Structural and functional analysis of a *Trichoderma virens* GAPDH associated with a secondary-metabolism related gene cluster

By

Shikha Pachauri

LIFE01201404005

Bhabha Atomic Research Centre, Mumbai

A thesis submitted to the Board of Studies in Life Sciences
In partial fulfillment of requirements
For the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

December 2020

Homi Bhabha National Institute[1]

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Shikha Pachauri entitled "Structural and functional analysis of a *Trichoderma virens* GAPDH associated with a secondary-metabolism related gene cluser" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Dr. S.J. Jambhulkar	
Guide / Convener - Dr. Prasun K. Mukherjee	
Co-guide - Dr. Vinay Kumar	
Examiner - Dr. Birinchi K. Sarma	
Member 1 - Dr. Ajay Saini	
Member 2 - Dr. J. Souframanien	

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 10.12.20

Place: Mumbai

Signature
Co-guide (if any)

Signature
Guide

[1] This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

7/12/2020

Shikha Pachauri

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

.

7/12/2020

Shikha Pachauri

List of Publications Arising from the Thesis

Journal

1. **Shikha Pachauri**, Suchandra Chatterjee, Vinay Kumar & Prasun K. Mukherjee (2019). A dedicated glyceraldehyde-3-phosphate dehydrogenase is involved in the biosynthesis of volatile sesquiterpenes in *Trichoderma virens* -evidence for the role of a fungal GAPDH in secondary metabolism. ***Current Genetics***, https://doi.org/10.1007/s00294-018-0868-y.

2. **Shikha Pachauri**, Pramod D. Sherkhane, Vinay Kumar & Prasun K. Mukherjee (2020). Whole genome sequencing reveals major deletions in the genome of M7, a gamma-ray induced mutant of *Trichoderma virens* that is repressed in conidiation, secondary metabolism and mycoparasitism. ***Frontiers in Microbiology***, https://doi.org/10.3389/fmicb.2020.01030.

3. **Shikha Pachauri**, Gagan D. Gupta, Prasun K. Mukherjee and Vinay Kumar (2020). Expression of a heptelidic acid-insensitive recombinant GAPDH from *Trichoderma virens*, and its biochemical and biophysical characterization. ***Protein Expression and Purification***, https://doi.org/10.1016/j.pep.2020.105697.

Book Chapter

1. **Shikha Pachauri,** Pramod D. Sherkhane and Prasun K. Mukherjee (2019) Secondary Metabolism in *Trichoderma*: Chemo- and Geno-Diversity. Volume 2. Soil & Agroecosystems. Microbial diversity in ecosystem sustainability and biotechnological applications. Publisher: ***Springer***, https://doi.org/10.1007/978-981-13-8487-5_17.

Symposia/Conference

1. "Role of a GAPDH in secondary metabolism in *Trichoderma virens*" presented at International Conference of *Trichoderma* and *Gliocladium* (TG2016) "Principles to Practice" in Nagpur, India (2016).

   (Received "**Best Young Researcher Award**")

2. "Fungal GAPDHs: role beyond glycolysis" presented at European fungal genetics Conference (ECFG14) in Haifa, Israel (2018).

   (Received "**FEMS-Early Career Scientist Award**")

3. "Detection of gene deletions in a mutant strain of *Trichoderma virens* using whole genome sequencing" presented at 88th The Society of Biological Chemists (India) (SBCI) Meet, Mumbai (2019).

7/12/2020

Shikha Pachauri

Publications from other topics

1. Gopi Muthukathan, Poulomi Mukherjee, Darshana Salaskar, **Shikha Pachauri**, Himanshu Tak, Thumballi Ganapathi, Prasun K. Mukherjee (2020). Secretome of *Trichoderma virens* induced by banana roots-identification of novel fungal proteins for enhancing plant defense. ***Physiological and Molecular Plant Pathology***, 10.1016/j.pmpp.2020.101476.

2. Gurjeet Kaur, Syed Azmal Ali, **Shikha Pachauri**, Dhruba Malakar, Jai Kaushik, Ashok Kumar Mohanty, Sudarshan Kumar (2017). Buffalo Leukemia Inhibitory Factor Induces Differentiation and Dome-Like Secondary Structures in COS-1 Cells. ***Cytogenetic Genome Res***, https://doi.org/10.1159/000465507.

7/12/2020

Shikha Pachauri

Dedicated to Almighty, my Mummy, Papa, Shekhar, Chiku and Rushit

ACKNOWLEDGEMENTS

In eleventh standard, I decided to be a RESEARCHER and after all these years while writing my acknowledgment for the doctorate thesis I am feeling so blissful. These 6 years journey of my Ph.D. has introduced me to many genuine and humble people. **Dr. Prasun Kumar Mukherjee**, Section Head, Nuclear Agriculture and Biotechnology Division, was not only my guide but also a father figure for me. I consider myself fortunate indeed to have a mentor like him who always encouraged me and allowed me to explore myself as a researcher. He not only shared his limitless knowledge with me but also guided me to be a nice person at the end of the day. The best reward a student can get from his guide is confidence and his confidence in me was enough to keep me energetic, motivated and focused throughout this journey. In this life, I may get another PhD but I guarantee that I would never get a guide like him.

I express my sincere thanks to my co-guide, **Dr. Vinay Kumar**, Ex-Head, RB&HSD, who always supported and provided valuable guidance. He always inspired me to work hard and efficiently at any hour of the day. Whenever I lost track, he was there to guide me towards the right path. I am thankful to Dr. Sanjay J Jambhulkar, Chairman of my doctoral committee (NA&BTD) for being positive and responsive towards me. I am also grateful to other doctoral committee members, Dr. Ajay Saini (MBD) and Dr. J. Souframanien (NA&BTD) for their critical and helpful suggestions every year towards my work progress. Special thanks to Ex. Chairperson of my doctoral committee, Dr. K. S. Reddy (NA&BTD), who constantly supported and encouraged me. My sincere thanks to Dr. Hema Rajaram, Dean, Life Sciences for her academic guidance and support.

I would like to thank my lab mates Mr. Pramod Sherkhane, Mr. Chavan, Dr. Ravindra Bansal, Dr. Poulomi Mukherjee, Dr. Sayaji Mehtre, Dr. Gopi Muthukathan, Dr. Darshana Salaskar, Mr. Tipre and Mr. Aniket Pokle for their constant support. My sincere thanks to Dr. Gagandeep Gupta (RB&HSD), Dr. Suchandra Chatterjee (FTD), Dr. Sachin Gupta (FTD), Dr. Y. S. Rajpurohit (MBD), Dr. R. Sashidhar (FTD), Prashant Mishra (FTD). I am also thankful to Dr. Kaushik Banerjee and Dr. Zareen Khan from ICAR-National Research Centre for Grapes, Pune for providing help with LC/MS analysis.

I acknowledge the Department of Atomic Energy, Bhabha Atomic Research Centre for financial assistance during my Ph.D

I would like to thank my seniors (Dr. Mahima Sharma, Dr. Neha Paraswani, Dr. Bharati Pandey) and juniors (Reema, Shweta, Shruti) for their support. My deepest

appreciation to my friends Shridhar Pranjape and Kirti Sawant for being part of my happiness and disappointments in BARC. Gayatri and Gunjan, thank you girls for making my hostel days peaceful and a happy place.

Dr. Ganesh Kumar Maurya, I am lucky and genuinely grateful to lord for having you as my senior in BARC. He always put faith in me whenever I felt low and disheartened as things were not working out. He is one of those genuine souls who taught me to help and support people selflessly.

Suyash Pandey, my only friend away from Mumbai, thank you much for listening to all my stories and dramas. He saved that bubbly girl in me who knows nothing but talking and giggling all day long for no reason. No doubt, he would be the happiest person after holding my thesis and definitely the only friend who would try to read and understand my research as well.

I am fortunate to thank **Rushit Shah**, my partner. This has just begun and I owe you many thanks for countless times of understanding. Thank you so much for respecting my dedication to research and supporting me unconditionally in all my decisions.

Lastly, with all my heart and soul I would like to thank God Almighty and my parents (**Mr. Yogesh Kumar Pachauri** and **Mrs. Neelam Pachauri**) for providing me strength and solidity to be able to achieve whatever I wanted. I would have never imagined my life so colorful and joyous without the support of my brothers (**Shekhar, Sagar and Tarun**) who stood by me all the time. I am thankful to my father for providing me immeasurable support, tough love, genuine advice, unending affection throughout this journey. I am lucky that I had you whenever I felt unsafe and I know I will always need you to have my back in this life. Without my mother's trust in me, I would have always underestimated myself. She always shared my achievements, failures, struggles from scratch to this day as if they were her dreams. She dedicated her life for me and I am merely dedicating my thesis to that proud smile of hers after seeing Dr. in front of my name.

Mummy this is for you!

July, 2020

7/12/2020

NA&BTD, BARC

Shikha Pachauri

HBNI, Mumbai

TABLE OF CONTENTS

Summary

Using gene knockout, the functional role of vGPD, a GAPDH associated with the "vir" cluster in *T. virens* was established. The vGPD mutant showed similar colony morphology and growth rate as wild type *T. virens*. Further, the vGPD mutant was characterised for secondary metabolites biosynthesis. Comparison of the Gas Chromatography–Mass Spectrometry (GC-MS) profile of the vGPD deletion mutant and the wild type *T. virens* showed that vGPD deletion mutant does not produce volatile sesquiterpenes metabolites. Similarly, comparison of the thin layer chromatography (TLC) and LC-MS profile of vGPD deletion with wild type *T. virens*, showed that vGPD deletion mutant is also not able produce non-volatile secondary metabolites. In the present thesis, we have documented the non-glycolytic role of a GAPDH in a *T. virens* and showed that vGPD gene which is a part of *vir* cluster, plays crucial role in volatile and non-volatile secondary metabolites biosynthesis. The expression of the *vir* cluster-specific genes in vGPD deletion mutant using RT-PCR showed that two genes (vir3 and vir4) were up-regulated, while other cytochrome P450s (*vir1, vir2, vir6* and *vir7*) and an MFS (*vir5*) were down-regulated. The regulatory mechanism of the *vir* cluster could only be identified if the mutant is downregulated in the expression of all the genes associated with the *vir* cluster. The M7 mutant displays similar secondary metabolism phenotype, i.e., lack of biosynthesis of volatile and non-volatile metabolites. With the aim to understand the regulatory mechanism of the *vir* cluster, we further characterized the M7 mutant at the genome level.

Characterization of M7 showed that this mutant, in addition to lacking conidiation, is defective in hydrophobicity. M7 is also non-mycoparasitic on *Rhizoctonia solani* and *Pythium aphanidermatum* and cannot utilize cellulose and chitin as sole carbon source. M7 is also not able to produce volatile and non-volatile metabolites as detected by GC-MS and TLC analysis.

The transcriptome analysis showed that 463 genes associated with secondary metabolism, gene regulation, mycoparasitism, carbohydrate metabolism, hydrophobicity and transportation were found to be down-regulated in the M7 mutant. All the *vir* cluster associated genes, including the vGPD gene, were downregulated in M7. In order to further identify the potential candidate genes involved in the regulation of the secondary metabolism in *T. virens*, whole genome sequencing of the M7 mutant was undertaken. The whole genome sequencing (WGS) analysis showed five deletions totalling about 250-Kb spread across three scaffolds. The deletions were further confirmed by PCR of a few representative genes. Total of 71 ORFs were found to be deleted and of these genes, eight genes were identified to be transcription factors. These transcription factors could be involved in the regulation of secondary metabolism, morphogenesis and biocontrol properties of *T. virens*.

The vGPD is an isoform of typical GAPDH (named as gGPD) which is a primarily involved in glycolysis (Sakai et al., 1990). Sakai et al. (1990) reported that vGPD is tolerant to heptelidic acid, an irreversible inhibitor of GAPDH. Under the condition of heptelidic acid production, vGPD protein takes over the glycolytic function in *T. virens*. Heptelidic acid is also an anti-cancer metabolite and is studied as a potential drug for cancer therapy. Characterizing the HA-tolerant form of GAPDH i.e., vGPD protein, has implications in cancer therapy. The cDNA for vGPD gene was cloned in pET-28a vector and expressed in *E. coli* BL21 Star (DE3) host cells. However, the recombinant protein was in inclusion bodies. Multiple approaches were employed to improve vGPD protein solubility, like varying IPTG concentrations, varying temperature, different expression systems, purification under denaturing condition etc. Other strategies like expression in *K. lactis* system and in the native host were also tried but protein either did not express or expressed at a very low concentration. Finally, we succeeded in expressing vGPD protein in the pNH-TrxT vector which possess a

poly-histidine tag and a thioredoxin tag. Thioredoxin tag provided stability to the vGPD protein by maintaining reduced environment, and thus allowed successful expression and purification of the vGPD protein using affinity chromatography. The recombinant vGPD-pNH-TrxT was sequenced and the protein molecular weight of the purified recombinant vGPD was confirmed by MALDI-TOF analysis. The biophysical techniques like thermo-flour shift assay and intrinsic fluorescence spectrum were used to confirm the proper folded structure of the recombinant vGPD protein. Phosphorylation activity assay of vGPD protein provided values for the Michaelis-Menten kinetics parameters towards glyceraldehyde-3 phosphate substrate and these values were estimated to be 0.38 mM (*K*m), 3.19 µmol/min/mg (*V*max) and 2.55 /sec (*k*cat). The size exclusion chromatography, dynamic light scattering and native-PAGE showed that vGPD protein exist as a tetramer in solution. Thus, biochemical and biophysical properties of the vGPD protein were similar to the glycolytic GAPDH (named as gGPD henceforth). Crystallization of vGPD protein was successful under two promising conditions, 1) Tris Na-citrate (0.1 M, pH 5.5), and PEG 2000 (20%) and 2) Calcium acetate (0.16 M), Na-cacodylate (0.08 M, pH 6.5), PEG 8000 (15%) and glycerol (20%), and small crystals were obtained, which were not suitable for X-ray diffraction. In order to assess whether vGPD acts as a transcription factor, DNA binding affinity of the vGPD protein with the upstream region of the putative terpene cyclase (*vir4*) gene was determined using EMSA. The results indicated that vGPD protein non-specifically interacts with the upstream region of the *vir4* gene (as it loses the interaction in the presence of cold non-specific DNA) and thus may not directly be involved in regulating the gene cluster.

Both gGPD and vGPD show more than 80% similarity (~73% identity) with essentially conserved NAD+ cofactor- and substrate-binding sites. To understand the difference in the affinity of vGPD and gGPD protein towards heptelidic acid (HA), we obtained highly reliable

structural models of the two proteins by homology modelling and probed HA binding by covalent docking. Interestingly, MM-GBSA (Molecular mechanics-generalized born surface area) binding energies determined using covalent docking suggested that vGPD protein model has lower affinity for HA as compared to gGPD protein model, in line with the available experimental data. A notable difference in the structure, however, was the presence of an indel (serine-143) in the loop preceding the active site and this is strictly specific to vGDP proteins. This indel is localized at the entrance to the active-site cavity. The role of this indel is not known but could be involved in interaction with other proteins or accessibility of the HA towards the active site as this indel.

Future prospective

The major contribution of the present study is assigning the role of the *vir* cluster associated GAPDH (vGPD), which showed that this gene cluster is involved in biosynthesis of volatile sesquiterpenes. The exact mechanism, however, needs to be elucidated. At this point of time, it is not clear if vGPD plays a biosynthesis role or a regulatory role. The downregulation of unrelated non-volatile metabolites like heptelidic acid (sesquiterpenes), viridin (steroid), gliovirin (non-ribosomal peptide) in vGPD knockout mutant may point to its role as a regulator of secondary metabolism. This study has also led to the identification of more than 400 genes that are downregulated in the M7 mutant. Among these are genes for secondary metabolism, mycoparasitism and plant-interactions. These genes could be candidates for future research. We also detected large deletions in the genome of M7, encompassing 71 genes. A few are transcription regulators which needs to be studied further for their role as global regulators of gene expression. We attempted purification and crystallization of the vGPD protein. Further studies are required to elucidate the finer differences in structures of vGPD and gGPD, which would reveal novel information on the structure-function relationship of GAPDH, an enzyme of enormous importance in metabolism and cancer therapy.

1.1. Overview:

Trichoderma spp. produce a wide range of volatile and non-volatile secondary metabolites, like non-ribosomal peptides, peptaibols, polyketides, pyrones, siderophores and several volatile secondary metabolites. These secondary metabolites are involved in various functions such as mycoparasitism, nutrient competition, induction of plant defense, antibiosis, symbiosis, etc. (Vinale et al., 2008). Identification of gene clusters involved in the biosynthesis of these metabolites is important to understand the biosynthetic pathways and associated regulatory mechanisms (Zeilinger et al., 2016). The first secondary metabolite gene cluster in a *Trichoderma* spp. was identified by Mukherjee and colleagues using suppression subtractive hybridization technique and a mutant of *Trichoderma virens* named as M7 (Mukherjee et al., 2006). M7 is a non-conidiating mutant that is not able to produce a detectable quantity of secondary metabolites. The role of the newly identified cluster was performed by knocking-out a terpene cyclase gene, *vir4* which is the core enzyme in the cluster. Further analysis of the volatile and non-volatile secondary metabolites biosynthesis using GC-MS and HPLC revealed that this cluster (named as *vir* cluster) is associated with volatile sesquiterpene metabolites biosynthesis (Crutcher et al., 2013).

The *vir* cluster is present in a few strains of *Aspergillus* spp. but not in other strains of *Trichoderma* spp. The cluster harbors 8 genes including one core enzyme i.e. terpene cyclase, four cytochrome P450s, one major facilitator superfamily (MFS), one monooxygenase, and a primary metabolism-related gene i.e. GAPDH (Fig 1.1). This GAPDH is also present in all the *Aspergillus* spp. *vir* clusters. The consistent association of this primary metabolite gene (*GAPDH*) with a secondary metabolism related gene cluster suggests a role of this GAPDH in secondary metabolism.

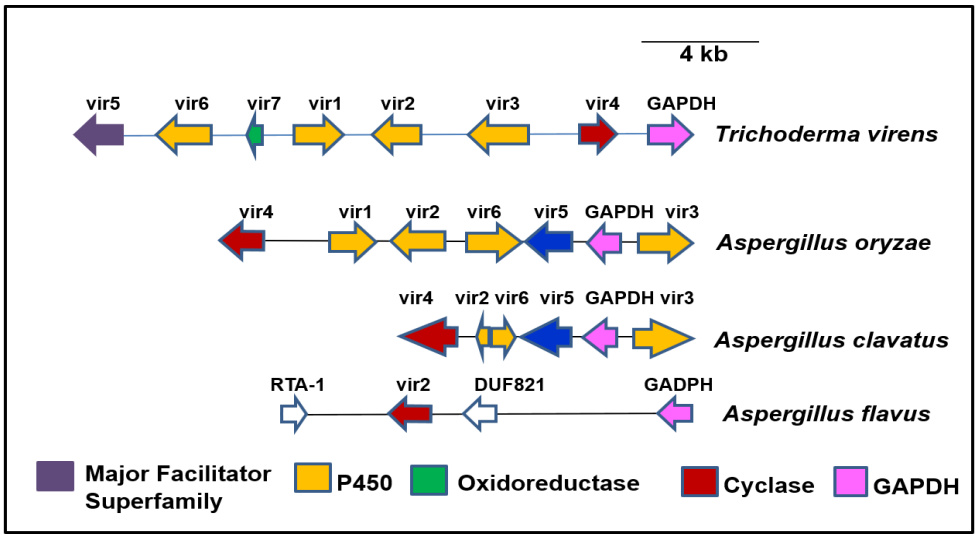


Figure 1.1 The *vir* gene cluster of *Trichoderma virens*, and its orthologue in *Aspergillus* spp. Direction of the arrows indicate 5'–3' coding orientation (Adapted from Bansal et al., 2018).

The *vir* cluster- associated GAPDH (named as vGPD) that is an isoform of GAPDH (named as gGPD) primarily involved in glycolytic function (Sakai et al., 1990). Sakai and group reported that this additional copy of GAPDH (vGPD) is a heptelidic acid-tolerant form of GAPDH that takes-over glycolysis under the condition of production of heptelidic acid, an irreversible inhibitor of GAPDH. Genome sequencing of *T. virens* provided evidence that heptelidic acid- tolerant form of GAPDH (vGPD) is a part of the volatile sesquiterpene metabolite producing *vir* cluster. The role of vGPD in secondary metabolism is not known yet. The major objective of the present thesis has been to study the function of this GAPDH (vGPD), biochemical and biophysical characterization, and get some insight into the regulation of the *vir* cluster.

1.2. Secondary metabolites:

Secondary metabolites are low molecular weight bioactive compounds that are not directly involved in primary functions, like energy production-dependent metabolic pathways, but is associated with biological activities that support the survival of the producing

organisms and are involved in interactions with other microorganisms (Brakhage, 2013). Secondary metabolites show restricted distribution among organisms i.e., a small group of organisms is involved in producing each metabolite at a defined developmental stage. Microorganisms are challenged by multiple biotic and abiotic stresses, and over millions of years, secondary metabolites have been used as a physiological response for survival and evolution. Microbes use secondary metabolites for communication, to inhibit the growth of competitors, for nutrient acquisition, for symbiotic interactions, and pathogenic interactions by acting as a virulence factor (Macheleidt et al., 2016).

More than one million natural compounds have been identified of which 2,50,000 compounds possess bioactivity and 22,000 compounds are of microbial origin (Solecka et al., 2012). About 17% of the metabolites of microbial origin are produced by unicellular bacteria (especially *Bacillus* spp. and *Pseudomonas* spp.). While actinomycetes (majorly *Streptomyces* genus) and fungi contribute 45% and 38% of the secondary metabolites produced by microorganisms (Solecka et al., 2012).

Some of the important secondary metabolites produced by fungi, bacteria, and plants include antibiotics, pharmaceuticals, toxins, bio-pesticides, and animal and plant growth factors. Both gram-positive (*Streptomyces, Saccharopolyspora, Nocardia, Corynebacterium, Bacillus,* and *Lactobacillus*) and gram-negative bacteria (*Escherichia, Pseudomonas, Serratia*) produce different secondary metabolites but the *Streptomyces* genus has the substantial ability to produce a wide range of secondary metabolites, such as antibacterial and antifungal substances (Sharma et al., 2014). Many secondary metabolites derived from fungi can be considered as potential drugs for targeting many diseases. Table 1.1 enlists some of the important secondary metabolites produced by fungi of industrial importance and commercial significance (Boruta, 2018). Some well-known plant secondary metabolites used as drugs are

artemisinin, paclitaxel, ginsenosode, lycopene, digitalis, morphine, quinine, and resveratrol (Song et al., 2014).

Table 1.1. Examples of secondary metabolites produced by fungi, with commercial applications

Secondary Metabolite	Producing Fungus	Application
Beta-carotene	*Blakeslea trispora*	Pigment
Cephalosporin C	*Acremonium chrysogenum*	cephalosporins production
Cyclosporine A	*Tolypocladium inflatum*	Immunosuppressant
Gibberelic acid	*Gibberella fujikuroi*	plant growth regulation
Lovastatin	*Aspergillus terreus*	cholesterol-lowering drug
Penicillin G	*Penicillium rubens*	Antibiotic
Taxol	*Taxomyces andreanae*	anticancer drug

1.3. Secondary metabolism gene clusters:

The secondary metabolite biosynthetic genes have two general features: 1) they are outstretched to more than thousand bases in a contiguous cluster and 2) the biosynthetic gene cluster (BGC) possess a gene encoding for a core enzyme, for example, polyketide synthase (PKS), non-ribosomal peptide synthetase (NRPS) and terpene cyclase (TC). Secondary metabolites are produced by the polymerization reaction of primary metabolites by these core enzymes of the BGC (Keller et al., 2005). PKS, NRPS, and TC utilize acyl CoAs, amino acids, and isoprene units respectively, as their building blocks for secondary metabolites

production. Other enzymes that are involved in further modifications of the initial secondary metabolite are known as tailoring enzymes coded by genes, for example oxidoreductases, cytochrome P450s, methyltransferases, GSTs, etc., are also part of BGC. Besides, transcription factors driving the expression of the gene cluster and transporters are also associated with BGCs (Hautbergue et al., 2018; Keller, 2019) The tremendous diversity in the structure of secondary metabolites is brought-about by the modifications (e.g. oxidoreduction, halogenation, hydroxylation, methylation/demethylation, and cyclization) to the basic structure/skeleton and sometimes by the condensation reaction with other metabolites. Apart from the typical series of genes, a BGC can also be present as a supercluster in which genes encoding secondary metabolites are present on two or more different chromosomes. For example, Some secondary metabolites are hybrid molecules such as polyketide-terpene, non-ribosomal peptide-polyketide and polyketide-fatty acid derived from two synthases or synthetases (separate or fused), and few secondary metabolites are generated from alternative pathways like ribosomally derived peptide ustiloxin, fatty-acid derived oxylipins and isocyanide xanthocillin generated from isocyanide synthase (Keller et al., 2005).

1.4. Cryptic and orphan gene clusters:

Analysis of the genome sequences of many fungi by cluster detection has uncovered many secondary metabolite biosynthetic gene clusters. Some bioinformatics algorithms used for the cluster detection are SMURF, antiSMASH, ClustScan, and PRISM. The genome analysis has identified many cryptic or orphan gene clusters whose corresponding secondary metabolites are still unknown. For example, *Aspergillus nidulans* possess 68 clusters but the product of only 20 clusters are known (Macheleidt et al., 2016). The secondary metabolite-biosynthesis genes within such BGCs are not expressed under the laboratory conditions but are induced under specific ecological niches and developmental

stages of the producing organisms. For example, expression of trichothecene, a virulence factor of *Fusarium graminearum* is induced during interaction of *Fusarium* spp. with plants (Hautbergue et al., 2018). Strategies like *in vitro* simulation of ecological niches, heterologous host transfer and stimulation of transcription factors and chromatin modification can be deployed for the identification of the metabolite(s) coded by the silent or cryptic gene clusters (Malik, 1980; Hautbergue et al., 2018).

1.5. Secondary metabolites produced by *Trichoderma* spp.:

The genus *Trichoderma* is a dominant flora in the soil, on decaying wood and other substrates and thus considered to be the most cosmopolitan filamentous ascomycetes (Druzhinina et al., 2011). As is well known, *Trichoderma* species are mycoparasitic, have high nutrient acquisition capacity, and are prolific producers of inhibitory compounds and degradative enzymes. These attributes have resulted in selection of *Trichoderma* spp. as biocontrol agents of choice, especially for soil-borne plant pathogens (Harman et al., 2004). Furthermore, the interaction with *Trichoderma* spp. results in many benefits to plants, like induction of plant defences, promotion of plant growth and tolerance to abiotic stresses (Djonović et al., 2007; Mukherjee et al., 2013b). The wide range of adaptability of *Trichoderma* species in dealing with challenging ecosystem is probably the reason why they produce array of secondary metabolites responsible for the evolutionary fitness of *Trichoderma* species on account of natural selection (Druzhinina et al., 2011). Many plant-beneficial interactions have been attributed to its ability to biosynthesize and secrete an array of secondary metabolites (Zeilinger et al., 2016; Mendoza-Mendoza et al., 2018).

Secondary metabolites produced by *Trichoderma* species include non-ribosomal peptides, peptaibols, siderophores, polyketides, pyrones, and terpenes (Vinale et al., 2008; Zeilinger et

al., 2016). The classification of secondary metabolites produced by *Trichoderma* species is given in the Table 1.2.

Table 1.2: Some examples of secondary metabolites produced by *T. reesei, T. atroviride* and *T. virens*

Chemical category	Compound	Species	Biological role
Non-ribosomal peptides			
Peptaibols	Trichovirin II	*T. virens*	Induces resistance in cucumber plants against a bacterial pathogen
Peptaibols	Trichorzianins Atroviridins A-C	*T. atroviride*	Atroviridins A-C are associated with conidiation
Peptaibols	Paracelsin, hypojecorins A and B	*T. reesei*	Paracelsin is reported to be hemolytic
Siderophore (intracellular)	Ferricrocin	*T. atroviride*, *T. virens*, *T. reesei*	Intracellular storage of iron, involved in gliotoxin biosynthesis and ISR
Siderophores (extracellular)	Fusarinine A-B Dimerum Acid Fusigen Coprogen	*T. virens*	Iron acquisition, competition
Diketopiperazine/NRP	Gliotoxin	*T. virens*	Antiviral, antibacterial, fungistatic activity, anti-cancer and immuno-suppressive properties
Diketopiperazine	Gliovirin	*T. virens*	Antimicrobial compound against oomycetes and *Staphylococcus aureus*, antitumor
Dipeptide	Trichodermamide A,B	*T. virens*	Cytotoxicity
Polyketides			
Polyketides	Trichodermatide B-	*T. reesei*	Cytotoxicity

	D		
Polyketide	Conidial pigment	*T. reesei*	Conidial pigmentation, stress tolerance
Polyketides	Trichorenins A-C	*T. virens*	Algicidal
Polyketide	Sorbicillin	*T. reesei*	antiviral, anti-inflammatory, and antimicrobial activities
Terpenes and steroidal compounds			
Sesquiterpene	Heptelidic acid (koningic acid)	*T. virens*	Potential activity against the human malaria parasite *Plasmodium falciparum*, antimicrobial, anticancer
Sesquiterpene	*β*-Farnesene	*T. atroviride*	Acts as an alarm pheromone in aphids
Sesquiterpene	*β*-Caryophyllene	*T. virens*	Attracts nematodes that prey on insect larvae
Monoterpene	*β*-Myrcene	*T. virens*	Regulates the expression of genes related to abiotic and biotic stresses
Monoterpenes	*cis*- and *trans*-*β*-Ocimene	*T. virens*	Induces expression of JA defense response-related genes in *A. thaliana*
Steroidal compound	Viridin	*T. virens*	Antifungal metabolite that alter the spore germination of *Botrytis allii*, *Colletotrichum lini* and *Fusarium caeruleum*
Other compounds			
Indolic compound	Indole-3-acetic acid (IAA)	*T. atroviride*, *T. virens*	Plant growth promotion
Indolic compound	Indole-3-acetaldehyde	*T. atroviride*, *T. virens*	Plant growth promotion
Indolic compound	Indole-3-carboxaldehyde	*T. atroviride*, *T. virens*	Induces adventitious root formation in *A. thaliana*
Carotanes	Trichocaranes A-D	*T. virens*	Inhibits the growth of etiolated wheat coleoptiles
Pyrone	6-Pentyl-2*H*-pyran-	*T. atroviride*	Antifungal, antinematode

	2-one		and plant growth-promoting activities in tomato and *A. thaliana*
Ketone	3-Octanone	*T. atroviride*	Induces condiation
Alcohol	1-Octen-3-ol	*T. atroviride*	Induces conidiation and defense responses in plants through jasmonic acid

(Source: Reino et al., 2008; Ruiz et al., 2013; Contreras-Cornejo et al., 2016)

1.5.1. Non-ribosomal peptides (NRPs):

The non-ribosomal peptides are synthesized by multi-modular non-ribosomal peptide synthetases (each module includes adenylation, peptidyl carrier and condensation domains) and contain both proteinogenic and non-proteinogenic amino acids. They may exist in linear or cyclic forms. The important NRPs produced by *Trichoderma* spp. are peptaibols, epidithiodioxopiperazines (ETPs) and siderophores (Keller et al., 2005).

1.5.2. Peptaibols:

Peptaibols are the most prominent NRPs produced by *Trichoderma* species and are reported to have antimicrobial property, cytotoxic activity and induced systemic resistance property in plants (Mukherjee et al., 2012a). Peptaibols are short peptides containing α-aminoisobutyric acid (aib) and a C-terminal alcohol. The antibiotic property of peptaibols is primarily due to amphipathic nature of peptaibols (Bortolus et al., 2013). Peptaibols are synthesized by peptaibol synthetases (NRPSs) consisting of different modules and each module adding a single amino acid. In *Trichoderma* genomes, there are 3 types of peptaibol synthetases present: 7-, 14- and 18-20 module peptaibol synthetases (Mukherjee et al., 2011; Wiest et al., 2002). The first peptaibol synthetase enzyme (Tex1) has been reported in *T. virens* by (Wiest et al., 2002). A 14-module peptaibol synthetase enzyme which produces two classes of peptaibols (the 11- and 14-residue peptaibols) due to module skipping mechanism

has also been identified in *T. virens* (Mukherjee et al., 2011). *T. atroviride* is also reported to produce 19-residue atroviridins produced by a 19-module peptaibol synthetase and *T. reesei* produce 11- and 14-residue peptaibols catalyzed by a 14-module peptaibol synthetase (Degenkolb et al., 2012).

1.5.3. Epidithiodioxopiperazines (ETPs):

Gliotoxin and gliovirin are members of the epipolythiodioxopiperazines (ETPs) which are cyclic, peptide-derived secondary metabolites and have characteristic diketopiperazine ring. Gliovirin is produced by “P” strains and gliotoxin is produced by “Q” strains of *T. virens* (Howell et al., 1993). In "P" strains of *T. virens*, the gliovirin biosynthetic gene cluster has been identified by Sherkhane and group and named it as *glv* cluster (Sherkhane et al., 2017). The gliotoxin biosynthetic gene cluster with 12 genes has been characterized in *A. fumigatus* (Fig 1.2). Gliotoxin is reported to act as a virulence factor in *A. fumigatus* (Scharf et al., 2016) and the synthetase involved in gliotoxin biosynthesis is a NRPS dioxopiperazine synthetase enzyme named as GliP. The functional characterization of the *gliP* gene was performed by partial deletion of it in *T. virens* and it was confirmed that *gliP* gene is involved in gliotoxin biosynthesis (Vargas et al., 2014). The gliotoxin biosynthetic gene cluster in *T. virens* possesses 8 genes but in *T. reesei* which is not involved in gliotoxin production harbours only 6 genes (Mukherjee et al., 2012a).

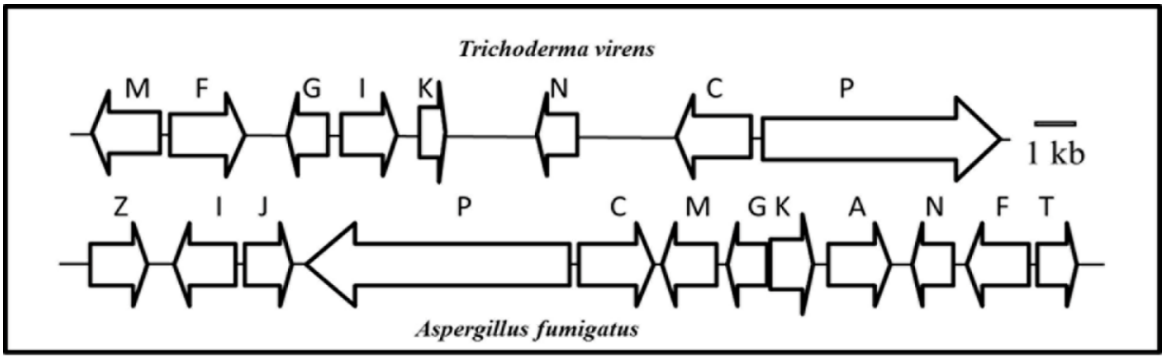


Figure 1.2 The gliotoxin biosynthesis cluster of *Trichoderma virens* and *Aspergillus*

fumigatus P: Non-ribosomal peptide synthetase (NRPS); G: Glutathione-S-transferase; J: Dipeptidase; N: N-methyl transferase; F and C: Cytochrome P450; M: O-methyl transferase; I: C-S-bond lyase; K: gamma-glutamate cyclotransferase: Z: Zinc Finger; T: Thioredoxin Reductase; A: Transporter (Adapted from Zeilinger et al., 2016).

1.5.4. Siderophores:

Siderophores are the class of secondary metabolites which are involved in iron metabolism by binding, transport and storage of it. Iron is an important metal ion for microbial competition, biocontrol and in interaction with other microorganisms (Saha et al., 2016). In fungi, the siderophores are classified into two types, extracellular and intracellular. Three NRPSs are reported to be associated with siderophore biosynthesis in *Trichoderma* species. NRPSs putatively involved in extracellular siderophore biosynthesis are NPS6 and SidD. NPS6 is found to be present in all the *Trichoderma* species and the gene cluster having SidD as a core enzyme is present only in *T. virens* and *T. reesei* (Mukherjee et al., 2012b). The function of NPS6 was confirmed by gene deletion in *T. virens* (Mukherjee et al., 2013a). The biosynthetic gene cluster for ferricrocin which is an intracellular siderophore was identified in *T. virens, T. atroviride* and *T. reseei,* but the functional characterisation has been done only in *T. virens* (Mukherjee et al., 2018).

1.5.5. Polyketides:

Polyketides are secondary metabolites, many having antimicrobial and anti-cancer properties. Some polyketides are important for competition for substrate, and for interaction with other organisms. Polyketides are synthesized by the polyketide synthase, which are complex enzymes with ketoacyl synthase, an acyl transferase, and a phosphopantetheine attachment site domains (Fujii, 2010). Only a few studies have been published on the biosynthesis and functional role of polyketides in *Trichoderma* species, although the genomes of *Trichoderma* species are rich in PKS-encoding genes. Orthologues of PKS genes

associated with the conidial pigment biosynthesis cluster have been identified in all the three *Trichoderma* species (Baker et al., 2012). Additional 20 putative PKS gene clusters have also been reported in these three species (Bansal and Mukherjee, 2016). The role of PKS genes in green pigmentation of conidia, teleomorphic structure, conidial cell wall stability and antagonistic abilities has been confirmed in *T. reesei* by the deletion of *pks4* gene which is an orthologue of pigment forming PKS in *Fusarium* spp. (Atanasova et al., 2013a).

1.5.6. PK/NRPs:

Several PKS-NRPS hybrid enzymes are present in *Trichoderma* genomes. Functional study provided indication for the role of PKS-NRPS hybrid (Tex13) in inducing the defense-related *pal* gene in maize seedlings. The metabolite produced by Tex13 cluster is still not known (Mukherjee et al., 2012b).

1.5.7. Terpenoids:

Terpenoids represent a diverse class of secondary metabolites produced by almost all the organisms including fungi (Schmidt-dannert, 2015). They are composed of five-carbon isoprene units (C5H8) producing hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), sesterterpenes (C25), triterpenes (C30), tetraterpenes (C40), or polyterpenes ((C5H8)n). Terpenoid biosynthetic gene clusters have terpene cyclase as the core enzyme. Only a few terpene cyclases may be responsible for the production of diverse form of terpenoids. *Trichoderma* species are reported to produce all forms of terpenoids like volatile terpenoids, diterpenes, sesquiterpenes and triterpenes (Zeilinger et al., 2016).

In *Trichoderma* spp., the mevalonate pathway is responsible for the formation of isoprene units (Zeilinger et al., 2016). Hydroxy-methylglutaryl-CoA reductase (HMGR) encoded by *hmgR* gene is the first enzyme in the mevalonate pathway involved in the conversion of hydroxy-methylglutaryl-CoA into mevalonate. Deletion of *hmgR* gene in

Trichoderma harzianum showed reduction in antifungal activity against *Rhizoctonia solani* and *Fusarium oxysporum* and decrease in ergosterol levels. Silencing of *erg1* gene (ergosterol encoding gene) also showed decrease in the ergosterol level in *T. harzianum*, whereas overexpression of *erg1* gene increased the antifungal activity of *T. harzianum* (Cardoza et al., 2006, 2014). The genome analysis of *T. reesei, T. atroviride* and *T. virens* revealed that *T. virens* (11) has the highest number of terpene cyclases followed by *T. atroviride* (7) and *T. reesei* (6), but the terpene cyclases associated with the biosynthetic gene cluster are six in *T. virens*, three in *T. atroviride* and two in *T. reesei* (Bansal and Mukherjee, 2016). The first terpenoid biosynthetic gene cluster was identified in *T. virens* using suppression subtractive hybridization (Mukherjee et al., 2006). The cluster was initially predicted to be associated with viridin production but later on found to be responsible for biosynthesis of volatile sesquiterpenes (Crutcher et al., 2013). Deletion of this terpene cyclase abolished the production of all the volatile sesquiterpenes. This cluster, named as "*vir*" cluster, was found to be present in *T. virens* and in few *Aspergillus* species but not in other species of *Trichoderma*. The existence of the *vir* cluster in distantly related *Trichoderma* and *Aspergillus* species could be explained by horizontal gene transfer (Mukherjee et al., 2006). A terpene cyclase (Tri5), responsible for the production of a phytotoxic agent trichodermin was characterized in *Trichoderma brevicompactum*. Overexpression of *tri5* enhanced the production of trichodermin in *Trichoderma brevicompactum* (Tijerino et al., 2011a, 2011b).

1.5.8. Steroids:

Viridin is a triterpene steroidal metabolite produced by both 'P' and 'Q' strains of *T. virens*. It has antifungal and anticancer properties (Howell et al., 1993). The reduced form of viridin is viridiol which has herbicidal properties (Jones and Hancock, 1987). Both viridin and viridiol are produced abundantly by *T. virens* and the *vdn* cluster for viridin biosynthesis has

recently been discovered (Fig 1.3). Interestingly, an orthologous gene cluster is also present in the bat white nose fungus *Pseudogymnoascus destructans* (Bansal et al., 2018).

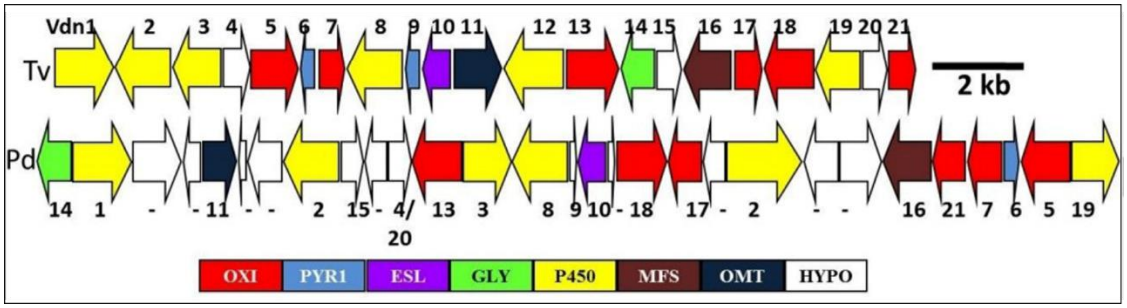


Figure 1.3 The viridin-biosynthesis gene cluster of *Trichoderma virens*, and its orthologue in *Pseudogymnoascus destructans*. Oxidoreducatse (Red); Transporter (Brown); Hypothetical protein (White); Glyoxalase (Green); Cytochrome P450 (Yellow); Esterase and lipase domain protein (Pink); O-methyl Transferase (Dark Blue); Pyrabactin-resistance 1 like protein (Light Blue) (Adapted from Bansal et al., 2018).

1.5.9. 6-Pentyl pyrone (6-PP):

6-Pentyl pyrone **(**6-PP) is a volatile secondary metabolite. The characteristic "coconut aroma" produced by some *Trichoderma* species is due to 6-PP metabolite production. 6-PP has antifungal and plant growth promoting properties (Vinale et al., 2008). There is not much information available on the biosynthetic pathway for the 6-PP production, but a lipoxygenase enzyme has been predicted to be involved in the 6-PP biosynthesis as it is present exclusively in *T. atroviride*, the producing species, but not in other *Trichoderma* species (Kubicek et al., 2011).

1.6. Regulation of secondary metabolism:

The regulation of secondary metabolism is very tightly held at multiple levels, for example, global transcription regulation, cluster specific transcription level, epigenetic regulation, signal transduction pathway-based regulation, regulation induced by interaction

with other microbes, and environmental regulation. There can be cross-talk and interconnection between the regulatory pathways.

1.6.1. Transcriptional regulation:

The transcription of the gene cluster can be regulated by cluster-specific transcription factors (Macheleidt et al., 2016), This can also be achieved by the global regulators which might regulate the secondary metabolite biosynthetic gene cluster, alongside other genes that are not involved in secondary metabolism.

1.6.1.1. Global regulators:

Global regulators exert transcriptional control on multiple gene clusters or unrelated genes. Some examples are PacC (regulated by pH), CCAAT-binding complex (CBC) (regulated by iron), AreA (nitrogen-regulated), Vel1 (light-regulated) and CreA (carbon catabolite regulator).

The pH regulation in fungi is controlled by the global regulator PacC which affects secondary metabolite production as well as virulence in many plant and animal pathogens. For example, PacC activates penicillin and alkaline phosphatase D production at alkaline pH in *A. nidulans* but represses aflatoxin production in *Aspergillus parasiticus* (Keller et al., 1997). PacC also regulates production of gliotoxin by *Aspergillus fumigatus*, fuminosins by *Fusarium verticillioides*, ochratoxin A by *Aspergillus ochraceus* (Bertuzzi et al., 2014; Flaherty et al., 2003; O'Callaghan et al., 2006). Deletion of PacC orthologue in *T. virens* repressed the expression of genes related to secondary metabolite biosynthesis iron transport. The PacC deletion mutant in *T. virens* also had reduced biocontrol activity (Trushina et al., 2013).

CBC complex possesses three subunits (HapB, HapC, and HapE) and requires HapX which is an iron regulating protein. CBC regulates the redox status and iron-induced responses and secondary metabolite biosynthesis. (Hortschansky et al., 2007; Gsaller et al., 2014). Example of redox induced response mediated by CBC complex is penicillin production in *Penicillium chrysogenum* and aflatoxin production in *A. parasiticus* (Brakhage et al., 1998; Reverberi et al., 2008). Secondary metabolite production is also influenced by carbon and nitrogen source. For example, Cre1 is a key regulator of carbon metabolism and upregulates aflatoxin production in *Aspergillus parasiticus* when glucose is present and bikaverin biosynthesis in *F. fujikuroi* when sucrose is the carbon source (Roze et al., 2004; Rodr´ıguez-Ortiz et al., 2010). But homolog of Cre1, CreA represses the production of cephalosporin in *Acremonium chrysogenum* when glucose is present (Janus et al., 2008; Jekosch et al., 2000).

There are two global nitrogen regulators which are GATA transcription factor, AreA and AreB. AreA is involved in regulating the nitrogen-induced response by repressing gibberellin biosynthesis in *F. fujikuroi* and by upregulating fumonisin B1 production in *F. verticillioides* (Kim et al., 2008; Tudzynski et al., 1999). AreB, in complex with AreA, regulates biosynthesis of gibberellin, apicidin F and fusaric acid in *F. fujikuroi* (Chang et al., 2000; Mihlan et al., 2003).

The velvet complex proteins regulate the light response in filamentous fungi and is well studied in *Aspergillus nidulans*. This is a hetrotrimeric complex consisting of two velvet proteins VeA (or Vel1) and VelB and the nuclear methyltransferase-domain protein LaeA (or Lae1) (Bayram et al., 2012). Deletion of *vel1* gene in *T. virens* ceased the production of gliotoxin and downregulated many secondary metabolism related genes (Mukherjee and Kenerley, 2010). In *T. reesei*, the *vel1* gene was found to be involved in secondary metabolism and mating partner sensing (Bazafkan et al., 2015). LaeA is an important regulator as it affects 50% of secondary metabolite biosynthetic gene clusters in fungi and the

deletion mutant of LaeA displayed decreased biosynthesis of secondary metabolites (Bok et al., 2004). Loss of Lae1 resulted in the reduced expression of lignocellulose-degrading enzymes in *T. reesei* and PKS-encoding genes and 6-PP-related lipoxygenase gene in *T. atroviride* (Karimi-Aghcheh et al., 2013).

1.6.1.2. Cluster-specific regulators:

Cluster specific regulators common in fungi are sequence-specific DNA binding proteins of the Zn2Cys6 type and Cys2His2 type (Chang et al., 2013; Tsuji et al., 2000). Some examples of Zn2Cys6 type of transcription factors are ApdR and AfoA in *A. nidulans*, GliZ in *A. fumigatus*, LovE in *Aspergillus terreus*, and Bik5 in *F. fujikuroi* (Bergmann et al., 2007; Bok et al., 2006; Kennedy et al., 1999; Wiemann et al., 2009). Other less common transcription factors are bZIP type and winged helix proteins (Hong et al., 2013). A regulatory gene may be encoded inside or outside the secondary metabolite biosynthetic gene cluster. The cluster specific transcription factors can be single or multiple in number. For example, the gliotoxin biosynthesis gene cluster encoding 13 members have single Zn-finger transcription factor GliZ while two regulatory genes (*aflR* and *aflS*) are present in the aflatoxin biosynthesis cluster in *Aspergillus flavus* and in the sterigmatocystin biosynthesis cluster in *A. nidulans* (Flaherty et al., 1997; Yin et al., 2011). A transcription factor normally upregulates all the other members in the BGC. For example, in *A. nidulans,* all the genes in the aspyridone biosynthesis cluster are regulated by ApdR, and AflR regulates the aflatoxin cluster in *A. flavus* and the sterigmatocystin cluster in *A. nidulans* (Bergmann et al., 2007).

In biosynthetic gene cluster, cross talk and cluster interactions have also been reported. For example, in *A. nidulans,* the ScpR regulates two different secondary metabolite clusters (inp and asperfuranone biosynthetic gene clusters) located on different chromosomes and the AflR also regulates both sterigmatocystin and asperthecin biosynthetic gene clusters (Yin et al., 2012).

1.6.2. Epigenetic regulators:

The clustered arrangement of secondary metabolite biosynthetic genes favours the regulation by chromatin modifications by altering transcriptional accessibility to BGCs (Li et al., 2011). Chromatin structure consists of chromatin remodelers, histone variants, and histone chaperones and the chromatin modifiers incorporate required modifications like methylation, acetylation, phosphorylation, ubiquitylation and sumoylation to the chromatin structure. Characterization of the histone modifying enzymes such as histone acetylases, deacetylases, methylases and demethylases by overexpression and deletion and by using small inhibitor molecules that target these enzymes provide an approach in identifying metabolites from silent gene clusters (Palmer et al., 2010; Gacek et al., 2012). Shwab and group achieved epigenetic activation of multiple BGCs in *A. nidulans* through the deletion of a histone deacetylase, *hdaA* (Shwab et al., 2007). LaeA is a global regulator of secondary metabolism and morphogenesis (Bayram et al., 2012). LaeA might also be involved in the epigenetic regulation as it contains sequence similar to histone and arginine methyltransferases. The LaeA deletion mutants in *A. nidulans* showed downregulaion of the sterigmatocystin and penicillin biosynthesis gene clusters and some indole alkaloid biosynthesis gene clusters (Reyes-Dominguez et al., 2010). Similarly, comparison of whole genome transcriptional profile of wild type, deletion mutant of *laeA* gene in *A. fumigatus* strains revealed that out of 22 secondary metabolism BGCs, 13 clusters down regulated. The deletion of *laeA* gene in *A. nidulans* decreased the HepA and H3K9me3 levels (specifying the euchromatin state) in sterigmatocystin biosynthesis gene cluster when the fungus is in the stationary growth phase that promotes secondary metabolism. However, during early growth phase of *A. nidulans*, the levels of H3K9me3 histone (H3 lysine-9-trimethylation) and HepA (heterochromatin protein 1) are high (specifying the heterochromatin state) (Reyes-Dominguez et al., 2010). The histone acetyltransferase GcnE of *A. nidulans* is a member of the SAGA–ADA complex (Spt–

Ada–Gcn5–acetyltransferase–ADA) and induces the expression of orsellinic acid BGC in *A. nidulans* by acetylation of histone H3 lysine 14 during interaction with the bacterium *Streptomyces rapamycinicus* (Schroeckh et al., 2009; Albright et al., 2015).

1.6.3. Signal transduction pathway regulators:

Signalling pathways transform the environmental stimulus to response, by regulating gene expression and activating secondary metabolite production. Some highly conserved signalling pathways in fungi are cyclic adenosine monophosphate (cAMP)/protein kinase A (PKA), calcineurin/calmodulin, TOR, and mitogen-activated protein kinase (MAPK). Association of signalling pathways and secondary metabolite production was shown for the first time in *A. fumigatus*. The deletion of α-subunit of G-protein and adenylate cyclase decreased the production of dihydroxynaphthalene-melanin production in *A. fumigatus* while the overexpression of protein kinase A catalytic subunit (PkaC1) increased the production of dihydroxynaphthalene-melanin (Grosse et al., 2008). In *T. atroviride,* the deletion of the adenylyl cyclase-inhibiting G-α subunit, which is encoded by *tga1* gene, repressed the production of 6-PP but induced the peptaibols biosynthesis (Reithner et al., 2005). However, the deletion of the *tga3* gene encoding adenylyl cyclase-stimulating G-α subunit abolished the synthesis of peptaibols in *T. atroviride* (Komon-Zelazowska et al., 2007). The role of *tac1* gene encoding adenylate cyclase in growth, germination, mycoparasitism and secondary metabolite production was studied in *T. virens* (Mukherjee et al., 2007). Similar to cAMP pathway, MAPK-dependent pathway is also reported to be involved in regulating secondary metabolism. For example, deletion of *tmk1* gene encoding for a MAPK in *T. atroviride* enhanced the production of peptaibols and 6-PP (Reithner et al., 2007); but in *T. virens* deletion of *tmk1* gene showed no alteration in secondary metabolite production (Mendoza-Mendoza et al., 2003).

1.6.4. Interaction with other microbes:

Fungi share habitat with different microorganisms and secondary metabolites provide protection by serving as a signal for competition and cooperation (Grosse et al., 2008). Analysis of transcriptome during interaction of *R. solani* with *T. atroviride*, *T. virens* and *T. reesei* revealed important alterations in expression of some genes. In *T. virens*, gliotoxin biosynthetic gene cluster got induced during interaction with *R. solani* (Atanasova et al., 2013b). The presence of *R. solani* also altered the expression of a 6-PP biosynthesis related lipoxygenase gene and 2 PKSs in *T. atroviride* (Atanasova et al., 2013b; Kubicek et al., 2011). Some microorganisms produce compounds which suppress the secondary metabolite production in fungi. For example botrydial which is a secondary metabolite and a virulence factor of *B. cinerea* is reported to repress the *tri* gene expression and the production of harzianum A in *T. arundinaceum* (Malmierca et al., 2015). Similarly, the presence of *Fusarium* is reported to increase the production of 1-octen-3-ol and decrease the production of 6-PP in *T. atroviride* while *Fusarium* itself overproduces a mycotoxin called fusaric acid (Stoppacher et al., 2010).

1.7. GAPDH and its moonlighting properties:

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH, EC 1.2.1.12) is an enzyme catalysing the sixth step of glycolysis. GAPDH is a house keeping gene for gene expression analysis. GAPDH is a model protein for enzyme kinetic analysis (Barber et al., 2005). GAPDH gene has an active promoter region, a strong enhancer region and possess CpG islands with active acetylation and absence of methylation in many tissues (Krasnov et al., 2013). GAPDH is one of the most well-known and most studied moonlighting proteins (Sirover, 1999; Sirover, 2011; Tristan, 2011) (Fig 1.4).

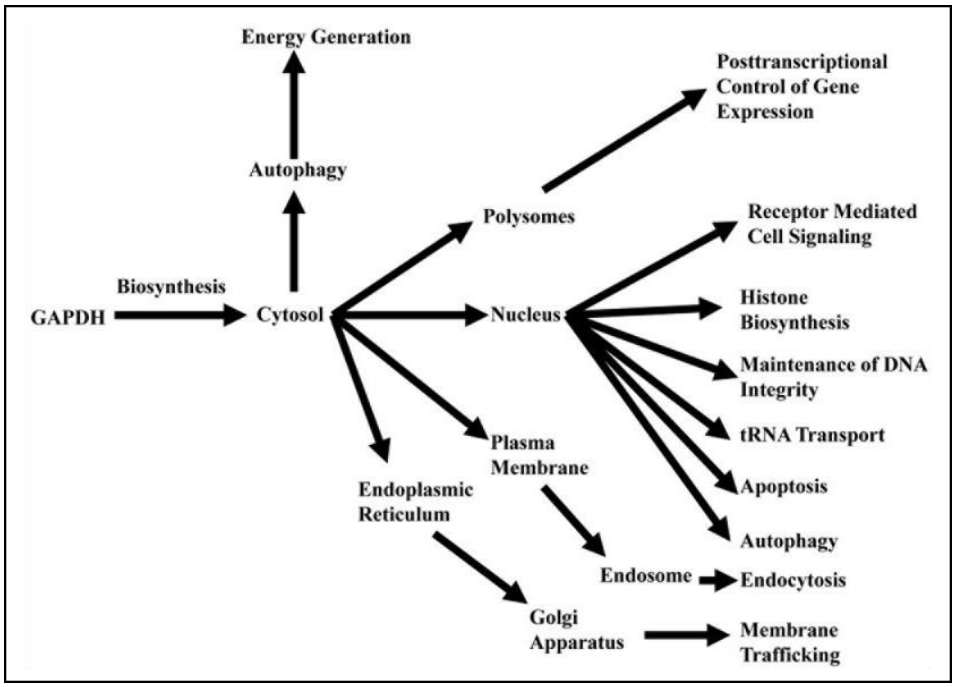


Figure 1.4 Subcellular dynamics of GAPDH and its functional diversity. (Adapted from Sirover et al., 2012)

Moonlighting proteins are the ones which perform additional functions other than the primary functions. In addition to glycolysis, GAPDH is reported to be involved in different functions like DNA replication and repair (Meyer-Siegler et al., 1991; Baxi and Vishwanatha, 1995; Azam et al., 2008), intracellular membrane trafficking (Tisdale, 2001), regulating RNA stability (Bonafé et al., 2005; Rodriguez-Pascual et al., 2008; Zhou et al., 2008; Backlund et al., 2009; Kondo et al., 2011; Ikeda et al., 2012), apoptosis (Hara et al., 2005; Sen et al., 2009), neurodegenerative disorders (Mazzola and Sirover, 2002; Bae et al., 2006; Shalova et al., 2007; Naletova et al., 2008; Butterfield, et al., 2010), oxidative stress response (Benhar and Stamler, 2005; Hara et al., 2005; Bae et al., 2006; Baek et al., 2008; Pierce et al., 2008; Nakamura and Lipton, 2009; Sen et al., 2009) and many others (Fig 1.5). The non-glycolytic role of GAPDH has been well explored in animal system as compared to the bacteria, fungi and plants. In animal system, only one copy of GAPDH gene is present but plants possess

multiple copies of GAPDH gene and the isoforms of GAPDH are dispersed in different subcellular compartments (Zaffagnini et al., 2013).

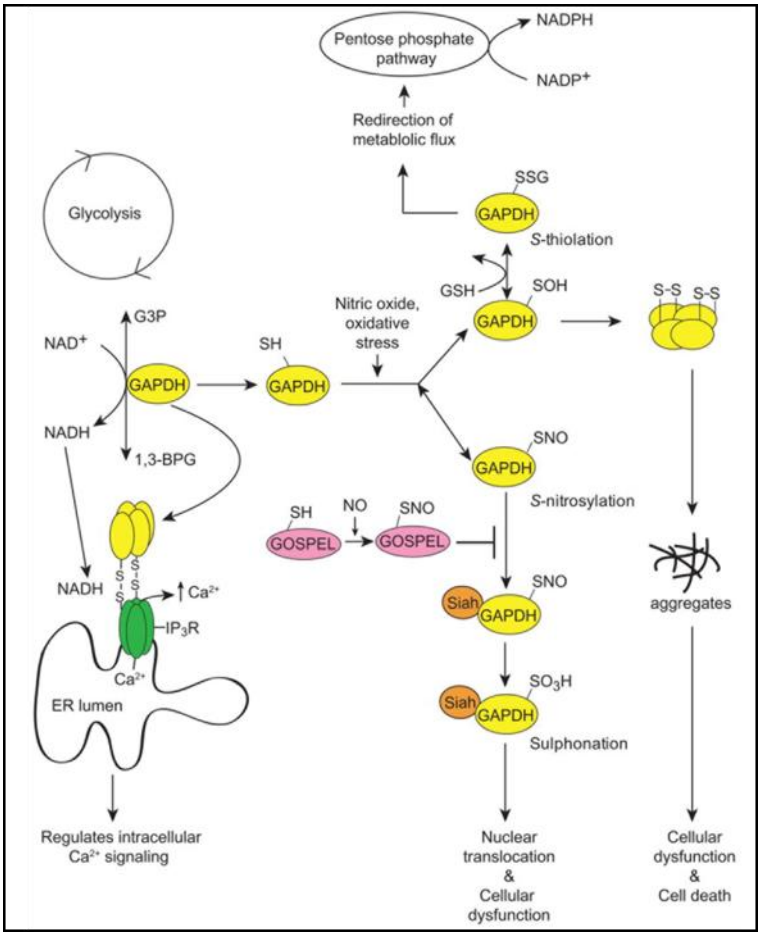


Figure 1.5 Moonlighting functions of GAPDH inside the cytoplasm Post-translational modifications of GAPDH under oxidative stress condition such as reversible S-thiolation (–SSG), reversible S-nitrosylation (–SNO) and irreversible sulphonation (–SO3H) provides additional functions to GAPDH (Adapted from Tristan et al., 2011).

In *Arabidopsis*, GAPDH is reported to regulate transpiration rate in drought condition (Guo et al., 2012) and phytopathogenic virus replication (Prasanth et al., 2011). In plants, GAPDH undergoes redox regulated post-translational modifications resulting in interaction with other proteins, and by relocating to different compartments inside the cell it performs non-glycolytic functions (Zaffagnini et al., 2013). In bacteria like *Staphylococcus aureus, Staphylococcus epidermidis, Escherichia coli, Streptococcus pneumonia* and *Streptococcus*

pyogenes, GAPDH has been identified as a virulence factor (Gil-Navarro et al., 1997; Gozalbo et al., 1998; Modun, et al., 2000; Taylor and Heinrichs, 2002; Egea et al., 2007). The functional diversity of GAPDH is due to oligomerization state of the protein, localization inside the cell, complex formation with other molecules and chemical modifications (Sirover et al., 2011).

1.8. GAPDH inside the nucleus:

GAPDH has been reported as a nucleic acid binding protein which results in moonlighting functions performed by it. The first study on GAPDH interaction with single stranded DNA was reported in 1977 by Perucho and group (Perucho et al., 1977). Subsequently many reports were published demonstrating the DNA-binding mediated functions of GAPDH in regulation of mRNA translation, binding with viral RNAs, nuclear tRNAs transport, telomere protection, gene regulation etc. (Sirover et al., 2005). Moreover, GAPDH is also associated with functions like apoptosis, DNA integrity maintenance and DNA repair mechanism which necessitates nuclear translocation. Nuclear functions of GAPDH are reported to be associated with oligomeric status, post translational modifications, oxidative stress and status of -SH groups. Furthermore, different proteins are also involved in forming complex with GAPDH and translocation of it inside the nucleus under certain conditions. Such as, siah1 forms complex with nitrosylated GAPDH under oxidative stress condition (Hara et al., 2005), and androgen receptor, which is a transcription factor, forms complex with GAPDH and translocates to the nucleus. (Harada et al., 2007) (Fig 1.6).

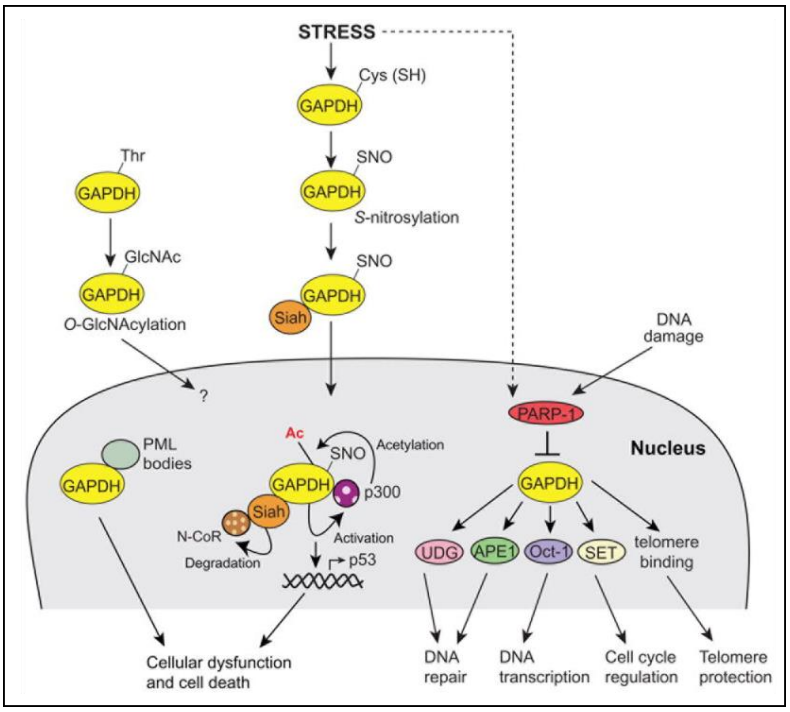


Figure 1.6 Moonlighting functions of GAPDH inside the nucleus (Adapted from Tristan et al., 2011).

Based on structural analysis, it is reported that GAPDH protein sequence harbours nuclear localisation domain (KKVVK, amino acids 259-263) at the catalytic site (Sirover, 1999). Brown and group proposed a nuclear export signal in the catalytic domain (KKVVKQQASEGPLK) at position 258–270 and identified the involvement of Lys-259 in binding with CRM1, a nuclear export receptor (Brown et al., 2004). GAPDH can also migrate to nucleus in association with other proteins. Moreover, post-translational modifications like acetylation of Lys-117, Lys-227 and Lys-251, NO-induced carbonylation and S-nitrosylation of the active site cysteine residue facilitates the nuclear localisation of GAPDH. PTMs and stress conditions generate different isoforms of GAPDH and these isoforms then perform non-glycolytic and diverse functions in different subcellular compartments (Sirover, 2011). Changes in oligomeric state also adds to the functional diversity in GAPDH as it is reported that addition of N-acetyl glucosamine residue disrupts the oligomeric status of the GAPDH from native tetramer to monomeric state which is preferable for nuclear translocation (Kosova et al., 2017).

The role of NAD^+ binding site or Rossman-fold as the putative DNA binding site in GAPDH has been demonstrated. The role of both the NAD^+-binding site and the catalytic site has been identified as a DNA binding motif by mutational analysis of Asp-32 and Cys-149 residues (Demarse et al., 2009). Additionally, mutational study and molecular modelling experiments identified phosphorylating residues (Y-94, S-98 and T-99) within NAD^+ binding site of GAPDH to be involved in interactions with unknown biomolecules. They concluded that the NAD^+ binding site is required for performing intra-nuclear functions (Phadke et al., 2015). For studying interactions with DNA/RNA, electrophoretic mobility shift assay (EMSA) has been used to identify GAPDH-DNA interaction (Demarse et al., 2009) as well as GAPDH-RNA interaction (Zang et al., 1998). For GAPDH-protein interaction different techniques like immunoprecipitation (Bae et al., 2006) and immunofluorescence microscopy (Kim et al., 2007) have been used. For studying interactions with DNA/RNA, EMSA has been used (Svedruzic and Spivey, 2006).

1.9. GAPDH isozymes in *Trichoderma virens*:

T. virens genome harbours two genes for GAPDH, one is involved in glycolytic function for energy metabolism (designated here as gGPD) while another one (designated here as vGPD) is associated with a secondary metabolism related-gene cluster (the *vir* cluster). Sakai and group identified GAPDH 1 (vGPD) protein as an isozyme of GAPDH 2 (gGPD) protein; GAPDH 1 was less sensitive to heptelidic acid (HA), a sesquiterpene lactone. They observed that HA production in *T. virens* (the study mis-identified as *T. koningii*) caused inhibition of gGPD protein and a concomitant increase in vGPD protein activity. They proposed that vGPD takes over the glycolytic function of gGPD when *T. virens* produces HA. For identification of the residue involved in HA binding, Sakai and group performed another experiment with radiolabelled HA complexed with GAPDH followed by trypsin digestion and

reverse-phase HPLC (Sakai et al., 1991). This analysis identified the role of sulfhydryl group of Cys-149 at active site of the pig muscle GAPDH (Sakai et al., 1991). Later, both the GAPDHs from *T. virens* were cloned and sequenced in order to determine the difference in sensitivity towards HA at amino acid sequence level (Watanabe et al., 1993). Their analysis showed the importance of S-loop region (aa 178-201) and Thr-174 and Thr-181 in gGPD substituted by alanine and serine in vGPD. Twenty-four amino acids in S-loop region showed variability in both GAPDHs which might be associated with the sensitivity to HA. Similarly, Thr-174 is reported to form H-bond with catalytic residue His-176 required for protonation of the glyceraldehyde-3-phosphate (Watanabe et al., 1993). In another report, amino acid sequences of GAPDH from different sources implicating difference in sensitivity towards HA were compared and possible candidates like S-loop region and few other amino acids in catalytic region were reported (Kato et al., 1992).

1.10. Heptelidic acid biosynthetic pathway:

Heptelidic acid (HA) or koningic acid (KA) is a sesquiterpene lactone majorly produced by filamentous fungi from *Ascomycetes* family including *Aspergillus oryzae*, *Anthostoma acocetta*, *Chaetomium globosum* and *Trichoderma virens*. The biosynthesis pathway was not discovered until recently (Shinohara et al., 2019). These authors demonstrated that the "vir" cluster that is responsible for biosynthesis of volatile sesquiterpenes in *T. virens* (Crutcher et al., 2013), is the heptelidic acid biosynthetic gene cluster in *A. oryzae*. *A. oryzae* and *A. sojae* are used as koji mold along with a salt tolerant lactic acid bacterium *Tetragenococcus halophilus* and salt tolerant yeast *Zygosaccharomyces rouxii* for the preparation of soy sauce. The growth of salt- tolerant lactic acid bacterium *T. halophilus* was found to be inhibited by HA (antibacterial compound) produced by *A. oryzae* while *A. sojae* does not inhibit *T. halophilus* growth as they are not involved in HA

biosynthesis and does not possess the HA biosynthesis gene cluster. Comparison of the HA biosynthesis cluster genes (hep A-H, hepR and hepS) of *A. oryzae* with the *vir* cluster genes (vir 1-7, gapdh) of *T. virens* revealed that all the orthologues from the HA biosynthesis cluster of *A. oryzae* are present in the *vir* cluster of *T. virens* except the transcriptional regulators (hepR and hepS). BlastP analysis revealed that hepR and hepS genes from *A. oryzae* showed 36% and 89% identity to the TRIVIDRAFT_50650 and TRIVIDRAFT_134018 of *T. virens* which are also putative transcriptional regulators but not the part of the *vir* cluster. Functional assessment of all the genes in the HA biosynthesis gene cluster of *A. oryzae* was performed by gene deletion and LC/MS analysis for monitoring HA production. The terpene cyclase (hepA) and three cytochrome P450s (hepD, hepE and hepF) were found to be essential for HA biosynthesis as the deletion of these genes completely terminated the HA production while deletion of a monooxygenase (hepB) and another cytochrome P450 (hepC) partially affected the HA biosynthesis by reducing the HA production. However, the deletion of MFS transporter (hepG) and GAPDH (hepF) gene showed that these genes are not involved in HA biosynthesis. They have not detected the volatile metabolites production. Functional characterization by gene deletion and overexpression of both the transcriptional regulators hepR and hepS confirmed that they are crucial for the HA production. LC-MS analysis showed no production of HA in hepR and hepS deleted strains of *A. oryzae*. However, LC-MS analysis of strains overexpressing hepR and hepS showed distinct profile; hepR overexpressing strain produced four times HA in comparison to the wild type while hepS overexpressing strain did not show any change in HA production. Further analysis by qRT-PCR of hepS and hepR overexpressing strains revealed that both the transcription factors regulate the HA biosynthesis independently with different mechanisms. They have also studied the role GAPDH (HepG) present in the cluster by biochemical characterization, revealing that HepG is resistant to HA (similar to vGDP in *T. virens*) and its expression is

essential for survival of *A. oryzae* under the condition of HA production. To the best of our knowledge, the heptelidic acid biosynthesis genes/cluster is not reported yet in *T. virens*.

1.11. Pentalenolactone and its biosynthetic pathway:

Pentalenolactone is another sesquiterpenoid antibiotic which is an irreversible inhibitor of GAPDH by alkylating an active site cysteine residue and is very similar in structure to HA (Cane and Sohng, 1989, 1994, Hartmann et al., 1978). Both the compounds have an epoxide ring. This compound is produced by several species of *Streptomycetes* such as *S. chromofuscus, S. griseochromogenes, S. baarnensis, S. arenae* and *S. roseogriseus* and is toxic to a broad range of prokaryotic and eukaryotic microbes (Keller-schlerlein et al., 1972, Takeuchi et al., 1969, Martin et al., 1970). Similar in trend with the expression of HA-tolerant vGPD protein under the condition of HA biosynthesis in *T. virens*, *S. arenae* also expresses a pentalenolactone-insensitive form of GAPDH for self-resistance under the condition of pentalenolactone biosynthesis (Maurer et al., 1983, 1986). Pentalenolactone as an antibiotic against bacteria has been widely considered but due to a strong inhibition of human GAPDH, it did not receive much attention (Seidler et al., 2013).

Biosynthesis of pentalenolactone has been widely investigated and the first committed step in pentalenolactone biosynthesis is the cyclisation of farnesyl diphosphate (FPP, substrate for all the sesquiterpene synthases) to pentalene which is the parent hydrocarbon for pentalenolactone and related compounds. Several studies have identified different intermediates in the conversion of pentalene to pentalenolactone such as 1-deoxypentalenic acid, pentalenolactone D, pentalenolactone E, pentalenolactone F and shunt metabolites like pentalenic acid, *epi*-pentalenolactone F, pentalenolactone G, pentalenolactone H, pentalenolactone A, pentalenolactone B, pentalenolactone P, pentalenolactone O and pentalenolactone D (Cane et al., 1979 1981 and 1990). Cloning and expression in *E. coli* and crystal structure of pentalenene synthase from *S. exfoliatus* UC5319 established the

mechanism of action of the most important enzyme involved in the first committed step of the pentalenolactone biosynthesis (Cane et al., 1994).

The biosynthetic gene cluster for pentalenolactone has been identified in *S. exfoliatus* (named as *pen* cluster), *S. arenae* (named as *pnt* cluster) and in *S. avermitilis* (named as *ptl* cluster) (Seo et al., 2011). A putative pentalenolactone biosynthetic gene cluster has also been identified in *S. bingchenggensi* but production of pentalenolactone from this organism has not been reported yet (Seo at el., 2011). Further experiments with the aim of exploring the *pnt* and *pen* clusters (in *S. exfoliatus* and *S. arenae* respectively) revealed that both the clusters are responsible for biosynthesis of pentalenolactone. However, the *ptl* cluster in *S. avermitilis* is responsible for the biosynthesis of a new metabolite neopentalenolactone. Neopentalenolactone biosynthetic gene cluster (*ptl* cluster) includes 13 unidirectionally transcribed ORFs outspread in a 13.4 kb region (Tetzlaff et al., 2006). Each gene in the *ptl* gene cluster has been functionally assigned by biochemical characterization using recombinant proteins, and functions were further validated using several deletion mutants of *S. avermitilis* containing different segments of the *ptl* gene cluster with ermE as the promoter to control the expression (You et al., 2006, 2007; Jiang et al., 2009). However, both the *S. exfoliatus* UC5319 *pen* gene cluster and *S. arenae* TU469.56 *pnt* gene cluster for pentalenolactone biosynthesis contain 11 unidirectionally transcribed ORFs spread over approximately 13 kb sequence. All the orthologues of previously biochemically characterized *ptl* gene cluster were present in *pen* and *pnt* cluster except for the *penM* and *pntM* genes encoding for a P450 monooxygenase that is present only in *pen* and *pnt* gene clusters (Seo et al., 2011; Zhu et al., 2011).

The elucidation of the biochemical pathway for neopentalenolactone and pentalenolactone biosynthesis established that all the three gene clusters (*pnt*, *pen* and *ptl*) have common initial steps until the formation of an intermediate called 1-deoxy-11-

oxopentalenic acid (Seo et al., 2011). These common multiple steps of farnesyl diphosphate (FPP) to 1-deoxy-11-oxopentalenic acid are performed by the orthologous genes from *pen*, *pnt* and *ptl* clusters. The first enzyme pentalenene synthase (PenA/PntA/PtlA) cyclize farnesyl diphosphate (FPP) to the parent hydrocarbon i.e., pentalenene. In the subsequent step, cytochrome P450 (PenI/PntI/PtlI) performs oxidation of pentalenene in two steps to 1-deoxypentalenol and 1-deoxypentalenal (Seo et al., 2011; Tetzlaff et al., 2006; Quaderer et al., 2006). The same cytochrome P450 performs another oxidation of 1-deoxypentalenal to 1-deoxypentalenic acid. Hydroxylation of 1-deoxypentalenic acid to 1-deoxy-11-hydroxypentalenic acid is brought about by a non-heme iron / α-ketoglutarate-dependent dioxygenase (PenH/PntH/PtlH) (You et al., 2006, 2007; Seo et al., 2011). The latter compound was then subjected to oxidation by NAD^+-dependent short chain dehydrogenase/ reductase (PenF/PntF/PtlF) to form 1-deoxy-11-oxopentalenic acid (You et al., 2007). Other ORFs with defined functions on the basis of sequence similarity are farnesyl diphosphate synthase (PenB/PntB/PtlB), transcriptional regulator (PenR/PntR/PtlR), a transmembrane efflux protein (PenG/PntG/PtlG) and a pentalenolactone-insensitive GAPDH (gapN/gapR/gap1) (Seo et al., 2011). After the formation of 1-deoxy-11-oxopentalenic acid, *pen/pnt* pathway and *ptl* pathway splits and produces pentalenolactone and neopentalenolactone respectively (Seo et al., 2011). In *pen* and *pnt* pathways, a regio-specific flavin dependent Baeyer-Villiger oxygenase (penE/pntE) catalyzes the oxidation of 1-deoxy-11-oxopentalenic acid to pentalenolactone D. Pentalenolactone D is then converted to pentalenolactone E and F in a two-step saturation/epoxidation reaction by an α-ketoglutarate- and non-heme iron containing dependent dehydrogenase/oxygenase (penD/pntD) enzymes, resepectively (Seo et al., 2011). The final step to generate pentalenolactone from pentalenolactone F by oxidative rearrangement is performed by a cytochrome P450 (penM/pntM) (Zhu et al., 2011). In the *ptl* pathway, the regio-specific flavin dependent

Baeyer-Villiger oxygenase (ptlE) oxidizes 1-deoxy-11-oxopentalenic acid to neopentalenolactone D (an isomer of pentalenolactone D), which is subjected to further oxidation by α-ketoglutarate- and non-heme iron containing dehydrogenase/oxygenase (pltD) to form neopentalenolactone E and neopentalenolactone F (Jiang et al., 2009; Seo et al., 2011).

The *pen*, *pnt* and *ptl* gene cluster harbors a MarR-family transcriptional regulator designated as penR, pntR and ptlR respectively, which are positioned adjacent to the pentalenolactone-insensitive GAPDH in each pentalenolactone or neopentalenolactone gene clusters (Zhu et al., 2013). The regulatory role of penR in *Streptomyces exfoliatus* UC5319 has been affirmed by *penR* gene deletion approach, complementation of the *penR* gene and RT-PCR for the analysis of the expression of biosynthetic cluster genes affected by the deletion of the *penR* gene (Zhu et al., 2013). The functional study of PenR and PntR proteins using EMSA revealed that they act as an activator for the biosynthesis of pentalenolactone by binding to the pntR-gapR intergenic region (Zhu et al., 2013).

1.12. GAPDH inhibitors as agents for cancer therapy:

Glycolytic enzymes are potential targets of cancer therapy. One of the hallmarks of the cancer cells is metabolic reprogramming. Under normal condition with no oxygen deficiency, cells perform glycolysis followed by mitochondrial oxidative phosphorylation of pyruvate. While during hypoxia condition, oxidative phosphorylation is not performed by the cells and glycolysis is followed by fermentation process in which pyruvate is converted to lactate (Zhang et al., 2015). But cancerous cells perform glycolysis followed by fermentation even in the presence of sufficient oxygen and this phenomenon is known as the Warburg effect or aerobic respiration/glycolysis (Fan et al., 2019). Due to the Warburg effect, cancer cells have metabolic phenotypes like increased glycolysis and high rate of glucose uptake that

are used for clinical diagnosis of cancer by PET imaging. But only a few subtypes of malignancies treatment are available while broad class of cancers are still untreated and catastrophic. Discovery of small molecules like natural inhibitors which can target and modulate the biological pathways associated with priming cancer phenotypes could form the basis of an alternative therapy (Ganapathy-Kanniappan et al., 2012). For instance, anti-GAPDH therapeutic approach with natural inhibitor i.e. koningic acid or heptelidic acid has been recently demonstrated by Liberti and group (Liberti et al., 2017). However, targeting glycolytic enzymes like hexokinase, PFK and lactate dehydrogenase would result in anticancer effect (Krasnov et al., 2013), but GAPDH as a target has received maximum attention because it performs the important rate limiting step of glycolysis and produces first essential redox molecule i.e. NADH (Ganapathy-Kanniappan and Geschwind, 2013). Secondly, metabolomics demonstrated that aerobic glycolysis in cancer cells flux through GAPDH and thus is important regulatory enzyme in promoting the Warburg effect (Shestov et al., 2014). Lastly, being moonlighting protein, GAPDH is also associated with many cancer progressing pathways like, protection against caspase-independent cell death and increased proliferative index and cell cycle progression, etc. (Krasnov et al., 2013; Zhang et al., 2015).

Apart from koningic acid, many other anti-GAPDH compounds like iodoacetate, methylglyoxal, saframycin A, 3-bromopyruvate and oligonucleotides like siRNA and shRNA have been explored as candidates for cancer treatment in preclinical investigation. But the failure of these anti-GAPDH compounds has been due to the metabolic plasticity in cancer cells and unavoidable systemic toxicity (Ganapathy-Kanniappan and Geschwind, 2013). The required therapy should have three important concerns that have to be addressed while considering anti-GAPDH approach for cancer treatment. Firstly, the inhibitor of GAPDH should be highly specific and selective in targeting GAPDH. Secondly, in order to prevent systemic toxicity, loco-regional therapy should be encouraged. Thirdly, combinatorial

therapies of targeting other metabolic pathways along with anti-GAPDH or anti-glycolytic approach will help in avoiding metabolic plasticity (Ganapathy-Kanniappan et al., 2012).

The major advantage with targeting GAPDH would be targeting different cancer cell types, as glycolysis is a universal metabolic pathway. Thus, cancer therapy by GAPDH targeting should involve approaches which would inhibit glycolytic activity of GAPDH without affecting its pro-apoptotic feature essential for cancer cells killing and limiting drug tolerance. Liberti and group presented that inhibition of GAPDH in cancerous cells by koningic acid is a very effective therapy to treat cancer. The thermodynamic and kinetic study of the glycolytic flux in cancerous cells and normal cells showed that GAPDH possess different regulatory properties in cancerous cells due to overexpression during the Warburg effect. They also identified that cancer treatment with HA also targets other metabolic pathway associated with the cancer development and thus HA is a promising candidate to use as drug for cancer treatment (Liberti et al., 2017).

Active thiol group is essential for a wide range of moonlighting functions of GAPDH and one of the major functions is the pro-apoptotic activity by nuclear translocation under stress condition in both cancerous cells and normal cells (Guo et al., 2013). Another moonlighting function performed by GAPDH irrespective of the cell type is the protection of telomeres of the chromosome and avoiding cell senescence (Demarse et al., 2009). Hence, the therapies involved in targeting cysteine residue at the active site of the GAPDH using natural products and synthetic products might affect these essential events as well as the enzymatic activity of GAPDH. Thus, thiol-based targeting for cancer treatment should be considered in future.

Aims and objectives:

Objective 1. Heterologous expression of virGAPDH from *T. virens* in *E. coli* / yeast

Objective 2. Purification and elucidation of structure of virGAPDH.

Objective 3. Establishing the role of the virGAPDH in secondary metabolism in *T. virens*.

2.1. Materials:

Analytical and molecular biology grade reagents were used to perform different experiments. Buffers for protein purification, protein-DNA interactions and other studies were prepared in autoclaved Milli-Q water which was double-distilled (Millipore, USA). All the buffer solutions were filtered through a 0.44-micron filter (Millipore, USA) and degassed before performing experiments.

2.1.1. Chemicals, plastic wares and glasswares, and media:

Molecular biology grade chemicals, inorganic and organic salts and organic solvents were purchased from Sigma-Aldrich (USA), Merck (Germany), Roche Diagnostics (UK), Pharmacia (Sweden), Sisco Research Laboratories SRL (India), Himedia (India), Bangalore Genei (P) Ltd. (India), Promega (US) and GE Healthcare (US). Sterile disposable plastic Petri dishes, microtubes (0.2 mL, 0.5 mL or 1.5 mL), polypropylene oak ridge tubes (15 mL or 50 mL), micropipette tips (10 µL – 1.0 mL) were purchased from Axygen (USA) and Tarsons (India). Glass beakers, conical flasks and measuring cylinders were obtained from Borosil, India. Luria-Bertani broth, Luria-Bertani agar, Potato dextrose broth and Potato dextrose agar were procured from Himedia (India) and Sigma-Aldrich (USA).

2.1.2. Molecular biology reagents:

Restriction enzymes, DNA ligation kits, PCR reagents, T4 polynucleotide kinase, Isopropyl β- d-1-thiogalactopyranoside (IPTG), Lysozyme (from chicken egg-white), DNase I (from bovine pancreas) and RNaseA (from bovine pancreas), protein markers and DNA molecular weight standards were purchased from New England Biolabs (USA) and Sigma-Aldrich (USA). Bacterial plasmid DNA isolation kits, PCR purification kit and agarose gel DNA extraction kit were from Qiagen (Germany), or Sigma-Aldrich (USA).

2.1.3. Protein purification resins:

Immobilized metal affinity chromatography matrices were obtained from GE Healthcare (USA) and pre-packed Nickel-nitrilotriacetic acid (Ni-NTA) agarose resins were purchased from Roche Diagnostics (UK) or Sigma-Aldrich (USA). Ion exchange and gel filtration columns were obtained from GE Healthcare (USA).

2.1.4. Instruments:

Thermal cycler (Eppendorf, Germany), Centrifuge (Eppendorf, Germany), Shaker incubator (Orbitek, India), Circular Dichroism polarimeter (JASCO 815, USA), Fluorescence spectrophotometer (Horiba, USA), Zetasizer Nano ZS90 (Malvern, UK), GelDoc (Syngene, England), Nanodrop (GE Healthcare, USA), LightCycler® 480 System (Roche Diagnostics, Germany), BioRad CFX96™ Real-time system (Hercules, USA) and cell sonicator (Branson, USA) were used for various experiments.

2.2. Molecular biology grade antibiotics:

All the antibiotics were prepared in the chilled condition and were filter sterilized through a 0.22-micron filter (Millipore, USA).

Table 2.1 Antibiotics used for the growth of recombinant strains of *E. coli* and *T. virens*

Antibiotic	Stock solution (mg/mL)	Working concentration (µg/mL)
Ampicillin (Sigma)	100 mg/mL	100 µg/mL
Kanamycin (Sigma)	50 mg/mL	50 µg/mL
Tetracycline (Sigma)	20 mg/mL	20 µg/mL
Chloramphenicol (Sigma)	35 mg/mL	20 µg/mL
Hygromycin B (Sigma)	50 mg/mL	100-200 µg/mL

2.3. Media preparation:

Luria Bertani Medium (LB/LA):

Bacto-tryptone, 10 g; Bacto-yeast extract, 5 g; and NaCl, 10 g were dissolved in distilled water; pH was adjusted to 7.0 with NaOH solution and volume made to 1 litre. For LB agar, Bacto-agar was added to a final concentration of 1.5%.

Potato Dextrose Broth (PDB/PDA):

Potato infusion, 200 g and Dextrose, 20 g were dissolved in distilled water and final volume made to 1 litre. For PD agar, agar was added to a final concentration of 1.5%.

2.4. Stock Solutions of different chemicals:

1 M Tris-HCl solution (pH 7.6) - 121.14 g of Trizma base was dissolved in 800 mL of double distilled water (DDW) and pH was adjusted using conc. HCl. Volume was made to 1 litre. The solution was autoclaved and stored at room temperature (RT).

Phosphate buffered saline - 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4 and 0.2 g KH_2PO_4 were dissolved in 800 mL of DDW. The pH was adjusted to 7.4 with HCl and volume made to 1 litre. The solution was autoclaved and stored at RT.

5 M Sodium chloride (NaCl) solution - 292 g of sodium chloride was dissolved in 700 mL of DDW and volume made to 1 litre. The solution was filtered, autoclaved and stored at RT.

2.5 M Potassium chloride (KCl) solution - 46.60 g of potassium chloride was dissolved in 700 mL of DDW and volume made to 1 litre. The solution was autoclaved and stored at RT.

10 N NaOH solution - 4 g of NaOH pellets were dissolved in 10 mL autoclaved ultrapure milliQ water, stored at RT.

8 M Imidazole solution - 54.46 g of imidazole was dissolved in 80 mL of DDW. The pH was adjusted to 8.0 using NaOH pellets and volume was made to 100 mL. The solution was autoclaved and stored at 4 ºC.

1 M Magnesium chloride ($MgCl_2$) solution - 9.52 g of $MgCl_2$ was dissolved in 80 mL of DDW and volume made to 100 mL. Solution was autoclaved and stored at RT.

1 M Magnesium sulphate ($MgSO_4$) solution - 12.03 g of $MgSO_4$ was dissolved in 70 mL of DDW and volume made to 100 mL. Solution was autoclaved and stored at RT.

1 M Calcium chloride ($CaCl_2$) solution - 14.7 g of $CaCl_2$ dihydrate was dissolved in 70 mL of DDW and volume made to 100 mL. Solution was autoclaved and stored at 4 ºC.

0.5 M Nickel chloride ($NiCl_2$) solution - 23.77 g of $NiCl_2$ hexahydrate was dissolved in 150 mL of DDW and volume made to 200 mL. Solution was autoclaved and stored at RT.

0.5 M EDTA solution - 186.1 g of disodium EDTA.$2H_2O$ was dissolved in 800 mL of DDW and pH adjusted to 8.0 using NaOH pellets. The final volume was made to 1 litre, autoclaved and stored at RT.

0.5 M Dithiothreitol (DTT) solution - 1.54 g of Dithiothreitol powder was dissolved in 20 mL of autoclaved DDW and the solution was filter sterilized. Aliquots of 1 mL were stored at -20 ºC.

50% Glycerol solution - 50 mL of glycerol was mixed with equal volumes of DDW. The solution was autoclaved and stored at RT.

100 mM phenylmethylsulfonyl fluoride (**PMSF)** - 348 mg of PMSF powder was dissolved in 20 mL isopropanol/ absolute ethanol and vortexed vigorously to dissolve the contents. It was stored at -20 ºC.

1 M Isopropyl β- d-1-thiogalactopyranoside (IPTG) solution - 2.38 g of IPTG powder was dissolved in 7 mL of autoclaved DDW and volume made to 10 mL. The solution was filter sterilized and aliquots of 1 mL stored at -20 ºC.

Lysozyme (100 mg/mL) - 1 g of lyophilised lysozyme powder was dissolved in 10 mL of 10 mM Tris-HCl (pH 8.0) and stored at -20 ºC.

10 mM deoxyribonucleotide triphosphate (dNTP) solution (for PCR) - 10 µL each of the four dNTPs solutions (100 mM stock) were mixed and the volume was adjusted to 100 µL with autoclaved ultrapure MQ water and stored at -20 ºC.

Reagents for protoplast transformation:

A. Sterile Glasswares & Plasticwares:

i. Sorvall (SS34) tubes with cap
ii. Funnels with three layers of cheesecloth
iii. Filter assembly (500 mL)
iv. Glass test tube with cap
v. 50 mL flasks with cotton plug
vi. Funnel with Nytrex 50 plastic mesh
vii. 1 mL wide bore tips
viii. Spatula

B. Sterile solutions:

i. STC (Sorbitol tris $CaCl_2$): (500 mL)

1.2 M Sorbitol	109.32 gm
10 mM Tris (pH 7.5)	5 mL of 1M
50 mM $CaCl_2$	25 mL of 1M

ii. Polyethylene Glycol (PEG) solution: **(100 mL)**

PEG (60%) (Mol.wt 3350)	60gm
10 mM Tris (pH 7.5)	1 mL of 1M
50 mM $CaCl_2$	5 mL of 1M

iii. Regeneration Media A (50 mL)

Yeast extracts	0.5 gm
Casein hydrolysate (enzymatic)	0.5 gm
Water up to	50 mL

iv. Regeneration Media B (450 mL)

Sucrose	171 gm
Agar	8 gm
Water up to	450 mL

v. Regeneration media Mix (A+ B) (500 mL)

Add 50 mL media A to 450 mL media B. The flask should be autoclaved separately, combined and kept at 55 ºC before use.

vi. Potato Dextrose Broth 100 mL

vii. 0.7 M NaCl 80 mL

viii. 1% Agar 100 mL

ix. Hygromycin B in Phosphate buffered saline (PBS) 50 mg/mL

Reagents for isolation of genomic DNA:

i. Extraction Buffer (100 mM Tris HCl, pH 8.0; 50 mM EDTA, pH 8.0; 500 mM NaCl; 10 mM 2-Mercaptoethanol)

ii. 20% SDS

iii. 5 M Potassium Acetate

iv. TE Buffer (10 mM Tris, 1 mM EDTA, pH 8.0)

v. Chloroform, AR

vi. Isopropanol, AR

vii. 70% Ethanol, AR

Reagents for Agarose Gel Electrophoresis:

10 x TBE (Tris borate EDTA) - 108 g of Tris base, 55 g of boric acid (borate) and 40 mL of 0.5 M EDTA (pH = 8.0) were resuspended in distilled water and the volume was made up to 1 litre.

50X TAE (Tris acetate EDTA) buffer - 242 g of Tris base was dissolved in 700 mL of DDW and 57.1 mL of glacial acetic acid was added to it. Further, 100 mL of 0.5 M EDTA at pH 8.0 was added and pH was adjusted to 8. Buffer was autoclaved and stored at RT.

6X DNA loading dye - 25 mg bromophenol blue or xylene cyanol and 4 g of sucrose were dissolved in autoclaved DDW and volume was made to 10 mL. The dye was autoclaved and aliquoted before storing at -20 ºC.

Ethidium bromide solution (10 mg/mL) - 100 mg of ethidium bromide (EtBr) was dissolved in 10 mL of autoclaved ultrapure MQ water, aliquoted and stored at RT.

Reagents for SDS-Polyacrylamide Gel Electrophoresis:

30% Acrylamide solution - 29.2 g of acrylamide and 0.8 g of N, N'-methylenebisacrylamide was dissolved in 60 mL of DDW and volume made to 100 mL. The solution was filtered in dark and stored at 4 ºC in dark bottle.

1.5 M Tris-Cl, pH 8.8 - 90.75 g of Tris base was dissolved in 400 mL of DDW. The pH was adjusted to 8.8 with conc. HCl and volume was made to 500 mL. The solution was autoclaved and stored at RT.

0.5 M Tris-Cl, pH 6.8 - 30.25 g of Tris base was dissolved in 400 mL of DDW. The pH was adjusted to 6.8 with conc. HCl and volume was raised to 500 mL. The solution was autoclaved and stored at RT.

10% Sodium dodecyl sulphate (SDS) solution - 10 g of SDS was dissolved in 100 mL of DDW and stored at RT.

10% Ammonium per sulphate (APS) - 1 g of APS was dissolved in 10 mL of autoclaved DDW, vortexed to mix and stored in dark at 4 ºC.

10X Tris-Glycine-SDS Buffer - 30 g of Tris base and 144 g of glycine was dissolved in 800 mL of DDW. The pH was adjusted to 8.8. Further, 10 g SDS was dissolved to this and volume was made to 1 litre, stored at RT.

2X Leammli dye (Cracking buffer) - 460 mg SDS, 7.6 mg ethylene glycol bis (2-aminoethyl ether) tetraacetic acid (EGTA), 20 mg sodium azide, 2 mL 100% glycerol, 2.5 mL 0.5 M Tris-HCl pH 6.8, 1 mL beta-mercaptoethanol, 112 µL of 100 mM PMSF and 40 µL of 5% bromophenol blue were dissolved in autoclaved DDW to a final volume of 10 mL, stored at RT.

Fixing solution - 500 mL methanol and 100 mL glacial acetic acid were added to 400 mL DDW and stored at RT.

Staining solution - Coomassie Brilliant Blue R250 at final concentration of 0.5% was dissolved in fixing solution with continuous shaking overnight. The solution was filtered through Whatman filter paper No. 1 and stored at RT.

Destaining solution - 100 mL methanol, 100 mL glacial acetic acid and 20 mL glycerol were mixed in 780 mL of DDW and stored at RT.

Reagents for blotting:

Transfer buffer for semi dry blot (10X) - 24.28 g of Tris base (200 mM) and 144.14 g of glycine (1920 mM) was dissolved in 800 mL of DDW and pH was adjusted to 8.3. Volume was raised to 1litre with DDW.

Tris buffered saline (TBS) - 20 mL of 1 M Tris-Cl pH 7.6 and 200 mL of 5 M NaCl solution were added to 780 mL DDW, mixed well and stored at 4 ºC.

Tween 20 Tris buffered saline (TTBS) - 0.1% Tween-20 was in TBS buffer before use.

Blocking buffer/ Antibody buffer - 3 g of skimmed powder was dissolved in TBS by vortexing. Prepare freshly before use.

Alkaline Phosphatase buffer; pH 9.5 (Developing buffer) - 100 mM Tris-HCl, pH 9.5, 100 mM NaCl and 50 mM $MgCl_2$ in 500 mL DDW were made and stored at RT.

Solutions for plasmid isolation:

Solution I- (glucose-tris-EDTA (GTE) solution; resuspension buffer) - GTE solution was prepared by diluting the respective stock solutions to 50 mM Glucose, 25 mM Tris-HCl (pH, 8.0); 10 mM EDTA (pH, 8.0). Solution was autoclaved and stored at 4 ºC.

Solution II – Lysis buffer - It was prepared freshly before use by diluting the respective stock solutions to 0.2 N NaOH (from stock of 10 N) and 1% SDS (from 10% stock).

Solution III - 5 M potassium acetate (Neutralization solution) - 59 g of potassium acetate was dissolved in 100 mL Milli Q grade water and 23 mL of glacial acetic acid was added to

get the final pH of 4.8. The volume was made up to 300 mL with sterile water. Solution was autoclaved and stored at 4 ºC.

2.5. Methods:

2.5.1. Recombinant DNA cloning:

Different recombinant constructs were made in the present study for protein expression and generation of the deletion mutant by following the protocols as described in (Green et al., 2014). The following steps were performed for the construction of these recombinant plasmids.

2.5.1.1. Plasmid purification:

Plasmid DNA was prepared using the High Pure Plasmid Isolation kit (Sigma-Aldrich, USA) which is based on the alkaline lysis method. In brief, the bacterial cells containing the desired plasmid were grown in 10 mL LB medium at 37 ºC overnight with antibiotics if required. The bacterial cells were pelleted and resuspended in 250 µL suspension buffer containing RNase. 250 µL of lysis buffer was added and mixed gently at room temperature followed by incubation for 5 minutes. Next, a chilled binding buffer was added and incubated on ice for 5 minutes. The cloudy suspension was then centrifuged at 13,000xg for 20 minutes and the supernatant was transferred to the high pure filter tube inserted into the collection tube. The tube was centrifuged at 13,000xg for 1 minute at room temperature and the supernatant was discarded. Washing of the tube was done using 700 µL of wash buffer and centrifuged at 13,000xg for 1 minute. Again, dry spin was performed to allow removal of the residual wash buffer. The filter tube was transferred into a fresh 1.5 mL microcentrifuge tube. The plasmid DNA was eluted by adding elution buffer (30 µL) to the filter tube and centrifuged at 13,000xg for 1 minute at room temperature. The plasmid DNA concentration

and purity were determined spectrophotometrically by measuring OD_{260} and OD_{280}. The A_{260} value indicates plasmid DNA concentration with A_{260} value 1 corresponds to 50 µg/mL DNA and the A_{260}/A_{280} ratio indicates the purity of the plasmid DNA.

2.5.1.2. Primer designing, synthesis and PCR amplification:

The corresponding nucleotide sequences for primer designing were taken from the fungal genome portal MycoCosm (http://www.genome.jgi-psf.org). Primers were designed manually and with the following parameters: primer length (18-25 bp), GC content (40-60 %), Tm (50-65 ºC), Self-dimer & hetero-dimer tendencies were considered using OligoAnalyzer Tool (https://www.idtdna.com/pages/tools/oligoanalyzer). The presence of restriction enzymes (RE) sites in the DNA insert was ascertained using NEB cutter online tool (http://nc2.neb.com/NEBcutter2/). Unique restriction enzyme sites were incorporated in the forward and reverse primer sequences as per requirement. Overhanging nucleotides required for restriction enzyme cutting efficiency were added at the 5′ ends of both the primers to facilitate digestion. The primers were commercially synthesized from manufacturers.

PCR amplification of the desired gene was performed with gene-specific primers using high fidelity Phusion DNA polymerase enzyme or AccuTaq™ LA DNA polymerase (Sigma, USA). Taq DNA polymerase enzyme was used for positive recombinant selection in 25 or 50 µL reaction mixture volume. The primer-template complementarity and GC content were used as a parameter for annealing temperature determination, and extension time were determined based on the length of the amplified product. The parameters used for PCR amplification were: denaturation at 95 ºC for 5 minutes, annealing at 55-65 ºC for 45 sec, extension at 72 ºC for 0.5 kb/minute, repeat cycles from step 2 to 4 for 32 cycles, final extension at 72 ºC for 10 minutes and hold at 4 ºC. The final amplified PCR product was

qualitatively analyzed in 0.8−1 % agarose gels in 1X TAE buffer with 0.5 µg/mL concentration of ethidium bromide.

2.5.1.3. Restriction digestion:

The restriction digestion of the PCR-amplified gene of interest and plasmid with the same restriction enzymes was performed as per protocol provided by the manufacturers (New England Biolabs, USA). The restriction digestion was performed using 10 units of restriction enzyme per µg DNA in 50 µL reaction for 3-4 h or overnight at 37 ºC. Single digestions with individual restriction enzymes and double digestion with both restriction enzymes was performed to ensure the proper digestion of the plasmid DNA. The double digested linearized plasmid DNA and PCR amplified gene of interest were agarose gel purified and used for further experiments.

2.5.1.4. Ligation:

Ligation of the digested and purified linearized plasmid DNA and the insert was done using T4 DNA ligase (New England Biolabs, USA). The insert to vector molar ratios was set to 3:1 to 5:1 and the maximum amount of insert used was 200 ng. Both plasmid DNA and insert were precipitated with sodium acetate salt and absolute ethanol. The precipitated DNA pellet was washed with 70% ethanol, dried and dissolved in autoclaved Milli-Q water. This DNA mixture was added to the ligation reaction mixture (15 µL in volume) containing T4 DNA ligase buffer (1X), T4 DNA ligase (1 µL) and 1 mM ATP. The ligation was carried out at 16 ºC overnight. The ligated product was used for transformation in a bacterial host.

2.5.1.5. Ligation independent cloning:

In the present study, apart from the traditional ligase based cloning method, gene cloning was also performed with the ligation independent cloning (LIC) method. In this

method instead of ligase enzyme, 15 base overhangs were used for annealing and the vector used was pNH-TtrxT (procured from Addgene, plasmid no. 26106; a gift from OpherGileadi). These overhangs were created by 3'−5' activity of T4 DNA polymerase in the expression vector and insert DNA. The primers were designed carefully to create overhangs in the insert complementary to the LIC vector. The final annealed product was transformed into competent cells and positive transformants were screened through colony PCR.

2.5.1.6. Bacterial transformation:

Competent cells required for the transformation of bacterial hosts were prepared using the calcium chloride method. In brief, the overnight grown primary culture of the desired *E. coli* strain (such as DH5α or NovaBlue strain) was diluted to 100-fold in 20 mL fresh LB medium and allowed to grow at 37 ºC (180 rpm) until $O.D_{600}$ 0.3-0.4. From this step onward all the steps were performed in laminar airflow and chilled condition only. The culture was then transferred to SS34 tubes and chilled on ice for 45-60 minutes. The chilled culture was centrifuged at 4,000xg for 5 minutes at 4 ºC and the pellet was gently resuspended in 10 mL of chilled 50 mM $CaCl_2$. The re-suspended bacterial cells were incubated on ice for 45 minutes and centrifuged at 4,000xg for 10 minutes at 4 ºC. Further, the bacterial cells were gently resuspended in 1 mL of 50 mM $CaCl_2$ and incubated on ice for 45 min. The competent cells so prepared were aliquoted (100 µL each) and stored at –70 ºC in 30% glycerol for a maximum of one month. For transformation, the aliquoted competent cells were mixed with the ligation mixture and incubated on ice for 45 minutes. Thereafter, heat shock was given to the transformation mixture for 2 minutes at 42 ºC followed by incubation on ice for 5 minutes. The transformation mixture was revived by adding 900 µL of 1X LB broth medium and incubated at 37 ºC for 45 minutes. Two different dilutions of transformation mixtures were

plated on LB agar plate with appropriate antibiotics and incubated at 37 ºC for overnight. The transformed bacterial cells were screened by colony PCR after growing on grid-plate.

2.5.1.7. Colony PCR:

After bacterial transformation, the positive transformants were selected by colony PCR. The transformants to be screened were streaked onto fresh LB agar plates (known as grid transfer) with appropriate antibiotics and incubated at 37 ºC overnight. The PCR reaction mixture components and conditions were kept the same as described in section 2.3.1.2 except that the template used was a colony of the bacterial transformant suspended into the PCR reaction mixture. Gene-specific primers were used for the amplification of the gene of interest for the screening of the positive transformants. Positive clones were selected after checking on 1% agarose gel and the plasmid was isolated and the presence of the correct insert was verified with restriction digestion and sequencing.

2.5.2. Protoplast transformation of *Trichoderma virens*:

The protoplast transformation of *T. virens* includes three major steps i.e. isolation of protoplast using enzyme mixture, the transformation of protoplasts with the required plasmid by the PEG-$CaCl_2$ mediated method and purification of the transformants, as detailed below.

2.5.2.1. Protoplast isolation:

The germinated conidia were collected from an overnight grown culture of *T. virens* in PDB at 28 ºC and filtered through three layers of sterile cheesecloth. The harvested conidia were washed twice with sterile distilled water. The enzyme mixture containing driselase: 400 mg, glucanase: 200 mg, chitinase: trace amount, lyticase: trace amount and cellulase: trace amount was suspended in 0.7 M NaCl, mixed properly, centrifuged at 13,000xg for 10

minutes and filter sterilized. The enzyme mixture was used for hydrolyzing the washed conidia in 50 mL flasks by incubating at 28 ºC with gentle shaking at 50-60 rpm for 2 hours. Microscopy was used to ascertain the release of protoplasts. Protoplasts were harvested by filtering through three layers of cheesecloth twice, followed by filtering through a single layer of Nytrex 50 membrane and then transferred to sterile SS 34 tubes for centrifugation at 9,000xg for 10 minutes at 4 ºC. The pellet was gently resuspended in 10 mL of ice-chilled 0.7 M NaCl using 1 mL cut tips and centrifuged at 9,000xg for 10 minutes at 10 ºC. The pellet was again gently resuspended and washed in 10 mL ice-cold STC medium using 1 mL cut tips and centrifuged at 9,000xg for 10 minutes at 10 ºC. Finally, the protoplasts were resuspended in 1 mL or 0.5 mL of ice-chilled STC medium and used for transformation with the PCR amplified linear DNA.

2.5.2.2. Transformation of protoplasts:

For protoplast transformation, 100 µL of protoplasts and 15-20 µL of linear DNA of both flanks in a 1:1 ratio (about 5 µg) were mixed in a glass tube and incubated on ice for 12 minutes. Two hundred µL of ice-chilled PEG was added to the solution and mixed gently with a 1 mL cut tip. Again 200 µL of ice-chilled PEG was added to the solution and mixed gently followed by the addition of 800 µL of ice-chilled PEG and mixed gently with 1 mL of cut tips. This solution was kept for 5-6 minutes on ice and then 1 mL of ice-chilled STC was added. The transformation mixture was then transferred to sterile Petri plates, overlaid with 20 mL of regeneration medium (with the temperature maintained at 52-55 ºC), mixed by gently swirling and finally incubated at 28 ºC overnight. The next day, the plates were overlaid with 10 mL of 1% agar containing 600 ppm of hygromycin B (making the final concentration of hygromycin B to be 200 ppm) and incubated at 28 ºC until the emergence of the transformed colonies from the surface of the agar. These colonies (putative transformants) were then transferred to PDA slants containing 200 ppm hygromycin B for further purification.

2.5.2.3. Purification of transformants:

Purification of the colonies was required to screen the homologously recombined transformants and eliminating the false-positive ectopically recombined transformants. Purification of the transformants was performed by repeated sub-culturing on PDA plates with 200 ppm hygromycin B. The colonies that survived were further sub-cultured on PDA plates without Hygromycin B and then back to PDA with hygromycin B. Only homologous recombined transformant colonies were able to grow for several repeated sub-culturing but the ectopically recombined transformed colonies were not able to sustain because of loss of hygromycin B resistance. Wild type was also grown on PDA with and without Hygromycin B as a control. The pure transformants were grown in PDB containing 50 ppm Hygromycin B for the isolation of genomic DNA which was analyzed further through PCR for the presence of homologous recombination.

2.5.3. Genomic DNA isolation from the transformants of *T. virens*:

Isolation of chromosomal DNA of the transformants was done for screening of transformants showing homologous recombination by PCR. Genomic DNA isolation was performed using the SDS-lysis-chloroform method (Green et al., 2014). In brief, the selected transformants were grown in PDB medium (100 mL) for 3 days at 28 ºC and 180 rpm. After 3 days, fungal mycelium was filtered through Whatman no.1 filter paper, dried with blotting paper, and crushed in a porcelain mortar and pestle with liquid nitrogen. Powdered mycelia were transferred to 1.5 mL microcentrifuge tubes containing 500 µL extraction buffer. The tubes were left open for a while and then vortexed vigorously. Seventy µL of 10 % SDS was then added and mixed gently by inversion followed by incubation on the heating block at 65 ºC for 10 minutes. 170 µL of 5 M potassium acetate was added and mixed gently by inversion followed by incubation of the tubes on ice for 20 minutes. Five hundred µL of chloroform

was added, mixed by inversion and centrifuged at 13,000xg for 10 minutes at room temperature. The supernatant was collected in a fresh microcentrifuge tube and 500 µL isopropanol was added to the supernatant and mixed gently. Precipitated DNA was then centrifuged at 1,000xg for 5 minutes and the supernatant was discarded. Washing of the DNA pellet was done with 70% ethanol followed by centrifugation at 13,000xg for 5 minutes at room temperature. The DNA pellet was air-dried and dissolved in TE buffer (pH 8.0) by incubating in a heating block at 50 ºC. The quality of the DNA was checked on 1% agarose gel in 1X TAE buffer and visualized with 0.5 µg/mL concentration of ethidium bromide.

2.5.4. Agarose gel DNA extraction:

Molecular biology grade low melting point agarose (Sigma, USA) was used to separate the DNA, and after electrophoreses the desired band of DNA was spliced out with a sterile blade under a UV lamp. Agarose gel extraction kit (Sigma, USA) was used for DNA extraction as described by the manufacturer. In brief, isolated gel pieces were dissolved in 3 volumes of suspension buffer and incubated at 50 ºC till the gel pieces got completely dissolved. One volume of isopropanol was added to the suspension and the mixture was transferred to the DNA filter column inserted in the collection tube. The column was centrifuged for one minute at room temperature and then washed twice with wash buffer. Residual wash buffer was removed by dry spinning the column for one minute and finally, DNA was eluted with pre-warmed autoclaved MilliQ water.

2.5.5. RNA isolation from *T. virens*:

TRI reagent (Sigma, USA) was used for RNA isolation. TRI reagent is an acid guanidium-phenol based reagent with guanidium salt involved in denaturation of proteins like a chaotropic agent, and phenol, which is involved in the extraction of nucleic acid and

proteins like an organic compound. Chloroform is used for phase separation of the solution into aqueous and organic phases. Low pH helps in the separation of RNA from DNA and protein. Subsequently, high-quality total RNA is recovered by alcohol precipitation.

The fungal cultures for RNA isolation were grown for 2-3 days in PDA and then fresh mycelium was used for inoculation in PDB medium (100 mL). The fungus was allowed to grow for 3 days at 28 ºC and 180 rpm. The mycelia were harvested by filtration on Whatman no.1 filter paper and blotted dry to remove excess water. The mycelium was grounded to a fine powder in liquid nitrogen. One mL Tri reagent was added to approximately 100 mg of the sample in a 2 mL Eppendorf tube and allowed to thaw for a while with a cap opened. Three hundred microliters of chloroform was added to the solution and mixed by inverting the tubes for 5-6 times. After 5 minutes incubation, the tubes were centrifuged at 13,000xg for 5 minutes. Aqueous phase was collected in another 1.5 mL tube followed by addition of 500 µL of isopropanol was added and centrifuged at 13,000xg for 15 mins at 4 ºC. The RNA pellet was washed with ice-cold 75% ethanol and dried at room temperature or 37 ºC. Lastly, the pellet was re-suspended in the nuclease-free DEPC treated water and stored at -20 ºC for future use.

2.5.6. Agarose gel electrophoresis:

Agarose gel electrophoresis was used for the separation of DNA. The buffer used for electrophoreses was 1X TAE (Tris-acetate EDTA) buffer. Agarose gel (0.8-1.5 %) was used for the separation of PCR amplified product and plasmid DNA. Agarose (Sigma-Aldrich, USA) powder was dissolved in 1X TAE buffer by heating, and after the solution reached the temperature around 45-50 ºC, ethidium bromide (0.5 µg/mL) was added before pouring the solution into the casting tray. The solution was allowed to solidify at room temperature and after solidification, the DNA samples were loaded with gel loading dye. The electrophoresis

apparatus was allowed to run at 90 volts till the loading dye reached 70% of the gel length. The gel documentation system was used for visualization and analysis of the DNA bands separated.

2.5.7. Protein expression and purification:

2.5.7.1. Overexpression of recombinant protein using bacterial host system:

A strong and controlled expression of the protein in a bacterial host system without inflicting toxicity to the host is a requirement for purification and biochemical, and structural characterization of any protein. Several bacterial host strains are available for efficient expression of proteins. The expression vectors are generally available with strong inducible promoters e.g., Lactose operon induced by the addition of IPTG (a non-hydrolyzable inducing agent and analog of allolactose). In the absence of the inducer in the growth medium, the lac promoter gets repressed by the Lac repressor protein (LacI). Otherwise, the presence of an inducer prevents the binding of Lac repressor from binding to the Lac operator, resulting in over-expression of the genes. In the present study, *E. coli* BL21 Star (DE3) strain was used as a protein expression host and pNH-TtrxT vector was used as an expression vector. The gene of interest was cloned in pNH-TtrxT by ligation independent cloning and transformed in the cloning host (DH5α or Nova Blue strains) followed by isolation of cloned plasmid for transformation in protein expression host i.e., *E. coli* BL21 Star (DE3) strain. *E. coli* BL21 Star (DE3) containing the recombinant plasmid was grown in 50 mL LB media supplemented with required antibiotics at 220 rpm and 37 ºC overnight. The overnight grown culture was diluted in fresh LB medium supplemented with a required antibiotic in 1:100 ratio and allowed to incubate at 37 ºC for 3-4 hours until all the cells reached mid-log phase, i.e., A_{600} approximately 0.6. The protein expression was induced by adding 0.5-1.0 mM IPTG followed by shaking at 22 ºC overnight. The induced cells were harvested by centrifugation for 15

minutes at 4,000xg in a pre-cooled centrifuge (a small aliquot of the cells was taken to check for the induction of the protein) and the cell pellet was stored at -80 ºC till further use. For protein expression analysis, induced and un-induced cells were re-suspended in 20 µL of TE buffer, and an equal volume of 2X Laemmli SDS dye was added to the solution. The solution was heated at 95 ºC for 10 minutes and centrifuged for 10 minutes at 13,000xg . Cell extracts of induced and un-induced cells along with the protein marker were resolved on 12% reducing SDS-PAGE followed by staining with coomassie brilliant blue.

2.5.7.2. SDS-PAGE analysis:

SDS-PAGE is used for resolving proteins by size, and estimation of the molecular weight of the proteins. Sodium dodecyl sulfate (SDS) is an ionic detergent that neutralizes the charge on the proteins and thus negatively charged SDS-protein complex migrates and resolves in the polyacrylamide gel electrophoresis (PAGE) irrespective of their charges but based on their masses only. SDS-PAGE analysis of proteins involves the following components as described in Table 2.2. Usually, 10-12% of the SDS-PAGE is used for protein separation. The protocol for SDS-PAGE preparation involved assembling of glass plates with 0.1 cm thick spacers in a vertical gel electrophoresis apparatus. Resolving gel solution was poured into the glass plate sandwich and overlaid with isopropanol to avoid contact with oxygen. After polymerization of the resolving gel, isopropanol was removed and the gel was washed with water. Next, stacking gel was poured into the glass plate sandwich and immediately a comb of similar size as the spacer was placed carefully and allowed to polymerize. The comb was removed and the glass plate was transferred to the electrophoresis chamber filled with electrophoresis buffer composed of 10% SDS-Tris-Glycine, pH 8.8. The samples were prepared by adding an equal volume of 2X SDS gel loading dye to the samples and heated for 10 minutes at 95 ºC. After a hard spin, the samples were loaded into the wells and the gel was run at 20 mA of constant current or 100 V until the BPB tracking dye entered

the separating gel and then at 30 mA or 200 V until the BPB dye reached the bottom of the gel. The gel was then stained with Coomassie Brilliant Blue R-250 (CBBR) staining solution (0.1% w/v CBBR, 50% methanol and 10% acetic acid in water) for ~1 hour on the shaker, followed by destaining in a solution containing 50% methanol and 10% acetic acid in water until clear protein bands appeared on the gel.

Table 2.2 Composition of 12% SDS-PAGE gel

	Resolving gel	**Stacking gel**
Components	10%	5%
Acrylamide (30%)	8.0 mL	830 µL
Tris-HCl	5.8 mL (pH 8.8; 1.5 M)	630 µL (pH 6.8; 1M)
10% SDS	200 µL	50 µL
10% Ammonium per sulphate	200 µL	50 µL
TEMED	20 µL	10 µL
Water	7.0 mL	3.4 mL
Total volume (mL)	20 mL	5 mL

2.5.7.3. Protein purification:

Purification of protein depends on protein properties like size, charge, affinity to its tag, etc. The most common technique used for protein purification from the cell lysate is affinity chromatography because it allows fractionation of the protein with high purity. The size-based purification approach is size exclusion chromatography, while charge based

purification approach is ion-exchange chromatography. Performing multiple purification techniques achieves higher purity of the protein suitable for downstream experiments.

Protein purification protocol involved re-suspension of the induced bacterial cell pellet in lysis buffer A (1 gm mL^{-1}) containing 25 mM Tris-HCl, pH 8.0; 300 mM NaCl; 1 mM PMSF, 10 % glycerol and 1 mM DTT. Lastly, lysozyme was added to the suspension to a final concentration of 1 mg mL^{-1}. The suspension was incubated in ice for 45 min, followed by sonication at 40% duty cycle for 15 minutes with 10 sec ON and 15 sec OFF mode. The suspension was centrifuged at 21,000xg for 40 minutes and the clear supernatant was collected in a fresh autoclaved Sorvall tube for further purification steps. The Ni-NTA resin was washed with a 5 column volume of water to remove residual ethanol because resin was stored in 20% ethanol at 4 ºC. Washed resin was then pre-equilibrated with 5 column volumes of the binding buffer A containing 15 mM imidazole (pH 8.0) to avoid nonspecific binding of the protein to the resin. The purification column (Biorad, USA) was then packed manually with the pre-equilibrated resins and was allowed to settle under gravity. Once the column was prepared with intact pre-equilibrated resins, the clear supernatant of the cell-free extract containing protein of interest with poly-histidine tag was allowed to flow through the column by gravity. The flow-through from the column was collected. The column was washed with 20 column volume of buffer A containing 50 mM imidazole to remove non-specifically bound proteins from the affinity column and the wash-through was collected. Elution of the target protein was performed by the gradient method by washing the affinity resins with buffer A containing an increasing concentration of imidazole from 100 mM to 500 mM. All the fractions including induced, uninduced, wash-through, flow-through, and different fractions of elution were loaded and resolved in 12% SDS-PAGE and stained with coomassie brilliant blue.

2.5.7.4. Estimation of protein concentration:

In the present study, nanodrop and UV spectrophotometer from JASCO were used which are based on measurement of the absorbance maxima of the proteins at 280 nm following Beer-Lamberts law. Other methods like the Bradford method and modified Folin-Lowry method based on the colorimetric method were also employed for determining the protein concentration.

Bradford's method was performed as described in the protocol provided by the manufacturers. In brief, a mixture of 200 µL of 5X Bradford dye solution and 800 µL of protein solution was incubated at room temperature for 30 minutes. The color that developed in the solutions after incubation was monitored in a spectrophotometer at a wavelength of 595 nm and the OD_{595} was compared with the standard curve made from a standard concentration of Bovine Serum Albumin protein for estimation of protein concentration in the solution.

2.5.7.5. Western blot analysis:

Levels of particular proteins or their post-translational modifications were assessed by western blot analysis using the appropriate antibody.

Electroblotting from SDS-PAGE: Protein (5-10 µg) was electroblotted from SDS PAGE gels to polyvinylidene fluoride or polyvinylidene difluoride (PVDF) membranes for western blot analysis. The transfer tank of electroblotting apparatus (Trans-Blot Cell, Bio-Rad) was filled with 1x transfer buffer. PVDF membrane was activated in 100% methanol for 5 sec. The activated membrane and SDS-PAGE gel were equilibrated in 1x transfer buffer. The gel membrane transfer sandwich was prepared and inserted into the transfer tank with gel on cathode side and membrane on anode side. Transfer was conducted at a constant current of 300 mA for 200 min. Proteins transferred onto the membrane were detected by staining with

Ponceau S (0.5% w/v Ponceau S in 1% v/v acetic acid) and destaining with several changes of water.

Immunoblot detection: Proteins transferred onto PVDF membrane were probed with antibodies and was incubated in blocking buffer in Tween 20/Tris-buffered saline (TTBS) for 1 hour at room temperature on orbital shaker. Blocking buffer was then replaced by recommended dilutions of primary antibodies in TTBS and incubated for 1 hour at room temperature in orbital shaker. The membrane was washed four times with TTBS for 20 minutes each at room temperature. Further the membrane was incubated in recommended concentrations of secondary antibodies in TTBS for 1 hour at room temperature on orbital shaker. The membrane was again washed four times with TTBS at room temperature and developed using Immobilon Western (Millipore, USA, cat#P90719). The membrane was exposed to X-ray film in dark room that was developed using Optimax X-ray film processor (Protec). Band intensities were analysed using ImageJ and GraphPad prism software.

2.5.7.6. Mass spectrometry analysis:

Matrix-Assisted Laser Desorption Ionization Time-of-flight (MALDI-TOF) is a method for the identification of the protein through peptide mass fingerprinting and molecular weight of the protein. The MALDI-TOF analysis was performed at ACTREC, Mumbai. The protein sample for MALDI-TOF was prepared by diluting the protein to a final concentration of 100 pmol/µL and mixing it with sinapinic acid matrix. The copper grid was used for the spotting of the above protein-matrix mixture and dried at room temperature. The dried copper grid was loaded on the spectrometer and subjected to external calibration using the Peptide Standard Calibration II. FlexAnalysis software was employed for spectra processing and analyzed as recommended by the manufacturer. The protein was separated in 12% SDS-PAGE followed by excision of the desired band. The excised protein band was trypsinized

and processed for Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry to ascertain the identity of the protein.

2.5.8. Protein oligomerization state:

The protein purified from affinity chromatography was further subjected to additional purification based on the size of the protein using gel filtration chromatography. For gel filtration chromatography, 1 mL of concentrated protein (about 1 mg mL^{-1}) was used for determining the molecular weight and oligomeric state of the protein. The column used in the present study was Superdex™ 200GL column (Pharmacia, Sweden) on AKTA purifier (GE Healthcare, USA) which was pre-calibrated with the known molecular weight markers (Chymotripsinogen- 25 kDa, Ovalbumin- 44 kDa, Bovine Serum Albumin – 66.5 kDa, Aldolase – 158 kDa, Catalase- 250 kDa (Sigma)). Based on elution volume, the standard curve was plotted with elution volume against the logarithm of the molecular weight of markers. The fractions eluted from the column by monitoring the UV absorbance at λ_{280} nm were collected in 1.5 mL Eppendorf tubes and checked on SDS-PAGE for further analysis.

2.5.9. Electrophoretic mobility shift assay (EMSA):

Electrophoretic mobility shift assay (EMSA) was used to study the interaction of the protein with nucleic acids. EMSA is based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is retarded compared with that of the free nucleic acid. Different variants of EMSA are available based on detection methods like chemiluminescence, immunohistochemical, fluorescence-based methods but detection using radioisotope-labeled nucleic acids is the most sensitive one and requires very low concentrations of protein and nucleic acid. The steps adopted for performing EMSA with protein and radiolabeled DNA are described below.

2.5.9.1. Radiolabeling of DNA:

Double-stranded DNA was labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase as described in Maurya et al. (2019). The dsDNA was PCR amplified and agarose gel-purified to avoid non-specific DNA contamination. DNA was labeled in a reaction mixture containing T4 polynucleotide kinase enzyme (1 µL), 1X T4 polynucleotide kinase buffer and 1 µCi activity of [γ-^{32}P] ATP, and incubated at 37 ºC for 1 hr. After radiolabelling, the dsDNA was purified using a PCR purification kit (Sigma-Aldrich, USA).

2.5.9.2. Preparation of native PAGE:

For separation of protein-DNA complex, non-denaturing native PAGE (6%) was utilized to maintain structures of macromolecules and preserve interactions between them. The polymerization mixture was prepared using components described in the table below and poured immediately into the glass plate assembly. Twelve well-forming comb was inserted and the gel was allowed to polymerize for 1-2 hr. After polymerization, the combs were removed carefully and wells were flushed with water to remove any residual polymerized agarose gel. The glass plate assembly with polymerized gel was inserted in the vertical electrophoresis apparatus and filled with chilled 0.5X TBE buffer. 5 µL of non-SDS dye was placed in each well and pre-electrophoresis was performed at 50 volts for 30 minutes.

Table 2.3 List of the composition of a native PAGE gel

Components	**Resolving gel**	
	6 % (20 mL volume)	**10 % (20 mL volume)**

Acrylamide (30%)	4.0 mL	6.67 µL
5X TBE	2 mL	2 mL
50% Glycerol	2 mL	2 mL
10% Ammonium per sulphate	200 µL	50 µL
TEMED	25 µL	25 µL
Water	11.8 mL	9.1 mL

2.5.9.3. Sample preparation and electrophoresis:

The samples for EMSA were prepared by mixing radiolabeled DNA (hot DNA) or non-radiolabeled DNA (cold DNA), 1X binding buffer (50 mM Tris–HCl (pH 8.0), 75 mM KCl, 5 mM $MgSO_4$ and 0.5 mM DTT), 0.5 mM DTT, and protein. The samples were incubated at 20 ºC for 20 minutes. A titration protocol was performed with an increasing concentration of protein and a constant amount of radiolabelled DNA. The reaction mixture of EMSA samples were mixed with 1X non-SDS loading dye and loaded in wells. The gel was allowed to run for 4-5 hours at 50 V.

2.5.9.4. Gel drying and film development:

The electrophoresis apparatus was dis-assembled carefully, as the samples were radiolabeled. The glass plates were gently separated and the gel was placed on 3 mm Whatman filter paper. The gel with filter paper at the bottom was placed on a gel dryer and wrapped with saran wrap while avoiding bubble formation. The vacuum drying was performed for 45 minutes, followed by exposure of the dried gel (wrapped with saran wrap from both sides) to X-ray film in X-ray cassette overnight at -20 ºC. After the exposure, the

film was developed in dark using developing and fixing buffer. The developed film was washed with water, air-dried, and documented for further analysis.

2.5.10. Gas Chromatography/Mass Spectrometry (GC-MS):

The Gas Chromatography/Mass Spectrometry (GC/MS) instrument is composed of two parts, the Gas Chromatography (GC), used for separation of a complex mixture of volatile compounds while the Mass Spectrometry (MS) part is used for identification of the compound based on the mass. Headspace solid-phase microextraction (HS-SPME) was used for sample extraction, which involves the use of an extracting phase-coated fiber. In brief, fungal cultures were grown in SPME vials for three days at 37 ºC with loose cap for proper aeration, followed by tightening of the caps on 4th day with continued incubation for 16 hrs at 37 ºC. The vials were further incubated at 37 ºC for 30 minutes for equilibration and the SPME fiber (PDMS/DVB) was exposed to the head region of the vial for the proper extraction of all volatile compounds. The fiber was injected into the injection port of the machine for thermal desorption. The default program was used for the GC-MS instrument run and result analysis.

2.5.11. Whole-genome sequencing:

The next-generation sequencing (NGS) was outsourced from Xcelris Labs. Ltd. Ahmedabad (India). Illumina platform was used for whole-genome sequencing and Illumina TruSeq Nano DNA HT Library Preparation Kit was used for paired-end sequencing library preparation. Illumina Nextseq 500 using 2 x 150 bp chemistry was used for sequencing of the generated libraries, and genome mapping was done using Burrows-Wheeler Aligner (BWA) (v. 0.7.5a) program with optimized mapping parameters. The genome mapping, coverage, gene prediction statistics, and a total number of SNPs and INDELs were analyzed and

compared with the reference genome *Trichoderma virens* GVW (accession number LQCH00000000) available in NCBI.

2.5.12. Transcriptome analysis:

The *de novo* transcriptome sequencing was performed on Illumina HiSeq 2500 platform at M/S Scigenom, Cochin, Kerala, India. In brief, a paired-end library was prepared for *de novo* transcriptome sequencing, and Trinity software was used for assembly of the cleaned reads produced by Data Pre-Processing methods including adaptor trimming, quality filtering, and end trimming. Gene expression estimation was performed by aligning the trimmed reads to the assembled transcriptome using the Bowtie2 program. DESeq program was employed for differential gene expression analysis and an in-house pipeline CANoPI (Contig Annotator Pipeline) was used for transcriptome annotation. FPKM values were considered for fold change calculation and $\log_2$-fold change ≥ 2 was considered for gene expression profile comparison.

2.5.13. Thin Layer Chromatography (TLC):

For TLC, silica-coated aluminum plates (Merck) were used and a mixture of acetone, chloroform, and formic acid (28:70:2) was used as a solvent. Samples were extracted by mixing an equal volume of sample (cell-free filtrate) and ethyl acetate. The mixture was vortexed for 5-6 times and the upper phase was passed through a sodium sulfate column and collected in a fresh glass tube. The samples were dried, reconstituted in methanol, and loaded immediately after reconstitution on the TLC plate pre-equilibrated with a solvent mixture. The plate was then transferred to the TLC chamber saturated with the solvent mixture and allowed

to run for 2-3 hrs. The separated samples were observed under short-wavelength UV at 254 nm and photographed.

2.5.14. Biophysical Techniques:

2.5.14.1. Dynamic light scattering:

Dynamic Light Scattering (DLS) measures the fluctuations of the scattering intensity from particles of the solution undergoing random brownian motion in a time-dependent manner. Dynamic light scattering is also used to determine the oligomeric status of the particle and several other biophysical parameters such as molecular weight, hydrodynamic radius, correlation coefficient, and translational diffusion constant, etc. Zetasizer Nano ZS90 (Malvern, UK) was used to determine the DLS pattern of the purified protein by scanning for 15 minutes at an interval of 5 minutes and the effective diameter of each measurement was recorded and plotted with percent intensity versus diameter (nm) values. The data analysis was done with the help Malvern Zetasizer software suite (http://www.malvern.com).

2.5.14.2. Fluorescence spectroscopy:

A well-folded protein shows a fluorescence emission spectrum which is a combination of the fluorescence emission of aromatic amino acids like phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) which have intrinsic fluorescence properties due to the presence of conjugated aromatic ring system. Emission from tryptophan residues is majorly involved in intrinsic fluorescence emissions of a folded protein because its quantum yield is high in comparison to tyrosine and phenylalanine. Disulfide bonds also have considerable absorption in the same wavelength range. The fluorescence spectrophotometer used in the present study

was JASCO spectrofluorometer (FP-8500). Fluorescence emission spectra were collected from λ_{310} to λ_{450} nm wavelength, following excitation at λ_{280} nm or λ_{295} nm wavelength. In both, we collected emission spectrum for the native and unfolded protein using a single wavelength to observe the structural changes.

2.5.14.3. Thermoflour shift assay:

A thermal shift assay determines the stability of the protein by evaluating the change or shift in the thermal denaturation temperature of a protein under different conditions. The thermoflour shift assay was performed in a 96-well plate containing protein concentration of 2 µM along with freshly diluted SYPRO orange dye (1:1000, v/v) in 25 µL reaction mixture, in triplicates. Lysozyme was also monitored as an internal control. PCR plates were then sealed and placed into a BioRadCFX96™ Real-time system in FRET mode. Thermal scanning was performed from 20 ºC to 95 ºC every 30 sec in 1 ºC increment. Subsequent analysis of the fluorescence data was done using software provided in the CFX96™ Real-time system.

2.5.15. Semiquantitative RT-PCR:

The semi-quantitative RT-PCR is a method to determine the relative amount of gene expression using the housekeeping gene as an internal control to normalize the expression levels of the gene of interest. The 20 µL semiquantitative PCR reaction volume comprised of 10 µL Taq polymerase Ready Mix™, 100 ng of cDNA, and 0.3 µM final concentration of gene-specific primers. Reactions were performed in triplicates for all three biological replicates using the following protocol: initial activation at 95 ºC for 10 min; 25 cycles of 95 ºC for 15 sec, 55 ºC for 30 sec, 72 ºC for 30 sec followed by a final extension at 72 ºC for 10 min.

2.5.16. Real-Time PCR:

Two microgram of RNA was used for cDNA synthesis with the help of RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA). Histone3 and actin were used as a housekeeping genes. The amplification assay was carried out in a LightCycler® 480 System (Roche Diagnostics, Germany). PCRs were performed in a 12.5 µL final volume containing 6.25 µL SYBR® Green JumpStart™ TaqReadyMix™, 100 ng of cDNA, and 0.3 µM final concentration of gene-specific primers. Reactions were performed in triplicates for all 3 biological replicates. PCR conditions were as follows: an initial step of 10 minutes at 95 ºC, followed by 40 cycles at 95 ºC for 10 sec, 60 ºC for 30 sec, and 72 ºC for 30 sec and a cooling phase of 30 sec at 40 ºC. Melting curves were generated by increasing the temperature from 65 ºC to 99 ºC at 0.29 ºC/sec. The relative gene expression was determined using the relative expression software tool. Results were considered positive when the fluorescent signal above the baseline was detected as determined by second-derivative analysis and were expressed in terms of the quantification cycle (C_q).

2.5.17. Mycoparasitic assay:

The mycoparasitic ability of *Trichoderma virens* was determined using confrontation assay. A disk of *Trichoderma virens* and the test pathogens were placed opposite each other on the same PDA plate. The cultures were allowed to grow for 5 days and examined for the ability of *Trichoderma virens* to overgrow and lyse the mycelia of pathogens. Micoparasitic ability was also observed under a light microscope. The confrontation of *Trichoderma* and pathogens was performed on a plate containing 1% agar. After 2-3 days the mycelium growth of *Trichoderma* and the pathogen was visualized under a light microscope. Images were acquired and documented. Since *Pythium aphanidermatum* and *Rhizoctonia solani* cell walls are primarily composed of cellulose and chitin, respectively, we assessed the ability of *Trichoderma* to utilize cellulose and chitin as carbon-source by growing them in Vogel's

minimal medium containing cellulose or chitin as a sole carbon source in liquid shake culture at 28 ºC and 150 rpm. For the initiation of growth, 0.1% sucrose was added to the medium.

2.5.18. Hydrophobicity assay:

Hydrophobicity is provided by the hydrophobins which are low molecular weight cysteine-rich secreted proteins in fungus. They assist in many ways in the fungal life cycle. For example, spores are covered with hydrophobin layer which helps in their dispersal. Hydrophobins also provide resistance and immunological resistance to spores against many stresses and pathogens (Bayry et al., 2012). Hydrophobin layer assists fungus to breach air-water interface. Hydrophobicity in the *T. virens* culture was determined by growing them for 5 days on PDA at 28 ºC. After 5 days, the hydrophobicity was tested by adding 0.5% aqueous rose bengal and recording the disappearance of the drop after one hour.

2.5.19. Molecular modeling:

The *in silico* modeling was performed by I-TASSER server (Iterative Threading ASSEmbly Refinement) (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang et al., 2014). The input is the FASTA arrangement of the protein sequence and the 3D model was generated using three phases. The first phase involves a meta-threading server LOMETS (Local Meta threading server) which identifies the template from the library of the PDB structure. Phase two involves replica-exchange Monte Carlo simulations for iterative structure assembly simulation. Lastly, the third phase involves the use of REMO7 and FG-MD8 for structure refinement and construction at the atomic-level. COFACTOR is employed for the prediction of the modeled structure in I-TASSER. Ramachandran plot was employed for

accessing stereochemical nature of the protein for the favoured and unfavoured regions in the predicted model. Modloop server was used for further refinement of the loops in the built model (https://modbase.compbio.ucsf.edu/modloop/; (Fiser et al., 2003). The refined model structure was finally verified by Verify3D and the validated model was visualized by the PyMOL visualization tool.

2.5.20. Covalent-Docking:

The molecular docking study was performed using Schrodinger suite 2019-4 (https://www.schrodinger.com/covdock). The modelled structure of the protein was pre-processed using Protein preparation wizard and the ligand structure was refined using the LigPrep module in Schrodinger suite 2019-4. The OPLS3 force field was used for the minimization of the protein and ligand structure. The CovDock protocol was performed as provided in the manual for the formation of a covalent bond between protein and ligand. Glide module was employed for the creation of the covalent docked complexes with multiple poses in Schrodinger suite 2019-4.

Chapter 3

Results

Chapter 3.1

Role of GAPDH (vGPD) in secondary metabolism of *Trichoderma virens*

3.1.1. Introduction:

The genus *Trichoderma* is a dominant soil fungus and used widely as biocontrol agents due to its ability to kill other fungi and induces resistance in plants against invading pathogens (Mukherjee et al., 2013a). Secondary metabolites are small organic compounds and are known to be involved in *Trichoderma*-fungal and *Trichoderma*-plant interactions (Druzhinina et al., 2011; Mukherjee et al., 2012b). *Trichoderma* produces a wide range of secondary metabolites, both volatile and non-volatile. Earlier, a gene cluster (designated as "*vir*" cluster) was reported to be involved in the biosynthesis of volatile sesquiterpene metabolites, which include ecologically important compounds like beta-caryophyllene and germacrenes in *Trichoderma virens* (Crutcher et al., 2013; Mukherjee et al., 2006). A Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was reported to be associated with the *vir* cluster in *T. virens* as well as in certain *Aspergillus* species harboring this cluster, even though, it was thought to be the enzyme involved in glycolysis (Mukherjee et al., 2006). Further analysis of the genome sequence of *T. virens* and *Aspergillus* species revealed that this GAPDH (named here as vGPD or GPD2) is consistently associated with the *vir* cluster, and is an additional gene different from the one involved in glycolysis (designated here as gGPD or GPD1). In this chapter, we present the data about the role of vGPD protein in secondary metabolism in *T. virens*, using genetic approaches.

3.1.2. Materials and methods:

3.1.2.1. Strains, media and culture conditions:

Trichoderma virens IMI 304061 isolated from soil (Mukherjee et al., 1993) was used as the wild type strain. The fungi (wild type or transformed) were grown in potato dextrose agar medium (prepared in the laboratory using infusion from 200 g potatoes, 20 g dextrose and 20 g

agar per liter water) at 28 ± 2 ºC and when required, hygromycin B was added to a concentration of 200 µg/mL for the screening of transformants. For long term storage, the fungi were maintained as glycerol stock at -80 ºC**.**

3.1.2.2. Phylogenetic analysis of *Trichoderma* and *Aspergillus* GAPDHs:

The protein sequences of vGPD and gGPD of *Trichoderma* and *Aspergillus* species were obtained from the fungal genome portal MycoCosm (www.genome.jgi-psf.org). The sequences were aligned using MUSCLE and the phylogenetic tree was built using the maximum likelihood method in MEGA 7 (https://megasoftware.net/; Kumar et al., 2016). The bootstrap consensus tree was inferred using 1,000 resamplings of the data. The near-neighbor interchange algorithm was used to obtain the tree.

3.1.2.3. Generation of vGPD deletion mutants:

Gene deletion was performed by using a split-marker technique with hygromycin resistance as a dominant selectable marker (Catlett et al., 2003). Plasmid pATBS (Mukherjee et al., 2003) was used to generate two flanks with part of the hygromycin cassette, designated as LF-HY and RF-YG, in a two-step process. The upstream (1.7 kb) and downstream flanking regions (2.1 kb) of vGPD gene were PCR amplified by AccuTaq™ LA DNA polymerase (Sigma-Aldrich, USA) using primers given in Table 3 of Annexure I that incorporate *Kpn*I/*Xho*I and *Xba*I/*Sac*I sites in the left flank and right flank, respectively. These PCR products were cloned in the pATBS plasmid (left flank at *Kpn*I-*Xho*I sites and right flank at *Xba*I-*Sac*I sites, separately, resulting in two plasmids designated as pATBS-LF and pATBS-RF, respectively). Each plasmid was used as a template for PCR amplification by AccuTaq™ LA DNA polymerase for the generation of LF-HY and RF-YG, respectively (using primer pairs vGLFkpn-NLC37 and NLC38-vGRFxba). The PCR products were purified using the High Pure PCR product purification kit (Roche Diagnostic, UK) and used for protoplast transformation.

3.1.2.4. *Trichoderma virens* transformation and screening of mutants:

The procedure used for protoplast isolation and transformation of *T. virens* was as described earlier (Mukherjee et al., 2003). The linear DNA of both the flanks fused to the part of the hygromycin resistance cassette, was used in a 1:1 molar ratio for transformation. The transformants were selected on PDA under the selection pressure of 200 µg/mL of hygromycin B. The stable transformants were obtained by diluting and replating the spores for 3 cycles and reconfirmed for successful homologous recombination by PCR using a primer upstream of both the flanks and a primer within the hygromycin resistance cassette (vGLOf and PtrpFR for left flank and TtrpRF and vGROrfor right flank). Finally, gene-specific primers were used to verify the replacement of the gene with hygromycin resistance cassette and the absence of any traces of wild type gene. The absence of any transcript of the gene was further confirmed by semi-quantitative RT-PCR.

3.1.2.5. Gas Chromatography-Mass Spectrometry (GC– MS) technique:

The volatile metabolites were detected by the Head Space Solid Phase Microextraction (HS-SPME) and Gas Chromatography-Mass Spectrometry (GC-MS) technique (Crutcher et al., 2013). The details of the technique are described in the Materials and methods section. Briefly, the wild type and deletion mutant were grown in PDA for 3 days at 28 ± 2 ºC. The mycelial disk of 5 mm diameter was transferred from a petri dish to a 10 mL PDA slant in 20 mL headspace vial and grown at 28 ± 2 ºC. Control vials without fungal mycelium were maintained under the same conditions. The headspace vial caps were kept loose to maintain aerobic growth conditions. After 4 days, the cultures were tightly sealed and incubated for 6 hours. Each headspace vial containing the sample was left for 30 min at 37 ºC for equilibration followed by exposure of SPME fiber (PDMS/DVB) to the headspace region of the vial for 30 min while maintaining the sample at 37 ºC. After sampling, the SPME fiber was injected for 30 s in the injector port of the GC to allow thermal desorption of analytes. A Shimadzu GC– MS

instrument (Shimadzu Corporation, Japan) equipped with a GC-17A gas chromatograph was used for GC–MS analysis. The instrument is supported with a DB-5 (J & W Scientific, CA, USA) capillary column [(5%-Phenyl) methylpolysiloxane]. Helium was used as the carrier gas at a flow rate of 0.9 mL min^{-1}. The oven temperature was set initially at 40 ºC for 5 min, then raised to 200 ºC using a temperature gradient of 4 ºC min^{-1}, held at 200 ºC for 2 min, increased to 280 ºC at 10 ºC min^{-1} and finally held at 280 ºC for 20 min. The injector and interface temperatures were 210 ºC and 230 ºC respectively. For mass spectroscopy, the transfer line was held at 280 ºC, the source at 230 ºC, and the quad at 150 ºC. Mass spectra were taken in EI mode (at 70 eV) in the range from 35m/z to 500m/z, with a scanning rate of 4.42 scans per second. Analyses were carried out in split-less mode. The separated major peaks in each fraction were identified by a standard databank of mass spectra available in the spectral library (Wiley/NIST Libraries) of the instrument, by running standards of some of the identified compounds and by comparing their retention indices with the literature values measured on the identical column. Each analysis was performed three times with independent biological replicates. Linear retention indices were calculated using GC data of a homologous series of saturated aliphatic hydrocarbons (C8 to C40) separated on the same column using the same conditions as for GC analysis of the volatile organic compounds (VOCs).

3.1.2.6. Thin Layer Chromatography and LC-MS analysis for detecting non-volatile compounds:

The non-volatile compounds were detected in wild type *T. virens* and vGPD deletion mutants using thin-layer chromatography as described earlier (Mukherjee et al., 2006). The wild type *T. virens* and vGPD deletion mutants were grown in potato dextrose broth (HiMedia, India) for three days at 30 ºC and 150 rpm. The filtrate was extracted with an equal volume of chloroform, dried and reconstituted in $1/10^{th}$ volume of methanol. A 100 µL aliquot of this was

spotted on the pre-coated TLC plate. TLC plate was developed in acetone: chloroform: formic acid (28:70:2) and visualized under short-wavelength UV (254 nm).

The presence of non-volatile compounds was further confirmed by Ultra-performance liquid chromatography (UPLC) coupled with QToF-MS for analysis of the compounds detected. Wild type *T. virens* and vGPD deletion mutant was allowed to grow for 3 days in the VMS medium. The filtrate was then subjected to LC-MS analysis for the presence of the metabolites.

3.1.2.7. Gene expression analysis using qRT-PCR:

The change in the expression profile of all the six *vir* cluster genes in vGPD knockout mutant *vis-a-vis* wild type was analyzed using real-time PCR (qRT-PCR). Total RNA was extracted from wild type and deletion mutant (grown in PDB for 3 days) using TRI-Reagent (Sigma-Aldrich, USA) and total RNA integrity was checked on formaldehyde gels. Traces of genomic DNA were removed by treating with DNase I (Thermo Scientific, USA). Two µg of RNA were used for cDNA synthesis with RevertAid first strand cDNA synthesis kit (Thermo Scientific, USA). Histone 3 and actin were used as a housekeeping gene. LightCycler® 480 System (Roche Diagnostics, Germany) was used for gene expression analysis. The 12.5 µL real-time qRT-PCR reactions were performed in triplicates for all 3 biological replicates. The relative gene expression was determined using the relative expression software tool.

3.1.3. Results:

3.1.3.1. Phylogenetic analysis:

Since the vGPD orthologues are present in the genomes of *Trichoderma* and *Aspergillus* (Fig. 3.1.1), we performed the phylogenetic analysis of GAPDH proteins from these two genera. All the species of *Aspergillus* considered here possess multiple isoforms of GAPDH protein while only 2 of the *Trichoderma* species (*Trichoderma virens* and *Trichoderma harzianum*) have another isoform of GAPDH protein (named as vGPD) in addition to GAPDH protein involved in glycolysis.

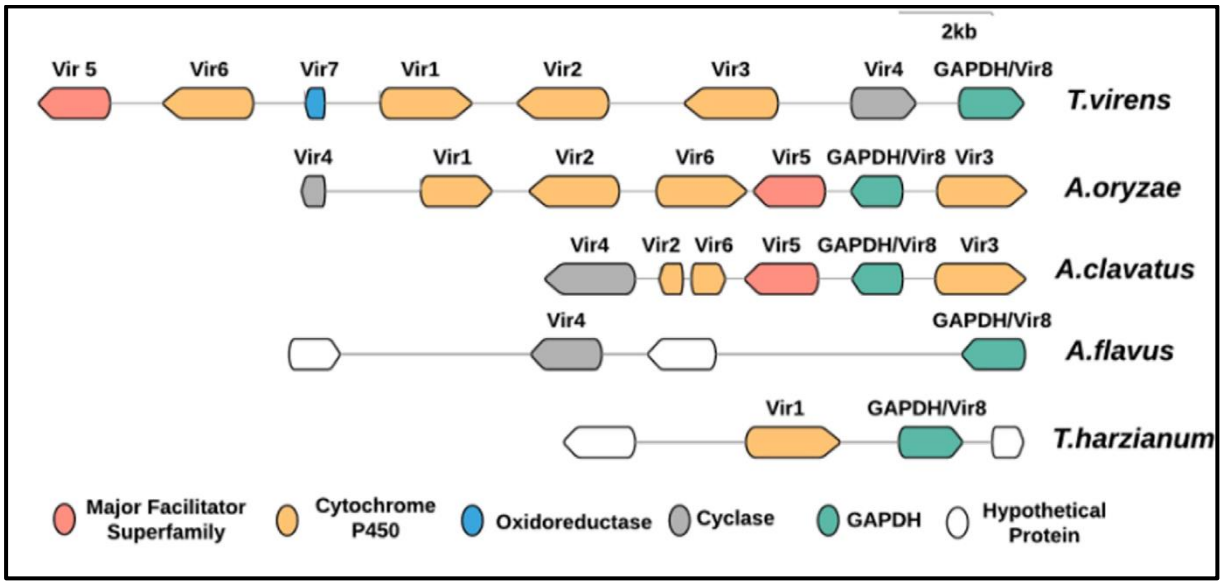


Figure 3.1.1. The *vir* cluster homologs present in *Trichoderma* and *Aspergillus* species. The arrow color represents the predicted function. Arrow direction indicates 5' to 3' coding orientation.

We constructed the molecular phylogenetic tree for 26 GAPDH proteins from 12 species of *Aspergillus* and 8 species of *Trichoderma* genus using the MEGA 7 program (Fig. 3.1.2). The maximum likelihood analysis of the GAPDH proteins of both *Aspergillus* and *Trichoderma* revealed the presence of distinct phylogenetic groups for both GAPDH types.

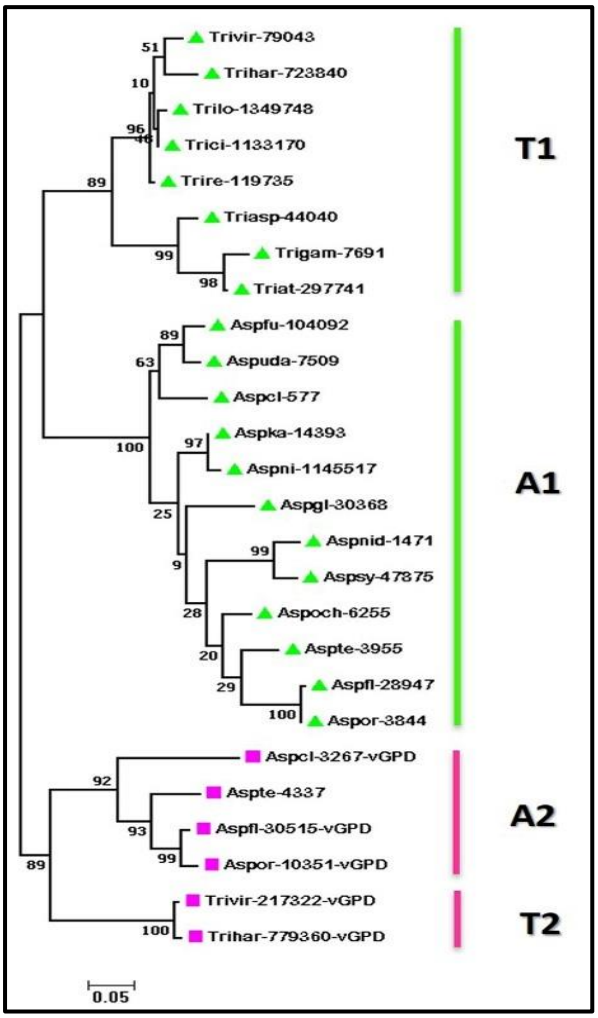


Figure 3.1.2. Phylogeny of the vGPD and gGPD homologue of *Trichoderma* and *Aspergillus* species. Molecular phylogenetic tree for 26 GAPDH proteins from 12 species of *Aspergillus* and 8 species of *Trichoderma* genus using the MEGA 7 program. Sequence alignment was performed using MUSCLE and the phylogenetic tree was built using the maximum likelihood method. The near-neighbor interchange algorithm was used to obtain the tree. The numbers on the branches are bootstrap values (%) obtained with 1000 pseudoreplicates. Fungal species utilized for the tree construction were: *Aspergillus clavatus* (Aspcl), *Aspergillus fumigatus* (Aspfu), *Aspergillus flavus* (Aspfl), *Aspergillus glaucus* (Aspgl), *Aspergillus kawachii* (Aspka), *Aspergillus nidulans* (Aspnid), *Aspergillus niger* (Aspni), *Aspergillus ochraceoroseus* (Aspoch), *Aspergillus oryzae* (Aspor), *Aspergillus sydowii* (Aspsy), *Aspergillus terreus* (Aspte), *Aspergillus udagawae* (Aspuda), *Trichoderma asperellum* (Triasp), *Trichoderma atroviride* (Triat), *Trichoderma citrinoviride* (Trici), *Trichoderma gamsii* (Trigam), *Trichoderma harzianum* (Trihar), *Trichoderma longibrachiatum* (Trilo),

Trichoderma reesei (Trire), *Trichoderma virens* (Trivir). Protein ID numbers are shown following the abbreviated names and the vGPD proteins are labelled corresponding to the protein IDs of *Trichoderma* and *Aspergillus* species.

GAPDHs of *Trichoderma* species form T1 (gGPD) and T2 (vGPD) form groups and GAPDHs from *Aspergillus* species form A1 (gGPD) and A2 (vGPD) form groups. The vGPD of *Trichoderma* and *Aspergillus* species form a group together. The same is true for gGPD, which is however, present in all life forms, being an essential protein for survival, and hence vertically inherited. Clustering of vGPD of *Trichoderma* species with the GAPDHs of *Aspergillus* species, but not with the *Trichoderma* species provides a clue about the horizontal transfer of vGPD between *Aspergillus* and *Trichoderma*.

3.1.3.2. Cloning for the generation of vGPD mutant of *T. virens*:

For the generation of vGPD mutant, two constructs were produced using a pBluescript II KS (+) based vector named pATBS (5161 bp) harboring a hygromycin resistance cassette (2200 bp) for selection of positive transformants. The upstream flanking region of vGPD gene of 1.7 kb was cloned at the 5' end of the *hph* gene while the downstream flanking region of vGPD gene was cloned at the 3' end of the *hph* gene in the MCS region as depicted in the Fig. 3.1.3.

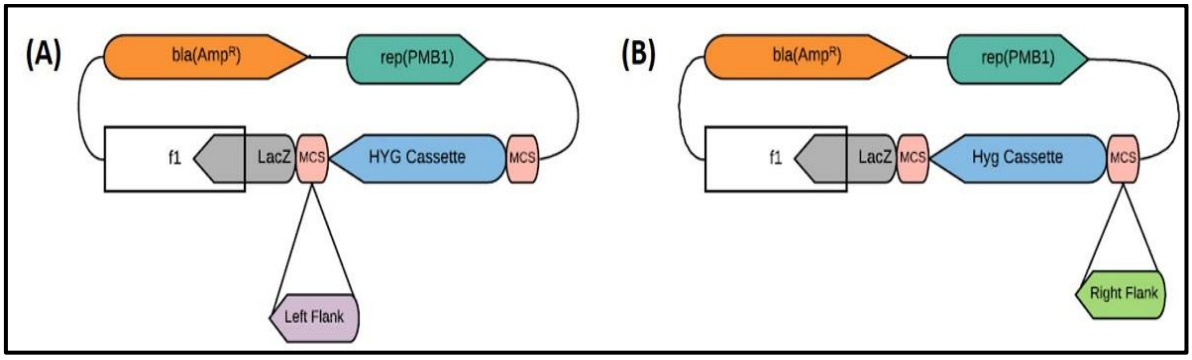


Figure 3.1.3. **Plasmids for *vGPD* gene deletion mutant generation.**
A). pATBS-LF is the pBLUESCRIPT (KS+) based pATBS plasmid (5.1 kb) containing Left flank of vGPD gene (1.7 kb) cloned at *Kpn*I/*Xho*I sites downstream of Hygromycin resistance cassette (2.2 kb). **B).** pATBS-RF is the pBLUESCRIPT (KS+) based pATBS plasmid (5.1 kb)

containing the Right flank of vGPD gene (2.1 kb) cloned at *XbaI*/*Sac*I sites upstream of Hygromycin resistance cassette (2.2 kb).

For plasmid construction, upstream and downstream flanking regions of vGPD gene were amplified with AccuTaq™ LA DNA Polymerase, and PCR products were purified. The PCR products and pATBS plasmids were digested with respective restriction enzymes and ligated for transformation in the DH5α strain *E. coli*. The positive transformants were detected by colony PCR using gene-specific primers (Fig. 3.1.4).

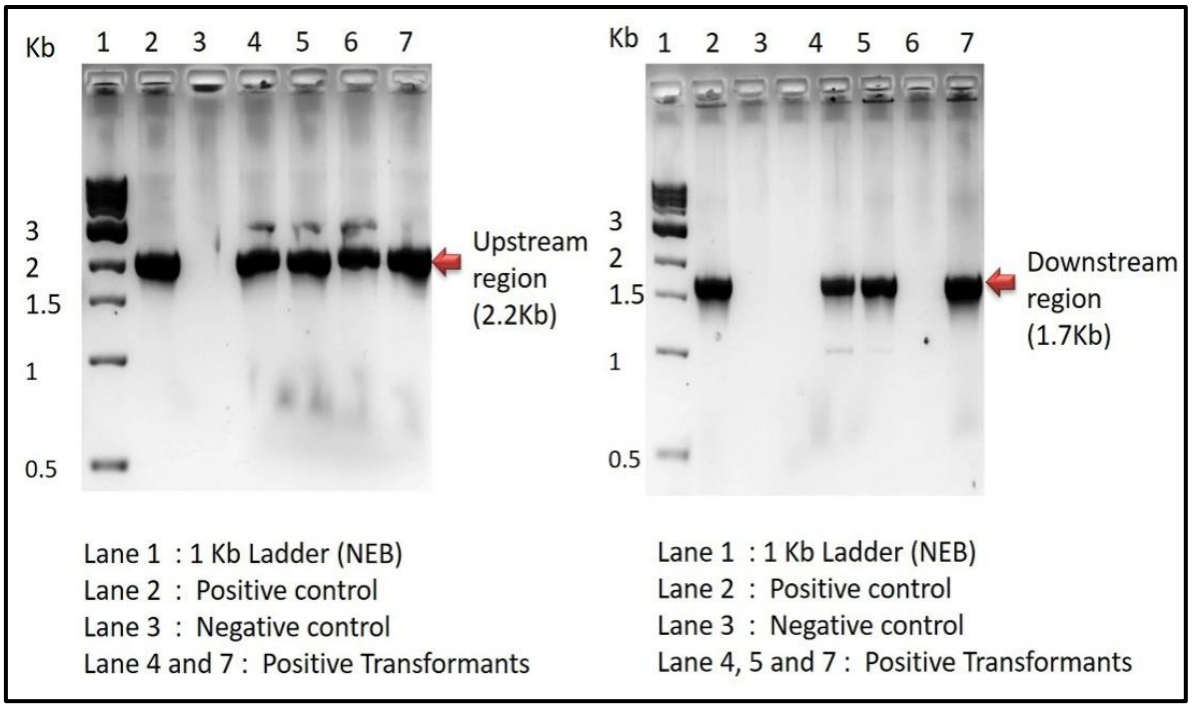


Figure 3.1.4. Agarose gel (1.5%) showing colony PCR of DH5α cells transformed with pATBS-LF and pATBS-RF constructs. The positive transformants showed amplification of upstream region from pATBS-LF construct and downstream region from pATBS-RF construct using respective primers for both regions.

3.1.3.3. Generation of vGPD deletion mutants of *T. virens*:

Using a split-marker strategy, we obtained three independent knockout mutants in vGPD locus. DNA isolation of these vGPD mutants was performed, followed by confirmation with PCR. Primers designed for the detection of locus-specific homologous integration event

and gene-specific primers for determining the absence of the vGPD gene were used for PCR analysis. PCR analysis showed amplification of the vGPD gene in WT but not in vGPD knockout mutant. Similarly, the absence of transcripts was also determined by semiquantitative RT-PCR. Biological triplicates of vGPD deletion mutants were used for semiquantitative RT-PCR and analysis showed that vGPD transcripts were present only in wild type but not in vGPD deletion mutants. (Fig. 3.1.5).

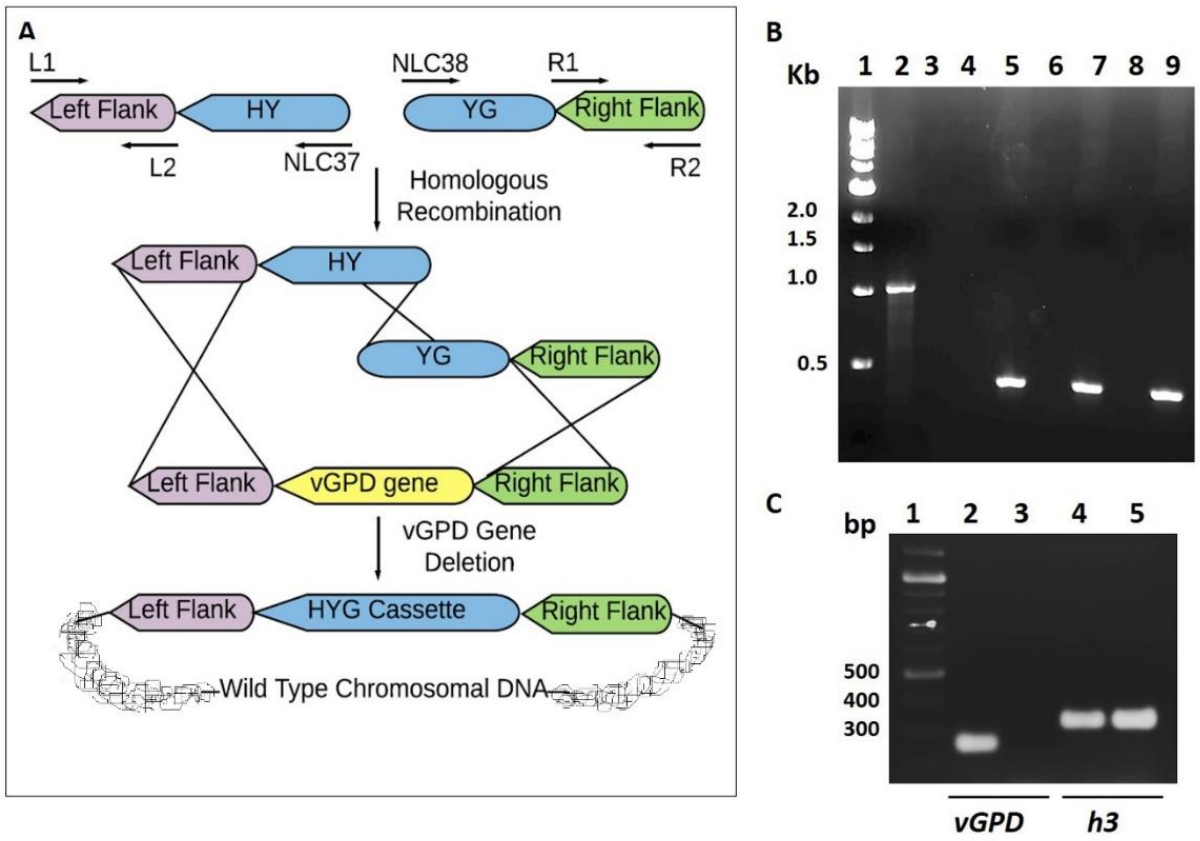


Figure 3.1.5. Split-marker strategy for *vGPD* gene deletion.

A). Primers L1/L2 and R1/R2 amplify *vGPD* gene left flank and right flank sequences respectively. Primers L1/NLC37 and NLC38/R2 amplify "LF-HY" and "YG-RF" fragments, respectively. The overlap between the "HY" and the "YG" marker sequences is 491 bp. Homologous recombination between the overlapping regions of the selectable marker (HYG), and between the flank regions and chromosomal DNA results in a directed replacement of *vGPD* gene with *HYG* cassette. **B).** 1.5% Agarose gel showing amplification of the *vGPD* gene (1.2kb) only in WT and amplification of hygromycin gene (0.4 kb) only in the deletion mutant.

(Lane 1: 1 Kb Ladder (NEB), Lane 2: WT genomic DNA amplified with *vGPD* gene-specific primers, Lane 3: WT genomic DNA amplified with Hph for and rev primers, Lane 4: KO-1 genomic DNA amplified with *vGPD* gene-specific primers, Lane 5: KO-1 genomic DNA amplified with Hph for and rev primers, Lane 6: KO-2 genomic DNA amplified with *vGPD* gene-specific primers, Lane 7: KO-2 genomic DNA amplified with Hph for and rev primers, Lane 8: KO-3 genomic DNA amplified with *vGPD* gene-specific primers, Lane 9: KO-3 genomic DNA amplified with Hph for and rev primers). **C).** Semi-quantitative analysis of vGPD transcript in WT and deletion mutant. 1.5% Agarose gel showing amplification of vGPD gene transcript (300 bp) only in WT and not in the deletion mutant. H3 is taken as housekeeping gene control. (Lane 1: 100 bp Ladder (NEB), Lane 2: WT RNA amplified with *vGPD* gene-specific primers, Lane 3: Deletion mutant RNA amplified with *vGPD* gene-specific primers, Lane 4: WT RNA amplified with Histone 3 gene-specific primers and Lane 5: Deletion mutant RNA amplified with Histone 3 gene-specific primers).

To analyze the growth rate of wild type and knockout mutants, all the strains were grown on PDA plates for 3 days at 28 ºC. The growth and colony morphology of the mutants were similar to the wild type. (Fig. 3.1.6).

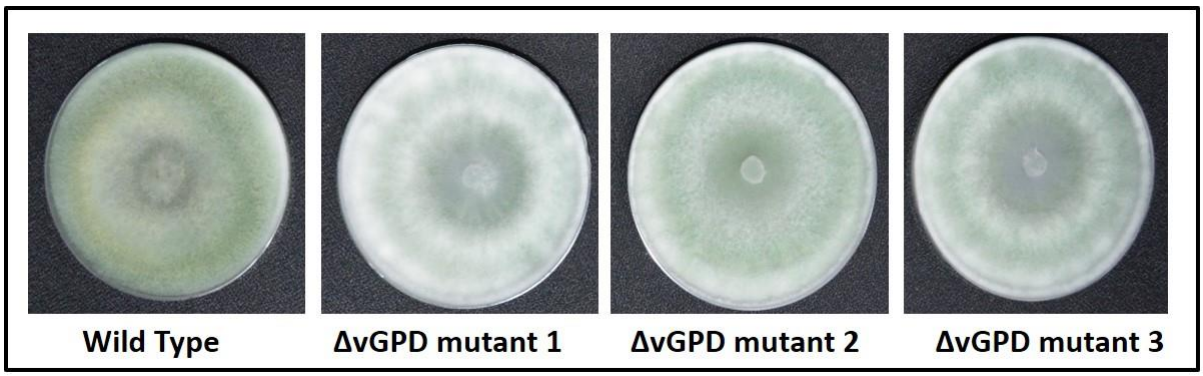


Figure 3.1.6. Comparison of colony morphology of wild type and vGPD deletion mutant in *Trichoderma virens* (with three independently isolated vGPD deletion mutants), 3 days after inoculation.

3.1.3.4. vGPD knockout mutants do not produce volatile sesquiterpenes:

The volatile metabolites profile of 4-day old wild type and deletion mutants as determined by HS-SPME-GCMS is presented in Fig. 3.1.7. We identified 32 VOCs comprising of 13 sesquiterpenes, 10 monoterpenes, and 9 other alkane compounds in the wild type strain (Table 1 of Annexure II). Monoterpenes constituted 65.71% of the volatile mixture while sesquiterpenes made up 23.26%. Germacrene D (17%), germacrene A (10%), ledene (5%), and beta-caryophyllene (0.7%) were the prominent sesquiterpene constituents whereas limonene (26.5%) was the major monoterpene followed by sabinene (4.8%). The vGPD knockout mutant did not produce any volatile sesquiterpene compounds; however, 7 monoterpenes (87.25%) and 6 alkane compounds were present in the deletion mutant. Both the wild type and the deletion mutant produced limonene (15.4%) and sabinene (2.5%) as the major monoterpenes.

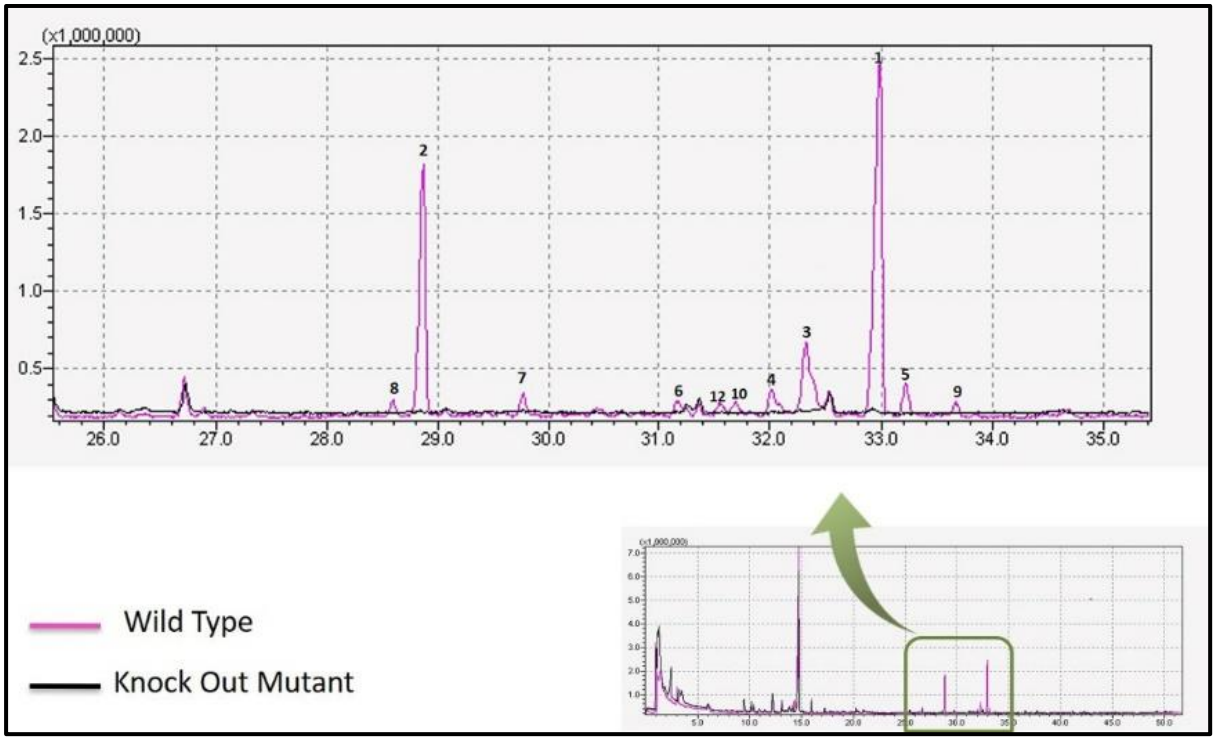


Figure 3.1.7. Gas chromatographic profile of volatile compounds produced by wild type and vGPD deletion mutant. Pink color peaks correspond to the volatile compounds produced by wild type and black peaks are for the volatile compounds from vGPD deletion mutant. The

sesquiterpene volatile compounds containing region (in the green box) is zoomed out. Peak numbering corresponds to the number of metabolites in Table 1 of Annexure II.

3.1.3.5. vGPD knockout mutants show downregulation in biosynthesis of non-volatile secondary metabolites:

Characterization of ΔvGPD mutants revealed that the deletion of the vGPD gene downregulates non-volatile metabolites (like heptelidic acid, viridin, viridiol, and gliovirin) production. TLC and LC-MS analysis of the ΔvGPD mutant showed reduction in the biosynthesis of non-volatile metabolites (Fig. 3.1.8 and Fig 3.1.9). This is similar to the role of vGPD in volatile sesquiterpene production. These data indicate a regulatory role of vGPD or an additional role of vGPD by an unknown mechanism in the biosynthesis of both volatile and non-volatile compounds synthesis. It is already known that HA production in *T. virens* suppresses the activity of glycolytic GAPDH and vGPD takes over the glycolytic function.

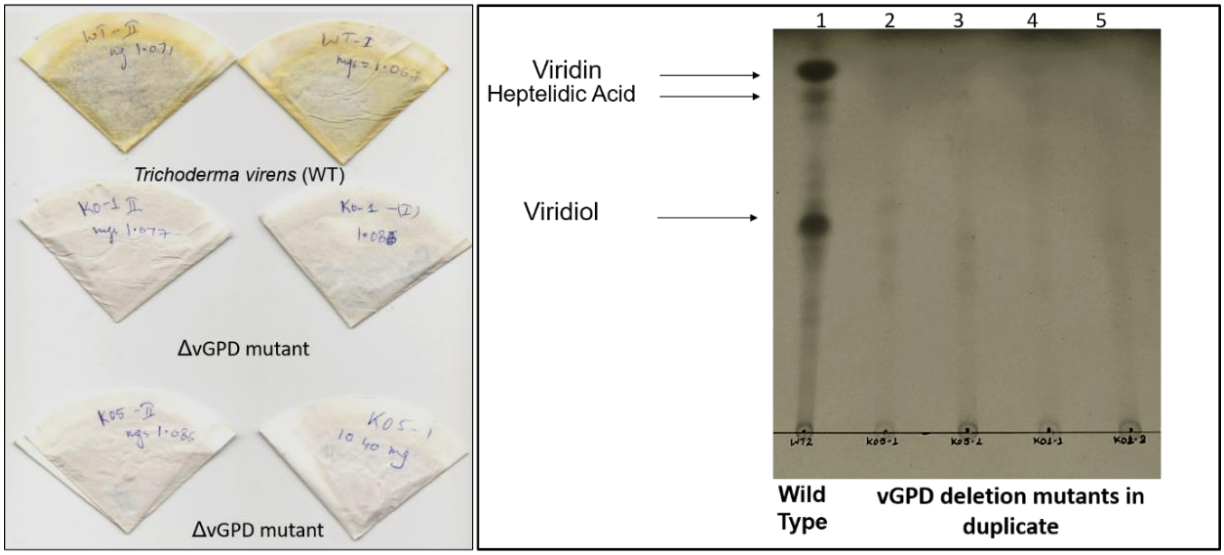


Figure 3.1.8 Comparison of non-volatile compounds produced by wild type *T virens* and vGPD deletion mutant A). Dry weight measurement of wild type *T. virens* and ΔvGPD mutants. The yellow color of the filtrate produced by wild type *T. virens* shows that, it is able

to produce non-volatile metabolites as compared to filtrate of ΔvGPD mutants which is colorless. **B).** TLC of wild type *T. virens* (Lane 1) and ΔvGPD mutant (Lane 2-5).

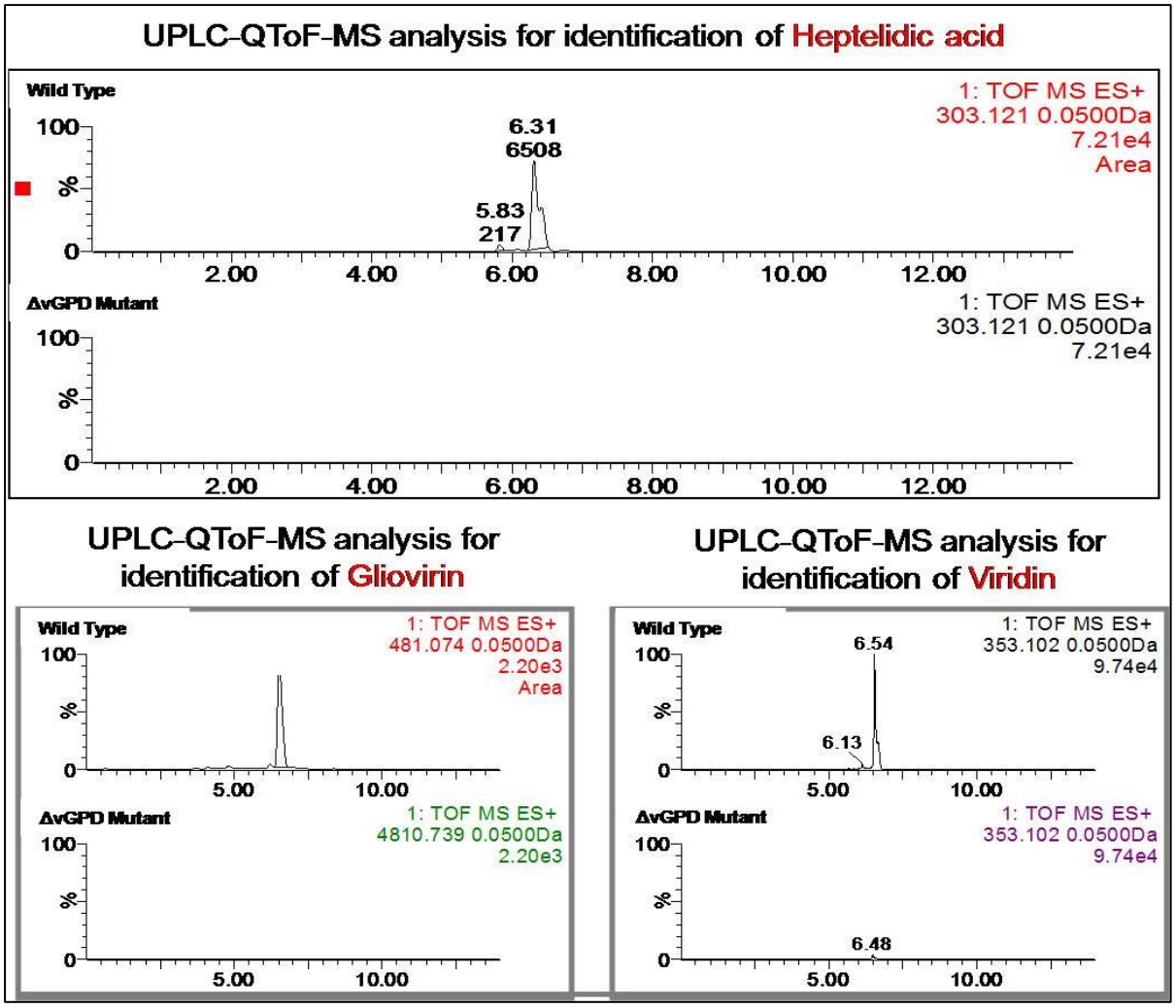


Figure 3.1.9. UPLC-QToF-MS analysis for identification of heptelidic acid, gliovirin and viridin in wild type *T. virens* and vGPD deletion mutant

3.1.3.6. Expression of *vir* cluster-specific genes altered in vGPD knockout mutants:

qRT-PCR was performed to analyze the effect of vGPD deletion on the different genes of the *vir* cluster. Among the seven genes (*vir1-vir7*) studied for expression in the knockout mutants by qRT-PCR analysis, two genes (*vir4* encoding a terpene cyclase and *vir3* for a cytochrome P450) were upregulated in the vGPD deletion mutant strain, while other

cytochrome P450s (*vir1*, *vir2*, *vir6*, and *vir7*) and MFS (*vir5*) were down-regulated in the mutant (Fig. 3.1.10).

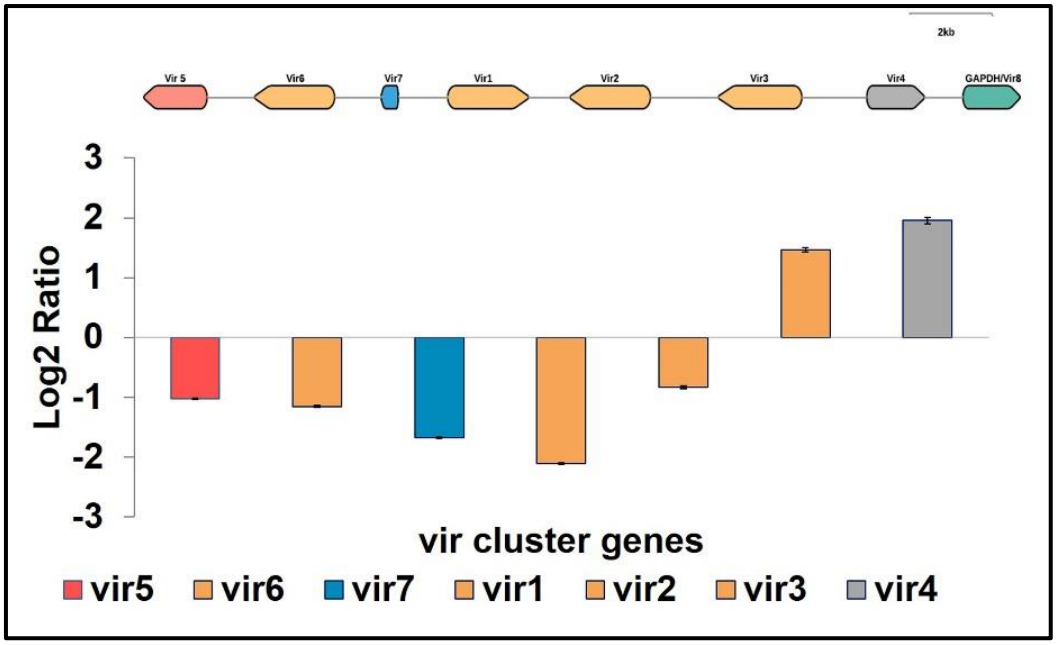


Figure 3.1.10. The relative expression of the *vir* cluster genes in vGPD deletion mutant as compared to wild type *T. virens*. Two genes were upregulated (*vir4* encoding a terpene cyclase and *vir3* for cytochrome P450), while four cytochrome P450s (*vir1*, *vir2*, *vir6*, and *vir7*) and an MFS (*vir5*) showed downregulation in the vGPD deletion mutant.

3.1.4. Discussion:

Using suppression subtractive hybridisation (SSH) and a viridin non-producing mutant, Mukherjee and group (2006) had earlier discovered a gene cluster in *T. virens* that was thought to be associated with viridin biosynthesis (Mukherjee et al., 2006). However, subsequently, using a gene knockout approach (of the terpene cyclase gene *vir4*) it was shown that this cluster is involved in the biosynthesis of volatile sesquiterpenes (Crutcher et al., 2013) and that viridin is biosynthesized by another cluster comprising of 21 genes, that does not include a typical secondary metabolism signature gene like NRPS, PKS or terpene cyclase (Bansal et al., 2018). An interesting feature of the *vir* cluster is the consistent association of a

GAPDH gene, which, on analysis of the genomes of *Trichoderma* and *Aspergillus* spp., revealed to be an additional GAPDH, different from the one involved in glycolysis. Organisms responsible for the production of toxins that irreversibly inhibits GAPDH activity have been identified to harbor an isozyme of GADPH resistant to the toxin. For example, a GAPDH isozyme associated with a gene cluster for biosynthesis of a sesquiterpene antibiotic pentalenolactone is found in *Streptomyces avermitilis.* This GAPDH is pentalenolactone insensitive and has been proposed to perform glycolysis under toxin biosynthesis conditions (Tetzlaff et al., 2006). *Trichoderma virens* (incorrectly identified as *T. koningii* in the earlier study) produces a sesquiterpene lactone koningic acid (heptelidic acid) which inhibits glycolysis by inactivating GAPDH. Sakai et al., (1990) identified an additional isoform of GAPDH, which is resistant to inactivation by koningic acid, and proposed that this additional GAPDH is responsible for carrying-out the glycolysis when koningic acid is being produced (Sakai et al., 1990). Due to the unavailability of the genome sequence, then, it was not known if this GAPDH is associated with the still unearthed heptelidic/koningic acid cluster. Moreover, now we know that this additional GAPDH is consistently associated with the *vir* cluster. We, therefore, hypothesized that this GAPDH might be responsible for the biosynthesis of the metabolites and studied its role by gene knockout. Our findings reported here proves that the *vir* cluster-associated GAPDH is involved in the biosynthesis of several sesquiterpene metabolites such as germacrene D, β-caryophyllene, α-amorphene, germacrene A, β-selene, aromadendrene, ledene, β-elemene, etc. The moonlighting effect of glycolytic GAPDH was earlier reported in animal systems and some yeasts and bacteria, but this is the first evidence on the non-glycolytic role of a dedicated GAPDH in a filamentous fungus and also, first, report on the involvement of a GAPDH in secondary metabolism in any organism. Sesquiterpenes are synthesized via a very complex and in many cases still unclear metabolic pathways. The first dedicated step is cyclization, which is followed by dehydrogenation and the action of modifying

enzymes like cytochrome P450s (Schmidt-dannert 2015). Since the GAPDH deletion mutants do not produce any of the 21 volatile sesquiterpene metabolites reported earlier to be products of *vir* cluster, we speculate that this GAPDH is involved in early dehydrogenation step that follows cyclization by *vir4* (terpene cyclase). Interestingly, the deletion of vGPD also suppressed the biosynthesis of viridin, viridiol, heptelidic acid, gliovirin, and many unknown non-volatile metabolites, further complicating the interpretation on whether vGPD is a biosynthetic enzyme or a regulatory protein. In *Aspergillus oryzae*, however, the function of the ortholog of the *vir* cluster genes has been reported and showed that this cluster is responsible for the biosynthesis of HA. They have studied the role of the vGPD ortholog gene and showed that it is not associated with the biosynthesis of HA but being tolerant to HA, performs glycolytic function when HA is produced by *A. oryzae* (Shinohara et al., 2019). To gain a better insight into the mechanism of regulation of secondary metabolism by GAPDH, we performed a detailed genome and transcriptome analysis of M7, a non-producer of all these metabolites (Chapter 3.2) and also performed an EMSA analysis to see if vGPD acts as a transcription factor (Chapter 3.3).

Chapter 3.2

Molecular characterization of M7, a radiation-induced mutant of *Trichoderma virens* that is deficient in secondary metabolites biosynthesis

3.2.1. Introduction:

The discovery of the "*vir*" cluster, to which vGPD belongs, came from isolation and analysis of a mutant M7 of *Trichoderma virens* (Mukherjee et al., 2006). Using this mutant and suppression subtraction hybridization (SSH), a set of genes was identified. Among these was a terpene cyclase that was designated as *vir4*. A gene cluster, named as "*vir*" cluster, was identified using a cosmid library screen (using *vir4* as a probe) and sequencing. A GAPDH was found to be associated with this cluster, which was speculated to be the housekeeping gene. After the whole genome sequence of *Trichoderma virens* Gv29-8 was published, the full gene cluster was identified (Mukherjee et al., 2012b). Knocking-out the *vir4* gene abolished the biosynthesis of 22 volatile sesquiterpenes, but not of non-volatile metabolites (Crutcher et al., 2013). Moreover, it was found that the gene cluster is also present in some *Aspergillus* spp. genomes, and GAPDH was consistently associated with this cluster. In Chapter 3.1, we demonstrated that the deletion of this GAPDH (vGPD) also abolished the biosynthesis of the volatile sesquiterpenes. Moreover, the biosynthesis of many non-volatile metabolites was also suppressed in the vGPD mutant, indicating a bigger role of vGPD beyond the biosynthesis of volatile sesquiterpenes, and pointing toward global regulation. Earlier, only the non-volatile profile of the M7 mutant in comparison with wild type was analyzed but in the present study expression of vGPD gene and the volatile profile has also been discussed. Since the metabolites profile of vGPD mutant largely overlaps with that of M7 mutant, we did a detailed genome and transcriptome analysis of M7 mutant with an intention of identifying the key regulatory gene(s) that might be global regulators of secondary metabolism in general, and vGPD in particular.

3.2.2 Materials and methods

3.2.2.1 Fungal strains and growth conditions

Trichoderma virens (IMI 304061), *Pythium aphanidermatum*, and *Rhizoctonia solani* (ITCC 4110) were used from our previous studies (Mukherjee et al., 2007). The non-conidiating mutant M7 of *T. virens* was obtained from our previous study (Mukherjee et al., 2006). The vGPD knockout mutant of *T. virens* was obtained in the present study (Chapter 3.1). The fungal cultures were grown in potato dextrose medium at 28 ºC and maintained at –80 ºC for long term storage.

3.2.2.2 Identification of secondary metabolites produced by wild type and M7

The non-volatile compounds were detected using thin-layer chromatography as described earlier (Mukherjee and Kenerley, 2010). WT and M7 cultures were grown for 3 days in potato dextrose broth and extracted with an equal volume of ethyl acetate. The extract was dried and reconstituted in 0.1 volume methanol and 100 µl was spotted on the pre-coated TLC plate. TLC plate was developed in acetone: chloroform: formic acid (28:70:2) and visualized under short-wavelength UV (254 nm). The spots were eluted with 100% methanol and subjected to LC-MS/MS analysis as described earlier (Bansal et al., 2018). For confirmation, LC-MS/MS analysis of standard viridin and viridiol (kind gift from Late Dr. CR Howell) and heptelidic acid (Cayman Chemical Company, USA) were also analysed and matched with the sample.

The Head Space Solid-Phase Microextraction (HS-SPME) and Gas Chromatography-Mass Spectrometry (GC-MS) technique were used for volatile compounds detection in wild type *T. virens* and M7. The instrument used for analysis was GC-MS 2010 Plus (Shimadzu, Kyoto, Japan). The GC-MS instrument is equipped with an injection port having SPME glass liner (Supelco) and RTX-5 column (5% diphenyl dimethylpolysiloxane, 10 m × 0.1 mm I.D.; Restek Corporation, Bellefonte, PA). The injection port was maintained at 270 ºC with no solvent cut. Helium was used as a carrier gas. GC column temperature was programmed as follows: 40 ºC for 5 min and then increased to 200 ºC at 4 ºC min^{-1}, held for 5 min and then increased to 280 ºC at 10 ºC min^{-1} with a final hold of 10 min. MS parameters were:

ionization voltage 70 eV, electron multiplier voltage 1 kV, and scan mode from 35 m/z to 350 m/z. Identification of the peaks was done by comparing their mass fragmentation pattern, Kovats retention indices, and from the data available in the spectral (Wiley/NIST) libraries of the instrument. The experiments were conducted in three biological replicates and repeated once.

3.2.2.3 Test of hydrophobicity

Wild type *T. virens* and M7 mutant were grown for 5 days on PDA at 28 ºC, and hydrophobicity was tested by adding 0.5% aqueous rose bengal or water and recorded for the disappearance of the drop after one hour (Mukherjee and Kenerley, 2010).

3.2.2.4 Assays for mycoparasitism

The mycoparasitic ability of wild type, M7 was assessed using confrontation assay by pairing the *Trichoderma* and the test pathogens simultaneously on the same PDA plate, placed opposite each other. Since *T. virens* wild type takes 4-5 days to completely overgrow these test pathogens (Mukherjee et al., 2007), ability of wild type or mutant to overgrow and lyse the mycelia of *P. aphanidermatum* and *R. solani* was recorded after 5 days of co-inoculation.

Mycoparasitic ability of wild type *T. virens* and M7 mutant was also observed under a microscope. The confrontation of WT and M7 with *P. aphanidermatum* and *R. solani* was performed on a plate containing 1% agar. After 2-3 days the mycelial growth of *Trichoderma* and the pathogen was visualized under a light microscope (Carl-Zeiss, USA). Images were acquired and documented.

P. aphanidermatum and *R. solani* cell walls are primarily composed of cellulose and chitin, respectively, and hence, we assessed the ability of wild type and M7 to utilize chitin and cellulose as carbon-source by growing them in Vogel's minimal medium containing chitin or

cellulose as a sole carbon source in liquid shake culture for 7-days at 28 ºC and 150 rpm. For the initiation of growth, 0.1% sucrose was added to the medium. All the experiments were conducted in four biological replicates.

3.2.2.5 Transcriptome analysis

Wild type and M7 were grown on PDA plates overlaid with the dialysis membrane (MWCO 12 kDa) opposite of *R. solani* for two days. For control, wild type *T. virens* and M7 were confronted with self. Mycelial mat from the zone of contact was harvested with a sterile spatula and frozen in liquid nitrogen and stored at -80 ºC until further use. Cultures were grown in four biological replicates and samples were pooled before RNA extraction. RNA was extracted with Tri Reagent and transcriptome sequencing was outsourced on Illumina HiSeq 2500 platform at M/S Scigenom, Cochin, Kerala, India as discussed earlier (Mukherjee et al., 2019). Trinity software was used for assembly of the cleaned reads produced by standard data pre-processing methods including adaptor trimming, quality filtering, and end trimming. Gene expression estimation was performed by aligning the trimmed reads to the assembled transcriptome using the Bowtie2 program. DESeq program was employed for differential gene expression analysis and an in-house pipeline CANoPI (Contig Annotator Pipeline) was used for transcriptome annotation. Fragments Per Kilobase of transcript per Million (FPKM) values were considered for fold-change calculation and log 2-fold change ≥2 was considered for comparison of gene expression.

3.2.2.6 Whole-genome sequencing and analysis

Total genomic DNA was extracted and whole-genome sequencing of the M7 mutant was performed using Illumina platform. Illumina TruSeq Nano DNA HT Library Preparation Kit was used for paired-end sequencing library preparation. The generated libraries were sequenced on Illumina Nextseq 500 using 2 x 150 bp chemistry. *T. virens* wild type sequence (LQCH00000000) was used as a reference genome for M7 genome mapping using Burrows-

Wheeler Aligner (BWA) (v. 0.7.5a) program with optimized mapping parameters. The genome coverage, gene prediction, and a total number of single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (INDELs) were analyzed and compared with the reference genome. The absence of the genes in the deleted regions was further confirmed by PCR amplification with gene specific primers of representative genes using the following conditions: denaturation at 95 ºC for 5 minutes, annealing at 55-65 ºC for 45 seconds, extension at 72 ºC for 0.5 kb $mins^{-1}$, repeat cycles from step 2 to 4 for 30 cycles and a final extension at 72 ºC for 10 minutes. The primers used are listed in Table 3 of Annexure I.

3.2.3 Results

3.2.3.1 M7 mutant downregulated in the biosynthesis of secondary metabolites

Similar to the previous report, the TLC profile of wild type and M7 culture filtrate shows that none of the non-volatile metabolites including viridin, viridiol and heptelidic acid are detected in M7 mutant (Mukherjee et al., 2006) (Fig. 3.2.1A). In the present study, GC-MS analysis of head-space gas shows that M7 mutant is also down-regulated in the biosynthesis of volatile metabolites that are present in wild type *T. virens*; only 13 of 73 metabolites being detected in M7 mutant (Fig. 3.2.1B and Table 1 of Annexure III). Similar to the M7 mutant, the volatile and non-volatile profile of the vGPD knockout mutant also revealed that the vGPD mutant is downregulated in the biosynthesis of secondary metabolites.

3.2.3.2 M7 mutant is deficient in hydrophobicity in addition to conidiation

M7 mutant does not conidiate while wild type *T. virens* produces green pigmented conidia (Fig. 3.2.2). The deficiency in conidiation is associated with loss of hydrophobicity. Aqueous rose bangal or water drops stay on the wild type colony for several hours but diffuse through the M7 colony immediately after application (Fig. 3.2.3).

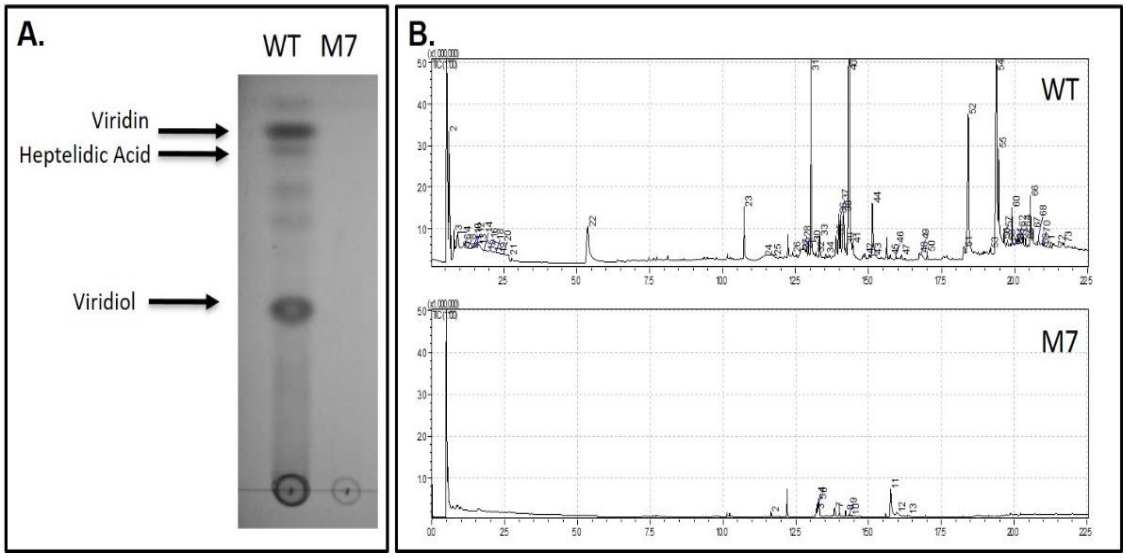


Figure 3.2.1 Secondary metabolite profile of *T. virens* wild type (WT) and M7.
A). Thin layer chromatography analysis of WT and M7 filtrate **B).** GC-MS analysis of WT and M7 head-space gas. The peak numbers correspond to the volatile metabolites detected as described in Table1 of Annexure III

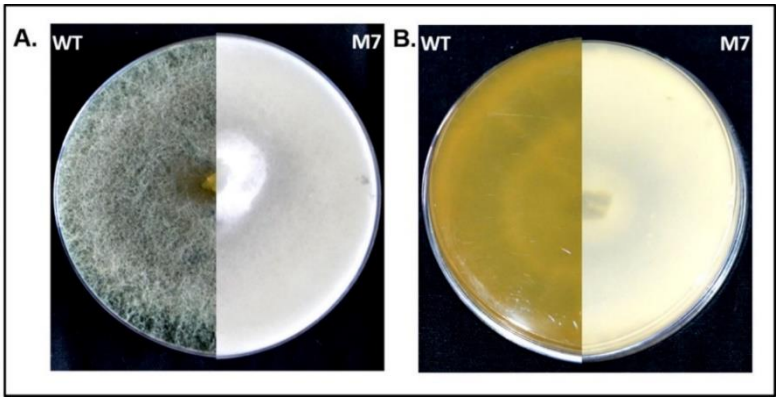


Figure 3.2.2 Culture of *T. virens* wild type and M7 grown on PDA for 5 days A). Front side B). Reverse side.

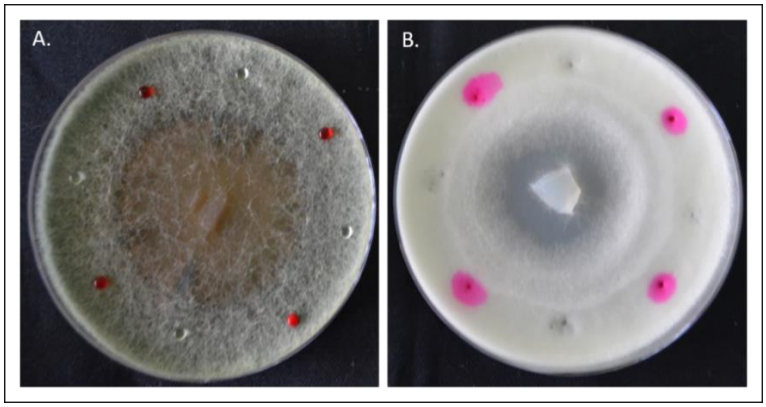


Figure 3.2.3 Culture of *T. virens* wild type and M7 with rose bengal and water placed on it A). Wild type *T. virens* **B).** M7

3.2.3.3 M7 is defective in mycoparasitism

Confrontation assay demonstrated that M7 is not able to overgrow the plant pathogens *R. solani* and *P. aphanidermatum* while wild type *T. virens* completely overgrew the colonies in 5 days (Fig. 3.2.4). Moreover, M7 could not utilize chitin and cellulose as a carbon source even when the medium was amended with 0.01% sucrose. (Fig. 3.2.5). While wild type showed mycoparasitic coiling on *R. solani* and *P. aphanidermatum*, the mycelium of M7 was also unable to coil over the mycelium of pathogens as shown in figure 3.2.6. These results confirm that M7 has no mycoparasitic activity.

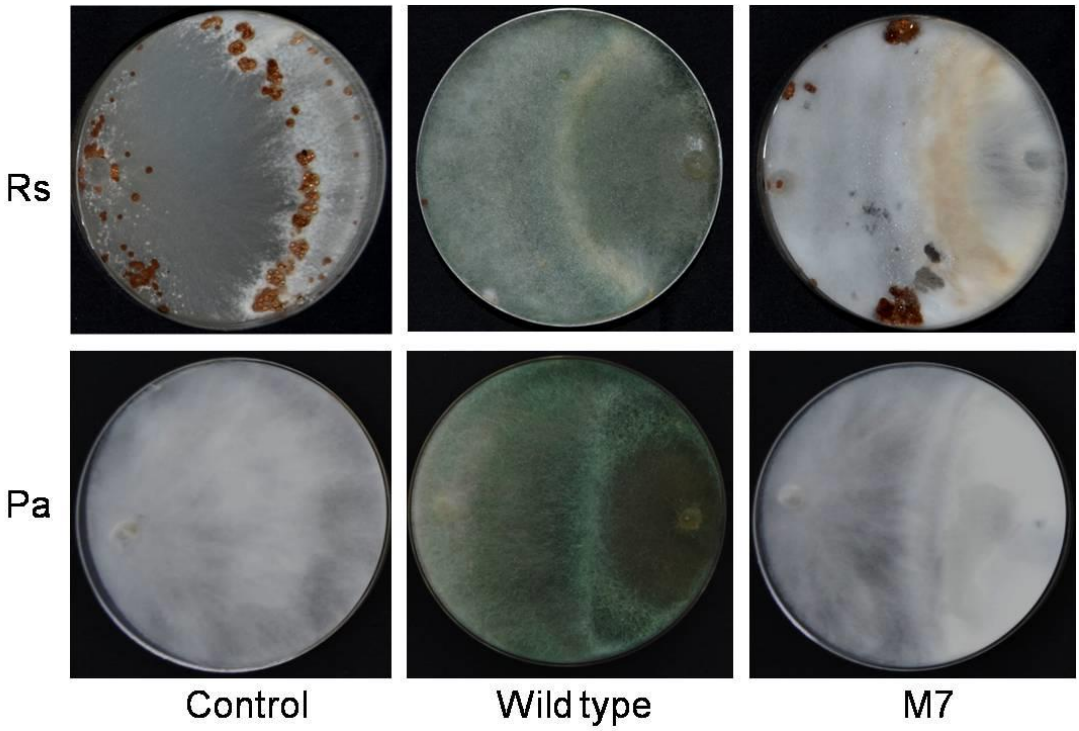


Figure 3.2.4 Confrontation assay demonstrating that M7 is non-mycoparasitic as it cannot overgrow *Rhizoctonia solani* (Rs) and *Pythium aphanidermatum* (Pa) after 5 days of incubation.

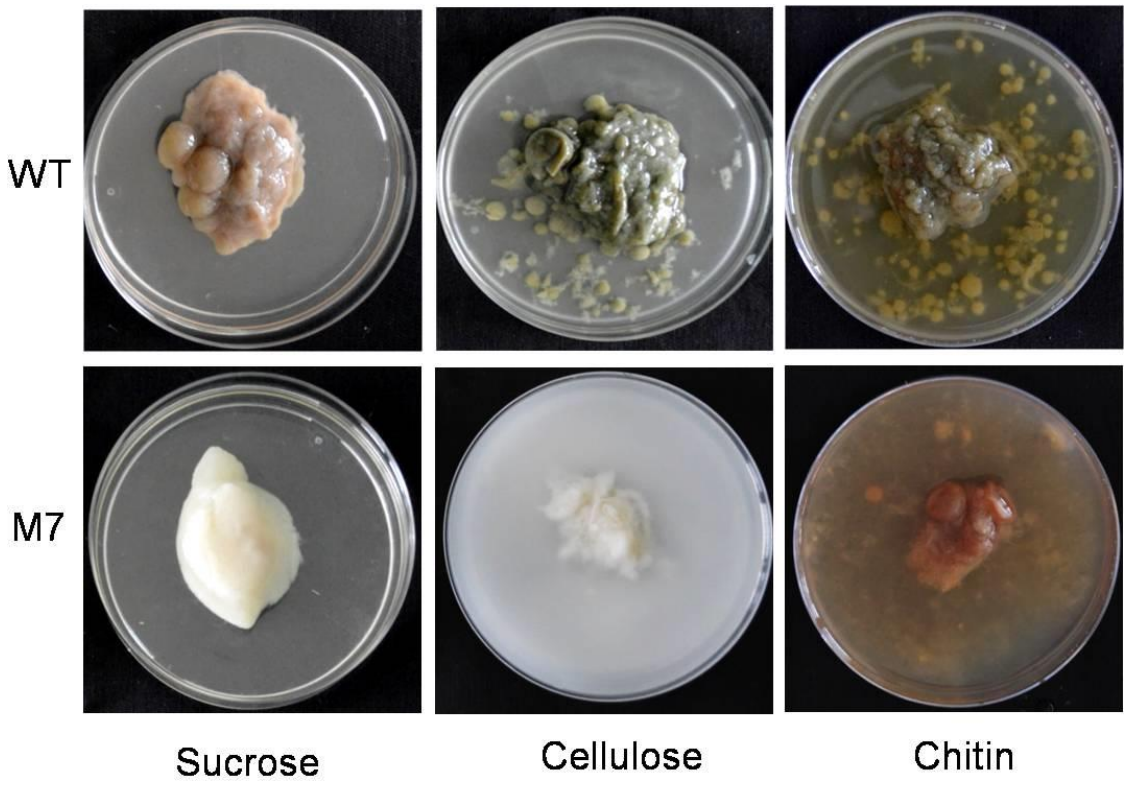
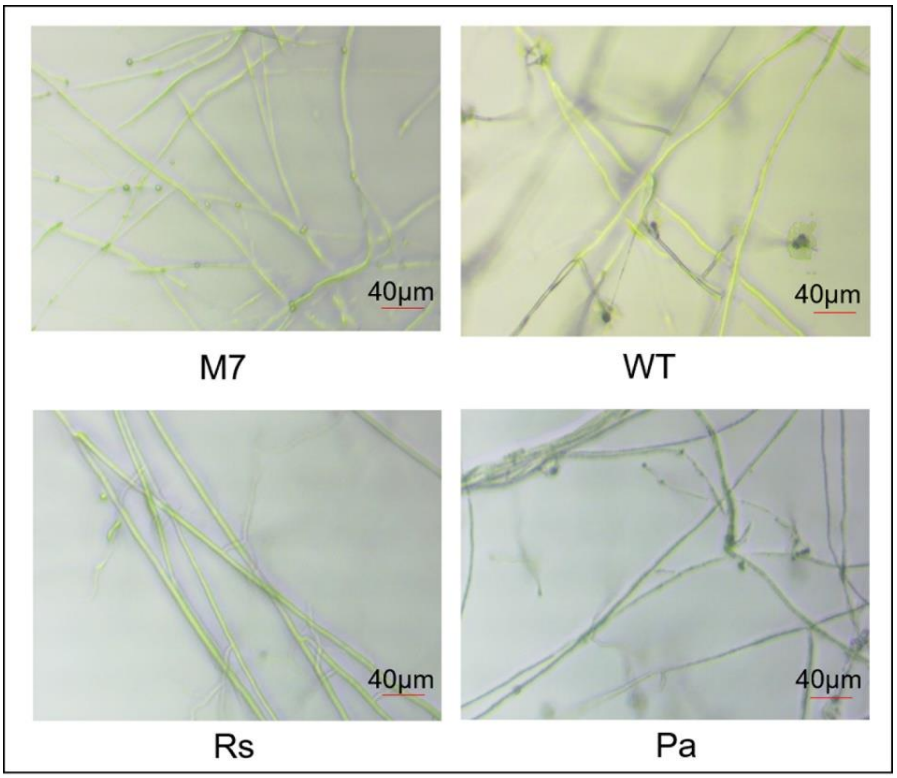


Figure 3.2.5 Seven-day old culture of *T. virens* wild type (WT) and M7 in Vogel's minimal medium containing sucrose, cellulose or chitin. Please note that WT can utilize chitin and cellulose but M7 can utilize sucrose only.



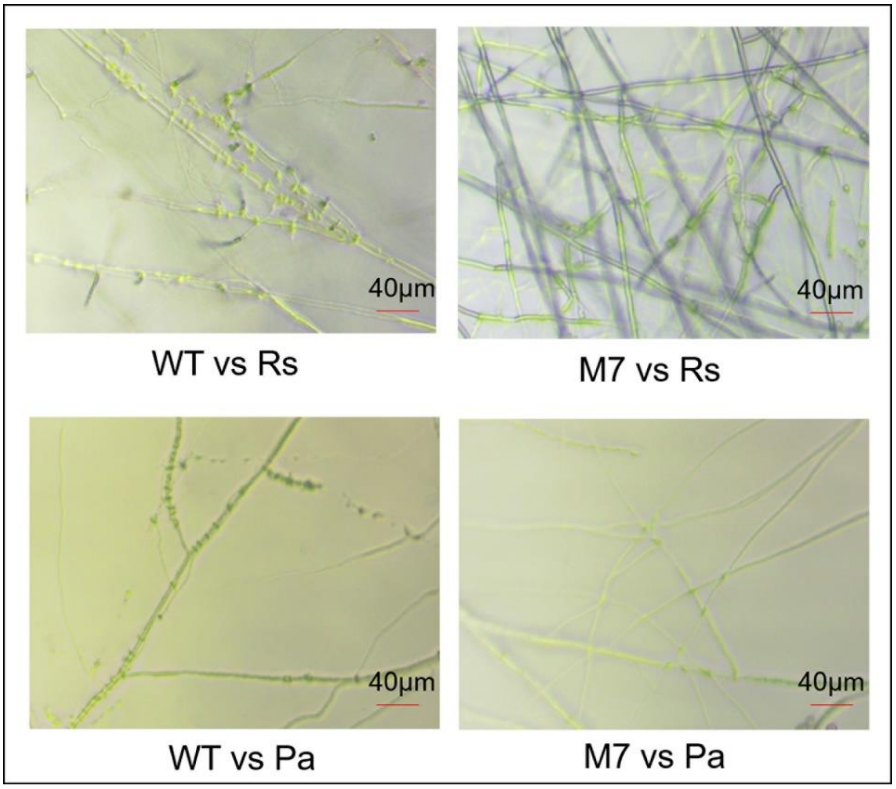


Figure 3.2.6 Images of mycoparasitic interaction of *T. virens* wild type (WT) and M7 with *Rhizoctonia solani* (Rs) and *Pythium aphanidermatum* (Pa) in 1% agar after 2 days (scale bars, 40 µm).

3.2.3.4 Several genes for secondary metabolism, hydrolytic enzymes and hydrophobicity are down-regulated in M7

De novo transcriptome analysis of the wild type and M7, while interacting with self or *R. solani* (Fig. 3.2.7), indicated that as many as 463 genes are down-regulated in M7 (Table 2 and Table 3 of Annexure III, Fig. 3.2.8, Fig. 3.2.9 and Fig. 3.2.10). Eleven of these genes which are under-expressed including hydrolases, oxidoreductases, transferases, transcription factors, cytochrome P450, are also deleted in the M7 genome (See section 3.2.3.5). In self-confrontation, 294 genes were down-regulated in M7 compared to wild type (Table 2 of Annexure III). Among these, are genes for secondary metabolism (including genes coding for NRPSs, PKSs, and terpene cyclases, modifying enzymes, transporters, and transcription factors), other cytochrome P450s (not associated with secondary metabolism gene clusters),

carbohydrate-active enzymes (CAZymes), peptidases and hydrophobins (Fig. 3.2.9). In the confrontation with the plant pathogen *R. solani*, a larger set of genes (375) was down-regulated in M7 (including genes deleted in the M7 genome) (Table 3 of Annexure III). A large number of genes involved in secondary metabolism (69), carbohydrate utilization (31), and transporters (30) are down-regulated in M7 in confrontation with *R. solani* (Fig. 3.2.10). This is in agreement with the fact that M7 does not produce a detectable amount of secondary metabolites and also does not parasitize the plant pathogens *P. aphanidermatum* and *R. solani*.Eighty-eight genes are exclusively downregulated in M7 while 169 genes are exclusively downregulated in M7 confronted with *R. solani*, 206 genes are common between M7 transcriptome alone and transcriptome of M7 confronted with plant-pathogen *R. solani* (Fig. 3.2.11). Six of the genes (protein IDs. TRIVIDRAFT_49849, TRIVIDRAFT_74289, TRIVIDRAFT_74291, TRIVIDRAFT_56195, TRIVIDRAFT_230740, TRIVIDRAFT_53375) that are downregulated in M7 were identified in our earlier study as under-expressed in M7 in RT-PCR analysis (Mukherjee et al., 2006).

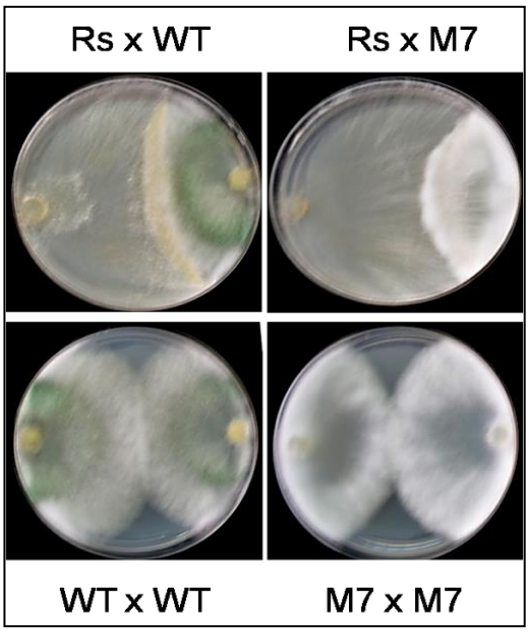


Figure 3.2.7. Confrontation of *T. virens* wild type (WT) and mutant (M7) with *Rhizoctonia solani.* The photograph was taken 48 hours after co-inoculation. The mycelia from contact point were used for RNA extraction and transcriptome analysis.

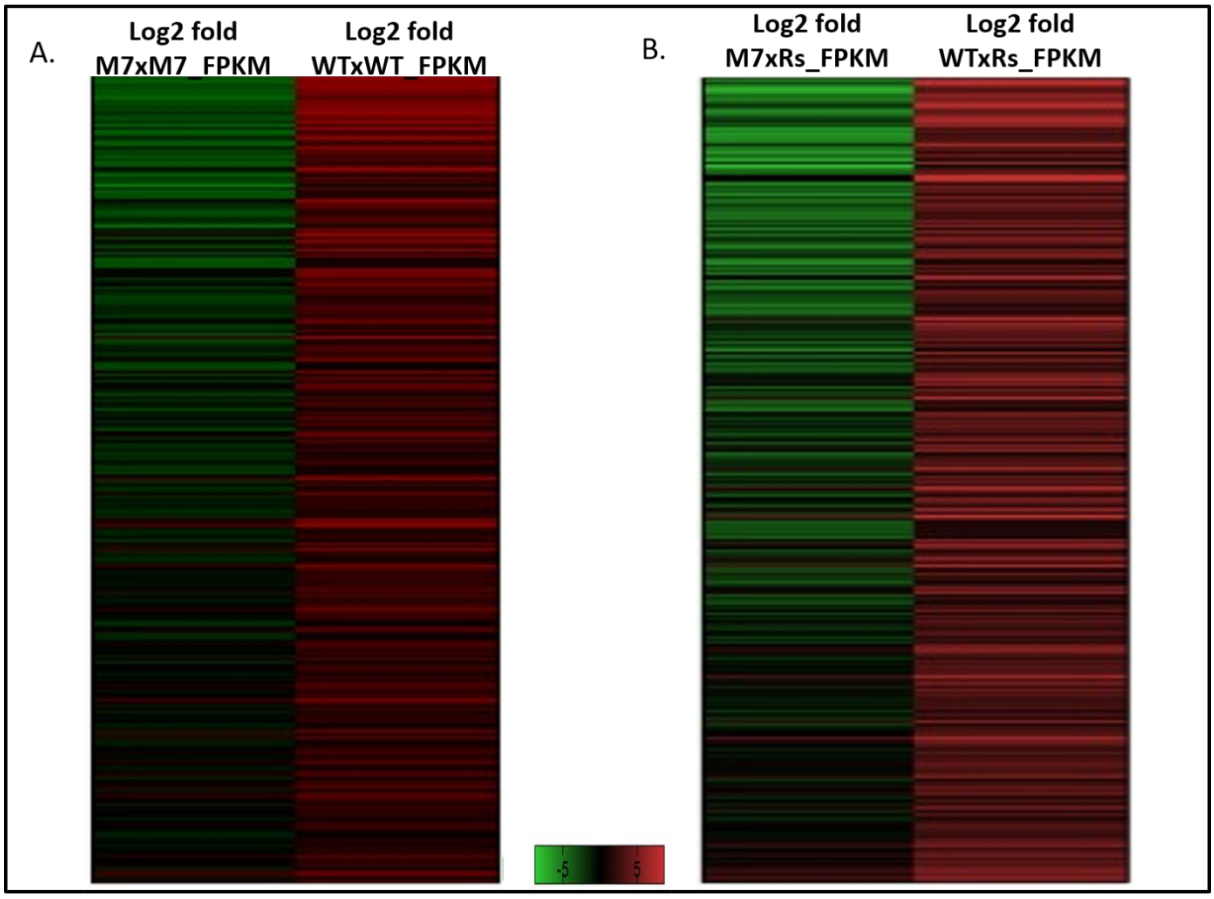


Figure 3.2.8. A. Heatmap of genes downregulated in M7XM7 < WTXWT. B. Heatmap of genes downregulated in M7XRs < WTXRs

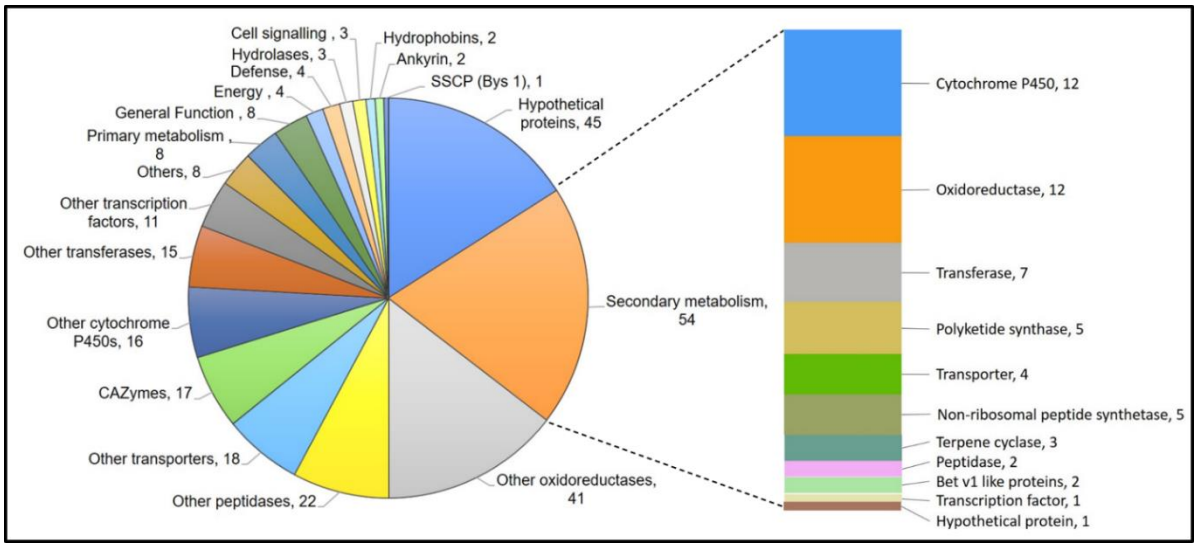


Figure 3.2.9. Pie chart representing genes down-regulated in M7 compared to the wild type *T. virens* including eleven genes deleted in the M7 genome.

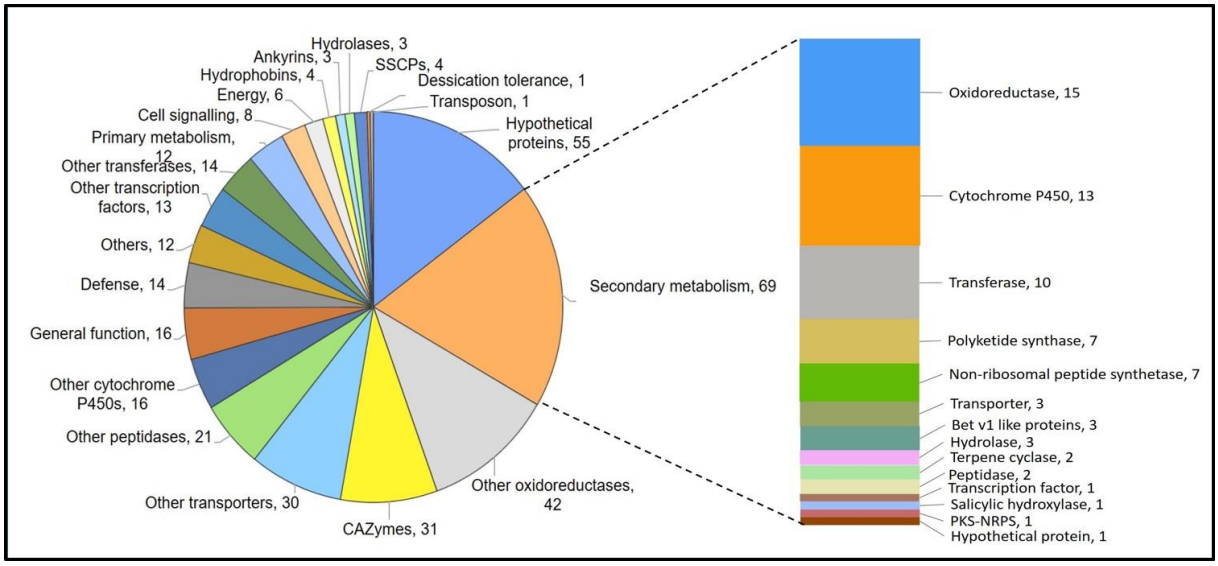


Figure 3.2.10. Pie chart representing genes down-regulated in M7 in confrontation with *Rhizoctonia solani* including eleven genes deleted in the M7 genome.

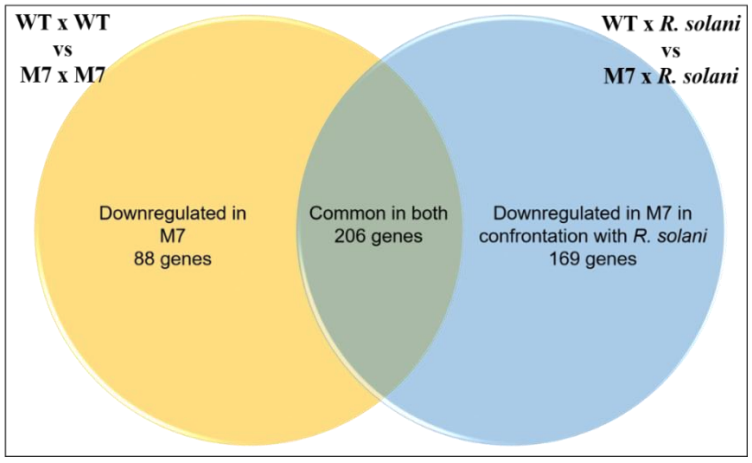


Figure 3.2.11. Venn Diagram presenting downregulated genes in M7 and M7 in confrontation with *Rhizoctonia solani*

3.2.3.5 M7 genome has large deletions

. Whole genome sequencing of the M7 mutant was performed using the Illumina platform to generate 2.1 GB data. The paired end sequencing library was of mean size 624 bp. The generated library was sequenced and mapped to *T. virens* wild type assembly as a reference genome. A genome alignment revealed five deletions across three scaffolds

(scaffold 1, 20, and 58) totalling about 250 Kb, comprising of 71 genes (Table 2 of Annexure II). The scaffold 1 contains three deletions (deletion 1 containing 167 kb region comprising of 44 genes, deletion 2 containing 35 kb region comprising of 13 genes and deletion 3 containing 0.3 kb region comprising of 1 gene). Scaffold 20 and 58 contain 1 deletion each of 8.5 kb and 40 kb region and comprising of 3 and 11 genes respectively (Fig. 3.2.12). Eight transcription factors are identified in the deleted region namely CDS378, CDS379, CDS390, CDS395, CDS401, CDS409, CDS419 and CDS1018. These transcription factors might be involved in global regulation or cluster-specific regulation of the genes which are downregulated in the M7 mutant. Apart from transcription factors, other important genes with putative functions in regulation, transportation, mycoparasitism, metabolism, etc are also deleted in the M7 genome. One Non-ribosomal peptide synthetase (NRPS), Tex9, and one putative polyketide synthase, PKS6 are also deleted in the M7 genome.

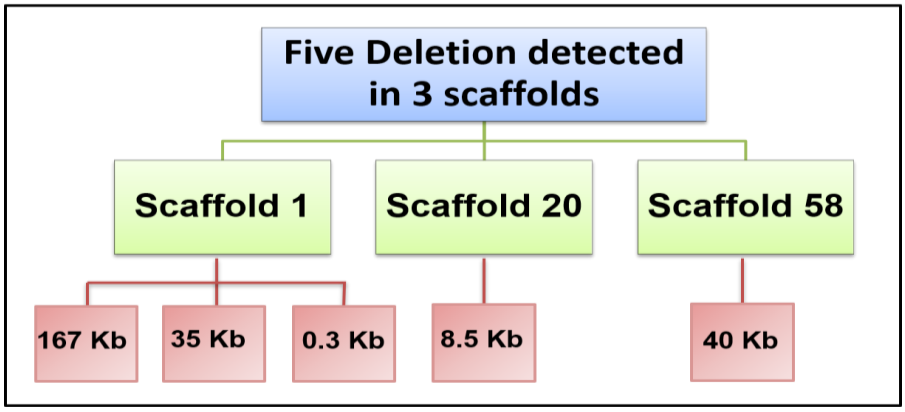


Figure 3.2.12. Deletions identified in *Trichoderma virens* mutant (M7) genome.

To further confirm these deletions, PCR of representative genes was performed using wild type and M7 genomic DNA as a template. No amplification of these genes was obtained in M7, confirming the gene deletions (Fig. 3.2.13 and Fig. 3.2.14). Besides, 118 SNPs and small 57 INDELs were also detected in M7 (Table 4 of Annexure III).

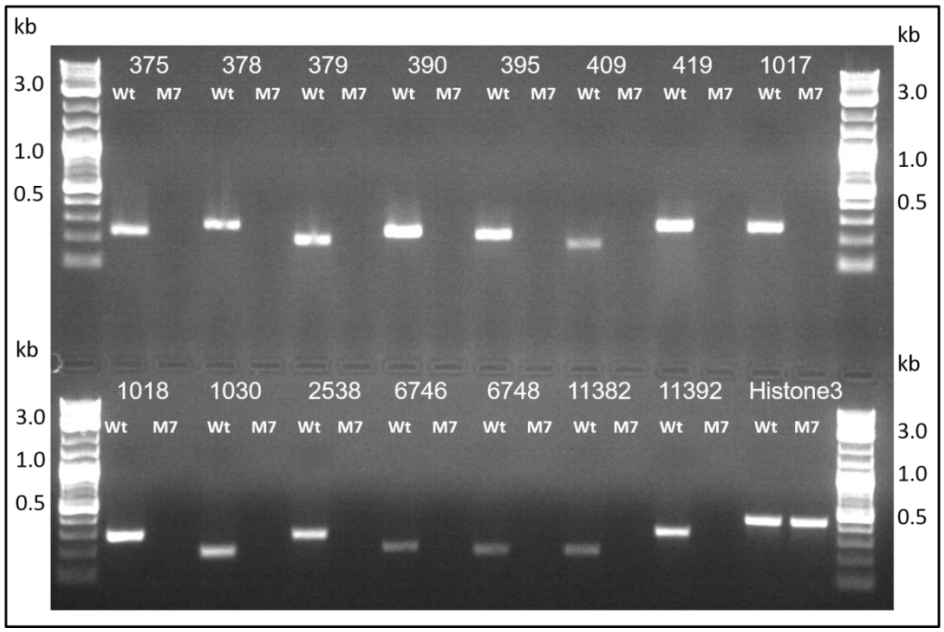


Figure 3.2.13 PCR analysis of representative genes that are deleted in M7, using genomic DNA as a template. Histone 3 was used as a positive control. The number corresponds to gene IDs as per Table 2 in Annexure II.

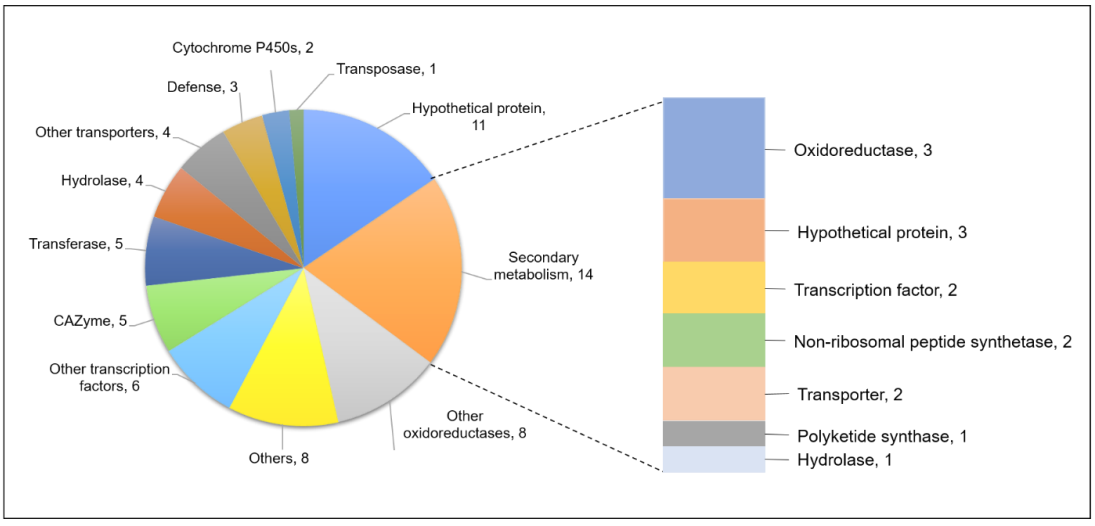


Figure 3.2.14. Pie chart representing genes deleted in M7 genome compared to the wild type *T. virens* genome.

3.2.4 Discussion:

Since there is an overlap in phenotypes (suppressed secondary metabolism) in both vGPD mutant and M7, we studied the genetics of M7 intending to get a better understanding

of global regulation of secondary metabolism in *T. virens*. Classically induced mutants are valuable tools in genetic research (Casselton and Zolan, 2002). Identification of gene function by mutation is a tedious strategy, due to the involvement of steps like complementation and plasmid rescue (Firon et al., 2003; Madi et al., 1994; Mattern et al., 1992; Navarro-Sampedro et al., 2008). Recent technological advances, especially high-throughput sequencing has facilitated the identification of such mutations (Le Crom et al., 2009; McCluskey et al., 2011). *Trichoderma* spp. produce a large number of useful secondary metabolites with diverse biological activities (Zeilinger et al., 2016). A large number of genes for secondary metabolism (69) were down-regulated in M7 during the confrontation with *R. solani*. Viridin is one of the oldest known metabolites produced by *T. virens* possessing anti-cancer and anti-fungal activities while viridiol is a phytotoxic agent produced by irreversible enzymatic conversion of viridin (Howell et al., 1993, Jones and Hancock, 1987). Gliovirin is a 'P' strain-specific non- ribosomal peptide with anti-oomycete activities (Howell et al., 1983). Non-volatile compounds like viridin, viridiol, and heptelidic acid were detected by TLC followed by LC-MS/MS analysis, while gliovirin was detected by LC-MS in *T. virens* wild type, the mutant does not produce these metabolites. Also, M7 is down-regulated in the biosynthesis of volatile organic compounds, including volatile sesquiterpenes, like germacrenes, caryophyllene, alloaromadendrene, and gamma-muurolene, all having important biological activities (Fig. 3.2.1B, Table 1 of Annexure III). Volatile compounds produced by *Trichoderma* spp. promote plant growth and have antimicrobial effects (shield) against plant pathogens (Hung et al., 2013; Vinale et al., 2008a). In transcriptome analysis, many secondary metabolism related genes and gene clusters were down-regulated in M7 (Table 2 and Table 3 of Annexure III). Viridin biosynthesis gene cluster comprises of 21 genes (Bansal et al., 2018), of which, 16 genes were detected to be down-regulated in M7. Similarly, gliovirin biosynthetic gene cluster has 22 genes (Sherkhane et al., 2017), of which 16 genes were

detected to be down-regulated in M7. All the eight genes from the "*vir*" cluster which are associated with volatile sesquiterpene compounds production (Crutcher et al., 2013) were down-regulated in M7 including the vGPD gene which is the part of the vir cluster. Also, eight NRPSs, one NRPS/PKS, seven PKSs, and three terpene cyclase genes were down-regulated in M7. The NRPS genes that are down-regulated in transcriptome analysis of M7 are Tex1, Tex2, Tex5, Tex7, Tex21 (NRPSs) and Tex14 (NRPS/PKS). In addition, certain genes belonging to Tex6, Tex8, Tex10, and Tex12 (NRPS/PKS) cluster were also down-regulated in M7. Tex1 and Tex2 code for peptaibol synthetase (Wiest et al., 2002; Mukherjee et al., 2011). Tex10 is involved in the biosynthesis of intracellular siderophore ferricrocin while Tex 21 is responsible for the biosynthesis of extracellular siderophores (Mukherjee et al., 2018). The metabolites produced by Tex5 and Tex7- NRPS genes and Tex12 and Tex14 - NRPS/PKS hybrid genes are not known. Of these NRPS genes, Tex5, Tex14, and Tex21 are down-regulated in M7 during the confrontation with *R. solani* (Table 3 of Annexure III). Seven PKS genes, namely PKS4, PKS8, PKS14, and PKS17, one un-annotated PKS (Trividraft_53518), and two orthologs of *Arthroderma benhamiae* (Acc. Nos. DAA76265 and EGE05473) PKSs were downregulated in M7 during confrontation with *R. solani*. The biosynthetic products of none of these PKS genes are known. Besides, some members of PKS3 and PKS6 gene clusters were also found to be down-regulated in M7. PKS4 is an orthologue of pigment-forming PKS from *Trichoderma reesei* responsible for green conidial pigment biosynthesis and is involved in mechanical stability and stress tolerance (Atanasova et al., 2013b). Since M7 does not produce conidia, it is not surprising that the pigment PKS4 gene is down-regulated in M7. In addition to PKS4, another gene (*con6*) that is associated with conidiation in *Neurospora crassa* (White and Yanofsky, 1993) was also down-regulated in M7. Interestingly, three genes of the PKS3 cluster were also down-regulated in M7 in interaction with *R. solani*.

Having noted that the mutant M7 has deficiency in conidiation, hydrophobicity, secondary metabolism and mycoparasitism (all traits that dictate the success of *Trichoderma* spp. as plant beneficial fungi), which was apparent from the phenotyping and transcriptome data analysis, we did a whole genome sequence analysis of the mutant vis-à-vis the wild type strain. It was interesting to note that M7 genome has 5 deletions comprising of as many as 72 genes. Since the mutant was generated using gamma-ray induced mutagenesis, deletions may be expected. Two large secondary-metabolism related gene clusters (PKS6 and Tex9) are deleted in the mutant. Four genes for glycoside hydrolase (GH) are also deleted, and so are several oxidoreductases. Since the dataset of genes that are under-expressed in M7 is much larger that the number of genes that are deleted in the genome, it is likely that one or a few of these genes that are deleted or affected due to other mutations (Table 2 in Annexure II, Table 4 of Annexure III) could have pleiotropic phenotype regulating a big set of genes. Our findings identify possible candidates for future research leading to identification of master regulator gene(s) in *Trichoderma*, a biotechnologically important fungal genus. Also, since the mutant is under-regulated in secondary metabolism and a large number of carbohydrate active enzymes, it could be possible to develop this mutant as a microbial cell factory for production of secondary metabolites and proteins.

We have recently reported isolation and characterization of a mutant (G2) of *T. virens* that is up-regulated in secondary metabolites biosynthesis (Mukherjee et al., 2019). A total of 140 genes were found to be up-regulated in G2 over wild type when grown on PDA medium. Of these, as many as 45 genes are down-regulated in M7, including genes for CAZymes (13), secondary metabolism (5), Cytochrome P450 (3), SSCPs (2), oxidoreductase (3), peptidases (3), metabolism (5), hydrophobins (3), hypothetical proteins (5), transporter (1) and others (2). G2 and M7 thus seem to be two contrasting “gene- regulation” mutants of

T. virens. While G2 has been developed as an improved biocontrol formulation (Mukherjee et al., 2019), M7 can serve as an excellent genetic tool for understanding *Trichoderma* biology.

We provide here evidence that, in addition to suppressed secondary metabolism, the mutant also lacks hydrophobicity and mycoparasitism. Hydrophobicity is imparted in fungi by small secreted cysteine-rich proteins known as hydrophobins. One hydrophobin (Trividraft_49849) was reported earlier to be down-regulated in M7 (Mukherjee et al., 2006). In the transcriptome data, we noted one additional hydrophobin to be down-regulated in the mutant. Hydrophobins are associated with conidiation (Bayry et al., 2012; Wessels, 1996), and the loss of hydrophobicity in this non-conidiating mutant is not unexpected. *Trichoderma* hydrophobins also play important roles in the attachment to root (Viterbo and Chet, 2006), elicitation of plant defense (Zhang et al., 2019), and tolerance to biotic and abiotic stresses (Przylucka et al., 2017). Hydrophobins are also imvolved in mycoparasitic interactions in *Trichoderma* (Guzman-Guzman et al., 2017). We have noticed two additional hydrophobins to be downregulaed in the mutant in interaction with *R. solani* (Fig. 3.2.10). These hydrophobins might be involved in mycoparasitism of *T. virens* on *R. solani*.

T. virens is an aggressive mycoparasite on many plant pathogenic fungi (Druzhinina et al., 2011). The mutant, however, has lost this property which can be explained by the lack of up-regulation of a large number of CAZymes (Carbohydrate-Active enZYmes, a database contains information about enzymes involved in the synthesis, metabolism, and recognition of complex carbohydrates) known to be involved in parasitism on other fungi (Carsolio et al., 1999; Djonović 2006, 2007; Haran et al., 1996). CAZymes of glycoside hydrolase 2 (GH-2), GH-3, GH-18, GH-30, GH-47, GH-55, GH-75, and GH-92 families were down-regulated in M7. Furthermore, the down-regulation of CAZymes of GH-16, GH-20, GH-78, and GH-79 families only during the confrontation with *R. solani* suggests the role of these CAZymes in

mycoparasitism. Atanasova et al. (2013a) reported the up-regulation of many gene families like metabolism, transporters, signal transduction, transcriptional regulators, defense, etc. in *T. virens* during the confrontation with *R. solani*. Similarly, in our study, we have detected down-regulation of different groups of genes involved in secondary metabolism (69), oxidoreductase (42), CAZymes (31), peptidases (21), transporters (30), transcription factors (13), transferases (14) and defense-related (14) in *T. virens* during confrontation with *R. solani*. Approximately, twice the number of genes for transporters (30) and carbohydrate utilization groups (31) were down-regulated in M7 during confrontation with *R. solani* as compared to confrontation with the self, where only 18 genes for transporter and 17 for carbohydrate utilisation were down-regulated (Fig. 3.2.9). These results indicate that the mutant, which cannot parasitize the host fungus *R. solani*, fails to respond to its presence. Role of most of these genes in mycoparasitism is not known and further studies on the role of these genes could throw novel insights into the phenomenon of fungus-fungus interactions.

Our studies point to the fact that M7 has a mutation(s) affecting the regulation of a wide set of genes involved in conidiation, secondary metabolism, and mycoparasitism. Even though the genome analysis revealed the deletion of about 250 kb encompassing 71 ORFs, the present data cannot predict the master regulator(s). However, eight transcription factors that are deleted in the M7 genome could be the chosen candidates. Further studies by gene knockout/complementation would be required to identify the gene(s) for global regulation.

Chapter 3.3

Expression of the heptelidic acid-insensitive GAPDH (vGPD) from *Trichoderma virens* in *E. coli* and its biochemical and biophysical characterization

3.3.1 Introduction

As detailed in Chapter 3.1, we have identified a secondary-metabolism gene cluster associated GAPDH (designated as vGPD) to be involved in secondary metabolism; deletion of the vGPD gene abolished the biosynthesis of both volatile and non-volatile secondary metabolites, including viridin, viridiol, heptelidic acid, gliovirin, and several volatile sesquiterpenes. This is rather unusual as glyceraldehyde-3-phosphate dehydrogenase (GAPDH, EC 1.2.1.12) is regarded as a glycolytic enzyme catalyzing the oxidative phosphorylation of glyceraldehyde 3-phosphate. Even though GAPDH is known to perform alternate non-metabolic functions, like DNA replication and repair, gene expression and signal transduction (Sirover, 1999, 2011; Tristan, 2011), its role in secondary metabolism is new. The multi-functionality of glycolytic GAPDH in performing other functions is attributed to oligomerization, differential localization, complex formation with other proteins, and chemical modifications (Sirover, 2005, 2011; Zaffagnini et al., 2013; Kosova et al., 2017). In microbes, the functional diversity is also brought about by the presence of isoforms of the GAPDH gene in some antibiotic-producing organisms. For instance, the *Streptomyces arenae* produces pentalenolactone antibiotic, a sesquiterpene metabolite, that irreversibly inactivates GAPDH by binding at the active site (Hartmann et al., 1978). *S. arenae* protects its own GAPDH from the antibiotic by the deployment of an alternate isoform of GAPDH (an additional copy present in the genome and associated with the pentalenolactone gene cluster) which is less sensitive to this metabolite (Maurer et al., 1983). However, earlier, the functions of these GAPDHs were not studied.

Trichoderma virens produces heptelidic acid (HA), a selective inhibitor of GAPDH. This is a sesquiterpene lactone that irreversibly inhibits GAPDH by binding at the active site to the

sulfhydryl group of the cysteine residue (Endo et al., 1985). *T. virens* also possesses an isoform of GAPDH which is tolerant to heptelidic acid (HA) and has been proposed to take over the glycolytic activity when heptelidic acid is being produced by *T. virens* (Sakai et al., 1990). The HA-tolerant GAPDH (vGPD) is encoded by a different gene and is not a modification of the glycolytic or heptelidic acid-sensitive GAPDH. The GAPDH is associated with a secondary metabolism-related gene cluster in *T. virens* and a few *Aspergillus* spp. genomes (Mukherjee et al., 2012b, Mukherjee et al., 2006; Crutcher et al., 2013).

In the present study, we have expressed the HA-insensitive GAPDH (vGPD) in *E. coli* and studied its biochemical and biophysical characteristics as a first step to understand the structural differences between these two isoforms- one sensitive and the one tolerant to heptelidic acid.

3.3.2 Materials and Methods:

3.3.2.1 Bacterial strains:

Escherichia coli strain DH5α (Novagen) was used for vector amplification. pNH-TrxT plasmid from Opher Gileadi (Addgene plasmid # 26106) and *E. coli* BL21 star (DE3) strain (Novagen) were used as an expression vector and host strain, respectively.

3.3.2.2 Cloning of the *vGPD* gene in *E. coli*:

The gene for *vGPD* was PCR-amplified from an EST clone obtained from the cDNA library of *Trichoderma virens* prepared by (Mukherjee et al., 2003). The primer sequences for the amplification of the full-length *vGPD* cDNA are given in Table 2.1 in the "Material and Methods" section. The PCR product was purified using the PCR product purification kit (Sigma-Aldrich, USA) and the ligation independent cloning method was utilized for cloning of PCR

amplified *vGPD* gene into pNH-TrxT expression vector yielding pNH-TrxT-vGPD construct. The recombinant pNH-TrxT-vGPD construct was transformed into the cloning host, *E. coli* DH5α competent cells, and the transformed cells were grown onto LB agar plates (supplemented with 25 µg mL^{-1} of kanamycin) at 37 ^{o}C. Colony PCR using *Taq* DNA polymerase was performed for the selection of transformed colonies and the recombinant plasmid was purified using a plasmid purification kit and sequenced.

3.3.2.3 Expression and purification of vGPD protein:

For vGPD protein expression, the purified pNH-TrxT-vGPD plasmid was transformed into the expression host BL21 Star (DE3). The transformed *E. coli* BL-21 cells harboring the recombinant plasmid construct pNH-TrxT-vGPD were cultured in 5 mL LB medium (containing 25 µg mL^{-1} of kanamycin) at 37 ºC overnight. The primary culture was diluted (1:100) into fresh LB containing 25 µg mL^{-1} of kanamycin. The recombinant cells were grown at 37 ºC in LB medium supplemented with 50 µg mL^{-1} of kanamycin until OD_{600} reached 0.6. Higher concentration of the antibiotic was used during large-scale protein expression to avoid bacterial contamination. Expression of protein was induced by adding 0.5 mM IPTG followed by shaking at 18 ºC overnight. The cells expressing recombinant proteins were harvested by centrifugation at 10,000xg for 5 min and re-suspended in buffer A (20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1 mM PMSF, 2 mM NAD+, 2 mM TCEP and 0.2 mM EDTA) and containing 1 mg mL^{-1} lysozyme. The re-suspended cells were stored at -80 ºC until further use. The cells were thawed at room temperature and then lysed using a sonicator on ice for 10 min at 30% duty cycle in pulse mode (3 sec ON, 3 sec OFF). The lysed cells were centrifuged at 21,000xg for 30 min at 4 ºC and the clear supernatant was collected in fresh autoclaved

centrifuge tubes for further purification steps. The recombinant vGPD protein was purified using Ni-iminodiacetic acid (Ni-IDA) matrix pre-equilibrated with buffer A containing 20 mM imidazole to avoid nonspecific binding of the protein. The clear supernatant of the cell-free extract containing recombinant vGPD protein with histidine-tag was loaded onto the pre-equilibrated column. The flow-through from the column was collected. The column was washed with 20 column volume of buffer A containing 50 mM imidazole to remove non-specifically bound proteins and the wash-through was also collected. The protein was eluted to near homogeneity using the same buffer containing 100 mM to 300 mM imidazole. Eluted fractions were subjected to SDS-PAGE analysis and the fractions containing the recombinant protein were pooled. Overnight dialysis was performed to remove imidazole from the protein. The protein was concentrated using Corning spin-X UF concentrators (10K MWCO, CLS431478 Sigma-Aldrich, USA) and stored in a storage buffer containing (20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10% glycerol, 1 mM PMSF, 2 mM NAD^+ and 2 mM TCEP). The protein concentration was determined using modified Folin-Lowry's method using bovine serum albumin (BSA, fermentas, USA) as a standard.

The identity of purified His-tagged vGPD protein was confirmed by the Western blotting. The purified vGPD protein was separated on 12% SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membranes by electro-blotting. Non-fat-dried milk bovine blocking solution (Sigma-Aldrich, USA) was used at working concentration of 3%. Membranes were washed with phosphate buffer saline (PBS) and then incubated for 1 hr at room temperature with a 1:10000 dilution of mouse monoclonal anti-His-tag antibody (SAB2702218-Sigma-Aldrich, USA). Blots were again washed with PBS buffer and incubated with alkaline phosphatase-conjugated rabbit anti-mouse IgG antibody (1:50 000 dilution; A4312-Sigma-Aldrich, USA), and the membranes

were developed with 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitro-blue tetrazolium (NBT) substrate (B1911-Sigma-Aldrich, USA). The molecular mass of the stable vGPD was estimated from MALDI-TOF analysis.

3.3.2.4 Phosphorylation activity assay:

The oxidative phosphorylation activity of vGPD was assayed using micro-plate reader at 25 ºC by monitoring NADH absorbance at 340 nm. The reaction mixture of 100 µL contained 15 mM sodium pyrophosphate buffer (pH 8.9), 30 mM sodium arsenate, 10 mM DTT, 7.5 mM β-nicotinamide adenine dinucleotide and 2 mM DL-glyceraldehyde 3-phosphate solution. All experiments and assays were carried out in triplicate. Initial velocities of the enzymatic reaction were determined by varying concentrations of the substrate D- glyceraldehyde 3-phosphate (G3P; from 0.02 to 4 mM). Kinetic parameters (*Km* and *Kcat*) were determined by fitting initial velocities with the Michaelis–Menten equation and the graph was plotted using Origin 8.0 software. Rabbit muscle GAPDH was also subjected to assay as a positive control.

3.3.2.5 Biophysical characterization of vGPD protein:

To probe the protein folding and thermal stability of vGPD, the intrinsic fluorescence spectrum and thermoflour shift assays were performed. Intrinsic fluorescence was determined on a JASCO spectrofluorometer (FP-8500) at 25 ºC. The intrinsic fluorescence spectrum of the vGPD protein (20 µM) was measured at an excitation wavelength of 280 nm and emission in the wavelength range of 300-400 nm with a 1 cm pathlength cuvette. Three independent spectral measurements were averaged for determining the intrinsic fluorescence spectrum of the vGPD protein.

The thermo-flour shift assay was performed in a 96-well plate containing protein concentration of 2 µM along with freshly diluted SYPRO orange dye (1:1000, v/v) in 25 µL reaction mixture,

in triplicates. Lysozyme was also monitored as an internal control. PCR plates were then sealed and placed into a BioRadCFX96™ Real-time PCR system in FRET mode. Thermal scanning was performed from 20 ºC to 95 ºC, every 30 secs, in 1 ºC increments. Subsequent analysis of the fluorescent data was done using software provided with the CFX96™ Real-time system. Three independent measurements were averaged for determining melting temperature for the tertiary structure of the vGPD protein.

3.3.2.6 Molecular weight determination:

Superdex™ 200GL column (Pharmacia, Sweden) on AKTA purifier (GE Healthcare, USA) was used for determining the molecular weight and oligomeric state of the vGPD protein. The column was pre-calibrated with standard molecular weight markers (Chymotrypsinogen- 25 kDa, Ovalbumin- 44 kDa, Bovine Serum Albumin – 66.5 kDa, Aldolase – 158 kDa, Catalase- 250 kDa) and the standard curve was plotted with elution volume *v/s* the logarithm of molecular weight of markers. Conformational integrity of vGPD and its molecular weight was also checked by electrophoresis on 10% non-denaturing polyacrylamide gel (Native-PAGE). Dynamic Light Scattering (DLS) was employed to characterize the hydrodynamics properties, and the DLS curve was analyzed for determining homogeneity and polydispersity of the recombinant vGPD protein.

3.3.2.7 Protein crystallization:

Initial crystallization attempts were made using the oil drop method in 96-well plate at 295 K using commercially available JCSG+ kit (Qiagen, Germany) and crystallization cryo kit (Sigma-Aldrich, USA). Hanging-drop and sitting-drop methods were used for performing crystallization with a single drop of 2.5 µL protein (10 mg mL^{-1}) and 2.5 µL buffer.

3.3.2.8 Molecular modeling:

Structural models of gGPD and vGPD proteins were constructed by the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/; Yang et al., 2015). Both gGPD (64.2%) and vGPD (62.8%) proteins share high sequence identity with human GAPDH (PDB ID, 3h9eB; Chaikuad et al., 2011) which was identified as the top-hit by I-TASSER protocol. The protein models were validated using the Ramachandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) for the stereochemical nature of the protein for favored and unfavored regions, and the loops in the unfavored region of the protein were further refined using ModLoop server (https://modbase.compbio.ucsf.edu/modloop/; (Fiser et al., 2003). The refined model structures with the best acceptable range of Ramachandran plot were validated by Verify3D (https://servicesn.mbi.ucla.edu/Verify3D/; Eisenberg et al., 1997) and visualized by PyMOL visualization tools.

3.3.2.9 Docking of heptelidic acid:

Heptelidic acid forms a covalent bond with active-site Cys-150 of GAPDH as confirmed by Sakai and group (Sakai et al., 1988). The atomic coordinates of heptelidic acid, an irreversible inhibitor of GAPDH, were obtained from the PubChem compound database (https://www.ncbi.nlm.nih.gov/pccompound/) and the modeled structures of gGPD and vGPD proteins (this work) were used for molecular docking studies. The gGPD and vGPD protein structures and ligand structure (heptelidic acid) were pre-processed using a Protein preparation wizard and a Ligprep module in Schrodinger suite 2019-4 (https://www.schrodinger.com/covdock; Zhu et al., 2014). Both the protein and the ligand structures were minimized using the OPLS3 force field. The CovDock protocol was used to

perform covalent docking between heptelidic acid and active-site of proteins. Covalent docking was carried out by specifying Cys-150 at the active site of the proteins as reactive residue and the epoxide opening reaction was simulated for covalent bond formation. Covalently docked complexes were created using the Glide module in Schrodinger suite 2019-4.

3.3.2.10 DNA binding activity of vGPD protein:

Electrophoretic mobility shift assay (EMSA) was performed for analyzing the interaction between the vGPD protein and the putative promoter region of the *vir4* gene. The upstream region of *vir4* gene (~1600 bp) was used as the target DNA (double-stranded; henceforth named as Pvir4) and ORF region of another gene of ~1700 bp (Trividraft_217322) was used as the non-specific DNA that was used for comparing the specificity of the vGPD protein for Pvir4 dsDNA. Both Pvir4 dsDNA and non-specific dsDNA were PCR amplified using sequence-specific primers (Table 3 of Annexure I) and gel extracted. Approximately, 1 pmol concentration of the purified Pvir4 dsDNA was radiolabeled using 1 µCi [γ-^{32}P] ATP by T4 polynucleotide kinase. Different concentrations of purified recombinant vGPD proteins were added in a 10 µL reaction mixture containing 20 fmol of labeled dsDNA in buffer B (50 mM Tris–HCl (pH 8.0), 75 mM KCl, 5 mM $MgSO_4$, and 0.5 mM DTT). For determination of specificity of vGPD protein for Pvir4 dsDNA, different concentrations of cold non-specific dsDNA were added in the reaction mixture containing a saturating concentration of vGPD protein incubated with the labeled Pvir4 dsDNA. All the reactions were incubated at 37 ºC for 20 minutes and loaded on 6% native-PAGE gels for separation of the bound and unbound fractions. Gels were dried using a vacuum gel dryer and were exposed to X-ray films overnight. Autoradiograms were developed and documented for image analysis. ImageJ 2.0 software was used for calculation of band intensity

of bound and unbound fractions from each autoradiogram. The % bound fraction of Pvir4 dsDNA against increasing concentration of recombinant vGPD protein was plotted using GraphPad Prism 7. The theoretical fit of a one-site specific binding kinetics model in GraphPad Prism 7 was used to calculate dissociation constant (*Kd*).

3.3.3 Results

3.3.3.1 Cloning, heterologous expression, and purification of the *T. virens* vGPD protein:

Initially, the vGPD gene was cloned and expressed in pET-28a vector and BL21 Star (DE3) host cells. Even though we were able to express the protein in the pET-28a expression vector, the recombinant protein was insoluble and was found in the inclusion bodies. Several strategies were employed to improve the solubility and stability of the recombinant protein, for example by varying IPTG concentration, temperature, media, different expression systems like BL21 (DE3), Rosetta (DE3), BL21-pLys (DE3), purification under denaturation conditions and addition of detergents. However, either the protein expression was not seen or the protein was found in inclusion bodies. We also tried purification under denaturing conditions of 8M urea, but the protein was partially refolded or aggregated. Other approaches like the expression in *K. lactis* system and the native host *Trichoderma virens* were also employed but protein either did not express or very low-levels of protein expression were observed.

The susceptibility of the free thiol group at the active site of the protein could account for the instability of the protein followed by aggregation (Nakajima et al., 2007). Thus, we employed the pNH-TrxT vector which has a thioredoxin tag (~12 kDa) that helps in maintaining a reduced environment in the bacterial cytoplasm and provides stability to redox sensitive-proteins. Full-

length vGPD was cloned into pNH-TrxT plasmid via ligation-independent cloning (Fig 3.3.1) and then transformed into *E. coli* BL-21 expression host for protein expression and purification. The best condition for the recombinant protein expression was found to be induction with 0.5 mM IPTG for 16 hrs at 18 ºC. The recombinant protein with N-terminal poly-His tag was purified to near homogeneity by immobilized metal affinity chromatography (IMAC). All the steps were performed at 4 ºC in the presence of 1 mM TCEP, a reducing agent, to maintain a reduced environment. Western blot analysis with anti-His antibody confirmed the purification of recombinant protein containing both poly-His tag and thioredoxin tag (Fig 3.3.2). MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) analysis of purified vGPD showed the mass of the protein to be 50.2 kDa corresponding to the estimated mass of 50 kDa of the recombinant construct with poly-His tag (~1.8 kDa) and thioredoxin tag (12 kDa).

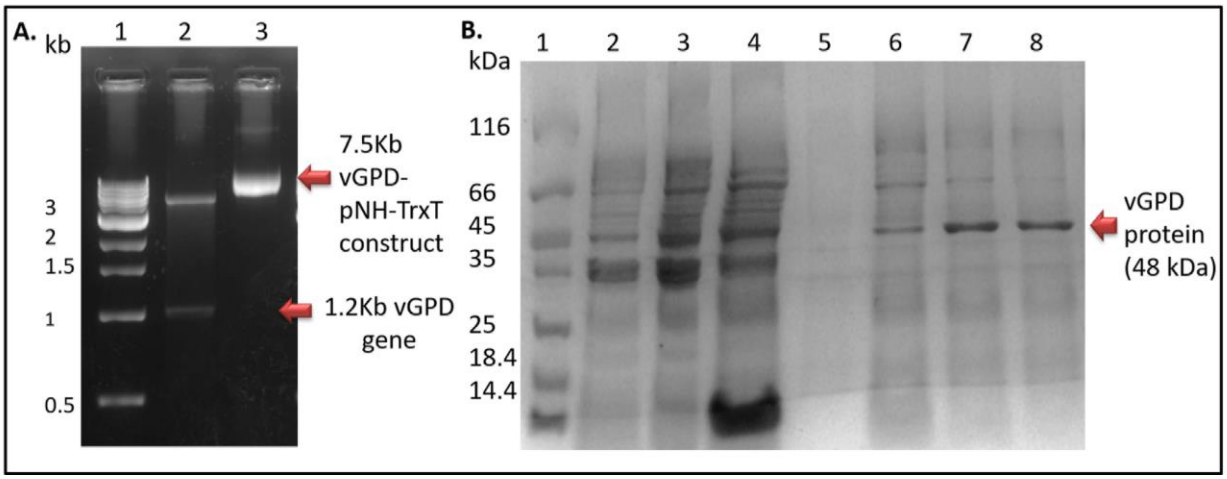


Figure 3.3.1. Cloning, expression and purification of vGPD protein.

A). Cloning of vGPD gene in pNH-TrxT plasmid via ligation-independent cloning **B).** Over expression of vGPD protein in *E. coli* BL-21 expression cells: Lane 1: Molecular weight marker, Lane 2 Uninduced BL21 star cells transformed with vGPD-pNH-TrxT plasmid, Lane 3: Induced BL21 star cells (0.5 mM IPTG), Lane 4: Unbound fraction from the Ni-NTA purification column, Lane 5: Wash through from the Ni-NTA column, Lane 6-10: Eluted vGPD protein with increasing concentration of (200, 250 and 300 mM) imidazole.

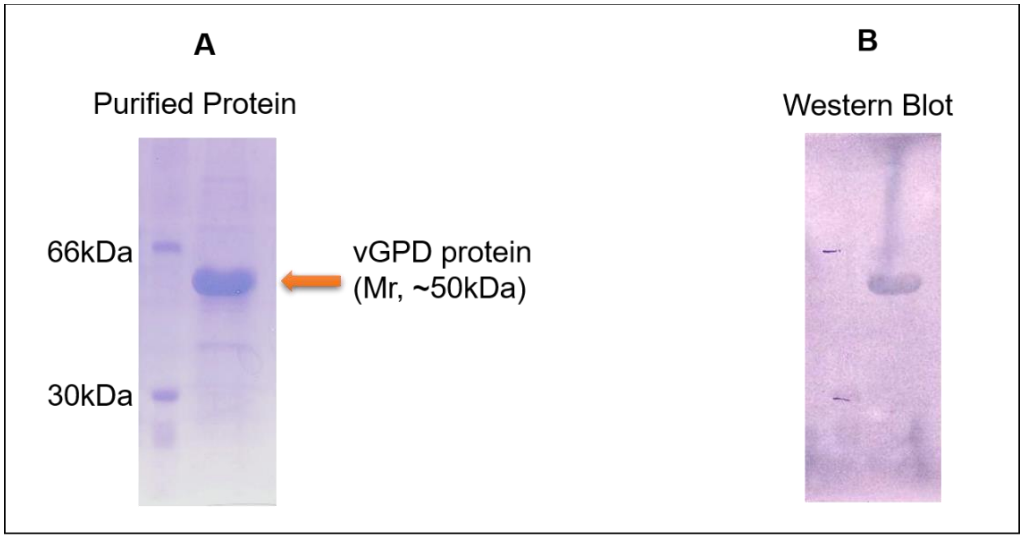


Figure 3.3.2. Purification of vGPD by affinity chromatography and Western blot with Anti-His tag antibody. A). 15% SDS-PAGE gel showing purified recombinant His6-vGPD protein (~50 kDa) using IMAC. **B).** Western blot analysis of the vGPD protein using monoclonal anti His-tag antibody.

3.3.3.2 vGPD protein exists as a tetramer in solution:

Purified vGPD protein was analyzed on a Superdex™ 200GL column. The protein eluted as a single major peak, coinciding with the UV absorbance peak at a molecular weight of 200 kDa. Rabbit muscle GAPDH was also analyzed on the same column and it eluted at the same elution volume (Fig 3.3.3A). We predict from this analysis that the vGPD protein has similar subunit interaction like the rabbit muscle GAPDH protein and thus vGPD protein exists as a tetramer in solution. The vGPD protein was also electrophoresed on 10% native-PAGE which showed that the protein migrated as a single band parallel to 200 kDa protein band (Fig 3.3.3B). The DLS profile with a narrow particle size distribution curve confirmed the homogeneity of the vGPD protein indicating the monodisperse nature of the oligomer. The hydrodynamic diameter was estimated to be 12.4 ± 2.5 nm. The polydispersity percentage of recombinant vGPD protein was

estimated to be 17.3 (Fig 3.3.3C). Taken together, with the subunit mass of 50.2 kDa, as observed in MALDI-TOF analysis, it can be inferred that vGPD exists as a stable homo-tetramer (oligomer) in solution, like rabbit muscle GAPDH. Tetramer as a biological significant oligomerization state have also been reported for 197 of the available 316 GAPDH entries in the Protein Data Bank (PDB)

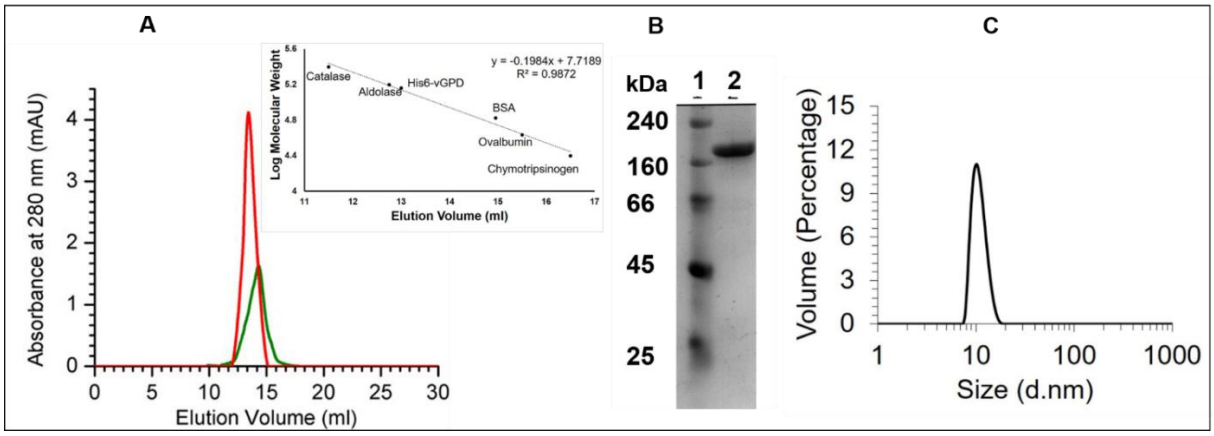


Figure 3.3.3 Determination of molecular weight and oligomeric state of the purified recombinant vGPD protein A). Gel-filtration column chromatography of vGPD protein (red) compared with rabbit muscle GAPDH protein (green) elution profile. Molecular weight marker proteins used as standard were chymotrypsinogen (25 kDa), BSA (66 kDa), ovalbumin (45 kDa), aldolase (160 kDa) and catalase (240 kDa). **B).** 10% Native page. Lane 1: protein markers for native-PAGE Lane 2: purified vGPD protein **C)**. Dynamic light scattering (DLS) curve showing the hydrodynamic radius of 6.2 ± 2.5 nm and the poly-dispersity percentage of 17.3% of recombinant vGPD protein.

3.3.3.3 Biophysical characterization of vGPD protein:

The thermal denaturation of vGPD was scanned from 20 °C to 95 °C. The melting temperature for the tertiary structure (*T*m) of vGPD was determined to be 56.5 °C (Fig 3.3.4A and 3.3.4B), which is following the secondary structure denaturation melting temperature of other GAPDH

proteins from different organisms (Bell et al., 2014). The *T*m of 69.5 °C for lysozyme, which was used as an internal control in this experiment, matched with the reported literature value. Intrinsic fluorescence spectrum analysis yielded information on the folded state of the vGPD protein. The spectrum result of vGPD protein shows emission at 330 nm which is characteristic of buried tryptophan residues (Fig 3.3.4C). Fluorescence spectrum of the 8 M urea treated denatured protein was found to show emission at 352 nm due to quenching of the fluorescence intensity by exposed tryptophan residues in the denatured protein (data not shown). The thermal denaturation and fluorescence studies thus confirm that recombinant vGPD protein adopts a globular fold similar to that of other known GAPDH proteins.

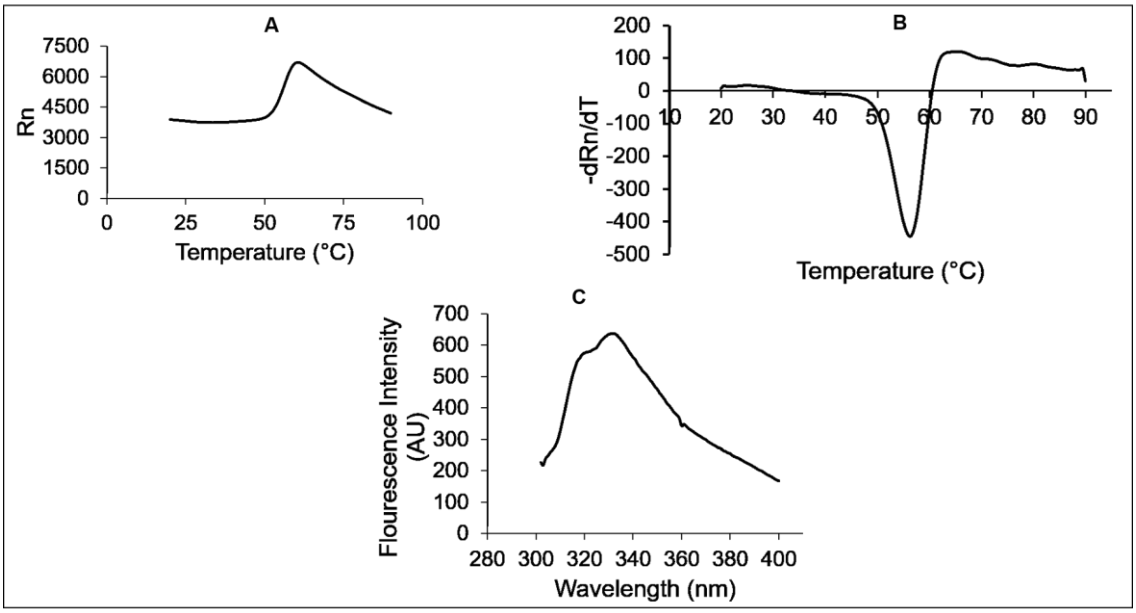


Figure 3.3.4 Biophysical characterization of the recombinant vGPD protein. A). Thermofluor-shift melt curve. **B).** The derivative plot of the thermofluor-shift melt curve **C).** Intrinsic Trp fluorescence spectrum. The vGPD protein (0.3 mg mL^{-1}) was dissolved in 25 mM Tris-HCl buffer (pH 7.6) and excitation wavelength was 280 nm. Intrinsic Trp fluorescence spectra were obtained by subtracting the spectra of the buffer. Three independent measurements were performed during the thermofluor-shift assay.

3.3.3.4 Kinetic analysis of the vGPD protein:

Values of the Michaelis-Menten parameters, were calculated for the oxidation of G3P by recombinant vGPD protein. Michaelis-Menten plot was used for determining apparent values of *K*m, *V*max, and *K*cat for vGPD protein; they were 0.38 mM, 3.19 µmol min^{-1} mg^{-1} and 2.55 sec^{-1}, respectively (Fig. 3.3.5). The *K*m for rabbit muscle GAPDH was determined to be 0.26 mM. A previous study reported *K*m values of 0.54 mM and 0.33 mM for vGPD and gGPD proteins (Sakai et al., 1990).

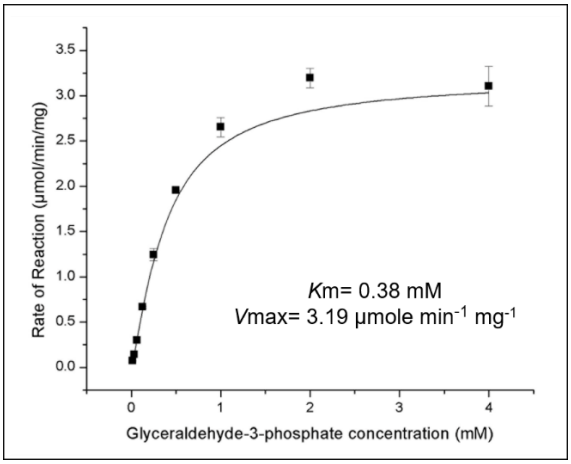


Fig 3.3.5 Kinetics of vGPD enzyme activity. The initial velocity curve was fitted to the Michaelis-Menten equation using Origin lab 8.0 and the plot was used to determine Km and Vmax of vGPD protein. The reactions were monitored by estimating NADH production with increasing concentration of substrate (0.01 mM to 4 mM) at 25 ºC, as described in the Material and Methods. Apparent Km, Vmax, and Kcat values for glyceraldehyde-3-phosphate were calculated as 0.38 mM, 3.19 µmol min^{-1} mg^{-1}, and 2.55 sec^{-1} respectively.

3.3.3.5 Crystallization of vGPD protein:

Crystals of vGPD were obtained by oil-drop method under two promising conditions containing [Tris Na-citrate (0.1 M, pH 5.5), and PEG 2000 (20%)] and [calcium acetate (0.16 M), Na-

cacodylate (0.08 M, pH 6.5), PEG 8000 (15%) and glycerol (20%)] as precipitants. Hanging-drop and Sitting-drop methods were also employed for crystallization under similar conditions. However, very small crystals of dimensions approx. 5 microns were obtained which were not suitable for diffraction experiments (Fig. 3.3.6). Further optimization of buffer and PEG concentration was also attempted to grow crystals of a size suitable for single-crystal diffraction analysis.

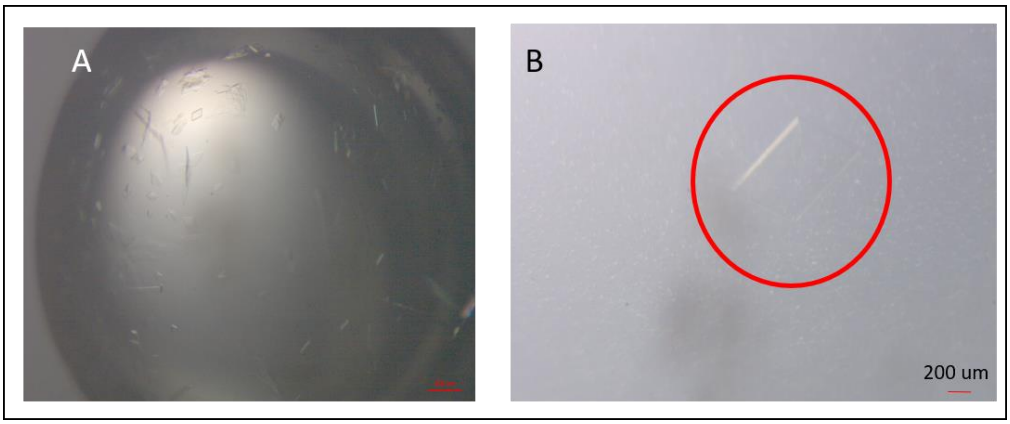


Figure 3.3.6 Crystallization of vGPD protein. Protein was crystallized using the sitting-drop vapor-diffusion method and commercial screens. Initial small crystals **A).** obtained were further optimized **B).** using optimized condition containing calcium Acetate (0.16 M), Na-cacodylate (0.08M, pH 6.5), PEG 8000 (15%) and glycerol (20%).

3.3.3.6 Molecular modeling of gGPD (glycolytic GAPDH) and vGPD proteins and docking with heptelidic acid.

The gGPD and vGPD protein model structures were derived using the GAPDH structure of *Homo sapiens* (PDB ID:3h9eB) as a template on the I-TASSER server. The models were validated and refined to best-fit by the on-line RAMPAGE (Ramachandran plot examination) tool. Ramachandran plot for gGPD model displayed that 94.9% residues are in the most favored

region, 4.5% in the additional allowed region and only 0.6% residues in the outlier region. Similarly, 95.2% residues in the most favored region, 3.9% in the additional allowed region and only 0.9% residues in the outlier region of the Ramachandran plot were observed for the model for vGPD protein. The model structures for gGPD and vGPD were also assessed using Verify3D. The Verfiy3D analysis showed that 96.45% residues of the gGPD model and 98.81% residues of the vGPD structure have averaged 3D-1D score >=0.2. These parameters showed good quality of vGPD and gGPD model structures.

The structural alignment of the selected models with each other and with the template in PyMol revealed that both the protein structures are highly conserved. The secondary structures generated with ESpript further revealed that gGPD and vGPD protein sequences from *Aspergillus* and *Trichoderma* share more than 80% similarity (~73% identity) with essentially conserved NAD+ cofactor- and substrate- binding sites (Fig 3.3.7 and Fig 3.3.8). Notably, the two proteins differ due to the presence of an INDEL in the loop-region placed at the entrance to active-site (Fig 3.3.7 and Fig 3.3.8).

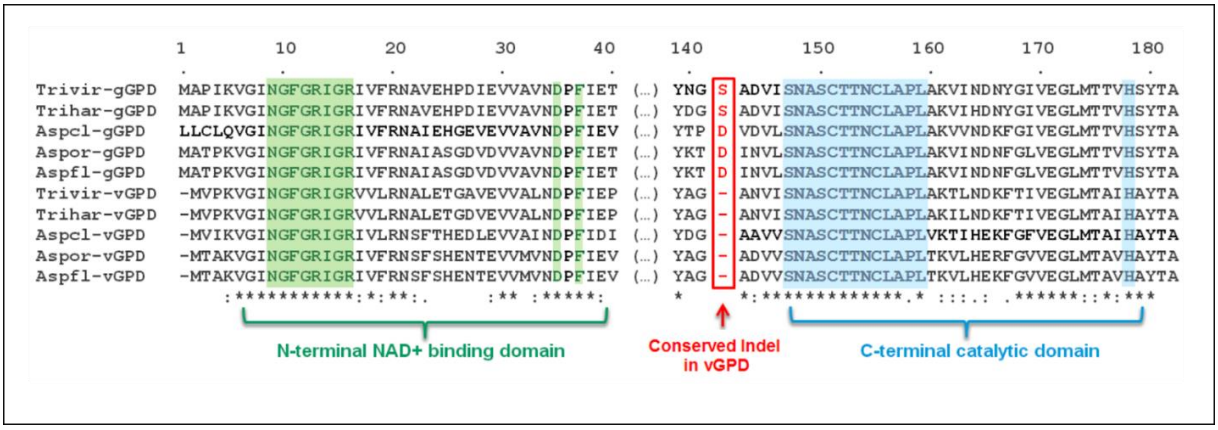


Figure 3.3.7: **Partial sequence alignment of GAPDH isoforms from *Trichoderma* and *Aspergillus* species**

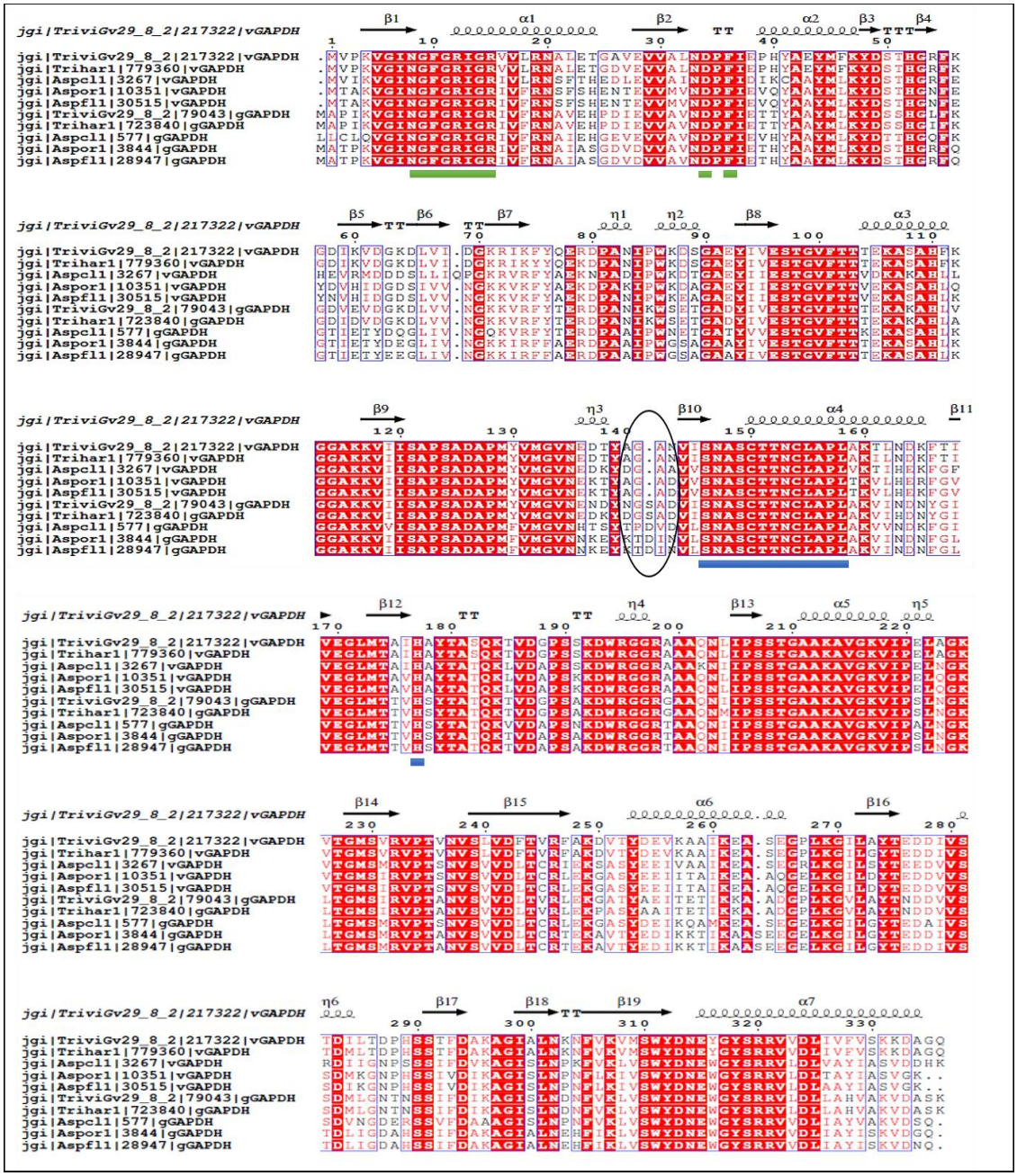


Figure 3.3.8: Multiple sequence alignment of GAPDH orthologs. gGPD and vGPD protein sequences from *Aspergillus* and *Trichoderma* species share more than 80% similarity. The secondary structure (β1, α1,…) is shown. Conserved indel (Ser-143) of vGPD orthologs is marked with a circle around it. Active-site scaffold (active site) and co-factor binding residues (cofactor) are identified with blue and green color bars respectively. The figure was prepared with Espript (http://espript.ibcp.fr).

This deletion-site (Ser-143 in gGPD) precedes the active-site scaffold of the GAPDH proteins and is situated on the surface loop in gGPD. The conserved indel could possibly introduce several modifications in the vGPD protein, like interaction with other proteins, change in subcellular location and post-translational modifications, and might add to the functional diversity of the protein (Tristan et al., 2011).

We performed covalent docking using CovDock (https://www.schrodinger.com/covdock) and constraint that the irreversible inhibitor, heptelidic acid, formed a covalent bond with the active-site cysteine residue, Cys-150 and Cys-152 of vGPD and gGPD proteins, respectively. This covalent binding mode of heptelidic acid in gGPD protein at the active site showed no interaction with the serine residue which is deleted in the vGPD protein. MM-GBSA binding energies were determined for both the covalently docked structures and heptelidic acid, showed marginally higher affinity for gGPD protein as compared to vGPD protein with binding energy values of -49.63 and -43.67, respectively (Fig 3.3.9). It is also likely, however, that the deletion of Ser-143 in vGPD isoform alters the conformation of the surface loop and restricts the approach of HA to the active site (Fig 3.3.10).

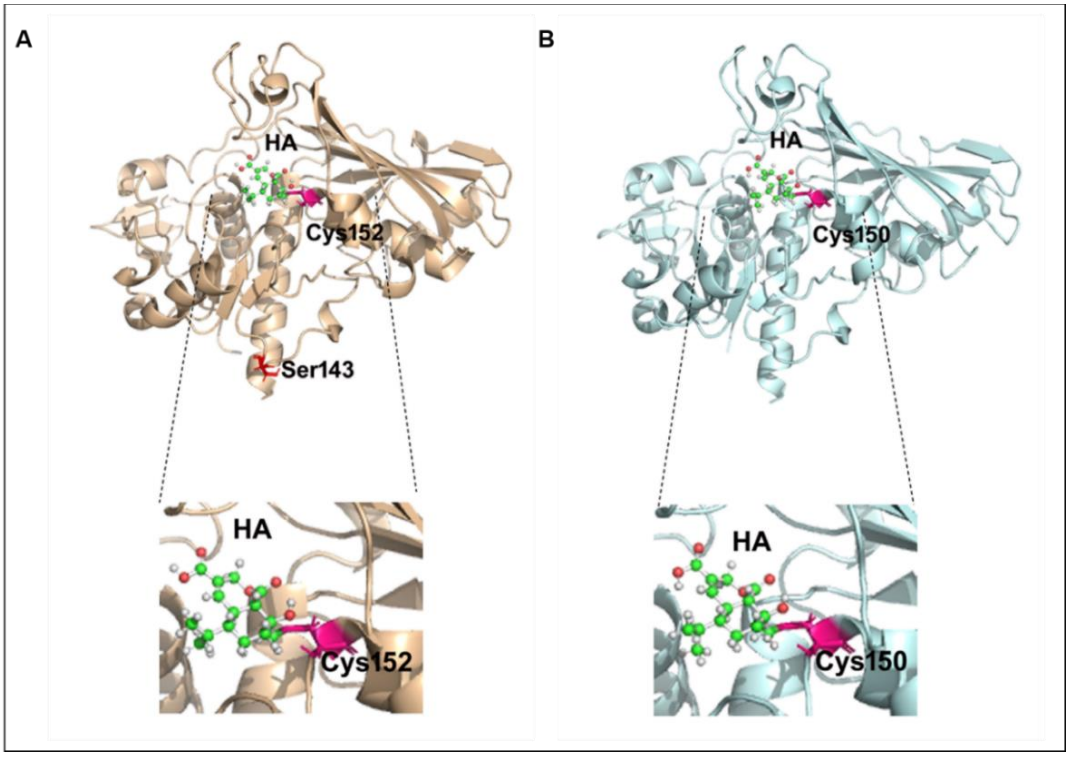


Figure 3.3.9 Molecular modeling of the two isoforms of GPDs (gGPD and vGPD) and docking of heptelidic acid ligand. A). gGPD-HA covalent docked structure. Residues Cys-152 and Ser-143 are labelled. **B**). vGPD-HA covalent docked structure with labelled Cys-150. Heptelidic acid (HA) is shown in ball-and-stick representation. The active site Cysteine residue and Serine residue are shown in red colour.

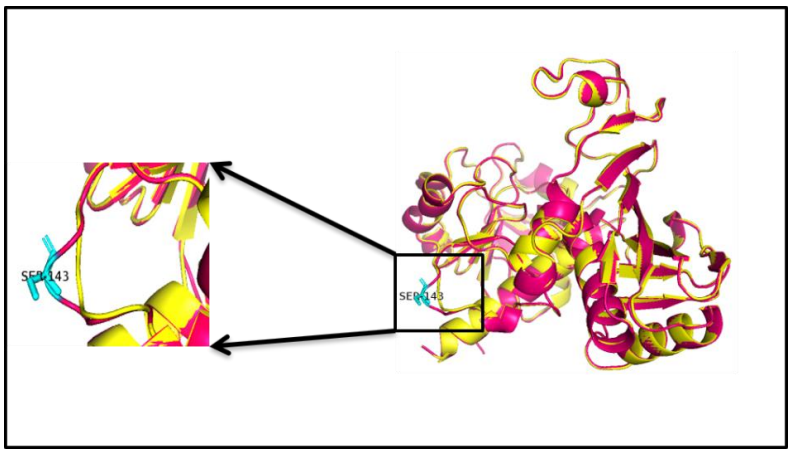


Figure 3.3.10 Structural comparison of vGPD and gGPD. Structural overlay of monomeric gGPD (pink) and vGPD (yellow). Deleted Ser143 forming a signature indel in vGPD protein is highlighted in cyan color.

3.3.3.7 vGPD protein binds non-specifically to the upstream region of *vir*4 gene:

GAPDH has been addressed as a nucleic acid binding protein in several studies. In section 3.1, knockout studies also indicated differential effect on genes of the 'vir' cluster, suggesting that vGPD may be involved in the biosynthesis of volatile sesquiterpenes in *T. virens*. The recombinant vGPD protein was purified to homogeneity from *E. coli* strain for performing electrophoretic mobility shift assays (EMSA) with [γ-^{32}P] ATP labeled Pvir4 dsDNA (~20 fmol). Binding assay with *vir4* gene was performed as it codes for the putative terpene cyclase, which is a core enzyme in the *vir* cluster, and its deletion also abolished volatile sesquiterpene compounds production like vGPD deletion mutant. The specificity of vGPD protein for the radiolabeled Pvir4 dsDNA was examined by adding an increasing amount of (1-12 molar excess) cold non-specific dsDNA (NS-DNA) to the pre-incubated reaction mixture containing saturating concentration of the vGPD protein and radiolabeled Pvir4 dsDNA. This competition assay estimated that vGPD protein and Pvir4 dsDNA binding is nonspecific with the dissociation constant (*K*d) of 3.4 ± 0.21 µM (Fig 3.3.11).

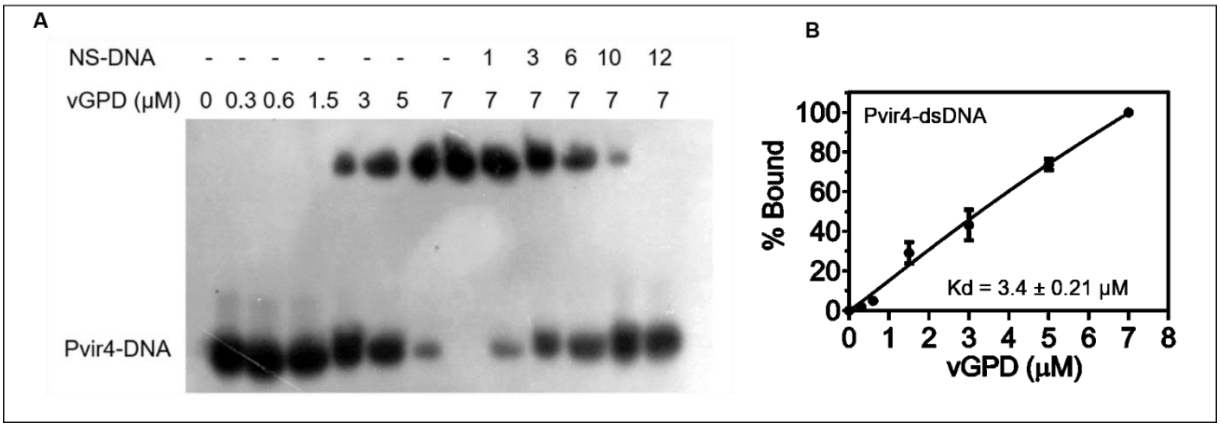


Figure 3.3.11 Interaction of vGPD protein with the promoter region of the *vir4* gene (Pvir4). A). The DNA-binding activity of vGPD protein with the upstream region of the *vir*4

gene in *Trichoderma virens* was determined using EMSA. PCR-amplified and gel purified Pvir4 DNA (20 fmol; ~1.6 kbp) was radiolabeled with [γ-^{32}P] ATP and mixed with varying concentration to vGPD (0-7 µM). The mixtures were resolved on 6% native-PAGE. The competition assay performed by adding an increasing amount of (1-12 molar excess) cold non-specific DNA to the pre-incubated reaction mixture containing saturating concentration of the vGPD protein and radiolabeled Pvir4-DNA, suggests non-specific DNA-binding of vGPD. **B**). Measurement of the equilibrium dissociation constant (*K*d) by gel mobility shift assay.

3.3.4 Discussion:

GAPDH is classically known as a housekeeping protein catalyzing the sixth step in glycolysis. However, research suggests that the same protein has many auxiliary functions, like DNA replication and repair, gene expression, and signal transduction (Sirover, 1999, and 2011; Tristan, 2011). An additional copy of GAPDH was earlier found in *T. virens* (Sakai et al., 1990). This GAPDH, which is tolerant of heptelidic acid, was proposed to takeover glycolysis function when this metabolite is produced by *Trichoderma* (Sakai et al., 1990). Maurer et al. (1983) found a copy of GAPDH to be associated with a secondary metabolism-related gene cluster responsible for the biosynthesis of pentalenolactone in *Streptomyces arenae*. Mukherjee et al. (2006) documented the association of additional copies of GAPDH with secondary metabolism-related gene clusters in *T. virens* and *Aspergillus* spp. This GAPDH is thus an important subject of study because of two reasons: 1. How and why this GAPDH (vGPD) is involved in secondary metabolism and 2. What makes it tolerant of heptelidic acid that irreversibly inhibits the glycolytic GAPDH. One of the hallmarks of cancer cells is metabolic reprogramming. Due to the Warburg effect, cancer cells have metabolic phenotypes like increased glycolysis (Zhang et al., 2015). GAPDH proved to be the most preferred candidate for targeting glucose metabolism because it performs the important rate-limiting step of glycolysis and produces the first essential

redox molecule i.e. NADH (Ganapathy-Kanniappan and Geschwind, 2013). Secondly, metabolomics demonstrated that aerobic glycolysis in cancer cells flux through GAPDH and thus is an important regulatory enzyme in promoting the Warburg effect (Shestov *et al.*, 2014). Lastly, being moonlighting protein, GAPDH is also associated with many cancer progressing pathways like protection against caspase-independent cell death, and increased proliferative index and cell cycle progression (Krasnov et al., 2013; Zhang et al., 2015). Anti-GAPDH therapeutic approach with natural inhibitor, heptelidic acid also known as koningic acid, for promoting anticancer effect has been recently demonstrated (Liberti et al., 2017; Ganapathy-Kanniappan and Geschwind, 2018). Even though the presence of HA-insensitive GAPDH in some fungal genomes has been known for decades, its mechanism is not known yet. The objective of the present study was to study in detail the HA-tolerant GAPDH at the protein level. However, several attempts (including different hosts, different vector, expression in yeast, and expression in a native host) did not yield a significant amount of protein in the soluble form. This is interesting, as several authors have earlier reported purification and crystallization of the glycolytic GAPDH from many organisms (Robien et al., 2006; Liu et al., 2012; Tien et al., 2012; White et al., 2015). We finally succeeded in getting the protein in soluble form by using the pNH-TrxT vector in *E. coli* expression system. The obtained value for *K*m of D-G3P for vGPD protein was approximately similar to the *K*m value reported by Sakai et al (Sakai et al., 1990) for the HA-tolerant GAPDH purified from *T. virens*. The structural integrity of the recombinant vGPD protein has been confirmed by the thermo-flour shift assay which showed the apparent melting temperature of vGPD protein to be 56.5 °C and from the intrinsic fluorescence analysis, which showed emission at 330 nm that is typical of partially buried tryptophan residues. Further,

the recombinant protein displayed a single sharp peak in DLS analysis and oligomeric status of a homo-tetramer similar to other characterized GAPDH enzymes.

GAPDH has been widely addressed as a nucleic acid-binding protein which is one of the moonlighting functions performed by it. The first study on GAPDH interaction with single-stranded DNA was reported in 1977 by Perucho and colleagues (1977). Subsequently, many reports were published demonstrating DNA-mediated functions of GAPDH in the regulation of mRNAs translation, binding with viral RNAs, nuclear tRNAs transport, telomere protection, gene regulation, etc. Moreover, GAPDH is also associated with functions like apoptosis, DNA integrity maintenance and DNA repair mechanism which demands nuclear translocation. Nuclear functions of GAPDH are due to its oligomeric status, PTMs, oxidative stress and status of active-site -SH groups. However, in the present study, we observed that vGPD protein can have nonspecific interaction with the upstream region of putative *vir4* ORF (Pvir4 dsDNA) of approximately 1.6 Kbp in *T. virens*. This study suggests that vGPD protein is not associated with the transcriptional regulation of the "*vir*" cluster. However, binding to this region mediated by other proteins cannot be ruled out as *K*d is high. It is known that different proteins, like Siah1 and androgen receptor, are involved in forming a complex with GAPDH and translocation of it inside the nucleus (Hara et al., 2005; Harada et al., 2007).

In silico analysis sheds some light on the possible crystal structure and differential affinity of the two GAPDHs towards HA. Both the *T. virens* proteins are expected to have similar tertiary structure and active-site scaffold, though from the partial amino acid sequences different molecular structures were advocated previously (Sakai et al., 1990). A notable difference in the two models is the location of conserved indel (residue Ser-143 of gGPD deleted in vGPD

orthologs). This site is located in the surface loop that could alter interaction of the two isoforms with other proteins. It has earlier been demonstrated that isoform-I of GAPDH in *T. virens* (misidentified as *T. koningii*) is inhibited to the extent of 50% by 1 mM heptedilic acid but it is not affected at 0.1 mM of HA, while 50% inhibition was observed for isoform-II at 0.01 mM (Sakai et al., 1990). The covalent-docking of HA with the gGPD protein showed higher affinity in comparison to the vGPD protein (binding energy values of -49.63 vs -43.67 kcal/mol, respectively). The lower affinity towards vGPD compared to gGPD may be responsible for the observed inhibition data and also indicates that vGPD is responsible for the glycolysis in *Trcihoderma* when heptedilic acid is produced, as suggested earlier (Sakai et al., 1990).

Trichoderma virens is a beneficial fungus and is appreciated for its role as a biocontrol agent, beneficial interactions with plants, antagonistic interactions with pathogenic microorganisms, production of important hydrolytic enzymes, and many others including biosynthesis of an array of secondary metabolites - both volatile and non-volatile (Mukherjee et al., 2013b; Vinale et al., 2008). Secondary metabolites are produced by core enzymes like, non-ribosomal peptide synthetases, polyketide synthases, terpene cyclases, etc. and other "decorating/tailoring" enzymes like cytochrome P450, dehydrogenases, transferases, etc., that provide structural modifications to the secondary metabolites (Zeilinger et al., 2016). These genes along with regulatory genes, like transcription factors and transporters, are present in the genome of the producing organism, more often in a clustered form known as secondary metabolite biosynthetic gene cluster. In *T. virens*, the 'vir' cluster is the first secondary metabolite biosynthetic gene cluster identified which was later on reported to be involved in the biosynthesis of volatile sesquiterpene compounds (Mukherjee et al., 2006; Mukherjee et al., 2012b; Crutcher et al., 2013). A GAPDH is consistently associated with the "vir" cluster in *Trichoderma* and *Aspergillus* spp. Moonlighting functions of GAPDH, the sixth enzyme in the glycolytic pathway, is considered important in both prokaryotic and eukaryotic systems including animals and plants (Tristan, 2012; Zaffagnini et al., 2013). There is no report for the non-glycolytic role of GAPDH in any filamentous fungus. *Trichoderma virens* genome harbors genes for two GAPDHs named here as gGPD that is the primary enzyme involved in glycolysis and vGPD that performs glycolytic function under heptelidic acid (an irreversible inhibitor of GAPDH) biosynthesis (Sakai et.al., 1990). Analysis of the whole-genome sequence of *T. virens* revealed that the "vir" cluster harbors 8 genes, encoding four cytochrome P450s, one monooxygenase, one MFS, and a GAPDH (vGPD). Furthermore, the vGPD is consistently present in the genome of few strains of *Aspergillus* spp. those harbour

the "vir" cluster. In the present study, using the gene knock-out approach, we established that vGPD is involved in the biosynthesis of volatile sesquiterpene metabolites and is functionally associated with the *vir* cluster. Deletion of the gene for vGPD abolished the production of volatile sesquiterpenes such as germacrene D, β-caryophyllene, α-amorphene, germacrene A, β-selene, aromadendrene, ledene, β-elemene, etc. The non-volatile profile of the vGPD mutant was also determined with TLC and LC-MS, which showed that vGPD deletion mutant is downregulated in the biosynthesis of non-volatile metabolites including viridin, viridiol, gliovirin, and heptelidic acid. This is the first functional analysis of a GAPDH (vGPD) isoform in any filamentous fungus, and provides confirmation of the involvement of vGPD in volatile and non-volatile secondary metabolites biosynthesis. The vGPD mutants were also characterized for other phenotypes. The mutants were similar to wild-type with respect to radial growth and sporulation.

The *vir* cluster was discovered by using suppression subtractive hybridization (SSH) technique where cDNA of a radiation-induced, secondary metabolism-downregulated mutant (M7) was subtracted from the wild type. Three genes of the *vir* cluster, i.e., *vir2*, *vir3* and *vir4*, were under-expressed in the M7 mutant (Mukherjee et al., 2006). In the present study, however, *vir3* and *vir4* were upregulated in vGPD mutant and other five genes were downregulated. Here, we report that M7 is also downregulated in the biosynthesis of volatile sesquiterpenes, much like that of our vGPD mutant. Due to this overlap, we further studied M7 at the genome and transcriptome level to gain further insight into the regulation of the *vir* cluster. In addition to the downregulation of secondary metabolism (volatile and non-volatile), we also found that M7 lacks hydrophobicity and mycoparasitism. RNAseq analysis supported the M7 phenotype, as many of the important secondary metabolite biosynthetic gene clusters were downregulated in M7 mutant including viridin biosynthetic gene cluster, gliovirin

biosynthetic gene cluster and *vir* cluster. The vGPD gene which is part of the vir cluster is also downregulated in the M7 mutant and thus M7 and vGPD deletion mutants show similarity in the metabolic profile.

As many as 463 genes including those involved in morphogenesis and biocontrol properties (including secondary metabolism and mycoparasitism) are downregulated (including genes with no expression in M7) in the mutant M7, of which 11 genes are also deleted in the genome of the mutant M7. The transcriptome and whole-genome sequencing study of the M7 mutant provide information about putative gens involved in globally affecting the biology of the *Trichoderma* spp. varying from hydrophobicity, secondary metabolites biosynthesis and mycoparasitism. Among 71 genes deleted in the M7 genome are genes for eight transcription factors that could be involved in global regulation of secondary metabolites biosynthesis. Apart from secondary metabolism, these transcription factors could also be involved in the regulation of morphogenesis and biocontrol properties in *T. virens*. Thus, our study identified novel candidate genes for regulation of secondary metabolites biosynthesis through transcriptome analysis and whole-genome sequencing of M7 mutant. Further studies shall be needed to ascertain the functions of these transcription factors.

Since this is the first report on the functional analysis of the second GAPDH in any filamentous fungus, and for the first time we assigned secondary-metabolism related functionalities of this GAPDH, we wanted to study in detail the structural characteristics of this protein. Initial exhaustive attempts for cloning and purification of the recombinant vGPD protein involved the pET28 cloning vector and BL21 expression host and were not completely successful. Different methods were performed to improve the solubility and stability of the recombinant vGDP protein such as different growth temperatures (30 ºC, 20 ºC and 15 ºC), different IPTG concentrations (varying from 0.02 mM-1 mM), different host strains (BL21

(DE3) codon plus cells, BL21(DE3)pLysS cells, Rosetta (DE3) cells). The expressed recombinant vGPD protein was found in inclusion bodies. Different modifications in growth conditions were also employed to make the protein in soluble form. Detergents like Sodium lauroyl sarcosinate, tritonX 100 and NP40 were also used in lysis buffer for protein purification. Purification under denaturing conditions (8M urea) was also performed but the protein was partially refolded. We also tried expression of truncated protein having only the N-terminal region of the vGPD protein in pET28a vector, but the expressed protein was not in soluble form. Another cloning vector independent of lac promoter like in pET28a vector was used i.e pBAD-HisB vector because it possesses a weak promoter (arabinose) which would avoid inclusion bodies formation due to overexpression of the protein. Although recombinant vGPD protein was in soluble form in pBAD-HisB vector but was very unstable after purification. Next, we attempted to express and purify the vGPD protein in the eukaryotic system to take care of the post-translational modifications. Recombinant vGPD protein was successfully cloned in the pKLAC2 vector and transformed in the *Kluyveromyces lactis* (yeast system), but the expression of the vGPD protein was not achieved. Expression of the recombinant vGPD protein in the native host i.e. *T. virens* was also tried, but a very small amount of protein was expressed and recovered after purification which was not enough to perform downstream experiments.

The major reason for the aggregation and instability of the purified recombinant protein was the susceptibility of the vGPD protein to the oxidizing environment. To maintain the stability of the protein, we cloned vGPD protein in the pNH-TrxT vector as this vector possesses a thioredoxin tag which helps in maintaining a reduced environment as well as provides solubility to the protein and expressed vGPD protein in BL21 Star (DE3). We purified and showed that biophysical and biochemical characteristics (oligomeric status, Tm,

Michaelis-menten parameters etc) of the vGPD protein are similar to the glycolytic GAPDH (gGPD). We hypothesized the possibility of vGPD protein as a transcription factor in the *vir* cluster as no typical transcription factor is associated with the *vir* cluster. We monitored the binding affinity of purified vGPD protein to the upstream region of the terpene cyclase (*vir4* gene). The results indicated that the vGPD protein shows non-specific interaction with the upstream region of the *vir4* gene and thus is not a transcription factor in the *vir* cluster. Another possible function of vGPD protein in the *vir* cluster could be dehydrogenation which requires experimental confirmation. The crystallization attempt provided very small size crystals which were not of a suitable for diffraction. Besides, we also wanted to gain insight into the tolerance of vGPD to heptelidic acid, a known GAPDH inhibitor. Phylogenetic analysis and homology modeling of the two GAPDHs (vGPD and gGPD) revealed that both the proteins are very similar in primary sequence as well as the conserved 3D structure which includes the active-site scaffold and NAD^+ cofactor binding site. The major difference is an indel located at the entrance to the active site of the vGPD protein. We further performed *in silico* analysis of both the proteins to understand the difference in the affinity of both the GAPDH types towards heptelidic acid binding. Based on homology modeling, covalent docking and binding free-energy estimations, we concluded that vGPD protein has a lower affinity for HA as compared to the glycolytic gGPD protein. The in-silico analysis is supported by the experimental data reported earlier.

Appendix

Annexure I

Table 1. List of plasmids used or derived in the present study

S.no	Name of plasmids	Size (bp)	Marker	Source
1.	pNH-TrxT	5671	Kanamycin	Addgene plasmid # 26106
2.	pNH-TrxT-vGPD	6771	Kanamycin	Present study
3.	pTrp-ATBS	5161	Ampicillin	Lab stock
4.	pTrp-ATBS-LF	6861	Ampicillin	Present study
5.	pTrp-ATBS-RF	7261	Ampicillin	Present study

Table 2. List of bacterial strains used during the present study

S.no	Strain	Genotype	Source
1.	*E. coli* DH5α	*F- / endA1 hsdR17 glnV44 thi-1 recA1 gyrA relA* Δ *(lacIZYA-argF) U169 deoR (Φ80dlac*Δ *(lacZ)M15)*	Lab stock; Cloning host
2.	XL10-Gold	*TetrΔ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZΔM15Tn10 (Tetr) Amy Camr]*	Stratagene; Cloning host
3.	BL21Star (DE3)	*F-ompT hsdSB (rB-mB-) gal dcm rne131(DE3)*	Novagen; expression host

Table 3. List of Primers used in the present thesis:

Primers	Sequence (5'→3')	Description
M1F	GTAAAACGACGGCCAGT	Forward and reverse primers for amplification of hygromycin B resistance gene cassette from pATBS vector.
M1R	CAGGAAACAGCTATGAC	
NLC8	CGTTGCAAGACCTGCCTGAA	Forward and reverse internal primers for amplification of hygromycin B resistance gene cassette from pATBS vector.
NLC7	GGATGCCTCCGCTCGAAGTA	
vGLFkpn	AAGGTACCGATCGACCAATCTACCTCAGAAC	Forward and reverse primers for amplification of 5' flanking region of vGPD gene.
vGLFxho	CGAAAAGCATCTTCTCTTTTTGGCCTCGAGATAT	
vGRFxba	ATCTAGAGAGTAGTTCACGAGTGTTAGTGCT	Forward and reverse primers for amplification of 3' flanking region of vGPD gene.
vGRFsac	TGCGTGTTCGTCAAATACAACCGAGCTCATA	
vGPDnde	AAACATATGGTTCCCAAAGTTGGCATC	Forward and reverse primers for amplification of vGPD gene.
vGPDbam	CAAAAAGGATGCCGGACAGTAAGGATCCAAA	
vGLOf	GAGGCTAGAAGTTGGCACC	Forward outer primer for amplication of 5' flanking region of vGPD gene
PtrpFR	CAAAAAGTGCTCCTTCAATATCATCTTC	Reverse of forward primer for trpC promoter region in pATBS vector.
TtrpRF	CACAGGTACACTTGTTTAGAGGTAATC	Reverse of reverse primer for trpC terminator region in pATBS vector.

vGROr	GGTATGAATGGCCAATCTTACTTC	Reverse outer primer for amplification of 3' flanking region of vGPD gene
hnRTf	CAGGAGTCCGTCGAGTCTTA	For real-time PCR of histone 3 gene
hnRTr	GGATGTCCTTGCTCTGGATG	
qVir1F	ATTCAGCTTCGACAGGGATG	For real-time PCR of Vir1 gene
qVir1R	CACCTCAGGTAAACTGTGGATAA	
qVir2F	CCAAGGTCTACGGCAAACTAA	For real-time PCR of Vir2 gene
qVir2R	CTCAACAACGCCAAGGAGATA	
qVir3F	CGACTTCTGGCTCACACTATC	For real-time PCR of Vir3 gene
qVir3R	GCAGCGTCTAGGATCTCTTTC	
qVir4F	CCCAATGAGGAGACAGACAAA	For real-time PCR of Vir4 gene
qVir4R	CTTCACCTCCTCGCCAATATC	
qVir5F	CCTCAAGTTCTCGCTGGTATTC	For real-time PCR of Vir5 gene
qVir5R	GGGCACTATTAGCGTACATGAG	
qVir6F	ACTGTTGCCATCCACACTAC	For real-time PCR of Vir6 gene
qVir6R	CAGCACACTCCTGATCTCATC	

qVir7F	GTGCATGTCAAGCTTGTTCG	For real-time PCR of Vir7 gene
qVir7R	GACTGCCAGCCAAGTGATTA	
qGapdhF	CAAGGCTGTTGGAAAGGTTATTC	For real-time PCR of vGPD gene
qGapdhR	CCGTGAAGTCAACGAGAGAAA	
LICrvGPD	TATCCACCTTTACTGTTA CTG GCC GGC ATC CTT T	For cloning in pNH-TrxT vector with ligation independent cloning method
LICfvGPD	TACTTCCAATCCATG GTT CCC AAA GTT GGC ATC	
375F	CACTCACTGGGCAGGATATAAT	Forward and reverse primers for amplification of CDS375 gene from genomic DNA.
375R	GCGTCATAGCCGTCTACATT	
378F	ATCCCGGCTAAGGAACAATATC	Forward and reverse primers for amplification of CDS378 gene from genomic DNA.
378R	TGGTCGAAGCTGTGGAATG	
379F	CCTCGATGGCAGTAACATCTC	Forward and reverse primers for amplification of CDS379 gene from genomic DNA.
379R	CACTGGAGTAAGACGTGGAATTA	
390F	AGAATGCATCTATCTCGCAGTG	Forward and reverse primers for amplification of CDS390 gene from genomic DNA.
390R	CGACGCGACCGAATTGT	
395F	TCCTTAACCGTCTCTTCAAACTC	Forward and reverse primers for amplification of CDS395 gene from genomic DNA.
395R	CAAGCAAATCTAAATCCGGCAA	

409F	ACCATCATATTTCTCAATTGAAGCC	Forward and reverse primers for amplification of CDS409 gene from genomic DNA.
409R	ATTTAGGTTATCCGCCCACAG	
419F	TATTCAGGACATCAGCGTTCTC	Forward and reverse primers for amplification of CDS419 gene from genomic DNA.
419R	CATGGTTCCTCTTGGAAGTCA	
1017F	TCTCACGCGAAGTCAACAA	Forward and reverse primers for amplification of CDS1017 gene from genomic DNA.
1017R	ATCCTGCCGAGCAAGATAAG	
1018F	ACTGGACTACCTGCCTTATACT	Forward and reverse primers for amplification of CDS1018 gene from genomic DNA.
1018R	GCTGCTACTCAGGAATTGACTAA	
1030F	GAACTTGAACGACGCTGCCT	Forward and reverse primers for amplification of CDS1030 gene from genomic DNA.
1030R	CAGAGTCGCAGACTGCATCAA	
2538F	AAAGGCAGCGGCTCAAA	Forward and reverse primers for amplification of CDS2538 gene from genomic DNA.
2538R	GGAGGGAGTGCAGGATCT	
6746F	ATTGCAGAGCAGGCCTTT	Forward and reverse primers for amplification of CDS6746 gene from genomic DNA.
6746R	CACATCCGTCCGATTTCCTT	
6748F	GCGGACGTGAGAAGGAAT	Forward and reverse primers for amplification of CDS6748 gene from genomic DNA.
6748R	CCAGAGTACAGTCAAGGGAAC	

11382F	CGTCAATCCGGTTGATAGTACA	Forward and reverse primers for amplification of CDS11382 gene from genomic DNA.
11382R	GCCTCTATGACTCTTCTGTGATT	
11392F	AATCCCGCATGGCGAAATA	Forward and reverse primers for amplification of CDS11392 gene from genomic DNA.
11392R	CAGGGTTATCCGAATAGCAAGAA	
Histone3F	C GTAAGCAGCTCGCCTCCAAG	Forward and reverse primers for amplification of Histone3 gene from genomic DNA.
Histone3R	CTTGCTCTGGATGGTGACAC	
Pvir4FP	TTTATTCAATTGCAGAGAGAGTTGAGCA	Forward and reverse primers for amplification of promoter region of vir4 gene
Pvir4RP	CAT GAG AGA TTT GTG TGA ATG ATGATG G	
NonSFP	AAACATATGGTTCCCAAAGTTGGCATC	Forward and reverse primers for amplification of nonspecific dsDNA.
NonSRP	CAAAAAGGATGCCGGACAGTAAGGATCCAAA	

Annexure II

Table 1. Volatile organic compounds produced by *T. virens* and vGPD deletion mutant as detected by SPME-GC-MS.

Compounds were identified based on NIST library searches.

S.No.	Name of the Compound	KI Calculated[A]	KI Reported[B]	WT	vGPD
				Sesquiterpenes (%)	
1	Germacrene D	1494.59	1480	17	N.D
2	Germacrene A	1371.98	1391	10	N.D
3	Ledene	1474.55	1482	5	N.D
4	Beta-selinene	1465.15	1485	1	N.D
5	Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1 methylethyl)-[1S-(1alpha,4a beta,8a alpha)]-	1501.81	1524	1	N.D
6	Alloaromadendrene	1439.8	1496	0.9	N.D
7	Beta-caryophyllene[C]	1397.4	1418	0.7	N.D
8	Beta-elemene	1364.41	1389	0.5	N.D
9	Naphthalene, 1,2,4a,5,6,8a-hexahydro-4,7- dimethyl-1-(1-methylethyl)(1 alpha,4a alpha, 8a alpha)-	1516.18	1499	0.5	N.D
10	Alpha-amorphene	1455.48	1474	0.4	N.D
11	Veridiflorol	1573.09	1590	0.4	N.D
12	(-)-Isoledene	1451.43	1421	0.2	N.D
13	Gamma-gurjunene	1450.92	1473	0.2	N.D
				Monoterpenes (%)	
14	l-Limonene	1014.52	1029	26.5	15.4
15	Sabinene[C]	957.16	975	4.8	2.5
16	Trans-beta-ocimene	993.59	1032	1.8	1.09
17	Alpha-thujene	911.7	924	1.6	1
18	Gamma-terpinene	1043.16	1062	1.4	0.93
19	Alpha-terpinene[C]	1000.34	1014	1.35	N.D
20	Alpha-pinene[C]	916.96	932	0.79	0.72
21	Terpinolene	1071.96	1086	0.59	0.25
22	Alpha-terpineol	1177.12	1189	0.1	N.D
23	Trans-ocimene	914.24		0.06	N.D

Table 2. Genes deleted in *Trichoderma virens* mutant (M7) genome. Tv: *Trichoderma virens*, Mg: *Metarhizium guizhouense*, Ap: *Aspergillus pseudocaelatus*, Tas: *Trichoderma asperellum*, Qs: *Quercus suber*, Ta: *Trichoderma atroviride*, Tg: *Trichoderma guizhouense*, Th: *Trichoderma harzianum*, Up: *Eutypalata*

Gene ID (IMI 304061)	Size (aa)	Protein ID (*T. virens* Gv 29-8/ orthologue)	Domains (NCBI CDD search)
Deletion 1, Scaffold 1 (167 Kb)			
CDS375	421	33554 (Tv)	Glycoside hydrolase family 28 protein
CDS376	558	55166 (Tv)	Aryl sulfatase
CDS377	223	34822 (Tv)	Glutathione S-transferase family (GST)
CDS378	430	34910 (Tv)	Transcription factor, GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain
CDS379	157	XP_024770685 (Th)	Transcription factor bZIP
CDS380	314	34863 (Tv)	Cysteine synthase family protein
CDS381	276	137438 (Tv)	Aromatic alcohol reductase
CDS382	447	34387 (Tv)	Acyl-CoA dehydrogenase
CDS383	612	140693 (Tv)	Abhydrolase 1 (esterase/lipase)
CDS384	1241	187765 (Tv)	Multi drug resistance-associated protein (MRP), ABC transporter
CDS385	1266	34120 (Tv)	ABC transporter
CDS386	201	XP_024768020 (Th)	Adenylate forming domain, Class I superfamily and peptide synthase
CDS387	153	62551 (Tv)	No Hit

CDS388	296	140688 (Tv)	Domain of unknown function (DUF3328)
CDS389	2475	62549 (Tv)	Putative polyketide synthase (PKS6)
CDS390	329	62548 (Tv)	Transcription factor, GAL4 and AflR domain-containing protein
CDS391/392	3374	70819 (Tv)	Non-ribosomal peptide synthetase (NRPS), Tex9
CDS393	378	53253 (Tv)	No hit
CDS394	590	191897 (Tv)	Multicopper oxidase with three cupredoxin domains
CDS395	505	2412 (Tv)	Transcription factor, GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain and middle homology region family
CDS396	318	131309 (Tv)	Domain of unknown function (DUF3328)
CDS397	98	222704 (Tv)	No hit
CDS398	382	222705 (Tv)	Class B metal beta-lactamase,
CDS399	386	OPB40689 (Tg)	No hit
CDS400	347	53264 (Tv)	Glycoside hydrolase family 10 protein
CDS401	534	127080 (Tv)	Transcription factor, GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain and middle homology region family
CDS402	506	151748 (Tv)	Amino acid permease (GABA permease)
CDS403	390	151736 (Tv)	Agmatinase, Arginase-like and histone-like hydrolases
CDS404	394	151735 (Tv)	Glycosyltransferase family A

CDS405	291	191903 (Tv)	Short-chain dehydrogenases/reductases (SDR)
CDS406	606	191904 (Tv)	Acetyltransferase (GNAT) family
CDS407	345	191905 (Tv)	Prostaglandin dehydrogenases
CDS408	293	216175 (Tv)	Type 1 glutamine amidotransferase (GATase1)-like domain
CDS409	680	191907 (Tv)	Transcription factor, GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain and middle homology region family
CDS410	635	230771 (Tv)	Ferric reductase and NADPH oxidase (NOX) like domain-containing protein
CDS411	78	---NA---	No hit
CDS412	179	53545 (Tv)	Copper transporter family protein
CDS413	431	151414 (Tv)	Transferase family
CDS414	361	53537 (Tv)	NAD/NADP octopine/ nopaline dehydrogenase and Glycerol-3-phosphate dehydrogenase like domain-containing protein
CDS416	513	59982 (Tv)	Cytochrome P450
CDS417	217	53542 (Tv)	Lipid A phosphoethanolamine transferase
CDS418	534	191806 (Tv)	Major Facilitator Superfamily
CDS419	672	XP_013944953 (Ta)	Transcription factor Middle homology region (MHR)
Deletion 2, Scaffold 1 (35 Kb)			
CDS1017	349	222457 (Tv)	Short-chain dehydrogenases/reductases (SDR)

CDS1018	469	222632 (Tv)	Transcription factor Middle homology region (MHR)
CDS1019	622	53406 (Tv)	Alpha/beta hydrolases (Abhydrolase)
CDS1020	455	53401 (Tv)	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases
CDS1021	136	222635 (Tv)	No hit
CDS1022	163	XP_023915254 (Qs)	Major Facilitator Superfamily
CDS1023	622	201619 (Tv)	Alpha/beta hydrolases (Abhydrolase)
CDS1024	368	213077 (Tv)	No hit
CDS1025	223	191847 (Tv)	Soluble inorganic pyrophosphatase
CDS1026	307	XP_024754708 (Tas)	Camphor resistance (CrcB) family protein
CDS1027	433	53420 (Tv)	Enolase superfamily
CDS1028	242	70901 (Tv)	Alginate lyase (Polysaccharide lyase family 7 protein)
CDS1029	177	191851 (Tv)	Short-chain dehydrogenases/reductases (SDR)
Deletion 3, Scaffold 1 (0.3 Kb)			
CDS2538	81	65480 (Tv)	Protein tyrosine phosphatase-like protein
Deletion 4, Scaffold 20 (8.5 Kb)			
CDS6746	809	90966 (Tv)	Cytochrome P450 (No methyl transferase)
CDS6747	667	202324 (Tv)	No hit
CDS6748	318	51663 (Tv)	Alpha/beta hydrolase family (Abhydrolase 6)
Deletion 5, Scaffold 58 (40 Kb)			
CDS11382	217	66994 (Tv)	No hit

CDS11383	410	128312 (Tv)	Lactonase
CDS11384	261	EMR68722 (Up)	No hit
CDS11385	410	70730 (Tv)	No hit
CDS11386	1167	201824 (Tv)	Catalytic domain of phospholipase D superfamily proteins
CDS11387	378	152027 (Tv)	Glycosyl hydrolase families: GH43, GH62, GH32, GH68, GH117, CH130 and Fungal-type cellulose-binding like domain-containing protein
CDS11388	390	52963 (Tv)	No hit
CDS11389	735	KID82061 (Mg)	hAT family C-terminal dimerization region
CDS11390	312	60120 (Tv)	S1/P1 nucleases
CDS11391	608	KAE8412673 (Ap)	Tetratricopeptide repeat
CDS11392	553	193258 (Tv)	Berberine and berberine and FAD-binding like domain-containing protein

Annexure III

Table 1A. Kovats retention indices for metabolites analysed in WT

S.no.	RT Time	Name of the compound	KI Calculated	KI Reported
1	0.539	Carbamic acid, monoammonium salt		
2	0.55	Pentane	500	
3	0.621	Ethanol	544.0993789	
4	0.711	n-Hexane	600	
5	0.805	Ammonium acetate	606.4032698	
6	0.924	Ethyl Acetate	614.5095368	608
7	1.067		624.2506812	
8	1.105		626.8392371	
9	1.132	1,3-Pentadiene, 2,3-dimethyl-	628.6784741	
10	1.274	dl-Alanyl-l-alanine	638.3514986	
11	1.361		644.2779292	
12	1.47	2,4(1H,3H)-Pyrimidinedione, dihydro-5-hydroxy-	651.7029973	
13	1.505		654.0871935	
14	1.551	Acetic acid, oxo-	657.2207084	
15	1.65	1-Butanol, 3-methyl-	663.9645777	
16	1.834		676.4986376	
17	1.904		681.26703	
18	2.027	2,2-Bis(ethylsulfonyl)propane	689.6457766	
19	2.146		697.7520436	
20	2.179	Heptane, 2,4-dimethyl-	700	
21	2.231		700.9613607	
22	2.331		702.8101313	
23	2.488		705.7127011	
24	2.67	Carbon dioxide	709.0774635	
25	5.379	2,4-Pentadien-1-ol, 3-pentyl-, (2Z)-	759.1606582	
26	7.588	Undecane	1000	
27	9.384	Undecane	1100	
28	10.733	Dodecane	1193.745657	
29	10.823	Dodecane	1200	
30	11.426		1248.047809	
31	11.7	Dodecane, 4,6-dimethyl-	1269.880478	
32	12.078	Tridecane	1300	
33	12.392	Cyclohexane, 1-ethenyl-1-methyl-2-(1-methylethenyl)-4-(1-methylethylidene)-	1330.604288	
34	12.692		1359.844055	
35	12.74		1364.522417	
36	12.788	2-Undecenal	1369.20078	
37	12.935	iso-beta-elemene	1383.528265	1379
38	13.028	iso-beta-elemene	1392.592593	1379
39	13.104	Tetradecane	1400	

40	13.203	1H-Cycloprop[e]azulene, 1a,2,3,4,4a,5,6,7b-octahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.alpha.,4a.beta.,7b.alpha.)]- / Aromadendrene	1409.574468	1439
41	13.323	Caryophyllene	1421.179884	1418
42	13.533	Alloaromadendrene	1441.489362	1461
43	13.873	Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	1474.371373	1472
44	13.974	Germacrene D	1484.139265	1480
45	14.038		1490.32882	
46	14.138	Pentadecane	1500	
47	14.147	.gamma.-Muurolene	1500.983607	1477
48	14.224	(-)-beta-Elemene	1509.398907	1570
49	14.335	Naphthalene, 1,2,3,4,4a,5,6,8a-octahydro-7-methyl-4-methylene-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)- / beta-cadinene	1521.530055	1524
50	14.455	(3aS,8aS)-6,8a-Dimethyl-3-(propan-2-ylidene)-1,2,3,3a,4,5,8,8a-octahydroazulene	1534.644809	
51	14.877	Palustrol	1580.765027	1563
52	15.027	Heptadecane	1597.15847	
53	15.043	Hexadecane	1598.823529	
54	15.131	1H-Cycloprop[e]azulen-4-ol, decahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.beta.,4a.beta.,7.alpha.,7a.beta.,7b.alpha.)]- / (+)-Viridiflorol	1609.176471	1590
55	15.745	3-Heptadecene, (Z)-	1681.411765	
56	15.903	Heptadecane	1700	
57	15.938	Heptadecane	1704.088785	
58	16.129	Pentadecanal-	1726.401869	
59	16.744	Tetradecanoic acid	1798.247664	
60	16.759	Octadecane	1800	
61	16.806	Heptadecane	1805.62201	
62	17.003		1829.186603	
63	17.595	Nonadecane	1900	

Table 1B. Kovats retention indices for metabolites analysed in M7

S.no	RT Time	Name of the compound	KI Calculated	KI Reported
1	0.494	Carbamic acid, monoammonium salt		
2	0.55	Pentane	500	
3	0.711	n-Hexane	600	
4	2.179	Heptane, 2,4-dimethyl-	700	
5	9.384	Undecane	1100	
6	10.823	Dodecane	1200	
7	11.647	Dodecane, 4,6-dimethyl-	1230.586488	
8	12.078	Tetradecane	1300	
9	13.104	Tetradecane	1400	
10	13.204	1H-3a,7-Methanoazulene, octahydro-3,8,8-trimethyl-6-methylene-, [3R-(3.alpha.,3a.beta.,7.beta.,8a.alpha.)]- / beta-Cedrene	1409.67118	1418
11	13.23		1412.185687	
12	13.277	1H-3a,7-Methanoazulene, octahydro-3,8,8-trimethyl-6-methylene-, [3R-(3.alpha.,3a.beta.,7.beta.,8a.alpha.)]-/ beta-Cedrene	1416.731141	1418
13	13.301	Naphthalene, 1,2,3,4,4a,7-hexahydro-1,6-dimethyl-4-(1-methylethyl)-	1419.052224	
14	13.831	(1R,4R,5S)-1,8-Dimethyl-4-(prop-1-en-2-yl)spiro[4.5]dec-7-ene	1470.309478	
15	14.138	Pentadecane	1500	
16	14.209	(1R,5S)-1,8-Dimethyl-4-(propan-2-ylidene)spiro[4.5]dec-7-ene	1507.845304	
17	14.336	Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	1521.878453	1524
18	14.408	(1S,2R,5R)-2-Methyl-5-((R)-6-methylhept-5-en-2-yl)bicyclo[3.1.0]hexan-2-ol /Sabinene hydrate	1529.834254	
19	15.043	Hexadecane	1600	
20	15.759	Spiro[4.5]dec-8-en-7-ol, 4,8-dimethyl-1-(1-methylethyl)-	1683.255814	
21	15.903	Heptadecane	1700	
22	15.973	Spiro[4.5]dec-6-en-8-one, 1,7-dimethyl-4-(1-methylethyl)-	1708.17757	
23	16.365	Tricyclo[5.4.0.0(2,8)]undec-9-ene, 2,6,6,9-tetramethyl-, (1R,2S,7R,8R)-	1753.971963	
24	16.759	Octadecane	1800	

Table 2. Genes downregulated in M7XM7 vs WTXWT (with genes not expressed in M7)

Entries in Bold are common in Table 2 and Table 3

Entries highlighted in Cyan color were confirmed to be downregulated by semi-quantitative RT-PCR in our previous study (Mukherjee et al., 2006)

S.no	Contig_ID	BLAST Hit	WXW_FPKM	7X7_FPKM	Log2 of fold change (C/D)	CDD Search
1	**c22066_g1_i1**	**TRIVIDRAFT_53536**	**19.332**	**0**	**NA**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
2	**c14010_g1_i1**	**TRIVIDRAFT_43673**	**6.8835**	**0**	**NA**	**Short-chain dehydrogenases/reductases (SDR)**
3	**c159_g1_i1**	**TRIVIDRAFT_59982**	**5.8383**	**0**	**NA**	**Cytochrome P450**
4	**c18127_g1_i1**	**TRIVIDRAFT_53537**	**5.7858**	**0**	**NA**	**NAD/NADP octopine/nopaline dehydrogenase**
5	**c6144_g1_i1**	**TRIVIDRAFT_216915**	**5.7803**	**0**	**NA**	**2OG-Fe dioxygenase**
6	**c17958_g1_i1**	**TRIVIDRAFT_222635**	**4.1072**	**0**	**NA**	**No hit**
7	**c4946_g1_i1**	**TRIVIDRAFT_231364**	**3.6796**	**0**	**NA**	**Carbohydrate-binding module family 13 protein**
8	**c14703_g1_i1**	**TRIVIDRAFT_77102**	**3.6389**	**0**	**NA**	**No hit**
9	**c32377_g1_i1**	**TRIVIDRAFT_140693**	**3.5314**	**0**	**NA**	**Alpha/beta hydrolase**
10	**c8205_g1_i2**	**TRIVIDRAFT_151414**	**3.5094**	**0**	**NA**	**Transferase family**
11	**c21741_g1_i1**	**TRIVIDRAFT_153900**	**3.3864**	**0**	**NA**	**Multicopper oxidase with three cupredoxin domains**
12	**c6354_g1_i1**	**TRIVIDRAFT_53350**	**3.1345**	**0**	**NA**	**Pimeloyl-ACP methyl ester carboxylesterase**
13	**c6429_g1_i2**	**TRIVIDRAFT_127080**	**1.8719**	**0**	**NA**	**Fungal specific transcription factor domain**
14	**c32361_g1_i1**	**TRIVIDRAFT_222494**	**1.8711**	**0**	**NA**	**Cytochrome P450**
15	**c954_g1_i2**	**TRIVIDRAFT_35498**	**1.6505**	**0**	**NA**	**Putative polyketide synthase**
16	**c6239_g1_i1**	**TRIVIDRAFT_71600**	**1.0436**	**0**	**NA**	**Glycoside hydrolase family 92 protein**
17	**c14129_g1_i1**	**TRIVIDRAFT_56195**	**194.85**	**0.0165**	**13.525**	**Isoprenoid Biosynthesis enzymes/Terpene cyclase**
18	**c7386_g1_i1**	**TRIVIDRAFT_74289**	**147.85**	**0.0335**	**12.107**	**Cytochrome P450**
19	**c11501_g1_i2**	**TRIVIDRAFT_39153**	**53.192**	**0.0137**	**11.918**	**Cytochrome P450**
20	**c28458_g1_i1**	**GAO88605-Aspergillus udagawae**	**117.3**	**0.0503**	**11.186**	**Thioredoxin reductase**
21	**c28328_g1_i1**	**GAO88609-Aspergillus udagawae**	**92.35**	**0.0491**	**10.879**	**Cytochrome P450**
22	**c11644_g1_i1**	**TRIVIDRAFT_151465**	**42.797**	**0.0249**	**10.746**	**S-adenosylmethionine-dependent methyltransferases**
23	**c10519_g1_i1**	**TRIVIDRAFT_191816**	**45.421**	**0.0339**	**10.39**	**Cytochrome P450**
24	**c12105_g1_i2**	**GAO88607-Aspergillus udagawae**	**20.112**	**0.0176**	**10.158**	**Cytochrome P450**
25	**c14274_g1_i1**	**TRIVIDRAFT_202324**	**13.963**	**0.0123**	**10.153**	**No hit**
26	c7380_g1_i1	GAO88597-Aspergillus udagawae	63.17	0.0588	10.069	Ent-kaurene oxidase

27	**c11410_g1_i1**	**GAO88608-Aspergillus udagawae**	**39.286**	**0.0385**	**9.9943**	**Sterigmatocystin 8-O-methyltransferase**
28	**c17841_g1_i1**	**GAO88603-Aspergillus udagawae**	**87.101**	**0.0877**	**9.9558**	**Nucleobase-cation-symport-1 (NCS1) transporter NRT1-like**
29	c1416_g1_i1	TRIVIDRAFT_83751	52.423	0.0535	9.9366	ABC transporter
30	**c7132_g1_i1**	**TRIVIDRAFT_88460**	**180.5**	**0.1853**	**9.9279**	**Peptidase S8 family domain in ProteinaseK-like proteins**
31	**c15403_g1_i1**	**TRIVIDRAFT_49849**	**36.92**	**0.0433**	**9.7358**	**Hydrophobin**
32	**c8390_g1_i2**	**TRIVIDRAFT_90966**	**28.129**	**0.0364**	**9.5945**	**Cytochrome P450**
33	**c28545_g1_i1**	**TRIVIDRAFT_230740**	**123.2**	**0.1664**	**9.5318**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
34	**c21185_g1_i1**	**TRIVIDRAFT_151590**	**8.2193**	**0.0115**	**9.4845**	**Putative polyketide synthase**
35	**c7528_g1_i3**	**TRIVIDRAFT_74291**	**152.09**	**0.2226**	**9.4162**	**Cytochrome P450**
36	**c10984_g1_i3**	**TRIVIDRAFT_230771**	**7.6083**	**0.012**	**9.3075**	**Ferric reductase and NADPH oxidase (NOX) domain-containing protein**
37	**c5945_g1_i1**	**GAO88606-Aspergillus udagawae**	**13.137**	**0.0277**	**8.8906**	**Membrane bound O-acyl transferase family**
38	**c28953_g1_i1**	**TRIVIDRAFT_91466**	**106.64**	**0.257**	**8.697**	**Hydrophobin**
39	**c28897_g1_i1**	**TRIVIDRAFT_213072**	**81.331**	**0.2016**	**8.6564**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
40	**c31890_g1_i1**	**TRIVIDRAFT_77334**	**7.1051**	**0.0198**	**8.4873**	**Peptidase S8 family domain in ProteinaseK-like proteins**
41	**c1705_g2_i1**	**TRIVIDRAFT_86763**	**5.8954**	**0.0173**	**8.4153**	**Peptidase M14 carboxypeptidase subfamily A/B-like**
42	**c500_g1_i1**	**TRIVIDRAFT_38198**	**114.1**	**0.3505**	**8.3465**	**Cytochrome P450**
43	**c31889_g1_i1**	**TRIVIDRAFT_55809**	**13.184**	**0.0481**	**8.0977**	**Ulp1 protease family**
44	**c32736_g1_i1**	**TRIVIDRAFT_91256**	**19.532**	**0.0739**	**8.0464**	**Domain of unknown function (DUF3237)**
45	**c25402_g1_i1**	**TRIVIDRAFT_156844**	**5.5829**	**0.0229**	**7.9305**	**Ubiquitin 3 binding protein**
46	**c14089_g1_i1**	**TRIVIDRAFT_111957**	**17.716**	**0.0789**	**7.8115**	**No hit**
47	**c9391_g1_i1**	**KID95422-Metarhizium majus**	**3.7186**	**0.0166**	**7.8072**	**Ankyrin repeat-containing domain protein**
48	**c8612_g1_i3**	**TRIVIDRAFT_230775**	**7.3828**	**0.0335**	**7.7856**	**Fungal GAL4-like Zn(2)Cys(6) binuclear cluster DNA-binding domain**
49	**c8643_g1_i1**	**GAO88614-Aspergillus udagawae**	**197.46**	**1.0348**	**7.576**	**ATP synthase subunit beta, mitochondrial**
50	**c10665_g1_i3**	**TRIVIDRAFT_64821**	**24.521**	**0.131**	**7.5481**	**Cytochrome P450**
51	**c17836_g1_i1**	**TRIVIDRAFT_56038**	**1.9876**	**0.0112**	**7.4705**	**Glycosyl hydrolase family 92**
52	**c483_g1_i2**	**TRIVIDRAFT_151142**	**14.104**	**0.0869**	**7.3431**	**SRPBCC super family, Bet v1-like protein**
53	**c18252_g1_i1**	**TRIVIDRAFT_216822**	**7.7022**	**0.0489**	**7.298**	**No hit**
54	**c5798_g1_i1**	**TRIVIDRAFT_81227**	**8.4**	**0.0554**	**7.2436**	**No hit**
55	**c2537_g1_i1**	**TRIVIDRAFT_64883**	**0.5449**	**0.004**	**7.1021**	**Acyl transferase domain in polyketide synthase (PKS) enzymes**
56	c4966_g1_i2	TRIVIDRAFT_231246	17.582	0.1502	6.8712	Sulfite oxidase
57	**c9828_g1_i2**	**TRIVIDRAFT_59507**	**2.2697**	**0.0196**	**6.8547**	**class I SAM-dependent methyltransferase**

58	**c25814_g1_i1**	**TRIVIDRAFT_223757**	**3.7788**	**0.0332**	**6.8285**	**Rossmann-fold NAD(P)(+)-binding proteins**
59	**c11930_g1_i4**	**TRIVIDRAFT_69442**	**3.2161**	**0.0288**	**6.8019**	**Cupredoxin superfamily**
60	c32291_g1_i1	TRIVIDRAFT_90003	123.08	1.1795	6.7052	Trypsin-like serine protease
61	**c7886_g1_i1**	**TRIVIDRAFT_27806**	**22.134**	**0.213**	**6.699**	**Short-chain dehydrogenases/reductases (SDR)**
62	**c16659_g1_i1**	**TRIVIDRAFT_119998**	**15.68**	**0.1537**	**6.6724**	**Methyltransferase domain**
63	**c18324_g1_i1**	**TRIVIDRAFT_84609**	**7.7854**	**0.0768**	**6.6641**	**Alpha/beta hydrolases**
64	**c21378_g1_i1**	**TRIVIDRAFT_48076**	**3.1328**	**0.0312**	**6.6507**	**Glycoside hydrolase family 2**
65	**c21353_g1_i1**	**TRIVIDRAFT_191877**	**4.5035**	**0.0477**	**6.5623**	**Glyoxalase I that uses Zn(++) as cofactor**
66	**c23169_g1_i1**	**TRIVIDRAFT_47547**	**17.487**	**0.1917**	**6.5113**	**Glycoside hydrolase family 18**
67	**c10733_g1_i2**	**CRG91957-Talaromyces islandicus**	**16.875**	**0.1874**	**6.4924**	**Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)**
68	**c11662_g1_i1**	**TRIVIDRAFT_134555**	**9.5204**	**0.106**	**6.4888**	**Cytochrome P450**
69	**c9794_g1_i1**	**TRIVIDRAFT_66940**	**0.5628**	**0.0063**	**6.4806**	**Non-robosomal peptide synthetase**
70	**c4934_g1_i1**	**TRIVIDRAFT_78735**	**35.243**	**0.3981**	**6.4682**	**SRPBCC super family, Bet v1-like protein, Pyrabactin resistance 1 (PYR1)-like**
71	c8154_g1_i5	TRIVIDRAFT_170504	23.866	0.2771	6.4282	Gamma-glutamyl cyclotransferase
72	**c19044_g1_i1**	**TRIVIDRAFT_230790**	**223.32**	**2.6983**	**6.3709**	**Cytochrome P450**
73	**c10825_g1_i2**	**TRIVIDRAFT_179915**	**4.4471**	**0.0545**	**6.3505**	**NAD(P)/FAD-dependent oxidoreductase**
74	**c25019_g1_i1**	**TRIVIDRAFT_9842**	**133.31**	**1.7078**	**6.2865**	**Hydrophobin**
75	**c8283_g1_i2**	**TRIVIDRAFT_53581**	**13.41**	**0.1907**	**6.1357**	**Lactoylglutathione lyase, Catechol 2,3-dioxygenase/ Glyoxalase**
76	**c17734_g1_i1**	**TRIVIDRAFT_209609**	**3.3682**	**0.0494**	**6.0919**	**Putative conidial pigment polyketide synthase PksP/Alb1**
77	**c7434_g1_i1**	**TRIVIDRAFT_53366**	**109.83**	**1.6373**	**6.0679**	**Cytochrome P450**
78	**c3520_g1_i2**	**TRIVIDRAFT_197528**	**3.0584**	**0.0475**	**6.0088**	**No hit**
79	**c24872_g1_i1**	**TRIVIDRAFT_230186**	**27.085**	**0.4286**	**5.9816**	**NAD(P)+-dependent aldehyde dehydrogenase superfamily**
80	c7579_g1_i1	TRIVIDRAFT_178991	1.2815	0.0204	5.9742	C6 transcription factor domain-containing protein
81	**c3720_g1_i1**	**TRIVIDRAFT_160041**	**2.3356**	**0.0386**	**5.919**	**Amidase**
82	c14474_g1_i1	TRIVIDRAFT_64824	1.8447	0.0312	5.8855	Gamma-glutamyltranspeptidase
83	**c15466_g1_i1**	**TRIVIDRAFT_166093**	**1.8675**	**0.033**	**5.8232**	**Fungal catalases**
84	**c9754_g2_i1**	**TRIVIDRAFT_151337**	**66.079**	**1.2233**	**5.7554**	**Cytochrome P450**
85	**c7590_g1_i2**	**TRIVIDRAFT_53582**	**31.26**	**0.5889**	**5.7301**	**NAD(P)/FAD-dependent oxidoreductase, Pyridine nucleotide-disulphide oxidoreductase**
86	**c1052_g1_i1**	**QBZ65333-Pyricularia oryzae**	**73.204**	**1.417**	**5.691**	**Cytochrome P450**
87	**c4882_g1_i1**	**TRIVIDRAFT_110896**	**6.7194**	**0.1328**	**5.6606**	**No hit**
88	**c12112_g1_i3**	**TRIVIDRAFT_191817**	**16.301**	**0.3257**	**5.6455**	**1-aminocyclopropane-1-carboxylate synthase, Aminotransferase class I and II**

89	c13369_g1_i5	TRIVIDRAFT_222581	55.734	1.138	5.614	Major Facilitator Superfamily
90	**c3248_g1_i1**	**TRIVIDRAFT_217588**	**4.6074**	**0.0968**	**5.5729**	**No hit**
91	c28396_g1_i1	TRIVIDRAFT_ 209137	16.708	0.3592	5.5397	DASH complex subunit Ask1
92	**c4567_g1_i1**	**TRIVIDRAFT_216703**	**9.0155**	**0.1954**	**5.5276**	**No hit**
93	**c25704_g1_i1**	**TRIVIDRAFT_231073**	**2.922**	**0.0638**	**5.5171**	**Zinc peptidases M18, M20, M28, and M42**
94	**c32081_g1_i1**	**TRIVIDRAFT_59698**	**3.5586**	**0.0826**	**5.4293**	**Flavocytochrome b2 (FCB2) FMN-binding domain**
95	**c8953_g1_i1**	**TRIVIDRAFT_60674**	**4.0675**	**0.0962**	**5.4025**	**No hit**
96	**c7762_g1_i1**	**TRIVIDRAFT_10003**	**7.7944**	**0.1873**	**5.3794**	**Non-robosomal peptide synthetase**
97	**c17816_g1_i1**	**TRIVIDRAFT_222495**	**11.477**	**0.2837**	**5.338**	**Cytochrome P450**
98	**c6152_g1_i1**	**TRIVIDRAFT_1900**	**8.5442**	**0.2231**	**5.2593**	**WSC domain**
99	**c33071_g1_i1**	**TRIVIDRAFT_205388**	**8.8264**	**0.2341**	**5.2368**	**No hit**
100	**c19025_g1_i1**	**TRIVIDRAFT_53637**	**3.3207**	**0.0917**	**5.1787**	**Cytochrome P450**
101	**c5044_g1_i2**	**TRIVIDRAFT_59236**	**18.651**	**0.5254**	**5.1496**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
102	c2588_g1_i2	TRIVIDRAFT_135139	36.76	1.0421	5.1406	Major Facilitator Superfamily
103	**c21626_g1_i1**	**TRIVIDRAFT_190071**	**3.238**	**0.093**	**5.1224**	**Cytochrome P450**
104	**c5474_g1_i1**	**TRIVIDRAFT_228753**	**2.7007**	**0.0798**	**5.0816**	**No hit**
105	**c22421_g1_i1**	**TRIVIDRAFT_35703**	**14.332**	**0.4272**	**5.0682**	**Transferase family**
106	**c12588_g1_i3**	**TRIVIDRAFT_42485**	**1.1241**	**0.0338**	**5.0577**	**2OG-Fe(II) oxygenase superfamily**
107	**c14667_g1_i1**	**TRIVIDRAFT_53375**	**223.92**	**6.8908**	**5.0222**	**Cytochrome P450**
108	c152_g1_i1	TRIVIDRAFT_50427	1.5525	0.0495	4.971	Cytochrome P450
109	**c17626_g1_i1**	**TRIVIDRAFT_150453**	**2.6111**	**0.0848**	**4.9452**	**Peptidase domain**
110	**c28445_g1_i1**	**TRIVIDRAFT_79838**	**12.976**	**0.4225**	**4.9407**	**WSC domain**
111	**c32849_g1_i1**	**TRIVIDRAFT_216106**	**9.5405**	**0.3205**	**4.8957**	**Blastomyces yeast-phase-specific protein (Bys1)**
112	**c25564_g1_i1**	**TRIVIDRAFT_83686**	**11.727**	**0.4033**	**4.8618**	**Phenylacetate-coenzyme A ligase PaaK**
113	**c18593_g1_i1**	**TRIVIDRAFT_228094**	**3.2714**	**0.1132**	**4.8532**	**Polyamine oxidase**
114	**c29573_g1_i1**	**TRIVIDRAFT_111274**	**9.8794**	**0.351**	**4.8147**	**No hit**
115	**c21140_g1_i1**	**TRIVIDRAFT_85599**	**50.301**	**1.8181**	**4.7901**	**Fungalysin metallopeptidase**
116	**c14014_g1_i1**	**TRIVIDRAFT_59078**	**2.2719**	**0.0837**	**4.7626**	**Major Facilitator Superfamily**
117	**c26273_g1_i1**	**TRIVIDRAFT_27800**	**1.1772**	**0.0439**	**4.7448**	**No hit**
118	**c8255_g1_i2**	**TRIVIDRAFT_40202**	**1.4563**	**0.0588**	**4.6306**	**Putative GprK-type G-protein coupled receptor protein**
119	c2630_g1_i2	TRIVIDRAFT_52608	48.484	1.9763	4.6166	Pleiotropic Drug Resistance (PDR) Family protein
120	c14328_g1_i1	TRIVIDRAFT_8190	1.9248	0.0786	4.6143	Domain of unknown function (DUF4203)
121	**c14950_g1_i1**	**TRIVIDRAFT_60010**	**23.524**	**0.9671**	**4.6043**	**Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)**
122	**c9321_g1_i2**	**TRIVIDRAFT_36679**	**2.5646**	**0.1095**	**4.5496**	**Short-chain dehydrogenases/reductases (SDR)**
123	**c29226_g1_i1**	**TRIVIDRAFT_211280**	**18.469**	**0.7903**	**4.5466**	**No hit**
124	**c7917_g1_i1**	**TRIVIDRAFT_151341**	**27.268**	**1.2289**	**4.4717**	**Alpha/beta hydrolase fold**

125	**c12606_g2_i2**	**KJK82553-Metarhizium anisopliae**	**6.3818**	**0.2887**	**4.4662**	**No hit**
126	**c29003_g1_i1**	**TRIVIDRAFT_83807**	**6.047**	**0.2751**	**4.4582**	**NAD(P)/FAD-dependent oxidoreductase**
127	c10107_g1_i3	TRIVIDRAFT_226303	1.1588	0.0535	4.4362	Rossmann-fold NAD(P)(+)-binding proteins
128	**c5893_g1_i2**	**TRIVIDRAFT_58307**	**11.106**	**0.5142**	**4.4327**	**Peroxidase**
129	**c32183_g1_i1**	**TRIVIDRAFT_55059**	**3.1313**	**0.145**	**4.4325**	**Domain of unknown function (DUF4965)**
130	**c28374_g1_i1**	**TRIVIDRAFT_84937**	**11.707**	**0.5472**	**4.4192**	**M14 family of metallocarboxypeptidases and related proteins**
131	**c1433_g1_i1**	**TRIVIDRAFT_45299**	**7.812**	**0.3657**	**4.4168**	**Aldo/keto reductase**
132	c18499_g1_i1	TRIVIDRAFT_177757	1.0307	0.049	4.394	No hit
133	**c5507_g1_i1**	**TRIVIDRAFT_83658**	**9.4212**	**0.4817**	**4.2897**	**Zinc peptidase like domain-containing protein**
134	**c14151_g1_i1**	**TRIVIDRAFT_214993**	**4.603**	**0.2391**	**4.2672**	**DOMON-like type 9 carbohydrate binding module**
135	**c9428_g1_i1**	**TRIVIDRAFT_170440**	**20.281**	**1.0565**	**4.2627**	**FAD-dependent oxidoreductase**
136	**c22822_g1_i1**	**TRIVIDRAFT_10277**	**2.6742**	**0.1414**	**4.2413**	**GLEYA domain-containing protein**
137	**c22308_g1_i1**	**TRIVIDRAFT_215947**	**13.986**	**0.758**	**4.2055**	**No hit**
138	**c27165_g1_i1**	**TRIVIDRAFT_119401**	**2.826**	**0.1535**	**4.2025**	**Acetyltransferase (GNAT) family**
139	**c9675_g1_i1**	**TRIVIDRAFT_68467**	**0.3203**	**0.0177**	**4.1749**	**Non-ribosomal peptide synthetase**
140	**c17692_g1_i1**	**TRIVIDRAFT_111469**	**16.183**	**0.9131**	**4.1476**	**No hit**
141	**c7152_g1_i1**	**TRIVIDRAFT_215894**	**40.4**	**2.2809**	**4.1467**	**No hit**
142	c7563_g1_i1	TRIVIDRAFT_60434	2.7378	0.1551	4.1418	No hit
143	**c32463_g1_i1**	**TRIVIDRAFT_67871**	**16.407**	**0.9318**	**4.1381**	**FMN-dependent oxidoreductase**
144	**c2200_g1_i1**	**TRIVIDRAFT_30537**	**2.7743**	**0.1593**	**4.1224**	**Acyl-CoA synthetase (AMP-forming)**
145	**c5495_g1_i1**	**TRIVIDRAFT_156714**	**2.9029**	**0.1687**	**4.105**	**Dynamin like protein family includes dynamins and Mx proteins**
146	**c5305_g1_i1**	**TRIVIDRAFT_52335**	**7.3749**	**0.4389**	**4.0708**	**Glycoside hydrolase family 18**
147	c28477_g1_i1	TRIVIDRAFT_47782	1.5969	0.0951	4.0697	Amidase
148	**c3606_g1_i1**	**TRIVIDRAFT_67390**	**3.6393**	**0.2172**	**4.0667**	**GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain**
149	**c21433_g1_i1**	**TRIVIDRAFT_89880**	**2.7878**	**0.1682**	**4.0511**	**Putative sialic acid transporter**
150	c11711_g1_i1	TRIVIDRAFT_61638	10.306	0.6294	4.0335	No hit
151	**c19085_g1_i1**	**TRIVIDRAFT_180603**	**3.307**	**0.2131**	**3.9558**	**Clavaminic acid synthetase**
152	c11735_g2_i2	YP_008757982-Tor sinensis	1.3284	0.0882	3.913	Cytochrome c oxidase subunit III (mitochondrion)
153	**c11156_g1_i2**	**TRIVIDRAFT_47297**	**5.4993**	**0.3669**	**3.9057**	**Domain of unknown function (DUF3292)**
154	c2793_g1_i1	TRIVIDRAFT_114654	1.0263	0.0685	3.9057	No hit
155	c22774_g1_i1	XP_002145932-Talaromyces marneffei	10.23	0.69	3.8902	No hit
156	**c21076_g1_i1**	**TRIVIDRAFT_217322**	**118.35**	**8.0138**	**3.8844**	**Glyceraldehyde-3-phosphate dehydrogenase**

157	c17932_g1_i1	TRIVIDRAFT_186579	1.8352	0.1259	3.8651	Cytochrome P450
158	**c5975_g1_i1**	**TRIVIDRAFT_63956**	**10.282**	**0.712**	**3.8521**	**Peptidase M35 family**
159	**c13975_g1_i1**	**TRIVIDRAFT_172252**	**2.0334**	**0.1416**	**3.8443**	**No hit**
160	**c12608_g1_i1**	**TRIVIDRAFT_69939**	**3.1226**	**0.2228**	**3.8088**	**Peptidase S8 family domain**
161	**c31887_g1_i1**	**TRIVIDRAFT_71936**	**33.44**	**2.4391**	**3.7772**	**Domain in Tre-2, BUB2p, and Cdc16p**
162	**c407_g1_i1**	**TRIVIDRAFT_124167**	**2.2714**	**0.1665**	**3.7704**	**Serine/threonine protein kinase**
163	**c29219_g1_i1**	**TRIVIDRAFT_83099**	**6.8614**	**0.5045**	**3.7655**	**Short-chain dehydrogenases/reductases (SDR)**
164	**c11292_g1_i1**	**TRIVIDRAFT_53086**	**3.9583**	**0.2956**	**3.743**	**Choline dehydrogenase**
165	**c14949_g1_i1**	**TRIVIDRAFT_75875**	**5.3301**	**0.4035**	**3.7234**	**NAD(P)+-dependent aldehyde dehydrogenase superfamily**
166	**c8990_g1_i1**	**TRIVIDRAFT_49**	**3.7385**	**0.2843**	**3.7169**	**Salmonella virulence plasmid 65kDa B protein**
167	**c25173_g1_i1**	**TRIVIDRAFT_48578**	**1.7222**	**0.1365**	**3.6574**	**WD40 domain**
168	**c8085_g1_i2**	**TRIVIDRAFT_215055**	**5.285**	**0.4223**	**3.6457**	**Sulfite oxidase**
169	c28607_g1_i1	TRIVIDRAFT_159594	2.1195	0.1736	3.6102	IBR domain, a half RING-finger domain
170	**c3708_g1_i2**	**TRIVIDRAFT_71258**	**36.232**	**2.9736**	**3.6069**	**GAL4-like Zn(2)Cys(6) binuclear cluster DNA-binding domain**
171	**c8574_g1_i1**	**TRIVIDRAFT_77829**	**32.074**	**2.6351**	**3.6055**	**Taurine catabolism dioxygenase TauD, TfdA family**
172	c7787_g1_i3	TRIVIDRAFT_91392	151.78	12.475	3.6049	Short-chain dehydrogenases/reductases (SDR)
173	**c12248_g1_i1**	**TRIVIDRAFT_118500**	**4.7123**	**0.3911**	**3.5907**	**Zinc finger, C2H2 type**
174	**c5355_g1_i2**	**TRIVIDRAFT_221764**	**2.2481**	**0.1976**	**3.5084**	**No hit**
175	**c24770_g1_i1**	**TRIVIDRAFT_111866**	**3.1357**	**0.2804**	**3.483**	**Glycoside hydrolase family 18**
176	c3299_g1_i1	TRIVIDRAFT_112283	7.0118	0.6506	3.43	Peptidase domain in the S53 family
177	c6482_g1_i1	TRIVIDRAFT_43959	1.061	0.1015	3.386	Major Facilitator Superfamily
178	**c2904_g1_i1**	**TRIVIDRAFT_64335**	**23.262**	**2.2393**	**3.3768**	**Major Facilitator Superfamily**
179	**c1399_g1_i1**	**TRIVIDRAFT_32683**	**4.4216**	**0.4275**	**3.3705**	**Glycoside hydrolase family 55**
180	**c11451_g3_i1**	**TRIVIDRAFT_73257**	**55.711**	**5.3965**	**3.3678**	**Heme-dependent peroxidases similar to plant peroxidases**
181	c8233_g1_i1	TRIVIDRAFT_90482	2.2986	0.2233	3.3638	Choline dehydrogenase
182	**c12649_g1_i1**	**TRIVIDRAFT_193258**	**8.681**	**0.8528**	**3.3476**	**Berberine and berberine like**
183	c3497_g1_i1	TRIVIDRAFT_232088	2.5045	0.2472	3.3406	Epoxide hydrolase N-terminus
184	c33243_g1_i1	TRIVIDRAFT_60814	1.5835	0.163	3.2801	No hit
185	**c29773_g1_i1**	**TRIVIDRAFT_52915**	**4.0141**	**0.4192**	**3.2593**	**Isoprenoid Biosynthesis enzymes, Terpene cyclase**
186	**c8944_g1_i1**	**TRIVIDRAFT_79145**	**60.167**	**6.3105**	**3.2531**	**Predicted oxidoreductase**
187	**c9061_g1_i2**	**TRIVIDRAFT_15465**	**9.4896**	**0.9997**	**3.2467**	**No hit**
188	c2045_g1_i1	TRIVIDRAFT_87510	4.0438	0.428	3.2399	Major Facilitator Superfamily
189	**c24836_g1_i1**	**TRIVIDRAFT_73911**	**5.5032**	**0.5864**	**3.2302**	**No hit**
190	**c6320_g1_i1**	**TRIVIDRAFT_111446**	**5.1795**	**0.5598**	**3.2097**	**No hit**
191	c25876_g1_i1	TRIVIDRAFT_111832	5.59	0.6071	3.2029	No hit

192	c4217_g1_i1	TRIVIDRAFT_32650	3.6736	0.4011	3.1952	Short-chain dehydrogenases/reductases (SDR)
193	**c9987_g1_i2**	**TRIVIDRAFT_141245**	**6.5707**	**0.7302**	**3.1698**	**Acetamidase/Formamidase family**
194	**c18224_g1_i1**	**TRIVIDRAFT_52906**	**16.812**	**1.8777**	**3.1624**	**Protein Kinases, catalytic domain**
195	c6198_g2_i1	KFA45407-Stachybotrys chartarum	2.9889	0.3339	3.162	Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)
196	c4978_g1_i1	TRIVIDRAFT_58920	25.186	2.9166	3.1103	NADPH-dependent FMN reductase
197	**c14150_g1_i1**	**TRIVIDRAFT_209800**	**5.1395**	**0.596**	**3.1083**	**Short-chain dehydrogenases/reductases (SDR)**
198	**c5079_g1_i1**	**TRIVIDRAFT_200088**	**3.0036**	**0.35**	**3.1014**	**L-type amino acid transporter**
199	**c808_g1_i1**	**TRIVIDRAFT_222800**	**5.8644**	**0.694**	**3.079**	**PP-binding and Condensation domain-containing protein**
200	c12941_g1_i4	TRIVIDRAFT_82254	11.428	1.3548	3.0764	Aldehyde dehydrogenase family protein
201	**c11151_g1_i1**	**TRIVIDRAFT_230947**	**28.163**	**3.3759**	**3.0605**	**Peptidases like domain-containing protein**
202	**c10607_g1_i2**	**TRIVIDRAFT_181642**	**6.6514**	**0.8011**	**3.0535**	**Glycoside hydrolase family 2 protein**
203	c6933_g1_i2	TRIVIDRAFT_30156	5.275	0.6387	3.046	Glutathione S-transferase
204	**c28172_g1_i1**	**TRIVIDRAFT_57302**	**6.1353**	**0.7456**	**3.0406**	**S-adenosylmethionine-dependent methyltransferases**
205	**c18000_g1_i1**	**TRIVIDRAFT_76895**	**1.4974**	**0.1849**	**3.0178**	**Glycoside hydrolase family 30 protein**
206	c25277_g1_i1	TRIVIDRAFT_228756	1.1275	0.1419	2.9906	SET domain-containing protein
207	**c28935_g1_i1**	**TRIVIDRAFT_53640**	**10.676**	**1.3481**	**2.9854**	**FAD/FMN-containing dehydrogenase**
208	c7701_g1_i3	TRIVIDRAFT_155537	14.295	1.8662	2.9374	Sulfotransferase family
209	c10272_g1_i1	TRIVIDRAFT_227483	1.277	0.172	2.8923	No hit
210	c3833_g1_i2	TRIVIDRAFT_190955	1.6859	0.235	2.8427	Cytochrome P450
211	c2976_g1_i1	TRIVIDRAFT_57301	1.1841	0.1651	2.8427	Ferredoxin reductase (FNR), an FAD and NAD(P) binding protein
212	c17631_g1_i1	TRIVIDRAFT_180483	5.6324	0.7885	2.8366	Fungal transcription factor regulatory middle homology region
213	**c11578_g1_i4**	**TRIVIDRAFT_58857**	**14.106**	**1.9942**	**2.8224**	**Cytochrome P450**
214	c12902_g2_i1	TRIVIDRAFT_28321	6.9312	0.9841	2.8163	Major Facilitator Superfamily
215	**c12503_g2_i1**	**TRIVIDRAFT_126156**	**3.6684**	**0.5421**	**2.7586**	**No hit**
216	**c31916_g1_i1**	**TRIVIDRAFT_39211**	**6.9488**	**1.0371**	**2.7442**	**Alpha/beta hydrolases**
217	c12878_g1_i2	TRIVIDRAFT_44742	1.9003	0.2874	2.7251	The superfamily of RING finger
218	c10973_g1_i3	TRIVIDRAFT_215037	13.932	2.1618	2.6881	Putative thiazole synthesis
219	c8084_g1_i2	TRIVIDRAFT_50383	1.0579	0.165	2.6806	Non-robosomal peptide synthetase
220	c10531_g1_i3	TRIVIDRAFT_53976	6.1627	0.9624	2.6789	Fungal transcription factor regulatory middle homology region
221	c8436_g1_i1	TRIVIDRAFT_224958	7.641	1.2091	2.6598	Isoprenoid Biosynthesis enzymes, Terpene cyclase
222	c18194_g1_i1	TRIVIDRAFT_81080	4.5457	0.7254	2.6477	Transcription factor bZIP

223	**c5388_g1_i2**	**TRIVIDRAFT_188402**	**2.0109**	**0.322**	**2.6427**	**Glycoside hydrolase family 3 protein**
224	**c24994_g1_i1**	**TRIVIDRAFT_87376**	**6.9584**	**1.1216**	**2.6332**	**Allantoate permease family MFS transporter**
225	**c13171_g1_i2**	**TRIVIDRAFT_54033**	**1.6014**	**0.2581**	**2.633**	**Glycoside hydrolase family 2 protein**
226	**c1136_g1_i1**	**TRIVIDRAFT_179276**	**6.0296**	**0.9756**	**2.6277**	**Neutral/alkaline ceramidase**
227	c21733_g1_i1	TRIVIDRAFT_58191	14.823	2.4017	2.6257	OPT family oligopeptide transporter
228	c25051_g1_i1	TRIVIDRAFT_180068	9.1423	1.4836	2.6235	F-box-like
229	**c8385_g3_i1**	**TRIVIDRAFT_58016**	**3.169**	**0.5156**	**2.6196**	**Major Facilitator Superfamily**
230	c13162_g1_i5	TRIVIDRAFT_34415	7.372	1.207	2.6106	Atrophin-1 family
231	**c11509_g1_i2**	**TRIVIDRAFT_167449**	**5.1085**	**0.8507**	**2.5861**	**Protease associated domain**
232	c6380_g1_i1	TRIVIDRAFT_45439	50.169	8.3812	2.5816	Glycosyl hydrolase family 115
233	**c2641_g1_i2**	**TRIVIDRAFT_78296**	**10.87**	**1.8244**	**2.5748**	**Domain of unknown function (DUF3455)**
234	c18011_g1_i1	TRIVIDRAFT_40764	2.0678	0.348	2.5707	Alpha/beta hydrolases
235	**c16055_g1_i1**	**TRIVIDRAFT_74115**	**9.0135**	**1.5778**	**2.5142**	**Reactive intermediate/imine deaminase**
236	c29251_g1_i1	TRIVIDRAFT_37166	1.6476	0.2886	2.5134	Deoxyribodipyrimidine photo-lyase/cryptochrome family protein
237	c4036_g2_i1	TRIVIDRAFT_45446	4.3845	0.7871	2.4777	Glycoside hydrolase family 31 protein
238	c26017_g1_i1	TRIVIDRAFT_77162	3.3787	0.6076	2.4753	Carbon-nitrogen hydrolase family protein
239	c10917_g1_i2	TRIVIDRAFT_193582	5.0502	0.9134	2.467	Cytochrome P450
240	c18373_g1_i1	TRIVIDRAFT_79679	2.4318	0.4439	2.4536	NAD(P)-dependent alcohol dehydrogenase
241	c7780_g1_i1	TRIVIDRAFT_196765	1.6602	0.3034	2.452	Major Facilitator Superfamily
242	**c9070_g1_i1**	**TRIVIDRAFT_54504**	**2.5802**	**0.4717**	**2.4516**	**Multidrug resistance protein (mdr1)**
243	**c12633_g1_i1**	**TRIVIDRAFT_216052**	**32.64**	**5.9702**	**2.4508**	**No hit**
244	**c9217_g1_i2**	**TRIVIDRAFT_78927**	**1.7133**	**0.3136**	**2.45**	**Lactonase**
245	**c14882_g1_i1**	**TRIVIDRAFT_140278**	**21.255**	**3.9201**	**2.4388**	**Catalase/hydroperoxidase HPI(I)**
246	c3748_g1_i2	TRIVIDRAFT_196192	1.9447	0.3593	2.4362	Aspergillopepsin_like domain-containing protein
247	**c102_g1_i1**	**TRIVIDRAFT_190045**	**1.7643**	**0.3285**	**2.4251**	**Cytochrome P450**
248	c4647_g1_i1	TRIVIDRAFT_55917	2.8546	0.5338	2.419	Cobalamin-independent methionine synthase II family protein
249	c32201_g1_i1	TRIVIDRAFT_189996	5.0713	0.9499	2.4165	Lactonase
250	c21575_g1_i1	TRIVIDRAFT_49552	8.957	1.6819	2.4129	No hit
251	**c14167_g1_i1**	**TRIVIDRAFT_213573**	**3.4282**	**0.6438**	**2.4128**	**Gamma-glutamyltranspeptidase**
252	**c9960_g1_i3**	**TRIVIDRAFT_70229**	**3.561**	**0.6701**	**2.4097**	**Major Facilitator Superfamily**
253	c4740_g1_i1	TRIVIDRAFT_185571	15.454	2.91	2.4089	No hit
254	c28338_g1_i1	TRIVIDRAFT_165573	7.6162	1.4367	2.4063	MBL fold metallo-hydrolase
255	**c24759_g1_i1**	**TRIVIDRAFT_111479**	**2.3347**	**0.4427**	**2.3987**	**Glycoside hydrolase family 75 protein**
256	c28485_g1_i1	TRIVIDRAFT_52450	2.2083	0.4198	2.3953	Phosphotransferase enzyme family
257	c10615_g1_i2	TRIVIDRAFT_75334	9.182	1.763	2.3808	Glutamate-cysteine ligase

258	**c14811_g1_i1**	**TRIVIDRAFT_192636**	**14.624**	**2.8248**	**2.3722**	**Domain of unknown function (DUF2235)**
259	c14031_g1_i1	TRIVIDRAFT_33553	1.1416	0.2211	2.368	Heterokaryon incompatibility protein (HET)
260	**c4698_g1_i1**	**TRIVIDRAFT_212656**	**7.5832**	**1.4826**	**2.3547**	**Major Facilitator Superfamily**
261	**c32281_g1_i1**	**TRIVIDRAFT_194303**	**6.3865**	**1.2513**	**2.3516**	**Cytochrome P450**
262	c23308_g1_i1	TRIVIDRAFT_91258	18.064	3.5414	2.3507	Glutathione S-transferase
263	**c5024_g1_i1**	**TRIVIDRAFT_209187**	**1.6303**	**0.3212**	**2.3435**	**No hit**
264	**c2922_g1_i1**	**TRIVIDRAFT_52839**	**10.71**	**2.1331**	**2.328**	**Glutathione S-transferase**
265	**c7046_g1_i1**	**TRIVIDRAFT_55803**	**20.946**	**4.1927**	**2.3207**	**Linoleate (8R)-dioxygenase and related enzymes**
266	**c31870_g1_i1**	**TRIVIDRAFT_204543**	**6.5254**	**1.3062**	**2.3207**	**No hit**
267	c58_g1_i1	TRIVIDRAFT_177810	6.2604	1.2561	2.3173	Nitrate reductase
268	c17694_g1_i1	TRIVIDRAFT_68015	0.7664	0.1549	2.3065	Ankyrin repeats
269	c5986_g1_i1	TRIVIDRAFT_10390	7.6316	1.552	2.2979	RNA dependent RNA polymerase
270	c5295_g1_i1	TRIVIDRAFT_63798	3.4056	0.7001	2.2823	Short-chain dehydrogenases/reductases (SDR)
271	c340_g1_i1	TRIVIDRAFT_74437	4.2277	0.8796	2.265	No hit
272	**c4413_g1_i2**	**TRIVIDRAFT_69335**	**4.515**	**0.943**	**2.2593**	**Amino acid permease (GABA permease)**
273	**c28394_g1_i1**	**TRIVIDRAFT_33825**	**1.5845**	**0.3315**	**2.257**	**L-lysine 6-monooxygenase (NADPH-requiring)**
274	**c18322_g1_i1**	**TRIVIDRAFT_60255**	**3.8269**	**0.8011**	**2.256**	**No hit**
275	c22140_g1_i1	TRIVIDRAFT_44127	6.0793	1.2767	2.2515	No hit
276	**c14505_g1_i1**	**TRIVIDRAFT_183528**	**6.1373**	**1.2938**	**2.246**	**Gametolysin peptidase M11**
277	c3813_g1_i1	TRIVIDRAFT_215959	2.6306	0.5566	2.2408	Methylmalonate semialdehyde dehydrogenase and ALDH family members
278	**c28619_g1_i1**	**TRIVIDRAFT_39958**	**1.1795**	**0.2513**	**2.2305**	**No hit**
279	c28439_g1_i1	TRIVIDRAFT_192994	1.4648	0.3172	2.2071	No hit
280	c25575_g1_i1	TRIVIDRAFT_172947	3.078	0.6683	2.2034	D-Lactate and related Dehydrogenases
281	c4473_g1_i1	TRIVIDRAFT_232505	3.4915	0.7612	2.1975	Amidase
282	**c25858_g1_i1**	**TRIVIDRAFT_157800**	**1.6089**	**0.3513**	**2.1952**	**Glycoside hydrolase family 47 protein**
283	c5022_g1_i2	TRIVIDRAFT_59338	3.9094	0.8664	2.1739	Glutathione S-transferase
284	c8849_g1_i1	TRIVIDRAFT_192178	3.8991	0.8758	2.1544	Multidrug resistance protein (mdr1)
285	**c1964_g1_i2**	**TRIVIDRAFT_156071**	**2.8576**	**0.654**	**2.1274**	**ATPase family associated with various cellular activities (AAA)**
286	c31921_g1_i1	TRIVIDRAFT_110629	15.582	3.5692	2.1262	Mannitol dehydrogenase (MDH)-like
287	c9888_g1_i1	TRIVIDRAFT_43	3.6767	0.8611	2.0942	Salmonella virulence plasmid 65kDa B protein
288	c32780_g1_i1	TRIVIDRAFT_58255	4.5177	1.0886	2.0531	Superfamily of metallo-dependent hydrolases
289	**c28703_g1_i1**	**TRIVIDRAFT_61863**	**6.4236**	**1.554**	**2.0474**	**Glycosyltransferase family 2 protein**
290	c17925_g1_i1	TRIVIDRAFT_132119	1.06	0.2566	2.0467	No hit

291	**c6337_g1_i1**	**TRIVIDRAFT_83973**	**20.42**	**4.9608**	**2.0413**	**Aminoglycoside 3'-phosphotransferase and Choline Kinase family**
292	c745_g1_i1	TRIVIDRAFT_37192	33.703	8.1901	2.0409	Glycine/D-amino acid oxidase (deaminating)
293	**c9685_g1_i2**	**TRIVIDRAFT_53518**	**1.7836**	**0.4434**	**2.008**	**Acyl transferase domain in polyketide synthase (PKS) enzymes**
294	c6460_g1_i1	TRIVIDRAFT_14210	10.165	2.5394	2.001	Nitrate transmembrane transporter

Table 3. Genes downregulated in M7XRs vs WTXRs (including genes not expressed in M7xRs)

Entries in Bold are common in Table 3 and Table 2

Entries highlighted in Cyan color were confirmed to be downregulated by semi-quantitative RT-PCR in our previous study (Mukherjee et al., 2006)

S.no	Contig_ID	BLAST Hit	WXR_FPKM	7XR_FPKM	Log2 of Fold change (C/D)	CDD Search
1	**c11644_g1_i1**	**TRIVIDRAFT_151465**	**58.868**	**0**	**NA**	**S-adenosylmethionine-dependent methyltransferases**
2	**c11501_g1_i1**	**TRIVIDRAFT_39153**	**55.292**	**0**	**NA**	**Cytochrome P450**
3	**c8390_g1_i1**	**TRIVIDRAFT_90966**	**34.932**	**0**	**NA**	**Cytochrome P450**
4	**c15403_g1_i1**	**TRIVIDRAFT_49849**	**29.577**	**0**	**NA**	**Hydrophobin**
5	**c29573_g1_i1**	**TRIVIDRAFT_111274**	**19.235**	**0**	**NA**	**No hit**
6	**c10984_g1_i1**	**TRIVIDRAFT_230771**	**12.734**	**0**	**NA**	**Ferric reductase and NADPH oxidase (NOX) domain-containing protein**
7	**c29226_g1_i1**	**TRIVIDRAFT_211280**	**12.373**	**0**	**NA**	**No hit**
8	c4854_g1_i1	KEY73627-Stachybotrys chartarum	11.743	0	NA	Amidohydrolase
9	**c4567_g1_i1**	**TRIVIDRAFT_216703**	**11.596**	**0**	**NA**	**No hit**
10	**c159_g1_i2**	**TRIVIDRAFT_59982**	**5.8422**	**0**	**NA**	**Cytochrome P450**
11	c15834_g1_i1	M431DRAFT_171314-Trichoderma harzianum	5.8365	0	NA	No hit
12	c32299_g1_i1	TRIVIDRAFT_83938	5.7139	0	NA	Alpha/beta hydrolases
13	**c18127_g1_i1**	**TRIVIDRAFT_53537**	**5.5826**	**0**	**NA**	**NAD/NADP octopine/nopaline dehydrogenase**
14	**c8255_g1_i1**	**TRIVIDRAFT_40202**	**5.271**	**0**	**NA**	**Putative GprK-type G-protein coupled receptor protein**
15	**c11930_g1_i1**	**TRIVIDRAFT_69442**	**5.1747**	**0**	**NA**	**Cupredoxin superfamily**
16	**c14703_g1_i1**	**TRIVIDRAFT_77102**	**5.0577**	**0**	**NA**	**No hit**
17	**c21741_g1_i1**	**TRIVIDRAFT_153900**	**4.5246**	**0**	**NA**	**Multicopper oxidase with three cupredoxin domains**
18	c10988_g1_i1	TRIVIDRAFT_213202	4.5003	0	NA	Glycoside hydrolase family 18 protein
19	**c12773_g2_i1**	**TRIVIDRAFT_217588**	**4.4252**	**0**	**NA**	**No hit**
20	**c32377_g1_i1**	**TRIVIDRAFT_140693**	**4.3228**	**0**	**NA**	**Alpha/beta hydrolase**
21	**c17958_g1_i1**	**TRIVIDRAFT_222635**	**3.9815**	**0**	**NA**	**No hit**
22	**c8205_g1_i1**	**TRIVIDRAFT_151414**	**3.8095**	**0**	**NA**	**Transferase family**
23	c34934_g1_i1	TRIVIDRAFT_192768	3.6565	0	NA	Glycoside hydrolase family 78 protein
24	c2359_g1_i1	LPUS_07394-Umbilicaria pustulata	3.3634	0	NA	Beta transducin-like protein het-e4s
25	**c9828_g1_i1**	**TRIVIDRAFT_59507**	**3.3116**	**0**	**NA**	**class I SAM-dependent methyltransferase**
26	c25824_g1_i1	TRIVIDRAFT_92793	3.2902	0	NA	No hit

27	c23941_g1_i1	TRIVIDRAFT_70901	3.0978	0	NA	Polysaccharide lyase family 7 protein
28	c21264_g1_i1	TRIVIDRAFT_78959	2.9045	0	NA	Bicupin, oxalate decarboxylase family
29	c34248_g1_i1	TRIVIDRAFT_69040	2.8948	0	NA	17-beta hydroxysteroid dehydrogenase-like
30	c11871_g1_i1	TRIVIDRAFT_64881	2.8094	0	NA	No hit
31	c33380_g1_i1	TRIVIDRAFT_36613	2.7967	0	NA	No hit
32	c30416_g1_i1	TRIVIDRAFT_80526	2.2316	0	NA	Major Facilitator Superfamily
33	c19668_g1_i1	TRIVIDRAFT_60531	2.2083	0	NA	Hydrophobin
34	c10796_g1_i1	TRIVIDRAFT_155454	2.2056	0	NA	No hit
35	c24974_g1_i1	TRIVIDRAFT_151736	2.1574	0	NA	Agmatinase-like family includes proclavaminic acid amidinohydrolase
36	c14523_g1_i1	TRIVIDRAFT_64889	2.1082	0	NA	Phosphopantetheine attachment site, PP-binding and AdoMet_MTases domain-containing protein
37	c14503_g1_i1	TRIVIDRAFT_147314	1.9808	0	NA	NADPH oxidase (NOX)
38	c2118_g1_i1	TRIVIDRAFT_194289	1.9548	0	NA	Short-chain dehydrogenases/reductases (SDR)
39	c3988_g1_i1	TRIVIDRAFT_6301	1.952	0	NA	No hit
40	c25311_g1_i1	TRIVIDRAFT_218089	1.8198	0	NA	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases
41	c22735_g1_i1	TRIVIDRAFT_69487	1.8065	0	NA	No hit
42	**c15466_g1_i1**	**TRIVIDRAFT_166093**	**1.6671**	**0**	**NA**	**Fungal catalases**
43	c18639_g1_i1	UCREL1_9593-Eutypa lata	1.6004	0	NA	Putative trichodiene oxygenase protein
44	c26408_g1_i1	TRIVIDRAFT_124686	1.5967	0	NA	Lipoxygenase
45	c23411_g1_i1	TRIVIDRAFT_200439	1.5407	0	NA	No hit
46	c26092_g1_i1	TRIVIDRAFT_191904	1.5372	0	NA	Acetyltransferase (GNAT) family
47	c14776_g1_i1	TRIVIDRAFT_33810	1.4196	0	NA	Major Facilitator Superfamily
48	c34618_g1_i1	TRIVIDRAFT_61959	1.4117	0	NA	Major Facilitator Superfamily
49	c31904_g1_i1	TRIVIDRAFT_16072	1.3415	0	NA	No hit
50	c3490_g1_i1	TRIVIDRAFT_59067	1.3257	0	NA	Major Facilitator Superfamily
51	c24660_g1_i1	TRIVIDRAFT_46591	1.3191	0	NA	Short-chain dehydrogenases/reductases (SDR)
52	c5194_g1_i1	TRIVIDRAFT_208681	1.271	0	NA	Basic leucine zipper (bZIP) domain of Yeast Activator Protein (YAP)
53	c8086_g1_i1	TRIVIDRAFT_219507	1.2422	0	NA	Domain of unknown function (DUF4188)
54	c19820_g1_i1	TRIVIDRAFT_70756	1.2335	0	NA	No hit
55	c16070_g1_i1	TRIVIDRAFT_192258	1.2273	0	NA	Predicted dehydrogenase
56	c33547_g1_i1	TRIVIDRAFT_52825	1.1986	0	NA	Peptidase M14 carboxypeptidase subfamily A/B-like
57	c31910_g1_i1	TRIVIDRAFT_206986	1.1933	0	NA	No hit
58	c369_g1_i1	TRIVIDRAFT_44039	1.1555	0	NA	Acetyltransferase (GNAT) domain
59	c31562_g1_i1	TRIVIDRAFT_44273	1.1547	0	NA	Non-ribosomal peptide synthetase
60	c10277_g1_i1	TRIVIDRAFT_111733	1.1386	0	NA	Chromosome segregation protein SMC
61	c32264_g1_i1	TRIVIDRAFT_59139	1.1315	0	NA	GLEYA domain
62	c2116_g1_i1	TRIVIDRAFT_48760	1.126	0	NA	No hit

63	c28443_g1_i1	TRIVIDRAFT_206874	1.1217	0	NA	No hit
64	c5596_g1_i1	THARTR1_06381-Trichoderma harzianum	1.0994	0	NA	Zinc-dependent metalloprotease
65	c25902_g1_i1	TRIVIDRAFT_77255	1.0916	0	NA	glycoside hydrolase family 79 protein
66	c26856_g1_i1	TRIVIDRAFT_53859	1.0908	0	NA	Major Facilitator Superfamily
67	c21780_g1_i1	TRIVIDRAFT_155696	1.0771	0	NA	No hit
68	c30042_g1_i1	TRIVIDRAFT_62770	1.041	0	NA	Polysaccharide lyase family 8 protein
69	c16508_g1_i1	TRIVIDRAFT_191945	1.0253	0	NA	Indoleamine 2,3-dioxygenase
70	**c7386_g1_i1**	**TRIVIDRAFT_74289**	**169.16**	**0.0095**	**14.113**	**Cytochrome P450**
71	**c7528_g1_i1**	**TRIVIDRAFT_74291**	**167.11**	**0.049**	**11.735**	**Cytochrome P450**
72	**c14129_g1_i1**	**TRIVIDRAFT_56195**	**293.54**	**0.0989**	**11.536**	**Isoprenoid Biosynthesis enzymes**
73	**c12105_g1_i1**	**GAO88607-Aspergillus udagawae**	**28.99**	**0.0141**	**11.011**	**Cytochrome P450**
74	c21195_g1_i1	TRIVIDRAFT_62540	9.1224	0.0052	10.79	Non-ribosomal peptide synthetase
75	**c14274_g1_i1**	**TRIVIDRAFT_202324**	**17.65**	**0.0105**	**10.718**	**No hit**
76	c1416_g1_i1	GAO88611-Aspergillus udagawae	91.917	0.0588	10.611	Non-ribosomal peptide synthetase 10
77	**c10665_g1_i1**	**TRIVIDRAFT_64821**	**44.794**	**0.0373**	**10.229**	**Cytochrome P450**
78	**c5945_g1_i1**	**GAO88606-Aspergillus udagawae**	**27.955**	**0.0237**	**10.207**	**Membrane bound O-acyl transferase family**
79	**c28953_g1_i1**	**TRIVIDRAFT_91466**	**93.738**	**0.0941**	**9.9603**	**Hydrophobin**
80	**c7132_g1_i1**	**TRIVIDRAFT_88460**	**380.62**	**0.4275**	**9.7983**	**Peptidase S8 family domain in ProteinaseK-like proteins**
81	**c17841_g1_i1**	**GAO88603-Aspergillus udagawae**	**106.17**	**0.1199**	**9.7903**	**Nucleobase-cation-symport-1 (NCS1) transporter NRT1-like**
82	**c5798_g1_i1**	**TRIVIDRAFT_81227**	**13.904**	**0.0158**	**9.7825**	**No hit**
83	**c22066_g1_i1**	**TRIVIDRAFT_53536**	**20.84**	**0.0254**	**9.6802**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
84	**c10519_g1_i1**	**TRIVIDRAFT_191816**	**54.335**	**0.0723**	**9.5535**	**Cytochrome P450**
85	c7380_g1_i1	GAO88595-Aspergillus udagawae	88.27	0.134	9.3635	Cytochrome P450
86	**c28458_g1_i1**	**GAO88605-Aspergillus udagawae**	**161.68**	**0.2581**	**9.2909**	**Thioredoxin reductase**
87	**c11410_g1_i1**	**GAO88608-Aspergillus udagawae**	**51.252**	**0.0823**	**9.283**	**Sterigmatocystin 8-O-methyltransferase**
88	**c28328_g1_i1**	**GAO88609-Aspergillus udagawae**	**126.16**	**0.2096**	**9.2336**	**Cytochrome P450**
89	**c10825_g1_i2**	**TRIVIDRAFT_179915**	**6.6548**	**0.0116**	**9.159**	**NAD(P)/FAD-dependent oxidoreductase**
90	c29787_g1_i1	TRIVIDRAFT_146040	10.84	0.0216	8.9683	Short-chain dehydrogenases/reductases (SDR)
91	**c17734_g1_i1**	**TRIVIDRAFT_209609**	**5.233**	**0.0105**	**8.9544**	**Putative conidial pigment polyketide synthase PksP/Alb1**
92	**c17626_g1_i1**	**TRIVIDRAFT_150453**	**9.723**	**0.0207**	**8.8763**	**Peptidase domain**
93	**c14010_g1_i1**	**TRIVIDRAFT_43673**	**7.0956**	**0.0154**	**8.8488**	**Short-chain dehydrogenases/reductases (SDR)**
94	**c11662_g1_i3**	**TRIVIDRAFT_134555**	**12.315**	**0.0287**	**8.7449**	**Cytochrome P450**
95	**c500_g1_i1**	**TRIVIDRAFT_38198**	**126.96**	**0.3106**	**8.6751**	**Cytochrome P450**
96	**c11292_g1_i1**	**TRIVIDRAFT_53086**	**8.2451**	**0.0202**	**8.6724**	**Choline dehydrogenase**

97	**c8612_g1_i1**	**TRIVIDRAFT_230775**	**12.173**	**0.03**	**8.6665**	**Fungal GAL4-like Zn(2)Cys(6) binuclear cluster DNA-binding domain**
98	**c17836_g1_i1**	**TRIVIDRAFT_56038**	**3.6528**	**0.0096**	**8.5754**	**Glycosyl hydrolase family 92**
99	**c32081_g1_i1**	**TRIVIDRAFT_59698**	**19.662**	**0.0529**	**8.5373**	**Flavocytochrome b2 (FCB2) FMN-binding domain**
100	c3248_g1_i1	TRIVIDRAFT_217588	4.7095	0.0138	8.4164	No hit
101	**c6427_g1_i1**	**TRIVIDRAFT_66940**	**0.7699**	**0.0023**	**8.3738**	**Non-robosomal peptide synthetase**
102	**c28897_g1_i1**	**TRIVIDRAFT_213072**	**79.316**	**0.2727**	**8.1842**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
103	**c9675_g1_i1**	**TRIVIDRAFT_68467**	**0.643**	**0.0023**	**8.1078**	**Non-ribosomal peptide synthetase**
104	**c18324_g1_i1**	**TRIVIDRAFT_84609**	**6.8608**	**0.0262**	**8.0306**	**Alpha/beta hydrolases**
105	c19191_g1_i1	TRIVIDRAFT_87418	14.507	0.0555	8.0306	Glycoside hydrolase family 75 protein
106	**c1705_g2_i1**	**TRIVIDRAFT_86763**	**15.271**	**0.059**	**8.0153**	**Peptidase M14 carboxypeptidase subfamily A/B-like**
107	**c8643_g1_i1**	**GAO88614-Aspergillus udagawae**	**320.99**	**1.2408**	**8.0152**	**ATP synthase subunit beta, mitochondrial**
108	**c28545_g1_i1**	**TRIVIDRAFT_230740**	**275.95**	**1.1376**	**7.9222**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
109	**c17816_g1_i1**	**TRIVIDRAFT_222495**	**37.076**	**0.1531**	**7.9197**	**Cytochrome P450**
110	c29795_g1_i1	TRIVIDRAFT_196109	4.2565	0.0176	7.9197	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I
111	**c14089_g1_i1**	**TRIVIDRAFT_111957**	**21.799**	**0.0943**	**7.8523**	**No hit**
112	c26198_g1_i1	TRIVIDRAFT_119427	12.467	0.0569	7.7743	No hit
113	**c16659_g1_i1**	**TRIVIDRAFT_119998**	**18.71**	**0.0876**	**7.7393**	**Methyltransferase domain**
114	**c2537_g1_i1**	**TRIVIDRAFT_64883**	**3.5335**	**0.0169**	**7.7041**	**Acyl transferase domain in polyketide synthase (PKS) enzymes**
115	c8154_g1_i5	PGH15375-Polytolypa hystricis	30.731	0.1496	7.6828	ATP11 protein
116	c2087_g1_i1	TRIVIDRAFT_89999	6.8017	0.0363	7.5501	Glycoside hydrolase family 18 protein
117	c33065_g1_i1	TRIVIDRAFT_232090	7.155	0.0417	7.4211	Taurine catabolism dioxygenase TauD
118	**c12112_g1_i3**	**TRIVIDRAFT_191817**	**27.707**	**0.1637**	**7.4033**	**1-aminocyclopropane-1-carboxylate synthase, Aminotransferase class I and II**
119	**c25402_g1_i1**	**TRIVIDRAFT_156844**	**13.231**	**0.0782**	**7.4023**	**Ubiquitin 3 binding protein**
120	**c6354_g1_i1**	**TRIVIDRAFT_53350**	**10.371**	**0.0623**	**7.3802**	**Pimeloyl-ACP methyl ester carboxylesterase**
121	**c13975_g1_i1**	**TRIVIDRAFT_172252**	**3.5778**	**0.022**	**7.3458**	**No hit**
122	**c6144_g1_i1**	**TRIVIDRAFT_216915**	**7.4737**	**0.046**	**7.3446**	**2OG-Fe dioxygenase**
123	**c4946_g1_i1**	**TRIVIDRAFT_231364**	**5.248**	**0.0339**	**7.2742**	**Carbohydrate-binding module family 13 protein**
124	**c31889_g1_i1**	**TRIVIDRAFT_55809**	**25.077**	**0.1645**	**7.2522**	**Ulp1 protease family**
125	c19536_g1_i1	TRIVIDRAFT_207931	4.8099	0.0322	7.2218	L,D-transpeptidases/carboxypeptidases
126	**c21185_g1_i1**	**TRIVIDRAFT_151590**	**35.493**	**0.2386**	**7.2171**	**Putative polyketide synthase**
127	c987_g1_i1	TRIVIDRAFT_59784	3.2592	0.0227	7.1673	FAD/NADP like domain-containing protein

128	**c23169_g1_i1**	**TRIVIDRAFT_47547**	**33.887**	**0.2366**	**7.1622**	**Glycoside hydrolase family 18**
129	c32476_g1_i1	TRIVIDRAFT_153908	5.3855	0.0381	7.1449	Taurine catabolism dioxygenase TauD
130	**c25704_g1_i1**	**TRIVIDRAFT_231073**	**22.786**	**0.1636**	**7.1222**	**Zinc peptidases M18, M20, M28, and M42**
131	**c7886_g1_i1**	**TRIVIDRAFT_27806**	**39.247**	**0.2832**	**7.1148**	**Short-chain dehydrogenases/reductases (SDR)**
132	c12902_g1_i3	TRIVIDRAFT_28006	16.212	0.118	7.1019	Photolyase, partial
133	c7516_g1_i1	TRIVIDRAFT_47944	4.9504	0.0362	7.0953	L-type amino acid transporter
134	c28239_g1_i1	TRIVIDRAFT_180069	1.4594	0.011	7.0518	Glycoside hydrolase family 20 protein, partial
135	**c32736_g1_i1**	**TRIVIDRAFT_91256**	**20.661**	**0.1578**	**7.0325**	**Domain of unknown function (DUF3237)**
136	c28396_g1_i1	TRIVIDRAFT_209137	18.663	0.1432	7.0258	DASH complex subunit Ask1
137	**c24872_g1_i1**	**TRIVIDRAFT_230186**	**42.296**	**0.3527**	**6.906**	**NAD(P)+-dependent aldehyde dehydrogenase superfamily**
138	**c25564_g1_i1**	**TRIVIDRAFT_83686**	**14.403**	**0.1241**	**6.8592**	**Phenylacetate-coenzyme A ligase PaaK**
139	c11227_g1_i1	TRIVIDRAFT_206945	2.6593	0.0229	6.8592	Salicylate hydroxylase
140	c14656_g1_i1	TRIVIDRAFT_47998	1.8467	0.0161	6.8454	Flavin-containing monooxygenase FMO GS-OX
141	**c21378_g1_i1**	**TRIVIDRAFT_48076**	**16.339**	**0.1421**	**6.8454**	**Glycoside hydrolase family 2**
142	c25031_g1_i1	DAA76265-Arthroderma benhamiae	2.5555	0.0225	6.8279	PKS
143	c21133_g1_i1	KEY73630-Stachybotrys chartarum	12.806	0.115	6.7996	FAD binding domain, Oxidoreductase
144	**c9321_g1_i2**	**TRIVIDRAFT_36679**	**3.8282**	**0.0351**	**6.7694**	**Short-chain dehydrogenases/reductases (SDR)**
145	**c25019_g1_i1**	**TRIVIDRAFT_9842**	**109.06**	**1.0032**	**6.7643**	**Hydrophobin**
146	**c21140_g1_i1**	**TRIVIDRAFT_85599**	**104.03**	**1.116**	**6.5425**	**Fungalysin metallopeptidase**
147	**c8953_g1_i1**	**TRIVIDRAFT_60674**	**2.9798**	**0.0329**	**6.5025**	**No hit**
148	**c21626_g1_i1**	**TRIVIDRAFT_190071**	**9.3531**	**0.1059**	**6.4646**	**Cytochrome P450**
149	**c5495_g1_i1**	**TRIVIDRAFT_156714**	**2.5287**	**0.0288**	**6.4548**	**Dynamin like protein family includes dynamins and Mx proteins**
150	**c5474_g1_i1**	**TRIVIDRAFT_228753**	**3.8429**	**0.0454**	**6.4023**	**No hit**
151	c25901_g1_i1	TRIVIDRAFT_215593	21.731	0.2716	6.3221	Polysaccharide lyase
152	c17837_g1_i1	TRIVIDRAFT_222452	12.43	0.1561	6.3149	Transposase
153	**c4882_g1_i1**	**TRIVIDRAFT_110896**	**8.9104**	**0.1135**	**6.2947**	**No hit**
154	**c7762_g1_i1**	**TRIVIDRAFT_10003**	**8.2516**	**0.1058**	**6.2858**	**Non-ribosomal peptide synthetase**
155	**c32361_g1_i1**	**TRIVIDRAFT_222494**	**6.6882**	**0.0866**	**6.2708**	**Cytochrome P450**
156	**c12588_g1_i1**	**TRIVIDRAFT_42485**	**2.298**	**0.03**	**6.2587**	**2OG-Fe(II) oxygenase superfamily**
157	c7236_g2_i2	TRIVIDRAFT_153126	5.7967	0.0802	6.1747	Lysozyme-like domains
158	**c9391_g1_i1**	**KID95422-Metarhizium majus**	**3.0272**	**0.0426**	**6.1524**	**Ankyrin repeat-containing domain protein**
159	**c15439_g1_i1**	**TRIVIDRAFT_48578**	**2.3603**	**0.0334**	**6.1449**	**WD40 domain**
160	**c19025_g1_i1**	**TRIVIDRAFT_53637**	**10.047**	**0.1436**	**6.1285**	**Cytochrome P450**
161	c28553_g1_i1	TRIVIDRAFT_72151	205.39	3.0184	6.0885	ABC transporter F family
162	**c6152_g1_i2**	**TRIVIDRAFT_1900**	**9.136**	**0.1355**	**6.0757**	**WSC domain**
163	c32291_g1_i1	TRIVIDRAFT_71933	205.98	3.0694	6.0684	Trypsin-like serine protease

164	**c10733_g1_i1**	**CRG91957-Talaromyces islandicus**	**27.926**	**0.4272**	**6.0306**	**Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)**
165	**c17692_g1_i1**	**TRIVIDRAFT_111469**	**25.078**	**0.3901**	**6.0065**	**No hit**
166	c6198_g1_i1	TRIVIDRAFT_64982	11.694	0.1836	5.9929	Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)
167	**c5044_g1_i2**	**TRIVIDRAFT_59236**	**16.485**	**0.2599**	**5.987**	**S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)**
168	**c12606_g2_i4**	**KJK82553-Metarhizium anisopliae**	**7.0299**	**0.1119**	**5.9728**	**No hit**
169	**c1399_g1_i1**	**TRIVIDRAFT_32683**	**24.691**	**0.3941**	**5.9692**	**Glycoside hydrolase family 55**
170	**c25814_g1_i1**	**TRIVIDRAFT_223757**	**3.5023**	**0.0568**	**5.9458**	**Rossmann-fold NAD(P)(+)-binding proteins**
171	**c5507_g1_i1**	**TRIVIDRAFT_83658**	**29.508**	**0.4882**	**5.9175**	**Zinc peptidase like domain-containing protein**
172	**c22421_g1_i1**	**TRIVIDRAFT_35703**	**6.424**	**0.1065**	**5.915**	**Transferase family**
173	**c2743_g1_i1**	**TRIVIDRAFT_35498**	**0.9936**	**0.0168**	**5.8864**	**Putative polyketide synthase**
174	c13369_g1_i8	GAO88602-Aspergillus udagawae	72.841	1.275	5.8362	Probable sterigmatocystin biosynthesis P450 monooxygenase stcL
175	c32596_g1_i1	TRIVIDRAFT_127573	1.4661	0.0265	5.7888	Heme-dependent peroxidases similar to plant peroxidases
176	c32122_g1_i1	TRIVIDRAFT_39626	6.3488	0.1152	5.784	Clavaminic acid synthetase (CAS) - like
177	**c8283_g1_i3**	**TRIVIDRAFT_53581**	**25.967**	**0.4731**	**5.7784**	**Lactoylglutathione lyase, Catechol 2,3-dioxygenase/ Glyoxalase**
178	**c14014_g1_i1**	**TRIVIDRAFT_59078**	**3.1154**	**0.0572**	**5.767**	**Major Facilitator Superfamily**
179	c18193_g1_i1	TRIVIDRAFT_111830	8.4387	0.1582	5.7374	Eliciting plant response-like protein
180	**c27165_g1_i1**	**TRIVIDRAFT_119401**	**3.0937**	**0.0583**	**5.7299**	**Acetyltransferase (GNAT) family**
181	**c18000_g1_i1**	**TRIVIDRAFT_76895**	**4.4619**	**0.0862**	**5.6944**	**Glycoside hydrolase family 30 protein**
182	c23397_g1_i1	TRIVIDRAFT_78354	22.286	0.4342	5.6815	Prokaryotic phospholipase A2
183	**c31890_g1_i1**	**TRIVIDRAFT_77334**	**29.216**	**0.5921**	**5.6248**	**Peptidase S8 family domain in ProteinaseK-like proteins**
184	**c1052_g1_i1**	**QBZ65333-Magnaporthe oryzae**	**124.51**	**2.5328**	**5.6194**	**Cytochrome P450**
185	c17438_g1_i1	TRIVIDRAFT_148464	20.087	0.4225	5.5712	Mitogen activated protein kinase
186	**c483_g1_i1**	**TRIVIDRAFT_151142**	**36.451**	**0.7724**	**5.5606**	**SRPBCC super family, Bet v1-like protein**
187	**c407_g1_i1**	**TRIVIDRAFT_124167**	**2.9704**	**0.0632**	**5.5543**	**Serine/threonine protein kinase**
188	**c4934_g1_i1**	**TRIVIDRAFT_78735**	**74.147**	**1.5872**	**5.5458**	**SRPBCC super family, Bet v1-like protein, Pyrabactin resistance 1 (PYR1)-like**
189	**c9217_g1_i2**	**TRIVIDRAFT_78927**	**7.8131**	**0.1674**	**5.5441**	**Lactonase**
190	**c21353_g1_i1**	**TRIVIDRAFT_191877**	**7.4959**	**0.1629**	**5.5243**	**Glyoxalase I that uses Zn(++) as cofactor**
191	**c7434_g1_i1**	**TRIVIDRAFT_53366**	**176.1**	**3.8601**	**5.5116**	**Cytochrome P450**
192	**c5355_g1_i2**	**TRIVIDRAFT_221764**	**3.8106**	**0.0844**	**5.4966**	**No hit**
193	c3698_g1_i1	TRIVIDRAFT_48261	2.9469	0.0662	5.4759	Taurine catabolism dioxygenase TauD
194	**c19085_g1_i1**	**TRIVIDRAFT_180603**	**5.0152**	**0.1159**	**5.4356**	**Clavaminic acid synthetase**

195	c28218_g1_i1	TRIVIDRAFT_211887	2.0548	0.0477	5.4303	NAD(P)/FAD-dependent oxidoreductase
196	c32847_g1_i1	TRIVIDRAFT_170512	1.4376	0.0339	5.4055	Catalytic domain of phosphoinositide-specific phospholipase C-like phosphodiesterases superfamily
197	**c28374_g1_i1**	**TRIVIDRAFT_84937**	**29.022**	**0.6883**	**5.3979**	**M14 family of metallocarboxypeptidases and related proteins**
198	c14921_g1_i1	TRIVIDRAFT_231266	18.382	0.4454	5.367	Protein-arginine deiminase (PAD)
199	**c2200_g1_i1**	**TRIVIDRAFT_30537**	**6.5214**	**0.1588**	**5.36**	**Acyl-CoA synthetase (AMP-forming)**
200	c18975_g1_i1	TRIVIDRAFT_155482	27.904	0.6952	5.3268	Pimeloyl-ACP methyl ester carboxylesterase
201	**c29003_g1_i1**	**TRIVIDRAFT_83807**	**9.8368**	**0.2586**	**5.2497**	**NAD(P)/FAD-dependent oxidoreductase**
202	c3787_g1_i1	TRIVIDRAFT_40363	20.181	0.5307	5.2489	Beta-lactamase
203	c18356_g1_i1	TRIVIDRAFT_55235	7.9737	0.21	5.2465	Glycoside hydrolase family 16 protein
204	c3299_g1_i1	TRIVIDRAFT_112283	17.141	0.4529	5.242	Peptidase domain in the S53 family
205	**c22822_g1_i1**	**TRIVIDRAFT_10277**	**3.2632**	**0.0863**	**5.2409**	**GLEYA domain-containing protein**
206	**c18252_g1_i1**	**TRIVIDRAFT_216822**	**9.3865**	**0.2509**	**5.2253**	**No hit**
207	c5940_g1_i1	TRIVIDRAFT_45236	31.513	0.8535	5.2063	Expansin-like protein
208	c14328_g1_i1	THAR02_01813-Trichoderma harzianum	2.8521	0.0783	5.1862	Domain of unknown function (DUF4203)
209	c9754_g1_i1	TRIVIDRAFT_53690	95.78	2.6785	5.1602	Cytochrome P450
210	**c28445_g1_i1**	**TRIVIDRAFT_79838**	**14.247**	**0.4027**	**5.1449**	**WSC domain; This domain may be involved in carbohydrate binding**
211	**c8574_g1_i1**	**TRIVIDRAFT_77829**	**45.438**	**1.289**	**5.1395**	**Taurine catabolism dioxygenase TauD, TfdA family**
212	**c3520_g1_i2**	**TRIVIDRAFT_197528**	**1.8547**	**0.0541**	**5.0991**	**No hit**
213	c6254_g1_i1	TRIVIDRAFT_70175	3.3392	0.0979	5.0914	Basic leucine zipper (bZIP) domain of General control protein GCN4
214	**c28394_g1_i1**	**TRIVIDRAFT_33825**	**5.6951**	**0.1678**	**5.0845**	**L-lysine 6-monooxygenase (NADPH-requiring)**
215	c10988_g2_i1	TRIVIDRAFT_131286	9.6684	0.2891	5.0638	No hit
216	c8233_g2_i1	TRIVIDRAFT_70659	11.929	0.3578	5.0592	Glycoside hydrolase family 75 protein
217	**c808_g1_i1**	**TRIVIDRAFT_222800**	**7.5467**	**0.2271**	**5.0545**	**PP-binding and Condensation domain-containing protein**
218	**c9754_g2_i1**	**TRIVIDRAFT_151337**	**113.64**	**3.4405**	**5.0457**	**Cytochrome P450**
219	c18488_g1_i1	TRIVIDRAFT_30456	8.9468	0.274	5.0293	The third Cupredoxin domain of multicopper oxidase Fet3p
220	c42_g1_i1	TRIVIDRAFT_193863	1.9712	0.061	5.0153	Saccharomyces cerevisiae DCR2 phosphatase and related proteins,
221	**c26273_g1_i1**	**TRIVIDRAFT_27800**	**40.356**	**1.265**	**4.9955**	**No hit**
222	c33408_g1_i1	TRIVIDRAFT_66077	7.3085	0.2299	4.9904	Sugar (and other) transporter
223	**c24770_g1_i1**	**TRIVIDRAFT_111866**	**16.994**	**0.5477**	**4.9555**	**Glycoside hydrolase family 18**
224	**c6320_g1_i1**	**TRIVIDRAFT_111446**	**5.4655**	**0.1772**	**4.9471**	**No hit**
225	**c19044_g1_i1**	**TRIVIDRAFT_230790**	**364.26**	**11.895**	**4.9365**	**Cytochrome P450**
226	**c32849_g1_i1**	**TRIVIDRAFT_216106**	**66.956**	**2.2593**	**4.8893**	**Blastomyces yeast-phase-specific protein (Bys1)**

227	**c3606_g1_i1**	**TRIVIDRAFT_67390**	**7.1941**	**0.2436**	**4.8845**	**GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain**
228	**c6239_g1_i1**	**TRIVIDRAFT_71600**	**2.7317**	**0.0939**	**4.8619**	**Glycoside hydrolase family 92 protein**
229	**c12649_g1_i1**	**TRIVIDRAFT_193258**	**14.697**	**0.5171**	**4.8289**	**Berberine and berberine like**
230	**c14882_g1_i1**	**TRIVIDRAFT_140278**	**29.738**	**1.0504**	**4.8232**	**Catalase/hydroperoxidase HPI(I)**
231	**c14151_g1_i1**	**TRIVIDRAFT_214993**	**12.196**	**0.434**	**4.8124**	**DOMON-like type 9 carbohydrate binding module**
232	**c29773_g1_i1**	**TRIVIDRAFT_52915**	**3.1189**	**0.1119**	**4.8003**	**Isoprenoid Biosynthesis enzymes, Terpene Cyclase**
233	**c7590_g1_i3**	**TRIVIDRAFT_53582**	**52.016**	**1.8906**	**4.782**	**NAD(P)/FAD-dependent oxidoreductase, Pyridine nucleotide-disulphide oxidoreductase**
234	**c14950_g1_i1**	**TRIVIDRAFT_60010**	**40.091**	**1.4774**	**4.7621**	**Saccharomyces cerevisiae Azole resistance protein 1 (Azr1p)**
235	**c18593_g1_i1**	**TRIVIDRAFT_228094**	**4.8575**	**0.1796**	**4.7574**	**Polyamine oxidase**
236	**c14667_g1_i1**	**TRIVIDRAFT_53375**	**332.98**	**12.386**	**4.7487**	**Cytochrome P450**
237	**c18224_g1_i1**	**TRIVIDRAFT_52906**	**10.544**	**0.3966**	**4.7326**	**Protein Kinases, catalytic domain**
238	c11464_g1_i1	TRIVIDRAFT_230590	1.0489	0.0423	4.6304	Heterokaryon incompatibility protein (HET)
239	**c5024_g1_i1**	**TRIVIDRAFT_209187**	**2.5723**	**0.1046**	**4.6207**	**No hit**
240	**c13171_g1_i5**	**TRIVIDRAFT_54033**	**2.228**	**0.0911**	**4.6125**	**Glycoside hydrolase family 2 protein**
241	c25154_g1_i1	TRIVIDRAFT_49743	2.7168	0.1112	4.6109	Major Facilitator Superfamily
242	c21484_g1_i1	TRIVIDRAFT_217363	1.6544	0.068	4.6044	Fungal specific transcription factor domain
243	c15089_g1_i1	TRIVIDRAFT_231370	3.5658	0.1466	4.6044	Laminin G domain
244	**c28619_g1_i1**	**TRIVIDRAFT_39958**	**1.4031**	**0.0586**	**4.5823**	**No hit**
245	**c7152_g1_i1**	**TRIVIDRAFT_215894**	**44.018**	**1.878**	**4.5508**	**No hit**
246	**c12503_g2_i2**	**TRIVIDRAFT_126156**	**4.4724**	**0.1942**	**4.5251**	**No hit**
247	c2588_g1_i1	TRIVIDRAFT_70971	64.627	2.8583	4.4989	SRPBCC superfamily, includes Bet v1-like protein
248	**c2904_g1_i1**	**TRIVIDRAFT_64335**	**26.583**	**1.1876**	**4.4844**	**Major Facilitator Superfamily**
249	c10107_g1_i1	TRIVIDRAFT_28610	2.5389	0.1148	4.4669	Putative PKS-NRPS protein
250	c17899_g1_i1	TRIVIDRAFT_34210	5.3289	0.2426	4.457	Major Facilitator Superfamily
251	**c21433_g1_i1**	**TRIVIDRAFT_89880**	**6.1674**	**0.2874**	**4.4236**	**Putative sialic acid transporter**
252	**c31887_g1_i1**	**TRIVIDRAFT_71936**	**62.399**	**2.9177**	**4.4186**	**Domain in Tre-2, BUB2p, and Cdc16p**
253	c22774_g1_i1	KFX52676-Talaromyces marneffei	9.9532	0.4762	4.3856	Ankyrin-3
254	**c102_g1_i1**	**TRIVIDRAFT_190045**	**3.7243**	**0.1825**	**4.3514**	**Cytochrome P450**
255	**c21076_g1_i1**	**TRIVIDRAFT_217322**	**149.73**	**7.3924**	**4.3401**	**Glyceraldehyde-3-phosphate dehydrogenase**
256	c28343_g1_i1	TRIVIDRAFT_45995	2.3382	0.1156	4.3376	Major Facilitator Superfamily
257	**c3720_g1_i1**	**TRIVIDRAFT_160041**	**2.8689**	**0.1429**	**4.3271**	**Amidase**
258	**c11151_g1_i1**	**TRIVIDRAFT_230947**	**53.173**	**2.6638**	**4.3191**	**Peptidases like domain-containing protein**
259	c22336_g1_i1	OPB37584-Trichoderma guizhouense	3.7631	0.1891	4.3149	NAD dependent epimerase/dehydratase

260	c26041_g1_i1	TRIVIDRAFT_46158	4.2202	0.212	4.3149	Short-chain dehydrogenases/reductases (SDR)
261	c28509_g1_i1	TRIVIDRAFT_190875	1.5229	0.0771	4.3048	OPT oligopeptide transporter protein
262	**c14167_g1_i1**	**TRIVIDRAFT_213573**	**5.2901**	**0.2686**	**4.2995**	**Gamma-glutamyltranspeptidase**
263	**c7917_g1_i1**	**TRIVIDRAFT_151341**	**41.453**	**2.1196**	**4.2896**	**Alpha/beta hydrolase fold**
264	c11711_g1_i1	TRIVIDRAFT_18950	14.584	0.7544	4.273	No hit
265	**c28935_g1_i1**	**TRIVIDRAFT_53640**	**41.365**	**2.1535**	**4.2636**	**FAD/FMN-containing dehydrogenase**
266	c7255_g1_i2	TRIVIDRAFT_135901	1.9197	0.1021	4.2324	Ribonuclease T2 (RNase T2)
267	**c24759_g1_i1**	**TRIVIDRAFT_111479**	**9.4411**	**0.5044**	**4.2263**	**Glycoside hydrolase family 75 protein**
268	c28654_g1_i1	TRIVIDRAFT_54057	2.7753	0.1506	4.2038	Short-chain dehydrogenases/reductases (SDR)
269	c4859_g1_i1	EGE05473-Arthroderma benhamiae	2.3027	0.1269	4.1816	polyketide synthase
270	c7635_g1_i3	TRIVIDRAFT_52801	2.0136	0.1115	4.1747	Major Facilitator Superfamily
271	**c11156_g1_i2**	**TRIVIDRAFT_47297**	**7.9151**	**0.4389**	**4.1726**	**Domain of unknown function (DUF3292)**
272	**c5305_g1_i1**	**TRIVIDRAFT_52335**	**4.2593**	**0.2386**	**4.1578**	**Glycoside hydrolase family 18**
273	**c33071_g1_i1**	**TRIVIDRAFT_205388**	**20.377**	**1.1715**	**4.1205**	**No hit**
274	c2187_g1_i1	TRIVIDRAFT_61110	2.5071	0.1443	4.1189	No hit
275	**c5975_g1_i1**	**TRIVIDRAFT_63956**	**11.078**	**0.6654**	**4.0573**	**Peptidase M35 family**
276	c7119_g1_i1	TRIVIDRAFT_112017	2.7923	0.1684	4.0518	No hit
277	**c10607_g1_i2**	**TRIVIDRAFT_181642**	**14.308**	**0.8688**	**4.0416**	**Glycoside hydrolase family 2 protein**
278	**c9960_g1_i3**	**TRIVIDRAFT_70229**	**6.8146**	**0.4164**	**4.0325**	**Major Facilitator Superfamily**
279	**c8085_g1_i1**	**TRIVIDRAFT_215055**	**4.8451**	**0.2969**	**4.0284**	**Sulfite oxidase**
280	c22549_g1_i1	TRIVIDRAFT_49075	3.2592	0.2016	4.0153	Ctr copper transporter family
281	**c12608_g1_i1**	**TRIVIDRAFT_69939**	**10.566**	**0.6538**	**4.0144**	**Peptidase S8 family domain**
282	**c14505_g1_i1**	**TRIVIDRAFT_183528**	**10.653**	**0.6694**	**3.9923**	**Gametolysin peptidase M11**
283	c10921_g1_i2	KEY73628-Stachybotrys chartarum	2.8552	0.1798	3.9889	Cytochrome P450
284	c10531_g1_i2	TRIVIDRAFT_221996	10.904	0.6956	3.9704	GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain
285	**c1964_g1_i1**	**TRIVIDRAFT_156071**	**3.2387**	**0.208**	**3.9609**	**ATPase family associated with various cellular activities (AAA)**
286	c1519_g1_i1	TRIVIDRAFT_76316	18.651	1.1985	3.9599	Amino acid permease
287	**c11451_g3_i1**	**TRIVIDRAFT_73257**	**42.871**	**2.7747**	**3.9496**	**Heme-dependent peroxidases similar to plant peroxidases**
288	c28441_g1_i1	TRIVIDRAFT_192166	41.355	2.7352	3.9184	Solute carrier families 5 and 6-like
289	c1558_g1_i1	TRIVIDRAFT_110580	10.421	0.7009	3.8942	Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases
290	**c24836_g1_i1**	**TRIVIDRAFT_73911**	**8.2506**	**0.5552**	**3.8933**	**No hit**
291	**c28172_g1_i1**	**TRIVIDRAFT_57302**	**3.895**	**0.2623**	**3.8922**	**S-adenosylmethionine-dependent methyltransferases**
292	**c16055_g1_i1**	**TRIVIDRAFT_74115**	**11.676**	**0.7864**	**3.8922**	**Reactive intermediate/imine deaminase**
293	**c8990_g1_i2**	**TRIVIDRAFT_49**	**4.7342**	**0.3249**	**3.8648**	**Salmonella virulence plasmid 65kDa B protein**
294	**c14949_g1_i1**	**TRIVIDRAFT_75875**	**11.714**	**0.815**	**3.8454**	**NAD(P)+-dependent aldehyde dehydrogenase superfamily**

295	**c24994_g1_i1**	**TRIVIDRAFT_87376**	**8.4922**	**0.5939**	**3.8378**	**Allantoate permease family MFS transporter**
296	**c31870_g1_i1**	**TRIVIDRAFT_204543**	**10.096**	**0.7102**	**3.8294**	**No hit**
297	**c5893_g1_i2**	**TRIVIDRAFT_58307**	**10.749**	**0.7689**	**3.8052**	**Peroxidase**
298	**c8944_g1_i1**	**TRIVIDRAFT_79145**	**112.09**	**8.05**	**3.7995**	**Predicted oxidoreductase**
299	c462_g1_i1	NP_570149-Trichoderma reesei	7.5722	0.5479	3.7888	NADH-ubiquinone oxidoreductase subunit 2 (mitochondrion)
300	c34712_g1_i1	TRIVIDRAFT_40370	22.019	1.6003	3.7824	Putative D-isomer specific 2-hydroxyacid dehydrogenases
301	**c22308_g1_i1**	**TRIVIDRAFT_215947**	**24.086**	**1.7727**	**3.7642**	**No hit**
302	**c14150_g1_i1**	**TRIVIDRAFT_209800**	**4.5445**	**0.3455**	**3.7172**	**Short-chain dehydrogenases/reductases (SDR)**
303	**c32463_g1_i1**	**TRIVIDRAFT_67871**	**20.009**	**1.5592**	**3.6818**	**FMN-dependent oxidoreductase**
304	**c18322_g1_i1**	**TRIVIDRAFT_60255**	**7.3696**	**0.5764**	**3.6763**	**No hit**
305	**c1433_g1_i1**	**TRIVIDRAFT_45299**	**16.27**	**1.2908**	**3.6559**	**Aldo/keto reductase**
306	**c12248_g1_i1**	**TRIVIDRAFT_118500**	**5.9952**	**0.4761**	**3.6546**	**Zinc finger, C2H2 type**
307	**c5388_g1_i1**	**TRIVIDRAFT_188402**	**5.2878**	**0.4208**	**3.6513**	**Glycoside hydrolase family 3 protein**
308	c32917_g1_i1	TRIVIDRAFT_141200	6.6304	0.5301	3.6448	Haem-degrading
309	c15146_g1_i1	TRIVIDRAFT_59417	5.9342	0.4811	3.6247	Methyltransferase domain
310	c7996_g1_i1	XP_002483852-Talaromyces stipitatus	8.4339	0.6884	3.6149	Cytochrome P450
311	c2002_g1_i1	TRIVIDRAFT_64185	2.4716	0.2021	3.6125	Major Facilitator Superfamily
312	c8019_g1_i2	TRIVIDRAFT_57186	7.5392	0.6282	3.5851	No hit
313	**c4698_g1_i1**	**TRIVIDRAFT_212656**	**7.4035**	**0.6334**	**3.547**	**Major Facilitator Superfamily**
314	**c25858_g1_i1**	**TRIVIDRAFT_157800**	**2.6276**	**0.2251**	**3.5449**	**Glycoside hydrolase family 47 protein**
315	**c9428_g1_i1**	**TRIVIDRAFT_170440**	**39.858**	**3.4401**	**3.5343**	**FAD-dependent oxidoreductase**
316	c4594_g1_i2	TRIVIDRAFT_132754	2.0595	0.1805	3.5121	MHYT domain, NO-binding membrane sensor
317	c6338_g1_i1	TRIVIDRAFT_140013	6.0959	0.5379	3.5025	CFEM domain
318	**c32183_g1_i1**	**TRIVIDRAFT_55059**	**6.3478**	**0.57**	**3.4773**	**Domain of unknown function (DUF4965)**
319	c6749_g1_i1	TRIVIDRAFT_183616	9.3108	0.8407	3.4692	L-type amino acid transporter
320	**c9061_g1_i2**	**TRIVIDRAFT_15465**	**11.649**	**1.063**	**3.4539**	**No hit**
321	c7787_g1_i1	TRIVIDRAFT_128161	108.95	9.9691	3.45	FAD/FMN-containing dehydrogenase
322	c12941_g1_i4	TRIVIDRAFT_176452	16.775	1.5421	3.4433	Fungal transcription factor regulatory middle homology region
323	**c12633_g1_i2**	**TRIVIDRAFT_216052**	**25.637**	**2.358**	**3.4426**	**No hit**
324	**c9987_g1_i1**	**TRIVIDRAFT_141245**	**9.0433**	**0.848**	**3.4147**	**Acetamidase/Formamidase family**
325	c5387_g1_i2	TRIVIDRAFT_66021	4.4319	0.4161	3.413	Homeobox KN domain
326	c2390_g1_i1	M441DRAFT_299104-Trichoderma asperellum	4.4853	0.4229	3.4069	No hit
327	c13507_g1_i1	TRIVIDRAFT_49765	12.643	1.1965	3.4014	Putative Pfs NB-ARC and ankyrin-domain-containing protein
328	c3631_g1_i2	TRIVIDRAFT_29732	4.7843	0.4584	3.3837	Peptidase S8 family domain,
329	c10929_g1_i1	TRIVIDRAFT_76620	8.0452	0.7715	3.3824	Predicted extracellular nuclease
330	**c8385_g3_i1**	**TRIVIDRAFT_58016**	**6.9942**	**0.6714**	**3.381**	**Major Facilitator Superfamily**
331	c7597_g1_i1	TRIVIDRAFT_114530	16.74	1.6182	3.3708	NDT80 / PhoG like DNA-binding family

332	**c2641_g1_i1**	**TRIVIDRAFT_78296**	**12.73**	**1.2744**	**3.3203**	**Domain of unknown function (DUF3455)**
333	**c11509_g1_i2**	**TRIVIDRAFT_167449**	**7.5274**	**0.7689**	**3.2914**	**Protease associated domain**
334	**c11578_g1_i1**	**TRIVIDRAFT_58857**	**5.7022**	**0.5859**	**3.2829**	**Cytochrome P450**
335	**c1136_g1_i1**	**TRIVIDRAFT_179276**	**14.31**	**1.4781**	**3.2752**	**Neutral/alkaline ceramidase**
336	c5927_g1_i2	TRIVIDRAFT_179700	31.061	3.2237	3.2683	PCI domain
337	**c9070_g1_i1**	**TRIVIDRAFT_54504**	**2.2436**	**0.2343**	**3.2593**	**Multidrug resistance protein (mdr1)**
338	c2570_g1_i2	TRIVIDRAFT_35658	6.4425	0.6862	3.2309	Acid sphingomyelinase and related proteins
339	c24653_g1_i1	TRIVIDRAFT_76832	5.7428	0.6215	3.208	Glycoside hydrolase family 71 protein
340	**c2922_g1_i1**	**TRIVIDRAFT_52839**	**17.068**	**1.8502**	**3.2056**	**Glutathione S-transferase**
341	**c32281_g1_i1**	**TRIVIDRAFT_194303**	**8.321**	**0.9025**	**3.2047**	**Cytochrome P450**
342	c27758_g1_i1	TGAMA5MH_00705-Trichoderma gamsii	20.857	2.2719	3.1986	Conidiation protein 6
343	**c5079_g1_i1**	**TRIVIDRAFT_200088**	**4.438**	**0.4945**	**3.1658**	**L-type amino acid transporter**
344	c28290_g1_i1	TRIVIDRAFT_112183	3.1887	0.3576	3.1565	Metal ion transporter CorA-like divalent cation transporter superfamily
345	c1838_g1_i1	TRIVIDRAFT_17045	8.6744	0.9904	3.1307	No hit
346	c11518_g1_i1	TRIVIDRAFT_210053	3.3139	0.3815	3.1186	Multidrug resistance protein
347	**c31916_g1_i1**	**TRIVIDRAFT_39211**	**19.29**	**2.2292**	**3.1132**	**Alpha/beta hydrolases**
348	**c14811_g1_i1**	**TRIVIDRAFT_192636**	**19.203**	**2.242**	**3.0985**	**Domain of unknown function (DUF2235)**
349	**c9685_g1_i1**	**TRIVIDRAFT_53518**	**2.5901**	**0.3061**	**3.0807**	**Acyl transferase domain in polyketide synthase (PKS) enzymes**
350	**c4413_g1_i1**	**TRIVIDRAFT_69335**	**10.96**	**1.3059**	**3.0691**	**Amino acid permease (GABA permease)**
351	c7462_g1_i1	TRIVIDRAFT_78530	3.5569	0.4296	3.0494	Glycoside hydrolase family 47 protein
352	c21228_g1_i1	TRIVIDRAFT_164079	2.6573	0.3277	3.0194	Adenylate forming domain
353	c8331_g1_i2	TRIVIDRAFT_171879	4.9344	0.6117	3.0121	Clavaminic acid synthetase (CAS) - like
354	**c29219_g1_i1**	**TRIVIDRAFT_83099**	**9.8505**	**1.2214**	**3.0117**	**Short-chain dehydrogenases/reductases (SDR)**
355	c21240_g1_i1	TRIVIDRAFT_90260	12.032	1.5036	3.0004	Aldo/keto reductase
356	**c6429_g1_i1**	**TRIVIDRAFT_127080**	**7.6049**	**0.9561**	**2.9917**	**Fungal specific transcription factor domain**
357	c10917_g1_i1	TRIVIDRAFT_193582	5.3939	0.6799	2.9878	Cytochrome P450
358	c32131_g1_i1	TRIVIDRAFT_150924	4.6365	0.6007	2.9483	Fungal transcription factor regulatory middle homology region
359	c5720_g1_i1	TRIVIDRAFT_229897	5.121	0.6676	2.9394	No hit
360	c25051_g1_i1	THARTR1_07001-Trichoderma harzianum	10.828	1.4349	2.9158	F-box-like
361	c883_g1_i2	TRIVIDRAFT_211568	15.935	2.1227	2.9082	Sugar (and other) transporter
362	c5827_g1_i1	NP_570148-Trichoderma reesei	35.038	4.6748	2.9059	Ribosomal protein S5 (mitochondrion)
363	c11371_g1_i1	TRIVIDRAFT_133587	4.1937	0.5611	2.902	Fungal transcription factor regulatory middle homology region
364	c12724_g1_i1	TRIVIDRAFT_49762	6.5539	0.882	2.8935	Glycoside hydrolase family 20 protein
365	c1412_g1_i2	TRIVIDRAFT_74688	15.267	2.0648	2.8863	Glycoside hydrolase family 3 protein

366	c17725_g1_i1	TRIVIDRAFT_82631	21.203	2.8726	2.8838	Phospholipid methyltransferase
367	**c28703_g1_i1**	**TRIVIDRAFT_61863**	**5.8552**	**0.7983**	**2.8747**	**Glycosyltransferase family 2 protein**
368	**c3708_g1_i1**	**TRIVIDRAFT_71258**	**29.598**	**4.0982**	**2.8524**	**GAL4-like Zn(2)Cys(6) binuclear cluster DNA-binding domain**
369	c3011_g1_i1	TRIVIDRAFT_62772	8.3337	1.1757	2.8254	Cation transport protein
370	c534_g1_i1	TRIVIDRAFT_90038	16.663	2.355	2.8229	Methylenetetrahydrofolate reductase (MTHFR)
371	**c7046_g1_i1**	**TRIVIDRAFT_55803**	**31.889**	**4.5233**	**2.8176**	**Linoleate (8R)-dioxygenase and related enzymes**
372	c7981_g1_i1	TRIVIDRAFT_83043	34.498	4.9988	2.7869	Glycoside hydrolase family 16 protein
373	**c6337_g1_i1**	**TRIVIDRAFT_83973**	**16.944**	**2.4563**	**2.7862**	**Aminoglycoside 3'-phosphotransferase and Choline Kinase family**
374	c29145_g1_i1	TRIVIDRAFT_72521	46.09	6.7341	2.7749	Acetyl-CoA C-acetyltransferase
375	c10541_g1_i1	TRIVIDRAFT_168035	20.989	3.1078	2.7557	No hit

Table 4A. Single nucleotide polymorphisms (SNPs) present in M7 genome

S.no.	Chromosome ID	SNP Position	Reference base	Altered base	Read Depth	Quality	TYPE
1	Scaffold_1	297880	T	G	25	24	Transversion
2	Scaffold_1	384597	A	T	17	24	Transversion
3	Scaffold_1	1E+06	A	C	23	25	Transversion
4	Scaffold_1	1E+06	C	T	53	24	Transition
5	Scaffold_1	2E+06	T	G	39	24	Transversion
6	Scaffold_1	3E+06	T	A	22	21	Transversion
7	Scaffold_1	3E+06	T	C	46	23	Transition
8	Scaffold_1	3E+06	T	G	21	25	Transversion
9	Scaffold_1	3E+06	A	C	15	25	Transversion
10	Scaffold_1	3E+06	C	G	33	24	Transversion
11	Scaffold_1	3E+06	C	A	22	25	Transversion
12	Scaffold_1	5E+06	T	G	36	25	Transversion
13	Scaffold_1	5E+06	C	A	45	24	Transversion
14	Scaffold_1	5E+06	T	G	55	25	Transversion
15	Scaffold_1	5E+06	T	G	33	25	Transversion
16	Scaffold_1	5E+06	T	A	40	23	Transversion
17	Scaffold_1	5E+06	A	C	52	25	Transversion
18	Scaffold_1	5E+06	T	G	22	17	Transversion
19	Scaffold_1	6E+06	T	A	58	24	Transversion
20	Scaffold_1	6E+06	T	G	43	24	Transversion
21	Scaffold_1	6E+06	T	G	41	25	Transversion
22	Scaffold_1	6E+06	A	T	35	21	Transversion
23	Scaffold_1	6E+06	C	T	52	25	Transition
24	Scaffold_1	7E+06	A	G	45	24	Transition
25	Scaffold_1	7E+06	G	C	86	24	Transversion
26	Scaffold_1	7E+06	T	C	47	25	Transition
27	Scaffold_1	7E+06	A	C	34	25	Transversion
28	Scaffold_1	8E+06	T	G	29	24	Transversion
29	Scaffold_1	8E+06	G	T	23	19	Transversion
30	Scaffold_1	8E+06	T	A	41	25	Transversion
31	Scaffold_1	9E+06	C	A	17	24	Transversion
32	Scaffold_1	1E+07	C	T	56	24	Transition
33	Scaffold_1	1E+07	G	T	49	24	Transversion
34	Scaffold_1	1E+07	C	T	50	22	Transition
35	Scaffold_1	1E+07	A	C	46	24	Transversion
36	Scaffold_1	1E+07	T	A	48	24	Transversion
37	Scaffold_1	1E+07	G	A	46	24	Transition
38	Scaffold_1	1E+07	A	T	46	24	Transversion
39	Scaffold_1	1E+07	A	C	37	24	Transversion
40	Scaffold_1	1E+07	A	G	44	24	Transition
41	Scaffold_1	1E+07	A	T	19	17	Transversion
42	Scaffold_1	1E+07	G	T	35	24	Transversion

43	Scaffold_1	1E+07	T	A	79	23	Transversion
44	Scaffold_1	1E+07	T	G	24	25	Transversion
45	Scaffold_1	1E+07	A	T	16	25	Transversion
46	Scaffold_1	1E+07	G	C	58	25	Transversion
47	Scaffold_1	1E+07	A	T	33	25	Transversion
48	Scaffold_1	1E+07	T	A	20	24	Transversion
49	Scaffold_1	1E+07	T	G	20	23	Transversion
50	Scaffold_1	1E+07	T	C	40	24	Transition
51	Scaffold_1	1E+07	A	C	53	25	Transversion
52	Scaffold_1	1E+07	C	T,A	54	24	Heterozygous
53	Scaffold_1	2E+07	T	G	51	25	Transversion
54	Scaffold_1	2E+07	G	C	44	24	Transversion
55	Scaffold_1	2E+07	A	C	31	25	Transversion
56	Scaffold_1	2E+07	T	G	35	25	Transversion
57	Scaffold_1	2E+07	T	C	38	25	Transition
58	Scaffold_1	2E+07	T	G	30	25	Transversion
59	Scaffold_1	2E+07	A	T	42	25	Transversion
60	Scaffold_1	2E+07	G	A	39	23	Transition
61	Scaffold_1	2E+07	A	C	52	25	Transversion
62	Scaffold_3	63549	A	T	41	24	Transversion
63	Scaffold_12	218885	A	C	27	24	Transversion
64	Scaffold_12	284685	A	C	48	24	Transversion
65	Scaffold_12	704898	G	C	23	23	Transversion
66	Scaffold_15	117051	G	T	59	24	Transversion
67	Scaffold_20	369820	C	A	63	25	Transversion
68	Scaffold_20	401671	T	A	40	24	Transversion
69	Scaffold_20	697618	T	G	39	23	Transversion
70	Scaffold_20	726338	A	T	26	24	Transversion
71	Scaffold_22	52499	A	T	36	23	Transversion
72	Scaffold_25	42910	C	G	52	24	Transversion
73	Scaffold_28	42653	A	C	40	25	Transversion
74	Scaffold_30	459476	A	C	57	25	Transversion
75	Scaffold_31	265466	C	A	19	22	Transversion
76	Scaffold_32	355065	C	A	47	25	Transversion
77	Scaffold_33	30625	T	A	23	24	Transversion
78	Scaffold_34	113780	A	C	61	25	Transversion
79	Scaffold_50	201613	A	T	31	23	Transversion
80	Scaffold_50	229364	A	C	32	25	Transversion
81	Scaffold_50	243204	C	T	46	24	Transition
82	Scaffold_50	416885	C	A	50	25	Transversion
83	Scaffold_50	444356	C	A	41	24	Transversion
84	Scaffold_50	444479	G	A	40	24	Transition
85	Scaffold_51	331438	A	C	36	25	Transversion
86	Scaffold_51	637602	A	G	43	25	Transition
87	Scaffold_53	195149	G	T,A	48	24	Heterozygous
88	Scaffold_53	333095	G	T	23	24	Transversion

89	Scaffold_54	443704	T	G	49	25	Transversion
90	Scaffold_54	474850	A	C	58	25	Transversion
91	Scaffold_54	523327	A	T	25	24	Transversion
92	Scaffold_54	530285	A	T	29	24	Transversion
93	Scaffold_54	530657	A	T	65	25	Transversion
94	Scaffold_54	533083	A	T	19	22	Transversion
95	Scaffold_55	235371	T	G	17	24	Transversion
96	Scaffold_55	247345	T	A	25	24	Transversion
97	Scaffold_56	210349	C	A	28	22	Transversion
98	Scaffold_56	215495	T	G	31	24	Transversion
99	Scaffold_56	327943	A	C	57	25	Transversion
100	Scaffold_57	423338	T	G	59	24	Transversion
101	Scaffold_57	788227	A	C	42	24	Transversion
102	Scaffold_57	820674	A	C	52	24	Transversion
103	Scaffold_61	101993	A	C	23	25	Transversion
104	Scaffold_64	337599	C	A	20	25	Transversion
105	Scaffold_71	1580	C	A	35	23	Transversion
106	Scaffold_71	1706	A	T	43	25	Transversion
107	Scaffold_71	1905	T	A	38	24	Transversion
108	Scaffold_72	118760	T	C	34	25	Transition
109	Scaffold_77	203189	A	C	15	23	Transversion
110	Scaffold_77	221269	A	G	42	23	Transition
111	Scaffold_77	290725	G	T	41	24	Transversion
112	Scaffold_78	417690	T	A	25	24	Transversion
113	Scaffold_86	138154	T	G	36	24	Transversion
114	Scaffold_86	193675	T	G	44	25	Transversion
115	Scaffold_90	154284	G	T	21	24	Transversion
116	Scaffold_99	78220	A	C	22	24	Transversion
117	Scaffold_99	91482	A	T	50	24	Transversion
118	Scaffold_99	97365	T	G	37	24	Transversion

Table 4B. Insertion/deletion polymorphisms (INDELs) present in M7 genome

S. no	Chromosome ID	INDEL Position	TYPE
1	Scaffold_1	518667	Insertion
2	Scaffold_1	518878	Deletion
3	Scaffold_1	1851882	Insertion
4	Scaffold_1	3429826	Insertion
5	Scaffold_1	4523295	Deletion
6	Scaffold_1	4732111	Insertion
7	Scaffold_1	4860534	Insertion
8	Scaffold_1	5681116	Deletion
9	Scaffold_1	5789115	Insertion
10	Scaffold_1	6496649	Insertion
11	Scaffold_1	7133533	Insertion
12	Scaffold_1	7817667	Insertion
13	Scaffold_1	8200132	Deletion
14	Scaffold_1	9338237	Insertion
15	Scaffold_1	9557544	Insertion
16	Scaffold_1	10402062	Deletion
17	Scaffold_1	11404036	Insertion
18	Scaffold_1	11697643	Insertion
19	Scaffold_1	11704938	Insertion
20	Scaffold_1	11897398	Insertion
21	Scaffold_1	12494983	Insertion
22	Scaffold_1	12523380	Insertion
23	Scaffold_1	13036146	Insertion
24	Scaffold_1	13366202	Deletion
25	Scaffold_1	14262366	Insertion
26	Scaffold_1	14538829	Insertion
27	Scaffold_1	14543651	Insertion
28	Scaffold_1	14616664	Insertion
29	Scaffold_1	14664600	Deletion
30	Scaffold_1	14762395	Insertion
31	Scaffold_1	15250500	Deletion
32	Scaffold_1	16136314	Insertion
33	Scaffold_3	149519	Insertion
34	Scaffold_5	232738	Insertion
35	Scaffold_12	838116	Deletion
36	Scaffold_22	48387	Deletion
37	Scaffold_32	273615	Deletion
38	Scaffold_33	2300	Insertion
39	Scaffold_34	79857	Insertion
40	Scaffold_44	22566	Insertion
41	Scaffold_45	110643	Insertion
42	Scaffold_45	230705	Insertion
43	Scaffold_54	532450	Insertion
44	Scaffold_54	532917	Deletion
45	Scaffold_54	533203	Deletion
46	Scaffold_54	533360	Insertion
47	Scaffold_57	460907	Insertion
48	Scaffold_57	521295	Insertion
49	Scaffold_57	720964	Deletion
50	Scaffold_57	819221	Insertion

51	Scaffold_64	264462	Insertion
52	Scaffold_71	3033	Deletion
53	Scaffold_71	3134	Deletion
54	Scaffold_86	168603	Deletion
55	Scaffold_86	179823	Insertion
56	Scaffold_90	67908	Deletion
57	Scaffold_99	154993	Insertion

Summary

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is an enzyme converting D-glyceraldehyde 3-phosphate (G3P) into 3-phospho-D-glyceroyl phosphate in glycolytic pathway. Some fungi harbours an additional gene for GAPDH in their genome, functional characterization of which remains elusive so far. Using gene knockout approach, we have been able to establish the role of a *Trichoderma virens* GAPDH isoform (vGPD), which is associated with a secondary metabolism-related gene cluster (*vir* cluster) involved in the biosynthesis of volatile sesquiterpene metabolites. The phenotypic characterization of the vGPD deletion mutant showed that it has radial growth, colony morphology and sporulation similar to the wild type *T. virens*. However, GC-MS analyses of the vGPD deletion mutant showed that it is unable to produce volatile sesquiterpene metabolites. TLC and LC-MS profile of non-volatile secondary metabolites showed that the vGPD deletion mutant was also unable to biosynthesize non-volatile compounds like viridin, viridiol and heptelidic acid, establishing a global role of GAPDH in biosynthesis of volatile and non-volatile secondary metabolites. The *vir* cluster was discovered by using suppression subtractive hybridization (SSH) technique. Three genes of the *vir* cluster, i.e., *vir2*, *vir3* and *vir4* were under-expressed in M7 while other four genes showed no change in expression (Mukherjee et al., 2006). While in our findings in the present study, the real-time PCR analysis of *vir* cluster associated genes showed that two genes (*vir4* encoding a terpene cyclase and *vir3* for a cytochrome P450) were upregulated in the vGPD deletion mutant, while the other five genes in the cluster were downregulated. Earlier, the expression of vGPD was not studied in M7, nor it was assessed whether M7 biosynthesizes volatile sesquiterpenes. In order to understand the regulation of *vir* cluster in *T. virens*, we undertook a detailed molecular characterization of the M7 mutant. TLC and GC-MS analysis showed that M7 was deficient in producing both non-volatile and volatile secondary metabolites, similar to the vGPD deletion mutant. Additional characterization of the M7 mutant revealed that the mutant colony was deficient in

hydrophobicity and mycoparasitism against plant pathogens. RNAseq analysis of the M7 transcriptome also provided evidence for the downregulation of many genes associated with mycoparasitism (like chitinase, chitosanase and beta-glucanases), secondary metabolite biosynthesis, regulation, hydrophobicity and conidiation. The whole genome sequencing of M7 mutant identified five deletions in the mutant genome, totaling about 250 kb (encompassing 71 predicted ORFs). The study of M7 genome using RNAseq and whole genome sequencing provides novel insight into the genetics of morphogenesis, secondary metabolism and mycoparasitism and could lead to the identification of novel gene regulators in *Trichoderma* spp. We also performed cloning, expression, purification and characterization of the vGPD protein of *T. virens*. The *vGPD* gene was cloned in the pNH-TrxT vector (with poly-histidine tag and a thioredoxin tag), expressed in *E. coli* and was purified using affinity chromatography. The vGPD protein displays biochemical and biophysical characteristics similar to the gGPD isoform. It exists as a tetramer with *T*m of about 56.5 ºC, and displays phosphorylation activity with *K*m and *K*cat of 0.38 mM and 2.55 sec^{-1}, respectively. In order to determine the regulatory role of vGPD protein as a transcription factor in the *vir* cluster, EMSA was performed with vGPD protein and promoter region of the *vir4* gene that codes for the core enzyme (a terpene cyclase) of the "*vir*" cluster. The EMSA analysis showed that vGPD may not act as a transcription factor driving the "*vir*" cluster, at least not by directly binding to the promoter region. We also succeeded in obtaining small crystals of this protein. We have constructed reliable structural models of vGPD and gGPD of *T. virens* and were statistically validated. The *in silico* constrained docking analysis was performed between heptelidic acid as an inhibitor, and vGPD and gGPD as structural models. The analysis reveals that heptelidic acid shows weaker binding with vGPD as compared to gGPD. Moreover, a conserved indel is specific to vGPD protein, both in *T. virens* and *Aspergillus* spp., but the effect of this indel in vGPD cannot be completely decoded yet.

List of figures:

List of Tables:

List of abbreviations:

LB	Luria Broth	EDTA	Ethylene diamine tetra acetic
PDB	Potato dextrose broth	dsDNA	Double stranded DNA
FISH	Florescence in-situ hybridization		
PCR	Polymerase chain reaction	MSA	Multiple sequence alignment

NBT/ BCIP	Nitro blue tetrazolium / 5-bromo-4-chloro-3-indolyl phosphate	W/V	Weight/Volume
RT	Room Temperature	Ni-NTA	Nickel-nitrilo acetic acid

Highlight of the thesis

Thesis title: Structural and functional analysis of a *Trichoderma virens* GAPDH associated with a secondary-metabolism related gene cluster

1. The present work provides insight into the structural and functional analysis of a glycolytic protein - GAPDH in the secondary metabolism of filamentous fungi *Trichoderma virens*.
2. Both non-volatile secondary metabolites and volatile secondary metabolites showed that deletion of vGPD gene affected the biosynthesis of the secondary metabolites in *Trichoderma virens*.
3. Both M7 mutant (gamma radiation-induced mutant) and vGPD deletion mutant possess similar secondary metabolism biosynthesis phenotype, thus detailed molecular characterization of the M7 mutant was performed to understand the regulatory mechanism associated with the vir cluster.
4. The M7 mutant was deficient in hydrophobicity, secondary metabolism biosynthesis and mycoparasitism against plant pathogens.
5. The study of the M7 genome using RNAseq and whole-genome sequencing provides novel insight into the genetics of morphogenesis, secondary metabolism and mycoparasitism and could lead to the identification of novel gene regulators in *Trichoderma* spp.
6. Cloning, expression, purification and characterization of the vGPD protein of *T. virens* was also performed.
7. The EMSA analysis of vGPD and promoter region of the *vir4* gene showed that vGPD protein may not act as a transcription factor driving the vir cluster.
8. We also succeeded in obtaining small crystals of this protein.
9. The *in silico* constrained docking analysis between heptelidic acid as an inhibitor and vGPD and gGPD as structural models showed that heptelidic acid shows weaker binding with vGPD as compared to gGPD.
10. A conserved indel is specific to vGPD protein, both in *T. virens* and *Aspergillus* spp., but the effect of this indel in vGPD cannot be completely decoded yet.