Lawrence Berkeley National Laboratory
Recent Work

Title

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

Permalink

https://escholarship.org/uc/item/17z281q8

Authors

Steindorff, Andrei S.
Noronha, Elilane F.
Ulhoa, Cirano J.
et al.

Publication Date

2017-12-07

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

Andrei S. Steindorff[1,2], Eliane F. Noronha[1], Cirano J. Ulhoa[1], Asaf A. Salamov[2], Sajeet Haridas[2], Robert W. Riley[2], Irina S. Druzhinina[3], Christian P. Kubicek[3], Igor Grigoriev[1]

[1]Universidade de Brasilia- Brasilia, DF, Brazil
[2]DOE Joint Genome Institute, Walnut Creek, CA,
[3]Institute of Chemical Engineering, Vienne University of Technology, Vienna, Austria

March 2015

DISCLAIMER

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

Steindorff, AS*[1,2]; Noronha, EF[2]; Ulhoa, CJ[2]; Kuo, A[1]; Salamov, AA[1]; Haridas, S[1]; Riley, RW[1]; Druzhinina, IS[3]; Kubicek, CP[3]; Grigoriev, IV[1]

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

*assteindorff@lbl.gov
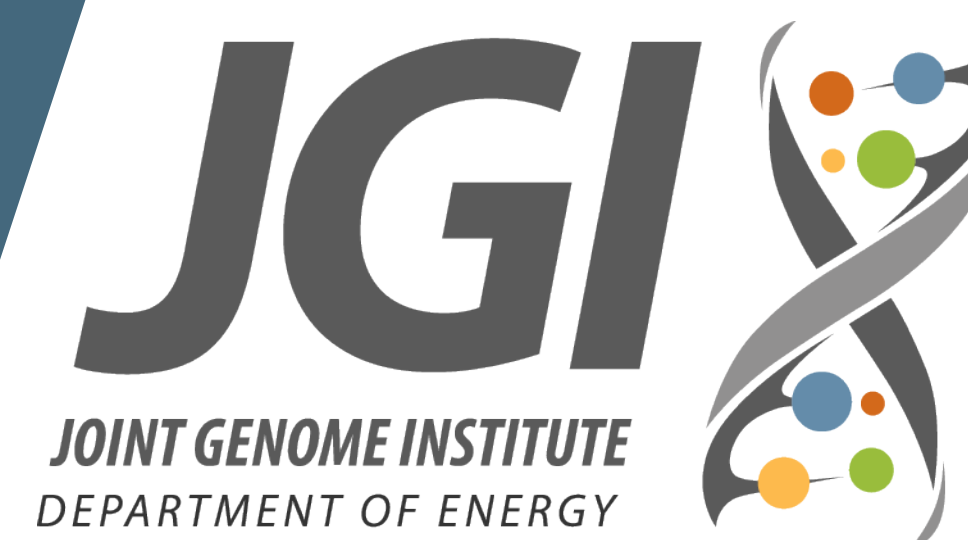
Abstract

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 Trichodema 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* stricto sensu.
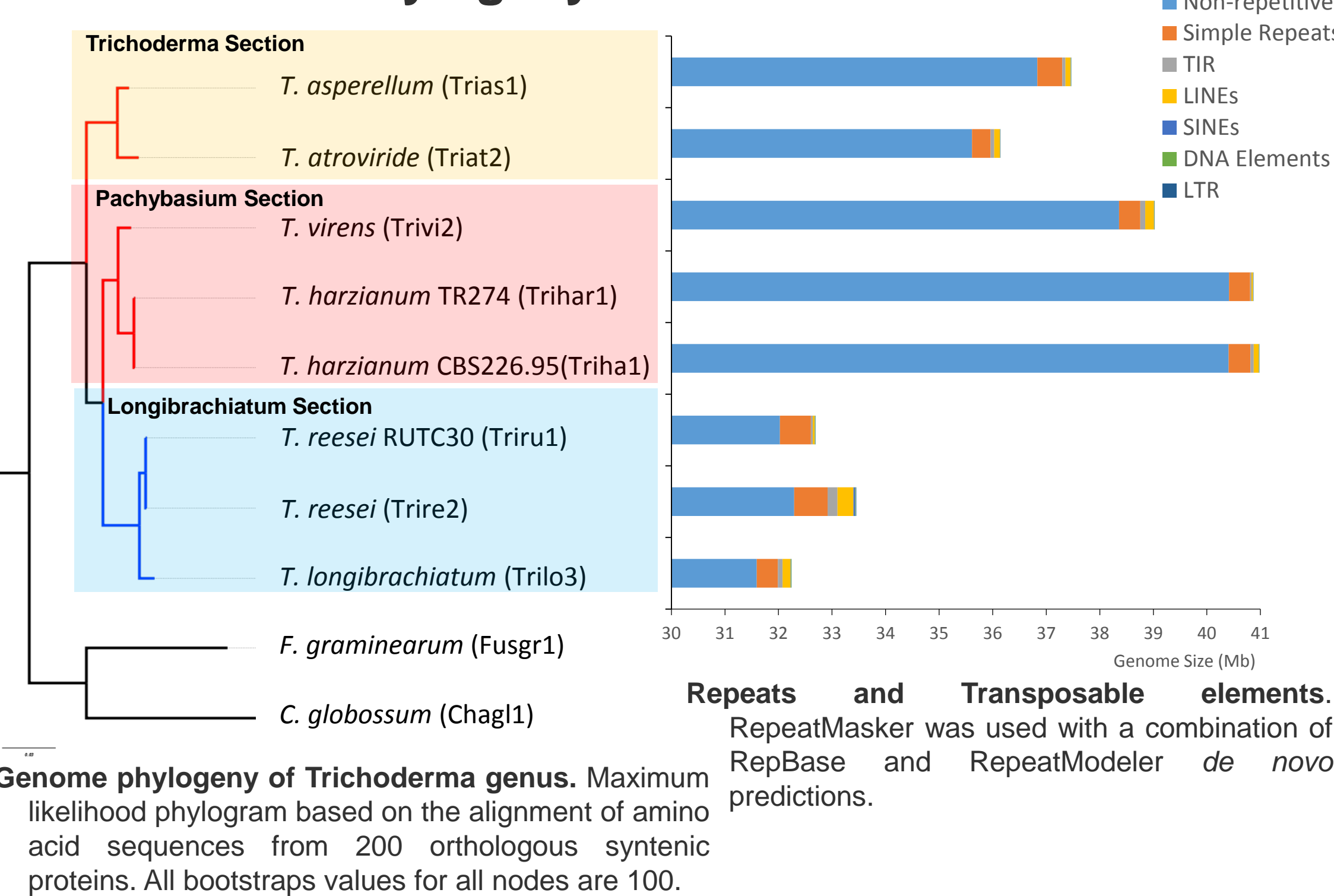
Conclusions

- Corroboration of the ancestral mycoparasite lifestyle idea previously described by Kubicek *et al*, 2011;
- Trichoderma and Pachibasium sections has diferent 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)
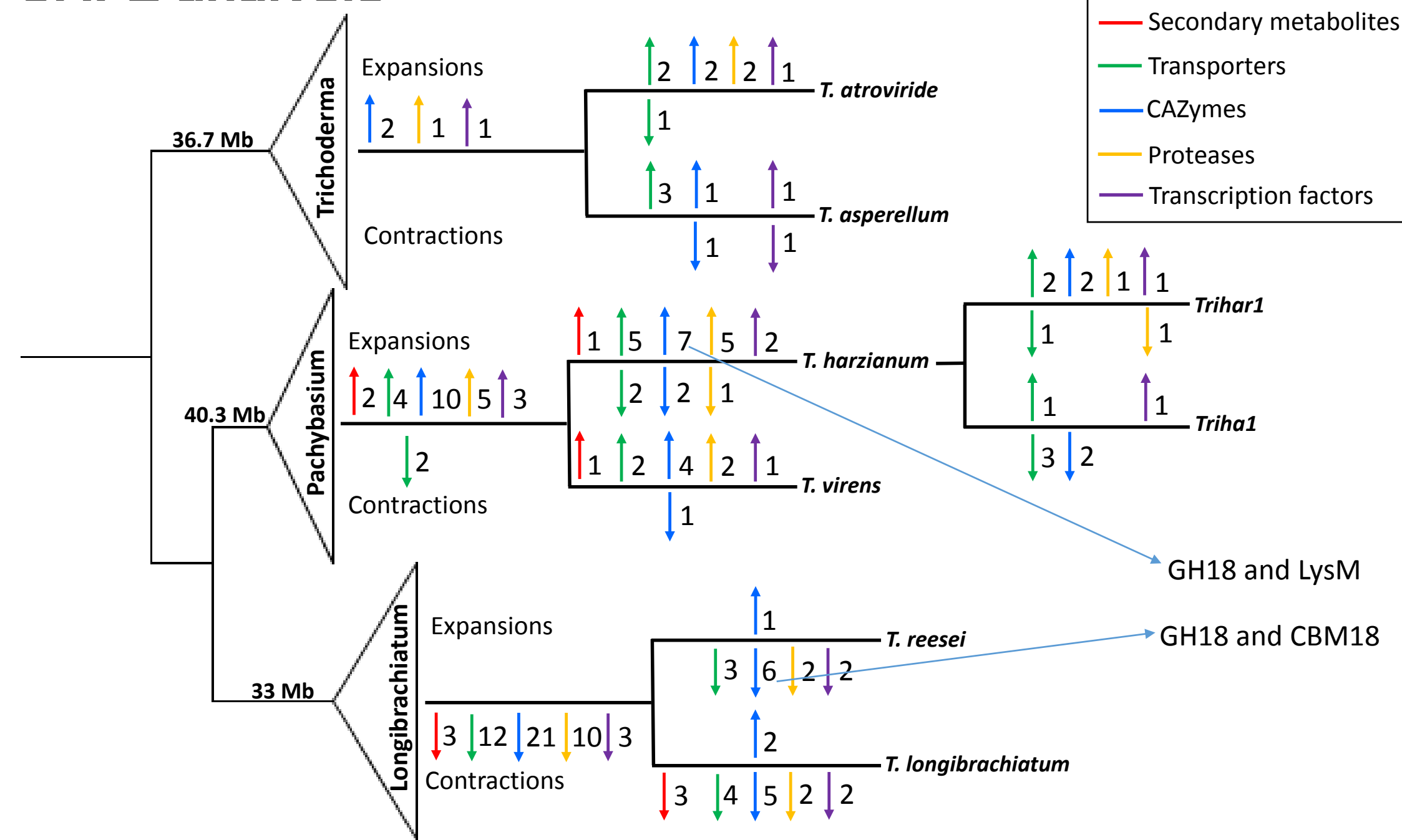
References

- 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.

Repeats and Transposable elements. RepeatMasker was used with a combination of RepBase and RepeatModeler *de novo* predictions.

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.
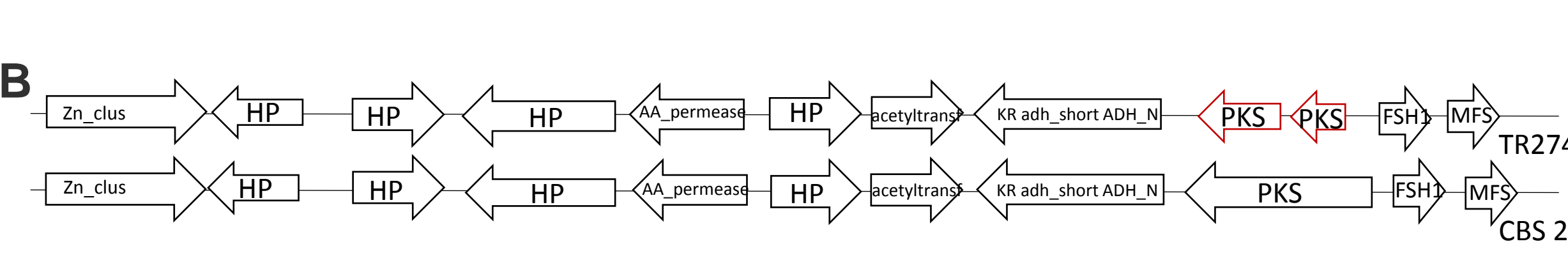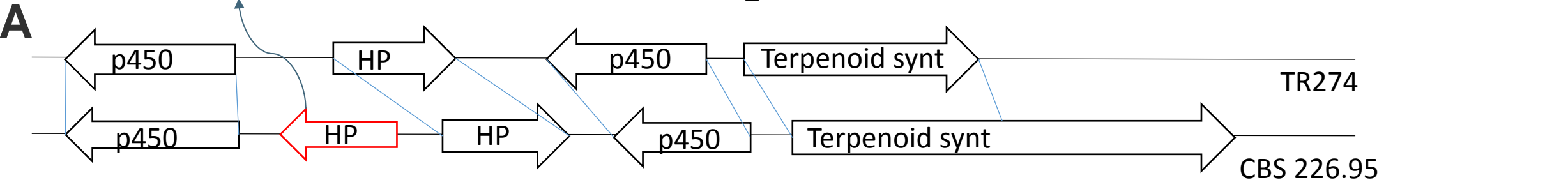
CAFE analysis

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.
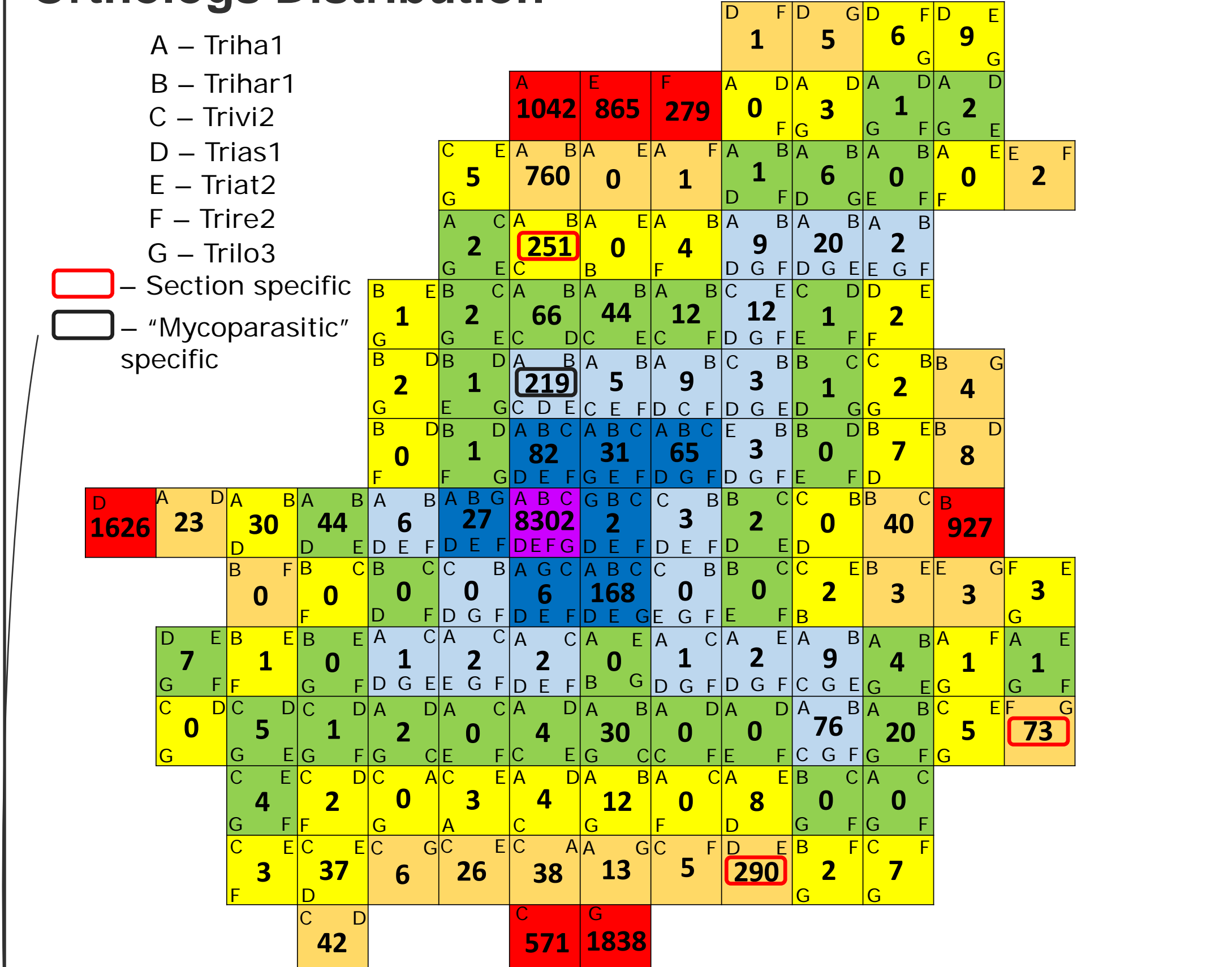
Two *T. harzianum* synteny

Distribution of PFAM domains on non-syntenic genes on two strains of *T. harzianum*. Table shows just PFAM that appeared at least two times.

Venn Diagram showing synteny of two *T. harzianum* isolates.

PFAM Domains	Triha1	Trihar1
Total PFAM domains	**475**	**522**
Ank	54	100
Zn_clus	21	30
MFS_1	15	16
Abhydrolase	11	11
adh_short	9	11
Fungal_trans_2	10	11
GST_N	14	11
RRM_1	11	11
ABC_membrane	8	10
FAD_binding	7	10
Fungal_trans	9	10
p450	7	10
AAA	6	9
NAD_binding	7	9
EF-hand	12	8
Pkinase	8	8
Pkinase_Tyr	8	8
WD40	6	8
zf-C2H2	9	8
Aldo_ket_red	4	6
NACHT	4	6
NmrA	6	6
Epimerase	5	5
KR	5	5
ADH_N	4	4

PFAM Domains	Triha1	Trihar1
ADH_zinc_N	3	4
CFEM	0	4
Glyco_hydro_47	3	4
ketoacyl-synt	3	4
PH	4	4
Acyl_transf_1	3	3
AMP-binding	3	3
Asp	3	3
Beta-lactamase	2	3
CBM_1	3	3
Cyt-b5	3	3
HATPase_c	3	3
Helicase_C	3	3
His_Phos_1	3	3
PNP_UDP_1	3	3
PP-binding	3	3
Response_reg	3	3
Sugar_tr	3	3
Dala_Dala_lig_C	2	2
Glyco_hydro_18	2	2
Na_Ca_ex	2	2
NMO	2	2
Tubulin	2	2
Dimer_Tnp_hAT	6	0
Exo_endo_phos	12	0

Example of non-syntenic gene model

A

B

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

A – Triha1
B – Trihar1
C – Trivi2
D – Trias1
E – Triat2
F – Trire2
G – Trilo3

🔲 – Section specific
⬛ – "Mycoparasitic" specific

Distribution of genes among all 7 Trichodermas 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 Veen 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 Pachybasium sections and absent on Longibrachiantum 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
Ankryrn	6	10	12	10	10
Acetyltransferase	5	5	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-FeII_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