific HET Domain Mediates Programmed Cell Death in *Podospora Anserina*’. *Eukaryotic Cell* 6 (11): 2001–8.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. ‘SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions’. *Nature Methods* 8 (10): 785.

Pond, Sergei L. Kosakovsky, and Spencer V. Muse. 2005. ‘HyPhy: Hypothesis Testing Using Phylogenies’. In *Statistical Methods in Molecular Evolution*, 125–81. Springer.

Ronquist, Fredrik, Maxim Teslenko, Paul Van Der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. 2012. ‘MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space’. *Systematic Biology* 61 (3): 539–42.

Roy, Helen E., Fernando E. Vega, Dave Chandler, Mark S. Goettel, Judith Pell, and Eric Wajnberg. 2010. *The Ecology of Fungal Entomopathogens*. Springer.

Santos, R. F. dos, E. Blume, G. B. P. da Silva, M. Lazarotto, L. E. Scheeren, P. B. Zini, B. O. Bastos, and C. Rego. 2014. 'First Report of Ilyonectria Robusta Associated with Black Foot Disease of Grapevine in Southern Brazil'. *Plant Disease* 98 (6): 845–845.

Sedgwick, Steven G., and Stephen J. Smerdon. 1999. 'The Ankyrin Repeat: A Diversity of Interactions on a Common Structural Framework'. *Trends in Biochemical Sciences* 24 (8): 311–16.

Seiboth, Bernhard, Christa Ivanova, and Verena Seidl-Seiboth. 2011. '*Trichoderma Reesei:* A Fungal Enzyme Producer for Cellulosic Biofuels'. In *Biofuel Production-Recent Developments and Prospects*. InTech.

Seidl, Verena, Christian Seibel, Christian P. Kubicek, and Monika Schmoll. 2009. 'Sexual Development in the Industrial Workhorse *Trichoderma Reesei'*. *Proceedings of the National Academy of Sciences* 106 (33): 13909–14.

Shriner, Daniel, David C. Nickle, Mark A. Jensen, and James I. Mullins. 2003. 'Potential Impact of Recombination on Sitewise Approaches for Detecting Positive Natural Selection'. *Genetics Research* 81 (2): 115–21.

Smith, Shirley Nash. 2007. 'An Overview of Ecological and Habitat Aspects in the Genus *Fusarium* with Special Emphasis on the Soil-Borne Pathogenic Forms'. *Plant Pathol Bull* 16: 97–120.

Sonnhammer, Erik LL, Gunnar Von Heijne, and Anders Krogh. 1998. 'A Hidden Markov Model for Predicting Transmembrane Helices in Protein Sequences.' In *Ismb*, 6:175–82.

Stricker, Astrid R., Robert L. Mach, and Leo H. De Graaff. 2008. 'Regulation of Transcription of Cellulases-and Hemicellulases-Encoding Genes in *Aspergillus Niger* and *Hypocrea Jecorina* (*Trichoderma Reesei*)'. *Applied Microbiology and Biotechnology* 78 (2): 211.

Suyama, Mikita, David Torrents, and Peer Bork. 2006. 'PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments'. *Nucleic Acids Research* 34 (suppl_2): W609–12.

Takehara, Munenori, Feng Ling, Shingo Izawa, Yoshiharu Inoue, and Akira Kimura. 1995. 'Molecular Cloning and Nucleotide Sequence of Purine Nucleoside Phosphorylase and Uridine Phosphorylase Genes from *Klebsiella* Sp.' *Bioscience, Biotechnology, and Biochemistry* 59 (10): 1987–90.

Team, R. Core. 2013. 'R: A Language and Environment for Statistical Computing'.

Tyrkalska, Sylwia, Sergio Candel, Ana B. Perez Oliva, Ana Valera, Francisca Alcaraz, Diana García-Moreno, Maria Luisa Cayuela, and Victoriano Mulero. 2017. 'Identification of an Evolutionarily Conserved Ankyrin Domain-Containing Protein, Caiap, Which Regulates Inflammasome-Dependent Resistance to Bacterial Infection'. *Frontiers in Immunology* 8: 1375.

Wei, Xiangying, Jianjun Chen, Chunying Zhang, and Dongming Pan. 2016. 'A New Oidiodendron Maius Strain Isolated from *Rhododendron Fortunei* and Its Effects on Nitrogen Uptake and Plant Growth'. *Frontiers in Microbiology* 7: 1327.

Wiemann, Philipp, Christian MK Sieber, Katharina W. Von Bargen, Lena Studt, Eva-Maria Niehaus, Jose J. Espino, Kathleen Huß, Caroline B. Michielse, Sabine Albermann, and Dominik Wagner. 2013. 'Deciphering the Cryptic Genome: Genome-Wide Analyses of the Rice Pathogen *Fusarium Fujikuroi* Reveal Complex Regulation of Secondary Metabolism and Novel Metabolites'. *PLoS Pathogens* 9 (6): e1003475.

Xiao, Guohua, Sheng-Hua Ying, Peng Zheng, Zheng-Liang Wang, Siwei Zhang, Xue-Qin Xie, Yanfang Shang, Raymond J. St Leger, Guo-Ping Zhao, and Chengshu Wang. 2012.

‘Genomic Perspectives on the Evolution of Fungal Entomopathogenicity in Beauveria Bassiana’. *Scientific Reports* 2: 483.

Yang, Dongqing, Kyle Pomraning, Alexey Kopchinskiy, Razieh Karimi Aghcheh, Lea Atanasova, Komal Chenthamara, Scott E. Baker, Ruifu Zhang, Qirong Shen, and Michael Freitag. 2015. ‘Genome Sequence and Annotation of *Trichoderma Parareesei*, the Ancestor of the Cellulase Producer *Trichoderma Reesei*’. *Genome Announcements* 3 (4): e00885-15.

Zheng, Peng, Yongliang Xia, Guohua Xiao, Chenghui Xiong, Xiao Hu, Siwei Zhang, Huajun Zheng, Yin Huang, Yan Zhou, and Shengyue Wang. 2012. ‘Genome Sequence of the Insect Pathogenic Fungus *Cordyceps Militaris*, a Valued Traditional Chinese Medicine’. *Genome Biology* 12 (11): R116.

Supplementary Materials

All the sequences and additional data including following supplementary materials are included in the attached compact disc provided with the thesis.

Supplementary material 1

	N° of ANKdc-proteins in ANKyrome	N° of ANK-repeats in ANKyrome	ANK-repeats per ANKdc-protein
Trichoderma guizhouense	134	1259	9,4
Trichoderma afroharzianum	126	1068	8,5
Metarhizium robertsii	125	1015	8,1
Fusarium fujikuroi	142	1130	8,0
Fusarium oxysporum	208	1518	7,3
Trichoderma atroviride	134	938	7,0
Trichoderma parareesei	77	537	7,0
Trichoderma virens	169	1177	7,0
Metarhizium acridium	81	561	6,9
Beauveria bassiana	82	519	6,3
Fusarium graminearum	98	591	6,0
Cordyceps militaris	53	318	6,0
Nectria haematococca	128	665	5,2
Trichoderma reesei	58	297	5,1
Trichoderma harzianum	151	754	5,0
Trichoderma longibrachiatum	77	365	4,7
Trichoderma asperellum	112	524	4,7
Trichoderma citrinoviride	82	345	4,2
Trichoderma spp.	1120	7264	6,5
NTHGs	917	6317	6,9
Total	2037	13581	6,7

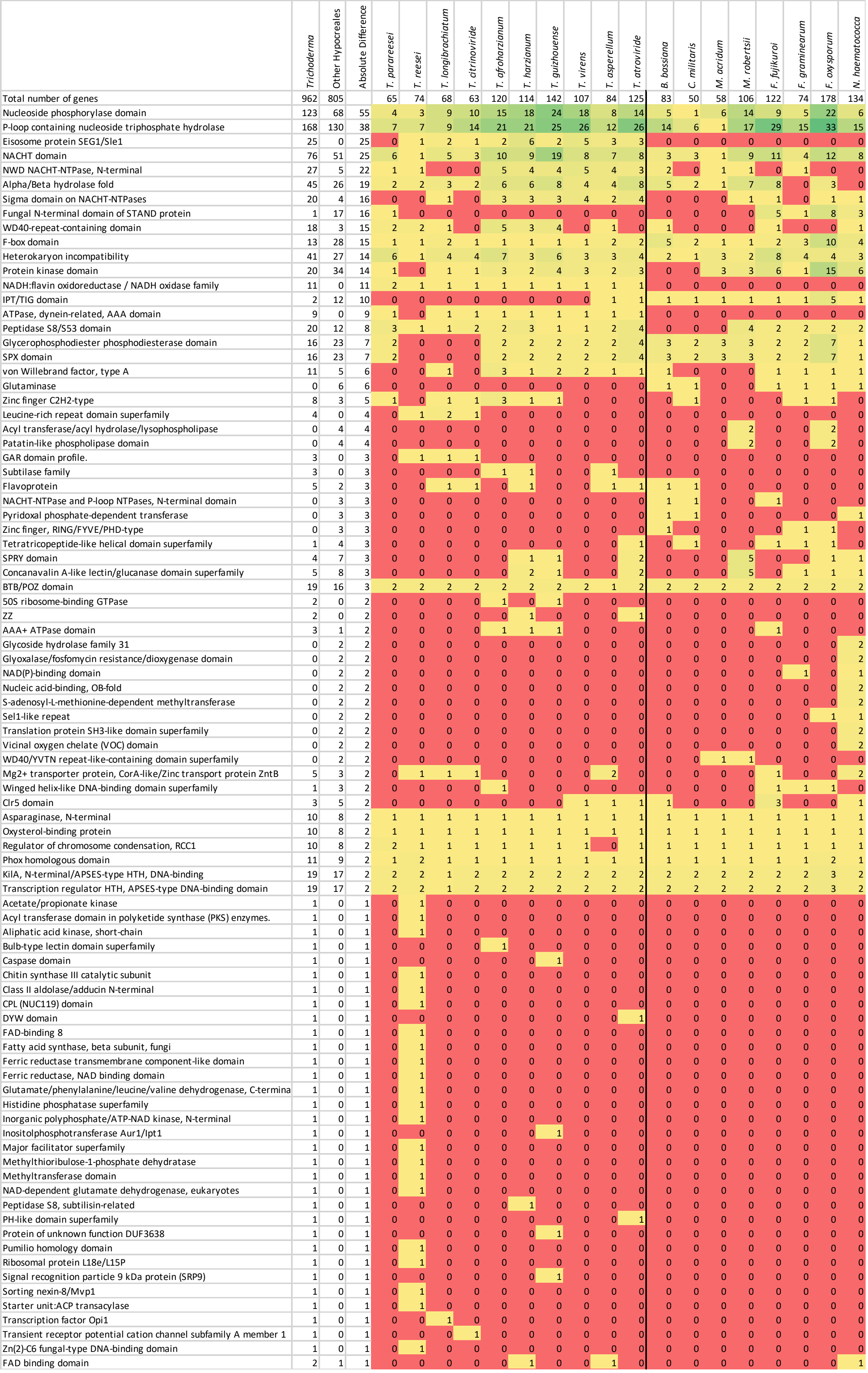
Supplementary table 1: Comparison of N° of ANK-repeats per ANKdc-proteins

Supplementary material 2

	Share of hdc-ANKdc-proteins	N° of hdc-ANKdc-proteins
Trichoderma guizhouense	63	87
Trichoderma afroharzianum	59	82
Trichoderma atroviride	57,1	88
Nectria haematococca	55,3	94
Trichoderma reesei	53	44
Beauveria bassiana	52	51
Fusarium fujikuroi	51,5	84
Cordyceps militaris	49,2	30
Trichoderma parareesei	48,4	46
Trichoderma longibrachiatum	47,9	46
Fusarium oxysporum	45,8	121
Fusarium graminearum	45,6	52
Trichoderma asperellum	45,2	61
Trichoderma citrinoviride	43,8	42
Metarhizium robertsii	43,7	59
Metarhizium acridium	42,4	39
Trichoderma harzianum	41,3	74
Trichoderma virens	37,1	72
Trichoderma spp.	49	642
NTHGs	48,3	530
Total	48,7	1172

Supplementary table 2: Share of all hdc-ANKdc-proteins in different ANKyromes

Supplementary material 3



	Trichoderma	Other Hypocreales	Absolute Difference	*T. parareesei*	*T. reesei*	*T. longibrachiatum*	*T. citrinoviride*	*T. afroharzianum*	*T. harzianum*	*T. guizhouense*	*T. virens*	*T. asperellum*	*T. atroviride*	*B. bassiana*	*C. militaris*	*M. acridum*	*M. robertsii*	*F. fujikuroi*	*F. graminearum*	*F. oxysporum*	*N. haematococca*
Total number of genes	962	805		65	74	68	63	120	114	142	107	84	125	83	50	58	106	122	74	178	134
Nucleoside phosphorylase domain	123	68	55	4	3	9	10	15	18	24	18	8	14	5	1	6	14	9	5	22	6
P-loop containing nucleoside triphosphate hydrolase	168	130	38	7	7	9	14	21	21	25	26	12	26	14	6	1	17	29	15	33	15
Eisosome protein SEG1/Sle1	25	0	25	0	1	2	1	2	6	2	5	3	3	0	0	0	0	0	0	0	0
NACHT domain	76	51	25	6	1	5	3	10	9	19	8	7	8	3	3	1	9	11	4	12	8
NWD NACHT-NTPase, N-terminal	27	5	22	1	1	0	0	5	4	4	5	4	3	2	0	1	1	0	1	0	0
Alpha/Beta hydrolase fold	45	26	19	2	2	3	2	6	6	8	4	4	8	5	2	1	7	8	0	3	0
Sigma domain on NACHT-NTPases	20	4	16	0	0	1	0	3	3	3	4	2	4	0	0	0	1	1	0	1	1
Fungal N-terminal domain of STAND protein	1	17	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	8	3
WD40-repeat-containing domain	18	3	15	2	2	1	0	5	3	4	0	1	0	1	0	0	0	1	0	0	1
F-box domain	13	28	15	1	1	2	1	1	1	1	1	2	2	5	2	1	1	2	3	10	4
Heterokaryon incompatibility	41	27	14	6	1	4	4	7	3	6	3	3	4	2	1	3	2	8	4	4	3
Protein kinase domain	20	34	14	1	0	1	1	3	2	4	3	2	3	0	0	3	3	6	1	15	6
NADH:flavin oxidoreductase / NADH oxidase family	11	0	11	2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
IPT/TIG domain	2	12	10	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	5	1
ATPase, dynein-related, AAA domain	9	0	9	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
Peptidase S8/S53 domain	20	12	8	3	1	1	2	2	3	1	1	2	4	0	0	0	4	2	2	2	2
Glycerophosphodiester phosphodiesterase domain	16	23	7	2	0	0	0	2	2	2	2	2	4	3	2	3	3	2	2	7	1
SPX domain	16	23	7	2	0	0	0	2	2	2	2	2	4	3	2	3	3	2	2	7	1
von Willebrand factor, type A	11	5	6	0	0	1	0	3	1	2	2	1	1	1	0	0	0	1	1	1	1
Glutaminase	0	6	6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1
Zinc finger C2H2-type	8	3	5	1	0	1	1	3	1	1	0	0	0	0	1	0	0	0	1	1	0
Leucine-rich repeat domain superfamily	4	0	4	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acyl transferase/acyl hydrolase/lysophospholipase	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0
Patatin-like phospholipase domain	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0
GAR domain profile.	3	0	3	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Subtilase family	3	0	3	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
Flavoprotein	5	2	3	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0
NACHT-NTPase and P-loop NTPases, N-terminal domain	0	3	3	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
Pyridoxal phosphate-dependent transferase	0	3	3	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1
Zinc finger, RING/FYVE/PHD-type	0	3	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
Tetratricopeptide-like helical domain superfamily	1	4	3	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	0
SPRY domain	4	7	3	0	0	0	0	0	1	1	0	0	2	0	0	0	5	0	0	1	1
Concanavalin A-like lectin/glucanase domain superfamily	5	8	3	0	0	0	0	0	2	1	0	0	2	0	0	0	5	0	1	1	1
BTB/POZ domain	19	16	3	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
50S ribosome-binding GTPase	2	0	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
ZZ	2	0	2	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
AAA+ ATPase domain	3	1	2	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0
Glycoside hydrolase family 31	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Glyoxalase/fosfomycin resistance/dioxygenase domain	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
NAD(P)-binding domain	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Nucleic acid-binding, OB-fold	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
S-adenosyl-L-methionine-dependent methyltransferase	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Sel1-like repeat	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Translation protein SH3-like domain superfamily	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Vicinal oxygen chelate (VOC) domain	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
WD40/YVTN repeat-like-containing domain superfamily	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Mg2+ transporter protein, CorA-like/Zinc transport protein ZntB	5	3	2	0	1	1	1	0	0	0	0	2	0	0	0	0	0	1	0	0	2
Winged helix-like DNA-binding domain superfamily	1	3	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0
Clr5 domain	3	5	2	0	0	0	0	0	0	0	1	1	1	1	0	0	0	3	0	0	1
Asparaginase, N-terminal	10	8	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oxysterol-binding protein	10	8	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Regulator of chromosome condensation, RCC1	10	8	2	2	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Phox homologous domain	11	9	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
KilA, N-terminal/APSES-type HTH, DNA-binding	19	17	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2
Transcription regulator HTH, APSES-type DNA-binding domain	19	17	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2
Acetate/propionate kinase	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acyl transferase domain in polyketide synthase (PKS) enzymes.	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aliphatic acid kinase, short-chain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bulb-type lectin domain superfamily	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Caspase domain	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Chitin synthase III catalytic subunit	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Class II aldolase/adducin N-terminal	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CPL (NUC119) domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DYW domain	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
FAD-binding 8	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fatty acid synthase, beta subunit, fungi	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ferric reductase transmembrane component-like domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ferric reductase, NAD binding domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glutamate/phenylalanine/leucine/valine dehydrogenase, C-termina	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Histidine phosphatase superfamily	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inorganic polyphosphate/ATP-NAD kinase, N-terminal	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inositolphosphotransferase Aur1/Ipt1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Major facilitator superfamily	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Methylthioribulose-1-phosphate dehydratase	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Methyltransferase domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAD-dependent glutamate dehydrogenase, eukaryotes	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Peptidase S8, subtilisin-related	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
PH-like domain superfamily	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Protein of unknown function DUF3638	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Pumilio homology domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ribosomal protein L18e/L15P	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Signal recognition particle 9 kDa protein (SRP9)	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Sorting nexin-8/Mvp1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Starter unit:ACP transacylase	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Transcription factor Opi1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Transient receptor potential cation channel subfamily A member 1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zn(2)-C6 fungal-type DNA-binding domain	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAD binding domain	2	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1

Supplementary table 3: Full results of ANKdc-proteins host-domain annotations part 1

	Trichoderma	Other Hypocreales	Absolute Difference	*T. parareesei*	*T. reesei*	*T. longibrachiatum*	*T. citrinoviride*	*T. afroharzianum*	*T. harzianum*	*T. guizhouense*	*T. virens*	*T. asperellum*	*T. atroviride*	*B. bassiana*	*C. militaris*	*M. acridum*	*M. robertsii*	*F. fujikuroi*	*F. graminearum*	*F. oxysporum*	*N. haematococca*
2-oxoglutarate dehydrogenase C-terminal	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2-oxoglutarate dehydrogenase N-terminus	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Amidase signature domain	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Amidohydrolase 3	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Aminoacyl-tRNA synthetase, class II (G/ P/ S/T)	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Aminotransferase class V domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Aminotransferase class-III	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Anticodon binding domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Armadillo-like helical	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Aspartic peptidase domain superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Beta-hexosaminidase	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
C2 domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cation-transporting P-type ATPase, C-terminal	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Centromere protein C/Mif2/cnp3	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Chitobiase/beta-hexosaminidase-like, domain 2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Citron homology (CNH) domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
DDHD domain	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Dehydrogenase, E1 component	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Dimeric alpha-beta barrel	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Enoyl-CoA hydratase 2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Epoxide hydrolase, N-terminal	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Galactose mutarotase, N-terminal barrel	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Glycoside hydrolase family 20, catalytic domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
HAD-like superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
HIT-like superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
HotDog domain superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Kinetochore CENP-C fungal homologue, Mif2, N-terminal	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Lactonase, 7-bladed beta propeller	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Lipase, secreted	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
MFS transporter superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Mif2/CENP-C cupin domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Mitochondrial substrate/solute carrier	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NB-ARC	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Nucleophile aminohydrolases, N-terminal	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NUDIX hydrolase domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Oligopeptide transporter, OPT superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Oligosaccharyl transferase complex, subunit OST3/OST6	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfan	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Poly(ADP-ribose) polymerase, catalytic domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Polyketide cyclase / dehydrase and lipid transport	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Proteasome alpha-subunit, N-terminal domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Proton-dependent oligopeptide transporter family	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Region found in RelA / SpoT proteins	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Ribosomal protein L2, domain 2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
RNA cap guanine-N2 methyltransferase	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
SAM-dependent O-methyltransferase class I-type profile.	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Scavenger mRNA decapping enzyme DcpS/DCS2	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Serine aminopeptidase, S33	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
SKP1 component, dimerisation	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
S-phase kinase-associated protein 1-like	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
SPRY_RanBP_like	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
START-like domain superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Synaptotagmin-like mitochondrial-lipid-binding domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Thioredoxin-like superfamily	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Threonine-tRNA ligase catalytic core domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Transketolase-like, pyrimidine-binding domain	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Translation elongation factor IF5A-like	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Triacylglycerol lipase	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ubiquitin-interacting motif (UIM) domain profile.	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Ureohydrolase	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Vacuolar sorting protein 39/Transforming growth factor beta recep	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Vam6/VPS39/TRAP1 family	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
YCII-related	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AH/BAR domain superfamily	3	2	1	0	1	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
Acyl-CoA N-acyltransferase	1	2	1	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0
Fungal specific transcription factor domain	1	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Putative GTPase activating protein for Arf	1	2	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
Peptidase A2A, retrovirus, catalytic	2	3	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	2	0	0
G-patch domain	5	4	1	1	1	1	0	0	1	0	1	0	0	1	0	1	1	0	0	0	1
bZIP_YAP	6	5	1	0	0	0	0	1	0	1	1	2	1	1	1	1	0	1	1	0	0
Putative serine esterase (DUF676)	8	7	1	1	0	1	0	1	0	2	1	1	1	0	0	1	1	2	1	1	1
Arginine and arginine-like N-methyltransferase domain profile.	9	8	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Allantoicase	10	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Dilute domain	10	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Palmitoyltransferase, DHHC domain	10	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
VPS9 domain profile.	10	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Pleckstrin homology domain	11	10	1	1	1	2	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1
Argininosuccinate lyase C-terminal	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Armadillo-type fold	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
B30.2/SPRY domain	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
Coatomer, WD associated region	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cyclin, C-terminal domain	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cyclin, N-terminal domain	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Fumarate lyase, N-terminal	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Galactose-binding-like domain superfamily	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
MaoC-like domain	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Zinc finger, MYND-type	2	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
Prion-inhibition and propagation, HeLo domain	3	3	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	2	0	1	0
MBF transcription factor complex subunit Mbp1/Res1/Res2	9	9	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Zinc finger, ZZ-type	9	9	0	2	0	0	0	3	0	3	0	0	1	0	0	2	2	2	1	2	0

Supplementary table 4: Full results of ANKdc-proteins host-domain annotations part 2

Supplementary material 4

	No. Of ANKdc-proteins orphans	**Total no. Of ANKdc-proteins**	**% putative ANKdc-proteins orphans**	**Absolute difference**
Nectria haematococca	120	170	70,6	50
Metarhizium robertsii	85	135	63,0	50
Trichoderma virens	119	194	61,3	75
Trichoderma asperellum	82	135	60,7	53
Fusarium graminearum	69	114	60,5	45
Beauveria bassiana	55	98	56,1	43
Metarhizium acridum	51	92	55,4	41
Trichoderma atroviride	84	154	54,6	70
Trichoderma harzianum	91	179	50,8	88
Cordyceps militaris	29	61	47,5	32
Fusarium fujikuroi	77	163	47,2	86
Trichoderma reesei	39	83	47,0	44
Trichoderma citrinoviride	42	96	43,8	54
Trichoderma longibrachiatum	41	96	42,7	55
Fusarium oxysporum	111	264	42,1	153
Trichoderma guizhouense	56	138	40,6	82
Trichoderma afroharzianum	51	139	36,7	88
Trichoderma parareesei	32	95	33,7	63
Trichoderma spp.	637	1309	48,7	672
Other Hypocreales	597	1097	54,4	500
Total	1234	2406	51,3	1172

Supplementary table 5: Detailed summary of orphan ANKdc-proteins in respective ANKyrome

Supplementary material 5

Orthogroup	Model	N° of taxa
16	JTT	57
17	JTT	59
20	LG	79
21	JTT	92
22	JTT	42
23	JTT	50
25	JTT	47
26	JTT	73
28	JTT	91
30	JTT	81
31	JTT	46
39	JTT	11
43	LG	61
48	JTT	11
50	JTT	18
58	JTT	13
64	JTT	10
66	JTT	20

Supplementary table 6: BIC substitution model selection results

Supplementary material 6

	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-30440,5**	**50**	**60981,1**	**Test**	**0.02 (91.53%)**	**0.61 (8.25%)**	**6.67 (0.21%)**
	Constrained model	-30441,7	49	60981,6	Test	0.02 (92.48%)	0.55 (3.82%)	1.00 (3.70%)
17	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-21455,8**	**50**	**43011,8**	**Test**	**0.01 (91.22%)**	**0.54 (8.40%)**	**29.69 (0.37%)**
	Constrained model	-21476,6	49	43051,3	Test	0.01 (90.57%)	0.13 (3.62%)	1.00 (5.81%)
20	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-15350,4**	**50**	**30801,1**	**Test**	**0.02 (90.28%)**	**0.38 (7.36%)**	**2.00 (2.36%)**
	Constrained model	-15351,6	49	30801,5	Test	0.02 (86.33%)	0.03 (7.29%)	1.00 (6.37%)
21	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-5505,7**	**48**	**11108,6**	**Test**	**0.01 (92.23%)**	**0.71 (7.14%)**	**7.75 (0.63%)**
	Constrained model	-5507,7	47	11110,4	Test	0.01 (92.87%)	1.00 (4.91%)	1.00 (2.22%)
22	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	-50431,8	48	100959,7	Test	0.01 (84.32%)	0.51 (15.17%)	4.27 (0.51%)
	Constrained model	**-50432,5**	**47**	**100959,2**	**Test**	**0.01 (84.50%)**	**0.36 (9.62%)**	**1.00 (5.88%)**
23	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	-5109,5	47	10314,3	Test	0.02 (81.89%)	0.38 (17.08%)	2.96 (1.03%)
	Constrained model	**-5109,9**	**46**	**10313**	**Test**	**0.02 (81.97%)**	**0.30 (13.15%)**	**1.00 (4.88%)**
25	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-26343,6**	**48**	**52783,6**	**Test**	**0.00 (60.50%)**	**0.17 (20.31%)**	**1.35 (19.19%)**
	Constrained model	-26347,5	47	52789,4	Test	0.00 (59.42%)	0.07 (14.74%)	1.00 (25.84%)
26	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-16352,7**	**47**	**32799,8**	**Test**	**0.01 (84.06%)**	**0.59 (13.80%)**	**5.73 (2.15%)**
	Constrained model	-16368,8	46	32830	Test	0.00 (85.27%)	1.00 (10.44%)	1.00 (4.29%)
28	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-28553,7**	**48**	**57203,7**	**Test**	**0.02 (93.66%)**	**1.00 (5.89%)**	**1790.43 (0.45%)**
	Constrained model	-28613,3	47	57320,8	Test	0.01 (92.52%)	1.00 (0.05%)	1.00 (7.43%)
30	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-16885,6**	**48**	**33867,6**	**Test**	**0.01 (91.89%)**	**0.54 (7.90%)**	**13.12 (0.21%)**
	Constrained model	-16890,8	47	33875,9	Test	0.01 (92.37%)	0.31 (3.72%)	1.00 (3.91%)
31	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-11490,5**	**48**	**23077,5**	**Test**	**0.01 (89.56%)**	**0.49 (9.28%)**	**8.30 (1.16%)**
	Constrained model	-11504,7	47	23103,9	Test	0.01 (89.91%)	0.01 (2.35%)	1.00 (7.74%)
39	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3

	Unconstrained model	**-6794,4**	**32**	**13653,2**	**Test**	**0.09 (57.25%)**	**0.15 (26.61%)**	**2.35 (16.14%)**
	Constrained model	-6798,4	31	13659,3	Test	0.00 (25.69%)	0.00 (36.04%)	1.00 (38.26%)
43	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-3161,5**	**42**	**6408,8**	**Test**	**0.02 (84.85%)**	**0.11 (9.84%)**	**2.37 (5.31%)**
	Constrained model	-3163,5	41	6410,7	Test	0.02 (76.37%)	0.03 (13.32%)	1.00 (10.31%)
48	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-10513,8**	**32**	**21092**	**Test**	**0.07 (89.82%)**	**1.00 (9.80%)**	**22.40 (0.37%)**
	Constrained model	-10518,5	31	21099,3	Test	0.04 (85.59%)	1.00 (0.29%)	1.00 (14.12%)
50	**Unconstrained model**	**-14199,5**	**43**	**28485,6**	**Test**	**0.00 (22.64%)**	**0.01 (47.64%)**	**3.49 (29.72%)**
					Background	**0.00 (52.50%)**	**1.00 (41.58%)**	**8.56 (5.92%)**
	Constrained model	-14225,1	42	28534,6	Test	0.00 (12.66%)	0.00 (45.45%)	1.00 (41.90%)
					Background	0.00 (52.56%)	1.00 (41.86%)	8.53 (5.57%)
58	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-19446,7**	**32**	**38957,5**	**Test**	**0.00 (64.86%)**	**0.83 (34.27%)**	**41.72 (0.87%)**
	Constrained model	-19464,2	31	38990,6	Test	0.00 (68.31%)	0.91 (0.00%)	1.00 (31.69%)
64	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-5922,8**	**30**	**11906,1**	**Test**	**0.07 (71.11%)**	**0.71 (26.26%)**	**9.24 (2.63%)**
	Constrained model	-5927,5	29	11913,5	Test	0.04 (63.83%)	0.05 (6.70%)	1.00 (29.48%)
66	Model	log L	#. params	AICc	Branch set	ω1	ω2	ω3
	Unconstrained model	**-23346,8**	**32**	**46757,7**	**Test**	**0.04 (80.34%)**	**0.91 (19.30%)**	**37.57 (0.36%)**
	Constrained model	-23356,4	31	46774,9	Test	0.02 (79.01%)	0.96 (0.00%)	1.00 (20.99%)

Supplementary table 7: BUSTED model selection. Selected models are bold.

16	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	70628.83	-35275.35	39			
	Global MG94xREV	61311.91	-30655.95	46	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0605
17	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	50224.31	-25073.08	39			
	Global MG94xREV	43463.51	-21731.76	46	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0586
20	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	35391.76	-17656.76	39			
	Global MG94xREV	31007.35	-15503.67	46	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0672
21	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	12836.64	-6380.99	37			
	Global MG94xREV	11126.57	-5563.29	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0652
22	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	114036.28	-56981.09	37			
	Global MG94xREV	101777.22	-50888.61	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0854
23	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	11676.71	-5801.97	36			
	Global MG94xREV	10269.49	-5134.74	43	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0883
25	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	56415.96	-28170.87	37			
	Global MG94xREV	53504.88	-26752.44	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.204
26	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	36862.66	-18395.21	36			
	Global MG94xREV	33217.44	-16608.72	43	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.113
28	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	65939.45	-32932.66	37			
	Global MG94xREV	57952.95	-28976.48	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.074
30	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	40277.68	-20101.73	37			
	Global MG94xREV	34082.99	-17041.49	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0481
31	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	26366.44	-13146.07	37			
	Global MG94xREV	23324.95	-11662.48	44	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0695
39	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	14184.62	-7071.21	21			
	Global MG94xREV	13672.43	-6836.22	28	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.335
43	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	7258.82	-3597.94	31			
	Global MG94xREV	6387.88	-3193.94	38	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.0894
48	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	22751.11	-11354.49	21			
	Global MG94xREV	21110.94	-10555.47	28	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.154
50	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	29304.29	-14625.05	27			
	Global MG94xREV	28701.21	-14350.6	34	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.505
58	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	41063.94	-20510.93	21			
	Global MG94xREV	39166.66	-19583.33	28	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.263
64	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	12504.37	-6233.08	19			
	Global MG94xREV	11917.33	-5958.66	26	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.265
66	Model	AICC	log L	Parameters	Rate distributions		
	Nucleotide GTR	49823.53	-24890.74	21			
	Global MG94xREV	46911.29	-23455.64	28	non-synonymous/synonymous rate ratio for *test*		
					100%	@	0.189

Supplementary table 8: MEME model selection

16	Model	AICC	log L	Parameters
	Nucleotide GTR	70637.61	-35279.74	39
17	Model	AICC	log L	Parameters
	Nucleotide GTR	50224.31	-25073.08	39
20	Model	AICC	log L	Parameters
	Nucleotide GTR	35391.76	-17656.76	39
21	Model	AICC	log L	Parameters
	Nucleotide GTR	12836.64	-6380.99	37
22	Model	AICC	log L	Parameters
	Nucleotide GTR	114036.28	-56981.09	37
23	Model	AICC	log L	Parameters
	Nucleotide GTR	11676.71	-5801.97	36
25	Model	AICC	log L	Parameters
	Nucleotide GTR	56415.96	-28170.87	37
26	Model	AICC	log L	Parameters
	Nucleotide GTR	36862.66	-18395.21	36
28	Model	AICC	log L	Parameters
	Nucleotide GTR	65939.45	-32932.66	37
30	Model	AICC	log L	Parameters
	Nucleotide GTR	40277.68	-20101.73	37
31	Model	AICC	log L	Parameters
	Nucleotide GTR	26366.44	-13146.07	37
39	Model	AICC	log L	Parameters
	Nucleotide GTR	14184.62	-7071.21	21
43	Model	AICC	log L	Parameters
	Nucleotide GTR	7258.82	-3597.94	31
48	Model	AICC	log L	Parameters
	Nucleotide GTR	22751.11	-11354.49	21
50	Model	AICC	log L	Parameters
	Nucleotide GTR	29304.29	-14625.05	27
58	Model	AICC	log L	Parameters
	Nucleotide GTR	41063.94	-20510.93	21
64	Model	AICC	log L	Parameters
	Nucleotide GTR	12504.37	-6233.08	19
66	Model	AICC	log L	Parameters
	Nucleotide GTR	49823.53	-24890.74	21

Supplementary table 9: FUBAR model selection

Supplementary material 7

GeneID	Annotation	FPKM	FPKM	log2 (fold change)	p-values
OPB36460	P-loop containing nucleoside triphosphate hydrolase Alpha/Beta hydrolase fold	1,70	10,84	2,67	5.00E-05
OPB45668	ANK-domain only	5,91	32,93	2,48	5.00E-05
OPB44758	NADH:flavin oxidoreductase / NADH oxidase family	3,66	0,74	-2,31	5.00E-05
OPB46923	**ANK-domain only**	**0,19**	**6,76**	**5,19**	**5.00E-05**
OPB40539	**ANK-domain only**	**0,90**	**12,66**	**3,82**	**5.00E-05**
OPB38680	**ANK-domain only**	**2,39**	**24,51**	**3,36**	**5.00E-05**
OPB44326	Protein kinase domain P-loop containing nucleoside triphosphate hydrolase	1,88	10,32	2,45	5.00E-05
OPB47160	NACHT domain	0,85	4,08	2,26	5.00E-05
OPB40466	NACHT domain Nucleoside phosphorylase domain	7,89	37,27	2,24	5.00E-05
OPB44812	Protein kinase domain	0,24	1,12	2,23	5.00E-05
OPB37151	Zinc finger, ZZ-type Alpha/Beta hydrolase fold,	11,87	52,07	2,13	5.00E-05
OPB37153	Peptidase S8/S53 domain	12,64	53,72	2,09	5.00E-05
OPB44870	Sigma domain on NACHT-NTPases	0,87	3,60	2,06	5.00E-05
OPB42534	ANK-domain only	2,64	0,66	-2,00	5.00E-05
OPB39582	ANK-domain only	12,76	2,75	-2,22	5.00E-05
OPB35975	NACHT domain Nucleoside phosphorylase domain	6,85	1,26	-2,45	5.00E-05
OPB42857	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain	3,60	0,58	-2,64	5.00E-05
OPB44958	ANK-domain only	1,58	0,23	-2,79	5.00E-05
OPB41799	P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain	4,40	0,60	-2,88	5.00E-05
OPB44602	**Nucleoside phosphorylase domain**	**20,79**	**0,74**	**-4,81**	**5.00E-05**
OPB42886	**P-loop containing nucleoside triphosphate hydrolase Nucleoside phosphorylase domain**	**2,50**	**0,03**	**-6,16**	**5.00E-05**

Supplementary table 10: Detailed RNAdeepseq results of confrontation of T. guizhouense *against itself*

GeneID	Annotation	FPKM	FPKM	log2 (foldchange)	p-values
OPB36460	P-loop containing nucleoside triphosphate hydrolase Alpha/Beta hydrolase fold	2,33	13,71	2,55	5.00E-05
OPB40005	Asparaginase, N-terminal	4,58	18,80	2,04	5.00E-05
OPB46923	**ANK-domain only**	**1,22**	**31,45**	**4,68**	**5.00E-05**
OPB46886	**ATPase, dynein-related, AAA domain**	**0,43**	**10,25**	**4,58**	**5.00E-05**
OPB40228	**P-loop containing nucleoside triphosphate hydrolase**	**1,28**	**16,79**	**3,71**	**5.00E-05**

	Nucleoside phosphorylase domain				
OPB44812	**Protein kinase domain**	**0,22**	**2,71**	**3,62**	**5.00E-05**
OPB37316	**von Willebrand factor, type A**	**5,74**	**49,35**	**3,10**	**5.00E-05**
OPB45571	ANK-domain only	0,75	4,58	2,62	5.00E-05
OPB36873	P-loop containing nucleoside triphosphate hydrolase	1,23	6,90	2,49	5.00E-05
OPB38680	ANK-domain only	3,76	20,72	2,46	5.00E-05
OPB42070	ANK-domain only	1,74	8,64	2,31	5.00E-05
OPB46656	NACHT domain Nucleoside phosphorylase domain	1,69	8,10	2,26	5.00E-05
OPB38898	P-loop containing nucleoside triphosphate hydrolase Putative serine esterase (DUF676) Alpha/Beta hydrolase fold	0,55	2,29	2,05	5.00E-05
OPB42501	NACHT domain P-loop containing nucleoside triphosphate hydrolase	3,11	12,52	2,01	5.00E-05
OPB44602	**Nucleoside phosphorylase domain**	**16,12**	**0,74**	**-4,44**	**5.00E-05**

Supplementary table 11: Detailed RNAdeepseq results of confrontation of T. guizhouense *against* F. oxysporum

GeneID	Annotation	FPKM	FPKM	log2 (foldchange)	p-values
OPB36460	**P-loop containing nucleoside triphosphate hydrolase** **Alpha/Beta hydrolase fold**	**2,05**	**19,54**	**3,25**	**5.00E-05**
OPB45668	**ANK-domain only**	**9,23**	**41,94**	**2,18**	**5.00E-05**
OPB45571	**ANK-domain only**	**0,90**	**28,14**	**4,97**	**5.00E-05**
OPB37316	**von Willebrand factor, type A**	**4,12**	**87,86**	**4,41**	**5.00E-05**
OPB40217	**Protein kinase domain**	**1,00**	**18,28**	**4,19**	**5.00E-05**
OPB38898	**P-loop containing nucleoside triphosphate hydrolase** **Putative serine esterase (DUF676)** **Alpha/Beta hydrolase fold**	**0,24**	**3,74**	**3,97**	**5.00E-05**
OPB40882	**SPRY domain** **P-loop containing nucleoside triphosphate hydrolase** **Concanavalin A-like lectin/glucanase domain** **Alpha/Beta hydrolase fold**	**0,76**	**11,21**	**3,88**	**5.00E-05**
OPB38680	**ANK-domain only**	**2,66**	**26,89**	**3,34**	**5.00E-05**
OPB36873	P-loop containing nucleoside triphosphate hydrolase	1,27	9,21	2,86	5.00E-05
OPB44326	Protein kinase domain P-loop containing nucleoside triphosphate hydrolase	2,05	14,42	2,81	5.00E-05
OPB40539	ANK-domain only	2,35	16,39	2,80	5.00E-05
OPB39071	NACHT domain	1,81	12,57	2,80	5.00E-05
OPB37151	Zinc finger, ZZ-type Alpha/Beta hydrolase fold,	12,00	75,08	2,65	5.00E-05
OPB42501	NACHT domain P-loop containing nucleoside triphosphate hydrolase	2,55	15,39	2,59	5.00E-05
OPB37153	Peptidase S8/S53 domain	12,11	70,30	2,54	5.00E-05
OPB47160	NACHT domain	0,84	4,66	2,48	5.00E-05

OPB46656	NACHT domain Nucleoside phosphorylase domain	1,37	7,31	2,42	5.00E-05
OPB42070	ANK-domain only	1,60	8,07	2,33	5.00E-05
OPB46030	ANK-domain only	2,12	9,69	2,19	5.00E-05
OPB40809	ANK-domain only	6,01	26,73	2,15	5.00E-05
OPB42080	NWD NACHT-NTPase, N-terminal	3,62	15,96	2,14	5.00E-05
OPB40466	NACHT domain Nucleoside phosphorylase domain	10,20	43,07	2,08	5.00E-05
OPB42886	**P-loop containing nucleoside triphosphate hydrolase** **Nucleoside phosphorylase domain**	**2,82**	**0,34**	**-3,07**	**5.00E-05**
OPB44602	**Nucleoside phosphorylase domain**	**26,51**	**1,42**	**-4,22**	**5.00E-05**

Supplementary table 12: Detailed RNAdeepseq results of confrontation of T. guizhouense *against* T. harzianum

GeneID	Annotation	FPKM	FPKM	log2 (foldchange)	p-values
jgi479222	**ANK-domain only**	**0,30**	**21,10**	**6,14**	**5.00E-05**
jgi145036	**ANK-domain only**	**0,56**	**15,65**	**4,80**	**5.00E-05**
jgi505476	**von Willebrand factor, type A**	**18,39**	**449,74**	**4,61**	**5.00E-05**
jgi476900	**ANK-domain only**	**5,23**	**82,10**	**3,97**	**5.00E-05**
jgi44122	**ANK-domain only**	**1,56**	**12,87**	**3,05**	**5.00E-05**
jgi67504	ANK-domain only	17,28	129,20	2,90	5.00E-05
jgi270473	Sigma domain on NACHT-NTPases	25,25	156,82	2,63	5.00E-05
jgi508483	Nucleoside phosphorylase domain	2,18	13,07	2,58	5.00E-05
jgi551129	ANK-domain only	2,74	16,29	2,57	5.00E-05
jgi515791	ANK-domain only	0,42	2,45	2,55	5.00E-05
jgi70585	ANK-domain only	4,22	20,22	2,26	5.00E-05
jgi488896	NWD NACHT-NTPase, N-terminal	3,91	16,93	2,12	5.00E-05
jgi427104	NACHT domain	0,43	1,84	2,11	5.00E-05
jgi39922	ANK-domain only	2,99	12,82	2,10	5.00E-05
jgi335626	P-loop containing nucleoside triphosphate hydrolase NACHT domain	3,01	12,83	2,09	5.00E-05
jgi78458	ANK-domain only	5,43	22,36	2,04	5.00E-05

Supplementary table 13: Detailed RNAdeepseq results of confrontation of T. harzianum *against* T. guizhouense

GeneID	Annotation	FPKM	FPKM	log2 (foldchange)	p-values
jgi479222	**ANK-domain only**	**0,38**	**9,25**	**4,59**	**5.00E-05**
jgi505476	**von Willebrand factor, type A**	**17,10**	**210,39**	**3,62**	**5.00E-05**
jgi551129	**ANK-domain only**	**1,90**	**18,12**	**3,25**	**5.00E-05**
jgi145036	**ANK-domain only**	**0,86**	**7,65**	**3,15**	**5.00E-05**
jgi63203	P-loop containing nucleoside triphosphate hydrolase	4,14	30,64	2,89	5.00E-05
jgi78458	ANK-domain only	4,23	25,55	2,60	5.00E-05
jgi70585	ANK-domain only	3,04	15,83	2,38	5.00E-05

jgi335626	P-loop containing nucleoside triphosphate hydrolase NACHT domain	2,68	12,89	2,27	5.00E-05
jgi508483	Nucleoside phosphorylase domain	1,93	8,56	2,15	5.00E-05
jgi270473	Sigma domain on NACHT-NTPases	28,02	120,33	2,10	5.00E-05
jgi476900	ANK-domain only	9,66	41,17	2,09	5.00E-05

Supplementary table 14: Detailed RNAdeepseq results of confrontation of T. harzianum *against* F. oxysporum

GeneID	Annotation	FPKM	FPKM	log2 (foldchange)	p-values
jgi479222	**ANK-domain only**	**0,22**	**26,89**	**6,96**	**5.00E-05**
jgi505476	**von Willebrand factor, type A**	**12,27**	**479,77**	**5,29**	**5.00E-05**
jgi145036	**ANK-domain only**	**0,47**	**9,10**	**4,27**	**5.00E-05**
jgi476900	**ANK-domain only**	**4,47**	**79,59**	**4,15**	**5.00E-05**
jgi515791	**ANK-domain only**	**0,21**	**3,31**	**3,95**	**5.00E-05**
jgi67504	**ANK-domain only**	**15,22**	**164,89**	**3,44**	**5.00E-05**
jgi270473	Sigma domain on NACHT-NTPases,	19,02	125,03	2,72	5.00E-05
jgi551129	ANK-domain only	1,69	10,65	2,65	5.00E-05
jgi44122	ANK-domain only	1,44	8,05	2,48	5.00E-05
jgi459023	WD40-repeat-containing domain Nucleoside phosphorylase domain	0,86	4,49	2,38	5.00E-05
jgi335626	P-loop containing nucleoside triphosphate hydrolase NACHT domain,	2,21	9,51	2,10	5.00E-05
jgi70585	ANK-domain only	3,07	12,79	2,06	5.00E-05
jgi98872	ANK-domain only	0,92	0,21	-2,12	5.00E-05
jgi505734	ANK-domain only	164,48	33,39	-2,30	5.00E-05
jgi479284	ANK-domain only	32,39	4,67	-2,80	5.00E-05

Supplementary table 15: Detailed RNAdeepseq results of confrontation of T. harzianum *against itself*