BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(1)

# In search of function for hypothetical proteins encoded by genes of SA-JA pathways in *Oryza sativa* by *in silico* comparison and structural modeling

**Indra Singh[1]*, Pragati Agrawal[1], Kavita Shah[2]**

[1]Bioinformatics, Mahila Mahavidyalaya, Banaras Hindu University, Varanasi-221 005, India; [2]Associate Professor, Environmental Biochemistry and Bioinformatics, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi-221 005, India. Indra Singh – Email: singhanjanabhu@gmail.com; Phone: +91-542-9307856111; * Corresponding author

Received December 01, 2011; Accepted December 05, 2011; Published January 06, 2012

**Abstract:**
Knowledge of rice genome brings new dimensions to the management of abiotic stresses; however, gene sequences in the rice genome are yet to be assigned structure and function. Hydrogen peroxide, salicylates and jasmonates act as signal molecules in plants employing common machinery to manage abiotic stress. The present work is primarily focused to assign a structure-function relationship by modeling of the hypothetical proteins of SA-JA signaling pathway known in *Arabidopsis thaliana* and compare them with corresponding proteins in rice *in silico*. Thirteen known gene sequences with their encoded proteins for SA/JA pathway in model plant *A. thaliana* were obtained and similar gene sequences from rice were retrieved at NCBI. Five rice gene sequences Os09g0392100, Os03g0233200, OsJ_33269, OsJ_23610 and Os01g0194300 resulted in hypothetical protein products with unknown structure and function. Modeling and comparison of 5 proteins from rice and *Arabidopsis* showed 73 - 98% identity with acceptable RMSD values of 0.6 - 1.7 upon superimposition. Results suggest conserved nature of these proteins during evolution. The hypothetical protein from rice contains similar functional protein domain as that in *A. thaliana* and therefore are likely to perform similar functions in rice. There is a cross talk between the genes in SA/JA pathway wherein Os09g0392100 or EDS1, Os03g0233200 or PR5, OsJ_33269 or PAD4 and OsJ_23610 or SFD-1 activates the pathway and Os01g0194300 or NPR1 inhibit the pathway. Further investigation through wet-lab experiments are in progress to look into suppression/activation of the genes of SA-JA signaling in rice plants exposed to abiotic stress.

**Keywords:** Abiotic stress, Gene ontology, Jasmonic acid, Rice genome, Salicylic acid

**Background:**
With the availability of rice genomic information **[1]** it is possible to exploit gene sequences from rice varieties, for stress management and enhanced crop yield **[2].** Rice plants respond to various abiotic and biotic stresses via a number of mechanisms **[3].** Salicylic acid and Jasmonic acid have been reported to act as signal molecules in abiotic stress amelioration in plants **[4-9].** Protective action of SA in the presence of abiotic stress is tied with cell membrane stabilization **[7]** and suppression of antioxidant enzyme activities **[9]**. Jasmonic acid a derivative of linolenic acid **[10]** is involved in the plant defense mechanism through a family of peptide ligands **[11-13]** resulting from enhanced antioxidant levels **[3]**. Since plants response to abiotic stress, salicylic acid and jasmonic acid is characterized by the altered formation of reactive oxygen species, therefore it is quite likely that genes involved in SA and

 © 2012 Biomedical Informatics

JA pathways collide in plants under abiotic stress. The authors used the gene sequences reported in SA/JA pathway from model plant *Arabidopsis thaliana* and searched for similar genes in rice genome. Surprisingly most of the resultant sequences for SA/JA pathway obtained upon similarity search from rice were unannotated. The present study therefore aims to annotate the selected unknown gene sequences of rice having similarity with *A. thaliana* genes involved in SA/JA pathway by using gene ontology, homology search, structural modeling and superimposition tools *in silico*, so as to better understand role of these genes in SA/JA pathway in rice plants.

**Methodology:**

***Sequence retrieval and data analysis***

Nucleotide sequences were obtained from National Centre for Biotechnology Information (http://ncbi.nlm.nih.gov). Full genome of model plant *Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000) **[14]** was initially chosen and retrieved from the nucleotide database at NCBI and the protein coding genes identified in the cross linking pathway of salicylate and jasmonate were retrieved from NCBI's database **[15].** The corresponding proteins of the genes retrieved from Arabidopsis were checked for the presence of functional domains and protein signatures using InterPro Scan sequence search **[16]** of EMBI (European Bioinformatics Institute).

***Prediction of Gene Ontology***

The ontology of the genes involved in SA/JA pathway was decided using TAIR (The Arabidopsis Information Resource), a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana* **[14].** Data available from TAIR includes the complete genome sequence along with gene structure, gene product information etc which is updated every two weeks from the latest published research literature and community data submissions. Protein products for the thirteen genes were obtained and their sequences were retrieved. Shown in SA/JA pathway **Figure 1** and listed in **Table A (see supplementary material).**

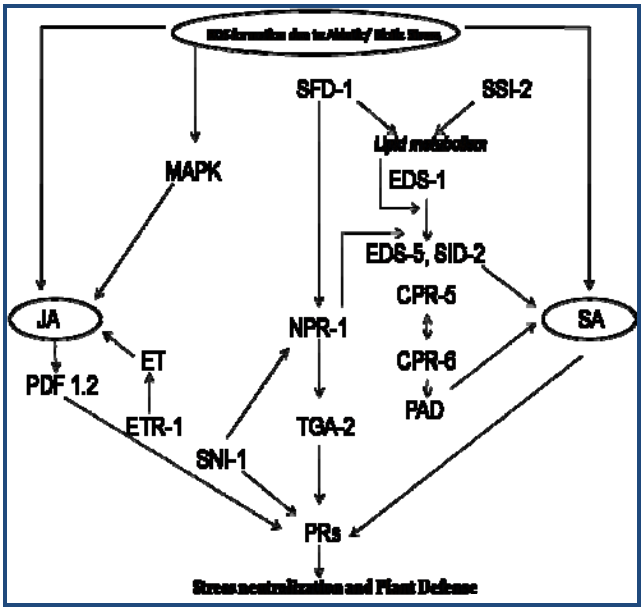



**Figure 1:** Genes involved in signaling via salicylate and jasmonate derivatives in plants exposed to abiotic/biotic stress and plant defense. EDS-1: Enhanced Disease Susceptibility1; CPR-6: Constitutive expressor of PR genes 6; PAD-4: Phytoalexin Deficient 4; EDS-5: Enhanced Disease Susceptibility 5; SID-2: Salicylic-acid-Induction Deficient 2; TGA-2: TGA-element binding protein 2; SFD-1: Suppressor of Fatty-acid-desaturase Deficiency 1; PR: Pathogenesis-Related gene.; CPR-5:Constitutive expressor of PR genes 5; NPR-1: Non-expressor of PR1; SNI-1: Transcriptional repressor of SAR genes; SSI-2: Suppressor of SA-Insensitivity.

***Homology Search and Structural Modeling***

Sequences obtained above were used as query for searching homologous sequences in *Oryza sativa* using BLASTp **[17].** Seven protein sequences (homologous) in rice that had not been assigned any function till date were obtained. Of these five proteins of *O. sativa* and *A. thaliana* were then modeled for their respective 3-dimensional structures using Geno3d **[18].** Proteins of *O. sativa* and corresponding protein of *A. thaliana* were superimposed **[19]** for structural similarity based upon RMSD, percent identity, and other parameters. The structures were viewed using UCSF Chimera tool **[20]** an interactive molecular graphics program with high extensibility, interactive molecular visualization and analysis system. The sequence for each protein from *O. sativa* and *A. thaliana* for best 20 structures were obtained.

***Validation of various physiological parameters***

The obtained proteins were also checked using ProtParam tool of ExPASy Proteomics Server [21]. This allows the computation of various physical and chemical parameters like theoretical pI, hydropathicity, estimated half-life, etc. for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence.

**Result & Discussion:**

***SA/JA pathway genes of Arabidopsis thaliana and Gene ontology prediction***

Overall 13 genes involved in SA/JA pathway were retrieved from *A. thaliana* database at NCBI and 13 homologous gene sequences were obtained for each gene from *O. sativa* genome by BLAST algorithm. Of these five gene sequences OSJ_014658, Os04g0689900, OsI_04142, OsJ_33269, OSJ_23610 and Os01g0194300 from rice having >70% identity and corresponding to genes of *A. thaliana* TAIR ID: AT3G48090, AT1G75040, AT3G52430, AT2G40690, AT1G64280 respectively were selected, which resulted in hypothetical proteins whose structure and functions are not known. Gene ontology study of *A. thaliana* genes using Interpro domain search program predicted presence of conserved gene specific domains and their function in SA-JA cross linking **(Table 1)** and this is still not reported in rice.

***Retrieval of predicted cross linking genes from O. sativa genome***

Protein known in *O. sativa* to that of *A. thaliana* were screened using BLOSUM62 matrix. Protein sequences with more than 70% identity with Arabidopsis genome were selected in O. sativa. Of which OSJ_014658, Os04g0689900, OsI_04142, OsJ_33269, OSJ_23610 and Os01g0194300 corresponding to EDS1, PR5, PAD4, SFD1, and NPR1 genes (TAIR ID: AT3G48090, AT1G75040, AT3G52430, AT2G40690, AT1G64280 respectively) of *A. thaliana* respectively were obtained as hypothetical protein and are still unannotated..

 © 2012 Biomedical Informatics

***Structural modeling of proteins of A. thaliana and O. sativa involved in SA/JA pathway***

Sequence similarity could be a measure for functional similarity; however for two proteins to be functionally similar it is important that the two proteins are also structurally similar. Model were created for the known proteins of SA/JA pathway retrieved for *A. thaliana* and corresponding proteins (BAF25000.1, NP_001049473.1, EAZ17725.1, EEE66834.1, NP_001042286) of *O. sativa*. Five unknown protein sequences of *O. sativa* were modeled for their structure using Geno3d tertiary structure prediction tool which resulted in three PDB files for each of the query protein, out of which the most appropriate structural coordinate file was chosen based on R-value, B-factor, Ramachandran plot etc. (data not included). The *O. sativa* proteins were aligned structurally with corresponding *A. thaliana* protein and their superimposed structures were evaluated **Table 1 (See supplementary material).** Reasonably good RMSD values were obtained that lie between 0.6-1.7 suggesting proper structural and functional similarities between the proteins of the two species. The functional similarity was further validated by Protparam **(Table B, see supplementry material)** that suggest the rice proteins to have similar estimated half life of >20 hrs (in yeast), a varied hydropathicity and all had a theoretical pI ranging between 4.75 and 9.83.

**Conclusion:**

In conclusion results suggest that similar to that in *Arabidopsis* there exist a SA- JA cross linking pathway in *O. sativa* as well. In SA-JA interacting pathways 13 genes, are found to be expressed that either activate (EDS-1, R-gene, CPR-6, PAD-4, EDS-5, SID-2, TGA-2, SFD-1, PR) or inhibit (CPR-5, NPR-1, SNI-1, SSI-2) the effect of abiotic stress in *A. thaliana*. Of these, the structure and function of five gene sequences from *O. sativa* genome are not known. In silico comparative structural modeling and comparison of physiological parameters of the proteins coded by these five genes in *A. thaliana* and *O. sativa* using different bioinformatics tools, suggest them to be conserved in nature. Results, predict that the hypothetical protein translated from these five unanotated genes in rice are similar in structure to the proteins obtained for similar genes from *A. thaliana* and therefore these protein probably have similar functions as that in *A. thaliana*. It is also likely that these genes are involved in SA- JA cross linking pathway where OsJ_014658 or EDS1, Os04g0689900 or PR4, OsI_04142 or TGA2, OsJ_33269 or PAD4 and OsI_23610 or SFD-1 might activate the SA-JA signaling pathway in rice whereas Os01g0194300 or NPR1 perhaps inhibit the pathway. Further investigation through wet-lab experiments are in progress that would provide a better insight into suppression or activation of these genes involved in SA-JA signaling in rice plants exposed to abiotic stress.

**Acknowledgement:**

We are grateful to the bioinformatics section at MMV from providing computational grid for carrying out this work.

**References:**

**[1]** Ohyanagi H *et al. Nuc Acid Res.* 2006 **34**: 714-744 [PMID:M16381971]
**[2]** Oncel I *et al. Environ Pollut.* 2000 **107**: 315 [PMID: 15092977]
**[3]** Kumar R G *et al. Plant Sci.* 2000 **156**:23 [PMID: 10908802]
**[4]** Waldemar M. *Acta Physiol Plant.* 2007 **29**: 177[DOI 10.1007/s11738-007-0036-3]
**[5]** Wildermuth MC *et al. Nature* 2001 **414**: 562 [PMID: 11734859]
**[6]** Krantev A *et al. J. Plant Physiology* 2008 **165**: 920 [PMID:17913285]
**[7]** Loutfy N et al. J. *Plant Res.* 2011 45: 215 [PMID: 21445718]
**[8]** Chen Z *et al. Science.* 1993 17: 262 [PMID: 8266079]
**[9]** Aydin G *et al. J. Plant Physiol.* 2007 164: 728 [PMID: 16690163]
**[10]** Vick BA & Zimmerman DC, *Plant Physiol.* 1986 **80**: 202 [PMID: 16664582]
**[11]** Cobbett S. Christopher *et al. Curr Opin Plant Biol.* 2000 3: 211 [PMID: 10837262]
**[12]** Waldemar M *et al. J Plant Physiol.* 2005 162: 1338 [PMID: 16425452]
**[13]** Doares S. H *et al. Plant Physiol.* 1995 **108:** 1741-1746 [PMID: 12228577]
**[14]** Arabidopsis Genome Initiative. *Nature.* 2000 **408**: 796 [PMID: 11130711]
**[15]** http://www.ncbi.nlm.nih.gov/
**[16]** http://www.ebi.ac.uk/Tools/pfa/iprscan/
**[17]** http://blast.ncbi.nlm.nih.gov/Blast.cgi
**[18]** Natarajan V *et al. In Silico Biol.* 2008 **8:** 471 [PMID: 19374132]
**[19]** http://www.cgl.ucsf.edu/chimera/
**[20]** http://expasy.org/tools/protparam.html

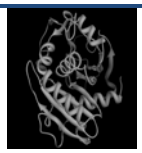
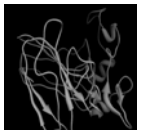
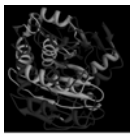
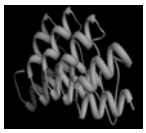
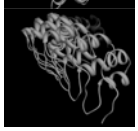
**Edited by P Kangueane**

**Citation: Singh *et al.*** Bioinformation 8(1): 001-005 (2012)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

## Supplementary material:

**Table 1**: The annotated genes involved in SA-JA signaling pathway in *A. thaliana* and corresponding unannotated genes in *O. sativa* genome, the modeled structure of the proteins and RMSD upon superimposition

| Name and TAIR Id of *A. thaliana* gene | Conserved gene specific Domain and Interpro Id | Domain specific Function | Modeled PDB Structure of protein domain of *A. thaliana* | Corresponding *O. sativa* gene and its hypothetical unannotated protein | Modeled PDB Structure of protein domain of *O. sativa* | Protein Sequence Similarity | Superimposed Structures | Structural Similarity (%) and RMSD between protein domain models |
|---|---|---|---|---|---|---|---|---|
| Enhanced Disease Susceptibility 1(EDS1) AT3G48090 | Lipase **IPR002921** | Systemic acquired resistance, salicylic acid mediated signaling pathway | | Os09g0392100 BAF25000.1 |  | 330/763 | | 77% , 1.7 |
| Pathogenesis-related Gene 5 (PR5) AT1G75040 | Thaumatin, Pathogenesis-related domain **IPR001938** | Response to cadmium ion, systemic acquired resistance, a network of signal transduction pathways as other stress-regulated genes |  | Os03g0233200 NP_001049473.1 | | 121/214 | | 98%, 0.6 |
| Phytoalexin deficient 4 (PAD-4) AT3G52430 1 | Lipase, class 3, lipid metabolism process **IPR002921** | Systemic acquired resistance, salicylic acid mediated signaling pathway, PAD4 can interact directly with EDS1, Required for the ssi2-dependent heightened resistance to GPA. | | OsJ_33269 EAZ17725.1 | | 138/360 |  | 80%, 1.7 |
| Suppressor of fatty-acid-desaturase deficiency 1 (SFD-1) AT2G40690 | Glycerol-3-phosphate dehydrogenase NAD-dependent **IPR006109** | Systemic acquired resistance. | | OsJ_23610 EEE66834.1 | | 291/378 | | 73%, 1.3 |
| Non-expressor of PR 1 (NPR1) AT1G64280 | Dehydrogenase BTB/POZ-like (Protein binding domain) **IPR000210** | Key regulator of the salicylic acid (SA) mediated systemic acquired resistance (SAR) pathway, enhance sensitivity to SA. |  | Os01g0194300 NP_001042286 |  | 259/565 | | 98%, 0.9 |

 © 2012 Biomedical Informatics

**Table A:** List of the 13 genes that crosslink the JA and SA signaling pathways as annotated and reported in *A. thaliana* [TAIR database] and corresponding genes in *O. sativa* genome sequence, as obtained by BLASTp search using protein specific domain sequences from *A. thaliana* as query. * indicate genes products yet to be annotated.

| Gene Found To Be Cross linked In *A. thaliana* | Conserved gene specific Domain | Domain specific Function | Similar *O. sativa* Gene obtained by BLASTp |
|---|---|---|---|
| **Enhanced Disease Susceptibility 1(EDS1)** | Lipase | Systemic acquired resistance, salicylic acid mediated signaling pathway | OsJ09g0392100* |
| **Pathogenesis-related Gene 5 (PR5)** | Thaumatin, Pathogenesis-related Domain | Response to cadmium ion, systemic acquired resistance, a network of signal transducing pathways as other stress-regulated genes | Os03g0233200* |
| **Constitutive Expressor of PR Genes 5 (CPR5)** | IQ calmodulin-binding region | Jasmonic acid and ethylene-dependent systemic resistance, jasmonic acid mediated signaling pathway, systemic acquired resistance | Os01g0918500 |
| **TGA-element binding protein (TGA2)** | Basic-leucine zipper ( BZIP) Transcription factor | Interacts with NPR1 and may regulate PR gene expression, systemic acquired resistance, salicylic acid mediated signaling pathway, its phosphorylation is found to be enhanced by salicylic acid treatment. | OsI_04142 * |
| **Constitutive Expressor of PR Genes 6 (CPR6)** | IQ calmodulin-binding region | Participates in signal transduction pathways involved in plant defense (systemic acquired resistance --SAR), salicylic acid mediated signaling pathway | putative NAC2 protein |
| **Resistance gene (R-gene)** | Toll-Interleukin receptor | Involved in defense response signaling pathway | - |
| **Phytoalexin deficient 4 (PAD-4)** | Lipase, class 3 | Systemic acquired resistance, salicylic acid mediated signaling pathway, PAD4 can interact directly with EDS1, Required for the ssi2-dependent heightened resistance to GPA. | OsJ_33269* |
| **Enhanced Disease Susceptibility 5 (EDS5)** | Multi antimicrobial extrusion protein MatE | Negative regulation of defense response, Essential component of salicylic acid-dependent signaling for disease resistance. | OsJ_05176 |
| **Salicylic-acid-induction deficient2 (SID-2)** | Isochorismate synthase | Salicylic acid biosynthetic process, systemic acquired resistance, jasmonic acid biosynthetic process | OsI_31053 |
| **Suppressor of SA insensitivity 2 (SSI-2)** | Fatty acid desaturase, type 2 | Regulation of jasmonic acid mediated signaling pathway, regulation of salicylic acid mediated signaling pathway. | Os01g0919900 |
| **Suppressor of fatty-acid-desaturase deficiency1(SFD-1)** | NAD-dependent glycerol-3-phosphate dehydrogenase | Systemic acquired resistance | OsJ_23610* |
| **Non-expressor of PR 1 (NPR1)** | BTB/POZ-like (Protein binding domain) | A Key regulator of the salicylic acid (SA)-mediated systemic acquired resistance (SAR) pathway, enhance sensitivity to SA. | Os01g0194300* |
| **Transcriptional repressor of SAR genes (SNI-1)** | - | Negative regulation of systemic acquired resistance, Many NPR1-dependent PR gene are specifically de-repressed in the sni1 mutant. | Os02g0312800* |

**Table B:** Verification of different physiological parameters of *A. thaliana* and *O. sativa* using Protparam tool.

| Protein | Organism | Theoretical pI | Estimated half –life (in yeast) | Grand Average of Hydropathicity(GRAVY) |
|---|---|---|---|---|
| **Eds1** | *A. thaliana* | 5.73 | >20 hrs | -0.473 |
| | *O .sativa* | 5.71 | >20 hrs | -0.160 |
| **Sfd1** | *A. thaliana* | 8.75 | >20 hrs | 0.135 |
| | *O .sativa* | 9.83 | >20 hrs | 0.063 |
| **Npr1** | *A. thaliana* | 5.72 | >20 hrs | -0.213 |
| | *O .sativa* | 5.44 | >20 hrs | -0.235 |
| **Pad4** | *A. thaliana* | 6.01 | >20 hrs | -0.192 |
| | *O .sativa* | 9.42 | >20 hrs | -0.173 |
| **Pr5** | *A. thaliana* | 4.75 | >20 hrs | -0.133 |
| | *O .sativa* | 5.86 | >20 hrs | -0.387 |
| **Tga2** | *A. thaliana* | 8.63 | >20 hrs | -0.576 |
| | *O .sativa* | 6.82 | >20 hrs | -0.426 |

Though TGA2 and PAD4 Structure models derived from protein domain sequences from *O. Sativa* had 88% and 80% similarity and RMSD 1.2 and 1.7 respectively with corresponding gene of *A. thaliana* superimposition, however a large difference in theoretical PI was noted for the 2 genes (Table 1). This observation was surprising as the percent similarity score obtained for the 2 genes upon BLASTp were 77.7% and 38.3% respectively.

 © 2012 Biomedical Informatics