

Genome Sequencing and Comparative Analysis of the Biocontrol Agent *Trichoderma harzianum sensu stricto* TR274

Andrei S. Steindorff^{1,2}, Eliane F. Noronha¹, Cirano J. Ulhoa¹, Asaf A. Salamov², Sajeet Haridas², Robert W. Riley², Irina S. Druzhinina³, Christian P. Kubicek³, Igor Grigoriev¹

¹Universidade de Brasilia- Brasilia, DF, Brazil

²DOE Joint Genome Institute, Walnut Creek, CA,

³Institute of Chemical Engineering, Vienne University of Technology, Vienna, Austria

March 2015

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Genome sequencing and comparative analysis of the biocontrol agent *Trichoderma harzianum sensu stricto* TR274

Steindorff, AS^{*1,2}; Noronha, EF²; Ulhoa, CJ²; Kuo, A¹; Salamov, AA¹; Haridas, S¹; Riley, RW¹; Druzhinina, IS³; Kubicek, CP³; Grigoriev, IV¹

¹Joint Genome Institute – Walnut Creek – CA – USA; ²Universidade de Brasilia – Brasilia – DF – Brazil; ³Institute of Chemical Engineering Vienna University of Technology – Vienna - Austria

*assteindorff@lbl.gov



Abstract

Biological control is a complex process which requires many mechanisms and a high diversity of biochemical pathways. The species of *Trichoderma harzianum* are well known for their biocontrol activity against many plant pathogens. To gain new insights into the biocontrol mechanism used by *T. harzianum*, we sequenced the isolate TR274 genome using Illumina. The assembly was performed using AllPaths-LG with a maximum coverage of 100x. The assembly resulted in 2282 contigs with a N50 of 37033bp. The genome size generated was 40.8 Mb and the GC content was 47.7%, similar to other *Trichoderma* genomes. Using the JGI Annotation Pipeline we predicted 13,932 genes with a high transcriptome support. CEGMA tests suggested 100% genome completeness and 97.9% of RNA-SEQ reads were mapped to the genome. The phylogenetic comparison using orthologous proteins with all *Trichoderma* genomes sequenced at JGI, corroborates the *Trichoderma* (*T. asperellum* and *T. atroviride*), Longibrachiatum (*T. reesei* and *T. longibrachiatum*) and Pachibasium (*T. harzianum* and *T. virens*) section division described previously. The comparison between two *Trichoderma harzianum* species suggests a high genome similarity but some strain-specific expansions. Analyses of the secondary metabolites, CAZymes, transporters, proteases, transcription factors were performed. The Pachibasium section expanded virtually all categories analyzed compared with the other sections, specially Longibrachiatum section, that shows a clear contraction. These results suggests that these proteins families have an important role in their respective phenotypes. Future analysis will improve the understanding of this complex genus and give some insights about its lifestyle and the interactions with the environment.

Objectives

- Perform assembly and annotation of the *T. harzianum* TR274 genome;
- Compare all *Trichoderma* genomes available on JGI structural and functionally;
- Analyze the expansions and contractions of gene families involved on their respective lifestyles.
- Narrow possible variations between two isolates of *Trichoderma harzianum sensu stricto*.

Conclusions

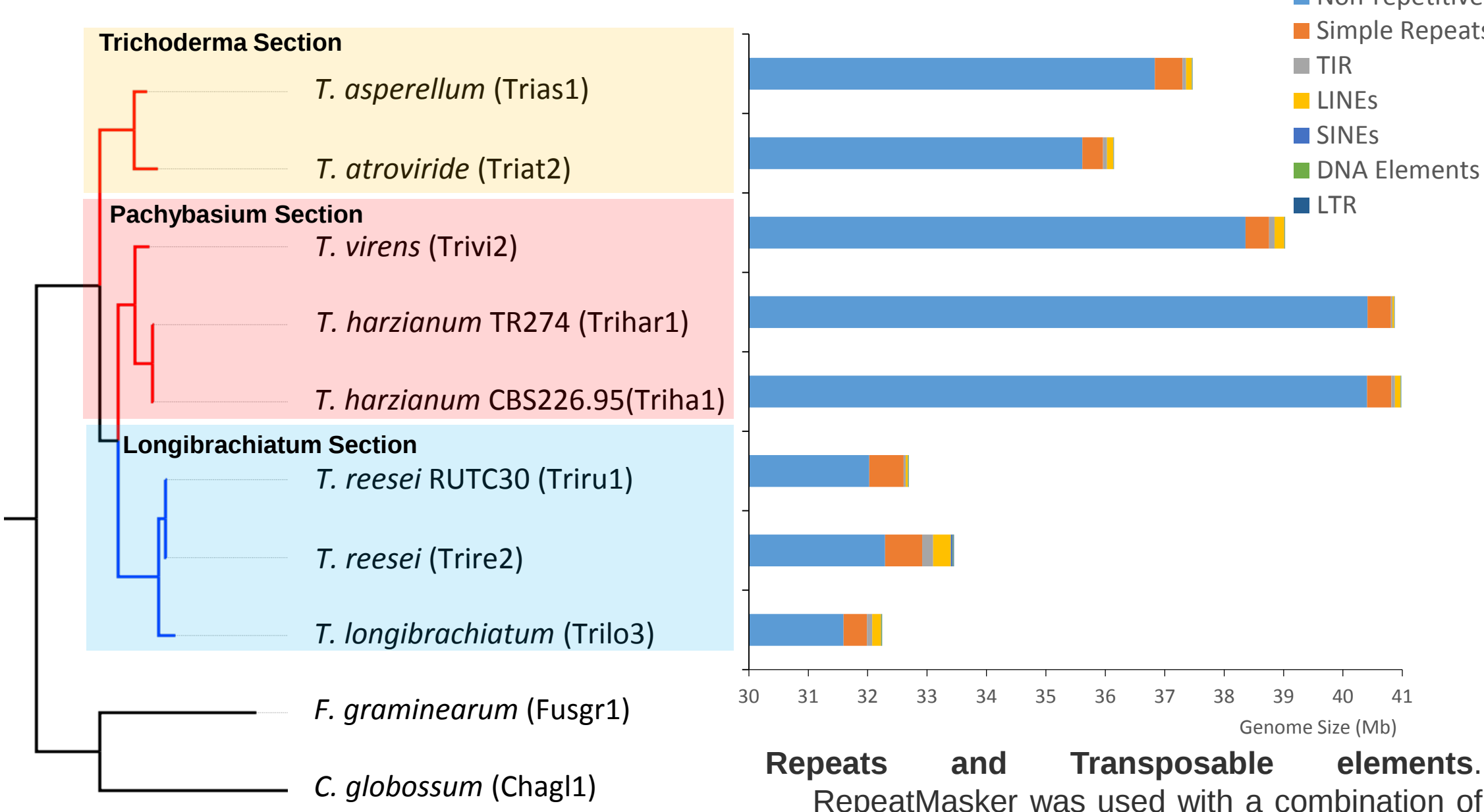
- Corroboration of the ancestral mycoparasite lifestyle idea previously described by Kubicek *et al*, 2011;
- *Trichoderma* and Pachibasium sections has different mycoparasite approaches reflected by the gene composition (i.e. Secondary metabolism expansions on Pachibasium section);
- Longibrachiatum section lost a high number of gene families leading to a different lifestyle.
- Even close related strains like *T. harzianum* TR274 and CBS226.95, have consistent differences (i.e. secondary metabolites clusters)

References

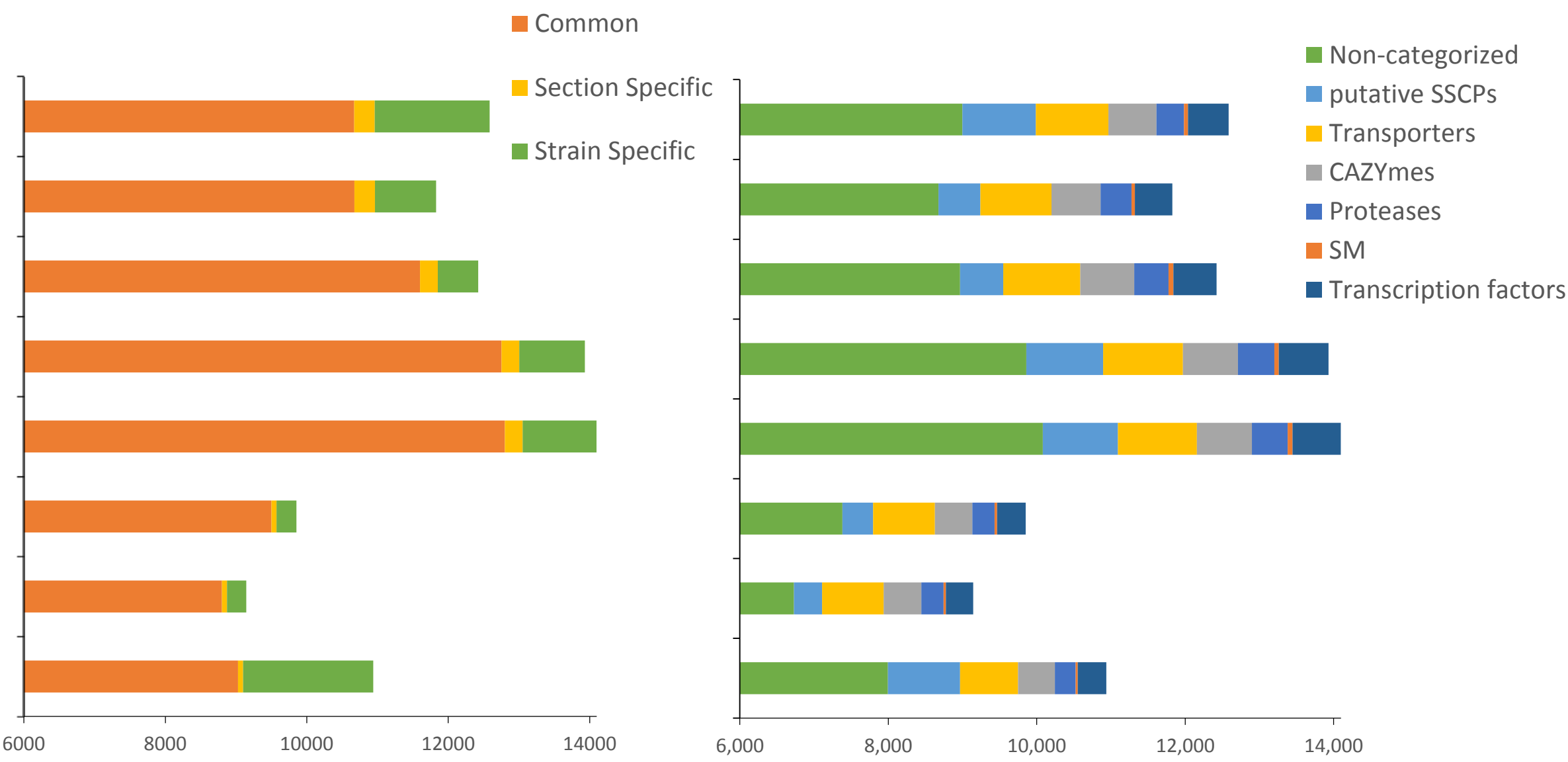
- Kubicek CP, Herrera-Estrella A, Martinez DA, Druzhinina IS, *et al*. Comparative genome sequence analysis underscores mycoparasitism as the ancestral lifestyle of *Trichoderma*. Genome Biology 2011.

Results

Trichoderma Phylogeny and Genome Features



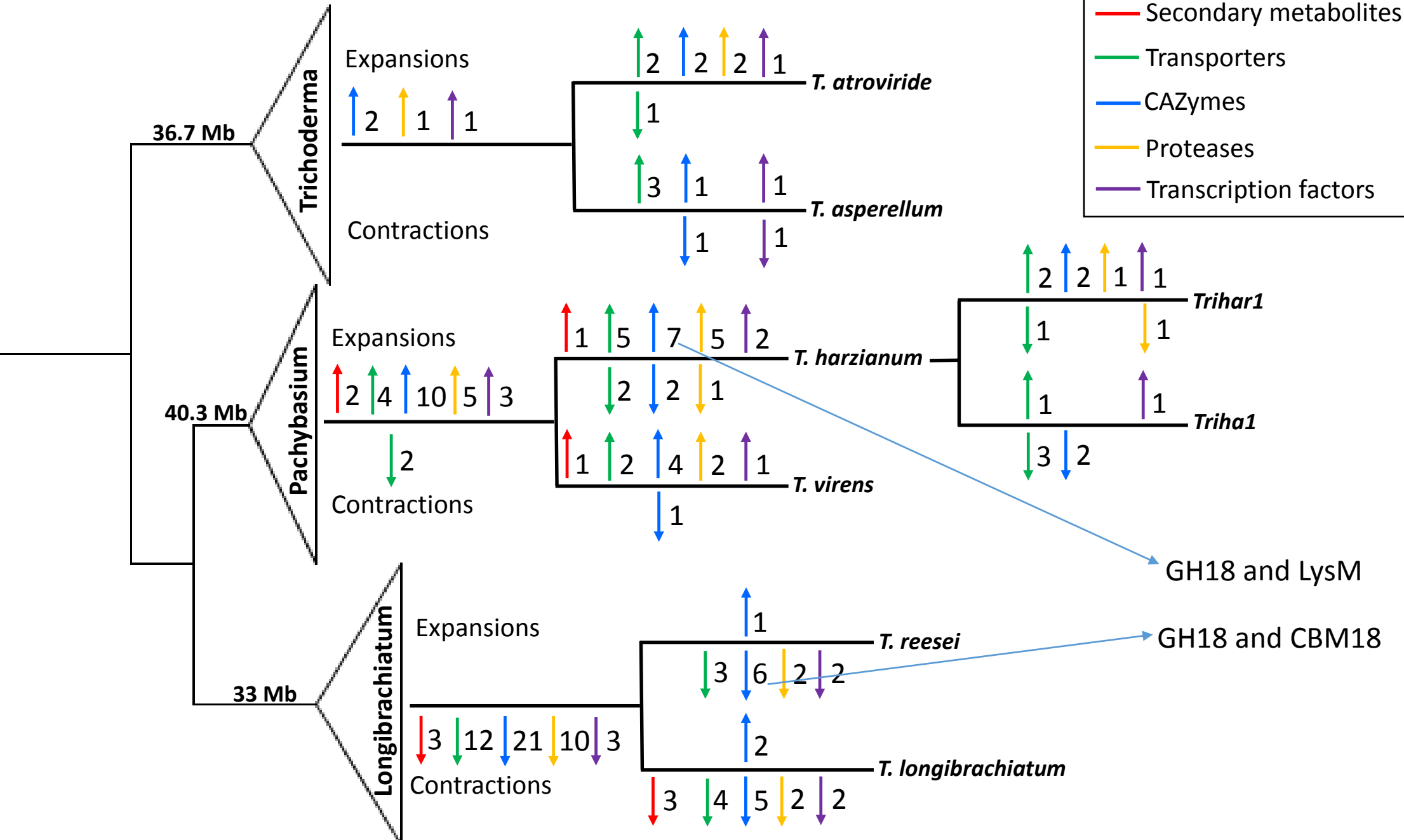
Genome phylogeny of *Trichoderma* genus. Maximum likelihood phylogram based on the alignment of amino acid sequences from 200 orthologous syntenic proteins. All bootstraps values for all nodes are 100.



Predicted genes, broken down by level of conservation. Division based on Best Bidirectional Blast analysis.

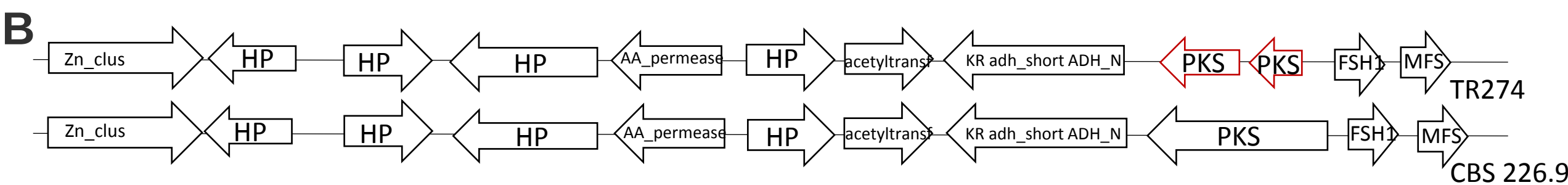
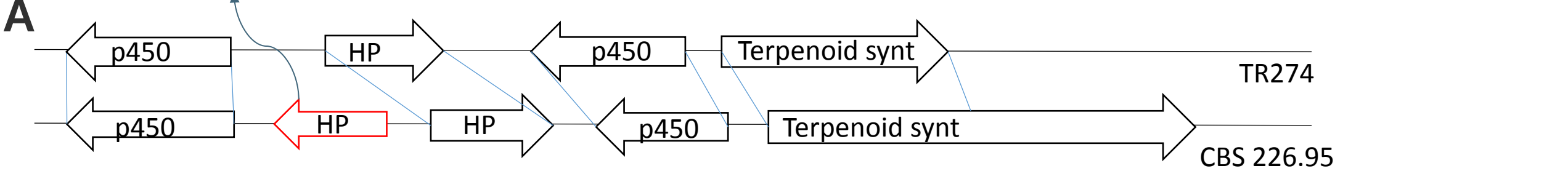
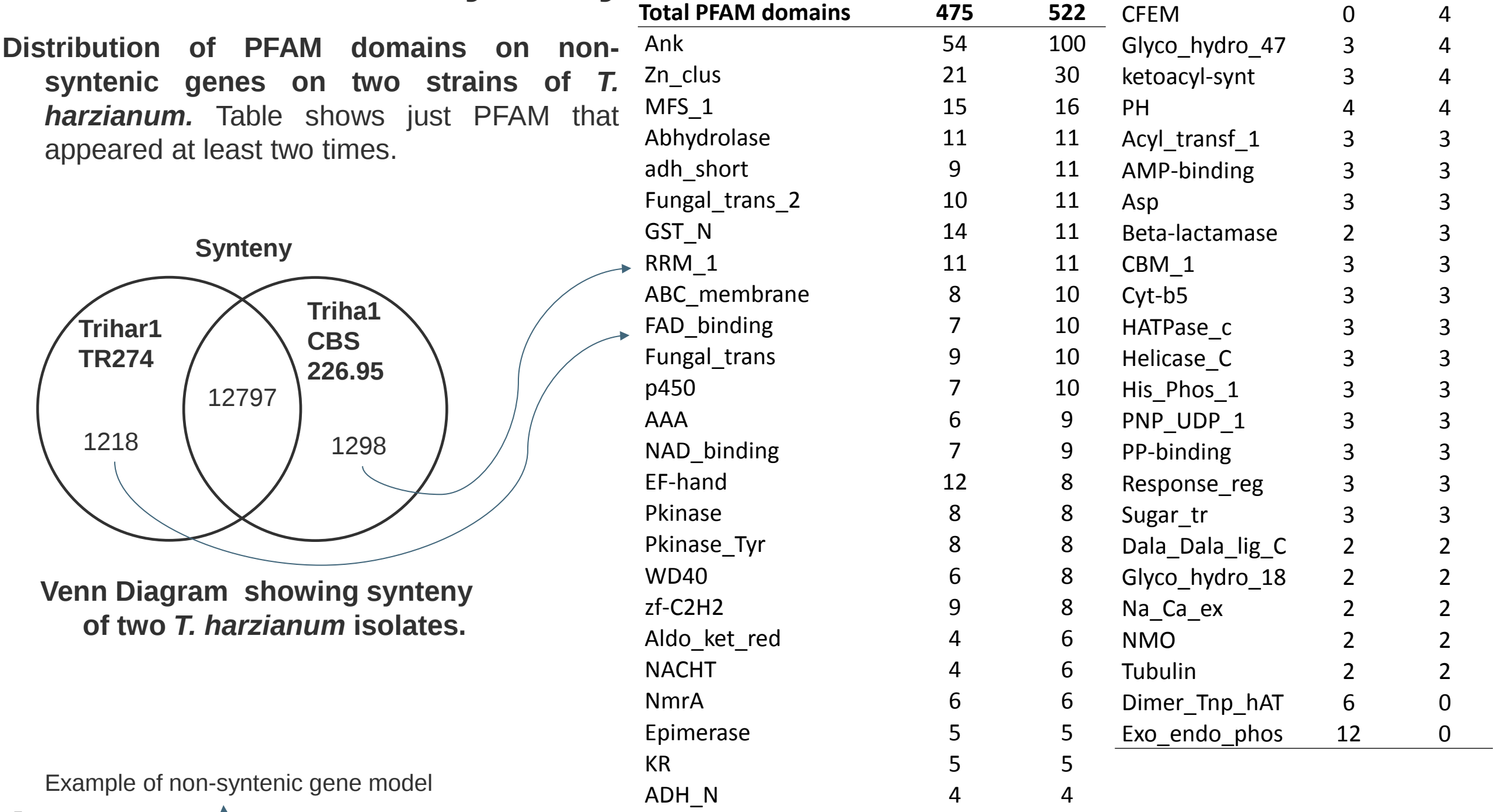
Effectors distributed by functions. Proteins were categorized based on homology with databases build specifically for each category.

CAFE analysis



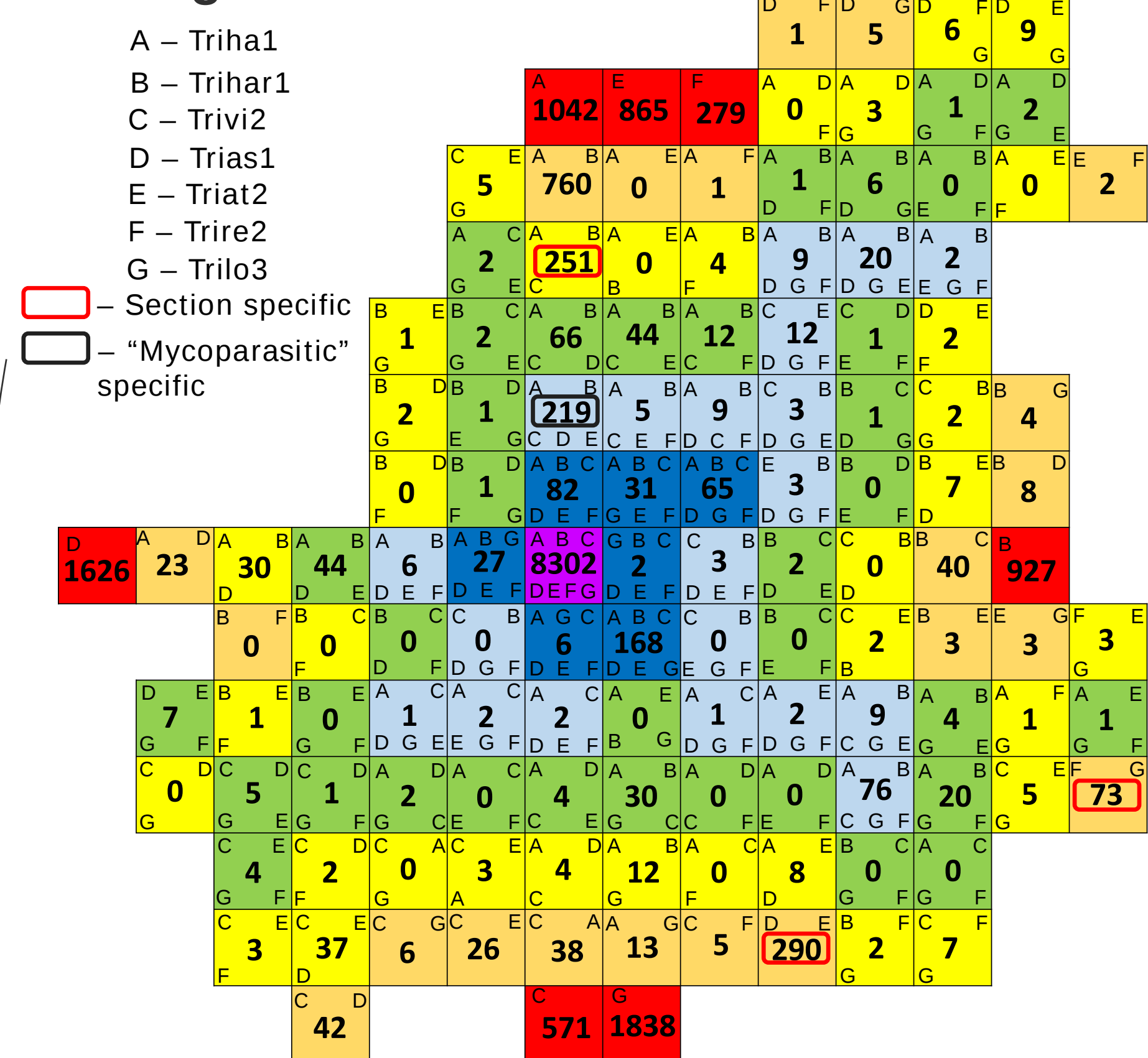
CAFE analysis showing expansions and contractions on *Trichoderma* genus. Number of gene families contracted and expanded on each node analyzed using CAFE software. Up arrow means expansions and down arrow means contractions. Numbers on left of sections means average of genome sizes within that branch. Blue arrows show two examples of gene families expanded on *T. harzianum* and contracted on *T. reesei* (Chitinases and Chitin binding domains). Values on branches means average of genome size in Mb.

Two *T. harzianum* synteny



Syntenic variation in two different secondary metabolite clusters on TR274 and CBS 226.95 *T. harzianum* isolates. A – Terpenoid synthase cluster showing a truncated gene and the absence of a gene model on the TR274 strain. B – Polyketide synthase (PKS) cluster showing two different PKS gene models on the same region.

Orthologs Distribution



Distribution of genes among all 7 *Trichoderma* strains analysed and comparison between two *T. harzianum* strains. Distribution of orthologs based on Best Bidirectional Blast hits. Each square corresponds to one intersection of the multidimensional 7-way Venn Diagram. Colors means the numbers of strains within the intersection: Red – 1, Beige – 2, Yellow – 3, Green – 4, Light Blue – 5, Dark Blue – 6, Purple – 7.

Distribution of PFAM domains present exclusively on all 5 mycoparasitic strains analyzed.

Orthologs present on *Trichoderma* and Pachibasium sections and absent on Longibrachiatum section. Table shows just PFAM that appeared at least two times.

PFAM Domains	Trias1	Triat2	Triha1	Trihar1	Trivi2
Total PFAM domains	103	102	98	99	103
Abhydrolase_6	14	14	14	13	13
DUF1295	14	13	11	11	11
Ankyryn	6	10	12	10	10
Acetyltransferase	5	6	5	5	6
Glycoside Hydrolase	4	4	3	3	4
GST	4	4	5	5	4
Methyltransferase	4	6	5	5	8
Tannase	4	3	2	1	2
2OG-Fell_Oxy	3	2	2	2	2
CHAT	3	1	1	1	1
Cu-oxidase	3	3	2	2	3
TPR	0	0	0	0	7
Isochorismatase	3	2	2	2	2
NUDIX	3	3	3	3	2
Pectinesterase	3	1	1	1	2
Tyrosinase	3	1	1	1	1
Bac_rhamnosid	2	2	2	2	2
bZIP_1	2	2	2	2	2
DeoC	2	2	2	2	2
Lipase_GDSL	2	2	2	2	2
NAD_binding_10	2	2	1	1	2
NmrA	2	2	1	1	2
Ricin_B_lectin	2	2	2	2	2
Transferase	2	2	2	2	2
UbiD	2	2	1	1	1

Acknowledgments

