ANALYSIS OF SECRETED PROTEINS OF *MAGNAPORTHE GRISEA* AND THE SEARCH FOR PROTEIN EFFECTORS

A Thesis

by

YUE SHANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2007

Major Subject: Plant Pathology

ANALYSIS OF SECRETED PROTEINS OF *MAGNAPORTHE GRISEA* AND THE SEARCH FOR PROTEIN EFFECTORS

A Thesis

by

YUE SHANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee, Daniel J. Ebbole
Committee Members, Charles M. Kenerley
Herman B. Scholthof
Wayne K. Versaw
Head of Department, Dennis C. Gross

May 2007

Major Subject: Plant Pathology

ABSTRACT

Analysis of Secreted Proteins of *Magnaporthe grisea* and the Search for Protein Effectors.

(May 2007)

Yue Shang, B.S., Lanzhou University, China

Chair of Advisory Committee: Dr. Daniel Ebbole

Magnaporthe grisea is a notorious pathogenic fungus that causes rice blast disease worldwide. Proteins secreted by the fungus are likely candidates for being effectors that are potentially recognized by determinants of resistance or susceptibility in host plants. However, knowledge of the role of secreted proteins of *M. grisea* is still limited. In this study, I identified 29 proteins that were secreted into culture filtrates from *M. grisea* strains expressing candidate proteins. I confirmed secretion of these proteins and tested them for elicitor activity on plants. Among them, I studied two groups: cell wall degrading enzymes (CWDEs) and small cysteine-rich proteins. Cysteine-rich proteins have been shown in other systems to function as elicitors. Initially, I expressed and purified proteins in *M. grisea* to obtain proteins by a homologous expression system. Although this was effective for a number of proteins, the need for greater amounts of protein led me to express several proteins in the *Pichia pastoris* system. Several candidate proteins were purified and found to induce symptoms on rice and maize. Hypothetical proteins MG10424.4 and MG09998.4 were both found to have elicitor activity. Lipase MG07016.4 did not induce response of plants and we concluded that the lipase activity of MG07016.4 does not function as an elicitor. I also purified a small cysteine-rich protein, which belongs to the group of cluster 180 proteins in *M. grisea,* MG10732.4 from *P.*

pastoris. It is able to cause yellowing symptoms and hydrogen peroxide production in plants and it might contain elicitor activity.

DEDICATION

I dedicate this thesis to my parents, Ying Ma and Delong Shang, who supported me through the challenges of graduate life.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Daniel Ebbole, who has been an outstanding both mentor and advisor. He taught me a lot about molecular biology and genetics during my graduate career at Texas A&M University. I really appreciate his support and help. I also thank my other committee members, Dr. Charles Kenerley, Dr. Herman B. Scholthof and Dr. Wayne Versaw for their support.

This thesis represents work that involved collaborations with other former members of the lab including Dr. Guodong Lu, Dr. Hanno Wolf, Dr. Cristina Flippi, Dr. Dan Li and current lab member Mr. Kiran Bhattarai**.** I also appreciate the helpful conversations and assistance from Mr. Dong Qi and Dr. Rustem Omarov.

Several members of the Department of Plant Pathology and Microbiology helped me during my time at Texas A&M. Dr. Jim Starr and Dr. Mike Kolomiets taught me a large amount of knowledge about plant pathology.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Rice blast disease, caused by the fungus *Magnaporthe grisea*, is one of the most serious diseases of cultivated rice throughout the world (Nicholas 2003) and can cause up to 30% crop loss (Wang et al. 2005). To develop improved methods for disease control, a better understanding of the host-pathogen interaction is needed. An exchange of molecular signals from both the fungus and plant sides is involved in plant defense responses. Secreted proteins of *Magnaporthe*, by virtue of their being present outside of fungus, are the most likely candidates for being effectors that potentially induce disease or defense responses in plants. I hypothesize that secreted proteins of *Magnaporthe grisea* might function as effectors of plant responses. To test this hypothesis, I purified several of these secreted proteins and tested them for their activity on plants.

Fungal pathogens establish intimate associations with their plant host. To invade the plant and complete their life cycle, specific proteins, known as effectors (Kamoun 2006), play an important role. Effectors can be thought of as suppressors of plant defense response. A simple way to think about this is that effectors may act as toxins to poison the cell (or inhibit the function of a specific target) to inhibit the defense response. Plants have evolved mechanisms to recognize the pathogen. Recognition of the effectors is an obvious way to recognize a pathogen, and there are now several examples of effector molecules that can be recognized by the plant to trigger resistance (Kamoun 2006). When the ability to recognize the pathogen varies within the host species this is

---

This thesis follows the style and format of Molecular Plant Microbe Interactions.

characteristic of the gene-for-gene resistance phenomenon (Keen 1990). Plants are capable of recognizing other proteins that are not effectors. For example, flagellin protein of *Pseudomonas syringe* can be recognized by *Arabidopsis* to trigger a defense response (Zipfel et al. 2004). This form of resistance, mediated by the recognition of Pathogen Associated Molecular Patterns (PAMPs) is one basis of innate immunity associated with non-host resistance (Parker 2003). At the molecular level, this appears to be very similar to the gene-for-gene mechanism of resistance-gene mediated recognition of specific pathogen molecules (Zipfel et al. 2004).

We can classify effectors into two groups according to the different location of target sites in plants: extracellular effectors, which are secreted by fungi into the plant extracellular space and function extracellularly; and cytoplasmic effectors, which are located within the plant cells. The *Avr pita* gene of *Magnaporthe* is a good example of cytoplasmic effector (Jia et al. 2000), but the mechanism used by this fungus to deliver the protein into plant cells is still unknown. In other fungi and oomycetes, some of the extracellular effectors have been well studied. For example, the *Avr4* gene product from *Cladosporium fulvum* can induce the HR response in tomato. It is also a cysteine–rich chitin-binding protein that has anti-chitinase activity by virtue of its ability to bind to fungal chitin and shield it from being degraded (van den Burg et al. 2004). A direct interaction has been shown between the product of the *NIP1* gene from *Rhynchosporium secalis* and the corresponding R gene product in barley **(**van't Slot et al. 2003**)**. The protein encoded by *NIP1* is an 82-amino acid protein that has a 22-amino acid signal peptide. Cleavage of the signal peptide yields a 60-amino acid mature protein that contains 10 cysteine residues that form five intramolecular disulfide bonds. These

proteins are presumably able to recognize and interact with the extracellular target in the host plant. NIP1 is known to activate the plasma membrane ATPase and this may be its target as a virulence factor (van't Slot et al. 2003).

With the help of genomic and bioinformatic methods, a series of predicted secreted proteins with various functional domains have been identified in *Magnaporthe*. These include small cysteine-rich proteins and proteins with cellulose/chitin binding domains, and homologs of cell wall degrading enzyme (CWDE). I examined representative proteins by focusing on three classes: cell wall degrading enzymes, several hypothetical proteins, and a family of small cysteine-rich proteins unique to *M. grisea* to help understand the roles of these secreted proteins in *Magnaporthe*.

The best-studied cellulose binding protein was found in *Phytophthora parasitica var. nicotianae*. CBEL (cellulose binding, elicitor, and lectin-like), a 34-kDa cell wall glycoprotein in *Phytophthora parasitica* binds to cellulosic substrates and elicits necrosis and defense responses in tobacco (Kamoun 2006). Immunogold-labelling showed that this glycoprotein was localized to the external and internal layers of the hyphal cell wall (Gaulin et al. 2002). CBDs (cellulose binding domains) are believed to enhance the efficiency of hydrolysis notably by attaching the enzymes to their substrate (Gilkes et al. 1991). Another good example of CBDs is *Avr4* gene in *Cladosporium fulvum* as mentioned above, which has chitin binding activity. It protects fungal chitin from degradation by binding and shielding it from the plant chitinases. In *M. grisea*, 15 genes encoding proteins with cellulose/chitin-binding motifs were identified. These genes await future characterization.

The plant cell wall is an important barrier to invasion. It contains polymers of sugars that can serve as a carbon source for an invading pathogen. Plant pathogenic fungi make a variety of enzymes that can degrade the polymers of the plant cell wall. Cellulases represent a large group of CWDEs. Many fungi use them to degrade plant cell wall polysaccharides (Ng 2004). Endoglucanses (endo-1-4-β-glucanase), exo-1,4-β-glucanase and β-glucosidase are the three major types of celluloytic enzymes. They hydrolyze 1,4-β bonds along the interior of the cellulose chain, cleave cellobiosyl units from the non-reducing ends of the cellulose chains and cleave glucosyl units from non-reducing ends of cello-oligosaccharides, respectively (Ng 2004). Of the CWDEs in *Magnaporthe,* we have identified that cellulases represent the largest group.

Another important group of CWDEs is pectinases. They are the only CWDEs capable of tissue maceration by disrupting the middle lamella in plants. Endopolygalacturonase and exopolygalacturonase are two pectinases that degrade the galacturonan backbone of pectin molecules (Cooper 1983). In the oomycete, *Phytophthora cinnamomi,* a polygalacturonase gene family has been characterized and this analysis demonstrated that degradation of pectin in the plant cell wall plays a major role in tissue invasion and maceration (Gotesson et al. 2002). *Botrytis cinerea*, an opportunistic plant pathogen, is able to weaken plant cell walls by producing various pectinases, including exo- and endopolygalacturonases, pectin methylesterases, and pectin and pectate lyases to hydrolyze pectin. The best known one is the endopolygalacturonase-encoding (*Bcpg*) family which contains at least six Bcpg genes. Five of these genes have been purified from *Pichia pastoris* and tested for biological activity (Kars et al. 2005).

Xylanases such as XYN22 and XYN33 of *M.grisea* (Wu et al. 1995) have been purified, cloned and characterized, and they are expressed when *M. grisea* is grown on rice cell walls or on oatspelt xylan, but not when grown on sucrose. These enzymes attack the side chain of hemicellulose fibrils to release oligosaccharides. Oligosaccharides released from plant cell can serve as endogenous elicitors of plant defense. In fact, a xylanase from *M. griesa* was shown to induce defense reactions when applied to rice plants.. A xylanase from *Trichoderma* spp. was also found to act as an elicitor when applied to plant leaves. However, a site-directed mutant that inactivated enzyme activity retained its elicitor activity (Enkerli et al. 1999). This suggests the protein itself was recognized by the plant to trigger the plant response. Thus, CWDEs can generate cell wall fragments to induce defense reactions or act as PAMPs. I tested several CWDEs to determine if they acted as elicitors towards rice (Chapter II).

Several fungal genes have been identified (Table 1) that encode small (<150 amino acids) secreted proteins with an even number of cysteine residues. Several of these have been found to induce defense responses when infiltrated into plants (Lauge and de Wit 1998, van't Slot 2002). *Cladosporium fulvum Avr2, Avr4*, and *Avr9*, *ecp1*, *ecp2*, *Rhychosporium secalis nip1*, and *Phytophthora* elicitins are well-studied examples of this type of cysteine-rich protein genes. The disulfide bridges increase the stability of the protein in the plant intercellular spaces that are rich in proteinases (Joosten et al. 1997, Kamoun et al. 1999, Kooman-Gersmann et al. 1997, Luderer et al. 2002). The cysteine residues, by virtue of their ability to stabilize protein structure, are often found in enzyme inhibitors, for example in the Kazal proteinase inhibitor domain (Laskowski and Qasim 2000, Tian M 2005). We have identified a gene family in *Magnaporthe*, which contains

at least fourteen genes unique to *M. grisea*. I hypothesize that these cysteine-rich proteins are inhibitors of rice plant enzymes or act as elicitors (Chapters IV and V). In order to prove our hypothesis, I expressed three of the genes in *Pichia* expression system and test one of them on plants.

Table 1. Small cysteine-rich proteins that act as virulence and/or avirulence factors

Pathogen species	Pathogen agent	Description	Reference
Cladosporium fulvum	*Avr2*	Small cysteine-rich protein	Joosten et al., 1997
	Avr4	Cysteine-rich chitin binding protein	Joosten, M.H *et al*., 1997
	Avr9	GATA-type transcriptional regulators binding protein	Kooman-Gersmann *et al*., 1997
	ecp1	Extracellular protein (virulence factor)	Luderer R *et al*., 2002
	ecp2	Extracellular protein (virulence factor)	Luderer R *et al* ., 2002
Rhychosporium secalis	*nip1*	Activates plant plasma membrane ATPase	van't Slot KA *et al*., 2003
Phytophthora	elicitin	Hypersensitive response inducing protein	Gotesson A *et al*., 2002

CHAPTER II

PROTEIN OVEREXPRESSION IN *MAGNAPORTHE GRISEA* AND ACTIVITY TEST ON PLANTS

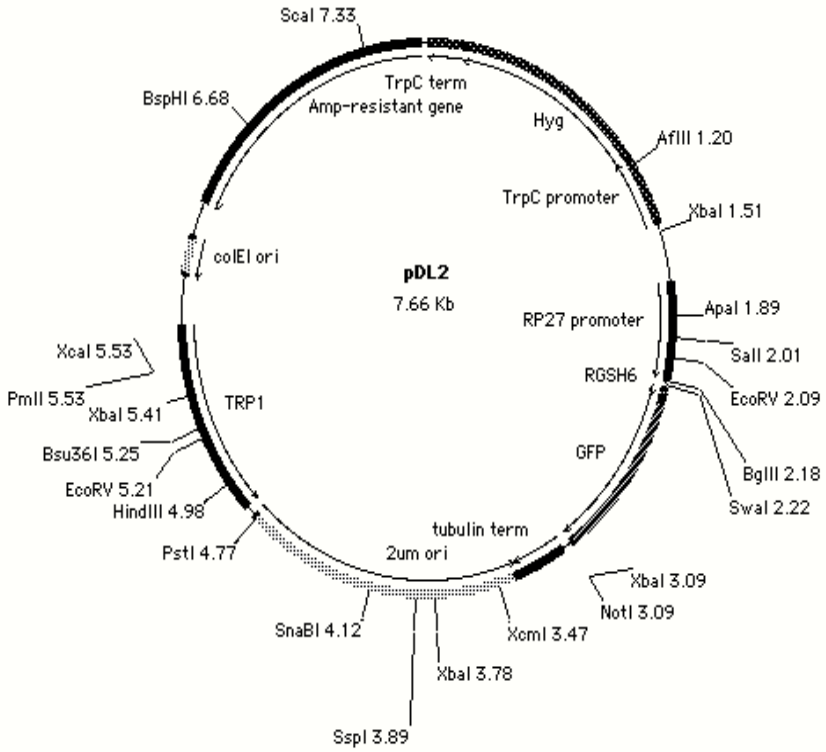
Introduction

A long term goal is to identify all *M. grisea* proteins that can serve as plant effectors as elicitors or as suppressors of the plant defense response. Toward this goal, a set of putative secreted proteins were defined by bioinformatic analysis of predicted genes in the *M. grisea* genome (Dean. et al 2005). Of the approximately 750 predicted secreted proteins, 300 were selected for cloning by amplification from *M. grisea* strain 70-15 with specific primers. The primers contained flanking sequences that shared homology with a vector (pDL1) designed for expression of the cloned genes in filamentous fungi (Lu et al. unpublished data). The 5' primer was designed to include 18 nucleotides homologous to promoter sequence in the vector including the ATG plus an additional 20 nucleotides matching the gene. The 3'primer contained 18 nucleotides matching the sequence RGSHHH codon in the vector (see below) with 20 nucleotides matching the gene sequence starting at the final codon of the coding region.

An oligonucleotide encoding the peptide sequence RGSHHHHHH ($RGSH_6$) tag was attached to the pTE11 vector to construct the pDL1 vector (Fig.1). Coding regions were cloned in-frame with the start codon in the vector and the C-terminal $RGSH_6$ tag. The vector contained the *M.grisea* RP27 (ribosomal protein 27) promoter to drive expression of the transgene in filamentous fungi. The vector was linearized and the vector and amplified products were co-transformed into *Saccharomyces cerevisiae* to reconstitute the circular plasmid by in vivo homologous recombination. A crude DNA

preparation from yeast was used to transform *Escherichia coli* to amplicillin resistance. *E. coli* colonies were screened by amplification with vector-specific primers to identify colonies containing appropriate clones. High fidelity thermostable polymerase was used for the amplification to minimize misincorporation of nucleotides that might lead to inactive alleles of the genes. To test this, 96 clones were sequenced and no mutations were found in the >50,000 high quality nucleotides reported (Lu et al. unpublished data)

I hypothesized that some fraction of proteins would be found to induce a visible symptom when exposed to plant tissue because they i) act as elicitors directly ii) they act as PAMPs or iii) they function to alter plant physiology (virulence factor). An important question is to address what fraction of pathogen secreted proteins are able to induce symptoms. I tested 29 genes to determine if they were able to produce secreted proteins as predicted. Large-scale preparations of 19 of these proteins from *M. grisea* culture filtrates were prepared and tested for their ability to induce plant responses (Table 2). I found several proteins that induced weak responses in rice plants.



Plasmid Name: pDL2
Plasmid size: 7.66 kb
Constructed by: Dan Li
Construction date: unknown

Figure 1. Map of vector pDL1. RP 27 promoter is to drive gene expression in filamentous fungi. SwaI site is where the vector was linearized before being transformed to *Saccharomyces cerevisiae.*

Table 2. Nineteen proteins purified from *M. grisea*

Protein #	Predicted vs exp size	Path: Prot: I: C	Biological activity
MG01247.4	Match	9S, 3S : N : N : D	endochitinase precursor
MG07901.4	Match	9HS, 9HS: N	hypothetical protein
MG08424.4	Match	NT: N	endo-beta-1,4-D-xylanase
MG10424.4	Match	5/9HS, 5/9HS, 9HS:	Hypothetical protein
MG09726.4	Match	9HS, 9HS : HR	Fungal Beta-1,4-Galactanases
MG07715.4	Match	9HS, 9HS : N	predicted protein
MG05344.4	Match	9HS, 9HS, 9HS : N	probable SnodProt1 PRECURSOR
MG00311.4	Two bands	5HS, 5HS, 7HS : N	acid protease
MG03746.4	Match	9HS,9HS,9HS : N	acetyl xylan esterase
MG05232.4	Match	7HS, 7HS : N; N; D	IgE-binding protein
MG00994.4	Match	9HS, 9HS: HR: N; D	mannosyl-oligosaccharide 1,2-alpha-mannosidase
MG06538.4	Match	9HS, 9HS, 9HS : N	hypothetical protein
MG08054.4	Match	7/9HS, 3S/9HS, 3S/3HS :HR: N:N	extracellular chitinase
MG00582.4	Match	3HS, 3HS, 5HS : N	endoglucanase C
MG09998.4	Match	7-9HS, 9HS, 7HS : N	Hypothetical protein
MG07965.4	Match	3S/5HS, 3-5S/3HS, 3S: HR : N : N	alkaline proteinase
MG01403.4	Match	9HS, 9HS, 7HS: HR: N: N	ferulic acid esterase A
MG07303.4	Match	3S, 7HS : N	predicted protein
MG00269.4	Match	NT : N	predicted protein

Table 2 footnotes:
Pathogenicity ratings: 1 = 0-5 % diseased leaf area; 3 = 5-25%; 5 = 25-50%; 7 = 50-75%; 9 = 75 – 100% relative to 70-15 control. HS = hypersensitive type lesion. S = susceptible lesion.
Leaf assay against purified protein (Prot). N = no response; HR = induced necrosis/browning.
Infection assay with incompatible 4091-5-8 strain co-inoculated with purified protein (I). N = no infection.
Infection assay with compatible 70-15 co-inoculated with purified protein (C). N = no infection; D = disease.

Results and Discussion

90 proteins were detected to be secreted proteins in *Magnaporthe grisea*. Previous work done by Drs. Guodong Lu and Hannon Wolf found 61 proteins to be secreted proteins. I found another 29 secreted proteins by screening additional transformed lines of *M. grisea*. Figure 2 is an example of SDS-PAGE Coomassie blue staining and western blot detection with $RGSH_6$ antibody.

Symptoms that the plants respond to the proteins after being tested repeatedly were not consistent. To determine for effector activity of proteins purified from *M. grisea*, we tested the purified proteins on rice leaf segments ('Materials and Methods'). Hydrogen peroxide detection was done to determine plant responses to the proteins. The experiments were repeated several times (>2) with different preparations of proteins. In these experiments, we used 1 mM Tris buffer as a negative control and crude elicitor extracted from *Magnaporthe grisea* was used as positive control, however, rice plants did not always show symptoms each time with the positive control and/or Tris buffer produced as much H_2O_2 as leaf segments treated with proteins or elicitor. Figure 3 shows rice leaf segments treated with two different secreted proteins: MG09726.4 (**A**) and MG08054.4 (**B**). Small lesions were observed similar to the elicitor treated leaf segment **(C)** in contrast to the negative control that displayed no necrosis/browning reaction **(D)**. The purification from *M. grisea* was found to work well, however, the low yield required several different preparations of protein to obtain the required amount. Additionally, the inconsistency of the results suggested that different batches of protein had different levels of activity. I also could not exclude the possibility that some batches of protein might be contaminated with *M. grisea* elicitors unrelated to the purified protein. Another

interpretation is that the elicitors being used had relatively weak elicitor activity and perhaps are not of sufficient biological relevance. This led me to seek a better positive control elicitor and an alternative expression system to *M. grisea*. These studies are detailed in Chapters III and IV.

Materials and Methods

Pipeline for the detection of 29 secreted proteins from *M.grisea* transformants. Three individual transformants of *M. grisea* for each gene were inoculated into 24-well plates with complete medium (Talbot et al. 1993). The plates were incubated at 25° for 4 days to obtain mycelia pads, and the mycelia pads were then transferred to fresh CM medium containing 50 ul Ni-NTA agarose (Qiagen) in new 24-well plates. The plates were incubated at 25° with gentle shaking for another 4 days. For secreted protein detection, the culture filtrate with Ni-NTA was added to 1.5 ml tubes, filtrate was centrifuged for a few seconds, the supernatant was removed, 40 ul of elution buffer was added to each tube to resuspend the agrose, and then 8 ul of protein loading buffer was added. The protein samples were incubated at 100° for 5 min to denature the protein and loaded into a 12-15% SDS-polyacrylamide gel for analysis. After the electrophoresis, protein was transferred to PVDF membrane and detected using RGS-His tag antibody following the supplied protocols (Qiagen).

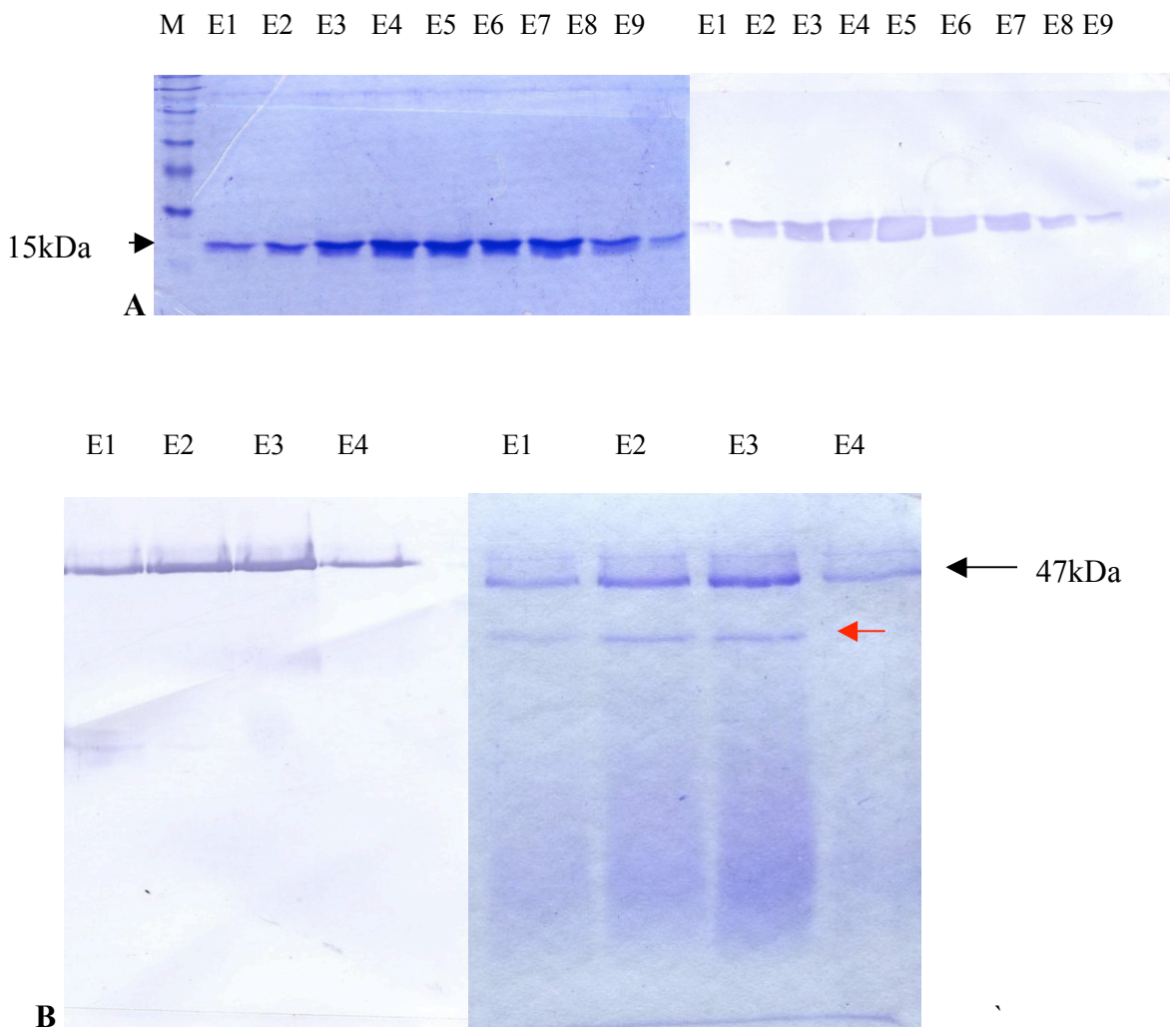


Figure 2. Examples of purified proteins from *M. grisea*. **A,** MG10424.4, the column was eluted 9 times; E1-E9 stands for the Elution fraction 1-Elution fraction 9. The elution fraction was detected by coomassie blue staining (left) and western blot (right). Protein started to come out from the first elution fraction and the size is about 15kDa. **B,** MG03746.4, column was eluted 4 times and detected by coomassie blue staining (left) and western blotting detected with RGS-His_6 antibodies (right). The smaller band (red arrow) might indicate that other unspecific protein was eluted out of the column but the absence of the band on the western blot membrane ruled out the possibility that this small protein form has His_6 tag.

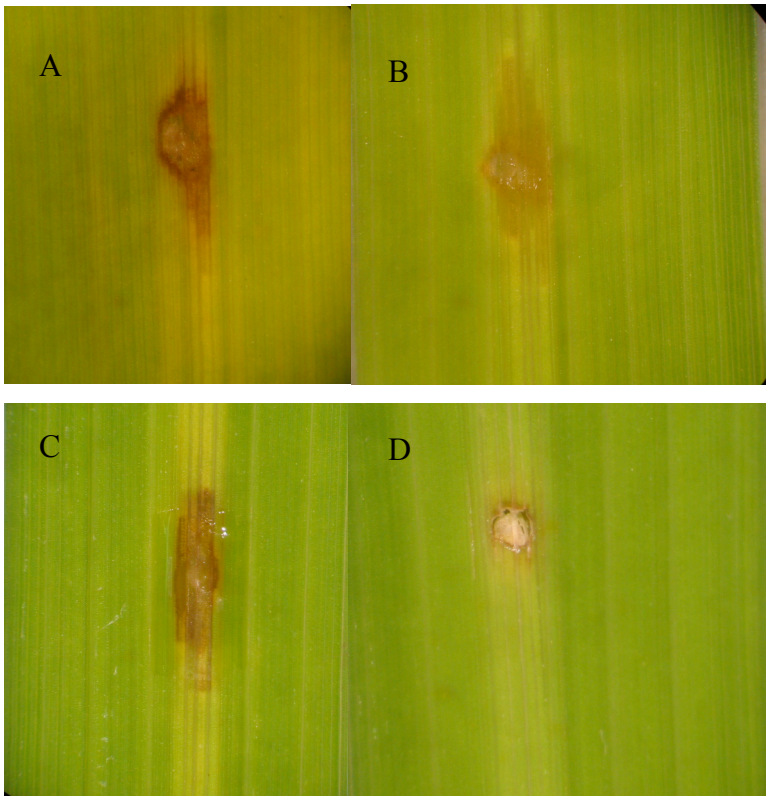


Figure 3. Activity test of proteins purified from *M. grisea* on rice leaf segments. Brown lesions on the leaf segment inoculated with protein MG09726.4 (**A**) and MG08054.4 (**B**). Elicitor purified from *Magnaporthe* caused brown lesion too, it worked as positive control(**C**); 1mM Tris buffer was used as negative control (**D**).

Large-scale preparation of 20 secreted proteins from *M. grisea*. The transformed strains and 70-15 control strain were pre-cultured in Petri dishes with 10 ml complete medium for 4-5 days. The mycelia was blended and transferred to 500 ml complete medium and incubated for 4 days at 25° with gentle shaking. The culture filtrate was filtered through filter paper to a new flask, and then transferred to a 500 ml separatory funnel followed by the addition of 50 ml starting buffer, 50 ml glycerol and 4 ml Ni-NTA. The solution was incubated at least 4 hours at 4° with occasional shaking. The Ni-NTA was allowed to settle and drained into a chromatography column. The column was washed twice with 4 ml washing buffer. The protein was eluted off the column with 1 ml elution buffer 9 times.

Desalting of purified protein. Salt in the purified protein was removed by Amicon Ultra Centrifugal Filter Devices (Millipore Corporation Bedford, MA, USA) with 1 mM Tris-HCl (pH 7.5) buffer as the exchange buffer. Alternatively, proteins were desalted by dialysis using 500 molecular weight cut-off Spectropor dialysis membrane (Spectrum Laboratories Inc., Rancho Dominguez, CA).

Plants. Four week old rice cultivar M202, susceptible to *M. grisea* 70-15 strain was used in this study to test protein activity. Maize plants were provided by Dr. Kolomiets' lab.

Elicitor extraction from *Magnaporthe grisea.* See protocol for production of *M. grisea* elicitor in Chapter III.

Activity test on plants. Two assays were performed for protein activity test: detached leaf assay and hydrogen peroxide assay.

1. Detached leaf assay

Leaves were detached from the plant by cutting with scissors or slicing with a razor blade. Detached leaves were wounded by piercing with a 21-gauge needle and then purified proteins were applied to the wound site; usually the detached leaves were incubated with 100% humidity for 72 hours under constant light at room temperature. Symptoms observed with test proteins were compared with the positive control, crude elicitor from *M. grisea* culture filtrates (Matsumura et al. 2003) and the negative control, 1 mM Tris-HCl buffer or 1 mg/ml bovine serum albumin (BSA) to assess whether the applied proteins had elicitor activity. Alternatively, protein and control samples were applied to the cut end of leaf segments as a method of applying protein to the plant.

2. Hydrogen peroxide assay

The detached leaves with applied protein were placed in water with 0.01% Triton-X-100 and DAB (3,3'-diaminobenzidine) at 1 mg/ml after incubation for 10 hours. The solution was then infiltrated with low vacuum pressure for 30 min and then incubated in the dark overnight. Leaves were fixed and cleared in alcoholic lacto-phenol at 65° for 30 min, rinsed with 50 % ethanol and finally rinsed with water. To visualize staining, whole leaf sections were incubated in 70% glycerol then mounted on slides.

CHAPTER III

DISCOVERY AND PRODUCTION OF A HIGH-ACTIVITY ELICITOR TO SERVE AS POSITIVE CONTROL IN PLANT ASSAYS

Introduction

To test the activity of purified proteins, an effective positive control is needed. According to Matsumura et al (2003), crude protein from *M. grisea* works well as an elicitor to induce resistance responses in rice plants. During the purification of proteins in the *Pichia pastoris* system, I also found a small molecular weight activity with a size between 500 Da and 5kDa that is able to induce strong watersoaking symptoms on rice and maize leaves. Therefore, crude protein extracts from *M. grisea* and the small molecule from *Pichia pastrois* were both used as positive control to treat plant. In this chapter I describe the protocol for production and testing of elicitor-active fractions from both fungi.

Results and Discussion

Activity of elicitor extracted from *M. grisea*. The elicitor activity from *M. grisea* produces a browning (see Chapter II, Fig. 3) and hydrogen peroxide production. This elicitor is useful, however, in some cases this positive control produces only very weak symptoms that are difficult to distinguish from negative controls. In addition, the concentration of material (100 mg/ml) raises concerns about the physical effects of such high solute concentrations on the plant's response. In addition, the complexity of constituents of the elicitor fraction raises concerns about batch-to-batch variation in elicitor activity. Another elicitor active material would be beneficial.

Activity of elicitor molecules extracted from *Pichia pastoris.* The PEF1, PEF3 and PEF4 elicitor fractions were used to treat maize leaf segments. Watersoaking symptoms were observed 20-36 h after treating leaf segments with PEF1 and PEF3, but no symptoms were observed using PEF4 (Fig. 4). These observations suggest that some small molecules from *P. pastoris* act as elicitors to induce responses in plants. This experiment was repeated three times, watersoaking symptom occurred each time with the PEF1 and PEF3 fractions. The lack of activity of PEF4 is critical, since the 5 kDa centrifugation step is used to purify proteins away from the *Pichia* elicitor and remove other small molecule contaminants. The PEF1 and PEF3 fractions should serve as excellent positive controls for elicitor activity.

Further investigation of the elicitor-active molecule would be useful. It is interesting that this elicitor was retained on the Ni-NTA column and eluted with 250 mM imidazole. Concentrated crude culture filtrate should also contain this activity and it would be interesting to know how this activity is retained by Ni-NTA. Possibly, molecular characterization would reveal if it is composed of a signal molecular species that could be obtained in a more simple way, such as chemical synthesis.

In addition to crude elicitors, other molecules have been shown to act as elicitors in rice and other plants. For example, chitin oligomers can serve as elicitors (van den Burg HA, 2004). However, the concentration of these oligomers required for activity is relatively high (~1 mg/ml). The *Pichia* elicitor was not detected by protein gel electrophoresis, and although it could be a peptide it is not likely to be a protein of ~ 5 kDa. Additional testing of the *Pichia* elicitor could include determining if heat or proteases inactivate elicitor activity. HPLC analysis would be useful in determining the

complexity of the material. If the material is relatively pure, Mass Spectrometry analysis and NMR might allow determination of its molecular structure.

Materials and Methods

Elicitor extraction from *M. grisea*. I followed the protocol of Matsumura (2003). Rice blast fungus, *M. grisea* 70-15 was cultured in 500 ml of medium containing potato extract, 20 gl^{-1} sucrose and 5 gl^{-1} yeast extract (Koga et al. 1998). Fungal mycelia were harvested and resuspended in 40 ml of 20 mM potassium phosphate buffer containing 0.1% Tween 20. Buffer-suspended mycelia were autoclaved at 120 C for 20 min after sonication. The autoclaved mycelia were centrifuged at 15000 x g for 1h. The supernatant was transferred to four 15 ml tubes and frozen at –80° and then lyophilized. The weight of the 15 ml polypropylene tubes was determined before addition of supernatant and after lyophilization to determine the weight of the elicitor (0.4 g to 1.3 g / tube). The dried elicitor fraction was suspended in 1 mM Tris-HCl pH 7.5 buffer at a final concentration of 100 mg ml^{-1}.

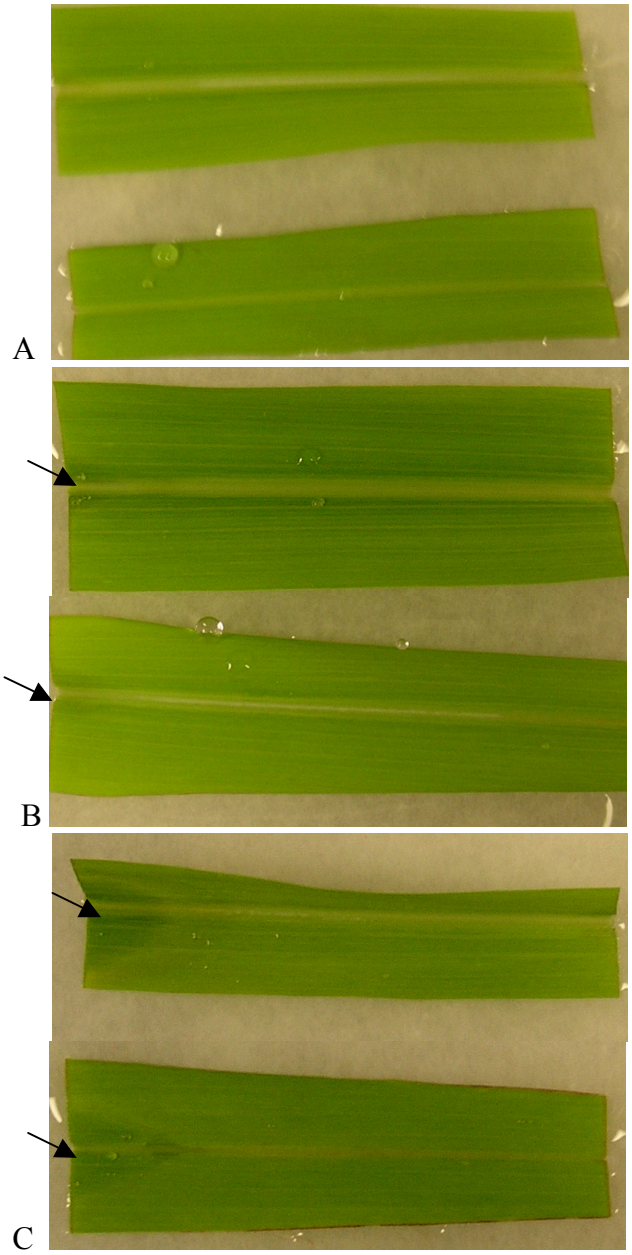


Figure 4. Elicitor from *Pichia pastoris* treated maize leaf segments. A, PEF4 treated leaf segments, no watersoaking was observed; B, PEF3 treated leaf segments, watersoaking showed up after 36 hours incubation; C, PEF1 treated leaf segments, watersoaking appeared after 36 hours.

Scale-up of protein expression from *P. pastoris*. A 10 ml BMGY (Buffered Glycerol-complex Medium) culture was inoculated with colony from a plate (MD medium) overnight at 30 ° with shaking at 220 rpm. This culture was transferred to 200 ml BMGY culture and grown as above until OD_{600} > 2. The culture was centrifuged at 1500-3000 x g for 5-10 min. The pellet was resuspended with 100 ml BMMY then transferred to a sterile flask covered with sterile cheesecloth. This culture was grown as above except that 0.5 ml of methanol was added every 24 h to maintain induction of the AOX promoter. After 3 to 4 days, the culture was centrifuged as above. Three to four ml of Ni-NTA (100mM) was added to the culture filtrate and incubated at 4° for 5 h in a separatory funnel. The beads were then loaded onto a chromatography column and washed with the washing buffer (4 ml) twice. The material bound to the beads was eluted with imidazole (250 mM). This fraction is *Pichia* Elicitor (PEF1). PEF1 was dialyzed in 500 molecular weight cut-off Spectropor dialysis membrane (Spectrum Laboratories Inc.) against 1 mM Tris, pH 7.5 (*Pichia* Elicitor Fraction 2, PEF2). The dialyzed solution was centrifuged through Amicon Ultra Centrifugal Filter Devices with a 5kDa molecular weight cut-off. The flow-through (PEF3) was collected for activity testing on plants The fraction retained by the 5 kDa filter (0.25 ml) was washed with 1mM Tris buffer, pH 7.5, with additional 1 mM Tris buffer (5 ml) three times. This resulted in an 8000-fold dilution of any <5000 Da molecule in the final retained 0.25 ml on the filter. Tris (1 mM, pH 7.5, 0.75 ml) was added to the retained 0.25 ml to produce 1 ml of solution (PEF4) that was 32,000-fold diluted for PEF3.

CHAPTER IV

EXPRESSION AND ACTIVITY TEST OF TWO HYPOTHETICAL PROTEINS

Introduction

The problem with overexpressing protein in *M. grisea* is that the amount of most of the purified proteins is usually only sufficient for performing preliminary analysis of the activity on plants. A higher yielding expression system would help to solve this problem. A second source of protein would also help to address concerns of potential co-purification of contaminant elicitors from *M. grisea*. As *P. pastoris* grows faster and vectors for high-level production of protein are commercially available, I decided to purify proteins from *Pichia* expression system for large amounts of proteins.

Two hypothetical proteins designated MG10424.4 and MG09998.4 were examined for elicitor activity. The best bidirectional blast hit of MG10424.4 is a putative riboflavin reductase in *Aspergillus oryzae (*46% identity). However, the *A. oryzae* protein is ~100 amino acids longer at the N-terminus. Careful analysis of the *M*. g*risea* sequence did not lead me to suspect that there is any problem in the annotation of the *M. grisea* sequence. Furthermore, the *M. grisea* protein is approximately the same length of other hypothetical proteins from sequenced fungal genomes. Thus, the *M. grisea* MG10424.4 might have enzymatic activity but its function is unclear. There are no EST sequences of *M. grisea* for this coding region in the databases. The *M. grisea* transformant expressing MG10424.4 did not have altered pathogenicity toward rice plants. Thus, the MG10424.4 protein encodes a small hypothetical protein that may have reductase-type activity that is predicted to be secreted and contains four cysteine residues that may form disulfide bridges. Although protein was expressed in *M. grisea,* direct

testing of the protein on rice plants was not performed. Thus, overexpression in *Pichia* was attempted to determine if the protein had activity. The coding region contains no introns, thus, facilitating its cloning for expression in *Pichia.*

MG09998.4 was also previously expressed in *M. grisea* and was not found to have activity in the rice symptom assay. The *M. griesa* strains expressing the protein also did not display altered symptoms on rice plants. Three ESTs for the gene were detected in when *M. griesa* is grown in rich medium, verifying it is an expressed gene. The gene encodes a small cysteine-rich protein that is unique to *M. grisea.*

I chose to use these two genes to test the *Pichia* protein expression system since neither gene has introns. Although these hypothetical proteins were not be expected to have activities based on the available data, I found that these proteins appear to cause leaf yellowing symptoms on rice leaves.

Results and Discussion

MG10424.4 and MG09998.4 are annotated as hypothetical proteins in *Magnaporthe grisea*. MG10424.4 has 137 amino acids including the signal peptides (Fig. 5). The signal sequence is predicted to be encoded by the first 18 amino acids. BLAST analysis (Althshul et al. 1990) of protein MG10424.4 revealed that it is related (~20-46% identity) to a group of proteins in other fungi, and the best hit with known function is annotated as being a riboflavin aldehyde-forming enzyme in *Aspergillus fumigatus*.

MG09998.4 is predicted as a 93 amino acid protein with an 18 amino acid secretion signal peptide (Fig. 6). The BLAST analysis showed that it has no credible homologues in other organisms.

mqlsvmtlaa lattalgsal pprhtpplst rstalhtgdi tyfhpalgac grtngdddli gslpqsffdr ytpggnpnln slcgtrvrvr rgdrhvdvev vdrcvgcadg didisiga hiadvgegrv ggsweqi

Figure 5. Sequence of hypothetical protein MG10424.4. The underlined amino acids are the secretion signal sequence, which will be cut after being secreted.

mkassilali fvgvavaapg tpvqgavleg rqtkptppkn tpkpsspptt ctpgkyrcsg sdiqvcnssk qwvlsakcsp kkcseqngga yci

Figure 6. Sequence of hypothetical protein MG09998.4. The underlined amino acids are the secretion signal sequence, which will be cut after being secreted.

Both proteins were overexpressed to high levels in the *P. pastoris* system. The crude culture filtrate contained a large number of proteins, however, the overexpressed proteins were detected in the crude culture filtrate by western blot analysis (Figs. 7 and 8). One-step purification by Ni-NTA affinity chromatography, followed by desalting using the Amicon 5000 Da Centricon system yielded a highly purified product (Figs 9 and 10). The total yield was 0.1 ug/ul of MG10424.4 and 0.1 ug/ul of MG09998.4. Thus, the *Pichia* expression system was readily adapted for use in expression of *M. grisea* proteins. Co-purifying material might be present, as overloading of gels revealed a faint smearing throughout the lane, however, this could also be trailing caused by overloading. These smears were absent in gels that contain lower amounts of protein to give sharp bands. Thus, if there is contamination with other proteins, they are present at much lower amounts than the target proteins.

To understand the activity of these two hypothetical proteins purified from *Pichia*, I applied approximately 1 ug of the protein solution in 15 ul at the end of rice leaf segments (Fig 11). Yellowing symptoms on rice leaf segments appeared 48 hours after inoculation with protein drop, leaf segments treated with 1 mM Tris buffer did not show any response.

The yellowing indicates that both hypothetical proteins may have senescence inducing activity. Jasmonic acid and ethylene are regulators of plant senescence and future studies to determine the levels of these plant compounds will be of interest. The yellowing symptoms of the rice leaves suggests that these proteins might play a role in fungal virulence. Further work will be focused on revealing their activity. For example, expression analysis will help to determine if the genes for these proteins are expressed

during growth of the fungus in planta. Since these are single copy genes and not members of a protein gene family, mutational analysis may prove useful. However, since multiple proteins may contribute to virulence, a lack of phenotype would not be sufficient to exclude a role for the proteins during infection. Once the symptom-inducing activity has been better characterized, these proteins may serve as useful tools for identifying their plant targets.

Materials and Methods

Plants and growth. Four week old rice cultivar M202, susceptible to *M. grisea* strain 70-15 was used in this study to test protein activity. Maize plants (*Zea mays* B73) were provided by Dr. Kolomiets (Texas A&M Univ.). Two week old barley (cv Bonanza) plants were grown from seed. Cotton (cv Atlas) plants were kindly provided by Dr. Kenerley (Texas A&M Univ.). And tobacco plants (cv *Nicotiana benthamiana*) were kindly provided by Dr. Scholthof (Texas A&M Univ.).

Expression of two hypothetical proteins (MG09998.4 and MG10424.4) in *Pichia pastoris.* The vector pIC3.5 (Invitrogen, Figure 12) vector was used for cloning genes for expression in *Pichia pastoris* strain KM71 (Invitrogen). The pDL1 constructs were used as template for PCR amplification of the inserts for cloning into pIC3.5. Universal primers for amplification incorporated a *Bam*HI restriction site in the 5' primer and an *Eco*RI restriction site in the 3' primer. The restriction sites were used to directionally clone DNA fragments into pIC3.5. Confirmation of clone inserts was performed by PCR with 5' AOX1 and 3' AOX1 primer (Invitrogen).

.

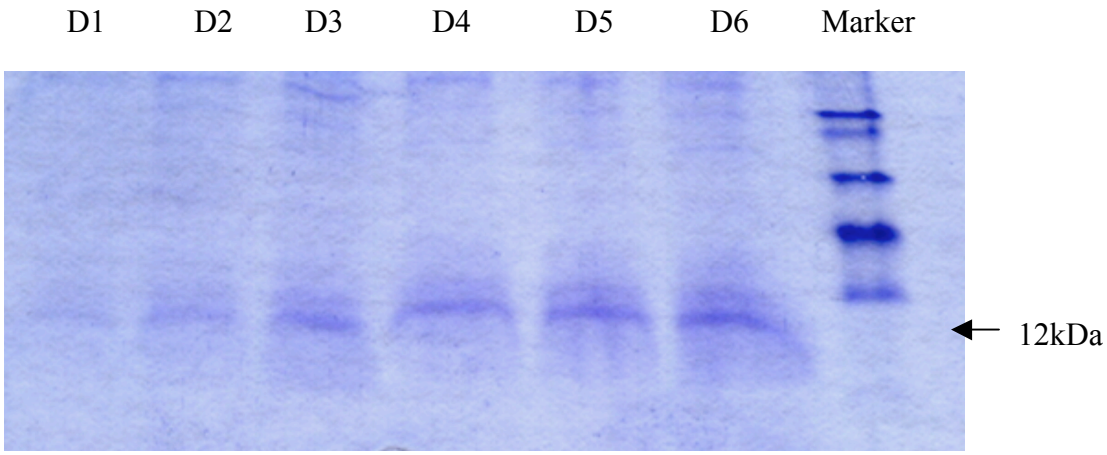


Figure 7. Overexpressed protein MG09998.4 on SDS-PAGE. The protein was from day1-day 6 culture filtrate.

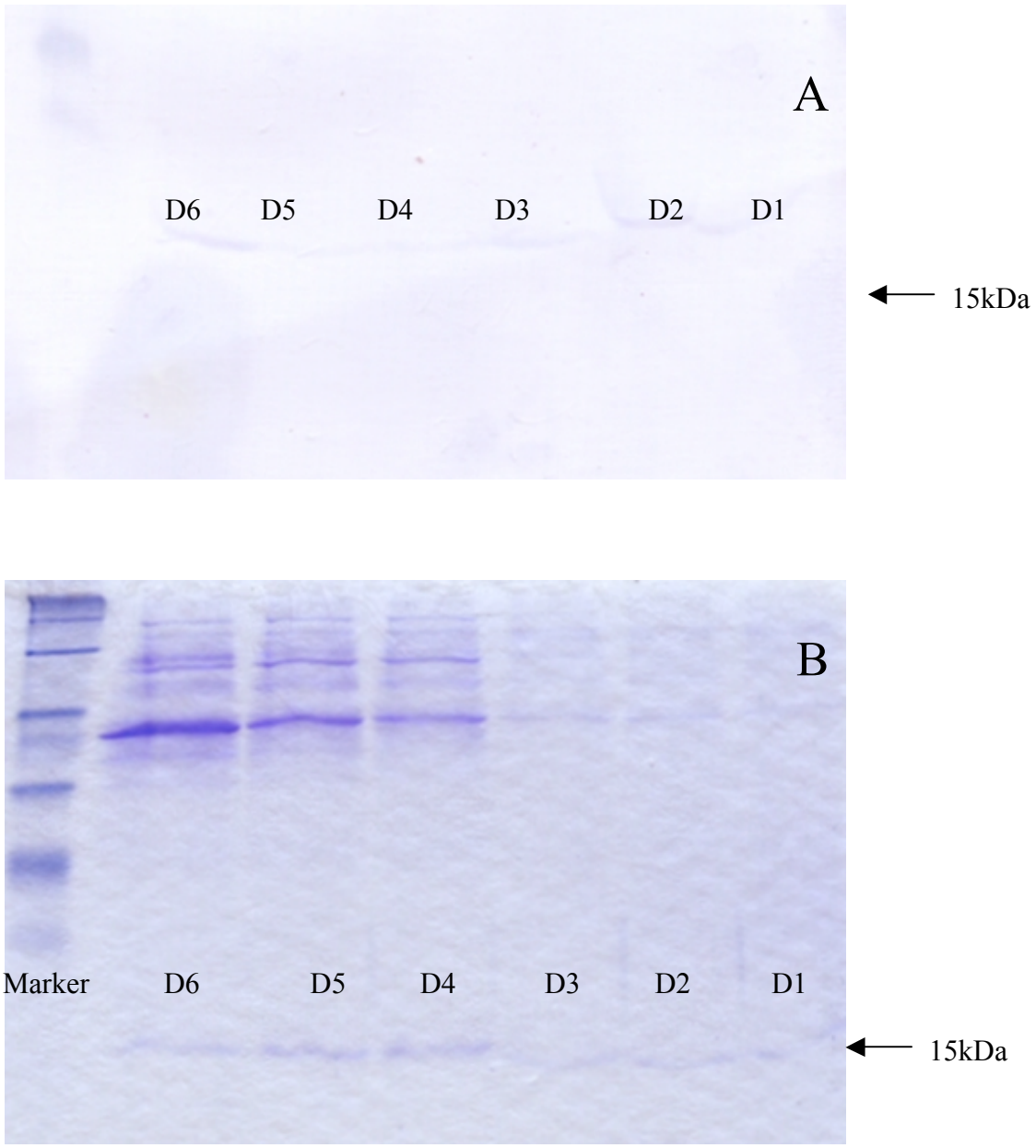


Figure 8. Overexpressed protein MG10424.4 on western blot membrane (**A**) and SDS-PAGE (**B**). The protein was from day1-day6 (D1-D6) transformed *Pichia* culture filtrate. Multiple bands were detected on coomassie blue staining gel, which might indicate that *Pichia* proteins were also secreted into culture filtrate.

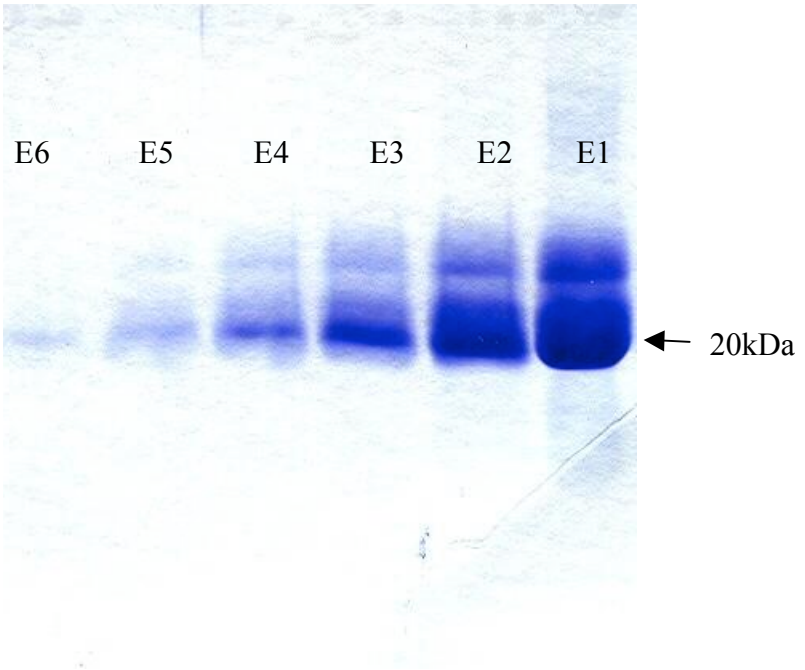


Figure 9. Purified protein: MG09998.4 by *Pichia* expression system. Detected by commassie blue staining. Column was eluted with six 1 ml aliquots of imadazole. A total of 25 ul of each elution was loaded on the gel. E1 is the first elution through which has the most protein detected.

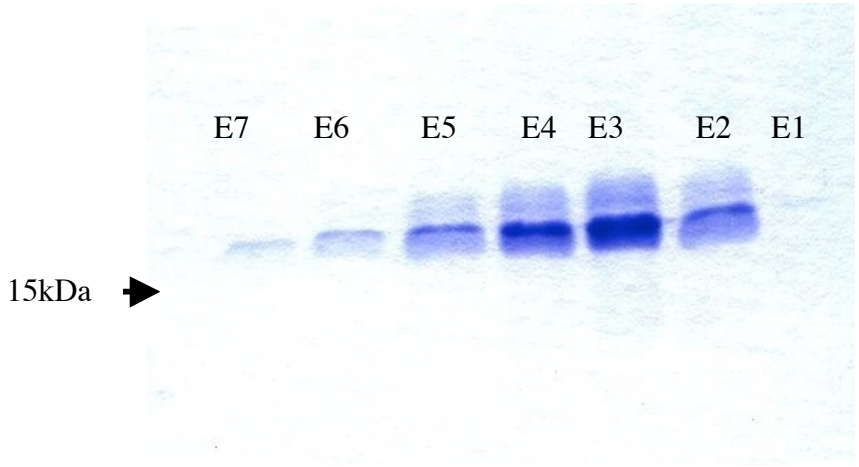


Figure 10. Hypothetical protein MG 10424.4 purified from *P. pastoris*. Column was eluted six times and protein started to come out from the first elution.

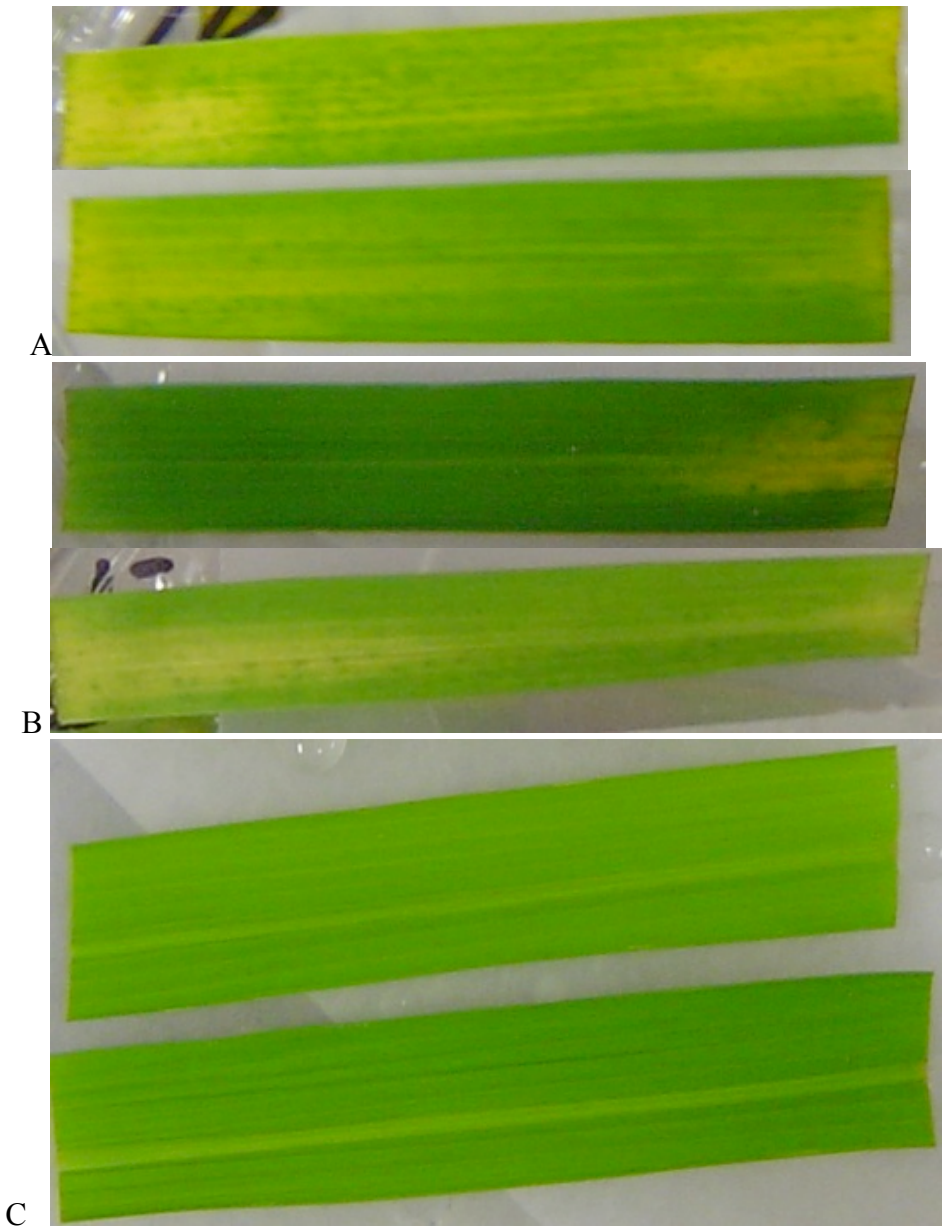


Figure 11. Yellowing symptoms of rice leaf segments infected with MG10424.4 (**A**) and MG09998.4(**B**), both are hypothetical proteins. Protein drop was applied at the right end of leaf segments and laid in a pre-wet Petri dish. After 48 hours, systems began to appear. **C**, 1mM Tris-HCl buffer was used as the control treatment.

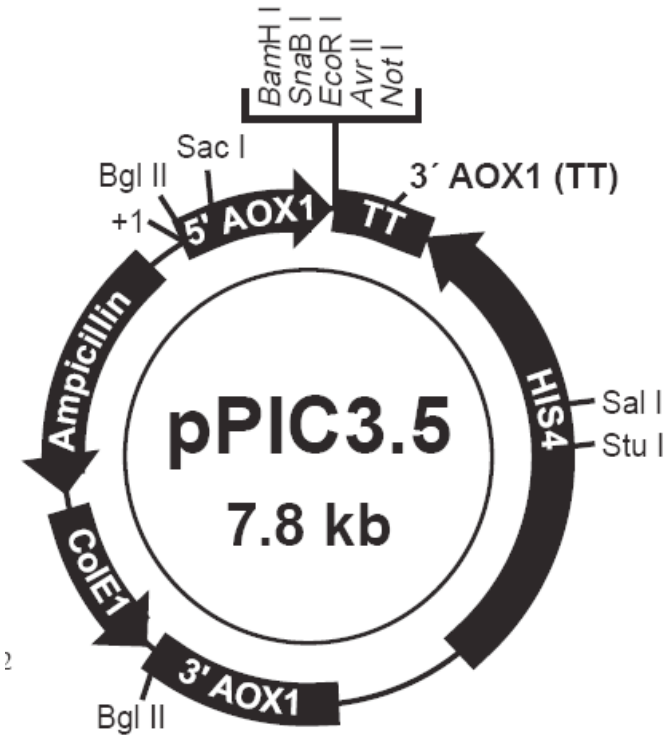


Figure 12. Map of pPIC3.5, which is a 7751 bp nonfusion vector. *Bam*HI and *Eco*RI are the restriction sites we used to integrate the *M. grisea* gene into the vector. HIS4 gene is to screen His+ colony on MD plates. The promoter of 5' AOX1 gene is to drive the integrated gene expression.

Transformation of genes with no-introns into *P. pastoris*. The constructed plasmids were propagated in *E.coli* DH5a cells. The plasmid DNA was linearized by digestion with *Stu*I (New England Biolabs, Pickerin, ON, Canada) to promote integration into the *his4* region in KM71 cells. The linearized plasmids were extracted with phenol: chloroform: isoamyl alcohol and ethanol precipitated, then dissolved in 1-20 microlitre Tris-EDTA (TE) buffer and stored at –20° C until ready to transform.

Transformation of *P. pastoris* strain KM71 cells was performed by electroporation using the "Gene Pulser II" electroporation system (Bio-Rad, Mississauga, ON, Cananda). Integration of the plasmid DNA into the yeast chromosomes was verified by PCR using AOX1 primers.

Expression assay. Two *P. pastoris* transformants were selected to test for protein expression. Protein expression at different times during growth was determined by addition of methanol to a final concentration of 0.5% at 24 h-intervals, and 1-ml samples were taken at selected times. The cell pellet and supernatant were both stored at -80°. Supernatants and cell pellets were analyzed by staining SDS-PAGE gels with Coomassie Blue and performing western blots with antibodies directed against the histidine tag.

Scale-up of protein expression from *P. pastoris*. A single colony was used to inoculate 10 ml BMGY [(Buffered Glycerol-complex Medium) medium containing 1% yeast extract, 2% peptone, 100mM potassium phosphate, pH6.0, 1.34% YNB (Yeast Nitrogen Base with Ammonium Sulfate without amino acids), 4x 10-5% biotin and 1% glycerol] in a 100 ml baffled flask and incubated at 28-30° in a shaking incubator (250-300 rpm) until culture reached an $OD_{600} > 2$. This 10 ml culture was used to inoculate 1 liter of BMGY in a 2 liter baffled flask and incubated at 30° in a shaking incubator (250-300 rpm) until the culture reached an $OD_{600} > 2$. The cells were harvested by centrifuging at 1500-3000 x g in Sorvall

RC-5B Refrigerated Superspeed Centrifuge for 5 min at room temperature. To induce expression, the pellet was resuspended in 1/5 to 1/10 of the original culture volume of BMMY medium. The culture was placed in a one liter flask and the flask opening was covered with 2 layers of sterile cheesecloth and returned to the incubator. Methanol was added to 0.5% every 24 hours until the optimal time of induction was reached. Cells were harvested by centrifuging at 1500-3000 x g for 5 min at room temperature. The supernatants were transferred to separatory funnel for protein purification by Ni-NTA affinity Chromatography.

Large-scale secreted proteins preparation from *P. pastoris.* After incubation for an optimal period of time (3-5 days), *P. pastoris* transformants were centrifuged at 5000x g and the supernatant was subjected to the same procedure as the culture filtrate from *M. grisea.*

Desalting purified protein. Please refer to 'Materials and Methods' in Chapter II.

Detached leaf assay and hydrogen peroxide assay. Please refer to 'Materials and Methods' in Chapter II.

CHAPTER V

PURIFICATION AND ACTIVITY TEST OF CLUSTER 180 PROTEINS

Introduction

Gene families unique to pathogenic fungi, or unique to *M. grisea* are strong candidates for being effectors based on BLAST analysis. Expression pattern during plant infection, is additional evidence that would indicate a role in virulence. One of the largest families of secreted proteins is represented by a set of small cysteine-rich proteins, called the cluster 180 proteins because of their cluster number assigned during automated annotation. There are 14 family members in the cluster and two of these were expressed in *M. grisea* and found to be secreted. I chose to use the *P. pastoris* expression system to purify them, however, since these genes have introns and no ESTs or cDNA clones were available, it was necessary to produce cDNA for each gene.

Characterization of the proteins reveals that they induce very weak visible symptoms on rice leaves, but do induce increased levels of hydrogen peroxide when applied to maize leaf segments. Thus, these proteins are candidates for virulence factors. Careful annotation of the gene family revealed that two genes contained frameshifts and would produce proteins that are likely nonfunctional. This suggests that there is selection in the population of *M. grisea* for elimination of some members of the gene family. This is commonly observed for genes that can act as avirulence genes via recognition by resistance genes in the host species.

Results and Discussion

***Magnaporthe grisea* has a group of 14 cysteine-rich proteins.** Many *avr* genes have been found to encode small secreted cysteine-rich proteins, and the cysteine-rich domains

have been found to interact with plant proteins and induce defense responses when infiltrated into plants (Lauge and de Wit 1998), (van't Slot 2002). A family of 10 related cysteine-rich proteins was found in *M. grisea* (Dean et al, 2005). On more careful inspection of the genome, four additional coding regions were found, two of which appear to have frameshifts, and thus may be pseudogenes. The translated nucleotide sequence for each protein is shown in appendix 1. Figure 13 shows the alignment of these 14 proteins where the genes with frameshifts have been manually altered by addition or deletion of a single nucleotide to shift the reading frame to produce the best alignment with the other family members (Altschul et al. 1990). All of them have two positions of double-cysteine residues. In addition, the cysteine residues have a well-conserved amino acid context in all the 14 proteins. A tree illustrating the relationship between the proteins was generated using Clustalw (Higgins et al 1994). This shows that some family members are more closely related to each other than to other members of the family based on the sequence similarity (Fig. 14).

I purified three cluster 180 proteins from *P. pastoris* (Fig. 15), and test one of them MG 10734.2 on maize leaf segments for elicitor test.

Cluster 180 protein MG10732.4 causes mild yellowing symptoms and H_2O_2 production on maize leaf. To determine the activity of the small cysteine-rich cluster 180 proteins on plants, we applied protein drops at the end of maize leaf segments. A very weak yellowing was observed 48 h after inoculation (Fig. 16, A, B and C). The 1mM Tris-HCl pH 7.5 treated leaf segments did not show yellowing at all after 48 hours.

To determine if H_2O_2 was produced during the reaction between the cluster 180 protein and plants, we stained the inoculated leaf segments with DAB (3, 3'-

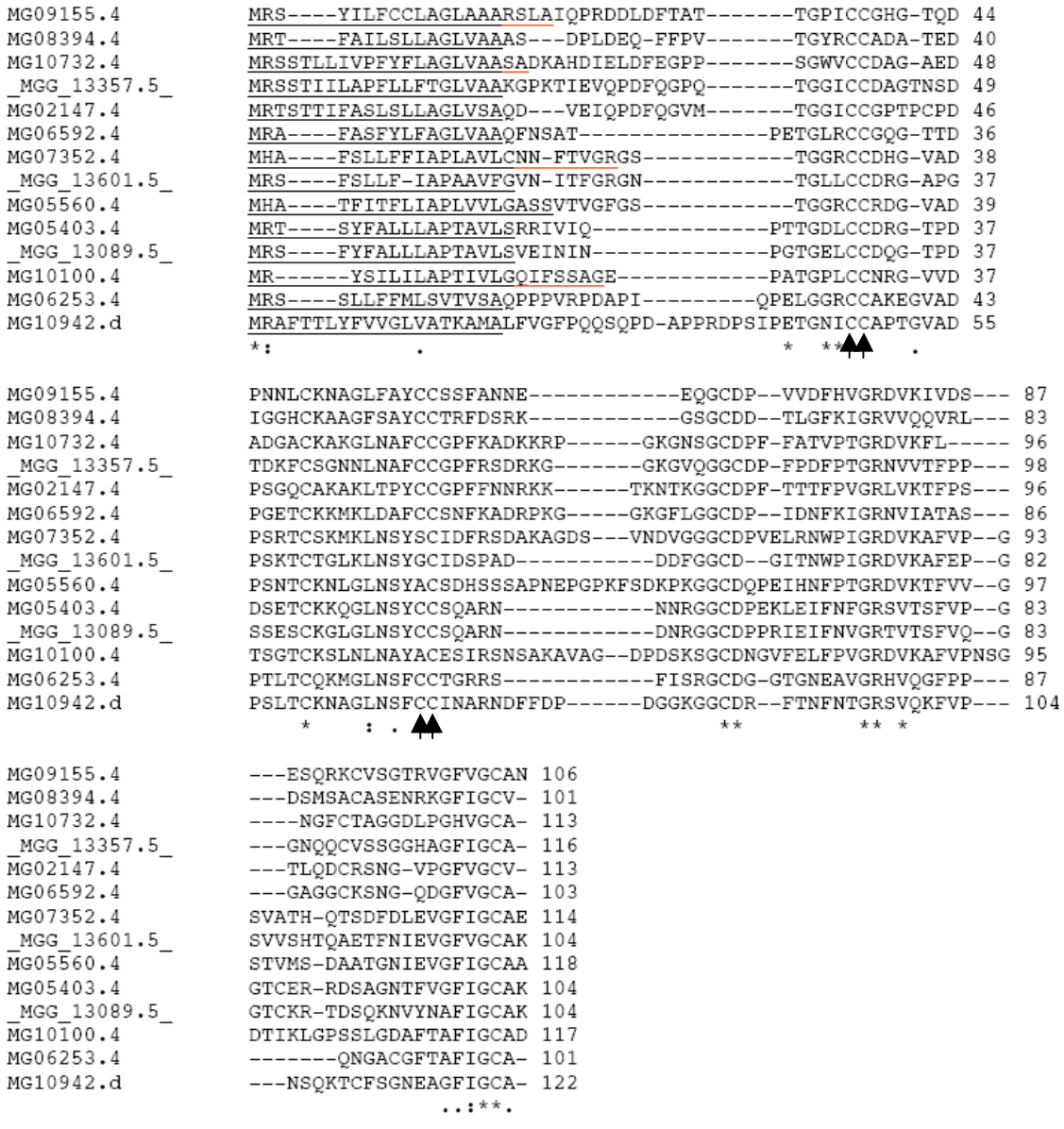


Figure 13. Alignment of 14 cluster 180 proteins. The alignment was done using CLUSTALW (Higgins et al. 1994). Identical residues in all fourteen sequences are marked by an asterisk, conserved substitutions are marked by a semicolon (:), and semiconserved substitutions by a dot. The position of double-cysteine residues is indicated by arrow.

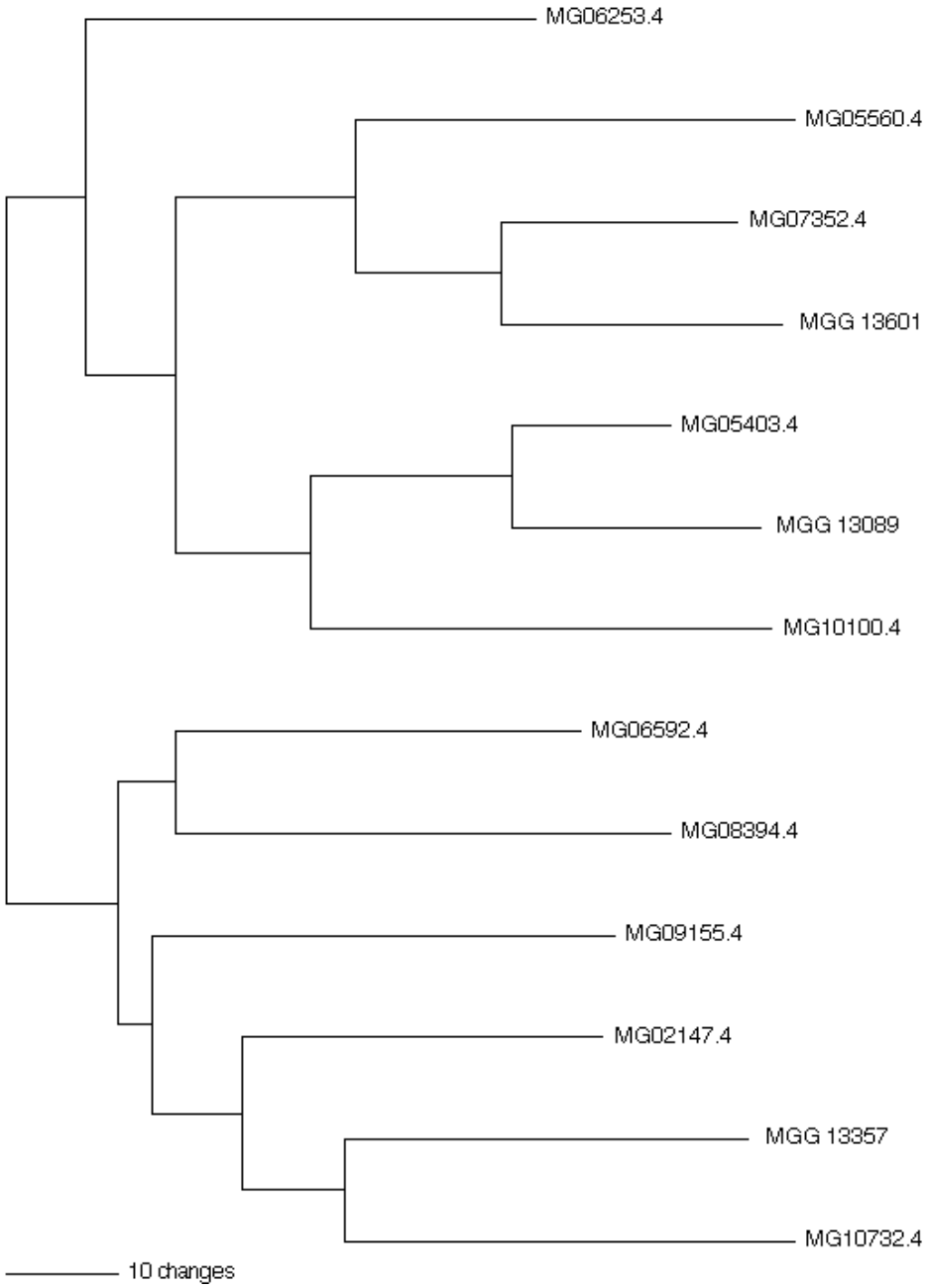


Figure 14. Phylogram showing the relationship of the cl180 gene family members.

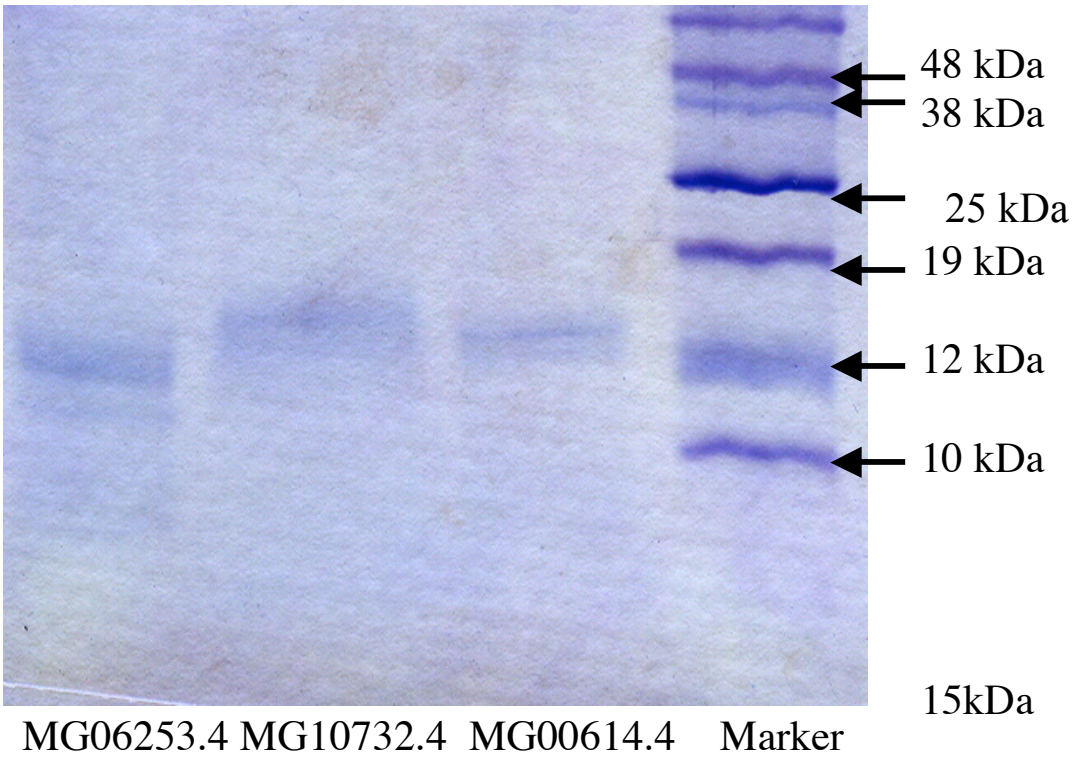


Figure 15. Commassie blue staining gel of three cluster 180 proteins purified from *Pichia.* MG06253.4, MG10732.4 and MG00614.4(left to right).

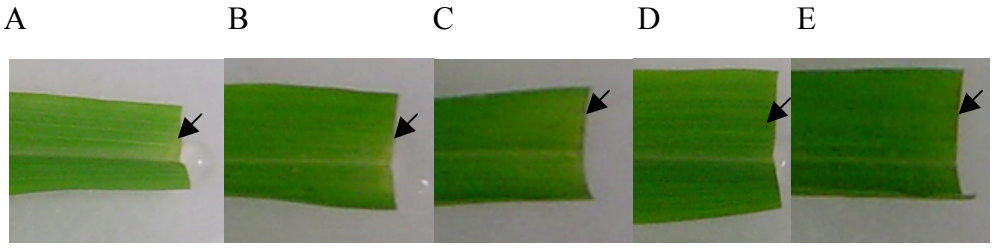


Figure 16. Yellowing symptoms of maize leaf segments inoculated with cluster 180 protein MG10732.4 (inoculated sites are indicated by arrow in **A**, **B and C**). **D and E**, 1mM Tris-HCl buffer was used as the control treatment.

diaminobenzidine tetrahydrochloride). Hydrogen peroxide is a product of the plant defense system to attack by pathogens. The chemical DAB precipitates in plant cells where H_2O_2 is produced (Fig. 17 A and B). In Figure 18-A and –B, DAB (dark brown) precipitated at the end of maize leaf segments where the protein drop was applied, indicating that leaves responded to the protein and defense reactions were induced by the protein. In contrast, leaf segment inoculated with 1mM Tris buffer did not show any response. In repeated experiments the reaction was observed but was weak.

The cluster 180 proteins represent an interesting group of small cysteine-rich proteins in *M. grisea*. Previous studies by other researchers with other fungi have shown that the cysteine-rich domains are able to interact with molecules in plants (Lauge and de Wit 1998),(van't Slot 2002). Some cysteine-rich proteins in *Phytophthora* behave as proteinase inhibitors according to Sophien Kamoun's (1999) work, these proteins can function in the proteinase-rich environment such as the plant intercellular spaces, inhibiting the proteinase activity. In *M. grisea*, the group of cluster 180 proteins has not been studied, and there is little known about their function. After the first step of testing one of the cluster 180 proteins, MG10732.4, on plants, we observed the mild yellowing and DAB precipitation in leaf segments. These reactions might indicate that plants respond to proteins MG10732.4 and the protein triggered the defense response. Thus, cluster 180 protein MG10732.4 may have elicitor activity. Further analysis of MG10732.4 and all other members of this gene family is warranted.

Mutational analysis of the family members would be a useful step in characterizing their role in virulence. However, as they may possess redundant function, it may be necessary to mutate multiple members of the gene family to discern their roles.

A B C

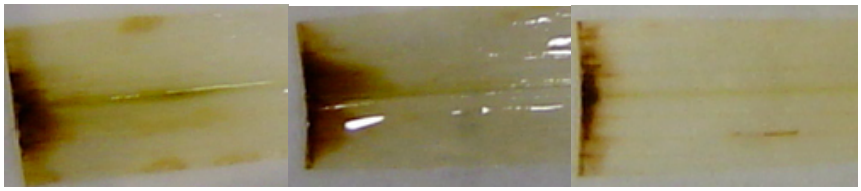


Figure 17. Hydrogen peroxide formed at the end of maize leaf segments (Dark brown). Cluster 180 protein MG10732.4 (A and B) and 1mM Tris buffer (C) were applied to the leaf ends, after 48 hours in the dark, the segments were stained by DAB(3,3'-Diaminobenzidine tetrahydrochloride) which would precipitate where H_2O_2 was produced. Ethanol: Lactic acid: Phenol (2:1:1) was used to distain the leaves.

Materials and Methods

cDNA cloning of three intron-containing cluster 180 genes. Four week-old rice plants were infected by *M. grisea* strain 70-15 according to the inoculation method of Koga and Nakayachi (2003). Leaf sheaths of the sixth leaves of rice plants were peeled off with leaf blades and roots. The leaf sheath was placed horizontally on a support and the sheath cavity filled with a suspension of spores (1000 spores ml^{-1}) of *M. grisea* using a needle and syringe. The leaf sheaths were incubated at 25° in glass trays under white fluorescent light with a 12 h light period. Total RNA was extracted from the sheath from sets of plants every day for six days. cDNA was synthesized using PowerScript Reverse Transcriptase (BD Biosciences) with Oligo-dT as the primer. Eleven of the 14 members of the gene family could be amplified. Intron-free MG06253.4, MG10732.4 and MG02147.4 genes were amplified with the synthesized cDNA (Fig. 18) as the template from RNA isolated from 6 day post-inoculation plants and cloned into pPIC3.5.

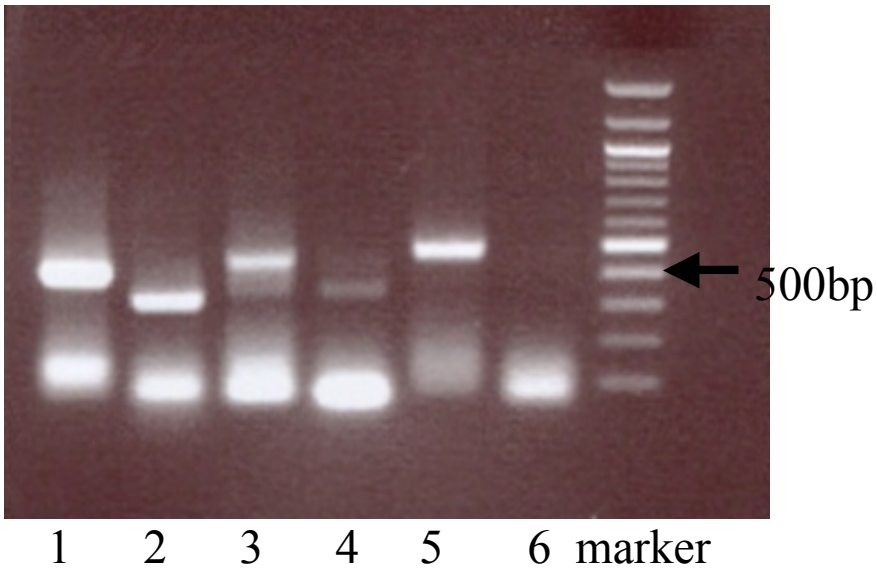


Figure 18. Result of RT-PCR of three cluster 180 genes. 1.MG06235.4 (gDNA); 2.cDNA of MG06235.4 ; 3.MG10732.4 (gDNA); 4.cDNA of MG10734.2; 5.MG00614.4 (gDNA); 6.cDNA of MG00614.4.

CHAPTER VI

PURIFICATION AND ACTIVITY TEST OF MG 07016.4 (LIPASE)

Introduction

Fungal pathogens are believed to secrete different extracellular enzymes to increase the virulence against host plants. However, the specific role in virulence of most of these enzymes is still under study. Lipases represent a large group of secreted enzymes in fungi. They are enzymes that are able to hydrolyze triacylglycerols into glycerol and free fatty acids (FFA). In nature, lipases are ubiquitous (Brockman 1984). They have been found in animals, plants, fungi, and bacteria (Jaeger and Reetz 1998; Mukherjee and Hills 1994). In *Fusarium graminearum,* the gene *FGL1,* encoding an extracellular lipase, was found to be a virulence factor. Disease severity was strongly reduced in Δfgl1 strains (Voigt et al, 2005). In *M. grisea*, we found a gene (MG07016.4) encoding a 348 amino acid protein, which has strongest identity to a lipase in *F. heterosporum* (Fig. 19). Because MG07016.4 is the ortholog of the *F. graminearum* lipase virulence factor, I wanted to test this protein to determine if the enzyme itself may induce symptoms on plants. I purified this protein of *M. grisea* in *P. pastoris* (Fig.20) and tested it for the lipase activity as well as an effector activity on plants.

Results and Discussion

MG07016.4 encodes a protein with lipase activity. BLASTP analysis of MG07016.4 revealed that it is homologous to lipases in other organisms. Amino acid identity values, using BLASTP to the lipase in *Fusarium heterosporum* and *Gibberella zeae* (*FGL1*) are 51% and 48% respectively. ProSite (http://ca.expasy.org/prosite/) analysis of the

```
Query  6    VLTLLATALTCSASVLPAGLTYTKTVEGRDVTVSETDLDNFRFYAQYSAATYCNDAAASG  65
            VL+LL+     +A  +P+    T+ +E R VTV+  DL NFRFY Q++ A YCN   A G
Sbjct  4    VLSLLSIIAFTAAGPVPSVDENTRVLEHRAVTVTTQDLSNFRFYLQHADAAYCNFNTAVG  63

Query  66   AAVACSNDGCPAVVANGAKIIRSLNQDTSTNTAGYLALDPKRKNIVLALRGSTSLRNWIT  125
              V CS   CP +  + A ++ S+   T T    Y+A D  RK IV+++RGS ++RNWIT
Sbjct  64   KPVHCSAGNCPDIEKDAAIVVGSV-VGTKTGIGAYVATDNARKEIVVSVRGSINVRNWIT  122

Query  126  NLTFLWTRCDFVQDCKLHTGFATAWSQVQADVLAAIADAKAQNPDYTVVVTGHSLGGAVA  185
            N  F    CD V  C +HTGF  AW +V A+V AA++ AK  NP +  VVTGHSLGGAVA
Sbjct  123  NFNFGQKTCDLVAGCGVHTGFLDAWEEVAANVKAAVSAAKTANPTFKFVVTGHSLGGAVA  182

Query  186  TVAGVYLRQLGYPVEVYTYGSPRIGNQEFVQWVSTQAGNVEYRVTHIDDPVPRLPPIFLG  245
            T+A  YLR+ G+P ++YTYGSPR+GN  F  +V+ Q G  EYRVTH DDPVPRLPPI  G
Sbjct  183  TIAAAYLRKDGFPFDLYTYGSPRVGNDFFANFVTQQTG-AEYRVTHGDDPVPRLPPIVFG  241

Query  246  YRHVTPEYWLNSGTSNTVNYTVADIKVCEGFANINCNGGSLGLDTNAHLYYLTDMIACGS  305
            YRH +PEYWLN G  +  +YTV +IKVCEG AN+ CNGG++GLD  AH+ Y   M  C
Sbjct  242  YRHTSPEYWLNGGPLDK-DYTVTEIKVCEGIANVMCNGGTIGLDILAHITYFQSMATCAP  300

Query  306  NKFVFRRDDANAISDAELEQRLTMYAQMDREFVAAL  341
               ++RD    +SD ELE++LT Y++MD+EFV  +
Sbjct  301  IAIPWKRD----MSDEELEKKLTQYSEMDQEFVKQM  332

Query: MG07016.4
Sbjet: lipase in Fusarium heterosporum
```

Figure 19. Alignment between MG7016.4 and lipase in *Fusarium heterosporum*. The lipase active site is between 173-182 amino acid (red underlined). This site is conserved in both proteins.

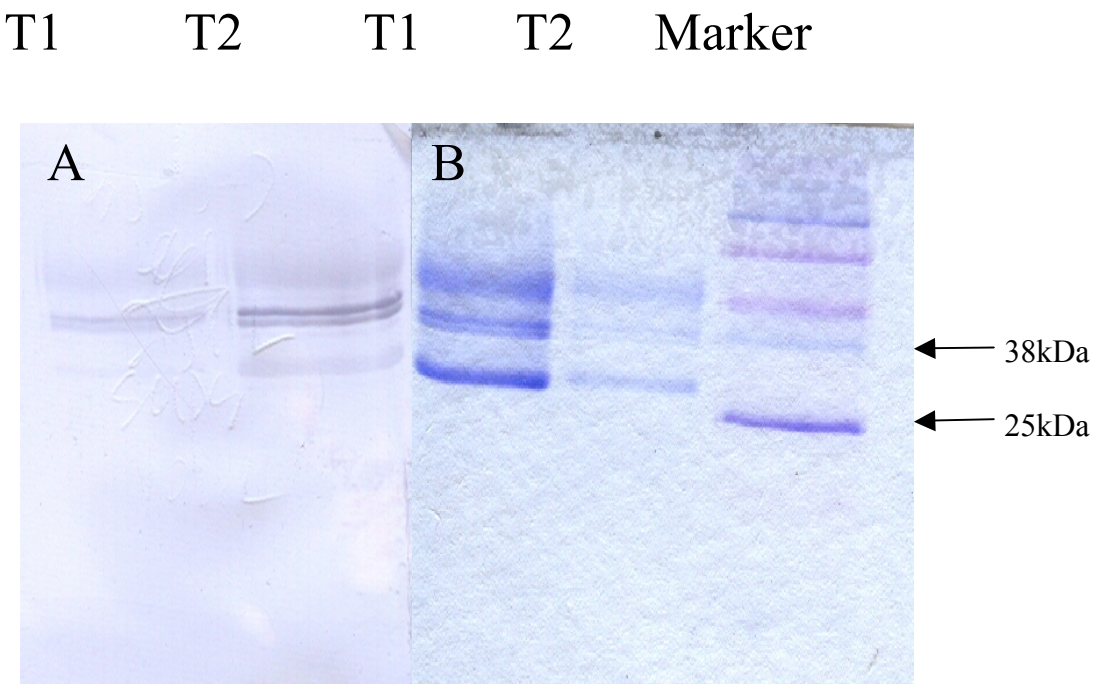


Figure 20. Lipase MG07016.4 purified by *Pichia* expression system. Two transformants (T1 and T2) of lipase MG07016.4 was detected by commassie blue staining (**B**) and western blot (**A**) using $RGSH_6$ antibody. 30 ul of protein solution was loaded.

MG07016.4 protein sequence predicated that it has a lipase activity domain (173-182): VVVTGHSLGG, which is also conserved in the lipases of other fungi.

I evaluated MG07016.4's lipase activity by performing the lipase activity assay described by Voigt,C.A, et al (2005) . The OD_{410} value of reaction buffer-suspended protein solution was measured at 60 one minute-intervals in the TECAN Spetrafluo Reader (MTX Lab Systems, Inc.). The graph of the enzyme activity time course is shown in Figure 21. Protein purified from two *Pichia* transformants were tested for lipase activity. T1 (tranformant 1) consumed the substrate to near completion after 14 minutes incubation and T2 had less enzyme and the reaction was complete after 25 minutes.

The protein that gene MG07016.4 encodes in *Magnaporthe grisea* contains lipase activity. Moreover, the protein purified by *Pichia* protein expression system still maintained the enzyme activity. Thus, the system can be used to obtain enzymatically active extracellular proteins from *M. grisea*.

Lipase (MG07016.4) from *Magnaporthe grisea* does not induce responses on plants. To test the response that plants show to protein MG07016.4, I applied the purified protein on leaf segments of rice (M202), barley (cv Bonanza), maize (B73), and cotton (Atlas) and tobacco (*Nicotiana benthamiana*) leaf disks with 1mM Tris buffer as negative control. Seventy-two hours after inoculation, no leaf segments showed symptoms (data not shown). I also performed the hydrogen peroxide detection assay, and did not observe DAB precipitation in leaf segments.

The absence of a response in plants to MG07016.4 may mean that the lipase cannot trigger a response or the protein is not recognized as an elicitor. Another possible explanation is that the plant intracellular environment contains proteinases and

MG07016.4 may get degraded and lose enzyme activity or the structure of the protein, which can trigger the response, was eliminated by plant proteases. In addition, the lipase may act as a virulence factor only when delivered into the plant cell. Finally, the lipase may simply be important in *F. graminearum* for nutrient acquisition and its reduced virulence may simply result from an inability to properly utilize host resources.

Materials and Methods

Purification. Please refer to 'Materials and Methods' in Chapter II.

MG07016.4 Lipase activity assay. A volume of 20 ul of reaction buffer [2 mM pNPP, 0.1% (v/v) Triton X-100, 0.1% (w/v) gum arabicum, 0.05 M Sorensen phosphate buffer pH8.0 was diluted 10-fold, and then mixed with purified lipase MG07016.4 (usually 0.3 to 3.0 microgram) (Voigt et al., 2005). The assay was carried out in 96-well microtiter plate at 37° in a TECAN Spetrafluo Reader (MTX Lab Systems, Inc.) Para-nitrophenol (pNP) production was determined photometrically at 410 nm at one-minute intervals. Lipase protein concentration was measured using the Bio-Rad protein assay kit (Bio-Rad, Inc.).

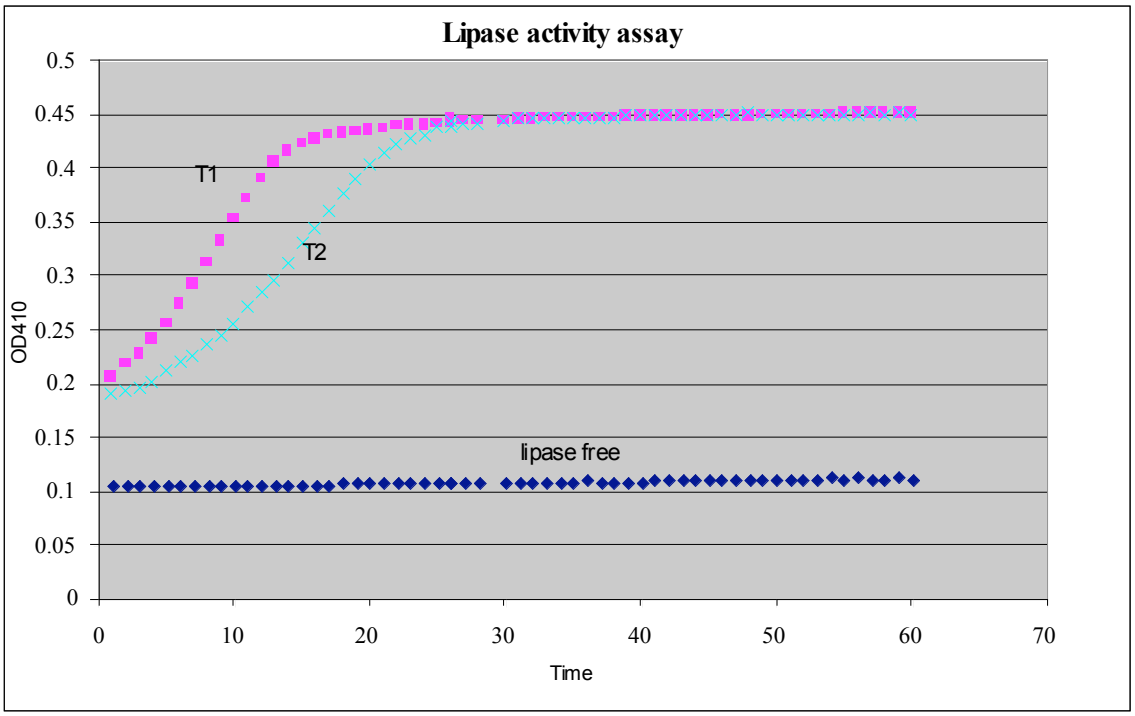


Figure 21. Lipase activity assay curve. The para-nitrophenol (pNP) amount was determined photometrically at 410 nm 60 times with one-minute interval against lipase free reaction solution (blue line). Lipase was purified from two *Pichia* transformants (T1 and T2). The lipase activity is expressed as the increase of OD_{410} value.

CHAPTER VII

SUMMARY

Magnaporthe grisea is the most economically important pathogenic fungus of rice and has been studied extensively. However, knowledge of secreted proteins of *M.grisea* is still limited. This study provided more insights about the function and biological activity of several secreted proteins. The proteins purified from *M. grisea* did not produce strong or consistent symptoms on plants. However, the two hypothetical proteins (MG09998.4 and MG10424.4) purified from *Pichia* were found to cause leaf yellowing, which might indicate the pre-mature senescencing of leaves. These two proteins may contain elicitor activity. MG09998.4 is particularly interesting since there are no homologs in other fungi based on BLAST searches. Additional analysis with these proteins is warranted. We found a 14-member group of small cysteine-rich proteins (cluster 180) in *M. grisea* and the purified cluster 180 protein MG10732.4 can induce mild yellowing and hydrogen peroxide in plants. The elicitor activity of small cysteine-rich proteins has been found in other fungi such as the elicitin in *Phytophthora infestans* and *avr4* gene product in *Cladosporium*. In the future, we are going to study all of the 14 proteins in this group. Protein MG07016.4 was found to contain lipase activity in this study. However, the enzyme activity seems not to function in the induction of plant responses. I successfully purified 6 proteins from *Pichia pastoris* with relatively higher production, which proves it to be an effective protein production system with less background proteins that could interfere with my interpretation.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Brockman HL. 1984. General features of lipolysis: reaction scheme, interfacial structure and experimental approaches. Pages 3-46 in: Lipases. Borgström B., and Brockman HL, eds. Elsevier, Amsterdam.

Cooper, R. M. 1983 The mechanisms and significance of enzymic degradation of host cell walls by parasites. Pages 101-135 in: Biochemical Plant Pathology. J. A. Callow, ed. John Wiley and Sons, New York.

Dean, RA., Talbot, N.J., Ebbole, D.J., Farman,, M.L., Mitchell, T.K., *et al.* 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434: 980-6.

Enkerli, J., Felix, G., and Boller, T. 1999. The enzymatic activity of fungal xylanases is not necessary for its elicitor activity. Plant Physiol. 121: 391-398.

Gaulin, E., Jauneau, A., Villalba, F., Rickauer, M., Esquerré Tugayé, M., and Bottin, A., 2002. The CBEL glycoprotein of *Phytophthora parasitica* var.*nicotianae* is involved in cell wall deposition and adhesion to cellulosic substrates. Journal of Cell Science 115: 4565-4575.

Gilkes, N.R., Henrissat, B., Kilburn, D. G., Miller, R. C., and Warren, R.A. 1991 Domains in microbial b1, 4-glycanases: sequence, conservation, function, and enzyme families. Microbiol Rev. 55: 303-315.

Gotesson A, M.J., Jones, D. A., and Hardham, A. R. 2002. Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete plant pathogen *Phytophthora cinnamomi*. Mol Plant Microbe Interact. 9: 907-921.

Higgins, D., Thompson J., Gibson, T., and Thompson, J. D. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Jaeger, K.-E., and Reetz, M.T. 1998. Microbial lipases form versatile tools for biotechnology. Trends Biotechnol. 16: 396–403.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. 15: 4004-4014.

Joosten, M.H., Vogelsang, R., Cozijnsen, T.J., Verberne, M.C. and De Wit, P.J. 1997. The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4-mediated resistance by producing unstable AVR4 elicitors. Plant Cell 9: 367-379.

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol. 44: 41-60.

Kamoun, S., Honee, G., Weide, Rob., Laugé, R,. Kooman-Gersmann, Miriam., and de Groot, Koen., et al., 1999. The fungal gene Avr9 and the oomycete gene inf1 confer avirulence to potato virus x on tobacco. Mol. Plant Microbe. Interact. 12: 459-462.

Kars, I., Krooshof,G.H., Wagemakers, L., Joosten, R., Benen, J.A., and van Kan, J.A. 2005. Necrotizing activity of five *Botrytis cinerea* endopolygalacturonases produced in *Pichia pastoris*. Plant J.12: 213-225.

Keen, N. T., 1990. Gene-for-gene complementarity in plant-pathogen interactions. Annu Rev Genet. 24: 447-63.

Koga, J., Oshima, K., Ogawa, N., Ogasawara, N. and Shimura, M., 1998. A new bioassay for measuring elicitor activity in rice leaves. Ann. Phytopathol. Soc. Jpn. 64: 97–101.

Koga, H.,and Nakayachi, O., 2003. Morphological studies on attachment of spores of *Magnaporthe grisea* to the leaf surface of rice. Journal of General Plant Pathology. 70: 11-15.

Kooman-Gersmann, M., Vogelsang, R., Hoogendijk, E.CM., and de Wit, P.J. 1997. Assignment of amino acid residues of the Avr9 peptide of *Cladosporium fulvum* that determine elicitor activity. Mol. Plant Microbe. Interact. 10: 821-829.

Laskowski, M., and Qasim, M., 2000. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? Biochim Biophys Acta. 1477: 324-337.

Lauge, R, and De Wit,P. J. 1998. Fungal avirulence genes: structure and possible functions. Fungal Genet. Biol. 24: 285-297.

Lu, G., Filippi, C. and Ebbole, D. 2002. Identification and characterization of secreted proteins from *Magnaporthe grisea*. Pages 47-56 in Rice blast: Interaction with Rice and Control. Shinji K, ed. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Luderer, R., De Kock, M.J.D., Dees, R.H.L., de Wit, P.J.M., and Joosten, M.H. 2002 Functional analysis of cysteine residues of ECP elicitor proteins of the fungal tomato pathogen *Cladosporium fulvum*. Mol. Plant Pathol. 2: 91-95.

Matsumura, H., Nirasawa, S., Kiba ,A., Urasaki, N., Saitoh, H., Ito, M., Kawai-Yamada, M., Uchimiya, H., and Terauchi, R. 2003. Overexpression of Bax inhibitor suppresses the fungal elicitor-induced cell death in rice (*Oryza sativa* L) cells. Plant Journal. 33: 425-434

Mukherjee, K.D., and Hills, M.J. 1994. Lipases from plants. Pages 49-75 in: Lipases: Their Structure, Biochemistry and Application.Woolley, P. and Petersen, S.B., eds. Cambridge: Cambridge University Press.

Ng,T.B. 2004. Peptides and proteins from fungi. Peptides 25: 1055-1073.

Nicholas, T. 2003. On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. Annu Rev Microbiol. 57: 177-202.

Parker, J.E. 2003. Plant recognition of microbial patterns. Trends in Plant Science. 8: 245-247.

Talbot, N.J., Ebbole, D.J., and Hamer, J.E. 1993. Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. Plant Cell 5: 1575-1590.

Tian, M., and Kamoun, S. 2005. A two disulfide bridge Kazal domain from *Phytophthora* exhibits stable inhibitory activity against serine proteases of the subtilisin family. BMC Biochem. 6: 15-24.

van den Burg, H. S. C., Boeren, S., Kennedy, M.A., Vissers, J.P., Vuister, G.W., de Wit, P.J., and Vervoort, J. 2004. Binding of the AVR4 elicitor of *Cladosporium fulvum*

to chitotriose units is facilitated by positive allosteric protein-protein interactions: the chitin-binding site of AVR4 represents a novel binding site on the folding scaffold shared between the invertebrate and the plant chitin-binding domain. J Biol Chem. 16: 16786-16796.

van't Slot, K. A. E., van den, Burg, H. A., Kloks, C. P., Hilbers, C. W., Knogge, W., and Papavoine, C. H. 2003. Solution structure of the plant disease resistance-triggering protein NIP1 from the fungus *Rhynchosporium secalis* shows a novel beta-sheet fold. J Biol Chem. 46: 45730-45736.

van't Slot K. A. E. 2002. A dual role for microbial pathogen-derived effector proteins in plant disease and resistance. Crit. Rev. Plant Sci. 21: 229-271.

Voigt, C. A., Schafer, W., and Salomon, S. 2005. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. Plant Journal. 42: 364-375

Wang, Z. Y., Jenkinson, J., Holcombe, L. J., Soanes, D. M., Veneault-Fourrey, C., Bhambra, G. K., and Talbot, N. J. 2005. The molecular biology of appressorium turgor generation by the rice blast fungus *Magnaporthe grisea*. Biochem Soc Trans. 33: 384-388.

Wu, S.C., Kauffmann, S., Darvill, A.G., and Albersheim, P. 1995. Purification, cloning and characterization of two xylanases from *Magnaporthe grisea*, the rice blast fungus. Mol. Plant Microbe Interact. 8: 506-514.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J.D., Felix, G. and Boller, T. 2004. Bacterial disease resistance in *Arabidopsis* through flagellin perception. Nature 428: 764-767.

APPENDIX

```
MG02147 from 1 to 1000

          ---------|---------|---------|---------|---------|
        1 AGGGCTTCAGGCCTATGACCTTCTGTGCGTTACGTCAGTCCCCCAACATC 50
F1      1 R  A  S  G  L  *  P  S  V  R  Y  V  S  P  P  T  S  11
F2      1  G  L  Q  A  Y  D  L  L  C  V  T  S  V  P  Q  H  Q 17
F3      1   G  F  R  P  M  T  F  C  A  L  R  Q  S  P  N  I   16

          ---------|---------|---------|---------|---------|
       51 AAGGCTTTTCATCGTATGGTGTTTCGGATTAAAGAGTTGATTAATAGGGT 100
F1     12  R  L  F  I  V  W  C  F  G  L  K  S  *  L  I  G  F 4
F2     18   G  F  S  S  Y  G  V  S  D  *  R  V  D  *  *  G   1
F3     17 K  A  F  H  R  M  V  F  R  I  K  E  L  I  N  R  V  33

          ---------|---------|---------|---------|---------|
      101 TTTAACTATTTGCAGATCAAAGGCTTAGCCTGACAATCAACTTTAGGGGG 150
F1      1   *  L  F  A  D  Q  R  L  S  L  T  I  N  F  R  G   15
F2      2 F  N  Y  L  Q  I  K  G  L  A  *  Q  S  T  L  G  G  6
F3     34  L  T  I  C  R  S  K  A  *  P  D  N  Q  L  *  G  G 2

          ---------|---------|---------|---------|---------|
      151 GTTCCATGATAGATGACACTGTCAGGTGTTGTGACAGCATGTGCAATAAG 200
F1     16 V  P  *  *  M  T  L  S  G  V  V  T  A  C  A  I  S  13
F2      7  F  H  D  R  *  H  C  Q  V  L  *  Q  H  V  Q  *  V 1
F3      3   S  M  I  D  D  T  V  R  C  C  D  S  M  C  N  K   18

          ---------|---------|---------|---------|---------|
      201 TTACCCCCAAAGTGAAATCTAAGTCGGCATGTCCACATGACTGACAAGCA 250
F1     14  Y  P  Q  S  E  I  *  V  G  M  S  T  *  L  T  S  I 4
F2      2   T  P  K  V  K  S  K  S  A  C  P  H  D  *  Q  A   2
F3     19 L  P  P  K  *  N  L  S  R  H  V  H  M  T  D  K  H  12

          ---------|---------|---------|---------|---------|
      251 TTGTGTCAAATATGTTTTGCGACTCAGTCATTTGGGATATTTTGGAACCC 300
F1      5   V  S  N  M  F  C  D  S  V  I  W  D  I  L  E  P   20
F2      3 L  C  Q  I  C  F  A  T  Q  S  F  G  I  F  W  N  P  19
F3     13  C  V  K  Y  V  L  R  L  S  H  L  G  Y  F  G  T  Q 29

          ---------|---------|---------|---------|---------|
      301 AGTACGTACCTAACATGTGTACAACACCCAGCTGTGAAGTATAAAAACAA 350
F1     21 S  T  Y  L  T  C  V  Q  H  P  A  V  K  Y  K  N  N  37
F2     20  V  R  T  *  H  V  Y  N  T  Q  L  *  S  I  K  T  M 5
F3     30   Y  V  P  N  M  C  T  T  P  S  C  E  V  *  K  Q   2

          ---------|---------|---------|---------|---------|
      351 TGGAGTCGCTTGCAGAAAACAAGAGTCTGAATCTGAAATCATCAAGCAGG 400
F1     38  G  V  A  C  R  K  Q  E  S  E  S  E  I  I  K  Q  V 54
F2      6   E  S  L  A  E  N  K  S  L  N  L  K  S  S  S  R   21
F3      3 W  S  R  L  Q  K  T  R  V  *  I  *  N  H  Q  A  G  5

          ---------|---------|---------|---------|---------|
      401 TTCTTAGTTCGCTGTGCGTACATCTCAGTCTACCAAATCATACGAATTGC 450
F1     55   L  S  S  L  C  V  H  L  S  L  P  N  H  T  N  C   70
F2     22 F  L  V  R  C  A  Y  I  S  V  Y  Q  I  I  R  I  A  38
F3      6  S  *  F  A  V  R  T  S  Q  S  T  K  S  Y  E  L  R 15
```

```
         ---------|---------|---------|---------|---------|
       451 GTTCTTACAGTCAATCTCTAGCATTCCATTTATAATCACCTTCCGTCAAC 500
F1      71 V  L  T  V  N  L  *  H  S  I  Y  N  H  L  P  S  T  10
F2      39  F  L  Q  S  I  S  S  I  P  F  I  I  T  F  R  Q  H 55
F3      16   S  Y  S  Q  S  L  A  F  H  L  *  S  P  S  V  N   5

         ---------|---------|---------|---------|---------|
       501 ATGCGTACCTCTACGACCATTTTCGCATCGCTGAGTCTTTTAGCTGGCCT 550
F1      11  C  V  P  L  R  P  F  S  H  R  *  V  F  *  L  A  W 3
F2      56   A  Y  L  Y  D  H  F  R  I  A  E  S  F  S  W  P   71
F3       6 M  R  T  S  T  T  I  F  A  S  L  S  L  L  A  G  L  22

         ---------|---------|---------|---------|---------|
       551 GGTTTCCGCCCAGGATGTAGAGATACAGCCGGATTTCCAAGGAGTTATGA 600
F1       4   F  P  P  R  M  *  R  Y  S  R  I  S  K  E  L  *   9
F2      72 G  F  R  P  G  C  R  D  T  A  G  F  P  R  S  Y  D  88
F3      23  V  S  A  Q  D  V  E  I  Q  P  D  F  Q  G  V  M  T 39

         ---------|---------|---------|---------|---------|
       601 CTGGAGGAATTTGTTGCGGCCCAACTCCCTGCCCAGATCCCAGTGGACAA 650
F1       1 L  E  E  F  V  A  A  Q  L  P  A  Q  I  P  V  D  N  17
F2      89  W  R  N  L  L  R  P  N  S  L  P  R  S  Q  W  T  M 105
F3      40   G  G  I  C  C  G  P  T  P  C  P  D  P  S  G  Q   55

         ---------|---------|---------|---------|---------|
       651 TGTGCCAAGGCCAAGCTGACCCCTTACTGCGTAAGTTCCCTGTTTCTGTT 700
F1      18  V  P  R  P  S  *  P  L  T  A  *  V  P  C  F  C  S 6
F2     106   C  Q  G  Q  A  D  P  L  L  R  K  F  P  V  S  V   121
F3      56 C  A  K  A  K  L  T  P  Y  C  V  S  S  L  F  L  F  72

         ---------|---------|---------|---------|---------|
       701 CCTTGTCTTTTTATTCTTCTTTTTTTTGTTCGTCTGGCAAAGCAAATGTC 750
F1       7   L  S  F  Y  S  S  F  F  C  S  S  G  K  A  N  V   22
F2     122 P  C  L  F  I  L  L  F  F  V  R  L  A  K  Q  M  S  138
F3      73  L  V  F  L  F  F  F  F  L  F  V  W  Q  S  K  C  L 89

         ---------|---------|---------|---------|---------|
       751 TTGGATGCGCGAATCTAATACTTTCTTTAGTGCGGTCCTTTCTTTAACAA 800
F1      23 L  D  A  R  I  *  Y  F  L  *  C  G  P  F  F  N  N  7
F2     139  W  M  R  E  S  N  T  F  F  S  A  V  L  S  L  T  T 155
F3      90   G  C  A  N  L  I  L  S  L  V  R  S  F  L  *  Q   1

         ---------|---------|---------|---------|---------|
       801 CCGAAAGAAAACAAAAAATACCAAGGGTGGATGTGACCCGTTTACGACCA 850
F1       8  R  K  K  T  K  N  T  K  G  G  C  D  P  F  T  T  T 24
F2     156   E  R  K  Q  K  I  P  R  V  D  V  T  R  L  R  P   171
F3       2 P  K  E  N  K  K  Y  Q  G  W  M  *  P  V  Y  D  H  5

         ---------|---------|---------|---------|---------|
       851 CTTTCCCTGTTGGCCGCCTCGTTAAGACCTTTCCCTCCACCTTGCAAGAC 900
F1      25   F  P  V  G  R  L  V  K  T  F  P  S  T  L  Q  D   40
F2     172 L  S  L  L  A  A  S  L  R  P  F  P  P  P  C  K  T  188
F3       6  F  P  C  W  P  P  R  *  D  L  S  L  H  L  A  R  L 9

         ---------|---------|---------|---------|---------|
       901 TGCAGGTCCAATGGGGTTCCTGGGTTTGTTGGATGCGTCTGAACTGAAGG 950
F1      41 C  R  S  N  G  V  P  G  F  V  G  C  V  *  T  E  G  3
F2     189  A  G  P  M  G  F  L  G  L  L  D  A  S  E  L  K  G 205
```

```
F3      10   Q  V  Q  W  G  S  W  V  C  W  M  R  L  N  *  R   1

          ---------|---------|---------|---------|---------|
        951 GGATCTCCATTTACTGTGGATGAGATCAGAGGATATTATGACCATGAAAA 1000
F1        4  D  L  H  L  L  W  M  R  S  E  D  I  M  T  M  K  M 20
F2      206   I  S  I  Y  C  G  *  D  Q  R  I  L  *  P  *  K   1
F3        2 G  S  P  F  T  V  D  E  I  R  G  Y  Y  D  H  E  N  18
```

Three frame translation of MG02147.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG05403 from 1 to 950

```
           ---------|---------|---------|---------|---------|
         1 AAATAATTATCCACGTTCAGGGACATGGCTTTTAACATGATCAAATCGCT 50
F1       1 K  *  L  S  T  F  R  D  M  A  F  N  M  I  K  S  L  15
F2       1  N  N  Y  P  R  S  G  T  W  L  L  T  *  S  N  R  Y 4
F3       1   I  I  I  H  V  Q  G  H  G  F  *  H  D  Q  I  A   5

           ---------|---------|---------|---------|---------|
        51 ATACTCTTCAAATAGAAAATAAAAAGTATATAAACGGCACCACATAACGA 100
F1      16  Y  S  S  N  R  K  *  K  V  Y  K  R  H  H  I  T  R 10
F2       5   T  L  Q  I  E  N  K  K  Y  I  N  G  T  T  *  R   1
F3       6 I  L  F  K  *  K  I  K  S  I  *  T  A  P  H  N  E  6

           ---------|---------|---------|---------|---------|
       101 GGTTTGTGTTGATTTTGTTCTTTACAGCGGCAGCGAATTCTTGACGAGTA 150
F1      11   F  V  L  I  L  F  F  T  A  A  A  N  S  *  R  V   2
F2       2 G  L  C  *  F  C  S  L  Q  R  Q  R  I  L  D  E  *  12
F3       7  V  C  V  D  F  V  L  Y  S  G  S  E  F  L  T  S  S 23

           ---------|---------|---------|---------|---------|
       151 GCAGCACCCCCTCGAGGCTTCTGAGGTTCTTGACGAACAAAGCTTTGTCT 200
F1       3 A  A  P  P  R  G  F  *  G  S  *  R  T  K  L  C  L  6
F2       1  Q  H  P  L  E  A  S  E  V  L  D  E  Q  S  F  V  F 17
F3      24   S  T  P  S  R  L  L  R  F  L  T  N  K  A  L  S   39

           ---------|---------|---------|---------|---------|
       201 TCTTTCGAGAATATCACACCTTTTTTTTCCCGACTTGCTATTTTCGTTCC 250
F1       7  L  S  R  I  S  H  L  F  F  P  D  L  L  F  S  F  P 23
F2      18   F  R  E  Y  H  T  F  F  F  P  T  C  Y  F  R  S   33
F3      40 S  F  E  N  I  T  P  F  F  S  R  L  A  I  F  V  P  56

           ---------|---------|---------|---------|---------|
       251 CATCTCCCCCAAAACCACTACTCGGTTCGGGCATTTGAATCAAAAAAAAA 300
F1      24   S  P  P  K  P  L  L  G  S  G  I  *  I  K  K  K   4
F2      34 H  L  P  Q  N  H  Y  S  V  R  A  F  E  S  K  K  K  50
F3      57  I  S  P  K  T  T  T  R  F  G  H  L  N  Q  K  K  N 73

           ---------|---------|---------|---------|---------|
       301 ATGCGCACCTCTTATTTCGCCCTCCTCCTGGCTCCCACCGCCGTCCTATC 350
F1       5 M  R  T  S  Y  F  A  L  L  L  A  P  T  A  V  L  S  21
F2      51  C  A  P  L  I  S  P  S  S  W  L  P  P  P  S  Y  P 67
F3      74   A  H  L  L  F  R  P  P  P  G  S  H  R  R  P  I   89

           ---------|---------|---------|---------|---------|
       351 CCGCAGGATCGTGATCCAACCCACGACCGGAGACTTGTGCTGCGACCGGG 400
F1      22  R  R  I  V  I  Q  P  T  T  G  D  L  C  C  D  R  G 38
F2      68   A  G  S  *  S  N  P  R  P  E  T  C  A  A  T  G   12
F3      90 P  Q  D  R  D  P  T  H  D  R  R  L  V  L  R  P  G  106

           ---------|---------|---------|---------|---------|
       401 GCACACCAGACGACAGCGAAACCTGCAAGAAGCAAGGCTTGAACTCTTAC 450
F1      39   T  P  D  D  S  E  T  C  K  K  Q  G  L  N  S  Y   54
F2      13 A  H  Q  T  T  A  K  P  A  R  S  K  A  *  T  L  T  3
F3     107  H  T  R  R  Q  R  N  L  Q  E  A  R  L  E  L  L  L 123

           ---------|---------|---------|---------|---------|
       451 TGCGTGAGTTGTGTAAACCAAAGTCCTACAGACCCAAGTGGCAAAGGGCA 500
```

```
F1      55 C  V  S  C  V  N  Q  S  P  T  D  P  S  G  K  G  Q  71
F2       4  A  *  V  V  *  T  K  V  L  Q  T  Q  V  A  K  G  K 12
F3     124   R  E  L  C  K  P  K  S  Y  R  P  K  W  Q  R  A  139

          ---------|---------|---------|---------|---------|
       501 AAAGGGAAAACAAGATTAACATAATTTGCTGCAGTGTTCCCAAGCTCGCA 550
F1      72  K  G  K  Q  D  *  H  N  L  L  Q  C  S  Q  A  R  N 11
F2      13   R  E  N  K  I  N  I  I  C  C  S  V  P  K  L  A   28
F3     140 K  G  K  T  R  L  T  *  F  A  A  V  F  P  S  S  Q  9

          ---------|---------|---------|---------|---------|
       551 ATAACAACCGCGGCGGGTGCGACCCGGAAAAACTAGAAATCTTCAACTTT 600
F1      12   N  N  R  G  G  C  D  P  E  K  L  E  I  F  N  F   27
F2      29 I  T  T  A  A  G  A  T  R  K  N  *  K  S  S  T  L  5
F3       1  *  Q  P  R  R  V  R  P  G  K  T  R  N  L  Q  L  W 16

          ---------|---------|---------|---------|---------|
       601 GGGCGCTCGGTCACGTCCTTTGTTCCGGGAGGCACTTGCGAAAGGCGCGA 650
F1      28 G  R  S  V  T  S  F  V  P  G  G  T  C  E  R  R  D  44
F2       6  G  A  R  S  R  P  L  F  R  E  A  L  A  K  G  A  T 22
F3      17   A  L  G  H  V  L  C  S  G  R  H  L  R  K  A  R   32

          ---------|---------|---------|---------|---------|
       651 CTCGGCTGGTAATACGTTTGTGGGATTCATTGGTTGTGCAAAGTGATTTT 700
F1      45  S  A  G  N  T  F  V  G  F  I  G  C  A  K  *  F  W 2
F2      23   R  L  V  I  R  L  W  D  S  L  V  V  Q  S  D  F   38
F3      33 L  G  W  *  Y  V  C  G  I  H  W  L  C  K  V  I  L  13

          ---------|---------|---------|---------|---------|
       701 GGTTGAATCGGCGAAATGAATGGTTGAACCGGGGTTTTGGCTAGTCTTTT 750
F1       3   L  N  R  R  N  E  W  L  N  R  G  F  G  *  S  F   2
F2      39 G  *  I  G  E  M  N  G  *  T  G  V  L  A  S  L  L  8
F3      14  V  E  S  A  K  *  M  V  E  P  G  F  W  L  V  F  C 11

          ---------|---------|---------|---------|---------|
       751 GTATAAATTATTAAGCCTTTATTGCGACGGAGCGAACAACAAGAGGTCTG 800
F1       3 V  *  I  I  K  P  L  L  R  R  S  E  Q  Q  E  V  W  15
F2       9  Y  K  L  L  S  L  Y  C  D  G  A  N  N  K  R  S  G 25
F3      12   I  N  Y  *  A  F  I  A  T  E  R  T  T  R  G  L   12

          ---------|---------|---------|---------|---------|
       801 GCGTCAACAACCGGTTATTTTGCTCCGAAAACTTACATCTAGCGCCTGAT 850
F1      16  R  Q  Q  P  V  I  L  L  R  K  L  T  S  S  A  *  C 1
F2      26   V  N  N  R  L  F  C  S  E  N  L  H  L  A  P  D   41
F3      13 A  S  T  T  G  Y  F  A  P  K  T  Y  I  *  R  L  M  3

          ---------|---------|---------|---------|---------|
       851 GTTCCATAGTACTTTTGGAGCGGGGGTCGCTGATCCTACGGGGAGGCGCC 900
F1       2   S  I  V  L  L  E  R  G  S  L  I  L  R  G  G  A   17
F2      42 V  P  *  Y  F  W  S  G  G  R  *  S  Y  G  E  A  P  6
F3       4  F  H  S  T  F  G  A  G  V  A  D  P  T  G  R  R  P 20
```

Three frame translation of MG05403.4. Exonic regions are underlined. The start codon is located at nucleotide position 301. Intron consensus sequences are underlined.

MG05560 SHOWORF of Magnaporthe from 1 to 1300

```
          ---------|---------|---------|---------|---------|
        1 AGGCATTTCGTTGTTAGATTTTCAGGGTGCCGAACAGGGGGGGTGCCGAA 50
F1      1 R  H  F  V  V  R  F  S  G  C  R  T  G  G  V  P  N  17
F2      1  G  I  S  L  L  D  F  Q  G  A  E  Q  G  G  C  R  T 17
F3      1   A  F  R  C  *  I  F  R  V  P  N  R  G  G  A  E   11

          ---------|---------|---------|---------|---------|
       51 CAGGGGGTACGCAACGTTAAAATGAAACTGCCAATCATACAAGCTGATAA 100
F1     18  R  G  Y  A  T  L  K  *  N  C  Q  S  Y  K  L  I  K 9
F2     18   G  G  T  Q  R  *  N  E  T  A  N  H  T  S  *  *   8
F3     12 Q  G  V  R  N  V  K  M  K  L  P  I  I  Q  A  D  K  28

          ---------|---------|---------|---------|---------|
      101 AATTATCCCCCTGGTGCCAGGTGGAGCTCTTTGAAAGGAAATCTTACATT 150
F1     10   L  S  P  W  C  Q  V  E  L  F  E  R  K  S  Y  I   25
F2      1 N  Y  P  P  G  A  R  W  S  S  L  K  G  N  L  T  L  17
F3     29  I  I  P  L  V  P  G  G  A  L  *  K  E  I  L  H  C 6

          ---------|---------|---------|---------|---------|
      151 GTCTAATACCGCTGTTATGAGACAGTAGGAAATTTGATTTATTTCTTCGT 200
F1     26 V  *  Y  R  C  Y  E  T  V  G  N  L  I  Y  F  F  V  15
F2     18  S  N  T  A  V  M  R  Q  *  E  I  *  F  I  S  S  S 5
F3      7   L  I  P  L  L  *  D  S  R  K  F  D  L  F  L  R   10

          ---------|---------|---------|---------|---------|
      201 CATTATTTATATTGGCCCTTGAAGCCAGGTGTTCTAGACGTTTCTCGAGC 250
F1     16  I  I  Y  I  G  P  *  S  Q  V  F  *  T  F  L  E  P 5
F2      6   L  F  I  L  A  L  E  A  R  C  S  R  R  F  S  S   21
F3     11 H  Y  L  Y  W  P  L  K  P  G  V  L  D  V  S  R  A  27

          ---------|---------|---------|---------|---------|
      251 CCCTCTTCTCATTAGCCCCCTTTGACATTTTTCAGCAACAGTTCTTGACG 300
F1      6   L  F  S  L  A  P  F  D  I  F  Q  Q  Q  F  L  T   21
F2     22 P  S  S  H  *  P  P  L  T  F  F  S  N  S  S  *  R  1
F3     28  P  L  L  I  S  P  L  *  H  F  S  A  T  V  L  D  E 9

          ---------|---------|---------|---------|---------|
      301 AACTCTTATTATTTTTTTTTCACCTCCCTGTCAGAAGCCGCCACTGTCTT 350
F1     22 N  S  Y  Y  F  F  F  T  S  L  S  E  A  A  T  V  F  38
F2      2  T  L  I  I  F  F  S  P  P  C  Q  K  P  P  L  S  L 18
F3     10   L  L  L  F  F  F  H  L  P  V  R  S  R  H  C  L   25

          ---------|---------|---------|---------|---------|
      351 TAACATTTATAAAATCTGCTTGGTCAAATTTCAAATAAGAACAGAAACAC 400
F1     39  N  I  Y  K  I  C  L  V  K  F  Q  I  R  T  E  T  Q 55
F2     19   T  F  I  K  S  A  W  S  N  F  K  *  E  Q  K  H   4
F3      1 *  H  L  *  N  L  L  G  Q  I  S  N  K  N  R  N  T  13

          ---------|---------|---------|---------|---------|
      401 AACAAAAATGCATGCCACCTTCATCACCTTTCTCATAGCCCCTCTCGTCG 450
F1     56   Q  K  C  M  P  P  S  S  P  F  S  *  P  L  S  S   4
F2      5 N  K  N  A  C  H  L  H  H  L  S  H  S  P  S  R  R  21
F3     14  T  K  M  H  A  T  F  I  T  F  L  I  A  P  L  V  V 30

          ---------|---------|---------|---------|---------|
      451 TTTTGGGTGCTTCATCCGTCACGGTTGGTTTTGGTTCGACTGGTGGGCGA 500
```

```
F1       5 F  W  V  L  H  P  S  R  L  V  L  V  R  L  V  G  D  21
F2      22  F  G  C  F  I  R  H  G  W  F  W  F  D  W  W  A  M 38
F3      31   L  G  A  S  S  V  T  V  G  F  G  S  T  G  G  R   46

         ---------|---------|---------|---------|---------|
       501 TGCTGCCGTGATGGTGTTGCGGACCCCTCCAATACCTGCAAAAACCTAGG 550
F1      22  A  A  V  M  V  L  R  T  P  P  I  P  A  K  T  *  V 1
F2      39   L  P  *  W  C  C  G  P  L  Q  Y  L  Q  K  P  R   13
F3      47 C  C  R  D  G  V  A  D  P  S  N  T  C  K  N  L  G  63

         ---------|---------|---------|---------|---------|
       551 TTTGAATTCTTATGCAGTAAGTGCTTTTGATCCTTCCAACAGGAAACAGG 600
F1       1   *  I  L  M  Q  *  V  L  L  I  L  P  T  G  N  R   10
F2      14 F  E  F  L  C  S  K  C  F  *  S  F  Q  Q  E  T  G  7
F3      64  L  N  S  Y  A  V  S  A  F  D  P  S  N  R  K  Q  E 80

         ---------|---------|---------|---------|---------|
       601 AGGCTTACCGGTTCCTCTTTTCGTCTTTTCTTTTTCTCCTCCTGCAGTGC 650
F1      11 R  L  T  G  S  S  F  R  L  F  F  F  S  S  C  S  A  27
F2       8  G  L  P  V  P  L  F  V  F  S  F  S  P  P  A  V  Q 24
F3      81   A  Y  R  F  L  F  S  S  F  L  F  L  L  L  Q  C   96

         ---------|---------|---------|---------|---------|
       651 AGCGACCACAGTAGCTCTGCTCCTAATGAGCCTGGACCCAAGTTTAGCGA 700
F1      28  A  T  T  V  A  L  L  L  M  S  L  D  P  S  L  A  T 44
F2      25   R  P  Q  *  L  C  S  *  *  A  W  T  Q  V  *  R   1
F3      97 S  D  H  S  S  S  A  P  N  E  P  G  P  K  F  S  D  113

         ---------|---------|---------|---------|---------|
       701 CAAGCCCAAAGGTGGCTGCGATCAGCCAGAGATTCATAACTTCCCGACAG 750
F1      45   S  P  K  V  A  A  I  S  Q  R  F  I  T  S  R  Q   60
F2       2 Q  A  Q  R  W  L  R  S  A  R  D  S  *  L  P  D  R  4
F3     114  K  P  K  G  G  C  D  Q  P  E  I  H  N  F  P  T  G 130

         ---------|---------|---------|---------|---------|
       751 GACGCGATGTGAAAACTTTCGTTGTCGGGTCCACTGTAAATGTCAGACGC 800
F1      61 D  A  M  *  K  L  S  L  S  G  P  L  *  M  S  D  A  4
F2       5  T  R  C  E  N  F  R  C  R  V  H  C  K  C  Q  T  Q 21
F3     131   R  D  V  K  T  F  V  V  G  S  T  V  N  V  R  R   146

         ---------|---------|---------|---------|---------|
       801 AGCAACGGGCAACATCGAGGTTGGATTCATCGGATGCGCGGCATGATAAA 850
F1       5  A  T  G  N  I  E  V  G  F  I  G  C  A  A  *  *  T 1
F2      22   Q  R  A  T  S  R  L  D  S  S  D  A  R  H  D  K   37
F3     147 S  N  G  Q  H  R  G  W  I  H  R  M  R  G  M  I  N  163

         ---------|---------|---------|---------|---------|
       851 CTGTAACAAGGCCCAAAGCAGAGCCGGGGACTCTGGAGAGGGGGCATTGA 900
F1       2   V  T  R  P  K  A  E  P  G  T  L  E  R  G  H  *   16
F2      38 L  *  Q  G  P  K  Q  S  R  G  L  W  R  G  G  I  E  15
F3     164  C  N  K  A  Q  S  R  A  G  D  S  G  E  G  A  L  N 180

         ---------|---------|---------|---------|---------|
       901 ACGGTGTCATTAATTAGAATTTTTGCGGAGATACATAGGTACCTAATCA 950
F1       1 T  V  S  L  I  R  I  F  C  G  D  T  *  V  P  N  Q  4
F2      16  R  C  H  *  L  E  F  F  A  E  I  H  R  Y  L  I  S 13
F3     181   G  V  I  N  *  N  F  L  R  R  Y  I  G  T  *  S   1
```

```
          ---------|---------|---------|---------|---------|
       951 GCAAGAAGGATCCCACTGCTGCCGTCAGAGATCAACCTCTGTTTAATAGT 1000
F1       5  Q  E  G  S  H  C  C  R  Q  R  S  T  S  V  *  *  S 1
F2      14   K  K  D  P  T  A  A  V  R  D  Q  P  L  F  N  S   29
F3       2 A  R  R  I  P  L  L  P  S  E  I  N  L  C  L  I  V  18

          ---------|---------|---------|---------|---------|
      1001 CAGTGTTTCCGGGTCCAATCCTTAGGGATCTCGGCAACTGTAAACTTGCC 1050
F1       2   V  F  P  G  P  I  L  R  D  L  G  N  C  K  L  A   17
F2      30 Q  C  F  R  V  Q  S  L  G  I  S  A  T  V  N  L  P  46
F3      19  S  V  S  G  S  N  P  *  G  S  R  Q  L  *  T  C  L 3

          ---------|---------|---------|---------|---------|
      1051 TGCCACTATACTTGACCCAGAAACGCAAGATAAGTATACATTAAGTCATA 1100
F1      18 C  H  Y  T  *  P  R  N  A  R  *  V  Y  I  K  S  Y  6
F2      47  A  T  I  L  D  P  E  T  Q  D  K  Y  T  L  S  H  I 63
F3       4   P  L  Y  L  T  Q  K  R  K  I  S  I  H  *  V  I   2

          ---------|---------|---------|---------|---------|
      1101 TTTGATCCCCCACATCTGGGGAGACATCAACAAAAGACGTCTACCTGCGC 1150
F1       7  L  I  P  H  I  W  G  D  I  N  K  R  R  L  P  A  P 23
F2       1   *  S  P  T  S  G  E  T  S  T  K  D  V  Y  L  R   15
F3       3 F  D  P  P  H  L  G  R  H  Q  Q  K  T  S  T  C  A  19

          ---------|---------|---------|---------|---------|
      1151 CAAAAAGCCATGCCACATAAAGGCCAGATGAACCACAAATCCATGAGCCA 1200
F1      24   K  S  H  A  T  *  R  P  D  E  P  Q  I  H  E  P   10
F2      16 Q  K  A  M  P  H  K  G  Q  M  N  H  K  S  M  S  Q  32
F3      20  K  K  P  C  H  I  K  A  R  *  T  T  N  P  *  A  N 2

          ---------|---------|---------|---------|---------|
      1201 ACTACCAAGCGTCCACACGTTCCACTCCAGCCAGCCTATAAAAGGAAATG 1250
F1      11 T  T  K  R  P  H  V  P  L  Q  P  A  Y  K  R  K  C  27
F2      33  L  P  S  V  H  T  F  H  S  S  Q  P  I  K  G  N  A 49
F3       3   Y  Q  A  S  T  R  S  T  P  A  S  L  *  K  E  M   3

          ---------|---------|---------|---------|---------|
      1251 CAGGGTTGGCTATAGCTGTGCGAGAATCGAAAAAGTCATGTCACCTCTTG 1300
F1      28  R  V  G  Y  S  C  A  R  I  E  K  V  M  S  P  L  G 44
F2      50   G  L  A  I  A  V  R  E  S  K  K  S  C  H  L  L   65
F3       4 Q  G  W  L  *  L  C  E  N  R  K  S  H  V  T  S  W  12
```

Three frame translation of MG05560.4. Exonic regions are underlined. The start codon is located at nucleotide position 408. Intron consensus sequences are underlined. Deletion of an A residue at position 790 is was used to generate a protein coding region to align with other members of the gene family.

```
MG06253 from 1 to 1150

          ---------|---------|---------|---------|---------|
        1 TTGGTTCAAGAAATAGAAATAACCACCAGCTGCTTCCCATTGAGACTCGA 50
F1      1 L  V  Q  E  I  E  I  T  T  S  C  F  P  L  R  L  D  17
F2      1  W  F  K  K  *  K  *  P  P  A  A  S  H  *  D  S  I 3
F3      1   G  S  R  N  R  N  N  H  Q  L  L  P  I  E  T  R   16

          ---------|---------|---------|---------|---------|
       51 TCCGGGACCTTTTGGCCGCGGACATCGCTAAACCGTAATAATAAATATTT 100
F1     18  P  G  P  F  G  R  G  H  R  *  T  V  I  I  N  I  * 6
F2      4   R  D  L  L  A  A  D  I  A  K  P  *  *  *  I  F   2
F3     17 S  G  T  F  W  P  R  T  S  L  N  R  N  N  K  Y  L  33

          ---------|---------|---------|---------|---------|
      101 AAAGGCGTGAACATGTATTATGAGCATACTTTTGTTTCCAAATATACAAA 150
F1      1   R  R  E  H  V  L  *  A  Y  F  C  F  Q  I  Y  K   9
F2      3 K  G  V  N  M  Y  Y  E  H  T  F  V  S  K  Y  T  K  19
F3     34  K  A  *  T  C  I  M  S  I  L  L  F  P  N  I  Q  K 14

          ---------|---------|---------|---------|---------|
      151 AAAAATGTATATTCAGCGCCCAAGTGCCACTTTTGAAACATTTTGTAACA 200
F1     10 K  N  V  Y  S  A  P  K  C  H  F  *  N  I  L  *  H  1
F2     20  K  M  Y  I  Q  R  P  S  A  T  F  E  T  F  C  N  T 36
F3     15   K  C  I  F  S  A  Q  V  P  L  L  K  H  F  V  T   30

          ---------|---------|---------|---------|---------|
      201 CTTCTCACACGCTACACTGTCAATTTGCCGCTGGATGTGGTAATTCCTTT 250
F1      2  F  S  H  A  T  L  S  I  C  R  W  M  W  *  F  L  C 3
F2     37   S  H  T  L  H  C  Q  F  A  A  G  C  G  N  S  F   52
F3     31 L  L  T  R  Y  T  V  N  L  P  L  D  V  V  I  P  L  47

          ---------|---------|---------|---------|---------|
      251 GCAACCTACCCACACAACATTCCTTGTTCTCTACCTACCTACCTATGTAA 300
F1      4   N  L  P  T  Q  H  S  L  F  S  T  Y  L  P  M  *   18
F2     53 A  T  Y  P  H  N  I  P  C  S  L  P  T  Y  L  C  K  69
F3     48  Q  P  T  H  T  T  F  L  V  L  Y  L  P  T  Y  V  N 64

          ---------|---------|---------|---------|---------|
      301 ATCTTGTACATCCTTTTTCAGCATCTTACATTCAGTTTAAATAGGTAGAC 350
F1      1 I  L  Y  I  L  F  Q  H  L  T  F  S  L  N  R  *  T  1
F2     70  S  C  T  S  F  F  S  I  L  H  S  V  *  I  G  R  H 4
F3     65   L  V  H  P  F  S  A  S  Y  I  Q  F  K  *  V  D   2

          ---------|---------|---------|---------|---------|
      351 ATCCCTTTAATTATTTCTAGAAATACTTGCCATGATATCCAACTCTTGAA 400
F1      2  S  L  *  L  F  L  E  I  L  A  M  I  S  N  S  *  T 1
F2      5   P  F  N  Y  F  *  K  Y  L  P  *  Y  P  T  L  E   5
F3      3 I  P  L  I  I  S  R  N  T  C  H  D  I  Q  L  L  N  19

          ---------|---------|---------|---------|---------|
      401 CTTGTAACTACTACTCCTTTTCATATTTTCTTTCCCCCGGCTGTGTATTT 450
F1      2   C  N  Y  Y  S  F  S  Y  F  L  S  P  G  C  V  F   17
F2      6 L  V  T  T  T  P  F  H  I  F  F  P  P  A  V  Y  L  22
F3     20  L  *  L  L  L  L  F  I  F  S  F  P  R  L  C  I  C 15

          ---------|---------|---------|---------|---------|
      451 GTTGATCGTTTAGCCGGCTAAACAAAAAAAAAAAAAAAAAAAAAAGAAA 500
```

```
F1      18 V  D  R  L  A  G  *  T  K  K  K  K  K  K  K  R  K  10
F2      23  L  I  V  *  P  A  K  Q  K  K  K  K  K  K  K  E  N 13
F3       1   *  S  F  S  R  L  N  K  K  K  K  K  K  K  K  K   15

         ---------|---------|---------|---------|---------|
       501 ATGCGCTCATCTCTTTTATTCTTTATGCTCTCGGTCACTGTTTCCGCGCA 550
F1      11  C  A  H  L  F  Y  S  L  C  S  R  S  L  F  P  R  N 27
F2      14   A  L  I  S  F  I  L  Y  A  L  G  H  C  F  R  A   29
F3      16 M  R  S  S  L  L  F  F  M  L  S  V  T  V  S  A  Q  32

         ---------|---------|---------|---------|---------|
       551 ACCACCCCCGGTCAGGCCCGATGCTCCCATCCAACCCGAACTTGGTGGGC 600
F1      28   H  P  R  S  G  P  M  L  P  S  N  P  N  L  V  G   43
F2      30 T  T  P  G  Q  A  R  C  S  H  P  T  R  T  W  W  A  46
F3      33  P  P  P  V  R  P  D  A  P  I  Q  P  E  L  G  G  R 49

         ---------|---------|---------|---------|---------|
       601 GATGCTGCGCCAAGGAGGGGGTGGCAGACCCTACATTAACCTGCCAAAAG 650
F1      44 D  A  A  P  R  R  G  W  Q  T  L  H  *  P  A  K  R  4
F2      47  M  L  R  Q  G  G  G  G  R  P  Y  I  N  L  P  K  D 63
F3      50   C  C  A  K  E  G  V  A  D  P  T  L  T  C  Q  K   65

         ---------|---------|---------|---------|---------|
       651 ATGGGTTTAAACTCGTTTTGTGTGAGTTTTTGGCTGTCCTGTTACTGAGA 700
F1       5  W  V  *  T  R  F  V  *  V  F  G  C  P  V  T  E  S 9
F2      64   G  F  K  L  V  L  C  E  F  L  A  V  L  L  L  R   79
F3      66 M  G  L  N  S  F  C  V  S  F  W  L  S  C  Y  *  E  1

         ---------|---------|---------|---------|---------|
       701 GTGAGTGAAAGAAGAAAAAAGAATTGTAAAAATAATAATCTGTTTTGTTT 750
F1      10   E  *  K  K  K  K  E  L  *  K  *  *  S  V  L  F   4
F2      80 V  S  E  R  R  K  K  N  C  K  N  N  N  L  F  C  F  96
F3       1  *  V  K  E  E  K  R  I  V  K  I  I  I  C  F  V  L 16

         ---------|---------|---------|---------|---------|
       751 TGCAGTGCACCGGCCGGAGATCTTTTATTTCAAGGGGGTGCGACGGTGGG 800
F1       5 C  S  A  P  A  G  D  L  L  F  Q  G  G  A  T  V  G  21
F2      97  A  V  H  R  P  E  I  F  Y  F  K  G  V  R  R  W  D 113
F3      17   Q  C  T  G  R  R  S  F  I  S  R  G  C  D  G  G   32

         ---------|---------|---------|---------|---------|
       801 ACAGGAAACGAAGCGGTTGGGCGCCATGTTCAAGGCTTTCCACCTCAAAA 850
F1      22  Q  E  T  K  R  L  G  A  M  F  K  A  F  H  L  K  T 38
F2     114   R  K  R  S  G  W  A  P  C  S  R  L  S  T  S  K   129
F3      33 T  G  N  E  A  V  G  R  H  V  Q  G  F  P  P  Q  N  49

         ---------|---------|---------|---------|---------|
       851 CGGCGCTTGTGGTTTTACTGCGTTTATTGGATGCGCATAAATGAATCATT 900
F1      39   A  L  V  V  L  L  R  L  L  D  A  H  K  *  I  I   2
F2     130 R  R  L  W  F  Y  C  V  Y  W  M  R  I  N  E  S  F  146
F3      50  G  A  C  G  F  T  A  F  I  G  C  A  *  M  N  H  F 4

         ---------|---------|---------|---------|---------|
       901 TCATTCCACCCATTAAACGCTTGACTGACAATTACATGACGCGCGATATT 950
F1       3 S  F  H  P  L  N  A  *  L  T  I  T  *  R  A  I  L  4
F2     147  H  S  T  H  *  T  L  D  *  Q  L  H  D  A  R  Y  * 7
F3       5   I  P  P  I  K  R  L  T  D  N  Y  M  T  R  D  I   20
```

```
          ---------|---------|---------|---------|---------|
       951 AACTATCGGGGGCACTGCAATTGGATTTCACATAACAAAAGAACCGTTTA 1000
F1       5  T  I  G  G  T  A  I  G  F  H  I  T  K  E  P  F  I 21
F2       1   L  S  G  A  L  Q  L  D  F  T  *  Q  K  N  R  L   5
F3      21 N  Y  R  G  H  C  N  W  I  S  H  N  K  R  T  V  Y  37

          ---------|---------|---------|---------|---------|
      1001 TTAGTACGAACATCACCTACAGGTGTAAAGGACTTTGTTCGTTTTCCTCA 1050
F1      22   S  T  N  I  T  Y  R  C  K  G  L  C  S  F  S  S   37
F2       6 L  V  R  T  S  P  T  G  V  K  D  F  V  R  F  P  H  22
F3       1  *  Y  E  H  H  L  Q  V  *  R  T  L  F  V  F  L  I 8

          ---------|---------|---------|---------|---------|
      1051 TTTTCCTACAGCACCTTCTTCACCTAAATCCAACCGTACCTAAGGTACCT 1100
F1      38 F  S  Y  S  T  F  F  T  *  I  Q  P  Y  L  R  Y  L  8
F2      23  F  P  T  A  P  S  S  P  K  S  N  R  T  *  G  T  W 3
F3       9   F  L  Q  H  L  L  H  L  N  P  T  V  P  K  V  P   24

          ---------|---------|---------|---------|---------|
      1101 GGGCTACAACATGGGGGAGGGATGGATTCCAGATTGTTTGAAGCACTTGC 1150
F1       9  G  Y  N  M  G  E  G  W  I  P  D  C  L  K  H  L  P 25
F2       4   A  T  T  W  G  R  D  G  F  Q  I  V  *  S  T  C   3
F3      25 G  L  Q  H  G  G  G  M  D  S  R  L  F  E  A  L  A  41
```

Three frame translation of MG06253.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG07352 from 351 to 1400

```
         ---------|---------|---------|---------|---------|
       351 TGTGATCCAGTGTTTGACCCACCACATAGCTCTCATCGGAAATTTACGAG 400
F1      16  V  I  Q  C  L  T  H  H  I  A  L  I  G  N  L  R  A 32
F2       1   *  S  S  V  *  P  T  T  *  L  S  S  E  I  Y  E   7
F3       2 C  D  P  V  F  D  P  P  H  S  S  H  R  K  F  T  S  18

         ---------|---------|---------|---------|---------|
       401 CTTTGCAGTATCGTCTCGGGGTTTTCCGAATCTAGAGATCGATCAAATGC 450
F1      33   L  Q  Y  R  L  G  V  F  R  I  *  R  S  I  K  C   5
F2       8 L  C  S  I  V  S  G  F  S  E  S  R  D  R  S  N  A  24
F3      19  F  A  V  S  S  R  G  F  P  N  L  E  I  D  Q  M  L 35

         ---------|---------|---------|---------|---------|
       451 TTGTTCCGGGTTGATCTCGTGCTATAAATAGTTCGTCCGCTATCAAAACA 500
F1       6 L  F  R  V  D  L  V  L  *  I  V  R  P  L  S  K  Q   8
F2      25  C  S  G  L  I  S  C  Y  K  *  F  V  R  Y  Q  N  N 7
F3      36   V  P  G  *  S  R  A  I  N  S  S  S  A  I  K  T   12

         ---------|---------|---------|---------|---------|
       501 ACGTGCAACCCCGGCTGCACCTATGTTGGCAGGGGCTATGATAAAACTGC 550
F1       9  R  A  T  P  A  A  P  M  L  A  G  A  M  I  K  L  P 25
F2       8   V  Q  P  R  L  H  L  C  W  Q  G  L  *  *  N  C   2
F3      13 T  C  N  P  G  C  T  Y  V  G  R  G  Y  D  K  T  A  29

         ---------|---------|---------|---------|---------|
       551 CAATCATACATGATATCAACCTTATTATATATACCTTGGGACTGGAAATG 600
F1      26   I  I  H  D  I  N  L  I  I  Y  T  L  G  L  E  M   41
F2       3 Q  S  Y  M  I  S  T  L  L  Y  I  P  W  D  W  K  C  19
F3      30  N  H  T  *  Y  Q  P  Y  Y  I  Y  L  G  T  G  N  V 13

         ---------|---------|---------|---------|---------|
       601 TCACTATAATCTCAACCTTTAATGGTTATTTTCTTCTAGTTGCAGAGTAT 650
F1      42 S  L  *  S  Q  P  L  M  V  I  F  F  *  L  Q  S  I  4
F2      20  H  Y  N  L  N  L  *  W  L  F  S  S  S  C  R  V  F 10
F3      14   T  I  I  S  T  F  N  G  Y  F  L  L  V  A  E  Y   29

         ---------|---------|---------|---------|---------|
       651 TTGAGTAAGCCTCGAGCAACCTTTCAATGAGGAAAGCCTCCCGTGTGTAT 700
F1       1  *  V  S  L  E  Q  P  F  N  E  E  S  L  P  C  V  S 16
F2      11   E  *  A  S  S  N  L  S  M  R  K  A  S  R  V  Y   14
F3      30 L  S  K  P  R  A  T  F  Q  *  G  K  P  P  V  C  I  7

         ---------|---------|---------|---------|---------|
       701 CGGCAGCTTTCGATTTGTTTCATCTGATAGTTCTTGAGGAACTGTTCTTT 750
F1      17   A  A  F  D  L  F  H  L  I  V  L  E  E  L  F  F   32
F2      15 R  Q  L  S  I  C  F  I  *  *  F  L  R  N  C  S  F  7
F3       8  G  S  F  R  F  V  S  S  D  S  S  *  G  T  V  L  F 5

         ---------|---------|---------|---------|---------|
       751 TCTTTGTTCACTTCTTTTCAACCACCACTTTCGTTTATAATTCAACTTTG 800
F1      33 S  L  F  T  S  F  Q  P  P  L  S  F  I  I  Q  L  C  49
F2       8  L  C  S  L  L  F  N  H  H  F  R  L  *  F  N  F  A 4
F3       6   F  V  H  F  F  S  T  T  T  F  V  Y  N  S  T  L   21

         ---------|---------|---------|---------|---------|
       801 CTGGTCAAATTTCCAATACAAGATTAATAAACAACAAAAAAAATGCACGC 850
```

```
F1      50  W  S  N  F  Q  Y  K  I  N  K  Q  Q  K  K  C  T  P 66
F2       5   G  Q  I  S  N  T  R  L  I  N  N  K  K  N  A  R   20
F3      22 L  V  K  F  P  I  Q  D  *  *  T  T  K  K  M  H  A  7

         ---------|---------|---------|---------|---------|
       851 CTTCTCCCTTCTCTTCTTCATAGCCCCTCTTGCCGTCCTGTGTAACAACT 900
F1      67   S  P  F  S  S  S  *  P  L  L  P  S  C  V  T  T   9
F2      21 L  L  P  S  L  L  H  S  P  S  C  R  P  V  *  Q  L  2
F3       8  F  S  L  L  F  F  I  A  P  L  A  V  L  C  N  N  F 24

         ---------|---------|---------|---------|---------|
       901 TCACCGTCGGCAGGGGCTCGACTGGCGGGCGCTGCTGCGATCATGGCGTT 950
F1      10 S  P  S  A  G  A  R  L  A  G  A  A  A  I  M  A  L  26
F2       3  H  R  R  Q  G  L  D  W  R  A  L  L  R  S  W  R  C 19
F3      25   T  V  G  R  G  S  T  G  G  R  C  C  D  H  G  V   40

         ---------|---------|---------|---------|---------|
       951 GCAGACCCCTCTCGCACATGCTCCAAGATGAAGCTGAATTCCTACAGTGT 1000
F1      27  Q  T  P  L  A  H  A  P  R  *  S  *  I  P  T  V  * 4
F2      20   R  P  L  S  H  M  L  Q  D  E  A  E  F  L  Q  C   35
F3      41 A  D  P  S  R  T  C  S  K  M  K  L  N  S  Y  S  V  57

         ---------|---------|---------|---------|---------|
      1001 AAGTGTAATTTCTTTACGTTTTTTTTCATCATGACAAGGAGAAGAAAAAA 1050
F1       1   V  *  F  L  Y  V  F  F  H  H  D  K  E  K  K  K   14
F2      36 K  C  N  F  F  T  F  F  F  I  M  T  R  R  R  K  K  52
F3      58  S  V  I  S  L  R  F  F  S  S  *  Q  G  E  E  K  K 6

         ---------|---------|---------|---------|---------|
      1051 AAAAGAGACAAGAATGCGAGGCGAGTACACAATTAAACCCTTTCTTTTGT 1100
F1      15 K  R  D  K  N  A  R  R  V  H  N  *  T  L  S  F  V   5
F2      53  K  E  T  R  M  R  G  E  Y  T  I  K  P  F  L  L  * 68
F3       7   K  R  Q  E  C  E  A  S  T  Q  L  N  P  F  F  C   22

         ---------|---------|---------|---------|---------|
      1101 AGTGCATCGATTTCAGAAGCGACGCAAAAGCAGGTGACTCGGTCAACGAC 1150
F1       6  V  H  R  F  Q  K  R  R  K  S  R  *  L  G  Q  R  R 5
F2       1   C  I  D  F  R  S  D  A  K  A  G  D  S  V  N  D   16
F3      23 S  A  S  I  S  E  A  T  Q  K  Q  V  T  R  S  T  T  39

         ---------|---------|---------|---------|---------|
      1151 GTGGGCGGTGGCTGCGATCCAGTAGAGCTCCGCAACTGGCCCATTGGGCG 1200
F1       6   G  R  W  L  R  S  S  R  A  P  Q  L  A  H  W  A   21
F2      17 V  G  G  G  C  D  P  V  E  L  R  N  W  P  I  G  R  33
F3      40  W  A  V  A  A  I  Q  *  S  S  A  T  G  P  L  G  A 9

         ---------|---------|---------|---------|---------|
      1201 CGACGTCAAAGCATTCGTTCCCGGTAGCGTGGCGACGCACCAGACTTCAG 1250
F1      22 R  R  Q  S  I  R  S  R  *  R  G  D  A  P  D  F  R  8
F2      34  D  V  K  A  F  V  P  G  S  V  A  T  H  Q  T  S  D 50
F3      10   T  S  K  H  S  F  P  V  A  W  R  R  T  R  L  Q   25

         ---------|---------|---------|---------|---------|
      1251 ACTTTGATCTCGAGGTTGGCTTTATCGGATGTGCAGAATAAAGAATCAAA 1300
F1       9  L  *  S  R  G  W  L  Y  R  M  C  R  I  K  N  Q  N 15
F2      51   F  D  L  E  V  G  F  I  G  C  A  E  *  R  I  K   3
F3      26 T  L  I  S  R  L  A  L  S  D  V  Q  N  K  E  S  K  42
```

```
          ---------|---------|---------|---------|---------|
      1301 ATATCAGGGATGGTCCTTGTTAGATCGTAATCTAGAGGTGATGATGTTCC 1350
F1      16   I  R  D  G  P  C  *  I  V  I  *  R  *  *  C  S   2
F2       4 I  S  G  M  V  L  V  R  S  *  S  R  G  D  D  V  P  7
F3      43  Y  Q  G  W  S  L  L  D  R  N  L  E  V  M  M  F  H 59

          ---------|---------|---------|---------|---------|
      1351 ATTTCACTAGCTTGCTCCATTCTGTAATCTTGTTTTCGTCCGAAAGTGTA 1400
F1       3 I  S  L  A  C  S  I  L  *  S  C  F  R  P  K  V  Y  8
F2       8  F  H  *  L  A  P  F  C  N  L  V  F  V  R  K  C  I 14
F3      60   F  T  S  L  L  H  S  V  I  L  F  S  S  E  S  V   75
```

Three frame translation of MG07352.4. Exonic regions are underlined. The start codon is located at nucleotide position 843. Intron consensus sequences are underlined.

MG06592 from 1 to 1350

```
           ---------|---------|---------|---------|---------|
         1 AATCAGCCCAGGACGGAAAGCCCAGGACGGAAAGCCCAGGACGGAAACCC 50
F1       1 N  Q  P  R  T  E  S  P  G  R  K  A  Q  D  G  N  P  17
F2       1  I  S  P  G  R  K  A  Q  D  G  K  P  R  T  E  T  P 17
F3       1   S  A  Q  D  G  K  P  R  T  E  S  P  G  R  K  P   16

           ---------|---------|---------|---------|---------|
        51 CAGGAAACAGCCATCAAAAGACGCATTGAAGAATATCTATAAACAACCCT 100
F1      18  R  K  Q  P  S  K  D  A  L  K  N  I  Y  K  Q  P  S 34
F2      18   G  N  S  H  Q  K  T  H  *  R  I  S  I  N  N  P   7
F3      17 Q  E  T  A  I  K  R  R  I  E  E  Y  L  *  T  T  L  3

           ---------|---------|---------|---------|---------|
       101 CGTGCAACAGCCATTTATGTCGACTCCAGCACTATAATCTCTGGCGCCCT 150
F1      35   C  N  S  H  L  C  R  L  Q  H  Y  N  L  W  R  P   50
F2       8 R  A  T  A  I  Y  V  D  S  S  T  I  I  S  G  A  L  24
F3       4  V  Q  Q  P  F  M  S  T  P  A  L  *  S  L  A  P  L 5

           ---------|---------|---------|---------|---------|
       151 TGATATAACTATTCCCTATTCGAGAGTACATATACTCGTCATGTATAAAT 200
F1       1 *  Y  N  Y  S  L  F  E  S  T  Y  T  R  H  V  *  I  1
F2      25  D  I  T  I  P  Y  S  R  V  H  I  L  V  M  Y  K  * 40
F3       6   I  *  L  F  P  I  R  E  Y  I  Y  S  S  C  I  N   14

           ---------|---------|---------|---------|---------|
       201 AATCGGGCTGCTTGGCCGTTTCAGACAGTCCTATAGCTAATTGCCGGAAT 250
F1       2  I  G  L  L  G  R  F  R  Q  S  Y  S  *  L  P  E  F 4
F2       1   S  G  C  L  A  V  S  D  S  P  I  A  N  C  R  N   16
F3      15 N  R  A  A  W  P  F  Q  T  V  L  *  L  I  A  G  I  5

           ---------|---------|---------|---------|---------|
       251 TTCTTATGTCGATCAACAAATTCGAGTAATGCATACATCCATATATAGCC 300
F1       5   L  M  S  I  N  K  F  E  *  C  I  H  P  Y  I  A   7
F2      17 F  L  C  R  S  T  N  S  S  N  A  Y  I  H  I  *  P  1
F3       6  S  Y  V  D  Q  Q  I  R  V  M  H  T  S  I  Y  S  L 22

           ---------|---------|---------|---------|---------|
       301 TCTTGGGAAACCTGGTGACGCTCCAGTTTGTTTTACTGCCGCCCATGATA 350
F1       8 S  W  E  T  W  *  R  S  S  L  F  Y  C  R  P  *  *  9
F2       2  L  G  K  P  G  D  A  P  V  C  F  T  A  A  H  D  R 18
F3      23   L  G  N  L  V  T  L  Q  F  V  L  L  P  P  M  I   38

           ---------|---------|---------|---------|---------|
       351 GGTAGGTATGAAACCAGTTCACCCAGTTCAGACATTAATACCGACACTTA 400
F1       1  V  G  M  K  P  V  H  P  V  Q  T  L  I  P  T  L  N 17
F2       1   *  V  *  N  Q  F  T  Q  F  R  H  *  Y  R  H  L   4
F3      39 G  R  Y  E  T  S  S  P  S  S  D  I  N  T  D  T  *  54

           ---------|---------|---------|---------|---------|
       401 ACTATGCTTGGCTGGGCCAACAGTACACTGTAAAACAGAGCCCTCCCCTC 450
F1      18   Y  A  W  L  G  Q  Q  Y  T  V  K  Q  S  P  P  L   33
F2       5 T  M  L  G  W  A  N  S  T  L  *  N  R  A  L  P  S  6
F3       1  L  C  L  A  G  P  T  V  H  C  K  T  E  P  S  P  R 17

           ---------|---------|---------|---------|---------|
       451 GCTTTGCGCAGAATCCCCAACTTAATCCTGTAATTTTTTTCCAACCACAA 500
```

```
F1      34 A  L  R  R  I  P  N  L  I  L  *  F  F  S  N  H  N  6
F2       7  L  C  A  E  S  P  T  *  S  C  N  F  F  P  T  T  I 9
F3      18   F  A  Q  N  P  Q  L  N  P  V  I  F  F  Q  P  Q   33

         ---------|---------|---------|---------|---------|
       501 TTAATTTCTTGCTTTTTTGTCTCTCCAGGACAACTACCGTTAAAATGCGT 550
F1       1  *  F  L  A  F  L  S  L  Q  D  N  Y  R  *  N  A  C 3
F2      10   N  F  L  L  F  C  L  S  R  T  T  T  V  K  M  R   25
F3      34 L  I  S  C  F  F  V  S  P  G  Q  L  P  L  K  C  V  50

         ---------|---------|---------|---------|---------|
       551 GCATTTGCAAGTTTTTACTTATTCGCTGGCCTGGTTGCCGCCCAATTCAA 600
F1       4   I  C  K  F  L  L  I  R  W  P  G  C  R  P  I  Q   19
F2      26 A  F  A  S  F  Y  L  F  A  G  L  V  A  A  Q  F  N  42
F3      51  H  L  Q  V  F  T  Y  S  L  A  W  L  P  P  N  S  T 67

         ---------|---------|---------|---------|---------|
       601 CTCCGCAACCCCCGAGACAGGACTGCGCTGTTGCGGCCAGGGCACGACAG 650
F1      20 L  R  N  P  R  D  R  T  A  L  L  R  P  G  H  D  R  36
F2      43  S  A  T  P  E  T  G  L  R  C  C  G  Q  G  T  T  D 59
F3      68   P  Q  P  P  R  Q  D  C  A  V  A  A  R  A  R  Q   83

         ---------|---------|---------|---------|---------|
       651 ATCCTGGCGAAACCTGCAAAAAAATGAAACTGGACGCTTTTTGCGTAAGT 700
F1      37  S  W  R  N  L  Q  K  N  E  T  G  R  F  L  R  K  L 53
F2      60   P  G  E  T  C  K  K  M  K  L  D  A  F  C  V  S   75
F3      84 I  L  A  K  P  A  K  K  *  N  W  T  L  F  A  *  V  1

         ---------|---------|---------|---------|---------|
       701 TGGCATTTATTTTCTTTTCTTCTTCTTCTTGTGCAAAAGAATTGAAAAAA 750
F1      54   A  F  I  F  F  S  S  S  S  C  A  K  E  L  K  K   69
F2      76 W  H  L  F  S  F  L  L  L  L  V  Q  K  N  *  K  K  2
F3       2  G  I  Y  F  L  F  F  F  F  L  C  K  R  I  E  K  K 18

         ---------|---------|---------|---------|---------|
       751 AGAACTTTCGGCGTTGGAAACTCGCGACTGACGCACAATTAACCTCAACT 800
F1      70 R  T  F  G  V  G  N  S  R  L  T  H  N  *  P  Q  L  3
F2       3  E  L  S  A  L  E  T  R  D  *  R  T  I  N  L  N  F 7
F3      19   N  F  R  R  W  K  L  A  T  D  A  Q  L  T  S  T   34

         ---------|---------|---------|---------|---------|
       801 TTAGTGCAGCAATTTCAAGGCGGACCGGCCAAAAGGAGGAAAGGGTTTCT 850
F1       1  *  C  S  N  F  K  A  D  R  P  K  G  G  K  G  F  L 16
F2       8   S  A  A  I  S  R  R  T  G  Q  K  E  E  R  V  S   23
F3      35 L  V  Q  Q  F  Q  G  G  P  A  K  R  R  K  G  F  L  51

         ---------|---------|---------|---------|---------|
       851 TGGGTGGATGTGATCCGATTGACAATTTCAAAATAGGACGCAATGTTATC 900
F1      17   G  G  C  D  P  I  D  N  F  K  I  G  R  N  V  I   32
F2      24 W  V  D  V  I  R  L  T  I  S  K  *  D  A  M  L  S  5
F3      52  G  W  M  *  S  D  *  Q  F  Q  N  R  T  Q  C  Y  R 10

         ---------|---------|---------|---------|---------|
       901 GCGACTGCTAGTGGCGCCGGCGGATGTAAATCCAATGGCCAGGATGGATT 950
F1      33 A  T  A  S  G  A  G  G  C  K  S  N  G  Q  D  G  F  49
F2       6  R  L  L  V  A  P  A  D  V  N  P  M  A  R  M  D  L 22
F3      11   D  C  *  W  R  R  R  M  *  I  Q  W  P  G  W  I   7
```

```
          ---------|---------|---------|---------|---------|
       951 TGTTGGATGTGCTTAAATTCGCATTTGAAGGGGCATGGTCTTGATATATA 1000
F1      50  V  G  C  A  *  I  R  I  *  R  G  M  V  L  I  Y  R  8
F2      23   L  D  V  L  K  F  A  F  E  G  A  W  S  *  Y  I   2
F3       8 C  W  M  C  L  N  S  H  L  K  G  H  G  L  D  I  *  23

          ---------|---------|---------|---------|---------|
      1001 GATACCCTATGTGCAGGCGACCTGATCTATATACTCTACACGCCCCTGTC 1050
F1       9   Y  P  M  C  R  R  P  D  L  Y  T  L  H  A  P  V   24
F2       3 D  T  L  C  A  G  D  L  I  Y  I  L  Y  T  P  L  S  19
F3       1  I  P  Y  V  Q  A  T  *  S  I  Y  S  T  R  P  C  P 9

          ---------|---------|---------|---------|---------|
      1051 CTAATTCCCGATGCAGAAGGATGTCACGCCCCCGCAACATGCACATTGGG 1100
F1      25 L  I  P  D  A  E  G  C  H  A  P  A  T  C  T  L  G  41
F2       1  *  F  P  M  Q  K  D  V  T  P  P  Q  H  A  H  W  V 16
F3      10   N  S  R  C  R  R  M  S  R  P  R  N  M  H  I  G   25

          ---------|---------|---------|---------|---------|
      1101 TCGATTTACCCGACGCCTGCCACAATTGAAACTTTTTTTTTTCCTCTTTT 1150
F1      42  R  F  T  R  R  L  P  Q  L  K  L  F  F  F  L  F  Y 58
F2      17   D  L  P  D  A  C  H  N  *  N  F  F  F  S  S  F   7
F3      26 S  I  Y  P  T  P  A  T  I  E  T  F  F  F  P  L  L  42

          ---------|---------|---------|---------|---------|
      1151 ATTCAGGCAACTGCCACTTCACAACACCCTCGTGCAAGCGCGACTGGGCA 1200
F1      59   S  G  N  C  H  F  T  T  P  S  C  K  R  D  W  A   74
F2       8 I  Q  A  T  A  T  S  Q  H  P  R  A  S  A  T  G  H  24
F3      43  F  R  Q  L  P  L  H  N  T  L  V  Q  A  R  L  G  T 59

          ---------|---------|---------|---------|---------|
      1201 CCAACTCAGTCCTGGCCGTGGGCGTGGCCGACTCGGGCACGGTGGTGGTG 1250
F1      75 P  T  Q  S  W  P  W  A  W  P  T  R  A  R  W  W  W  91
F2      25  Q  L  S  P  G  R  G  R  G  R  L  G  H  G  G  G  G 41
F3      60   N  S  V  L  A  V  G  V  A  D  S  G  T  V  V  V   75

          ---------|---------|---------|---------|---------|
      1251 GGCGCGACGCCGGCGGTCAGGGTCGCGGCGCCACTGCCGACCTGACCCTC 1300
F1      92  A  R  R  R  R  S  G  S  R  R  H  C  R  P  D  P  R 108
F2      42   R  D  A  G  G  Q  G  R  G  A  T  A  D  L  T  L   57
F3      76 G  A  T  P  A  V  R  V  A  A  P  L  P  T  *  P  S  2

          ---------|---------|---------|---------|---------|
      1301 GACGGCGAGAAAGACGGGCTGCTGGGCGCCGAGGGGCAGGCCGGCGCGTC 1350
F1     109   R  R  E  R  R  A  A  G  R  R  G  A  G  R  R  V   124
F2      58 D  G  E  K  D  G  L  L  G  A  E  G  Q  A  G  A  S  74
F3       3  T  A  R  K  T  G  C  W  A  P  R  G  R  P  A  R  R 19
```

Three frame translation of MG06592.4. Exonic regions are underlined. The start codon is located at nucleotide position 545. Intron consensus sequences are underlined.

```
SHOWORF of MG08394 from 1 to 1200

           ---------|---------|---------|---------|---------|
         1 CTTTCCGCATCGAATTTATAAGGTGTTTGTGTAAAGGTATTCTTAATGTG 50
F1       1 L  S  A  S  N  L  *  G  V  C  V  K  V  F  L  M  *  9
F2       1  F  P  H  R  I  Y  K  V  F  V  *  R  Y  S  *  C  D 2
F3       1   F  R  I  E  F  I  R  C  L  C  K  G  I  L  N  V  16

           ---------|---------|---------|---------|---------|
        51 ATTATTTTGGTCAGTGCAAAAGGCCAACTCCTCACGAGATCACAAATCTC 100
F1       1  L  F  W  S  V  Q  K  A  N  S  S  R  D  H  K  S  Q 17
F2       3   Y  F  G  Q  C  K  R  P  T  P  H  E  I  T  N  L  18
F3      17 I  I  L  V  S  A  K  G  Q  L  L  T  R  S  Q  I  S  33

           ---------|---------|---------|---------|---------|
       101 AGATGCAATTCTAGAGAACTTTTGGCGGGGAGAAGGCAATGCGTGAGGAC 150
F1      18   M  Q  F  *  R  T  F  G  G  E  K  A  M  R  E  D  12
F2      19 R  C  N  S  R  E  L  L  A  G  R  R  Q  C  V  R  T  35
F3      34  D  A  I  L  E  N  F  W  R  G  E  G  N  A  *  G  P 2

           ---------|---------|---------|---------|---------|
       151 CCGTGTGGACCCTTGAAGTGATCATGTAAGTGGAATGTGATTTGCCTAGA 200
F1      13 P  C  G  P  L  K  *  S  C  K  W  N  V  I  C  L  E  10
F2      36  R  V  D  P  *  S  D  H  V  S  G  M  *  F  A  *  R 1
F3       3   V  W  T  L  E  V  I  M  *  V  E  C  D  L  P  R  7

           ---------|---------|---------|---------|---------|
       201 GGTTGGAGATCAACTAGGATCGCCCTTCTTAAGGCATGGATAATGTAAGG 250
F1      11  V  G  D  Q  L  G  S  P  F  L  R  H  G  *  C  K  D 3
F2       2   L  E  I  N  *  D  R  P  S  *  G  M  D  N  V  R  6
F3       8 G  W  R  S  T  R  I  A  L  L  K  A  W  I  M  *  G  1

           ---------|---------|---------|---------|---------|
       251 ATGCGGTAAAGCCATTACTCAAAACTATGTTGAAAAATAAGATTACAGTT 300
F1       4   A  V  K  P  L  L  K  T  M  L  K  N  K  I  T  V  19
F2       7 M  R  *  S  H  Y  S  K  L  C  *  K  I  R  L  Q  F  6
F3       2  C  G  K  A  I  T  Q  N  Y  V  E  K  *  D  Y  S  S 4

           ---------|---------|---------|---------|---------|
       301 CCAGTTCAAACCGGGGGAGATTTTTATAAATAGAAAAAAAAAACGGATACA 350
F1      20 P  V  Q  T  G  G  D  F  Y  K  *  K  K  K  R  I  H  6
F2       7  Q  F  K  P  G  E  I  F  I  N  R  K  K  N  G  Y  I 23
F3       5   S  S  N  R  G  R  F  L  *  I  E  K  K  T  D  T  7

           ---------|---------|---------|---------|---------|
       351 TACGCAATGAACAAAAGGTTTTGGGAATATAAATCATGGCATTCCCCGTA 400
F1       7  T  Q  *  T  K  G  F  G  N  I  N  H  G  I  P  R  T 14
F2      24   R  N  E  Q  K  V  L  G  I  *  I  M  A  F  P  V  6
F3       8 Y  A  M  N  K  R  F  W  E  Y  K  S  W  H  S  P  Y  24

           ---------|---------|---------|---------|---------|
       401 CGTAAACCCCAACCTGTTGAAATGAGCTACAACTGCCTTGTATTGACCCA 450
F1       1   *  T  P  T  C  *  N  E  L  Q  L  P  C  I  D  P  10
F2       7 R  K  P  Q  P  V  E  M  S  Y  N  C  L  V  L  T  H  23
F3      25  V  N  P  N  L  L  K  *  A  T  T  A  L  Y  *  P  I 2

           ---------|---------|---------|---------|---------|
       451 TTCGCCCTCCTATATCGTCTTTGACTCTCTTTTCGTTCAAGTCACAAAAA 500
```

```
F1      11 F  A  L  L  Y  R  L  *  L  S  F  R  S  S  H  K  K  9
F2      24  S  P  S  Y  I  V  F  D  S  L  F  V  Q  V  T  K  N 40
F3       3   R  P  P  I  S  S  L  T  L  F  S  F  K  S  Q  K   18

         ---------|---------|---------|---------|---------|
       501 ATGCGCACATTTGCTATTCTCAGCTTACTAGCTGGCCTGGTTGCCGCTGC 550
F1      10  C  A  H  L  L  F  S  A  Y  *  L  A  W  L  P  L  H 7
F2      41   A  H  I  C  Y  S  Q  L  T  S  W  P  G  C  R  C   56
F3      19 M  R  T  F  A  I  L  S  L  L  A  G  L  V  A  A  A  35

         ---------|---------|---------|---------|---------|
       551 ATCCGACCCACTTGATGAACAGTTCTTCCCAGTTACGGGATACAGGTGCT 600
F1       8   P  T  H  L  M  N  S  S  S  Q  L  R  D  T  G  A   23
F2      57 I  R  P  T  *  *  T  V  L  P  S  Y  G  I  Q  V  L  11
F3      36  S  D  P  L  D  E  Q  F  F  P  V  T  G  Y  R  C  C 52

         ---------|---------|---------|---------|---------|
       601 GCGCCGACGCCACGGAAGACATAGGCGGCCATTGCAAGGCGGCTGGCTTT 650
F1      24 A  P  T  P  R  K  T  *  A  A  I  A  R  R  L  A  F  9
F2      12  R  R  R  H  G  R  H  R  R  P  L  Q  G  G  W  L  F 28
F3      53   A  D  A  T  E  D  I  G  G  H  C  K  A  A  G  F   68

         ---------|---------|---------|---------|---------|
       651 TCCGCATATTGTGTAGGTCTTGTTTCTTATTTTTGTTCGATTAACTTGTG 700
F1      10  P  H  I  V  *  V  L  F  L  I  F  V  R  L  T  C  A 12
F2      29   R  I  L  C  R  S  C  F  L  F  L  F  D  *  L  V   2
F3      69 S  A  Y  C  V  G  L  V  S  Y  F  C  S  I  N  L  C  85

         ---------|---------|---------|---------|---------|
       701 CAATTATACTTGATATTCCCCGTGGTCAAATTTCGTGTATTGGACATTTG 750
F1      13   I  I  L  D  I  P  R  G  Q  I  S  C  I  G  H  L   28
F2       3 Q  L  Y  L  I  F  P  V  V  K  F  R  V  L  D  I  C  19
F3      86  N  Y  T  *  Y  S  P  W  S  N  F  V  Y  W  T  F  V 13

         ---------|---------|---------|---------|---------|
       751 TTAATACCTTTTCTAGTGCACTCGTTTCGACTCTCGCAAGGGAAGTGGAT 800
F1      29 L  I  P  F  L  V  H  S  F  R  L  S  Q  G  K  W  M  45
F2       1  *  Y  L  F  *  C  T  R  F  D  S  R  K  G  S  G  C 12
F3      14   N  T  F  S  S  A  L  V  S  T  L  A  R  E  V  D   29

         ---------|---------|---------|---------|---------|
       801 GTGATGATACCCTCGGTTTCAAAATAGGACGCGTAGTTCAACAGGTTCGA 850
F1       1  *  *  Y  P  R  F  Q  N  R  T  R  S  S  T  G  S  T 15
F2      13   D  D  T  L  G  F  K  I  G  R  V  V  Q  Q  V  R   28
F3      30 V  M  I  P  S  V  S  K  *  D  A  *  F  N  R  F  D  5

         ---------|---------|---------|---------|---------|
       851 CTTGATTCAATGTCAGCTTGTGCTTCTGAAAATCGCAAGGGATTTATTGG 900
F1       1   *  F  N  V  S  L  C  F  *  K  S  Q  G  I  Y  W   7
F2      29 L  D  S  M  S  A  C  A  S  E  N  R  K  G  F  I  G  45
F3       6  L  I  Q  C  Q  L  V  L  L  K  I  A  R  D  L  L  D 22

         ---------|---------|---------|---------|---------|
       901 ATGCGTCTAAAGCATATTCCACGGTGGTGATACGGTGTTAGCTGGAAGGG 950
F1       8 M  R  L  K  H  I  P  R  W  *  Y  G  V  S  W  K  G  7
F2      46  C  V  *  S  I  F  H  G  G  D  T  V  L  A  G  R  E 14
F3      23   A  S  K  A  Y  S  T  V  V  I  R  C  *  L  E  G   3
```

```
          ---------|---------|---------|---------|---------|
       951 AGCTGAAAATCGGGGGCTTGTGGGTTTGGGGCCGGCACCCTAACAAGCGC 1000
F1       8  A  E  N  R  G  L  V  G  L  G  P  A  P  *  Q  A  Q 3
F2      15   L  K  I  G  G  L  W  V  W  G  R  H  P  N  K  R   30
F3       4 S  *  K  S  G  A  C  G  F  G  A  G  T  L  T  S  A  15

          ---------|---------|---------|---------|---------|
      1001 AAATTCTAATGGGCCAATCACAAAGGCTAAAAATGCTTTACAGCAACCGG 1050
F1       4   I  L  M  G  Q  S  Q  R  L  K  M  L  Y  S  N  R   19
F2      31 K  F  *  W  A  N  H  K  G  *  K  C  F  T  A  T  G  7
F3      16  N  S  N  G  P  I  T  K  A  K  N  A  L  Q  Q  P  A 32

          ---------|---------|---------|---------|---------|
      1051 CAATGTTCATGATAACCGTTTGGATATCAAAAGACGCAATGTCTTTTCCA 1100
F1      20 Q  C  S  *  *  P  F  G  Y  Q  K  T  Q  C  L  F  Q  12
F2       8  N  V  H  D  N  R  L  D  I  K  R  R  N  V  F  S  R 24
F3      33   M  F  M  I  T  V  W  I  S  K  D  A  M  S  F  P   48

          ---------|---------|---------|---------|---------|
      1101 GAACGCTATTAATTCCCTAAAATAAAACAATATCGTTTTACGCCTAAAAG 1150
F1      13  N  A  I  N  S  L  K  *  N  N  I  V  L  R  L  K  A 9
F2      25   T  L  L  I  P  *  N  K  T  I  S  F  Y  A  *  K   1
F3      49 E  R  Y  *  F  P  K  I  K  Q  Y  R  F  T  P  K  S  13

          ---------|---------|---------|---------|---------|
      1151 CTTATTTCCGGTGCTCTTACAGATAACAGACTGTCGGGAAGGCTTGATTT 1200
F1      10   Y  F  R  C  S  Y  R  *  Q  T  V  G  K  A  *  F   1
F2       2 L  I  S  G  A  L  T  D  N  R  L  S  G  R  L  D  L  18
F3      14  L  F  P  V  L  L  Q  I  T  D  C  R  E  G  L  I  Y 30
```

Three frame translation of MG08394.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG09155 from 1 to 1001

```
           ---------|---------|---------|---------|---------|
         1 ATGTTTTGCCCGTTCGTATCTATCCCTTACGGGTTTTAGTGTATGTGTTT 50
F1       1 M  F  C  P  F  V  S  I  P  Y  G  F  *  C  M  C  L  4
F2       1  C  F  A  R  S  Y  L  S  L  T  G  F  S  V  C  V  C 17
F3       1   V  L  P  V  R  I  Y  P  L  R  V  L  V  Y  V  F   16

           ---------|---------|---------|---------|---------|
        51 GCATATAGTTTGAGGTTTCTGAGACGTCAATGGTTTGTTCGTATTGTCCA 100
F1       5  H  I  V  *  G  F  *  D  V  N  G  L  F  V  L  S  T 10
F2      18   I  *  F  E  V  S  E  T  S  M  V  C  S  Y  C  P   14
F3      17 A  Y  S  L  R  F  L  R  R  Q  W  F  V  R  I  V  H  33

           ---------|---------|---------|---------|---------|
       101 CCTGAACGAATTGCATTTCATTCAGGGTAACATCCGTAAGTAATATTAGC 150
F1       1   *  T  N  C  I  S  F  R  V  T  S  V  S  N  I  S   15
F2      15 P  E  R  I  A  F  H  S  G  *  H  P  *  V  I  L  A  4
F3      34  L  N  E  L  H  F  I  Q  G  N  I  R  K  *  Y  *  R 1

           ---------|---------|---------|---------|---------|
       151 GTCTGCTAAAACCGTGCCGCAAACTTTAAAGCTTAATCCCTGTCAAAGCC 200
F1      16 V  C  *  N  R  A  A  N  F  K  A  *  S  L  S  K  P  5
F2       5  S  A  K  T  V  P  Q  T  L  K  L  N  P  C  Q  S  P 21
F3       2   L  L  K  P  C  R  K  L  *  S  L  I  P  V  K  A   7

           ---------|---------|---------|---------|---------|
       201 CCGGTGGAAATAGCCAAGGTAGTATATATAGAAACACACCGAGATGGATC 250
F1       6  R  W  K  *  P  R  *  Y  I  *  K  H  T  E  M  D  L 7
F2      22   G  G  N  S  Q  G  S  I  Y  R  N  T  P  R  W  I   37
F3       8 P  V  E  I  A  K  V  V  Y  I  E  T  H  R  D  G  S  24

           ---------|---------|---------|---------|---------|
       251 TACAGCAAAAGCTTGGCAATATAAAACCACGTTTGTCGCTTGCGTTGGGA 300
F1       8   Q  Q  K  L  G  N  I  K  P  R  L  S  L  A  L  G   23
F2      38 Y  S  K  S  L  A  I  *  N  H  V  C  R  L  R  W  D  9
F3      25  T  A  K  A  W  Q  Y  K  T  T  F  V  A  C  V  G  T 41

           ---------|---------|---------|---------|---------|
       301 CAAACGCCGGTTTAGAGCTATCACTTGGAACCCAAGCGGCTTTTCATCTA 350
F1      24 Q  T  P  V  *  S  Y  H  L  E  P  K  R  L  F  I  Y  12
F2      10  K  R  R  F  R  A  I  T  W  N  P  S  G  F  S  S  T 26
F3      42   N  A  G  L  E  L  S  L  G  T  Q  A  A  F  H  L   57

           ---------|---------|---------|---------|---------|
       351 CGTCTTTACCCTTTGATATCCTGCGCTTTGATTTCCTGCCTCGTCTTCTT 400
F1      13  V  F  T  L  *  Y  P  A  L  *  F  P  A  S  S  S  L 7
F2      27   S  L  P  F  D  I  L  R  F  D  F  L  P  R  L  L   42
F3      58 R  L  Y  P  L  I  S  C  A  L  I  S  C  L  V  F  F  74

           ---------|---------|---------|---------|---------|
       401 TGTTCTACTCATACAAACAAAAAAAATGCGTTCATACATTCTTTTTTGCT 450
F1       8   F  Y  S  Y  K  Q  K  K  C  V  H  T  F  F  F  A   23
F2      43 C  S  T  H  T  N  K  K  N  A  F  I  H  S  F  L  L  59
F3      75  V  L  L  I  Q  T  K  K  M  R  S  Y  I  L  F  C  C 91

           ---------|---------|---------|---------|---------|
       451 GTTTAGCTGGCCTGGCTGCCGCTAGATCGCTTGCAATCCAGCCCCGGGAC 500
```

```
F1      24 V  *  L  A  W  L  P  L  D  R  L  Q  S  S  P  G  T  15
F2      60  F  S  W  P  G  C  R  *  I  A  C  N  P  A  P  G  R 9
F3      92    L  A  G  L  A  A  A  R  S  L  A  I  Q  P  R  D   107

         ---------|---------|---------|---------|---------|
       501 GATTTGGACTTTACGGCAACCACGGGACCCATCTGCTGCGGCCATGGCAC 550
F1      16  I  W  T  L  R  Q  P  R  D  P  S  A  A  A  M  A  H 32
F2      10    F  G  L  Y  G  N  H  G  T  H  L  L  R  P  W  H   25
F3     108 D  L  D  F  T  A  T  T  G  P  I  C  C  G  H  G  T  124

         ---------|---------|---------|---------|---------|
       551 ACAAGATCCAAACAACCTTTGCAAAAATGCTGGCCTCTTTGCTTATTGCG 600
F1      33    K  I  Q  T  T  F  A  K  M  L  A  S  L  L  I  A   48
F2      26 T  R  S  K  Q  P  L  Q  K  C  W  P  L  C  L  L  R  42
F3     125  Q  D  P  N  N  L  C  K  N  A  G  L  F  A  Y  C  V 141

         ---------|---------|---------|---------|---------|
       601 TAAGCTTGTTCTTTCATTCTTGTTCATCTACCATACGCCGTTGGGTCTCT 650
F1       1 *  A  C  S  F  I  L  V  H  L  P  Y  A  V  G  S  L  16
F2      43  K  L  V  L  S  F  L  F  I  Y  H  T  P  L  G  L  W 59
F3     142    S  L  F  F  H  S  C  S  S  T  I  R  R  W  V  S   157

         ---------|---------|---------|---------|---------|
       651 GGACAGGTTCAGTACCTCGGTGCCTTTGGTTGGAGTTGGTCGAATCATCA 700
F1      17  D  R  F  S  T  S  V  P  L  V  G  V  G  R  I  I  I 33
F2      60    T  G  S  V  P  R  C  L  W  L  E  L  V  E  S  S   75
F3     158 G  Q  V  Q  Y  L  G  A  F  G  W  S  W  S  N  H  H  174

         ---------|---------|---------|---------|---------|
       701 TTAACAACTTCTTTAGTGTTCCTCCTTTGCAAATAACGAAGAGCAAGGAT 750
F1      34    N  N  F  F  S  V  P  P  L  Q  I  T  K  S  K  D   49
F2      76 L  T  T  S  L  V  F  L  L  C  K  *  R  R  A  R  M  5
F3       1  *  Q  L  L  *  C  S  S  F  A  N  N  E  E  Q  G  C 12

         ---------|---------|---------|---------|---------|
       751 GTGATCCAGTCGTTGATTTCCACGTTGGGCGCGACGTCAAAATAGTCGAC 800
F1      50 V  I  Q  S  L  I  S  T  L  G  A  T  S  K  *  S  T  2
F2       1  *  S  S  R  *  F  P  R  W  A  R  R  Q  N  S  R  Q 12
F3      13    D  P  V  V  D  F  H  V  G  R  D  V  K  I  V  D   28

         ---------|---------|---------|---------|---------|
       801 AGCGAATCACAGCGGAAATGTGTTTCCGGGACACGCGTTGGATTTGTTGG 850
F1       3  A  N  H  S  G  N  V  F  P  G  H  A  L  D  L  L  D 19
F2      13    R  I  T  A  E  M  C  F  R  D  T  R  W  I  C  W   28
F3      29 S  E  S  Q  R  K  C  V  S  G  T  R  V  G  F  V  G  45

         ---------|---------|---------|---------|---------|
       851 ATGTGCTAACTGAAGTGTATCTTGGCGGAAATGGTTTTTTCTTGTTCTTT 900
F1      20    V  L  T  E  V  Y  L  G  G  N  G  F  F  L  F  F   35
F2      29 M  C  *  L  K  C  I  L  A  E  M  V  F  S  C  S  F  14
F3      46  C  A  N  *  S  V  S  W  R  K  W  F  F  L  V  L  L 13

         ---------|---------|---------|---------|---------|
       901 TGCTCCGGAAGCGAAGCTCGTATACCCGAACAGGAGAGATGGCGGCTGGA 950
F1      36 C  S  G  S  E  A  R  I  P  E  Q  E  R  W  R  L  E  52
F2      15  A  P  E  A  K  L  V  Y  P  N  R  R  D  G  G  W  K 31
F3      14    L  R  K  R  S  S  Y  T  R  T  G  E  M  A  A  G   29
```

```
          ---------|---------|---------|---------|---------|
        951 AATACGGAAGCCCTGTTGAATTCTCACATCTGGAAATTGTAAAAAGGGCA 1000
F1       53  I  R  K  P  C  *  I  L  T  S  G  N  C  K  K  G    10
F2       32   Y  G  S  P  V  E  F  S  H  L  E  I  V  K  R  A   47
F3       30 N  T  E  A  L  L  N  S  H  I  W  K  L  *  K  G  Q  3
```

Three frame translation of MG09155.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG10100.4 from 1 to 1200

```
           ---------|---------|---------|---------|---------|
         1 AGGAGCCGTTGTAAGCAAGTGTAAATTGCAAAAGTACTCCACGAAGCTTG 50
F1       1 R  S  R  C  K  Q  V  *  I  A  K  V  L  H  E  A  C  9
F2       1  G  A  V  V  S  K  C  K  L  Q  K  Y  S  T  K  L  V 17
F3       1   E  P  L  *  A  S  V  N  C  K  S  T  P  R  S  L  12

           ---------|---------|---------|---------|---------|
        51 TGTATATTTTCTCTGCCAGCTGTTTCCTCAAAGCTTTCCGCCTATAAAGC 100
F1      10  V  Y  F  L  C  Q  L  F  P  Q  S  F  P  P  I  K  P 26
F2      18   Y  I  F  S  A  S  C  F  L  K  A  F  R  L  *  S   1
F3      13 C  I  F  S  L  P  A  V  S  S  K  L  S  A  Y  K  A  29

           ---------|---------|---------|---------|---------|
       101 CTAAGTGAAAGCGACACTTGCTAAATGTCAGTGCATTTGACCATTAAATA 150
F1      27   K  *  K  R  H  L  L  N  V  S  A  F  D  H  *  I   1
F2       2 L  S  E  S  D  T  C  *  M  S  V  H  L  T  I  K  *  8
F3       1  *  V  K  A  T  L  A  K  C  Q  C  I  *  P  L  N  S 4

           ---------|---------|---------|---------|---------|
       151 GCTTGTAACAGATCTGTACCATTTGTAAAGTTCCATGTTTAGACGCTGAT 200
F1       2 A  C  N  R  S  V  P  F  V  K  F  H  V  *  T  L  I  3
F2       1  L  V  T  D  L  Y  H  L  *  S  S  M  F  R  R  *  S 1
F3       5   L  *  Q  I  C  T  I  C  K  V  P  C  L  D  A  D  14

           ---------|---------|---------|---------|---------|
       201 CATACACTCAGCCGTCCGATCGCGTCTTGGAACTCCTGGAATCGTAGCAC 250
F1       4  I  H  S  A  V  R  S  R  L  G  T  P  G  I  V  A  P 20
F2       2   Y  T  Q  P  S  D  R  V  L  E  L  L  E  S  *  H   1
F3      15 H  T  L  S  R  P  I  A  S  W  N  S  W  N  R  S  T  31

           ---------|---------|---------|---------|---------|
       251 CCCTGTCCTTGCAAGGAATTTGTGACCGCTGTCCCGTACAATATTCTTTG 300
F1      21   L  S  L  Q  G  I  C  D  R  C  P  V  Q  Y  S  L   36
F2       2 P  C  P  C  K  E  F  V  T  A  V  P  Y  N  I  L  C  18
F3      32  P  V  L  A  R  N  L  *  P  L  S  R  T  I  F  F  A 9

           ---------|---------|---------|---------|---------|
       301 CCCATTGCCATTGGACTCGAATTACATAAACTGCACTGGATCCCTGGTTT 350
F1      37 P  I  A  I  G  L  E  L  H  K  L  H  W  I  P  G  L  53
F2      19  P  L  P  L  D  S  N  Y  I  N  C  T  G  S  L  V  * 34
F3      10   H  C  H  W  T  R  I  T  *  T  A  L  D  P  W  F   7

           ---------|---------|---------|---------|---------|
       351 AATCCCGGCTCAGATACTATTTCCGTACCTCTGTCGCTTAACTTGATACT 400
F1      54  I  P  A  Q  I  L  F  P  Y  L  C  R  L  T  *  Y  S 2
F2       1   S  R  L  R  Y  Y  F  R  T  S  V  A  *  L  D  T   3
F3       8 N  P  G  S  D  T  I  S  V  P  L  S  L  N  L  I  L  24

           ---------|---------|---------|---------|---------|
       401 CGTTTTCAAGAAATCGGTATCGCGAGAACGATTACTGCAGTTTCTTGACT 450
F1       3   F  S  R  N  R  Y  R  E  N  D  Y  C  S  F  L  T   18
F2       4 R  F  Q  E  I  G  I  A  R  T  I  T  A  V  S  *  L  1
F3      25  V  F  K  K  S  V  S  R  E  R  L  L  Q  F  L  D  S 41

           ---------|---------|---------|---------|---------|
       451 CCCTTGTCTTACACATCCCTTTTCAACTCTCTGCTGGTCAGCTGGAAGCA 500
```

```
F1      19 P  L  S  Y  T  S  L  F  N  S  L  L  V  S  W  K  Q  35
F2       2  P  C  L  T  H  P  F  S  T  L  C  W  S  A  G  S  N 18
F3      42   L  V  L  H  I  P  F  Q  L  S  A  G  Q  L  E  A  57

         ---------|---------|---------|---------|---------|
       501 ATGCGCTACTCTATTCTCATTTTGGCCCCGACCATAGTGCTGGGCCAAAT 550
F1      36  C  A  T  L  F  S  F  W  P  R  P  *  C  W  A  K  S 5
F2      19   A  L  L  Y  S  H  F  G  P  D  H  S  A  G  P  N   34
F3      58 M  R  Y  S  I  L  I  L  A  P  T  I  V  L  G  Q  I  74

         ---------|---------|---------|---------|---------|
       551 CTTCTCCTCTGCCGGCGAACCAGCGACCGGCCCGCTGTGCTGCAACAGGG 600
F1       6   S  P  L  P  A  N  Q  R  P  A  R  C  A  A  T  G   21
F2      35 L  L  L  C  R  R  T  S  D  R  P  A  V  L  Q  Q  G  51
F3      75  F  S  S  A  G  E  P  A  T  G  P  L  C  C  N  R  G 91

         ---------|---------|---------|---------|---------|
       601 GTGTTGTAGACACCAGTGGGACATGCAAGAGTTTGAACTTGAACGCTTAT 650
F1      22 V  L  *  T  P  V  G  H  A  R  V  *  T  *  T  L  M  3
F2      52  C  C  R  H  Q  W  D  M  Q  E  F  E  L  E  R  L  C 68
F3      92   V  V  D  T  S  G  T  C  K  S  L  N  L  N  A  Y   107

         ---------|---------|---------|---------|---------|
       651 GCAGTGAGTTATCCGCATTTATATCTGAAAACCGCAAGGGGCAATGGAAA 700
F1       4  Q  *  V  I  R  I  Y  I  *  K  P  Q  G  A  M  E  T 8
F2      69   S  E  L  S  A  F  I  S  E  N  R  K  G  Q  W  K   84
F3     108 A  V  S  Y  P  H  L  Y  L  K  T  A  R  G  N  G  N  124

         ---------|---------|---------|---------|---------|
       701 CCAGCTCGAACCTCTTTTTTTTTAGTGTGAATCTATCAGGAGCAACAGTG 750
F1       9   S  S  N  L  F  F  F  S  V  N  L  S  G  A  T  V   24
F2      85 P  A  R  T  S  F  F  L  V  *  I  Y  Q  E  Q  Q  C  7
F3     125  Q  L  E  P  L  F  F  *  C  E  S  I  R  S  N  S  A 9

         ---------|---------|---------|---------|---------|
       751 CCAAAGCTGTGGCGGGCGACCCTGACTCGAAAAGTGGATGCGACAATGGC 800
F1      25 P  K  L  W  R  A  T  L  T  R  K  V  D  A  T  M  A  41
F2       8  Q  S  C  G  G  R  P  *  L  E  K  W  M  R  Q  W  R 9
F3      10   K  A  V  A  G  D  P  D  S  K  S  G  C  D  N  G   25

         ---------|---------|---------|---------|---------|
       801 GTATTCGAATTATTTCCGGTTGGACGAGATGTTAAAGCATTCGTTCCGAA 850
F1      42  Y  S  N  Y  F  R  L  D  E  M  L  K  H  S  F  R  T 58
F2      10   I  R  I  I  S  G  W  T  R  C  *  S  I  R  S  E   5
F3      26 V  F  E  L  F  P  V  G  R  D  V  K  A  F  V  P  N  42

         ---------|---------|---------|---------|---------|
       851 CTCGGGCGACACGATTAAACTTGGACCAAGCTCTCTTGGCGACGCCTTTA 900
F1      59   R  A  T  R  L  N  L  D  Q  A  L  L  A  T  P  L   74
F2       6 L  G  R  H  D  *  T  W  T  K  L  S  W  R  R  L  Y  11
F3      43  S  G  D  T  I  K  L  G  P  S  S  L  G  D  A  F  T 59

         ---------|---------|---------|---------|---------|
       901 CTGCATTCATTGGATGCGCAGACTAGATATTGTGGTTTAAATGCATCGGG 950
F1      75 L  H  S  L  D  A  Q  T  R  Y  C  G  L  N  A  S  G  91
F2      12  C  I  H  W  M  R  R  L  D  I  V  V  *  M  H  R  G 4
F3      60   A  F  I  G  C  A  D  *  I  L  W  F  K  C  I  G   8
```

```
          ---------|---------|---------|---------|---------|
       951 GAGGGGTTATTCGCATCCTGTATGCTGCAGTCTTTGTTCTGTCCCCAATG 1000
F1      92  R  G  Y  S  H  P  V  C  C  S  L  C  S  V  P  N  G 108
F2       5   G  V  I  R  I  L  Y  A  A  V  F  V  L  S  P  M   20
F3       9 E  G  L  F  A  S  C  M  L  Q  S  L  F  C  P  Q  W  25

          ---------|---------|---------|---------|---------|
      1001 GGCACATAATGCTTTGATCTCTTTCCGCGCCAAGTATTCCTCGATGGACC 1050
F1     109   H  I  M  L  *  S  L  S  A  P  S  I  P  R  W  T   11
F2      21 G  T  *  C  F  D  L  F  P  R  Q  V  F  L  D  G  P  14
F3      26  A  H  N  A  L  I  S  F  R  A  K  Y  S  S  M  D  L 42

          ---------|---------|---------|---------|---------|
      1051 TGCGAGCAGAAGTTGGGGATACAATCATGTTAGTTTTAAATAACCAAATG 1100
F1      12 C  E  Q  K  L  G  I  Q  S  C  *  F  *  I  T  K  C  4
F2      15  A  S  R  S  W  G  Y  N  H  V  S  F  K  *  P  N  V 3
F3      43   R  A  E  V  G  D  T  I  M  L  V  L  N  N  Q  M   58

          ---------|---------|---------|---------|---------|
      1101 TACGATTATTGGCAAGCTGAAAAAAAAGCCTTGCTTTGAACAAGAATCAA 1150
F1       5  T  I  I  G  K  L  K  K  K  P  C  F  E  Q  E  S  T 21
F2       4   R  L  L  A  S  *  K  K  S  L  A  L  N  K  N  Q   10
F3      59 Y  D  Y  W  Q  A  E  K  K  A  L  L  *  T  R  I  N  4

          ---------|---------|---------|---------|---------|
      1151 CTCGTCTCGGTCAAACGTTCCTCTACTTAACAACGGACCGAAGCCTTGTA 1200
F1      22   R  L  G  Q  T  F  L  Y  L  T  T  D  R  S  L  V   37
F2      11 L  V  S  V  K  R  S  S  T  *  Q  R  T  E  A  L  *  6
F3       5  S  S  R  S  N  V  P  L  L  N  N  G  P  K  P  C  N 21
```

Three frame translation of MG10100.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

```
MG10732 from 1 to 1200

          ---------|---------|---------|---------|---------|
        1 GCTTTGATCTGTGTTTCGGAACCTGATGCGTATAGCCTCTTTTTATATTT 50
F1      1 A  L  I  C  V  S  E  P  D  A  Y  S  L  F  L  Y  L  17
F2      1  L  *  S  V  F  R  N  L  M  R  I  A  S  F  Y  I  * 14
F3      1   F  D  L  C  F  G  T  *  C  V  *  P  L  F  I  F   5

          ---------|---------|---------|---------|---------|
       51 AGGAGACAGATTAAATACAACCAAGTGTAGCAGTCCCCACTACTTATTTA 100
F1     18  G  D  R  L  N  T  T  K  C  S  S  P  H  Y  L  F  T 34
F2      1   E  T  D  *  I  Q  P  S  V  A  V  P  T  T  Y  L   12
F3      6 R  R  Q  I  K  Y  N  Q  V  *  Q  S  P  L  L  I  Y  7

          ---------|---------|---------|---------|---------|
      101 CCAAGCCCAAAGTCTTATTATAAACGTAGCCGAGGTTTATGAATACTTGC 150
F1     35   K  P  K  V  L  L  *  T  *  P  R  F  M  N  T  C   7
F2     13 P  S  P  K  S  Y  Y  K  R  S  R  G  L  *  I  L  A  3
F3      8  Q  A  Q  S  L  I  I  N  V  A  E  V  Y  E  Y  L  Q 24

          ---------|---------|---------|---------|---------|
      151 AAATATTCAAGATGAAAATCCCCACTTCTGGTAGTATAGATCTGGTTTTA 200
F1      8 K  Y  S  R  *  K  S  P  L  L  V  V  *  I  W  F  Y  4
F2      4  N  I  Q  D  E  N  P  H  F  W  *  Y  R  S  G  F  T 6
F3     25   I  F  K  M  K  I  P  T  S  G  S  I  D  L  V  L   40

          ---------|---------|---------|---------|---------|
      201 CGCATCCTCAAACAAAACGTGTTCGGCGATCTTGCAAAAACATACTAATT 250
F1      5  A  S  S  N  K  T  C  S  A  I  L  Q  K  H  T  N  Y 21
F2      7   H  P  Q  T  K  R  V  R  R  S  C  K  N  I  L  I   22
F3     41 R  I  L  K  Q  N  V  F  G  D  L  A  K  T  Y  *  L  1

          ---------|---------|---------|---------|---------|
      251 ATAGAATGTGTAAAACTAATTCTTATTTTACAACGTTTCAATAACGTTTT 300
F1     22   R  M  C  K  T  N  S  Y  F  T  T  F  Q  *  R  F   2
F2     23 I  E  C  V  K  L  I  L  I  L  Q  R  F  N  N  V  L  39
F3      1  *  N  V  *  N  *  F  L  F  Y  N  V  S  I  T  F  C 11

          ---------|---------|---------|---------|---------|
      301 GCCGGCTGATTAACCAACCTACTTCCAACTTCATAATCCCAAGTACATTT 350
F1      3 A  G  *  L  T  N  L  L  P  T  S  *  S  Q  V  H  F  5
F2     40  P  A  D  *  P  T  Y  F  Q  L  H  N  P  K  Y  I  F 13
F3     12   R  L  I  N  Q  P  T  S  N  F  I  I  P  S  T  F   27

          ---------|---------|---------|---------|---------|
      351 TTATAAATTTAACTGCCGGTTAAAAACCAATGGAATCCCCTCTTTGGCAT 400
F1      6  Y  K  F  N  C  R  L  K  T  N  G  I  P  S  L  A  F 22
F2     14   I  N  L  T  A  G  *  K  P  M  E  S  P  L  W  H   9
F3     28 L  *  I  *  L  P  V  K  N  Q  W  N  P  L  F  G  I  13

          ---------|---------|---------|---------|---------|
      401 TTTATCCCTGGCTTGTAATGCTTAACCATATTTTACATTCGCTGTGCGTA 450
F1     23   Y  P  W  L  V  M  L  N  H  I  L  H  S  L  C  V   38
F2     10 F  I  P  G  L  *  C  L  T  I  F  Y  I  R  C  A  Y  11
F3     14  L  S  L  A  C  N  A  *  P  Y  F  T  F  A  V  R  I 9

          ---------|---------|---------|---------|---------|
      451 TACCTCACTCCTTTCTAAAAACTTTTTTTATTTGCCCCTTCATCGTCAAC 500
```

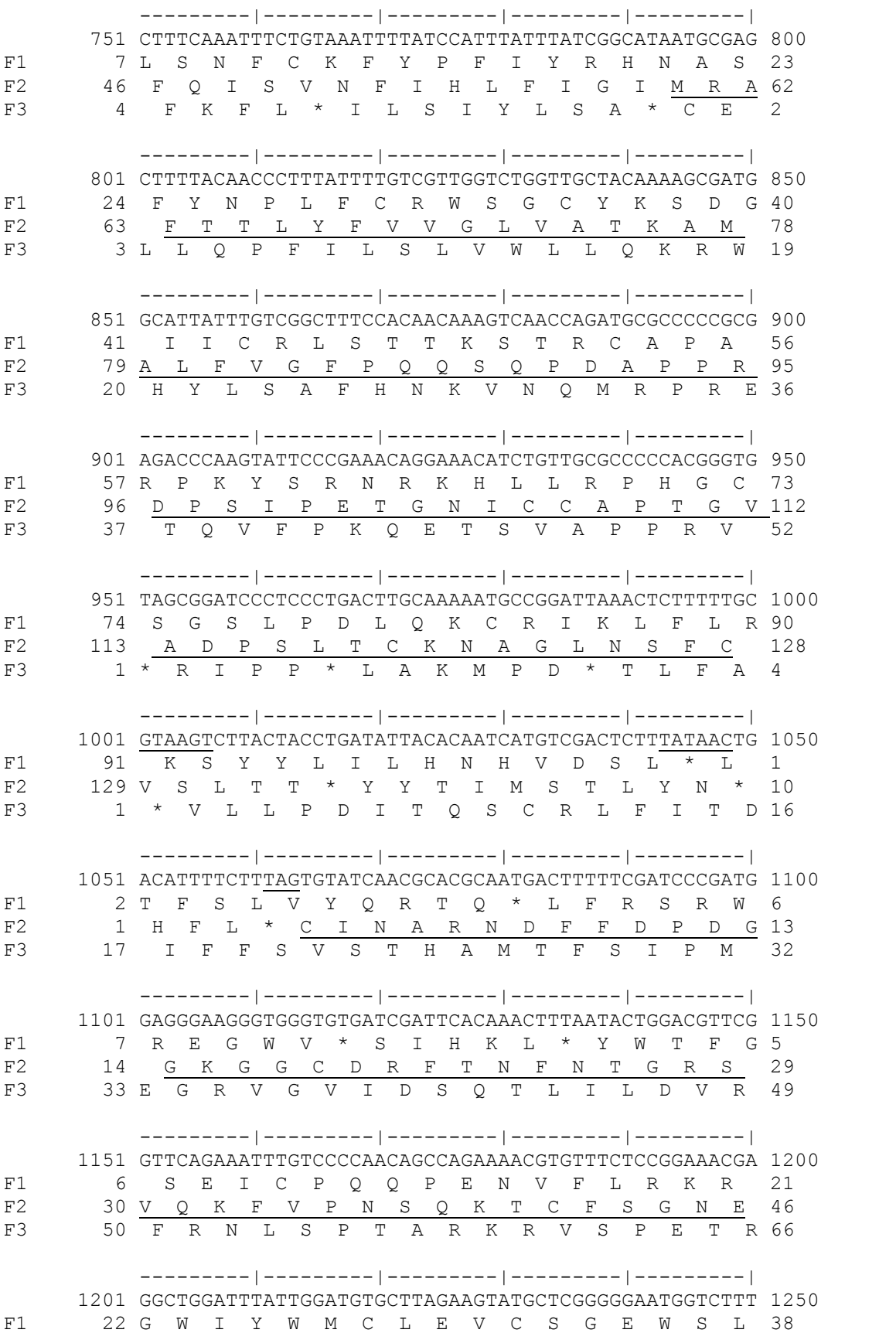
```
F1      39 Y  L  T  P  F  *  K  L  F  L  F  A  P  S  S  S  T  11
F2      12  T  S  L  L  S  K  N  F  F  Y  L  P  L  H  R  Q  H 28
F3      10   P  H  S  F  L  K  T  F  F  I  C  P  F  I  V  N   25

         ---------|---------|---------|---------|---------|
       501 ATGCGTTCTTCCACACTTCTTATTGTTCCTTTTTACTTCCTAGCCGGCCT 550
F1      12  C  V  L  P  H  F  L  L  F  L  F  T  S  *  P  A  W 3
F2      29   A  F  F  H  T  S  Y  C  S  F  L  L  P  S  R  P   44
F3      26 M  R  S  S  T  L  L  I  V  P  F  Y  F  L  A  G  L  42

         ---------|---------|---------|---------|---------|
       551 GGTTGCCGCCTCGGCCGACAAGGCCCACGACATAGAGCTCGACTTCGAGG 600
F1       4   L  P  P  R  P  T  R  P  T  T  *  S  S  T  S  R   5
F2      45 G  C  R  L  G  R  Q  G  P  R  H  R  A  R  L  R  G  61
F3      43  V  A  A  S  A  D  K  A  H  D  I  E  L  D  F  E  G 59

         ---------|---------|---------|---------|---------|
       601 GGCCTCCGTCGGGATGGGTTTGTTGCGACGCCGGCGCAGAGGACGCTGAC 650
F1       6 G  L  R  R  D  G  F  V  A  T  P  A  Q  R  T  L  T  22
F2      62  A  S  V  G  M  G  L  L  R  R  R  R  R  G  R  *  R 1
F3      60   P  P  S  G  W  V  C  C  D  A  G  A  E  D  A  D   75

         ---------|---------|---------|---------|---------|
       651 GGAGCTTGCAAGGCCAAGGGTCTAAACGCCTTTTGTGTAAGTTGCCTTTT 700
F1      23  E  L  A  R  P  R  V  *  T  P  F  V  *  V  A  F  C 4
F2       2   S  L  Q  G  Q  G  S  K  R  L  L  C  K  L  P  F   17
F3      76 G  A  C  K  A  K  G  L  N  A  F  C  V  S  C  L  L  92

         ---------|---------|---------|---------|---------|
       701 GTTTTATTTGCTTGGGCTGGGGCAAAAAGATAAACAAAAAAAAAGAAGGC 750
F1       5   F  I  C  L  G  W  G  K  K  I  N  K  K  K  E  G   20
F2      18 V  L  F  A  W  A  G  A  K  R  *  T  K  K  K  K  A  6
F3      93  F  Y  L  L  G  L  G  Q  K  D  K  Q  K  K  R  R  L 109

         ---------|---------|---------|---------|---------|
       751 TAATCCTTTCGTTCTCTTGTTCCAGTGCGGTCCTTTCAAGGCAGACAAGA 800
F1       1 *  S  F  R  S  L  V  P  V  R  S  F  Q  G  R  Q  E  16
F2       7  N  P  F  V  L  L  F  Q  C  G  P  F  K  A  D  K  K 23
F3     110   I  L  S  F  S  C  S  S  A  V  L  S  R  Q  T  R   125

         ---------|---------|---------|---------|---------|
       801 AGCGGCCCGGCAAGGGCAACAGTGGATGCGACCCGTTTTTCGCAACTGTC 850
F1      17  A  A  R  Q  G  Q  Q  W  M  R  P  V  F  R  N  C  P 33
F2      24   R  P  G  K  G  N  S  G  C  D  P  F  F  A  T  V   39
F3     126 S  G  P  A  R  A  T  V  D  A  T  R  F  S  Q  L  S  142

         ---------|---------|---------|---------|---------|
       851 CCCACGGGACGTGATGTTAAATTCCTCAACGGGTTTTGCACCGCCGGTGG 900
F1      34   H  G  T  *  C  *  I  P  Q  R  V  L  H  R  R  W   10
F2      40 P  T  G  R  D  V  K  F  L  N  G  F  C  T  A  G  G  56
F3     143  P  R  D  V  M  L  N  S  S  T  G  F  A  P  P  V  V 159

         ---------|---------|---------|---------|---------|
       901 TGATTTGCCTGGACATGTTGGATGTGCTTAAATCACGACGGATGGATCTT 950
F1       1 *  F  A  W  T  C  W  M  C  L  N  H  D  G  W  I  F  16
F2      57  D  L  P  G  H  V  G  C  A  *  I  T  T  D  G  S  S 7
F3     160   I  C  L  D  M  L  D  V  L  K  S  R  R  M  D  L   175
```

```
         ---------|---------|---------|---------|---------|
       951 CGTGCGGAGATGTTCCAATAACAGTGTCTCGCGTTTGGCATGGGATGCCT 1000
F1      17  V  R  R  C  S  N  N  S  V  S  R  L  A  W  D  A  Y 33
F2       8   C  G  D  V  P  I  T  V  S  R  V  W  H  G  M  P   23
F3     176 R  A  E  M  F  Q  *  Q  C  L  A  F  G  M  G  C  L  10

         ---------|---------|---------|---------|---------|
      1001 ACCCTGGTAGATGAGTGGATGCACGAGCGATCCGGGCGCTTGTAAAGCAT 1050
F1      34   P  G  R  *  V  D  A  R  A  I  R  A  L  V  K  H   12
F2      24 T  L  V  D  E  W  M  H  E  R  S  G  R  L  *  S  I  2
F3      11  P  W  *  M  S  G  C  T  S  D  P  G  A  C  K  A  F 14

         ---------|---------|---------|---------|---------|
      1051 TCATAGATGCACGGAACCTTTAGTCTGGTTTTTTCTTTTTTTTTGTTTTT 1100
F1      13 S  *  M  H  G  T  F  S  L  V  F  S  F  F  L  F  F  15
F2       3  H  R  C  T  E  P  L  V  W  F  F  L  F  F  C  F  F 19
F3      15   I  D  A  R  N  L  *  S  G  F  F  F  F  F  V  F   9

         ---------|---------|---------|---------|---------|
      1101 TTTTTAGGGATCGATTACACCTTTACTTCTATGCTTGGAGGACGGAACTC 1150
F1      16  F  *  G  S  I  T  P  L  L  L  C  L  E  D  G  T  R 15
F2      20   F  R  D  R  L  H  L  Y  F  Y  A  W  R  T  E  L   35
F3      10 F  L  G  I  D  Y  T  F  T  S  M  L  G  G  R  N  S  26

         ---------|---------|---------|---------|---------|
      1151 GGCCAATTGTTTAATATAATTTAGCATCAAGATCAGACTTGACTTTTGAA 1200
F1      16   P  I  V  *  Y  N  L  A  S  R  S  D  L  T  F  E   12
F2      36 G  Q  L  F  N  I  I  *  H  Q  D  Q  T  *  L  L  N  3
F3      27  A  N  C  L  I  *  F  S  I  K  I  R  L  D  F  *  T 1
```

Three frame translation of MG10732.4. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG10942 from 301 to 1500

```
         ---------|---------|---------|---------|---------|
       301 AAATGTGGCAGAGCTGCCGAGCGATAAGGATGCCCAGCATACAATTGTAG 350
F1      24 K  C  G  R  A  A  E  R  *  G  C  P  A  Y  N  C  R  8
F2      15  N  V  A  E  L  P  S  D  K  D  A  Q  H  T  I  V  E 31
F3       1   M  W  Q  S  C  R  A  I  R  M  P  S  I  Q  L  *   15

         ---------|---------|---------|---------|---------|
       351 AAAAGATATTCTTGGGGGATTGCGTTGAATGCTGCCAATCAAACAAATAG 400
F1       9  K  D  I  L  G  G  L  R  *  M  L  P  I  K  Q  I  A 8
F2      32   K  I  F  L  G  D  C  V  E  C  C  Q  S  N  K  *   46
F3       1 K  R  Y  S  W  G  I  A  L  N  A  A  N  Q  T  N  S  17

         ---------|---------|---------|---------|---------|
       401 CTCTTTTGCTACCAGGCTCAGGACCTGACGAACTATACCCAATTAATACC 450
F1       9   L  L  L  P  G  S  G  P  D  E  L  Y  P  I  N  T   24
F2       1 L  F  C  Y  Q  A  Q  D  L  T  N  Y  T  Q  L  I  P  17
F3      18  S  F  A  T  R  L  R  T  *  R  T  I  P  N  *  Y  Q 2

         ---------|---------|---------|---------|---------|
       451 AGGAAATAAATAAATGCGGATTTGTAAATAAAGATTAATCGTTTCAAAGG 500
F1      25 R  K  *  I  N  A  D  L  *  I  K  I  N  R  F  K  G  8
F2      18  G  N  K  *  M  R  I  C  K  *  R  L  I  V  S  K  V 7
F3       3   E  I  N  K  C  G  F  V  N  K  D  *  S  F  Q  R   4

         ---------|---------|---------|---------|---------|
       501 TTGAAGAAAGAGAGTGCTAAAATTAGAAGAAGAAAATAGAAATTAAACAT 550
F1       1  *  R  K  R  V  L  K  L  E  E  E  N  R  N  *  T  F 2
F2       8   E  E  R  E  C  *  N  *  K  K  K  I  E  I  K  H   8
F3       5 L  K  K  E  S  A  K  I  R  R  R  K  *  K  L  N  I  4

         ---------|---------|---------|---------|---------|
       551 TTAAACGTTCAAAAGTTAAAACTCTAAGTATAACTGACTAATAGTTAATA 600
F1       3   K  R  S  K  V  K  T  L  S  I  T  D  *  *  L  I   2
F2       9 L  N  V  Q  K  L  K  L  *  V  *  L  T  N  S  *  *  4
F3       1  *  T  F  K  S  *  N  S  K  Y  N  *  L  I  V  N  R 5

         ---------|---------|---------|---------|---------|
       601 GAGTTTGCTATAGTAAAGTATATGTAGCGTTGTAACCCCCGTGAAATTTA 650
F1       3 E  F  A  I  V  K  Y  M  *  R  C  N  P  R  E  I  Y  8
F2       1  S  L  L  *  *  S  I  C  S  V  V  T  P  V  K  F  T 12
F3       6   V  C  Y  S  K  V  Y  V  A  L  *  P  P  *  N  L   2

         ---------|---------|---------|---------|---------|
       651 CATATAGTGAAGATAGTTCAACTCACTTGTTCTTTACTGGTTATAGACTT 700
F1       9  I  *  *  R  *  F  N  S  L  V  L  Y  W  L  *  T  C 2
F2      13   Y  S  E  D  S  S  T  H  L  F  F  T  G  Y  R  L   28
F3       3 H  I  V  K  I  V  Q  L  T  C  S  L  L  V  I  D  L  19

         ---------|---------|---------|---------|---------|
       701 GTTCCAACTTTTCTTTTCATCTTTTTTTTTAAAAAAATACTAGCAGTTTA 750
F1       3   S  N  F  S  F  H  L  F  F  *  K  N  T  S  S  L   6
F2      29 V  P  T  F  L  F  I  F  F  F  K  K  I  L  A  V  Y  45
F3      20  F  Q  L  F  F  S  S  F  F  L  K  K  Y  *  Q  F  T 3
```


751 CTTTCAAATTTCTGTAAATTTTATCCATTTATTTATCGGCATAATGCGAG 800
F1 7 L S N F C K F Y P F I Y R H N A S 23
F2 46 F Q I S V N F I H L F I G I M R A 62
F3 4 F K F L * I L S I Y L S A * C E 2
801 CTTTTACAACCCTTTATTTTGTCGTTGGTCTGGTTGCTACAAAAGCGATG 850
F1 24 F Y N P L F C R W S G C Y K S D G 40
F2 63 F T T L Y F V V G L V A T K A M 78
F3 3 L L Q P F I L S L V W L L Q K R W 19
851 GCATTATTTGTCGGCTTTCCACAACAAAGTCAACCAGATGCGCCCCCGCG 900
F1 41 I I C R L S T T K S T R C A P A 56
F2 79 A L F V G F P Q Q S Q P D A P P R 95
F3 20 H Y L S A F H N K V N Q M R P R E 36
901 AGACCCAAGTATTCCCGAAACAGGAAACATCTGTTGCGCCCCCACGGGTG 950
F1 57 R P K Y S R N R K H L L R P H G C 73
F2 96 D P S I P E T G N I C C A P T G V 112
F3 37 T Q V F P K Q E T S V A P P R V 52
951 TAGCGGATCCCTCCCTGACTTGCAAAAATGCCGGATTAAACTCTTTTTGC 1000
F1 74 S G S L P D L Q K C R I K L F L R 90
F2 113 A D P S L T C K N A G L N S F C 128
F3 1 * R I P P * L A K M P D * T L F A 4
1001 GTAAGTCTTACTACCTGATATTACACAATCATGTCGACTCTTTATAACTG 1050
F1 91 K S Y Y L I L H N H V D S L * L 1
F2 129 V S L T T * Y Y T I M S T L Y N * 10
F3 1 * V L L P D I T Q S C R L F I T D 16
1051 ACATTTTCTTTAGTGTATCAACGCACGCAATGACTTTTTCGATCCCGATG 1100
F1 2 T F S L V Y Q R T Q * L F R S R W 6
F2 1 H F L * C I N A R N D F F D P D G 13
F3 17 I F F S V S T H A M T F S I P M 32
1101 GAGGGAAGGGTGGGTGTGATCGATTCACAAACTTTAATACTGGACGTTCG 1150
F1 7 R E G W V * S I H K L * Y W T F G 5
F2 14 G K G G C D R F T N F N T G R S 29
F3 33 E G R V G V I D S Q T L I L D V R 49
1151 GTTCAGAAATTTGTCCCCAACAGCCAGAAAACGTGTTTCTCCGGAAACGA 1200
F1 6 S E I C P Q Q P E N V F L R K R 21
F2 30 V Q K F V P N S Q K T C F S G N E 46
F3 50 F R N L S P T A R K R V S P E T R 66
1201 GGCTGGATTTATTGGATGTGCTTAGAAGTATGCTCGGGGGAATGGTCTTT 1250
F1 22 G W I Y W M C L E V C S G E W S L 38

```
F2      47  A  G  F  I  G  C  A  *  K  Y  A  R  G  N  G  L  C 9
F3      67   L  D  L  L  D  V  L  R  S  M  L  G  G  M  V  F   82

         ---------|---------|---------|---------|---------|
      1251 GTTTCCGGTTTCCTTTTTTTTTTTTTCGAAACGGGCCAAACAAAAGCTT 1300
F1      39  F  P  V  S  F  F  F  F  F  E  T  G  Q  T  K  A  L 55
F2      10   F  R  F  P  F  F  F  F  S  K  R  A  K  Q  K  L   25
F3      83 V  S  G  F  L  F  F  F  F  R  N  G  P  N  K  S  F  99

         ---------|---------|---------|---------|---------|
      1301 TACCGATAGTACAGCCTGTGGCTTTTGTCAATTGACTTGACTTTTATCGG 1350
F1      56   P  I  V  Q  P  V  A  F  V  N  *  L  D  F  Y  R   5
F2      26 Y  R  *  Y  S  L  W  L  L  S  I  D  L  T  F  I  G  14
F3     100  T  D  S  T  A  C  G  F  C  Q  L  T  *  L  L  S  V 4

         ---------|---------|---------|---------|---------|
      1351 TAAAAACTTGAATACAATGGTTCCAGCACAAGATGCAGCACCTGTAACCA 1400
F1       1 *  K  L  E  Y  N  G  S  S  T  R  C  S  T  C  N  Q  16
F2      15  K  N  L  N  T  M  V  P  A  Q  D  A  A  P  V  T  S 31
F3       5   K  T  *  I  Q  W  F  Q  H  K  M  Q  H  L  *  P   1

         ---------|---------|---------|---------|---------|
      1401 GTTGCGCTTTCTGTTTTTCCAACTTTGATTCGCAGTGGCGCCAAATTTCT 1450
F1      17  L  R  F  L  F  F  Q  L  *  F  A  V  A  P  N  F  * 7
F2      32   C  A  F  C  F  S  N  F  D  S  Q  W  R  Q  I  S   47
F3       2 V  A  L  S  V  F  P  T  L  I  R  S  G  A  K  F  L  18

         ---------|---------|---------|---------|---------|
      1451 AAATGACCCGGACATCACTCAAGAGCTGTTCAAGGCAGGGTAAGTAGATG 1500
F1       1   M  T  R  T  S  L  K  S  C  S  R  Q  G  K  *  M   1
F2      48 K  *  P  G  H  H  S  R  A  V  Q  G  R  V  S  R  *  14
F3      19  N  D  P  D  I  T  Q  E  L  F  K  A  G  *  V  D  D 3
```

Three frame translation of MG10942.4. Exonic regions are underlined. The start codon is located at nucleotide position 794. Intron consensus sequences are underlined. The Broad Institute annotation for MGG_10942.2 has the incorrect start codon, bad intron calls and other problems and predicts a 1028 amino acid polypeptide.

```
SHOWORF of MG13089.5 from 1 to 1050

           ---------|---------|---------|---------|---------|
         1 AGCTTAGCGTTTGTTCTGTTGAGAGTGTTTAGTCAGTTTTTCAATATTAT 50
F1       1 S  L  A  F  V  L  L  R  V  F  S  Q  F  F  N  I  I  17
F2       1  A  *  R  L  F  C  *  E  C  L  V  S  F  S  I  L  F 10
F3       1   L  S  V  C  S  V  E  S  V  *  S  V  F  Q  Y  Y   6

           ---------|---------|---------|---------|---------|
        51 TTATTAGTTAGCTTCTTTTTATAACATTCGCGCTGATAGTCATGTACCCT 100
F1      18  Y  *  L  A  S  F  Y  N  I  R  A  D  S  H  V  P  F 15
F2      11   I  S  *  L  L  F  I  T  F  A  L  I  V  M  Y  P   13
F3       7 L  L  V  S  F  F  L  *  H  S  R  *  *  S  C  T  L  4

           ---------|---------|---------|---------|---------|
       101 TTGGAAGCCCCGTTGATAGATATCCACTTTCAGATGGTCAGAGACATGGG 150
F1      16   G  S  P  V  D  R  Y  P  L  S  D  G  Q  R  H  G   31
F2      14 L  E  A  P  L  I  D  I  H  F  Q  M  V  R  D  M  G  30
F3       5  W  K  P  R  *  *  I  S  T  F  R  W  S  E  T  W  G 11

           ---------|---------|---------|---------|---------|
       151 GTTTGACATAAGTATATCACTATACTCTTCAAACCCAACGTATTTAAACC 200
F1      32 V  *  H  K  Y  I  T  I  L  F  K  P  N  V  F  K  P  15
F2      31  F  D  I  S  I  S  L  Y  S  S  N  P  T  Y  L  N  R 47
F3      12   L  T  *  V  Y  H  Y  T  L  Q  T  Q  R  I  *  T   1

           ---------|---------|---------|---------|---------|
       201 GCACCAAATTACGAGTTTTTTTTGCATTGGTTATCAAAGCCATAGCAAGT 250
F1      16  H  Q  I  T  S  F  F  C  I  G  Y  Q  S  H  S  K  F 32
F2      48   T  K  L  R  V  F  F  A  L  V  I  K  A  I  A  S   63
F3       2 A  P  N  Y  E  F  F  L  H  W  L  S  K  P  *  Q  V  2

           ---------|---------|---------|---------|---------|
       251 TCTTGACAAAAAAAACATCTTTAGTCACAGCTTTCAAAGTTCTTGAGGAA 300
F1      33   L  T  K  K  T  S  L  V  T  A  F  K  V  L  E  E   48
F2      64 S  *  Q  K  K  H  L  *  S  Q  L  S  K  F  L  R  N  9
F3       3  L  D  K  K  N  I  F  S  H  S  F  Q  S  S  *  G  T 2

           ---------|---------|---------|---------|---------|
       301 CTTCCCTTTATTGTTTCAGAGATATTACTCCTTTTCCTAACCAACTCTGT 350
F1      49 L  P  F  I  V  S  E  I  L  L  L  F  L  T  N  S  V  65
F2      10  F  P  L  L  F  Q  R  Y  Y  S  F  S  *  P  T  L  S 4
F3       3   S  L  Y  C  F  R  D  I  T  P  F  P  N  Q  L  C   18

           ---------|---------|---------|---------|---------|
       351 CCATTTCCTTCTATTAAGAAACACATACCTGCACAGGCATTCAGAGAAAA 400
F1      66  H  F  L  L  L  R  N  T  Y  L  H  R  H  S  E  K  N 82
F2       5   I  S  F  Y  *  E  T  H  T  C  T  G  I  Q  R  K   11
F3      19 P  F  P  S  I  K  K  H  I  P  A  Q  A  F  R  E  K  35

           ---------|---------|---------|---------|---------|
       401 ATGCGCAGCTTTTATTTCGCTCTCCTCTTGGCTCCAACCGCCGTCCTTTC 450
F1      83   A  Q  L  L  F  R  S  P  L  G  S  N  R  R  P  F   98
F2      12 M  R  S  F  Y  F  A  L  L  L  A  P  T  A  V  L  S  28
F3      36  C  A  A  F  I  S  L  S  S  W  L  Q  P  P  S  F  Q 52

           ---------|---------|---------|---------|---------|
       451 AGTTGAGATCAACATTAACCCTGGAACCGGAGAGCTATGCTGCGACCAGG 500
```

```
F1      99 S  *  D  Q  H  *  P  W  N  R  R  A  M  L  R  P  G  11
F2      29  V  E  I  N  I  N  P  G  T  G  E  L  C  C  D  Q  G 45
F3      53   L  R  S  T  L  T  L  E  P  E  S  Y  A  A  T  R   68

         ---------|---------|---------|---------|---------|
       501 GTACACCAGACAGTTCCGAATCCTGCAAAGGTTTAGGTTTGAACTCATAC 550
F1      12  Y  T  R  Q  F  R  I  L  Q  R  F  R  F  E  L  I  L 28
F2      46   T  P  D  S  S  E  S  C  K  G  L  G  L  N  S  Y   61
F3      69 V  H  Q  T  V  P  N  P  A  K  V  *  V  *  T  H  T  3

         ---------|---------|---------|---------|---------|
       551 TGTGTAAGTTGTGTAAAGATTATAACGGCGAGAAGCAAAGGAAAAAAAAT 600
F1      29   C  K  L  C  K  D  Y  N  G  E  K  Q  R  K  K  N   44
F2      62 C  V  S  C  V  K  I  I  T  A  R  S  K  G  K  K  I  78
F3       4  V  *  V  V  *  R  L  *  R  R  E  A  K  E  K  K  L 9

         ---------|---------|---------|---------|---------|
       601 TAATCACCCTTCTCCGCAGTGTTCCCAAGCTCGGAATGACAACCGCGGGG 650
F1       1 *  S  P  F  S  A  V  F  P  S  S  E  *  Q  P  R  G  4
F2      79  N  H  P  S  P  Q  C  S  Q  A  R  N  D  N  R  G  G 95
F3      10   I  T  L  L  R  S  V  P  K  L  G  M  T  T  A  G   25

         ---------|---------|---------|---------|---------|
       651 GATGTGATCCCCCCAGAATAGAAATCTTCAACGTCGGGCGCACTGTCACG 700
F1       5  M  *  S  P  Q  N  R  N  L  Q  R  R  A  H  C  H  V 15
F2      96   C  D  P  P  R  I  E  I  F  N  V  G  R  T  V  T   111
F3      26 D  V  I  P  P  E  *  K  S  S  T  S  G  A  L  S  R  10

         ---------|---------|---------|---------|---------|
       701 TCTTTTGTTCAGGGAGGCACTTGTAAAAGGACTGACTCGCAAAAGAATGT 750
F1      16   F  C  S  G  R  H  L  *  K  D  *  L  A  K  E  C   5
F2     112 S  F  V  Q  G  G  T  C  K  R  T  D  S  Q  K  N  V  128
F3      11  L  L  F  R  E  A  L  V  K  G  L  T  R  K  R  M  C 27

         ---------|---------|---------|---------|---------|
       751 GTATAATGCATTCATTGGGTGCGCAAAGTGATTTGGGTTTCAATCGGCAA 800
F1       6 V  *  C  I  H  W  V  R  K  V  I  W  V  S  I  G  K  15
F2     129  Y  N  A  F  I  G  C  A  K  *  F  G  F  Q  S  A  N 7
F3      28   I  M  H  S  L  G  A  Q  S  D  L  G  F  N  R  Q   43

         ---------|---------|---------|---------|---------|
       801 ATTGCATAGTGTGCACGGGCATATGCAGCGGCACTTGTCTTACGCGATAG 850
F1      16  L  H  S  V  H  G  H  M  Q  R  H  L  S  Y  A  I  G 32
F2       8   C  I  V  C  T  G  I  C  S  G  T  C  L  T  R  *   22
F3      44 I  A  *  C  A  R  A  Y  A  A  A  L  V  L  R  D  R  14

         ---------|---------|---------|---------|---------|
       851 GTATTATGGAGTGGCACGCCTGGGCGACTGCCAACCGTGACTAGCTCAAC 900
F1      33   I  M  E  W  H  A  W  A  T  A  N  R  D  *  L  N   2
F2       1 V  L  W  S  G  T  P  G  R  L  P  T  V  T  S  S  T  17
F3      15  Y  Y  G  V  A  R  L  G  D  C  Q  P  *  L  A  Q  P 4

         ---------|---------|---------|---------|---------|
       901 CTCACGTGACAAGGAGCCAGGCTAAACCTGTTATTGTTTGCTTAGGCGCT 950
F1       3 L  T  *  Q  G  A  R  L  N  L  L  L  F  A  *  A  L  2
F2      18  S  R  D  K  E  P  G  *  T  C  Y  C  L  L  R  R  L 9
F3       5   H  V  T  R  S  Q  A  K  P  V  I  V  C  L  G  A   20
```

```
          ---------|---------|---------|---------|---------|
       951 TGTACGCGATCGGCACTGCGTGCCTGTCTGTCTCTTTTATGCTTAGATTT 1000
F1       3  V  R  D  R  H  C  V  P  V  C  L  F  Y  A  *  I  S 2
F2      10   Y  A  I  G  T  A  C  L  S  V  S  F  M  L  R  F   25
F3      21 C  T  R  S  A  L  R  A  C  L  S  L  L  C  L  D  F  37

          ---------|---------|---------|---------|---------|
      1001 CAATTTTCGAATACAATCGCCCCTATGTTGCAGGGTCATCTTTTTGGCAT 1050
F1       3   I  F  E  Y  N  R  P  Y  V  A  G  S  S  F  W  H   18
F2      26 Q  F  S  N  T  I  A  P  M  L  Q  G  H  L  F  G  I  42
F3      38  N  F  R  I  Q  S  P  L  C  C  R  V  I  F  L  A  L 54
```

Three frame translation of MG13809.5. Exonic regions are underlined. The start codon is located at nucleotide position 401. Intron consensus sequences are underlined. The annotation of MGG_13809.5 at the Broad Institute is incorrect at the 3' splice site predicted for the intron.

MG13357.5 from 1 to 1237

```
           ---------|---------|---------|---------|---------|
         1 AAACTCTGGGACAAGGCATGCGAGAGTATCCACCGAGACCAAACATTTAA 50
F1       1 K  L  W  D  K  A  C  E  S  I  H  R  D  Q  T  F  K  17
F2       1  N  S  G  T  R  H  A  R  V  S  T  E  T  K  H  L  K 17
F3       1   T  L  G  Q  G  M  R  E  Y  P  P  R  P  N  I  *   15

           ---------|---------|---------|---------|---------|
        51 AGCCCATTAATAGCACATCAGCCCCATGGCCCCCATCAACCTTACAATTT 100
F1      18  A  H  *  *  H  I  S  P  M  A  P  I  N  L  T  I  S 13
F2      18   P  I  N  S  T  S  A  P  W  P  P  S  T  L  Q  F   33
F3       1 S  P  L  I  A  H  Q  P  H  G  P  H  Q  P  Y  N  F  17

           ---------|---------|---------|---------|---------|
       101 CTTACTGTGGCGGATATTTTTATATGCATAAATGGATTTACGCTCGCAAT 150
F1      14   Y  C  G  G  Y  F  Y  M  H  K  W  I  Y  A  R  N   29
F2      34 L  T  V  A  D  I  F  I  C  I  N  G  F  T  L  A  M  50
F3      18  L  L  W  R  I  F  L  Y  A  *  M  D  L  R  S  Q  C 7

           ---------|---------|---------|---------|---------|
       151 GTAATTACTCATGGACCGCAACAGGTCAAAGCACAAGAGACCCATTATGA 200
F1      30 V  I  T  H  G  P  Q  Q  V  K  A  Q  E  T  H  Y  D  46
F2       1  *  L  L  M  D  R  N  R  S  K  H  K  R  P  I  M  I 16
F3       8   N  Y  S  W  T  A  T  G  Q  S  T  R  D  P  L  *   22

           ---------|---------|---------|---------|---------|
       201 TAGCTACATGGGATCTAGGTACCCAACAGAACGAACTAGGTTTGTCGCAG 250
F1      47  S  Y  M  G  S  R  Y  P  T  E  R  T  R  F  V  A  D 63
F2      17   A  T  W  D  L  G  T  Q  Q  N  E  L  G  L  S  Q   32
F3       1 *  L  H  G  I  *  V  P  N  R  T  N  *  V  C  R  R  4

           ---------|---------|---------|---------|---------|
       251 ATTTGTTTTGCAAACCGGTTGTACGAACCACTTTCGGACACAAGAACGAG 300
F1      64   L  F  C  K  P  V  V  R  T  T  F  G  H  K  N  E   79
F2      33 I  C  F  A  N  R  L  Y  E  P  L  S  D  T  R  T  S  49
F3       5  F  V  L  Q  T  G  C  T  N  H  F  R  T  Q  E  R  V 21

           ---------|---------|---------|---------|---------|
       301 TATGTGAAAATGCAACGCTCGCTTAGCGTTGGTATATATCAAGCATTGGA 350
F1      80 Y  V  K  M  Q  R  S  L  S  V  G  I  Y  Q  A  L  D  96
F2      50  M  *  K  C  N  A  R  L  A  L  V  Y  I  K  H  W  T 15
F3      22   C  E  N  A  T  L  A  *  R  W  Y  I  S  S  I  G   8

           ---------|---------|---------|---------|---------|
       351 CTATATAACAACACAAATCGCCTGCATGAGATATCAATCATAACCTCTAG 400
F1      97  Y  I  T  T  Q  I  A  C  M  R  Y  Q  S  *  P  L  A 3
F2      16   I  *  Q  H  K  S  P  A  *  D  I  N  H  N  L  *   6
F3       9 L  Y  N  N  T  N  R  L  H  E  I  S  I  I  T  S  S  25

           ---------|---------|---------|---------|---------|
       401 CTTCCAAACACTTTGTCTGATCGCTGTGCGTACATCTCAGATTTCTTTTT 450
F1       4   S  K  H  F  V  *  S  L  C  V  H  L  R  F  L  F   10
F2       1 L  P  N  T  L  S  D  R  C  A  Y  I  S  D  F  F  F  17
F3      26  F  Q  T  L  C  L  I  A  V  R  T  S  Q  I  S  F  F 42

           ---------|---------|---------|---------|---------|
       451 TCTTGATACCCTATTGTTTTTATTTACTTCTCTCACGACCTGGCGTCAAC 500
```

```
F1      11 S  *  Y  P  I  V  F  I  Y  F  S  H  D  L  A  S  T  15
F2      18  L  D  T  L  L  F  L  F  T  S  L  T  T  W  R  Q  H 34
F3      43   L  I  P  Y  C  F  Y  L  L  L  S  R  P  G  V  N   58

        ---------|---------|---------|---------|---------|
       501 ATGCGTTCATCTACGATTATTCTGGCCCCTTTCCTTCTTTTCACTGGCCT 550
F1      16  C  V  H  L  R  L  F  W  P  L  S  F  F  S  L  A  W 32
F2      35   A  F  I  Y  D  Y  S  G  P  F  P  S  F  H  W  P   50
F3      59 M  R  S  S  T  I  I  L  A  P  F  L  L  F  T  G  L  75

        ---------|---------|---------|---------|---------|
       551 GGTTGCCGCCAAAGGCCCGAAGACAATCGAAGTACAGCCGGATTTCCAAG 600
F1      33   L  P  P  K  A  R  R  Q  S  K  Y  S  R  I  S  K   48
F2      51 G  C  R  Q  R  P  E  D  N  R  S  T  A  G  F  P  R  67
F3      76  V  A  A  K  G  P  K  T  I  E  V  Q  P  D  F  Q  G 92

        ---------|---------|---------|---------|---------|
       601 GTCCTCAGACGGGAGGTATTTGCTGCGACGCTGGTACAAATTCTGATACC 650
F1      49 V  L  R  R  E  V  F  A  A  T  L  V  Q  I  L  I  P  65
F2      68  S  S  D  G  R  Y  L  L  R  R  W  Y  K  F  *  Y  R 2
F3      93   P  Q  T  G  G  I  C  C  D  A  G  T  N  S  D  T   108

        ---------|---------|---------|---------|---------|
       651 GACAAGTTTTGCTCAGGCAACAATTTAAACGCTTTCTGCGTAAGCTGCAT 700
F1      66  T  S  F  A  Q  A  T  I  *  T  L  S  A  *  A  A  F 3
F2       3   Q  V  L  L  R  Q  Q  F  K  R  F  L  R  K  L  H   18
F3     109 D  K  F  C  S  G  N  N  L  N  A  F  C  V  S  C  I  125

        ---------|---------|---------|---------|---------|
       701 TTTTTTTCCTTTTTTTTCCCTTTTTTTATATCAAGCGACATGAAAAGCAT 750
F1       4   F  F  L  F  F  P  F  F  Y  I  K  R  H  E  K  H   19
F2      19 F  F  S  F  F  F  P  F  F  I  S  S  D  M  K  S  I  35
F3     126  F  F  P  F  F  S  L  F  L  Y  Q  A  T  *  K  A  S 3

        ---------|---------|---------|---------|---------|
       751 CAGCGGGGTAGAACTAACCTTCTTATAGTGCGGCCCCTTTCGCTCGGACC 800
F1      20 Q  R  G  R  T  N  L  L  I  V  R  P  L  S  L  G  P  36
F2      36  S  G  V  E  L  T  F  L  *  C  G  P  F  R  S  D  R 8
F3       4   A  G  *  N  *  P  S  Y  S  A  A  P  F  A  R  T   11

        ---------|---------|---------|---------|---------|
       801 GAAAGGGCGGCAAGGGAGTGCAGGGTGGATGTGACCCATTCCCAGATTTT 850
F1      37  K  G  R  Q  G  S  A  G  W  M  *  P  I  P  R  F  S 6
F2       9   K  G  G  K  G  V  Q  G  G  C  D  P  F  P  D  F   24
F3      12 E  R  A  A  R  E  C  R  V  D  V  T  H  S  Q  I  F  28

        ---------|---------|---------|---------|---------|
       851 CCAACAGGACGCAATGTTGTAACTTTCCCCCCCGGCAACCAACAATGCGT 900
F1       7   N  R  T  Q  C  C  N  F  P  P  R  Q  P  T  M  R   22
F2      25 P  T  G  R  N  V  V  T  F  P  P  G  N  Q  Q  C  V  41
F3      29  Q  Q  D  A  M  L  *  L  S  P  P  A  T  N  N  A  F 10

        ---------|---------|---------|---------|---------|
       901 TTCCTCCGGTGGGCATGCTGGATTTATTGGATGTGCATAAGTTGAGCGGA 950
F1      23 F  L  R  W  A  C  W  I  Y  W  M  C  I  S  *  A  E  2
F2      42  S  S  G  G  H  A  G  F  I  G  C  A  *  V  E  R  N 4
F3      11   P  P  V  G  M  L  D  L  L  D  V  H  K  L  S  G   26
```

```
          ---------|---------|---------|---------|---------|
       951 ATTGCCTGTTATTTCCTTGTGGATAAAATATGGGGATATCAGGACGCAAA 1000
F1       3  L  P  V  I  S  L  W  I  K  Y  G  D  I  R  T  Q  R 19
F2       5   C  L  L  F  P  C  G  *  N  M  G  I  S  G  R  K   8
F3      27 I  A  C  Y  F  L  V  D  K  I  W  G  Y  Q  D  A  K  43

          ---------|---------|---------|---------|---------|
      1001 GGTTTCCTCAAAAAGGCTGTATCATTTATCTAGACCTGTTTAATATCTAG 1050
F1      20   F  P  Q  K  G  C  I  I  Y  L  D  L  F  N  I  *   34
F2       9 G  F  L  K  K  A  V  S  F  I  *  T  C  L  I  S  S  6
F3      44  V  S  S  K  R  L  Y  H  L  S  R  P  V  *  Y  L  V 3

          ---------|---------|---------|---------|---------|
      1051 TATCTCCGTCATCCTTTGCCGCAGCCAGGAGTAATGTGCAGCTATCATGG 1100
F1       1 Y  L  R  H  P  L  P  Q  P  G  V  M  C  S  Y  H  G  17
F2       7  I  S  V  I  L  C  R  S  Q  E  *  C  A  A  I  M  A 6
F3       4   S  P  S  S  F  A  A  A  R  S  N  V  Q  L  S  W   19

          ---------|---------|---------|---------|---------|
      1101 CGTCCACAATATTATTATTATGGTGTGAGTACTGGAAAATTATGTACTTT 1150
F1      18  V  H  N  I  I  I  M  V  *  V  L  E  N  Y  V  L  L 8
F2       7   S  T  I  L  L  L  W  C  E  Y  W  K  I  M  Y  F   22
F3      20 R  P  Q  Y  Y  Y  Y  G  V  S  T  G  K  L  C  T  F  36

          ---------|---------|---------|---------|---------|
      1151 TGGGAGTCTCAGGTAGCCTGAGCAGGTGCAACGGTATCGAACTGTGATGC 1200
F1       9   G  V  S  G  S  L  S  R  C  N  G  I  E  L  *  C   1
F2      23 W  E  S  Q  V  A  *  A  G  A  T  V  S  N  C  D  A  10
F3      37  G  S  L  R  *  P  E  Q  V  Q  R  Y  R  T  V  M  R 12

          ---------|---------|---------|-------
      1201 GCAGTCGCTTTTGCTGCGTCAGTGATTTATTCCATAC 1237
F1       2 A  V  A  F  A  A  S  V  I  Y  S  I    13
F2      11  Q  S  L  L  L  R  Q  *  F  I  P  Y   4
F3      13   S  R  F  C  C  V  S  D  L  F  H     23
```

Three frame translation of MG13357.5. Exonic regions are underlined. The start codon is located at nucleotide position 501. Intron consensus sequences are underlined.

MG13601.5 from 1 to 1176

```
           ---------|---------|---------|---------|---------|
         1 CCCACGGACTAGCTCGGCAACCTTGAGAAGCGCCGAGCATGCTATATATA 50
F1       1 P  T  D  *  L  G  N  L  E  K  R  R  A  C  Y  I  *  12
F2       1  P  R  T  S  S  A  T  L  R  S  A  E  H  A  I  Y  S 17
F3       1   H  G  L  A  R  Q  P  *  E  A  P  S  M  L  Y  I   8

           ---------|---------|---------|---------|---------|
        51 GTATATTCAGATTGGCGCGGTGGCATCTTTTACAAACTAGTGTCTGAGTT 100
F1       1  Y  I  Q  I  G  A  V  A  S  F  T  N  *  C  L  S  W 4
F2      18   I  F  R  L  A  R  W  H  L  L  Q  T  S  V  *  V   1
F3       9 V  Y  S  D  W  R  G  G  I  F  Y  K  L  V  S  E  L  25

           ---------|---------|---------|---------|---------|
       101 GGGAGGTTTGGTTTCAAGATCGGAACGTGAAACAAGTGATTTTGGATCTT 150
F1       5   E  V  W  F  Q  D  R  N  V  K  Q  V  I  L  D  L   20
F2       2 G  R  F  G  F  K  I  G  T  *  N  K  *  F  W  I  F  4
F3      26  G  G  L  V  S  R  S  E  R  E  T  S  D  F  G  S  S 42

           ---------|---------|---------|---------|---------|
       151 CGGAATACTAATAAGGGGGGTTTTCTTAGGATCTTTATTACCTATACATA 200
F1      21 R  N  T  N  K  G  G  F  L  R  I  F  I  T  Y  T  *  36
F2       5  G  I  L  I  R  G  V  F  L  G  S  L  L  P  I  H  N 21
F3      43   E  Y  *  *  G  G  F  S  *  D  L  Y  Y  L  Y  I   7

           ---------|---------|---------|---------|---------|
       201 ACTATACAATGTAAAACAAAGCAGATCACTAAGGCCAAATTCAGTTTCAC 250
F1       1  L  Y  N  V  K  Q  S  R  S  L  R  P  N  S  V  S  H 17
F2      22   Y  T  M  *  N  K  A  D  H  *  G  Q  I  Q  F  H   6
F3       8 T  I  Q  C  K  T  K  Q  I  T  K  A  K  F  S  F  T  24

           ---------|---------|---------|---------|---------|
       251 ATGAATTAAAACTTAGCAAACTGAGACATACTGAGTTTTAATATTATAAG 300
F1      18   E  L  K  L  S  K  L  R  H  T  E  F  *  Y  Y  K   3
F2       7 M  N  *  N  L  A  N  *  D  I  L  S  F  N  I  I  S  9
F3       1  *  I  K  T  *  Q  T  E  T  Y  *  V  L  I  L  *  A 1

           ---------|---------|---------|---------|---------|
       301 CTTATCGAATACCCTAATTTCTTACATATTTAAAGTCTTCATTCTACCCA 350
F1       4 L  I  E  Y  P  N  F  L  H  I  *  S  L  H  S  T  H  6
F2      10  L  S  N  T  L  I  S  Y  I  F  K  V  F  I  L  P  T 26
F3       2   Y  R  I  P  *  F  L  T  Y  L  K  S  S  F  Y  P   11

           ---------|---------|---------|---------|---------|
       351 CTCTTCAACAGCTTTCTGTGCTATATTAATGGCTTGTATTCTTTTTTCCT 400
F1       7  S  S  T  A  F  C  A  I  L  M  A  C  I  L  F  S  C 23
F2      27   L  Q  Q  L  S  V  L  Y  *  W  L  V  F  F  F  P   7
F3      12 L  F  N  S  F  L  C  Y  I  N  G  L  Y  S  F  F  L  28

           ---------|---------|---------|---------|---------|
       401 GCAACTCTAAACGCAATTTTTCCCGAATTTCCATGTTACGATTTGCTGGC 450
F1      24   N  S  K  R  N  F  S  R  I  S  M  L  R  F  A  G   39
F2       8 A  T  L  N  A  I  F  P  E  F  P  C  Y  D  L  L  A  24
F3      29  Q  L  *  T  Q  F  F  P  N  F  H  V  T  I  C  W  P 14

           ---------|---------|---------|---------|---------|
       451 CAAATTTGAAAAAGAAAAAAAACCATAAAAACAATGCGCTCCTTCTCCCT 500
```

```
F1      40 Q  I  *  K  R  K  K  T  I  K  T  M  R  S  F  S  L  14
F2      25  K  F  E  K  E  K  K  P  *  K  Q  C  A  P  S  P  F 8
F3      15   N  L  K  K  K  K  N  H  K  N  N  A  L  L  L  P   30

            ---------|---------|---------|---------|---------|
        501 TTGTTTATCGCCCCTGCTGCAGTTTTTGGTGTAAATATCACCTTCGGCAG 550
F1      15  C  L  S  P  L  L  Q  F  L  V  *  I  S  P  S  A  G 6
F2       9   V  Y  R  P  C  C  S  F  W  C  K  Y  H  L  R  Q   24
F3      31 L  F  I  A  P  A  A  V  F  G  V  N  I  T  F  G  R  47

            ---------|---------|---------|---------|---------|
        551 GGGTAACACCGGTCTGCTGTGCTGCGATAGAGGCGCTCCCGGCCCGAGCA 600
F1       7   V  T  P  V  C  C  A  A  I  E  A  L  P  A  R  A   22
F2      25 G  *  H  R  S  A  V  L  R  *  R  R  S  R  P  E  Q  7
F3      48  G  N  T  G  L  L  C  C  D  R  G  A  P  G  P  S  K 64

            ---------|---------|---------|---------|---------|
        601 AGACCTGCACAGGTCTCAAGTTGAATTCTTATGGTGTAAGTGTTTTCTCT 650
F1      23 R  P  A  Q  V  S  S  *  I  L  M  V  *  V  F  S  L  4
F2       8  D  L  H  R  S  Q  V  E  F  L  W  C  K  C  F  L  S 24
F3      65   T  C  T  G  L  K  L  N  S  Y  G  V  S  V  F  S   80

            ---------|---------|---------|---------|---------|
        651 CAACGTTATATTAGATGAGGAAACAACTTGGCAACGGCGAATTAAAAAGC 700
F1       5  N  V  I  L  D  E  E  T  T  W  Q  R  R  I  K  K  L 21
F2      25   T  L  Y  *  M  R  K  Q  L  G  N  G  E  L  K  S   12
F3      81 Q  R  Y  I  R  *  G  N  N  L  A  T  A  N  *  K  A  2

            ---------|---------|---------|---------|---------|
        701 TAAACAACTCCTTCCAGTGCATTGATAGTCCCGCCGATGACGACTTTGGT 750
F1      22   N  N  S  F  Q  C  I  D  S  P  A  D  D  D  F  G   37
F2       1 *  T  T  P  S  S  A  L  I  V  P  P  M  T  T  L  V  16
F3       3  K  Q  L  L  P  V  H  *  *  S  R  R  *  R  L  W  W 4

            ---------|---------|---------|---------|---------|
        751 GGCTGCGACGGCATTACAAACTGGCCTATTGGACGGGATGTTAAAGCCTT 800
F1      38 G  C  D  G  I  T  N  W  P  I  G  R  D  V  K  A  F  54
F2      17  A  A  T  A  L  Q  T  G  L  L  D  G  M  L  K  P  S 33
F3       5   L  R  R  H  Y  K  L  A  Y  W  T  G  C  *  S  L   2

            ---------|---------|---------|---------|---------|
        801 CGAGCCCGGTAGCGTGGTGTCGCACACTCAAGCGGAAACCTTCAACATTG 850
F1      55  E  P  G  S  V  V  S  H  T  Q  A  E  T  F  N  I  E 71
F2      34   S  P  V  A  W  C  R  T  L  K  R  K  P  S  T  L   49
F3       3 R  A  R  *  R  G  V  A  H  S  S  G  N  L  Q  H  *  12

            ---------|---------|---------|---------|---------|
        851 AGGTTGGATTTGTTGGATGTGCGAAATAATTAATGATAAGGGACACTTGA 900
F1      72   V  G  F  V  G  C  A  K  *  L  M  I  R  D  T  *   6
F2      50 R  L  D  L  L  D  V  R  N  N  *  *  *  G  T  L  E  4
F3       1  G  W  I  C  W  M  C  E  I  I  N  D  K  G  H  L  R 17

            ---------|---------|---------|---------|---------|
        901 GAGAGATTAAGGGAGGGACAAATGAGGGTCATTAATACTTTTAGGTAGAC 950
F1       1 E  R  L  R  E  G  Q  M  R  V  I  N  T  F  R  *  T  1
F2       5  R  D  *  G  R  D  K  *  G  S  L  I  L  L  G  R  P 9
F3      18   E  I  K  G  G  T  N  E  G  H  *  Y  F  *  V  D   2
```

```
          ---------|---------|---------|---------|---------|
       951 CAACGGCATTGTAACTGTAAAGTACCTGGGTGCAACAGTGTACAAAGTTG 1000
F1       2  N  G  I  V  T  V  K  Y  L  G  A  T  V  Y  K  V  E 18
F2      10   T  A  L  *  L  *  S  T  W  V  Q  Q  C  T  K  L   10
F3       3 Q  R  H  C  N  C  K  V  P  G  C  N  S  V  Q  S  *  18

          ---------|---------|---------|---------|---------|
      1001 AAAGGGGACGAAACTGGGTAAAAGAAGGTTCTATTCTTCCGTTCATCTGT 1050
F1      19   R  G  R  N  W  V  K  E  G  S  I  L  P  F  I  C   34
F2      11 K  G  D  E  T  G  *  K  K  V  L  F  F  R  S  S  V  10
F3       1  K  G  T  K  L  G  K  R  R  F  Y  S  S  V  H  L  S 17

          ---------|---------|---------|---------|---------|
      1051 CAATATTGCATATGATCATTGGACACATCCCGGTCTCTGTGGGTCAACTT 1100
F1      35 Q  Y  C  I  *  S  L  D  T  S  R  S  L  W  V  N  L  12
F2      11  N  I  A  Y  D  H  W  T  H  P  G  L  C  G  S  T  Y 27
F3      18   I  L  H  M  I  I  G  H  I  P  V  S  V  G  Q  L   33

          ---------|---------|---------|---------|---------|
      1101 ACTGTTTGATACGTGCATACAGCGGGAAATTATACCGATGACTGTGATAA 1150
F1      13  L  F  D  T  C  I  Q  R  E  I  I  P  M  T  V  I  K 29
F2      28   C  L  I  R  A  Y  S  G  K  L  Y  R  *  L  *  *   1
F3      34 T  V  *  Y  V  H  T  A  G  N  Y  T  D  D  C  D  K  14

          ---------|---------|------
      1151 AGCAGGGGCTTTTACTACCTCCTCAT 1176
F1      30   Q  G  L  L  L  P  P  H   37
F2       1 S  R  G  F  Y  Y  L  L     8
F3      15  A  G  A  F  T  T  S  S    22
```

Three frame translation of MG13601.5. Exonic regions are underlined. The start codon is located at nucleotide position 485. Intron consensus sequences are underlined. A frameshift occurs at position at nucleotide 500 and an insertion of 1 nucleotide restores the reading frame. However, a 4 nucleotide insertion would fill the gap of a single amino acid in the alignment with all other members of the family. It is most likely that a four nucleotide deletion is responsible for the mutation.

VITA

Yue Shang

Peterson Building Room 120, Texas A&M University
College Station, Texas, 77843

Education

Master of Science in Plant Pathology and Microbiology, 2007
Texas A&M University, College Station, TX
Bachelor of Science in Biology, 2003
Lanzhou University, Lanzhou, P. R. China

Research Experience

Department of Plant Pathology and Microbiology, Texas A&M University
Research Assistant 2003-2006
College of Life Science, Lanzhou University, P. R. China
Research Assistant 2002-2003

Publication

Yue Shang, Zuhao Huang, and Naifa Liu (2003), "Molecular Evolution among Four Species of *Phasianidae* Based on mtDNA Sequence" ***Journal of Liaoning University***, 3: 275-280 (Chinese).

Poster

Yue Shang, Kiran Bhatterai, Guodong Lu, Hanno Wolf, Dan Li, and Daniel J. Ebbole, 'Functional Analysis of Secreted Proteins from *Magnaporthe grisea*' **The Meeting of the American Phytopathological Society**, Austin, Texas (2005).

Presentation

'Functional Analysis of Secreted Proteins from *Magnaporthe grisea*', **Program for the Biology of Filamentous Fungi (PBOFF),** Texas A&M University (2005)

Honors & Awards

Research Assistantship, Texas A&M University (2003-2006)
Third-Class Excellent Undergraduate Scholarship, Lanzhou University, P. R. China, 1999
First-Class Excellent Undergraduate Scholarship, Lanzhou University, P. R. China, 2000
First-Class Excellent Undergraduate Scholarship, Lanzhou University, P. R. China, 2001
Third-Class Excellent Undergraduate Scholarship, Lanzhou University, P. R. China, 2002

Leadership

Associate Vice President of **Chinese Students and Scholars Association**, Texas A&M University