Genome-wide Identification and Comparative Analysis of polygalacturonases in different fungal species

Laiba Amin

Regn Number 328971

Master of Science in Plant Biotechnology

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology

Islamabad, Pakistan

2022

Genome-wide Identification and Comparative Analysis of polygalacturonases in different fungal species

By

Laiba Amin

Regn Number 328971

Supervisor

Prof. Dr. Muhammad Faraz Bhatti

Head of the Department

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology

Islamabad, Pakistan

2022

Genome-wide Identification and Comparative Analysis of polygalacturonases in different fungal species

By

Laiba Amin

Regn Number 328971

A thesis submitted to the National University of Sciences and Technology Islamabad in partial fulfilment of the requirements for the degree of

Master of Science in

Plant Biotechnology

Thesis Supervisor

Prof. Dr. Muhammad Faraz Bhatti

Dr. Muhammad Faraz Bhatti
Tenured Professor
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Thesis Supervisor's Signature

Scanned with CamScanner

iii

FORM TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by:

Laiba Amin Reg no:00000328971

Titled: Genome Wide Identification and Comparative Analysis of Polygalacturonases In Different Fungal species be accepted in partial fulfillment of the requirements for the award of MS Degree in Plant Biotechnology degree with (A grade).

Examination Committee Members

1. Name: Dr. Roohi Aslam Signature: _____

2. Name: Dr. Alvina Gul Signature: _____

3. Name: Dr. Fazal Adnan Signature: _____

Supervisor's name: Dr. Muhammad Faraz Bhatti Signature: _____

Date: 20-10-22

Dr. Muhammad Faraz Bhatti
Tenured Professor
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

20-10-22

Head of Department Date

Dr. Muhammad Faraz Bhatti
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

COUNTERSINGED

Dr. Hussnain Ahmad Janjua
Principal
Atta-ur-Rahman School of Applied Biosciences (ASAB)
NUST, Islamabad

Date: 20-10-22 Dean/Principal

Scanned with CamScanner

THESIS ACCEPTANCE CERTIFICATE

Certified that the contents and form of the thesis entitled "Genome-wide Identification and Comparative Analysis of polygalacturonases in different fungal species" submitted by Laiba Amin have been found satisfactory for the requirement of the degree.

Supervisor: _____

Prof. Dr. Muhammad Faraz Bhatti

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Head of Department: _____

Prof. Dr. Muhammad Faraz Bhatti

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Principal: _____

Prof. Dr. Hussnain Janjua

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Dated: 26/10/22

Scanned with CamScanner

CERTIFICATE FOR PLAGIARISM

It is certified that MS Thesis titled " **Genome-wide Identfication and Comparative Analysis of polygalacturonases in different funagl species**" by **Laiba Amin** has been examined:

I undertake the follows:

a) Thesis has significant new work/knowledge as compared to already publish or are under consideration to be published work. No sentence, equation, diagram. Table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.

b) The work presented is original and own work of the author (i.e. there is no plagiarism) No ideas, processes, results or words of others have been presented as Author own work.

c) There is no fabrication of data or results, which have been compiled /analyzed.

d) There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record

e) The thesis has been checked using TURNTTIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions.

Dr. Muhammad Faraz Bhatti
Tenured Professor
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

(Supervisor)

Prof. Dr. Muhammad Faraz Bhatti

Scanned with CamScanner

Declaration

I certify that this research work titled "**Genome-wide Identification and Comparative Analysis of polygalacturonases in different fungal species**" *is* my work. The work has not been presented elsewhere for assessment. The material used from other sources has been properly acknowledged/referred to.

Signature of Student

Laiba Amin

Regn Number: 00000328971

Master of Science in Plant Biotechnology

Scanned with CamScanner

Copyright Statement

• Copyright in the text of this thesis rests with the student Laiba Amin Copies (by any process) both in full or in extracts, may be made only in accordance with instructions given by the Laiba Amin (author); and lodged in the Library of Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST. The Librarian may obtain details; This page must form part of any such copies made. Further copies (by any process) may not made without the permission (in writing) of the author.

• The ownership of any intellectual property rights which may be describe in this thesis is vested in Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the ASAB, which will prescribe the terms and conditions of any such agreement.

• Further information on the conditions under which disclosures and exploitation may take place is available from the Library of Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad.

Laiba Amin

Acknowledgements

Firstly, I would like to thank Almighty Allah for everything in my life.

I would like to express deep gratitude to my supervisor **Prof. Dr. Muhammad Faraz Bhatti** and GEC members especially **Dr. Roohi Aslam** whose guidance was throughout my research. I am also thankful to Principal ASAB, **Dr. Hussnain Ahmed Janjua** for providing me the opportunity and possible research facilities to study in NUST.

I am grateful to my parents and siblings for their support.I am thankful to my friends Oushna Fajer, Isra Farooq , Nabia Khan, Noor-ul-Huda and Qurat- ul-ain for staying with me through thick and thin. I am thankful to my senior Tehsin Fatima whose guidance help me during the entire research.

Last but not least a special thanks to my family for the encouragement, love and support

Dedication

I dedicated my work to my Mama, Baba and Brothers for their unconditional love, support, prayers and motivation.

Contents

Declaration vii

Copyright Statement vii

List of Figures xiii

List of Tables xv

List of abbreviations xvi

Abstract xvii

Chapter 2: Literature Review 6

2.1 Overview of Enzymes 6
2.1.1 Importance of Microbial Enzymes 7
2.1.2 Enzymes of Fungal Origin 8
2.2 Pectinases: Pectin Degrading Enzymes 9
2.2.1 Classification of Pectic Enzymes 10
2.2.2 Polygalacturonases 10
2.3 Glycoside Hydrolase 28 Domain of Polygalacturonases 12
2.4 Fungal Strains as a Source for Polygalacturonases 13
2.5 Substrate for Polygalacturonase Production 14

Chapter 3: Materials and Methods 15

3.1 Laboratory Guidelines 15
3.2 Sequence Retrieval of Fungal Polygalacturonases 15
3.3 Multiple Sequence Alignment and Phylogenetic Analysis of Selected Fungal Species 16
3.4 Chromosomal Mapping of Polygalacturonases 16
3.5 Gene Structure Analysis and Conserved Motif Analysis of Polygalacturonase 16
3.6 Determination of Physiochemical Properties and Subcellular Localization 16
3.7 Gene Ontology Analysis 17

3.8 Protein Modelling of Polygalacturonases 17
3.9 Secondary Structure Prediction of Fungal Polygalacturonases 17
3.9 Tertiary Structure Prediction of Enzyme 18
3.10 Superimposition of Polygalacturonases for Structural Analogy 18
3.11 Differential Gene Expression Analysis of Polygalacturonase in Model Fungi *Neurospora crassa* :.....19
3.12 Sample Collection of Fungal Strain *A. flavus* 19
3.13 Preparation of Fungal Slants 20
3.14 Inoculum Preparation.....20
3.15 Submerged Fermentation of Polygalacturonases Enzyme 20
3.16 Production of Enzyme Polygalactuoranase.....20
3.17 Enzyme Activity Analysis of Polygalacturonase 21
3.18 Protein Estimation by Bradford Reagent 21
3.19 Procedure of Standard Curve 21
Chapter 4 : Results 22

4.1 Identification of Fungal Polygalacturonases 22
4.2 Multiple Sequence Alignment and Phylogenic Analysis of Fungal Polygalacturonases 23
4.3 Chromosomal Mapping of Fungal Polygalacturonases 24
4.4 Gene Structure Analysis, Conserved Motif analysis and Domain Analysis for Polygalacturonases 28
4.5 Physiochemical Properties and Subcellular Localization Prediction 30
4.6 Gene Ontology Analysis of Fungal Polygalacturonases.....34
4.7 Signal Peptide Prediction of Polygalacturonases 36
4.8 Homology Modelling of Fungal Polygalacturonases 37
4.9 Secondary Structure Prediction of Polygalacturonases 38
5.0 Prediction of Tertiary Structure of Polygalacturonases 39
5.1 Superimposition of Selected Fungal Polygalacturonase with *A. aceleatus* 40
5.2 Differential Gene Expression of Polygalacturonase in Model Fungi *Neurospora crassa*..42
5.3 Production of Enzyme by Submerged Fermentation.....43
5.4 Enzyme Activity Analysis.....43
5.5 Protein Estimation through Bradford Assay 44
Chapter 5: Discussion 45

Chapter 6 : Conclusion and Future Aspects 49

References 51

List of Figures

Figure 2.1 Mode of action of Polygalacturonase. The primary component of the entire pectin structure, C6H10O7, is held together by the -1,4 linkage, which is broken down by the enzyme polygalacturonase (PG). 22

Figure 4.1 The boxes A,B,C,D,E and F shows the conserved domain of PG. Different similarities are indicated by distinct shading colours (red: 100%, blue: ≥ 80%, green: ≥ 60%). 23

Figure 4.2 Phylogenetic relationship of forty four fungal PGs .Clustal W was used for the multiple sequence alignment and and a phylogenetic tree was built using the neighbour joining method with 1000 bootstrap values. 24

Figure 4.3 Chromosomal location of *A.oryzae (A) ,A.flavus(B),N.crasaa(C) and R.solani*. The location of each protein is indicated by lines drawn on the chromosomes that are joined by coloured circles. 27

Figure 4.4 The intron exon structure predicted by Gene Structure Display Server 2.0.The scale at the bottom determine the gene length in bp. 28

Figure 4.5 Conserved motif identification from MEME moif finder server using default parameters and than viusulized through TB tool 29

Figure 4.6 Heat map generated to indicate the presence of proteins through graphical representation. Red colour indicate the higher expression of proteins in extracellular region. 33

Figure 4.7 Gene ontology analysis.Green colour indicates the biological functions,blue colour indicates the molecular function and GO distribution level of top 20 proteins indicates Biological,Molcular and Cellular processes. 36

Figure 4.8 Signal peptide prediction by signal p. 37

Figure 4.9 PSIPRED predicts the secondary structure of proteins.It graphically indicates the coil,helix and beta strands.Different colours distingushed strands, helix ,coils, Membrane heklix and transmembrane helix. 39

Figure 4.10 3D structure refinement by 3D refine webserver. The 3D refined structure of Ao9 ,Af7, Nsc1 and Rz6 represents as A,B ,C and D.Red colour indiactes the helix,Yellow sheet and green loops.....40
Figure 4.11 A,B,C and D represents the superimposition of Ao9 (Model 4) ,Af7 (Model 4),NSc1 (Model 4) and Rz6(Model 5) with A.Aceleatus.The purple and sky blue colour used for Ao9(Model 4) and Af7 (Model 4) and Green and sky-blue colours indicates NSc1 (Model 4) and Rz6(Model 5).....41
Figure 4.12 Gene expression. Red color represents the downstreaming.Black color represents the non significant part. Green color represents the up-regulation. Gene of interest is represented by green positive symbol.Gene of *Neurospora crassa* shows down-regulation.....42
Figure 4.13 Picture showing pure colony of *A.flavus* 44
Figure 4.14 A standard BSA curve for protein estimation 45

List of Tables

Table 4.1 The information of forty-four proteins identified from four fungal species are listed below i.e *A.oryzae, A.flavus, N.crassa and R.solani*.....31

Table 4.2 Theoretical pI , Molecular weight and amio acids are determined by expassy tool 33

Table 4.3 The values of ERRAT ,Disallowed region, Qmean score and Favourable region are given below. 39

Table 4.4 Spcies with their RMSD value are given below. 41

Table 4.5 A standard curve was plotted based on observed values .....….44

List of abbreviations

Ao	*Aspergillus oryzae*
Af	*Aspergillus flavus*
Rz	*Rhizoctonia solanai*
NSc1	*Neurospora crassa*
PG	Polygalacturonase
Pgases	Polygacturonases
PDA	Potato Detxtrose Agar
UV	Ultraviolet rays
SSF	Solid state fermentation
SmF	Submerged fermentation
OD	Optical density
KH2PO4	Potassium dihydrogen phosphate
SM	Screening media

Abstract

Pectinases, an enzyme family with many varied uses, which also include polygalacturonases having major industrial applications. These enzymes are employed in the textile sector for bioscoring cotton, the food industry for softening of fruits , in degumming of plant rough fiber and wastewater treatment. Polygalacturonases catalyze the breakdown of glycosidic linkages between two non-esterified galacturonic acid units. There were diverse sources of polygalacturonases however the potential of better catalyzing ability of fungal polygalacturonases were not explored. Therefore, the goal of the current work was to use bioinformatics tools and techniques to discover the fungal strains that have greater catalytic activity. A total of 4 different fungal species were considered for this research including *A. oryzae*, *A. flavus*, *N. crassa* and *R. solani*. The isolates mentioned were known to produce polygalacturonase. All proteins (44 sequences) shared the GH28 domain. Phylogenetic study confirmed the evolutionary relationship between *A. flavus* and *A. oryzae*. Later, chromosomal mapping of 44 sequences were done and these protein sequences were distributed on chromosome number 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 13 and 16. Subcellular localization predicted the extracellular nature of the enzyme. The Robetta server identified the four best enzyme models of polygalacturonase. The gene structure display server indicates the presence of only one exon in Af1 and Ao1 and one intron in NSc1. Five motifs were conserved in all the sequences predicted by motif analysis. The gene ontology analysis was performed to provide information about molecular cellular and biological functions of these proteins. Molecular functions indicated the major role of polygalacturonase in hydrolase activity. Biological functions indicated their key role in cell wall organization .Gene expression analysis of polygalacturonase revealed that only one gene of *N. crassa* downregulated under provided growth condition. The downregulation of gene indicated the inability of polygalacturonase to break pectin due to the

absence of transcriptional factor pdr1. Hence the gene expression analysis reveals the mechanism of action of polygalacturonase and mechanisms action of transcriptional factor pdr1. To verify the enzyme activity assay, a laboratory evaluation of one of the identified fungus species *A.flavus* was also optimized for polygalacturonase production.

1. Introduction

One of the most polluting industries is textile chemical processing. High temperature, acidity, pressure and large capital investment all occur in chemical processing. It is very costly to eradicate unwanted by-products. As a result, the environment suffers as a result of chemical and energy consumption, as well as harmful by-products. The textile industry consumes six to nine billion liters of water and the emission of greenhouse gases approximately about four million tons and chemicals released are about six million tones according to Sistema Mode Iitalia (SMI) (Magni & Carlo, 2019). The consumption non-phase in terms of waste production clothing, tanning textiles, accounts for non-hazardous manufacturing waste for 37 percent (Desiato et al., 2012). According to a report generated by the fashion sector, CO2 emissions will rise by up to 63% by 2030 (Tedesco & Montacchini, 2020). With the introduction of biotechnology, the textile industry is expanding at a reduced cost while maintaining quality and functionality by integrating enzymes. Attempts have been made to replace polluting chemical substances with enzymes which act as a biocatalyst. In the textile industry, enzymatic pretreatment is most widely used (Buchert et al., 2000; Csiszar et al., 2001; Dhiman et al., 2008; Haki & Rakshit, 2003)**.** Enzymes used to replace traditional chemical processes are xylanase, cellulase, pectinase and ligninases in biotechnological processes (rühlmann et al., 2000). In all stages of textile processing enzymes are being used by researchers such as in pretreatment effluent treatment dying, finishing and bleaching dye. Some applications have become well-established routines, while others, due to technical or financial constraints, have yet to be successfully industrialized environment-friendly approach is being adopted in bioscoring. Non-cellulosic impurities are released when enzymes textiles target them.

Pectin is a non-cellulosic component of cotton fiber that is mostly found in the primary wall's cuticle (Hardin & Kim, 1998). Cuticle properties is responsible for hydrophobic nature of cotton. Pectin acts as a network in the cuticle, attracting and retaining smaller hydrophobic molecules such as waxes (Sakai et al., 1993). To give a suitable finishing procedure, the hydrophobic layer of cotton must be reduced during the refinement process. Currently, biotechnological approaches in cotton fiber pretreatment are being used to improve and develop new textile industrial processes (desizing, scouring and bleaching). Among these pre-treatment methods (desizing) has been used in the industry, and it is based on biological products such as amylases (Hoondal et al., 2002). Cotton fiber is made up of structural and reserve polysaccharides. The outer layer is the cuticle, which is made up of waxes, pectin, and proteins. It is a soft cover that serves as a water-resistant fiber-protecting layer (Calafell & Garriga, 2004). The scouring process removes this hydrophobic outer layer. Alkaline and acidic pectinases are categorized on the basis of enzymatic activity. In the bioscoring of cotton, alkaline pectinases are particularly used (Kashyap et al., 2001). It gives more strength, and less weight of cotton is lost by keeping its cellulose structure intact that employs enzymes capable of specifically targeting and releasing noncellulosic impurities. The bleaching process becomes easier because fiber has greater wetting, penetration and dye uptake ability. Non-cellulosic molecules are removed in an environment-friendly way in bioscoring according to Li and Hardin (1998).Several classes of enzymes are identified for scouring including lyases, both exo and endo polygalacturonase. Proteins breakage into low molecular water weight molecules by effective action of pectinases which enhance the fiber absorbency and whiteness (McNeil et al., 1984). By the use of alkaline pectinases the fiber and cellulose backbone remain

unharmed (Ahlawat et al., 2009). Due to the rising demand for pectinolytic enzymes microbial systems are required. Because of their short lifespan, great production, low cost, and lack of toxic compounds that are present in enzymes from plant and animal sources, microorganisms are favored as an enzyme source (Chaplin & Bucke, 1990).

High-weight molecular proteins which act as catalysts are characterized as industrial fungal enzymes (Baur, 2005; Quirce et al., 1992). They are increasing day by day due to the development of a new bioeconomy. Besides their ability to survive in extreme environmental condition and high substrate selectivity as compared to bacteria (Souza et al., 2015). Numerous fungi have been employed for this purpose but Aspergillus species, *Aspergillus japonicus*, *Rhizopus stolonifer*, *Fusarium oxysporum*, *Alternaria mali*, *Mucor hiemenis*, and *Neurospora crassa* are some of the most popular. In terms of polygalacturonase production fungi are thought to be the most productive microorganisms. They have advantages over other microbial enzymes in terms of easier purification, high production potency, particularly for filamentous fungi, the ability to act as an efficient catalyst under harsh conditions with high stability, an acceptable shelf life due to protein stability, fulfilment of regulatory requirements, provide customers solution and meet end users need. Enzymatic activity occurs in intracellular and extracellular in fungal cells. A large number of extracellular enzymes are produced by further pathways that are identified as natural decomposers and they help in the bioconversion of complexes and substrates (Berbee et al., 2017).

In decaying products and the soil where peels of vegetables and fruits are buried pectin degrading fungi found there. Nutrient medium which has high concentrations of pectin

microorganisms grow there and are precisely used for PG synthesis. Utilizing the pectin in a media as a carbon source, PG-producing microorganisms create a lytic zone around their colonies, which may be seen using various chemical components like iodine solution (Amilia et al., 2017).

PGs have many advantages. They play important rule in an ecology by making interaction between hosts and pathogens (Idnurm & Howlett, 2001). They help in maintaining the carbon cycle and biodegradation of pectic chemicals as act as a pollution management (Kashyap et al., 2001). They are use in food technology for clarification of juices , in the preparation of alcoholic drinks and oil extraction (Lang & Dörnenburg, 2000).In transgenic plants their overgrowth leads to change in phenotype and induce growth and development (Atkinson et al., 2002).

Our study focuses on the identification of fungal strains that produce polygalacturonase and which specie have better catalytic activity. By examining the genome-wide identification of fungal PGs, we primarily concentrate on the analysis of fungal strains that are the top producers of PGs. Insilico analysis offers deeper insights to comprehend the connections between various fungal polygalacturonase and help us in understanding of chromosomal position, multiple sequence alignment motif analysis, primary, secondary and tertiary structure prediction of enzymes. Our discovery will serve as a springboard for further enzyme investigation and provide advanced level knowledge on the mode of action and mechanisms of PG through various gene expression studies.

Aims and Objectives

Fungal polygalacturonase has been widely used due to its high rate of enzyme activity, low pH range for optimum activity, and suitability for most fruit and vegetable processing applications (Zheng & Shetty, 2000). The overall process can be enhanced with the help of fungal enzymes. Due to the large demand for fungal PG, it is important to determine the novel fungal species for its production and to identify which specie has the better catalytic activity. However, not all enzymes can be employed directly in industrial processes. These enzymes undergo a variety of improvements to increase productivity and compete with chemical catalysts.

Keeping in view the importance of fungal polygalacturonase our current study focusses on the identification of novel fungal strains for the production of polygalacturonase by using bioinformatics tools and technique.

The following objectives were kept in mind in the proposed study:

- Genome-wide identification of fungal polygalacturonase.
- Insilico characterization of fungal polygalacturonase.
- Differential gene expression of polygalacturonase in model fungi *Neurospora crassa*.

Chapter 2: Literature Review

Fungal enzymes, importance of fungal enzymes, its family, structure, strains and characterization of polygalacturonase via in silico methods are studied in this chapter. For years, scientists have been experimenting with numerous ways to create PG. Through each passing year new techniques have been developed. Additionally, some researchers looked for the cheapest sources of carbon and nitrogen for the synthesis of PG. On its characterization and purification, experimental work was done.

2.1 Overview of Enzymes

Biological macromolecules or biological substances are called enzymes that act as a catalyst to fasten the biochemical reactions produced by living organisms. Another definition of enzymes is "large biological molecules which sustain life by interconversion of chemicals" (Gurung et al., 2013).

Biochemical reactions outside and inside the cell are fastened with the help of these chemical catalysts. Biocatalyst is their other name. The term enzyme was introduced in 1877 by physiology professor named as Wilhelm Friedrich Kühne at the University of Heidelberg. It originates from the Greek word "ενζυμον" (Gurung et al., 2013). Biological systems are used by enzymes for different purposes. DNA technology can increase the yield of an enzyme using recombinant DNA technology and accord to consumer needs. They help in a chemical process to occur for the necessity of life. Among many processes undergoing enzymes only accelerate the few because of their substrate specificity by allowing only a few metabolic processes to take place. The catalysis of almost 4,000 biological processes by enzymes has been demonstrated (Bairoch, 2000). Emil Fischer, a

Nobel Prize laureate, postulated in 1894 that the reason enzymes are so specific is because the substrate and enzyme both have distinctive complementary geometric geometries that are ideal for one another (Gurung et al., 2013). In more than 500 industrial goods enzymes are used. A growing need for environmentally friendly solutions is the primary driver of this upward trend in industrial enzyme demand. One of the most significant and beneficial sources of many enzymes has been the microbial world. Fungal mycelia can be easily separated in broth by filtration and high centrifugation is not required as in the case of bacteria. So fungi have many advantages over bacteria.

2.1.1 Importance of Microbial Enzymes

Enzymes of microbial origin take into account because industrial demands cannot meet by animals and plants (Guerrand, 2018). Enzymes from microbial sources have many advantages as they are regulated at a faster speed, easy genetic manipulation, high yield and simple handling. Another reason for their increased demand is due to their specificity, stability, Eco-friendliness, ease of manufacture and high catalytic activity. Under controlled conditions, microbial enzymes act as biocatalysts for the generation of several products. During prolonged fermentation at a commercial scale, the chances of microbial contamination decrease in thermostable enzymes (Adrio & Demain, 2014). DNA technology, metagenomics, and protein engineering are being used to make use of several microbial enzymes in bioprocesses. With the aid of several molecular approaches, the efficacy of certain microbial enzymes is enhanced in the pharmaceutical, paper, textile, and biotechnology sectors. Recombinant microbial enzymes are used in the industrial manufacture of a wide variety of goods. On the current worldwide market for industrial microbial enzymes, there is intense rivalry.

2.1.2 Enzymes of Fungal Origin

Fungal enzymes occupied more than half of the total enzymes market (Kango et al., 2019). Enzymes with desirable industrial are producing day by day due to rapidly growing enzyme market. Most scientists desire for discovering new bioactive secondary metabolites revolves around plants. But 170,000 herbal products had been described (Pérez-Victoria et al., 2016; Seyedsayamdost & Clardy, 2014), in which metabolites are produced twenty-two thousand five hundred by microorganisms with approximately 45 percent as products of actinomycetes fermentation and approximately 38 percent as products of fungal origin (Berdy, 2005; Demain & Sanchez, 2009). There are three major advantages of filamentous fungi in host production due to which it is widely used. Proper protein folding and glycosylation occur because fungi possess posttranscriptional modification machinery (Punt et al., 2002). A huge amount of proteins are secreted by fungi (Peberdy, 1994). Regular authorities considered proteins of fungi safe. They are termed as GRAS. For the production of enzymes of non-fungal and fungal origin, filamentous fungi enzymes me indispensable due to their exceptional ability to secrete and express proteins. The fungal species for the production of the native primary producers and combatant industrial enzymes are *Aspergillus niger*, *Aspergillus oryzae*, and *Trichoderma reesei* (Ma et al., 2016).

The fungal defense rule is also played by these extracellular enzymes against toxic components that occur the due to substrate hydrolysis process or naturally. The most effective organisms for producing polygalacturonase are thought to be fungi. They have wider substrate specificity and their ability to survive on harsh chemical conditions make them dominant over bacteria. However, the bacteria which are cited in literature are

Agrobacterium tumefaciens, *Bacteroides thetaiotamicron*, *Ralstonia solanacearum*, and *Bacillus spp* (Bhardwaj and Garg, 2014).

According to earlier study by Jayani et al. (2010), fungi are effective producers of pectinolytic enzyme. He used bacterial and fungal isolates in an experiment. 35% of the pectinase came from bacterial species, while 50% came from fungal sources. In comparison to bacteria, fungal pectinase is more effective, easier to extract, and more stable. The *Bacillus sphaericus* strain PG-31 has the highest level of enzyme activity (Jayani et al., 2010). Incubation time provided by 72 h ,160 rpm and 16 h old inoculum to check the enzyme activity. 35% of enzymes originate from bacteria, 15% are either from animals or plants and 50% of the enzymes that are currently in use are from fungus and yeast and the remaining microorganisms of filamentous origin are used for the enzyme pectinase production (Chaplin & Bucke, 1990).

2.2 Pectinases: Pectin Degrading Enzymes

Enzymes named as “pectinases” which help in the degradation of pectin. Pectin made components with cellular components such as lignin and cellulose. The division of pectin is into two regions i.e hairy and smooth regions. Rhamnogalacturonan and Xylofacturonan surround the hairy region and dominancy of Homogalacturonan in smooth region. Several specific enzymes cleaved these two regions of pectin and as a result esterases and depolymerase come under two major classes. Acetylation and methylation varied up to certain extents.

2.2.1 Classification of Pectic Enzymes

Pectic acids are those which have low acetylation and methylation. Several enzymes influence pectin. Pectin forms amorphous polysaccharides. Pectinase enzymes break down the galactouronase component of pectin molecules, and thus is the basis for classification of pectinases enzymes. They are distinguished into pectin depolymerase and methylesterase pectin.

- Glycosidic linkage is broken between methyl esterase and it is termed "Pectin depolymerase".
- Desterifiation into low pectic acids or low pectin methoxyl is called "Pectin methylesterases".
- Hydrolyzation of the glycosidic bond surrounding the carboxyl group occur by polygalacturonase meanwhile b elimination of the glycosidic bond occurs in pectate lyase adjacent to the carboxyl group.

Polygalacturonase and Phenylalanine ammonia-lyase end demonstration when random breaks occur in a pectin chain (EC 3.2.1.15 and EC 4.2.2.2). Non-reducing end of the chain liberates dimer and monomer in the case of Exo polygalacturonase. From the reducing end, the unsaturated dimer is liberated in the case of exoPAL and when these three combine i.e PAL, PG and PE a highly methylated pectin is released (Pilnik & Voragen, 1993; Sarkanen, 1991).

2.2.2 Polygalacturonases

Depolymerizing pectolytic enzymes are called as PGase (Polygalacturonase) that break the glycosidic bond of two galacturonic acid units (Figure 2.1; Fogarty & Griffin, 1974).

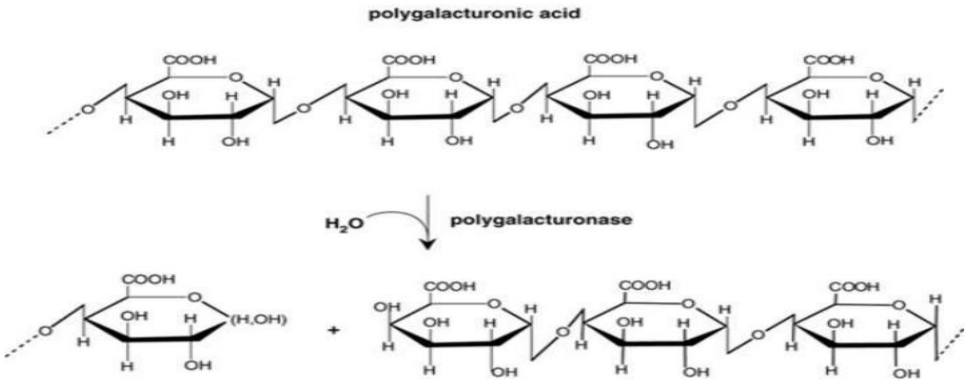


Figure 2.1. Mode of action of Polygalacturonase. The primary component of the entire pectin structure, C6H10O7, is held together by the -1,4 linkage, which is broken down by the enzyme polygalacturonase (PG).Adapted from (Amin et al., 2017)

Post-translational changes in fungal polygalacturonases result in the interaction of specific polypeptides with sugar molecules. In fact, PGs produced by fungi are monomeric. Fungal PGs contain five to ten percent carbohydrates. (Rao et al. 1996; Niture et al. 2001; Yan & Liou 2005).

Degradation of pectic molecules occurs by Endo and Exo PGs of fungi. Unsaturated mono or galacturonic acids are produced in exoPG by the hydrolyzation of 1,4 glycosidic linkages. Internal glycosidic links are broken by EndoPGs (Nakkeeran et al., 2010; Rombouts & Pilnik, 1980).

Applications for polygalacturonases in the industry are numerous. The fruit processing sector uses the majority of polygalacturonase enzyme preparations, and pectic enzymes alone produce nearly a fifth of all food enzymes worldwide.

2.3 Glycoside Hydrolase 28 Domain of Polygalacturonases

Pectin, a complex polymer, contributes to the production of plant cell walls. Pectin accounts for 35% of plant cell walls. Its backbone is commonly made up of 1,4 linked -d-galactosyluronic acid monomers, though it can contain up to 17 monosaccharides (Voragen et al., 2009). Due to the diverse variety of pectin's structural components, microorganisms have created a wide range of extracellular enzymes to aid in pectin's breakdown (Glass et al., 2013; Manucharova, 2009; Prade et al., 1999). The glycoside hydrolases that break down pectin have been grouped into family 28 based on sequence similarity (Henrissat, 1991). The members of this family are PG (polygalacturonase), exoPG (exopolygalacturonase), EPGD (exo-poly-α-galacturonosidase), RG (rhamnogalacturonase) and endo-xylogalacturonan are currently the members of this family. α-1,4 glycosidic bonds in pectate and other galacturonic cleavage occur in the presence of PG (Henrissat, 1991). Endopolygalacturonase is a newly introduced class of PG found in glycohydrolase 28 (Van der Vlugt-Bergmans et al., 2000). An inverting mechanism is employed by all the members of the family GH28 gene distribution and abundance is particularly significant for ecosystem functioning because of its function in the breakdown and plant pathogenicity (Glass et al., 2013; Manucharova, 2009). The GenBank (Benson et al., 2000) and SWISS-PROT (Bairoch & Apweiler, 2000) sequence databases contain more than 100 amino acid sequences for the family 28 glycoside hydrolases. Thus, it is a major family of the enzyme. The identification of three-dimensional structures, RG A from *Aspergillus aculeatus*, PG A from *Erwinia carotovora ssp*. carotovora, and PG II from *Aspergillus niger* already been reported. In pectate lyase

the 'parallel -helix structural domain,' was identified earlier (Glass et al., 2013; Jenkins et al., 1998).

Glyco hydro 28 domain contains right handed parallel helix sheets and twelve coils that form three to four parallel sheets. However, efficient methods for determining how frequently GH28 genes are found in natural populations have not been created, and there has not yet been a thorough examination of all the glycoside hydrolase family 28's amino acid sequences.

2.4 Fungal Strains as a Source for Polygalacturonases

Solid state and submerged fermentation are used for the production of PG from fungal sources. The earlier approach is chosen since it is simple to manage. When microbial growth and product synthesis occur on the surfaces of solid substrates with little to no free water present, this is considered a solid-state fermentation. Due to the limited water supply, fungi are the type of microorganisms that are most usually used in solid-state bioprocessing. However, one of the main barriers to solid state fermentation's ability is the lower mass transfer techniques in terms of temperature, gases and nutrient diffusion. Several agro-industrial waste byproducts, including sugar cane bagasse (Pandey et al., 2000), orange bagasse (Martins et al., 2002), and other food processing waste, act as a substrate for solid-state fermentation-based depolymerizing enzyme production. Different microorganisms have been discovered that can generate pectinase in solid-state cultures. It has recently been proven and tested that these pectinases can dissolve polysaccharides (Soares et al., 2001). While submerged fermentation is simpler to manage, requires less workspace and manpower, is automatable, exhibits less contamination, and yields a higher volume of product.

Infected host tissue of plants, soil, decaying plant parts, diverse fungal species, and mangrove/estuarine settings are only a few of the places where different researchers have isolated PGs from different fungi. The formation of PGs in submerged culture and laboratory settings typically depends on the medium's components, including its pectin supply, nitrogen source, starting pH, temperature, and agitation. Since fungi prefer to grow on semi-solid media, semi-solid fermentation is the optimum method in which the moisture content, raw agricultural source and pH are the primary determinants.

Paula et al., (2006) reported that the use of agro-industrial wastes for polygalacturonase production by newly isolated thermophili Monascus sp N8 and thermotolerant Aspergillus sp N12 was feasible in solid-state fermentation (Leh et al., 2017). A mixture of wheat bran, orange bagasse and sugar cane bagasse were used as substrate, and it was seen that fungi produced high levels of PG in the solid-state fermentation. The enzyme had good thermal stability and high tolerance to low pH. These properties could make these PGs useful for fruit juices processing in food industries.

2.5 Substrate for Polygalacturonase Production

For generating PG at low costs and making the biosynthesis more economical agriculture by products together with fruit peels used worldwide variety of substrates including lemon,orange ,musami as well as rice bran ,soybean is used as a substrate for the production of PG. In addition to helping to reduce pollution from agricultural and agro-industrial wastes, the use of these natural waste substrates helps to make the manufacture of PG more cost-effective (Siddiqui et al., 2015).

Chapter 3: Materials and Methods

3.1 Laboratory Guidelines

Safe lab procedures are essential to performing research. Effective procedures can reduce exposure to risks. These safe lab procedures cover many of the typical exposure routes, although they are by no means an exhaustive list. Drinking and eating is not allowed in lab .Raising the chance of experiment contamination increase by consuming food. Wear gloves and lab coats before handling any chemical. The workstation is constantly sterilized with ethanol for a secure and clean workplace. To prevent contamination and dangerous gases, solutions are always made in safety cabinets and fume hoods. To prevent any harm, each experiment's trash is disposed of properly at the end. To prevent any negative chemical interactions with the environment, the lab apparatus should never be removed. No food or beverage of any kind should be brought into the lab. To prevent any accidents, the workplace should always be clearly marked.

3.2 Sequence Retrieval of Fungal Polygalacturonases

Complete coding sequences (CDS) sequences (both genomic sequence and protein sequences) of fungal polygalacturonases were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/). Total of *Aspergillus oryzae ,Aspergills flavus, Neurospora crassa* and *R. solanai* .These species were characterized for the production of polygalacturonase. To perform BLAST P these sequences were used as queries against ten selected reference organisms. The Pfam server was used to find the conserved domain against the obtained proteins. Glyco hydro 28 domain was confirmed by using the software Pfam (http://pfam.xfam.org/).

Proteins having significant domains were selected and those which contain non-significant domains were removed.

3.3 Multiple Sequence Alignment and Phylogenetic Analysis of Selected Fungal Species

All sequence alignment of forty-four proteins of *A. oryzae*, *A. flavus*, *R. solani* and *Neurospora crassa* were performed using the CLUSTAL W in MEGAX software (Thompson et al., 1994). The evolutionary tree was constructed using the neighbor-joining approach by MegaX (Saitou & Nei, 1987). MEGA software (https://www.megasoftware.net/) used with 1000 bootstrap replicates. Tree was viewed using the application Tree view (Page, 1996). It was also possible to search for sequence similarities using the BLAST tool (Altschul et al., 1990). iTOL online tool visualized the phylogeny tree (Letunic & Bork, 2016). Editing of the tree was performed using it.

3.4 Chromosomal Mapping of Polygalacturonases

PhenoGram plot (https://ritchielab.org/software/phenogram-downloads) predicted the position of chromosomes of *A.oryzae, A.flavus, N.crassa* and *R.solanai*. Files of gene length and chromosomal length were uploaded as input in a phenogram as a result output was generated.

3.5 Gene Structure Analysis and Conserved Motif Analysis of Polygalacturonase

For motif analysis, Multiple Expectation Maximization for Motif Elicitation (MEME) suite webserver (https://meme-suite.org/meme/tools/meme) used to predict the 44 polygalacturonase sequeneces. In default settings with 10 motif hits was set.

To predict the introns and exons organization gene structure display server (http://gsds.gao-lab.org/) used. The number of the untranslated region(UTRs), exons and introns were identified.

3.6 Determination of Physiochemical Properties and Subcellular Localization

ExPASy ProtParam tool (https://web.expasy.org/protparam/) computes the physiochemical properties of polygalacturonase protein including molecular weight, amino acid and theoretical PI (Gasteiger et al., 2005) This tool predicted the different properties of all the forty-four proteins

were subjected to it . WoLF PSORT tool (https://wolfpsort.hgc.jp/) was used. This tool predicts the localization in different subcellular compartments. Then TBtools (https://github.com/CJ-Chen/TBtools) software used to generate the heatmap of 44 protein sequences.

3.7 Gene Ontology Analysis

Functional analysis was done after protein sequences are aligned using the BLASTp. Default parameters were set for analysis (Götz et al., 2008). Biology, molecular and cellular functions were determined by using the OmicsBox (https://www.biobam.com/omicsbox/). In gene ontology annotations connect genes and gene products to GO terms. Each annotation in the GO has a source and a database entry ascribed to it. The source can be a database reference or literature reference (Consortium, 2004; Smith et al., 2007).

3.8 Protein Modelling of Polygalacturonases

Protein modelling of fungal polygalacturonase was performed by using Robetta (https://robetta.bakerlab.org/) web servers. In comparative modelling structural model of the protein of interest was generated by using the templates (structures which were evolutionarily related).

3.9 Secondary Structure Prediction of Fungal Polygalacturonases

Web server PSIPRED were used for the prediction of secondary structure of a protein including alpha helix, beta sheets, coils and extended strands. Swiss model web server helps in modelling study (http://swissmodel.expasy.org/). It generated a three-dimensional structure of proteins (Kiefer et al., 2009).

3.9 Tertiary Structure Prediction of Enzyme

The fungal polygalacturonase protein sequences were submitted to web server Swiss model (https://swissmodel.expasy.org/qmean/). Three different approaches adopted under the heading of QMEAN models. Protein geometric features were identified through QMEAN. QMEAN generated when QMEAN combined with distance constraint score to enable greater quality prediction. It carried out statistical capabilities aimed at a quality estimate of membrane protein in their naturally existing oligomeric state. ERRAT server used to predict the statistics of non-bonded interactions between various atom types (services.mbi.ucla.edu/ERRAT). Validation of protein crystallographic structure through the outcome.

3.10 Superimposition of Polygalacturonases for Structural Analogy

Various methods and tools are required to compare structure analogy between protein folds. It is crucial for the analysis of proteins with a narrow range of sequences and aids in the discovery of evolutionary relationships. Superimposition of predicted models was performed by Superpose(http://superpose.wishartlab.com/). Based on sequential lengths, biological interests, conformational changes and modelling methods. Based on biological interest, conformational changes, and modelling methods a set of protein structures were selected (Bernstein et al., 1977). From the protein data bank the coordinated files of the protein Aspergillus Aculeatus were downloaded. PyMOL was used to overlay and view the anticipated 3D structures of the *A. flavus*, *A. oyzea*, *N. crassa*, and *R. solanai*.

Unexpectedly, the model organisms *A. aculeatus* and the structures *A. flavus* and *A. oryzae* superimposed quite well, and the computed root mean square deviation (RMSD) was only 1.26 and 1.55, respectively.

3.11 Differential Gene Expression Analysis of Polygalacturonase in Model Fungi

***Neurospora crassa* :**

A particularly popular tool R (https://bioconductor.org/packages/release/bioc/html/edgeR.html) used for analyzing the transcriptional activity of biological systems and its analysis of RNA sequencing (RNA-seq). Finding genes or biochemical pathways that exhibit differential expression (DE) between two or more biological situations is the goal of RNA-seq profiling (Robinson et al., 2010).

In this study RNA seq of data of pectin degradation regulator-1 (PDR-1) of *Neurospora crassa* were retrieved from NCBI GEO database (https://www.ncbi.nlm.nih.gov/gds). A transcription factor mutant screen in *N. crassa* led to the discovery of Pectin Degradation Regulator-1 (PDR-1). All of the pectin-related poly- and monosaccharides that were examined as well as pectin itself showed severe growth defects in the pdr-1 mutant. Here sucrose, pectin, rhamnose and control in the absence of carbon were the different growth conditions for *N. crassa* . A total of 11 samples were used for RNA sequencing under different growth conditions and their files were downloaded in ftp format. Data presents SRA and their accessions as well as the various growth media used by *N. crassa* strains. Edge R Bioconductor software was used for the analyses of read counts (Gentleman et al., 2004). The data was read through the library created. Each file contains a table with the tag IDs and counts for each tag. 8 samples were considered as treatments and 3 samples were considered as non-treatments (No carbon source).

3.12 Sample Collection of Fungal Strain *A. flavus*

Fungus strain *A. flavus* (Af1)were collected from Fungal Infections and Virology Group (FIV-G) in the form of stocks slants at ASAB (NUST).

3.13 Preparation of Fungal Slants

For later usage, the Af1 strain was kept on slants. Yeast 1 g/litre, Peptone 1g/litre and Glucose 1g/litre all were mixed in 800ml of water. pH adjusted to 7 by adding agar 220g/litre. The final volume of 1 litre was obtained by adding distill water. The medium was poured into test tubes and autoclaved. Inverting position of test tubes was employed. In laminar flow the strain Af1 was streaked over a solidified test tube. For microbial growth, test tubes were kept in an incubator at 37C for four to seven days.

3.14 Inoculum Preparation

Distilled water was poured in cultural slants. Using an inoculating loop, conidia on the surface of the culture media were gently and aseptically scraped off. Homogenized mixture of conidial suspension were prepared and added to the fermentation medium as an inoculant (Mandhania et al., 2010).

3.15 Submerged Fermentation of Polygalacturonases Enzyme

The conidial inoculum added to the fermentation medium, and then incubated at 30°C for 72 hours. The fermentation media contain 0.20% KH_2PO_4, 0.14%$(NH_4)_2SO_4$ and 0.6% K_2HPO_4 by maintaining a pH of 5. Following incubation, the enzyme was removed by centrifuging fermentation broth for four days at 6000 rpm. PG is an extracellular enzyme, hence the supernatant was used to calculate its activity (Vivek et al., 2019).

3.16 Production of Enzyme Polygalactuoranase

The fermentation media was filtered using filter paper .The supernatant was used as an enzyme source.

3.17 Enzyme Activity Analysis of Polygalacturonase

The enzyme activity was assessed by the reducing sugars released by citrus pectin.0.8ML of 1% citrus pectin,0.2ml of crude enzyme , copper sulphate solution and 0.2M acetate buffer were mixed and incubated for 10min at 40C.

3.18 Protein Estimation by Bradford Reagent

For the protein estimation Bovine serum albumin (BSA) and Bradford reagent was utilized. A standard curve was plotted to estimate the protein content.

3.19 Procedure of Standard Curve

Different dilutions of stocks were prepared for the standard curve BSA having concentration 1mg/ml. Then, various dilutions were made by adding 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1 mL from the stock solution. Then, distilled water added to each tube to bring the volume up to 1 mL, with the exception of the tube containing the 1 mL stock solution. The Bradford reagent was then poured 5 mL into each tube. 5 mL of Bradford reagent and 1 mL of distilled water were put to a blank. The tubes should be incubated at room temperature for 5 minutes. The standard curve was produced using absorbance measurement at 595 nm.

Chapter 4 : Results

4.1 Identification of Fungal Polygalacturonases

When obtaining extremely similar (>90% sequence coverage) protein sequences from various fungal species from the NCBI non-redundant protein sequence database, the amino acid sequence of fungal polygalacturonase was considered as a reference using BLASTp (Altschul et al., 1990). Glyco hydro 28 domain conformation was carried out using the Pfam service. In the conclusion, 44 sequences of fungal polygalacturonase were retired. The chromosomal position was used to rename the fungal proteins. Ao, Af, NSc, and Rz names were given in place of *Aspergillus oryzae*, *Asperfillus flavus*, *Neurospora crassa*, and *Rhizoctonia solani*. List of species with their RNA IDS,Protein IDS ,Gene IDS and Chromosome numbers are indicated in Table A1.

4.2 Multiple Sequence Alignment and Phylogenic Analysis of Fungal Polygalacturonases

Sequences were aligned using a software ClustalW .It is viewed in Jal view .It indicates the level of amino acid sequence similarity among different fungal proteins. In the Figure 4.1 multiple sequence alignment of polygalcturonases are shown.

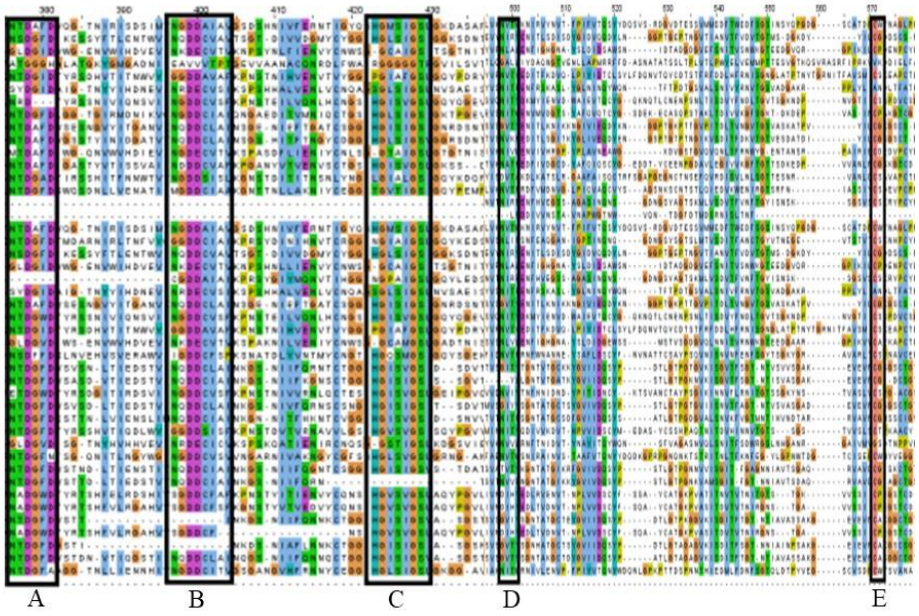


Figure 4.1. Multiple alignment of polygalacturonase. A, B, C, D, E and F show the conserved domain of PG. Different similarities are indicated by distinct shading colors (red: 100%, blue: ≥ 80%, green: ≥ 60%).

To explore the evolutionary relationship and linkage among fungal species an unrooted tree was built using mega X with 1000 bootstrap using the neighbor joining method. The proportion of clustered associated taxonomic tree replicates in the bootstrap test of 1,000 repetitions is displayed

next to each branch of the tree. The forty-four PGs were grouped into four different clades. Light green (A) ,dark green (B) ,purple (C) and sky blue (D) could be differentiated clearly by phylogeny analysis. The results indicate the clade B contains the largest number of fungal PGs and clade A contains the fewer number of PGs as indicated in Figure 4.2. Ao4 does not evolutionary relationship and indicated as a separate node in group D.NSc1 show close evolutionary relationship with Ao7, Rz3 and Af8 and Rz14, Rz12, Rz11 and Rz6 are closely linked.

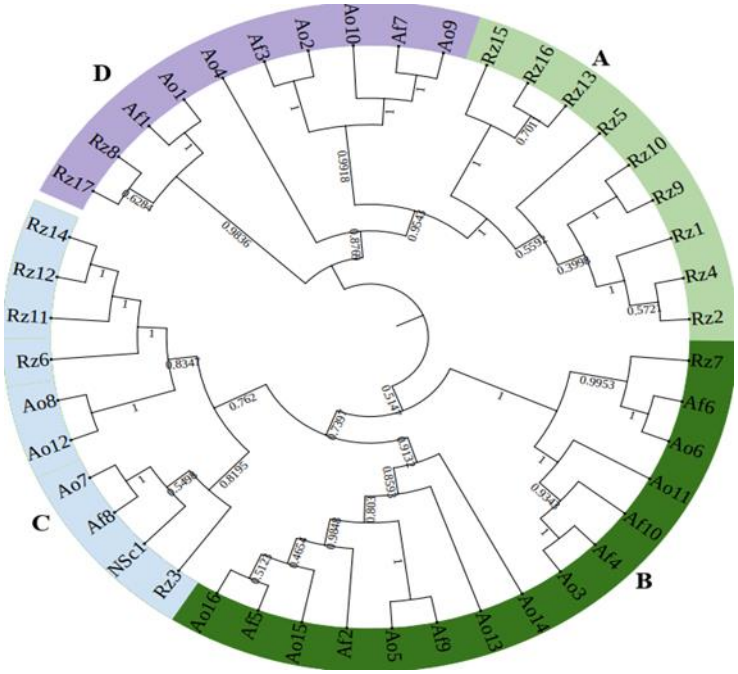


Figure 4.2. Phylogenetic relationship of forty four fungal PGs .Crustal W was used for the multiple sequence alignment and a phylogenetic tree was built using the Neighbor Joining method with 1000 bootstrap values.

4.3 Chromosomal Mapping of Fungal Polygalacturonases

Ideograms visualized regions in the form of regions or chromosomal base pairs. RMagick graphics library create phenogram by Ruby programmers. An ideogram of 16, 10, 1 and of 17 chromosomes are plotted for Ao,Af,NSc1 and Rz along X and Y axis. Through it, we anticipate a closer look at

a particular chromosome's location. The highest number of genes was observed in chromosome 1 and chromosome 3, an equal number of genes were observed in chromosomes 2,4, and 6 and only a single gene was predicted at chromosome 5 in Ao. In Af, chromosome1 indicates the highest number of genes and chromosomes 2 and 3 indicate an equal number of genes that are two only. In NSc1 chromosome 4 indicates one gene only. In Rz highest number of genes were indicated by chromosomes 9 and 13 and the lowest number of genes were indicated by 3,2,6 and 8.

Table 4.1 The information of forty-four proteins identified from four fungal species are listed below i.e *A. oryzae*, *A. flavus*, *N. crassa* and *R. solani*.

s

Specie	Label	Chromosomes	Start point	End point
Aspergillus oryzae	Ao1	1	2712584	2713897
	Ao2	1	5928717	5930357
	Ao3	1	6257823	6259667
	Ao4	1	1465444	1467585
	Ao5	2	297131	298588
	Ao6	2	3485993	3487833
	Ao7	3	2999700	3000671
	Ao8	3	4779098	4780315
	Ao9	3	1031380	1032605
	Ao10	3	406605	407835
	Ao11	3	4408511	4409964
	Ao12	4	4510396	4511670
	Ao13	4	4179347	4180918

	Ao14	5	4510271	4511626
	Ao15	6	4062566	4063206
	Ao16	6	4061382	4062213
	Af1	1	2576668	2577981
Aspergillus flavus	Af2	1	1237301	1238763
	Af3	1	5809593	5811233
	Af4	1	6134358	6136202
	Af5	1	170923	174561
	Af6	1	395231	3473103
	Af7	3	1082439	1083664
	Af8	3	2959828	2961658
	Af9	6	3606171	3607628
	Af10	6	2062357	2064041
Neurospora crassa	NSc1	4	4669933	4671737
Rhizoctonia solani	Rz1	0	525190	526779
	Rz2	3	484716	486289
	Rz3	4	2006965	2008696
	Rz4	6	1903100	1904639
	Rz5	7	888816	890515
	Rz6	7	1099760	1101502
	Rz7	8	1655418	1657289
	Rz8	9	502651	504318

	Rz9	9	987033	988451
	Rz10	9	1054420	1055828
	Rz11	10	1708593	1710220
	Rz12	11	1787624	1789250
	Rz13	12	1473178	1474999
	Rz14	13	300879	303419
	Rz15	13	1947278	1948954
	Rz16	13	2103770	2105440
	Rz17	16	710384	711990

The Figure 4.3 gives the visual representation of chromosomes of different fungi species.

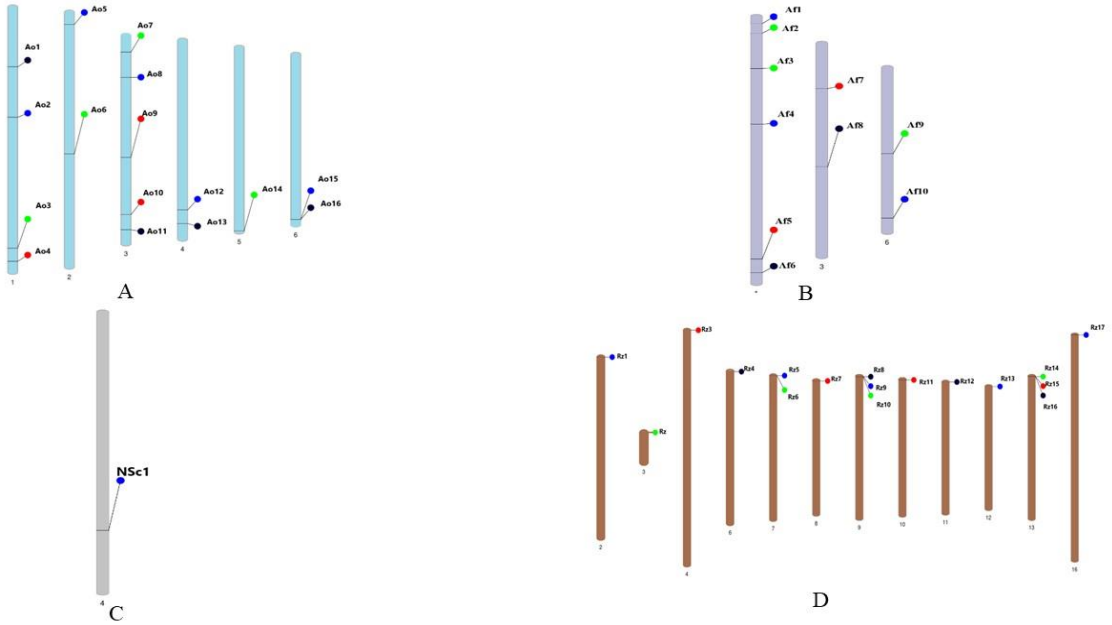


Figure 4.3. Chromosomal location of *A.oryzae* (A), *A.flavus* (B),*N.crasaa* (C) and *R.solani*. The location of each protein is indicated by lines drawn on the chromosomes that are joined by coloured circles.

4.4 Gene Structure Analysis, Conserved Motif analysis and Domain Analysis for Polygalacturonases

Intronic, exonic and upstream regions were predicted by the Gene Structure Display Server (GSDS). The genetic structure was similar to the fungal species which have closer linkage. The number 44 proteins have different gene structures with varying numbers of introns and exons as indicated in figure 4.4. In Af, the number of exons ranges from 0bp to 3500bp. In NSc1 exon ranged from 0 to 2000bp and only one intron was present. Ten unique motifs were observed**.** In Rz number of exon range from 0bp to 2700bp (Figure 4.4).

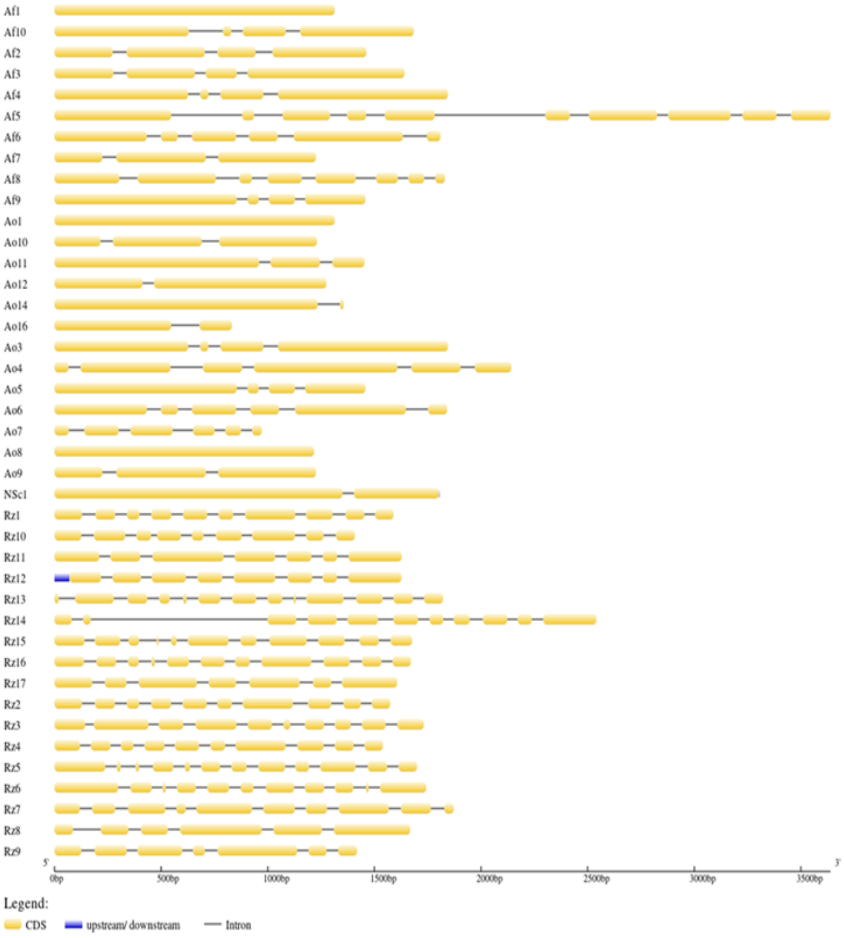


Figure 4.4. The intron-exon structure predicted by Gene Structure Display Server 2.0. The scale at the bottom determines the gene length in bp.

Meme identifies the conserved motifs shown in Figure 4.5. We identified that members of groups A, B, C and D show similar motif patterns. The member of group A Rz15, Rz16, Rz13, Rz5, Rz10, Rz9, Rz1, Rz4 and Rz2 contains 7, 7, 6, 7, 5 ,7, 7 ,7, 7 number of motifs .The member of group B Rz14,Rz12,Rz11,Rz6,Ao8,Ao12,Ao7,Af8,Ns1 and Rz3 contains four common motifs .The members of group C are largest in number and only Ao4 in shows a single motif because it forms a separate clade in a tree. In group D only Ao4 contains one motif.

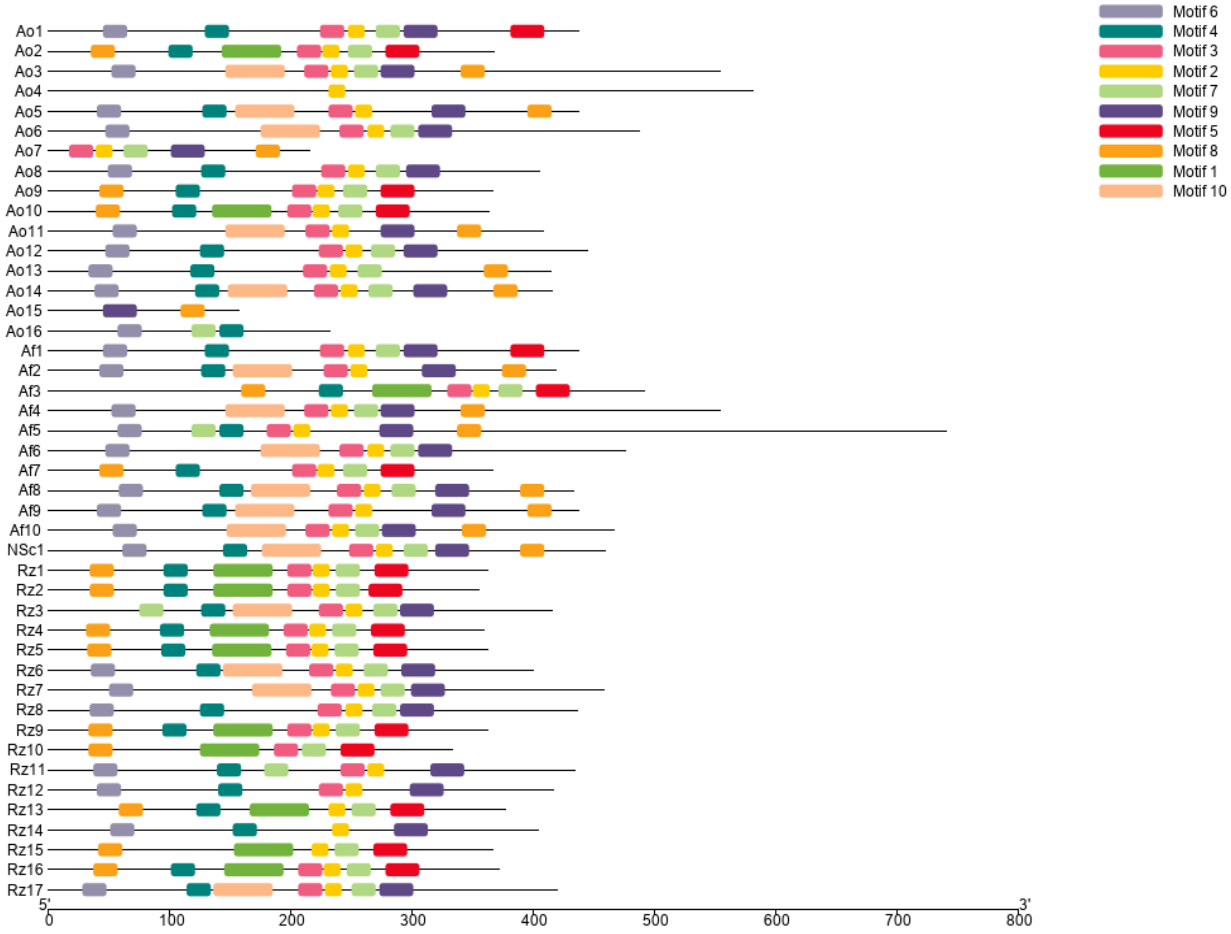


Figure 4.5. Conserved motif identification from MEME motif finder server using default parameters and then visualized through TB tool.

4.5 Physiochemical Properties and Subcellular Localization Prediction

Physiochemical properties of 44 proteins were predicted by the Expasy tool. Three parameters were considered for the determination of physiochemical properties i.e amino acid, molecular weight and theoretical PI (Table 4.2). Ao7 was the shortest protein and theoretical PI.I.e with 217 amino acids and 4.4 Tpi. Af5 was the largest protein with 742 amino acids. The molecular weight of Ao15 was the lowest and Ao4 was with the highest molecular weight. The largest Tpi was identified in Rz8. Graphical representation of data were illustrated by heat map in Figure 4.6. The size of each monitoring value is represented by a different colour. Protein abundance was found in the extracellular region. Red colour indicates high expression analysis and blue colour indicates low expression analysis.

Table 4.2. Theoretical pI , Molecular weight and amino acids were determined by Expassy tool.

Species	Theoretical pI	Molecular weight	No of Amino acids
Ao1	5.27	47626.1	437
Ao2	4.08	38397.67	368
Ao3	4.63	59426.81	556
Ao4	6.29	63263.32	583
Ao5	5.17	48362.92	439
Ao6	4.61	52808.14	489
Ao7	4.4	23539.04	217

Ao8	6.43	42746.44	407
Ao9	8.52	38106.55	369
Ao10	4.51	37959.71	365
Ao11	4.41	44697.65	410
Ao12	4.33	46530.55	447
Ao13	5.67	44416.52	416
Ao14	4.45	46454.59	418
Ao15	7.02	16924.61	159
Ao16	5.56	26248.59	234
Af1	5.27	47844.35	439
Af2	4.8	46045.73	421
Af3	4.37	50976.92	494
Af4	4.73	59229.58	556
Af5	5.65	82719.16	742
Af6	4.58	51324.45	478
Af7	8.52	38016.42	369
Af8	4.38	47031.32	435
Af9	5.17	48272.8	439
Af10	4.63	49612.77	469
NSc1	6.54	50965.27	462
Rz1	7.08	37108.78	364
Rz2	5.94	36485.02	357
Rz3	6.32	45394.43	418

Rz4	4.39	37017.23	361
Rz5	6.38	38201.69	364
Rz6	8.25	43006.15	402
Rz7	8.98	49511.44	460
Rz8	9.19	47963.49	438
Rz9	8.89	37523.65	364
Rz10	9.24	34894.09	335
Rz11	6.84	47127.08	436
Rz12	8.18	45685.17	419
Rz13	8.42	38559.05	379
Rz14	8.1	45001.68	406
Rz15	8.12	38226.35	369
Rz16	4.62	37673.29	374
Rz17	8.14	46077.65	422

The visual representation of heat map is given below in Figure 4.6 :

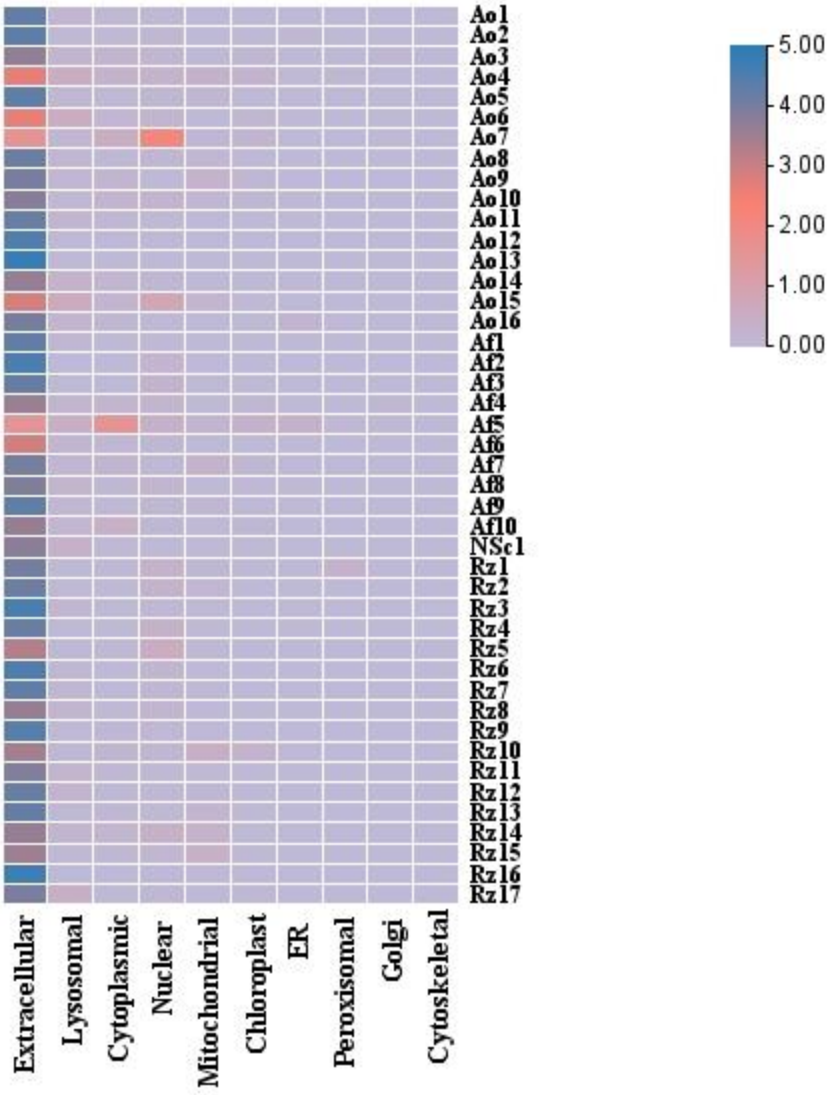
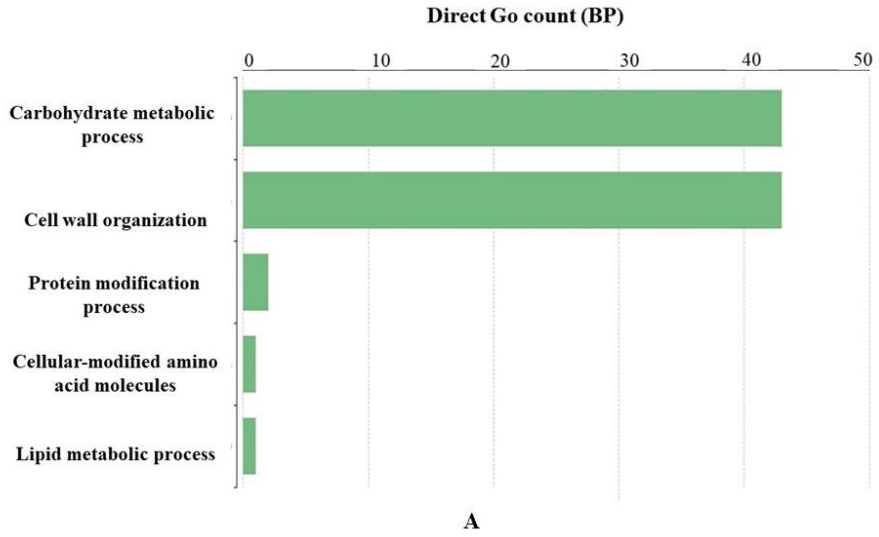


Figure 4.6.Heat map generated from Tbtool to indicate the presence of proteins through graphical representation. Blue colour indicate the higher expression of proteins in extracellular region. WoLf Psort confirms the extracellular nature of enzyme polygalacturonase.

4.6 Gene Ontology Analysis of Fungal Polygalacturonases

The Blast2GOtool was used to examine 44 fungal polygalacturonase gene ontology (GO) annotations. The three primary categories of biological processes, molecular activities, and cellular components were used to organize the target proteins into several functional groupings shown in Figure 4.7. The biological process includes protein synthesis, protein modification carbohydrate metabolism, and lipid metabolism. Hydrolase, ligase, catalytic, oxidoreductase, and transmembrane, activity appears to be involved, according to molecular processes and activity. Cellular processes indicates their role in Cellular anatomy.



A

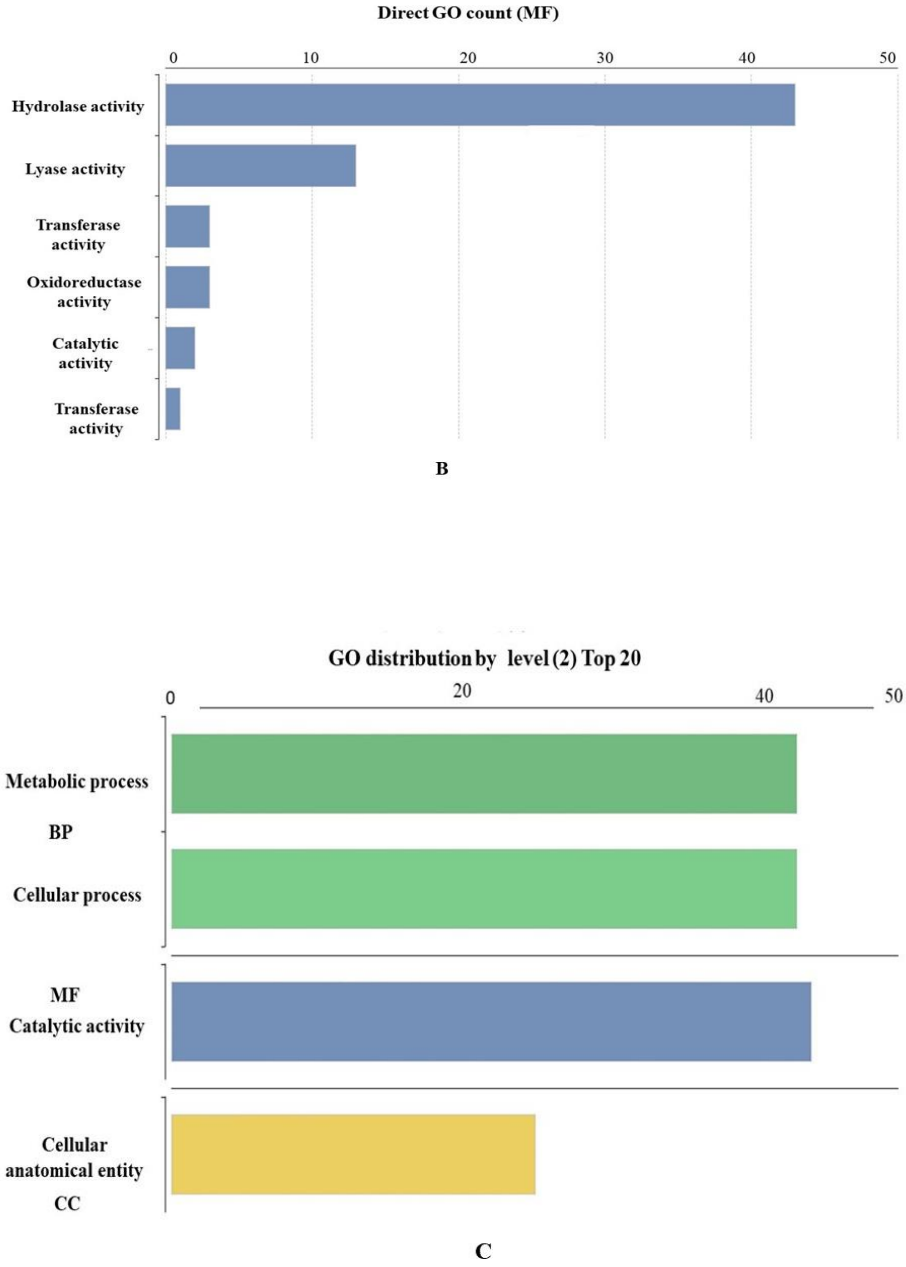


Figure 4.7.Gene ontology analysis. Green colour indicates the biological functions, the blue colour indicates the molecular function and GO distribution level of top 20 proteins indicates Biological, Molecular and Cellular processes.

4.7 Signal Peptide Prediction of Polygalacturonases

Many generated proteins contain SP (Signal peptides) that direct their passenger proteins toward membrane transfer or integration. Through signal peptide, 4 server precursors of polygalacturonase were identified (http://www.cbs.dtu.dk/-services/SignalP). No cleavage sites are shown by Ao2, Ao7, Ao15, Rz13 and Rz14. Ao1, Ao8 and Af1 shows the cleavage site between 21 and 22.Ao3,Rz7 and Af4 shows the cleavage site between 18 and 19.Ao4, Ao9, Af6, Rz9, Rz10, Rz11 and Rz17 shows the cleavage site between 17 and 18. Ao5, Ao10, and Rz5 shows the cleavage site between 20 and 21.Cleavage sites between 15 and 16 is shown by Ao6 ,Af6 and Ao14.Cleavage site between 21 and 22 shown by Ao8, Af1 and Af8.Seven proteins shows the cleavage site between 19 and 20 named as Ao11,Af10,Rz1,Rz2,Rz3,Rz8 and Rz4 (Table 4B). S score steep helps in differentiation of Y score .S score is called as signal peptide score which distinguishes between protein regions that contains signal peptides and proteins that do not contain signal peptide. In a Figure 4.8 indicates pattern of cleavage site of proteins.

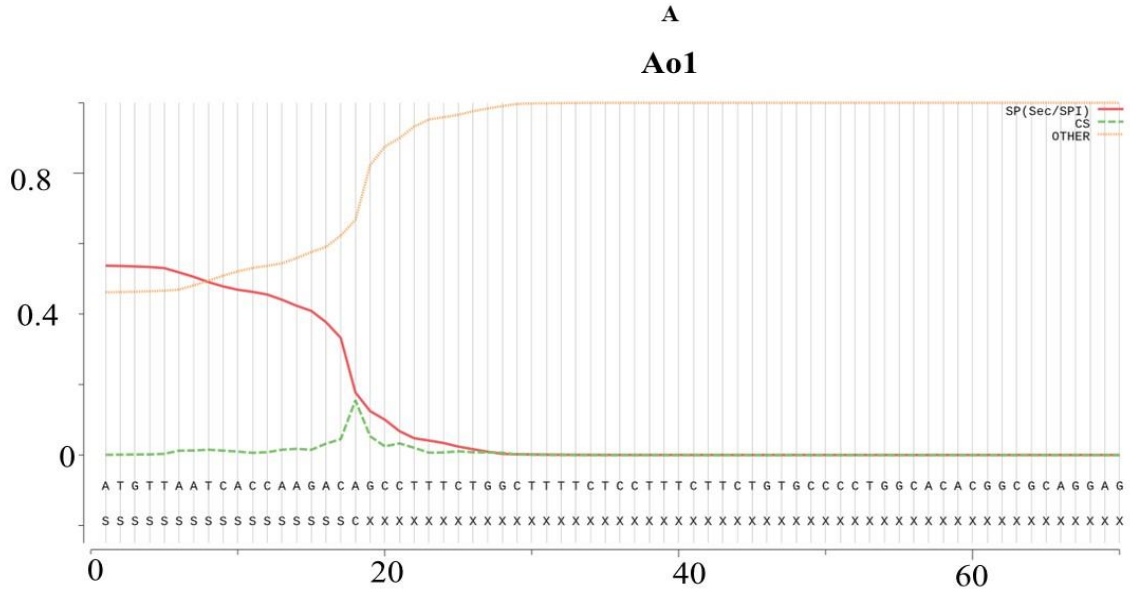


Figure 4.8. Signal peptide prediction by signal p.The cleavage site of Ao1 was between 21 to 22. The same pictogram used for other proteins but with different cleavage sites.The presence of signal peptide(Sec/SPI) was indicated by red line.Orange line indicates others .Green line depics CS.Probability is represented by Y.axis and amino acid sequence is represented by x.axis.

4.8 Homology Modelling of Fungal Polygalacturonases

QMEAN (Qualitative Model Energy Analysis) webserver used to estimate the quality of the protein structure model. This composite scoring function describes the attributes of protein structural geometry and can be used to derive both global (i.e., for the entire structure) and local (i.e., per residue) absolute quality estimates from a single model (Benkert et al., 2011). The scores provided by the final model were chosen from a sizable number of alternative models that were produced. Thus, the most trustworthy model against which the consensus structural scores were generated was chosen using the QMEAN score.

Table 4.3 The values of ERRAT ,Disallowed region, Qmean score and Favourable region are given below.

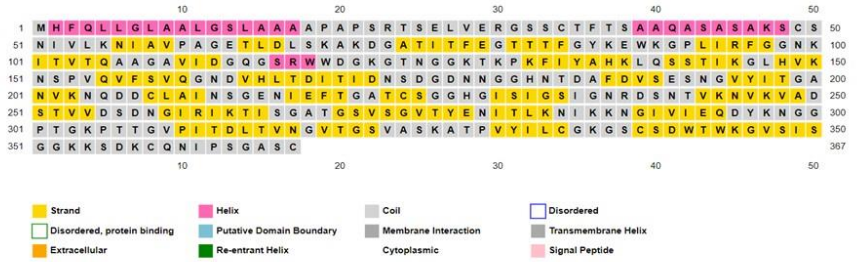
Species	**ERRAT value**	**Disallowed region**	**Qmean score**	**Favourable region**
Ao9	93.12	1	0.77	86.50%
Af7	94	0	0.83	88.90%
Rz6	96.93	0	0.72	88%
NSc1	88.3	0	0.68	84.50%

4.9 Secondary Structure Prediction of Polygalacturonases

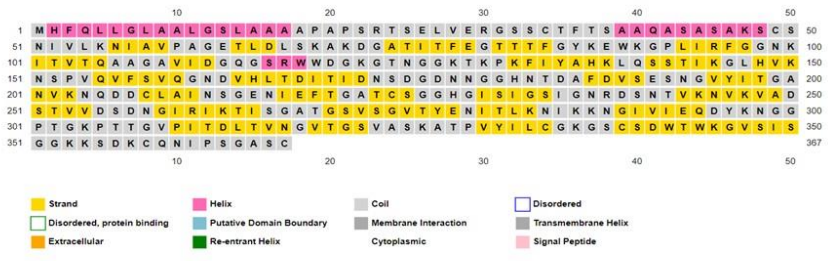
The secondary structure of proteins was predicted by the PSIPRED server when the amino acid sequence was submitted or the protein sequence fasta format submission (https://www.predictprotein.org/). The presence of structural elements like Coiled regions of the proteins, Helix and beta sheets extended strands were predicted by the PSIPRED indicated in Figure 4.9.

Proteins sequences of Ao9, Af7, Rz6 and NSc1 were visible through viewer layout predicted features. Extended strands of random coil 174 (47.41 percent), alpha helix 29 (7.90 percent), and beta-strand 171 are all indicated in the Ao9 Af7 sequence (46.5 percent). Indicate coil 337 (73.4%), alpha helix 37 (9.25%), and beta-strand 177 in the Rz6 sequence (38.6 percent). Indicate coil 207 (45 percent), alpha helix 37 (8%), and beta-strand 178 in the Nsc1 sequence (38.8 percent).Coils were larger in number among all sequences.

A.Ao9 (Model 2)



B.Af7(Model 4)



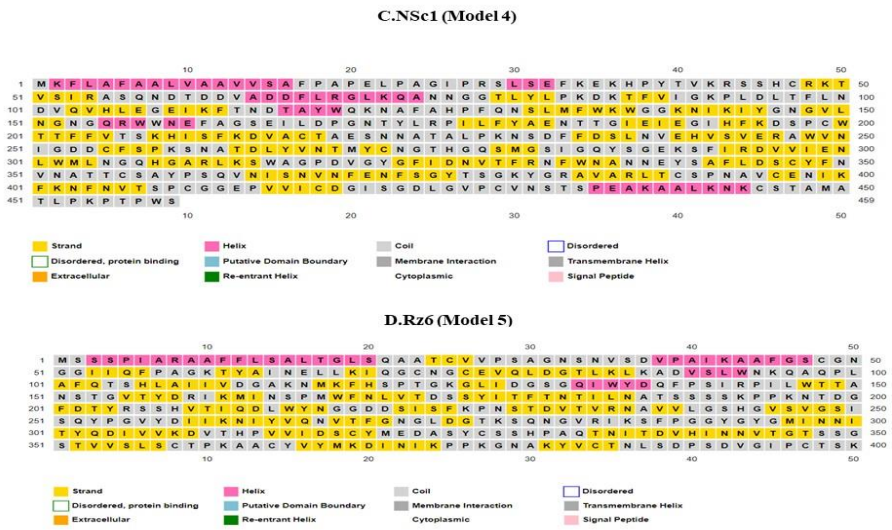


Figure 4.9. PSIPRED predicts the secondary structure of proteins. It graphically indicates the coil, helix and beta strands. Different colours distinguished strands, helix ,coils, Membrane helix and transmembrane helix.

5.0 Prediction of Tertiary Structure of Polygalacturonases

All of the models were assessed using the Qmean, the Errat score, and the Ramachandran plot. Based on the highest score, the top model was chosen. The predicted model was further improved using the 3D refined web server. Through PyMOL, graphics and visualization optimization were carried out (Figure 4.10)

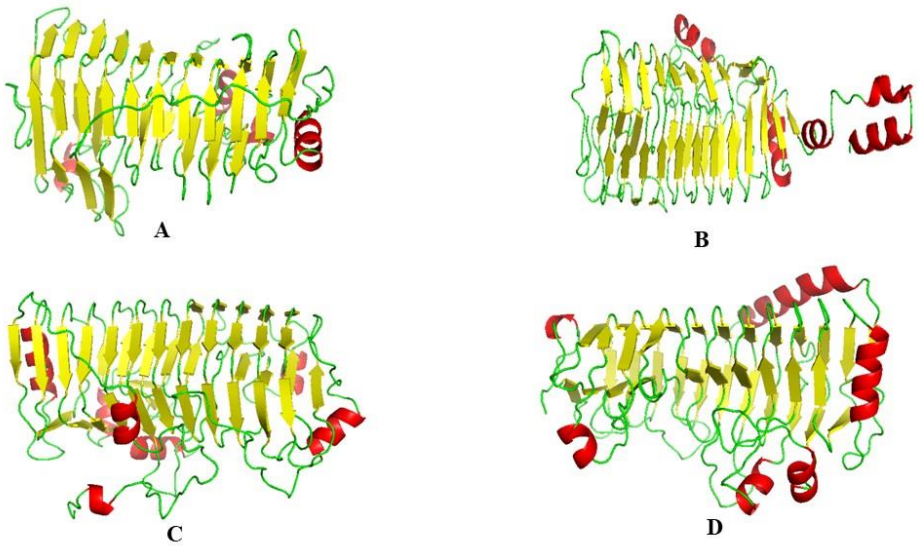


Figure 4.10. 3D structure refinement by 3D refine webserver. The 3D refined structure of Ao9 ,Af7, Nsc1 and Rz6 represents as A,B ,C and D. Red colour indicates the helix, Yellow sheet and green loops.

5.1 Superimposition of Selected Fungal Polygalacturonase with *A. aceleatus*

The protein structures were compared through superimposition. There was more structural similarity of Ao9(model 2) and Af7 (model 4)with *A. aceleatus* (Table 4.4). Proteins from the same family are more similar in their three-dimensional structure than in their basic sequences. As a result, structural similarity can typically be assumed between two proteins if similarity can be found at the sequence level (Lesk & Chothia, 1980). All the proteins were fully superimposed with each other.

Table 4.4 Species with their RMSD values .

Species	**RMSD value**
Ao	1.26Å
Af4	1.55Å
NSc1	2.289Å
Rz6	1.750Å

Following Figure 4.11 indicates the superimposed structures of four proteins.

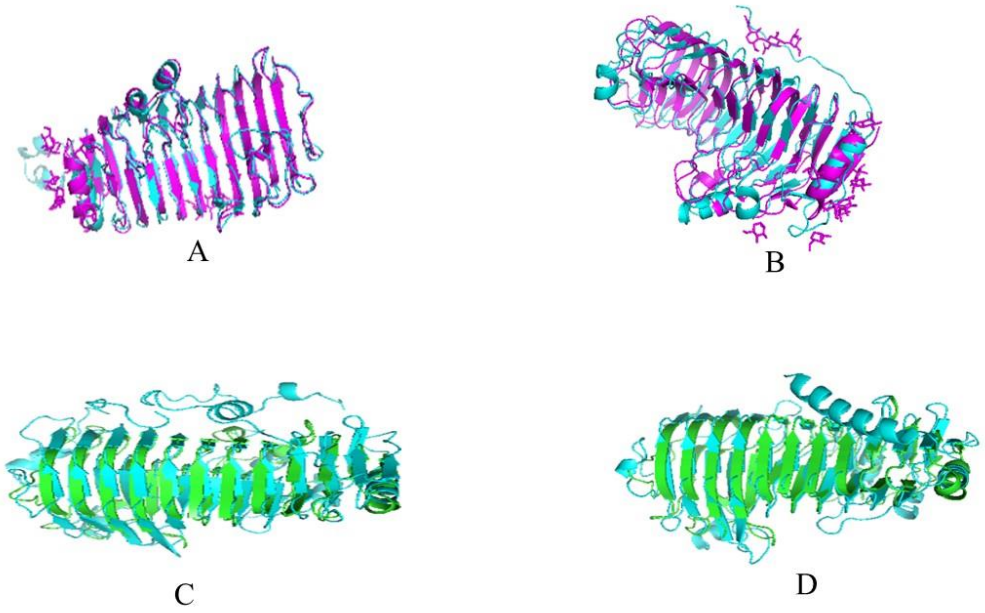


Figure 4.11. A,B,C and D represents the superimposition of Ao9 (Model 4) ,Af7 (Model 4),NSc1 (Model 4) and Rz6(Model 5) with A.aceleatus.The purple and sky blue colour used for Ao9(Model 4) and Af7 (Model 4) and Green and sky-blue colours indicates NSc1 (Model 4) and Rz6(Model 5).Water molecules were removed from sequences to obtained a refined superimposed structure.

5.2 Differential Gene Expression of Polygalacturonase in Model Fungi *Neurospora crassa*

Fungal pectinases are essential for industrial processes. However, little is currently known about how the pectinase gene's expression is regulated. A very specific collection of genes is induced by *N. crassa* which has been demonstrated to grow effectively on pectin. Our study focusses on the gene expression analysis of *N. crassa* under different growth conditions by using bioinformatic approach Edge R. Wild type strains were used by providing different media for growth. For treatments sucrose, pectin and rhamnose were used by providing different incubation time for growth . No carbon source was used for non-treated samples. Our findings showed that when given treatments, our gene Nsc1 exhibits downregulation. It demonstrates that our gene were not able degrade the pectin, sucrose and rhamnose under the transcriptional factor pdr1.

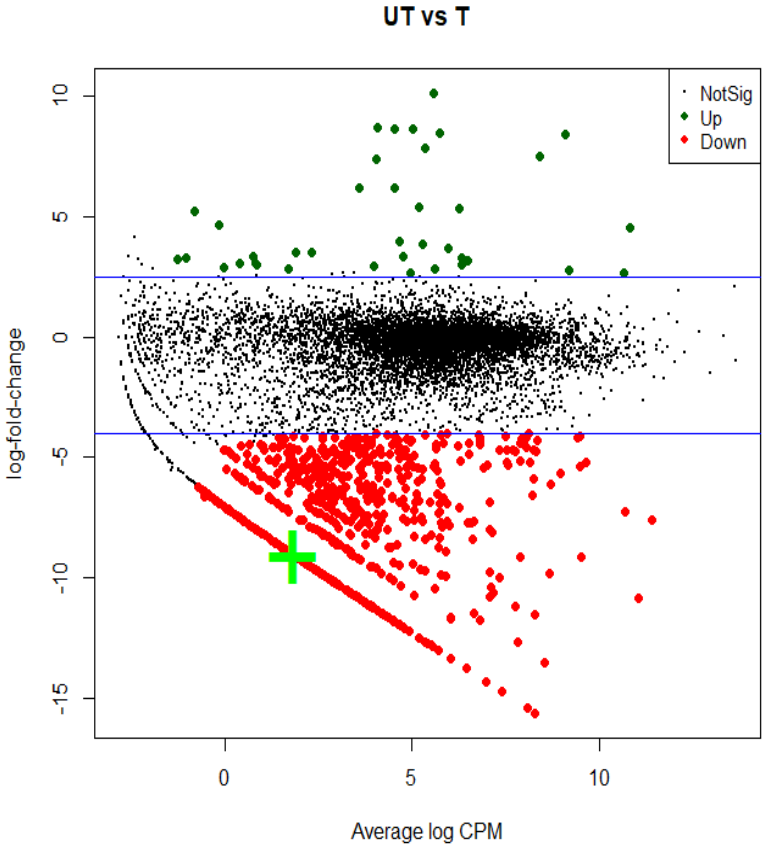
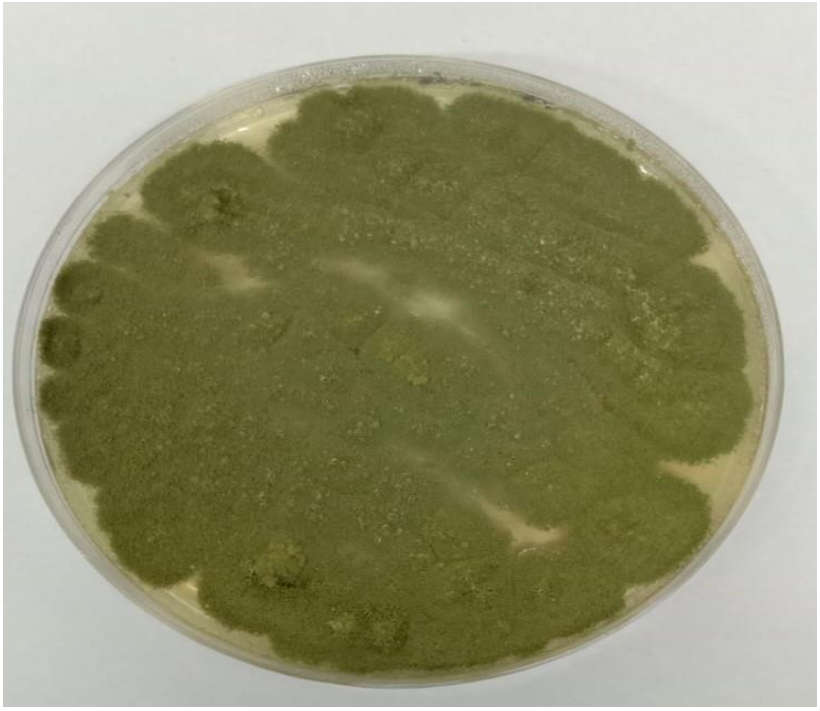


Figure 4.12. Gene expression. Red color represents the down streaming. Black color represents the non-significant part. Green color represents the up regulation. Gene of interest is represented by green positive symbol. Gene of *Neurospora crassa* shows down-regulation.

5.3 Production of Enzyme by Submerged Fermentation

A.flavus were selected for the production of enzyme PG from fungal virology lab.



Figur 4.13.Picture showing a pure colony of *A.flavus.*

5.4 Enzyme Activity Analysis

Enzyme activity assay was performed by submerged fermentation from *A.flavus* and the fermentation media recipe was mentioned in previous chapter 3.

5.5 Protein Estimation through Bradford Assay

By concentration of protein, we can estimate enzyme assay. Bradford reagent is used for the estimation of protein which acts as a standard. A graph of BSA is plotted (Figure 4.14).

Table4.1 A standard curve was plotted based on observed values.

Sample	Absorbance	Conc	Dilution factor	Actual conc	Vol for 25 ug	Vol for 50 ug
Af1	0.507	0.43	20.00	8.70	2.87	5.75

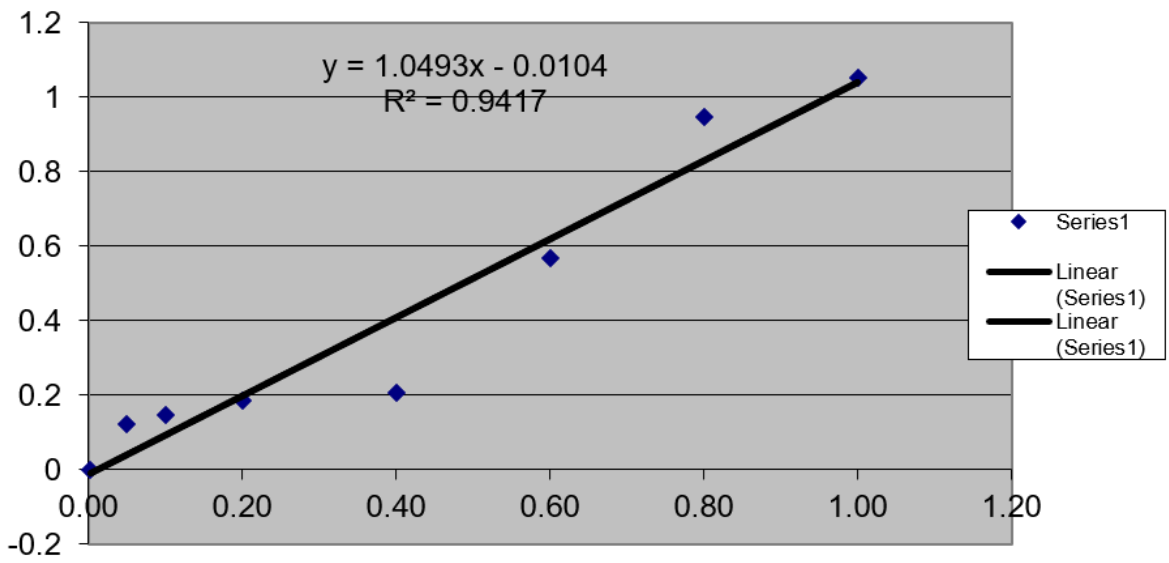


Figure 4.14. Standard BSA graph for protein estimation.

Chapter 5: Discussion

Plant cell walls contain the pectin polysaccharide which is rich in galacturonic acid residues. (Castilho et al., 2000; Mohnen, 2008).The dry mass of plants up to 25% and 35% are made by pectin. Pectin catalysis the reaction at various locations on the pectin molecule. The enzyme polygalacturonase hydrolyzes the polygalacturonic acid backbone of pectin (Devi & Rao, 1996). All polygalacturonase sequences are stored in NCBI-searchable databases. Prior research has concentrated on the heterologous production of polygalacturonase as a moderator of cell wall of plant and pathogen defense bioregulator wall pectin structure(Ferrari et al., 2008; Lionetti et al., 2010).Our research emphasizes the industrial importance of enzymes polygalacturonase. Concerning mechanism and sequence similarity, this family neatly separates into subfamilies that either comprise exo- or endo-PG (Markovič & Janeček, 2001).

Our Insilico based study identify 44 fungal polygalacturonases from four novel species. In earlier polygalcturonase from bacteria and plants were reported. Depending on the kind of species it was identified in polygalacturonase exhibits variability (Dubey et al., 2010; Yadav et al., 2009).In our study we identify *Aspergillus oryzae*, *Aspergillus flavus*, *Rhizoctonia solani* and *Neurospora crassa*. The 44 polygalcturonase were renamed based on chromosomal numbers. One of the largest hydrolase families is PGs(Markovič & Janeček, 2001). The 340 amino acids long GH28 domains present in all known PGs account for roughly 75% of the typical PG coding sequence described in the literature(Kim et al., 2006). For the confirmation of the presence of GH28 domains in the protein sequence, all fungal PG candidates were extensively analyzed using the protein family's database (Pfam).In previous study Bacillus fungi were found to produce PG. It was found enzymes which belong to GHs can degrade polysaccharides and carbohydrates. The majority of GHs that

have been discovered and studied earlier were from bacteria. In our work, it is suggested that the GHs family can be identified from different fungal species.

Every eukaryotic genome has considerable amounts of duplicated genes, and prior research has shown that the presence of gene families is one of the features of the genomes of higher eukaryotes (Horan et al., 2005).We identified that *N. crassa* has smaller number of genes and larger number of genes were identified in *R. solani*. In order to identify the evolutionary relationship among fungal PGs, a phylogeny tree was constructed using the NJ method. The tree was presented using TreeView for Windows and the bootstrapping process was carried out 100 times. It was divided into four major clade groups named A,B,C and D. In group D *A. oryzae* showed close linkage with *A. flavu*s. Group D also indicates the close evolutionary linkage of *A. oryzae* with *A. flavus*. Close evolutionary relationship among these two may indicates that *A. oryzae* may have descended from wild plant pathogen like *A. flavus* and may therefore have some enzyme that breaks down plant cell walls. Members of gene families that were created via tandem and genome duplication from a common ancestor typically retain similar functions (Chapman et al., 2004; Prince & Pickett, 2002).

The 44 proteins are split up across different cellular compartments, according to subcellular localization. Since the majority of them are likely to be located outside of cells (extracellular) they must be secretory proteins involved in the destruction of cell walls. Six conserved motifs were identified through 44 sequences. Among 44 proteins the physiochemical properties varied greatly. In determining the physiochemical properties three parameters were considered. Amino acid , Theoretical PI and molecular weight of proteins. We found that fungal polygalcturonases were acidic in nature. Whereas polygalcturonases of peal millet and guava were stable in nature (Shivashankar et al., 2010). Further to analyze the gene structure of forty-four polygalacturonase

GSDC servers illustrate the exonic, intronic, CDS and UTR regions. Every protein has a different number of introns and exons, according to research on the intronic and exonic regions. Most of the PGs' coding sequences were interrupted by the introns. In *Zea mays* the exons vary in number ranging from one to nine (Lu et al., 2021).The same results were predicted earlier in gene structure prediction of soybean PG genes (Wang et al., 2016).GO enrichment and GO modification process were performed to indicate the regulatory functions. Three groupings come under the GO terms. They are named molecular function, biological and cellular functions. In past studies for the genome wide analysis of PG in grapevine the results of Go terms indicates PG activity ,hydrolase activities and metabolic activities (Van Wyk & Divol, 2009). In our studies the molecular functions indicate the ligase, catalytic, oxireductase ,lyase and transferees activity whereas carbohydrate, lipid metabolic process and protein modification process were indicated by biological functions.All the GO terms validate the PG activity in identified fungal species. The main sources of endo PGs are fungi, bacteria, and yeast, and the Go term from fungal sources (P:GO:005975) revealed the same conclusion. To further understand the molecular basis of fungal PGs from the retrieved sequences, the 3D structure was modelled using the SWISS-MODEL service (Biasini et al., 2014). This information is useful for rational experimental design to improve enzymatic properties in future studies. The model's quality was assessed using the PROCHECK methodology. Model 2, Model 4,Model 4 and Model 5 were selected from Ao9, Af7, Nsc1 and Rz6.3D structure of *A. niger* polygalcturonase were reported before (Serrat et al., 2002) .Regulation of polygalcturonase were poorly known. Gene expression analysis of polygalacturonase in *N.crassa* were performed. The PDR1 transcriptional factor's rule were identified. Due to polygalacturonase, PDR1 aids in the breakdown of pectin. Transcriptomic data were gathered from wild type strains (Thieme et al., 2017).The transcriptional factor pdr1 were

absent in wild type strains. Because polygalacturonase was unable to dissolve the pectin, the downregulation is evident.

Further in wet lab experiments were performed to determine protein estimation and the enzyme PG activity in *A.flavus.* In previous studies polygalcturonase synthesis were also reported from *A. flavus (Cotty et al., 1990)*

Chapter 6 : Conclusion and Future Aspects

In the present study, we conclude that fungi can be a novel source of PGs. The main purpose of this study was to identify the fungal species which are rich in PGs for industrial purposes .In our current study we identify four novel fungal local strains which have the ability to produce PGs. Various computational tools help in the identification of fungal strains. *A. flavus* and *A. oryzae* show closed evolutionary relationship as compared to *N. crassa* and *R. solani*. GH28 were found to be common domain in all the 44 fungal polygalacturonase.42 proteins were found to be localized in extracellular region and other two in nuclear and cytoplasmic region. Prediction of only one gene in NSc1 by intro exon analysis. Homology modelling predicts best models of proteins. The four best models of Ao, Af, NSc1 and Rz were future superimposed with *A. aceleutus*.

Pectinases are essential for industrial processes. However, little is currently known about how the pectinase genes expression is regulated. Hence, in this study, we opted very specific collection of genes in *N. crassa* which has been demonstrated to grow effectively on pectin. This study focused on the gene expression analysis of *N. crassa* under different growth conditions by using bioinformatic approach Edge R. Gene expression of *N. crassa* under different growth medium show downregulation under the absence of transcriptional factor pdr1.

The combined phylogenetic classification and transcriptome analysis in this study provides a crucial reference for future biochemical characterization of new fungal strains. The transcriptomic data helps in understanding the rule of polygalacturonase in different fungal media. The genome-wide investigation described here is crucial to better understand the relationship among different fungal strains. The study will help in the rational engineering of industrial fungal species.

To conclude, this study can be implemented to other filamentous and non-filamentous fungal species to expand our knowledge on pectinase gene regulation. After that we proceed to the wet lab to identify the fungal strains that have the potential to produce PGs .Enzyme assay and protein estimation were performed.

References

Adrio, J. L., & Demain, A. L. (2014). Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules*, *4*(1), 117-139. https://www.mdpi.com/2218-273X/4/1/117

Ahlawat, S., Dhiman, S. S., Battan, B., Mandhan, R., & Sharma, J. (2009). Pectinase production by Bacillus subtilis and its potential application in biopreparation of cotton and micropoly fabric. *Process Biochemistry*, *44*(5), 521-526.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Amilia, K., Sari, S., & Setyaningsih, R. (2017). Isolation and screening of pectinolytic fungi from orange (Citrus nobilis Tan.) and banana (Musa acuminata L.) fruit peel. IOP Conference Series: Materials Science and Engineering,

Amin, F., Bhatti, H. N., Bilal, M., & Asgher, M. (2017). Multiple parameter optimizations for enhanced biosynthesis of exo-polygalacturonase enzyme and its application in fruit juice clarification. *International Journal of Food Engineering*, *13*(2).

Atkinson, R. G., Schröder, R., Hallett, I. C., Cohen, D., & MacRae, E. A. (2002). Overexpression of polygalacturonase in transgenic apple trees leads to a range of novel phenotypes involving changes in cell adhesion. *Plant Physiology*, *129*(1), 122-133.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Research*, *28*(1), 304-305. https://doi.org/10.1093/nar/28.1.304

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic acids research*, *28*(1), 45-48.

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343-350.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L. (2000). GenBank. *Nucleic acids research*, *28*(1), 15-18.

Berdy, J. (2005). Bioactive microbial metabolites. *The Journal of antibiotics*, *58*(1), 1-26.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer Jr, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of molecular biology*, *112*(3), 535-542.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., & Bordoli, L. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, *42*(W1), W252-W258.

Castilho, L. R., Medronho, R. A., & Alves, T. L. (2000). Production and extraction of pectinases obtained by solid state fermentation of agroindustrial residues with Aspergillus niger. *Bioresource Technology*, *71*(1), 45-50.

Chaplin, M., & Bucke, C. (1990). The large-scale use of enzymes in solution. *Enzyme technology*, 138-166.

Chapman, B. A., Bowers, J. E., Schulze, S. R., & Paterson, A. H. (2004). A comparative phylogenetic approach for dating whole genome duplication events. *Bioinformatics*, *20*(2), 180-185.

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, *32*(suppl_1), D258-D261.

Cotty, P., Cleveland, T., Brown, R., & Mellon, J. (1990). Variation in polygalacturonase production among Aspergillus flavus isolates. *Applied and Environmental Microbiology*, *56*(12), 3885-3887.

Demain, A. L., & Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *The Journal antibiotics*, *62*(1), 5-16.

Desiato, F., Fioravanti, G., Fraschetti, P., Perconti, W., & Piervitali, E. (2012). ISPRA (Istituto Superiore per la protezione e la ricerca ambientale, Roma). *Stato dell'Ambiente*, *35*(2011), 162.

Dubey, A. K., Yadav, S., Kumar, M., Singh, V. K., Sarangi, B. K., & Yadav, D. (2010). In silico characterization of pectate lyase protein sequences from different source organisms. *Enzyme Research*, *2010*.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*, 571-607.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., & Gentry, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, *5*(10), 1-16.

Glass, N. L., Schmoll, M., Cate, J. H., & Coradetti, S. (2013). Plant cell wall deconstruction by ascomycete fungi. *Annual review of microbiology*, *67*, 477-498.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, *36*(10), 3420-3435.

Guerrand, D. (2018). Economics of food and feed enzymes: Status and prospectives. In *Enzymes in human and animal nutrition* (pp. 487-514). Elsevier.

Gurung, N., Ray, S., Bose, S., & Rai, V. (2013). A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. *BioMed Research International*, *2013*, 329121. https://doi.org/10.1155/2013/329121

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical journal*, *280*(2), 309-316.

Horan, K., Lauricha, J., Bailey-Serres, J., Raikhel, N., & Girke, T. (2005). Genome cluster database. A sequence family analysis platform for Arabidopsis and rice. *Plant physiology*, *138*(1), 47-54.

Idnurm, A., & Howlett, B. J. (2001). Pathogenicity genes of phytopathogenic fungi. *Molecular Plant Pathology*, *2*(4), 241-255.

Jayani, R. S., Shukla, S. K., & Gupta, R. (2010). Screening of bacterial strains for polygalacturonase activity: its production by Bacillus sphaericus (MTCC 7542). *Enzyme research*, *2010*.

Jenkins, J., Mayans, O., & Pickersgill, R. (1998). Structure and evolution of parallel β-helix proteins. *Journal of structural biology*, *122*(1-2), 236-246.

Kango, N., Jana, U. K., & Choukade, R. (2019). Fungal enzymes: sources and biotechnological applications. In *Advancing frontiers in mycology & mycotechnology* (pp. 515-538). Springer.

Kashyap, D., Vohra, P., Chopra, S., & Tewari, R. (2001). Applications of pectinases in the commercial sector: a review. *Bioresource technology*, *77*(3), 215-227.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic acids research*, *37*(suppl_1), D387-D392.

Lang, C., & Dörnenburg, H. (2000). Perspectives in the biological function and the technological application of polygalacturonases. *Applied Microbiology and Biotechnology*, *53*(4), 366-375.

Leh, D. S., Biz, A., De Paula, D., Richard, P., Gonçalves, A. G., Noseda, M. D., Mitchell, D. A., & Krieger, N. (2017). Conversion of citric pectin into D-galacturonic acid with high substrate loading using a fermented solid with pectinolytic activity. *Biocatalysis and Agricultural Biotechnology*, *11*, 214-219.

Lesk, A. M., & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *Journal of molecular biology*, *136*(3), 225-270.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, *44*(W1), W242-W245.

Lu, L., Hou, Q., Wang, L., Zhang, T., Zhao, W., Yan, T., Zhao, L., Li, J., & Wan, X. (2021). Genome-Wide Identification and Characterization of Polygalacturonase Gene Family in Maize (Zea mays L.). *International journal of molecular sciences*, *22*(19), 10722.

Ma, Y., Sun, S., Hao, H., & Xu, C. (2016). Production, purification and characterization of an exo-polygalacturonase from Penicillium janthinellum sw09. *Anais da Academia Brasileira de Ciências*, *88*, 479-487.

Magni, A., & Carlo, N. (2019). *Innovazione e sostenibilità nell'industria tessile*. goWare & Guerini Next.

Mandhania, S., Jain, V., & Malhotra, S. (2010). Culture optimization for enhanced production of microbial pectin methylesterase under submerged conditions. *Asian J Biochem*, *5*(1), 12-22.

Manucharova, N. (2009). The microbial destruction of chitin, pectin, and cellulose in soils. *Eurasian Soil Science*, *42*(13), 1526-1532.

Markovič, O., & Janeček, Š. (2001). Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Engineering*, *14*(9), 615-631.

Martins, E., Silva, D., Da Silva, R., & Gomes, E. (2002). Solid state production of thermostable pectinases from thermophilic Thermoascus aurantiacus. *Process Biochemistry*, *37*(9), 949-954.

McNeil, M., Darvill, A. G., Fry, S. C., & Albersheim, P. (1984). Structure and function of the primary cell walls of plants. *Annual review of biochemistry*, *53*(1), 625-663.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current opinion in plant biology*, *11*(3), 266-277.

Nakkeeran, E., Subramanian, R., & Umesh-Kumar, S. (2010). Purification of polygalacturonase from solid-state cultures of Aspergillus carbonarius. *Journal of bioscience and bioengineering*, *109*(2), 101-106.

Page, R. D. (1996). Tree View: An application to display phylogenetic trees on personal computers. *Bioinformatics*, *12*(4), 357-358.

Pandey, A., Soccol, C. R., Nigam, P., & Soccol, V. T. (2000). Biotechnological potential of agro-industrial residues. I: sugarcane bagasse. *Bioresource technology*, *74*(1), 69-80.

Peberdy, J. F. (1994). Protein secretion in filamentous fungi—trying to understand a highly productive black box. *Trends in biotechnology*, *12*(2), 50-57.

Pérez-Victoria, I., Martín, J., & Reyes, F. (2016). Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta medica*, *82*(09/10), 857-871.

Pilnik, W., & Voragen, A. G. (1993). Pectic enzymes in fruit and vegetable juice manufacture. *Enzymes in food processing*, *3*, 363-399.

Prade, R. A., Zhan, D., Ayoubi, P., & Mort, A. J. (1999). Pectins, pectinases and plant-microbe interactions. *Biotechnology and genetic engineering reviews*, *16*(1), 361-392.

Prince, V. E., & Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics*, *3*(11), 827-837.

Punt, P. J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J., & van den Hondel, C. (2002). Filamentous fungi as cell factories for heterologous protein production. *Trends in biotechnology*, *20*(5), 200-206.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, *26*(1), 139-140.

Rombouts, F. M., & Pilnik, W. (1980). Pectic enzymes. In *Economic Microbiology, Vol. 5: Microbial Enzymes and Bioconversions* (pp. 227-282). Academic Press.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406-425.

Sarkanen, S. (1991). Enzymatic lignin degradation: an extracurricular view. In. ACS Publications.

Serrat, M., Bermúdez, R. C., & Villa, T. G. (2002). Production, purification, and characterization of a polygalacturonase from a new strain of Kluyveromyces marxianus isolated from coffee wet-processing wastewater. *Applied biochemistry and biotechnology*, *97*(3), 193-208.

Seyedsayamdost, M. R., & Clardy, J. (2014). Natural products and synthetic biology. In (Vol. 3, pp. 745-747): ACS Publications.

Shivashankar, S., Thimmareddy, C., & Roy, T. K. (2010). Polygalacturonase inhibitor protein from fruits of anthracnose resistant and susceptible varieties of Chilli (Capsicum annuum L).

Siddiqui, N. H., Azhar, I., Tarar, O. M., Masood, S., & Mahmood, Z. A. (2015). Influence of pectin concentrations on physico-chemical and sensory qualities of jams. *World Journal of Pharmacy and Pharmaceutical Sciences*, *4*(6), 68-77.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., & Mungall, C. J. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, *25*(11), 1251-1255.

Soares, M., Da Silva, R., Carmona, E., & Gomes, E. (2001). Pectinolytic enzyme production by Bacillus species and their potential application on juice extraction. *World Journal of Microbiology and Biotechnology*, *17*(1), 79-82.

Souza, P. M. d., Bittencourt, M. L. d. A., Caprara, C. C., Freitas, M. d., Almeida, R. P. C. d., Silveira, D., Fonseca, Y. M., Ferreira Filho, E. X., Pessoa Junior, A., & Magalhães, P. O. (2015). A biotechnology perspective of fungal proteases. *Brazilian Journal of Microbiology*, *46*, 337-346.

Tedesco, S., & Montacchini, E. (2020). From Textile Waste to Resource: A Methodological Approach of Research and Experimentation. *Sustainability*, *12*(24), 10667.

Thieme, N., Wu, V. W., Dietschmann, A., Salamov, A. A., Wang, M., Johnson, J., Singan, V. R., Grigoriev, I. V., Glass, N. L., & Somerville, C. R. (2017). The transcription factor PDR-1 is a multi-functional regulator and key component of pectin deconstruction and catabolism in Neurospora crassa. *Biotechnology for biofuels*, *10*(1), 1-21.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, *22*(22), 4673-4680.

Van der Vlugt-Bergmans, C., Meeuwsen, P., Voragen, A., & Van Ooyen, A. (2000). Endo-xylogalacturonan hydrolase, a novel pectinolytic enzyme. *Applied and environmental microbiology*, *66*(1), 36-41.

Van Wyk, H., & Divol, B. (2009). Recovery of endo-polygalacturonase activity in wine yeast and its effect on wine aroma. *FEMS yeast research, 10*(1), 58-71.

Vivek, K., Mishra, S., & Pradhan, R. C. (2019). Optimization of ultrasound-assisted enzymatic extraction of Sohiong (Prunus nepalensis) juice. *Journal of Food Process Engineering*, *42*(1), e12948.

Voragen, A. G., Coenen, G.-J., Verhoef, R. P., & Schols, H. A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry, 20*(2), 263-275.

Wang, F., Sun, X., Shi, X., Zhai, H., Tian, C., Kong, F., Liu, B., & Yuan, X. (2016). A global analysis of the polygalacturonase gene family in soybean (Glycine max). *PLoS One*, *11*(9), e0163012.

Yadav, P., Singh, V., Yadav, S., Yadav, K., & Yadav, D. (2009). In silico analysis of pectin lyase and pectinase sequences. *Biochemistry (Moscow), 74*(9), 1049-1055.

Zheng, Z., & Shetty, K. (2000). Solid state production of polygalacturonase by Lentinus edodes using fruit processing wastes. *Process biochemistry*, *35*(8), 825-830.

Appendix

Table 4A Fungal proteins with their RNA IDs ,Gene IDs and chromosomal location.

Protein IDs	Specie name	RNA IDs	Gene IDs	Chromosomal location
XP_001818327.1	*A.oryzae*	XM_001818275.1	395231	1
XP_023088847.1	*A.oryzae*	XM_023234227.1	5989233	1
XP_001817187.1	*A.oryzae*	XM_001817135.1	5990500	1
XP_023088704.1	*A.oryzae*	XM_023235763.1	5988924	1
XP_001818685.1	*A. oryzae*	XM_001818633.1	5990656	2
XP_001819695.1	*A. oryzae*	XM_001819643.1	5992583	2
XP_023091173.1	*A.oryzae*	XM_023235910.1	5994218	3
XP_001821616.1	*A. oryzae*	XM_001821564.1	3	
XP_001820953.1	*A.oryzae*	XM_001820901.3	5992955	3
XP_001820739.1	*A. oryzae*	XM_001820687.3	5992741	3
XP_001821735.1	*A.oryzae*	XM_001821683.1	5993763	3
XP_001822260.1	*A.oryzae*	XM_001822208.1	5994305	4
XP_023091689.1	*A.oryzae*	XM_023236963.1	5994418	4
XP_023092652.1	*A. oryzae*	XM_023237868.1	5996020	5
XP_001825581.3	*A. oryzae*	XM_001825529.3	5997682	6
XP_023093237.1	*A.oryzae*	XM_023238562.1	35119820	6
XP_041140580.1	*A.flavus*	XM_041284017.1	64843054	1
XP_041140162.1	*A. flavus*	XM_041287908.1	64842636	1
XP_041141640.1	*A.flavus*	XM_041288979.1	64844114	1
XP_041141746.1	*A.flavus*	XM_041289074.1	64844220	1

XP_041141904.1	*A.flavus*	XM_041289241.1	64844422	1
XP_041142957.1	*A.flavus*	XM_041284780.1	64845475	1
XP_041144219.1	*A.flavus*	XM_041290610.1	64846783	3
XP_041144806.1	*A. flavus*	XM_041291041.1	64847370	3
XP_041149670.1	*A. flavus*	XM_041294346.1	64852365	6
XP_041149181.1	*A.flavus*	XM_041286854.1	64851876	6
XP_962601.3	*N. crassa*	XM_957508.3	3878749	4
XP_043177002.1	*R.solani*	XM_043324583.1	67027046	0
XP_043177842.1	*R.solani*	XM_043322346.1	67024809	3
XP_043179490.1	*R.solani*	XM_043320478.1	67022941	4
XP_043181268.1	*R.solani*	XM_043325836.1	67028299	6
XP_043181699.1	*R.solani*	XM_043326267.1	67028730	7
XP_043181767.1	*R.solani*	XM_043326335.1	67028798	7
XP_043182617.1	*R. solani*	XM_043329783.1	67032246	8
XP_043182893.1	*R.solani*	XM_043327508.1	67029971	9
XP_043183052.1	*R.solani*	XM_043327667.1	67030130	9
XP_043183074.1	*R.solani*	XM_043327689.1	67030152	9
XP_043183898.1	*R.solani*	XM_043328513.1	67030976	10
XP_043184494.1	*R.solani*	XM_043330397.1	67032860	11
XP_043185038.1	*R.solani*	XM_043331528.1	67033991	12
XP_043185246.1	*R. solani*	XM_043326774.1	67029237	13
XP_043185738.1	*R.solani*	XM_043327266.1	67029729	13
XP_043185788.1	*R.solani*	XM_043327316.1	67029779	13

XP_043187080.1	*R.solani*	XM_043321257.1	67023720	16

Table 4B Cleavage sites and domain labels of fungal proteins .

Specie	**Cleavage sites**		**Labels**
Ao1	21	22	1
Ao3	18	19	2
Ao4	17	18	3
Ao5	20	21	4
Ao6	15	16	5
Ao8	21	22	6
Ao9	17	18	7
Ao10	20	21	8
Ao11	19	20	9
Ao12	18	19	10
Ao13	18	19	11
Ao14	15	16	12
Ao16	24	15	13
Af1	21	22	14
Af2	23	24	15
Af3	16	17	16
Af4	18	19	17
Af5	24	25	18
Af6	15	16	19
Af7	17	18	20
Af8	21	22	21
Af9	20	21	22
Af10	19	20	23
NSc1	16	17	24
Rz1	19	20	25
Rz2	19	20	26
Rz3	19	20	27
Rz4	16	17	28
Rz5	20	21	29
Rz6	22	23	30
Rz7	18	19	31
Rz8	19	20	32
Rz9	17	18	33
Rz10	17	18	34
Rz11	17	18	35

Rz12	16	17	36
Rz15	16	17	37
Rz16	16	17	38
Rz17	17	18	39

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Laiba Amin
Assignment title:	Manuscript MS Laiba
Submission title:	MS thesis Laiba Amin
File name:	Final_thesis_Refined_version_3-11-22.docx
File size:	2.11M
Page count:	76
Word count:	13,613
Character count:	78,470
Submission date:	04-Nov-2022 03:44AM (UTC-0700)
Submission ID:	1944346461

Genome-wide Identification and Comparative Analysis of polygulacturonases in different fungal species

Laiba Amin

Regn Number 328971

Master of Science in Plant Biotechnology

Atta-Ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology

Islamabad, Pakistan

2022

Dr. Muhammad Faraz Bhatti
Tenured Professor
Head of Department (HoD)
Plant Biotechnology
Atta-Ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Copyright 2022 Turnitin. All rights reserved.

Scanned with CamScanner

MS thesis Laiba Amin

ORIGINALITY REPORT

17%	14%	10%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Higher Education Commission Pakistan Student Paper	3%
2	ru.scribd.com Internet Source	1%
3	Submitted to University of Warwick Student Paper	1%
4	Submitted to University of Leeds Student Paper	1%
5	mafiadoc.com Internet Source	1%
6	www.ncbi.nlm.nih.gov Internet Source	<1%
7	Luis J. del Valle, Marta Oños, Pere Garriga, Margarita Calafell, Wolfgang Schnitzhofer, Georg M. Guebitz. "Bioscouring of Cotton Fiber with Polygalacturonase Induced in Sclerotium rolfsii using Cellulose and Glucose-pectin", Textile Research Journal, 2016 Publication	<1%

Dr. Muhammad Faraz Bhatti
Assistant Professor

CS Scanned with CamScanner

biochemical, thermodynamic, and fruit juice clarification properties of lignocellulosic biomass-derived exo-polygalacturonase from Penicillium paxilli", Biomass Conversion and Biorefinery, 2022

Publication

93 Perry Ridge, Hyrum Carroll, Dan Sneddon, Mark Clement, Quinn Snell. "Large Grain Size Stochastic Optimization Alignment", Sixth IEEE Symposium on BioInformatics and BioEngineering (BIBE'06), 2006 <1%

Publication

Exclude quotes Off

Exclude bibliography On

Exclude matches Off

Dr. Muhammad Faraz Bhatti
Tenured Professor
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Scanned with CamScanner