9-2019

Identification and Cloning of Putative Serine Protease Inhibitor (Serpin) Genes in Rice (Oryza sativa) and a Preliminary Approach to Generate RNAi using the Cloned Sequences

Santanu De
Nova Southeastern University, sde@nova.edu

Identification and Cloning of Putative Serine Protease Inhibitor (Serpin) Genes in Rice (*Oryza sativa*) and a Preliminary Approach to Generate RNAi using the Cloned Sequences

Santanu De

ABSTRACT

Programmed Cell Death (PCD) is an important mechanism of plant immune response against diseases. Serine protease inhibitors ('serpins') are a conserved superfamily of proteins that inhibit serine protease targets and prevent programmed cell death (PCD) in plants, in absence of pathogen infection. In this project, putative serpin genes in rice (*Oryza sativa*) which are homologous to necrotic/'Nec' serpin genes in *Drosophila* are identified by bioinformatic analyses. This is followed by cloning of specific exon sequences of the rice serpin genes identified. Finally, one of the cloned sequences is utilized in a series of steps to produce interfering RNA (RNAi), to block the gene expression. Results of this study would help elucidate the importance of serpins in the regulation of PCD in rice. The work could lead to further research aimed at generating disease-resistant transgenic plants.

KEYWORDS:

Serine protease inhibitor, serpin, Nec, plant innate immunity, programmed cell death (PCD), cloning, RNAi disease-resistant, transgenic, rice.

INTRODUCTION

It has been studied for long that plants and animals have evolved elaborate mechanisms to combat the wide range of potential microbial pathogens that invade them. Animal pathogens often target and suppress programmed cell death (PCD) pathway components to manipulate their hosts. On the contrary, plant pathogens often trigger PCD (Greenberg & Yao, 2004). In cases where disease resistance is accompanied by plant PCD, an event called the hypersensitive response, the plant surveillance system has developed for detecting pathogen-secreted molecules in order to initiate a defense response.

Usually, most plant species are resistant to majority of potential microbial invaders. This phenomenon is known as 'non-host' or 'species' resistance/immunity (Numberger & Brunner, 2004). When a plant gets infected with a pathogen, it may rapidly activate an immune response to effectively prevent the pathogen from growing and developing. Defense mechanisms in plants against pathogens are of two main classes; the ones present constitutively, and those which are induced upon exposure to a pathogen. Many microbe-associated products called 'general elicitors' (including polypeptides, glycoproteins, lipids and oligosaccharides) exhibit proven ability to induce plant species-specific immunity upon infiltration into leaf tissue (Numberger & Scheel, 2001).

Plant defense mechanisms and Programmed Cell Death

Plant defense mechanisms comprise processes resulting from transcriptional activation of pathogenesis-related genes; for example, production of lytic enzymes (proteases, chitinases and glucanases), anti-microbial proteins (defensins) or anti-microbial secondary metabolites (phytoalexins) or, more importantly, extracellular production of reactive oxygen intermediates (ROI's) during oxidative burst of plants depending on transient increases of cytosolic calcium ion levels, catalyzed by an NADPH oxidase protein complex. The most prominent plant defense response is the frequently observed, highly localized, hypersensitive response (HR) similar to apoptotic (programmed) cell death in animal cells (Lam et al., 2001) and associated with a phenomenon termed the Resistance Response (RR). A RR is triggered when the host has a dominant *R* gene that corresponds to a dominant *avr* gene in the pathogen. This gene-for-gene interaction results from direct or

indirect interaction between the R and avr gene products, depending on the R-avr gene pair (Fluhr & Kaplan-Levy, 2002).

Intracellular signal transduction in plant innate immunity

Signal transduction cascades link recognition and defense responses through second messengers conserved among most eukaryotes. Activation of innate immune responses in plants are commonly signaled by changes in cytoplasmic calcium ion levels, production of ROIs and nitric oxide (NO) and post-translational activation of MAPK cascades. Signaling pathways mediating the innate immunity in insects, plants and animals have been reported to be much conserved (Ausubel, 2005).

The Toll receptor pathway in *Drosophila*

Toll is a transmembrane receptor present in *Drosophila*, with a leucine-rich repeat (LRR) domain. It exerts significant roles in embryonic development and in immune response. The Spaetzle gene product, a cytokine-like polypeptide, acts as an extracellular ligand (after being formed by cleavage of pre-Spaetzle), that binds to the Toll receptor and activates the Toll pathway in *Drosophila*. The recognition of pathogen-associated molecular patterns (PAMPs) like bacterial lipo-polysaccharides and fungal mannans, propagates an immune reaction by proteases regulated by serine protease inhibitors called serpins (Medzhitov & Janeway, 2002). The immune-induced cleavage of Spaetzle involving a proteolytic cascade has been described by the discovery of Spn43Ac, a hemolymph serpin, which downregulates the Toll pathway. A Toll-like domain homologous to the *Drosophila* Toll or the human interleukin receptor is present in the plant disease resistance genes (R genes). In insects and mammals, a family of conserved transmembrane Toll-like receptors (TLRs) functions as PRRs (pathogen- or pattern-recognition receptors) for microbe-associated molecules. A human homologue of Toll was recently cloned and shown to activate signal transduction through NF-κB, leading to the production of pro-inflammatory cytokines.

In animals, apoptosis often involves proteases called caspases (Green & Reed, 1998). Though clear homologues of caspases have not been found in the genome of the model plant *Arabidopsis*, caspase-like activities in plants have been noted biochemically or from inhibitor studies.

Serpins

The serpins (serine protease inhibitors) are a superfamily of proteins (350-500 amino acids in size) that fold into a conserved structure (**Fig. A**) and act as suicide-inhibitors i.e. the serpin-enzyme complex, once formed, is irreversible (Silverman et al., 2001). The superfamily includes antithrombin, α1-antitrypsin and PAI-1, which control coagulation, inflammation and fibrinolysis respectively. Serpins have evolved as a conformational trap (Pike et al., 2002): the reactive center loop (RCL, i.e. the region of the molecule that interacts with the target proteinase) is held exposed as "bait" for the target proteinase, which cleaves the RCL after docking. However, upon cleavage, the RCL rapidly inserts into the A β-sheet, "dragging" the proteinase with it and conformationally distorting the catalytic triad (and the entire molecule). This distortion prevents the final breakdown of the acyl-enzyme intermediate, leaving the proteinase "trapped" in limbo unable to escape.

Plant serpins inhibit serine protease targets and thus inhibit the pathway, thereby preventing PCD in absence of pathogen infection. Marked differences in basal expression levels of eight selected rice serpin genes during development have been reported, indicating a potential range of functions in regulation and in plant defense for the corresponding proteins (Francis et al., 2012). The study of serpin function using different biological platforms should help identify the role of these molecules in development, homeostasis and host defense.

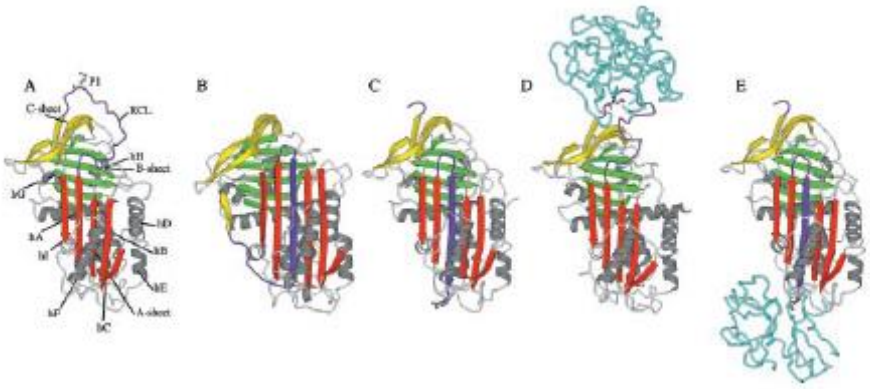


Fig. A. Serpin structures. From *left* to *right*: *A*, native _1AT (Protein Data Bank (PDB) entry 1QLP); *B*, latent ATIII (PDB entry 2ANT); *C*, cleaved _1AT (PDB entry 7API); *D*, Michaelis complex between Serpin 1 (Alaserpin from *Manduca sexta*) and trypsin (PDB entry 1I99); and *E*, covalent complex between _1AT and trypsin (PDB entry 1ezx). In all structures the A-sheet is in *red*, the B-sheet is in *green*, the C-sheet is in *yellow*, and the RSL(RCL) is in *purple*. The helices are in *gray* and are labeled on the structure of native _1AT. Trypsin is shown as a *cyan* coil.

The Drosophila 'Nec' gene

In *Drosophila*, a cluster of serpin transcripts called necrotic ('nec') gene is present at 43A on the second chromosome (Green et al, 2003). The Toll-mediated immune response is to fungal infections is constitutively activated in 'nec' mutants. The 'nec' protein comprises a serpin core having sequence homology with α1-antitrypsin, and a polyglutamine-rich N-terminal extension of 79 amino acids absent in other serpins. 'Nec' in *Drosophila* controls innate immunity and is homologous to human α1-antitrypsin (Green et al., 2003).

RNAi

When antisense RNA (aRNA) is introduced into a cell, it binds to the already present sense RNA to inhibit gene expression. The new sense RNA suppresses gene expression, similar to aRNA. While this may seem like a contradiction, it can be easily resolved by further examination. The cause is rooted in the prepared sense RNA. It turns out that preparations of sense RNA actually contain contaminating strands of antisense RNA. The sense and antisense strands bind to each other, forming a helix. This double helix is the actual suppressor of its corresponding gene. The suppression of a gene by its corresponding double stranded RNA is called RNA interference (RNAi), or post-transcriptional gene silencing (PTGS). The gene suppression by aRNA is likely also due to the formation of an RNA double helix, in this case formed by the sense RNA of the cell and the introduced antisense RNA.

Since the only RNA found in a cell should be single stranded, the presence of double stranded RNA signals is an abnormality. The cell has a specific enzyme (in *Drosophila* it is called Dicer) that recognizes the double stranded RNA and chops it up into small fragments between 21-25 base pairs in length. These short RNA fragments (called small interfering RNA, or siRNA) bind to the RNA-induced silencing complex (RISC). The RISC is activated when the siRNA unwinds and the activated complex binds to the corresponding mRNA using the antisense RNA. The RISC contains an enzyme to cleave the bound mRNA (called Slicer in *Drosophila*) and therefore cause gene suppression. Once the mRNA has been cleaved, it can no longer be translated into functional protein (**Fig. B**).

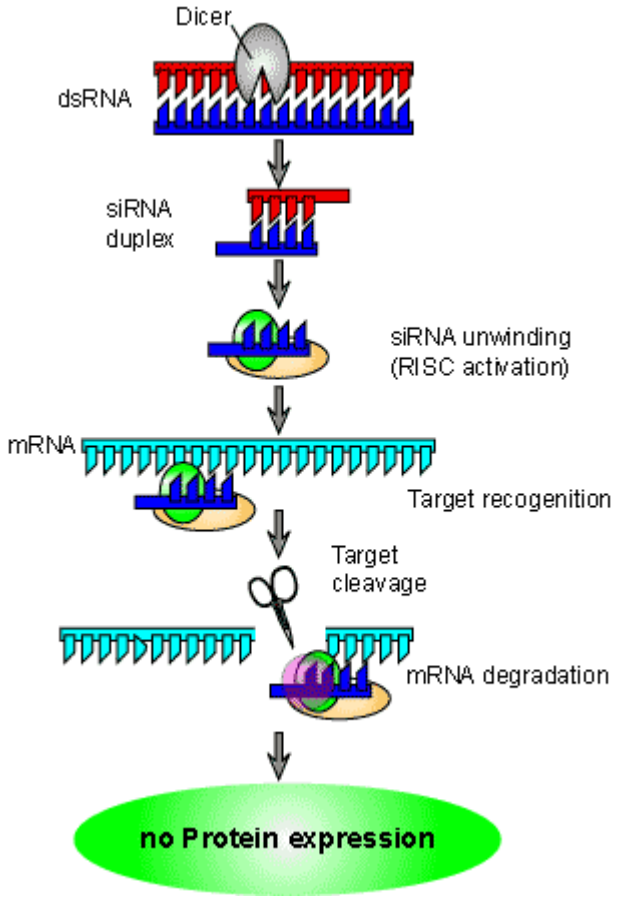


Fig. B: Mechanism of action of RNAi. Double stranded RNA is introduced into a cell and gets chopped up by the enzyme dicer to form siRNA. The siRNA then binds to the RISC complex and is unwound. The antisense RNA complexed with RISC binds to its corresponding mRNA which is the cleaved by the enzyme slicer rendering it inactive (Antler, 2003).

In this project, with the help of Bioinformatic analyses, 'nec'-homologous sequences in rice are identified and appropriate primers are designed to amplify specific exons of selected rice serpin genes, which then have to be cloned. Attempts would be made to generate a RNAi construct with the cloned sequences, in order to knockout and thus downregulate or inhibit expression of the corresponding gene in rice plants so as to identify its regulatory role in the plant's immune response to infection.

OBJECTIVES

The project is a part of the research aiming at identifying the signal transduction events during plant-pathogen interaction. Programmed Cell Death (PCD) in plants is intricately associated with plant defense responses. The model plant *Arabidopsis* is used to elucidate the role of a novel class of serine protease inhibitor (serpin) in regulating PCD associated with defense responses in one of the most economically important crops, rice.

The general objective of this study is to identify evolutionarily conserved serpin in rice involved in innate immunity, and to define its role in PCD and plant defense responses by under-expressing in transgenic plants, using the RNAi technique.

The area of this project involved identification of homologous sequences of the 'Nec' (Necrotic: a serpin gene found in *Drosophila*) gene in rice, by their homology with the Nec-homologue in *Arabidopsis.* The serpin family member genes in rice, homologous to Nec, will be found out by BLAST of rice genomic

sequence with *Arabidopsis* Nec homologue. By systemic approach, these have to be analyzed to note the candidate serpin gene involved in innate immunity in rice.

The general objectives of the work are as follows:

- Identification of evolutionary conserved rice serpin genes involved in innate immunity and programmed cell death by Bioinformatic tools and procedures.
- Construction of a phylogenetic tree for the genes identified.
- Primers will be designed to specifically amplify one exon of a given gene or a part of it and it will be made sure that the primers used do not bind with any other part of the rice genome.
- Total rice genomic DNA will be isolated and using the specific set of primers, a part of an exon from each serpin gene will be amplified.
- The amplified fragments will be cloned an analyzed for their orientation in the vector.
- Preliminary steps would be carried out to introduce the cloned fragment into a RNAi vector for downregulating the selected gene in rice via RNAi–mediated gene silencing, so as to determine the gene's function.

MATERIALS AND METHODS

The work was performed serially in the following order:

I. Designing primers to amplify a part of an exon of a specific gene by Bioinformatic studies.
II. Isolation of genomic DNA from rice.
III. Amplification of target sequences by PCR using designed primers.
IV. Cloning of target sequences in *E. coli.*
V. Plasmid isolation from transformed cells.
VI. Restriction digestion of isolated plasmids for confirmation of the insert.
VII. Restriction digestion for determination of orientation of the insert in the vector with respect to the T7 promotor in the vector.

I. Designing primers to amplify part of an exon of a specific gene through Bioinformatic analyses:

a) Identification of genomic, cDNA and protein sequences of the rice serpin genes homologous to 'nec' in *Arabidopsis*, from NCBI using BLAST (tblastn) function (*ref. i*).
b) Construction of phylogenetic tree for the cDNA sequences of the identified rice serpin genes using CLUSTAL-W function in 'Biology Workbench' software (to depict the extent of homology among the rice serpin gene sequences which were accordingly divided into 7 groups) (*ref. iii*).
c) Pairwise alignment of cDNA sequences with genomic sequences of rice serpin genes using ALIGN function in Biology Workbench software. This gives an idea of the number and distribution of exons in the particular gene, since the cDNA sequence includes only the exons (*ref. iii*).
d) Determination of exon lengths in TIGR database by subtracting the 5'-nucleotide number from the 3'-nucleotide number of the exon (*ref. ii*).
e) Designing a pair of primers unique for a part of an exon (cDNA sequence) of a selected gene in 'Biology Workbench' using PRIMER-3 function (*ref. iii*).

II. Isolation of genomic DNA from rice:

- The variety of rice used was *Oryza sativa* TP309 (Japonica) & *Oryza sativa* (indica) grown in laboratory green house.
- **Requirements:**

Extraction buffer:	Tris HCl (pH 7.5)	200mM
	EDTA (pH 8)	250mM
	NaCl	250mM
	SDS	0.5%

Liquid Nitrogen

Phenol:Chloroform	24:1
Isopropanol	
Ethanol	70%
TE Buffer	

- The rice DNA was isolated using the protocol given in *ref a*).

III. <u>Amplification of target sequences by Polymerase Chain Reaction using designed primers and isolated rice genomic DNA:</u>

Components of Master Mix:

Taq buffer	: 2.5 µL
Primer F	: 0.5 µL
R	: 0.5 µL
dNTP	: 0.5 µL
DNA	: 3.0 µL
Taq. pol	: 0.5 µL
Distilled H_20	: 17.5 µL
Total	: 25.0 µL

The PCR reactions were carried out with the following program, for 30 cycles:

Initial heating:	95°C for 4 minutes
Denaturation:	95°C for 20 seconds
Annealing:	70°C for 60 seconds
Extension:	72°C for 45 seconds
Final extension:	72°C for 10 minutes.

All the amplification products (of length 561 bp) were verified (using DNA ladder marker) by agarose (1.5%) gel electrophoresis (***Fig. 1***).

IV. <u>Cloning of the amplified target gene sequences of rice DNA in *E. coli*:</u>

- DH5α was the *E. coli* strain used.
- Fermentas InsT/A Clone PCR Product Cloning Kit #K1214 was used to clone the PCR products (*ref d*)) using the cloning vector pTZ57R/T.
- **LB media**, used for culturing bacterial cells, was prepared with:

Tryptone:	10 g/L
Yeast Extract:	5 g/L
NaCl:	10 g/L

The pH of LB media was adjusted to 7.0 with 5N NaOH using pH meter.
1.5% agar was added to prepare LB-agar. The antibiotic used in LB-agar plates for growth of transformed single colonies was ampicillin of concentration 1µl/ml.

<u>Ligation of the amplified gene sequence (PCR product) into cloning vector:</u>

The **ligation mix** was prepared with the following components-
-ligation buffer
-PEG (poly-ethylene glycol)
-plasmid vector pTZ57R/T
-PCR product
-T4 ligase
-sterilized distilled water.

The ligation mix was incubated at 22°C for 1 hour.

Preparation of competent cells:

Chemicals required:

1) 100 mM calcium chloride
2) 100 mM magnesium chloride.

E. coli DH5α cells were cultured by inoculating single colony of normal DH5α into LB-broth in laminar air-flow chamber and incubating overnight at 37°C in shaker incubator. They were made competent according to the protocol given in *ref. c*). The glycerol stocks were prepared and stored at -80ºC.

Transformation:

The prepared competent *E. coli* DH5α cells were transformed with the pre-incubated ligation mix (i.e. comprising the cloning vector ligated with the PCR product insert), following the protocol mentioned in *ref. c*) and transformation was confirmed by growth of single colonies on LB-agar plates containing ampicillin.

V. Plasmid isolation (minipreparation) from the transformed cells:

Chemicals required:

- Solution I (TGE): Tris, 25mM (pH 8)
  Glucose, 50mM
  EDTA, 10mM (pH 8)
- Solution II: SDS (1%) and NaOH (0.2N)
- Solution III: Potassium acetate (3M), pH 5.2
- Phenol:chloroform (24:1)
- Sodium acetate (3M), pH 5
- Isopropanol
- 70% Ethanol
- TE buffer [10:1 Tris-Cl (pH 8.0): EDTA (pH 8.0)].

The plasmids containing the expected insert were isolated from the transformed single colony *E. coli* DH5α cells (cultured by inoculating into LB-amp and incubating overnight in shaker incubator at 37°C) using the protocol given in *ref. c*). The isolated plasmids were analyzed by agarose (1.5%) gel electrophoresis (***Fig. 2***).

VI. Restriction digestion of the isolated plasmids for the confirmation of the insert:

By analyzing the multiple cloning sites of the plasmid vector pTZ57R/T, we chose EcoRI and BamHI for digestion of the isolated plasmid to release the expected insert. The reaction mixture included the following components:

Isolated plasmid DNA (with insert):	6.0 µl
Buffer (BamHI)	: 2.0 µl
BSA (100X)	: 0.2 µl
EcoRI	: 0.5 µl
BamHI	: 0.5 µl
Sterilized distilled water	: 10.8 µl
Total	: 20.0 µl

- The reaction mix was incubated at 37°C for 2 hours.
- Presence of the insert was confirmed by agarose (1.5%) gel electrophoresis (***Fig. 3***).

VII. <u>Restriction digestion for determination of the orientation of the insert in the vector relative to the T7 promoter present in the vector:</u>

"TAGC-Analysis of the nucleotide sequence for the presence of restriction enzyme sites", is a Bioinformatic tool in Biology workbench with the help of which we analyzed our cloned sequences for the presence of various restriction sites. Analyzing the enzyme profile, we identified the appropriate enzyme (Sac I) which can be used to digest the isolated plasmids with the insert, at two specific sites (one in the insert and the other in the MCS). The reaction mixture comprised:

DNA	5.0 µl
Buffer (Sac I)	6.0 µl
Sac I	1.2 µl
H_2O	37.8 µl
Total	50.0 µl

The reaction mix was then incubated at 37°C for 2 hours. The orientation of the insert (forward or reverse) in the cloning vector relative to the vector's promotor can be known from the predicted size of the products of digestion, confirmed by agarose (1.5%) gel electrophoresis (***Fig. 4***).

VIII. <u>Preliminary steps to produce an RNAi construct using the isolated insert:</u>

- The RNAi vector used was pTCK303.
- The antibiotic kanamycin (1µl/ml concentration) was used for growing single colony DH5α cells transformed with the pTCK303 vector.

Procedure:

1) Transformation of the *E. coli* DH5α cells with pTCK303 (*ref. c*) and verification by growth of single colonies in LB-kanamycin plates.
2) Isolation of plasmids (pTZ57R/T+11757 insert) (maxipreparation) from culture of transformed *E. coli* DH5α cells (*ref. a*).
3) Restriction digestion of the RNAi vector pTCK303 and the cloning vector pTZ57R/T with the 11757 insert with the restriction enzymes BamHI and KpnI, and analysis by agarose gel electrophoresis (*Figs. 5,6*).

IX. <u>Elution from gel and purification of the restriction digestion products (11757 insert and the cut pTCK303 vector) for further ligation to initiate RNAi construct formation, using the freeze-thaw technique</u> <u>(*ref. a*).</u>

RESULTS AND DISCUSSION

I. <u>Bioinformatic analyses and primer designing:</u>

a) i. <u>Identification of 'Nec' amino acid sequence in *Drosophila*:</u>

First, from the NCBI database, the amino acid sequence of the 'Nec' protein in *Drosophila* was noted to be as follows:

LOCUS AAF59285 **476 aa** linear INV 13- JAN-2006
DEFINITION CG1857-PA [*Drosophila melanogaster*].
ACCESSION AAF59285

ORIGIN
1 maskvsilll ltvhllaaqt faqeliawqr qqqqqqqqql qlqqqlllqq qqhqrnprpe
61 lglrslpgnp wtqnnqeais dvvavdltkr epvtpppnrp ppvfsymdrf sselfkeiik
121 sqsqqnvvfs pfsvhallal iygasdgktf relqkagefs knamavaqdf esvikykkhl
181 egadltlatk vyynrelggv nhsydeyakf yfsagteavd mqnakdtaak inawvmdttr

241 nkirdlvtpt dvdpqtqall vnavyfqgrw ehefatmdts pydfqhtngr iskvammfnd
301 dvyglaelpe lgatalelay kdsatsmlil lpnettglgk mlqqlsrpef dlnrvahrlr
361 rqsvavrlpk fqfefeqdmt eplknlgvhq mftpnsqvtk lmdqpvrvsk ilqkayinvg
421 eagteasaas yakfvplslp pkptefvanr pfvfavrtpa svlfighvey ptpmsv

ii. <u>Identification of 'Nec' homologue in *Arabidopsis*:</u>

Consequently, using the program **BLASTP- AA QUERY** in the TAIR database, the amino acid sequence of the protein in *Arabidopsis* (one of the most important experimental plant species) maximally homologous to 'Nec', was identified:

>P1;**At1g47710**
Serpin, putative
MDVRESISLQ NQVSMNLAKH VITTVSQNSN VIFSPASINV VLSIIAAGSA GATKDQILSF LKFSSTDQLN SFSSEIVSAV LADGSANGGP KLSVANGAWI DKSLSFKPSF KQLLEDSYKA ASNQADFQSK AVEVIAEVNS WAEKETNGLI TEVLPEGSAD SMTKLIFANA LYFKGTWNEK FDESLTQEGE FHLLDGNKVT APFMTSKKKQ YVSAYDGFKV LGLPYLQGQD KRQFSMYFYL PDANNGLSDL LDKIVSTPGF LDNHIPRRQV KVREFKIPKF KFSFGFDASN VLKGLGLTSP FSGEEGLTEM VESPEMGKNL CVSNIFHKAC IEVNEEGTEA AAASAGVIKL RGLLMEEDEI DFVADHPFLL VVTENITGVV LFIGQVVDPL H*
C; Length **391 aa**

iii. <u>Noting the genomic, cDNA and protein sequences of putative rice serpin genes:</u>

With the help of the program BLAST (tblastn) in the TIGR website, all the putative serpin family genes in rice homologous to 'Nec'-like protein in *Arabidopsis*, were identified. They were 22 in number with 4 overlapping sequences, thus rendering 18 independent genomic sequences. For each, the genomic, cDNA and protein sequences were retrieved from the TIGR database and saved, which, for example in case of the gene 05493, were found to be as follows:

<u>GENE: 05493</u>

Attributes	
Chromosome:	11
Coordinates (5' - 3'):	7416470 - 7415097 on assembly LOC_Os11g13530
Nucleotide length:	1275
Predicted protein length:	425
Predicted molecular weight:	45593.14
Predicted pI:	5.94

Genomic sequence length: 1374 nucleotides
CDS length: 1275 nucleotides
Protein length: 424 amino acids

Genomic sequence:
>11981.t01188
ATGGAAGACAACGCCGGCGACTGCGGCGGCATGACGGCCTTCGCGCTCCGCCTGGCGAAG

CGTCTCGCCGACGTCGGCGTCAGCAGCAACAAGAACCTCGTCTTCTCGCCGGCGTCCCTG
TACGCCGCGCTGGCGCTGGTGGCGGCCGGGGCACGGGGCACCACCCTGGACGAGCTCCTC
GCGCTGCTCGGCGCCGCGTCGCTCGACGACCTCGAGGAGTCCGTACGCCGCGCCGTGGAG
GTCGGCCTCGCCGACGAGTCCGCGTCCGGCGGGCCGCGCGTCTCCGACGCCTGCGGCGTC
TGGCACGACGAGACGCTCGAGCTTAAGCCGGCCTACCGCGCCGCCGCCGCCGGCACCTAC
AAGGCCGTGACGCGCGCCGCCAACTTCCAAAGACAGGTCAGTCAGTCCAGGGTTTCAATA
TGACGATTTGTGAAATTTCGACCAACATATGTAGTAAATTCGATCAATCTTGGTATGTAA
AACTTTCATATGCAGCCAAAGAGATCAAGAAAGAAGATCAACAAGTGGGTGTCCAAGGCG
ACGAACAAGCTCATCCCCGAGATCCTCCCAGATGGATCAGTTCATGTCGACACCGCCCTC
GTGCTCGTGAACGCCATCTACTTCAAGGGCAAATGGTCCAATCCCTTCCCCAGGTCGAGC
ACCACCACCGGCAAGTTCCACCGCCTCGACGGCAGCTCCGTCGACGTCCCGTTCATGAGC
AGCCGGGAGGATCAGTACATCGGCTTCCACGACGGCTTCACGGTGCTCAAGCTGCCGTAC
CACCACCGGACCATGAAGAACCATGGCGATGGTGGCGACACTATCACCAACAGTAGCATC
ACTCGGGCCATCCTCGAGCACTACGGCGGCGAGAACGTAGGGCTGTCGATGTACATCTTC
CTTCCGGACGAGCGCGACGGCCTGCCGGCGCTCGTCGACAAGATGGCGGCGTCGTCGTCG
TCGTCCTCCTTCCTGCGCGACCACCGGCCGACGCGGCGTCGCGAGGTCGGCGACCTCAGG
GTGCCGAGGTTCAAGGTGTCGTTCTACAGCCAGATCAACGGGGTTCTGCAGGGGATGGGG
GTGACGGCCGCCTTCGACGCCGGCGAGGCCGACCTGTCCGGCATGGCGGAGGGCGTGGAT
CAGCGGGGCGGCGGGCTGGTGGTGGAGGAGGTGTTCCACAGGGCGGTCGTGGAGGTGAAC
GAGGAAGGCACGGAGGCGGCGGCGTCCACGGCCTGCACGATCAGGCTCTTGAGCATGAGC
TACCCGGAGGACTTCGTCGCGGACCATCCGTTCGCGTTCTTCGTGGTGGAGGAGACGTCG
GGCGCGGTGCTCTTCGCCGGCCACGTCCTTGACCCAACAAGCAGCTCAGAGTGA

CDS:

>11981.m05493

ATGGAAGACAACGCCGGCGACTGCGGCGGCATGACGGCCTTCGCGCTCCGCCTGGCGAAG
CGTCTCGCCGACGTCGGCGTCAGCAGCAACAAGAACCTCGTCTTCTCGCCGGCGTCCCTG
TACGCCGCGCTGGCGCTGGTGGCGGCCGGGGCACGGGGCACCACCCTGGACGAGCTCCTC
GCGCTGCTCGGCGCCGCGTCGCTCGACGACCTCGAGGAGTCCGTACGCCGCGCCGTGGAG
GTCGGCCTCGCCGACGAGTCCGCGTCCGGCGGGCCGCGCGTCTCCGACGCCTGCGGCGTC
TGGCACGACGAGACGCTCGAGCTTAAGCCGGCCTACCGCGCCGCCGCCGCCGGCACCTAC
AAGGCCGTGACGCGCGCCGCCAACTTCCAAAGACAGCCAAAGAGATCAAGAAAGAAGATC
AACAAGTGGGTGTCCAAGGCGACGAACAAGCTCATCCCCGAGATCCTCCCAGATGGATCA
GTTCATGTCGACACCGCCCTCGTGCTCGTGAACGCCATCTACTTCAAGGGCAAATGGTCC
AATCCCTTCCCCAGGTCGAGCACCACCACCGGCAAGTTCCACCGCCTCGACGGCAGCTCC
GTCGACGTCCCGTTCATGAGCAGCCGGGAGGATCAGTACATCGGCTTCCACGACGGCTTC
ACGGTGCTCAAGCTGCCGTACCACCACCGGACCATGAAGAACCATGGCGATGGTGGCGAC
ACTATCACCAACAGTAGCATCACTCGGGCCATCCTCGAGCACTACGGCGGCGAGAACGTA
GGGCTGTCGATGTACATCTTCCTTCCGGACGAGCGCGACGGCCTGCCGGCGCTCGTCGAC
AAGATGGCGGCGTCGTCGTCGTCGTCCTCCTTCCTGCGCGACCACCGGCCGACGCGGCGT
CGCGAGGTCGGCGACCTCAGGGTGCCGAGGTTCAAGGTGTCGTTCTACAGCCAGATCAAC
GGGGTTCTGCAGGGGATGGGGGTGACGGCCGCCTTCGACGCCGGCGAGGCCGACCTGTCC
GGCATGGCGGAGGGCGTGGATCAGCGGGGCGGCGGGCTGGTGGTGGAGGAGGTGTTCCAC
AGGGCGGTCGTGGAGGTGAACGAGGAAGGCACGGAGGCGGCGGCGTCCACGGCCTGCACG
ATCAGGCTCTTGAGCATGAGCTACCCGGAGGACTTCGTCGCGGACCATCCGTTCGCGTTC
TTCGTGGTGGAGGAGACGTCGGGCGCGGTGCTCTTCGCCGGCCACGTCCTTGACCCAACA
AGCAGCTCAGAGTGA

Protein sequence:

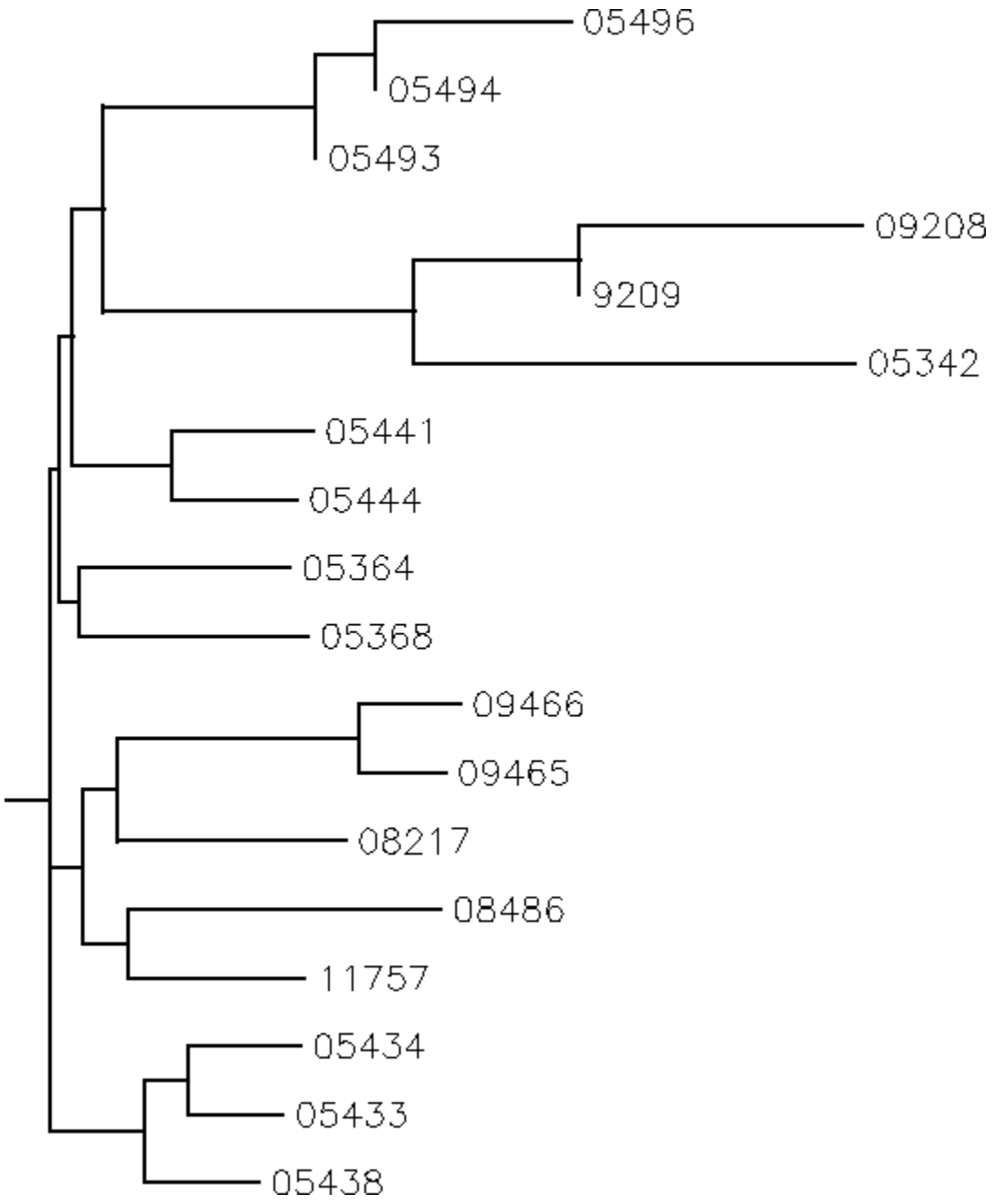
>11981.m05493

MEDNAGDCGGMTAFALRLAKRLADVGVSSNKNLVFSPASLYAALALVAAGARGTTLDELL
ALLGAASLDDLEESVRRAVEVGLADESASGGPRVSDACGVWHDETLELKPAYRAAAAGTY
KAVTRAANFQRQPKRSRKKINKWVSKATNKLIPEILPDGSVHVDTALVLVNAIYFKGKWS
NPFPRSSTTTGKFHRLDGSSVDVPFMSSREDQYIGFHDGFTVLKLPYHHRTMKNHGDGGD
TITNSSITRAILEHYGGENVGLSMYIFLPDERDGLPALVDKMAASSSSSSFLRDHRPTRR
REVGDLRVPRFKVSFYSQINGVLQGMGVTAAFDAGEADLSGMAEGVDQRGGGLVVEEVFH
RAVVEVNEEGTEAAASTACTIRLLSMSYPEDFVADHPFAFFVVEETSGAVLFAGHVLDPT
SSSE*

b) <u>Formation of Phylogenetic tree:</u>

With the help of the function CLUSTALW (also used for multiple sequence alignment) in SDSC-Biology Workbench software, the cDNA sequences of all the 18 putative rice serpin genes were used to generate a rooted and an unrooted genetic tree depicting the extent of sequence-homology among the sequences, as presented below.

ROOTED PHYLOGENETIC TREE (GENERATED BY PHYLIP'S DRAWGRAM)



The phylogenetic tree was used to divide the putative rice serpin gene superfamily into seven groups:

Grouping of the genes from the phylogenetic tree:

GROUP I - 05342, 09209, 09208

GROUP II - 05438, 05434, 05433

GROUP III - 05493, 05496, 05494

GROUP IV - 09466, 09465, 08217

GROUP V - 05444, 05441

GROUP VI - 05368, 05364

GROUP VII - 08486, 11757

c) Finding out the exon sequences of the putative rice serpin genes:

Since cDNA sequences (CDS) contain only exons (i.e. the coding regions) and no intron, the cDNA and the genomic sequences of all the Nec-homologous genes in rice were aligned pairwise using the ALIGN function in the SDSC-Biology Workbench software. This step provided all the exons present in the genomic sequence of each of those genes, which would further be required for designing primers for the genes, as shown below for the gene 05493.

Exon sequences of 05493 gene:

ATGGAAGACAACGCCGGCGACTGCGGCGGCATGACGGCCTTCGCGCTCCGCCTGGCGAAG
CGTCTCGCCGACGTCGGCGTCAGCAGCAACAAGAACCTCGTCTTCTCGCCGGC
GTCCCTGTACGCCGCGCTGGCGCTGGTGGCGGCCGGGGCACGGGGCACCACC
CTGGACGAGCTCCTCGCGCTGCTCGGCGCCGCGTCGCTCGACGACCTCGAGGA
GTCCGTACGCCGCGCCGTGGAGGTCGGCCTCGCCGACGAGTCCGCGTCCGGCGGGCCGCG
CGTCTCCGACGCCTGCGGCGTCTGGCACGACGAGACGCTCGAGCTTAAGCCGGCCTACCG
CGCCGCCGCCGCCGGCACCTACAAGGCCGTGACGCGCGCCGCCAACTTCCAAAGACAGCC
AAAGAGATCAAGAAAGAAGATCAACAAGTGGGTGTCCAAGGCGACGAACAAGCTCATCCC
CGAGATCCTCCCAGATGGATCAGTTCATGTCGACACCGCCCTCGTGCTCGTGAACGCCAT
CTACTTCAAGGGCAAATGGTCCAATCCCTTCCCCAGGTCGAGCACCACCACCGGCAAGTT
CCACCGCCTCGACGGCAGCTCCGTCGACGTCCCGTTCATGAGCAGCCGGGAGGATCAGTA
CATCGGCTTCCACGACGGCTTCACGGTGCTCAAGCTGCCGTACCACCACCGGACCATGAA
GAACCATGGCGATGGTGGCGACACTATCACCAACAGTAGCATCACTCGGGCCATCCTCGA
GCACTACGGCGGCGAGAACGTAGGGCTGTCGATGTACATCTTCCTTCCGGACGAGCGCGA
CGGCCTGCCGGCGCTCGTCGACAAGATGGCGGCGTCGTCGTCGTCGTCCTCCTTCCTGCG
CGACCACCGGCCGACGCGGCGTCGCGAGGTCGGCGACCTCAGGGTGCCGAGGTTCAAGGT
GTCGTTCTACAGCCAGATCAACGGGGTTCTGCAGGGGATGGGGGTGACGGCCGCCTTCGA
CGCCGGCGAGGCCGACCTGTCCGGCATGGCGGAGGGCGTGGATCAGCGGGGCGGCGGGCT
GGTGGTGGAGGAGGTGTTCCACAGGGCGGTCGTGGAGGTGAACGAGGAAGGCACGGAGGC
GGCGGCGTCCACGGCCTGCACGATCAGGCTCTTGAGCATGAGCTACCCGGAGGACTTCGT
CGCGGACCATCCGTTCGCGTTCTTCGTGGTGGAGGAGACGTCGGGCGCGGTGCTCTTCGC
CGGCCACGTCCTTGACCCAACAAGCAGCTCAGAGTGA

d) Determination of exon lengths:

In the TIGR database, the lengths of all the exons of the putative rice serpin genes were calculated by subtracting its 5'-end nucleotide number from the 3'-end nucleotide number. The following table reports the number of exons in each of these genes and the lengths of the corresponding exons.

GENE	NO. OF EXONS	EXON NO.	EXON LENGTH
9209	**3**	**1**	**482**
		2	**266**
		3	**684**
22765	**2**	**1**	**394**
		2	**878**
5493	**2**	**1**	**395**
		2	**878**
11757	**2**	**1**	**410**
		2	**1086**
5433	**3**	**1**	**144**
		2	**571**
		3	**815**
5494	**2**	**1**	**359**
		2	**1018**
22764	**2**	**1**	**359**
		2	**854**
5342	**2**	**1**	**413**
		2	**926**
8217	**1**	**1**	**1271**
5438	**2**	**1**	**422**
		2	**824**
5444	**2**	**1**	**401**
		2	**935**
5434	**3**	**1**	**150**
		2	**592**
		3	**830**
5441	**5**	**1**	**1424**
		2	**134**
		3	**71**
		4	**484**
		5	**1839**
5368	**1**	**1**	**1358**
9465	**2**	**1**	**344**
		2	**833**
9208	**5**	**1**	**410**
		2	**86**
		3	**128**
		4	**755**
		5	**906**
5364	**2**	**1**	**161**
		2	**854**
17941	**1**	**1**	**938**
8486	**1**	**1**	**938**
9466	**2**	**1**	**345**
		2	**811**
5496	**2**	**1**	**245**
		2	**299**
22762	**2**	**1**	**245**

		2	**299**

e) <u>Primer designing:</u>

Each of the rice serpin genes have to be amplified through PCR for further experimental procedures, by appropriately designing a forward and a reverse primer for it. My work included designing primers for the gene, 05493.

In Biology Workbench software, using the function "Primer 3- Formation of Primers and Probes", primers were designed for the exon (cDNA) sequence of the selected gene. All required inputs were supplied viz. number of primer sets to be generated, starting and ending nucleotide positions on the DNA for PCR amplification by the primer, product size, primer length, melting temperature etc. Accordingly, sets of left and right primer of the required length were shown. Considering each of these, such a set was chosen, for which, the region in that gene where the primers would bind, is maximally non-homologous with either of the other gene sequences in that group. This step was undertaken by consulting the pairwise alignment of the cDNA sequence of that gene 05493 with those of the other genes (05494 and 05496) in the same phylogenetic group (group III) (depicted below). It was done to ensure that the selected primer set would amplify only the gene under study and not any other.

Next, in NCBI database, the selected left and right primers of the gene were used to BLAST with the genomic sequence of rice (*Oryza sativa*), to mark the uniqueness of the primer set for the particular gene. A number of combinations were shown, presenting the probable binding of different lengths of the primer with complementary sequences of the rice genome. It was ensured if the whole length of the primer could bind with at least one portion of the genomic sequence <u>which had not been cloned previously</u>.

The study aims to inhibit the expression of the serpin family genes in rice to be verified by Northern analysis. Therefore, the primers are designed to specifically amplify the target sequences but the amplified product will hybridize to all the genes in the group, so that the expression of a given group of genes can be determined by a single Northern analysis.

<u>Pairwise alignment between cDNA sequences of genes 05493 and 05494:</u>

ALIGN calculates a global alignment of two sequences
version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
05493 1275 nt vs.
05494 1215 nt
scoring matrix: DNA, gap penalties: -16/-4
85.7% identity; Global alignment score: 4668

```
                10        20        30        40        50        60
05493
ATGGAAGACAACGCCGGCGACTGCGGCGGCATGACGGCCTTCGCGCTCCGCCTGGCGAA
G
          ::::::::: :::::::: :::::::::: ::::::: ::::::::::::::::::::
05494
ATGGAAGACGACGCCGGCAACTGCGGCGGCCTGACGGCGTTCGCGCTCCGCCTGGCGAA
G
                10        20        30        40        50        60

                70        80        90       100       110       120
05493
CGTCTCGCCGACGTCGGCGTCAGCAGCAACAAGAACCTCGTCTTCTCGCCGGCGTCCCTG
          :::::::::::::: ::::: :: :::::::: :::: ::::::::::::::: :::::::
```

```
05494
CGTCTCGCCGACGACGGCGACAACAGCAACAGGAACGTCGTCTTCTCGCCGGTGTCCCTG
                70       80       90      100      110      120

               130      140      150      160      170      180
05493
TACGCCGCGCTGGCGCTGGTGGCGGCCGGGGCACGGGGCACCACCCTGGACGAGCTCCT
C
         :::::::::::::::::::::::: :::::::::::::::::::::::::::::::: ::
05494
TACGCCGCGCTGGCGCTGGTGGCGTCCGGGGCACGGGGCACCACCCTGGACGAGCTCGTC
               130      140      150      160      170      180

               190      200      210      220      230      240
05493
GCGCTGCTCGGCGCCGCGTCGCTCGACGACCTCGAGGAGTCCGTACGCCGCGCCGTGGAG
         ::::::::::::::::::::::::::::::::::::::::: ::::::::::::::::::
05494
GCGCTGCTCGGCGCCGCGTCGCTCGACGACCTCGAGGAGTCAGTACGCCGCGCCGTGGAG
               190      200      210      220      230      240

               250      260      270      280      290      300
05493
GTCGGCCTCGCCGACGAGTCCGCGTCCGGCGGGCCGCGCGTCTCCGACGCCTGCGGCGTC
         ::::::::::::::::::::: :::::::::::::::::::::: ::::::::::::::
05494
GTCGGCCTCGCCGACGAGTCCGAGTCCGGCGGGCCGCGCGTCTCCTACGCCTGCGGCGTC
               250      260      270      280      290      300

               310      320      330      340      350      360
05493
TGGCACGACGAGACGCTCGAGCTTAAGCCGGCCTACCGCGCCGCCGCCGCCGGCACCTAC
         :::::::::::::: ::::: ::: ::::::::::::::::::::::
05494       TGGCACGACGAGAGGCTCGCGCTGAAGCCGGCCTACCGCGCCGCCG----------
               310      320      330      340

               370      380      390      400      410      420
05493
AAGGCCGTGACGCGCGCCGCCAACTTCCAAAGACAGCCAAAGAGATCAAGAAAGAAGAT
C
                          :::::::::::::::::::: ::::::::::::::
05494       ----------------------ACTTCCAAAGACAGCCAAAGAGTTCAAGAAAGAAGATC
                            350      360      370      380

               430      440      450      460      470      480
05493
AACAAGTGGGTGTCCAAGGCGACGAACAAGCTCATCCCCGAGATCCTCCCAGATGGATC
A
         :::::::::::::::::::::::::::::::::::: ::::::::::::::::::::
```

```
05494
AACAAGTGGGTGTCCAAGGCGACGAACAAGCTCATCCGCGAGATCCTCCCAGATGGATC
A
          390       400       410       420       430       440

             490       500       510       520       530       540
05493
GTTCATGTCGACACCGCCCTCGTGCTCGTGAACGCCATCTACTTCAAGGGCAAATGGTCC
       :::::::: :: :::::::::::::::::: ::::::::::::::::::::::::::::::
05494
GTTCATGGCGGCACCGCCCTCGTGCTCGTCAACGCCATCTACTTCAAGGGCAAATGGTCC
          450       460       470       480       490       500

             550       560       570       580       590       600
05493
AATCCCTTCCCCAGGTCGAGCACCACCACCGGCAAGTTCCACCGCCTCGACGGCAGCTCC
       ::::::::::::::::  ::: ::::::::::::::::::::::::::::::::::::::
05494
AATCCCTTCCCCAGGGAGAGAACCACCACCGGCAAGTTCCACCGCCTCGACGGCAGCTCC
          510       520       530       540       550       560
             610       620       630       640       650       660
05493
GTCGACGTCCCGTTCATGAGCAGCCGGGAGGATCAGTACATCGGCTTCCACGACGGCTTC
       :::::::  :::::::::::::::::::::::::::::::::::::::: ::::::::::::
05494
GTCGACGCTCCGTTCATGAGCAGCCGGGAGGATCAGTACATCGGCTTCTACGACGGCTTC
          570       580       590       600       610       620

             670       680       690       700       710       720
05493
ACGGTGCTCAAGCTGCCGTACCACCACCGGACCATGAAGAACCATGGCGATGGTGGCGA
C
       : ::::::::::::::::::::   :::::::::::::::::::::::::::::::: :::
05494       AAGGTGCTCAAGCTGCCGTACCA---
CCGGACCATGAAGAACCATGGCGATGGTGGTGAC
          630       640         650       660       670       680

             730       740       750       760       770       780
05493
ACTATCACCAACAGTAGCATCACTCGGGCCATCCTCGAGCACTACGGCGGCGAGAACGTA
       :               :::::: :::::::::: : :::::::::   ::::::::
05494       A------------------TCACTCCGGCCATCCTCAAACACTACGGC---GAGAACGTG
                            690       700       710         720

             790       800       810       820       830       840
05493
GGGCTGTCGATGTACATCTTCCTTCCGGACGAGCGCGACGGCCTGCCGGCGCTCGTCGAC
       ::: :::::::::::::::::: ::::::: ::::::::::::::::::::::::::
05494
GGGTTGTCGATGTACATCTTCCTGCCGGACGCGCGCGACGGCCTGCCGGCGCTCGTCGAC
```

```
              730       740       750       760       770       780

                850           860       870       880       890
05493         AAGATGGC---GGCGTCGTC------
GTCGTCGTCCTCCTTCCTGCGCGACCACCGGCCG
          ::::::::   :::::::::      : : :: :::::::::::::: ::::::::::::
05494
AAGATGGCAGTGGCGTCGTCCGGCACGGCCTCCTCCTCCTTCCTGCGTGACCACCGGCCG
              790       800       810       820       830       840

             900         910       920       930       940
05493         ACGCGGCGTCG---
CGAGGTCGGCGACCTCAGGGTGCCGAGGTTCAAGGTGTCGTTCTAC
           ::: :: ::   : :::::::::::: ::::::::::::::::::::::::::::::
05494
GGGCGACGGCGGATCAAGGTCGGCGACCTGAGGGTGCCGAGGTTCAAGGTGTCGTTCTA
C
              850       860       870       880       890       900

          950       960       970       980       990      1000
05493
AGCCAGATCAACGGGGTTCTGCAGGGGATGGGGGTGACGGCCGCCTTCGACGCCGGCGA
G
          ::: :::: :::: :::::::: :::::::::::: :   :::::::::::::: :::: ::
05494
AGCGAGATGAACGAGGTTCTGAAGGGGATGGGGATTGGGGCCGCCTTCGACGTCGGCAA
G
              910       920       930       940       950       960

         1010      1020      1030      1040      1050      1060
05493
GCCGACCTGTCCGGCATGGCGGAGGGCGTGGATCAGCGGGGCGGCGGGCTGGTGGTGGA
G
          : :::::: :::::::::   :: ::::      ::: ::         ::::::::::::
05494         GTCGACCTATCCGGCATGATCGACGGCG------AGCTGG---------TGGTGGTGGAG
              970       980           990             1000

         1070      1080      1090      1100      1110      1120
05493
GAGGTGTTCCACAGGGCGGTCGTGGAGGTGAACGAGGAAGGCACGGAGGCGGCGGCGTC
C
           ::::: : :::::::::::::::::::::::::::::::::::::::::::::::: ::
05494
AAGGTGATGCACAGGGCGGTCGTGGAGGTGAACGAGGAAGGCACGGAGGCGGCGGCGG
CC
          1010      1020      1030      1040      1050      1060

         1130      1140      1150      1160           1170      1180
05493         ACGGCCTGCACGATCAGGCTCTTGAGCATGAG------
CTACCCGGAGGACTTCGTCGCG
```

```
          :::::::::::::::: : : ::::: :  : :       ::  :: : ::::::::::::::
05494
ACGGCCTGCACGATGAAGTTCTTGTGTTTAACTTTAACCTCTCCTGTGGACTTCGTCGCG
          1070     1080     1090     1100     1110     1120

            1190     1200     1210     1220     1230     1240
05493
GACCATCCGTTCGCGTTCTTCGTGGTGGAGGAGACGTCGGGCGCGGTGCTCTTCGCCGGC
          :::::::::::::::::::::::::::::::::::: :::::  :::::::::::: :: :::
05494
GACCATCCGTTCGCGTTCTTCGTGGTGGAGGAGAAGTCGGATGCGGTGCTCTTTGCTGGC
          1130     1140     1150     1160     1170     1180

            1250     1260     1270
05493       CACGTCCTTGACCCAACAAGCAGCTCAGAGTGA
          ::::::::::::::::::::   ::  :::: :
05494       CACGTCCTTGACCCAACAAG---CTTGGAGTAA
          1190     1200       1210
```

Similarly, the cDNA sequences of 05493 & 05496 genes were aligned pairwise.

Primer Design

GENE USED FOR PRIMER DESIGNING: 05493

	START	LENGTH	Tm	GC%	
LEFT PRIMER	715	21	60.05	52.38	(GGCGACACTATCACCAACAGT)
RIGHT PRIMER	1175	20	60.10	55.00	(AAGTCCTCCGGGTAGCTCAT)

PRODUCT SIZE: 461

II. Amplification of targeted exon sequences after rice DNA isolation:

The rice DNA was isolated and used to amplify the target exon sequence (belonging to 11757 gene) through PCR using primers designed beforehand. The amplification products were confirmed by agarose (1.5%) gel electrophoresis using DNA ladder mix as the marker. The results are as follows (**Fig.1**).

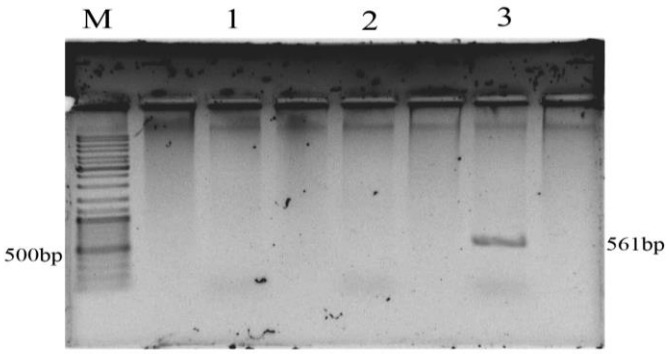


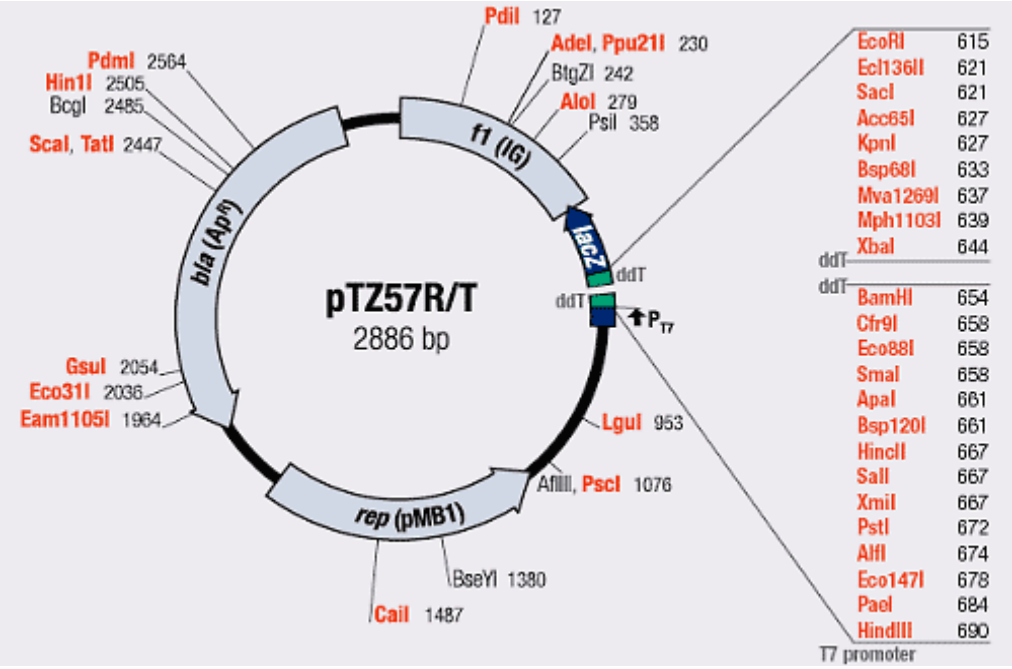
Fig.1: PCR amplification of 11757-exon

Lane- 1- TP309 DNA
2- Pusa basmati DNA
3- TP309 DNA M-Marker

III. Cloning of the PCR-amplified target sequences in *E. coli*:

Ligation of amplified sequences into cloning vector, preparation of competent *E. coli* cells and their transformation:

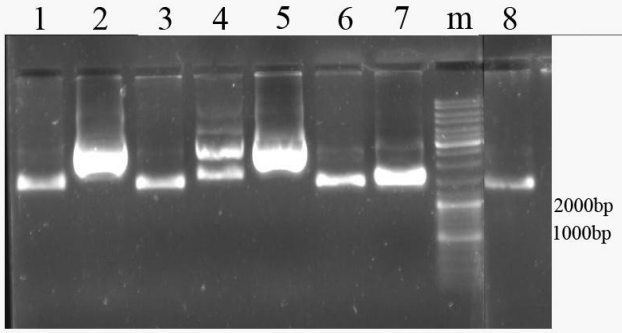
The nucleotide sequences amplified by the PCR have a 3' "A" overhang. Hence, the T/A cloning vector pTZ57R/T (2886bp, provided by the Fermentas InsT/Aclone PCR Product Cloning Kit #K1214) is designed with a T overhang, so as to ligate the amplified product directly without any post PCR treatment. The amplified fragments of both the genes were ligated with the vector molecule by individual ligation reactions. *E. coli* DH5α cells cultured in LB-broth by overnight incubation in shaker incubator at 37°C, were made competent (as mentioned before). They were then transformed with the cloning vector ligated with the expected insert (*ref. c*) and plated on LB-amp plates for overnight incubation. A total of 8 single colonies grew in the LB-amp plates. Cells from different colonies were taken, inoculated in culture tubes with LB-amp, incubated at 37°C overnight and used for plasmid isolation.



Map of cloning vector pTZ57R/T

IV. Plasmid isolation (miniprep.) from the transformed *E. coli* DH5α cells:

The plasmids containing the expected insert were isolated from the transformed *E. coli* DH5α cells (*ref. c*). The result of an agarose (1.5%) gel electrophoresis of the isolated plasmids is presented in **Fig. 2**.

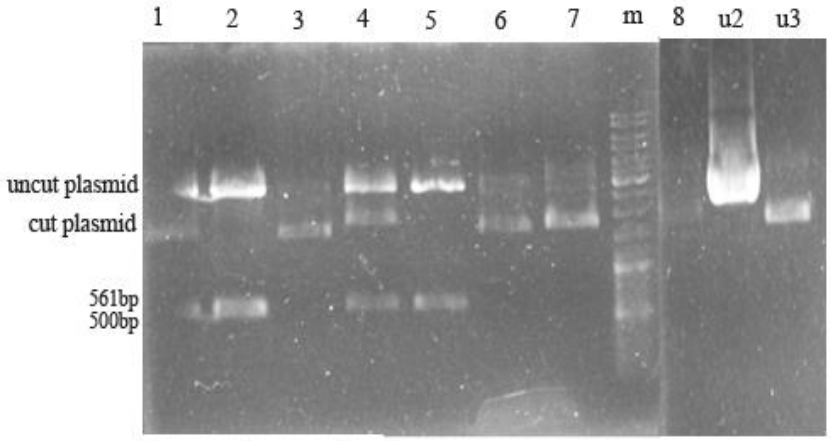


Lanes 1-8: isolated plasmids (pTZ57R/T+11757 insert)
m:marker

Fig.2: Agarose Gel electrophoresis of plasmids isolated from DH5α transformants (pTZ57R/T+11757 insert)

V. Restriction digestion of the plasmids isolated for confirmation of the insert:

The isolated plasmids were restriction digested with BamHI and EcoRI restriction enzymes (selected by analyzing the MCS of the plasmid vector pTZ57R/T) expected to produce the cut vector and the 11757 insert. The result of the digestion was reflected in agarose (1.5%) gel electrophoresis (**Fig. 3**).



Lanes u2,u3: uncut plasmids
m: marker
1-8: isolated plasmids (pTZ57R/T+11757 insert) digested by EcoR1 & BamH1

Fig.3: Agarose gel electrophoresis after restriction digestion of isolated plasmids (pTZ57R/T+11757 insert) by EcoR1 &BamH1 for confirmation of the insert

VI. Restriction analysis to identify the direction of the insert in the cloning vector with respect to the T7 promotor:

By analyzing the cloned target sequences for the available restriction sites (as presented below) using the "TAGC-Analysis of the nucleotide sequence for the presence of restriction enzyme sites" function in Biology Workbench, we identified the appropriate enzyme (Sac I) which can be used to digest the isolated plasmids containing the insert, at two specific sites (one in the insert and the other in the MCS).

Total Number of Hits per Enzyme:

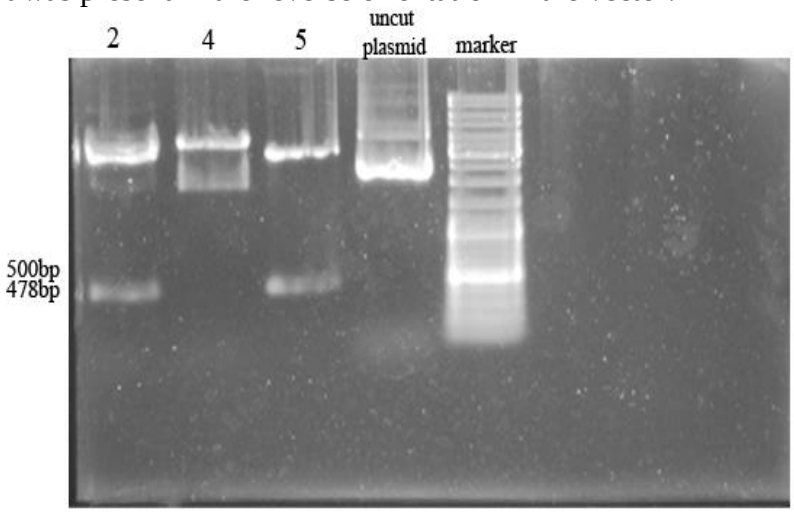
AlwI	1	BsrI	1	Hin4I	2	RsaI	7
ApaI	2	BsrBI	2	HinP1I	31	**SacI**	2
AscI	1	BsrFI	6	HincII	1	SacII	5

....and so on. Of all the restriction enzyme sites, only those for SacI were chosen for reasons discussed below.

Sites Cut by the Enzyme:

AatII	G_ACGT'C (0 Err) - 1 Cut(s)	591				
SacI	G_AGCT'C (0 Err) - 2 Cut(s)	9	833			
SacII	CC_GC'GG (0 Err) - 5 Cut(s)	160	201	283	712	718

....and so on. In order to check the orientation of the insert we had to choose an enzyme that has a single restriction site in the nucleotide sequence and also a same one in the vector MCS. It was found that the enzyme SacI has a restriction site at the 405bp position in the 561bp length of the 11757 insert. The vector also has the SacI restriction site in the 621bp position in the vector's MCS. If the insert is in the forward orientation relative to the T7 promotor of the vector, the restriction digestion will produce a fragment of approximately 178bp but a fragment of approximately 428bp if in reverse direction. The plasmids containing the 11757 inserts were digested with the SacI enzyme. An agarose (1.5%) gel electrophoresis of the digested products showed the presence of the band at approximately 428bp position (**Fig. 4**). Therefore, it was concluded that the insert was present in the reverse orientation in the vector.



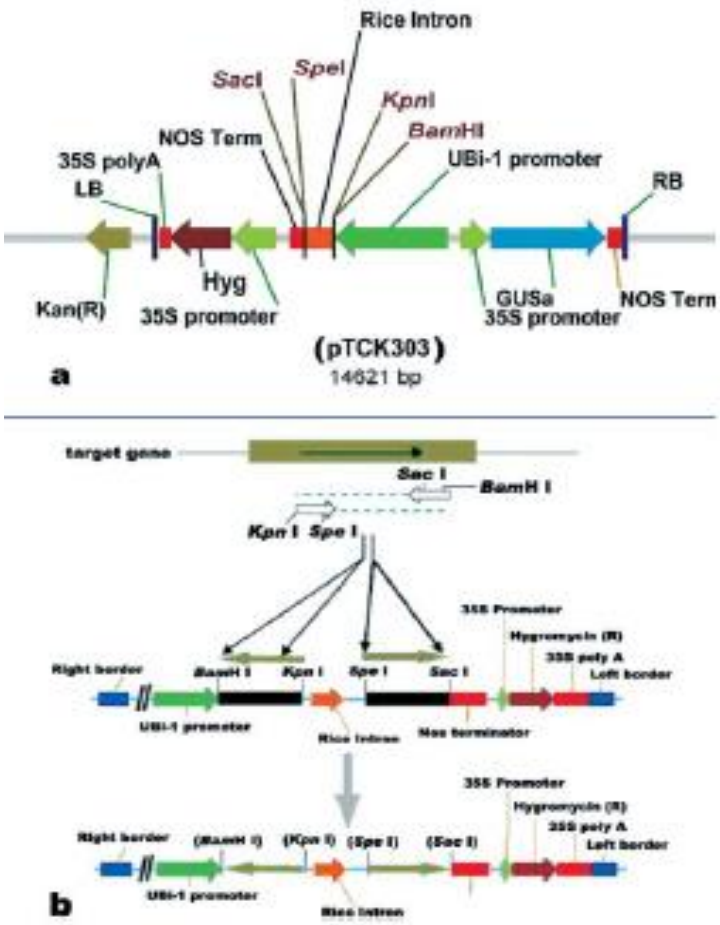
Lanes 2,4,5: plasmids from colonies 2,4,5
with 11757 insert, restriction digested with Sac1

Fig. 4: Agarose gel eletrophoresis after restriction digestion of isolated plasmids (pTZ57R/T+11757 insert) by Sac1

VII. Preliminary approach to generate a RNAi using the cloned insert:

The RNAi vector used was pTCK303 (as shown below). It is a PCR-based RNAi vector (14621bp) with a maize ubiquitin promoter, and a rice intron flanked by a specific multiple cloning site on either of its two ends, which is specifically designed for monocot gene silencing. With this vector, only one PCR product (but

to be inserted in opposite orientation in the vector with respect to the promotor), amplified by a single pair of primers, and two ligation reactions, are needed to create an RNAi construct.



(a) The core structure of RNAi (RNA interference) vector pTCK303.
(b) An example of the application of vector pTCK303 in RNAi construct.

1) <u>Transformation of the DH5α cells with pTCK303 and verification by growth of single colonies on LB-kanamycin plates:</u>

The *E. coli* DH5α cells were transformed with the vector pTCK303 (*ref. c*) and cultured overnight on LB-kanamycin plates, after which, the single colonies that grew, were further cultured for plasmid isolation (maxiprep.) (*ref. a*).

2) <u>Isolation of plasmids (pTZ57R/T+11757 insert) (maxipreparation) from culture of transformed *E. coli* DH5α cells:</u>

The DH5α colonies that were confirmed to have been transformed with the insert in the reverse direction with respect to the T7 promotor in pTZ57R/T (Fig. 4), were cultured in LB media, and used for plasmid isolation from them (maxiprep.) using the protocol as in *ref. a*) and verified by agarose gel electrophoresis.

3) <u>Restriction digestion of the RNAi vector pTCK303 and the pTZ57R/T vector containing the 11757 insert, with the restriction enzymes BamHI and KpnI:</u>

By comparing the available maps of both the vectors, the enzymes BamHI and KpnI were chosen for inserting the cloned segment in the forward orientation with respect to the ubiquitin promoter in pTCK303

vector. Therefore, both the vectors were digested with these enzymes and the results were analyzed by agarose gel electrophoresis (***Figs. 5, 6***).

Lane 1: undigested pTCK303 (14621bp)
Lane 2: pTCK303 digested with BamH1 and Kpn1
Fig.5: Agarose (1%) gel electrophoresis after restriction digestion of RNAi vector pTCK303 with BamH1 and Kpn1

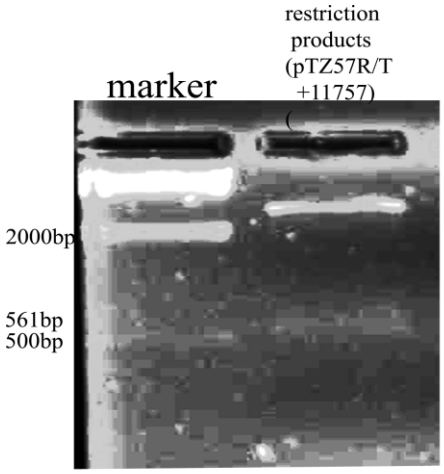


Fig. 6: Agarose(1%) gel electrophoresis after restriction digestion of isolated (maxiprep) plasmid (pTZ57R/T with 11757 insert) by BamHI & KpnI for confirmation of insert

VIII. <u>Elution from gel and purification of the restriction digestion products (11757 insert and the cut pTCK303 vector) for further ligation to initiate RNAi construct formation, using the freeze-thaw technique</u> (*ref. a*)**<u>:</u>**

Both the products of the restriction digestion mentioned above were run in agarose (1%) gel followed by separation of the corresponding bands containing the insert and the cut pTCK303 vector with a scalpel, which were then purified by the freeze-thaw technique as given in *ref. a*) for the ligation of the insert into the RNAi vector in order to generate the prospective RNAi construct.

SUMMARY AND CONCLUSION

Programmed Cell Death (PCD) plays a crucial role in plant immune response against diseases. Though widely observed, it has not been studied extensively at either biochemical or genetic level. Hence, analyzing the factors that regulate the process would help explain various aspects of PCD and plant-pathogen interaction. Taking advantage of *Arabidopsis* and rice genomic data, sequence homology of a particular gene family to a serpin (serine protease inhibitor) was noted, that is conserved in evolution.

The study began with identification of rice genomic sequences homologous to 'nec' (a protein in the serpin superfamily) in *Drosophila*, with the help of the 'nec' homologue found in *Arabidopsis.* Through bioinformatic analyses, primers unique for a selected gene exon in the rice serpin superfamily were designed, using which the exon was amplified by PCR of genomic DNA isolated from rice leaves. After that, the amplification products (i.e. the amplified target sequence) were cloned in *E. coli* DH5α cells. Plasmids containing the insert were isolated from the transformed *E. coli* cells and used for restriction digestion with suitable restriction enzymes (analyzed from the map of the vector used) for the confirmation of the insert. The orientation of the inserts in each clone with respect to the promotor in the vector were then verified by restriction digestion analyses. Preliminary steps required to introduce the insert (with its identified direction in the vector) extracted from the isolated plasmids into an appropriate RNAi vector were performed, as preparatory to the successful construction of an interfering RNA (RNAi). Using RNAi, knock-out lines of the selected gene would be produced in transgenic rice plants to show the potential regulatory role of the specific rice serpin gene in plant-pathogen interaction related to PCD.

The project reports bioinformatic analyses required for identification of the putative serpin genes in rice and the preliminary experiments required to make an RNAi construct of a single gene in the serpin superfamily. Results of this study would help a better understanding of the evolutionarily conserved innate immunity pathways regulated by serpin genes in rice. The research indicates a potential strategy to generate disease-resistant transgenic plants.

REFERENCES

Articles:

1) Greenberg, J.T, & Yao, N. (2004). The role and regulation of programmed cell death in plant pathogen interactions. *Cellular Microbiology, 6*(3):201-11.
2) Numberger, T., & Brunner, F. (2004). Innate immunity in plants and animals: Striking similarities and obvious differences. *Immunological Reviews, 198*:249-266.
3) Numberger, T., & Scheel, D. (2001). Signal transmission in the plant immune response. *Trends Plant Science, 6*:372-379.
4) Lam, E., Kato, N., & Lawton, M. (2001). Programmed cell death, mitochondria and the plant hypersensitive response. *Nature, 411*:848-853.
5) Fluhr, R. & Kaplan-Levy, R.N. (2002). Plant disease resistance: commonality and novelty in multicellular innate immunity. *Curr Top Microbiol Immunol., 270*:23-46.
6) Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J., & Whisstock, J.C. (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. *The Journal of Biological Chemistry,* 276(36):33293-6.
7) Francis S.E., Ersoy, R.A., Ahn, J., Atwell, B.J., & Roberts, T.H. (2012). Serpins in rice: protein sequence analysis, phylogeny and gene expression during development. *BMC Genomics, 13*: 449.
8) Green, C., Brown, G., Daffron, T.R., Reichhaart, J., Morley, T., Lomas, D.A., & Gubb, D. (2003). Drosophila necrotic mutations mirror disease-associated variants of human serpins. *Development 130*: 1473-1478.
9) Pike, R.N., Bottomley, S.P., Irving, J.A., Bird, P.I., Whisstock, J.C. (2002). Serpins: Finely Balanced Conformational Traps. *IUBMB Life, 54(1):1-7.*
10) Ausubel, M.F. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nature Immunology, 6*(10):973-9.
11) Janeway, C.A., Jr. & Medzhitov, R. (2002). Innate Immune Recognition. *Annu. Rev. Immunol., 20*:197–216.
12) Green, D.R. & Reed, J.C. (1998). Mitochondria and apoptosis. *Science, 281*(5381):1309-12.
13) Antler, C. (2003). Antisense RNA. *The Science Creative Quarterly*.

Books and Kits:

a) Molecular Cloning: A Laboratory Manual: Sambrook and Russell, 3rd edition, CSHL press.
b) Molecular Biology of The Cell – Baltimore, Lodish, Darnell.
c) Fermentas Ins T/A Clone PCR Product Cloning Kit #K1214.

Other Internet Resources:

i. http://www.ncbi.nih.gov
ii. http://www.tigr.org/tdb/e2k1/osa1
iii. http://workbench.sdsc.edu