CHARACTERIZATION OF SUBTRACTED cDNA LIBRARY FOR DETECTING EXPRESSED SEQUENCE TAGS (ESTs) SPECIFIC TO DROUGHT TOLERANCE IN BLACK PEPPER (*Piper nigrum* L.)

By

MALI GAJANAN VISHVAMBHAR

(2008-11-122)

THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agriculture

(PLANT BIOTECHNOLOGY)

Faculty of Agriculture

Kerala Agricultural University, Thrissur

Centre for Plant Biotechnology and Molecular Biology

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2010

DECLARATION

I hereby declare that this thesis entitled **"Characterization of subtracted cDNA library for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)"** is a bonafide record of research work done by me during the course of research and that it has not been previously formed the basis for the award to me of any degree, diploma, fellowship or other similar title, of any other University or Society.

Vellanikkara

Mali Gajanan Vishvambhar
(2008-11-122)

CERTIFICATE

Certified that this thesis entitled **"Characterization of subtracted cDNA library for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)"** is a bonafide record of research work done independently by **Mr. Mali Gajanan Vishvambhar** under my guidance and supervision and that it has not formed the basis for the award of any degree, diploma, fellowship or associateship to him.

Vellanikkara

Dr. P.A. Nazeem
(Chair Person, Advisory Committee)
Professor and Head
Centre for Plant Biotechnology and Molecular Biology
College of Horticulture
Vellanikkara, Thrissur

CERTIFICATE

We, the undersigned members of the advisory committee of **Mr. Mali Gajanan Vishvambhar** (2008-11-122), a candidate for the degree of **Master of Science in Agriculture**, with major field in **Plant Biotechnology**, agree that the thesis entitled **"Characterization of *subtracted* cDNA *library* for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)"** may be submitted by **Mr. Mali Gajanan Vishvambhar** in partial fulfillment of the requirement for the degree.

Dr. P.A. Nazeem
(Major advisor)
Professor and Head
Centre for Plant Biotechnology and Molecular Biology
College of Horticulture
Vellanikkara

Dr. R. Kesavachandran
(Member, Advisory committee)
Professor
Centre for Plant Biotechnology and
Molecular Biology (CPBMB)
College of Horticulture
Vellanikkara.

Dr. D. Girija
(Member, Advisory committee)
Professor
Centre for Plant Biotechnology
and Molecular Biology
(CPBMB) College of
Horticulture, Vellanikkara

Dr. K. Nandini
(Member, Advisory committee)
Professor and Head,
Dept. of Plant Physiology
College of Horticulture, Vellanikkara

External Examiner

Dr. PH Ramanjini Gowda
Professor
Dept of Plant Biotechnology
UAS, GKVK Bangalore
-560 065

Acknowledgement

Words cannot express my deep sense of gratitude and indebtdness to ***Dr. P.A. Nazeem****, Professor and Head, CPBMB and chairperson of my Advisory Committee. I wish to place my heartfelt thanks to her for her inspiring guidance, untiring help, patience, encouragement, constructive criticism and valuable suggestions during the period of the investigation and preparation of the thesis. My sincere and heartfelt gratitude ever remains with her.*

I am ever grateful to ***Dr. D. Girija****, Professor, CPBMB and member of my advisory committee for her invaluable help, guidance and critical assessment throughout the period of the work. I thank her for all the help and cooperation she has extended to me.*

I express my sincere gratitude to ***Dr. R. Keshavachandran,*** *Professor, CPBMB and member of my advisory committee for his valuable suggestions and guidance rendered me for the completion of the research programme and preparation of the thesis.*

I am deeply obliged to ***Dr. K. Nandini****, Professor and Head, Department of Plant Physiology and member of my advisory committee for her unfailing support and enthusiasm, relevant suggestions and whole hearted cooperation throughout the period of investigation.*

I also avail this opportunity to pay my sincere obligations and heartfelt thanks to ***Dr. Sujatha*** *and* ***Dr.Valsala*** *of CPBMB for their encouragement and kind help offered at different stages of the study.*

I owe special thanks to ***Dr. P. Sureshkumar*** *who has helped me in learning and handling Radioactive elements during Colony hybridization and also for suggestions given by them time to time.*

I wish to express my sincere thanks to all the non-teaching staff members and labourers of CPBMB for their whole hearted cooperation and timely assistance.

I express my deep sense of gratitude to ***Divya Matrin, Sherin, Praveena, Anju, Deepa, Firoz, Sunil, Simi chechi, Shailaja chechi, Sara chechi and others*** *of CPBMB, who helped me in several ways for the completion of this venture.*

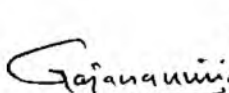
I am grateful to ***Reena, Alishya, Reemya, Shidi and, RA's*** *of DIC for the bioinformatics lessons, which helped me a lot to carry out the in-silico analysis of the sequences.*

I have no words to express my deep sense of gratitude to all my senior friends ***Dinesh Labade, Renu, Kiran, Shivaji, Suja, Rahul, Dhinesh, Vijith, Deviprasad, Madhusudan, Hanumant and Manikantan*** *for their valuable assistance and guidance extended throughout the investigation.*

Words fail short as I place on record my indebtedness to my class mates and friends, ***Sunil, Prathemesh, Arya, Sampath, Tamarai, Navneet, Suda, Suma, Laxmi, Shabana, Neethu, Shijini, Ambili, Randeep and Mrudula*** *for their prompt help and cooperation for the entire period of study.*

I am in dearth of words to thank my friends, ***Vishal, Ram, Vivek, Suhas, Nitin, Sonthosh, Deepak, Balu, Raj, Balaji, Vijay, Manjiri, Vaishali, Rohini*** *and my juniors* ***Rakesh, Ashok, Pradeep, Manjunath, Achintya, Sarwesher, Lubina, Sahana, Somya and others*** *for their encouragement and unliving support.*

With gratitude and affection, I recall the boundless affection, constant encouragement, warm blessings and motivation from my brothers ***Datta, Dayanand*** *and* ***my loving sister Jyoti*** *without which this endeavour would never have become a reality.*



Gajanan

Dedicated to
My
Parents
and Brothers

ABBREVIATIONS

ABA	Abscisic Acid
akthr	Aspartokinase-homoserine dehydrogenase
Amp.	Ampicillin
β	Beta
bp	Base pair
BLAST	Basic Local Alignment Search Tool
CPBMB	Centre for Plant Biotechnology and Molecular Biology
cDNA	Complementary DNA
CDD	Conserved Domain Database
°C	Degree Celsius
DREB	Dehydration Responsive Element Binding
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribo Nucleoside Triphosphate
DIC	Distributed Information Centre
ds	Double stranded
EDTA	Ethylene diamine tetra acetic acid
ESTs	Expressed Sequence Tags
epcrf	Eukaryotic peptide chain release factor
GS	Glutamine synthetase
HSP	Heat Shock Protein
pH	Hydrogen ion concentration
IP	Imaging plate
Kb	Kilo base
LB	Luria Bertani
L	Litre
LEA	Late Embryogenesis Abundant
mRNA	messenger RNA
µg	Micro gram
µl	Micro Litre

ml	**Milli litre**
mM	**Milli Molar**
M	**Molar**
NCBI	**National Centre for Biotechnology information**
ng	**Nano gram**
nm	**Nano meter**
OD	**Optical density**
ORF	**Open Reading Frame**
%	**Percentage**
PCR	**Polymerase Chain Reaction**
PPIas	**Peptidyl proyl isomerases**
PMT	**Photo-Multiplier Tube**
PNK	**Pepper nigrum var. kalluvally**
QTL	**Quantitative Trait Loci**
RT- PCR	**Reverse Transcriptase Polymerase Chain Reaction**
rpm	**revolutions per minute**
RNA	**Ribonucleic acid**
RNase	**Ribonuclease**
SSH	**Suppression subtractive Hybridization**
tRNA	**transfer RNA**
UV	**Ultra violet**
UTR	**Untranslated Region**
v/v	**volume/volume**
X-gal	**5bromo-4-chloro-3-indoyl-β-D-galactosidase**

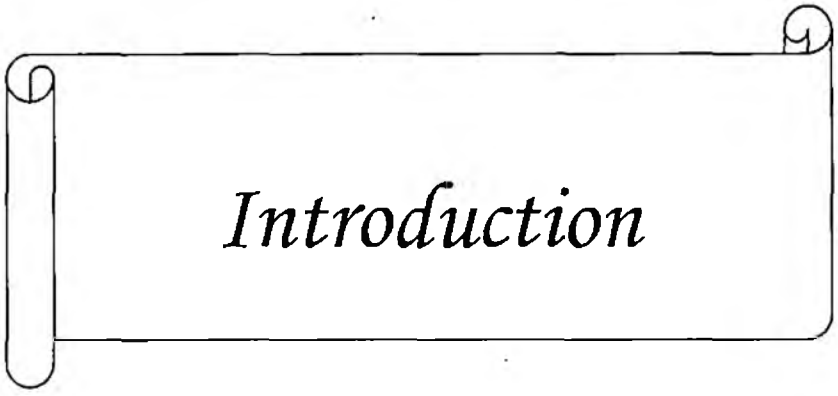
TABLE OF CONTENTS

List of tables

List of tables

List of figures

List of plates



Introduction

1. INTRODUCTION

Black pepper is the most important spice of the world and is commonly known as the 'black gold'. Therefore it is rightly called 'King of Spices', as the volume of international trade is of the highest among all the known spices. Ninety percent of the area under black pepper is rainfed. These areas frequently experience severe water deficit due to uncertain and uneven rainfall distribution pattern and the occurrence of drought in such areas may devastate the plantations of black pepper.

Drought is the most important limiting factor for crop production and it is becoming an increasingly severe problem in many regions of the world. The percentage of drought affected land areas have doubled from the 1970s to the early 2000s in the world. It is a world-wide phenomena seriously influencing crop production and quality; with increasing population and global climate change making the situation more serious. This problem will probably worsen in the next decades, according to climate change models which predict longer, more frequent and more intense drought periods. In this situation, breeding for stress tolerance in crop plants have become urgent for the future of agriculture and food production.

Over the last few years, many research groups have isolated and characterised different genes, involved in mechanisms of plant response to stress; to be used as biotechnological tools to reach the goal. Despite the fact that many of these genes actually confer variable levels of tolerance to different types of abiotic stress when expressed in transgenic plants, their practical usefulness has been questioned. In fact, no crop cultivar with sufficient tolerance levels, from an agronomic point of view, has been obtained by molecular breeding (Blum, 1996).

Identification of genes involved in environmental stress response provide new target for genetic engineering of crops for better tolerance.

If the current trends do not change in the near future, it is estimated that up to 50 percent of the land cultivated at present will be completely lost for agriculture by 2050 (Wang *et al.*, 2003).

Genotypic variations for drought tolerance has been reported in black pepper and the variety Kalluvally is one among the six genotypes identified as drought tolerant (Thankamani *et al.*, 2003) based on physiological and biochemical parameters. Unraveling the mechanisms involved in drought tolerance in the resistant genotypes, could help the molecular breeding of economic crops in a long way.

Tolerance to drought involves a complex mechanism working in combination to avoid or tolerate water deficits. Many genes are reported to be involved in drought management and they function directly in protecting the cells and also in the regulation of gene expression and signal transduction (Shinozaki and Yamaguchi, 1997). The molecular basis of plant tolerance to water stress remains unclear because of the reason that several regulatory mechanisms are involved in the stress signal pathways thus making it a multigenic character. Therefore, for producing drought tolerant plants it is important to understand this complex network, for which it becomes necessary to identify and characterize the genes that respond to water stress.

One of the powerful techniques of molecular biology is to identify these differentially expressed genes by Suppression Subtractive Hybridization (SSH). It is a method of detecting and isolating gene sequences that are differentially expressed. Way *et al.* (2005) have successfully used this technique to identify genes expressed in wheat undergoing gradual water deficit stress.

The Expressed Sequence Tags (EST), generated via partial sequencing of cDNA clones contain a vast amount of information about the genes that are expressed in higher plants at different developmental stages and under a variety of environmental conditions.

At the Centre for Plant Biotechnology and Molecular Biology (CPBMB), College of Horticulture, Vellanikkara a cDNA library has been constructed in black pepper through SSH in relation to drought. The present study was undertaken to characterize the library for ESTs and unravel the genes involved in drought tolerance in black pepper.

Review of Literature

2. REVIEW OF LITERATURE

Black pepper (*Piper nigrum* L.), renowned as 'King of spices' is one of the important spice crops of the world and is an export oriented crop in India with a substantial share in foreign exchange earnings.

Black pepper occupies an area of 2,01,000 ha in India with production of 45,000 tonnes during 2008-2009 (Spices Board, 2010). However, in productivity, India occupies the last position among the leading black pepper producing countries in the world and it is only 310 kg/ha.

More than 80 percent of the cultivable land in the world is under rainfed condition. In such areas, drought is the major constraint for the crop productivity and the picture is same in plantation crops since they are widely grown in different soil types such as sandy, sandy loam, laterite and forest soils in the state of Kerala, Karnataka, Tamilnadu and Andhra Pradesh (Rajagopal and Nareshkumar, 2003).

Even though Kerala receives good rains, rainfall is not distributed uniformly and the crop suffers due to moisture stress from December to May (Ravindran *et al.*, 2003). Increasing the crop area under irrigation has several limitations, water resources being the major one and water requirement of black pepper being fairly high. Since, the availability of water has become a limiting factor; further increase in crop area under irrigation is very difficult. Then, the approach to alleviate such problems has to be the use of available water source with high productivity efficiency.

Thus, it becomes necessary to identify the varieties, which are tolerant to water stress conditions in the field and to evolve management strategies for conserving available water resources to overcome adverse effects of drought. In the breeding strategies, exploitation of the identified varieties which are not only drought tolerant, but also have the potential for high yield under limited water

conditions, would be an important step for overall improvement of crop productivity in drought prone areas.

The variety *Kalluvally* was selected for the present study as this variety was reported to be one of the hardy genotypes that can thrive over water stress in a better way (Thankamani *et al.*, 2003).

Drought continues to be a challenge to agricultural scientists in general and to plant breeders in particular, despite many decades of research. The development through breeding of cultivars with higher harvestable yield under drought would be a major breakthrough in crop production. Severity, timing and duration of drought will vary from year to year and cultivars successful in one dry year may fail in another. To make matters worse, drought seldom occurs in isolation and it often interacts with other abiotic and biotic stress. Areas with a high risk of drought generally have low-input agriculture (Ceccarelli and Grando, 1996).

2.1 Crop losses due to Drought

Water deficit is one of the prevalent causes of crop loss and increased plant tolerance to water deficit is considered among the most important abiotic parameters that can contribute to increased grain production (Araus *et al.*, 2003).

The loss due to drought in the tropics alone is thought to exceed 20 million tons of grain per year or approximately 17 percent of well-watered production with losses reaching up to 60 percent in severely affected regions (Ribaut *et al.*, 2002). To improve plant growth and performance under water-limited conditions, drought tolerant crops must be developed.

Water stress is one of the major environmental conditions that adversely affect plant growth and crop yield (Bartels and Nelson, 1994). To cope with environmental stress, plants execute a number of physiological and metabolic responses (Bray, 1993; Bohnert *et al.*, 1995; Passioura, 2007.).

During the drought in 2004 in India, 62 million hectares (30 % of total crop area) was affected and food grains production decreased to 24 million tonnes from normal 12 percent. A survey was conducted during the 6th Asia regional maize workshop in India in November 1995 to determine the effects of drought and low nitrogen on Asia maize yields. The results revealed that half of the total maize area in Asia has been affected by both stresses and yield loss due to drought was estimated to be 10 to 75 percent (Logroiio and Lothrop, 2006). The estimate of maize production losses due to drought in Indonesia was about 15 percent or about 1.2 million tonnes.

An annual estimated loss in peanut production equivalent to US $ 208 million could be recovered through genetic enhancement for drought resistance with a benefit:cost ratio of 5:2 (Johansen and Nigam, 1994).

2.2 Response of plants to drought conditions

Interest in understanding the molecular basis of plant responses to water stress is driven by a desire to understand the mechanism that plant evoke to tolerate environmental stresses and the prospect that such knowledge might provide new strategies to improve the stress tolerance of agriculturally important plants (Xiong *et al.*, 2002).

The cellular and molecular responses of plants to environmental stress have been investigated mainly using *Arabidopsis thaliana* as a model plant (Ingram and Bartels, 1996; Shinozaki and Yamaguchi-Shinozaki, 1997; Zhu, 1997, 2002). Various genes are involved in water stress response in plants. These genes encode two groups of proteins. The first group include proteins that probably function in stress tolerance, such as the enzymes required for osmolyte biosynthesis, chaperones, LEA proteins, mRNA binding proteins, water channel proteins, sugar and proline transporters, detoxification enzymes and various proteases (Ingram and Bartels, 1996; Seki *et al.*, 2002).

The second group include protein factors involved in the regulation of signal transduction and gene expression that probably function in stress response-such as protein kinases, transcription factors and enzymes in phospholipid metabolism (Shinozaki and Yamaguchi-Shinozaki, 1997). The existence of a variety of water-inducible genes suggests that the responses of plants to water stress are rather complex (Zhu, 2002 and Xiong *et al.*, 2002).

Drought decreases cuticular transpiration in seedlings (Bengstson *et al.*, 1978) and increase plasma membrane permeability (Quartacci and Navarilzzo, 1992). It reduces leaf area and protoplast volume (Saradadevi *et al.*, 1996) and decreases photosynthesis reorganizing the photosystem II (PSII) through increased phosphorylation of the PSII core coupled with increased D1 protein synthesis (Giardi *et al.*, 1996). Some plants turn from C3 to C4 photosynthetic pathway so as to increase water use efficiency.

2.2.1 Drought response in Black pepper

Thomas *et al.* (1990) studied the proline accumulation potential of ten selected cultivars of black pepper under PEG induced stress. Accumulation of proline was higher and earlier in cvs. Kottandan and Neelamundi than control. Maximum decline in the accumulated proline was observed during recovery from water stress. Vasantha *et al.* (1991) reported that absolute proline content during water stress in black pepper ranged from 1.96 to 3.48 µmoles.

During moisture stress, expansion of leaves in black pepper decreased before the soil moisture reached a critical level (Ramadasan and Vasantha, 1994). Root to shoot ratio was higher in cv. Kalluvally as compared to other cultivars. The content of epicuticular wax ranged from 1.46 to 2.08 mg/cm^3. The activity of nitrate alone was found to decline in all cultivars under water stress. Various physiological and biochemical parameters such as membrane leakage, relative water content (Rahman *et al.*, 1999) enzyme activities *viz.* catalase, superoxide dismutase, peroxidase and polyphenol oxidase and lipid peroxidation

(Chempakam *et al.*, 1993) have been worked out to assess the effect of water stress on crop and its relative tolerance.

Krishnamurthy *et al.* (2000) reported that in different black pepper genotypes, relative water content decreased and membrane leakage increased due to water stress. The activity of catalase and superoxide dismutase decreased while that of peroxidase and polyphenol oxidase increased with the intensity of stress in all the genotypes.

Thankamani *et al.* (2003) found that the drought resistant cultivars showed a higher percentage increase in enzyme activities of peroxidase and superoxide dismutase on exposure to water stress than the drought susceptible cultivars. Chlorophyll and leaf epicuticular wax content of four black pepper varieties at mild, moderate, severe and zero water stress was studied by Thankamani and Ashokan (2002). They observed that under moderate water stress, chlorophyll-a content decreased in all the varieties except Poonjarmunda and Uthirankotta -2. Severe stress decreased chlorophyll a, b and the total content in all the varieties. Maximum wax content was observed under severe water stress, which was relatively high in varieties Kalluvally–4, Padarpan, Poonjarmunda, Kumbakodi, Panniyur-5 and Uthirankotta.

2.2.2 Response to drought in other plants

Liu and Li (2005) studied the effects of soil water regime in two wheat cultivars, differing in drought tolerance with respect to root respiration, photosynthesis and grain yield. They found that the drought tolerant cultivar (Dingxi 24), had a higher biomass and metabolic activity under severe drought stress compared to the sensitive cultivar (Longchun 139-2), which resulted in further limitation of grain yield. Results showed that root respiration, carbohydrates allocation and grain yield were closely related to soil water status in wheat cultivars.

Pincheiro *et al.* (2004) studied drought tolerance in four clones of *Coffea canephora* (Robusta coffee) in relation to protection against oxidative stress when subjected to long-term drought. They found that drought tolerant clones were better able to maintain their leaf water status than drought sensitive clones after withholding irrigation. Regardless of the clones irrigated, the net carbon assimilation rate decreased under drought stress. There was an increase in superoxide dismutase, ascorbate peroxidase, catalase, and guaiacol peroxidase and also in glutathione reductase and dehydroascorbate reductase.

Yin *et al.* (2004) found that exogenous ABA application in two Poplar species significantly affected morphological and physiological properties such as decreasing dry matter accumulation, specific leaf area and gas exchange and increasing at least relatively dry matter allocation into the root fraction, endogenous ABA content and δ13 C, under both well watered and water stressed conditions.

Susihoto and Berninger (2007) studied the interactions between morphological and physiological drought responses in *Eucalyptus microtheca* and found that the growth of the seedlings decreased and allocation patterns changed so that allocation to the roots increased. However, changes in photosynthesis and stomatal conductance under drought were modest.

Zayed and Zeid (1998) described that drought stressed mung bean (*Vigna radiata*) seedlings have increased concentration of hydrolytic enzymes such as α-amylase that resulted in an increased content of soluble polysaccharides. The accumulation of such organic solutes resulted improvement in cytoplasmic osmoregulation and thus increased plant tolerance to drought.

Cui *et al.* (2004) studied the effect of drought on physiological aspects of Crassulacean Acid Metabolism (CAM) in Doritaenopsis, an orchid. They found that under light and dark photosynthesis, stomatal conductance and transpiration decreased with increasing the duration of drought. Also, there was an increase in reduced sugar content while total sugar and starch content decreased. On

imposition of drought there was an increase in phosphoenol pyruvate carboxylase and peroxidase activity while, superoxide dismutase activity decreased.

2.3 Molecular basis of drought tolerance

Abiotic stresses including drought can significantly reduce crop yields and restrict the latitudes and soils on which commercially important species can be cultivated (Blum, 1985). Identifying and understanding mechanisms of drought tolerance is crucial to the development of tolerant commercial cultivars. Thus, the responses of plants to various stresses have for decades been the focus of physiological studies (Levitt, 1980) and of molecular genetics studies (Grover *et al.* 1999; Forster *et al.* 2000). A large and increasing number of genes, transcripts and proteins have been correlatively implicated in stress response pathways, while their precise functions in either tolerance or sensitivity remain unclear (Bray, 1993). Extensive efforts have also been devoted to the characterization of genes induced or upregulated by drought (Close *et al.* 1989, 1993).

Water deficit occurs when water potentials in the rhizosphere are sufficiently negative to reduce water availability to sub-optimal levels for plant growth and development. When plants are subjected to drought stress, a number of physiological responses have been observed (Ludlow and Muchow 1990). In some cultivated cereals, osmotic adjustment has been found to be one of the physiological mechanisms associated with plant resistance to water deficit (Morgan1984; Blum 1988; Zhu *et al.* 1997).

The complex plant response to abiotic stress involves many genes and biochemical- molecular mechanisms. A complex of signal transduction events that have not been clearly delineated induces changes in gene expression. Various genes respond to drought stress in several species and functions of their gene products have been predicted from sequence homology with known proteins. Such genes that are induced during drought stress conditions provides tolerance to plant by functioning directly in protecting cells from water deficit by the

production of important metabolic proteins and indirectly by regulating other genes for signal transduction (Shinozaki *et al.*, 2003).

2.4 Drought associated genes and proteins

2.4.1 Heat shock proteins and chaperones

Heat shock proteins (Hsps) and molecular chaperons, as well as late embryogenesis abundant (LEA) protein families, are reported to be involved in plant drought stress tolerance (Wang *et al.*, 2003; 2004). High temperature and drought stress can cause denaturation and diysfunction of many proteins.

Hsps and LEA proteins help to protect against stresses by controlling the proper folding and conformation of both structural (i.e. cell membrane) and functional (i.e. enzymes) proteins (Almoguera and Jordano, 1992; Sachs and David, 1986; Vierling, 1991). It has been shown that two of hsps, hsp 70 in maize and hsp 27 in soybean can also be induced by water stress. Small hsps are also found to be associated with plant desiccation tolerance. These act as molecular chaperones during seed dehydration and first few days of rehydration (Hoekstra *et al.*, 2001).

Over expression of HVA1, a group 3 LEA protein isolated from barley (*Hordeum vulgare*) conferring dehydration tolerance to transgenic plants was reported (Chandra Babu *et al.*, 2001). Over expression of LEA proteins was correlated in several cases with desiccation tolerance, although the actual function of these proteins is still unknown (Villalobos *et al.*, 2004).

2.4.2 Reactive oxygen species

Stress induced production of Reactive Oxygen Species (ROS) is another aspect of environmental stress in plants (Mittler, 2002). Alleviation of oxidative damage by the use of different antioxidants and ROS scavengers can enhance plant resistance to drought. Over expression of aldehyde dehydrogenase AtALDH3 gene in Arabidopsis conferred tolerance to drought (Sonkar *et al.*,

2003). The transgenic plants showed improved tolerance to dehydration as well as to other type of stress suggesting that aldehyde dehydrogenase can maintain membrane integrity under osmotic stress. Aldehyde dehydrogenase catalyzes the oxidation of toxic aldehydes, which accumulates as a result of side reactions of ROS with lipids and proteins (Kirch *et al.*, 2004).

2.4.3 ABA responsive genes

Several different stresses may trigger the same or similar transduction pathways. The plant hormone ABA also accumulates in response to the physical phenomena of loss of water due to drought or other stresses and alleviation in endogenous ABA content is known to induce certain water deficit induced genes. Therefore, ABA accumulation is a step in one of the signal transduction pathways during water deficit. Various protein kinases have been reported in plants and are thought to function in phosphorylation processes in various signal transduction pathways, including water stress and ABA responses.

Anderberg and Walker-Simmons (1992) reported that 9-cis-epoxy carotenoid dioxygenase gene (NCED), involved in ABA synthesis was strongly induced under water deficit in the eight days old cowpea plants. A cDNA, pKABA1, is corresponding to a protein kinase, which is induced by ABA.

Available evidences suggests that stress induced responses may be ABA mediated or independent of ABA (Shinozaki and Yamaguch-Shinozaki, 1997). ABA mediated gene response may or may not require protein synthesis to take place. The induction of mRNA of rd22 gene by ABA, which showed homology to an unidentified seed protein of *Vicia faba*, required protein synthesis to take place since cycloheximide inhibited induction of the gene (Yamaguch-Shinozaki and Shinozaki, 2000).

2.5 Signaling cascades and transcriptional control

Genes involved in signaling cascades and in transcriptional control in drought tolerance such as kinases, phospholipases (Thiery *et al.*, 2004) and transcription factors (e.g. heat shock factor) and the C- repeat binding factor /dehydration responsive element binding protein (CBF/DREB) and ABA responsive element binding factor/ ABA responsive element (ABF/ABRE) families have been extensively studied (Zhang *et al.*, 2004).

Enzymes of the eukaryotic protein kinase superfamily catalyze the reversible transfer of the y-phosphate from ATP to amino acid side chains of proteins. Protein kinase function can be counteracted by the action of phosphoprotein phosphatases. Phosphorylation status of a protein can have profound effects on its activity and interaction with other proteins. An estimated 1 to 3 per cent of functional eukaryotic genes encode protein kinases, suggesting that they are involved in many aspects of cellular regulation and metabolism (Stone and Walker, 1995)

In plants, protein phosphorylation has been implicated in response to many signals, including light, pathogen invasion, hormones, temperature stress, and nutrient deprivation. Activities of several plant metabolic and regulatory enzymes are also controlled by reversible phosphorylation. As might be expected from this diversity of function, there is a large array of different protein kinases. Purification of protein kinases and their subsequent cloning, facilitated by the PCR and advances in homology-based cloning techniques, as well as functional analyses, including complementation of conditional yeast mutants and positional cloning of mutant plant genes, has already led to identification of more than 70 plant protein kinase genes. However, the precise functional roles of specific protein kinases and phosphatases during plant growth and development have been elucidated.

There are some transcriptional factors such as MYB and MYC that are involved in drought tolerance. These factors function as transcription activators in the dehydration (Abe *et al.*, 1997).

Upon over expression of DREB1A (a dehydration responsive element binding protein) under the control of rd29a promoter in *A. thaliana*, a number of stress tolerance genes were expressed and resulted in an improved tolerance under drought and several stresses (Kasuga *et al.*, 1999).

Analysis of another gene of DRE- binding protein DREB 2 showed that its promoter was induced under water stress in transgenic Arabidopsis (Nakasimha *et al.*, 2000). These genes do not require ABA for their expression but do respond to exogenous ABA.

2.6 Plant Protein Kinases and Signal Transduction

Comparison of protein kinases found in plants and other eukaryotes reveals some interesting parallels and incongruities between plant and animal signal transduction mechanisms. Many plant protein kinases are found ubiquitously in other eukaryotes (e.g. SNF1, CKII, MAPK, and CDK). Other kinases are conspicuously absent from plants, such as the cyclic nucleotide-dependent protein kinases and conventional PTKs, in spite of numerous attempts to isolate them by homology-based methods. Conversely, higher plants have unique protein kinases distinct from those found in most eukaryotes (e.g. RLKs, CDPKs, Tsl, and the PVPKI family).

An intriguing pattern emerges by comparing the plant protein kinases with the entire eukaryotic superfamily. In plants, the protein kinases implicated in the earlier events of signal transduction are mediated by unique protein kinases, but these signals converge into pathways utilizing more highly conserved protein kinases that are universal in eukaryotes. The differences observed in the initial steps of signal transduction pathways may reflect the divergence of developmental and environmental signals to which plants must respond. Protein kinases

responsible for signal recognition appear to differ (RLKs versus RTKs), as do the second-messenger-regulated protein kinases (CDPKs versus PKC or CaMK). For example, a common mechanism for transmitting signals across the plasma membrane in eukaryotes, with the notable exception of plants and yeast, involves the activation of RTKs. Ligand binding results in autophosphorylation on specific Tyr residues (Vander *et al.*, 1994).

Plant signal transduction is and will continue to be, a very exciting field. Further identification and characterization of plant protein kinases and their interactions will lead to insights into the mechanisms controlling plant growth and development. Genetic approaches, which were instrumental in identifying *CTRZ, Pto, Fen,* and *Tsl,* provide clues to function and are likely to yield a plethora of new kinases and other signaling molecules. Biochemical approaches will be useful in the identification of the specific kinases involved in regulation of the activities of plant enzymes (Huber *et al.*, 1994)

2.7 Aquaporins

Apart from the above proteins, assisting in maintaining the dehydration status; drought or osmotically stressed plants, synthesize several genes which produce water channel proteins and water transport proteins such as membrane proteins of family aquaporins that can alter the cellular water potential and thus, protect against water deficit (Chrispeels and Agre, 1994).

Neale *et al.* (2000) reported the involvement of tonoplast intrinsic protein (TIP) a plant aquaporins during severe drought stress in resurrection grass (*Sporobolus stapfianus*) along with other proteins.

2.8 Isolation of Differentially Expressed Genes through SSH

In higher eukaryotes, programs of differential gene expression mediate biological processes such as cellular growth and organogenesis. To understand the molecular regulation of these processes, the relevant differentially expressed

genes of interest must be identified, cloned and studied in detail. Subtractive cDNA hybridization has been a powerful approach to identify and isolate cDNAs of differentially expressed genes (Hara *et al.*, 1991; Hendrick *et al.*, 1984). Numerous cDNA subtraction methods have been reported, although these methods have been successful in some cases, they require several rounds of hybridization and are not well suited for the identification of rare messages (Sargent and David, 1983; Davis *et al.*, 1984). The new PCR based cDNA subtraction method, termed as suppression subtractive hybridization (SSH) overcomes the technical limitations of traditional subtraction methods (Diatchenko *et al.*, 1996).

SSH is a powerful procedure that offers the advantage of being able to compare two mRNA populations and isolate the cDNAs of genes that are either over expressed or exclusively expressed in one population and not in the other (Ablett *et al.*, 2000).

Park *et al.* (2001) have constructed subtractive cDNA library and generated EST database for the genes involved in secondary metabolism in Camellia sinensis (tea), using suppression subtractive hybridization.

Ok *et al.* (2003) have successfully utilized SSH technique to identify differentially expressed genes during flower development in Carnation. Chang *et al.* (2006) identified specially transcribed genes in sterile anther of wheat through suppression subtractive hybridization.

Way *et al.* (2004) identified and characterized genes involved in wheat adaptation to water deficit. Differentially expressed genes in wheat plants subjected to a progressive drought stress were isolated through suppression subtractive hybridization.

Gazeudam and Oilofse (2007) used SSH technique to identify and isolate the genes conferring drought tolerance in cowpea and constructed a subtracted cDNA drought expression library.

Dicto and Manjula (2005) have successfully utilized SSH in identification of genes from *Piper colubrinum* that were differentially expressed in response to the signaling molecule, Salicylic acid.

2.9 Development of Expressed Sequence Tags (ESTS)

Defining the transcriptome of a complete, multicellular eukaryote is a daunting challenge. The two most widely used and comprehensive approaches are whole genome sequencing coupled with application of gene prediction algorithms (Mathe *et al.*, 2002) and single pass sequencing of cDNAs to obtain expressed sequence tags (Adams *et al.*, 1991). Among newer approaches that have not been used as widely are targeted sequencing of gene rich regions, identified either as being hypomethylated (Rabinowicz *et al.*, 1999; Bedell *et al.*, 2005) or enriched in single copy sequences (Peterson *et al.*, 2002) and several analysis of gene expression. No one methodological approach however is sufficient and in a practical sense capable of identifying every potentially expressed gene.

Among the available approaches development of ESTs provides a number of substantial advantages (Pratt *et al.*, 2005). It is much less expensive route to gene discovery than whole genome sequencing. It offers unambiguous identification of transcribed genomic sequences. It results in a cDNA resource that can serve broad scientific community. It provides at no additional cost the templates suitable for cDNA based microarray applications. It also provides information about gene expression as a function of developmental stage, organ and /or environmental parameters at the time plant material is harvested for RNA isolation. It can reveal information about several transcript properties, including untranslated region (UTR) structures, polyadenylation signals and alternate splicing.

2.10 ESTs for abiotic stress

Diab *et al.* (2004) have developed ESTs to identify drought inducible genes and differentially expressed genes by QTL mapping in barley. Sharma and Kumar

(2005) identified three drought modulated expressed sequence tags in tea using differential display mRNA technique.

Jin *et al.* (2006) generated ESTs by large scale sequencing of cDNA clones to identify the genes and characterize the complexity of adaptation of *Leymus chinensis* to high pH sodic soil.

Iturriaga *et al.* (2006) generated EST database to understand the molecular basis of dehydration tolerance by constructing a cDNA library from *Selaginella lepidophylla,* a resurrection plant.

A collection of ESTs from cDNA libraries from leaves and roots of Burma mangrove treated with NaCl, mannitol, phytohormones and dehydration was developed and analyzed (Miyama *et al.*, 2006).

Mishra *et al.* (2007) developed ESTs to identify differentially regulated transcripts in response to abiotic (salinity, drought and cold) stresses from subtracted cDNA libraries of Pearl millet.

Diab *et al.* (2008) have developed ESTs for drought tolerance in durum wheat by QTL mapping and mapping candidate genes derived from differentially expressed genes.

2.11 Other applications of ESTs

Besides use in identification of genes involved in abiotic stress, ESTs has also been developed for use in other applications.

ESTs provide a valuable tool that can be used to identify genes in secondary metababolite synthesis. Jung *et al.* (2003) have developed Ginseng (Panax ginseng) ESTs to identify genes involved in biosynthesis of ginsenosides.

For refined annotation of the Arabidopsis genome Zhu *et al.* (2003) developed ESTs by complete EST mapping.

Suiba *et al.* (2005) have developed EST database for Lathyrus sativus to identify the gene sequences similar to other plant genes functionally associated with disease response.

Peng *et al.* (2007) generated ESTs in *Vitis vinifera*, wine grape and table grape to identify new candidate genes that are involved in the development of berry.

Schlueter *et al.* (2004) have successfully utilized EST database to resolve evolutionary events like finding out gene duplications in major crop species such as *Zea mays, Oryza sativa, Sorghum bicolor, Hordeum vulgare, Solanum tuberosum, Lycopersicon esculentum, Medicago truncatula and Glycine max.*

The growth pattern of plants can be easily studied using ESTs. Lee *et al.* (2005) generated ESTs to study the changes in gene expression during the growth phase of Poplar in suspension culture.

MicroRNAs (miRNAs) are a class of non coding RNAs that regulate gene (post-transcriptional) expression in animals and plants. Comparatively genomic computational methods have been developed to predict new miRNAs in worms, humans and Arabidopsis. Yang *et al.* (1999) successfully combined SSH and cDNA microarray technology to develop a high throughput screening procedure to identify differentially expressed genes. A cDNA microarray was used to monitor globel gene expression in response to several abiotic stresses in higher plants (Seki *et al.*, 2001, 2003). In addition, cDNA macroarray has also been proved to be a powerful tool in gene expression profiling (Ji *et al.*, 2003).

An EST-GSS (Genomic survey sequence) based combine approach for the detection of novel miRNAs in *Gossypium hirsutum* was carried out by Qiu *et al.* (2007).

Physiological, morphological and developmental changes that confer drought tolerance in plants must have a molecular genetic basis. Identification and

genetic mapping of QTL for specific drought tolerance traits combined with the mapping of candidate genes is a useful approach for dissecting the genetic basis of drought tolerance. Large expressed sequence tag (ESTs) databases have been obtained at different developmental stages from tissues and organs of plants exposed to a variety of environmental conditions.

Considering the consequences of global warming and the importance of black pepper as an important spice crop, efforts were made in the present study; to characterise the genes involved in drought tolerance in black pepper.

Kushwah (2008) has developed cDNA library through Suppression Subtractive Hybridization (SSH) specific to water stress in Black Pepper.

Materials and Methods

3. MATERIALS AND METHODS

The study on 'Characterization of subtracted cDNA library for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)' was carried out at the Centre for Plant Biotechnology and Molecular Biology (CPBMB), IT – BT Complex, College of Horticulture, Vellanikkara during the period from August, 2008 to June, 2010. Materials used and methodologies adopted for the studies are described in this chapter.

3.1 Materials

3.1.1 Chemicals, glass wares and plastic wares

All the common chemicals used in the study were of good quality (AR grade) procured from various firms such as Merck India Ltd., Sisco Research Laboratories (SRL), HIMEDIA and Sigma Research Laboratories. Molecular Bioloy reagents such as *Taq* DNA polymerase, dNTPs, buffers and molecular weight marker (λDNA /*Hind*III +*Eco*RI double digest) were obtained from Bangalore Genei. The AxyPrep DNA Gel Extraction Kit obtained from Axygen, Biosciences. The megaprime DNA labelling kit was supplied by Amersham Biosciences, USA. The plastic wares used for the study were purchased from Axygen, USA and Tarsons India Ltd.

3.1.2 Laboratory equipments

Instruments available at CPBMB were utilized for the present study. Thermal cycler (Eppendorf) was used for colony PCR, Gel DOC-It ^{TM} Imaging system UVP (USA) was used for imaging the gel, UV cross linker, Hybridization oven (Amersham life science,UK) and Phosphorimager (Fuji,FLA-5100) were used for preparation and documentation of hybridized membrane. Facilities at Radio Tracer Laboratory were utilized for radiolabelling during colony hybridization. Bioinformatics softwares were accessed from Distributed Information Centre (DIC)

3.2 Methods

The differentially expressed sequences screened out through SSH, cloned in pGEMT vector and maintained in *E. coli* (JM 109) cells with proper clone ID were designated. These transformed colonies were conserved and utilized for characterization. Since few of these ESTs were already characterized colony hybridization was carried out to detect the identical colonies in order to avoid duplication.

3.2.1 Conservation of EST library provided

EST library containing 280 clones maintained on LB plates with ampicillin were used for the study (Plate 1). In order to conserve the library, stabs and glycerol cultures were prepared.

Luria bertani (LB) medium was prepared (composition given in annexure I), autoclaved and stored at room temperature. Medium was melted and cooled to 42^0C, added ampicillin ($50mgl^{-1}$) and poured in petriplates aseptically. Colonies were streaked on solidified LB plates using flame sterile loop and incubated at 37^0C overnight. The single colonies developed on the plate were inoculated in cryo vials containing the same media, incubated overnight and stored as stabs at 4^0C. For glycerol culture, single colonies were inoculated in LB/amp ($50mgl^{-1}$). It was incubated at 37^0C overnight in a shaker at 120 rpm. The culture growth was ensured on the next day and aliquots (800 µl) of cell culture was added to autoclaved glycerol (200 µl) aseptically and stored at -80^0C for long term storage.

3.2.2 Colony hybridization

3.2.2.1 Preparation of probes

The already characterized colonies (details provided in annexure II) were randomly selected for preparation of probe. Recombinant plasmid was isolated; the

insert amplified with specific primers and eluted after gel electrophoresis for the purpose of radioactive labeling.

3.2.2.1.1 Isolation of recombinant plasmid DNA

Twenty already characterized colonies were selected randomly from the cDNA library. Plasmid DNA was isolated by alkaline mini prep method as described by Birnboim and Doly (1979). Details of reagents used are provided in Annexure III and procedure followed is as follows,

Reagents

Solution I - Resuspension buffer
Solution II - Lysis Buffer
Solution III- Neutralization buffer
LB medium
Ampicillin

Procedure

1. The cultures were revived on LB/Amp plate so as to obtain single colony.
2. Transferred single colony into 3 ml LB broth/Amp in autoclaved test tubes.
3. Incubated overnight at 37°C; 160 rpm.
4. Poured 1.5 ml of culture into a microcentrifuge tube (1.5 ml) and centrifuged at 12,000 rpm for 1 min. at 4°C.
5. The supernatant was discarded leaving the bacterial pellet as dry as possible.
6. The bacterial pellet was resuspended in 100 µl of ice cold Solution I by vigorous vortexing.
7. 200µl of freshly prepared lysis buffer was added, mixed well by inverting the tube rapidly five times.
8. Added 150µl of ice cold Solution III vortexed gently in an inverted position for 10 seconds and kept on ice for 5 min.

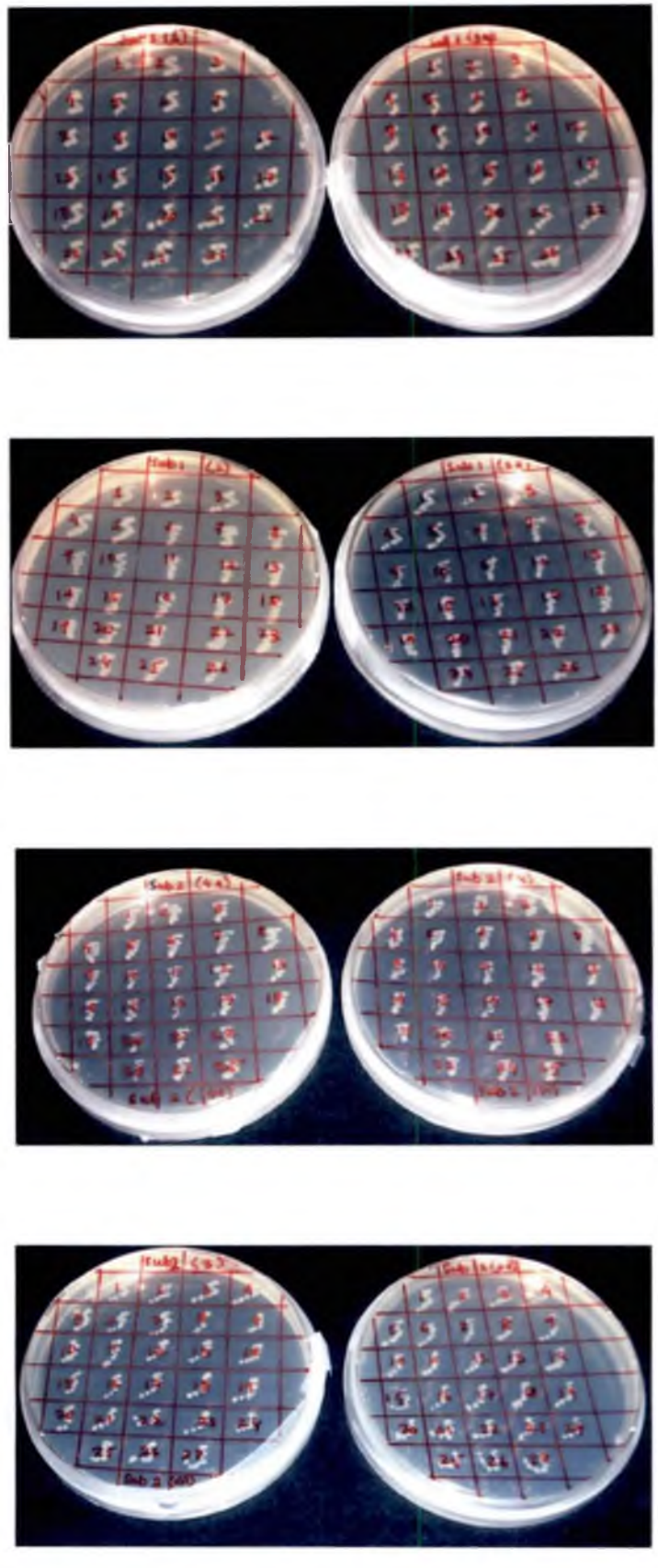


Plate 1 cDNA library established for the study

9. The contents were centrifuged at 12,000 rpm for 5 min. at 4°C and the supernatant was transferred to a fresh microcentrifuge tube.
5. To the supernatant, 0.6 volume of ice cold isopropanol was added and kept at room temperature for 2 min.
10. The contents were centrifuged at 12,000 rpm for 5 min. at 4°C and the supernatant was discarded.
11. The pellet was rinsed with one ml of 70 percent (v/v) ethanol.
12. The pellet was air dried and suspended in 30µl autoclaved distilled water.
13. The quality was checked on 0.8 percent agarose gel.

3.2.2.1.2 PCR amplification of isolated recombinant plasmid

The isolated plasmid was amplified to obtain the insert with the help of T7 and SP6 primer since it was cloned in a pGEMT vector. The primer details are as follows

T7 (Forward) TAATACGACTCACTATAGGG

SP6 (Reverse) ATTTAGGTGACACTATAGAA

Procedure followed for amplification was,

1. Diluted 1µl (≈25 ng) of plasmid DNA (step 12 of previous section) to 10 µl with sterile H_2O.
2. 1 µl of diluted plasmid DNA was added in a new PCR tube.
3. A Master Mix was prepared by addin the reagents in the sequential order as shown in the table 1.
4. The contents were mixed well and briefly centrifuged.
5. 24 µl of Master Mix was added into PCR tube of step 2.
6. The following PCR programme was run immediately:

Step1: 94°C for 2 min - Initial denaturation

Step2: 94°C for 45sec - Denaturation

Step3: 52°C for 1 min - Annealing } 30 cycles

Step 4: 72°C for 2 min - Extension

Step 5: 72°C for 10 min - Final extension

Step 6: 4°C for 10 min - Cooling of samples

Table 1 Details of PCR reaction mix prepared

Components	Vol. per reaction (µl)
10X Taq buffer	2.5
dNTP Mix (10mM)	1.0
T7 Primer (27.5nM)	1.0
SP6 Primer(31.3nM)	1.0
Taq DNA polymerase (0.3 units)	2.0
Sterile H_2O	16.5
Total volume	**24.0**

At the end of the PCR programme, the tubes were taken out and analysed on 0.8% agarose gel and the results documented.

3.2.2.1.3 Agarose gel electrophoresis

The quality of PCR amplified DNA insert was checked through agarose gel electrophoresis following the procedure as described by Sambrook *et.al.* (1989). Materials used are as follows:

1) Agarose – 0.8 percent

2) 50X TAE buffer (pH 8.0)

3) 6X Loading/ Tracking dye

5) Ethidium bromide (0.5 µg/ml)

4) Electrophoresis unit with power pack

6) UV transilluminator (Herolab ^{R})

7) Gel documentation and analysis system

Chemical composition of buffers and dye are given in Annexure IV.

Procedure

1. 1X TAE buffer was prepared from the 50X TAE stock solution. This buffer was used to fill the electrophoresis tank and to prepare the gel.
2. The open ends of gel casting tray were sealed with cello tape and placed on a horizontal platform. The comb was placed towards one end properly.
3. Agarose (0.8%) was dissolved properly in 1X TAE buffer by boiling in a microwave oven.
4. The solution was cooled to 42^{0}C and 2µl ethidium bromide solution (0.5 µg/ml) was added as an intercalating dye, to help visualization of DNA under UV.
5. The solution was poured into the gel casting tray and allowed to solidify.
6. After the gel was completely set (30 to 45 min. at room temperature), the comb and cello tape were carefully removed.
7. The gel was placed in the electrophoresis tank containing 1X TAE buffer (with the wells near the cathode).
8. The plasmid DNA (5µl) along with tracking dye (1µl) was loaded in the wells with the help of a micropipette. 2µg/ml DNA */EcoR*I + *Hind*III double digest marker was also added in the first well.
9. The cathode and anode were connected to the power pack and the gel was run at a constant voltage of 100 volts.
10. The power was turned off when the loading dye moved 2/3rd length of the gel.

11. The amplicons were visualized in a transilluminator and the gel image was documented.

3.2.2.1.4 Elution of PCR amplified fragments

Specific band corresponding to the insert size was eluted from the gel using AxyPrep DNA Gel Extraction Kit (Axygen, Biosciences). Procedure adopted as per the manufacturer's guideline is as follows

1. Band was excised from the gel using a sterile, sharp scalpel avoiding much exposure to UV on a transilluminator.
2. Gel slice was weighed in a 1.5 ml microcentrifuge tube.
3. Added 3 times gel volume of solubilization buffer (DEA buffer).
4. The gel was resuspended in gel solubilization buffer by vortexing. The mix was heated at 75°C until the gel was completely dissolved. Intermittent vortexing was given every 2 to 3 min. to enhance gel solubilization.
5. Added 0.5X (solubilized gel volume) of binding buffer (DEB buffer) and mixed properly.
6. A spin column was placed in a 2 ml collection tube. The solubilized gel was transferred into the spin column that was assembled in the 2 ml collection tube and centrifuged at 10,000 rpm for 1 min.
7. The filtrate was discarded. Added 500 µl of wash buffer (W1) to the spin column and centrifuged at 10,000 rpm for 30 sec.
8. The filtrate was discarded and 700 µl of desalting buffer (W2) was added and centrifuged at 10,000 rpm for 30 sec.

9. A second wash was given by adding 700 µl of desalting buffer (W2), followed by centrifugation at 10,000 rpm for 30 sec. to ensure the complete removal of salt. The filtrate was discarded and spin column was again placed in collection tube.

10. Column was again centrifuged for 1 min. at 10,000 rpm to remove any residual buffer.

11. Spin column was transferred to a fresh 1.5 ml microcentrifuge tube. The eluent was prewarmed at 65°C to improve the elution efficiency.

12. To elute the DNA, 7 µl of eluent was added to the centre of the spin column. It was allowed to stand for 1 min. at room temperature and then centrifuged at 10,000 rpm for 1 min.

13. Eluted fragments were checked on 0.7 percent agarose gel and stored at – 20°C for further use.

3.2.2.2 Blotting on nylon membrane

The colonies in the library were blotted to nylon membrane for hybridisation with the labelled probes in order to avoid duplication

Materials used

1. Soln A (10% SDS)
2. Soln B (0.5N NaOH, 1.5M NaCl)
3. Soln C (0.5M Tris-Cl, 1.5M NaCl pH7.4)
4. Soln D (2X SSC)
5. Nylon-backed Membrane (Hybond XL, Amersham), cut to size of 100 mm petri dish
6. Whatman 3MM or other Blotting Paper (5 no.)

7. 15 x 150 mm petri dish (4no.)

Composition of the solutions provided in annexure V

Procedure:

1. Recombinant colonies were grown in grid on LB/Amp. ($50\mu l^{-1}$) agar plate.
2. Nylon membrane was cut exactly to the size of the petri plate.
3. Meanwhile pieces of blotting paper were soaked in different solutions *viz.* A, B, C, and D (one piece of paper in each solution kept in 15×150 mm petri dish.)
4. Nylon membrane was kept on the prepared grid for two minutes.
5. Marked the membrane with pin pricks to identify orientation.
6. The membrane was taken out and placed sequentially in soaked blotting paper with colony side up and incubation time as mentioned below

    a. Soln A- 3 min.

    b. Soln B - 5 min.

    c. Soln C - 5 min.

    d. Soln D - 5 min.

7. The membrane was placed over a dry piece of blotting paper with the DNA side up and air dried for 30 min.

8. The transferred DNA was immobilized on the membrane by exposing the membrane to UV light (254 nm) for 7 min. in a UV crosslinker.

9. The membrane was dried thoroughly, kept in between filter papers and stored in a desiccator for further use.

3.2.2.3 Pre- Hybridization (Blocking)

The blocking solution will preferentially bind to the +vely charged membrane except in places where DNA is present thus making these regions unavailable for the

probe. The reagents provided by Bangalore GeNei were used for the purpose. The pre-hybridization solution consisted of Sodium citrate, denhardt's reagent, SDS and sonicated salmon sperm DNA. This solution was used at 0.2 ml per sq cm membrane area. The solution was aliquoted as 30 ml and stored at -20^{0}C.

The procedure followed for pre-hybridization was,

1. The hybridization oven was set at 68^{0}C.
2. 30ml pre hybridization buffer was added to the clean and dry hybridisation tray and kept in the hybridization oven.
3. The blotted nylon membrane was placed in warm pre -hybridization buffer and kept for pre-hybridization for 2 $^1/_2$ hr.

3.2.2.4 Probe labeling

Eluted DNA fragments (3.2.2.1.4) was labeled using Megaprime DNA labeling system (Amersham Biosciences, USA).

Reagents

- Klenow fragment (3 units/µl)
- Random primer
- Reaction buffer
- dGTP, dCTP, dTTP
- Nuclease free water
- α-^{32}P labeled dATP
- Probe DNA

The details of the reagent composition are provided in Annexure VI.

The procedure for radiolabelling the probe is as follows

1. The eluted DNA fragment was diluted to a final concentration of 5 ng/µl in nuclease free water.

2. 5 µl of diluted probe DNA was taken into a sterile eppendorf tube holed at cap followed by addition of 5 µl random primer.
3. The content was denatured by heating in boiling water bath for 5 min.
4. It was followed by brief spinning to recollect the content at the bottom of the tube and kept at room temperature.

The reagents were added to the above content one by one as follows

a. Unlabeled nucleotides - 4 µl each of dGTP, dCTP, dTTP
b. Reaction buffer - 5 µl
c. Klenow fragment - 2 µl
d. Nuclease free water- 16 µl (to make final volume 50 µl)

5. The reaction mix was spun to recollect the content at the bottom.
6. Finally, 5 µl of α-^{32}P dATP was carefully added and mixed briefly by pipetting in and out.
7. The reaction mix was spun briefly and incubated for 1hr at room temperature.
8. Following incubation, the reaction was stopped by adding 5 µl of EDTA (0.2 M).

Before adding into hybridization solution, the radiolabeled probe was denatured by heating in boiling water bath for 5 min. followed by chilling on ice.

3.2.2.5 Hybridization

The prehybridization solution in the plastic container was replaced by 50 ml of prewarmed 1X hybridization solution (Bangalore Genei, Bangalore, India). The denatured radiolabled probe of previous section was recollected at the bottom of the tube by brief spinning and then added to the prewarmed hybridization solution in the box. Different probes prepared were mixed and used for hybridization. The solution was agitated to distribute the probe evenly into the solution. The prehybridized nylon membrane was placed into the tray containing the above solution and incubated

overnight at 68^{0}C with constant agitation in hybridization oven after resealing the box with the lid.

3.2.2.6 Post hybridization washings

Post hybridization washings were given to remove the nonspecifically bound probes from the membrane and to minimize the background signal interference. Each of the stringency wash was monitored with a radioactivity monitor to retain sufficient activity on the membrane. The washings were given as follows.

1. The membrane was taken out of hybridization oven and placed in a solution containing 2X SSC and 0.5 per cent SDS for 5 min. at room temperature with intermittent shaking.
2. Second washing was given with a solution containing 2X SSC and 0.1 per cent SDS for 15 min. at room temperature with intermittent shaking.
3. The membrane was then transferred to a solution containing 0.1X SSC and 0.5 per cent SDS and incubated at 68^{0}C for 30 min. If necessary, this was repeated once again.

Finally the membrane was rinsed with 0.1X SSC at room temperature and allowed to air dry on a pad of filter paper for about 30 min. to one hour.

3.2.2.7 Image analysis by phosphorimager

The air dried membrane was exposed to an erased Imaging plate (IP, Fugifilm) for 2 h. The membrane was removed from the IP and stored in dry desiccators for further use.

The Phosphorimager (Fugifilm) was switched on 1 hr before the image analysis to warm it up.

The invisible image of the nylon membrane imitated on IP was analyzed with the Phosphorimager using FLA-5100 image analysis software (Fugifilm) platform for Windows®.

3.3 Confirmation of presence of insert

Unhybridized colonies having novel insert were picked from cDNA library and presence of the insert was confirmed by performing colony PCR amplification of recombinant plasmid DNA using T7 and SP6 Primers. The procedure followed is as follows:

1. Colonies were taken with the help of sterile inoculating loop into 20 µl H_2O in PCR tubes.
2. Spun and denatured at 94°C.
3. 2 µl of supernatant was collected in a new PCR tube.
4. A Master Mix was prepared by adding the required quantity of the reagents referred in Table 2 in sequential order.
5. The contents were mixed well and briefly centrifuged.
6. 24 µl Master Mix was added into reaction tube from step 2.
7. The PCR programme was run immediately as described in section 3.2.2.1.2.

3.4 Sequencing of cDNA clones

Stabs were prepared for the novel clones in which the presence of the insert was confirmed. Stabs were sent for sequencing at Bioserve Biotechnology Pvt. Ltd. India with instruction to use SCF format file in chromas software only. This firm used automated sequencing.

3.6. *In silico* analysis of sequences

The cDNA sequences obtained from the firm were analyzed with various online bioinformatics tools like VecScreen, Blastn, Blastx, InterProScan, CDD, Pfam.

3.6.1 Vector screening

To remove the vector regions present in the sequences, vector screening was performed using VecScreen tool (www.ncbi.nlm.nih.gov/VecScreen) provided by NCBI. The vector and the adaptor sequences present were removed using Bioedit- Biological sequence alignment editor tool.

3.6.2 Search for Homology

The nucleotide sequence of all the sequences were compared with the sequences available in nucleotide database using BLAST tool (www.ncbi.nlm. nih.gov/Blast/; Altschul *et al.*, 1997) provided by NCBI. Nucleotide- Nucleotide sequence comparison was done using Blastn tool. While nucleotide-protein sequence comparison was done using Blastx and Blastp. The best sequence alignment results were noted and saved.

3.6.3 Detection of Open Reading Frame (ORF)

To find the open reading frame of the insert nucleotide sequence, the programmed 'ORF finder' (www.ncbi.nlm.nih.gov/gorf/gorf.html) of NCBI was used. The displayed web page showed ORF sequence in all reading frames. Open reading frames available in the entire region were noted and saved.

3.6.4 Amino acid analysis

The nucleotide sequences were translated into amino acid sequences using the tool Translate (http://www.expasy.ch/tools/dna.html) provided by ExPAsy. The amino acid sequences obtained were analyzed to locate functional domains and active sites of protein using 'InterProScan' (www.ebi.ac.uk/InterProScan/; Zdobnov *et al.*, 2001).The conserved regions were detected using Conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtmlCDD) provided by NCBI. Functional motifs, domains and Hidden Markol Model (HMM) sequence were detected using Pfam tool (http://pfam.sanger.ac.uk/).

Results

4. RESULTS

The results of the investigation conducted on the 'Characterization of subtracted cDNA library for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)' undertaken during the period from 2008 to 2010 at the Centre for plant Biotechnology and Molecular Biology, College of Horticulture, Kerala Agricultural University, Vellanikkara are presented in this chapter under different subheadings. The research work included mainly the following aspects.

4.1 Maintenance of cDNA library

4.2 Preparation of probes

4.3 Colony hybridization

4.4 Selection of novel clones and sequencing

4.5 *In Silico* analysis with Bioinformatics tools

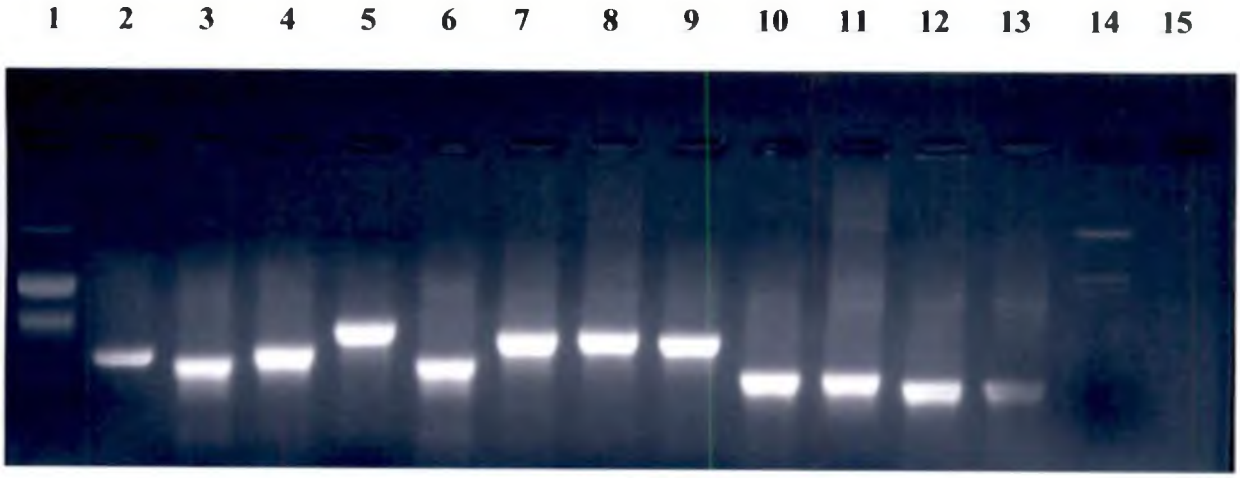
4.1 Maintenance of cDNA library

Out of the 280 clones streaked on LB/Amp. medium in petriplates, only 215 were revived. The live colonies were stored as stabs (4^0C) and as glycerol culture (-80^0C) for long term use (Plate 2).

4.2 Preparation of probes

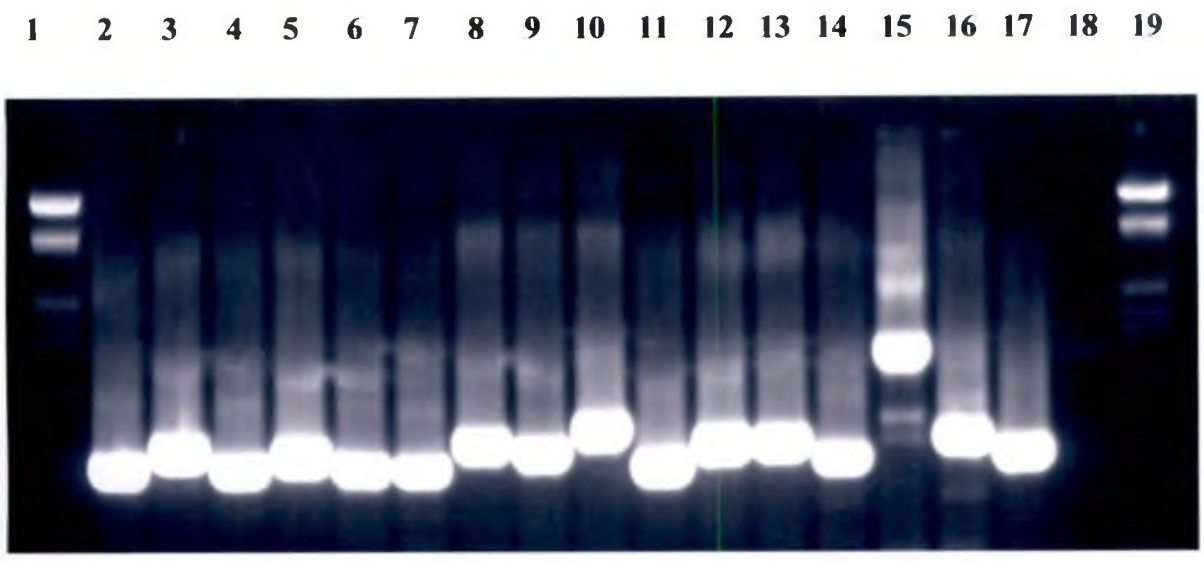
The inserts in the already characterized clones were to be used as probes for detecting identical clones in the library. The procedure followed could isolate the plasmid DNA from the selected clones. Discrete bands of 3 to 4 kb size were observed (Plate 3). These plasmid DNA when amplified with T7 and SP6 primers, provided amplicons of 500 to 900 bp size (Plate 4). The amplified PCR product was efficiently eluted by using AxyPrep DNA Gel Extraction. Purified product was good enough for further use as a probe in colony hybridization.

Plate 2 Conserved cDNA library in stabs and glycerol culture



Lane 1 & 14: λDNA /*Hind*III +*Eco*RI double digest Marker
Lane 2 to 13: Isolated plasmid
Lane 15: blank

Plate 3 Plasmid isolated from already characterized clones



Lane 1 & 19: λDNA /*Hind*III +*Eco*RI double digest Marker
Lane 2 to 17: PCR product of plasmid isolated
Lane 18: blank

Plate 4 PCR amplification of plasmid isolated from already characterized clones

4.3 Colony hybridization

Colonies to be screened were transferred on +vely charged nylon membrane. The transfered colonies were treated with different solutions as mentioned in section 3.2.2.2. The transferred DNA was immobilized on the nylon membrane by exposing to UV light in a UV crosslinker. The blotted nylon membrane was hybridized with probe and analyzed using phosphorimager. The results are represented in Plate 5. The colonies representing the already characterized clones appeared black and the others were not visible. The results are summarized as follows

Table 2 Colony hybridization to detect duplicate clones

No. of probes used	No. of colonies used for hybridization	No. of hybridized colonies	No. of colonies unhybridized
20	215	78	137

Among the 215 colonies, only 78 were hybridized and 137 were unhybridized. Thirty five unhybridized colonies were selected as novel ones for further analysis.

4.4 Selection of novel clones and sequencing

The novel clones (unhybridized colonies) were checked for the presence of insert by colony PCR using T7 and SP6 primers. The PCR products when run on 0.8 % agarose gel, showed amplicons of different sizes (500 to 900 bp). Higher molecular weight bands were obtained in PCR product of white colonies than blue colony (Plate 6). This confirmed the presence of insert in the selected clones.

The outsourcing facility available at Bioserve, Hyderabad was utilized for sequencing the novel inserts. The sequence data was obtained from the firm in the form of electropherogram and nucleotide sequence along with the clone id. The sequences thus obtained were referred to as EST and numbered as 1 to 35. The electropherograms are provided in annexure VII.

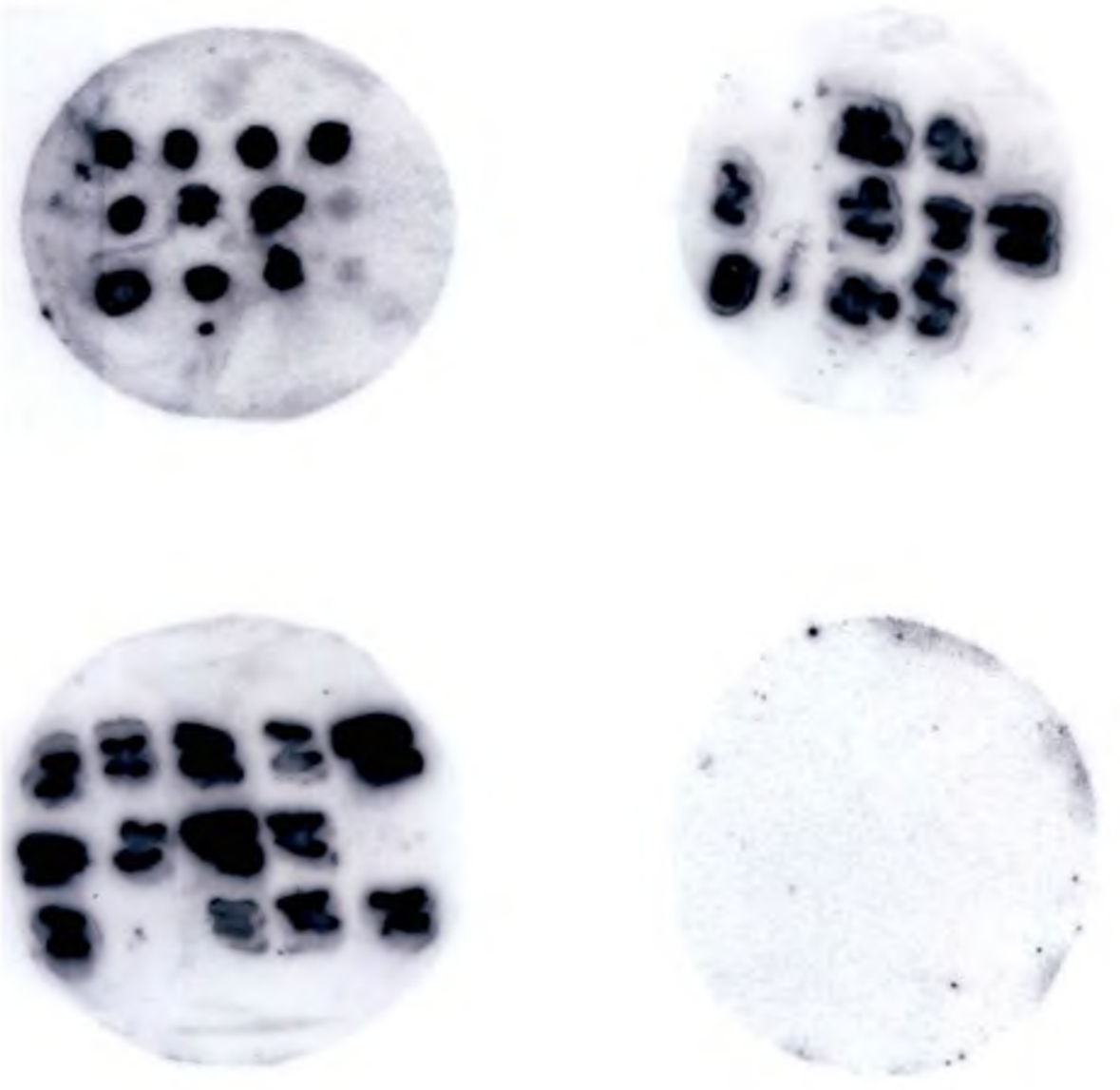
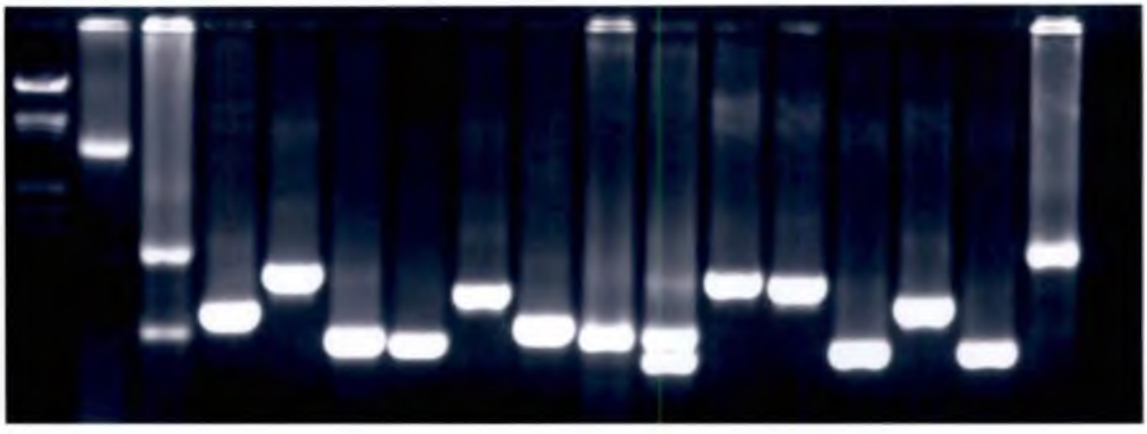


Plate 5 Signal of colonies after colony hybridization

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18



Lane 1: λDNA /HindIII +EcoRI double digest Marker
Lane 2 to 17: white colonies
Lane 18: blank

Plate 6 Confirmation of insert though colony PCR

4.5 *In Silico* analysis with Bioinformatics tools

EST- 1

The sequence data for EST-1 was obtained from the clone PNK 1(2)32 and it was of 625 bp in size. The vector screening indicated vector sequence from 1 to 18 bp. The results obtained are presented in Table 3 and Fig. 1. After vector and adaptor screening the total sequence obtained for further analysis was 608 bp as detailed below,

5'ACTCTAACGAGCTGATCATGGCCATAAAAGAAAGGCTCTTTTCGAA
ATATCATCCCAGCAAACACGCAGCCAAGGCTCCATAAATCTAAAGAGTAATC
ATAGTCTTGAAGATCAACCAAAAGTTCAGGTCCCTTGAAGTATCTGGACGCT
ACGCGAACATTGTACTCTTTGCCAGGGTGATAAAACTCAGCCAGTCCCCAGT
CTATCAACCGCAGCTTCCGCAGCTCATGATCTATCATGACGTTATGTGGCTT
GACATCTCGATGCATAATTCCTTGTGAATGACAGTAATCCAGTGCCTTCAGA
AGCTCATAAATGTAATAACGGATGTCATAATCAGTCCGGGTAGGATACAGGA
TCTTGAAATCAGTGCTGTTCACATATTCAAATATAAGGCTAGGAGTCTTTGA
ATGCTGATCTCTGACAATGTCCATCAGTTTAACAATATTTGGACCTCCGCGA
AAGATTTTGTAAACATTTTGATTTCCCTTTTTATCTTCTTTTTCTTCACGGG
CTTGAGGATTTTGATGATGCATGGAGGTCACATTGATGCCTTCAACAAACCC
CTGCACCGGGCGGC3'

Blastn results indicated 81 percent homology for 93 per cent sequence with full length cDNA clone of *Zea mays* and 89 percent of the sequence showed 82 percent homology with CK2 protein kinase (mRNA) in *Zea mays*. Blastx results indicated homology with casein kinase II alpha chain of *Ricinus communis* (bit score 216) and protein kinase 2 (bit score 212) of *Nicotiana tabacum*. The sequence also showed similarity in Blastp with protein kinase (bit score 311) and casein kinase II (bit score 312) in different plants like *Arabidopsis* and *Nicotiana tabacum*.

The sequence was found to have three ORFs with the longest one being 536 bp in length. The Interproscan results indicated the presence of active site for the protein kinase and conserved domain was also present for this protein. Pfam tools also indicated domain of the same protein with HMM sequence (40 to 201).

Table 3 Results of sequence data analysis for the EST-1 [PNK 1(2)32]

1. **Initial Sequence length** : 625
2. **Vector sequence** : 1 to18
3. **Sequence length after Vector/Adaptor Screening** : 608
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
BT035859.1	*Zea mays* full length cDNA clone	93%	2e-146	81%
NP_568098.1	*Zea mays* CK2 protein kinase (mRNA)	89%	1e-143	82%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
XP_002533161.1	casein kinase II, alpha chain, putative *(Ricinus communis)*	1e-54	216
CAD27341.1	protein kinase CK2 alpha chain *(Nicotiana tabacum)*	1e-54	212
BAC02728.1	casein kinase 2 catalytic subunit (*Nicotiana tabacum)*	1e-54	204

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_201539.2	CKA1 (CASEIN KINASE ALPHA 1); kinase (*Arabidopsis thaliana*)	1e-83	312
CAD27341.1	protein kinase CK2 alpha chain (*Nicotiana tabacum)*	2e-83	311
NP_001105632.1	casein kinase II subunit alpha *Zea mays*	2e-82	308

7. **ORFs Available:**

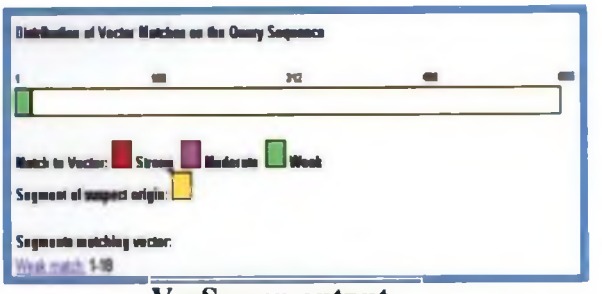
ORF Location	ORF length	Frame
1-536	536	-3

8. **Functional domains available:**

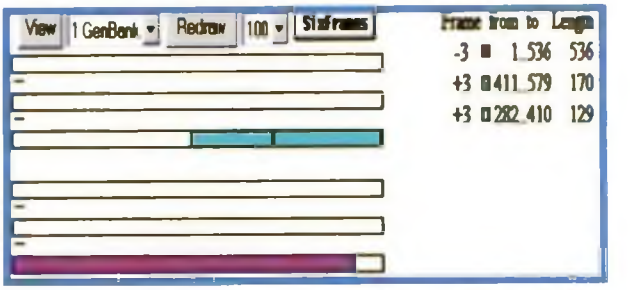
Tools	Accession no.	HMM	Domains details
CDD	c/009925	-	PKc like superfamily
Pfam	-	40-201	Protein kinase domain

9. **Interproscan scan results:**

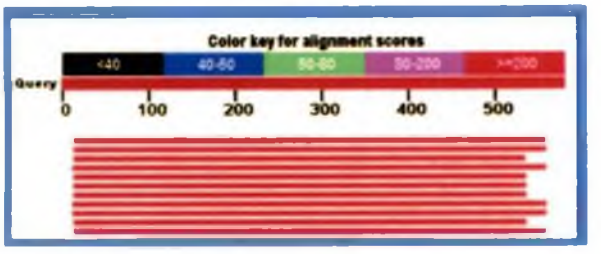
Database entry	Description
PS50011	PROTEIN KINASE DOM
PS00108	PROTEIN KINASE ST
SSF56222	Protein kinase like
PF0069	PKinase



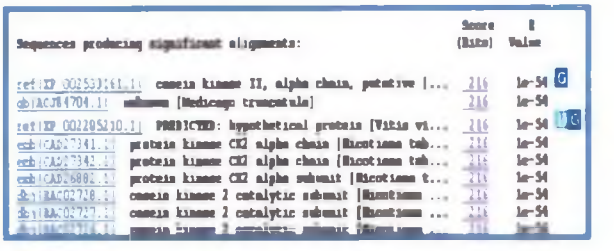
VecScreen output



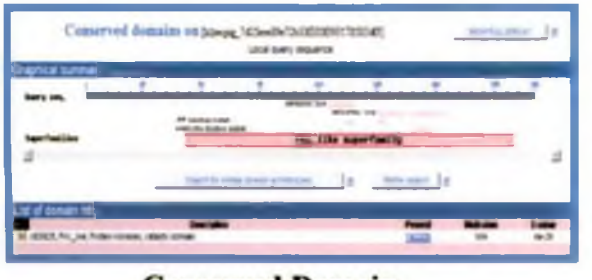
Open reading frames in the cloned sequence



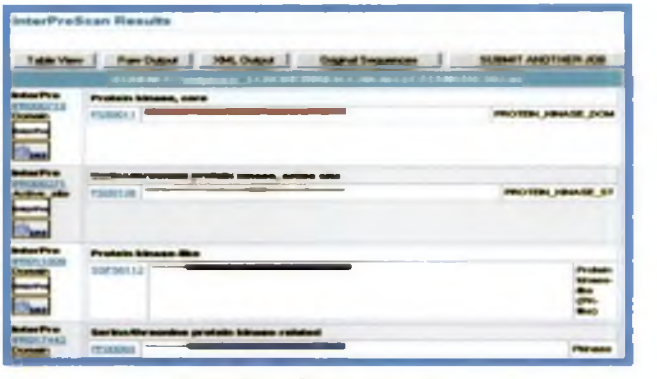
Blastn output



Blastx output



Conserved Domain



InterProScan result

PFAM results

Fig. 1 Results of sequence data analysis for the EST-1 [PNK 1(2)32]

EST - 2

The sequence data obtained for the EST- 2 was obtained from the clone PNK 1(2) 29 and it was of 700 bp in size and vector screening indicated vector sequence from 1to 22 bp. The results obtained are presented in Table 4 and Fig. 2. After vector and adaptor screening the total sequence obtained for further analysis was 678 bp., as shown below

5' CACTGAAGCTCTAAATGAACTTCTAGAATCTGATGATAAATTTGGGTTCA
TAATTATGGATGGTAATGGGGCTCTATTTGGCACCCTCAGTGGAAATGCACG
AGAAGTCTTACATAAGTTCACTGTTGACTTGCCTAAGAAGCATGGAAGAGGA
GGGCAATCTGCACTTCGTTTTGCACGTCTTCGAATGGAAAAGCGTCATAATT
ATGTTCGCAAGACAGCTGAGCTCGCTACTCAATATTTCATTAATCCGGCAAC
TAGTCAGCCGAATGTTGCAGGTCTGATATTGGCAGGTAGTGCCGATTTCAAG
ACAGAGTTGAGTCAGTCGGACATGTTTGACCCAAGACTTCAAGCTAAGATTC
TGAATGTGGTGGATGTATCGTATGGTGGTGAAAATGGTTTTAATCAGGCAAT
AGAGCCTTCTTCAGAGATCCTTTCAAATGTGAAATTCATTCAAGAGAAGAAG
TTGATTGGAAAGTATTTTGAGGAAATAAGCCAAGACACAGGGAAATATGTTT
TTGGTGTAGAGGATACGCTGAAGGCTCTAGAAATGGGTGCAGTGGAGATTCT
TATTGTCTGGGAAAACTCTTGACATTAGTCGTTATGAGCTTAAAAATAGTGT
GACTGGAGAAGTTTTCATCCACAGCCATTTGAGTAGGGGCCACAAAACCTGA
TCCA3'

Blastn results indicated that 93 percent of the sequence showed 79 percent homology with hypothetical protein and eukaryotic peptide chain release factor subunit-1 in *Sorghum bicolor* and *Zea mays* respectively. Blastx analysis indicated homology with eukaryotic peptide chain release factor subunit (bit score 367) and eukaryotic release factor 1-3 (bit score 365) in *Ricinus cummunis* and *Arabidopsis thaliana*. The sequence also showed similarity in Blastp for the same proteins with 348 to 349 bit score.

The sequence was found to have two ORFs with the longest one being 537 bp in length. The Interproscan results indicated the presence of active site for eukaryotic peptide chain release factor subunit-1. The sequence has ERF_2 super family conserved domain. Pfam tools indicated HMM sequence (1 to 133) for the same protein family

Table 4 Results of sequence data analysis for the EST-2 [PNK 1(2)29]

1. Initial Sequence length : 700
2. Vector sequence : 1 to 22
3. Sequence length after Vector/Adaptor Screening: 678
4. Blastn Results:

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM00246659.1	Sorghum bicolor hypothetical protein, mRNA	93%	2e-154	79%
NM001158066.1	*Zea mays* eukaryotic peptide chain release factor subunit-1	93%	9e-152	79%

5. Blastx Results:

Accession No.	Description	Evalue	Score (Bits)
XP_002512711.1	Eukaryotic peptide chain release factor subunit *(Ricinus communis)*	2e-70	367
NP_189295.3	ERF1-3 (eukaryotic release factor 1-3); translation release factor (*Arabidopsis thaliana*)	7e-69	365

6. Blastxp Results:

Accession No.	Description	Evalue	Score (Bits)
XP_002512711.1	eukaryotic peptide chain release factor subunit, putative (*Ricinus communis)*	6e-95	349
ACZ71034.1|	eukaryotic release factor 1-2 (*Brassica oleracea var. botrytis*)	9e-95	348
NP_172752.1|	ERF1-3 (eukaryotic release factor 1-3); translation release factor (*Arabidopsis thaliana*)	1e-94	348

7. ORFs Available:

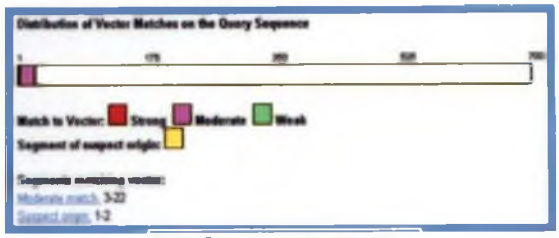
ORF Location	ORF length	Frame
56-592	537	+1

8. CDD available:

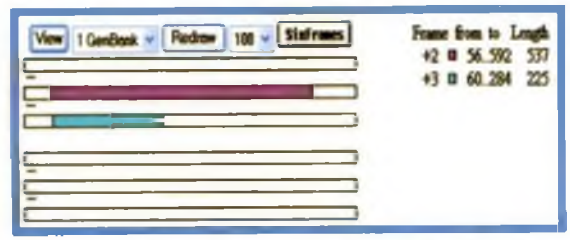
Tools	Accession no.	HMM	Domains details
CDD	03464, 03465	-	eRF 1 super family
Pfam	-	1- 133	eRF 1 domain

9. Interproscan scan results:

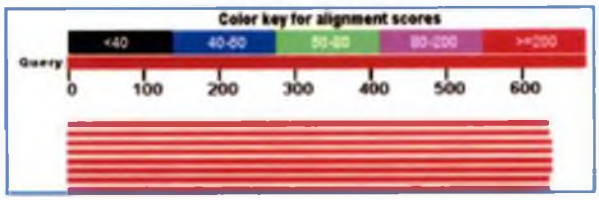
Database entry	Description
PTHR10113, PTHR10113:SF1	eukaryotic peptide chain release factor subunit1
PF03464, PF03465	ERf1_2
SSF55315	L 30e- like
SSF3137	Translational Machinery Components

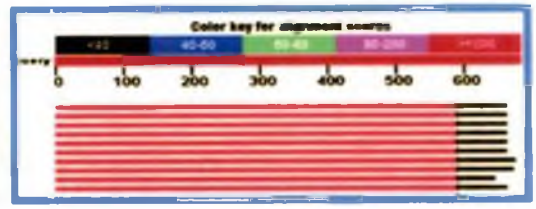


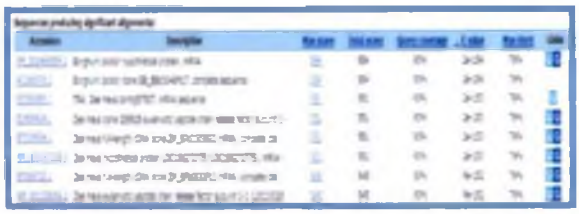
VecScreen output



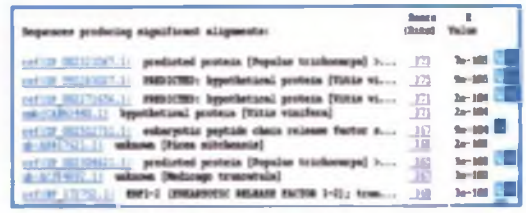
Open reading frames in the cloned sequence





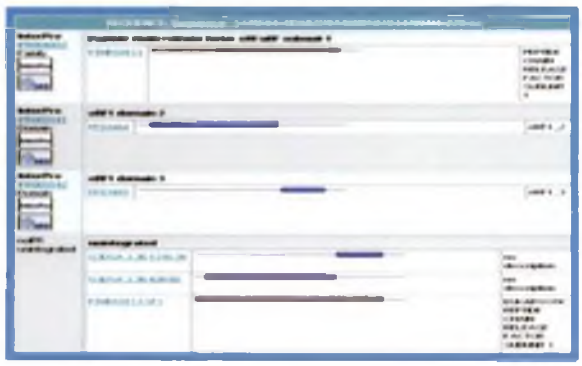


Blastn output



Blastx output

Conserved Domain



InterProScan result

Pfam result

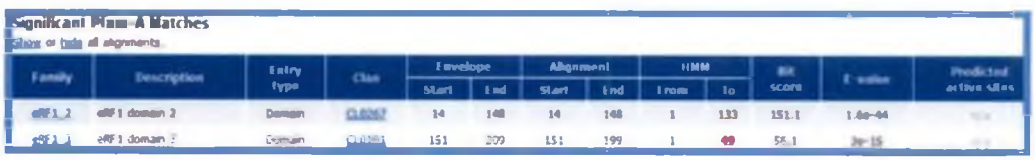


Fig. 2 Results of sequence data analysis for the EST- 2 [PNK 1(2)29]

EST-3

The sequence data obtained for the EST-3 was obtained from the clone PNK 1(2) 30 and it was of 539 bp in size and vector screening indicated vector sequence from 452 to 539 bp. The results obtained are presented in Table 5 and Fig. 3. After vector and adaptor screening the total sequence obtained for further analysis was 441 bp., as indicated below

5′ CAGTCCTGACNGCGGATTCGAGCGGCGCCGCGGCAGGTACNGGCAT
TGAGCAGGAATACACTTTGCTCCAAACAAACGTCCAATGGCCCCTGGGTTGG
CCAGTAGGAGGCTACCCTGGTCCACAGGGACCTTATTACTGTGGAAGTTGGA
GCAGACAAATCATATGGTCGTGACATATCAGATGCTCATTACAAAGCCTGCT
TATATGCTGGAATCAACATTAGTGGAACTAATGGTGAAGTCATGCCTGGCCA
GTGGGAGTATCAAGTTGGGCCAAGCGTTGGAATTGAAGCGGGAGACGACATC
TGGTGCTCAAGATATATCCTTGAGCGGATCACTGAACAAGCTGGCGTTGTCC
TCTCTCTTGATCCAAAACCAATCCAGGTGACTGGAATGGAGCTGGATGTCAC
ACTAACTACAGTACCTCGGCCGCGACCACGC3′

In Blastn analysis, 86 percent of the sequence showed 83 percent homology with glutamine synthetase mRNA in *Phragmites austrails* and *Ricinus communis*. Blastx and Blastp results also indicated homology with glutamine synthetase in different plants like *Cucumis melo*, *Nicotiana attenuate* and *Solanum tuberosum* with bit scores ranging from 128 to 164.

The sequence was found to have five ORFs with the longest one being 327 bp in length. The Interproscan results indicated the presence of active site glutamine synthetase domain.

The sequence had conserved domain for same protein. Pfam tools indicated HMM sequence (16 to 63) for the same protein.

EST-4

The sequence data obtained for the EST-4 was obtained from the clone PNK 1(2)65 and it was of 608 bp in size and vector screening indicated vector sequence from 573 to 608 bp. The results obtained are presented in Table 6 and Fig. 4. After vector and adaptor screening the total sequence obtained for further analysis was 572 bp.

Table 5 Results of sequence data analysis for the EST-3 [PNK 1(2)30]

1. **Initial Sequence length** : 539
2. **Vector sequence** : 452 to 539
3. **Sequence length after Vector/Adaptor Screening:** 441
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AB16176.1	phragmites austrails GS2 mRNA for plastidic glutamine synthetase	86%	2e-108	83%
XM002516755.1	*Ricinus communis* glutamine synthetase mRNA	86%	2e-100	83%

5. **BlastX Results:**

Accession No.	Description	Evalue	Score (Bits)
AAX35343.1	glutamine synthetase (*Cucumis melo)*	3e-46	146
AAO62992.1	chloroplast glutamine synthetase *(Nicotiana attenuate)*	1e-44	134
AAG40236.2	plastid glutamine synthetase GS2 *(Solanum tuberosum)*	7e-44	128

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
AAX35343.1	glutamine synthetase (*Cucumis melo)*	2e-39	164
AAK07678.1	glutamine synthetase GS2 (*Beta vulgaris)*	5e-38	160
AAO62992.1	chloroplast glutamine synthetase (*Nicotiana attenuate)*	2e-37	158

7. **ORFs Available:**

ORF Location	ORF length	Frame
155-440	327	+1

8. **Functional domains available:**

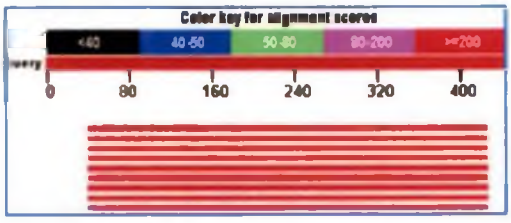
Tools	Accession no.	HMM	Domains details
CDD	CL0286	-	glutamine synthetase superfamily
Pfam	-	16-63	glutamine synthetase catalytic domains

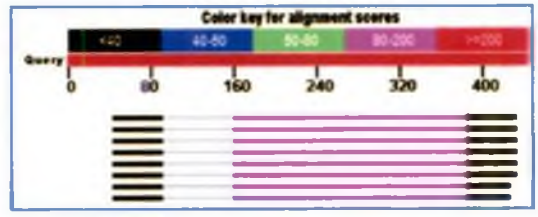
9. **Interproscan scan results:**

Database entry	Description
PTHR20852, PTHR20852.SF14	glutamine synthetase, glutamine synthetase (Glutamaleammonia-ligase)

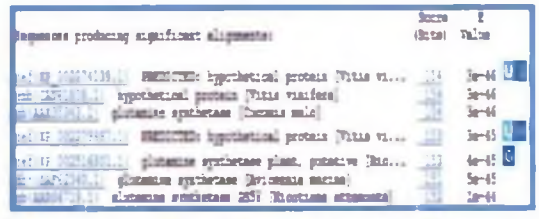
VecScreen output

Open reading frames in the cloned sequence



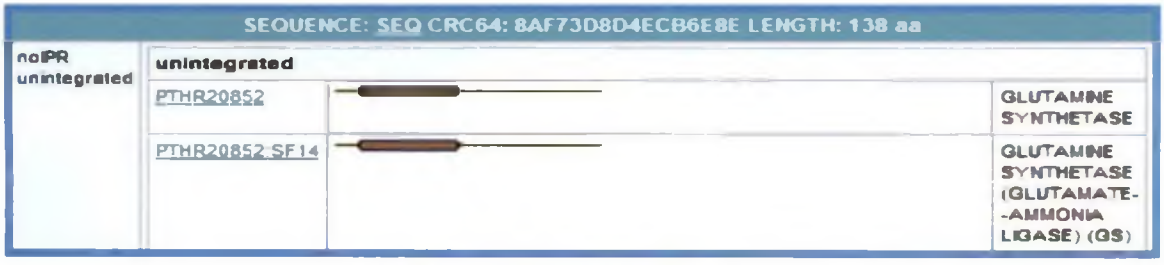


Blastn output

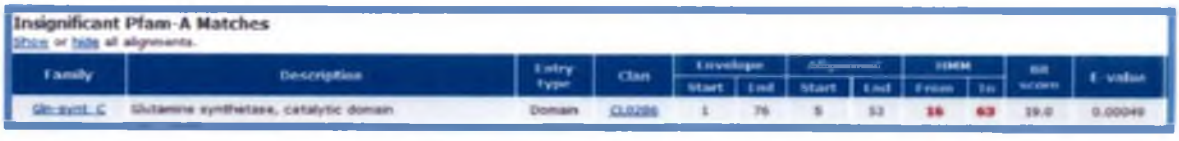


Blastx output

Conserved Domain



InterProScan result



Pfam result

Fig. 3 Results of sequence data analysis for the EST-3 [PNK 1(2)30]

Table 6 Results of sequence data analysis for the EST-4 [PNK 1(2) 65]

1. **Initial Sequence length** : 608
2. **Vector sequence** : 573 to 608
3. **Sequence length after Vector/Adaptor Screening** : 572

4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM002268662.1	Vitis vinifera hypothetical protein	50%	2e-21	69%
XM002518058.1	transcription factor *(Ricinus communis)*	28%	1e-16	73%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
ABK94861.1	unknown (*Populus trichocarpa)*	8e-26	120
XP_002518104.1	transcription factor, putative *(Ricinus communis)*	3e-25	118
NP_566744.1	myb family transcription factor (*Arabidopsis thaliana*)	4e-11	71.6

6. **Blastp Results:**

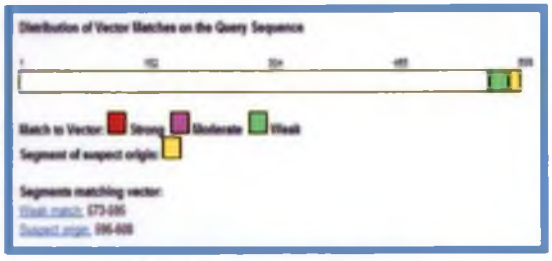
Accession No.	Description	Evalue	Score (Bits)
NP_974356.1	myb family transcription factor (*Arabidopsis thaliana*)	1e-16	113

7. **ORFs Available:**

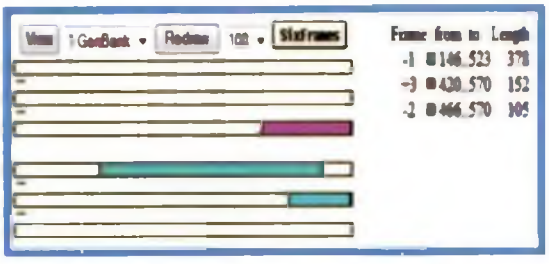
ORF Location	ORF length	Frame
420-570	378	+3

8. **Functional domains available :**

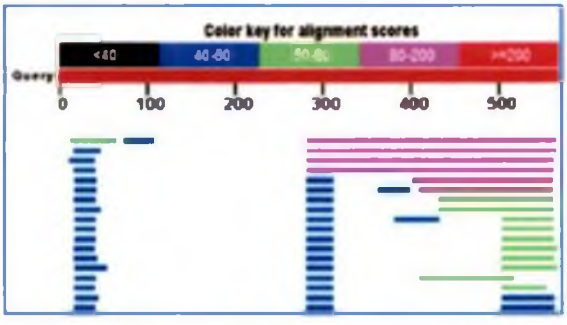
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
Pfam	-	Nil	Nil

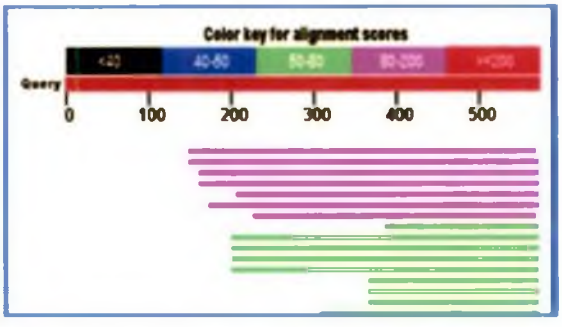


VecScreen output

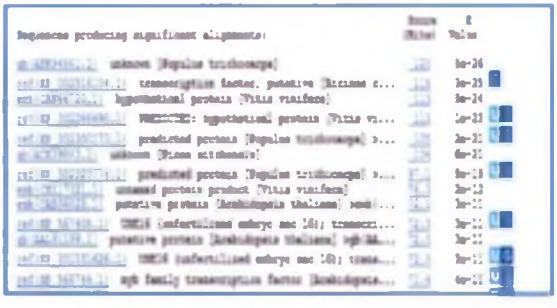


Open reading frames in the cloned sequence





Blastn output



Blastx output

Fig. 4 Results of sequence data analysis for the EST-4 [PNK 1(2)65]

5′GCCGCAGGTTCGACGGGATAGCGTGGTCGCGGCCGAGGTACAAACTATTC
ACCCATTTTAACTAGGGGAAAACCATTTTGTCATCACACACAATTGCAGCAG
CACATTTGCCGGCCATGACATCCAACGTCGCACTTTAGATAGCACTGCTTAC
CCACTGTGGTTCTTGCTTCATATCACCGTCCCAAGCCACAACTTCGCCTTCA
CCAAAAACAGGCCTATGCCGTTTTTTAAGAGCATTCTGCTGACCAACCATGG
ACACCCTTGGGCTTTCTGTTACTGATGTCAAGCAGCTTTCCACAGAACAGTC
ACCAATCTGAGCCGCTGGGGCATTTACAAATTTTTCACCAAGACAAGCTGCC
ATAATATCAGACAAAGATGGCAGTTTTGATGCAACTTCTGGAGAGTCATTGG
AAACCTTGGTAACAAGATCAGACAGCTCTTGTCTTGGGGTTTCTATCCCAGC
AGATATGACAACCTGCTCAGTGAGTGCTTTGCAGGCCTTCTCCAGCATATCC
TGTAGATATTTTCCTTGGGCTTCAATGCGAAGATGGAGATGGCG3′

Sixty nine percent homology was observed for 50 per cent sequence with hypothetical protein from *vitis vinifera* and 73 per cent homology with transcriptional factor from *Ricinus communis* in Blastn analysis. Blastx and Blastp results showed homology with transcriptional factor (bit score 118) in *Ricinus communis* and MYB family Transcriptional factor (bit score 71.6 to 113) in *Arabidopsis thaliana*.

The sequence was found to have three ORFs with the longest one being 378 bp in length. The Interproscan and CDD analysis indicated no positive results.

EST-5

The sequence data obtained for the EST-5 was obtained from the clone PNK 1(2) 55 and it was of 688 bp in size. Vector screening indicated vector sequence from 1 to 21 bp and 400 to 566. The results obtained are presented in Table 7 and Fig. 5. After vector and adaptor screening the total sequence obtained for further analysis was 380 bp., as shown below

5′ACAGAACAGAATCCTCTTATTTTGTCGAATGGATCCCTAACAATGTGAAA
TCCAGCATCTGTGATATTCCGCCAAAGGGTCTGAAAATAGCTTCGACCTTCA
TTGGTAACTCGACCTCTATTCAGGAGATGTTCAGGAGGGTTAGTGAACAGTT
CACTGCCATGTTCAGCAGGAAGGCTTTCTTGCACTGGTACAATTAATGGCTG
TTTTTTGTATCGAACTGTGGTAAACGCAATGATGTTGTCTGATCCAGAAAGC
TGAGCAAAAGGGTGCTCTTTCTTGTATCTCTTTAGTTCCACCCGCCCCTCGG

Table 7 Results of sequence data analysis for the EST- 5 [PNK (2)55]

1. **Initial Sequence length** : 688
2. **Vector sequence** : 1 to 21, 400 to 566
3. **Sequence length after Vector/Adaptor Screening** : 380

4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AK285306.1	Glycine max clone	49%	3e-54	86%
XM 002863542.1	Arabidopsis lyrata sub sp. Tubulin beta-4 chain	51%	6e-50	84%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
ACM89774.1	beta-tubulin 14 *(Gossypium hirsutum)*	6e-04	125
XP_002510540.1	tubulin beta chain, putative (*Ricinus communis)*	5e-27	125

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
BAA11417.1	aspartate kinase-homoserine dehydrogenase (*Oryza sativa)*	1e-18	95.9
AAC05981.1|	aspartokinase-homoserine dehydrogenase *(Glycine max)*	4e-17	90.9

7. **ORFs Available:**

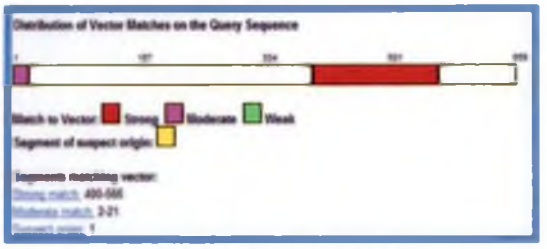
ORF Location	ORF length	Frame
1-344	344	-3

8. **Functional domains available :**

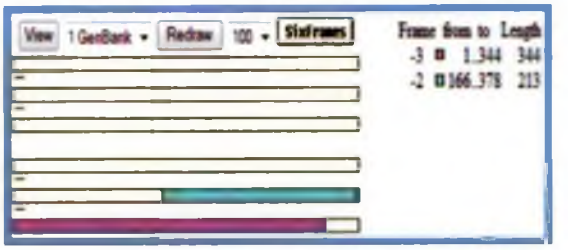
Tools	Accession no.	HMM	Domains details
CDD	0742	-	dehydrogenase dh superfamily
Pfam	-	105-159	homoserine dehydrogenase

9. **Interproscan scan results:**

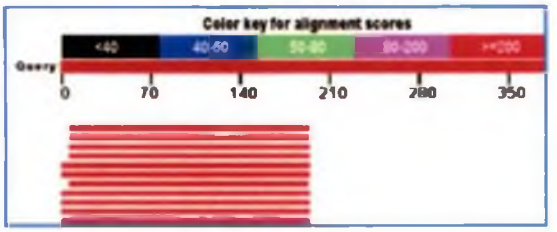
Database entry	Description
PF00742	Homoserine dh
PTHR21499, ;SF1	Aspartate kinase

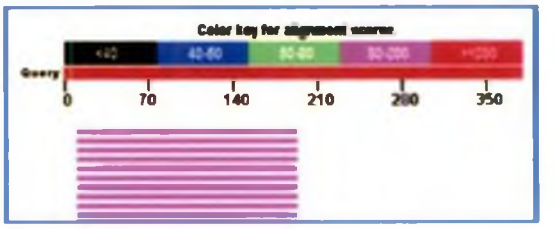


VecScreen output



Open reading frames in the cloned sequence

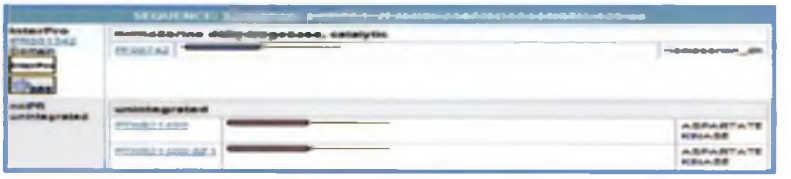




Blastn output

Blastx output

Conserved Domain



InterProScan result

Pfam result

Fig. 5 Results of sequence data analysis for the EST-5 [PNK 1(2)55]

AGTTGATAACATCCACCACACCCACGTATCTCAAAACCTCTCCAGTACCTGC
CCGGGCGGCCGCTCGAA3'

Blastn results indicated that about 50 percent of the sequence had over 80 percent homology with clone from *Glycine max* and beta tubulin in *Arabidopsis lyrata*. In Blastx results the mRNA encoding beta-tubulin in *Gossypium hirsutum* (score bits 125) and in *Ricinus cummunis* (score bits 125). The sequence showed similarity in Blastp with aspartate kinase-homoserine dehydrogenase from *Oryza sativa* (score bits 95.9) and *Glycine max* (score bits 90.9).

The sequence was found to have two ORFs with the longest one being 344 bp in length. The Interproscan results indicated the presence of active site for the aspartate kinase. The sequence has Homoserine dehydrognase superfamily conserved domain. Pfam tools indicated domain for the same protein with HMM sequence 105 to 159.

EST- 6

The sequence data obtained for the EST-6 was obtained from the clone PNK 3 and it was of 799 bp in size. Vector screening indicated vector sequence from 1to60 and 599 to799 bp. The details of results obtained are presented in Table 8 and Fig. 6. After vector and adaptor screening the total sequence obtained for further analysis was 539 bp., as indicated below

5'TCGAGCGGCCGCCCGGGCAGGTACATCAAATACAACGGTGGTCTCGACAC
AGAGGAGTCCTATCCATACGCCGGTGTCAATGGCATCTGCGGATACAAGATA
GAAAACATTGGTGTCAAGGTCGCTGAGTCCGTGAATATCACAGAGGGCGCCG
AAGATGAATTGAAACATGCAGTCGCTTTGGTCCGTCCCGTCAGTATTGCGTT
CCAGGTTGTGCACGACTTCCGCTCATACAAAGGAGGGGTTTACACAAGTCAA
GAGTGTGGTAGCGCTCCCATGGATGTAAACCATGCTGTTCTGGCTGTCGGTT
ATGGTGTGGAGAATGGCGTGCCATATTGGCTGGTCAAGAATTCATGGGGAAA
TGATTGGGGGGTTGATGGTTATTTTAAGATCGAGCTCGGAAAGAACATGTGT
GGTGTTGCTACTTGTGCATCTTATCCTATTCTCTCTCTGTAATTGACTGAGC
TCGTCGGGGCATCAACCCCAAACACTGCATCGTTTATCTATTCTGTATCATC
CATTCAAATGTGCTCGTGATTGTACCTCGGCCGCGACCACGCTA3'

Table 8 Results of sequence data analysis for the EST-6 [PNK 3]

1. **Initial Sequence length** : 799
2. **Vector sequence** : 1 to 60, 599 to799
3. **Sequence length after Vector/Adaptor Screening** : 539
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
ABR19829.1	cysteine proteinase (*Elais quineensis*)	76%	2e-66	82%
ABW71226.1	Cysteine protease (*Nicotina tabacum*)	76%	1e-63	78%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
ABR19829.1	cysteine proteinase (*Elais quineensis*)	2e-25	118
NP001105479.1	Cysteine protease (*Zea mays*)	6e-25	117

6. **ORFs Available:**

ORF Location	ORF length	Frame
1-210	210	-2

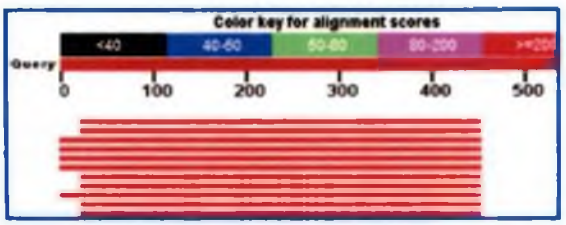
7. **Functional domains available :**

Tools	Accession no.	HMM	Domains details
CDD	Cl0298	-	Peptidase family
Pfam	-	166-220	Papain family cysteine protease

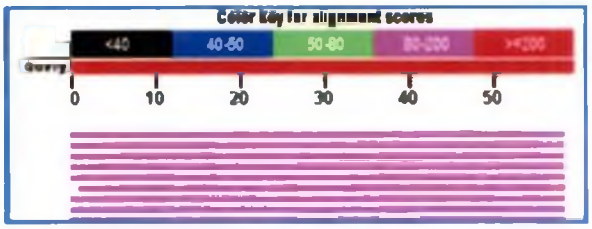
8. **Interproscan scan results:**

Database entry	Description
PS00639	cysteine proteinase active site
PF00112	Peptidase C1A

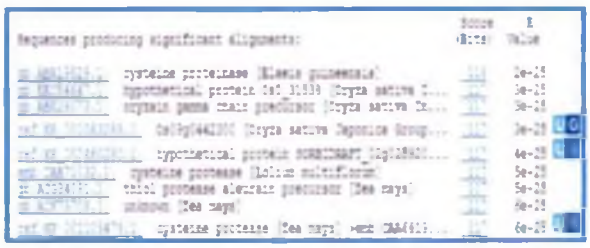
VecScreen output



Open reading frames in the cloned sequence

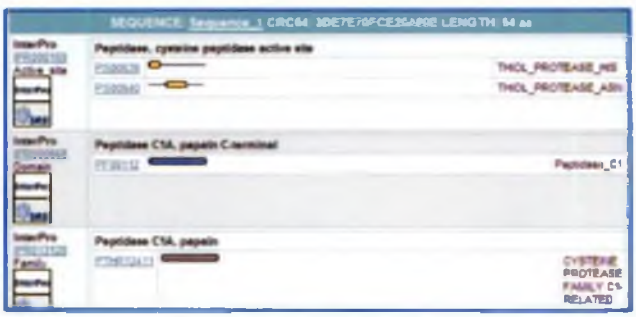


Blastn output



Blastx output

Conserved Domain



InterProScan result

Pfam result

Fig. 6 Results of sequence data analysis for the EST- 6 [PNK 3]

In Blastn analysis 76 percent of the sequence showed over 80 percent homology with cysteine proteinase in *Elais quineensis* and cysteine protease in *Nicotina tabacum*. Blastx results also gave homology with the mRNA encoding same protein from *Elais quineensis* and *Zea mays* with bit score 117 to 118.

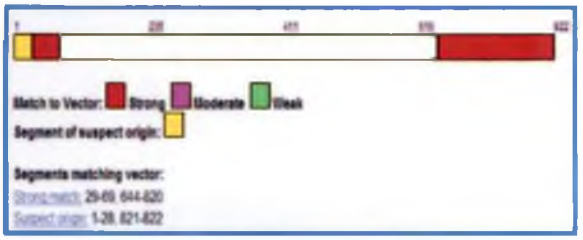
The sequence was found to have six ORFs with the longest one being 210 bp in length. The Interproscan results indicated the presence of active sites for the cysteine proteinase. The sequence had conserved domain for the same protein. Pfam tools indicated HMM sequence (166 to 220) for the same protein.

EST- 7

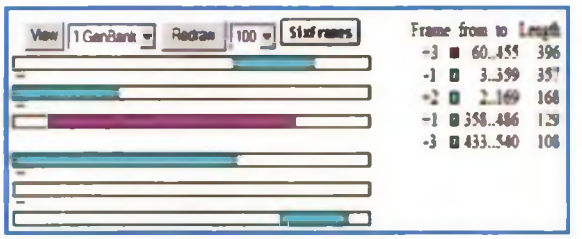
The sequence data obtained for the EST-7 was obtained from the clone PNK 5 and it was of 822 bp in size. Vector screening indicated vector sequence from 1-69 to 644-822 bp. The details of results obtained are presented in Table 9 and Fig. 7. After vector and adaptor screening the total sequence obtained for further analysis was 575 bp., as shown below

5' CGTGGTCGCGGCCGAGGTACAACGACACGCAGCGCCACCGGTTGGGTCCG
AACTATCTGATGCTGCCGGTGAACGCACCCAAGTGTGCCTACCACAACAATC
ACTACGACGGTGCCATGAATTTCATGCACAGGGACGAAGAGGTGGACTACTT
CCCATCGAGGTATGATCCGGTTCGTCACGCCGAGAAGTTCCCCATCCCAACC
AGGCTCATCACTGGCAAGCGAGAGAAGGCCACTATAAACAAGGAGAACAACT
TCAAGCAGGCCGGGGATCGTTACCGGTCCTTCGATCCTGCCAGGCAAGAGAG
GTTCATCACTCGTGTAGTTGAAGGGCTGTCGGACCCACGCCTCACACATGAG
CTTCGAAACATCTGGATCTCATATTGGACTCAGTGTGACCAGTCACTGGGCC
AGAAAATCGCTTCTCGCCTCAGCATGCGATCCAACATTTGAGGATTTCAGAA
GGGTGGAGATGGCCAGTTGAATATGCTGGAATTACCTGTACCTTTGATATGA
TACGATGTTTAATTATACTCATGTAATTATATGTACCTGCCCGGGCGGCCGC
TCGAA3'

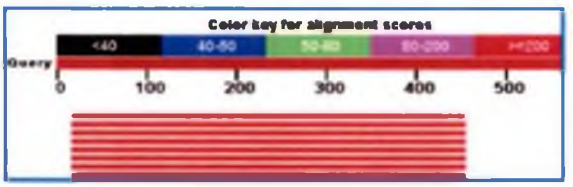
In Blastn analysis 75 per cent of the sequence showed 77 percent homology with catalase isozymes and catalase from *Oryza sativa*. Blastx results indicated mRNA encoding catalase isozymes and catalase in *Oryza sativa* with 225 bit score.

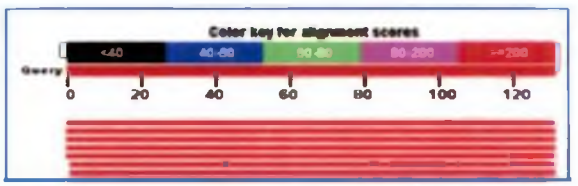


VecScreen output

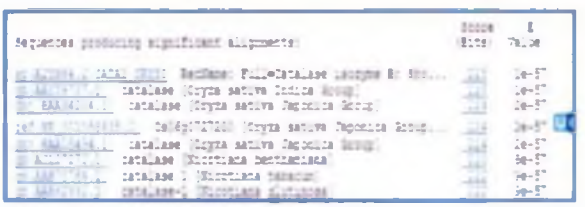


Open reading frames in the cloned sequence



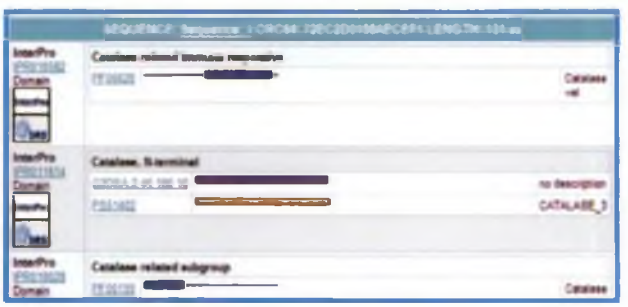


Blastn output

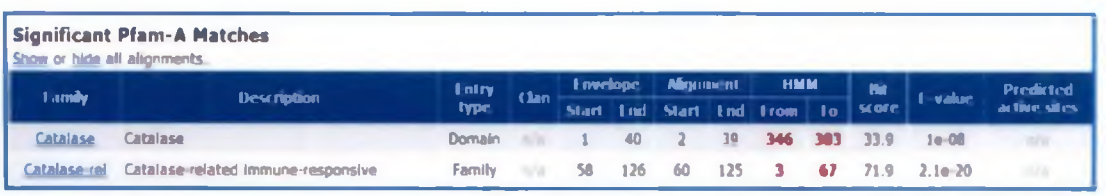


Blastx output

Conserved Domain



InterProScan result



Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope Start	Envelope End	Alignment Start	Alignment End	HMM from	HMM to	Bit score	E-value	Predicted active sites
Catalase	Catalase	Domain		1	40	2	39	346	383	33.9	1e-08	
Catalase-rel	Catalase-related immune-responsive	Family		58	126	60	125	3	67	71.9	2.1e-20	

Pfam result

Fig. 7 Results of sequence data analysis for the EST-7 [PNK 5]

Table 9 Results of sequence data analysis for the EST-7 [PNK 5]

1. Initial Sequence length : 822
2. Vector sequence : 1 to 69, 644 to 822
3. Sequence length after Vector/Adaptor Screening : 575
4. Blastn Results:

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
A2YH64.2	Catalase isozymes*(Oryza sativa)*	75%	8e-66	77%
BAA05494.1	Catalase *(Oryza sativa)*	75%	1e-63	77%

5. Blastp Results:

Accession No.	Description	Evalue	Score (Bits)
A2YH64.2	Catalase isozymes	1e-57	225
AAY59707.1	Catalase *(Oryza sativa)*	1e-57	225

6. ORFs Available:

ORF Location	ORF length	Frame
60-455	396	+3

7. Functional domains available :

Tools	Accession no.	HMM	Domains details
CDD	CI09506	-	Catalase super family
Pfam	-	346-383	Catalase

8. Interproscan scan results:

Database entry	Description
PF06628	Catalase

The sequence was found to have five ORFs with the longest one being 396 bp in length. The Interproscan results indicated the presence of active sites for the catalase.

The sequence has catalase like superfamily conserved domain. Pfam tools indicated HMM sequence (346 to 383) for the same domain.

EST- 8

The sequence data obtained for the EST-8 was obtained from the clone PNK 21 and it was of 778 bp in size. Vector screening indicated vector sequence from 1-57 bp. The results obtained are presented in Table 10 and Fig. 8. After vector and adaptor screening the total sequence obtained for further analysis was 721 bp.

5' CGTGGTCGCGGCCGAGGTGTCAGACTTTGGATTGGCTCGCCAAGGGCCTG
GTGAAGGACTTAGTCATGTCTCAACTGCGGTTGTTGGGACTATAGGGTATGC
AGCACCAGAATATATTCAAACTGGACGTTTGACGGCTAAAAGTGATGTTTGG
AGCTATGGAGTTGTCCTCTATGAACTGATCACTGGTAGAAAGCCAATAGACC
GTAATCGTCCTAAGGGTGAACATAAACTACTTGAGTGGGTGAAGCCTTATTT
ATCAGAGAAGCGTTTTTCCATGATTCTTGATCCAAGAATACAAGAATAGATT
TCCCTGAAGTCGGCCCAAAAGCTTGCCGCTGTTGCAAACAAATGCCTTGTAC
GACAGCAGAAGGCACGTCCAAAGATGAGTGAGATTTTAGCGATTGTGAAGCA
AATTGTGGAGTCATCAACAACTGGGTCTCCTCAACCTGCCATCAAAATCCTA
GATTCAAGTGAAGTCTCCAATGAAGAACACAAAAACAGAAGCCGATTGAAAC
GAGCCATTGCTTATCTTAATATGGGTGGAGATGGCAAGAGGTTTTACTGGTG
GAAGTTCAAATCCAAGCTCAGTAAAGTTAAGTTGTGGAAGACGAAGATAGGA
CTAATGGTTTCTAGGCTAGCGTCATGATAAGAAACCAGGCATAGGAGAAGTT
GGGAAGTTTTTAGAGATGGCCTAAAGAATCCACCATAAAGTGTCAGT3'

In Blastn results 71 per cent of the sequence showed over 60 percent homology with Protein kinase and protein serine theonin kinase from *Lophopyran elongation* and *Ricinus communis*. Blastx results indicated homology with the same protein with 35.4 to 43.5 bit score.

The sequence was found to have four ORFs with the longest one being 264 bp in length. The CDD and Interproscan analysis indicated negative results.

Table 10 Results of sequence data analysis for the EST-8 [PNK 21]

1. **Initial Sequence length** : 778
2. **Vector sequence** : 1 to 57
3. **Sequence length after Vector/Adaptor Screening** : 721
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XP002530935.1	Protein kinase APK1B (*Ricinus communis*)	71%	4e-59	67%
AAF43496.1	Protein serine/threonin kinase (*lophopyran elongation*)	81%	3e-56	57%

5. **Blastx Results:**

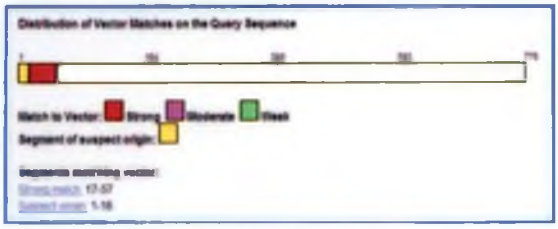
Accession No.	Description	Evalue	Score (Bits)
AAF43496.1	Protein serine/threonin kinase (*lophopyran elongation*)	0.007	43.5
XP002530935.1	Protein kinase APK1B (*Ricinus communis*)	2.2	35.4

5. **ORFs Available:**

ORF Location	ORF length	Frame
386-649	264	+2

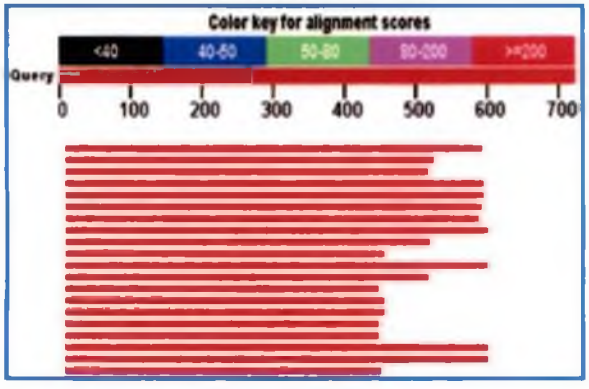
6. **Functional domains available :**

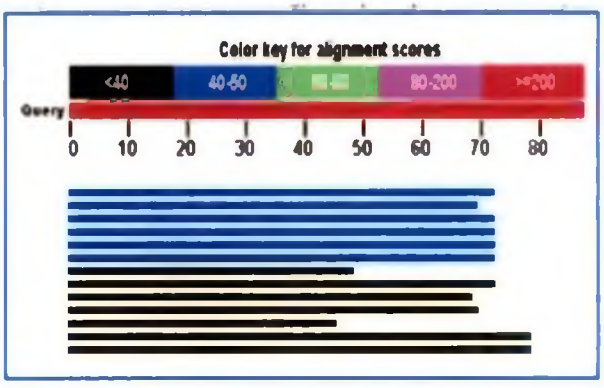
Tools	Accession no.	HMM	Domains details
IFS	-	-	nil
CDD	-	-	nil
Pfam	-	nil	nil

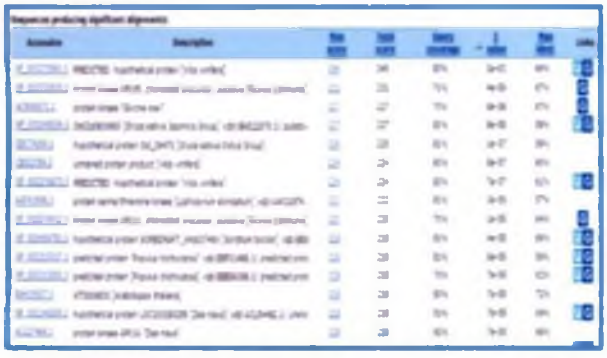


VecScreen output

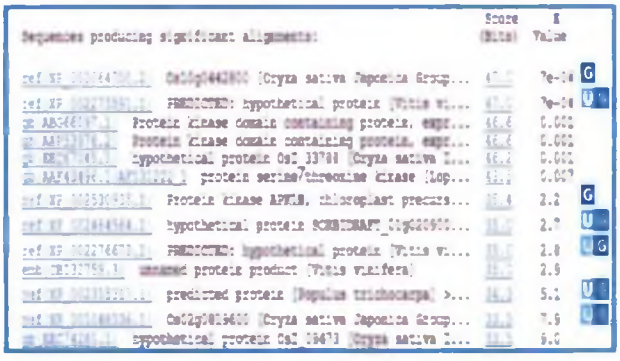
Open reading frames in the cloned sequence







Blastn output



Blastx output

Fig. 8 Results of sequence data analysis for the EST- 8 [PNK 21]

EST- 9

The sequence data obtained for the EST-9 was obtained from the clone PNK 1(2) 48 and it was of 687 bp in size. Vector screening indicated vector sequence from 658 to 687 bp. The results obtained are presented in Table 11 and Fig. 9. After vector and adaptor screening the total sequence obtained for further analysis was 657 bp.

5'TTCGAGCGGCGCCCGGGCAGGTACTGGCCGAGCTTCGCTTGCATCGAAGG
CCCAAACTTCAAGGAGTGGGAGCAGACTTGGATCGACAACGGCGACTGCACC
GGCCTGCCGGAGAGAGACTACTTTAGCAGGGCACTTGAACTTCGCAAGAGGC
CAACGTCCTCGATGTCCTCTCAAGGAATGGAATCAGGCCATCTCACTACAAT
ACATACCGGACGATCGATGTGGTGAACGCTATCAACCAAGGTCTGGGAGTGG
TGAATGCTACGGTTCTGCACTGCGGATCGACCAAGTTGGAGAGGATCATGAA
TGTGAGGTTGTGTGTGGACATCACCGCCACCAAATTCATCACCTGCCCGACG
CCGCGGAACTCGACCTGTCCGGCGTCTGTCAAGCTAGATGTGCTGCCATCTA
GCACCGACACTTTGGAGGTAGAGGACAACCCCATCGCAATGGTGGTTGCCCA
GTGATCGATGCTAGCCATCAAATGGAAAGCTTTAATATTATGTCACAAGTGT
GATCATGGATATGACCATACTTGGTGCGTTAGCTAATAATGTTAGTAAGGTG
GAATTAGAATAAGAAGAATAAATTTCCTCTTTATTTGTTTCATGATATGAAG
CAAATAATGGTTTTCAAAAAAAAAA3'

In Blastn analysis 98 per cent of the sequence showed 81 per cent homology with hypothetical protein from *Ricinus communis* and *Vitis vinifera*. Blastx results indicated homology with unknown protein from *Glycine max* (178 bit score) and the NAD-dependent epimerase/dehydratase in *Zea mays* (bit score 167). The sequence also showed similarity in Blastp with binding / catalytic/ coenzyme binding in *Arabidopsis thaliana* (bit score 213) and NAD-dependent epimerase/dehydratase in *Zea mays* (bit score 205).

The sequence was found to have four ORFs with the longest one being 339 bp in length. Interproscan results indicated the presence of active site for the NAD Binding Rossamann Super family and also conserved domain for the same protein was observed.

Table 11 Results of sequence data analysis for the EST-9 [PNK 1(2)48]

1. **Initial Sequence length** : 687
2. **Vector sequence** : 658 to 687
3. **Sequence length after Vector/Adaptor Screening** : 657

4. **Blastn results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM00253094.1	Hypothetical protein (*Ricinus communis*)	98%	3e-124	81%
AAF43496.1	Hypothetical protein (vitis vinifera)	98%	3e-124	81%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
ACU21370.1	unknown *[Glycine max]*	2e-43	178
NP_001148959.1	NAD-dependent epimerase/dehydratase *(Zea mays)*	4e-40	167

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_565868.1	binding / catalytic/ coenzyme binding (*Arabidopsis thaliana*)	6e-54	213
NP_001148959.1	NAD-dependent epimerase/dehydratase *(Zea mays)*	1e-51	205
ABK22551.1|	unknown (*Picea sitchensis*)	3e-44	181

6. **ORFs Available:**

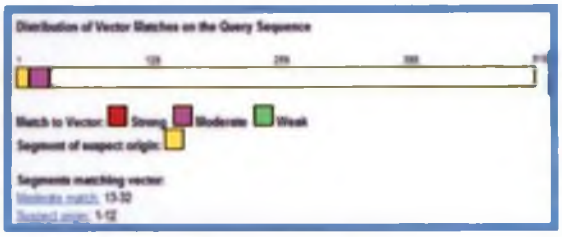
ORF Location	ORF length	Frame
148-485	339	+1

7. **Functional domains available :**

Tools	Accession no.	HMM	Domains details
CDD	CI09506	-	NAD Binding Rossamann Super family
pfam	-	nil	Nil

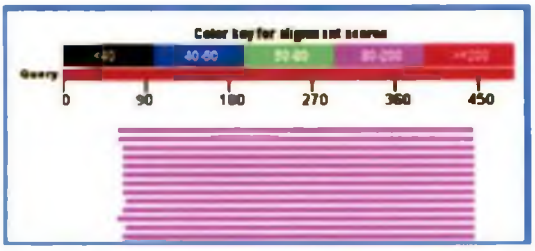
8. **Interproscan scan results:**

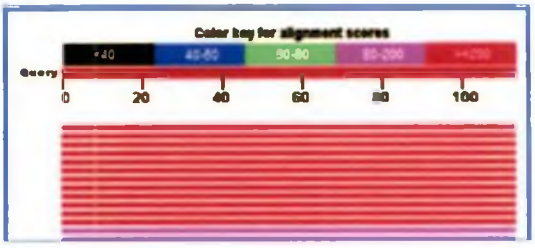
Database entry	Description
SSF51735	NAD Binding Rossamann fold domain

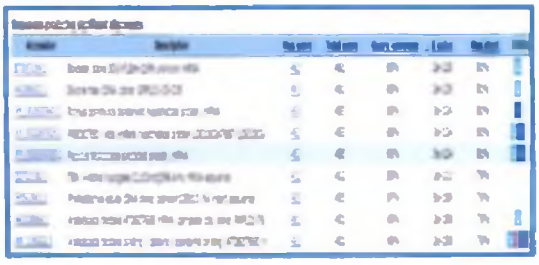


VecScreen output

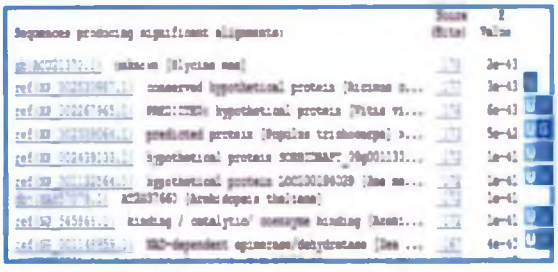
Open reading frames in the cloned sequence





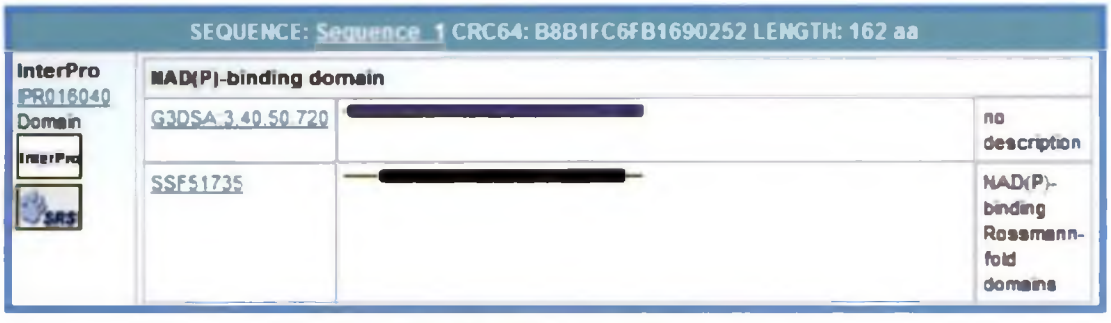


Blastn output



Blastx output

Conserved Domain



InterProScan result

Fig. 9 Results of sequence data analysis for the EST-9 [PNK 1(2) 48]

EST-10

The sequence data obtained for the EST-10 was obtained from the clone PNK 1(2)36 and it was of 608 bp in size. Vector screening indicated vector sequence from 1 to 31 bp. The results obtained through *in silico* anlysis are presented in Table 12 and Fig. 10. After vector and adaptor screening the total sequence obtained for further analysis was 577 bp., as shown below

5' GTACTATCAATCAGGAGCAGAAAAAGATGGGTAACTTTTTCCGAGTGACC
TGCATTTTGAATAATGCAGCTTGCAAACTGAAACTGAATGAATTCCGGGAAG
CTGCGAAGCTCTGTGCAAAGGCTTTGCAATTAGAGCCTGACAACGTGAAGGC
TCTCTTACCGTAGAGCTCAGTCTTACATTGAGACAAGTGATCTGGACTTGGCA
GAAATAGACTTATTGGCTGCAAAGGAAATTGAACCAGACAATGAGCCATTAA
ATCAACTACTGGAGAGAGTGAAAAAACTAAAGACACGAGCCAGTAAAATGGA
TATGAAATTTTACGCAAACATGCTGGCAAAGGTTAACATTGCACACTGTAAA
TAACTCAGAAATATCTTTATACTGCTCATTTTTGTTGAGTTAAAACAGATCT
GATTAGCGTTGGAAAAATAAGCATTGAAGAGATTGTTTGCCTCAGCGGCTCT
GTATCATTCTATAGGCGAGAGCTTTTAATTTCTTTTGGTGATTAGGTGATAA
AAGAAGGGACGAGATAAATAATTTGGGGGGATTGTTGGAATGTTTTCTTGTG
ATATGG3'

Blastn results indicated showed 74 percent homology for 26 percent of the sequence, with the cDNA clone in *Triticum asetivum* and 73 percent homology with the mRNA clone in *Oryza sativa*. Blastx results showed homology with peptidyl prolyl isomerase (bit score 96.7) and FK506-binding protein (bit score 96.3) from *Ricinus communis* and *Arabidopsis thaliana*. The sequence also showed similarity in Blastp with peptidyl prolyl isomerase in *Arabidopsis* (bit score 108) and *Triticum* (bit score 103).

The sequence was found to have three ORFs with the longest one being 339 bp in length. The Interproscan results indicated the presence of active sites for Peptidyl prolyl isomerase and the sequence also have transmenbrane helices.

The sequence possessed conserved domain for TPR super family. Pfam tools indicated domain for the same protein with HMM sequence (6 to 34).

Table 12 Results of sequence data analysis for the EST- 10 [PNK 1(2)36]

1. **Initial Sequence length** : 608
2. **Vector sequence** : 1 to 31
3. **Sequence length after Vector/Adaptor Screening** : 577
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AK332751.1	*Triticum asetivum* cDNA clone	26%	e-14	74%
AF140495.2	*Oryza sativa* mRNA clone	25%	2e-13	73%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
XP_002534361.1	Peptidyl prolyl isomerase *(Ricinus communis)*	1e-18	96.7
BAB02082.1	Peptidyl prolyl isomerase; FK506-binding protein (*Arabidopsis thaliana)*	2e-18	96.3
AAB82061.1	rof1(*Arabidopsis thaliana)*	5e-18	94.7

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
BAB02082.1	peptidylprolyl isomerase; FK506-binding protein (*Arabidopsis thaliana)*	2e-22	108
NP_001118695.1|	ROF1 (ROTAMASE FKBP 1); FK506 binding / calmodulin binding / peptidyl-prolyl cis-trans isomerase (*Arabidopsis thaliana)*	6e-22	107
CAA68913.1	Peptidyl prolyl isomerase (*Triticum aestivum)*	7e-21	103

7. **ORFs Available:**

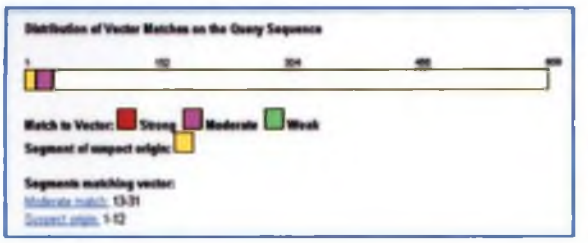
ORF Location	ORF length	Frame
28-366	339	+1

8. **Functional domains available :**

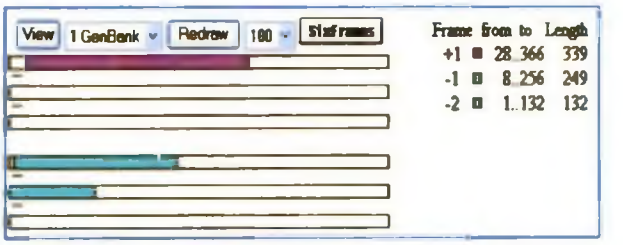
Tools	Accession no.	HMM	Domains details
CDD	c/02429	-	TPR superfamily
Pfam	-	6-34	TPR 1

9. **Interproscan scan results:**

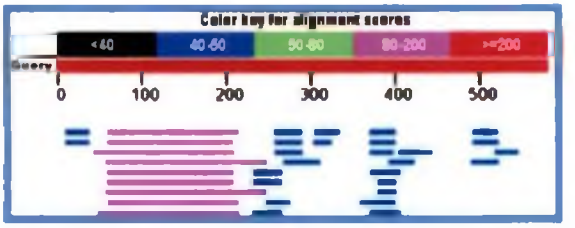
Database entry	Description
PTHR10516	FK506 Binding protein
PF00515	TPR 1
PTHR10516SF8	FK506 Binding protein(PASTICCINO)
tmhmm	Transmembrane helices

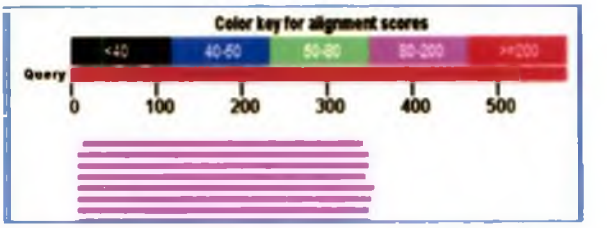


VecScreen output

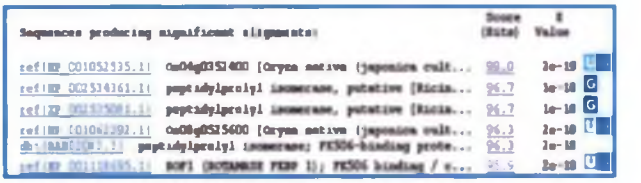


Open reading frames in the cloned sequence



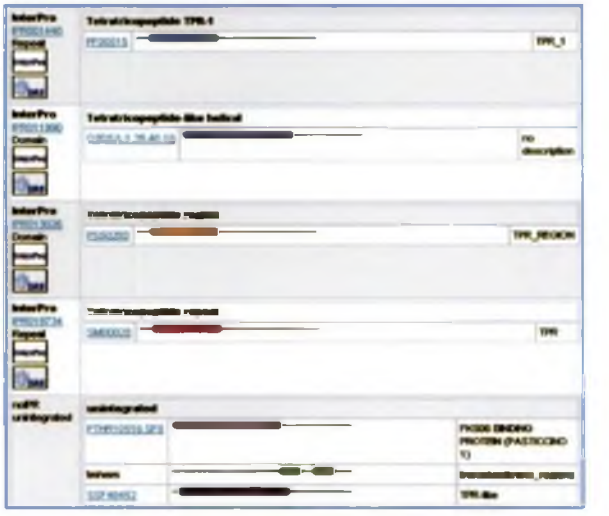


Blastn output



Blastx output

Conserved Domain



InterProScan result

Pfam result

Fig. 10 Results of sequence data analysis for the EST-10 [PNK 1(2)36]

EST-11

The sequence data obtained for the EST-11 was obtained from the clone PNK 1(2)33 and it was of 688 bp in size and vector screening indicated vector sequence from 530 to 688 bp. The results obtained are presented in Table 13 and Fig. 11. After vector and adaptor screening the total sequence obtained for further analysis was 529 bp., as indicated below

5′GCGGCCGCCCGCGGCAGGTTCTGGCCGAGCTTCGCTTGCATCGAAGGCCC
AAACTTCAAGGAGTGGGAGCAGACTTGGATCGACAACGGCGACTGCACCGGC
CTGCCGGAGAGAGACTACTTTAGCAGGGCACTTGAACTTCGCAAGAAGGCCA
ACGTCCTCGATGTCCTCTCAAGGAATGGAATCAGGCCATCTCACTACAATAC
ATACCGGACGATCGATGTGGTGAACGCTATCAACCAAGGTCTGGGAGTGGTG
AATGCTACGGTTCTGCACTGCGGATCGACCAAGTTGGAGAGGATCAAGAATG
TGAGGTTCTGTGTGGACAGCACCGCTACCAACTTCATCACCTGCCCGAGGCA
GCAGGCCTCTGTGTGTCCTGACAGAGTCAAGCTGGATGTGCTGCCATACGAC
GCTGCCTTGCTAAAGAACCCCATCGAGATGGTGGTTGCCCAGTGATGTTGCT
AAAGAGAAAAATGGTTTCAGTAGTATAGATGAGTCGTAAGAGGTACCTCGGC
CGCGACCACGC3′

Blastn results indicated 88 per cent homology for 6 percent of the sequence, with the mRNA b-like cysteine protein in *Vigna unguiculata.* In Blastx and Blastp analysis, the sequence showed homology to ribonuclease 1, ribonuclease and drought induced S- like ribonuclease in different plants like *Zea mays, Oryza sativa* and *Hordeum valgare* with 81.6 to 92.0 bit score.

The sequence was found to have three ORFs with the longest one being 240 bp in length. The Interproscan results indicated the presence of active sites for ribonuclease 1 with the signal peptide. The sequence has conserved domain for the same protein. Pfam tools indicated HMM sequence (72-189) for the same protein family.

EST- 12

The sequence data obtained for the EST-12 was obtained from the clone PNK 1(2) 41 and it was of 816 bp in size. Vector screening indicated vector sequence from 1-62 bp. The results obtained are presented in Table 14 and

Table 13 Results of sequence data analysis for the EST-11 [PNK 1(2)33]

1. **Initial Sequence length** : 688
2. **Vector sequence** : 530 to 688
3. **Sequence length after Vector/Adaptor Screening** : 529
4. **Blastn results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AM748427.1	*Vigna unguiculata* partial mRNA b-like cysteine protein	6%	0.049	88%
EU923562.1	Uncultured prokaryotes clone	9%	0.17	100%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_001151299.1	Ribonuclease 1 (*Zea mays*)	8e-17	90.1
AAL33377.1	drought induced S- like ribonuclease (*Oryza sativa*)	4e-15	82.0
AAMB0567.1	Rnase-S- like protein (*Hordeum valgare*)	2e-14	80.5

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_001151299.1	Ribonuclease 1 (*Zea mays*)	2e-17	92
AAL33776.1	drought-induced S-like ribonuclease (*Oryza sativa*)	7e-15	83.6
NP_563941.1	ribonuclease T2 family protein (*Arabidopsis thaliana*)	3e-14	81.6

7. **ORFs Available:**

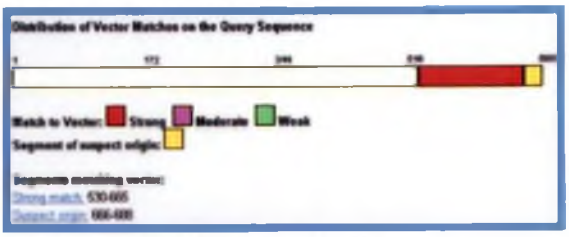
ORF Location	ORF length	Frame
198-437	240	-3

8. **Functional domains available:**

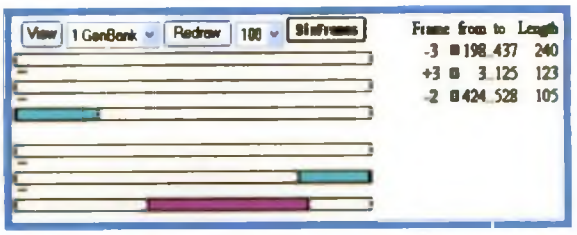
Tools	Accession no.	HMM	Domains details
CDD	C100208	-	Rnase 2 super family
Pfam	-	72-189	Ribonuclease T 2 family

9. **Interproscan scan results:**

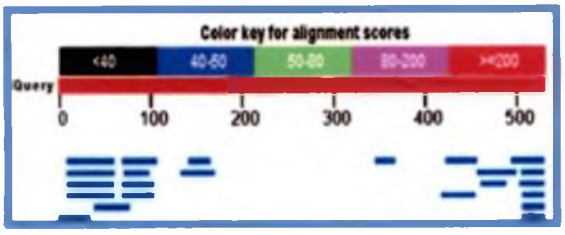
Database entry	Description
PTHR11240, PF00445	Ribonuclease T2
SSF55895	Ribonuclease –Rh like
Signalip	Signal peptide

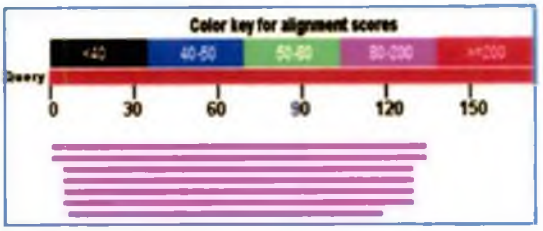


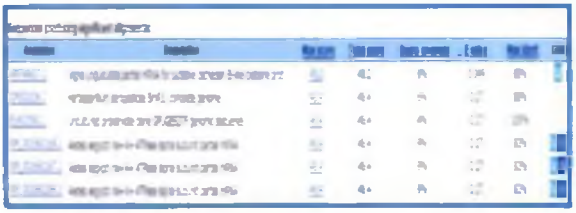
VecScreen output



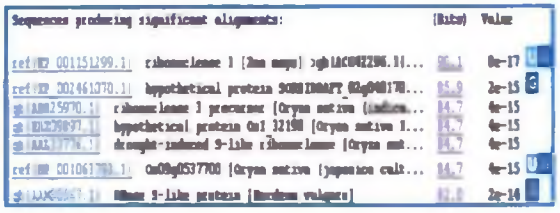
Open reading frames in the cloned sequence





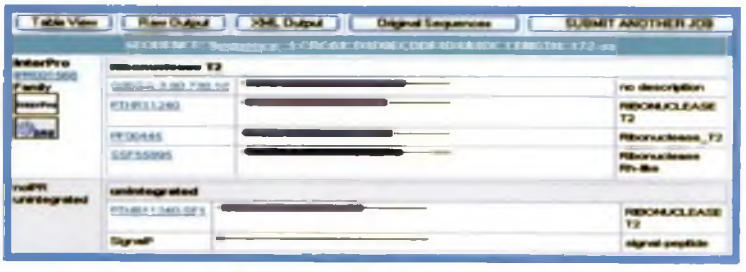


Blastn output



Blastx output

Conserved Domain



Pfam result

Fig. 11 Results of sequence data analysis for the EST-11 [PNK 2(1)33]

Table 14 Results of sequence data analysis for the EST-12 [PNK 1(2)41]

1. **Initial Sequence length** : 816
2. **Vector sequence** : 1-62
3. **Sequence length after Vector/Adaptor Screening** : 754

4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AJ277077.2	Rattus norvgicus mRNA damage specific DNA binding protein	5%	5e-04	88%

5. **Blastp Results:**

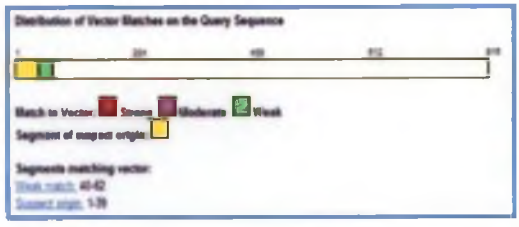
Accession No.	Description	Evalue	Score (Bits)
XP00305886.1	Protein kinase (Trichomonas vaginalis)	6.2	35.0

6. **ORFs Available:**

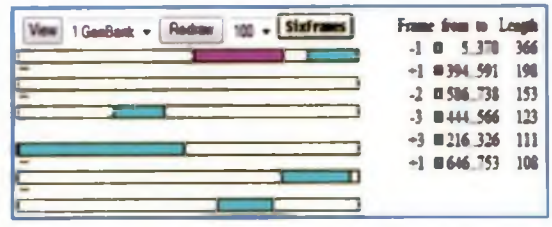
ORF Location	ORF length	Frame
5-370	366	-1

7. **Functional domains available :**

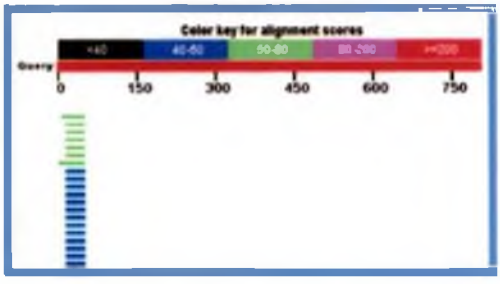
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
Pfam	-	Nil	Nil

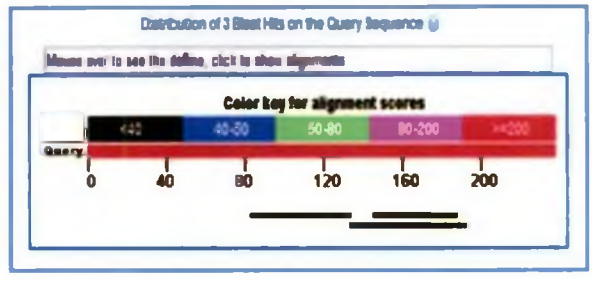


VecScreen output

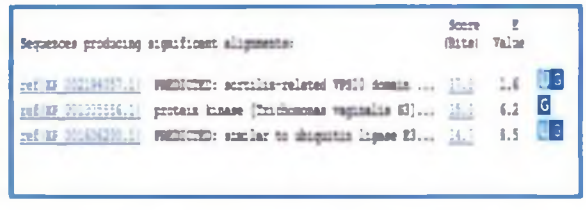


Open reading frames in the cloned sequence





Blastn output



Blastx output

Fig. 12 Results of sequence data analysis for the EST-12 [PNK 1(2)41]

Fig. 12. After vector and adaptor screening the total sequence obtained for further analysis was 754 bp., as shown below

5' AGCGCTAAATAGCACTGAGGGCTAGGGGCCGGGGCGGGAGGCATAGCGAGAG
TCGGTGCCCAAATTCNGCGCCTATAGTCGAGTATGTGATTAGAGGATAGNTACA
CCGGACGAAAGCGANCGCCAAAGTGGTGGCTCCTATTGACAAGAGCAGAGTGGC
ATGTGTCCACACGACGTGCTTTTTGGTTTTTGTTAACTTATTTCCACACTCACC
GATGCAGGCACATGTACGAACAAACTGGCAACTCCTCTTCATCTCCAAGGTGGG
TGTTCGAACTTCTGACACTGGCAGATCAACTCCTGTAGGTTTACCAGTGTTTGC
NTGAAGGCTTAGATGCCCGACNGCACAGCCAGCTTCTTGATCATCCATTTCTGG
ACGCCACCATCATTANCATGCATCAACTATTAATGTGTTATTCAATTTACTCGA
TTTTANTGCGCAGTTACTCTGCCCGGGCCCGCGCCCGCTCGCGCGCCCCAAGGG
GGCTTACACGGATGTCCACAGCGCACAAAGTCAAAAAAACGATAACTTCACCCN
AATTGGGGAAAGAGGAAAAGGACATCATAAAAAATGGTTTGTTTCAATTATAAC
AAATCGCGTTTCTTAANGGTGAAAAAGNATTTCGGAATCCACTTTAGAGCGCTA
TGAAATTTAGAGAATTGCCAAGGAAANTAAATTTTAGTNTTTTTCGAAATACAC
CCAACCCTACTGGTTTTTTTCCCCGATTTTGGCCACATTATACCGCTATATTAG
3'

No significant similarity/homology was detected for this sequence in Blastn, Blastx and Blastp analysis. However the sequence was found to have six ORFs with the longest one being 366 bp in length. Interproscan and CDD analysis also indicated negative results.

EST- 13

The sequence data obtained for the EST-13 was obtained from the clone PNK 1(2)56 and it was of 731bp in size and vector screening indicated vector sequence from 620 to 672 bp. The details of results obtained are presented in Table 15 and Fig. 13. After vector and adaptor screening the total sequence obtained for further analysis was 619 bp.

5' CGTGGTCGCGGCGCAGGTCATTTTGTTGTAAGTGAAGGACGTAGACCTTG
ACAACTCAACTGACTGATACACGTTGAGGTTGAAAATTTCACCTCTTCAAAA
TGGAACCTTTCTTCCGCAAAATATTGAGGACAAATAATGTCAACGAAGCCCA
AAATAACAATCGAGCTTTTGATGCCATAAATTGACCAAGTTTAGTCTTCTTC
AGAAGAAATTGAGTGGACTGAAGGCCCTGACAGGCTTTTCCATGCAACGGAT
GTTTTGTTGGCAACAGAAGATACAGTATTGTCCTCATGCTCAGGTAACCCAG
AAGCTGGTTTGGATGCTTTTGAAGCTATCAATGCGCGTAATTTATCCAATAG

Table 15 Results of sequence data analysis for the EST-13 [PNK 1(2)56]

1. Initial Sequence length : 731
2. Vector sequence : 620 to 670
3. Sequence length after Vector/Adaptor Screening : 619
4. Blastn Results:

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AY972077.1	Synthetic construct RLS gene	5%	1e-04	100%
AM 748496.1	Vigna unguiculata partial mRNA	5%	0.005	93%

5. Blastx Results:

Accession No.	Description	Evalue	Score (Bits)
EDZ39948.1	Conserved hypothetical protein (*Leptospirillum sp.*)	6.7	38

6. Blastp Results:

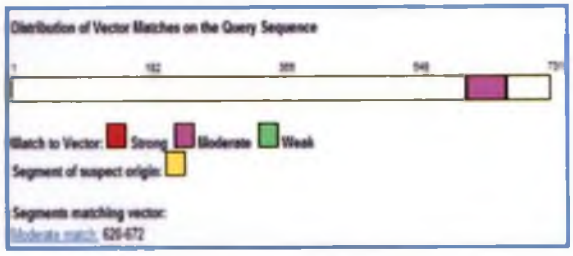
Accession No.	Description	Evalue	Score (Bits)
YP_001740508.1	Conserved proline rich protein (*Mycobacterium tuberculosis*)	9.4	33.1

7. ORFs Available:

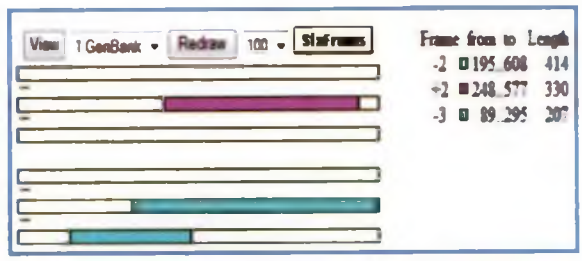
ORF Location	ORF length	Frame
248-577	414	+2

8. Functional domains available :

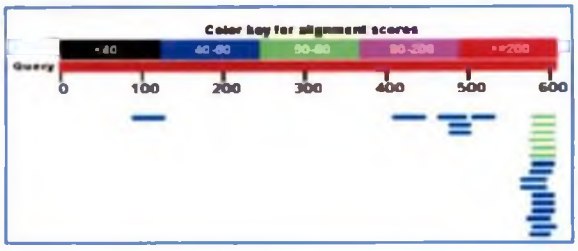
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
Pfam	-	Nil	Nil

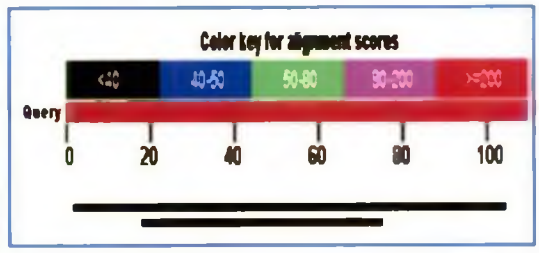


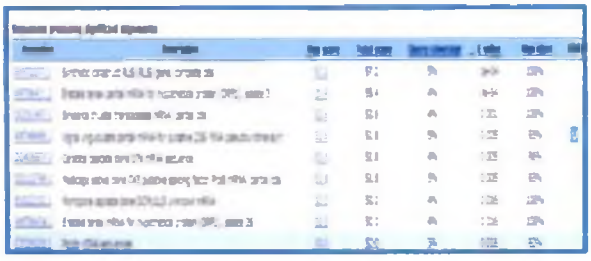
VecScreen output



Open reading frames in the cloned sequence







Blastn output

Blastx output

Fig. 13 Results of sequence data analysis for the EST-13 [PNK 1(2)56]

TTCCTGACGTCTTGCATCTGGAGTATGGGCTGCCTCTATCCAAGAAGCAGCA
TGATCAGTATTTCCATACACATTTCCCAAAAGCAATACTGAATATGCCTCTG
CAACTTCCAAATATTCATCAATGCTTAAAGGAAAAGGTGAAGAACATCCATT
TAATTTATCCACAACATGATTATACGACAGTTTTATCAGGTGAAGGTTTCCA
ACTGTAAAATACCTGCCCGGGCGGCCGCTCGAAATCACT3'

In Blastn analysis the sequence had 100 percent homology with Synthetic construct RLS gene. Blastx results indicated homology with the Conserved hypothetical protein in *Leptospirillum sp.* with 38 bit score.

The sequence was found to have three ORFs with the longest one being 414 bp in length. Interproscan and CDD analysis indicated negative results

EST- 14

The sequence data obtained for EST-14 was obtained from the clone PNK 1(2)39 and it was of 518 bp in size. The vector screening indicated vector sequence from 1 to 32 bp. The results obtained are presented in Table 16 and Fig. 14. After vector and adaptor screening the total sequence obtained for further analysis was 486 bp.

5'CCTAAAGATGAACCTGGGTTTGATCCAAGCAAAGGAGAGAGACCTGAATT
CTATTTTGAAGATGGTGCCTATCCTGAGCAGGTTGACTGGATTGGTCAAAAA
AACCAAATAGATGCTGCCAAAGCTGTAGAAGTTAAGCAGATTGTGTTGGTTG
GGTCAATGGGTGGAACAGACCTCAATCATCCTTTGAACAGCTTAGGAAATGG
GAATATATTGATTTGGAAAAGAAAAGCTGAGCAGTATTTGGCTGATTCTGAC
ATACCATACACAATAATAAGGGCTGGAGGATTACAAGACAAGGATGGAGGCA
TACGTGAGTTGGTAGTTGGGAAAGATGATGAGCTTCTGAAAACAGAAACAAG
AACAATTGCTAGACCGGACGTTGCCGAGATCTGCATTCAGGCTTTGCAATTT
GAGGAGGCTAAGTTCAAAGCATTTGACCTAGCATCAAAACCAGAAGGCACTG
GGACACCAACAAAGAACTTC3'

In Blastn analysis, 98 percent of the sequence showed over 80 percent homology with unknown mRNA from soybean and the binding/catalytic/coenzyme binding in *Arabidopsis thaliana.*

Table 16 Results of sequence data analysis for the EST-14 [PNK 1(2)39]

1. **Initial Sequence length** : 518
2. **Vector sequence** : 1 to 32
3. **Sequence length after Vector/Adaptor Screening** : 486
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
BT097194.1	Soybean clone (unknown RNA)	98%	2e-126	81%
NM12322.3	binding / catalytic/ coenzyme binding protein *(Arabidopsis thaliana)*	98%	2e-108	78%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_001148959.1	NAD-dependent epimerase/dehydratase(*Zea mays*)	4e-40	167
NP_568098.1	binding / catalytic/ coenzyme binding *(Arabidopsis thaliana)*	1e-36	155

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_565868.1|	binding / catalytic/ coenzyme binding (*Arabidopsis thaliana)*	6e-54	213
ACD03605.1	NAD-dependent epimerase/dehydratase (*Zea mays)*	1e-51	205
ZP_03156369.1	3-beta hydroxysteroid dehydrogenase/isomerase (*Cyanothece* sp.)	5e-30	133

7. **ORFs Available:**

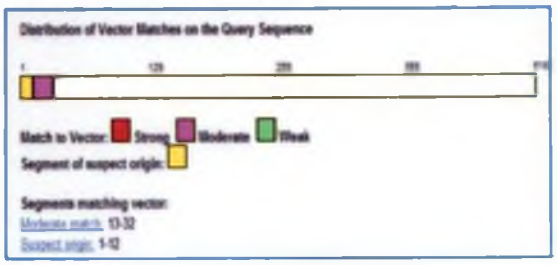
ORF Location	ORF length	Frame
148-485	339	+1

8. **Functional domains available :**

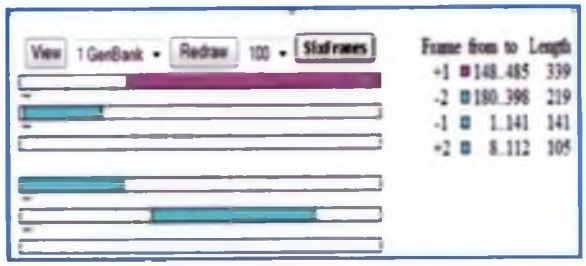
Tools	Accession no.	HMM	Domains details
CDD	CI00194	-	Ycf 39
pfam	-	nil	nil

9. **Interproscan scan results:**

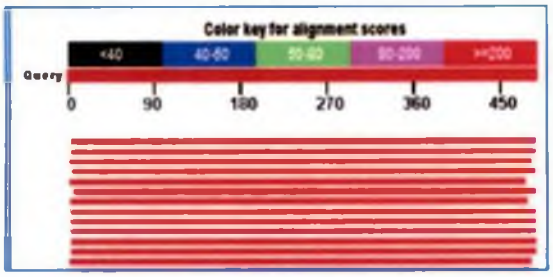
Database entry	Description
SFS51735	NAD(P)-binding Rossmann fold domains

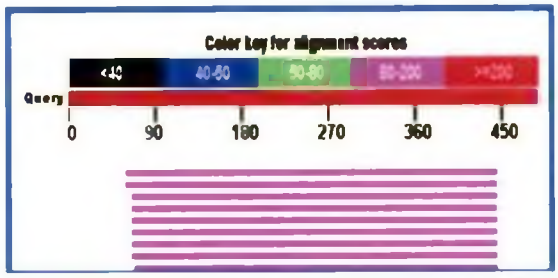


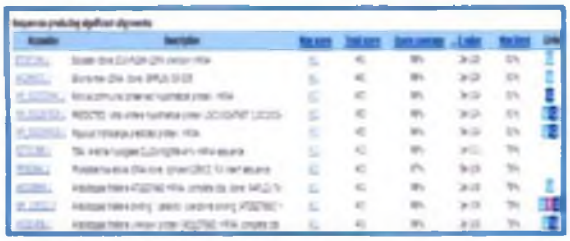
VecScreen output



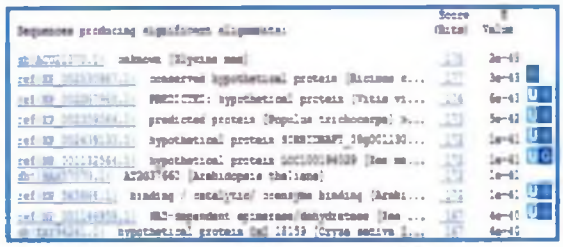
Open reading frames in the cloned sequence





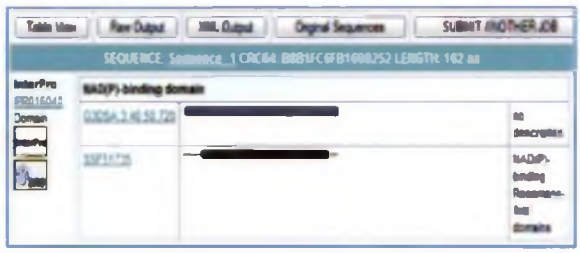


Blastn output



Blastx output

Conserved Domain



InterProScan result

Fig. 14 Results of sequence data analysis for the EST-14 [PNK 1(2)39]

Blastx and Blastp results, indicated similarity with the NAD-dependent epimerase/dehydratase and binding/catalytic/coenzyme binding in *Arabidopsis thaliana* and *Zea mays* with bit score 155 to 205.

The sequence was found to have four ORFs with the longest one being 339 bp in length. The Interproscan results indicated the presence of active sites for the NAD (P) binding rossmann fold domain. The sequence had the Ycf 39 conserved domain.

EST- 15

The sequence data obtained for the EST-15 was obtained from the clone PNK 1(2)57 and it was of 726 bp in size. Vector screening indicated vector sequence from 588 to 726 bp. The results obtained are presented in Table 17 and Fig. 15. After vector and adaptor screening the total sequence obtained for further analysis was 588 bp.

5' GAGCGGCGCCGCGCAGGATCGTTTTGCTTTTGTCGGTAATAGACGTTGAA
TTAGGTGGGCCAATCGAAACAGAGATGACAACATTATTGATCATGTCCACAC
GCTCAGATACAATACGCTGTCTAGCATCAGGAGACAATGGTGCTGCAGAAGA
ATAACGTTCAGGTTTTCTGGATTCAACCCACTTAAAGGCAAACTTTTTTGAA
GGAAGTTCGACAGGATAGAGAAATAAATAAGCATTAATTCCGTCTACTAATG
GAACCATTTTGATGCCATCACTCTGCTCAACCTCGGCAATTGAAGCAGGCGG
ATGCAAACTGGGAGAAGCCGGAAACGAGAGACTGAGCAGCACAAGCTCCCTC
CTTCTGAACGGCACCGACAGGCCTTTACACCCACAAAGAACGACACTTTTCG
ACACCGGAAACCTACTCCGCCGGATAGAAGAGCTTGGGCAGCCATTCCAATG
GAAGCGGGGAGAGGGATGAGGGAAAATGGAGGAGGGGAGCGATGGAGAGGGT
GAAGAGAGAAGGCAAGTTGCCGCCATGAATGATCTCCGGTACCTCGGCCGCG
ACCACGC3'

Blastn results indicated that about 37 percent of the sequence showed 77 percent homology with cDNA clone in *Oryza sativa* and *Zea mays*. Blastx and Blastp results showed homology of the sequence with the hypothetical protein from *Oryza sativa* and oxygen-evolving complex-related protein from *Arabidopsis thaliana* with 133 to 144 bit score.

Table 17 Results of sequence data analysis for the EST-15 [PNK 1(2)57]

1. **Initial Sequence length** : 726
2. **Vector sequence** : 588 to 726
3. **Sequence length after Vector/Adaptor Screening** : 577
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
NM001050121.1	*Oryza sativa* cDNA clone	38%	2e-40	77%
BT060570.1	Zea mays full length cDNA clone	37%	2e-38	77%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
EEC71094.1	hypothetical protein Osl_02877 *(Oryza sativa* Indica Group*)*	3e-32	141
NP_196706.2	oxygen-evolving complex-related protein (*Arabidopsis thaliana)*	1e-25	119

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_00113150.1	Hypothetical protein (*Zea mays*)	5e-33	144
NP_196706.2	oxygen-evolving complex-related protein(*Arabidopsis thaliana)*	8e-30	133

7. **ORFs Available:**

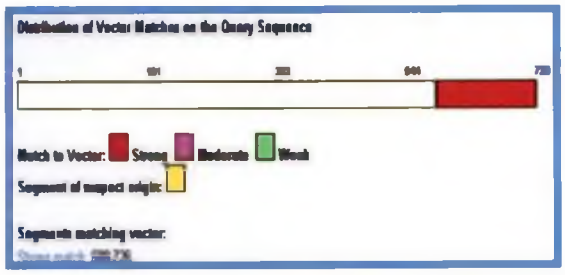
ORF Location	ORF length	Frame
270-485	543	+2

8. **Functional domains available :**

Tools	Accession no.	HMM	Domains details
CDD	-	-	nil
pfam	-	nil	nil

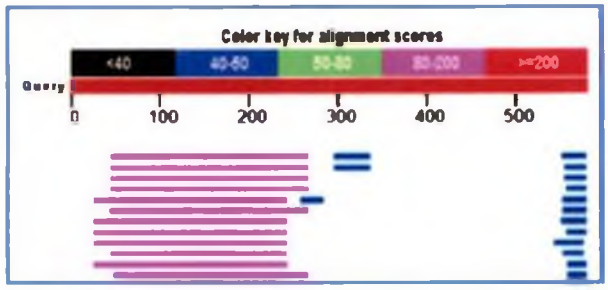
9. **Interproscan scan results:**

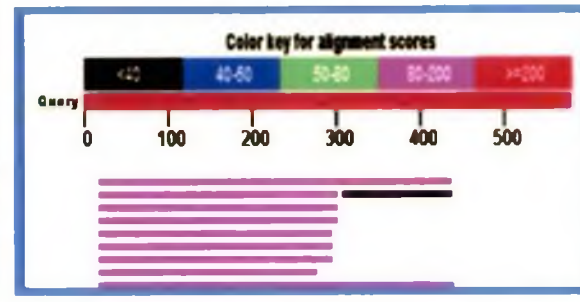
Database entry	Description
signalip	Signal peptide

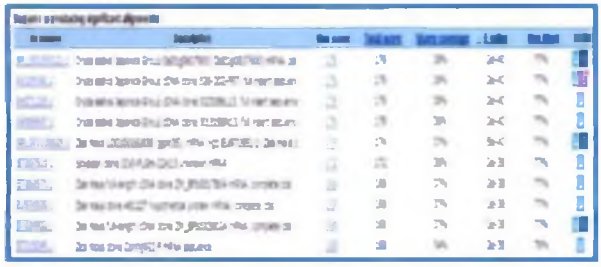


VecScreen output

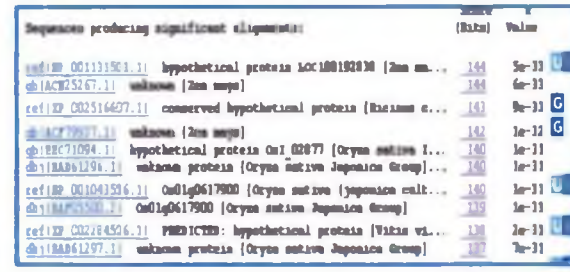
Open reading frames in the cloned sequence



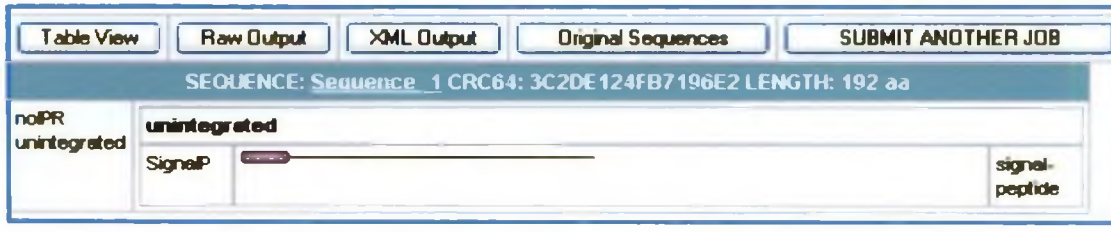




Blastn output



Blastx output



InterProScan result

Fig. 15 Results of sequence data analysis for the EST-15 [PNK 1(2)57]

The sequence was found to have six ORFs with the longest one being 543 bp in length. The Interproscan results indicated the presence of signal peptide. The sequence had no conserved domain.

EST-16

The sequence data obtained for the EST-16 was obtained from the clone PNK 1(2)47 and it was of 688 bp in size. Vector screening indicated vector sequence from 1–36 and 530 - 688 bp. The results obtained are presented in Table 18 and Fig. 16. After vector and adaptor screening the total sequence obtained for further analysis was 626 bp.

5'ACTGGCCGAGCTTCGCTTGCATCGAAGGCCCAAACTTCAAGGAGTGGGAG
CAGACTTGGATCGACAACGGCGACTGCACCGGCCTGCCGGAGAGAGACTACT
TTAGCAGGGCACTTGAACTTCGCAAGAGGCCAACGTCCTCGATGTCCTCTCA
AGGAATGGAATCAGGCCATCTCACTACAATACATACCGGACGATCGATGTGG
TGAACGCTATCAACCAAGGTCTGGGAGTGGTGAATGCTACGGTTCTGCACTG
CGGATCGACCAAGTTGGAGAGGATCATGAATGTGAGGTTGTGTGTGGACATC
ACCGCCACCAAATTCATCACCTGCCCGACGCCGCGGAACTCGACCTGTCCGG
CGTCTGTCAAGCTAGATGTGCTGCCATCTAGCACCGACACTTTGGAGGTAGA
GGACAACCCCATCGCAATGGTGGTTGCCCAGTGATCGATGCTAGCCATCAAA
TGGAAAGCTTTAATATTATGTCACAAGTGTGATCATGGATATGACCATACTT
GGTGCGTTAGCTAATAATGTTAGTAAGGTGGAATTAGAATAAGAAGAATAAA
TTTCCTCTTTATTTGTTTCATGATATGAAGCAAATTATGGTTTTCAAAAAAA
AA**3'**

In Blastn analysis, 10 per cent of the sequence showed 78 percent homology with mRNA clone from *Vits vinifera*. Blastx results indicated similarity with the ribonuclease-1 from *Zea mays* (bit score 47.4) and RNA binding /endoribonuclease/ribonuclease T2 from *Arabidopsis thaliana* with 50.1 bit score and S-like orphan receptor 2 from *Prunus dulcis* with 40.1 bit score. Blastp results indicated similarity with Hypothetical protein from *Verrucomicrobium spinosum* and Receptor tyrosine kinase-like orphan receptor from *Sus scrofa* with 33.5 to 35 bit score.

The sequence was found to have five ORFs with the longest one being 393 bp. Interproscan and CDD analysis indicated negative results.

Table 18 Results of sequence data analysis for the EST-16 [PNK 1(2) 47]

1. **Initial Sequence length** : 688
2. **Vector sequence** : 1 to 36, 662 to 688
3. **Sequence length after Vector/Adaptor Screening** : 626
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AM436145.2	Vitis vinifera mRNA clone	10%	0.005	78%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
NP_001151299.1	Ribonuclease 1(*Zea mays*)	2e-10	47.4
NP_564264.1	RNS3 (RIBONUCLEASE 3); RNA binding / endoribonuclease/ ribonuclease T2 *(Arabidopsis thaliana)*	3e-09	50.1
AAG09465.1	S-like ribonuclease *(Prunus dulcis)*	7e-07	40.1

6. **Blastp Results:**

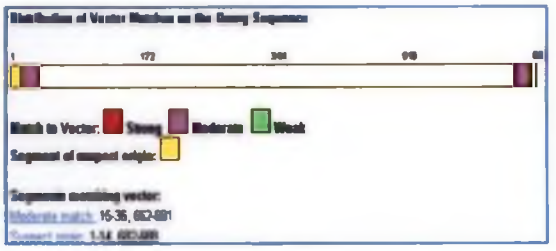
Accession No.	Description	Evalue	Score (Bits)
ZP_02930947.1	hypothetical protein VspiD_29930 (*Verrucomicrobium spinosum* DSM 4136)	2.6	35
XP_00192629.1	PREDICTED: similar to Receptor tyrosine kinase-like orphan receptor 2 (*Sus scrofa)*	2.9	33.5

7. **ORFs Available:**

ORF Location	ORF length	Frame
1-393	393	+1

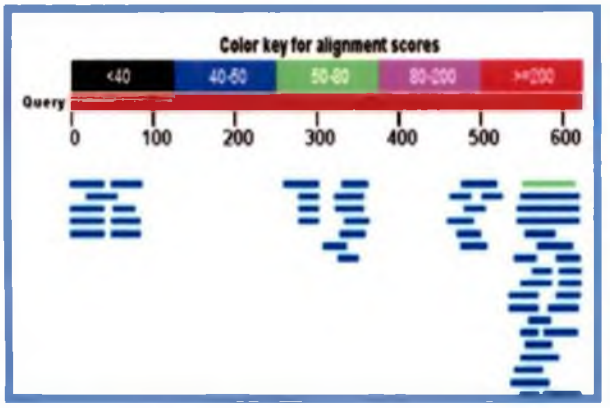
8. **Functional domains available :**

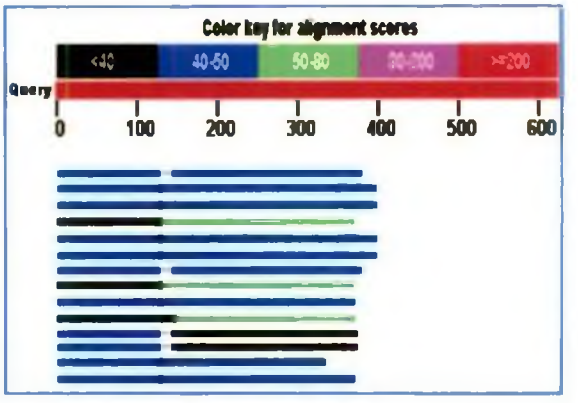
Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	-	-	nil
pfam	-	nil	nil



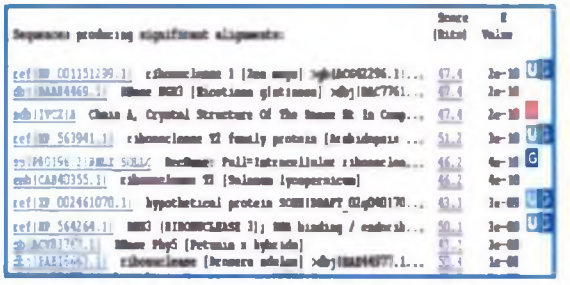
VecScreen output

Open reading frames in the cloned sequence





Blastn output



Blastx output

Fig. 16 Results of sequence data analysis for the EST-16 [PNK 1(2)47]

EST- 17

The sequence data obtained for the EST-17 was obtained from the clone PNK 1(2)54 and it was of 556 bp in size. Vector screening indicated vector sequence from 1-67 and 511-556 bp. The results obtained are presented in Table 19 and Fig. 17. After vector and adaptor screening the total sequence obtained for further analysis was 444 bp.

5' TCGAGCGGCCGCCCGGGCAGGTACATAAACCCATTGCAAAGCATCCTGAACC
TGAAACTGCAATCTCCAAAATGGAAGTTAACTGCCTTCATTTGATCTTTAAGCA
GCCTGGATTCACATTTTATGGCCAATGCACGCCACCCATAACTCAAGTCATAGC
TCAGCTATTACCAATAAGTTTGCCGCATAAAGCAGTGCAGTTGGCAATTTTTTT
CTTGTAGAGGATGCTTTTTCTTAGCTTCGGAACAAAGTTTTAACAAAGAGCAAC
TACCTTTTAGCGATCATATGACCCACCTGACGGATTGCCAGAATTTTGATCACA
ATGACCATTTCCGCTCCTCCACGCATTTCCACCATCTGAGCTGCTACAAGCCTG
ATTGGGAAGGATATTGTTCCCAAGATCTCCATCCAAGCTCATTTGTTGTACCTC
GGCCGCGACCACGC3'

Blastn results indicated that 20 per cent of the sequence showed 83 percent homology with auxin responsive protein from *Zea mays* and hypothetical protein from *Sorghum bicolor*. Blastx result showed similarity with hypothetical protein from *Sorghum bicolour* (bit score 86.0) and the Auxin response factor from *Oryza stiva (*bit score 78).

The sequence was found to have five ORFs with the longest one being 164 bp in length. Interproscan and CDD analysis indicated negative results.

EST-18

The sequence data obtained for the EST-18 was obtained from the clone PNK 1(2)44 and it was of 661bp in size. Vector screening indicated vector sequence from 462 to 661 bp. The results obtained are presented in Table 20 and Fig. 18. After vector and adaptor screening the total sequence obtained for further analysis was 461bp.

5' TTCAGGCGGCGCCGGGCAGGTACCATTGGCGACAACCTTAGTGCTTCTAA
TGCCTCTGGAGGTGTGCATGATCACGTGTATAATTTACAGTTACTCGATGCT
GCTTACTGCAAACTGCCTCAACCTAAAGACTCGGAACGTCCAAAAAGCTATA

Table 19 Results of sequence data analysis for the EST-17 [PNK 1(2)54]

1. **Initial Sequence length** : 556
2. **Vector sequence** : 1 to 67, 511 to 556
3. **Sequence length after Vector/Adaptor Screening** : 444
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
HM004535.1	auxin responsive protein (*Zea mays*)	20%	9e-17	83%
XM002453239.1	hypothetical protein (*Sorghum bicolor*)	20%	9e-17	83%

5. **Blastx Results:**

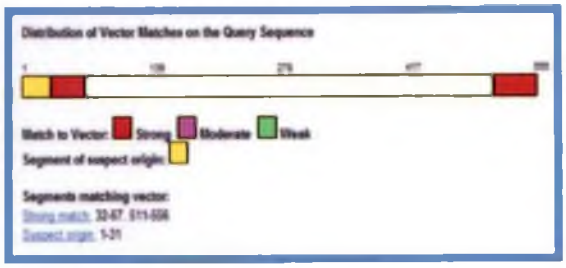
Accession No.	Description	Evalue	Score (bits)
XP_002437545.1	hypothetical protein SORBIDRAFT *(Sorghum bicolor)*	1e-17	86
ABR25348.1	auxin responsive protein [*Oryza sativa*]	1e-74	78

6. **ORFs Available:**

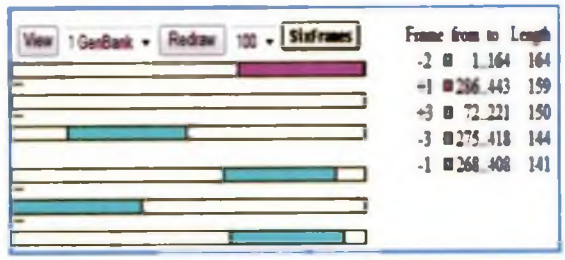
ORF Location	ORF length	Frame
1-164	164	+2

7. **Functional domains available :**

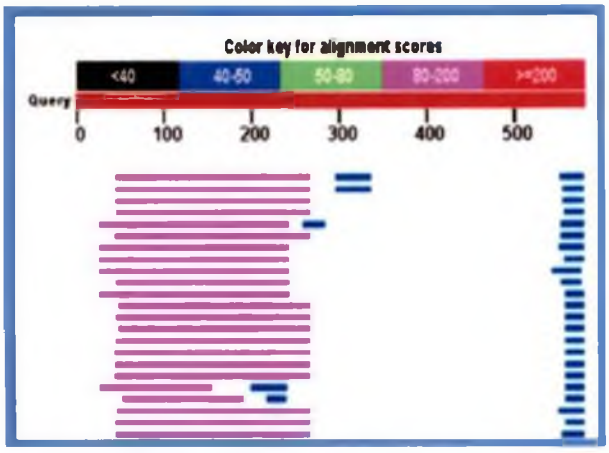
Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	-	-	Nil
pfam	-	nil	Nil

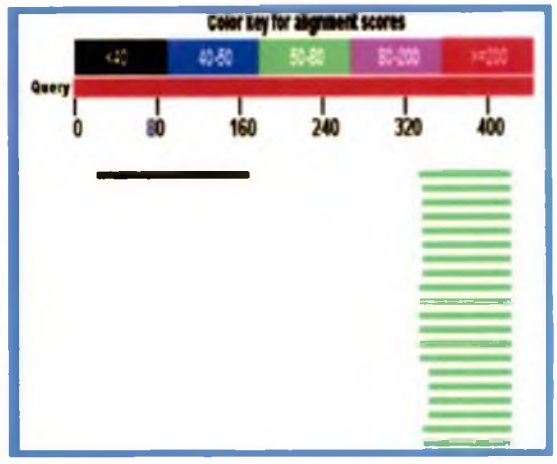


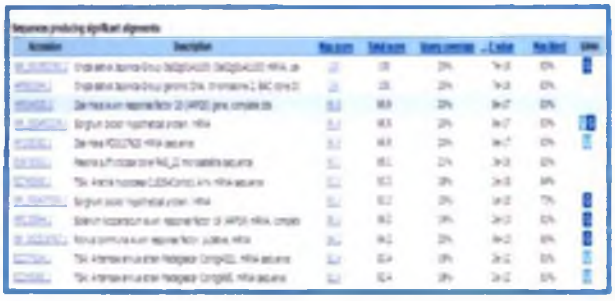
VecScreen output



Open reading frames in the cloned sequence







Blastn output

Blastx output

Fig. 17 Results of sequence data analysis for the EST-17[PNK 1(2)54]

Tabl: 20 Results of sequence data analysis for the EST-18 |PNK 1(2)44|

1. **Initial Sequence length** : 661
2. **Vector sequence** : 462 to 661
3. **Sequence length after Vector/Adaptor Screening:** 461
4. **Blastn Results**:-

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XR078117.1	*Vitis vinifera* misc RNA	84%	7e-69	75%
NM001138999.1	Arabidopsis thaliana transcription regulator (NOT2,3,5) family protein	82%	1e-28	68%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
CBI29924.1	unnamed protein product (*Vitis vinifera)*	3e-34	147
ABF97930.1	Not1 N-terminal domain, CCR4-Not complex component family protein, expressed (*Oryza sativa)*	9e-34	146
001154716.1	transcription regulator NOT2/NOT3/NOT5 family protein (*Arabidopsis thaliana*)	4e-28	127

6. **Blastp Results**:-

Accession No.	Description	Evalue	Score (Bits)
ABF97930.1	Not1 N-terminal domain, CCR4-Not complex component family protein, expressed (*Oryza sativa)*	9e-46	186
NP_001063802.1	transcription regulator NOT2/NOT3/NOT5 family protein (*Arabidopsis thaliana*)	5e-32	140
NP_001132471.1	hypothetical protein LOC100193927(*Zea mays)*	4e-28	127

7. **ORFs Available:**

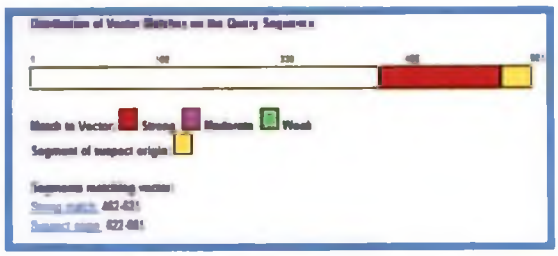
ORF Location	ORF length	Frame
1-399	399	+1

8. **Functional domains available :**

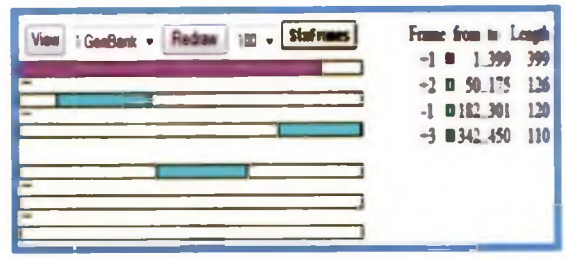
Tools	Accession no.	HMM	Domains details
CDD	-	-	Not 2 3 5 super family
Pfam	-	2- 86	Not 2 3 5

9. **Interproscan scan results:**

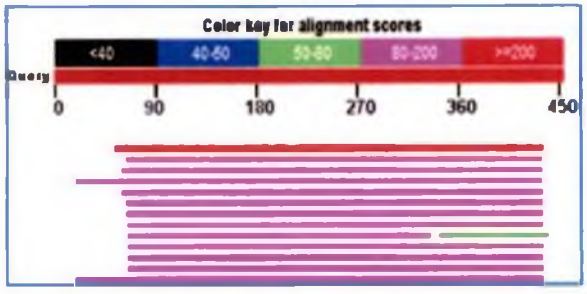
Database entry	Description
PF04153	Not 2 3 5
PTHR23326	CCR4 NOT-Releted

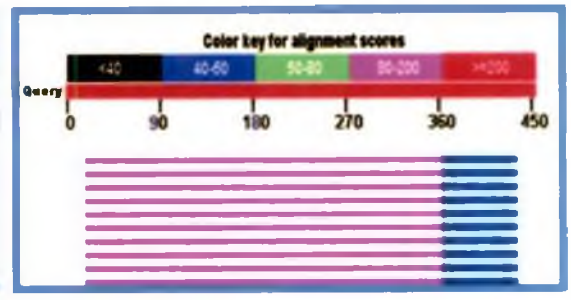


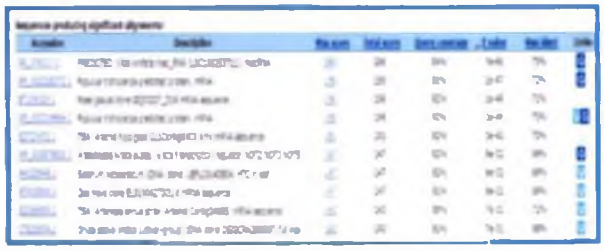
VecScreen output



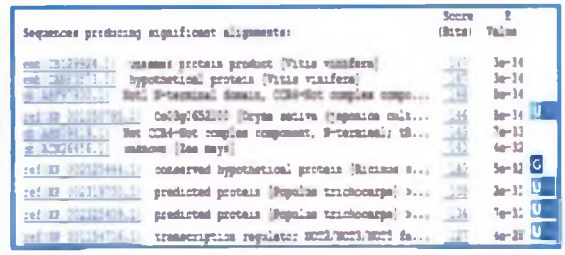
Open reading frames in the cloned sequence



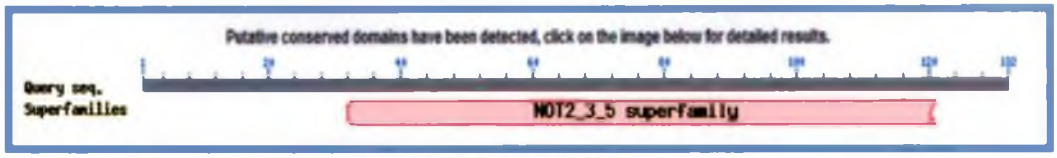




Blastn output



Blastx output



Conserved Domain

InterProScan result

Pfam result

Fig. 18 Results of sequence data analysis for the EST-18 [PNK 1(2)44]

TTCCAAAACACCCTGCTGTGACGCCTCTTAGCTTTCCTCAGGCCCCAGCCCC
AATATTTGACAATCCAACATTTTGGGAGCGATTAGGTGGTGATAATTTTGGG
ACCGACATCCTATTTTTTGCATTTTATTACCAGCAGAACACATACCAGCAGT
ATTTGGCAGCCAGAGAATTAAAGAGGCATTCTTGGAGATACCATCGCAAATA
CATACATGGTTTCAAAGGCACGAGGAACCTCTTATAACAAATGAAGAATTTG
AGCAAGGCACCTATGTGTACCTCGGCCGCGACCACGC**3'**

Blastn results indicated that about 82 percent of the sequence showed over 70 percent homology with RNA from *Vitis vinifera* and the transcription regulator (NOT2, 3, 5) family protein from *Arabidopsis thaliana*. Blastx and Blastp results indicated similarity with unnamed protein from *Vitis vinifera*, the NOT1 N-terminal domain CCR4-NOT complex from *Oryza sativa* and transcription regulator NOT-2,3,5 in *Arabidopsis thaliana* with bit score 127 to 186.

The sequence was found to have four ORFs with the longest one being 399 bp in length. The Interproscan results indicated the presence of active sites for the Not2, 3, 5. The sequence has conserved domain for the same protein. Pfam tools indicated HMM sequence 2 to 86 for the same protein.

EST- 19

The sequence data obtained for the EST-19 was obtained from the clone PNK 1(2)74 and it was of 645 bp in size. Vector screening indicated no vector sequence. The results obtained are presented in Table 21 and Fig. 19. The total sequence obtained used for further analysis.

5' TCAACGCGTCGGGAGCTNCTACCCATATGGTCGACCTGGCAGGCGAGCCA
GCGCTAATCGACTGACGGGCAGTGGCCGGGGCGGGAGCATCGACGACGTCGG
GCCCAATTCCGCGCCTCATAAGTTGAGNCTCGTATCAATTATTATAGTGATA
TATGTGTGGGTGCCATTGCTTGCCGCTTGGAGAAGAAAGAAGGGGGAGCTAT
TCGGGTTTATATAATGACTCTTGGTGTTTTGGCTCCATACCGTGGATTGGGT
GTTGGAACTAGGCTGCTGAATCATGTTTTGGAAGCCGCCTCAAAGCAAAATA
TTTCTGAGATCTACTTGCATGTGCAAACTAACAATGAGGAAGCTATAAACTT
CTACAACAAATTCGGTTTCGAAATCACAGAAACCATACAGAACTACTATATT
AACATCACCCCACCAGATTGCTATGTTGTTACAAAGATTATTGCCCCAGCTC
AACCCAAAAAGTAAGCCACATTCAAGTTTATTAGGGTTATGTTTCAGGCCAT
GTTACGTTCTGCCGTGCTTGAGATGGAACTATTGCCAAACCATTTCATTAAT

Table 21 Results of sequence data analysis for the EST-19 [PNK 1(2) 74]

1. Initial Sequence length : 645
2. Vector sequence : no match found
3. Sequence length after Vector/Adaptor Screening: 645

4. Blastn Results:

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
BT090284.1	Soybean clone mRNA	47%	6e-72	80%
XM00214314.1	Splicing factor CWC 24 *(Ricinus communis)*	50%	6e-59	76%
NM001153396.1	Zea mays pre-mRNA splicing factor	47%	3e-56	76%

5. Blastx Results:

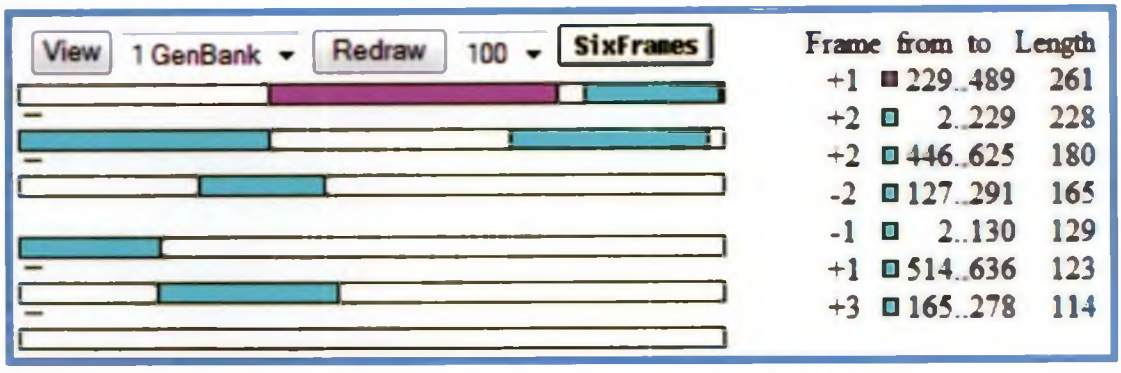
Accession No.	Description	Evalue	score (Bits)
XP_002284766.1	hypothetical protein (Vitis vinifera)	3e-51	205
NP01148868.1	Pre- mRNA splicing factor (*Zea mays*)	2e-50	199

6. ORFs Available:

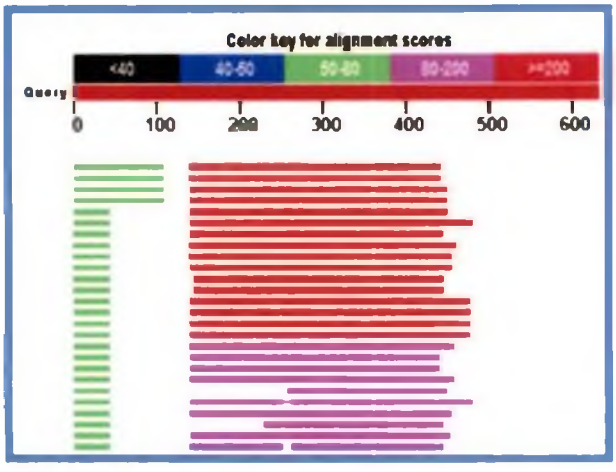
ORF Location	ORF length	Frame
229-489	261	+1

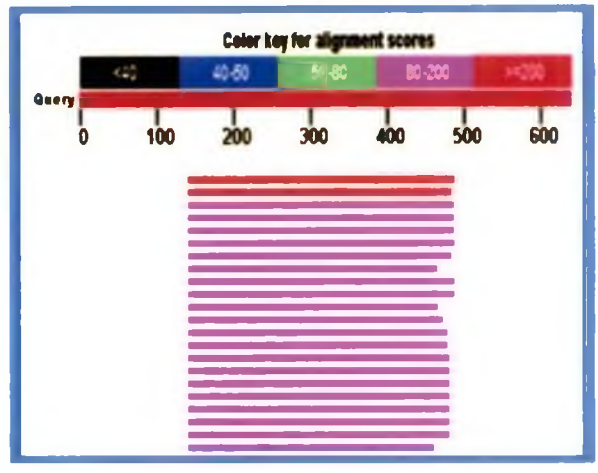
7. Functional domains available :

Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	-	-	nil
pfam	-	nil	nil

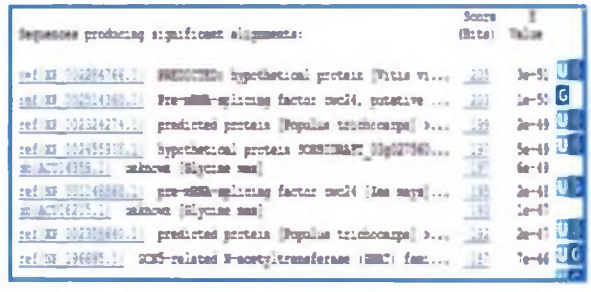


Open reading frames in the cloned sequence





Blastn output



Blastx output

Fig. 19 Results of sequence data analysis for the EST-19 [PNK 1(2)74]

ACATCATTTCAGGGAAATACTAAAGCACGAGCTATTTCATTTGTAGCGAAGT
TTCTAT3'

Blastn results indicated that 47 per cent of the sequence showed 80 per cent homology with the mRNA clone from soybean and 50 percent of the sequence has 76 percent homology with Splicing factor from *Ricinus communis* and *Zea mays*. Blastx results, indicated similarity with hypothetical protein from *Vitis vinifera*(bit score 205) and the Pre-mRNA splicing factor from *Zea mays* (bit score 199).

The sequence was found to have seven ORFs with the longest one being 261 bp in length. Interproscan and CDD analysis showed negative results.

EST-20

The sequence data obtained for the EST-20 was obtained from the clone PNK 1(2) 37 and it was of 788 bp in size. Vector screening indicated vector sequence from 1-66 bp. The results obtained are presented in Table 22 and Fig. 20. After vector and adaptor screening the total sequence obtained for further analysis was 722 bp.

5'TCGAGCGGCCGCCCGGGCAGGTACAAGTTTACATATATATACTCCCAGCAGC
CTGTGTATTATTATACATATCATCATCCCCTTCAAGCATTTATCAACAACAGCT
CATCATGCAGAATACATTCCAAACATATGCCCACCAAATTATTCTATCTCGTCC
CTTCTTTTATCACCTAATCACAAAAGAAATTAAAAGCTCTCGCCTATAGAATGA
TACAGAGCCGCTGAGGCAAACAATCTCTTCAATGCTTATTTTTCCAACCCTAAT
CAGATCTGTTTTAACTCAACAAAAATGAGCAGTATAAAGATATTTCTGAGTTAT
TTACAGTGTGCAATGTTAACCTTTGCCAGCATGTTTGCGTAAAATTTCATATCC
ATTTTACTGGCTCGTGTCTTTAGTTTTTTCACTCTCTCCAGTAGTCGATTTAAT
GGCTCATTGTCTGGTTCAATTTCCTTTGCAGCCAATAAGTCTATTTCTGCCAAG
TCCAGATCACTTGTCTCAATGTAAGACTGAGCTCTACGGTAGAGAGCCTTCACG
TTGTCAGGCTCTAAATTGCAAAGCCCTTTGCACAGAGCTTCGCCAGCTTCCCGG
AATTCATTCAGTTTCAGTTTGCAAGCTGCATTATTCAAAATGCAGGTCACTCGG
AAAAGTACCATCTTTTTCTGCTCCTGATTGAAAGTACCTCGGGCCGCGACCACA
GCTATCCGGGCCATGGGGGCGG3'

Table 22 Results of sequence data analysis for the EST- 20 [PNK 1(2) 37]

1. **Initial Sequence length** : 788
2. **Vector sequence** : 1to 66
3. **Sequence length after Vector/Adaptor Screening:** 722
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AK327113.1	Solanum lycopersicum cDNA clone	23%	3e-12	72%
Y07636.1	*Triticum astivum* mRNA for peptidylproyl isomerase	21%	2e-08	71%

5. **Blastx Results:**
6.

Accession No.	Description	Evalue	score (Bits)
XP_002534361.1	peptidylproyl isomerase (*Ricinus communis)*	6e-16	88.6
NP 001118695.1	FK506 binding protein	8e-21	85.5

7. **Blastp Results:**

Accession No.	Description	Evalue	score (Bits)
XP_002534361.1	peptidylproyl isomerase (*Ricinus communis)*	9e-23	110
NP 001118695.1	FK506 binding protein	8e-21	104

8. **ORFs Available:**

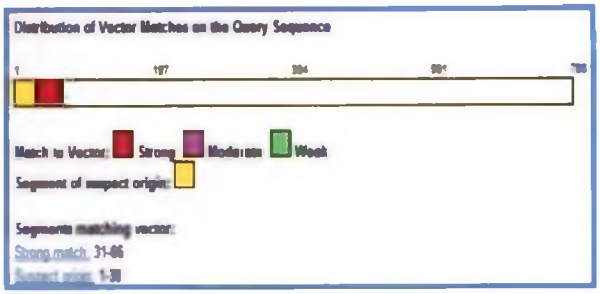
ORF Location	ORF length	Frame
319-714	396	-3

8. **Functional domains available :**

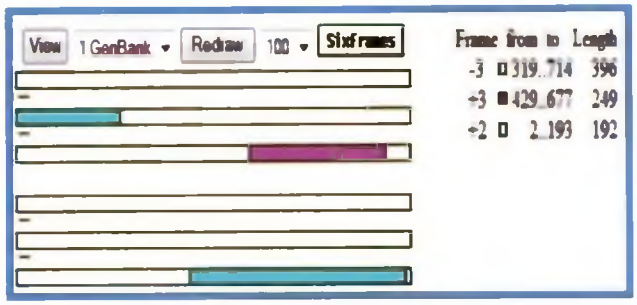
Tools	Accession no.	HMM	Domains details
CDD	-	-	nil
pfam	-	nil	nil

9. **Interproscan scan results:**

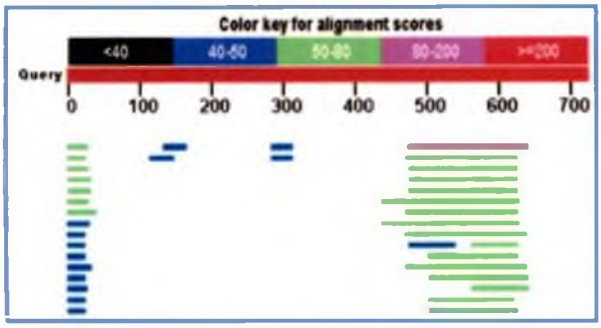
Database entry	Description
PTHR10516	FK506 binding protein

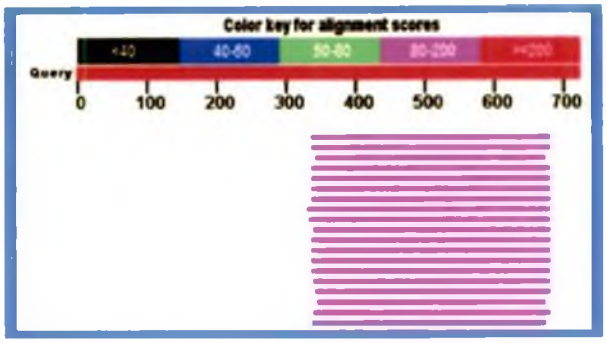


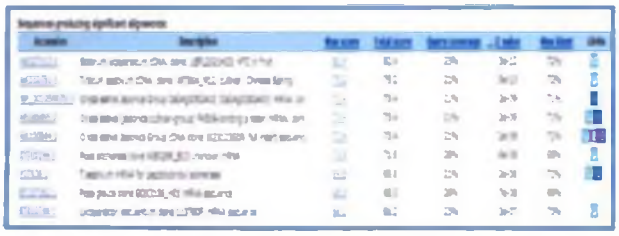
VecScreen output



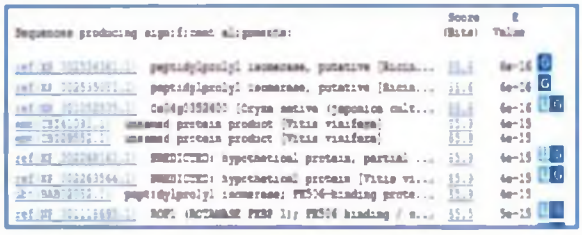
Open reading frames in the cloned sequence



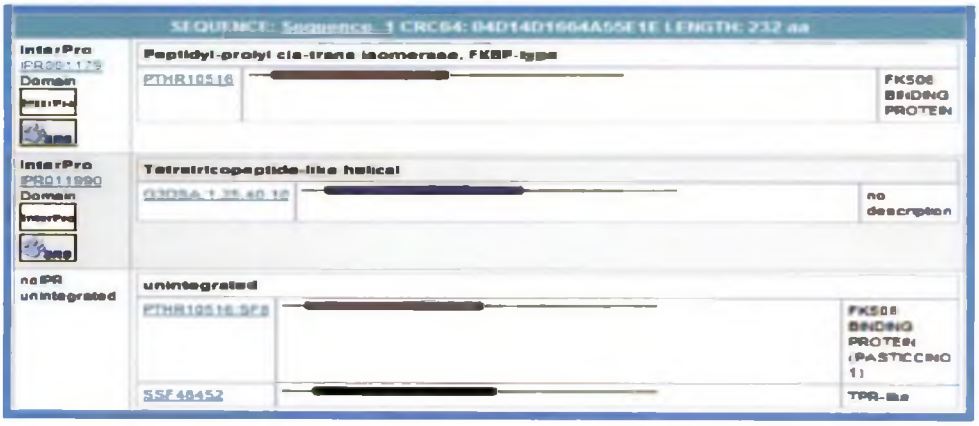




Blastn output



Blastx output



Interproscan result

Fig. 20 Results of sequence data analysis for the EST- 20 [PNK 1(2)37]

Blastn results indicated that about 21 percent of the sequence showed homology with cDNA clone from *Solanum lycopersicum* and the peptidyl proyl isomerase from *Triticum astivum*. Blastx and Blastp results showed homology with the peptidyl prolyl isomerise from *Ricinus cumminus* and FK 506 binding protein with 85.5 to 110 bit score.

The sequence was found to have three ORFs with the longest one being 396 bp in length. Interproscan indicated the presence of active site for the FK 506 binding protein. Conserved domain was not reported for this sequence.

EST-21

The sequence data obtained for the EST-21 was obtained from the clone PNK 1(2) 73 and it was of 616 bp in size. Vector screening indicated vector sequence from 1-57 and 576-616 bp. The details of results obtained are presented in Table 23 and Fig. 21. After vector and adaptor screening the total sequence obtained for further analysis was 499 bp.

5′ TGATTTCGAGCGGCCGCCCGGGCAGGTACAATGTCGCCAAGCGTTTTCTCAT
CACTAGAAGAACACGCAAATTTGTTACCCGCTCTTGAACTTTCTCCCTGAATAA
TGGGGCAAGAATGACTAGGTTGGGGTGCTTCACCTTTCTTTCCTATATAACTCA
TAATTCTAGGACCTGAAATTGGGCAAGTTTTAAAGCCAACTTCAAGATCACATA
CAGGTTGTTCCAATAATCTATATGGACAAGAAATAGCCCAATGGGTAAGCCTGA
AACAACGTATGCATAGGCATGATGCATCCTCTTCACCACTACATGAATTGAAGT
TGCCCAACAAATCCTCTAGTTCAGATTCATCTATCTCACTGCAATCTCGCAGTA
TGTGACCTCTCATGCCACAGAAAAGGCAGATTGTATTAGTTTCTTTGCCTCCTC
TAACTTTTGATGATGTAATGTGCTTAAGTGCATCCAATCTTCTCGCAAAAATAG
AAGCAATGTCATCGGTACCTCGGCCGCGACCACG3′

Blastn results indicated that 5 percent of the sequence showed 100 per cent homology with the hypothetical protein from *Eristalis tenax*. Blastx results indicated similarity with the transcription factor in *Ricinus cumminus* (bit score 50.0) and hypothetical protein from *Vitis vinifera* (bit score 46.1). In Blastp results showed homology with hypothetical protein from *Vitis vinifera* (bit score 85.5).

Table 23 Results of sequence data analysis for the EST-21 [PNK 1(2) 73]

1. **Initial Sequence length** : 616
2. **Vector sequence** : 1-57, 576-616
3. **Sequence length after Vector/Adaptor Screening:** 499
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AM706411.1	(Eristalis tenax) hypothetical protein	5%	0.001	100%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score(Bits)
AAG50695.1	Transcriptional activator (*Ricinus communis)*	5e-15	50.0
EEC68988.1	hypothetical protein (Vitis vinifera)	9e-14	46.1

6. **Blastp Results:**

Accession No.	Description	Evalue	Score(Bits)
XP_002280338.1	hypothetical protein (Vitis vinifera)	2e-15	85.5

7. **ORFs Available:**

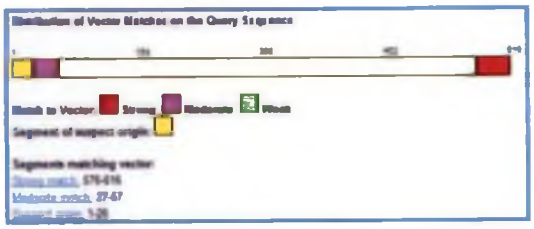
ORF Location	ORF length	Frame
3-467	465	-1

8. **Functional domains available :**

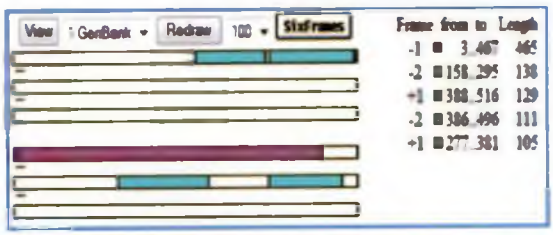
Tools	Accession no.	HMM	Domains details
CDD	-	-	Nil
pfam	-	nil	Nil

9. **Interproscan scan results:**

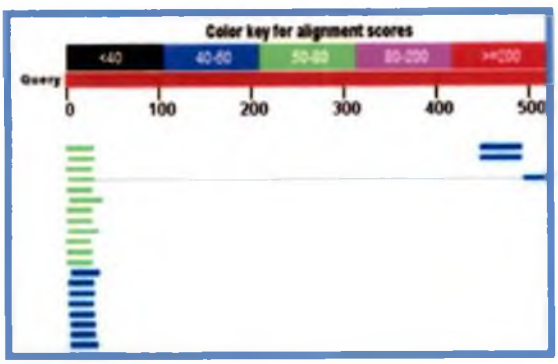
Database entry	Description
SM00343	Zinc finger domain

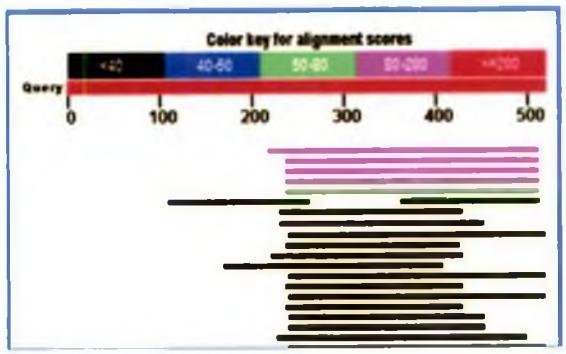


VecScreen output

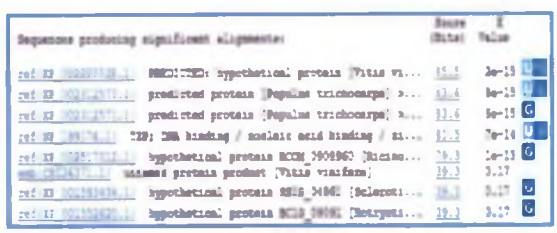


Open reading frames in the cloned sequence

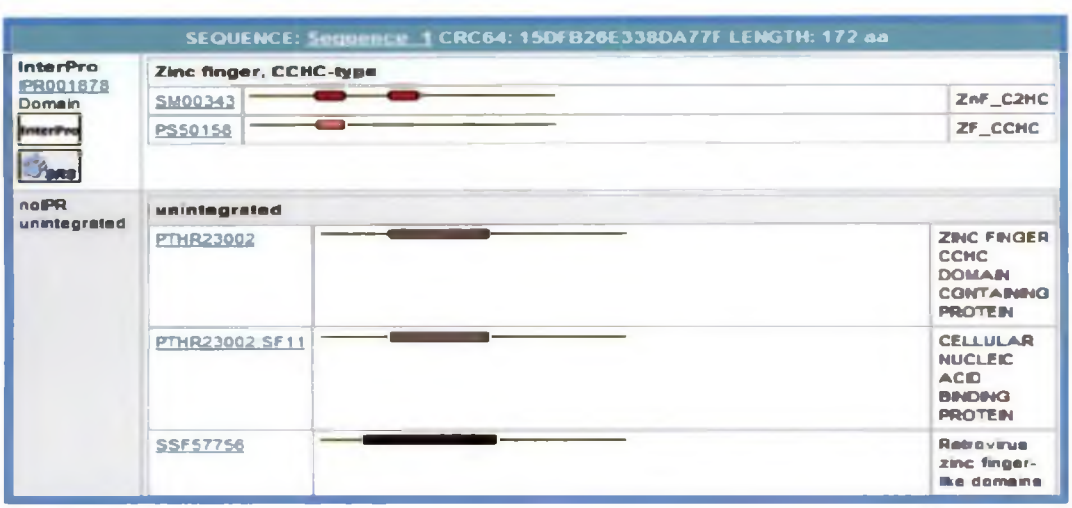




Blastn output



Blastx output



Interproscan result

Fig. 21 Results of sequence data analysis for the EST-21 [PNK 1(2)73]

The sequence was found to have five ORFs with the longest one being 465 bp in length. Interproscan results indicated presence of the Zinc finger domain and CDD analysis showed negative results.

EST-22

The sequence data obtained for the EST-22 was obtained from the clone PNK 1(2) 31 and it was of 496 bp in size. Vector screening indicated vector sequence from 369 to 496 bp. The results obtained are presented in Table 24 and Fig. 22. After vector and adaptor screening the total sequence obtained for further analysis was 368 bp.

5' GATTTCGAGCGGCGCCGGGCAGGTCCTAGAAGTCTGGTGAATACCTTCGC
TCCATTAAAGGGGCAACGGGAAAGGAAAACGTTGATGCCAGTGCCTTACAAA
CCATAAGTTTTGTCACCAACATGAAAACTTACGGACCGTACGGCTCGACACT
AGAAGGTGCTCTCCCCTTTGACGCTCCAGTGAGAAATGGGAAGATCACTGGC
TTTTCTGGCAGAGGATCCTGGTGGGTTGATGCAATTGCCATTAAGGTCGCTC
CAGTGGAGGACTGATCTACTACTAGTATAGACATATGAGGTTAATGGTGTTT
GCACTTGGCAGGTCCAGTGCAACACGTGTCTGTACCTCGGCCGCGACCACGC
TAATCC3'

Blastn results indicated that 7 percent of the sequence showed 100 percent homology with Synthetic construct (RLS) gene. Blastx results indicated homology with the Jacalin like lectin(bit score 44) from *Ananus comosus* and mannose binding lectin from *Musa acuminata* and *Morus nigra* with bit score 40 to 41. In Blastp results showed similarity with the jacalin - domain protein from *Plantago majar* with 48.1 bit score and mannose-binding lectin from *Musa acuminate* with 47.0 bit score.

The sequence was found to have three ORFs with the longest one being 216 bp in length. The Interproscan results indicated the presence of the active sites for the jacalin and mannose binding lectins. The sequence has conserved domain for the same protein. Pfam tools indicated HMM sequence (2 to 86) for the same protein.

EST-23

Table 24 Results of sequence data analysis for the EST- 22 [PNK 1(2)31]

1. **Initial Sequence length** : 496
2. **Vector sequence** : 369 to 496
3. **Sequence length after Vector/Adaptor Screening:** 368
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AY972077.1	Synthetic construct (RLS) gene	7%	0.003	100%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score (Bits)
AAQ07258.1	jacalin-like lectin (*Ananas comosus)*	9e-10	44
ABS86034.1	mannose-binding lectin (*Musa acuminata* AAA Group)	2e-08	41
AAL10685.1	mannose-binding lectin (*Morus nigra*)	8e-08	40

6. **Blastp Results:**

Accession No.	Description	Evalue	Score (Bits)
CAJ38387.1	jacalin-domain protein (*Plantago major*)	3e-04	48.1
ABS86034.1|	mannose-binding lectin (*Musa acuminata* AAA Group)	6e-04	47

7. **ORFs Available:**

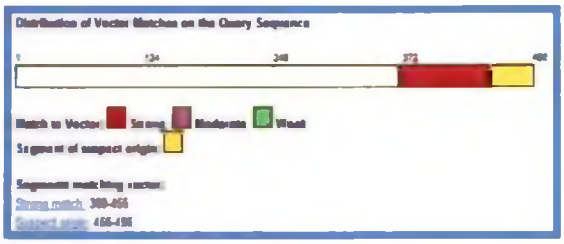
ORF Location	ORF length	Frame
123-272	216	+3

8. **Functional domains available:**

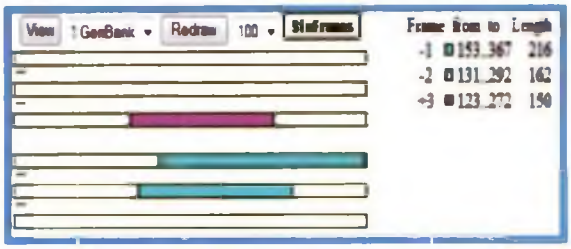
Tools	Accession no.	HMM	Domains details
CDD	01419	-	jacalin family
Pfam	-	2- 86	jacalin like protein domain

9. **Interproscan scan results:**

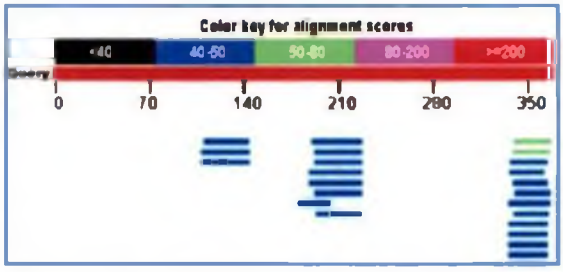
Database entry	Description
PF01419	jacalin
SSF51101	mannose-binding lectin

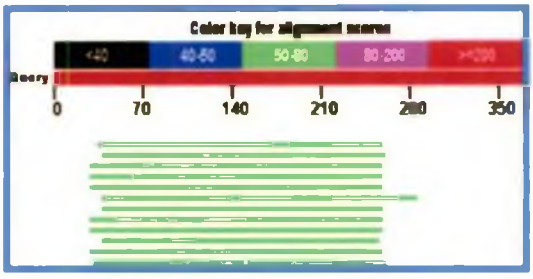


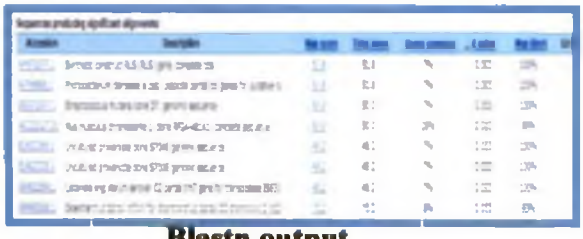
VecScreen output



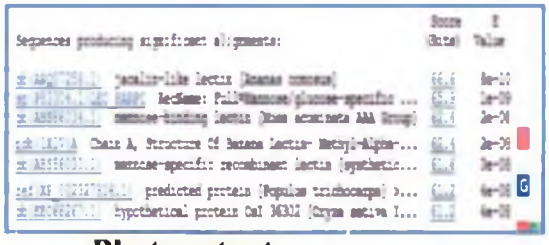
Open reading frames in the cloned sequence







Blastn output



Blastx output

Conserved Domain

InterProScan result

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry Type	Clan	Envelope		Alignment		HMM		Bit score	E-value
				Start	End	Start	End	From	To		
Jacalin	Jacalin-like lectin domain	Domain	n/a	8	85	11	85	60	130	49.6	2.7e-13

Pfam result

Fig. 22 Results of sequence data analysis for the EST-22 [PNK 1(2)31]

The sequence data obtained for the EST-23 was obtained from the clone PNK 1(2) 85 and it was of 676 bp in size. Vector screening indicated vector sequence from 1-69 and 653-676 bp. The results obtained are presented in Table 25 and Fig. 23. After vector and adaptor screening the total sequence obtained for further analysis was 584 bp., as detailed below

5' CAGCGTGGTCGCGGCCGAGGTACCGGAGATCGTTCATGGCGGAACTTGCC
TTCTCTCTTCACCCTCTCCTTCGCTCCCCTCCTCCATTTTCCCTCATCCCTC
TCCCCGCTTCCATTGGAATGGCTGCCCAAGCTCTTCTATCCGGCGGAGTAGG
TTTCCGGTGTCGAAAAGTGTCGTTCTTTGTGGGTGTAAAGGTCTGTCGGTGC
CGTTCAGAAGGAGGGAGCTTGTGCTGCTCAGTCTCTCGTTTCCGGCTTCTCC
CAGTTTGCATCCGCCTGCTTCAATTGCCGAGGTTGAGCAGAGTGATGGCATC
AAAATGGTTCCATTAGTAACGGAATTAATGCTTATTTATTTCTCTATCCTGT
CGAACTTCCTTCAAAAAAGTTTGCCTTTAAGTGGGTTGAATCCAGAAAACCT
GAACGTTATTCTTCTGCAGCACCATTGTCTCCTGATGCTAGACAGCGTATTG
TATCTGAGCGTGTGGACATGATCAATAATGTTGTCATCTCTGTTTCGATTGG
CCCACCTAATTCACGTCTATTACCGACTAAAGACAAAAGTACCTGCCCGGGC
GGCCGCTCGAAAT3'

In Blastn analysis indicated that 38 percent of the sequence showed 77 percent homology with the full length cDNA clone from *Zea mays* and hypothetical protein mRNA. Blastx results also indicated homology with hypothetical protein from *Zea mays*.

The sequence was found to have seven ORFs with the being 291 bp in length. Interproscan showed presence of transmembrane and conserved domain was not reported.

EST-24

The sequence data obtained for the EST-24 was obtained from the clone PNK 1(2) 24 and it was of 784 bp in size. Vector screening indicated vector sequence from 1-53 bp. The results obtained are presented in Table 26 and Fig. 24. After vector and adaptor screening the total sequence obtained for further analysis was 731 bp.

5' CGTGGTCGCGGCCGAGGTGTCAGACTTTGGATTGGCTCGCCAAGGGCCTG
GTGAAGGACTTAGTCATGTCTCAACTGCGGTTGTTGGGACTATAGGGTATGC

Table 25 Results of sequence data analysis for the EST-23 [PNK 1(2) 85]

1. **Initial Sequence length** : 676
2. **Vector sequence** : 1 to 69, 653 to 676
3. **Sequence length after Vector/Adaptor Screening:** 584
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
BT060570.1	Zea mays full lenght cDNA clone	38%	1e-41	77%
EU9744930.1	Hypothectical Protein mRNA	38%	1e-41	77%

5. **Blastx Results:**

Accession No.	Description	Evalue	score Bits
XP_002284766.1	hypothetical protein (*zea mays*)	2e-32	142

6. **ORFs Available:**

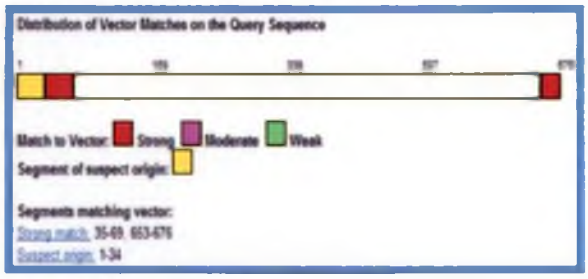
ORF Location	ORF length	Frame
36-326	291	+3

7. **Functional domains available :**

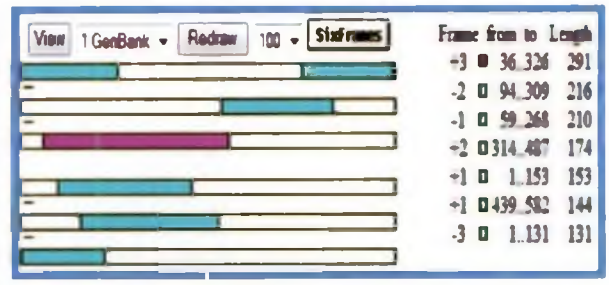
Tools	Accession no.	HMM	Domains details
CDD	-	-	nil
pfam	-	nil	nil

8. **Interproscan scan results:**

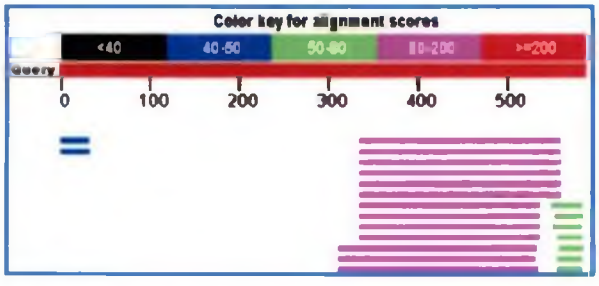
Database entry	Description
tmhmm	Transmembrane regions

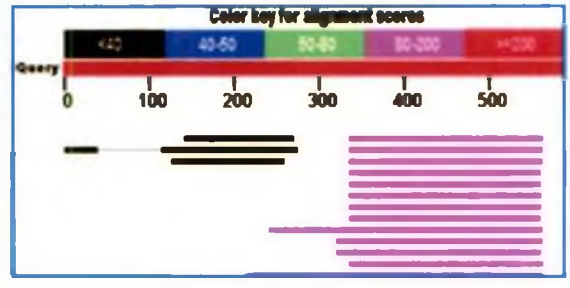


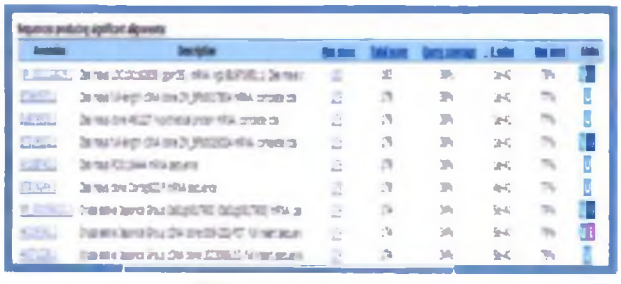
VecScreen output



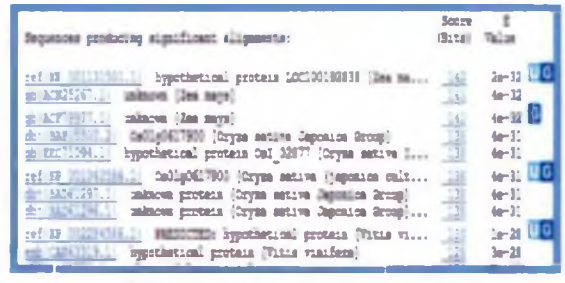
Open reading frames in the cloned sequence







Blastn output



Blastx output

InterProScan result

Fig. 23 Results of sequence data analysis for the EST-23 [PNK 1(2)85]

Table 26 Results of sequence data analysis for the EST-24 [PNK 24]

1. **Initial Sequence length** : 784
2. **Vector sequence** : 1 to 53
3. **Sequence length after Vector/Adaptor Screening** : 731
4. **Blastx Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XP002274190.1	Hypothetical protein (*Vitis vinifera*)	58%	8e-59	61%
AAT81727.1	expressed protein (*Oryza ativa*)	58%	1e-53	53%

5. **Blast p Results :**

Accession No.	Description	Evalue	Score (Bits)
XP002274190.1	Hypothetical protein (*Vitis vinifera*)	3e-57	204
AAT81727.1	expressed protein (*Oryza sativa*)	2e-42	175

6. **ORFs Available:**

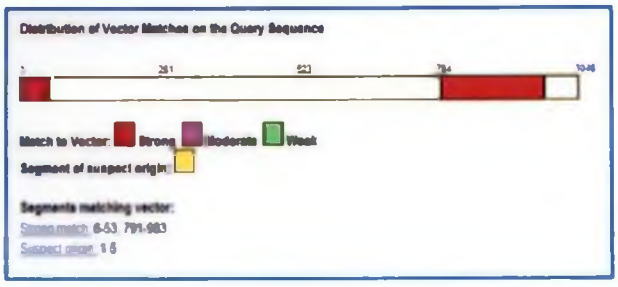
ORF Location	ORF length	Frame
90-605	516	+3

7. **Functional domains available :**

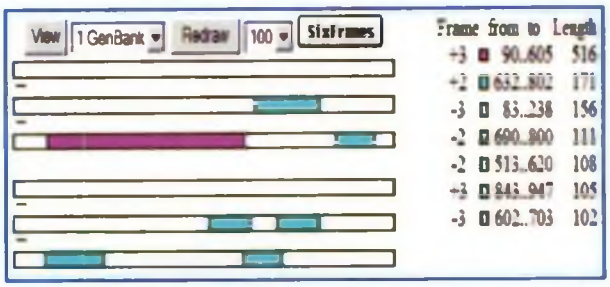
Tools	Accession no.	HMM	Domains details
CDD	CI02429	-	TPR super family
pfam	-	nil	Nil

8. **Interproscan scan results:**

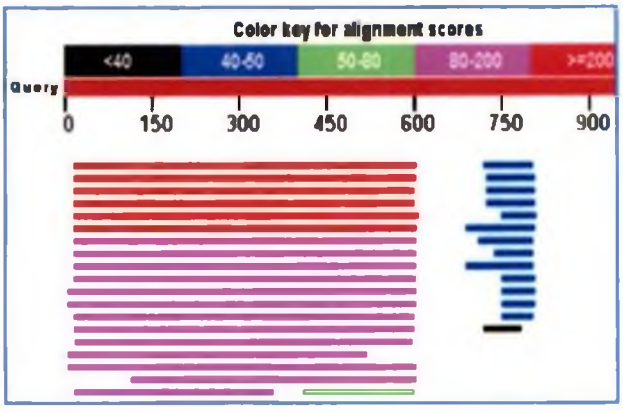
Database entry	Description
signalip	Signal peptide

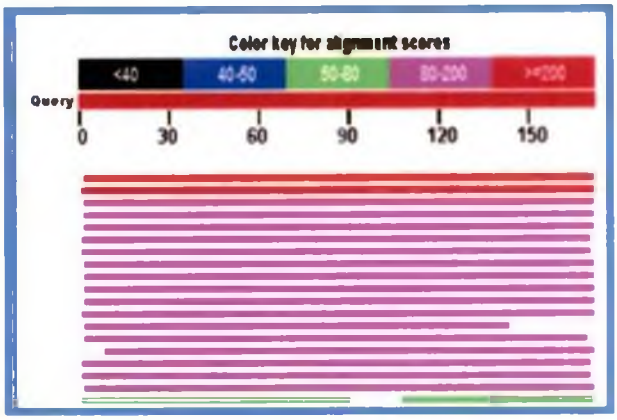


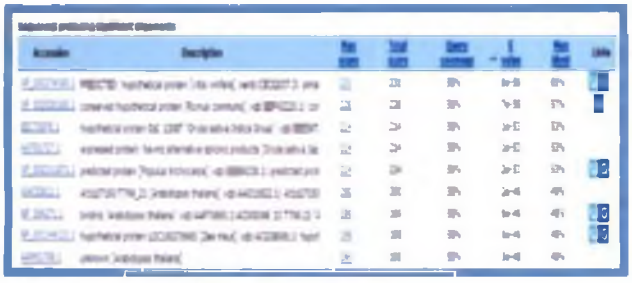
VecScreen output



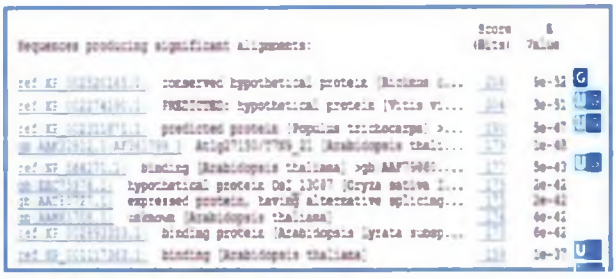
Open reading frames in the cloned sequence







Blastn output



Blastx output

Conserved Domain

InterProScan result

Fig. 24 Results of sequence data analysis for the EST-24 [PNK 24]

AGCACCAGAATATATTCAAACTGGACGTTTGACGGCTAAAAGTGATGTTTGG
AGCTATGGAGTTGTCCTCTATGAACTGATCACTGGTAGAAAGCCAATAGACC
GTAATCGTCCTAAGGGTGAACATAAACTACTTGAGTGGGTGAAGCCTTATTT
ATCAGAGAAGCGTTTTTCCATGATTCTTGATCCAAGAATACAAGAATAGATT
TCCCTGAAGTCGGCCCAAAAGCTTGCCGCTGTTGCAAACAAATGCCTTGTAC
GACAGCAGAAGGCACGTCCAAAGATGAGTGAGATTTTAGCGATTGTGAAGCA
AATTGTGGAGTCATCAACAACTGGGTCTCCTCAACCTGCCATCAAAATCCTA
GATTCAAGTGAAGTCTCCAATGAAGAACACAAAAACAGAAGCCGATTGAAAC
GAGCCATTGCTTATCTTAATATGGGTGGAGATGGCAAGAGGTTTTACTGGTG
GAAGTTCAAATCCAAGCTCAGTAAAGTTAAGTTGTGGAAGACGAAGATAGGA
CTAATGGTTTCTAGGCTAGCGTCATGATAAGAAACCAGGCATAGGAGAAGTT
GGGAAGTTTTTAGAGATGGCCTAAAGAATCCACCATAAAGTGTCAGT3'

Blastn results indicated that 58 percent of the sequence showed 61 percent homology with the hypothetical protein from *Vitis vinifera* and 53 percent homology with expressed protein from *Oryza sativa.* Blastx results indicated similarity of the sequence with hypothetical protein from *Vitis vinifera* (bit score 204) and the expressed protein from *Oryza sativa* (bit score 175).

The sequence was found to have seven ORFs with the longest one being 516 bp in length. Interproscan indicated presence of the signal peptide. A conserved domain for TPR superfamily was reported.

EST-25

The sequence data obtained for the EST-25 was obtained from the clone PNK 1(2) 42 and it was of 662 bp in size. Vector screening indicated vector sequence from 503 to 662 bp. The results obtained are presented in Table 27 and Fig. 25. After vector and adaptor screening the total sequence obtained for further analysis was 502 bp., as detailed below

5'TTCGAGCGGCGCCGGGCAGGTACAAACTGAGTGCACGACGAAATGAAAGG
GTAGTAGGGAATGCCAGCTCTGGGATGCTATGGCATCTGCTGTATGCCACTA
CCCAACCCTCATTGATGTTTTGATTTTTGGTCCATTCTCAGTCGTGCAGAAC
TATCGGGAAGAAACATCTGGCAGGCGAGCGACAACCAATCCTATCTGGAATC
TGATACATTGGGCTAGCGCTTACTGGGTAGGCTAAGAGCTGTCCACTCCATT
ATTTTTCCACCATCGGGAAGAATTCTTGTTGTGCTGTTCGTTTCAGTGGGAT

Table 27 Results of sequence data analysis for the EST-25 |PNK 1(2)42|

1. **Initial Sequence length** : 662
2. **Vector sequence** : 503 to 662
3. **Sequence length after Vector/Adaptor Screening:** 502
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XP_002286915.1	Predicted protein *(Thalassiosira pseudonana)*	34%	4.8	41%

5. **ORFs Available:**

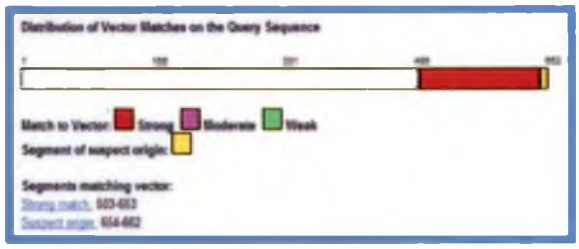
ORF Location	ORF length	Frame
80-241	162	+2

6. **Functional domains available :**

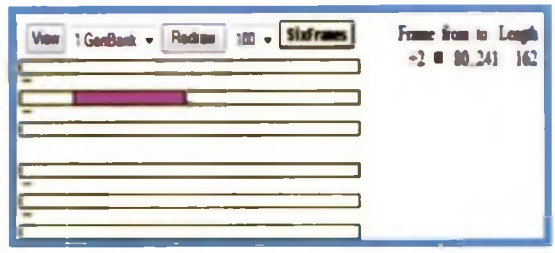
Tools	Accession no.	HMM	Domains details
CDD	-	-	Nil
pfam	-	nil	Nil

7. **Interproscan scan results:**

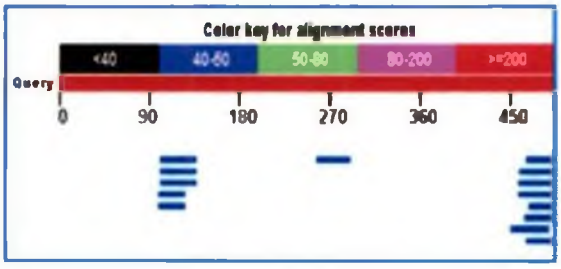
Database entry	Description
signalip	Signal peptide

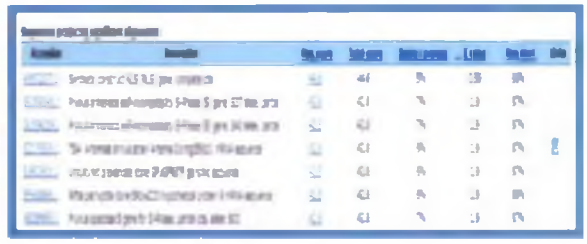


VecScreen output

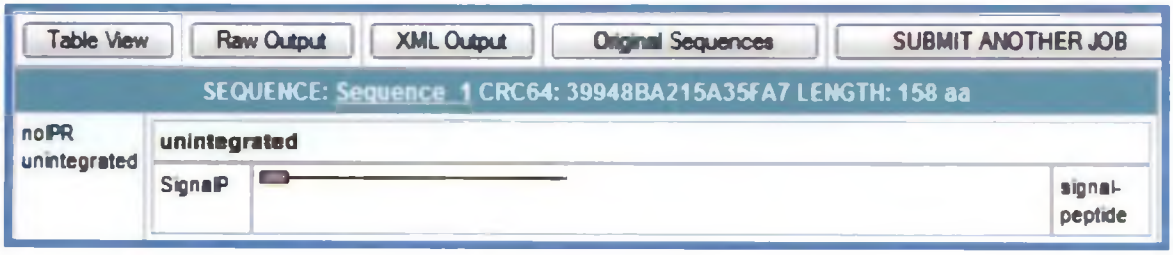


Open reading frames in the cloned sequence





Blastn output



Interproscan result

Fig. 25 Results of sequence data analysis for the EST-25 [PNK 1(2)42]

CGATCCCCTGCCTTAAGGATGTCCAAAACAGCAGAGTTATACCACTGGGTTA
TGACAACGTTGACAACTCATTATGCACATTGAGCAAACTGAACGTGGTTGTT
CTTAGAAAGTAAGATGTGAGCATATAGGCAATGCCCATATCAGGGTGATGTT
AAAGTGTACCTCGGCCACGACCACG3'

Blastn results indicated that 34 percent of the sequence showed 41 percent homology with the Predicted protein from the *Thalassiosira pseudonana*.

The sequence was found one ORF with 162 bp in length. The Interproscan results indicated the presence signal peptide. The sequence has no any conserved domain.

EST-26

The sequence data obtained for the EST-26 was obtained from the clone PNK i(2) 69 and it was of 851 bp in size. Vector screening indicated vector sequence from 1-44 bp. The results obtained are presented in Table 28 and Fig. 26. After vector and adaptor screening the total sequence obtained for further analysis was 807 bp.

5'TTCGAGCGGCCGCCCGGGCAGGTACAGACCTATGTGGGCAGAGCCAGATGGG
ATAAACCTCCATTCGTCAGCGATCATTCAAAGCTCTACACGAGCGATCGCCTAA
TCCTTGTAAGATGGGGCACCAAACATAATCTTGAAGTCAGCACTGCATTCTCAT
GGTAATAGTTCCGAATAGCCTCACCTTGGGTAGCCTTAAGGTAAGGAAACCTTT
GGTGGCATCAGCCAACCATAAAGCCCAATCGGTAAGCCACCACTGGTCAGCAAA
GAGGGAGAATCAGCTTTCCCTCCCAATACAAGCACCTTGTATAGACCTGTGGTA
ATACGTCCCGAGTAGGAATGTATTCAACGAATTCACCCTTCTTGAAAGCGAGCT
TCCGCGTGTTGATAGAACCATGGGCTTGGACACCACCATCCGACGATTGTTGAT
AACATTCCCCATCAGCTTTTGNTTCTTCCTTTGATGCATATCTGCCAGTCATCA
TCACTCCTTCGGAGGCATTATAAATCATGCCTCAACGAAATCAGGAACCTTTGG
CGGGTCACATCTTTTGGGAAGCAAAAGGGAATCCATATTAGCTTGGGGCTTACA
AATTCTGCATGTCCCCCCTCGCANGGTTGTATAAGTTCAACCTTCAATGTTATT
CCTTTAAACATGGATGCAGCCTTATTCTCAGGGGCACTAAAAGGGCTAATGGTT
GCCTGTGAACATTGGGGATGGTTAAAACAAGTCAAGTATTCGTTTTCTTGGNAG
CTCGAAACAAAGNCCATCACCGAAAAAATCTCATAACGCAAAATGGGTCCGAG3
'

Table 28 Results of sequence data analysis for the EST-26[PNK 1(2) 69]

1. **Initial Sequence length** : 851
2. **Vector sequence** : 1 to 44
3. **Sequence length after Vector/Adaptor Screening** : 807
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AK226335.1	Cell elongation protein (Arabidopsis thaliana)	68%	6e-60	72%
XM002520558.1	Cell elongation protein (*Ricinus communis)*	47%	3e-58	75%

5. **Blastx Results:**

Accession No.	Description	Evalue	score (Bits)
XP_002520604.1	Cell elongation protein (*Ricinus communis)*	1e-24	104
XP_002271846.1	hypothetical protein (Vitis vinifera)	2e-22	102

6. **Blastp Results:**

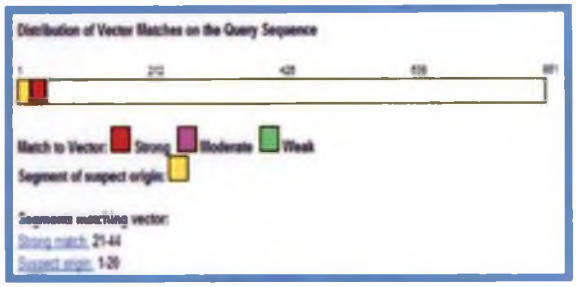
Accession No.	Description	Evalue	Score(Bits)
XP_002520604.1	Cell elongation protein (*Ricinus communis)*	0.006	104
XP_002271846.1	hypothetical protein (Vitis vinifera)	0.013	95.1

7. **ORFs Available:**

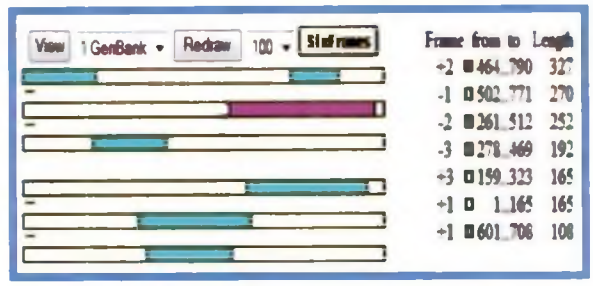
ORF Location	ORF length	Frame
464-790	327	+2

8. **Functional domains available :**

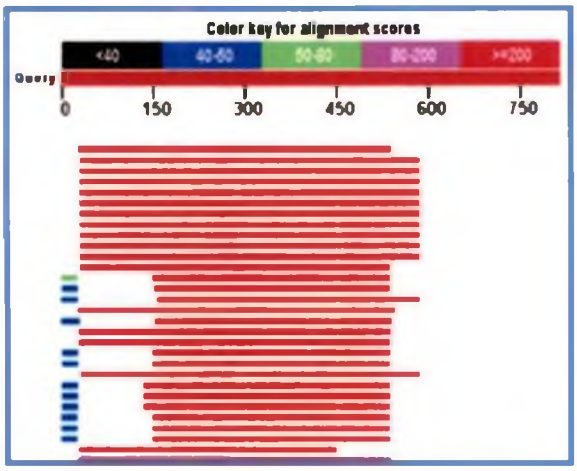
Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	-	-	nil
pfam	-	nil	nil

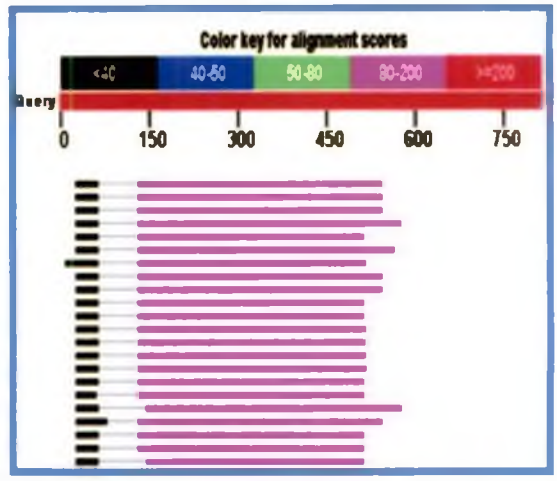


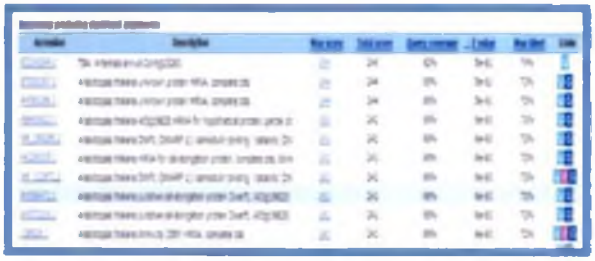
VecScreen output



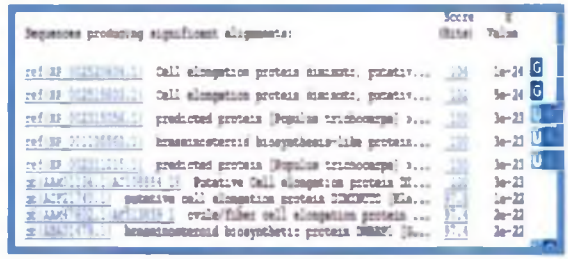
Open reading frames in the cloned sequence







Blastn output



Blastx output

Fig. 26 Results of sequence data analysis for the EST-26 [PNK 1(2)69]

Blastn results indicated that about 60 percent of the sequence showed over 72 percent homology with the Cell elongation protein from *Arabidopsis thaliana* and *Ricinus communis*. Blastx results indicated homology with the Cell elongation protein in *Ricinus communis* (bit score 104) and hypothetical protein *from Vitis vinifera* (bit score 102).

The sequence was found to have seven ORFs with the longest one being 327 bp in length. Interproscan and CDD analysis indicated negative results.

EST-27

The sequence data obtained for the EST-27 was obtained from the clone PNK 1(2) 70 and it was of 715 bp in size. Vector screening indicated vector sequence from 1-45 bp. The results obtained are presented in Table 29 and Fig. 27. After vector and adaptor screening the total sequence obtained for further analysis was 670 bp as detailed below,

5' TCGAGCGGCCGCCCGGGCAGGTACTAGTATCCTTGTTATTTAAAGGCAGGCC
GGCGAATTGGGAATCCAGGTCTTTACTGCTACTGCGTCTTTGATGCTTAGAAGC
TTGTTGCTCCTGTGGTTGAACATGCTGCTGTGCTGGCTGCTGTGGCAGTTGAAG
TTGGTAGAGTGCAACCCGGAGGCGTGATATCTCTCTCTCTCTAGCACTTCATGC
TCCATATATTTGATGAGTTGCTCCTGTGCCAAACTATCCAACTGTTGCTTCAAG
GCCTTATTTTCCATGTTCAATATAAGATTTAGTTGGTCTAGAAAC
CCTATTTCCGCTGCAATTTCAGAACCTTCTGCCTGCAATGCTTGCACATGTCGT
TCCAGATCAGCCAATGTATTGAAGTTTTCGAACGCGGGACCTCTGAGCATACTG
CTGTTTTGATCGTTTCATATCTGCTTCAAATGAAAAATTAAAGTATCCTGGCAT
GTTCAATTTTTCACCGGAAACCTCTCAAATNCATGCAGAACTGGACCCCCCCTC
CTGGTCATTGCTTCACCAAGTGATGATAGATGGGATAACCCGGTCTTTGCTCCC
TGAGGTAAATATTGGCTGTGATGGTACCTCGGGGCACGGAACACACGACAATCC
CGCGGGCCATGGGGCGGAGCTTGGAAGTCGGGC3'

Blastn results indicated that about 37 percent of the sequence showed over 70 percent homology with the Transcriptional activator from *Ricinus communis* and *Glycine max*. Blastx and Blastp results indicated homology with the transcriptional activator from *Ricinus cumminus* and hypothetical protein from *Vitis vinifera* with bit score 51.5 to 76.3.

Table 29 Results of sequence data analysis for the EST-27 [PNK 1(2) 70]

1. **Initial Sequence length** : 715
2. **Vector sequence** : 1 to 45
3. **Sequence length after Vector/Adaptor Screening** : 670
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM002532975.1	Transcriptional activator (*Ricinus communis)*	37%	1e-49	78%
EEC68988.1	Transcriptional activator (*Glycine max*)	38%	3e-38	74%

5. **Blastx Results:**

Accession No.	Description	Evalue	Score(Bits)
AAG50695.1	Transcriptional activator (*Ricinus communis)*	5e-15	51.6
EEC68988.1	hypothetical protein (*Vitis vinifera*)	9e-14	53.5

6. **Blastp Results:**

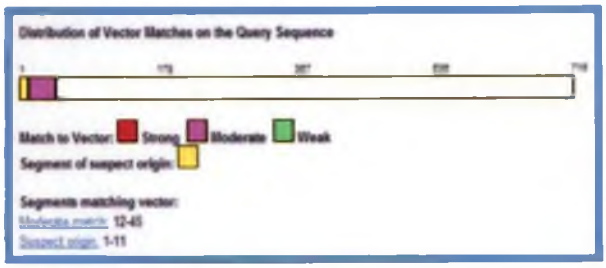
Accession No.	Description	Evalue	Score(Bits)
XP_00253302.1	Transcriptional factor (*Ricinus communis)*	1e-21	76.3
EEC67888.1	hypothetical protein (Vitis vinifera)	2e-14	53.5

7. **ORFs Available:**

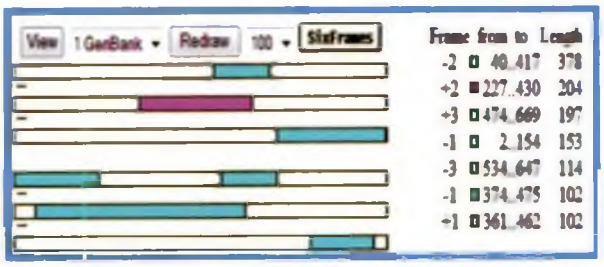
ORF Location	ORF length	Frame
40-417	378	-2

8. **Functional domains available :**

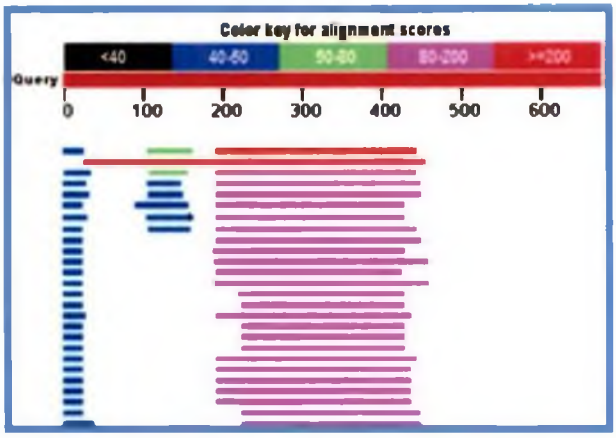
Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	DUF 972	-	DUF superfamily
pfam	-	nil	nil

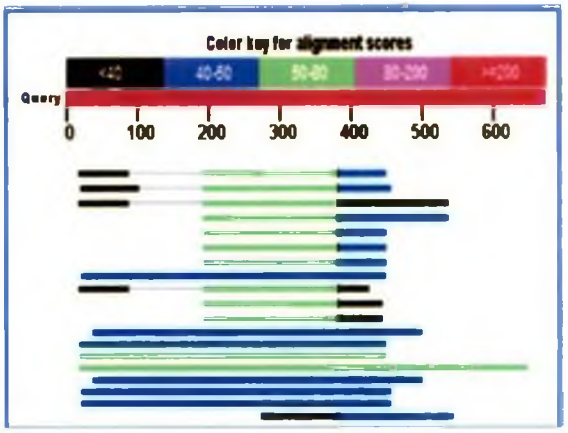


VecScreen output

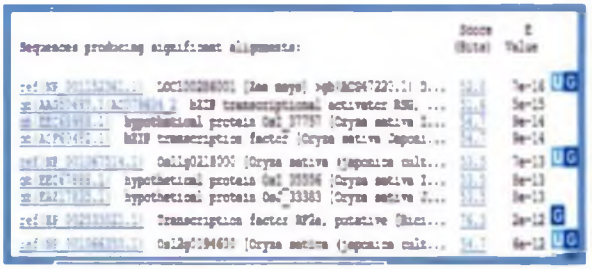


Open reading frames in the cloned sequence

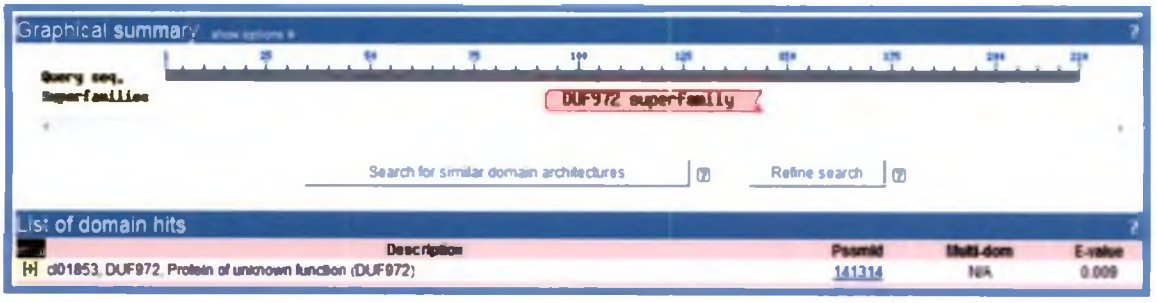




Blastn output



Blastx output



Conserved Domain

Fig. 27 Results of sequence data analysis for the EST-27 [PNK 1(2)70]

The sequence was found to have seven ORFs with the longest one being 378 bp in length. DUF superfamily conserved domain has been identified for this query sequence. Interproscan and pfam analysis indicated negative results.

EST-28

The sequence data obtained for the EST-28 was obtained from the clone PNK 1(2) 76 and it was of 506 bp in size. Vector screening indicated vector sequence from 1-67 and 455-506 bp. The results obtained are presented in Table 30 and Fig. 28. After vector and adaptor screening the total sequence obtained for further analysis was 389 bp.

5' CGAGCGGCCGCCCGGGCAGGTACTGCTCTATTATTCCAAATAAAGAAGTA
TATCACATTATACAAATTTATCCGAAAAGTTGTGATCAACTAAAACGAATAA
TGGACTTAACCTACCTGCAAAAAACAAAGACTACCAAACCTCGCAAATGCCT
TTGTCATGACAGCTTAACTCATGCAAATTACGACTACTAAACAATAATTAGT
AATAACTTCACTTTGTTAAACTAACATAACGCAACAGATTTGGAGCAGAGCA
GGCATCCTTCTTTCCATTAGCACCACTAATTCCTTCCACATGCAGTCAAACA
GCTGCTTGCACAGGACCATCTGATCTTCCCCGCCCGCCGCGTCCTCGGCCTC
GGCCATGACCTCGGCCGCGACCACG3'

In Blastn results indicated that 88 percent homology with hypothectical protein from *Sorghum bicolor*. The sequence was found to have one ORF with the being 147 bp in length. CDD and Interproscan analysis showed negative results.

EST- 29

The sequence data obtained for the EST-29 was obtained from the clone PNK 1(2) 40 and it was of 583 bp in size. Vector screening indicated no vector sequence. The details of results obtained are presented in Table 31 and Fig. 29. The total sequence obtained used for further analysis.

5' GTGAGTTGAAANTATGGTCGACTGCAGGCGGCGGAANTNATGATAAGTGC
GATGGTCGCGGCCGGAGGGTTCAGAACAGCGTCTCAATCGCATACTCTTGAT
GTAATGAGTTGGCTCACACGTGGACATAACAAGTAGCACGGACGTATCACAA
AAACAAATACTTCCCACGTCCAACATAGTCCATAATCTTCTTCGCTTCAACC

Table 30 Results of sequence data analysis for the EST-28 |PNK 1(2) 76|

1. **Initial Sequence length** : 506
2. **Vector sequence** : 1-67,455-506
3. **Sequence length after Vector/Adaptor Screening:** 389
4. **Blastn Results:**

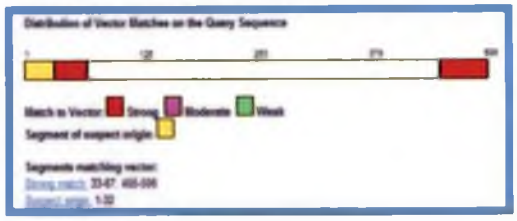
Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM002437195.1	hypothectical protein (*Sorghum bicolor*)	10%	53.6	88%

5. **ORFs Available:**

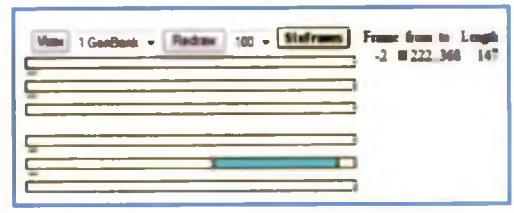
ORF Location	ORF length	Frame
222-368	147	-2

6. **Functional domains available :**

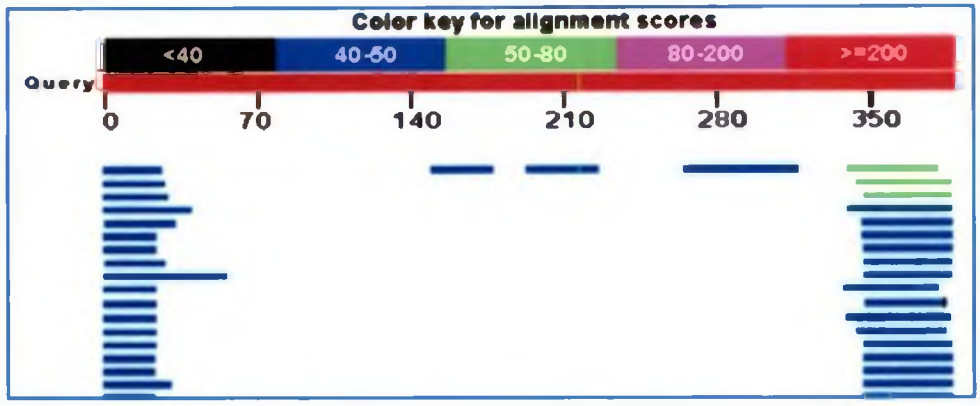
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	Nil	Nil

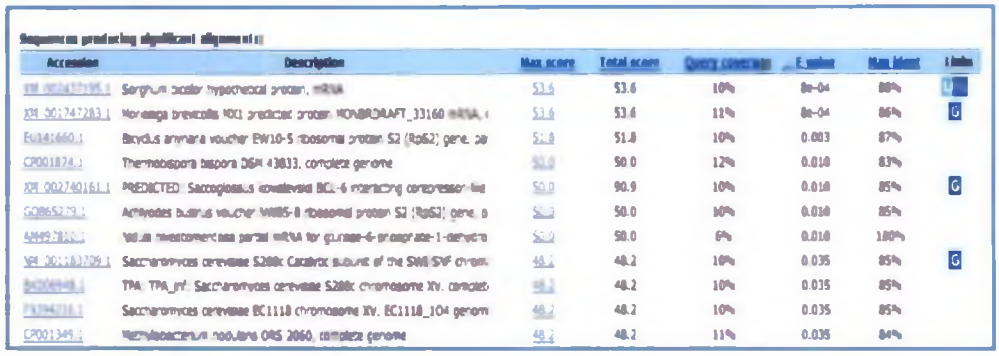


VecScreen output



Open reading frames in the cloned sequence





Blastn output

Fig. 28 Results of sequence data analysis for the EST-28 [PNK 1(2)76]

Table 31 Results of sequence data analysis for the EST-29 [PNK 1(2) 40]

1. **Initial Sequence length** : 583
2. **Vector sequence** : no match found
3. **Sequence length after Vector/Adaptor Screening:** 583
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AB182918.1	*Citrullus lanatus* mRNA for type -2 metalothionein	14%	5e-15	83%
AJ277599.2	Quercus suber mRNA for metalothionein protein	18%	5e-15	78%

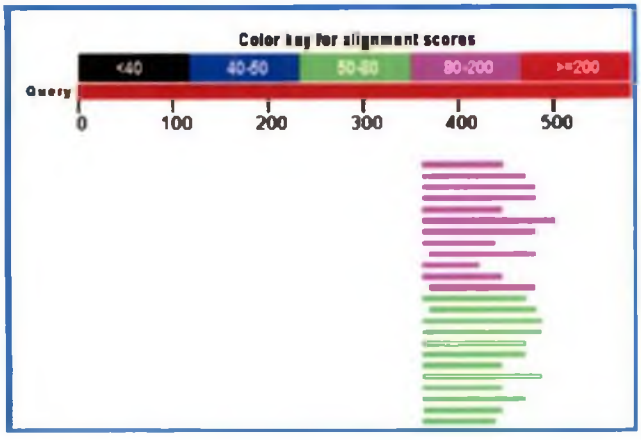
5. **ORFs Available:**

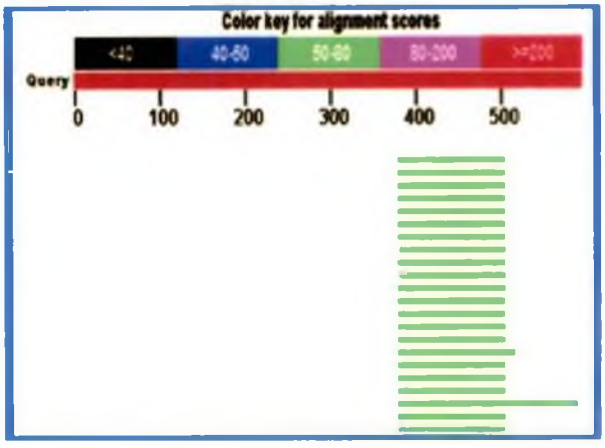
ORF Location	ORF length	Frame
2-223	222	-1

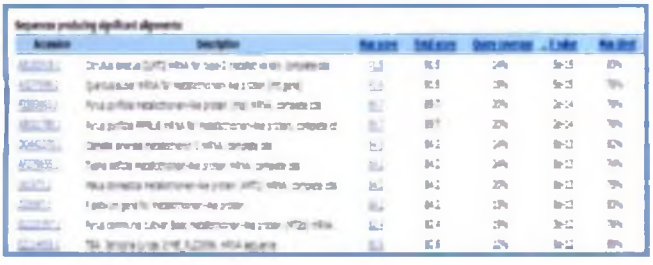
5. **Functional domains available :**

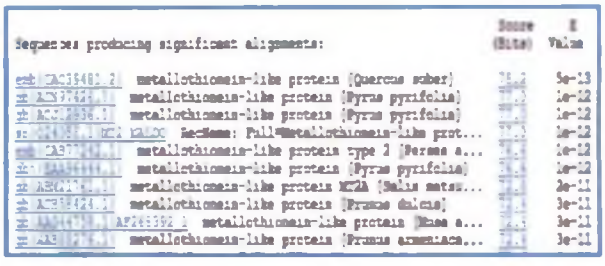
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	Nil	Nil

Open reading frames in the cloned sequence









Blastn output

Blastx output

Fig. 29 Results of sequence data analysis for the EST-29 [PNK 1(2)40]

Table 32 Results of sequence data analysis for EST-30 [PNK 1(2)61]

1. **Initial Sequence length** : 452
2. **Vector sequence** : 1 to 61, 405 to 452
3. **Sequence length after Vector/Adaptor Screening** : 344

4. **Blastn results :**

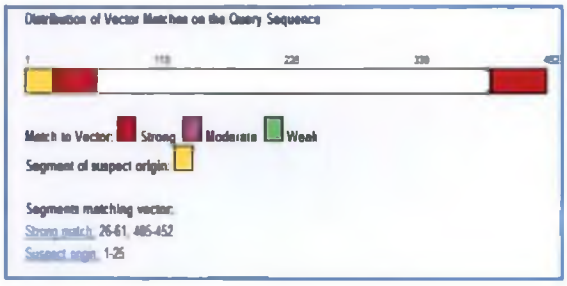
Accession No.	Description	Q.Coverage	Evalue	Max. Id.
Ct 028282.1	Poplar cDNA sequence	9%	7e-04	96%

5. **ORFs Available:**

ORF Location	ORF length	Frame
132-341	210	-3

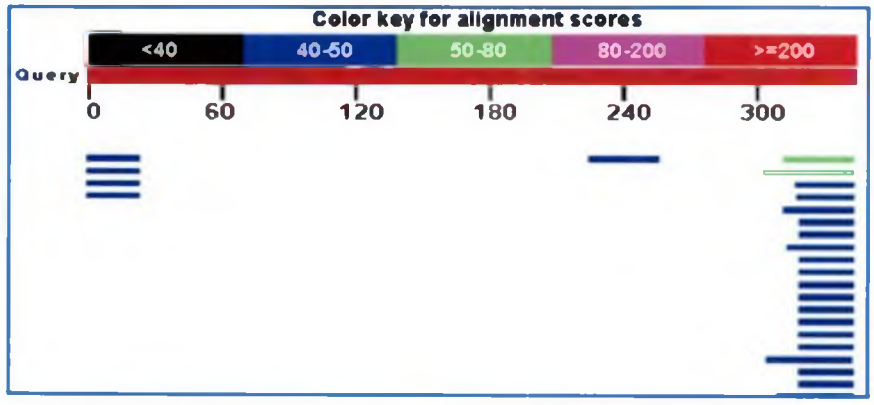
6. **Functional domains available :**

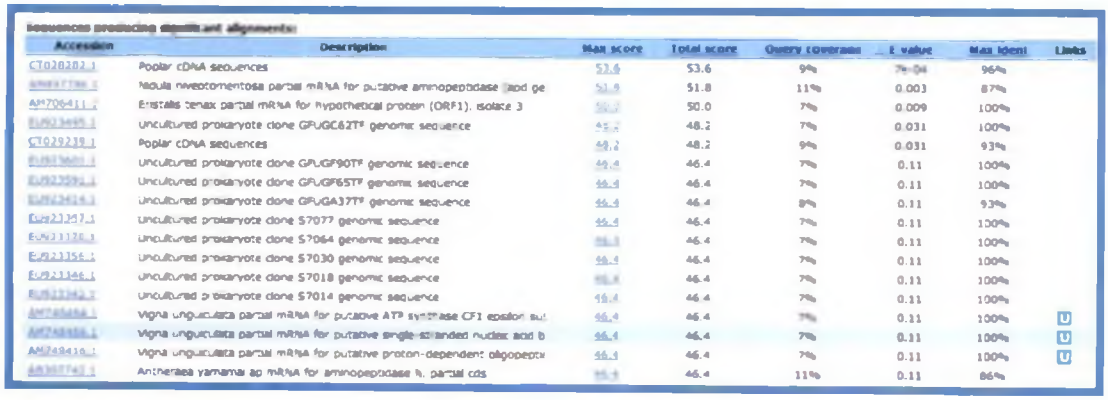
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	Nil	Nil



VecScreen output

Open reading frames in the cloned sequence





Blastn output

Fig. 30 Results of sequence data analysis for the EST-30 [PNK 1(2)61]

EST- 31

The sequence data obtained for the EST-31 was obtained from the clone PNK 1(2) 58 and it was of 441 bp in size. Vector screening indicated vector sequence from 1-36 bp. The results obtained are presented in Table 33 and Fig. 31. After vector and adaptor screening the total sequence obtained for further analysis was 405 bp.

5'ACGCAGTTCTAATTTACACAAGGATGATATATCTAATTAGGTCTAATACAAG
AGATTGATCAACTCTTGAATCTTAAGTTATGGGCTTGTGTTTGACAACTCTCTC
TGGTGAGTGTTACTATTGGCCTTCTTCACCGGACTCGGTGGCAGCGATTTTCCC
GCCGCCACCAGGCGACGCACAAAGACCTCCAGCCTCTCACCATTGGTCTCCTTC
AACAGCCGCTTGCCGCCAAATTGCACCACCTCGGCCGCGACCACGCTCTGAGTT
ATTTACAGTGTGCAATGTTAACCTTTGCCAGCATGTTTGCGTAAAATTTCATAT
CCATTTTACTGGCTCGTGTCTTTAGTTTTTTCACTCTCTCCAGTAGTCGATTTA
ATGGCTCATTGTCTGGTTCAATTTCCTTTGCAGCCAATAAGTCTATTTCTGCCA
AGTCCAGA3'

Blastn results indicated that 70 percent homology with the Hypothetical protein from *Vitis vinifera* and mRNA for metalothionein protein from *Quercus suber*. Blastx results showed similarity with unknown protein from *Vitis Vinifera* with bit score 60.8.

The sequence was found to have two ORFs with the longest one being 372 bp in length. CDD and Interproscan results indicated negative results.

EST- 32

The sequence data obtained for the EST-32 was obtained from the clone PNK 1(2) 46 and it was of 583bp in size. Vector screening indicated vector sequence from 553 to 583 bp. The results obtained are presented in Table 34 and Fig. 32. After vector and adaptor screening the total sequence obtained for further analysis was 552 bp.

5'ATTAGCGTGGTCGGGCCGAGGTACAAGGGTATCACTACCTTGCGTGAGCA
AAAAGGTTTTCTTTGGGATGACAAGTTTCACAGGGTTGAAGCTGATTATGAA
GTTTGGAGGGAGTTTATCAAGAAAAACGATTGGGCGGCTTCATTTAGAGAAA
AATCATTTCCCTTGTATGACAAGCTATCAATACTTTTTGTTAATTCCTCTGA
TATAGAAGCCATGGAAGCACGTTTTGAAGATAATGCTGAAGCGCTTCTTACT

Table 33 Results of sequence data analysis for the EST- 31 [PNK 1(2) 58]

1. **Initial Sequence length** : 441
2. **Vector sequence** : 1 to 36
3. **Sequence length after Vector/Adaptor Screening:** 405
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
XM 002277881.1	Hypothetical protein (*Vitis vinifera*)	42%	3e-07	72%
AM 454315.1	Quercus suber mRNA for metalothionein protein	42%	3e-07	72%

5. **Blastp Results:**

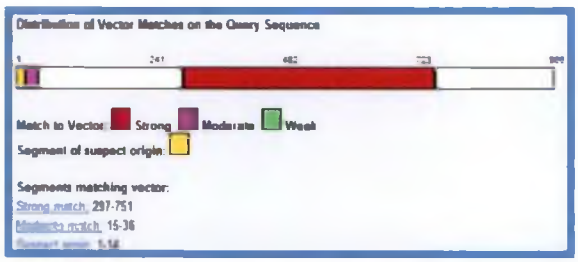
Accession No.	Description	Evalue	Score (Bits)
CTI 28689.1	Unknown protein (*Vitis vinifera*)	5e-08	60.8

5. **ORFs Available:**

ORF Location	ORF length	Frame
82-452	372	+3

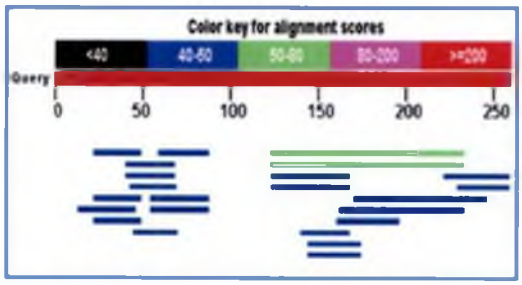
9. **Functional domains available :**

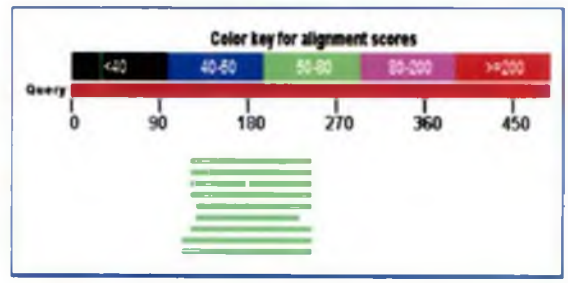
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	Nil	Nil



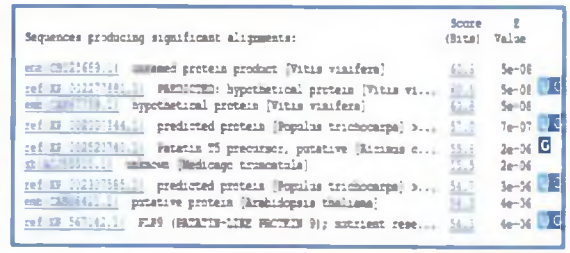
VecScreen output

Open reading frames in the cloned sequence





Blastn output



Blastx output

Fig. 31 Results of sequence data analysis for the EST-31 [PNK 1(2)58]

Table 34 Results of sequence data analysis for the EST-32 [PNK 1(2)46]

1. **Initial Sequence length** : 583
2. **Vector sequence** : 553 to 583
3. **Sequence length after Vector/Adaptor Screening** : 542
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
CT027578.11	Zebrafish DNA sequence	9%	1e-04	83%

5. **Blastp Results:**

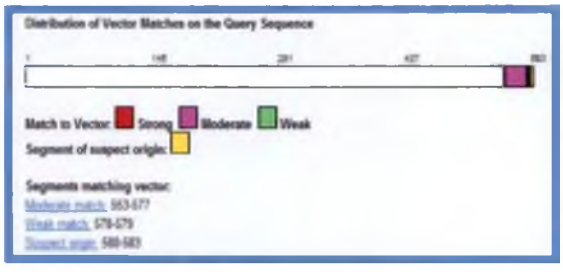
Accession No.	Description	Evalue	Score (Bits)
CAN71765.1|	Hypothetical protein (*Vitis vinifera*)	4e-04	47.8

6. **ORFs Available:**

ORF Location	ORF length	Frame
40-540	501	+1

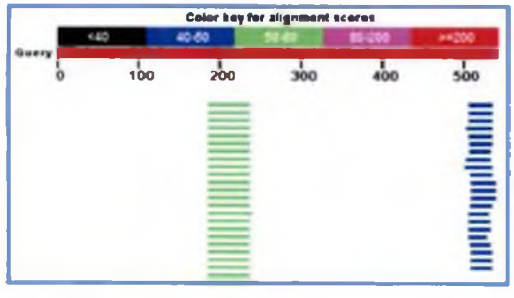
7. **Functional domains available :**

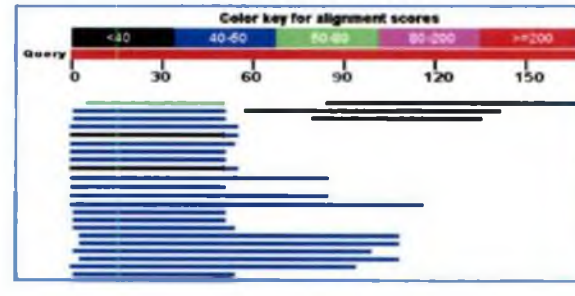
Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	Nil	Nil

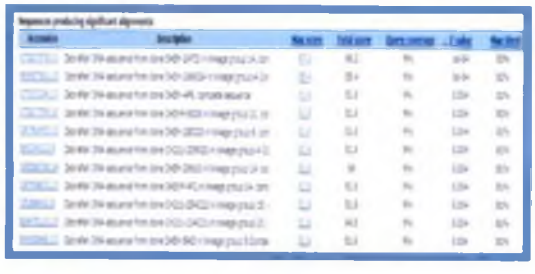


VecScreen output

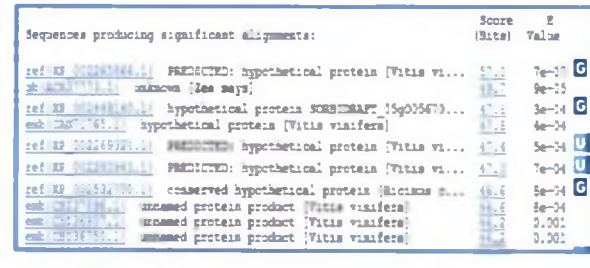
Open reading frames in the cloned sequence







Blastn output



Blastx output

Fig. 32 Results of sequence data analysis for the EST-32 [PNK 1 (2)46]

CAAATAGATTCAAGCGGTCATGGAGTCGATGCAGTAACAAAGCAGGATTATC
CACCCCTTGACACTGTTGATCCTGCTAGAAATGTAACAGAACTTCAGGCCAT
GCCTTCTCGCCGCCGCCGTCGATTCTCTCCTATAAAAACAGATGATCCCGCT
CTGCCATTGAGGCAGCGACAGAGGTTTTGTCCACCTCCCGAACAACCAGAGT
TTATGTCCAACTATGTGCGGGAGGTAACACATGCGATTGCAAGGGTACCTGC
CCGGGCGGCCGGTCGAAAATCACA3'

Blastn results indicated that 9 percent of the sequence showed 83 percent homology with the Zebrafish DNA sequence and Blastx indicated homology with the hypothetical protein from *Vitis vinifera* (bit score 47.8).

The sequence was found to have five ORFs with the longest one being 501 bp in length. The Interproscan and CDD analysis indicated negative results.

EST- 33

The sequence data obtained for the EST-33 was obtained from the clone PNK 1(2) 66 and it was of 779 bp in size. Vector screening indicated vector sequence from 1-61 bp. The results obtained are presented in Table 35 and Fig. 33. After vector and adaptor screening the total sequence obtained for further analysis was 718 bp.

5'TCGAGCGGCCGCCCGGGCAGGGTATTTGTGTGTTTCTTGTGAATCTGCNAAA
GATNCAAATNGAATGTATATCCCACATTAATGACCATATACTATTAATCTCGAA
AAAGTTCAGTCACGATAAATTGAAGAACTGATACTTAANGCGCATTATAAAACA
CATCACGCATAGTAAATACAAAACATAGAAACGAAACATACACCGAAAAANCCG
ACAGACAAAAGAGTGCCTCTCTAGACTCTNGTGAAACCAGACTTTTTATCTGTC
TAGTGCTACACATTAAACAGAACATGNCTTAAATAACAATAATGACACTTAGTA
AAAGTAACCTTTACAATATGAGATGNAAAACTAACCAATATACGCCCAAACGGA
GAAATTTGGAGCACGCACATAAGCCCAGGGACTGTCCACGTGTTTGTTTCTCTC
ATTTCACTCTATCTCCAATTANATGGCCCTTTCGCTAGCTATGGTCGAGTANCC
AAATAGCAGACTTGTCTCGCGATCCGGAGCGATACCTCGATACTCGCTATGGCT
TTATCCCCGCCCCCTGCTCGGCNTGTTCCCTCTGCATCGCGTCGGTGACACTGA
GCCATCCGTGTCCTGTCGAACCCTAGCGCCCTCTTTTGCCCCCTGGGTGTCCCA
ATGAGTGCGCCCGTGTGCACGTCCAATGACGAACCCGATTGGCGCCCCCATGAT
ATCTCGGCGCATATCTAG3'

Blastn results indicated that 5 percent of the sequence showed 86 percent homology with *Schistosoma mansoni*.

Table 35 Results of sequence data analysis for the EST-33 [PNK 1(2)66]

1. Initial Sequence length : 779
2. Vector sequence : 1 to 61
3. Sequence length after Vector/Adaptor Screening : 718

4. Blastn Results:

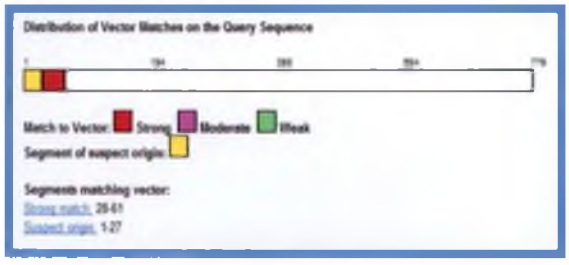
Accession No.	Description	Q.Coverage	Evalue	Max. Id.
FN 357688.1	Schistosoma mansoni	5%	0.24	86%

5. ORFs Available:

ORF Location	ORF length	Frame
453-650	198	+3

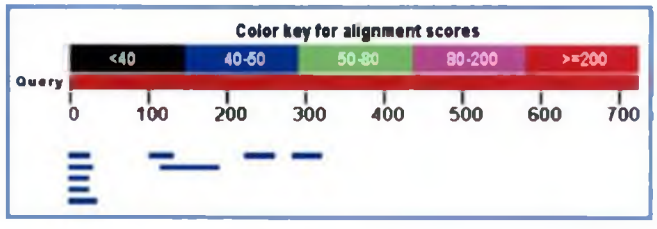
6. **Functional domains available :**

Tools	Accession no.	HMM	Domains details
IPS	-	-	Nil
CDD	-	-	Nil
pfam	-	nil	Nil



VecScreen output

Open reading frames in the cloned sequence



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
FN357688.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000397	46.4	46.4	5%	0.24	86%	
KN263366.1	Amborella trichopoda mRNA for ARF4 protein, short variant	44.6	44.6	4%	0.83	96%	
XM_745210.1	Aspergillus fumigatus Af293 PXA domain protein (AFUA_1G05540), partial	44.6	44.6	4%	0.83	90%	G
AJ270850.1	Mus musculus partial mRNA for hypothetical protein, clone mvx2011	44.6	44.6	4%	0.83	93%	U E G
AM706434.1	Entamoeba terax mRNA for hypothetical protein (ORF1), isolate 26	42.8	42.8	3%	2.9	96%	
AC138277.16	Mus musculus chromosome 1, clone RP23-6711, complete sequence	42.8	42.8	5%	2.9	85%	
AC137825.26	Medicago truncatula clone mth2-231l, complete sequence	42.8	42.8	10%	2.9	74%	
AF095771.1	Homo sapiens PTH-responsive osteosarcoma B1 protein (B1) mRNA, com	42.8	42.8	3%	2.9	92%	U E G
AB013375.1	Homo sapiens mRNA for SUI1 protein translation initiation factor	42.8	42.8	5%	2.9	84%	U E G

Blastn output

Fig. 33 Results of sequence data analysis for the EST-33 [PNK 1(2)66]

The sequence was found to have seven ORFs with the longest one being 198 bp in length. Interproscan and CDD analysis indicated negative results for this query sequence.

EST- 34

The sequence data obtained for the EST-34 was obtained from the clone PNK 1(2) 55 and it was of 583 bp in size. Vector screening indicated vector sequence from 1-53 bp. The results obtained are presented in Table 36 and Fig. 34. After vector and adaptor screening the total sequence obtained for further analysis was 530 bp.

5'GTGATTAGCGGCCGCCCGGGCAGGTTACAAAGACACACANAGAGTGTGTGCT
CTCTCCAACACAGATGTGTGAGAACACACACAACACATCTCATTTATTTGTGAG
AGAGAGGTGTGACACGGCTTATATATATCAGCAACATATATGAAGATGTCTCGC
AANACAGAAAAAAANCTACGAACACGAGGGGCCTTTAGAGAGGGCAAGAAAAAA
ANTTTTTTATNCGGTTCTCTCTTTGTCTCACACAAAAACAAAAAGTGTTTTACA
CACCAACACACACAACCCTTTAATATACTCAAAGAAATATTTCAAGAAAACTCT
TCTCAAACACATNTAACCNCGCTGTCTTTCGTGTTTAACGCGTTTTAAAGAGTA
TANCCCTTGCGGNGCGCGGGCGAAACCCAAGCGCCCTTATATCTCCCGCGCGCG
GGGCCATGTGGGGCGGGGGCGCGGGAAGCACATTGTGTCGAAGCGTGTCGGGGC
CCCCAATTTTTCGCCCCCCATTAG3'

Blastn results indicated that 9 percent of the sequence showed 96 percent homology with the Poplar DNA sequence.

The sequence was found to have three ORFs with the longest one being 225 bp in length. Interproscan and CDD analysis reported negative results for this query sequence.

EST- 35

The sequence data obtained for the EST-35 was obtained from the clone PNK 1(2) 78 and it was of 565 bp in size.Vector screening indicated no vector sequence. The results obtained are presented in Table 37 and Fig. 35. The total sequence obtained used for further analysis as,

Table 36 Results of sequence data analysis for the EST-34 [PNK 1(2) 51]

1. **Initial Sequence length** : 583
2. **Vector sequence** : 1 to 53
3. **Sequence length after Vector/Adaptor Screening** : 530

4. **Blastn Results:**

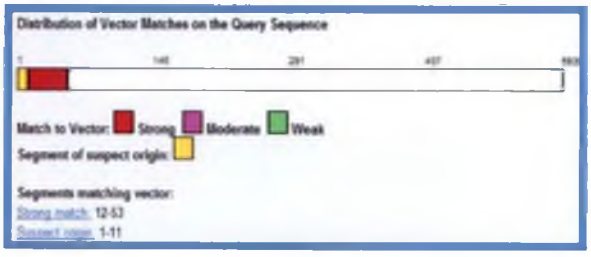
Accession No.	Description	Q.Coverage	Evalue	Max. Id.
Ct029573.1	Poplar DNA sequence	9%	2e-08	87%

5. **ORFs Available:**

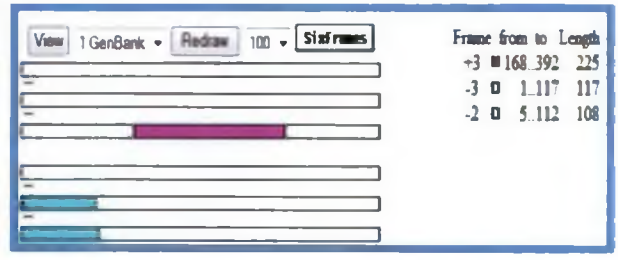
ORF Location	ORF length	Frame
168-392	225	+3

6. **Functional domains available :**

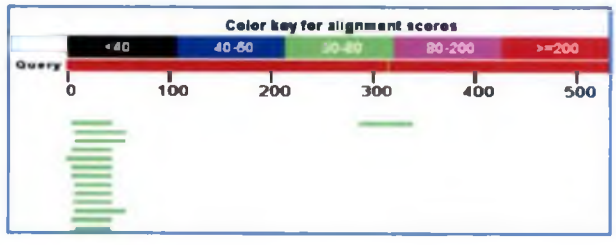
Tools	Accession no.	HMM	Domains details
IPS	-	-	nil
CDD	-	-	Nil
pfam	-	nil	Nil

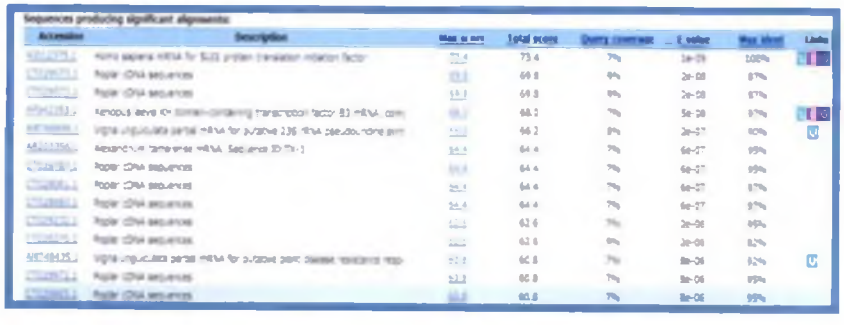


VecScreen output



Open reading frames in the cloned sequence





Blastn output

Fig. 34 Results of sequence data analysis for the EST-34 [PNK 1(2)51]

Table 37 Results of sequence data analysis for the EST-35 [PNK 1(2) 78]

1. **Initial Sequence length** : 565
2. **Vector sequence** : no match found
3. **Sequence length after Vector/Adaptor Screening:** 565
4. **Blastn Results:**

Accession No.	Description	Q.Coverage	Evalue	Max. Id.
AT972877.1	Synthetic construct RLS	17%	8e-13	96%

5. **ORFs Available:**

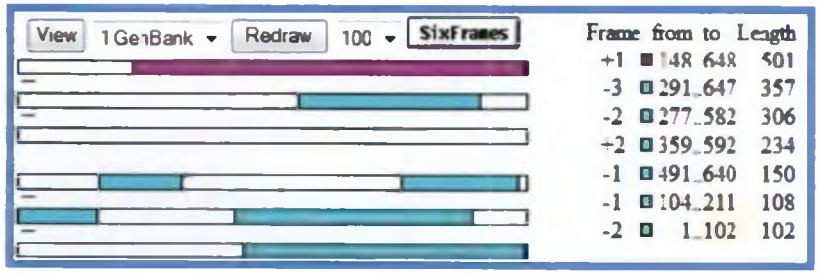
ORF Location	ORF length	Frame
148-648	501	+1

6. **Functional domains available :**

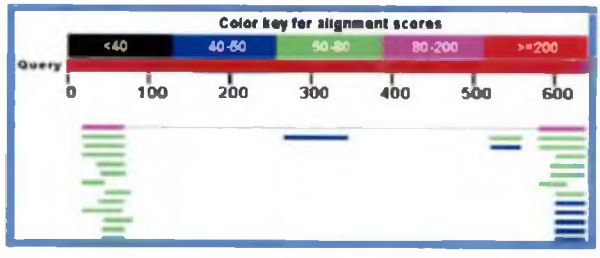
Tools	Accession no.	HMM	Domains details
CDD	-	-	Nil
pfam	-	nil	Nil

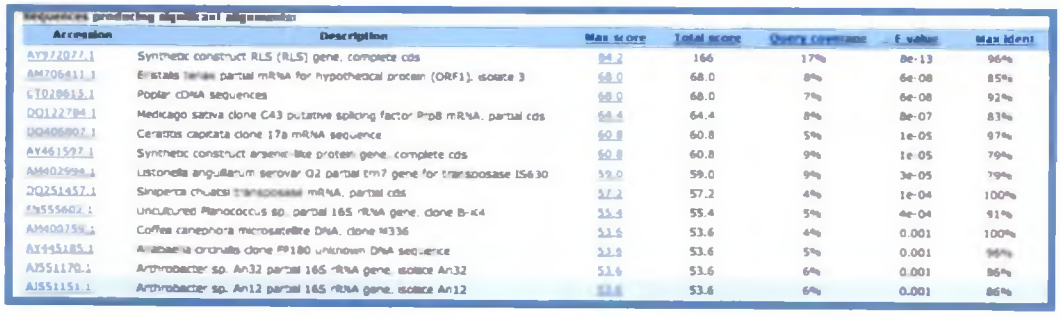
7. **Interproscan scan results:**

Database entry	Description
signalip	Signal peptide



Open reading frames in the cloned sequence





Accession	Description	Max score	Total score	Query coverage	E value	Max ident
AY972077.1	Synthetic construct RLS (RLS) gene, complete cds	84.2	166	17%	8e-13	96%
AM706411.1	Elaeis [illegible] partial mRNA for hypothetical protein (ORF1), isolate 3	68.0	68.0	8%	6e-08	85%
CT020615.1	Poplar cDNA sequences	68.0	68.0	7%	6e-08	92%
DQ122784.1	Medicago sativa clone C43 putative splicing factor Prp8 mRNA, partial cds	64.4	64.4	8%	8e-07	83%
DQ406807.1	Ceratos clapcata clone 17a mRNA sequence	60.8	60.8	5%	1e-05	97%
AY461597.1	Synthetic construct arsenic-like protein gene, complete cds	60.8	60.8	9%	1e-05	79%
AM402994.1	Listonella anguillarum serovar O2 partial trn7 gene for transposase IS630	59.0	59.0	9%	3e-05	79%
DQ251457.1	Sinipeta chuatsi transposase mRNA, partial cds	57.2	57.2	4%	1e-04	100%
AM555602.1	Uncultured Planococcus sp. partial 16S rRNA gene, clone B-K4	55.4	55.4	5%	4e-04	91%
AM400759.1	Coffea canephora microsatellite DNA, clone M336	53.6	53.6	4%	0.001	100%
AY445185.1	Arabidopsis cronulla clone P8180 unknown DNA sequence	53.6	53.6	5%	0.001	96%
AJ551170.1	Arthrobacter sp. An32 partial 16S rRNA gene, isolate An32	53.6	53.6	6%	0.001	86%
AJ551151.1	Arthrobacter sp. An12 partial 16S rRNA gene, isolate An12	53.6	53.6	6%	0.001	86%

Blastn output

Interproscan result

Fig. 35 Results of sequence data analysis for the EST-35 [PNK 1(2)78]

5' CCGTGCGACTGCTTCGTATAGGTCGACTGCAGGCGGCGCACTAGTGATTT
CGAGCGGCCGCCCGGGCAGGTACATGTATAACACTAAGTGCAAATCAAGCCA
CGTCTAAGGCAACTCTCCAGCTATACATTACACTTCTTGATATATAGGGTGC
CTCCAAAGAGAAACCAACTATGCAATATATTTCTAGAAGACCGCACATCCCT
TATCACAGACAGCTGCTTATCTCTTAGCACACAAACAGCTGCTCGTAGCAGC
TTGATGAAATTATGGTATCCCAATTATTGGAATGAAGGGGAAGTAGTAACTG
GTTGGAGCCAAGCCCCATTCGCGCGGCACAGGAAGATCGATGGTTTTAGAGT
ATTTGAAACCTGTAACCGAACCTTTGGCAGATGTGCACTTGGAGGAGAAGGT
GAGAGGAGTGGCGATTGTGAAAGAATTTGCTTCTTTGGTCTTGACAACCTTG
GAAATGAGTGCTTTCTTAGAGGAACAAAGCTGCTTGGTGCCGCCGAGAAGCA
TGGCGAAGCAGTTGGGGGAAGTGCTCTTCGGGAAGCTCGGTCTCGAAGTAGC
CATCTTTGTCTGACCTCGGCCGCGACCACGCTNATCCCGCGGGCATGGCGGN
CGGGAACATGCGAAGTC3'

Blastn analysis indicated that 17 percent of the sequence showed 96 percent homology with the Synthetic constructs RLS.

The sequence was found to have seven ORFs with the being 501 bp in length. Interproscan reported the presence of the signal peptide. CDD and pfam analysis indicated negative results.

Discussion

5. DISCUSSION

Expressed Sequence Tags (ESTs) are short, complementary DNA sequences usually 200 to 800 nucleotides long that represent the expressed portion of genes. Large-scale single-pass sequencing of cDNAs prepared from specific plant species or tissues has evolved as an efficient gene-discovery tool that can be used to identify novel cDNAs encoding enzymes of specific plant metabolic pathways. Collections of expressed sequence tags from metabolically active tissues can provide quantitative estimates of gene expression levels and thus are being exploited to unravel plant metabolic and regulatory networks (Ohlrogge and Benning, 2000).

A large number of ESTs have been deposited in public databases such as the US National Center for Biotechnology Information (NCBI) database for expressed sequence tags (dbEST) and also in private domains. Large-scale EST sequencing has emerged as a direct way to catalogue the expressed genes in rice and many other plant species. They serve as central resources in studies of global gene expression through high-density microarrays and analysis of complex traits such as drought and salinity tolerance governed by multiple genes, as demonstrated in *Arabidopsis*, barley and rice (Kawasaki *et al.* 2001; Seki *et al.* 2001; Öztürk *et al.* 2002). The different groups generated a large number of 'UniGene' sets of EST collections for different cereals such as barley (Michalek *et al.* 2002), maize (Fernandes *et al.* 2002) and wheat (Echenique *et al.* 2002) that will serve as powerful tools for comparative analysis and study of evolution of gene families among cereals (Bennetzen 2002).

Several works have been reported earlier on EST development for identification of genes induced during abiotic in secondary metabolism related to bud dormancy stress (Iturriaga *et al.*, 2006; Jin *et al.*, 2006; Diab *et al.*, 2008).

The present study was aimed at screening of cDNA library developed using Suppression Subtractive Hybridization (SSH) and characterization of differentially expressed genes (ESTs) in black pepper var: Kalluvally with special reference to drought tolerance. Present scenario of climate change projects drought as the most significant limiting factor in crop production (Bacsi *et al.*, 2009) and hence the study

taken up in black pepper, the prestigious spice crop of the country is of great relevance and significance. The cDNA library made available for the study was the one developed through subtracting the cDNA of water stressed plants against the cDNA of normal ones so as to fish out the stress induced genes in the resistant genotype, Kalluvally (Kushwah, 2008).

5.1 Maintenance of cDNA library

The differentially expressed genes developed through Suppression subtractive hybridization specific to drought tolerance in black pepper were cloned in pGEMT vector and maintained in *E. coli* (JM 109) bacteria on LB/amp. medium.

The ESTs developed through SSH are normally fragments of cDNA since the procedure include *Ras*I digestion and PCR amplication involving specific adaptors. Hence the fragments could be efficiently cloned in pGEMT vector which is universally used for the cloning of PCR products (Singh *et al.*, 2009).

The revived clones (215) were conserved as stabs and glycerol cultures for further studies.

5.2 Colony hybridization

There is every chance for having duplicates of the same fragments in an EST library. In order to avoid sequencing of such clones colony hybridization was performed in the present study.

The inserts derived from the already characterized 20 EST clones served as the probe. Colony hybridization is a usual practice to screen out desirable ones based on complementary of the probe with the sequences in the colony (Michael and David, 1975; Diab *et al.*, 2004).

The ^{32}P labeled ATP served to provide signals which was detected through the Phosphoimager FLA-5100 in the present study. The imaging plate (IP) is quite a new radiation energy memory type two-dimensional senor, which has an image recording layer consisting of polyester base material densely coated with accelerated

phosphorescent florescent material of fine crystals. An IP accumulates and stores radiation while it is exposed. The recording surface of an exposed IP is scanned with a laser beam inside the scanner and emits fluorescent light according to the exposure level. A photo-multiplier tube (PMT) detects the fluorescent light and converts it into electric signals. A radiation image recorded on the IP during exposure read as digital image information at the maximum resolution of 25µm/ pixel (40 pixels/mm) and was recorded in the analyzer unit (Fuji, FLA-5100).

The colony hybridization helped to detect 78 out of 215 clones as duplicates of already characterized ones. The presence of insert in the rest of the clones were confirmed through colony PCR with T7 and SP6 specific primers (Plate 6). The multiple cloning sites in pGEMT vector are flanked with T7 and SP6 forward and reverse universal primers (Promega, USA). Hence the size of region amplified in between these primers can give amplicons for the size of the insert present in that particular clone. From among 178 novel colonies, 35 clones which possessed larger inserts were selected through colony PCR for further analysis.

5.3 Sequence analysis

The sequence data for the insert shall possess the adaptor sequences and part of vector sequence which may mislead the interpretation of data. Thus any type of sequence analysis warrants deletion of such sequences which are not part of the insert. Different bioinformatics tools were used for editing the raw data obtained from the sequencing firm (Bioserve Biotechnology Pvt. Ltd. India) and to further analyse the sequence for final interpretation.

5.4 *In Silico* analysis

Broadly defined, ESTs are single-pass partial sequences of cDNA clones sequenced from one end. They are short, unedited, randomly selected single-pass sequence reads derived from cDNA libraries, providing a low-cost alternative to whole genome sequencing and gives a glimpse of the transcriptome of an organism at various stages of development or under different experimental conditions. ESTs are highly error prone and require several computational methods for pre-processing,

clustering, assembly and annotation to yield biological information (Nagaraj *et al.*, 2007).

5.4.1 Editing the sequence data

This software developed by NCBI for quickly identifying segments of a nucleic acid sequence that may be of vector origin. It helps to combat the problem of vector contamination in public sequence databases. If vector sequences are not removed, it may lead to erroneous conclusions about the biological significance of the sequence, time waste and effort in analysis of contaminated sequence, delay the release of the sequence in a public database and also pollute public databases with contaminated sequence. The programme is available at http://www. ncbi.nlm. nih.gov/ VecScreen/VecScreen.html. Editing after VecScreen is performed through Bioedit tool.

5.4.2 Searching for similarity

Sequence analysis is performed to trace out the biological significance of the nucleotide sequences. The basic step in any similarity search is an alignment of two or more sequences. This is done by searching the database. The search provides a list of database sequences with which a query sequence can be aligned. A common reason for performing a database search is to find a related gene/protein. A matched gene (or any other sequence) may provide a clue to its function. Sequence comparison is the most powerful and reliable method to determine evolutionary relationship between genes.

Similarity searching methods involve matching of the query sequences to sequences deposited in the database. A similarity score is calculated by measuring the closeness between the residues. The closeness is nothing but the number of nucleotide bases or amino acid residues that are similar between the compared sequences. Better and faster methods which look for local similarities between sequences were evolved in due course. The most widely used tools are BLAST and FASTA (Pearson, 2003). These two algorithms function in a similar fashion. They differ only in ranking the similarity or the differences between the sequences

(Altschul *et al.*, 1997). The sequences derived in the present study were analyzed for similarity using the BLAST tools.

5.4.2.1 BLAST (Basic Local Alignment Search Tool)

BLAST program uses a heuristic approach (a mathematical problem solving method). BLAST first creates a dictionary of three or four letter words from the query sequence (depending on Word size) and searches for matching words in the database sequence. A substitution matrix is then constructed for assigning scores. It matches its dictionary words one after the other and looks for match score (Altschul *et al.*, 1997).

The ESTs were annotated on the basis of the existing annotation by non-redundant databases at the NCBI using BLASTN and BLASTX. Homologies that showed e-value less than 1e-08 with more than 100 nucleotides were considered significant (Yang *et al.*, 1999).

Blast programs:

BLASTn - Analyzes nucleotide query sequence against a nucleotide sequence database. It provides identity status, query coverage and E-value with respect to the accessions already available in the database. E-value relates to the probability of the alignment occurring by chance and It is a statistical calculation based on the quality of alignment (the score) and the size of the database. Lower the E-value, better is the homology. An E-value of 1e-3 means that there is a 0.001 chance that alignment would exist in the database by chance, that is, if the database contains 10000 sequences, then we might expect that alignment to occur maybe 10 times..

BLASTx - Analyzes translation products of a nucleotide query sequence against a protein sequence database. The output indicates E-value and bit score in relation to the accession referred. The bit score is a normalized score expressed in bits that helps to estimate the magnitude of the search space. Higher the bit score, better the homology.

BLASTp – Analyzes protein sequence against a protein sequence database. The output is same as in BLASTx.

5.4.3 ORF search

An open reading frame (ORF) is a portion of an organism's genome which contains a sequence of bases that could potentially encode a protein. The start-points and end-points of a given ORF are not equivalent to the ends of the mRNA, but they are usually contained within the mRNA sequence. In a gene, ORFs are located between the start-code sequence (initiation codon) and the stop-code sequence (termination codon).

The ORF Finder is a graphical analysis tool available at the homepage of NCBI, which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database. This tool identifies all open reading frames using the standard or alternative genetic codes. The deduced amino acid sequence can be saved in various formats and searched against the sequence database using the WWW BLAST server. The ORF Finder is helpful in preparing complete and accurate sequence submissions.

5.4.4 Conserved Domain Search

The Conserved Domain Database (CDD) is compilation of the multiple sequence alignments representing protein domains conserved in molecular evolution. It has been populated with alignment data from the public collections Pfam and SMART, as well as with contributions from colleagues at NCBI. CDD alignments are linked to protein sequence and structure data in Entrez. CDD can be accessed at http://www.ncbi.nlm.nih.gob.structure/cdd/cdd.shtml.

Protein query sequences may be compared against databases of position specific score matrices derived from alignments in CDD search.

Protein domains may be thought of proteins structural and functional building blocks, dividing the primary and tertiary structure of a chain into distinct units.

Domains are also mobile genetic units, rearranging in various combinations throughout the molecular evolution of proteins (Aron *et al.*, 2001).

5.4.5 InterProScan Search

InterProScan (Zdobnov and Apweiler, 2001) is a tool that combines different protein signature recognition methods from the InterPro (Mulder *et al.*, 2005) consortium member databases into one resource. Protein as well as DNA sequences can be analysed with the web based version (http:// www.ebi.ac.uk/InterProScan/).

5.5 Interpretation of sequence data after *in silico* analysis

The sequence data obtained for all the 35 clones were edited to remove the vector and adaptor sequences and were named as EST 1 to 35 related to water stress in black pepper.

The ESTs were analyzed using different bioinformatics tools for further characterization and summarized results are presented in table 38. The 35 EST sequences ranged between 344 and 807 bp and were good enough for further analysis. All the 35 sequences indicated presence of one or more open reading frame which confirmed the expression status of the genes evaluated. The presence of ORFs confirms them as coding sequences, thus forming part of expressed gene (Flavia *et al.*, 2007).

The homology search with different Blast tools(X, N & P) revealed 30 to 100 per cent identity with several reported sequences in the database. Twelve out of 35 clones were found to have identity with different genes involved in water stress management in crop plants (Table 38). Such genes included Protein kinase, Eukaryotic peptide chain release factor, Glutamine synthetase, Transcription factor, Aspartokinase-homoserine dehydrogenase, Cysteine protease, Catalase isozymes, Protein serine/threonin kinase, NAD-dependent epimerase/dehydratase, Peptidyl prolyl isomerase and Ribonuclease. It was interesting to observe homology to such genes reported in drought tolerant crop like *Ricinnus communis*.

Table 38 Consolidated results of *in silico* analysis of ESTs developed from black pepper var. Kalluvally

Sl. No	EST no, Clone id & size (bp)	Blastn, Blastx & Blastp output						ORFs available	Length of longest ORF (bp)	CDD	Pfam	IPS	Whether related to water stress
		Description	Crop	Coverage (%)	Identity	E value	Bit score						
1.	EST-1 [PNK 1(2) 32] (608)	Protein kinase	*Ricinus communis, Zea mays*	>89	>80	1e-54 to 1e-83	204 to 312	3	536	yes	40-201	yes	yes
2.	EST-2 [PNK 1(2)29] (678)	Eukaryotic peptide chain release factor	*Ricinus communis, Arabidopsis thaliana*	>90	>78	2e-70 to 9e-95	348 to 367	2	537	yes	1-133	yes	yes
3.	EST-3 [PNK 1(2)30] (441)	Glutamine synthetase	*Ricinus communis, Nicotiana attenuate*	>85	>80	1e-44 to 7e-44	128 to 164	5	327	yes	16-63	yes	yes
4.	EST-4 [PNK 1(2)65] (572)	Transcription factor	*Ricinus communis, Arabidopsis thaliana*	>25	>65	1e-16 to 8e-26	71.6 to 118	3	378	Nil	-	nil	yes
5.	EST-5 [PNK 1(2) 55] (380)	aspartokinase-homoserine dehydrogenase	*Oryza sativa, Glycine max*	>48	>80	1e-18 to 6e-04	95.9 to 125	2	344	yes	104-159	Yes	yes
6.	EST-6 [PNK 3] (539)	Cysteine protease	*Nicotina tabacum, Zea mays*	>75	>78	1e-63 to 6e-25	117 to 118	6	210	yes	166-220	Yes	yes
7.	EST-7 [PNK 5] (575)	Catalase isozymes	*Oryza sativa*	>75	>75	1e-57 to 8e-66	225	5	396	yes	346-383	Yes	yes
8.	EST-8 [PNK 21] (721)	Protein serine/threonin kinase	*lophopyran elongation, Ricinus communis*	>70	>55	0.007 to 2.2	35.4 to 43.5	4	264	nil	-	Nil	yes

Sl. No	EST no, Clone id & size (bp	Blastn, Blastx & Blastp output						ORFs available	Length of longest ORF (bp)	CDD	Pfam	IPS	Whether related to water stress
		Description	Crop	Coverage (%)	Identity	E value	Bit score						
9.	EST-9 [PNK 1(2) 48] (380)	NAD-dependent epimerase/dehydratase	*Zea mays*	>95	>80	2e-43 to 3e-124	167 to 178	4	339	yes	nil	Yes	yes
10.	EST-10 [PNK 1(2)36] (577)	Peptidyl prolyl isomerase	*Ricinus communis, Arabidopsis thaliana*	>25	>70	e-14 to 5e-18	94.7 to 96.7	3	339	yes	6-34	Yes	Yes
11.	EST-11 [PNK 1(2)33] (529)	Ribonuclease	*Zea mays*	>5	>85	8e-17 to 2e-14	80.5 to 90.1	3	240	yes	72-189	Yes	Yes
12	EST-12 [PNK 1(2) 41] (754)	Protein kinase	*Trichomonas vaginalis*	>5	>85	6.2	35	6	366	no	-	No	Yes
13.	EST-13 [PNK 1(2) 56] (619)	Conserved hypothetical protein	*Leptospirillum sp, Mycobacterium tuberculosis*	>5	>95	6.7 to 9.4	167 to 178	3	414	Nil	Nil	Nil	?
14.	EST-14 [PNK 1(2)39] (486)	Binding / catalytic/ coenzyme binding protein	*Arabidopsis thaliana, Zea mays*	>95	>70	1e-36 to 6e-54	133 to 213	4	339	yes	Nil	yes	?
15.	EST-15 [PNK 1(2)57] (588)	hypothetical protein	*Zea mays*	>35	>75	1e-25 to 3e-32	133 to 144	6	543	Nil	Nil	Yes (Signal peptide)	?
16	EST-16 [PNK 1(2) 47] (626)	RNA binding / endoribonuclease/ ribonuclease	*Arabidopsis thaliana*	>10	>75	2e-10 to 7e-07	33.5 to 50.1	6	366	Nil	Nil	Nil	?
17.	EST-17 [PNK 1(2) 54] (444)	Auxin responsive protein	*Zea mays*	>20	>80	1e-17 to 9e-17	78 to 86	5	164	Nil	Nil	Nil	?

Sl. No	EST no, Clone id & size (bp	Blastn, Blastx & Blastp output						ORFs available	Length of longest ORF (bp)	CDD	Pfam	IPS	Whether related to water stress
		Description	Crop	Coverage (%)	Identity	E value	Bit score						
18.	EST-18 [PNK 1(2)44] (461)	Transcription regulator (NOT2,3,5) family protein	*Arabidopsis thaliana, Vitis vinifera*	>80	>65	1e-28 to 9e-34	127 to 186	4	399	yes	2-86	yes	?
19.	EST-19 [PNK 1(2)74] (645)	Pre- mRNA splicing factor	*Zea mays*	>45	>75	3e-56 to 6e-72	199 to 205	7	261	Nil	Nil	Nil	?
20.	EST-20 [PNK 1(2) 37] (722)	peptidylproyl isomerase	*Ricinus communis, Triticum astivum*	>20	>70	2e-08 to 8e-21	85.5 to 88.6	3	396	Yes (Binding protein)	Nil	Nil	?
21.	EST-21 [PNK 1(2) 73] (499)	Transcriptional activator	*Ricinus communis*	>5	100	5e-15 to 9e-14	46.1 to 50.0	5	465	Nil	Nil	Yes (Zinc finger)	?
22.	EST-22 [PNK 1(2)31] (368)	Jacalin-like lectin	*Ananas comosus, Plantago major*	>5	100	2e-08 to 9e-10	40 to 48.1	3	216	yes	2-86	yes	?
23.	EST-23 [PNK 1(2)85] (584)	hypothetical protein	*Zea mays*	>35	>75	1e-41 to 2e-32	142	7	291	Nil	Nil	Yes (thmm)	?
24.	EST-24 [PNK 24] (731)	Hypothetical protein	*Vitis vinifera*	>55	>50	1e-53 to 8e-59	175 to 204	7	516	Yes (TPR family)	Nil	Yes (signal peptide)	?
25.	EST-25 [PNK 1(2) 42] (502)	Predicted protein	*Ricinus communis*	>30	>40	4.8	-	1	162	Nil	Nil	Yes (signal peptide)	?
26.	EST-26 [PNK 1(69] (807	Cell elongation protein	*Ricinus communis*	>45	>70	1e-24 to 6e-60	95.1 to 104	7	327	Nil	Nil	Nil	?

Sl. No	EST no, Clone id & size (bp	Blastn, Blastx & Blastp output						ORFs available	Length of longest ORF (bp)	CDD	Pfam	IPS	Whether related to water stress
		Description	Crop	Coverage (%)	Identity	E value	Bit score						
27.	EST-27 [PNK 1(2)70] (670)	Transcriptional activator	*Ricinus communis, Glycine max*	>35	>75	1e-41 to 2e-32	142	7	291	Nil	Nil	Yes (thmm)	?
28.	EST-28 [PNK 1(2)76] (389)	Hypothetical protein	*Sorghum bicolor*	>9	>85	53.6	-	1	147	nil	Nil	Nil	?
29.	EST-29 [PNK 1(2) 40] (583)	mRNA for type -2 metalothionein	*Citrullus lanatus*	>12	>75	5e-15	-	5	222	Nil	Nil	Yes	?
30.	EST-30 [PNK 1(61] (344)	cDNA sequence	*Poplar sp.*	>8	>95	7e-04	-	4	210	Nil	Nil	Nil	?
31.	EST-31 [PNK 1(2)58] (405)	Hypothetical protein	*Vitis vinifera*	>40	>80	1e-04 to 4e-04	47.8	2	372	Nil	Nil	Yes	?
32.	EST-32 [PNK 1(2)46] (542)	DNA sequence	Zebrafish	>8	>80	1e-4	-	5	501	Nil	Nil	Nil	?
33.	EST-33 [PNK 1(2) 66] (718)	Schistosoma mansoni	-	5	86	0.24	-	7	198	Nil	Nil	Yes	?
34.	EST-34 [PNK 1(51] (530)	DNA sequence	-	9	87	2e-08	-	3	225	Nil	Nil	Nil	?
35.	EST-35 [PNK 1(2)78] (565)	Synthetic construct RLS	-	17	96	8e-13	-	7	501	Nil	Nil	Yes (signal peptide)	?

5.6.1 Water stress related genes

5.6.1.1 Protein kinase

Protein kinase is reported to be involved in signal transduction and at least five signal transduction pathways are controlled by these genes: two are dependent on abscisic acid (ABA), and the others are ABA-independent. A novel *cis*-acting element involved in one of the ABA-independent signal transduction pathways has been identified. In addition, a number of genes for protein kinases and transcription factors thought to be involved in the stress signal transduction cascades have been reported to be induced by environmental stresses (Shinozaki, K. and Yamaguchi-Shinozakif, K. 1996).

Involvement of various kinases and phosphatases in stress signal transduction has been well documented in numerous genetic and biochemical studies (Hirt, 2000). Among the kinases there are serine/threonine kinase, receptor related kinase and protein phosphatases that are involved in transcriptional regulation of downstream stress responsive genes.

The EST-8 showed over 55 percent homology with Serine/theronine kinase reported in *Ricinus communis* which is a drought tolerant plant. Thus gene fragments confirms it have significance with drought.

The sequence analysed in black pepper showed over 80 per cent homology in plants like *Ricinus communis* and *Zea mays*.

When the sequence (EST-1 and EST-12) was subjected to Blastn, Balstx and Blastp homology search, revealed the identity present with protein kinase from different plants like *Arabidopsis thaliana*, *Zea mays* with highest bit score (311). Similar results were obtained in rice described by Zhou, *et al.*, 2007. The sequence had three ORFs with the longest one 536 bp in length.

The sequence also possess conserved domain. The homology with protein kinase of the ESTs analyse was confirmed through the results obtained on Pfam and InterProScan.

5.6.1.2 Eukaryotic peptide chain release factor (*epcrf*)

The sequence (EST-2) has 78 per cent similarity with the reported sequence for *epcrf* in *Ricinus communis* and *Arabidopsis thaliana* with 367 bit score and 2e-70 evalue. It also possess conserved domain and was also confirmed with InterProScan and Pfam. The sequence has two ORFs with longest being 537 bp.

. Eukaryotic peptide chain release factor terminates the biosynthesis by recognizing stop codons at the A site of the ribosome and stimulating tRNA bond hydrolysis at the peptidyl transferase studied in *Hordeum vulgare* (Peiguo *et al.*, 2009).

5.6.1.3 Glutamine synthetase(GS)

The sequence (EST-3) analysed in black pepper showed over 80 per cent homology with GS reported in *Ricinus communis, Cucumis melo* and *Nicotiana.* The sequence also possessed conserved domain. The homology with Glutamine synthetase of the EST analysed was further confirmed through the results obtained in Pfam and InterProScan.

Glutamine synthetase is reported to be involved in nitrogen metabolism during drought condition and ammonia released as a result of photorespiration and the breakdown of proteins and nitrogen transport compounds. GS is distributed in different subcellular locations (chloroplast and cytoplasm) and in different tissues and organs (Miflin and Habash 2001).

5.6.1.4 Transcription factor

The sequence EST-4, EST-21 and EST-18; analysed in black pepper showed over 65 per cent homology with transcription factors in different crops *Ricinus communis* and *Arabidopsis thaliana* with 118 bit score. The sequence also found to have three ORFs with longest being 537 bp.

These factors belong to variable families including AP2, MYB, HB and Zinc finger. These factors could also be divided into several groups depending on their expression patterns.

The expression of different transcription factors play significant role in common or organ specific gene expression in response to drought conditions (Zhou *et al.*, 2007).

5.6.1.5 Aspartokinase-homoserine dehydrogenase (*akthr*)

The sequence analysed EST-5 exhibited over 80 per cent homology with *akthr* from different crop plants like rice and soybean with 125 bit score. It possess conserved domain and the homology with *akthr* of the EST analyse further confirmed through the results obtained in Pfam and InterProScan.

The aspartate kinase-homoserine dehydrogenase *(akthr)* sensitive to threonine feedback inhibition. Aspartate kinase is the first enzyme of the aspartate pathway leading to the synthesis of lysine, threonine, isoleucine and methionine, while homoserine dehydrogenase initiates the branch of the pathway for threonine synthesis. The expression pattern described by Zhu-Shimoni *et al.* (1997) of the first *akthr1* gene isolated from *Arabidopsis* by Ghislain *et al.* (1994) reveal some differences in the expression of these two genes that encode isozymes with a similar function.

5.6.1.6 Cystein protease

The sequence analysed as EST-6 has 78 per cent similarity with that reported in tobacco and maize with 118 bit score and 1e-63 evalue. It also possess conserved domain and was confirmed with InterProScan and Pfam. The sequence has six ORFs with longest being 210 bp.

Different cysteine proteases have been characterized in plants where they participate in various proteolysis activities. The best characterized plant cysteine proteases include actinidin (Kamphuis *et al.* 1985), papain (Cohen *et al.* 1986), and aleurain (Rogers *et al.* 1985). Various elements have been shown to cause

changes in the expression of the genes encoding cysteine proteases. These include environmental stresses such as drought (Koizumi *et al.* 1993) and developmental processes such as seed germination. It has been suggested that protease activity leading to protein degradation can be considered as a regulatory mechanism in plants (Callis 1995). It has been also reported that proteases could alter metabolism by increasing protein turnover rates or proteolytically activating specific proteins. Guerrero *et al.,* (1990) suggested that proteases may be required for degradation of polypeptides denatured because of cellular stress, or of storage proteins. The mobilized amino acids would then be available for the synthesis of new proteins in response to stress or for osmotic adjustment. Genes encoding cysteine proteases whose mRNA is induced by water stress have been isolated from pea and Arabidopsis (Koizumi *et al.* 1993).

5.6.1.7 Catalase

The sequence EST-7 has 75 per cent similarity with catalase reported in rice with 225 bit score and 1e-57 evalue. It also possess conserved domain and was confirmed with InterProScan and Pfam analysis. The sequence has five ORFs with longest one being 396 bp.

Hydrogen peroxide is eliminated by catalases (CAT) and ascorbate peroxidases (Chen and Asada, 1989; Scandalios *et al.*, 1997). These enzymes rapidly destroy the vast majority of H_2O_2 produced by metabolism, but they allow low steadystate levels to persist presumably to maintain redox signalling pathways (Noctor and Foyer, 1998).

Catalase is essential for the removal of H_2O_2 produced in the peroxisomes by photorespiration. Catalase activities decrease under conditions that suppress photorespiration, such as elevated CO2 (Azevedo *et al.*, 1998). The importance of CAT in photosynthetic cells is demonstrated by observations in CAT-deficient mutants (Kendall *et al.*, 1983).

5.6.1.8 NAD- dependent epimerase/dehydrogenase

A large family of proteins share a Rossmann-fold NAD(P)/NAD(P)(+) binding (NADB) domain. The NADB domain is found in numerous dehydrogenases of metabolic pathways such as glycolysis and many other redox enzymes. NAD binding involves numerous hydrogen-bonds and van der Waals contacts, in particular H-bonding of residues in a turn between the first strand and the subsequent helix of the Rossmann-fold topology.

The sequence EST-9 showed homology with NAD- dependent epimerase/ dehydrogenase with over 80 percent similarity in *Zea mays* with 178 bit score and also possessed conserved domain. The InterProScan also confirmed homology with this protein.

5.6.1.9 Peptidyl prolyl isomerase

The sequence analysed as EST-10 and EST-20 in black pepper showed over 70 percent homology with Peptidyl prolyl isomerase in *Ricinus communis* and *Arabidopsis thaliana* with bit score over 95.

Peptidyl prolyl isomerases (PPIas) are involved in inhibiting the activity of cylosporin-A during water stress condition (Sharma *et al.*, 2005). Up to 90 per cent of the PPIase activity was inhibited by cyclosporin-A. Maximal increase in specific PPIase activity in the 3-day old seedlings was observed in response to osmotic stress and it was transient in nature. The stress-induced enhancement in PPIase activity, depending upon tissue and stress treatment, was due to induction of cyclophilins as well as other PPIases.

5.6.1.10 Ribonuclease

The sequence EST-11 showed over 80 per cent homology with Ribonuclease in maize with 90.1 bit score. It possessed conserved domain and was also confirmed with the results obtained in Pfam and InterProScan.

Ribonuclease is an enzyme which degrades RNA. It is ubiquitous in living organisms and is exceptionally stable. By catalyzing the synthesis or degradation of RNA, two classes of enzymes control this flow. RNA synthesis is catalyzed by RNA polymerases and RNA degradation is catalyzed by RNA depolymerases, which are most often called ribonucleases.

Ribonucleasae interacts selectively and non-covalently with RNA molecules or a portion there of ribonuclease T2 activity. It is also involved as catalytic agent in endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides with 2',3'-cyclic phosphate intermediates.

5.6.1.11 Hypothetical proteins

The sequences from EST-13, EST-15, EST-23, EST-24, EST-28, EST-29 and EST-32 showed homology with hypothetical protein reported in different crops like *Ricinus communis, Vitis vinifera, Sorghum bicolor* and *Zea mays.* All the plants are drought tolerant ones. These sequences expressed during water stress in black pepper have similarity with seuences in other drought tolerant plants including Vitis which is a perennial climber like black pepper. Similar results were reported by Diab *et al.*, (2004) in barley.

These expressed sequences of black pepper though do not possess homology with any reported genes, are related to hypothetical coding sequences in different crops and may be unique to water stressed conditions.

5.6.1.12 other genes

EST-14, EST-16, EST-17, EST-19, EST-22, EST-25, EST-26, EST-30, EST-33, EST-34 and EST-35 showed identity with known sequences like Binding/catalytic/coenzyme binding protein, RNA binding/endoribonuclease, Auxin responsive protein, Pre-mRNA splicing factor, Jacalin-like lectin, Cell elongation protein, Metalothionein, Schistosoma mansoni and Synthetic construct RI.S. All the sequences possessed ORFs and showed over 80 per cent homology. However, their role in stress tolerance is yet to be discussed.

Through the present study, it was made possible to unravel 35 ESTs in black pepper, related to drought tolerance. The ESTs developed include seventeen clones with over 80 per cent homology with other reported genes out of which six were similar to stress related genes already reported. All the ESTs characterized possessed ORFs and many of them had CDDs highlighting their significance.

The results are highly useful in characterizing full length native genes that could be further exploited for crop improvement/ better management.

Future line of work:

- Study of expression levels through cDNA microarray or Real time- PCR
- ESTs developed can be used to isolate full length genes
- Data generated could be further utilised for better expression of these genes for drought management

Summary

6. SUMMARY

The present study entitled "Characterization of subtracted cDNA library for detecting Expressed Sequence Tags (ESTs) specific to drought tolerance in black pepper (*Piper nigrum* L.)" was carried out at the Centre for Plant Biotechnology and Molecular Biology (CPBMB), College of Horticulture (COH), Vellanikkara during the period from August, 2008 to June, 2010. The study was intended to characterize the novel ESTs clones in cDNA library developed through SSH in water stressed black pepper var. kalluvally. Salient findings of the study are summarized in this chapter.

1. The cDNA library containing 280 clones developed through SSH technique was used for the study.

2. Out of 280 clones, 215 were revived and maintained as stabs and glycerol cultures for further study.

3. Twenty probes were prepared from already characterized clones of different sizes (500 to 900 bp). These probes were used in the colony hybridization with the 215 clones.

4. Among the 215 colonies, 137 were unhybridized and detected as novel ones. Thirty five unhybridized colonies were selected for further analysis.

5. The selected clones were checked for the presence of insert by colony PCR using T7 and SP6 primers and were sequenced through out-sourcing.

6. The sequence data was obtained from the firm in the form of electropherogram and nucleotide sequence along with the clone id. The sequences thus obtained were referred as ESTs and were subjected to *In silico* analysis.

7. After vector screening and adaptor edition, the sequences were analysed using various bioinformatics tools such as Blastn, Blastx, Blastp, CDD, InterProScan, Pfam and ORF search.

8. ESTs-1 to 12 showed 55 to 85 per cent homology with already reported drought related protein from different crop plants. These included Protein kinase, Eukaryotic peptide chain release factor, Glutamine synthetase, Transcription factor, Aspartokinase-homoserine dehydrogenase, Cysteine protease, Catalase isozymes, Protein serine/threonin kinase, NAD-dependent epimerase/dehydratase, Peptidyl prolyl isomerase and Ribonuclease All the sequences possessed conserved domain and ORFs.

9. The sequences from EST-13, EST-15, EST-23, EST-24, EST-28, EST-29 and EST-32 showed homology with hypothetical protein reported in different crops like *Ricinus communis, Vitis vinifera, Sorghum bicolor* and *Zea mays.*

10. EST-14, EST-16, EST-17, EST-19, EST-22, EST-25, EST-26, EST-30, EST-33, EST-34 and EST-35 showed low identity with known sequences and therefore it was difficult to assign any role to these. But the presence of ORFs in all the sequences signifies that they code for various proteins. Homology was observed with catalytic/coenzyme binding protein, RNA binding/endoribonuclease, Auxin responsive protein, Pre-mRNA splicing factor, Jacalin-like lectin, Cell elongation protein, Metalothionein, Schistosoma mansoni and Synthetic construct RLS.

11. The results generated could be well exploited for characterization of full length native genes from black pepper.

References

REFERENCES

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. 1997. Role of Arabidopsis WC and MYB homologs in Drought and Abscisic Acid regulated gene expression. *The Plant Cell.* 9: 1859-1868

Ablett, E., Scaton, G., Scott, K., Shelton, D., Graham, M.W., Baverstock, P., Leu, L.S. and Henry, H. 2000. Analysis of grape ESTs: global gene expression patterns in leaf and berry. *Plant Sci.* 159: 87-95

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B. and Moreno, R.P. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Sci.* 252: 1651-1656

Almoguera, C. and Jordano, J. 1992. Developmental and environmental concurrent expression in sunflower dry seed stored low-molecular weight heat shock protein and lea mRNAs. *Plant Mol. Bio.* 19: 781-792

Altschul, S.F., Thomas, L., Madden, A.A., Schaffer, Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402

Anderberg, R.J. and Walker-Simmons, M.K. 1992. Isolation of wheat cDNA clone for an abscisic acid inducible transcript with homology to protein kinases. *Proc. Nat. Acad. Sci.* USA. 89: 10183-10187

*Araus, J.L., Bort, J., Steduto, P., Villegas, D., and Royo, C. 2003. Breeding cereals for Mediterranean conditions: Ecophysilogy clues for biotechnology application. *Ann Appl Biol.* 142:129–141

Aron, M.B., Anna, R.P., Benjamin, A.S., Paul, A.T., Lewis, Y.G. and Stephen, H.B. 2001. CDD: a database of conserved domain alignments with links to domain three- dimensional structure. *Bioinformatics* 12: 123-126

Azevedo, R.A., Alas, R.M., Smith, R.J. and Lea P.J. 1998. Response of antioxidant enzymes to transfer from elevated carbon dioxide to air and ozone fumigation, in the leaves and roots of wild-type and a catalase-deficient mutant of barley. *Physiologia Plantarum* 104:280–292

Bacsi, Z., Thornton P. K. and Dent J. B. 2009. Impacts of future climate change on Hungarian crop production: An application of crop growth simulation models. *Agric. Systems* 37(4): 435-450

Bartels, D. and Nelson, D.E. 1994. Approaches to improve stress tolerance using molecular genetics. *Plant Cell Environ.* 17: 659–667

*Bedeil, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J. and Bradford, K. 2005. Sorghum genome sequencing by methyl filtration. *Biol.* 3: 103-115

Bengstson, C., Larsson, S. and Liljenberg, C. 1978. Effects of water stress on cuticular transpiration rate and amount and composition of epicuticular wax in seedlings of six oat varieties. *Physiologia Plantarum.* 44: 319-324

Bennetzen J. 2002. Opening the door to comparative plant biology. *Sci.* 296: 60–62

Birnboim, H.C. and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA . *Nucleic Acids Res.* 7: 1513-1523

Blum, A. 1985. Breeding crop varieties for stress environment. *Plant Sci.* 2:199–238

Blum, A. 1988. Plant breeding for stress environments. CRC, Boca Raton, 1–223 p

Blum, A. 1996. Crop responses to drought and the interpretation of adaptation. *Plant Growth Regulation* 20: 135-148

Bohnert, H.J., Nelson, D.E. and Jensen, R.G. 1995. Adaptation to environmental stresses. *Plant Cell.* 7: 1099–1111

Bray, E.A. 1993. Molecular responses to water deficit. *Plant Physiol.* 103: 1035-1040

Callis, J. 1995. Regulation of protein degradation. *Plant Cell* 7: 845–857

Ceceарelli, S. and Grando, S. 1996. Drought as a challenge for the plant breeder. *Plang growth regulation.* 20: 149-155

Chandra Babu, R., Zhang, J., Blom, A., Mod, T.H.D., Wue, R. and Ngiyen, H.T. 2001. HVA 1, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. *Plant Sci.* 166: 855-862

Chang, Q.S., Zhov, R.H., Kong, K.Y., Lu, Z.L. and Jia, J.Z. 2006. Identification of Differentially expressed genes during anther abortion of Taigo Genic Mala sterile wheat by combining suppression subtractive hybridization and cDNA array. *J. of Integrative Plant Bio.* 48(11): 1348-1354

Chempakam, B., Kasturi, B.K. and Rajagopal, V. 1993. Lipid peroxidation in reletion to drought tolerance in coconut (*Cocos nocifera* L.) *Plant Physio. Biochem.* 20: 5-10

*Chen, G, and Asada K. 1989. Ascorbate peroxidase in tea leaves: occurrence of two isozymes and the differences in their enzymatic and molecular properties. *Plant Cell Physiol.* 30: 987–998

Chrispeels, M.J. and Agre, P. 1994. Water channel proteins of plant and animal cells. *Trends Biochem. Sci.* 19: 421-425

Close, T.J., Fenton, R.D. and Moonan, F. 1993. A view of plant dehydrins using antibodies specific to the carboxy-terminal peptide. *Plant Mol Biol.* 23:279–285

Cohen, L.W., Coghlan, V.M., and Dihel, L.C. 1986. Cloning and sequencing of papain encoding cDNA. *Gene* 48: 219–227

Cui, Y.Y., Pandey, D.M., Hahn, E.J. and Pack, K.Y. 2004. Effect of drought on physiological aspects of Crassulacean Acid Metabolism in Doritaenopsis. *Plant Sci.* 167: 1219- 1226

Davis, M.M., Cohen, D.I., Neilsen, E.A., Steinmetz, M., Paul, W.E. and Hood, L. 1984. Cell type-specific cDNA probes and the mineral region: the localization and orientation of Ad alpha. *Proc. Nat. Acad. Sci. USA* 81(7): 2194-2198

Diab, A.A., Merah, B.T., This, D., Ozturk, N.Z., Sorrells, M. and Benscher, D. 2004. Identification at drought-unducible genes and differentially expressed sequence tags in barely, *Apply Genet.* 109: 1417-1425

Diab, A.A., Kautety, R.V., Ozhirk, N.Z., Benscher, D., Machit, M.M. and Sarrells, M.E. 2008. Drought-inducible genes and differentially expressed sequence tags associated with components of drought tolerance in durum wheat. *Scient. Res. Essay.* 3(1): 9-26

Diatchenko, L., Chrislau, Y.F., Campbell, A.P., Chenchik, A., Mooadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. and Siebert, P.D. 1996. Suppression subtractive hybridization: a method for generating differentially regulatedor tissue- specific cDNA probe. *Proc. Nat. Acad. Sci. USA.* 93: 6025–6030

Dicto, J. and Manjula, S. 2005. Identification of elicitor-induced PR5 gene homologue in *Piper colubrinum* Link by suppression substractive hybridization. *Curr. Sci.* 88(4): 624-627

Echenique V., Stamova B., Wolters P., Lazo G., Carollo V. And Dubcovsky J. 2002. Frequencies of Ty1-copia and Ty3-gypsy retroelements within the *Triticeae* EST databases. *Theor. Appl. Genet.* 104: 840–844

Fernandes J., Brendel V., Gai X., Lal S., Chandler V. L., Elumalai R. P. *et al.* 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and microarray hybridization. *Plant Physiol.* 1283: 896–910

Flavia, F., Raffaella, C., Luca, L., Silvia, C., Lorenza, V., Federica, F., Paolo, C., Maria, Z. and Pierluigi, S. 2007. Systematic analysis of mRNA 5' coding sequence incompleteness in *Danio rerio*: an automated EST-based approach. *Biol.* 2-34

Forster, B.P., Ellis, R.P., Newton, A.C., Morris, W.L., Moir, J., Lyon, J., Keith, R., Tuberosa, R., Talame, V., This, D., Teulat, B., El-Enein, R.A., Bahri, H. and Salem, M. 2000. *Stable yield in Mediterranean barley: application of molecular technologies in improving drought tolerance and mildew resistance*. In: Proceedings of the 8th international barley genetics symposium, Adelaide, 22–27 October 2000, pp 273–274

Gazeudam, J. and Oilofse, O. 2007. Isolation of cowpea genes conferring drought tolerance: Construction of a cDNA drought expression library. *Water SA* 33(3): 387-391

Ghislain, M., Frankard, V., Vandenbossche, D., Matthews, B. and Jacobs M. 1994. Analysis of the expression pattern of a novel aspartokinase-homoserine dehydrogenase gene. *Plant Mol. Biol.* 24: 835-851

*Giardi, M.T., Cona, A., Geiken, B., Kucera, T., Masojidek, J. and Mattoo, A.K. 1996. Long-term drought stress induces structural and functional reorganization of photo system II. *Planta.* 199: 118-125

Grover, A. 1999. A novel approach for raising salt tolerant transgenic plants based on altering stress signaling through Ca_2+/ calmodulin-dependent protein phosphatase calcineurin. *Curr. Sci.* 76:136–137

Guerrero, F.D., Jones, J.T. and Mullet, J.E. 1990. Turgorresponsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. *Plant Mol. Biol.* 15:11–26

Hara, E., Kato, T., Nakada, S., Sekiya, S. and Oda, K. 1991. Subtractive cDNA cloning using oligo(dT)30 Latex and PCR: Isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. *Nucleic Acids Res.* 19(25):7097-7104

*Hendrick, S.M., Cohen, D.I. Neilson, E.A. and Davis, M.M. 1984. Isolation of cDNA clones encoding T cell specific membrane associated proteins. *Nature* 308(5955): 149-153

Hirt, H. 2000. Connecting oxidative stress, auxin and cell cycle regulation though a plant mitogen activated protein kinase pathway. *Proc. Natl Acad Sci* USA 97:2405-2407

Hoekstra, F.A., Golovina, F.A. and Buitink, J. 2001. Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6(9): 430-439

Huber, S.C., Huber. J. L. and McMichael R.W. 1994. Control of plant enzyme activity by reversible protein phosphorylation. *Int Rev Cytol* 149: 47-98

Ingram, J. and Bartels, D. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. *Plant Physiol. Plant Mol. Biol.* 47: 377–403

Iturriaga, G., Cushman, M.F. and Cusman, J.C. 2006. An EST catalogue from the resurrection plant *Selaginella lepidophylla* reveals abiotic stress-adaptive genes, *Plant Sci.* 170: 1173-1184

Ji, S.J., Lu, Y.C., Feng, J.X., Wei, G., Li, J., Shi, Y.H., Fu, Q., Liu, D., Luo, J.C. and Zhu, Y.X. 2003. Isolation and analyses of genes preferentially expressed during earlycotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res.* 31: 2534–43

Jin, H., Plaha, P., Park, J.Y., Hong, C.P., Lee, I.S., Yang, L.H., Jiang, G.B., Kwak, S.S., Liu, S.K., Lee, J.S., Kim, Y.A. and Lim, Y.P. 2006. Comparative EST profiles of leaf and root of *Leymus chinensis*, a xerophilous grass adapted to high pH sodic soil. *Plant Sci.* 170: 1081-1086

Johansen, C. and Nigam, S.N. 1994. Importance of drought stress and its alleviation in legumes. *Crop Sci.* 24: 17-19

Jung, J.D., Park, H.W., Hahn, Y., Hur, C.G., In, D.S., Chung, H.J., Liu, J.R. and Choi, D.W. 2003. Discovery of genes for giusenoside biosynthesis by analysis of giuseng expressed sequence tag. *Plant Cell Rep.* 22: 224-230

Kamphuis, I.G., Drenth, G., and Baker, E.N. 1985. Thiol proteases: comparative studies based on the high resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* 182: 317–320

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Skimoraki, K. and Shinozaki, K. 1999. Improving plant drought, salt and freesing tolerance by gene transfer of a single stress inducible transcription factor. *Nature Biotech.* 17: 287-291

Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S. and Kawai K. 2001 Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13: 889–906

Kendall, A.C., Kys, A.J., Turner, J.C., Lea, P.J. and Miflin, B.J. 1983. The isolation and characterization of a catalase-deficient mutant of barley (*Hordeum vulgare* L.). *Planta.* 159, 505–511

Kirch, H.H., Bartels, D., Loci, Y., Schnable, P. and Wood, A. 2004. The aldehyde dehydrogenase gene super family of *Arabidopsis thaliana. Trends Plant Sci.* 9: 371-377

*Koizumi, M., Yamaguchi-Shinozaki, K., Tsuji, H., and Shinozaki, K. 1993. Structure and expression of two genes that encode distinct drought inducible cysteine proteinases in *Arabidopsis thaliana. Gene* 129: 175–182

Krishnamurthy, K.S., Ankegowda, S.J. and Saji, K.V. 2000. Water Stress Effects on Membrane Damage and Activities of Catalase, Peroxidase and Superoxide Dismutase Enzmes in Black Pepper (*Piper nigrum* L.). *J. Plant Biol.* 27(1): 39-42

Kushwah, R.. 2008. Development and analysis of ESTs (Expressed Sequence Tags) in Black pepper (*piper nigrum* L.) M.Sc (Ag) thesis, Kerala Agricultural University, Thrissur, 1-155p

Lee, H., Lee, J.S., Noh, E.W., Bae, E.K., Choi, Y.I. and Han, M.S. 2005. Generation and analysis of expressed sequence tags from poplar (*Populus alba* x *P. tremula* var. *glandulosa*) suspension cells. *Plant Sci.* 169: 1118-1124

Levitt, J. 1980. *Chilling, freezing, and high temperature stress.* In: Responses of plants to environmental stress, Academic, New York, vol. 1pp. 3–56

Liu, H.S. and Li, F.M. 2005. Root respiration, photosynthesis and grain yield of two spring wheat in response to soil drying. *Plant Growth regulation* 46(3): 233-240

Logroiio, M.L. and Lothrop, J.E. 2006. Impacts of drought and low nitrogen on maize production in South Asia. Developing Drought and low nitrogen tolerant maize symposium oral presentation – 5

*Ludlow, M. M. and Muchow, R. C. 1990. A critical evaluation of traits for improving crop yields in water-limited environments. *Adv. Agron.* 43:107–153

Mathe, C., Sagot, M.F., Schiex, T. and Rouze, P. 2002. Current methods of gene prediction, their strengths and weaknesses. *Nucleic acids Res.* 30: 4103-4117

Michael, G. and David, S.H. 1975. Colony hybridization: A method for the isolation of clofied DNAs that contain a specific gene. *Proc. Nat. Acad. Sci.* USA. 72:10 pp. 3961-3965

Michalek W., Weschke W., Pleissner K.P. and Graner A. 2002 EST analysis in barley defines a unigene set comprising 4,000 genes. *Theor. Appl. Genet.* 104: 97–103

Miflin B. J. and. Habash, D. Z. 2001. The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.* 370(53) pp. 979–987

Mishra, R.N., Reddy, P.S., Nair, S., Markandeya, G., Reddy, A.R., Sopory, S.K. and Reddy, M.K. 2007. Isolation and characterization of expressed sequence tags (ESTs) from substracted cDNA libraries of *Pennisetum glaucum* seedlings. *Plant Mol. Biol.* 64: 713-732

*Mittler, R. 2002. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 7: 405-410

Miyama, M., Shimizu, H., Sugiyama, M. and Hanagata, N. 2006. Sequencing and analysis of 14,842 expressed sequence tags of burma mangrove, *Bruguiera gymnorrhiza*. *Plant Sci.* 171: 234-241

Morgan, J.M. 1984. Osmoregulation and water stress in higher plants. *Ann. Rev. Plant Physiol.* 35:299–319

Mulder, N. J., Apweiler, R., Attwood,T .K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P. and Cerutti, L. 2005. InterProScan. *Nucleic Acids Res.* 33:201–205

Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. 2007. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Bioinformatics* 8: 6-21.

Nakasimha, K., Shinwari, Z.K., Sakuma, Y., Seki, M., Miura, S., Shinoraki, K. and Yanaguchi-Shinorali, K. 2000. DREB2 genes encoding DRE-binding proteins involved in dehydration and high salinity responsive gene expression. *Plant Mol. Biol.* 42: 657-665

Neale, A.D., Blomstedt, C.K., Bronson, P., Le, T.N., Guthridge, K., Evans, J., Gaff, D.F. and Hamill, J.D. 2000. The isolation of genes from the resurrection grass sporobulus stap fianus which are induced during severe drought stress. *Plant Cell Environment.* 23: 265- 277

*Noctor G, Foyer CH. 1998. Ascorbate and glutathione: keeping active oxygen under control. A. Rev. *Plant Physiol. Plant Mol. Biol.* 49: 249–279

Ohlrogge, J. and Benning, C. 2000. Unraveling plant metabolism by EST analysis. *Curr Opin Plant Biol.* 3(3):224-228

Ok, S.H., Park, H.M., Kim, J.Y., Bahn, S.C., Bac, J.M., Suh, M.C., Jeong, J.U., Kim, K.N. and Shin, J.S. 2003. Identification of differentially expressed

genes during flower development in carnation (*Dianthus caryophylus*). *Plant Sci.* 165: 291-297

*Öztürk Z. N., Talamé V., Michalowski C. B., Gozukirmizu N., Tuberosa R. and Bohnert H. J. 2002 Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.* 48: 551–573

Park, J.S., Kim, J.B., Hahn, B.S., Kim, K.H., Ha, S.H., Kim, J.B. and Kim, Y.H. 2001. In *Camellia Sinesis* (tea), using Suppression Substractive Hybridization. *Plant Sci.* 166: 953-961

Passioura, J.B. 2007. The drought environment: Physical, Biological and Agricultural perspectives. *J. Exp. Bot.* 58: 113-117

Pearson, W.R. 2003. Finding protein and nucleotide similarities with FASTA. *Curr. Bioinformatics.* 3(9):1-23

Peiguo, G., Michael, B., Stefania G., Salvatore, C., Guihua, B., Ronghua L. Maria, K., Rajeev, K. and Jan V. 2009. Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *J. Exp. Bot.* 60:12, pp. 3531–3544

Peng, F.Y., Reid, K.E., Liao, N., Schlosser, J., Lijavetzky, D., Holt, R., Zapater, J.M., Jones, S., Marra, M., Bohlmann, J. and Lund, S.T. 2007. Generation of ESTs in *Vitis vinifera* wine grape and table grape and discovery of new candidate genes with potential roles in berry development. *Gene.* 402: 40-50

Peterson, D.G., Schulze, S.R., Sciara, E.B., Lee, S.A., Bowers, J.E., Nagel, A., Jiang, N., Tibbits, D.C., Wessler, S.R. and Paterson, A. H. 2002. Integration of cot analysis, DNA cloning and high - throughout sequencing facilitates genome characterization and gene discovery. *Genome Res.* 12: 795-807

Pincheiro, H.A., DaMatta, F.M., Chaues, A.R.M., Fonter, E.P.B. and Lourciro, M.E. 2004. Drought tolerance in relation to protection against oxidative stress in clones of *Coffea canephora* subjected to long term drought. *Plant Sci.* 167: 1307-1314

Pratt, L.H., Liang, C., Shah, M., Sun, F., Wang, H., Reid, S.P., Gingle, A.R., Paterson, A.H., Wing, R., Dean, R., Klein, R., Nguyen, H.J., Ma, H.M., Zhao, X., Morrishige, D.T., Mullet, J. E. and Pratt, M.M. 2005. Sorghum Expressed Sequence Tags: Identify signature genes for Drought, pathogenesis and Sktomorphogenesis from a milestone set of 16,801 unique transcripts. *Plant Physiol.* 139: 89-884

Qiu, C.X., Xie, F.L., Zhu, Y.Y., Guo, K., Huang, S.Q., Nie, L. and Yang, Z.M. 2007. Computational identification of micro RNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene.* 395: 49-61

Quartacci, M.F. and Navarilzzo, F. 1992. Water stress and free radical mediated changes in sunflower seedlings. *J. Plant Physiol.* 139: 621-625

Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., McCombie, W.R., Martienssen, R.A. 1999. Differential methylation of genes and retrotransposons facilitates shotgun sequencing at the maize genome. *Nat Genet* 23: 305-308

Rahman, S.M.L., Nawata, E. and Sakuratani, T. 1999. Effects of water stress on superoxide dismutase and water content of tomato cultivars at different plant ages. *Hort. Sci.* 34: 490

Rajagopal, V. and Nareshkumar, K. 2003. Drought and its management in Plantation crops. Proceedings of the Brain storming session on Drought management strategies. 3rd March, KAU, Vellanikkara, pp.6-10

Ramadasan, A. and Vasantha, S. 1994. Environment stress reaction of black pepper. *Spice India* 7(9): 12-15

Ravindran, P.N., Nirmal Babu, Sasikumar, B. and Krishnamurthy, K.S. 2003. Botany and Crop improvement of Black Pepper. In: Ravindran, P.N. (ed.) *Black pepper*, Harwood Academic Publishers, Netherlands, pp. 23-142

Ribaut, J.M., Banziger, M. and Hoisington, D. 2002. Genetic dissection and plant improvement under abiotic stress conditions: drought tolerance in maze as an example. *JIRCAS Working Rep.* 23: 85–92

Rogers, J.C., Dean, D., and Heck, G.R. 1985. Aleurain: a barley CysP closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci.* U.S.A. 82: 6512–6516

Sachs, M.M. and David, T. 1986. Alterations of gene expression during environmental stress in plants. *A. Rev. Plant Physiol.* 37: 363-376

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York, USA, 1322p

Saradadevi, K., Padmasree, K., and Raghavendra, A.S. 1996. Correlation between the inhibition of photosynthesis and the decrease in area of detached leaf discs or volume/absorbance of protoplasts under osmotic stress in pea (*Pisum sativum*), *Physiologia Planatarum.* 96: 395-400

Sargent, T.D. and Dawid, I.B.1983. Differential gene expression in the gastrula of *Xenopus laevis. Sci.* 2(1): 135-400

Scandalios, J.G., Guan, L. and Polidoros, A.N. 1997. Catalases in plants: gene structure, properties, regulation and expression. In: Scandalios, J.G. ed.

Oxidative *stress and the molecular biology of antioxidants defenses.* New York: Cold Spring Harbor Laboratory Press, 343–406

Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J. and Shoemaker, R.C. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome.* 47: 868-876

Seki M., Narusaka M., Abe H., Kasuga M., Yamaguchi- Shinozaki K., Carninci P., Hayashizaki Y. and Shinozaki K. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses using a full-length cDNA microarray. *Plant Cell* 13: 61–72.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A. and Sakurai, T. 2002. Monitoring the expression profiles of 7,000 Arabidopsis genes under drought, cold and highsalinity stresses using a full-length cDNA microarray. *Plant.* 31: 279–292

Seki, M., Kamei, A., Yamaguchi-Shinozaki, K. and Shinozaki, K. 2003. Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Curr. Opinion Biotech.* 14: 194-199

Sharma, P. and Kumar, S. 2005. Differential display - mediated identification of three drought-responsive expressed sequence tags in tea [*Camellia sinensis* (L.) O. Kuntze], *J. Biosci.* 30(2): 231-235

Shinozaki, K. and Yamaguchi-Shinozakif, K. 1996. Molecular responses to drought and cold stress. *Curr. Biol.* 7: 161-167

Shinozaki, K. and Yamaguchi-Shinozaki, K. 1997. Gene expression and signal transduction in water stress response. *Plant Physiol.* 115: 327–334

Shinozaki, K. and Yamaguchi-Shinozaki, K. and Seki, M. 2003. Regulatory network of gene expression in drought and cold stress responces. *Curr. Opin. Plant Biol.* 6: 410-417

Singh, K., Kumar, S., Kumar, S. Singh, Y. and Ahuja, P. 2009. Characterization of dihydroflavonol 4-reductase cDNA in tea [*Camellia sinensis* (L.) O. Kuntze]: *Plant Biotech.* 3: 95–101

Sonkar, R., Bartels, D. and Kirch, H.H. 2003. Over expression of a stress-inducible aldehyde dehydrogenase gene from *Arabidopsis thaliana* in transgenic plants improves stress tolerance. *Plant* 35: 452-464

Spices Board. 2010. *Spice Wise Area and Production* [online]. Available:http://www.indianspices.com/pdf/spicewisearprd.xls, 2nd August, 2010

Stone, J.M. and Walker, J.C. 1995. Plant Protein Kinase Families and Signal Transduction. *Plant Physiol.* 108: 451-457

*Suiba, B., Ford, R. and Pang, E.C. 2005. Construction of cDNA library of *Lathyrus sativa* inoculated with *Mycosphaerella pinodes* and the expression of potential defence releted expressed sequence tags (ESTs). *Physiol. Mol. Plant Path.* 66: 55-67

Susihoto, S. and Berminger, F. 2007. Interactions between morphological and physiological drought responses in *Eucalyptus microtheca. Silva Fennica.* 41(2): 221-233

Thankamani, C.K. and Ashokan, P.K. 2002. Chlorophyll and leaf epicuticular wax contents of black pepper (*Piper nigrum*) varieties in response to water stress. *J. Med. Aromat. Plant Sci.* 24(4): 943-946

Thankamani, C.K., Chempakam, B. and Ashokan, P.K. 2003. Water stress induced changes in enzyme activities and lipid peroxidation in black pepper (*Piper nigrum*). *J. Med. Aro. Plant Sci.* 25(3): 646-650 Thiery, L., Leprince, D., Ghars, M. A., Debarbieux, E. And Savoure, A. 2004. Phospholipase D is a negative regulator of proline biosynethesis *in Arabidopsis thaliana. J. Bio. Chem.* 279: 14812-14818

Thiery, L., Leprince, A.S., Lefebure, D., Ghars, M.A., Debarbieux, E. and Savoure, A. 2004. Phospholipase D is a negative regulator of proline biosynthesis in *Arabidopsis thaliana. J. Bio. Chem.* 279: 14812-14818

Thomas, T.V., John Zachariah, T. and Ramadasan, A. 1990. Proline accumulation under PEG induced water deficit stress in the leaf discs of selected black pepper. *Ann. Plant. Physiol.* 4: 233-236

Vander, G.P., Hunter, T. and Lindberg, R.A.1994. Receptor proteintyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol* 10:251-337

Vasantha, S., Gopalam, A. and Ramadasan, A. 1991. Aminoacids in black pepper (*Piper nigrum* L.) cultivars with an emphasis on endogenous proline. *J. Plantn. Crops* 18: 101-103

Villalobos, M.A., Bartels, D. and Iturriaga, G. 2004. Stress tolerance and glucose insensitive phenotypes in Arabidopsis over expressing the CpMYBIO transcription factor gene. *Plant Physiol.* 135: 309-324

Vierling, E. 1991. The roles of heat shock proteins in plants. *Ann. Rev. Plant Physiol. and Plant Mol. Biol.* 42: 579-620

Wang, W., Vinocur, B. and Altman, A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta.* 218: 1-14

Wang, w., Vinocur, B. and Altman , A. 2004. Role of plant heat shock proteins and molecular chaperones in the abiotic stress response. *Trends plant Sci.* 9: 244-252

Way, H., Chapman, S., McIntyre, L., Xue, G.P., Manners, J. and Shorter, R. 2004. Identification of differentially expressed genes in wheat undergoing gradual water deficit stress using a subtractive hybridization approach. *Plant Sci.* 168(3): 661-670

Way, H., Chapman, S., McIntyre, L., Casu, R., Xue, G.P., Manners, J. and Shorter, R. 2005. Identification of differentially expressed genes in wheat undergoing gradual water deficit stress using a subtractive hybridization approach. *Plant Sci.* 168(3): 661-670

Xiong, L.M., Schumaker, K.S. and Zhu, J.K. 2002. Cell signal during cold, drought, and salt stress. *Plant Cell suppl.* 83: 165 –183

Yamaguchi-Shinozaki, K. and Shinozaki, K. 2000. Molecular responses to dehydration and low temperature: differences and cross talk between two stress signaling pathways. *Curr. Opin. Plant Biol.* 3: 217-223

Yang, G.P., Ross, D.T., Kang, W.W., Brown, P.O. and Weigel, R.J. 1999. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.* 27: 1517–1523

Yin, C., Duan, B., Wang, X. and Li, C. 2004. Morphological and physiological responses of two contrasting Poplar species to drought stress and exogenous abscisic acid application. *Plant Sci.* 167: 1091-109

Zayed, M.A. and Zeid, I.M. 1998. Effect of water and salt stresses on growth, chlorophyll, mineral ions and organic solute contents, and enzymes activity in mung bean seedlings. *Biol. Plant.* 40: 351-356

Zdobnov, E.M. and Apweiler, R. 2001. InterProScan-an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.

Zhang, J.Z. Creelman, R.A. and Zhu, J.K. 2004. From laboratory to fileld. Using information from Arabidopsis to enigeee salt, cold and drought tolerance in crops. *Plant Physiol.* 135: 615-621

*Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., Chen, C., Ma, L., Wang, J., Zhang, Q., Fan, L. and Deng, X. 2007. Globale genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf and panicle. *Plant Mol Biol.* 63:591-608

Zhu, J.K. 1997. Molecular aspects of osmotic stress in plants. Crit. Rev. *Plant Sci.* 16: 253–277

Zhu, J.K. 2002. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53: 247–273

Zhu, W., Schlueter, S.B. and Brendel, V. 2003. Refined Annotation of the Arabidopsis Genome by Complete Expressed Sequence Tag Mapping. *Plant Physiol.* 132: 459-484

Zhu-Shimoni, J.X., Lev-Yadun, S., Matthews B. and Gallili G. 1997. Analysis of the expression pattern of a novel aspartokinase-homoserine dehydrogenase gene. *Plant Physiol.* 113: 695-706

*Originals not seen

Annexures

Annexure I

Composition of different media used in the study

1. Luria Bertani (LB) broth

Tryptone	-	10 g
Yeast Extract	-	5 g
NaCl	-	5 g
pH adjusted to	-	7± 0.2
Distilled water	-	to make up to 1000ml

2. Luria Bertani agar medium

Tryptone	-	10 g
Yeast Extract	-	5 g
NaCl	-	5 g
Agar	-	20g
pH adjusted to	-	7± 0.2
Distilled water	-	make up to 1000ml

Annexure II

Already characterized clones provided for the study

Sr. no.	Clone ID	Description
1.	PNK11(2)	Heat shock proteins
2.	PNK 1(1)	MYB transcription factor
3.	PNK 3(1)	Ribonuclease
4.	PNK 2(1)	Secretory carrier membrane protein family
5.	PNK 1(2)	Fatty acid desaturase
6.	PNK 10(2)	Gamma thionins
7.	PNK 4(2)	Peptidylprolyl isomerase
8.	PNK 14(1)	Targeting protein TPX2
9.	PNK 1(3)	Dehydroquinate synthetase
10.	PNK 5(2)	DUF protein family
11.	PNK 10(3)	NADH-ubiquinone oxidoreductase family protein
12.	PNK 4(2)	Hypothetical protein
13.	PNK 14(1)	Unnamed protein
14.	PNK 1(3)	Hypothetical protein
15.	PNK 5(2)	Hypothetical protein
16.	PNK 12(1)	Polyprotein
17.	PNK 3(2)	Seven TM receptor family
18.	PNK 4(3)	Zinc finger protein
19.	PNK 2(3)	Zinc finger protein
20.	PNK 7(2)	Hypothetical protein

Annexure III

Chemical composition of reagents used for plasmid isolation described by Birnboim and Doly (1979).

a) **Solution I (Resuspension buffer)**

50 mM glucose

25 mM Tris

10 mM EDTA

b) **Solution II (Lysis buffer)**

2N NaOH

1 per cent SDS

c) **Solution III (Neutralization buffer)**

5M Potassium acetate (60 ml)

Glacial acetic acid (11.5 ml)

Distilled water (28.5 ml)

Annexure IV

Composition of Buffers and Dyes used for gel electrophoresis

1. 6x Loading/ tracking dye

Bromophenol blue	-	0.25%
Xylene cyanol	-	0.25%
Glycerol	-	30%

The dye was prepared and kept in fridge at 4^0C

2. Ethidium bromide (intercalating dye)

The dye was prepared as a stock solution of 10 mg/ ml in water and was stored at room temperature in a dark bottle.

3. 50x TAE buffer (pH 8.0)

Tris base	-	242.0 g
Glacial acetic acid	-	57.1 ml
0.5M EDTA (pH 8.0)	-	100 ml
Distilled water	-	1000 ml

The solution was prepared and stored at room temperature

Annexure V

Composition of reagents used for blotting of nylon membrane

1. Denaturation solution (400 ml)

NaCl (1.5 M)	:	35.06 g
NaOH (0.5 M)	:	8.0 g

2. Neutralization solution (pH 7.4, 400 ml)

NaCl (1.5 M)	:	35.06 g
TrisHCl (1 M)	:	63.04 g

3. SSC (Standard Saline Citrate) (20X, 1000 ml)

Sodium citrate	: 88.23 g
NaCl	: 175.32 g

Sodium citrate was dissolved in 800 ml of sterile distilled water followed by dissolving NaCl. The pH was adjusted to 7 and the final volume was made to 1000 ml. The buffer was autoclaved and stored at 4^0C.

Annexure VI

Details of the reagents provided with Megaprime DNA (probe) labeling system (Amersham Biosciences, USA)

1. Primer solution

   Random nonamer primers in an aqueous solution

2. Nucleotide solutions

   dATP, dCTP, dGTP and dTTP in Tris HCl, pH- 8, 0.5 M EDTA

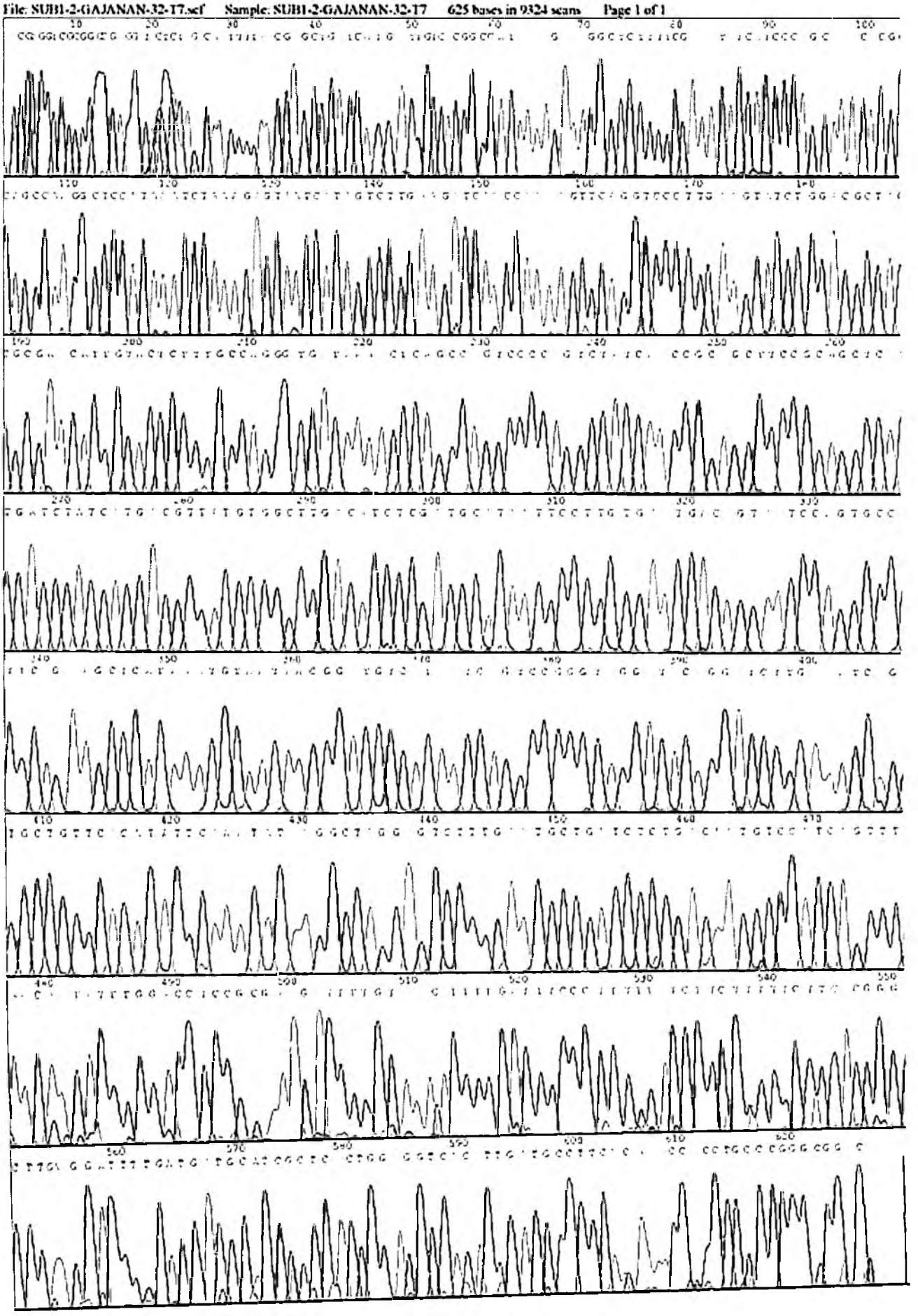
3. Reaction buffer

   10X concentrated buffer containing Tris HCl, pH-7.5, β-mercaptoethonol and $MgCl_2$

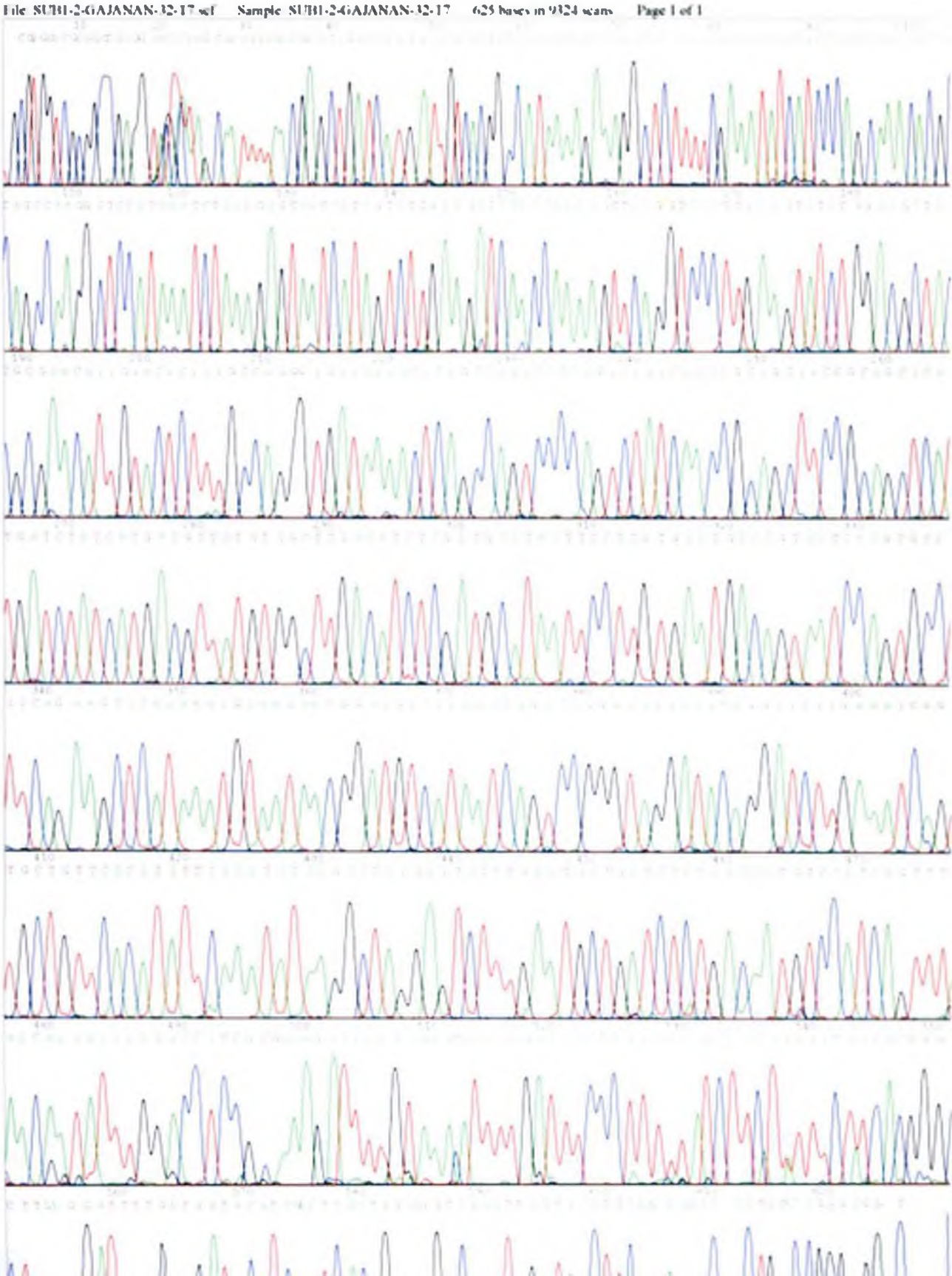
4. Enzyme solution

   1 unit/ µl DNA polymerase 1 Klenow fragment cloned in 100 mM potassium phosphate, pH 6.5, 10 mM β-mercaptoethanol and 50 per cent glycerol

Annexure VII



Annexure VII



CHARACTERIZATION OF SUBTRACTED cDNA LIBRARY FOR DETECTING EXPRESSED SEQUENCE TAGS (ESTs) SPECIFIC TO DROUGHT TOLERANCE IN BLACK PEPPER (*Piper nigrum* L.)

By

MALI GAJANAN VISHVAMBHAR

(2008-11-122)

ABSTRACT OF THE THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agriculture

(PLANT BIOTECHNOLOGY)

Faculty of Agriculture

Kerala Agricultural University, Thrissur

Centre for Plant Biotechnology and Molecular Biology

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2010

ABSTRACT

Despite many decades of research, drought continues to be a major challenge to agriculture. This is due to the unpredictability of its occurrence, duration and interaction with other biotic and abiotic stresses. So, it becomes necessary to identify the varieties which are tolerant to water stress conditions in the field and to evolve management strategies for conserving available water resources. Various genotypes of black pepper are reported to vary in their response to water stress and the variety '*Kalluvally*' has been identified as a drought tolerant one among the cultivated genotypes.

Plants respond to stress through various physiological and biochemical adaptations which are governed by different genes and these genes are differentially expressed during drought conditions. Suppression subtractive hybridization (SSH) is one of the powerful techniques to identify genes that are differentially expressed during water stress conditions. A subtracted cDNA library was constructed in black pepper by SSH using mRNA from stress induced drought tolerant plant.

Colony plates containing 280 independent clones were screened by colony hybridization with already characterized 20 clones as probes. DNA sequencing and *in silico* analysis of 35 novel sequences revealed good homology with known sequences which play important role during water stress conditions directly or indirectly. These genes include protein kinase, MYB transcription factor, Ribonuclease enzyme, peptidylprolyl isomerase, beta tubulin, NADH-ubiquinone oxidoreductase family protein, Eukaryotic peptide chain release factor, Glutamine synthetase, etc. Also these sequences had conserved domains for the above mentioned proteins. In addition to this, all the sequences possessed Open Reading Frames (ORFs) many had transmembrane helices and some were found to have signal peptide. The rest of the clones did not show any homology and therefore it was difficult to assign any reported role to these.

The results thus obtained could be very well utilized for improving drought tolerance in other susceptible varieties of black pepper through molecular breeding.