INTER-PHYLA COMPARATIVE GENOMICS OF PLANT PATHOGENIC FUNGUS REVEALS GENOMICS SIMILARITY AND COPY NUMBER VARIATIONS

KENNETH TAN LEE SHEAN

(SGJ130006)

SUBMITTED TO

INSTITUTE OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

IN PARTIAL FULFILMENT

OF THE REQUIREMENT FOR

THE DEGREE OF MASTER OF BIOINFORMATICS

2014

UNIVERSITY MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: KENNETH TAN LEE SHEAN(I.C/Passport No: 890516-14-5815)

Registration/Matric No: SGJ130006

Name of Degree: MASTER OF BIOINFORMATICS

Inter-Phyla Comparative Genomics of Plant Pathogenic Fungus Reveals Genomics Similarities and Copy Number Variations ("this Work")

Field of Study: Bioinformatics, Comparative Genomics, Fungal Genomics, Plant Pathogenic Fungus

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:

Designation:

ABSTRACT

The impact of plant pathogenic fungus towards the industry of agriculture causes massive destruction of crops worldwide and thus stirring interest around the world for research in these plant pathogenic funguses. The vast development of sequencing technologies enabled many public efforts to decipher the genomics information about these funguses and example of such effort is the Fungal Genome Initiative (FGI) by Broad Institute. By using public genomics and annotation data from FGI for four different fungus species from two different phyla, inter-phyla comparative genomics between the fungus species revealed important common features of plant pathogenic funguses. Inter-Phyla comparative genomics results showed that there are 1,388 homologous protein-coding genes between Basidiomycota and Ascomycota. Also done was discovery of candidate protein-coding genes from both pathogen-host interaction-related genes and carbohydrate-active enzymes, which are known as protein-coding genes that are related to fungus pathogenicity. A total of 159 common candidate protein-coding PHI-base genes and 64 common candidate protein-coding genes for carbohydrate-active enzymes were identified between fungus from Basidiomycota and Ascomycota. Genes Copy Number Variation was also observed in both pathogenicity related genes discovery, with 5 candidate PHI-related genes and 3 candidates CAZy showed to have variation in terms of genes copy number. Also discovered is the significant difference in total number of pathogenicity genes between Ascomycetes and Basidiomycetes where Ascomycetes is found to have more copy number of pathogenicity-related genes thatn Basdiomycetes. This research could lead to development of broad-spectrum antifungal solution for the agricultural industry by targeting the common genes identified.

ABSTRAK

Kesan tumbuhan kulat patogen ke atas industri pertanian menyebabkan kemusnahan besar tanaman di seluruh dunia dan dengan itu kepentingan kacau di seluruh dunia untuk penyelidikan dalam loji kulat patogenik. Pembangunan luas teknologi penjujukan membolehkan banyak usaha awam untuk mentafsirkan maklumat genomik yang mengenai funguses dan contoh usaha itu Inisiatif Genom Kulat (FGI) oleh Broad Institute. Dengan menggunakan genomik awam dan data anotasi dari FGI selama empat spesies kulat yang berbeza dari dua Filum berbeza, antara Filum genomik perbandingan antara spesies kulat mendedahkan ciri-ciri biasa yang penting tumbuhan kulat patogenik. Hasil kajian daripada perbandingan genomik antara Filum yang sama menunjukkan bahawa terdapat 1,388 homolog protein-pengekodan gen antara Basidiomycota dan Ascomycota. Juga dilakukan adalah penemuan calon protein-pengekodan gen daripada kedua-dua patogen-tuan rumah interaksi yang berkaitan dengan gen dan enzim karbohidrat-aktif, yang dikenali sebagai protein-pengekodan gen yang berkaitan dengan kulat pathogenik Sebanyak 159 calon biasa protein-pengekodan gen PHI-asas dan 64 calon biasa protein-pengekodan gen untuk enzim karbohidrat-aktif telah dikenal pasti antara kulat dari Basidiomycota dan Ascomycota. Perbezaan dalam bilangan gen juga diperhatikan dalam kedua-dua pathogenicity berkaitan penemuan gen, dengan 5 gen calon PHI Berkaitan dan 3 calon CAZy menunjukkan untuk mempunyai variasi dari segi gen menyalin nombor. Juga mendapati perbezaan yang ketara dalam jumlah bilangan gen pathogenik antara Ascomycetes dan Basidiomycetes mana Ascomycetes didapati mempunyai bilangan salinan lebih daripada pathogenik berkaitan gen thatn Basdiomycetes. Kajian ini boleh membawa kepada pembangunan spektrum luas penyelesaian antikulat untuk industri pertanian dengan mensasarkan gen yang dikenal pasti.

ACKKNOWLEDGEMENT

I would like to thank my supervisor Dr Saharuddin bin Mohamad for his guidance and insight throughout the course of my preparation, execution, and compilation of my research work and report writing. Also would like to thank Professor Amir F. Merican for his approval of my project proposal.

Beside I would also like to thank ACGT Sdn Bhd for providing compute facilities for many bioinformatics analysis taken during the course of the project which includes genome assembly, genome mapping, comparative bioinformatics, copy number variation analysis, single nucleotide polymorphism analysis, homology searches and etc.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

BGI	Beijing Genomics Institute
DOE	Department of Energy
JGI	Joint Genomics Institute
NCBI	National Center for Biotechnology Information
SRA	Sequence Reads Archive
CAZy	Carbohydrate-Active Enzyme Database
SNP	Single Nucleotide Polymorphism
INDEL	Insertion and Deletions
CNV	Copy Number Variation
BWA	Burrows Wheeler Aligner
PHI-base	Pathogen Host Interaction Database
HSP	Highest Scoring Pair

CHAPTER 1

INTRODUCTION

1.1 Research Interest and Objectives

Recent advancement in genome sequencing technologies and bioinformatics tools and applications allows new research initiatives such as Genome 10K (Genome 10K, 2009), 10,000 Microbial Genome Project (BGI, 2011) as well as the Fungal Genome Initiative (Broad Institute, 2014) to embark into sequencing projects of different organisms, understanding the importance of genome data towards molecular studies. Importance of genomics studies triggered large scale of sequencing of broad range of organisms, from human to bacteria and fungal genomics is one of the major group of organisms gathering enormous interest in genomics studies due to their impact and importance to the ecosystem. Research initiative and institute such as the Fungal Genome Initiative (Broad Intitute, 2014) and Joint Genome Institute, United States Department of Energy (DOE, JGI, 2014) have made genome data and the annotation data accessible to the scientific community, which helps to accelerate genomics studies on various organisms of interest and the National Center for Biotechnology Information also maintains the Sequence Read Archive (NCBI, 2009) that stores sequence data from wide range of sequencing projects.

Bioinformatics tools and applications had been developed to cater for various research requirement and objectives, facilitating the continuous development in genomics research. With various available tools high impact bioinformatics research on public available data could lead to important biological findings, these findings then can serve as a guide for experimental validation.

Fungal pathogenicity had long been a difficult issue to tackle but with the help of genome sequencing technologies and bioinformatics tools and applications more insights to the pathogenicity could be revealed. The objective of this research is to uncover relationship between fungal pathogens originating from different phyla via comparative genomics.

CHAPTER 2

LITERATURE REVIEW

2.1 Fungal Pathogenicity Overview

Over the years fungus had been widely studied for various purposes due to the benefit of consumption of funguses for health enhancement. However fungus also had been a key agent to many diseases in human, animal, and plants, which affect a broad range of organisms. About 300 of 1.5 million different species (Hawksworth, D.L., 2001) of fungi on earth are known to cause diseases in human (Garcia-Solache, M.A. *et al*, 2010) and in plants particular agricultural important crops, the effects of fungi inflicted plant diseases cause massive destruction of important crops.

Each year fungal infection destroys approximately 125 million tons of world top five food crops: rice, wheath, maize, potatoes, and soybean (Fisher, M.C. *et al*, 2011) and causes loss of billions of dollars in agriculture industry. One example of such devastating impact caused by fungus is the Rice Blast, which is caused by an ascomycete fungus *Magnaporthe oryzae* (Dean, R.A. *et al*, 2005). Study of fungal pathogenicity in plants is vital for eradication of plant fungal infections with then could prevent massive destruction of crops, which is key for the survival of human race.

These pathogenic funguses have been widely studied for their role in diseases and are known to originate from two major phyla in the kingdom of fungi, namely Basidiomycota and Ascomycota. Members of these two major phyla had collectively contributed to numerous plant diseases, infecting wide range of plants including a

number of important staple food stock for human population such as maize, wheat, rice, potatoes and etc.

A recent review of plant pathogenic fungus (Dean, R. *et al* 2012) revealed the top 10 fungal pathogens in molecular plant pathology, also shortlisting other pathogenic fungus that caused damages to different types of plant species. The ranking, voted by the international community, which combined to a total of 495 votes (Dean, R. *et al*, 2012) resulting in a top 10 ranking in Table 2.1:

Table 2.1: Ranking of Top 10 Fungal Pathogens (Dean, R. *et al*, 2012)

Ranking	**Fungus**
1	*Magnaporthe oryzae*
2	*Botrytis cinerea*
3	*Puccinia* spp.
4	*Fusarium graminearum*
5	*Fusarium oxysporum*
6	*Blumeria graminis*
7	*Mycosphaerella graminicola*
8	*Colletotrichum* spp.
9	*Ustilago maydis*
10	*Melamspora lini*

This list of Top 10 Fungal Pathogens comprises of members from both Basidiomycota as well as the Ascomycota, indicating different and wide range of infectious model observable in nature. Of these 10 listed fungal pathogens, *Magnaporthe oryzae*, *Botrytis cinerea*, *Puccinia graminis,* and *Ustilago maydis* are shortlisted for further study.

2.1.1 *Magnaporthe oryzae*

Magnaporthe oryzae is well known for its role in the outbreak of rice blast disease,, causing destruction of rice that could feed 60 million people a year (Dean, R.A. *et al*, 2005). Various studies has given a glimpse of the mode of infection of *Magnaporthe oyzae* revealing clues on what maybe the root cause of the rice blast infection to rice. From the study of life cycle of *Magnaporthe oryzae* infections happens as the spores of the fungus lands and adhere to the leaves by releasing an adhesive from the top of the spores. (Hamer, J.E. *et al*, 1988). These spores then germinates and develop into a special infection cell known as appressorium that causes extremely high turgor pressure that would cause the left cuticle to rupture, then allowing invasion of the infection cell in to the rest of the leaf tissues (Dean, R.A., 1997). Colonization of the leaf will lead to disease lesions where the fungus sporulates, spreding the disease to other plants.

The genome sequences of *Magnaporthe oryzae* had been sequenced and published (Dean, R.A. *et al*, 2005) which reveals genome details of this fatal pathogenic fungus as shown in Table 2.2. These genomics discovery provides an opportunity for researches to have an in-depth understanding about the genomics features of the fungus, increasing the resolution of research to pinpoint the disease-causing factor behind the fatal fungus.

Table 2.2: Genome Details of *Magnaporthe oryzae* (Dean, R.A. *et al*, 2005).

General genome features	Value
Size (bp)	37,878,070
Chromosomes	7
(G+C) percentage	51.6
Protein-coding genes	11,109
tRNA genes	316
Per cent coding	40.5
Average gene size (bp)	1,683
Average intergenic distance (bp)	1,503
Conserved hypothetical proteins	8,868 (79%)
Predicted proteins	2,233 (20%)

In comparison to *Neurospora crassa* and *Aspergillus nidulans*, which are both related pyrenomycete and non-plant pathogenic in nature, *Magnaporthe oryzae* contains more genes in comparison with those two species.

2.1.2 *Botrytis cinerea*

Also known as grey mould, *Botrytis cinerea* is known as a nectrotroph that infects host through programmed cell death pathways (van Baarlen *et al*, 2007). Difficulties in pinpointing cost and effect of infections inflicted by *Botrytis cinerea* is difficult as the fungus known to have a broad host range and specially effective in infecting mature or senescent tissues of dicotyledonous hosts (Dean, R. *et al*, 2012). The fungus may remain dormant until external environment becomes favorable and infections happen throughout the stage of development of a plant, from seedling to fruiting.

The control of *Botrytis cinerea* is vital to world economy due to the ability of the fungus to infect a broad range of plant hosts, particularly economically important crops. Fungicide remains important measure for *Botrytis cinerea* containment, however with the extensive use of fungicide against the pathogen it has been observed that the fungus had obtained resistance against fungicides (Leroch *et al*, 2011).

Genome sequencing of *Botrytis cinerea* was completed and published in year 2011 revealing a genome of 38.8 Mbp (Amselem, J. *et al*, 2011) which has a overall genome GC contents of around 41.8-43.2% however GC contents in the exonic regions are higher than the GC content in the intronic regions by 6%. Number of genes of *Botrytis cinerea* is comparable but slightly high with 16,360 genes predicted. The summary of the genome details of *Botrytis cinerea* can be found in Table 2.3.

Table 2.3: Details of *Botrytis cinerea* Genome Sequencing, Assembly and Annotation (Amselem, J. *et al*, 2011)

General genome features	**Value**
Coverage	4.5X
Assembly size (Mb)	42.3
Total contig length (Mb)	38.8
Scaffolds	588
Scaffolds N50 (kb)	257
Contigs	4,534
Contig N50 (kb)	16.4
>= Q40 (%)	98.0
GC (%)	43.1
Predicted protein-coding genes	16,448
Dubious genes	2,784
High-confidence genes	13,664
Median coding sequence length (nt)	744,
Median exon length (nt)	190
Median intron length (nt)	74
Median intergenic length (nt)	958
GC Exonic (%)	46.2
GC Intronic (%)	40.9
tRNAs	195
Transposable elements	0.9%

2.1.3 *Ustilago maydis*

Ustilago maydis (also known as smut fungus) is a pathogenic fungus that infects maize and it's seen as a model fungus for study due to the ability of the fungus to grow in controlled environment, for instance in culture of defined media and that the fungus is haploid and grows by budding which then forms compact colonies that allow direct replication of colonies (Dean, R. *et al*, 2012). The pathogenicity of smut fungus is straight forward as it corresponds to its sexual development, thus formation of dikaryotic filament is the most obvious symptom of infection.

The fungus then invades host plant cells via appressorium, eventually forming large tumors resulting from fungus-induced changes in plant growth. Genome analysis of haploid *Ustilago maydis* (Kamper, J. *et al*, 2006) resulted in an assembled genome of 19.8 Mbp with an estimated genome size of 20.5 Mbp with an overall GC content in the region of 54.03%. Number of prediceted genes resulting from genome annotation is estimated to be 6,522, much lesser than 3 other fungus in included in this study. All genome statistics of *Ustilago maydis* is summarized in Table 2.4.

Table 2.4: Details of *Ustilago maydis* Genome Sequencing, Assembly and Annotation (Kamper, J., 2006)

General genome features	Value
Coverage	12.92X
Assembly size (Mb)	19.8
Total contig length (Mb)	19.68
Scaffolds	274
Scaffolds N50 (kb)	127.49
Contigs	274
Contig N50 (kb)	127.49
>= Q40 (%)	98.91
GC (%)	54.03
Predicted protein-coding genes	6,522

2.1.4 *Puccinia graminis*

Puccinia graminis is known to cause rust disease on wheat and together with other its other siblings in the order of *Puccinia* collectively causing 3 rust diseases on wheat, which are the stem rust, stripe rust, and leaf rust. The fungus is an obligate, biotriphic badisiomycete with a heteroecious life cycles (Bolton *et al*, 2008). Mechanism of infection of these biotrophic basidiomycete fungi in host plants involve differentiation of specialized infection structures that couple as a suppressor to suspend host defense response mechanism as well as a media to obtain nutrients through specially differentiated feeding structures that extends into the plant cells known as haustoria (Voegele, R.T. *et al*, 2011).

Throughout the history there had been severe outbreaks of rusts in wheat that includes damages to crops in North America (Hodson, D.P., 2011), Europe and China (Leonard, K.J., *et al*, 2005) and many other damaging crops across different demographic areas. Genomics research was carried out to study and analyze the genomics of the fungus to uncover underlying factors to this highly effective pathogens.

The genome of *P. graminis* was sequenced by Sanger whole-genome shotgun sequencing (Duplessis, S. *et* al, 2011) with an assembled haploid genome size of 88.6 Mbp, with GC content estimated to be around 43.3%. The number of predicted protein coding genes is 17,773 with an average sequence length of 1,075 bp. All genome statistics of *P. graminis* is summarized in Table 5.

Table 2.5: Details of *Puccinia graminis* Genome Sequencing, Assembly and Annotation (Duplessis, S. *et al*, 2011)

General genome features	Value
Sequence coverage	12
Scaffold total (Mbp)	88.6
Scaffolds	392
Scaffold N50 length (Mbp)	0.97
Scaffold N50	30
Assembly in scaffolds > 50kb (%)	97.1
Contig sequence total (Mbp)	81.5
Contigs	4,557
Contig N50 length (kbp)	39.5
Contig N50	546
Base quality >= Q40 (%)	96.3
Gap content (%)	8
GC content (%)	43.3
Protein coding genes	17,773
Mean coding sequence length (bp)	1,075
Mean exon number per gene	4,7
Mean exon length (bp)	175
Mean intron length (bp)	133
Mean intron length (bp)	3,328
Mean intergenic length (bp)	3,328
tRNAs	428

2.1.5 Fungal Pathogenicity-related Genes

The study of fungal pathogenicity-related genes is the key to understand root cause of fungal inflicted plant diseases. These genes may play an important role in fungal life cycle development particularly when the mode of infection requires fungal vegetative growth for instance *Ustilago maydis* (Kamper, J., 2006), in wood decaying process during infection of host plants, the signaling pathway and etc.

2.1.5.1 Cell Wall Degrading Enzyme

Unique feature of the existence of rigid cell wall protects the plant cell from external invasion thus fungal pathogens secretes cell wall degrading enzyme to surpass the plant cell wall therefore penetrating the plant cell for nutrients (Choi, J. *et al*, 2013). Cutinase is an example of cell wall degrading enzyme where studies showed that cutinase is involved in cuticle penetration of apple leaves (Koller, W. *et al*, 1991). *Magnaporthe oryzae* is known to produce cutinase to facilitate penetration in rice and barley via hydrophobic surface sensing, differentiation and virulence (Skamnioti, P. *et al*, 2007).

Carbohydrate-active enzymes is another group of enzymes related to fungal pathogenicity and similar to cell wall degrading enzymes the group of enzymes participates in plant cell walls degrading activities (Suzuki, H., 2012) by digesting cell plant cell wall materials such as cellulose, hemicellulose, and pectin. An effort to compile sequences of publicly available carbohydrate-active enzymes resulting in the formation of CAZy (Lombard, V. *et al*, 2013), a database that stores curated sequence information of more than 340,000 CAZymes. CAZy classify carbohydrate-active enzymes into five major groups:

Table 2.6: CAZymes Grouping according to CAZy (Lombard, V. *et al*, 2013)

Grouping	Description
Glycoside Hydrolases (GHs)	Hydrolysis and/or rearrangement of glycosidic bonds
GlycosylTransferases (GTs)	Formation of glycosidic bonds
Polysaccharide Lyases (PLs)	Non-hydrolytic cleavage of glycosidic bonds
Carbohydrate Esterases (CEs)	Hydrolysis of carbohydrate esters
Auxiliary Activities (AAs)	Redox enzymes that act in conjunction with CAZymes
Carbohydrate-Binding Modules (CBMs)	Adhesion to carbohydrates

The availability of genes and proteins sequences from well-annotated fungal pathogen's genome enabled identification of candidate cell wall degrading enzymes or carbohydrate-active enzymes in wide range of pathogenic fungus, which provide important clues for fungal pathogenicity.

2.1.5.2 Signaling proteins

Signaling proteins is vital in host-pathogen interaction in the early stages of infection (Tudzynski, P. *et al*, 2003) in reception of extracellular signals from the host to pathogens to activate effector proteins for initiation of infection into the host. Example of such gene is the heterotrimeric G proteins where the G proteins activate other effector proteins such as kinases, adenylate cyclases, phospholipases and ion channels (Kronstadt, J.W., 1997) and this includes the MAPK gene. Receptor proteins recognize surface protein of the host and initiates infection mechanisms towards the host. GTP-biding proteins is another candidate gene responsible for fungal pathogens' pathogenicity where research had shown that absence of these proteins results in reduced growth rate and morphological changes. Furthermore GTP-binding protein is connected to MAP kinases cascades for cAMP pathway that triggers the development of appressorium formation (Tudzynski, B. *et al*, 2001)

2.2 Fungal Bioinformatics Research and Analysis

The emergence of sequencing technologies had increased the resolution of research into molecular causative factors in molecular plant pathology. Through genome sequencing of plant pathogens like *Magnaporthe oryzae* (Dean, R.A. *et al*, 2005), *Botrytis cinerea* (Amselem, J. *et al*, 2011), *Ustilago maydis* (Kamper, J. *et al*,

2006), and *Puccinia graminis* (Duplessis, S. *et al*, 2011) coupling with improving bioinformatics methodology genome assembly, genome annotation, comparative genomics enabling pathologist to identify genomics features in fungal pathogens that plays important role in fungal pathogenicity.

Whole genome sequencing of plant fungal pathogens allows high quality genome assembly to identify reveal-underlying sequences of the fungus. Genome annotation of the assembled genome then predicts gene models based on *ab initio* prediction as well as homology searches (Yandell, M. *et al*, 2012) to known nucleotide or protein sequences. Availability of an annotated genome allows downstream bioinformatics analysis such as polymorphic markers identification through genome mapping (Davey, J.W. *et al*, 2011) and comparative genomics (Wei, L. *et al*, 2002).

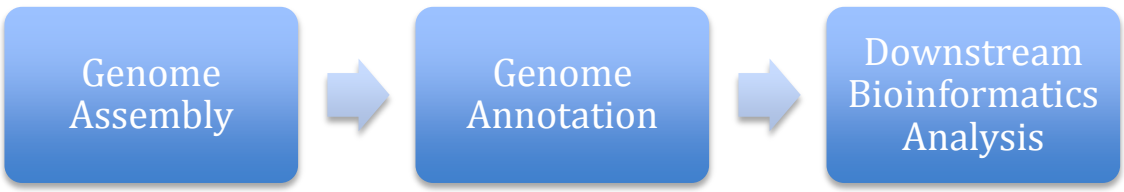

Figure 2.1: Typical Bioinformatics Workflow. Whole genome sequencing enabling genome assembly, thus enabling annotation of the assembled genome, which serve as a foundation for a series of downstream bioinformatics analysis.

2.2.1 Comparative Genomics

Comparative genomics is a technique used to compare genome sequences between two or more organisms to identify similarities and differences between organisms of study, comparing genome features of the organisms such as gene content, similarities and differences in genes sequences, number of genes presence, types of genes presence and etc (Wei, L. *et al*, 2002). Comparative genomics revolves around the comparison of genome sequences between organisms thus availability of sequences is a must before comparative genomics can be done.

Recent development of Next Generation Sequencing Platform for instance Illumina HiSeq (Illumina Inc., 2014) allows high throughout sequencing of whole genome sequences as well as targeted genomics region of interest which then becomes an enabling technology for discovery of high confidence polymorphic markers including single nucleotide polymorphism (Vignal, A. *et al*, 2002), SSR (Toth, G. *et al*, 2000), insertion, deletion, copy number variation, and other structural variation (Shigemizu, D. *et al*, 2013).

Sequence alignment of two sequence of interest is the simplest method in comparative genomics. By looking into the similarities and differences in sequence composition phylogenetic relationship between organism can be derived and thus determining level of either genomic similarity or genes sequence similarity.

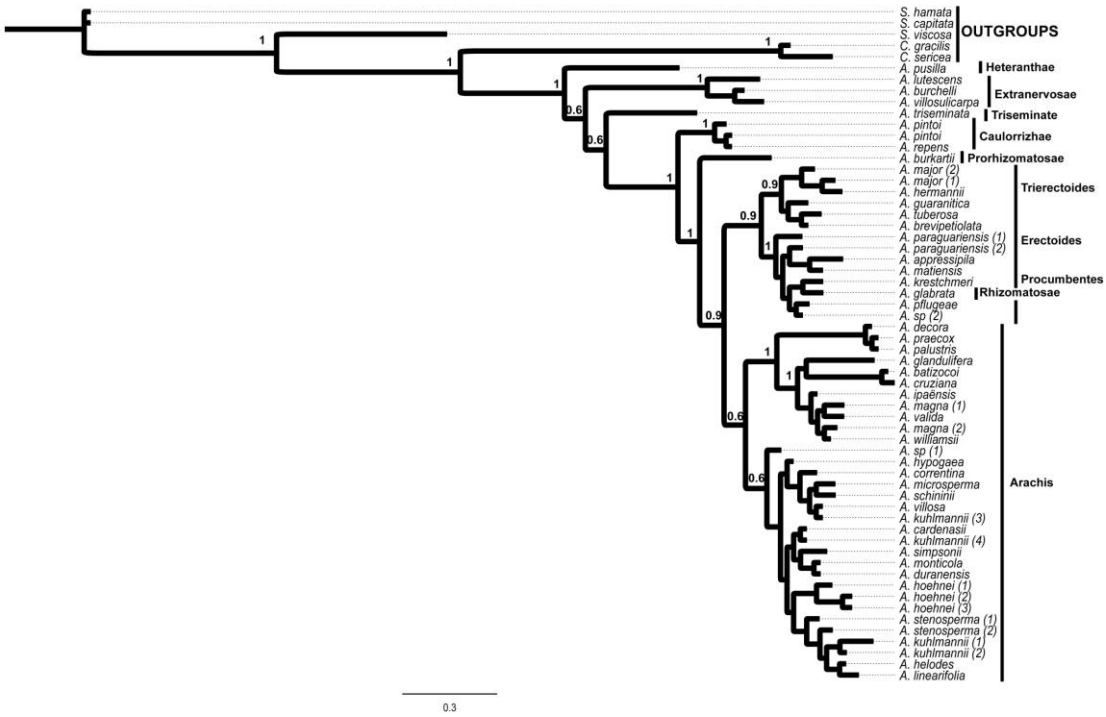

Figure 2.2: Example of phylogenetic analysis in genus of *Arachis* using ITS and 5.8S rDNA sequences (Bechara, M.D. *et al*, 2010)

2.2.1.1 Polymorphic marker identification

Sequencing technologies serves as an enabling platform for various downstream research and development, particularly setting the foundation for bioinformatics research and development. Discovery of different polymorphic markers such as Single Nucleotide Polymorphism, Insertions and Deletions, Copy Number Variations as well as presence of genes is important as each of these polymorphisms plays important roles in causing pathogenicity in fungus which could confers pathogenicity to pathogenic isolates as it is shown in human research.

2.2.2 Fungal Genome initiative

Spearheaded by the Fungal Genomics group at Broad Institute (Broad Institute, 2014) aims to sequence and analyze broad range fungus that plays vital role in medicine, agriculture as well as industrial application and this effort is supported by the National Human Genome Research Institute, the National Science Foundation, the National Institute of Allergy and Infectious Disease, and the US Department of Agriculture.

The initiative's emphasis in genome sequencing had resulted in sequencing of more than 100 fungal genomes (Broad Institute, 2014), and large fraction of these fungal whole genome sequencing had enabled high quality genome assembly and genome annotation. The Broad Institute had allowed download of these sequence data for scientific community for further study and research, resulting in an collective effort in the study of fungal genomics.

CHAPTER 3

MATERIAL AND METHODOLOGY

3.1 Data Source and Databases

Comparative genomics carried out between four different pathogenic fungal species from two different phyla with two fungus species from Basidiomycota (*Ustilago maydis, Puccinia graminis*) and the other two fungus species from Ascomycota (*Magnaporthe oryzae, Botrytis cinerea*). Genome sequencing reads for all four fungal species are obtained from the Sequencing Reads Archive hosted at the National Center of Biotechnology Information. Assembled genome sequence and annotation files that include FASTA sequences for genes and proteins are obtained from Fungal Genome Initiative by Broad Institute, United States of America. Details of each genomic data obtained are summarized in Table 3.1.

Table 3.1: Details of Genomics Data Source for Subjects of Study

Species	Description	Genome assembly	Genome annotation
Magnaporthe oryzae	FASTQ	Broad Institute	Broad Institute
Botrytis cinerea	FASTQ	Broad Institute	Broad Institute
Ustilago maydis	FASTQ	Broad Institute	Broad Institute
Puccinia graminis	FASTQ	Broad Institute	Broad Institute

Each set of the data will then be used for comparative genomics between all four fungal species in the workflow to be described in 3.2. Two databases were selected to be incorporated into the Comparative Genomics Workflow that includes CAZy (Lombard, V. *et al*, 2013) and PHI-base (Winnenburg, R. *et al*, 2006) where both databases contains proteins sequences of carbohydrate-active enzymes and pathogens host interaction-related proteins. CAZy contains carbohydrate-active enzyme sequences based on conserved domain search from known sequences, and PHI-base contains pathogen host interaction-related sequences from various organisms. These protein sequences were downloaded and used to build local CAZy and PHI-base databases.

3.2 Workflow of Comparative Genomics

A novel Comparative Genomics Workflow for Inter-Pyla Plant Pathogenic Fungal Comparative Genomics was constructed by incorporating various Bioinformatics tools and Application listed below:

- BLAST (Basic Local Alignment Search Tool) (Altschul, S.F. *et al*, 1990)
- MUMmer 3.0 (Kurtz, S. *et al*, 2004)
- HMMER (Durbin, R. *et al*, 1998)
- dbCAN (Yin, Y. *et al*, 2012)
- VENNY (Oliveros, J.C., 2007)

Shell scripting was required to establish the comparative genomics workflow. Global comparative and Local Comparative were combined to produce comparative genomics results from this workflow.

Global comparative workflow involves genome-scale comparison with MUMmer by aligning genome sequences of plant pathogenic fungus within the same phylum, followed by genome mapping of whole genome sequencing reads downloaded to genome sequence of another fungus within the same phylum using BWA (Burrows Wheeler Aligner) and variant calling with SAMtools using default settings. Local comparative workflow involves homologous protein coding genes analysis, focusing on three types of searches listed below:

- General homology search
- PHI-base
- CAZy Database

Thus combining the Global comparative workflow and the Local comparative workflow a detailed workflow for Inter-Phyla Plant Pathogenic Fungus Comparative Genomics is established. Comparative Genomics Workflow for Inter-Phyla Plant Pathogenic Fungal Comparative Genomics is separated into two major phases: the first phase of the comparative workflow involves intra-phyla comparison whereas the second phase of the comparative workflow involves inter-phyla comparison.

3.2.1 First Phase: Intra-Phyla Comparison

The first phase of the workflow involves intra-phyla comparison of plant pathogenic fungus where genomic sequences from plant pathogenic fungus of the same phylum were aligned against each other using MUMmer 3.0 with default parameters

and outputting results in postscript format to provide a global comparative view of sequence similarities between fungus of the same phylum.

Protein sequences from respective fungus in the same phylum were aligned against each other with BLASTP and BLAST hits were filtered with E-value cut-off at 1e-5 and HSP percentage at 80%. Protein sequences from each fungus of the same phylum were aligned to the PHI-base respectively with BLASTP and pathogen host interaction proteins were identified with an E-value cut-off at 1e-5 and HSP percentage at 80%. HMMER-based carbohydrate-active enzymes annotation tool dbCAN was used to carry out domain search to all protein entries in the CAZy with default settings.

Resulting set of homologous protein coding genes, candidates of protein coding genes of carbohydrate-active enzymes, and candidates of pathogen host interaction-related protein coding genes were then subject to customized shell scripting to produce consensus genes set for all three category for Second Phase Comparison. First phase of the plant pathogenic fungus workflow is represented in Figure 3.1

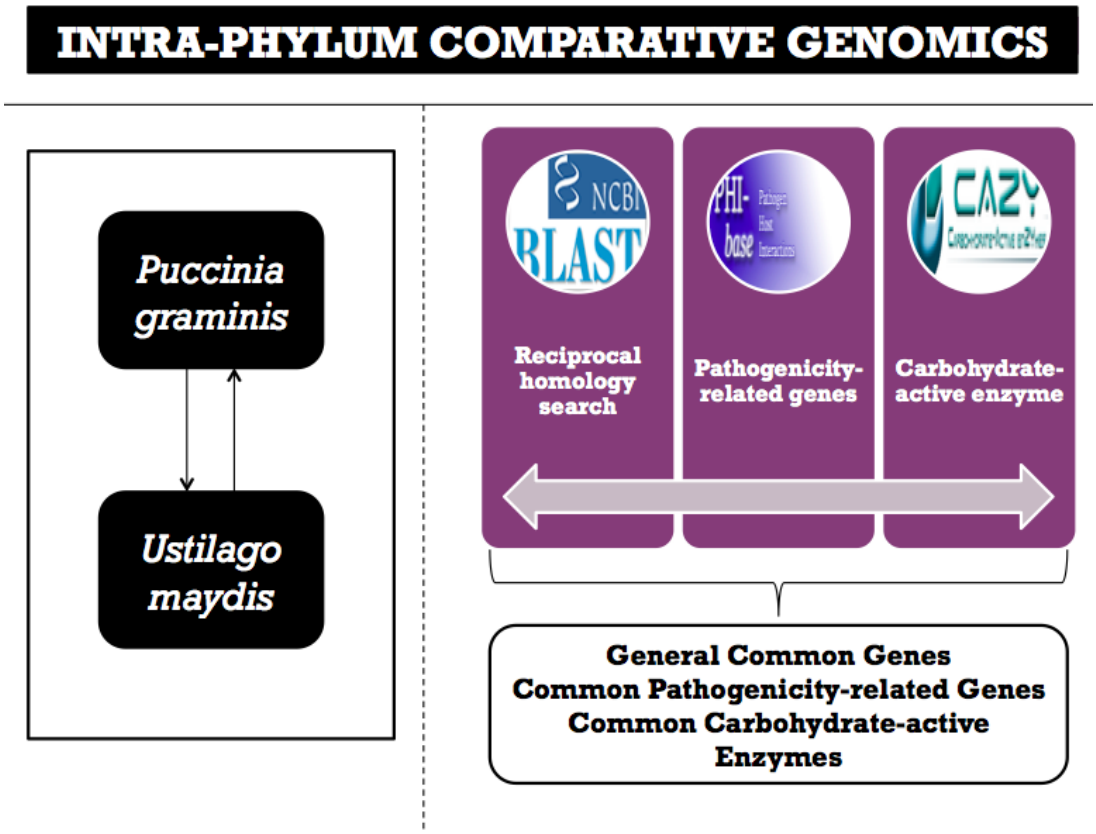

Figure 3.1: Illustrated Workflow for First Phase of Plant Pathogenic Fungus Comparative Genomics

3.2.2 Second Phase: Inter-Phyla Comparison

The second phase of the inter-phyla plant pathogenic fungus comparative genomics workflow involves analysis of consensus genes sets resulting from first phase of the workflow. Each homologous or candidate protein-coding genes are consolidated into a consensus list of genes which are then based on respective unique ID that ties to a gene, a common ground can be established between species. For instance by aligning each of the annotated protein-coding genes from each plant pathogenic fungus to the PHI-base and CAZy the corresponding aligned sequence will be assigned a corresponding ID resulting from alignment hits. This corresponding ID will be used as the basis of comparison to identify common and unique pathogenicity-related genes between plant pathogenic fungus from the same phylum and common and unique

pathogenicity-related genes between funguses of different phyla. The summarized inter-phyla comparative genomics can be seen in Figure 3.2.

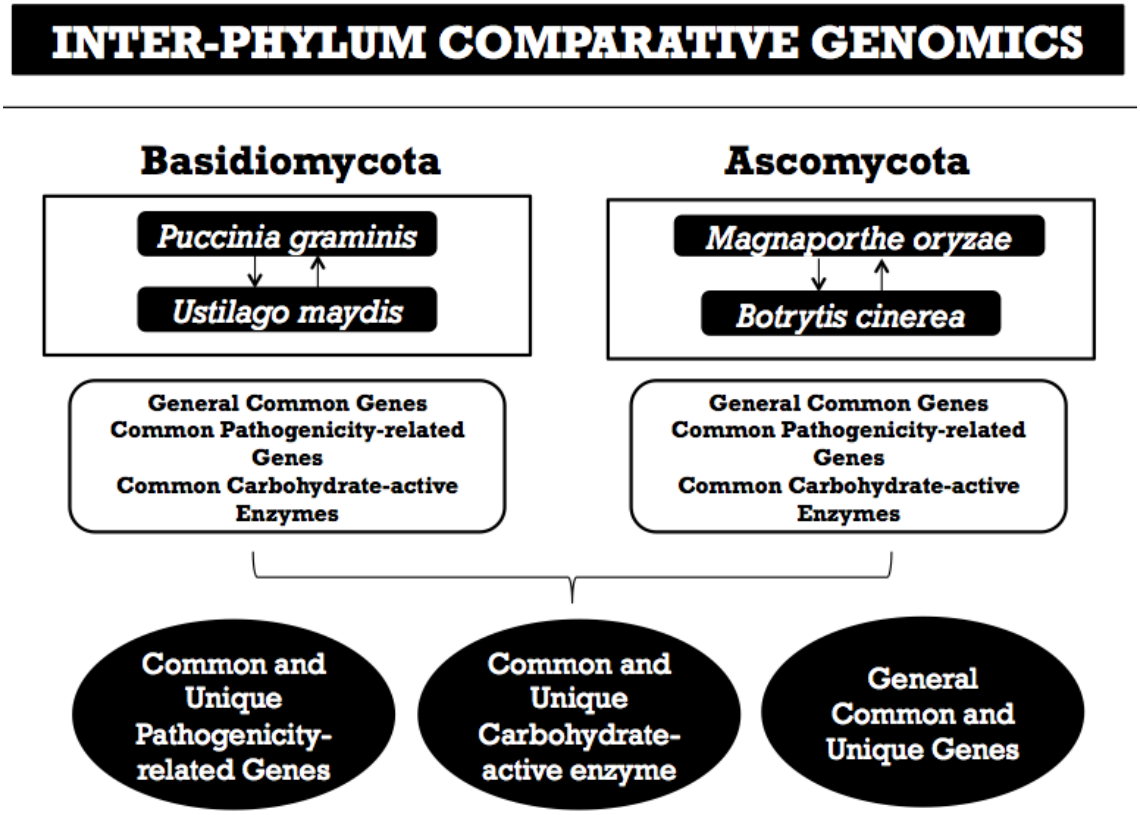

Figure 3.2: Illustrated Workflow for Inter-Phyla Comparative Genomics

3.3 Genes Copy Number Variation analysis

First step is to check correlation of genes copy number intra-phylum before going to second step to identify the copy number ratio between one phyla to the other. Assuming *a* and *b* are two different species from the same phylum, the optimum correlation between the two species within the same phylum is taken by the ratio of *a* to *b*, with 1 being maximum correlation, and in this analysis standard deviation of 0.25 is labeled as highly correlated.

For genes to be considered as potential genes company number variation, same genes have to be highly correlated between individuals within the same phylum. Inter-phyla copy number variation then can be determined by taking ratio of mean genes copy number of one phylum to the other. Genes showing ratio of more or equal to 2 times the copy number of another phylum will be shortlisted as potential genes copy number variation.

CHAPTER 4

RESULTS

4.1 Whole Genome Alignment Analysis

Whole genome alignment analysis of plant pathogenic fungus was completed between plant pathogenic fungus from within the same phylum. Thus alignment results were obtained from MUMmer alignment of whole genome sequence of *Botrytis cinerea* and *Magnaporthe oryzae* and alignment results were visualized with mummerplot. Scattered plot of alignment were plotted in Figure 4.1.

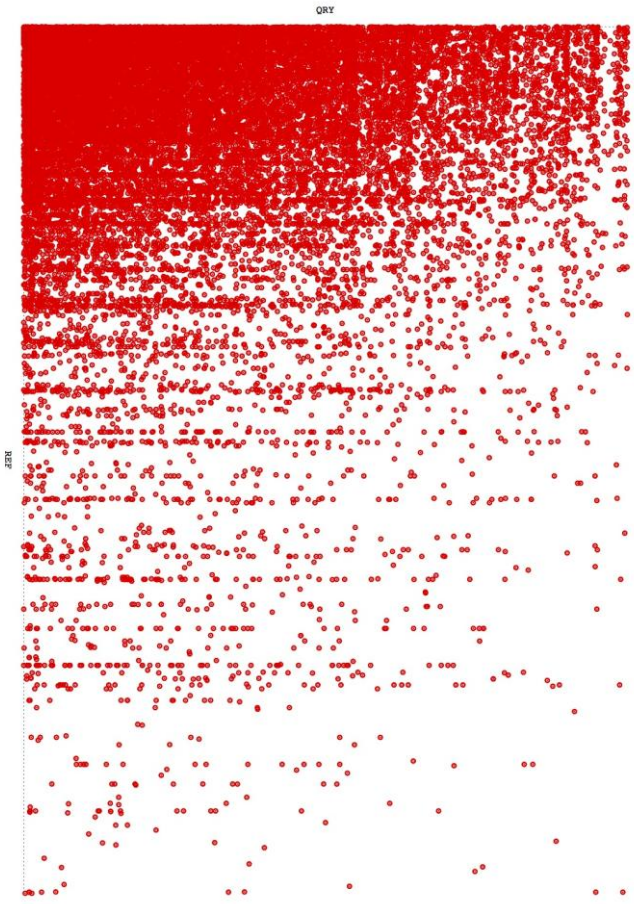

Figure 4.1: MUMMERPLOT of Whole Genome Alignment Results of *B. cinerea* and *M. oryzae*

Alignment of the genome sequence of the two fungus isolates did not result in a good alignment. The same steps were used to align whole genome sequence between basidiomycetes *Ustilago maydis* and *Puccinia graminis.* The alignment results of basidiomycetes is visualized in Figure 4.2.

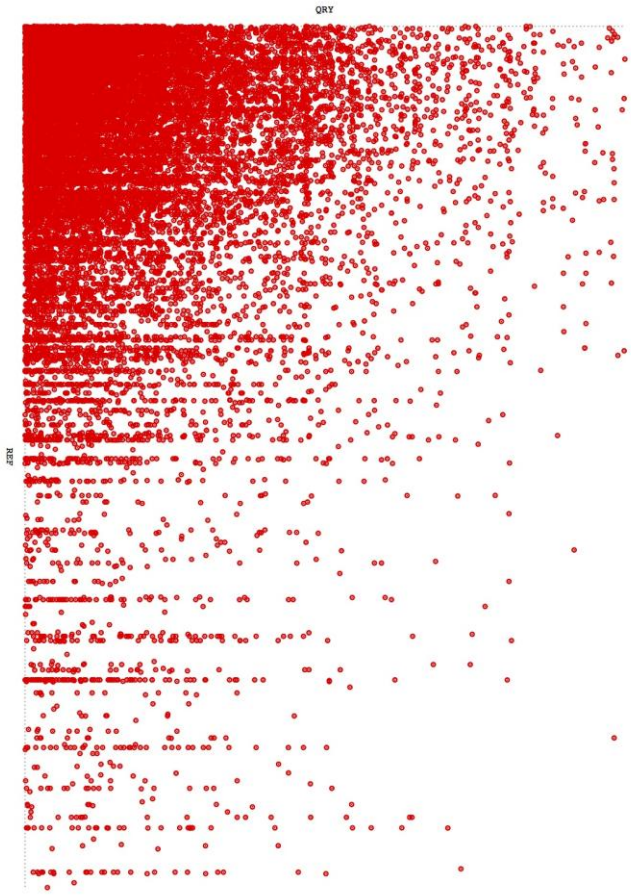

Figure 4.2: MUMMERPLOT of Whole Genome Alignment Results of *U. maydis* and *P. graminis*

Similar results were observed from alignment between the basidiomycetes as it was observed in ascomycetes, no notable alignments in intra-phyla analysis for both ascomycetes and basidiomycetes.

4.2 Reciprocal Homology Search

First phase of workflow reveals from the total number of homologous protein coding genes of both ascomycetes and basidiomycetes. From the total number of protein coding genes for plant pathogenic fungus from the phylum of Ascomycota *B. cinerea* (16,448) and *M. oryzae* (12,991) BLASTP alignment was executed and filtered based on the cut-off value for e-value of 1e-5 and HSP percentage of 80% a total of 5,508 homologous protein coding genes were identified in the phylum of Ascomycota. On the other hand, from the total number of protein-coding genes for plant pathogenic fungus from the phylum of Basidiomycota *U. maydis* (6,522) and *P. graminis* (15,979) BLASTP alignment was executed and filtered also based on cut-off value for e-value of 1e-5 and HSP percentage of 80%. In the comparison for the phylum of basidiomycetes a total of 2,433 homologous protein-coding genes were identified.

Second phase of the workflow involves comparison of homologous protein-coding genes identified for respective phylum into a single consensus list of homologous protein-coding genes between Basidiomycota and Ascomycota. By comparing 5,508 and 2,433 homologous protein-coding genes based on the cut-off value for e-value of 1e-5 and HSP percentage of 80%, a total 1,388 inter-phyla homologous protein-coding gene were identified of which 798 of these homologous genes were previously annotated as hypothetical protein and the remaining are genes essential for survival of fungus such as kinases, ribosomal proteins and etc.

4.3 Pathogenicity-related Genes Analysis

Search of pathogen-host interaction related pathogenicity genes is also divided to two phases for intra-phylum and inter-phyla comparison. BLASTP search of protein-coding genes protein sequences of all four plant pathogenic fungus species with a cut-off for e-value of 1e-5 and HSP percentage of 80% the candidates of pathogen-host interaction-related genes were identified for *B. cinerea* with 1,339 candidate genes, *M. oryzae* with 1,402 candidate genes, *P. graminis* with 626 candidate genes, and *U. maydis* with 533 candidate genes. Please refer to Appendix for full list of all homologus Pathogen-Host Interaction-Related genes (Table 4.8)

First phase intra-phylum comparison using customized shell scripting from list of identified candidate pathogen-host interaction-related genes for both Basidiomycota and Ascomycota resulting in 203 homologous candidate pathogen-host interaction-related genes in in the phylum of Basidiomycota and 534 homologous candidate pathogen-host interaction-related genes in the phylum of Ascomycota. Second phase inter-phyla comparison of homologous candidate pathogen-host interaction-related genes from the phylum of Basidiomycota and the phylum of Ascomycota using customized shell scripting resulted in 159 homologous candidate pathogen-host interaction-related genes between fungus from both Basidiomycota and Ascomycota.

Number of unique and common candidate pathogen-host interaction-related genes was identified between all four fungus species as shown in Figure 4.3 and the full list of identified candidate pathogen-host interaction-related genes can be found in Appendix (Table 4.7)

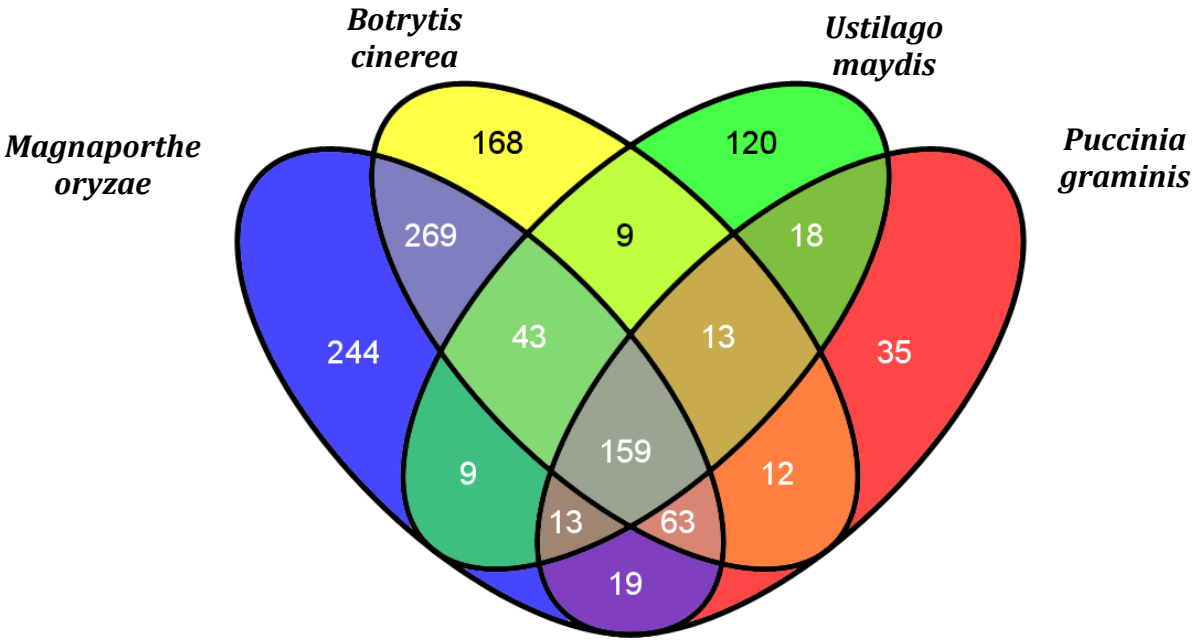

Figure 4.3: Venn diagram of Homologous Candidate Pathogen-Host Interaction-Related Genes between Four Plant Pathogenic Fungus

Genes copy number variation were observed across the two phylum. Genes copy number is determined by the number of identical PHI-base annotated for candidate pathogen-host interaction-related genes. From 159 homologous candidate pathogen-host interaction-related genes across all four fungus species, genes copy number variation were observed in 5 out of the 159 candidate genes identified based on methodology described in Chapter 3.3. The most promising genes that show high confidence copy number variation is PHI:2968, Hxs1 gene which participate in transmembrane transport activity. Clustering of this gene from all four fungus species also shows phyla specific clustering shown in Appendix (Figure 4.4).

Table 4.1: Statistics of Candidate Genes Showing Genes Copy Number Variation

	B. cinerea	*M. oryzae*	*P. graminis*	*U. maydis*
PHI:2968	33	36	9	10
PHI:2096	2	2	1	1
PHI:2171	3	4	1	1
PHI:2530	2	2	1	1
PHI:447	2	2	1	1

4.4 Carbohydrate Active Enzyme Analysis

Identification of candidate protein coding genes for carbohydrate active enzymes is divided into two phases for intra-phylum and inter-phyla comparison. HMM-based domain search to Carbohydrate-Active Enzyme Database (CAZy Database) using dbCAN with filtering of high confidence candidate domains based on default settings to all four plant pathogenic fungus species resulting in identification of candidates of protein coding genes for carbohydrate-active enzymes for *B. cinerea* with 134 candidate genes, *M. oryzae* with 137 candidate genes, *P. graminis* with 98 candidate genes, and *U. maydis* with 97 candidate genes. Please refer to Appendix for full list of all homologous Carbohydrate-Active Enzymes for all fungus (Table 4.9).

Classification of carbohydrate-active enzymes consists of seven distinct classes based on conserved domains. Identification of different classes of candidate protein coding genes for carbohydrate-active enzyme for all four plant pathogenic fungus species resulting in distribution of different classes of carbohydrate-active enzymes based on domain search in Table 4.2:

Table 4.2: Classification of Candidate Protein Coding Genes for Carbohydrate Active Enzymes

CAZy Families	***B. cinerea***	***M. oryzae***	***P. graminis***	***U. maydis***
AA	85	105	24	30
CBM	69	120	16	10
CE	119	136	74	64
GH	250	272	161	117
GT	104	103	102	69
PL	11	5	7	3

First phase of the comparative genomics workflow involves intra-phylum comparison for carbohydrate-active enzymes analysis for all four plant pathogenic fungus species with customized shell scripting for both Basidiomycota and Ascomycota resulting in 70 common candidate protein coding genes for carbohydrate active enzymes in Basidiomycota and 116 common candidate protein coding genes for carbohydrate active enzymes in Ascomycota and breakdown of classification for common candidate protein coding genes for carbohydrate active enzymes to different CAZy families is summarized in Table 4.3.

Table 4.3: Classification of Candidate Protein Coding Genes for Carbohydrate Active Enzymes for Common Intra-Phylum Candidate Genes

CAZy Families	**Basidiomycota**	**Ascomycota**
AA	6	9
CBM	4	13
CE	6	11
GH	29	52
GT	24	29
PL	1	2

Second phase of the comparative workflow involves inter-phyla comparison between candidate protein coding genes for carbohydrate-active enzymes of the fungus from Basidiomycota and Ascomycota respectively. Customized shell scripting was written to process results from first phase of the comparative workflow, thus from the 70 common candidate protein coding genes for carbohydrate-active enzyme from Basidiomycota and 116 from Ascomycota 64 candidate protein coding genes were found to be common between Basidiomycota and Ascomycota. Of these 64 common candidates protein coding genes identified between Basidiomycota and Ascomycota 6 are classified in the auxiliary activities family, 3 classified in the carbohydrate binding module family, 6 classified in the carbohydrate esterases, 25 classified in the glycoside hydrolases family, 23 classified in the glycosyl transferases family, and 1 classified in

the polysaccharide lyases family. The classification of the candidates protein coding genes for carbohydrate-active enzymes is summarized in Table 4.4, common and unique intra-phylum and inter-phyla candidate protein coding genes for carbohydrate-active enzymes are visualized in Figure 4.5. List of all common inter-phylum candidate genes can be found in Appendix (Table 4.7)

Table 4.4: Classification of Candidate Protein Coding Genes for Carbohydrate Active Enzymes for Common Inter-Phylum Candidate Genes

CAZy Families	Inter-Phyla
AA	6
CBM	3
CE	6
GH	25
GT	23
PL	1

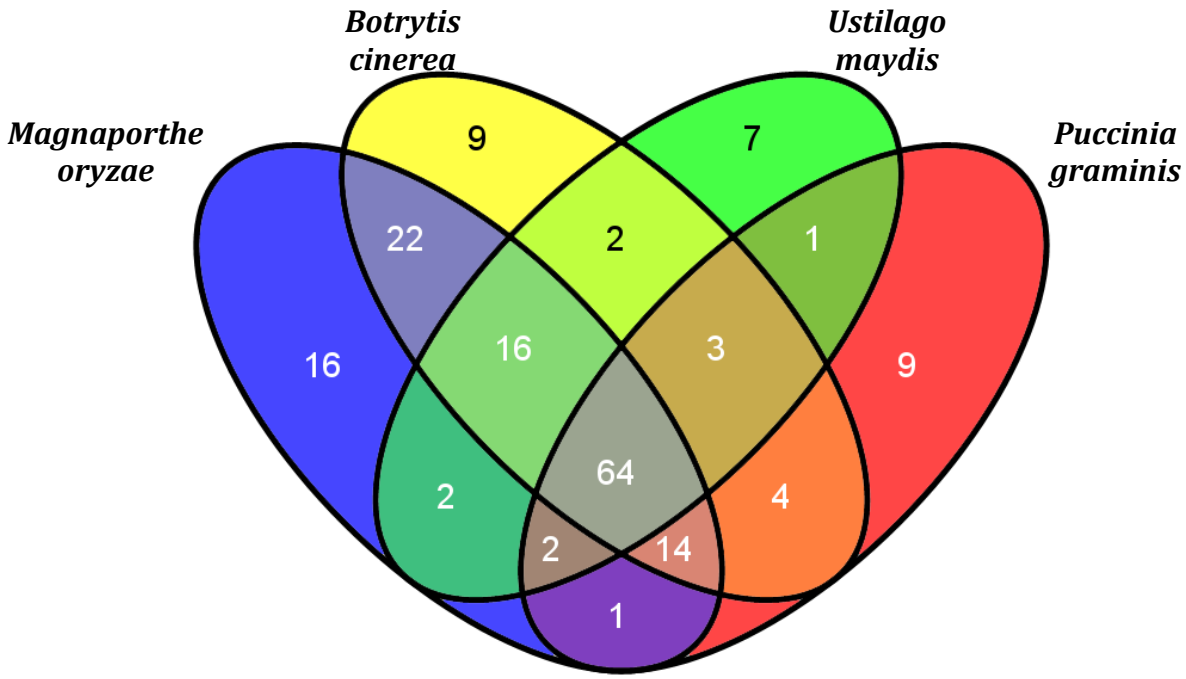

Figure 4.5: Venn diagram of Candidate Protein Coding Genes for Carbohydrate-Active Enzymes between Four Plant Pathogenic Fungus

Genes copy number variation analysis was done based on methodologies and criteria described in Chapter 3.3. Of the 64 common candidate protein coding genes for carbohydrate-active enzymes, only 3 candidate genes fulfills the criteria showing genes copy number variation as described in Table 4.5.

Table 4.5: Statistics of Candidate Genes Showing Genes Copy Number Variation

Family ID	*B. cinerea*	*M. oryzae*	*P. graminis*	*U. maydis*
AA7	26	33	6	6
GH72	6	5	1	1
GH79	2	2	1	1

AA7 was taken as an example for further analysis to study inter-phyla sequence similarities. Auxiliary Activities Family 7 belongs to a group carbohydrate-active enzyme which includes Glucooligosaccharide oxidase and Chitooligosaccharide oxidase and clustering of the protein coding genes annotated with the AA7 domain showed intra-phylum clustering closer than inter-phyla clustering in Appendix (Figure 4.6)

CHAPTER 5

DISUCSSION

5.1 Global Comparative Workflow and Local Comparative Workflow

The Plant Pathogenic Fungus Comparative Genomics Workflow comprises of two major parts of comparison, which is the Global Comparative Workflow and the Local Comparative Workflow where in Global Comparative Workflow genome-wide comparison was made between plant pathogenic funguses belonging to the same phylum. Local Comparative Workflow compares protein-coding genes of individual species to each other as well as to pathogenicity genes-related databases such as PHI-base and CAZy. Genome-scale alignment with MUMmer between funguses belonging to the same phylum did not reveal genomics sequence similarities as the resulting dot plot do not show clear aligned region as showed in an example of good quality alignment between two sequence in Appendix (Figure 5.1). This shows genomic variation between fungus species is relatively large although they are classified within the same phylum, which is observable in this study in both Basidiomycota and Ascomycota.

This phenomenon could be due to current taxonomical classification of the Kingdom of fungus was based on phenotype rather than genotype (Guarro, J. *et al*, 1999), thus explaining vast genome variation between fungus of same and different phylum. This however does not affect the results of Local Comparative Workflow as genes are much more conserved than whole genome sequences. Protein sequences were used for homology searches among 4 fungus species because it is more conserved than nucleotide sequences as nucleotide sequence consists intronic regions, which are more

variable than coding sequences. Local Comparative Workflow showed promising results in identifying homologous genes between species in the same phylum as well as homologous genes between species of different phylum. Besides, although genome sizes and number of annotated genes for all four fungus of this study varies the number of candidate protein-coding genes of pathogenicity-related genes are relatively uniform between funguses within the same phylum as listed in the Appendix (Table 5.1 & Table 5.2). This phenomenon agrees with a theory that eukaryotes have core proteins that must exists to ensure the survival of the species such as the list of core proteins listed in the Eukaryotic Orthologous Group Database (KOG) (Tatusov, *et al*, 2003) and the number of candidate protein-coding genes for Carbohydrate-Active Enzyme and Pathogen-Host Interaction-Related genes identified were conserved and specific within the based on the results obtained from this study.

Global Comparative may not be the best methodology for inter-phyla comparison as the whole genome variation does not represents the relationship between funguses as genome sequences of various species has a high degree of variation as seen from MUMmer genome-scale alignment results. On the other hand local comparative genomics may be a more accurate alternative methodology to measure phylogenetic clustering and relationship between funguses as these sequence are relatively more conserved compared to whole genome sequences. Results from this project supports such deduction and phylogenetics relationship based on selected common pathogenicity-related protein coding genes sequences showed clustering of funguses in concordance to their taxonomical relationship.

The availability of bioinformatics tools for genome annotation allows annotation of protein-coding sequences to their identity and functionality. Though the availability

of such tool is at the convenience of researchers due to the ease and availability of bioinformatics tools online, functionalities and identify of many protein-coding genes are still yet to be identified and is annotated as hypothetical proteins and thus domain search could be a key to annotate these hypothetical proteins.

5.2 Intra-Phylum and Inter-Phylum Comparison

Intra-Phylum comparison is a common application to study importance development in fungal pathogenicity (Manning, V.A. *et al*, 2013). Due to high degree of variation among fungus species minimal effort had been carried out for inter-phyla comparative genomics. Plant pathogenic fungus originates from two major phylum of Basidiomycota and Ascomycota and top ten fungal pathogens based on study by Dean, R. *et al* reveals that all top 10 funguses listed in the study were originated from either Basidiomycota or Ascomycota. Although taxonomically these funguses are classified in different phyla but their shares certain level of similarities in terms of host plant that these funguses infects. For example *Magnaporthe oryzae* and *Puccinia graminis*, belonging to the Ascomycota and Basidiomycota respectively and both of these funguses causes plant fungal plant diseases in wheat with *Magnaporthe oryzae* causing head blast disease in wheat (Figure 5.2 in Appendix) whereas *Puccinia graminis* causes stem rust disease in wheat (Figure 5.3 in Appendix) (Dean, R. *et al*, 2012). With funguses from different phyla infesting the same host plant it is possible for a development of a broad-spectrum antifungal agent to counter fungal pathogens on various plants and crops. As expected the number of homologous genes are greater during intra-phylum comparison compared to the number of homologous genes identified from inter-phyla comparison which agrees to the taxonomical relationship of these funguses.

Interestingly inter-phyla comparison resulted in a set of common pathogenicity-related genes that is found to be common between plant pathogenic fungus from Basidiomycota as well as plant pathogenic fungus from Ascomycota, suggesting that although phenotypically and morphologically these funguses are different the mechanisms behind their pathogenicity may draw high level of similarities with the presence of common sets of pathogenicity-related genes while phylogenetic analysis of an example of such a gene like PHI:2389 and AA7 (Figure 4.4 and Figure 4.5 in Appendix) show that these genes still maintain the phylum specific clustering.

5.3 Pathogenicity-related Genes Content

One of the most important objectives of this study is to identify common or unique pathogenicity-related genes among these four plant pathogenic fungus and also to identify intra-phylum and inter-phyla similarities and differences in gene numbers. Results revealed that number of pathogenicity-related genes for fungus in the phylum of Ascomycota is greater than the number pathogenicity-related genes identified from fungus in the phylum of Basidiomycota. The reason behind the differences in identified pathogenicity-related gene number is unclear, however duplication and expansion of gene families had showed to play a role in pathogenicity (Pendleton, A.L. *et al*, 2014) thus genes copy number variation found in different funguses may play a role in altering level of pathogenicity of plant pathogenic fungus.

Also number of protein-coding genes for Carbohydrate-Active Enzyme and Pathogen-Host Interaction-Related genes is much similar and closer within a phylum as it was seen from the results. The results may explain the fungal activities and the

requirement and need of the number of genes for the funguses to survive and grow on host, or may be due to morphological differences as certain pathogenic fungus requires switch of morphology to induce virulence factors (Magee, P.T., 2010). These pathogenicity-related genes identified from the study are genes that plays a direct or indirect role in pathogenicity and are defined as genes necessary for disease development but not compulsory for fungal pathogen life cycle development (Idnurm, A. *et al*, 2001) and genes found common in this studies are important genes in fungal pathogenicity.

5.3.1 Copy Number Variation Analysis Result

PHI:2968 was identified as one of the pathogenicity related genes that has large copy number variation between Ascomycota and Basidiomycota and the genes was identified as Hx1 gene which is a protein-coding gene that codes for High Affiinity Glucose Transporter which is needed for fungus to resist to oxidative stress as well as required in fungal virulence activities (Liu, T.B., *et al*, 2013) thus suppressing the genes could results in reduced virulence activities in plant pathogenic fungus across both phylum of Basidiomycota and Ascomycota.

One of the common Carbohydrate-Active Enzyme protein families across both phylum were found to be AA7 (Auxiliary Activities Family 7) and notable member in this family of Carbohydrate-Active Enzyme is glucooligosaccharide oxidase. Glucooligosaccharide oxidase involves in oxidation of glucooligosaccharide, an important component of the plant cell wall (Zemkova, Z., *et al*, 2012) thus the Glucooligosaccharide plays a vital role in degrading the plant cell wall for the fungal to penetrate into the plant cell. This gene is found to be common across the two phylum

thus providing another good target for development of a broad-spectrum antifungal agent.

5.4 Deduction of Fungal Pathogenicity

Identification of pathogenicity-related genes suggests that these genes may play an important role in pathogenicity development of these funguses in host plants. However knowing the list of pathogenicity genes is not adequate to deduce degree of virulence of plant pathogenic fungus as expression of these genes are more important than the existence of these genes. Expressed Sequence Tags and RNA Sequencineg are latest technology, which helps in identifying expression level of pathogenicity genes as reported in various studies of fungal pathogens (Lakshman, D.K. *et al*, 2012).

Pathogenicity-related genes that shows high expression profile during point of infection thus can be deduced as important causative factor that cause plant fungal infection

5.5 Public Genome Data and Bioinformatics Development

The lowering of experimental cost in genome projects had allowed more researchers to utilize the ability of whole genome sequencing technologies to sequence many species of plant pathogenic fungus for an effort to understanding the genomics reasoning behind the pathogenicity mechanisms of these pathogenic fungus to the host plants in order to identify molecular causative factors and to develop important tools for future usage. And public genome data provides an opportunity for bioinformaticians to analyze these genome data from public domain, which could lead to discovery of

important features of biological importance without having to carry out large-scale experiments, which might be costly and time-consuming. The drawbacks of using genome data from public domain is that the quality of the public domain data. Although most public domain data is good in quality, using the suitable and right data for analysis is vital as the quality of results generated is only as good as the quality of the data. Requirement of high performance computer is also required due to the massive amount of genome data that is involved in bioinformatics analysis.

The comparative workflow involves writing of customized shell scripting in order to process large dataset that is involved and generated from various bioinformatics tools and applications. The difficulty in analyzing large dataset lies not only in the performance of computing hardware as well as tuning and customization of software.

CHAPTER 6

CONCLUSION

A Plant Pathogenic Fungus Comparative Genomics Workflow had been developed for inter-phyla comparison of plant pathogenic fungus from two major phylum of fungus that constituting most plant pathogenic fungus, the phylum of Basidiomycota and the phylum of Ascomycota. By aligning the genome data from public domain to databases containing pathogenicity-related genes such as the Pathogen-Host Interaction-Related protein-coding genes and the Carbohydrate-Active Enzyme Database, candidate pathogenicity-related protein-coding genes for *B. cinerea, M, oryzae, P. graminis,* and *U. maydis* was identified. The list of candidate pathogenicity-related protein-coding genes then are screened separately according to the phylum of fungus, which is the Basidiomycota and Ascomycota before proceeding to inter-phyla comparison of the candidate protein-coding genes for pathogenicity-related function. The analysis resulted in the identification of 1,388 homologous genes across Basidiomycota and Ascomycota, 159 common Pathogen-Host Interaction-Related genes between Basidiomycota and Ascomycota, 64 common candidate protein-coding genes for Carbohydrate-Active Enzyme. Also identified from copy number variation analysis is 5 genes copy number variations and 3 genes copy number variations respectively for Pathogen-Host Interaction-Related genes and Carbohydrate-Active Enzymes.

REFERENCES

Garcia-Solache, M.A. & Casadevail, A. (2010). Global Warming will bring New Fungal Diseases for Mammals. *mBio*. 1.

Hawksworth, D.L. (2001). The Magnitude of Fungal Diversity: the 1.5 million species estimate revisted. *Mycol Res.* 105(1422-1432).

Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. & Gurr, S.J. (2012). Emerging Fungal Threats to Animal, Plant and Ecosystem Health. *Nature*. 484(186-194). doi: 10.1038/nature10947

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughian, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcel, S., Nusbaum, C., Galagan, J.E. & Birren C.W. (2005). The Genome Sequence of the Rice Blast Fungus *Magnaporthe grisea. Nature*. 434(980-986). doi: 10.1038/nature03449

Ustilago maydis Sequencing Project. Broad Institute of MIT and Harvard (http://www.broad.mit.edu)

Hamer, J.E., Howard, R.J., Chumley, F.G. & Valent, B. (1988). A Mechanism for Surface Attachment in Spores of a Plant Pathogenic Fungus. *Science*. 239(288-290).
Dean, R.A. (1997). Signal Pathways and Appressorium Morphogenesis. *Annu. Rev. Phyopathol.* 35(211-234).

van Baarlen, P., Woltering, E.J., Staats, M. and van Kan, J.A.L. (2007). Histochemical and Genetic Analysis of Host and Non-Host Interactions of *Aarabidopsis* with three *Botrytis* species: an Important Role for Cell Death Control. *Mol. Plant Pathol.* 8, 41-54.

Dean, R., van Kan, J.A.L., Pretorious, Z.A., Hammond-Kosack, K.E., Di Pietros, A., Spanu, A.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J. and Foster, G.D. (2012). The Top 10 Fungal Pathogens in Molecular Plant Pathology. *Molecular Plant Pathology*. 13(4), 414-430.

Leroch, M., Kretschmer, M. and Hahn, M. (2011) Fungicide resistance phenotypes of Botrytis cinerea isolates from commercial vineyards in South West Germany. *J. Phytopathol*. 159, 63–65.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P.M., de Vries, R.P., Dyer, P.S., Filinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K.M., Pradier, J.M., Quevillon, E., Sharon, A., Simon, A., ten Have, A., Tudzynski, B., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collemare, J., Cotton, P., Danchin, E.G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Guldener, U., Henrissat, B., Howlett, B.J., Kodira, C., Krestchmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuveglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Segurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D.M., Talbot, N.J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J.A, Lebrun, M. & Dickman, M. (2011). Genomic Analysis of the Necrotrophic

Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics*. 7(8). 1-27.

Bolten, N.D., Kolmer, J.A. & Garvin, D.F. (2008). Wheat Leaf Rust caused by *Puccinia triticina*. *Mol. Plant Pathol.* 9, 563-575.

Voegele, R.T. & Cubeta, M.A. (1994). Molecular Systematics and Population Biology of *Rhizoctonia*. *Annu. Rev. Phytopathol.* 32, 135-155.

Hodson, D.P. (2011). Shifting Boundaries: Challenges for Rust Monitoring. *Euphytica*. 179, 93-104.

Leonard, K.J. & Szabo, L.S. (2005). Stem Rust of Small Grains and Grasses Caused by *Puccinia graminis*. *Mol. Plant Pathol.* 6, 99-111.

Duplessis, S., Cuomo, C.A., Lin, Y., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., Chiu, R., Coutinho, P.M., Feau, N., Field., M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M.G., Kodira, C.D., Kohler, A., Kues, U., Lindquist, E.A., Lucas, S.M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J.L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A.A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G.A., Henrissat, B., Van de Peer, Y., Rouze, P., Ellis, J.G., Dodds, P.N., Schein, J.E., Zhong, S., Hamelin, R.C., Grigoriev, I.V., Szabo, L.J. & Martin, F. (2011). Obligate Biotrophy Features Unraveled by the Genomic Analysis of Rust Fungi. *PNAS*. 108(22), 9166-9171.

Kamper, J., Kahmann, R., Bolker, M., Ma, L., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E. Muller, O., Perlin, M.H., Wosten, H.A.B., de Vries, R., Ruiz-Herrera, J., Reynaga-Pena, C.G., Snetselaar, K., McCann, M., Perez-Martin, J., Feldbrugge, M., Basse, C.W., Steinberg, G., Ibeas, J.I., Holloman, W., Guzman, P., Farman, M., Stajich, J.E., Sentandreu, R., Gonzalez-Preito, J.M., Kennell, J.C., Molina, L., Schirawski, J., Mendoza-Mendoza, A., Greilinger, D., Munch, K., Rossel, N., Scherer, M., Vranes, M., Ladendorf, O., Vincon, V., Fuchs, U., Sandrock, B., Meng S., Ho, E.C.H., Cahill, M.J., Boyce, K.J., Klose, J., Klosterman, S.J., Deelstra, H.J., Ortiz-Castellanos, L., Li, W., Sanchez-Alonso, P., Schreier, P.H., Hauser-Hahn, I., Vaupel, M., Koopmann, E., Friedrich, G., Voss, H., Schluter, T., Margolis, J., Platt, D., Swimmer, C., Gnirke, A., Chen, F., Vysotskaia, V., Mewes, H., Mauceli, E.W., DeCaprio, D., Wade, C.M., Butler, J., Young, S., Jaffe, D.B., Calvo, S., Nusbaum, C., Galagan, J. & Birren, B.W. (2006). Insights from the Genome of the Biotrophic Fungal Plant Pathogen *Ustilago maydis*

Mutz, K., Heilkenbrinker, A., Lonne, M., Walter, J. & Stahl, F. (2012). Transcriptome Analysis using Next-Generation Sequencing. *Current Opinion in Biotechnology*. 21(1), 22-30.

Yandell, M. & Ence, D. (2012). A Beginner's Guide to Eukaryotic Genome Annotation. *Nature Reviews Genetics*. 13, 329-342.

Davey, J.W., Hoheniohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter M.L. (2011). Genome-wide Genetic Marker Discovery and Genotyping using Next-Generation Sequencing. *Nature Reviews Genetics*. 12, 499-510.

Wei, L., Liu, Y., Dubchak, I., Shon, J. & Park, J. (2002). Comparative Genomics Approaches to Study Organism Similarities and Differences. *Journal of Biomedical Informatics*. 35, 142-150.

Broad Institute (2014). Fungal Genomics. Retrieved from: http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genomics

Illumina Inc (2014). HiSeq 2500. Retrieved from: http://systems.illumina.com/systems/hiseq_2500_1500.html

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. (2002). A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics. *Genet. Sel. Evol.* 34, 275-305.

Toth, G., Gaspari, Z. & Jurka, J. (2000). Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Research*. 10(7), 967-981.

Shigemizu, D., Fujimoto, A., Akiyama, S., Abe, T., Nakano, K., Boroevich, K.A., Yamamoto, Y., Furuta, M., Kubo, M., Nakagawa, H. & Tsunoda, T. (2013). A Practical Method to Detect SNVs and Indels from Whole Genome and Exome Sequencing Data. *Scientific Reports*. 3(2161).

Bechara, M.D., Moretzsohn, M.C., Palmieri, D.A., Monteiro, J.P., Bacci, M., Martins, J., Valls, J.F.M., Lopes, C.R. & Gimenes, M.A. (2010). Phylogenetic Relationships in Genus *Arachis* based on ITS and 5.8S rDNA Sequences. *BMC Plant Biology*. 10, 255.

Genome 10K (2009). Genome 10K Project. Retrieved from: http://genome10k.soe.ucsc.edu/

BGI (2011). 10,000 Microbial Genome Project. Retrieved from: http://ldl.genomics.org.cn/page/M-research.jsp

DOE Joint Genome Institute (2014). Retrieved from: http://jgi.doe.gov/

NCBI (2009). Retrieved from: http://www.ncbi.nlm.nih.gov/sra

Tudzynski, P. & Sharon, A., (2003). Fungal Pathogenicity Genes. Retrieved from: http://www2.tau.ac.il/lifesci/plantsci/as/articles/Virulence%20genes%20Tudz%20&%20Sharon%202003.pdf

Kronstadt, J.W. (1997). Virulence and cAMP in Smuts, Blasts and Blights. *Trends Plant Sci*. 2, 193-199.

Tudzynski, B. & Tudzynski, P. (2001). Pathogenicity Factors and Signal Transduction in Plant-Pathogenic Fungi. *Prog. Bot.*. 63, 163-188.

Koller, W., Parker, D.M. & Becker, C.M. (1991). Role of Cutinase in the Penetration of Apple Leaves by *Venturia inaequalis*. *Phytopathology*. 81, 1375-1379.

Choi, J., Kim, K., Jeon, J. & Lee, Y. (2013). Fungal Plant Cell Wall-Degrading Enzyme Database: A Platform for Comparative and Evolutionary Genomics in Fungi and Oomycetes. *BMC Genomics*. 14, S7.

Skamnioti, P. & Gurr, S.J. (2007). *Magnaporthe oryzae* Cutinase2 Mediates Appressorium Differentiation and Host Penetration and is Required for Full Virulence. *Plant Cell*. 19(8), 2674-2689.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M. & Henrissat, B. (2013). The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Research*. 42.

Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J. and Hammond-Kosack, K.E. 2006. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Research*. 34(Database issue):D459-D464

Delcher, A.L., Phillippy, A., Carlton, J. and Saizberg, S.L. (2002). Fast Algorithms for Large-Scale Genome Alignment and Comparison. *Nucleic Acids Research*. 30(11), 2478-2483.

Guarro, J. Gene, J. & Stchigel, A.M. (1999). Developments in Fungal Taxonomy. *Clinical Microbiology Reviews*. 12(3), 454-500.

Pendleton, A.L., Smith, K.E., Feau, N., Martin, F.M., Grigoriev, I.V., Hamelin, R., Nelson, C.D. Burleigh, J.G. & Davis, J.M. (2014). Duplications and Losses in Gene Families of Rust Pathogens Highlight Putative Effectors. *Frontiers in Plant Science*. 5(299).

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Kryiov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdiov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. & Natale, D.A. (2003). The COG Database: An Updated Version Includes Eukaryotes. *BMC Bioinformatics*. 4(41).

Magee, P.T. (2010). Fungal Pathogenicity and Morphological Switches. *Nature Genetics*. 42, 560-561.

Idnurm, A. & Howlett, B.J. (2001). Pathogenicity Genes of Phytoathogenic Fungi. *Mol Plant Pathol*. 2, 241-255.

Manning, V.A., Pandelova, I., Dhillon, B., Wilhelm, L.J., Goodwin, S.B., Berlin, A.M., Figueroa, M., Freitag, M., Hane, J.K., Henrissat, B., Holman, W.H., Kodira, C.D., Martin, J., Oliver, R.P., Robbertse, B., Schakwitz, W., Schwartz, D.C., Spatafora, J.W., Turgeon, B.G., Yandava, C., Young, S., Zhou, S., Zeng, Q., Grigoriev, I.V., Ma, L. & Ciuffetti, L.M. (2013). Comparative enomics of a Plant-Pathogenic Fungus, *Pyrenophora tritici-repentis*, Reveals Transduplication and the Impact of Repeat Elements on Pathogencity and Population Divergence. *Genetics Society of America*. 3(1), 41-63.

Lakshman, D.K., Alkharouf, N., Roberts, D.P., Natarajan, S.S. & Mitra, A. (2012). Gene Expression Profiling of the Plant Pathogenic Basidiomycetous Fungus *Rhizoctonia solani* AG 4 Reveals Putative Virulence Factors. *Mycologia*. 104(5), 1020-1035.

Liu, T.B., Wang, Y., Baker, G.M., Fahmy, H., Jiang, L. & Xue, C. (2013). The Glucose Sensor-like Protein Hxs1 is a Highly-Affinity Glucose Transporter and Required for Virulence in *Cryptococcus neoformans*. *PLoS One*. 8(5).

Zemkova, Z., Garajova, S., Flodrova, D., Rehulka, P., Zelko, I., Vadkertiova, R., Farkas, V. & Stratilova, E. (2012). Incorporation of Beta-(1,6)-linked Glucooligosaccharides (Pustilooligosaccharides) into Plant Cell Wall Structures. *Chemical Papers*. 66(9).

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. (2012). dbCAN: a Web Resource for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res*. 40, 445-451.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403-410.

Kurtz, S., Phillyppy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S.L. (2003). Versatile and Open Software for Comparing Large Genomes. *Genome Biology*. 5.

Oliveros, J.C. (2007). VENNY. An Interactive Tool for Comparing Lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html

Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. *Cambridge University Press*.

APPENDIX

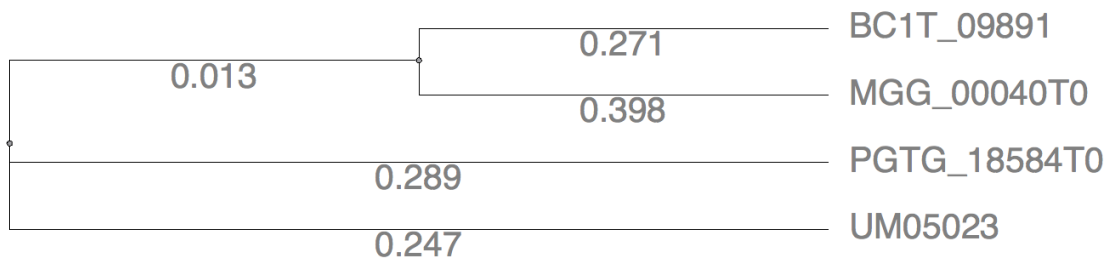

Figure 4.4: Phylogenetic Analysis of PHI:2968 from 4 Plant Pathogenic Fungus.

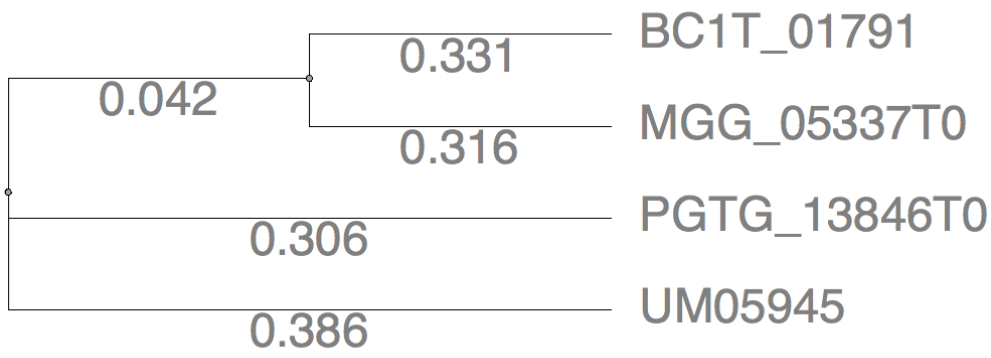

Figure 4.6: Phylogenetics Analysis of Candidate Carbohydrate-Active Enzyme of AA7 Family

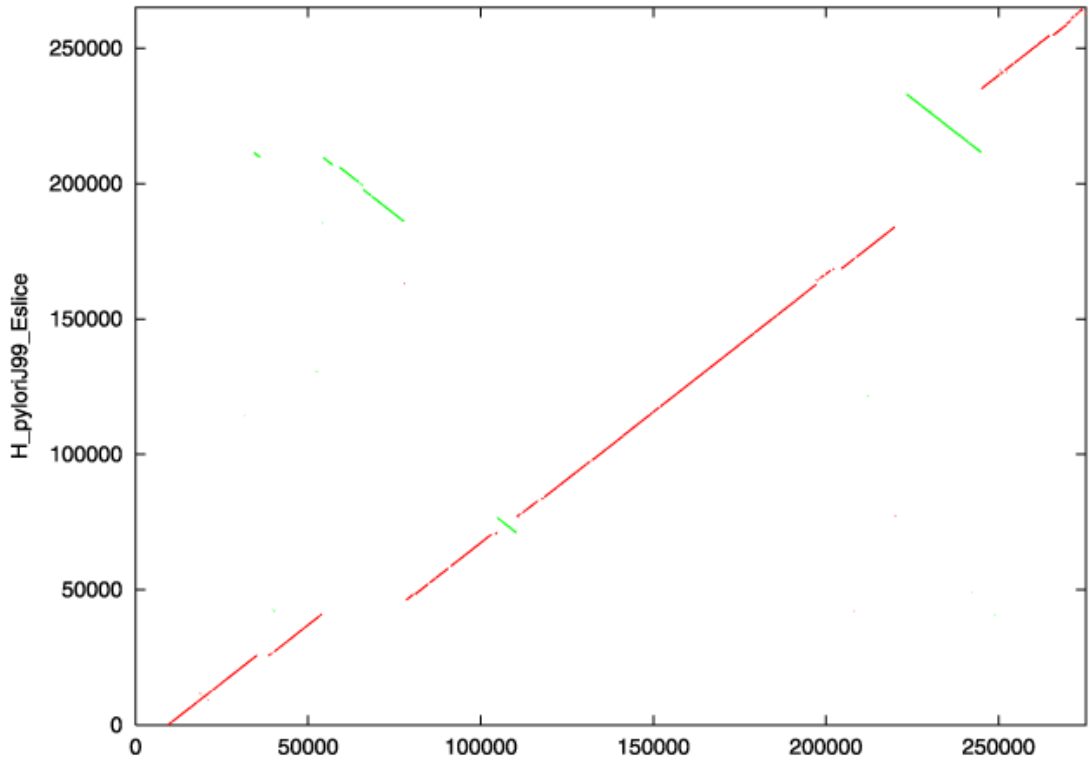

Figure 5.1: Example of Quality Alignment between Two Sequences using MUMmer. 3.0.

Table 5.1: Number of Pathogen-Host Interaction-Related Genes Identified from Plant Pathogenic Fungus of Study.

Phylum	**Species**	**Number of PHIbase Protein-coding genes**
Ascomycota	*Botrytis cinerea*	1339
	Magnaporthe oryzae	1402
Basidiomycota	*Puccinia graminis*	626
	Ustilago maydis	533

Table 5.2: Number of Pathogen-Host Interaction-Related Genes Identified from Plant Pathogenic Fungus of Study.

Family	*Botrytis cinerea*	*Magnaporthe oryzae*	*Puccinia graminis*	*Ustilago maydis*
AA	85	105	24	30
CBM	69	120	16	10
CE	119	136	74	64
GH	250	272	161	117
GT	104	103	102	69
PL	11	5	7	3
Total	**638**	**741**	**384**	**293**

Table 4.6: List of Common Carbohydrate-Active Enzymes in Four Plant Pathogenic Fungus from Basidiomycota and Ascomycota

	B. cinerea	*M. oryzae*	*P. graminus*	*U. maydis*
AA1.hmm	5	3	1	2
AA2.hmm	4	13	2	5
AA3.hmm	25	19	7	10
AA5.hmm	5	4	4	4
AA6.hmm	1	1	1	1
AA7.hmm	26	33	6	6
CBM13.hmm	4	3	1	2
CBM43.hmm	1	2	1	1
CBM48.hmm	1	1	2	1
CE10.hmm	55	50	15	32
CE14.hmm	1	1	3	1
CE1.hmm	22	35	13	15
CE4.hmm	6	12	18	8
CE5.hmm	12	19	9	4
CE8.hmm	5	1	9	1
GH105.hmm	2	3	2	2
GH109.hmm	6	8	2	6
GH10.hmm	2	6	5	2
GH13.hmm	9	9	4	3
GH15.hmm	3	2	3	1
GH16.hmm	22	18	10	27
GH17.hmm	5	7	2	2
GH18.hmm	10	17	16	4
GH20.hmm	1	3	2	2
GH27.hmm	4	3	6	1
GH28.hmm	21	4	1	1
GH2.hmm	2	8	8	1
GH31.hmm	5	6	3	3
GH32.hmm	3	5	2	2
GH37.hmm	1	2	3	2
GH38.hmm	1	2	1	2
GH3.hmm	16	18	2	3
GH43.hmm	6	19	2	2
GH47.hmm	10	9	11	3
GH5.hmm	16	13	29	13
GH63.hmm	1	1	2	1
GH72.hmm	6	5	1	1
GH74.hmm	2	5	2	2
GH76.hmm	10	8	7	1
GH79.hmm	2	2	1	1
GT15.hmm	3	4	2	2
GT1.hmm	12	12	6	3

Table 4.6, continued.

GT20.hmm	3	3	2	4
GT21.hmm	1	2	3	1
GT22.hmm	4	4	6	4
GT24.hmm	1	1	1	1
GT2.hmm	18	12	12	14
GT31.hmm	3	3	6	4
GT32.hmm	6	10	8	2
GT33.hmm	1	1	1	1
GT39.hmm	3	3	3	3
GT3.hmm	1	1	1	1
GT48.hmm	1	1	1	1
GT4.hmm	5	4	4	4
GT50.hmm	1	1	1	1
GT57.hmm	1	3	2	4
GT59.hmm	1	1	1	1
GT66.hmm	1	1	1	1
GT69.hmm	4	4	4	3
GT71.hmm	5	4	1	2
GT76.hmm	1	1	1	1
GT8.hmm	6	2	1	2
GT90.hmm	5	7	12	5
PL1.hmm	7	2	2	1
Total	433	467	301	247

Table 4.7: List of Common Pathogen-Host Interaction-Related Genes in Four Plant Pathogenic Fungus from Basidiomycota and Ascomycota

	B. cinerea	*M. oryzae*	*P. graminus*	*U. maydis*
PHI:1047	5	2	1	3
PHI:1048	1	3	5	1
PHI:1057	1	1	1	1
PHI:1061	1	1	1	1
PHI:1133	2	1	2	1
PHI:1161	9	3	2	2
PHI:1172	1	1	1	1
PHI:1178	1	1	1	1
PHI:1200	1	1	1	1
PHI:1232	1	1	1	1
PHI:1234	1	1	1	1
PHI:1244	1	1	1	1
PHI:1248	1	1	1	1
PHI:1288	1	1	1	1
PHI:1289	1	1	1	1
PHI:1375	1	1	1	1
PHI:1388	1	1	1	1
PHI:1454	2	1	2	2
PHI:1467	1	1	1	1
PHI:1530	1	1	1	1
PHI:1552	2	1	9	1
PHI:1555	9	18	7	6
PHI:1562	1	2	2	1
PHI:1566	3	4	3	4
PHI:1567	1	1	1	1
PHI:1572	1	1	1	1
PHI:1577	1	1	1	1
PHI:1579	5	5	6	2
PHI:1582	1	1	1	1
PHI:1584	1	1	1	1
PHI:1587	1	1	1	1
PHI:1595	1	1	1	1
PHI:1602	2	1	1	1
PHI:1603	1	1	1	1
PHI:1604	1	1	2	1
PHI:1618	2	1	2	1
PHI:1662	10	21	1	6
PHI:1670	1	1	1	2
PHI:178	2	1	1	1
PHI:182	1	1	1	1
PHI:194	1	1	2	1
PHI:195	1	1	1	1

Table 4.7, continued.

PHI:200	1	1	1	1
PHI:2020	4	5	6	3
PHI:2034	4	4	3	1
PHI:2038	8	9	3	6
PHI:2075	2	1	1	1
PHI:2084	1	1	1	1
PHI:2086	1	1	1	1
PHI:2087	2	1	2	1
PHI:2096	2	2	1	1
PHI:2097	1	2	2	2
PHI:2100	1	1	2	1
PHI:2101	3	2	2	1
PHI:213	1	1	1	1
PHI:2155	1	1	1	1
PHI:2171	3	4	1	1
PHI:2179	1	1	2	1
PHI:2183	2	4	1	3
PHI:2194	1	1	1	1
PHI:220	2	1	1	1
PHI:2203	1	2	1	1
PHI:2205	1	29	14	4
PHI:2244	2	2	9	3
PHI:2248	3	1	1	1
PHI:2255	2	2	4	2
PHI:2256	10	5	2	2
PHI:2259	1	2	1	1
PHI:2267	1	1	1	1
PHI:2269	9	5	6	2
PHI:2293	1	1	1	1
PHI:2321	9	8	2	5
PHI:2322	6	5	3	4
PHI:2329	10	4	1	1
PHI:2336	3	2	1	2
PHI:235	2	1	1	1
PHI:2351	1	1	1	1
PHI:2356	1	1	1	1
PHI:2357	15	13	7	2
PHI:2382	3	3	4	3
PHI:2393	4	5	3	2
PHI:244	1	1	1	2
PHI:2474	1	1	1	1
PHI:2491	2	1	1	1
PHI:2510	4	4	4	1
PHI:2513	1	1	1	1
PHI:2517	1	1	2	1
PHI:2520	2	2	1	2

Table 4.7, continued.

PHI:2522	1	1	1	1
PHI:2524	1	1	1	1
PHI:2525	1	1	1	1
PHI:2529	2	2	2	2
PHI:2530	2	2	1	1
PHI:2531	1	1	1	1
PHI:2533	3	1	1	1
PHI:2537	1	1	1	1
PHI:254	1	1	2	1
PHI:2540	1	1	1	1
PHI:2545	1	1	2	1
PHI:2546	1	1	2	1
PHI:2553	1	1	2	1
PHI:2568	1	1	1	1
PHI:2570	5	6	1	2
PHI:2604	1	1	2	1
PHI:262	1	1	1	1
PHI:2625	1	1	1	1
PHI:2638	1	2	1	2
PHI:2640	1	1	1	1
PHI:267	2	2	3	2
PHI:2728	3	3	3	1
PHI:280	1	1	1	1
PHI:2802	5	6	3	3
PHI:2915	1	1	1	1
PHI:2920	1	2	4	1
PHI:2959	2	3	2	1
PHI:2960	1	1	1	1
PHI:2961	2	1	1	1
PHI:2968	33	36	9	10
PHI:2969	1	1	3	1
PHI:2970	1	1	1	1
PHI:2976	7	5	15	4
PHI:305	1	1	2	1
PHI:336	1	1	1	1
PHI:339	10	7	9	7
PHI:358	3	3	1	2
PHI:367	1	1	1	1
PHI:391	1	1	4	2
PHI:419	11	8	2	2
PHI:423	2	3	4	2
PHI:424	1	1	1	1
PHI:435	1	1	1	1
PHI:438	30	44	2	4
PHI:440	5	3	4	4
PHI:442	1	1	1	1

Table 4.7, continued.

PHI:443	4	4	2	1
PHI:445	1	1	1	1
PHI:447	2	2	1	1
PHI:454	1	1	1	1
PHI:465	1	1	1	1
PHI:504	4	6	1	6
PHI:508	4	2	2	4
PHI:511	8	2	1	6
PHI:538	11	7	3	1
PHI:541	10	2	2	2
PHI:55	4	3	1	2
PHI:598	2	3	2	3
PHI:668	1	1	1	2
PHI:697	4	2	3	2
PHI:748	3	2	2	1
PHI:784	9	7	4	4
PHI:806	1	1	1	1
PHI:807	1	1	2	1
PHI:823	2	1	1	2
PHI:854	1	1	2	1
PHI:877	1	1	2	1
PHI:881	6	3	2	4
PHI:901	4	4	4	4
PHI:911	3	1	4	3
PHI:922	19	10	8	9
Total	484	474	342	283

Figure 5.2: Head Blast Disease caused by *Magnaporthe oryzae* (B), (A) *Magnaporthe oryzae* causing panicle blast on rice. (Dean, R. *et al*, 2012)

Figure 5.3: Stem Rust Disease caused by *Puccinia graminis* in Wheat. (Dean, R. *et al*, 2012)

Table 4.8: List of All Homologous Pathogen-Host Interaction-Related Genes for all four fungus.

M.oryzae	***B. cinerea***	***U. maydis***	***P. graminis***
PHI:1006	PHI:1006	PHI:100	PHI:1026
PHI:1008	PHI:101	PHI:1030	PHI:1028
PHI:1023	PHI:1022	PHI:1047	PHI:1029
PHI:1028	PHI:1023	PHI:1048	PHI:1030
PHI:1034	PHI:1024	PHI:1052	PHI:1035
PHI:1046	PHI:1025	PHI:1057	PHI:1047
PHI:1047	PHI:1028	PHI:1061	PHI:1048
PHI:1048	PHI:1029	PHI:1071	PHI:105
PHI:1049	PHI:1030	PHI:1133	PHI:1052
PHI:105	PHI:1031	PHI:1161	PHI:1055
PHI:1051	PHI:1032	PHI:1167	PHI:1057
PHI:1052	PHI:1034	PHI:1172	PHI:106
PHI:1054	PHI:104	PHI:1178	PHI:1061
PHI:1057	PHI:1046	PHI:1187	PHI:1133
PHI:1058	PHI:1047	PHI:1190	PHI:1161
PHI:1061	PHI:1048	PHI:1197	PHI:1172
PHI:1063	PHI:1049	PHI:12	PHI:1178
PHI:1071	PHI:105	PHI:1200	PHI:1190
PHI:112	PHI:1051	PHI:1218	PHI:1193
PHI:113	PHI:1055	PHI:1232	PHI:1200
PHI:1133	PHI:1056	PHI:1234	PHI:1201
PHI:1135	PHI:1057	PHI:1244	PHI:121
PHI:115	PHI:106	PHI:1248	PHI:1210
PHI:1161	PHI:1061	PHI:1288	PHI:1212
PHI:1162	PHI:1071	PHI:1289	PHI:1222
PHI:1167	PHI:112	PHI:133	PHI:1227
PHI:1172	PHI:1133	PHI:1375	PHI:1228
PHI:1174	PHI:1135	PHI:1388	PHI:1232
PHI:1175	PHI:114	PHI:1397	PHI:1234
PHI:1177	PHI:115	PHI:14	PHI:1235
PHI:1178	PHI:1159	PHI:1454	PHI:1244
PHI:1180	PHI:1161	PHI:1456	PHI:1247
PHI:1182	PHI:1162	PHI:1467	PHI:1248
PHI:1184	PHI:1164	PHI:1527	PHI:1249
PHI:1187	PHI:1166	PHI:1530	PHI:1251
PHI:1190	PHI:1167	PHI:1551	PHI:1287
PHI:1192	PHI:1172	PHI:1552	PHI:1288
PHI:1193	PHI:1173	PHI:1555	PHI:1289
PHI:1198	PHI:1175	PHI:1560	PHI:1371
PHI:12	PHI:1176	PHI:1561	PHI:1375
PHI:1200	PHI:1177	PHI:1562	PHI:1376
PHI:1201	PHI:1178	PHI:1566	PHI:138
PHI:1203	PHI:1179	PHI:1567	PHI:1383`

Table 4.8, continued.

PHI:1205	PHI:1180	PHI:157	PHI:1388
PHI:1206	PHI:1181	PHI:1571	PHI:14
PHI:1207	PHI:1184	PHI:1572	PHI:1411
PHI:1208	PHI:1185	PHI:1577	PHI:1419
PHI:121	PHI:1187	PHI:1579	PHI:143
PHI:1210	PHI:1193	PHI:1582	PHI:144
PHI:1214	PHI:1197	PHI:1584	PHI:1454
PHI:1216	PHI:1198	PHI:1585	PHI:1458
PHI:1218	PHI:1200	PHI:1587	PHI:1467
PHI:1219	PHI:1203	PHI:1588	PHI:1468
PHI:1220	PHI:1204	PHI:1595	PHI:1526
PHI:1222	PHI:1207	PHI:160	PHI:1529
PHI:1223	PHI:1208	PHI:1602	PHI:1530
PHI:1225	PHI:121	PHI:1603	PHI:1533
PHI:1226	PHI:1216	PHI:1604	PHI:1542
PHI:1227	PHI:1218	PHI:1608	PHI:1551
PHI:1228	PHI:1220	PHI:1618	PHI:1552
PHI:1232	PHI:1222	PHI:1627	PHI:1555
PHI:1234	PHI:1223	PHI:1629	PHI:1562
PHI:1235	PHI:1224	PHI:1637	PHI:1563
PHI:1236	PHI:1225	PHI:1649	PHI:1566
PHI:1237	PHI:1226	PHI:1662	PHI:1567
PHI:1238	PHI:1227	PHI:1670	PHI:1569
PHI:1240	PHI:1228	PHI:178	PHI:1570
PHI:1241	PHI:1232	PHI:182	PHI:1572
PHI:1242	PHI:1233	PHI:187	PHI:1573
PHI:1244	PHI:1234	PHI:191	PHI:1577
PHI:1246	PHI:1236	PHI:194	PHI:1579
PHI:1247	PHI:1240	PHI:195	PHI:158
PHI:1248	PHI:1242	PHI:200	PHI:1582
PHI:1249	PHI:1243	PHI:2019	PHI:1584
PHI:125	PHI:1244	PHI:2020	PHI:1587
PHI:1251	PHI:1246	PHI:2034	PHI:159
PHI:1252	PHI:1247	PHI:2038	PHI:1590
PHI:1253	PHI:1248	PHI:2042	PHI:1595
PHI:1255	PHI:1249	PHI:2052	PHI:1597
PHI:1256	PHI:1251	PHI:2054	PHI:1602
PHI:1257	PHI:1252	PHI:2075	PHI:1603
PHI:1263	PHI:1253	PHI:2084	PHI:1604
PHI:1267	PHI:1255	PHI:2086	PHI:1605
PHI:1268	PHI:1257	PHI:2087	PHI:1608
PHI:1269	PHI:1259	PHI:2094	PHI:1618
PHI:1273	PHI:1260	PHI:2096	PHI:1662
PHI:1277	PHI:1267	PHI:2097	PHI:1670
PHI:128	PHI:1268	PHI:2098	PHI:1681
PHI:1285	PHI:1273	PHI:2099	PHI:1685

Table 4.8, continued.

PHI:1287	PHI:1287	PHI:2100	PHI:178
PHI:1288	PHI:1288	PHI:2101	PHI:182
PHI:1289	PHI:1289	PHI:2117	PHI:188
PHI:1299	PHI:1291	PHI:213	PHI:194
PHI:1300	PHI:1294	PHI:2155	PHI:195
PHI:1302	PHI:1296	PHI:2171	PHI:200
PHI:1303	PHI:1299	PHI:2179	PHI:2020
PHI:1304	PHI:1300	PHI:2183	PHI:2025
PHI:1309	PHI:1302	PHI:2194	PHI:2034
PHI:1313	PHI:1303	PHI:22	PHI:2038
PHI:1316	PHI:1304	PHI:220	PHI:2050
PHI:1317	PHI:1309	PHI:2203	PHI:2054
PHI:1318	PHI:131	PHI:2205	PHI:2060
PHI:1319	PHI:1310	PHI:221	PHI:2074
PHI:132	PHI:1316	PHI:2217	PHI:2075
PHI:1325	PHI:1317	PHI:2219	PHI:2079
PHI:1326	PHI:1319	PHI:2220	PHI:208
PHI:1327	PHI:1326	PHI:2222	PHI:2084
PHI:133	PHI:133	PHI:2224	PHI:2086
PHI:1333	PHI:1333	PHI:2226	PHI:2087
PHI:1343	PHI:1337	PHI:2227	PHI:2091
PHI:1344	PHI:1338	PHI:2228	PHI:2096
PHI:1348	PHI:1344	PHI:2229	PHI:2097
PHI:1350	PHI:1348	PHI:2230	PHI:2099
PHI:1354	PHI:1353	PHI:2231	PHI:2100
PHI:1356	PHI:1356	PHI:2233	PHI:2101
PHI:1357	PHI:1357	PHI:2234	PHI:2104
PHI:1358	PHI:1358	PHI:2237	PHI:2109
PHI:1363	PHI:1359	PHI:2239	PHI:211
PHI:1369	PHI:1367	PHI:2240	PHI:2113
PHI:1371	PHI:1374	PHI:2244	PHI:2114
PHI:1374	PHI:1375	PHI:2246	PHI:213
PHI:1375	PHI:1376	PHI:2248	PHI:2140
PHI:1376	PHI:1377	PHI:2255	PHI:2155
PHI:1379	PHI:1378	PHI:2256	PHI:2171
PHI:1382	PHI:1379	PHI:2259	PHI:2179
PHI:1388	PHI:1388	PHI:2266	PHI:2183
PHI:1389	PHI:1393	PHI:2267	PHI:2189
PHI:1397	PHI:1402	PHI:2269	PHI:2194
PHI:1406	PHI:1403	PHI:2275	PHI:220
PHI:1407	PHI:1406	PHI:2293	PHI:2203
PHI:1408	PHI:1407	PHI:23	PHI:2205
PHI:141	PHI:1408	PHI:2309	PHI:221
PHI:1410	PHI:1410	PHI:2310	PHI:2224
PHI:1411	PHI:1411	PHI:2318	PHI:2227
PHI:1412	PHI:1413	PHI:2321	PHI:2228

Table 4.8, continued.

PHI:1413	PHI:1415	PHI:2322	PHI:2230
PHI:1414	PHI:1419	PHI:2329	PHI:2231
PHI:1419	PHI:1422	PHI:2336	PHI:2239
PHI:1420	PHI:1423	PHI:235	PHI:2244
PHI:143	PHI:143	PHI:2351	PHI:2246
PHI:1439	PHI:1432	PHI:2356	PHI:2248
PHI:144	PHI:144	PHI:2357	PHI:2255
PHI:1441	PHI:1440	PHI:2363	PHI:2256
PHI:1447	PHI:1447	PHI:2378	PHI:2257
PHI:1449	PHI:1449	PHI:2381	PHI:2259
PHI:1451	PHI:1451	PHI:2382	PHI:2267
PHI:1453	PHI:1452	PHI:2384	PHI:2269
PHI:1454	PHI:1453	PHI:2393	PHI:2271
PHI:1456	PHI:1454	PHI:24	PHI:2293
PHI:1457	PHI:1456	PHI:244	PHI:2296
PHI:1458	PHI:1457	PHI:2441	PHI:2297
PHI:1461	PHI:1458	PHI:2474	PHI:2305
PHI:1463	PHI:1460	PHI:2491	PHI:2321
PHI:1464	PHI:1467	PHI:2497	PHI:2322
PHI:1466	PHI:1472	PHI:2498	PHI:2329
PHI:1467	PHI:1475	PHI:2510	PHI:2336
PHI:1475	PHI:1478	PHI:2513	PHI:2338
PHI:1479	PHI:1498	PHI:2517	PHI:2339
PHI:1486	PHI:1500	PHI:2518	PHI:235
PHI:1487	PHI:1506	PHI:2520	PHI:2350
PHI:1489	PHI:1515	PHI:2522	PHI:2351
PHI:1492	PHI:1517	PHI:2524	PHI:2356
PHI:15	PHI:1520	PHI:2525	PHI:2357
PHI:1500	PHI:1522	PHI:2529	PHI:2359
PHI:1503	PHI:1525	PHI:2530	PHI:236
PHI:1515	PHI:1526	PHI:2531	PHI:237
PHI:1517	PHI:1527	PHI:2533	PHI:2382
PHI:1519	PHI:1529	PHI:2535	PHI:2384
PHI:1525	PHI:1530	PHI:2537	PHI:2386
PHI:1526	PHI:1531	PHI:254	PHI:2393
PHI:1529	PHI:1535	PHI:2540	PHI:2414
PHI:153	PHI:1539	PHI:2544	PHI:244
PHI:1530	PHI:1542	PHI:2545	PHI:2441
PHI:1531	PHI:1543	PHI:2546	PHI:2453
PHI:1533	PHI:1550	PHI:2553	PHI:2474
PHI:1542	PHI:1551	PHI:2558	PHI:2488
PHI:1543	PHI:1552	PHI:2568	PHI:249
PHI:1550	PHI:1553	PHI:2570	PHI:2491
PHI:1552	PHI:1554	PHI:26	PHI:2503
PHI:1554	PHI:1555	PHI:260	PHI:2510

Table 4.8, continued.

PHI:1555	PHI:1559	PHI:2602	PHI:2513
PHI:1559	PHI:1562	PHI:2603	PHI:2515
PHI:1562	PHI:1563	PHI:2604	PHI:2517
PHI:1563	PHI:1565	PHI:2605	PHI:2520
PHI:1564	PHI:1566	PHI:2607	PHI:2521
PHI:1565	PHI:1567	PHI:2608	PHI:2522
PHI:1566	PHI:1568	PHI:2609	PHI:2524
PHI:1567	PHI:1569	PHI:262	PHI:2525
PHI:1568	PHI:157	PHI:2625	PHI:2528
PHI:1569	PHI:1570	PHI:2638	PHI:2529
PHI:157	PHI:1571	PHI:2640	PHI:2530
PHI:1571	PHI:1572	PHI:265	PHI:2531
PHI:1572	PHI:1573	PHI:2651	PHI:2532
PHI:1573	PHI:1575	PHI:267	PHI:2533
PHI:1574	PHI:1576	PHI:269	PHI:2537
PHI:1575	PHI:1577	PHI:270	PHI:254
PHI:1576	PHI:1578	PHI:2700	PHI:2540
PHI:1577	PHI:1579	PHI:2728	PHI:2544
PHI:1578	PHI:1581	PHI:2744	PHI:2545
PHI:1579	PHI:1582	PHI:2748	PHI:2546
PHI:1580	PHI:1584	PHI:280	PHI:2553
PHI:1581	PHI:1585	PHI:2802	PHI:256
PHI:1582	PHI:1587	PHI:2832	PHI:2568
PHI:1584	PHI:1588	PHI:2839	PHI:257
PHI:1585	PHI:1589	PHI:2852	PHI:2570
PHI:1587	PHI:159	PHI:2853	PHI:2597
PHI:1589	PHI:1591	PHI:2854	PHI:2601
PHI:159	PHI:1592	PHI:2855	PHI:2602
PHI:1590	PHI:1595	PHI:2856	PHI:2604
PHI:1592	PHI:1596	PHI:2894	PHI:2607
PHI:1595	PHI:1598	PHI:290	PHI:2611
PHI:1601	PHI:1599	PHI:2911	PHI:262
PHI:1602	PHI:160	PHI:2915	PHI:2625
PHI:1603	PHI:1601	PHI:2920	PHI:2636
PHI:1604	PHI:1602	PHI:2928	PHI:2638
PHI:1605	PHI:1603	PHI:2959	PHI:2640
PHI:1608	PHI:1604	PHI:2960	PHI:2643
PHI:1610	PHI:1610	PHI:2961	PHI:2645
PHI:1611	PHI:1611	PHI:2968	PHI:2656
PHI:1614	PHI:1612	PHI:2969	PHI:267
PHI:1615	PHI:1614	PHI:2970	PHI:269
PHI:1618	PHI:1615	PHI:2976	PHI:2710
PHI:1621	PHI:1618	PHI:299	PHI:2728
PHI:1622	PHI:1621	PHI:3	PHI:274
PHI:1627	PHI:1625	PHI:305	PHI:280

Table 4.8, continued.

PHI:1628	PHI:1627	PHI:307	PHI:2802
PHI:1629	PHI:1629	PHI:310	PHI:281
PHI:1630	PHI:1630	PHI:317	PHI:2821
PHI:1632	PHI:1631	PHI:319	PHI:2822
PHI:1633	PHI:1632	PHI:323	PHI:2826
PHI:1634	PHI:1633	PHI:336	PHI:2841
PHI:1635	PHI:1635	PHI:338	PHI:2843
PHI:1637	PHI:1637	PHI:339	PHI:2844
PHI:1643	PHI:1638	PHI:345	PHI:286
PHI:1644	PHI:1640	PHI:346	PHI:290
PHI:1645	PHI:1643	PHI:352	PHI:2909
PHI:1647	PHI:1644	PHI:358	PHI:2911
PHI:1648	PHI:1645	PHI:367	PHI:2915
PHI:1649	PHI:1647	PHI:37	PHI:2920
PHI:1651	PHI:1648	PHI:370	PHI:2933
PHI:1653	PHI:1649	PHI:376	PHI:2959
PHI:1655	PHI:1651	PHI:386	PHI:296
PHI:1657	PHI:1653	PHI:387	PHI:2960
PHI:1658	PHI:1655	PHI:389	PHI:2961
PHI:1662	PHI:1658	PHI:391	PHI:2964
PHI:1666	PHI:1659	PHI:392	PHI:2968
PHI:167	PHI:1662	PHI:394	PHI:2969
PHI:1670	PHI:1666	PHI:397	PHI:2970
PHI:1671	PHI:167	PHI:411	PHI:2976
PHI:1673	PHI:1670	PHI:413	PHI:305
PHI:1674	PHI:1671	PHI:419	PHI:316
PHI:1675	PHI:1673	PHI:420	PHI:323
PHI:1676	PHI:1674	PHI:423	PHI:324
PHI:1677	PHI:1675	PHI:424	PHI:33
PHI:1682	PHI:1676	PHI:432	PHI:336
PHI:1683	PHI:1677	PHI:435	PHI:339
PHI:1685	PHI:1678	PHI:436	PHI:346
PHI:1695	PHI:1681	PHI:438	PHI:352
PHI:1707	PHI:1682	PHI:440	PHI:355
PHI:1713	PHI:1683	PHI:442	PHI:358
PHI:1726	PHI:1685	PHI:443	PHI:362
PHI:1742	PHI:1690	PHI:445	PHI:367
PHI:1752	PHI:1695	PHI:447	PHI:386
PHI:1753	PHI:1707	PHI:454	PHI:387
PHI:1760	PHI:1712	PHI:460	PHI:391
PHI:1763	PHI:1713	PHI:464	PHI:397
PHI:1765	PHI:1721	PHI:465	PHI:419
PHI:177	PHI:1724	PHI:477	PHI:420
PHI:1772	PHI:174	PHI:478	PHI:423
PHI:1774	PHI:1746	PHI:489	PHI:424

Table 4.8, continued.

PHI:1775	PHI:1751	PHI:494	PHI:435
PHI:178	PHI:1752	PHI:497	PHI:436
PHI:1785	PHI:1753	PHI:502	PHI:438
PHI:1787	PHI:1760	PHI:504	PHI:440
PHI:179	PHI:1764	PHI:506	PHI:442
PHI:1790	PHI:1767	PHI:508	PHI:443
PHI:1792	PHI:177	PHI:511	PHI:445
PHI:1793	PHI:1773	PHI:512	PHI:447
PHI:1799	PHI:1776	PHI:513	PHI:451
PHI:1807	PHI:178	PHI:515	PHI:454
PHI:181	PHI:1784	PHI:524	PHI:465
PHI:1812	PHI:1789	PHI:528	PHI:47
PHI:1816	PHI:1792	PHI:538	PHI:497
PHI:182	PHI:1793	PHI:541	PHI:504
PHI:1825	PHI:1797	PHI:544	PHI:506
PHI:1833	PHI:1798	PHI:55	PHI:508
PHI:1838	PHI:180	PHI:57	PHI:510
PHI:1857	PHI:181	PHI:598	PHI:511
PHI:1862	PHI:1810	PHI:599	PHI:538
PHI:1869	PHI:1812	PHI:616	PHI:541
PHI:1878	PHI:1816	PHI:668	PHI:547
PHI:1879	PHI:182	PHI:675	PHI:55
PHI:1880	PHI:1821	PHI:68	PHI:566
PHI:1881	PHI:1843	PHI:693	PHI:577
PHI:1887	PHI:1856	PHI:697	PHI:598
PHI:1893	PHI:1861	PHI:72	PHI:612
PHI:19	PHI:1869	PHI:747	PHI:616
PHI:1902	PHI:1878	PHI:748	PHI:62
PHI:191	PHI:1895	PHI:749	PHI:668
PHI:1915	PHI:19	PHI:750	PHI:672
PHI:1917	PHI:191	PHI:751	PHI:674
PHI:1921	PHI:1915	PHI:752	PHI:697
PHI:1923	PHI:1917	PHI:753	PHI:716
PHI:1931	PHI:1918	PHI:754	PHI:72
PHI:1934	PHI:1920	PHI:755	PHI:747
PHI:1935	PHI:1921	PHI:756	PHI:748
PHI:194	PHI:1924	PHI:757	PHI:76
PHI:1941	PHI:1931	PHI:76	PHI:77
PHI:1943	PHI:1933	PHI:77	PHI:78
PHI:195	PHI:1935	PHI:78	PHI:784
PHI:1953	PHI:1938	PHI:784	PHI:794
PHI:1954	PHI:194	PHI:785	PHI:796
PHI:1957	PHI:1941	PHI:79	PHI:804
PHI:1959	PHI:1947	PHI:806	PHI:806
PHI:196	PHI:195	PHI:807	PHI:807

Table 4.8, continued.

PHI:1960	PHI:1953	PHI:811	PHI:812
PHI:1961	PHI:1954	PHI:823	PHI:820
PHI:1964	PHI:1957	PHI:825	PHI:823
PHI:1967	PHI:1958	PHI:832	PHI:832
PHI:197	PHI:1959	PHI:853	PHI:838
PHI:1974	PHI:1960	PHI:854	PHI:854
PHI:1982	PHI:1961	PHI:860	PHI:876
PHI:1984	PHI:1963	PHI:877	PHI:877
PHI:1986	PHI:1969	PHI:881	PHI:881
PHI:1988	PHI:197	PHI:896	PHI:888
PHI:199	PHI:1974	PHI:897	PHI:901
PHI:1991	PHI:1984	PHI:898	PHI:903
PHI:1999	PHI:1986	PHI:899	PHI:911
PHI:200	PHI:1987	PHI:900	PHI:922
PHI:2002	PHI:1988	PHI:901	PHI:923
PHI:2005	PHI:199	PHI:902	PHI:97
PHI:2008	PHI:1996	PHI:903	PHI:98
PHI:201	PHI:1998	PHI:904	
PHI:2016	PHI:200	PHI:905	
PHI:2017	PHI:2016	PHI:906	
PHI:2018	PHI:2018	PHI:907	
PHI:2019	PHI:2020	PHI:908	
PHI:2020	PHI:2022	PHI:910	
PHI:2021	PHI:2025	PHI:911	
PHI:2022	PHI:2028	PHI:912	
PHI:2025	PHI:203	PHI:913	
PHI:2027	PHI:2030	PHI:914	
PHI:2028	PHI:2032	PHI:915	
PHI:2029	PHI:2034	PHI:917	
PHI:2030	PHI:2038	PHI:918	
PHI:2032	PHI:2042	PHI:919	
PHI:2033	PHI:2052	PHI:921	
PHI:2034	PHI:2055	PHI:922	
PHI:2037	PHI:2058	PHI:923	
PHI:2038	PHI:2060	PHI:924	
PHI:2039	PHI:2062	PHI:925	
PHI:2042	PHI:2065	PHI:926	
PHI:2050	PHI:2067	PHI:927	
PHI:2051	PHI:207	PHI:928	
PHI:2052	PHI:2074	PHI:929	
PHI:2054	PHI:2075	PHI:930	
PHI:2055	PHI:2076	PHI:931	
PHI:2058	PHI:2078	PHI:932	
PHI:2059	PHI:2079	PHI:933	
PHI:2060	PHI:208	PHI:934	

Table 4.8, continued.

PHI:2062	PHI:2080	PHI:935	
PHI:2064	PHI:2082	PHI:939	
PHI:2065	PHI:2083	PHI:940	
PHI:2067	PHI:2084	PHI:941	
PHI:2068	PHI:2085	PHI:942	
PHI:2069	PHI:2086	PHI:943	
PHI:207	PHI:2087	PHI:945	
PHI:2072	PHI:2089	PHI:946	
PHI:2074	PHI:2090	PHI:947	
PHI:2075	PHI:2091	PHI:948	
PHI:2076	PHI:2092	PHI:949	
PHI:2077	PHI:2094	PHI:950	
PHI:2078	PHI:2096	PHI:951	
PHI:2079	PHI:2097	PHI:952	
PHI:208	PHI:2098	PHI:953	
PHI:2080	PHI:210	PHI:954	
PHI:2081	PHI:2100	PHI:955	
PHI:2082	PHI:2101	PHI:957	
PHI:2083	PHI:2104	PHI:958	
PHI:2084	PHI:2105	PHI:959	
PHI:2085	PHI:2107	PHI:960	
PHI:2086	PHI:2109	PHI:961	
PHI:2087	PHI:2112	PHI:962	
PHI:2088	PHI:2114	PHI:98	
PHI:2089	PHI:2117		
PHI:2090	PHI:2118		
PHI:2091	PHI:2121		
PHI:2092	PHI:2127		
PHI:2093	PHI:2128		
PHI:2094	PHI:213		
PHI:2095	PHI:2140		
PHI:2096	PHI:2155		
PHI:2097	PHI:2158		
PHI:2098	PHI:216		
PHI:2099	PHI:2161		
PHI:210	PHI:2167		
PHI:2100	PHI:217		
PHI:2101	PHI:2171		
PHI:2103	PHI:2174		
PHI:2104	PHI:2175		
PHI:2105	PHI:2176		
PHI:2107	PHI:2177		
PHI:2109	PHI:2179		
PHI:211	PHI:2182		
PHI:2110	PHI:2183		

Table 4.8, continued.

PHI:2112	PHI:2186		
PHI:2113	PHI:2189		
PHI:2114	PHI:2191		
PHI:2117	PHI:2194		
PHI:2118	PHI:2195		
PHI:2119	PHI:2196		
PHI:2120	PHI:2197		
PHI:2121	PHI:220		
PHI:2128	PHI:2200		
PHI:2129	PHI:2201		
PHI:213	PHI:2203		
PHI:2130	PHI:2205		
PHI:2131	PHI:2215		
PHI:2133	PHI:222		
PHI:2134	PHI:2240		
PHI:2136	PHI:2244		
PHI:2137	PHI:2247		
PHI:2140	PHI:2248		
PHI:2141	PHI:2251		
PHI:2142	PHI:2255		
PHI:2147	PHI:2256		
PHI:2150	PHI:2257		
PHI:2152	PHI:2259		
PHI:2155	PHI:226		
PHI:2156	PHI:2260		
PHI:2158	PHI:2266		
PHI:2160	PHI:2267		
PHI:2161	PHI:2269		
PHI:2167	PHI:2270		
PHI:2168	PHI:2275		
PHI:2169	PHI:2279		
PHI:217	PHI:2290		
PHI:2170	PHI:2292		
PHI:2171	PHI:2293		
PHI:2172	PHI:2301		
PHI:2173	PHI:2302		
PHI:2174	PHI:2309		
PHI:2175	PHI:2315		
PHI:2176	PHI:2321		
PHI:2177	PHI:2322		
PHI:2178	PHI:2324		
PHI:2179	PHI:2328		
PHI:2180	PHI:2329		
PHI:2183	PHI:2334		
PHI:2184	PHI:2336		

Table 4.8, continued.

PHI:2185	PHI:2337		
PHI:2186	PHI:2339		
PHI:2187	PHI:2341		
PHI:2188	PHI:2342		
PHI:2189	PHI:235		
PHI:2190	PHI:2350		
PHI:2191	PHI:2351		
PHI:2192	PHI:2353		
PHI:2193	PHI:2354		
PHI:2194	PHI:2356		
PHI:2195	PHI:2357		
PHI:2196	PHI:2358		
PHI:2197	PHI:2359		
PHI:2198	PHI:2361		
PHI:2199	PHI:2368		
PHI:220	PHI:2370		
PHI:2200	PHI:2371		
PHI:2201	PHI:2375		
PHI:2202	PHI:2376		
PHI:2203	PHI:2377		
PHI:2205	PHI:2378		
PHI:2215	PHI:2379		
PHI:2216	PHI:2382		
PHI:2239	PHI:2383		
PHI:2240	PHI:2387		
PHI:2244	PHI:2393		
PHI:2247	PHI:2394		
PHI:2248	PHI:2396		
PHI:2251	PHI:24		
PHI:2255	PHI:2407		
PHI:2256	PHI:2419		
PHI:2257	PHI:242		
PHI:2259	PHI:243		
PHI:2266	PHI:244		
PHI:2267	PHI:2474		
PHI:2269	PHI:2487		
PHI:2270	PHI:2488		
PHI:2275	PHI:2491		
PHI:2279	PHI:2502		
PHI:2290	PHI:2504		
PHI:2292	PHI:2506		
PHI:2293	PHI:2510		
PHI:2296	PHI:2511		
PHI:2297	PHI:2512		
PHI:2299	PHI:2513		

Table 4.8, continued.

PHI:2300	PHI:2515		
PHI:2301	PHI:2517		
PHI:2302	PHI:2518		
PHI:2304	PHI:2519		
PHI:2305	PHI:2520		
PHI:2309	PHI:2522		
PHI:2315	PHI:2524		
PHI:2321	PHI:2525		
PHI:2322	PHI:2526		
PHI:2324	PHI:2528		
PHI:2329	PHI:2529		
PHI:2334	PHI:2530		
PHI:2336	PHI:2531		
PHI:2341	PHI:2532		
PHI:235	PHI:2533		
PHI:2351	PHI:2537		
PHI:2353	PHI:2538		
PHI:2354	PHI:2539		
PHI:2356	PHI:254		
PHI:2357	PHI:2540		
PHI:2376	PHI:2542		
PHI:2377	PHI:2543		
PHI:2378	PHI:2544		
PHI:2382	PHI:2545		
PHI:2383	PHI:2546		
PHI:2385	PHI:2547		
PHI:2387	PHI:255		
PHI:2388	PHI:2550		
PHI:2393	PHI:2553		
PHI:24	PHI:256		
PHI:2404	PHI:2560		
PHI:2405	PHI:2568		
PHI:2406	PHI:257		
PHI:2407	PHI:2570		
PHI:242	PHI:259		
PHI:2425	PHI:2597		
PHI:2428	PHI:26		
PHI:244	PHI:2600		
PHI:2441	PHI:2601		
PHI:2451	PHI:2602		
PHI:2452	PHI:2604		
PHI:2474	PHI:2605		
PHI:2476	PHI:2606		
PHI:2488	PHI:2608		
PHI:249	PHI:261		

Table 4.8, continued.

PHI:2491	PHI:2611		
PHI:2502	PHI:262		
PHI:2510	PHI:2625		
PHI:2513	PHI:2638		
PHI:2515	PHI:2639		
PHI:2517	PHI:2640		
PHI:2519	PHI:2643		
PHI:2520	PHI:2645		
PHI:2521	PHI:265		
PHI:2522	PHI:2653		
PHI:2524	PHI:2656		
PHI:2525	PHI:2668		
PHI:2526	PHI:267		
PHI:2528	PHI:270		
PHI:2529	PHI:2710		
PHI:2530	PHI:2714		
PHI:2531	PHI:2715		
PHI:2532	PHI:2728		
PHI:2533	PHI:273		
PHI:2537	PHI:2731		
PHI:2539	PHI:2735		
PHI:254	PHI:277		
PHI:2540	PHI:280		
PHI:2542	PHI:2802		
PHI:2543	PHI:2807		
PHI:2545	PHI:2808		
PHI:2546	PHI:281		
PHI:255	PHI:2817		
PHI:2550	PHI:2822		
PHI:2553	PHI:2838		
PHI:2558	PHI:2839		
PHI:256	PHI:284		
PHI:2560	PHI:2844		
PHI:2563	PHI:2847		
PHI:2568	PHI:286		
PHI:257	PHI:287		
PHI:2570	PHI:2884		
PHI:259	PHI:2895		
PHI:26	PHI:2896		
PHI:2600	PHI:2897		
PHI:2601	PHI:2901		
PHI:2604	PHI:2908		
PHI:2605	PHI:2911		
PHI:2609	PHI:2915		
PHI:2611	PHI:2916		

Table 4.8, continued.

PHI:262	PHI:2920		
PHI:2625	PHI:2927		
PHI:2638	PHI:2928		
PHI:2639	PHI:2959		
PHI:2640	PHI:2960		
PHI:2645	PHI:2961		
PHI:265	PHI:2962		
PHI:2651	PHI:2964		
PHI:2653	PHI:2968		
PHI:2656	PHI:2969		
PHI:267	PHI:2970		
PHI:269	PHI:2976		
PHI:2693	PHI:2978		
PHI:27	PHI:2981		
PHI:270	PHI:2989		
PHI:2700	PHI:3005		
PHI:2710	PHI:3011		
PHI:2712	PHI:304		
PHI:2714	PHI:305		
PHI:2728	PHI:31		
PHI:273	PHI:310		
PHI:2731	PHI:314		
PHI:277	PHI:323		
PHI:280	PHI:325		
PHI:2802	PHI:329		
PHI:2808	PHI:33		
PHI:281	PHI:330		
PHI:2819	PHI:335		
PHI:2821	PHI:336		
PHI:2822	PHI:337		
PHI:2829	PHI:339		
PHI:2834	PHI:352		
PHI:2837	PHI:358		
PHI:2838	PHI:361		
PHI:2839	PHI:367		
PHI:2844	PHI:384		
PHI:2849	PHI:387		
PHI:286	PHI:391		
PHI:2884	PHI:392		
PHI:2895	PHI:399		
PHI:2896	PHI:40		
PHI:2897	PHI:404		
PHI:2915	PHI:413		
PHI:2916	PHI:419		
PHI:2920	PHI:420		

Table 4.8, continued.

PHI:2921	PHI:423		
PHI:2924	PHI:424		
PHI:2926	PHI:435		
PHI:2927	PHI:436		
PHI:2928	PHI:438		
PHI:2930	PHI:440		
PHI:2940	PHI:441		
PHI:2959	PHI:442		
PHI:2960	PHI:443		
PHI:2961	PHI:445		
PHI:2962	PHI:447		
PHI:2964	PHI:454		
PHI:2968	PHI:455		
PHI:2969	PHI:465		
PHI:2970	PHI:469		
PHI:2976	PHI:480		
PHI:2978	PHI:482		
PHI:2981	PHI:485		
PHI:2983	PHI:486		
PHI:2985	PHI:487		
PHI:2986	PHI:489		
PHI:2987	PHI:491		
PHI:2988	PHI:492		
PHI:2989	PHI:494		
PHI:2990	PHI:496		
PHI:2991	PHI:501		
PHI:2992	PHI:502		
PHI:2994	PHI:503		
PHI:2995	PHI:504		
PHI:2996	PHI:505		
PHI:304	PHI:508		
PHI:305	PHI:510		
PHI:31	PHI:511		
PHI:311	PHI:512		
PHI:315	PHI:513		
PHI:317	PHI:519		
PHI:325	PHI:538		
PHI:33	PHI:541		
PHI:336	PHI:542		
PHI:337	PHI:543		
PHI:339	PHI:544		
PHI:346	PHI:545		
PHI:350	PHI:547		
PHI:355	PHI:55		
PHI:358	PHI:552		

Table 4.8, continued.

PHI:36	PHI:566		
PHI:361	PHI:573		
PHI:367	PHI:574		
PHI:386	PHI:58		
PHI:391	PHI:59		
PHI:393	PHI:598		
PHI:397	PHI:61		
PHI:399	PHI:616		
PHI:40	PHI:650		
PHI:401	PHI:651		
PHI:404	PHI:652		
PHI:405	PHI:668		
PHI:413	PHI:672		
PHI:419	PHI:68		
PHI:423	PHI:69		
PHI:424	PHI:693		
PHI:429	PHI:695		
PHI:433	PHI:697		
PHI:434	PHI:716		
PHI:435	PHI:747		
PHI:438	PHI:748		
PHI:440	PHI:777		
PHI:441	PHI:783		
PHI:442	PHI:784		
PHI:443	PHI:785		
PHI:445	PHI:789		
PHI:447	PHI:792		
PHI:454	PHI:796		
PHI:455	PHI:800		
PHI:465	PHI:803		
PHI:471	PHI:804		
PHI:474	PHI:806		
PHI:477	PHI:807		
PHI:479	PHI:811		
PHI:482	PHI:812		
PHI:485	PHI:815		
PHI:487	PHI:819		
PHI:489	PHI:822		
PHI:490	PHI:823		
PHI:491	PHI:825		
PHI:496	PHI:831		
PHI:502	PHI:837		
PHI:504	PHI:84		
PHI:505	PHI:854		
PHI:508	PHI:860		

Table 4.8, continued.

PHI:510	PHI:862		
PHI:511	PHI:875		
PHI:512	PHI:876		
PHI:513	PHI:877		
PHI:518	PHI:881		
PHI:519	PHI:882		
PHI:538	PHI:886		
PHI:541	PHI:887		
PHI:544	PHI:888		
PHI:547	PHI:893		
PHI:55	PHI:901		
PHI:551	PHI:903		
PHI:552	PHI:911		
PHI:566	PHI:922		
PHI:576	PHI:96		
PHI:578	PHI:97		
PHI:58			
PHI:594			
PHI:598			
PHI:668			
PHI:673			
PHI:68			
PHI:690			
PHI:693			
PHI:697			
PHI:698			
PHI:713			
PHI:714			
PHI:716			
PHI:72			
PHI:734			
PHI:748			
PHI:769			
PHI:777			
PHI:781			
PHI:784			
PHI:785			
PHI:789			
PHI:790			
PHI:792			
PHI:793			
PHI:794			
PHI:795			
PHI:796			
PHI:798			

Table 4.8, continued.

PHI:799			
PHI:800			
PHI:801			
PHI:803			
PHI:804			
PHI:806			
PHI:807			
PHI:809			
PHI:81			
PHI:811			
PHI:812			
PHI:813			
PHI:814			
PHI:815			
PHI:816			
PHI:817			
PHI:818			
PHI:819			
PHI:82			
PHI:820			
PHI:822			
PHI:823			
PHI:825			
PHI:83			
PHI:831			
PHI:84			
PHI:854			
PHI:858			
PHI:859			
PHI:860			
PHI:871			
PHI:874			
PHI:875			
PHI:877			
PHI:878			
PHI:879			
PHI:88			
PHI:880			
PHI:881			
PHI:882			
PHI:883			
PHI:885			
PHI:887			
PHI:888			
PHI:890			

Table 4.8, continued.

PHI:891			
PHI:893			
PHI:901			
PHI:911			
PHI:922			
PHI:923			
PHI:96			
PHI:97			
PHI:99			

Table 4.9: List of All Homologous Carbohydrate-Active Enzymes for all four Fungus.

M. oryzae	***B. cinerea***	***U. maydis***	***P. graminis***
AA10.hmm	AA1.hmm	AA10.hmm	AA1.hmm
AA1.hmm	AA2.hmm	AA1.hmm	AA2.hmm
AA2.hmm	AA3.hmm	AA2.hmm	AA3.hmm
AA3.hmm	AA4.hmm	AA3.hmm	AA5.hmm
AA4.hmm	AA5.hmm	AA4.hmm	AA6.hmm
AA5.hmm	AA6.hmm	AA5.hmm	AA7.hmm
AA6.hmm	AA7.hmm	AA6.hmm	AA9.hmm
AA7.hmm	AA8.hmm	AA7.hmm	CBM12.hmm
AA8.hmm	AA9.hmm	CBM13.hmm	CBM13.hmm
AA9.hmm	CBM13.hmm	CBM18.hmm	CBM20.hmm
CBM13.hmm	CBM18.hmm	CBM35.hmm	CBM21.hmm
CBM18.hmm	CBM1.hmm	CBM43.hmm	CBM32.hmm
CBM19.hmm	CBM20.hmm	CBM48.hmm	CBM43.hmm
CBM1.hmm	CBM21.hmm	CBM4.hmm	CBM48.hmm
CBM20.hmm	CBM24.hmm	CBM50.hmm	CBM63.hmm
CBM21.hmm	CBM32.hmm	CBM63.hmm	CBM67.hmm
CBM23.hmm	CBM35.hmm	CE10.hmm	CE10.hmm
CBM32.hmm	CBM37.hmm	CE13.hmm	CE12.hmm
CBM35.hmm	CBM42.hmm	CE14.hmm	CE14.hmm
CBM40.hmm	CBM43.hmm	CE1.hmm	CE16.hmm
CBM42.hmm	CBM46.hmm	CE4.hmm	CE1.hmm
CBM43.hmm	CBM48.hmm	CE5.hmm	CE4.hmm
CBM48.hmm	CBM50.hmm	CE8.hmm	CE5.hmm
CBM50.hmm	CBM51.hmm	CE9.hmm	CE7.hmm
CBM52.hmm	CBM66.hmm	GH105.hmm	CE8.hmm
CBM61.hmm	CBM67.hmm	GH109.hmm	GH105.hmm
CBM63.hmm	CE10.hmm	GH10.hmm	GH109.hmm
CBM66.hmm	CE12.hmm	GH115.hmm	GH10.hmm
CBM67.hmm	CE14.hmm	GH11.hmm	GH12.hmm
CBM6.hmm	CE16.hmm	GH125.hmm	GH131.hmm
CE10.hmm	CE1.hmm	GH128.hmm	GH13.hmm
CE12.hmm	CE2.hmm	GH13.hmm	GH15.hmm
CE14.hmm	CE3.hmm	GH15.hmm	GH16.hmm
CE15.hmm	CE4.hmm	GH16.hmm	GH17.hmm
CE16.hmm	CE5.hmm	GH17.hmm	GH18.hmm
CE1.hmm	CE7.hmm	GH18.hmm	GH20.hmm
CE2.hmm	CE8.hmm	GH20.hmm	GH23.hmm
CE3.hmm	CE9.hmm	GH23.hmm	GH26.hmm
CE4.hmm	GH105.hmm	GH25.hmm	GH27.hmm
CE5.hmm	GH106.hmm	GH26.hmm	GH28.hmm
CE8.hmm	GH109.hmm	GH27.hmm	GH2.hmm

Table 4.9, continued.

CE9.hmm	GH10.hmm	GH28.hmm	GH31.hmm
GH105.hmm	GH114.hmm	GH2.hmm	GH32.hmm
GH106.hmm	GH115.hmm	GH30.hmm	GH35.hmm
GH109.hmm	GH117.hmm	GH31.hmm	GH37.hmm
GH10.hmm	GH11.hmm	GH32.hmm	GH38.hmm
GH114.hmm	GH125.hmm	GH35.hmm	GH3.hmm
GH115.hmm	GH127.hmm	GH37.hmm	GH43.hmm
GH11.hmm	GH128.hmm	GH38.hmm	GH47.hmm
GH125.hmm	GH12.hmm	GH3.hmm	GH5.hmm
GH127.hmm	GH131.hmm	GH42.hmm	GH63.hmm
GH128.hmm	GH132.hmm	GH43.hmm	GH65.hmm
GH12.hmm	GH13.hmm	GH45.hmm	GH71.hmm
GH131.hmm	GH15.hmm	GH47.hmm	GH72.hmm
GH132.hmm	GH16.hmm	GH51.hmm	GH74.hmm
GH13.hmm	GH17.hmm	GH55.hmm	GH76.hmm
GH15.hmm	GH18.hmm	GH5.hmm	GH79.hmm
GH16.hmm	GH1.hmm	GH62.hmm	GH7.hmm
GH17.hmm	GH20.hmm	GH63.hmm	GH81.hmm
GH18.hmm	GH23.hmm	GH72.hmm	GH85.hmm
GH1.hmm	GH25.hmm	GH74.hmm	GT10.hmm
GH20.hmm	GH26.hmm	GH76.hmm	GT15.hmm
GH27.hmm	GH27.hmm	GH79.hmm	GT1.hmm
GH28.hmm	GH28.hmm	GH85.hmm	GT20.hmm
GH29.hmm	GH2.hmm	GH8.hmm	GT21.hmm
GH2.hmm	GH31.hmm	GH92.hmm	GT22.hmm
GH30.hmm	GH32.hmm	GH9.hmm	GT24.hmm
GH31.hmm	GH35.hmm	GT15.hmm	GT25.hmm
GH32.hmm	GH37.hmm	GT17.hmm	GT26.hmm
GH33.hmm	GH38.hmm	GT1.hmm	GT2.hmm
GH37.hmm	GH3.hmm	GT20.hmm	GT31.hmm
GH38.hmm	GH43.hmm	GT21.hmm	GT32.hmm
GH39.hmm	GH45.hmm	GT22.hmm	GT33.hmm
GH3.hmm	GH47.hmm	GT24.hmm	GT39.hmm
GH43.hmm	GH51.hmm	GT28.hmm	GT3.hmm
GH45.hmm	GH53.hmm	GT2.hmm	GT43.hmm
GH47.hmm	GH54.hmm	GT31.hmm	GT44.hmm
GH51.hmm	GH55.hmm	GT32.hmm	GT48.hmm
GH53.hmm	GH5.hmm	GT33.hmm	GT49.hmm
GH54.hmm	GH62.hmm	GT39.hmm	GT4.hmm
GH55.hmm	GH63.hmm	GT3.hmm	GT50.hmm
GH5.hmm	GH64.hmm	GT48.hmm	GT57.hmm
GH62.hmm	GH65.hmm	GT4.hmm	GT58.hmm
GH63.hmm	GH6.hmm	GT50.hmm	GT59.hmm
GH64.hmm	GH71.hmm	GT57.hmm	GT5.hmm
GH67.hmm	GH72.hmm	GT58.hmm	GT66.hmm
GH6.hmm	GH74.hmm	GT59.hmm	GT68.hmm

Table 4.9, continued.

GH71.hmm	GH76.hmm	GT60.hmm	GT69.hmm
GH72.hmm	GH78.hmm	GT66.hmm	GT71.hmm
GH74.hmm	GH79.hmm	GT69.hmm	GT76.hmm
GH75.hmm	GH7.hmm	GT71.hmm	GT8.hmm
GH76.hmm	GH81.hmm	GT76.hmm	GT90.hmm
GH78.hmm	GH88.hmm	GT8.hmm	GT93.hmm
GH79.hmm	GH89.hmm	GT90.hmm	PL14.hmm
GH7.hmm	GH92.hmm	PL12.hmm	PL15.hmm
GH81.hmm	GH93.hmm	PL1.hmm	PL1.hmm
GH88.hmm	GH95.hmm	PL22.hmm	PL20.hmm
GH92.hmm	GT15.hmm		PL21.hmm
GH93.hmm	GT17.hmm		
GH94.hmm	GT1.hmm		
GH95.hmm	GT20.hmm		
GT15.hmm	GT21.hmm		
GT17.hmm	GT22.hmm		
GT1.hmm	GT24.hmm		
GT20.hmm	GT25.hmm		
GT21.hmm	GT26.hmm		
GT22.hmm	GT28.hmm		
GT24.hmm	GT2.hmm		
GT25.hmm	GT31.hmm		
GT28.hmm	GT32.hmm		
GT2.hmm	GT33.hmm		
GT31.hmm	GT34.hmm		
GT32.hmm	GT39.hmm		
GT33.hmm	GT3.hmm		
GT34.hmm	GT48.hmm		
GT35.hmm	GT4.hmm		
GT39.hmm	GT50.hmm		
GT3.hmm	GT57.hmm		
GT41.hmm	GT59.hmm		
GT48.hmm	GT5.hmm		
GT4.hmm	GT62.hmm		
GT50.hmm	GT65.hmm		
GT57.hmm	GT66.hmm		
GT58.hmm	GT68.hmm		
GT59.hmm	GT69.hmm		
GT5.hmm	GT71.hmm		
GT62.hmm	GT76.hmm		
GT66.hmm	GT8.hmm		
GT69.hmm	GT90.hmm		
GT71.hmm	GT92.hmm		
GT76.hmm	PL1.hmm		
GT8.hmm	PL22.hmm		
GT90.hmm	PL3.hmm		

Table 4.9, continued.

PL1.hmm	PL7.hmm		
PL20.hmm			
PL3.hmm			
PL4.hmm			