GENE IDENTIFICATION IN *PHYTOPHTHORA CAPSICI* THROUGH EXPRESSED SEQUENCE TAGS

N Reena[*], A Chandrasekar[+], A Riju[@], Nima P L[#], Eapen S J[$] and M Anandaraj [£]

Indian Institute of Spices Research

(Indian Council of Agricultural Research)

Calicut, Kerala- 673012

04952730906

[*]reena_narayanan@yahoo.co.in, [+]chandrugcg_77@yahoo.co.in, [@]riju.aikkal@gmail.com, [#]nimanov1@gmail.com, [$]sjeapen@spices.res.in, [£]anandaraj@spices.res.in.

ABSTRACT

Expressed Sequence Tags (ESTs) are a rich source of information for gene discovery. In this paper, we describe the annotation of ESTs of *Phytophthora capsici* whose complete genome is not yet available. *P. capsici* is an Oomycete plant pathogen capable of infecting wide range of plants including cucumber, squash, melons, pumpkin, pepper, tomato and eggplant. In India it causes severe economic losses in black pepper, chillies and cocoa. Towards the understanding of gene function in *P.capsici*, we undertook the functional annotation of ESTs available at NCBI. A total of 56,457 ESTs were downloaded from NCBI and assembled into 5966 contigs. Functional categorization of these ESTs was performed using database similarity search. By functional analysis, we estimated that 84.73% of transcripts encode significant proteins. The most prominent sequences corresponds to members of metabolic pathways, avirulence-associated protein, beta-tubulin, calcium/calmodulin dependent protein kinase 3, catalase, endo-1,4-betaglucanase, cyst germination specific acidic repeat protein precursor, elicitin-like protein, glucanase inhibitor protein, heat shock protein, Kazal-like serine protease inhibitor, mitogen-activated protein kinase, ribosomal protein, serine/threonine kinase, syntaxin and ubiquitin. This EST-gene discovery information can be used to design sequence specific markers for *P.capsici* identification.

Categories and Subject Descriptors

A.0 Conference proceedings

General Terms

Experimentation.

Keywords

Phytophthora capsici, Expressed Sequence Tags, gene functional annotation.

1. INTRODUCTION

The genus *Phytophthora* causes most of devastating plant diseases worldwide [7]. Although Oomycetes exhibit fungal-like morphology, recent molecular and biochemical research have demonstrated that these organisms are actually more closely related to heterokont algae than they are to fungi [4]. *Phytophthora capsici* is known to infect many species of pepper, tomato and other agronomic and ornamental crops of the *Solanaceae* and *Cucurbitacea*e families [17]. This is a serious pathogen of black pepper [2] and cocoa [6]. The pathogen grows within the host and produces sporangia on the surface of diseased tissue, especially leaves. Sporangia are spread by splashing water from irrigation or rain. With moisture present, zoospores released from sporangia swim for a few minutes to more than an hour before encysting. The pathogen survives in the soil in host debris for months [18]. *Phytophthora* has been widely acknowledged as a taxonomically 'difficult' genus [5]. Traditional methods to detect or isolate these pathogens involve plating infected plant parts or soil on selective medium and conducting a pathogenicity assay. However, this is limited by its lack of sensitivity, as *P. capsici* shares similar morphology with certain other *Phytophthora* spp. Based on the morphology and pathogenicity a new species *P. tropicalis* has been described [3]. Recent studies based on r-DNA analysis and predicted ITS rDNA secondary structure profile showed that Indian isolate of *P. capsici* shares the characters of both *P. capsici* and *P. tropicalis* [13].

In this study, we have used EST sequence data for gene discovery. This EST-gene discovery information can be used to design sequence specific markers. Finally the ease with which EST-gene discovery can be mapped may facilitate the identification of primers of particular interest, to identify *P. capsici.*

ESTs are sequenced portions of messenger RNA. In recent years, EST projects have been initiated for numerous plant and animal species, and have generated a vast amount of sequence

information that can be used for gene discovery, functional genetic studies, and marker development [11]. EST databases represent a valuable resource for the identification of genes in organisms with uncharacterized genomes and for development of molecular markers. These are useful in genetic and evolutionary studies because they are located in transcribed genes and a putative function can often be inferred from homology searches [10]. Large-scale cDNA sequencing and EST analyses have been a rapid method to identify novel cDNAs which afford to identify genes of various physiological functions [16].

2. MATERIALS AND METHODS

2.1 EST processing

EST sequences of *P.capsici* were downloaded from dbEST of NCBI (Accession No: FG042267 - DY985170). EST and mRNA sequences often have poly-A tails at their ends. Poly-A tails of EST sequences were trimmed using the online tool Trimest, using default parameters. Trimest reads one or more nucleotide sequences and writes them out again but with any 3' poly-A/T tail removed.

2.2 EST assembly

EST assembling is done to cluster the ESTs into putative genes and to annotate these genes with sequence similarity searches. A foremost stage in short gun sequencing strategy approach is to assemble short reads into long sequences. Contig Assembly Program 3 (CAP3) [9] with the default parameters, is used to cluster the overlapping ESTs into contigs. This assembly algorithm consists of three major phases. In the first phase, 5' and 3' poor regions of each read are identified and removed. Overlaps between reads are computed. False overlaps are identified and removed. In the second phase, reads are joined to form contigs in decreasing order of overlap scores. Then, forward–reverse constraints are used to make corrections to contigs. In the third phase, a multiple sequence alignment of reads is constructed and a consensus sequence along with a quality value for each base is computed for each contig.

2.3 Annotation of *P. capsici* ESTs

Each of 5966 contigs were translated in all six reading frames and compared with the *Phytophthora* database at the National Centre for Biotechnology (NCBI) using the BLASTX program, with an E value cut-off $\leq 10^{-5}$ [1] which compares translated nucleotide sequences with protein database. Putative functions to ESTs were assigned based on the BLASTX matches. Results of BLASTX matches were categorized based on the molecular functions and biological process.

3. RESULTS AND DISCUSSION

A total of 56,457 initial sequences were downloaded from the NCBI database and trimmed to remove the poly A/T tails. CAP3, the contig assembly program helped to remove the redundant EST sequences and to correct the sequencing error. A total 56,457 ESTs were assembled into 5966 contigs and 3865 singletons making a total of 9831 unigenes.

Table1: Expressed Sequence Tags (EST) Summary

Total No of ESTs	56,457
No. of Contigs	5966
No of Singletons	3865
Total No of Unigenes	9831

3.1 Gene Annotation

To assign putative functions to the clustered ESTs, database similarity search were performed comparing every contigs sequence to *Phytophthora* database. Of the total 5966 contigs, 5054 contigs have homology to proteins with significant functions. Based on the biological process the ESTs are classified into different functional groups. Using the BLASTX program, 3.57% ESTs were assigned to hypothetical proteins of unknown function, and 11.7% were designed as "no hit." On the other hand, 84.73 % of the ESTs displayed significant similarity to known sequences in GenBank. These orphan sequences may well represent sequences unique to the pathogen and may be useful as targets for chemical control. Thus, many as yet uncharacterized genes may possess functions in virulence [15]. The gene information is categorized according to functional behavior like ADP/ATP translocase, acetyl-CoA carboxylase, avirulence-associated protein, beta-tubulin, calcium/calmodulin dependent protein kinase 3, catalase, cell 5A endo-1,4-betaglucanase, cyst germination specific acidic repeat protein precursor, elicitin-like protein, glucanase inhibitor protein, heat shock protein 90, Kazal-like serine protease inhibitor, mitogen-activated protein kinase, ribosomal protein, serine/threonine kinase, Syntaxin and ubiquitin. Based on the functional categories the ESTs are classified as below. (Figure 1)

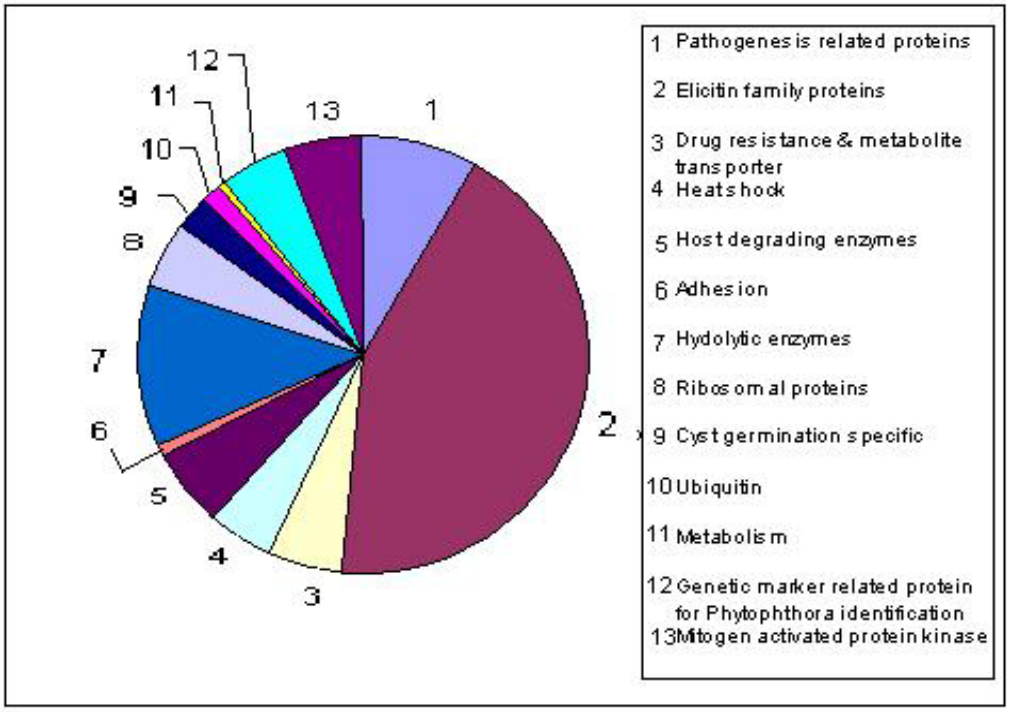

Figure 1 Classification of ESTs according to function prediction by BLASTX match.

Table 2 Functional distribution of ESTs showing best BLASTX match

(1) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with pathogenesis

Contig No	Best Blastx match	Organism	E value	Score
Contig 4801	calcium/calmodulin dependent protein kinase	*P. sojae*	0	684
Contig5635	Cutinase	*P. infestans*	2E-58	219
Contig 752	Avh	*P. sojae*	7E-24	103

(2) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Elicitin family

Contig No	Best Blastx match	Organism	E value	Score
Contig 1985	elicitin-like INF5	*P .infestans*	3E-53	201

(3) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Drug resistance and metabolite transport protein

Contig No	Best Blastx match	Organism	E value	Score
Contig 5288	Pleiotropic drug resistance	*P. sojae*	0	775
Contig 935	ABC Transporter	*P .infestans*	4e-61	227

(4) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Host degrading enzymes inhibitors

Contig No	Best Blastx match	Organism	E value	Score
Contig 5794	Kazal-like serine protease inhibitor	*P .infestans*	1E-70	260

(5) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Genetic marker related proteins for *Phytophthora* identification

Contig No	Best Blastx match	Organism	E value	Score
Contig 334	Beta- tubulin	*P .capsici*	0	845
Contig 200	Enolase	*P .infestans*	3E-131	460
Contig 2132	Translation elongation factor	*P. parasitica*	0	863

(6) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Stress

Contig No	Best Blastx match	Organism	E value	Score
Contig 9	heat shock protein	*P. nicotianae*	0	677

(7) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Adhesion

Contig No	Best Blastx match	Organism	E value	Score
Contig 4569	CBEL protein	*P. nicotianae*	8e-65	241

(8) *Phytophthora capsici* ESTs corresponding to cDNAs potentially associated with Hydolytic enzymes

Contig No	Best Blastx match	Organism	E value	Score
Contig 1517	beta-glucosidase	*P .infestans*	0	908
Contig 2783	endo-1,4-betaglucanase	*P. ramorum*	5e-149	489
Contig 739	exo-1,3-beta-glucanase	*P .infestans*	0	1211
Contig 5885	cathepsin-like cysteine protease	*P .infestans*	0	1199

(9) *Phytophthora capsici* ESTs corresponding to Ribosomal cDNAs

Contig No	Best Blastx match	Organism	E value	Score
Contig 3467	Ribosomal protein L2	*P. sojae*	5e-136	476

(10) *Phytophthora capsici* ESTs corresponding to Ubiquitin

Contig No	Best Blastx match	Organism	E value	Score
Contig 258	ubiquitin	*P .infestans*	1E-127	449

(11) *Phytophthora capsici* ESTs corresponding to Mitogen Activated Protein Kinase

Contig No	Best Blastx match	Organism	E value	Score
Contig4432	Mitogen Activated Protein Kinase	*P. parasitica*	*724*	0

ESTs assigned to a member of the Elicitin gene family are the most represented in the functional classification. This corroborates the findings of EST analyses for *P.parasitica* [8]. A variety of genes were identified that may be associated with pathogenesis. They include molecules that counteract plant defense responses, such as *P. sojae* member of a family encoding glucanase inhibitors that interacts with the soybean endo-beta-1,3-glucanases during infection [13], and an EST similar to Kazal-like proteinase inhibitor from *P. infestans* that inhibits host proteases [14]. Other candidates may be grouped in several categories (Table 2). They comprise a large array of hydrolytic enzymes, genes involved in host defense response, sugar transporters which may function in the uptake of host degradation products, ribosomal proteins. Other products may be secreted, such as molecules involved in the adhesion to host, surface binding proteins, Genetic marker related protein for *P.capsici* identification and Ubiquitin. All these genes are candidate on the sole basis of similarity reports and analogies to other *Phytophthora* sequences. Hence, they may constitute quantitative factors that contribute to the overall virulence of *P. capsici*. Thus main cellular functions were identified on the basis of EST annotation and functional analogies. The accumulation of sequences, as well as functional analyses will be necessary for a deeper investigation of the various metabolic pathways and cellular processes [8].

4. CONCLUSION

EST analysis provides a powerful and rapid means of restructuring the transcriptome and remains a useful means of gene discovery. In this work 5966 contigs from *P.capsici* were analysed to study the function of EST. The analysis of ESTs in this study identified a range of genes likely to be involved in

pathogenesis, drug resistance, stress, host degradation and genetic marker related protein for *Phytophthora* identification. These functional annotation results can help to design gene specific primers for *P. capsici* identification and understand host pathogen interaction. We found a substantial population of *P.caps*ici ESTs share similarity with *P.infestans* and *P.sojae* sequences. These identification reflects that *P.capsici* shares orthology with *P.infestans* and *P.sojae* genomes.

5. ACKNOWLEDGEMENTS

This work was funded by a grant from ICAR, Government of India. The support received from the Bioinformatics Centre, Indian Institute of Spices Research, Calicut-12 is gratefully acknowledged.

6. REFERENCES

[1] Altschul, S.F., Warren, G., Webb, M., Eugene, W. M., and David, J. L. Basic local alignment search tool. J. Mol. Biol. 215 (1990), 403-10.

[2] Anandaraj, M. Diseases of black pepper, In: Black pepper (*Piper nigrum* L). Ravindran, P.N. (Eds) Hardwood Academic Publishers, Netherlands. (2000), 239-268.

[3] Aragaki, M., and Uchida, J.Y. Morphological distinctions between *Phytophthora capsici* and *P. tropicalis* sp. Mycologia, 93 (2001), 137–45.

[4] Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolitle, W.F. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290, (2000), 972–977.

[5] Brasier, C.M. Problems and prospects in *Phytophthora* research. In: Erwin DC, Bartnicki-Garcia S, Tsao PH, eds. Phytophthora: its Biology, Taxonomy, Ecology and Pathology. St Paul, MN, USA: APS Press, (1983), 351–64.

[6] ChandraMohanan, R., and Chowdappa, P. Important diseases of cocoa and their management. In: IPM system in Agriculture Vol.6– Cash Crops. Eds. R.K. Upadhyay, K.G. Mukerji and O.P. Dubey. Aditya Books Pvt. Ltd., New Delhi, India (1999), 401- 421.

[7] Erwin, D.C., and Ribeiro, O.K. *Phytophthora* diseases worldwide. In: M.N. St. Paul (Ed.), American Phytopathological Society, 1996.

[8] Franck, P., Joelle, A., Eric, G., Jo-Yanne. Gene identification in the oomycete pathogen *Phytophthora parasitica* during in vitro vegetative growth through expressed sequence tags. Fungal Genetics & Biology. 42 (2005), 611–623

[9] Huang, X., and Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.*, 9, (1999), 868-877.

[10] Maria, D..M, Christine, O.D., Gabriel, A.B., and Alex, W. A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. BMC Genomics (2009), 10: 243.

[11] Pashley, C.H., Ellis, J.R., McCauley, D.E., and Burke, J. M. EST databases as a source for molecular markers: lessons from *Helianthus*. Journal of Heredity, 97 (2006), 381–388.

[12] Rose, J.K., Ham, K.S., Darvill, A.G., Albersheim, P. Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens. Plant Cell 14(2002), 1329–1345.

[13] Sheji, C., Renu, S. G., Balaji, S., and Anandaraj, M. Ribosomal DNA analysis of three Phytophthora species occurring in India. Indian Phytopathology 62(2) (2009), 155-162.

[14] Tian, M., Huitema, E., Da Cunha, L., Torto-Alalibo, T., Kamoun, S. A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. J. Biol. Chem. 279 (2004), 26370–26377.

[15] Trudy, A., Sucheta, T., Brian, M. S., Felipe, D.A., Lecong, Z., Hua, L., Marcus C. C., Dinah, Q., Mark, G., Chunhong, M., Bruno, W. S., Sobral., Mark E. W., Thomas, K. M., Ralph, A.D., Brett, M. T. Expressed Sequence Tags from *Phytophthora sojae* Reveal Genes Specific to Development and Infection. MPMI. 20 (2007), 781–793.

[16] Uno, Y., Suzuki, Y., Wakaguri, H., Sakomoto, Y., Sano, H., Osada, N., Hashimoto,K., Sugano, S., Inoue,I. Expressed sequence tags from cynomolgus monkey (*Macaca fascicularis*) liver. A systematic identification of drug-metabolizing enzymes. FEBS Letters 582 (2008), 531-358.

[17] G. Zhang, Y. Q. Li, H. Fan, Y. C. Wang, X. B. Zheng. (2006) Molecular detection of Phytophthora capsici in infected plant tissues, soil and water. Plant Pathology 55 (2007), 770–775.

[18] Zheng, X.B. Phytophthora and Methods in Phytophthora. Beijing, China: Agricultural Press. 1997.