

**IMPLEMENTATION OF GENOMICS AND BIOINFORMATICS  
APPROACHES FOR IDENTIFICATION AND CHARACTERIZATION OF  
TOMATO RIPENING-RELATED GENES**

A Dissertation

by

ZHANGJUN FEI

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

**DOCTOR OF PHILOSOPHY**

December 2003

Major Subject: Molecular and Environmental Plant Sciences

**IMPLEMENTATION OF GENOMICS AND BIOINFORMATICS  
APPROACHES FOR IDENTIFICATION AND CHARACTERIZATION OF  
TOMATO RIPENING-RELATED GENES**

A Dissertation

by

ZHANGJUN FEI

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

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

---

James Giovannoni  
(Co-Chair of Committee)

---

Marla Binzel  
(Co-Chair of Committee)

---

Carol Loopstra  
(Member)

---

Keerti Rathore  
(Member)

---

Tim Davis  
(Head of Department)

---

Marla Binzel  
(Chair of MEPS Faculty)

December 2003

Major Subject: Molecular and Environmental Plant Sciences

## ABSTRACT

Implementation of Genomics and Bioinformatics Approaches for Identification and Characterization of Tomato Ripening-Related Genes. (December 2003)

Zhangjun Fei, B.S.; M.S., China Agricultural University;

Co-Chairs of Advisory Committee: Dr. James J. Giovannoni  
Dr. Marla L. Binzel

Initial activities were focused on isolation and characterization of fruit ripening-related genes from tomato. Screening of four tomato cDNA libraries at low stringency with 10 fruit development and ripening-related genes yielded ~3000 positives clones. Microarray expression analysis of half of these positives in mature green and breaker stage fruits resulted in eight ripening-induced genes. RNA gel-blot analysis and previously published data confirmed expression for seven of the eight. One novel gene, designated LeEREBP1, was chosen for further characterization. LeEREBP1 encodes an AP2/ERF-domain transcription factor and is ethylene inducible. The expression profiles of LeEREBP1 parallel previously characterized ripening-related genes from tomato. Transgenic plants with increased and decreased expression of LeEREBP1 were generated and are currently being characterized to define the function of LeEREBP1.

A large public tomato EST dataset was mined to gain insight into the tomato transcriptome. By clustering genes according to the respective expression profiles of individual tissues, tissue and developmental expression patterns were generated and genes with similar functions grouped together. Tissues effectively clustered for

relatedness according to their profiles confirming the integrity of the approach used to calculate gene expression. Statistical analysis of EST prevalence in fruit and pathogenesis-related libraries resulted in 333 genes being classified as fruit ripening-induced, 185 as fruit ripening-repressed, and 169 as pathogenesis-related. We performed a parallel analysis on public EST data for grape and compared the results for ripening-induced genes to tomato to identify similar and distinct ripening factors in addition to candidates for conserved regulators of fruit ripening.

An online interactive database for tomato gene expression data – Tomato Expression Database (TED) was implemented. TED contains normalized expression data for approximately 12,000 ESTs over ten time points during fruit development. It also contains comprehensive annotation of each EST. Through TED, we provide multiple approaches to pursue analysis of specific genes of interest and/or access the larger microarray dataset to identify sets of genes that may behave in a pattern of interest. In addition, a set of useful data mining and data visualization tools were developed and are under continuing expansion.

## **DEDICATION**

I dedicate this dissertation to my wife, Xuemei Tang and my daughter, Lindsay Fei, whose love has encouraged me during my struggle to complete this research, and to my parents and parents-in-law for their moral support.

## ACKNOWLEDGMENTS

I sincerely thank Dr. Jim Giovannoni not only for his patient mentoring in science but also for his impressive kindness. I feel fortunate to have the opportunity to work with such a nice person.

I would like to acknowledge the committee members: Dr. Marla Binzel, Dr. Carol Loopstra, and Dr. Keerti Rathore for their advice, effort and time toward my thesis work.

I am very grateful to Dr Julia Vrebalov, Dr. Zhibiao Ye, Dr. Yongsheng Liu, Dr. Paxton Payton, Dr. Rob Alba, Dr. Cornelius Barry, Ms. Ruth White, Ms. Yang Liu, and Ms. Xuemei Tang for their kind help during my Ph.D study. I would like to thank all the other members in the Giovannoni laboratory for their support.

Finally I would like to thank my family for their encouragement and endless support.

## TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
DEDICATION.....	v
ACKNOWLEDGMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
 CHAPTER	
I INTRODUCTION.....	1
Tomato Fruit Ripening and Its Regulation.....	1
cDNA Microarray Technology and Its Application.....	10
Digital Expression Analysis of EST Libraries.....	20
Thesis Rationale.....	28
II IDENTIFICATION AND CHARACTERIZATION OF TOMATO FRUIT DEVELOPMENT REGULATORY GENES.....	31
Summary.....	31
Introduction.....	32
Materials and Methods.....	33
Results.....	39
Discussion.....	52
III CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF LeEREBP1, A RIPENING-INDUCED EREBP TRANSCRIPTION FACTOR.....	57
Summary.....	57
Introduction.....	58
Materials and Methods.....	61
Results.....	67

CHAPTER	Page
Discussion.....	82
IV DIGITAL EXPRESSION ANALYSIS OF TOMATO EST LIBRARIES.....	86
Summary.....	86
Introduction.....	87
Methods.....	89
Results.....	92
Discussion.....	112
V IMPLEMENTATION OF A TOMATO EXPRESSION DATABASE (TED).....	121
Summary.....	121
Introduction.....	122
Methods.....	123
Results .....	126
Conclusion.....	143
VI SUMMARY.....	144
LITERATURE CITED .....	146
APPENDIX A.....	160
APPENDIX B.....	249
VITA.....	280



## LIST OF FIGURES

FIGURE		Page
1	Schema of cDNA microarray for gene expression analysis ....	12
2	cDNA library screening.....	42
3	Microarray expression analysis of library screening.....	46
4	Scatter plot of average expression intensities.....	48
5	RNA gel-blot analyses to confirm microarray data.....	53
6	LeEREBP1 sense construct.....	64
7	RNAi construct of LeEREBP1.....	66
8	cDNA and predicted amino acid sequence of LeEREBP1.....	69
9	LeEREBP1 has homology to AP2/ERF domain plant DNA-binding proteins.....	70
10	Alignment of the amino acid sequences of LeEREBP1 ORF with those of closely related proteins .....	71
11	RNA gel-blot analysis of LeEREBP1.....	72
12	DNA gel-blot analysis of LeEREBP1 .....	75
13	Genetic mapping of LeEREBP1.....	76
14	PCR confirmation of transgenic plants.....	78
15	DNA gel-blot confirmation of transgenic plants.....	80
16	RNA gel-blot analysis of LeEREBP1 expression in transgenic plants.....	81
17	Number of specific genes in each tissue.....	97
18	Correlated expression analysis of tomato EST data.....	100
19	Expanded region of two sub-clusters derived from Figure 18..	101

FIGURE		Page
20	Distribution of ripening-related genes by functional category.....	104
21	Distribution of pathogenesis-related genes by functional category .....	111
22	Functional classification of ESTs on the TOM1 array.....	129
23	Entity relation (ER) diagram of TED .....	130
24	The TED result page obtained through a clone identifier search .....	132
25	TED normalized relative expression data.....	134
26	Result page returned by TED for finding similar or inverse expression patterns with a specific query EST .....	135
27	TED result page from keyword search.....	138
28	Sample TED expression ratio search form and result.....	139
29	Data mining tools in TED.....	140
30	TED TC mining query form and its results. ....	142

**LIST OF TABLES**

TABLE		Page
1	Microarray databases .....	19
2	Probes used for cDNA screening.....	41
3	Number of positives recovered from library screening .....	43
4	Identification of gene family members from cDNA library screening.....	45
5	Ripening-induced genes identified by microarray analysis .....	50
6	Ripening-repressed genes identified by microarray analysis ...	51
7	Tomato cDNA library statistics.....	93
8	The ten most abundant genes in tomato EST database.....	95
9	The most abundant genes in various tissues in the EST collection.....	96
10	Genes induced both by tomato and grape ripening.....	108
11	Tissue source of ESTs on the TOM1 cDNA array .....	128

## CHAPTER I

### INTRODUCTION

#### **Tomato Fruit Ripening and Its Regulation**

Fruit ripening is a complex process influenced by numerous factors including light, hormones, temperature and genotype. Ripening associated events are brought about by developmentally and physiologically regulated changes in gene expression which ultimately lead to alterations in color, texture, flavor, and aroma of fruit (Gray et al., 1994). Fruit is a major component of the human diet contributing a large portion of the vitamins, minerals and fiber. Fruit can be classified as climacteric or non-climacteric depending on the presence or absence of massive ethylene production during ripening and on their response to exogenous ethylene (Biale and Young, 1981). Even in climacteric fruit, ethylene production is generally low until the commencement of ripening. At the onset of ripening, climacteric fruit exhibit an increase in respiration, with a concomitant burst of ethylene production.

Tomato is a typical climacteric fruit and has long served as a model organism for plant genetics, physiology and development, resulting in the accumulation of substantial information regarding the biology of this economically important organism. In recent years, the most widely studied aspect of tomato biology is the development and ripening of their fleshy fruits. Considerable attention has been directed toward elucidating the

---

This dissertation follows the style and format of Plant Physiology.

molecular basis of fruit ripening in tomato during recent years (Hobson and Grierson, 1993; Giovannoni et al., 1998; Giovannoni, 2001). Numerous fruit ripening-related genes have been isolated via differential gene expression patterns (Gray et al., 1994) and *in vivo* functions of fruit development and ripening-related genes including HMG-CoA reductase, polygalacturonase (PG), pectin methyl-esterase, ACC synthase, ACC oxidase, phytoene synthase and the NR ethylene receptor have been tested via antisense gene repression and/or mutant complementation in tomato.

### ***Ethylene biosynthesis and signaling***

The critical role of ethylene in coordinating tomato fruit ripening at the molecular level was first observed via analysis of ethylene-inducible ripening-related gene expression (Lincoln et al., 1987; Maunders et al., 1987). Reduced ethylene evolution resulting in ripening inhibition occurred with ACC synthase and ACC oxidase antisense repression (Hamilton et al., 1990; Oeller et al., 1991) while introduction of a dominant mutant allele of the NR ethylene receptor resulted in tomato plants inhibited in virtually every measurable ethylene responses including fruit ripening (Lanahan et al., 1994; Wilkinson et al., 1995; Yen et al., 1995).

Although the regulation of ethylene biosynthesis has been studied extensively, much less is known about the regulation of ethylene perception. Several *Arabidopsis* mutants deficient in the classic triple response of etiolated seedlings to ethylene, or constitutively displaying the triple response without exogenous ethylene, have been identified (Bleecker et al., 1988; Guzman and Ecker, 1990; Roman et al., 1995) and some mutant genes responsible for these phenotypes such as ETR1 (Chang et al., 1993),

CTR1 (Kieber et al., 1993), EIN2 (Alonso et al., 1999), EIN3 (Chao et al., 1997) have been cloned. ETR1 acts as an ethylene receptor and is a member of a gene family consisting of five members: ETR1, ERS1, ETR2, EIN4, and ERS2. All are homologous to bacterial two-component histidine kinases (Hua et al., 1995; Hua et al., 1998; Sakai et al., 1998). Bacterial two-component systems typically consist of a sensor molecule with a histidine-kinase domain that autophosphorylates in response to an environmental stimulus, and a response regulator with a receiver domain that accepts the phosphate from the histidine residue of the sensor on its conserved aspartate residue (Pirrung, 1999). Among ethylene receptors, only ETR1, ETR2, and EIN4 contain a receiver domain incorporated into the C terminus. Since homodimerization of ETR1 and ERS1 was observed (Schaller et al., 1995), it is postulated that ERS1 and ERS2 (which lack receiver domains), signal through the receiver domain of other members of the family by forming heterodimers (Hua et al., 1998). Ethylene receptors are in a state of active suppression of the signal transduction pathway in the absence of ethylene and ethylene binding inactivates this suppression. Arabidopsis ethylene receptors are clearly redundant in function as disruption of one or two receptor genes has little effect. However, when three or four receptors are inactivated simultaneously, ethylene response defects become apparent and manifest as a constitutive ethylene response (Hua and Meyerowitz, 1998).

CTR1 encodes a Raf-like Ser/Thr protein kinase acting downstream of the ethylene receptor family and negatively regulates ethylene response. The negative regulatory effect of CTR1 was determined on the basis that loss-of-function *ctr1* mutants

constitutively exhibit the seedling triple response to ethylene (Kieber et al., 1993). Suggesting the possibility of direct interaction with the receptors, the N-terminal putative regulatory domain of CTR1 was found to associate with both ETR1 and ERS1 in yeast two-hybrid and in vitro protein association assays (Clark et al., 1998).

In animals, Raf kinases typically act through a mitogen-activated protein (MAP) kinase phosphorylation cascade (Pelech and Sanghera, 1992). Thus, the ethylene signal transduction pathway has features of both the bacterial two-component system and eukaryotic MAP kinase pathways. Efforts to identify new MAPK/MEK proteins involved in ethylene signaling have not proven successful until recently Ouaked et al. (2003) isolated two ACC-activated MAPKs from both Arabidopsis (MPK6 and MAPK) and Medicago (SIMK and MMK3). Another kinase, SIMKK of medicago was identified in an effort to identify upstream members of the ethylene-activated MAPKs. SIMKK specifically mediated ACC-activated SIMK and MMK3 kinase activity. Transgenic Arabidopsis plants overexpressing SIMKK have constitutive MPK6 activation and phenotypically resemble *ctr1* mutants. In addition, MPK6 is not activated by ACC in *Etr1* mutants while *ein2* and *ein3* mutants (which act downstream of the MAP kinase pathway) show normal activation profiles. In contrast, *ctr1* mutants show constitutive activation of MPK6 supporting the hypothesis that these new MAP kinases are components of the ethylene cascade (Ouaked et al., 2003).

Genetics epistasis analysis has shown that EIN2 acts downstream of CTR1. Mutation in the EIN2 locus leads to ethylene insensitivity, indicating that EIN2 is an essential positively acting component of the ethylene signaling pathway (Chao et al.,

1997; Alonso et al., 1999). EIN2 encodes an integral membrane protein whose amino-terminal region shows similarity to the disease related Nramp family of metal-ion transporters. The C-terminal domain is novel with the exception of a coiled-coil motif that has been implicated in protein-protein interactions in other organisms (Alonso et al., 1999). Overexpression of the EIN2 C-terminal domain in an *ein2* null background results in a partial ethylene response phenotype and restores the ability of the mutant to respond to paraquat and jamonic acid, but not to ethylene (Alonso et al., 1999). These results suggest that the amino-terminal domain of EIN2 is required for sensing the ethylene signal from upstream signaling molecules, while the carboxy-terminal domain is necessary for transducing the ethylene signal to downstream component of the ethylene pathway (Wang et al., 2002).

Cellular localization data suggest that EIN3 acts downstream of EIN2 (Chao et al., 1997; Alonso et al., 1999). EIN3 encodes a nuclear-localized protein and serves as a transcription factor that, in response to the ethylene signal, binds to specific sequences in the promoters of target genes and activates their transcription (Solano et al., 1998). EIN3 belongs to a multigene family in Arabidopsis. Besides EIN3, two EIN3 paralogs, EIL1 and EIL2, can rescue the *ein3* mutant phenotype. This indicates that EIN3, EIL1, and EIL2 are all involved in downstream ethylene signal transduction. EIN3 and EILs are functionally redundant because null mutation in *ein3* causes only partial ethylene insensitivity. By searching for target promoters for the EIN3 family, to date, only one ethylene-inducible gene, ERF1, has been identified as a direct target of EIN3/EILs (Solano et al., 1998). ERF1 encodes a protein that belongs to the Ethylene Response



Element Binding Protein (EREBP) family of DNA-binding proteins. EREBPs bind to a cis-acting element known as the GCC-box found in promoters of ethylene inducible genes (Ohme-takagi and Shinshi, 1995). Overexpression of ERF1 leads to constitutive activation of many ethylene responses, suggesting ERF1 is a fundamental positive regulator in the ethylene signaling pathway (Solano et al., 1998). The fact that the triple response of ERF1-overexpressing plants is not complete suggests other factors, such as other EREBPs, may be required for the full spectrum of ethylene responses (Solano et al., 1998). Five novel EREBPs were isolated recently by screening an Arabidopsis leaf cDNA library (Fujimoto et al., 2000). All of these genes were differentially regulated by ethylene and displayed GCC box-specific binding activity. Interestingly, among these five EREBPs, two act as transcriptional repressors for GCC box-dependent transcription. This observation suggests that a system utilizing antagonistic mechanisms for controlling GCC box-dependent transcription operates in plants (Fujimoto et al., 2000).

#### ***Analysis of ethylene signaling in tomato***

Tomato also contains a family of putative ethylene receptor genes: LeETR1, LeETR2, NR, LeETR4, LeETR5, and LeETR6 (reviewed in Klee, 2002). It has been shown that all except LeETR6, which was cloned recently and has not yet been functionally tested, are ethylene receptors, as defined by their ability to bind ethylene (Klee, 2002). Among these receptors, NR and LeETR4 showed increased expression during fruit ripening (Wilkinson et al., 1995; Tieman et al., 2000). Mutation in the putative ethylene binding domain of the NR gene results in global ethylene insensitivity, including inhibition of fruit ripening (Lanahan et al., 1994; Wilkinson et al., 1995). NR

and LeETR4 are also negative regulators of ethylene response. But unlike the ethylene receptors in *Arabidopsis*, the repression of LeETR4 alone can result in a constitutive ethylene response phenotype (Tieman et al., 2000).

Wilkinson et al. (1997) demonstrated that transgenic tomato and petunia plants containing a mutated *Arabidopsis* ETR1 gene show repression of ethylene phenotypes including flower senescence and fruit ripening. This result indicates that ethylene receptor structure and activity is highly conserved.

Recently, Leclercq et al. (2002) introduced LeCTR1, a member of the tomato CTR family (which contains at least two members) (Lin et al., 1998; Zegzouti et al., 1999), into the *Arabidopsis ctr1* mutant under the direction of the 35S promoter. They demonstrated that ectopic expression of LeCTR1 in the *Arabidopsis ctr1* mutant restores normal ethylene signaling, suggesting this component of ethylene signaling machinery is also well conserved. Interestingly, unlike CTR1, which is constitutively expressed (Kieber et al., 1993), LeCTR1 is ethylene inducible (Zegzouti et al., 1999). LeCTR1 mRNA accumulation was up-regulated during tomato fruit ripening and in other tissues synthesizing large amounts of ethylene (e.g. senescing flowers and abscission zones; Zegzouti et al., 1999). The role of LeCTR1 in fruit ripening and the function of other members of the tomato CTR family are still under investigation by other members of our lab.

Three genes with homology to the *Arabidopsis* EIN3 gene were cloned from tomato and designated LeEIL1-3. Each of these genes can complement the *ein3-1* mutation in transgenic *Arabidopsis* (Tieman et al., 2001). Transgenic tomato plants with

reduced expression of a single LeEIL gene did not exhibit significant changes in ethylene response, while reduced expression of multiple tomato LeEIL genes was necessary to reduce ethylene sensitivity significantly, indicating that LeEILs are functionally redundant (Tieman et al., 2001).

EREBPs represent a large gene family in plants (Riechmann and Meyerowitz, 1998). Several EREBPs involved in plant defense were identified in tomato and their functions were documented (Gu et al., 2000; Gu et al., 2002) but to date no EREBPs involved in fruit ripening have been characterized. Using differential display, our group isolated an EREBP gene from ripening fruit termed DDTFR10/A which is induced by ripening (Kannan and Giovannoni, unpublished).

### ***Light and fruit pigmentation***

Light also plays a critical role during fruit ripening, especially as related to pigmentation (Alba et al., 2000). Numerous loci involved in light signal transduction have been identified by genetic screens performed primarily in Arabidopsis. They include upstream signaling components, like the phytochrome photoreceptor family members designated *phy* (Neff et al., 2000) and cryptochrome photoreceptor family members designated *cry* (Ahmad and Cashmore, 1996), in addition to intermediate factors likely to interact with photoreceptors such as FIN2 (Soh et al., 1998) and RSF1 (Fankhauser and Chory, 2000). Numerous downstream components acting as negative regulators of photomorphogenesis (the development of seedlings in response to light) have been isolated from genetic screens to identify mutants that display characteristics of

light-grown seedlings in complete darkness. This group of genes is collectively referred to as the pleiotropic COP/DET/FUS loci of Arabidopsis (Wei and Deng, 1999).

In tomato, two mutants *hp-1* and *hp-2*, which display increased accumulation of both lycopene and beta-carotene during fruit ripening in addition to higher leaf chlorophyll have been identified. The normal HP-1 and HP-2 gene products act as negative regulators of phytochrome signal transduction in tomato and the HP-2 gene has been cloned and encodes the tomato homolog of the nuclear protein DEETIOLATED1 (DET1) originally defined in Arabidopsis (Peters et al., 1992; Mustilli et al., 1999). Our lab has recently isolated the HP-1 gene and functional analysis is in progress (Liu and Giovannoni, unpublished).

### ***Developmental regulation of fruit ripening***

Tomato geneticists and breeders have identified several mutants including *ripening-inhibitor (rin)* and *non-ripening (nor)* that dramatically impact the overall ripening progress. The *rin* and *nor* mutants are single locus mutations representing genes which control virtually all measurable ripening phenomena including fruit ripening-specific ethylene biosynthesis, and thus likely represent the earliest genetically defined steps of ripening. The single reported mutation at the *rin* allele is also associated with a large sepal phenotype termed *macrocalyx (mc)*. RIN and MC were cloned and shown to be tandem MADS-box genes that were mutated through a deletion in the *rin* mutant (Vrebalov et al., 2002). RIN acts upstream of both ethylene and nonethylene-mediated ripening control and may represent a global developmental regulator of ripening potentially shared among climacteric and non-climacteric species (Vrebalov et al.,

2002). The NOR gene has also recently been cloned and shares features suggestive of a transcription factor family other than the MADS family (Vrebalov and Giovannoni, unpublished).

### **cDNA Microarray Technology and Its Application**

Recent years have seen the rapid progress in whole genome and large-scale EST sequencing projects. Completed genome sequences of dozens of organisms, of which two are plants (*Arabidopsis* and rice; *Arabidopsis* Genome Initiative, 2000; Goff et al., 2002; Yu et al., 2002), have been launched or completed. Although the resulting billions of bases of DNA sequence give us new insights into areas such as genome architecture and organization, in addition to allowing cross-genome comparisons, they do not tell us the respective functions of genes, how they bring about living cells nor how said cells are organized into tissues and whole organisms. Such questions are the basis of functional genomics. The main goal of functional genomics is not only to assign function to each gene in an organism, but also to understand how they work together to comprise functioning cells and organisms.

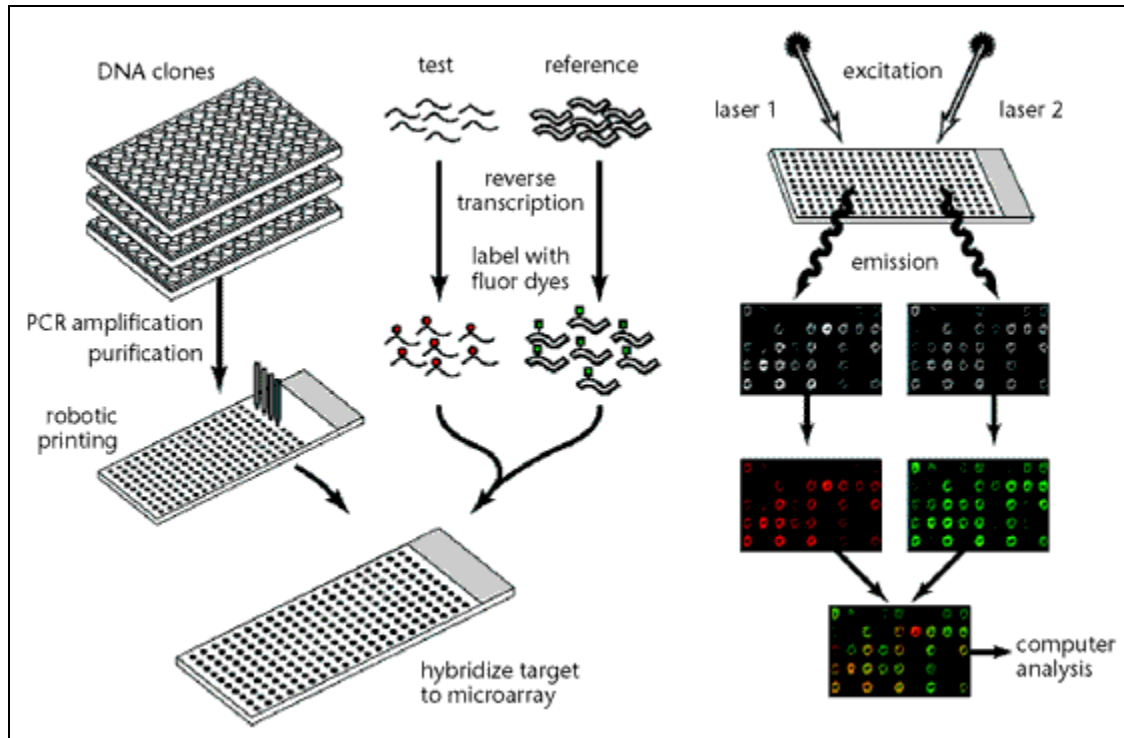
DNA microarrays are one of the most powerful technologies to take advantage of the large and rapidly increasing sequence data generated by large-scale sequencing projects. DNA arrays can be used for many different purposes including expression profiling (Schena et al., 1995), identification and genotyping of mutant alleles and DNA marker polymorphisms (Wang et al., 1998), and DNA resequencing (Hacia, 1999). Currently the major types of DNA arrays are cDNA and oligonucleotide arrays. For the

purposes of this dissertation I will focus primarily on cDNA arrays and associated data analysis and management.

### ***cDNA microarray fabrication and use***

cDNA microarrays are generated by physically depositing PCR amplification products of cloned cDNAs to known locations on glass surfaces (typically microscope slides coated with any of a number of substances that foster binding to DNA). cDNA arrays use the same DNA probe detection method used for single gene hybridization studies based on the principle of hybridization between nucleic acids, one of which is immobilized on a matrix (Southern, 1975). Through the use of high fidelity robotics in DNA spotting, tens of thousands of genes can be detected and quantified for expression in multiple tissues simultaneously via microarrays (Schena et al., 1995).

A two-probe cDNA microarray experiment consists of several general steps: array fabrication, probe RNA isolation, probe labeling and hybridization, image analysis, data extraction, and data management and analysis (Figure 1; Duggan et al., 1999). In the first step, cDNA clones are amplified through PCR and then spotted and immobilized onto a glass slide bearing chemically modified surfaces (usually poly-L-lysine or amino-silane) using a robotic printing device. Tens of thousands of cDNA inserts can be printed onto a single glass slide. Next, pools of purified RNA from two sources are reverse transcribed into cDNA and labeled separately with one of two fluorescent dyes defined by different fluorescence spectra (typically Cy3 and Cy5; Shalon et al., 1996). The cDNA pools are simultaneously hybridized to the same array. When sampled cDNAs in the pools find homologous sequences on the array they will



**Figure 1. Schema of cDNA microarray for gene expression analysis.** cDNA clones are PCR amplified and robotically printed onto chemically modified glass slides. Total pool messenger RNA samples from each of two (test and reference) tissues are reverse transcribed and labeled with fluorescent dye. mRNAs from test and reference tissues are labeled with different fluorophores to allow comparison. The two labeled cDNAs are simultaneously hybridized with the cDNA array. The array is scanned and specific gene expression levels are monitored through signal intensities using specialized computer programs. Adopted from Duggan et al., 1999.

bind while unbound probe is washed off prior to signal analysis. The array is then scanned at the two appropriate wavelengths (depending on label molecules employed) to detect the bound probes. Two images corresponding to the two fluorescent dyes (i.e. two probes) result. The signal intensity along with background and error measurements for each gene (spot) is collected from the images by a spot detection and grid alignment algorithm.

### ***Data analysis and interpretation***

After background subtraction, microarray data is typically normalized. The purpose of data normalization is to remove systematic errors introduced at various stages of a microarray experiment. Normalization can adjust for differences in labeling and detection efficiencies for the fluorescent labels and for the differences in the quantity of initial RNA from the two samples examined in the assay (Quackenbush, 2001), in addition to differences in printing pin quality or glass chip uniformity.

Following normalization, the data for each gene is typically reported as the logarithm of the expression ratio. The expression ratio is the normalized value of the expression level for a particular gene in the test sample divided by its normalized value for the reference (Quackenbush, 2001).

One of the most powerful and common uses for cDNA microarrays is identification of differentially expressed genes. A post-normalization cut-off of a two or three-fold expression change is commonly used to identify differentially expressed genes (Schena et al., 1996; Khan et al., 1999). Unfortunately, application of such thresholds completely fails to identify biologically important genes that may have a small fold



change yet sufficient replication to make a statistical argument for significance. In a recent study on sex-dependent gene expression and evolution of the *Drosophila* transcriptome, Ranz et al., (2003) found that 56% of the statistically significant differences have associated means that differ by less than a factor of two, and for the smallest significant difference the means differ by a factor of 1.13. Some groups have recently employed a fold change cutoff value combined with a value of statistical significance, such as P value obtained by standard t-test or ANOVA, to identify differentially expressed genes (Aharoni et al., 2000; Arbeitman et al., 2002).

Another powerful use of cDNA microarrays is the development of global expression profiles (DeRisi et al., 1997; Iyer et al., 1999; Hughes et al., 2000; Schenk et al., 2000; Arbeitman et al., 2002). In this kind of analysis, the goal is to identify genes that show similar expression patterns because genes with similar expression behavior are more likely to be related functionally (Quackenbush, 2001). This process is usually referred to as “clustering”. In this way, genes with unknown functions can be given annotation in terms of expression pattern related to genes in the same expression cluster of known function.

There exist many algorithms for clustering gene expression data, but generally they can be divided into two groups: supervised clustering and unsupervised clustering. Unsupervised clustering techniques are widely used in microarray analysis. Currently, the most commonly used unsupervised clustering techniques include aggregative hierarchical clustering (Eisen et al., 1998), Self-Organising Maps (SOM; Tamayo et al., 1999), k-means (Tavazoie et al., 1999), and the Self-Organising Tree Algorithm (SOTA;

Herrero et al., 2001). Aggregative hierarchical clustering has been extensively used but it suffers from its reliability and is computationally intensive. SOTA and SOM, which are unsupervised neural networks with linear running times, provide a more reliable framework appropriate for clustering large amounts of noisy data. They can deal in a natural way with missing values which frequently occur in DNA array datasets (Mateos et al., 2002). The most obvious disadvantage of SOM and k-means, which are non-hierarchical clustering methods, is that the number of clusters has to be fixed from the beginning of the procedure which can skew ultimate data interpretation. In addition, SOM may perform very poorly if irrelevant data (e.g. flat profiles) or a particular type of profile is over represented. In conclusion, SOTA is generally recognized as performing best for clustering gene expression data, though it is typical at this stage to try different algorithms and compare the results and associated variability. Most microarray data analysis software (including both commercial and freewares) include most if not all of the above algorithms.

Supervised clustering methods such as support vector machines (SVM; Brown et al., 2000), and principal component analysis (PCA; Khan et al., 2001) are not widely used for microarray analysis. SVM are able to use existing information regarding, for example, the gene functional class studied (Brown et al., 2000). PCA has been used to reduce the number of statistically significant genes from an array experiment for full analysis (Khan et al., 2001).

All of the above clustering methods are based on the assumption that genes with similar expression profiles are most likely to have similar functions in cells. Recently

Zhou et al. (2002) extended this coexpression concept to the more general “transitive coexpression” term. Transitive coexpression recognizes that an important observation of biological processes is that often two genes involved in the same process may not be strongly correlated in expression, yet can be strongly correlated with the more variant members of a single cluster. They developed what is termed a shortest-path (SP) algorithm and tested this algorithm by applying it to the analysis of the *Saccharomyces cerevisiae* gene expression profiles (Hughes et al., 2000). They were able to identify not only functionally related genes (according to gene ontology) with correlated expression profiles but also those with less well correlated expression profiles. They also showed that the SP algorithm could be used to reliably predict the function of unknown genes by relationship to known genes lying on the same shortest path.

There has been much recent interest in incorporating gene ontology (GO) annotation into microarray analyses and data output. Gene ontology refers to “a dynamic, structured, precisely defined, common, and controlled vocabulary for describing the roles of genes and gene products that can be applied to all eukaryotes even as knowledge of gene and protein roles in cells is accumulating and changing” (Ashburner et al., 2000; Ashburner et al., 2001). Gene ontology contains three basic categories: molecular function, biological process, and cellular component. Each gene or gene product can be described by one or more of these categories. Molecular function is defined as the biochemical activity (including specific binding to ligands or structures) of a gene product; biological process refers to a biological process to which the gene or

gene product contributes; and cellular component refers to the place in the cell where a gene product is active (Ashburner et al., 2000; Ashburner et al., 2001).

Bringing GO into the analysis of high throughput biological data such as that recovered from microarray experiments can be extremely useful. Storch et al. (2002) mapped various clusters of genes based on their microarray experiments to GO hierarchies and found that clock-regulated genes in liver and heart participated in many related or overlapping process, even though the two sets of genes have almost no overlap in temporal expression pattern. Coomes et al. (2002) used biological information, in this case both GO annotation and chromosomal location. They were able to get a more comprehensive picture of different kinds of cancers and information about the importance of different chromosomes and biological processes for distinguishing cancers.

### ***Data storage and management***

The use of cDNA arrays is a complicated process that generates large amounts of information. Effective data management and storage is critical to successful cDNA microarray experimentation. The ability to make comparisons between different experiments is highly dependent on the existence of a common, well-defined storage structure. Consequently, the large amounts of complex data generated by microarray experiments necessitate an efficient relational database management system.

The Microarray Gene Expression Data (MGED, <http://www.mged.org/>) group, “an international organization of biologists, computer scientists, and data analysts that aims to facilitate the sharing of microarray data generated by functional genomics and

proteomics experiments”, has been putting effort toward standardizing the way microarray data is stored, managed, and exchanged. They developed a standard definition of minimum information required to unambiguously interpret microarray data and to subsequently allow independent verification of this data at a later stage. The result is the Minimum Information About a Microarray Experiment (MIAME) guidelines which have been recently recognized as the acceptable standard by many researchers, research organizations, granting agencies and scientific publications (Brazma et al., 2001). It is an annotation standard for microarray data and includes six sections: experimental design, array design, samples, hybridizations, measurements, and normalization controls (Brazma et al., 2001). Many journals (for example, *Nature*, *Cell*, and *Lancet*) have begun to endorse or encourage MIAME compliance for papers describing the results of microarray experiments. The MGED group also developed a standard data exchange format (MAGE-ML, MicroArray Gene Expression Makeup Language) for microarray experiments (Spellman et al., 2002). MAGE-ML is an XML-based data format and provides a common platform for MIAME-compliant data exchange.

Currently dozens of microarray databases including public and commercial databases have been developed or are under development (Table 1). All the databases listed in Table 1 except AMAD follow MIAME standards and will support MAGE-ML standards when they are finalized (Gardiner-Garden and Littlejohn, 2001). All the listed databases (except GEO which only serves as a public repository) include or are linked with data analysis tools including clustering methods mentioned above. Most databases

**Table 1.** Microarray databases

<b>Database</b>	<b>Organization</b>	<b>URL</b>
GEO	NCBI	<a href="http://www.ncbi.nlm.nih.gov/geo">http://www.ncbi.nlm.nih.gov/geo</a>
ArrayExpress	EBI	<a href="http://www.ebi.ac.uk/arrayexpress">http://www.ebi.ac.uk/arrayexpress</a>
GeneX	NCGR	<a href="http://www.ncgr.org/genex/">http://www.ncgr.org/genex/</a>
SMD	Stanford University	<a href="http://genome-www5.stanford.edu/MicroArray/SMD/">http://genome-www5.stanford.edu/MicroArray/SMD/</a>
GXD	The Jackson Laboratory	<a href="http://www.informatics.jax.org/mgihome/GXD/aboutGX&lt;br/&gt;XD.shtml">http://www.informatics.jax.org/mgihome/GXD/aboutG XD.shtml</a>
READ	RIKEN	<a href="http://read.gsc.riken.go.jp/">http://read.gsc.riken.go.jp/</a>
ChipDB	WhiteHead Institute	<a href="http://staffa.wi.mit.edu/young_public/chipdb/">http://staffa.wi.mit.edu/young_public/chipdb/</a>
ArrayDB	NHGRI	<a href="http://genome.nhgri.nih.gov/arraydb/">http://genome.nhgri.nih.gov/arraydb/</a>
mAdb	National Cancer Institute	<a href="http://nciarray.nci.nih.gov/">http://nciarray.nci.nih.gov/</a>
NOMAD	UCSF	<a href="http://ucsf-nomad.sourceforge.net/">http://ucsf-nomad.sourceforge.net/</a>
AMAD	UCSF	<a href="http://www.microarrays.org/software.html">http://www.microarrays.org/software.html</a>
BASE	Lund University	<a href="http://base.thep.lu.se/">http://base.thep.lu.se/</a>
maxd	University of Manchester	<a href="http://www.bioinf.man.ac.uk/microarray/maxd/">http://www.bioinf.man.ac.uk/microarray/maxd/</a>
RAD	University of Pennsylvania	<a href="http://www.cbil.upenn.edu/RAD2/">http://www.cbil.upenn.edu/RAD2/</a>
ExpressDB	Harvard University	<a href="http://arep.med.harvard.edu/ExpressDB/">http://arep.med.harvard.edu/ExpressDB/</a>
YMD	Yale University	<a href="http://info.med.yale.edu/microarray/">http://info.med.yale.edu/microarray/</a>
2HAPI	San Diego Supercomputer Center	<a href="http://array.sdsc.edu/">http://array.sdsc.edu/</a>
GeNet	Silicon Genetics	<a href="http://www.silicongenetics.com/cgi/SiG.cgi/Products/G&lt;br/&gt;eNet/index.smf">http://www.silicongenetics.com/cgi/SiG.cgi/Products/G eNet/index.smf</a>
GeneTraffic	Iobion	<a href="http://www.iobion.com/">http://www.iobion.com/</a>
GeneDirector	BioDiscovery	<a href="http://www.biodiscovery.com/genedirector.asp">http://www.biodiscovery.com/genedirector.asp</a>

were created by large research groups and typically only store data generated within the group (Sherlock et al., 2001), or from analysis of a particular organism (Ringwald et al., 2001). In a situation analogous to that for GenBank/EMBL/DDBJ for molecular sequence data, there is an urgent need for public repositories for the large amounts of microarray data currently being generated. The publications *Nature* and *Cell* will require authors to submit their microarray data to a public repository as a precondition of publication. To fulfill this need, groups at NCBI developed GEO (Edgar et al., 2002) and groups at EBI developed ArrayExpress (Brazma et al., 2003). Both have web-based online submission function. In addition, GEO allows the user to directly submit files in SOFT (simple omnibus format) format and ArrayExpress allows the user to deposit data directly via MAGE-ML files. Besides serving as a public repository, ArrayExpress is also associated with the microarray data analysis tool Expression Profiler (<http://ep.ebi.ac.uk/EP/>). Expression Profiler allows the user to perform cluster analysis, pattern discovery, pattern visualization, and search Gene Ontology categories.

### **Digital Expression Analysis of EST Libraries**

Expressed sequence tags (ESTs) are created by partially sequencing randomly isolated gene transcripts that have been converted into cDNA and cloned (Adams et al., 1991). Recent developments in high volume biotechnology combined with advanced DNA sequencing technology have made it feasible to perform large-scale EST projects. Currently there are approximately 17 million ESTs in the NCBI EST collection – dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST/>). The EST collections for human and

mouse are the two largest and are far ahead of those from other organisms, with approximately 5.4 and 3.7 million ESTs, respectively, while the third largest collection is from *Ciona intestinalis*, with approximately 500,000 ESTs. With many large-scale EST sequence projects in process and new projects being initiated, the number of ESTs in the public domain will be likely to increase substantially for the foreseeable future.

As a result of rapid advances in computational molecular biology and biostatistics, it is possible to mine and analyze large EST datasets efficiently and exhaustively. ESTs have been successfully applied to accelerating gene discovery including gene family expansion (Papadopoulos et al., 1994; Rogaev et al., 1995; Semple et al., 2000; Bourdon et al., 2002), facilitating large-scale gene expression analysis (Ewing et al., 1999; Ewing and Claverie, 2000; Ogihara et al., 2003), elucidating phylogenetic relationships (Nishiyama et al., 2003), and building physical (Deloukas et al., 1998) and genetic maps (Schuler et al., 1996) of the human genome. One obvious outcome of large-scale EST projects is that ESTs provide the most comprehensive window into the transcriptome as contrasted to whole genome sequences which yield information about genome content and organization but not expression. In the following sections the focus is primarily on gene expression analysis using large-scale EST datasets.

### ***Coordinated gene expression***

Expression levels can be measured for one gene at a time using traditional methods such as RNA gel-blots, reverse transcriptase PCR (RT-PCR), or nuclease protection assays. These methods tend to be limited by the number of transcripts that can



be analyzed simultaneously. Recently emergence of microarray technology allows one to detect expression levels of tens of thousands genes simultaneously (Schena et al., 1995). It's a powerful and efficient approach to monitor global transcriptional profiles and has been generating a lot of valuable information.

Microarray is a hybridization-based technology, thus it suffers from hybridization artifacts. Cross-hybridization of closely related members in the same gene family and short stretches of nucleotide homologies might occur in cDNA arrays and in oligonucleotide arrays (Kachalo et al., 2002), respectively. In addition, secondary structures can interfere with hybridization in cDNA arrays which also can not distinguish different splicing forms. Tag sampling approaches, such as serial analysis of gene expression (SAGE; Velculescu et al., 1995) and ESTs (Ewing et al., 1999), provide alternative approaches for large-scale expression profile analysis. The successful use of ESTs for expression analysis depends on the availability of large unbiased cDNA libraries. Normalized and subtracted cDNA libraries, or cDNA libraries with insufficient size, should be excluded for expression analysis.

The average size of EST libraries ranges from between 1000 and 10,000 entries and therefore an EST library cannot be regarded as faithfully representing the gene expression pattern of a tissue (Vingron and Hoheisel, 1999). It has been estimated that between 10,000 and 30,000 different genes are expressed in a given cell, with an average of about 300,000 mRNA molecules per cell (Bishop et al., 1974). However, if large numbers of transcripts were sequenced, a library pool would be a proportional representation of all abundant and moderately expressed genes (Schmitt et al., 1999).

The frequencies of these genes could become statistically significant (Okubo et al., 1992). However, some reports have shown that relatively small EST collections are sufficient to provide statistically significant results for a subset of genes (Hwang et al., 1997; Bortoluzzi et al., 2000).

Ewing et al. (1999) used a relatively small collection of rice ESTs (~27,000 ESTs) to generate digital expression profiles by counting the occurrence of ESTs for different genes from different cDNA libraries. In order to obtain statistical significance, only cDNA libraries with >800 ESTs and genes with at least 5 ESTs were used in this analysis. By clustering gene expression profiles or whole cDNA library profiles, they showed that genes with similar functions or in the same pathway, or cDNA libraries expected to share pattern of gene expression, often grouped together. A simple statistical test (Pearson correlation) was used to associate genes or libraries having similar expression profiles (Ewing et al., 1999). This approach was further validated by using a much larger sample of ESTs from assorted UniGene projects (human, mouse, and rat ESTs) (Ewing and Claverie, 2000). By rejecting contigs (genes) with fewer than 10 constituent sequences and cDNA libraries for which fewer than 1000 ESTs have been sampled, Ewing and Claverie (2000) created digital expression profiles for 5624 genes across 129 human cDNA libraries, 3889 genes across 48 mouse cDNA libraries, and 1295 genes across 22 rat cDNA libraries, respectively. They found that genes with correlated profiles are frequently genes whose products either physically interact (as in a multi-subunit complex) or function in the same pathway. Furthermore, by taking the dendrograms derived from the hierarchical clustering of human and mouse genes, and

cross-referencing on said dendrograms known human/mouse orthologs, “conserved correlations” between human and mouse can be identified.

Recently, this approach was successfully applied to analyze the potato (Ronning et al., 2003) and wheat (Ogihara et al., 2003) EST collections.

### ***In-silico differential display***

Many biologists are interested in finding genes that are differentially expressed in different tissues, during different developmental stages, after different drug treatments, or during specific pathogen responses. Such genes may be of potential biological or pharmaceutical relevance. Various approaches, such as differential display (Liang and Pardee, 1992), suppression subtractive hybridization (SSH; Diatchenko et al., 1996), cDNA microarray (Schena et al., 1995), SAGE (Velculescu et al., 1995), and large-scale cDNA sequencing (Okubo et al., 1992), have been developed to identify differentially expressed genes or tissue specific genes.

Detecting differentially expressed genes from two conditions (such as normal and tumor tissues) using large-scale EST data has been extensively used in human research (Schmitt et al., 1999; Leerkes et al., 2002), and most notably by the Cancer Genome Anatomy Project (CGAP; <http://cgap.nci.nih.gov>; Strausberg et al., 2001). When dealing with large amounts of expression data derived from large-scale cDNA sequencing, the basic requirement is to develop efficient and reliable statistical methods to identify differentially expressed genes. Unlike microarray experiments, large-scale EST sequencing experiments are usually not replicated. This implies that the standard error associated with each expression measurement cannot be estimated from its

dispersion and that none of the standard tests requiring variance estimates (such as Student's t-test, ANOVA test) can be used (Claverie, 1999).

Based on the Poisson distribution property of large-scale EST expression data (total number of ESTs in one library  $\geq 1000$  and each transcript representing a small fraction (less than 5%) of the library), Audic and Claverie (1997) derived a new significance test (AC statistic) specifically adapted to the analysis of EST sampling data from two conditions. Suppose that a given gene (e.g. lipoxygenase) will occur  $x$  times in a library with size  $N_1$ , then the probability of this gene occurring  $y$  times in another library with size  $N_2$  is given by:

$$p(y|x) = \left(\frac{N_2}{N_1}\right)^y \frac{(x+y)!}{x!y! \left(1 + \frac{N_2}{N_1}\right)^{(x+y+1)}}$$

As for this equation, small  $p(y|x)$  are expected to characterize the genes exhibiting regulated expression, the relative abundance of which is unlikely to be the same in the two libraries (Audic and Claverie, 1997). In the same study, this significance test was validated by computing the rate of false alarm using Monte-Carlo simulation on the basis of two real datasets. This significance test also indicates that it's not appropriate to use ratios (such as 2-fold) in EST tag expression analysis, as a variation in counts from 1-2 (assuming identical size of the compared libraries) is not indicative of a significant ( $P < 0.05$ ) increase, whereas a variation from 20-40 is significant at  $P < 0.01$ .

Another commonly used significance test for pair-wise comparison of gene expression data in EST sampling experiments is Fisher's  $2 \times 2$  exact test. This test is traditionally used for the analysis of  $2 \times 2$  contingency tables arising from treatment

versus control experiments. The digital differential display (DDD) tool at the CGAP website adopted this test to evaluate differentially expressed genes between normal and cancer tissues ([http://www.ncbi.nlm.nih.gov/UniGene/info\\_ddd.shtml](http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml)). As pointed out by Audic and Claverie (1997), Fisher's  $2 \times 2$  exact test is too conservative for this kind of analysis and many biologically important genes may fail to be identified. In addition, Fisher's  $2 \times 2$  exact test requires that the total number of data values in the contingency table to be fixed, as well as both the row marginal total and the column marginal total. This is not the case for EST sampling data.

Another method used for pair-wise comparison of gene expression data in EST sampling experiments is the  $\chi^2$   $2 \times 2$  test statistic. This method is seldom used (one example, Leerkes et al., 2002). The purpose of the  $\chi^2$   $2 \times 2$  test is to evaluate whether the two conditions (cDNA libraries in this case) significantly differ as a whole, using the entire expression intensities in both conditions (Claverie, 1999). But what we need is to test whether the expression of each gene significantly differs between the two conditions. Thus, the  $\chi^2$   $2 \times 2$  test statistic may not be appropriate for this type of data analysis.

To detect differentially expressed genes from multiple conditions, we can compare gene expression levels between any pair of conditions by using the statistical tests described above. However, the proper Bonferroni correction will have to be applied to assess the statistical significance of the results (Claverie, 1999). This is very inconvenient and seldom used (45 comparisons will be needed for just 10 conditions). To solve this problem, several statistical tests were developed to identify differentially

expressed genes among multiple condition comparisons in a single test. Greller and Tobin (1999) developed a robust computational method that identifies genes expressed selectively (the expression of the gene is markedly high, or markedly low, in one particular tissue compared with its level in other tissues or sources) among multiple conditions. This method combines assessments of the reliability of expression quantitation with a statistical test of expression distribution patterns. This algorithm has been successfully implemented to analyze large databases of gene abundances routinely. A number of biologically and pharmacologically interesting selectively expressed genes have been identified (Greller and Tobin, 1999). But this method is only useful for the detection of single outliers and only applicable to a very large dataset (at least 10 different conditions). So it does not extend to more general patterns of differing gene expression.

Recently, Stekel et al. (2000) derived a log likelihood ratio statistical test (R statistic) which tends asymptotically to a  $\chi^2$  distribution. It is a more general test that compares the abundance of genes in any number of cDNA libraries. The ratio gives a measure of the extent to which the differences in gene expression correspond to heterogeneity of the libraries as opposed to random sampling variability. Like the AC statistic, the R statistic is also derived based on the fact that large-scale EST data closely follows the Poisson distribution. Since the test is to be used repeatedly on many thousands of genes, the authors deliberately use a rank-like value obtained from a randomization approach instead of P value to the test statistic. The rank value that is associated with a reliability value is dependent on the particular dataset used. This

analysis method was tested with prostate cDNA library data and a number of genes that appear to be biologically relevant were identified, as well as a number of unknown genes (Stekel et al., 2000).

Because EST sampling data closely follow a Poisson distribution, theoretically both the AC statistic and R statistic, which are based on Poisson distribution assumption, would have an advantage over other methods. Romualdi et al. (2001) compared the efficiency of several statistical methods (the ones described above plus general  $\chi^2$  test) used for detection of differentially expressed genes by using both a simulation approach and the analysis of datasets obtained from the UniGene database. They found that when dealing with pair-wise comparisons, as expected, the AC statistic was the most efficient method for detecting differences in gene expression, while the general  $\chi^2$  test was unexpectedly the most efficient in multitag sampling experiments, especially when dealing with variations affecting weakly expressed genes. On the other hand, the R statistic may be useful to search for differentially expressed genes when dealing with the mostly highly or moderately expressed genes (Romualdi et al., 2001).

### **Thesis Rationale**

The aim of this dissertation is to identify and characterize novel tomato fruit ripening-related genes. The process of fruit ripening is of great agricultural importance. Thus the benefits of elucidating the mechanism of fruit ripening in economically important crops such as tomato would be numerous. It has been shown that genetic information can be easily shared between tomato and other Solanaceae species, and

comparative mapping and hybridization studies have demonstrated that tomato and potato share basically the same gene repertoire and colinear chromosomes (Tanksley et al., 1992). Thus information generated from the study of tomato could be easily exploited to gain information about other economically important solanaceous crops.

In order to provide a comprehensive report, the dissertation was organized into five chapters as follows:

- Chapter I provides the background and recent advances in research on tomato fruit ripening, and the background of genomics and bioinformatics approaches taken in this study to identify ripening-related genes.
- Chapter II describes the isolation of additional novel ripening-related genes from gene families of known ripening-related genes using a cDNA microarray based strategy.
- Chapter III describes the characterization and functional analysis of a novel ripening-related EREBP gene (LeEREBP1), identified in chapter II.
- Chapter IV describes the identification of ripening-related and pathogenesis-related genes by analysis of the complete tomato EST dataset. This chapter also describes the large-scale digital transcriptome analysis of tomato and comparison of ripening-induced genes from tomato and grape.
- Chapter V describes the implementation of an online tomato expression database containing normalized expression data of ~11,049 ESTs for ten time points with nine pair-wise comparisons during wild type fruit development. A set of tools



developed through this database to manage and retrieve the expression data are also described in this chapter.

**CHAPTER II**  
**IDENTIFICATION AND CHARACTERIZATION OF TOMATO FRUIT**  
**DEVELOPMENT REGULATORY GENES**

**Summary**

The ripening of fleshy fruit is a coordinated regulatory process involving genetically, hormonally, and environmentally controlled interactions of complex gene expression patterns. Tomato has long served as a model system for fruit ripening. Considerable attention has been directed toward elucidating the molecular basis of fruit ripening in tomato during recent years and numerous fruit ripening-related genes have been identified and biochemical and *in vivo* functions of many of these genes have been tested. In this study I am attempting to identify additional fruit development regulatory genes. Others have constructed and I screened ordered tomato flower, callus, carpel and breaker fruit cDNA libraries at low stringency with 10 fruit development, ethylene-signaling, light-signaling and ripening-related genes to identify additional family members that may be involved in ripening. Half of the ~3000 positives obtained from screening were put onto cDNA chips and expression patterns of these positives in mature green and breaker stage fruit were determined by using cDNA microarray analysis. This screen revealed 89 positives representing eight different genes that were induced by fruit ripening. RNA gel-blot analysis and previously published data confirmed the microarray results for seven of the eight genes. Among these seven genes, four represent some of the same genes encoded by the probes used for library screening (specifically, NR, RIN,

NOR, DDTFR10/A) and one each are novel members of the RIN, NOR and AATFR10/A families.

## **Introduction**

Fruit ripening is a complex process influenced by numerous factors including light, hormones, temperature and genotype. Ripening associated events are brought about by developmentally and physiologically regulated changes in gene expression which ultimately lead to alterations in color, texture, flavor, and aroma of fruit (Gray et al., 1994). Tomato is a typical climacteric fruit and has long served as a model organism for fruit ripening. Numerous fruit ripening-related genes have been isolated via differential gene expression patterns, associated with ethylene synthesis and response, and in relation to developmental and environmental factors impacting ripening (Gray et al., 1994; Giovannoni, 2001). *In vivo* functions of many fruit development and ripening-related genes have been tested via antisense gene repression and/or mutant complementation in tomato (Please refer to Chapter I for more details).

cDNA microarray technology can be used to detect and quantify expression levels of tens of thousands of genes simultaneously (Schena et al., 1995). One of the most common uses of this technology is to identify novel genes involved in some processes or responses of interest. Using cDNA microarrays, Aharoni et al. (2000) identified the SAAT gene involved in strawberry flavor biogenesis in the first example of microarray analysis of fruit.

In this study, I screened four arrayed tomato cDNA libraries with probes derived from genes involved in ethylene signaling (NR, LeCTR1, EIN3, DDTFR10/A), light signaling (TCOP1, TCOP11, COP9), and developmental regulation (RIN, MC, NOR) of ripening in an attempt to identify additional members of these gene families that may be associated with ripening. The relative expression levels of thousands of positive clones from screening were investigated in mature green fruit (MG) and breaker fruit (BB) by cDNA microarray analysis with the hope of identifying novel fruit ripening-related genes.

## **Materials and Methods**

### ***Plant materials***

Tomato (*Lycopersicon esculentum*) wild type cultivar TA496 was used for ovary, callus, and breaker cDNA library construction, while E6203 was used for mixed flower library construction. cDNA libraries were constructed by various members of the Giovannoni or Tanksley labs and kindly provided for this project (van der Hoeven et al., 2002). Wild type cultivar Ailsa Craig was used for RNA analysis. All plants were grown under standard greenhouse condition at either Texas A&M or Cornell University.

### ***cDNA library screening***

cDNA clones from 384-well plates were robotically arrayed onto 22 × 22 cm nylon membranes using a Q-bot (Genomix Inc.) by Ruth White. These high density filters (approximately 18,400 double-spotted clones per filter) were screened by hybridization with probes at low stringency. All probes were obtained by PCR

amplification of the corresponding cDNA clones using gene specific primers. PCR products were column purified and random hexamer labeled with  $^{32}\text{P}$ . The filters were pre-hybridized at 37 °C overnight in a solution of 5X SSC, 0.5% (w/v) SDS, 5mM Na-P (pH 7.5), 5X Denhardt, and 0.1 mg/ml denatured salmon sperm DNA. Hybridization with heat-denatured probes was carried out at 37 °C overnight in 5X SSC, 0.5% (w/v) SDS, 5mM Na-P (pH 7.5), and 5X Denhardt's solution. After hybridization, filters were washed at 37 °C first in 2X SSC, 0.1% (w/v) SDS, then in 1X SSC, 0.05% (w/v) SDS, and finally in 0.5X SSC, 0.025% (w/v) SDS, for 20 min each time and exposed to X-ray film.

1478 of the 2944 positive clones obtained from screening were randomly selected within each probe and re-arrayed manually into 96 well plates and stored at -80 °C. Specifically, cDNA clones were picked into 96-well plates containing 135 ul LB medium, 15ul 10X freeze medium and antibiotics (50 ug/ml kanamycin for the flower library and 50 ug/ml ampicillin for the others) and these plates were stored at -80°C.

#### ***cDNA clone sequencing and sequence analysis***

Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (QIAGEN). cDNA sequencing primers were universal vector primers. To sequence cDNA clones from the 5' end, the M13 reverse primer (5'-GGAAACAGCTATGACCATG-3') was used for clones from the carpel, callus, and breaker fruit libraries, while the T3 primer (5'-AATTAACCCTCACTAAAGGG-3') was used for clones from the flower library. To sequence cDNA clones from the 3' end, the M13 forward primer (5'-GTAAAACGACGGCCAGT-3') was used for clones from the carpel, callus, and

breaker fruit libraries, while the T7 primer (5'-GTAATACGACTCACTATAGGGC-3') was used for clones from the flower library. 10µl sequencing reactions were performed containing 100ng cDNA template, 2µl BigDye mix, and 2.5pmole primer. Sequence reactions were performed with a 30 sec initial denaturation step at 96 °C, followed by 35 cycles of 96 °C for 15sec, 50 °C for 30 sec, and 60 °C for 3min. Sequencing reaction products were column purified and loaded on an ABI PRISM 377 DNA Sequencer (Applied Biosystems, Foster City, CA) to generated sequence trace files used in subsequent analysis.

Sequences and their corresponding quality scores were obtained by base calling the trace files using Phred software (Ewing and Green, 1998; Ewing et al., 1998). Sequences were “cleaned” by removing low quality regions, vector and linker sequences, and vector contaminations (Chou and Holmes, 2001). Cleaned high quality sequences were assembled using CAP3 software (Huang and Madan, 1999), blasted against the TIGR tomato gene index (<http://www.tigr.org/tdb/tgi/lgi/>) and/or analyzed by Sequencher 4.1 (GeneCodes, Ann Arbor, MI).

#### ***Total RNA isolation and RNA gel-blot analysis***

Tomato leaves and pericarp of fruits from different developmental stages [immature green fruits (IM), mature green fruits (MG), breaker fruits (BB), and red ripe fruits (RR)] were collected from tomato (*Lycopersicon esculentum* cv. *Ailsa craig*) plants grown under standard greenhouse conditions at Texas A&M University and Cornell University. Leaves and fruit pericarp tissue were immediately frozen in liquid nitrogen and homogenized with a mortar and pestle for RNA extraction. Total RNA was

extracted as follows: homogenized leaf or fruit tissue (1-2g) was added to 10 mL of 80°C phenol/extraction buffer containing 100 mM LiCl, 100 mM Tris-HC (pH 8.0), 10mM EDTA, 1% SDS, and 5ml DEPC-water saturated phenol. The extraction mixture was vortexed prior to addition of 5 mL chloroform/isoamyl alcohol (24:1), vortexed again and centrifuged at 4°C at 9000 rpm for 15 minutes. Following centrifugation, the aqueous phase was removed and RNA was precipitated with an equal volume of 4M LiCl for 1 h at -80°C. The precipitated pellet was washed with 70% ethanol and air-dried. The dried pellet was re-suspended in DEPC H<sub>2</sub>O.

Total RNA (10µg) was denatured at 65°C for 15min with denaturing solution consisting of 25% (v/v) formamide, 8.25% (v/v) formaldehyde (pH 7.0), 20mM sodium phosphate buffer (pH 6.5), and 5mM EDTA (pH 8.0). Following denaturation, total RNA was fractionated through 1% (w/v) agarose gels containing 8% (v/v) formaldehyde. Gels were blotted onto Hybond N nylon membrane (Amersham-Pharmacia Biotech, Uppsala) according to the manufacturer's instructions. Filters were fixed by UV-crosslinking with 150 mJoules energy and baked at 80°C for 2 h. The procedure for Northern hybridization was same as that for cDNA library screening described above except that the temperature for hybridization and wash was 65°C.

#### ***cDNA microarray preparation***

Re-arrayed cDNA clones described above were inoculated in LB + antibiotics (amp or kan depending on library) and incubated overnight at 37 °C. cDNA clones were amplified using 2 µl of overnight culture directly as template added to 98 µl PCR master mix (10mM Tris pH 8.3, 50mM KCl, 2mM MgCl<sub>2</sub>, 0.1% gelatin, 1mM dNTPs, 160nM

T7 and T3 universal primers describe above, 5 $\mu$ l DMSO, and 0.25 $\mu$ l Taq polymerase). PCR reactions were performed with an initial denaturation of 5 min at 94°, followed by 35 cycles of 94° for 1 min, 58° for 1 min and 72° for 1 min, with a final elongation step of 72 for 7 min. All PCR products were checked on 1% agarose gels to insure relative uniformity in DNA target concentration and then transferred to 384-well plates. PCR products were then ethanol precipitated and resuspended in 25 $\mu$ l 3X SSC with 0.1% sarcosyl.

Poly-L-lysine coated slides were made as follows: microscope glass slides (Corning, NY) were washed for 2 hours in a solution made by mixing 280 ml of 6.25 M NaOH with 420 ml of 95% ethanol. After rinsing with fresh MilliQ H<sub>2</sub>O, the slides were washed for one hour in a solution containing 10% (v/v) poly-L-lysine, 0.08% (w/v) NaCl, 0.002% (w/v) KCl, 0.0144% (w/v) Na<sub>2</sub>HPO<sub>4</sub>, and 0.0024% (w/v) KH<sub>2</sub>PO<sub>4</sub>. Slides were then rinsed again with MilliQ H<sub>2</sub>O. Water was removed immediately by centrifugation (5 min at 500 rpm in 50 ml tubes). The slides were dried at 45 °C for 10 min and stored at room temperature in a microscope slide box.

cDNA (PCR product) was printed in duplicate on poly-L-lysine-coated slides using a GMS 417 Microarrayer (Genetic Microsystems, Woburn, MA) using 4 printing pins at a pitch of 257  $\mu$ M. The array consists of 8 sub-grids and each sub-grid contains 384 spots (in a 32 by 12 format). Slides were “Snap-dried” on a hot plate at 80-90°C for 5-10 seconds and then rehydrated by hand-waving over a 65°C H<sub>2</sub>O bath for 5 seconds. cDNA was fixed to the poly-L-lysine-coated slides by UV-crosslinking with 550 mJoules of energy. Array fabrication was followed by a 20 min wash in 0.2% SDS



solution, 4-5 rinses in MilliQ H<sub>2</sub>O, 2 min in 100°C MilliQ H<sub>2</sub>O (to denature DNA), and a final 2 min rinse in ice-cold 95% ethanol. Slides were immediately dried by centrifugation at 500 rpm for 5 min.

***Fluorescent probe preparation, hybridization, and scanning***

The 3DNA Submicro EX Expression Array Detection Kit from Genisphere, Inc. (New Jersey, USA) was used for cDNA synthesis and labeling. 50µg total RNA was used for cDNA synthesis using either cy3 or cy5 RT primers according to the Genisphere protocol, Appendix B. Following cDNA synthesis, 4.0 µl of each cDNA pool was combined with 2.5 µl of cy3 capture reagent, 2.5 µl cy5 capture reagent, and 25 µl of hybridization buffer. Capture reagent was hybridized to the cDNA for 30-45 minutes at 55°C to bind one dye to each probe. Prior to cDNA hybridization, slides were blocked by pre-hybridizing polyT oligomers for 30 min at 55°C. The labeled cDNAs were then combined, applied to the blocked array, and covered with a 22 × 22 mm<sup>2</sup> glass coverslip (Corning, NY). The arrays were then allowed to hybridize in an incubator at 60°C for 12-18 hours. Following hybridization, slides were washed for 10 min at 55°C in 2X SSC with 0.2% SDS, followed by a 10 min wash at room temperature in 2X SSC and in 0.2X SSC. Slides were dried via brief centrifugation and immediately scanned with a ScanArray 5000 (GSI Lumonics, Oxnard, CA) laser scanner. Scans were conducted at a resolution of 10µm with laser power set between 90-95% and PMT set between 75%-85% (according to the intensity of each individual probe).

### ***Data acquisition and analysis***

Spot intensities from scanned slides were quantified using IMAGEONE 4.2 software from BioDiscovery (Marina del Rey, CA, USA). Grids were predefined and manually adjusted to ensure optimal spot recognition. Raw data was normalized and analyzed with GeneSpring 5.0 software from SiliconGenetics. Because each cDNA was spotted in duplicate, the average value for each gene was used for analysis in GeneSpring. Intensity dependent normalization (non-linear or LOWESS normalization) was applied to eliminate dye-related artifacts that cause the Cy3/Cy5 ratio to be affected by the total intensity of the spot. This normalization attempts to correct for artifacts caused by non-linear rates of dye incorporation as well as inconsistencies in the relative fluorescence intensity between some dyes. After normalization and background subtraction, the ratio of each spot was averaged. Clones with a BB/MG ratio greater than 2.0 and signal intensity in the BB channel greater than 400 were selected as putative ripening up-regulated clones, while clones with a BB/MG ratio less than 0.5 and signal intensity in the MG channel greater than 400 were selected as putative ripening down-regulated clones.

## **Results**

### ***cDNA library screening***

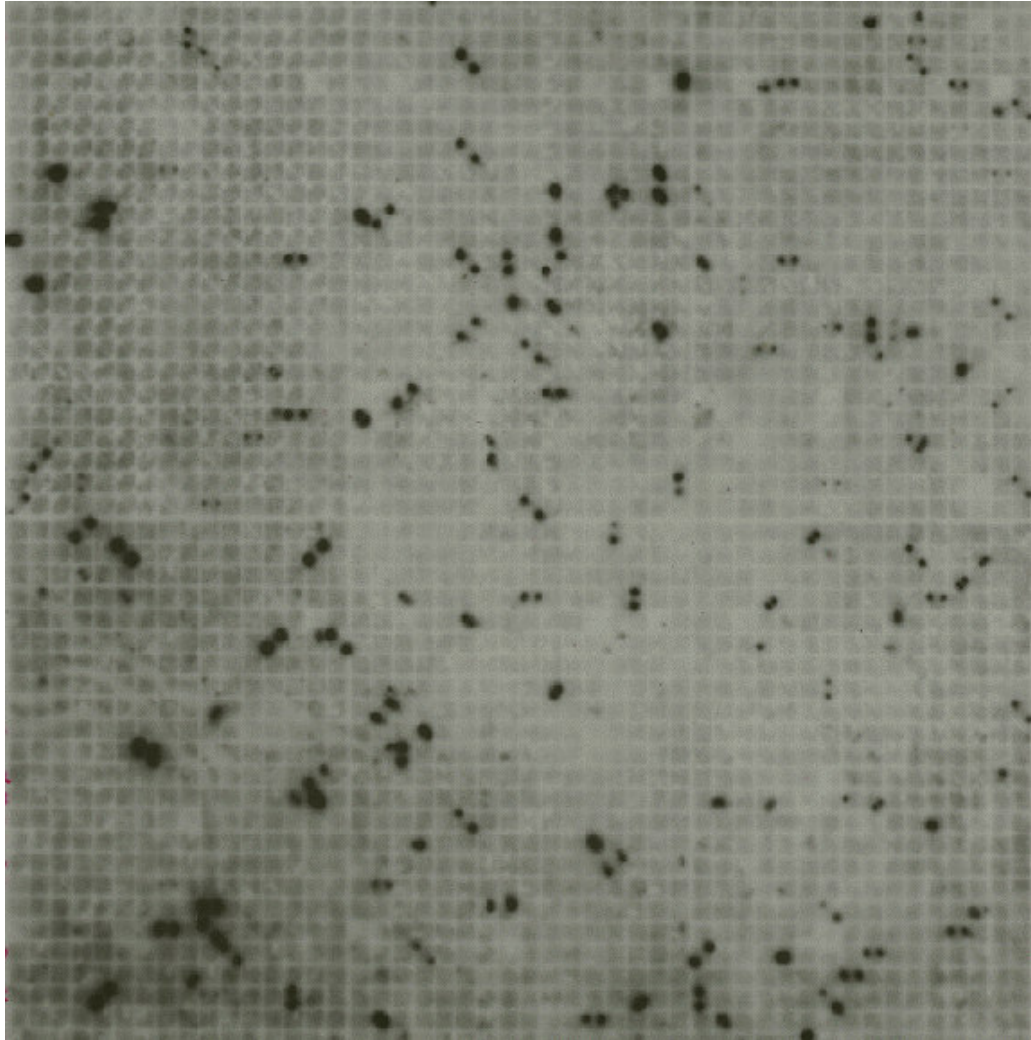
Four ordered cDNA libraries representing four different tomato tissues were used for screening. The cDNA libraries were from tomato flower (a mixture of flowers from different developmental stages including developing flower buds and open flowers),

ovary (carpel at 5 days pre-anthesis to 5 days post-anthesis), breaker fruit and callus. A total of ten genes were used as probes to screen each of the four cDNA libraries at low stringency to identify additional gene family members possibly related to ripening. Four probes represent genes involved in the ethylene signal transduction pathway [NR (Wilkinson et al., 1995), LeCTR1 (Leclercq et al., 2002), EIN3 (Chao et al., 1997), and DDTFR10/A (Kannan and Giovannoni, unpublished)], three participate in the light signaling pathway [TCOP1 (Deng et al., 1991), TCOP11 (Staub et al., 1996), and COP9 (Wei et al., 1994)], and three are transcription factors which regulate fruit ripening or flower development [RIN and MC (Vrebalov et al., 2002), and NOR (Vrebalov and Giovannoni, unpublished)] (Table 2). All probes were obtained by PCR amplification of the corresponding cDNA clones using gene specific primers. Approximately 300,000 ordered cDNA clones were screened in total with each probe. Approximately 105,000 were derived from flower, 80,000 from ovary, 55,000 from breaker fruit, and 52,000 from the callus cDNA libraries, respectively. In an attempt to recover other members of multigene families, screening was performed at low stringency with both hybridization and wash temperatures set at 37°C. An image of cDNA library screening using RIN cDNA as the probe to screen the ovary library is shown in Figure 2.

2,944 positives were identified in screening all four libraries with all ten probes (Table 3). The majority of positives (88%) hybridized to the DDTFR10/A, RIN, MC, or NOR probes. These four genes represent large multigene families in tomato and other plants. DDTFR10/A is a member of the EREBP AP2-like transcription factor family

**Table 2.** Probes used for cDNA screening

<b>Probes</b>	<b>Description</b>	<b>Source of probe</b>
NR	Tomato ethylene receptor	Tomato
LeCTR1	Negative regulator of ethylene signaling	Tomato
EIN3	Positive regulator of ethylene signaling	Arabidopsis
DDTFR10/A	Member of EREBP transcription factor family.	Tomato
TCOP1	Negative regulator of light signaling	Tomato
TCOP11	Negative regulator of light signaling	Tomato
COP9	Negative regulator of light signaling	Arabidopsis
RIN	MADS-box transcription factor. Control fruit ripening	Tomato
MC	MADS-box transcription factor. Control sepal development	Tomato
NOR	transcription factor (Vrebalov and Giovannoni, unpublished)	Tomato



**Figure 2. cDNA library screening.** cDNA libraries were robotically printed in duplicate on nylon membranes and screened with different probes at low stringency. This picture shows a subset of the ovary library (one filter, 18432 clones printed in duplicate) screened with the RIN cDNA as probe.

**Table 3.** Number of positives recovered from library screening

PROBE	LIBRARY				TOTAL
	Flower	Ovary	Callus	Breaker Fruit	
NR	5 (5) <sup>b</sup>	6 (6)	2 (2)	23 (23)	36 (36)
TCTR1	4 (4)	17 (17)	6 (6)	2 (2)	29 (29)
EIN3	9 (9)	6 (6)	20 (20)	5 (5)	40 (40)
DDTFR10A	24 (24)	495 (96)	154 (96)	224 (96)	897 (312)
RIN/MC <sup>a</sup>	150 (150)	475 (183)	116 (116)	235 (235)	976 (684)
NOR	45 (45)	506 (96)	76 (76)	92(92)	719 (309)
COP9	4 (4)	68 (10)	32 (10)	92 (10)	196 (34)
TCOP11	1 (1)	17 (0)	5 (0)	4 (4)	27 (5)
TCOP1	2 (2)	8 (6)	10 (10)	4 (4)	24 (22)
TOTAL	244 (244)	1598 (420)	421 (336)	681 (471)	2944 (1471)

<sup>a</sup> Because both RIN and MC are MADS-box genes, positives from these two probes were combined into one group

<sup>b</sup> Number in parenthesis indicates quantity of clones selected for microarray analysis

(Riechmann and Meyerowitz, 1998), RIN and MC are members of the MADS-box gene family (Vrebalov et al., 2002), and NOR is a member of a large gene family not previously associated with ripening (Vrebalov and Giovannoni, unpublished). In order to identify the rate of false positives, 130 random positive clones from across the spectrum of probes and libraries were subjected to single-pass sequencing from the 5' end. 104 (80%) clones were found to be members of the same gene family as the corresponding probes, while 26 (20%) were deemed as false positives based on this analysis. At least some of these 26 clones were chimeric and others were likely false positives resulting from very weak signals (Figure 1) or mis-scoring of the clone addresses. For most probes, two or more members from the respective gene family were identified (Table 4).

#### ***Expression analysis of library screening positives by cDNA microarray***

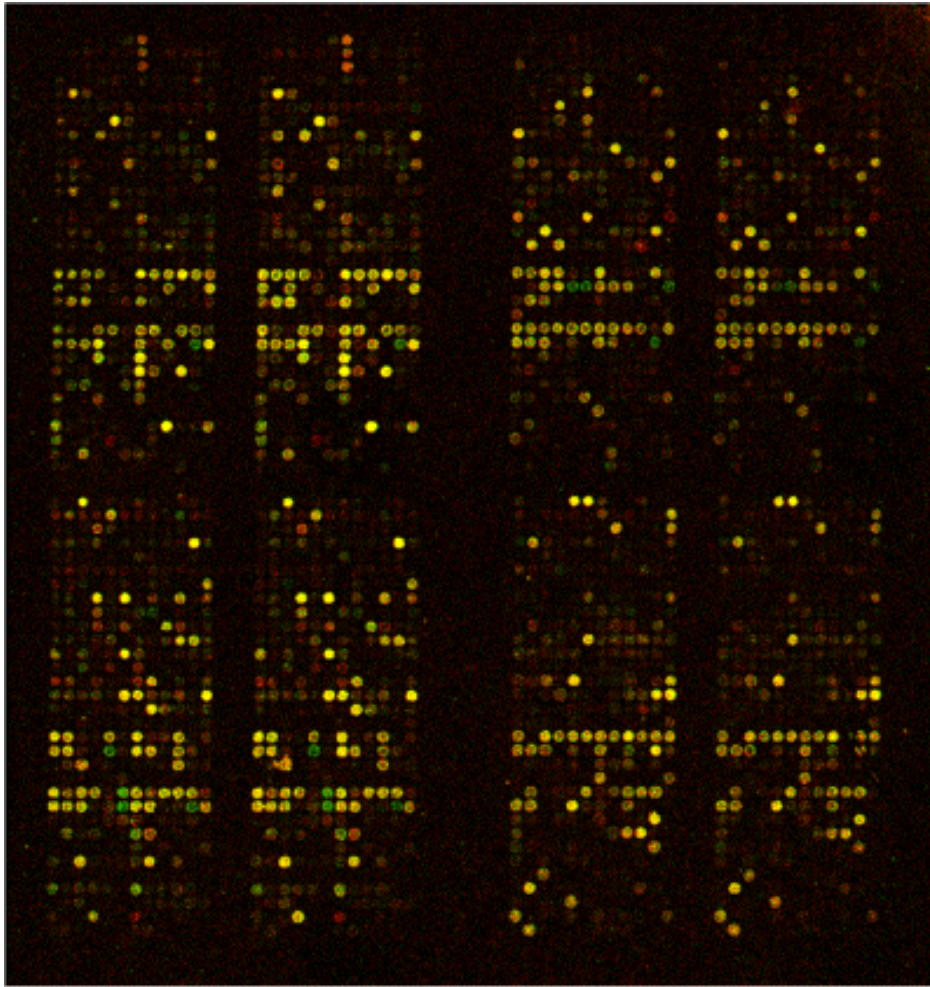
cDNA microarrays were used to quantitatively examine gene expression of positives from the library screen. In order to keep within a manageable number, 1471 positives were chosen semi-randomly for expression analysis (Table 3). These clones were amplified by polymerase chain reaction (PCR), and arrayed in duplicate on poly-L-lysine-coated glass slides using a robotic printing device (see methods).

In this study, cDNA microarray analyses were performed by competitive hybridization using probes from mature green and breaker fruit to identify those that may be related to ripening. Breaker is a critical stage in tomato fruit development as it defines the visual onset of fruit ripening and most changes related to fruit ripening such as color, texture, and flavor become noticeable in this stage. Figure 3 shows a composite

**Table 4.** Identification of gene family members from cDNA library screening

<b>Probe</b>	<b>Number of sequenced positives</b>	<b>Number of family members</b>
NR	13	1
TCTR1	11	3
EIN3	13	2
DDTFR10/A	15	5
RIN/MC	12	2
NOR	22	9
COP1	7	1
COP11	11	2





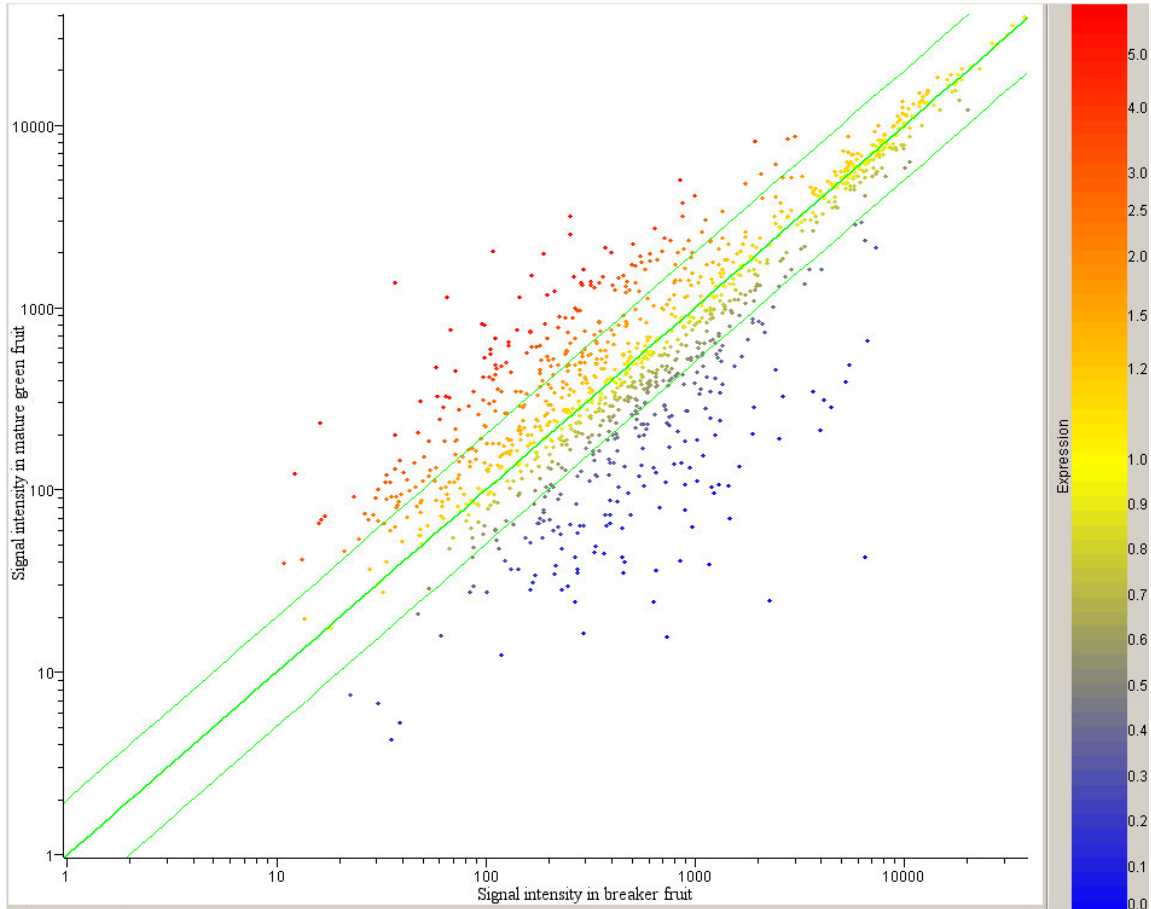
**Figure 3. Microarray expression analysis of library screening.** 1471 cDNA clones resulting from library screening were amplified by PCR and robotically arrayed in duplicate onto poly-l-lysine coated glass slides. Fluorescently labeled cDNA probes were prepared from total RNA isolated from mature green and breaker fruit. Shown is a composite false color image of the probed microarray. Genes in red are induced in breaker fruit; those in green are induced in mature green fruit, while those expressed at similar levels in both tissues are yellow.

false color image of a microarray in which the breaker fruit probe is red and mature green fruit probe is green. For each clone, there are two replicas within the array. Following background correction and LOWESS normalization, the signal intensity of each spot and for each probe from the two replicas was averaged. A scatter plot of the average expression intensities of all genes on the microarray is shown in Figure 4.

Clones with average signal intensities registering less than 400 in both channels were eliminated for further analysis. Following this selection, clones which showed more than a two-fold expression increase in breaker fruit compared to mature green fruit were classified as ripening-induced genes, while those with more than two-fold decreased expression in breaker were classified as ripening-repressed. Using the above criteria, 119 ripening-induced clones and 117 ripening-repressed clones were classified in this study by microarray analysis.

#### ***Sequence analysis of ripening-induced and repressed clones***

All 119 ripening-induced clones and 117 ripening-repressed clones were single-pass sequenced from the 5' end (using the M13 reverse primer). The sequences were compared with the TIGR tomato unigene set by performing the sequence similarity search and the TC identifier for each clone was identified (Appendix A Table I and II). Of the 119 ripening-induced clones, 30 (25.2%) are false positives (clones share no sequence similarity with the corresponding probes) or resulted in failed sequencing reactions (Appendix A Table II). Of the 117 ripening-repressed genes, 52 (44.4%) are false positives (Appendix A Table IV).



**Figure 4. Scatter plot of average expression intensities.** Data is shown for all spots on the array for breaker fruit and mature green fruit after background correction and LOWESS normalization. The center diagonal line represents similar expression between breaker and mature green fruits. The flanking lines represent the 2-fold differences in either direction. Differences in gene expression are shown by color according to the scale on the right.

The remaining 89 ripening-induced clones hybridized to one of the following five probes: DDTFR10/A, RIN, MC, NOR, or NR (Appendix A Table I). I did not identify any ripening-induced clones homologous to LeCTR1, EIN3, TCOP1, TCOP11, or COP9 in this collection. The 89 ripening-induced clones represented eight different genes including two EREBPs (TC99346, TC99621), three related to NOR (TC100002, TC101914, TC99988), one ethylene receptor (TC99771), and two MADS box genes (TC100261, TC98600) (Table 5).

The 65 ripening-repressed clones are related to the following four probes: DDTFR10/A, RIN, MC, and NOR (Appendix A Table III). The 65 ripening-repressed clones represent fifteen different genes including seven EREBPs (TC100180, TC99660, TC99989, TC99677, TC100620, TC103863, AW220395), three related to NOR (TC106246, TC102618, TC99219), and five MADS box genes (TC99375, TC99978, TC101162, TC101989, TC99899) (Table 6). The ripening-repressed clones were not characterized further and focus was directed to the ripening-induced set of genes.

### ***Confirmation of gene induction***

Among the eight ripening-induced genes, TC99771 (NR) and TC98600 (LeMADS-RIN) were published and the microarray data is consistent with these published results. TC99346 (DDTFR10/A; Kannan and Giovannoni, unpublished) and TC101914 (NOR; Vrebalov and Giovannoni, unpublished) were also identified in this screen and have been cloned previously in our lab. Microarray data reported here was also consistent with RNA gel-blot analyses showing induction of both of these genes.

**Table 5.** Ripening-induced genes identified by microarray analysis

TC number	Annotation	E value
TC99346	ripening regulated protein DDTFR10/A [ <i>Lycopersicon esculentum</i> ]	0.0
TC99621	AP2/ERF-domain protein [ <i>Solanum tuberosum</i> ]	e-116
TC100002	AtNAC2 [ <i>Arabidopsis thaliana</i> ]	2e-092
TC101914	NOR [ <i>Lycopersicon esculentum</i> ]	0.0
TC99988	nam-like protein 10 [ <i>Petunia x hybrida</i> ]	e-125
TC99771	ethylene receptor - tomato	0.0
TC100261	TDR4 transcription factor [ <i>Lycopersicon esculentum</i> ]	0.0
TC98600 <sup>a</sup>	LeMADS-RIN [ <i>Lycopersicon esculentum</i> ]	0.0

<sup>a</sup> TC98600 represents a chimeric clone where the 3' end is identical to RIN and the 5' end corresponds to a predicted protein of unknown function.

**Table 6.** Ripening-repressed genes identified by microarray analysis

<b>CloneID</b>	<b>Corresponding TC number and annotation</b>	<b>E value</b>
TC99375	MADS-box protein, TM29 [ <i>Lycopersicon esculentum</i> ]	0.0
TC99978	MADS box protein [ <i>Capsicum annuum</i> ]	e-107
TC101162	MADS box transcription factor MADS1 [ <i>Capsicum annuum</i> ]	7e-097
TC101989	TDR5 [ <i>Lycopersicon esculentum</i> ]	0.0
TC99899	APETALA3 homolog LeAP3 [ <i>Lycopersicon esculentum</i> ]	0.0
TC106246	NAM (no apical meristem)-like protein [ <i>Arabidopsis thaliana</i> ]	6e-083
TC102618	putative NAC domain protein [ <i>Solanum tuberosum</i> ]	e-149
TC99219	NAC domain protein NAC2 [ <i>Phaseolus vulgaris</i> ]	5e-084
TC100180	ethylene response factor 3 [ <i>Lycopersicon esculentum</i> ]	0.0
TC99660	AP2 domain transcription factor [ <i>Arabidopsis thaliana</i> ]	4e-021
TC99989	ethylene response factor 2 [ <i>Lycopersicon esculentum</i> ]	0.0
TC99677	ethylene responsive element binding factor [ <i>Nicotiana tabacum</i> ]	6e-046
AW220395	ethylene response factor 1 [ <i>Lycopersicon esculentum</i> ]	0.0
TC100620	ethylene-responsive element binding factor [ <i>Nicotiana sylvestris</i> ]	8e-053
TC103863	Pathogenesis-related genes transcriptional activator PTI5 [ <i>Lycopersicon esculentum</i> ]	0.0

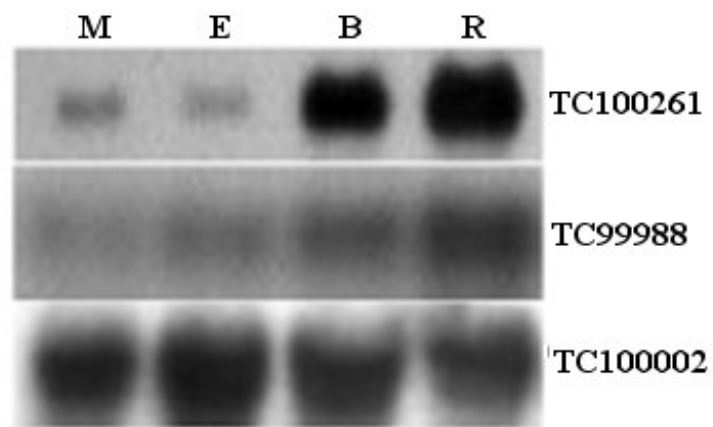
TC100002 (NOR homolog), TC99988 (NOR homolog), and TC100261 (TDR4 MADS gene) are currently being pursued by other members of the lab. Figure 5 shows RNA gel-blot analysis of these last three genes. TC100261 accumulates to low levels in mature green fruit but is induced in breaker fruit. Similarly, expression of TC99988 was very low in mature green fruit and was induced slightly in breaker fruit. The accumulation of TC100002 RNA was high in mature green fruit and was not induced in breaker fruit according to this RNA gel-blot analysis. This is the only gene tested where microarray and RNA gel-blot results were not consistent.

One ripening-induced gene (TC99621) has not been reported previously, with the first hint of a relationship to ripening coming through this microarray analysis. This gene encodes an AP2/ERF-domain protein and I have named it LeEREBP1. RNA gel-blot analysis (see chapter III) confirmed LeEREBP1 is induced during the onset of fruit ripening. This gene was chosen for further characterization and functional analysis described in the next chapter.

## **Discussion**

The primary objective of the study described here was to identify novel genes involved in tomato fruit ripening based on the assumption that ripening-associated members of large gene families impacting this process had not been fully explored.

In this study, I screened four tomato cDNA libraries with a total of 300,000 clones at low stringency, to obtain additional members of multigene families containing



**Figure 5. RNA gel-blot analyses to confirm microarray data.** Expression of three ripening-induced genes identified by cDNA microarray analysis in mature green (M), M after 12 hours of exposure to 10 ppm ethylene (E), breaker (B), and B + 7 days (R).



known or suspected members participating in fruit ripening or associated signal transduction processes (Table 2).

Approximately 3000 positives were obtained in the entire cDNA library screening effort. Sequencing of randomly selected clones indicates that the positives represent a number of gene family members in addition to those sequences represented by the original probes (Table 3).

In order to identify novel ripening-related genes from the cDNA positives, approximately half were selected and their relative expression levels were assayed in mature green and breaker fruit by microarray analysis (Figure 3). Through this analysis I identified eight ripening-induced genes (i.e. expressed higher in breaker fruit than in mature green fruit) and fifteen ripening-repressed genes (i.e. expressed higher in mature green fruit than in breaker fruit) (Tables 5 and 6, respectively). Of the fifteen ripening-repressed genes, seven were previously described but none in the context of ripening. TC99375 (TM29) encodes a MADS box protein expressed during early fruit development (Ampomah-Dwamena et al., 2002). Reduced expression of TM29 causes parthenocarpic fruit development and floral reversion, leading to the hypothesis that TM29 functions in floral organ development, fruit development, and maintenance of floral meristem identity in tomato (Ampomah-Dwamena et al., 2002). TC101989 (TDR5), another MADS box gene, was reported to be involved in mediating organ differentiation in the three inner whorls of tomato flowers (Pnueli et al., 1994). TC99989 (LeERF2) is an EREBP reported to be involved in resistance to cold stress (Tournier et al., 2003). TC103863 (Pti5), another EREBP, is a key player in plant disease resistance

(Gu et al., 2002). The functions of the remaining previously reported genes [TC99899 (LeAP3), TC100180 (LeERF3), and AW220395 (LeERF1)] remain to be characterized. The fifteen ripening-repressed genes were not characterized further.

Among the eight ripening-induced genes identified in this study, four are identical to the probes used for cDNA library screening. They are TC99346 (DDTFR10/A), TC101914 (NOR), TC99771 (NR), and TC98600 (LeMADS-RIN). All have been previously confirmed to be induced during fruit ripening by RNA gel-blot analysis (Wilkinson et al., 1995; Vrebalov et al., 2002; Kannan and Giovannoni, unpublished; Vrebalov and Giovannoni, unpublished) and RIN, NOR and NR have been previously shown to play significant roles in fruit ripening (Wilkinson et al., 1995; Vrebalov et al., 2002; Vrebalov and Giovannoni, unpublished).

Among the four novel ripening-induced genes identified in this study are an EREBP (TC99621), a MADS box gene (TC100261), and two related to the NOR transcription factor (TC100002 and TC99988). To confirm the microarray results, expression of all four genes in mature green and breaker fruit was investigated by RNA gel-blot analysis (Figure 5 and Chapter III). All except TC10002 (which is constitutively and highly expressed in both stages) were confirmed for ripening-induced expression (Figure 5).

One major drawback of cDNA microarray technology is that closely related members in the same gene family can cross-hybridize. However, RNA gel-blot analysis using 3' UTR region as probe shows that cross-hybridization was not a major problem in

that the cDNA microarray results of seven of eight putative ripening-induced genes was consistent with the corresponding RNA gel-blot analysis.

The fact that the majority of RNA gel-blot tested library positives gave results consistent with the respective array analysis indicates the screening was fairly robust. Further supporting the integrity of the screening is the fact that the lists of ripening-induced and ripening-repressed clones (Appendix A Table I, II, III, and IV) have no overlap between each other.

Of the three novel expression verified ripening-related genes identified in this study, one each related to RIN and NOR were passed on to other lab members for further functional characterization. The ripening-induced EREBP gene, TC99621, now termed LeEREBP1, was chosen for further characterization and functional analysis as detailed in Chapter III. The recently cloned Arabidopsis gene ERF1, also a member of EREBP gene family, is directly involved in the ethylene signal transduction pathway as a regulator of ethylene regulated genes (Solano et al., 1998). While the EREBP transcription factor family is a large gene family in plants, a member with confirmed involvement in climacteric ripening remains to be identified.

### CHAPTER III

## CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF LeEREBP1, A RIPENING-INDUCED EREBP TRANSCRIPTION FACTOR

### Summary

A previously undescribed ripening-induced gene, LeEREBP1, identified by cDNA microarray analysis as described in Chapter II, was selected for functional characterization. LeEREBP1 encodes a member of the AP2/ERF-domain transcription factor family specific to plants. AP2/ERF-domain transcription factors are associated with regulation of ethylene responses but none participating in fruit ripening have been identified to date. The amino acid sequence of LeEREBP1 has high similarity to CIP353, which is a cold-inducible protein in potato. In wild type tomato plants, LeEREBP1 transcripts were not detectable in leaf, persist at low levels in flowers and immature fruits but are highly induced during ripening. Inducibility of LeEREBP1 with exogenous ethylene suggests that the induction during ripening may be due at least in part to climacteric ethylene production. The transcript levels of LeEREBP1 in mutant *rin*, *nor*, and *Nr* fruit (which are all impaired in ripening ethylene synthesis or response) were also investigated and are reduced compared to normal controls. In summary, LeEREBP1 expression information suggests LeEREBP1 is a fairly typical ripening-related gene. DNA gel-blot analysis revealed that LeEREBP1 is a single copy gene in tomato.

In order to define function, LeEREBP1 was first mapped in an effort to identify possible linkage to known ripening mutations. The position of LeEREBP1 on tomato

chromosome one is not linked to any known and mapped ripening mutants. Transgenic plants containing the full open reading frame of LeEREBP1 expressed in the sense orientation under control of the CaMV 35S promoter were generated, as were transgenic plants containing an RNAi construct of LeEREBP1 for repression of the endogenous gene. Transgenic plants for both constructs are coming through tissue culture and will be available for phenotype characterization in the near future.

## **Introduction**

The plant hormone ethylene regulates complex processes of stress responses and developmental adaptation in higher plants. It is well known that ethylene participates in physiological processes as diverse as fruit ripening, senescence, germination, cell elongation, sex determination, pathogen defense, wounding, nodulation, and determination of cell fate (Tanimoto et al., 1995; O'Donnell et al., 1996; Penmetsa and Cook, 1997; Johnson and Ecker, 1998). The diverse roles that ethylene plays in plant biology suggest complexity in regulation of its biosynthesis and its signal transduction pathway. The precursor of ethylene, S-adenosyl-L-methionine, is converted to 1-aminocyclopropane-1-carboxylic acid (ACC) by ACC synthase. Ethylene is produced from ACC by ACC oxidase (Yang and Hoffman, 1984). Both the key enzymes, ACC synthase and ACC oxidase, have been cloned and encode multigene families (Kende and Zeevaart, 1997). As outlined in Chapter I, many components of ethylene signaling have been identified including transcription factors likely to represent the terminal steps in ethylene signaling.

By analyzing the promoter regions of ethylene-inducible pathogenesis-related genes, an ethylene-responsive element (ERE) was identified (Ohme-Tagaki and Shinishi, 1995; Sato et al., 1996). ERs contain an 11-bp conserved motif, TAAGAGCCGCC, which has been designated the GCC box (Eyal et al., 1993; Hart et al., 1993; Zhou et al., 1997). ERE binding factor proteins (EREBP or ERF) were first isolated as GCC box binding proteins from tobacco and their RNA levels were also shown to be up-regulated by ethylene (Ohme-Tagaki and Shinishi, 1995). ERBPs are a group of novel DNA-binding proteins that appear to be specific to plant species. ERBPs contain a highly conserved 58-59 amino acid region (designated the ERF domain) within which GCC box-specific binding activity has been demonstrated (Ohme-Tagaki and Shinishi, 1995). Cloning of the Arabidopsis ERF1 gene through direct interaction with EIN3 demonstrated integral involvement of ERBPs in the ethylene signaling pathway (Solano et al., 1998). Tests for epistatic interactions demonstrated that ERF1 acts downstream of EIN3 and all other components of the ethylene signaling pathway. Thus EIN3 and ERF1 act sequentially in a cascade of transcriptional regulation initiated by the hormone ethylene (Solano et al., 1998).

It has been suggested that the ERF domain is closely related to the AP2 domain of floral regulatory genes. However phylogenetic analysis shows that AP2-like proteins are grouped together as a monophyletic group, or subfamily, apart from the ERBP-like proteins (Riechmann and Meyerowitz, 1998). Moreover, the DNA-binding activity of AP2 has not been proven nor has its target sequence been identified and AP2 contains two putative DNA-binding domains whereas ERBPs have only one (Weigel, 1995).

The above observations indicate that the AP2 and EREBP-like proteins belong to distinct subfamilies. The AP2/EREBP superfamily can be classified into the AP2 and EREBP subfamilies based on sequence similarities and the number of putative DNA binding domains. All known AP2 subfamily members are involved in plant development, whereas the EREBP subfamily members are likely to be involved in responses to biotic and abiotic stress (Riechmann and Meyerowitz, 1998) and possibly other ethylene responses. Indeed, some or all of the specificity of ethylene responses could be regulated through the variability available in this transcription factor family.

Target gene over-expression or sense technology, has been extensively used for gene functional analysis. Briefly, this technology involves the insertion of a cDNA sequence encoding the full open reading frame of a given target gene in the sense orientation under the direction of a promoter (e.g. CAMV 35S for plants) and transcription termination sequences (e.g. NOS terminator) followed by plant transformation (e.g. using *Agrobacterium tumefaciens*). The presence of the resulting sense transgene can yield a wide range expression of levels, from decreased expression (co- or sense-suppression) to substantial overexpression. The ability to overexpress a target gene makes this approach especially robust for functional analysis of genes belonging to multigene families in which the functions of members may be redundant.

A recent and powerful reverse genetic approach for determining gene function is RNA interference (RNAi), or RNA induced gene silencing. It is also termed post-transcriptional gene silencing (PTGS) in plants, quelling in fungi and RNA interference in animals (Waterhouse and Helliwell, 2003). RNA interference is a process by which a

double-stranded RNA inserted into an organism, or cell, deactivates the gene encoding its complementary sequence by triggering the destruction of its mRNA (Waterhouse and Helliwell, 2003). Unlike other post-transcriptional silencing technology, such as anti-sense or sense co-suppression, which result in only a small proportion of silenced individuals, RNAi constructs generally result in a majority (as much as 90-100%) of independent transgenic lines with gene silencing effects (Wesley et al., 2001).

In this section, I report the characterization of a ripening-induced EREBP gene, LeEREBP1, which is ripening-related, ethylene regulated and a candidate ripening transcription factor mediating ripening effects of this hormone in fruit. Construction of sense and RNAi constructs for LeEREBP1 are described in addition to the creation of corresponding transgenic plants for future functional analysis.

## **Materials and Methods**

### ***Plant materials***

Tomato (*Lycopersicon esculentum*) wild type cultivar Ailsa Craig and nearly isogenic mutants *rin*, *nor*, and *Nr* were used for RNA analysis. Ailsa Craig wild type was also used for sense and RNAi transformation. Wild type cultivar M82 and *L. pennellii* (LA716) were used for genomic DNA analysis. For genetic mapping, the mapping population employed is the introgression line (IL) population composed of 75 lines each containing a single introgression from *L. pennellii* (LA716) in the genetic background of the processing tomato variety M82 (Pan et al., 2000). All the plants were



grown under standard greenhouse conditions at either Texas A&M or Cornell University.

***cDNA clone sequencing and sequence analysis***

cDNA clone sequencing and sequence analysis were as described in Chapter II.

***Total RNA isolation and RNA gel-blot analysis***

Total RNA isolation and RNA gel-blot analysis were as described in Chapter II.

***DNA isolation and DNA gel-blot analysis***

Young expanding leaf tissue (1-2g) from individual plants was collected and total DNA was extracted as follows. The tissue was homogenized with 20 – 25 mL of extraction buffer (100 mM Tris, 0.35 M Sorbitol, and 5 mM EDTA, pH 7.5; sodium bisulfite was added before use) in a virtishear homogenizer (Cyclone) and spun at 3200 rpm for 20 – 30 min in a Sorvall T 6000B (DuPont). Each pellet was resuspended in 1.25 mL extraction buffer, 1.75 mL nuclei lysis buffer (2% CTAB, decyltrimethylammonium bromide, Sigma; 2M NaCl, 50 mM EDTA, and 200 mM Tris), and 0.6 mL of 5% sarcosyl and heated to 65 °C for 20–30 min. After chloroform/isoamyl alcohol (24:1 v/v) extraction, the aqueous phase was mixed with an equal volume of isopropanol to precipitate the DNA. The precipitated DNA was resuspended in 100-300 µL of 10 mM Tris – 1 mM EDTA.

3 mg of genomic DNA was digested with the desired enzymes and treated with RNase and the restriction fragments were separated on 0.8% agarose gels and transferred to nylon membranes (Amersham-Pharmacia Biotech, Uppsala) according to

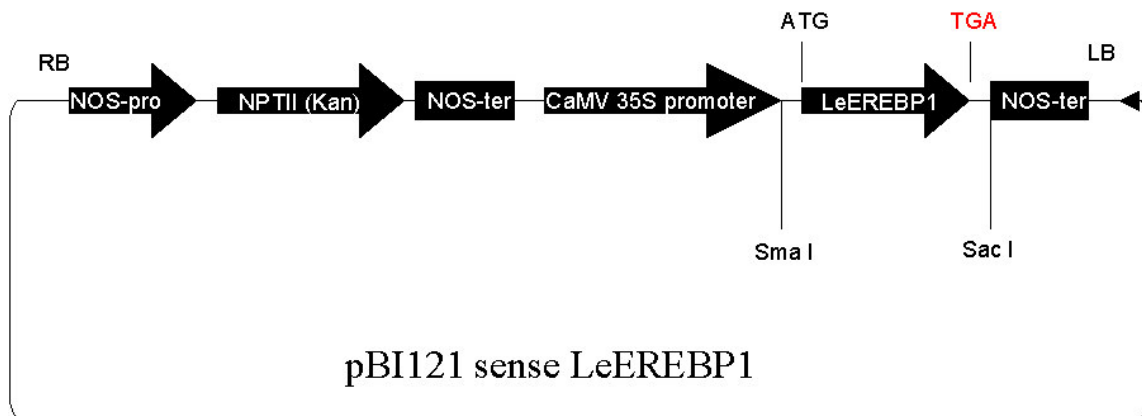
the manufacturer's instructions. DNA gel-blot hybridization was as previously described for RNA gel-blot.

### ***Genetic mapping of LeEREBP1***

76 *Lycopersicon pennellii*-derived introgression lines (ILs) were used to map LeEREBP1. Each IL differs in a single defined *L. pennellii* chromosome segment introgressed into the *L. esculentum* cv M82 background and the collection of 76 ILs provides overlapping coverage of the entire *L. pennellii* genome. A DNA gel-blot containing genomic DNA from all 76 ILs and their parents (M82 and *L. pennellii*) digested with Hae III was generated and hybridized with the 3' UTR region of LeEREBP1 under high stringency conditions as described in Chapter II.

### ***Development of LeEREBP1 sense and RNAi constructs***

A PCR product containing the full open reading frame of LeEREBP1 was obtained by amplifying the cDNA clone cLEG51D19 (GenBank accession #: BM410119) using the following primers: 5'- TTTTCCCGGGAAAAGATG TGTGGAGGTGCC-3' and 5'- TTTTGAGCTCCATGCTAAGGCACGCAACT-3'. The PCR product contains LeEREBP1 DNA sequences from position 59 to position 971 (Figure 8) with a Sma I restriction site at one end and Sac I at the other. The PCR product was digested with Sma I and Sac I and then gel-purified using the QIAquick gel extraction kit (QIAGEN). The gel purified PCR product was subcloned into pBI121 (Clontech, CA) following vector digestion with Sma I and Sac I to remove the GUS gene from pBI121 (Figure 6).

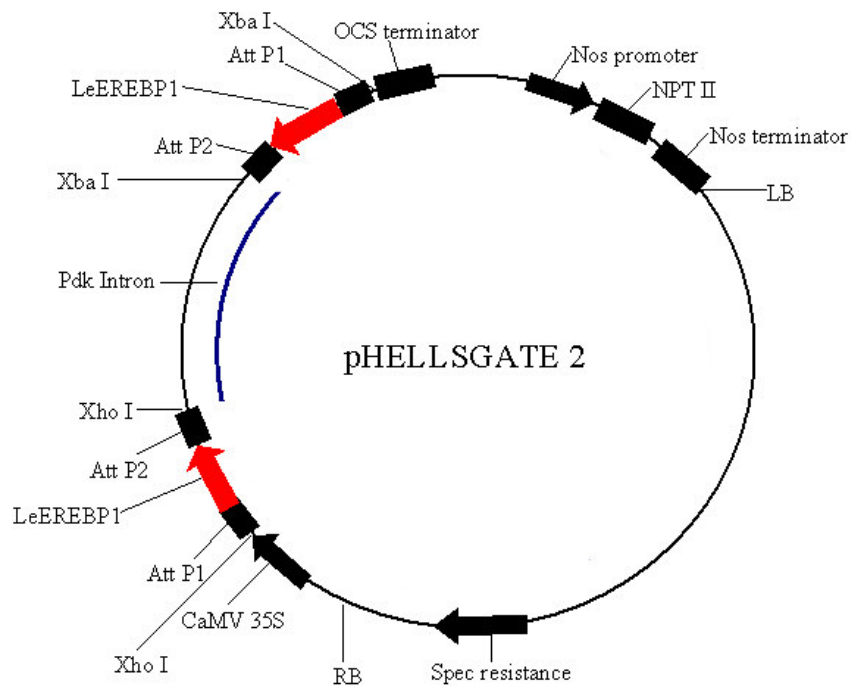


**Figure 6. LeERE BP1 sense construct.** A binary plant transformation vector pBI121 was digested with restriction enzymes SmaI and SacI to replace the GUS gene with the LeERE BP1 cDNA containing the full open reading frame in the sense orientation. NPTII provides selection of transformants on kanamycin. Nopaline synthase promoter and terminator are represented by Nos-pro and Nos-ter while the right and left border of the Ti plasmid is denoted by RB and LB, respectively.

In order to generate an LeEREBP1 RNAi construct, the full length cDNA was amplified with the following primers: 5'- GGGACAAGTTTGTACAA AAAAGCAGG CTACCACCAGCTAAGCGACAGT-3' and 5'- GGGGACCACTTTGTACAAGAAGC TGGGTGCATAAACACAGCACAAGACA-3'. The primers contain LeEREBP1 gene-specific sequences with the addition of flanking attB1 and attB2 sequences (used for Gateway site specific recombination), respectively. The resulting PCR product contained LeEREBP1 cDNA sequences from position 516 to 997 with attB1 and attB2 sequences at either end. The PCR product was gel-purified using the QIAquick gel extraction kit (QIAGEN) and subcloned into vector pHELLSGATE 2 vector (Wesley et al., 2001) by site-specific recombination using the Gateway BP Clonase Enzyme Mix according to manufacturer's instructions (Invitrogen). During this reaction the 5' ccdB unit in the vector was replaced with a sense orientation PCR product and the 3' ccdB unit with an antisense orientation PCR product (Figure 7).

Complete sense and RNAi constructs were transformed into *E. coli* DH 10B cells and the resulting plasmid DNA was isolated and electroporated into *Agrobacterium tumefaciens* strain LBA 4404 (Gibco BRL, MD). The resulting *Agrobacterium* was transferred into wild type tomato plants (*Lycopersicon esculentum* Mill. cv. Alisa Craig) as described in Fillatti et al. (1987).

The sense insert in pBI121 was confirmed by Sac I and Sma I digestion to release the insert and by DNA sequencing of the whole insert in pBI121. The RNAi inserts in



**Figure 7. RNAi construct of LeERE BP1.** PCR products were amplified from LeERE BP1 cDNA using primers that add attB1 and attB2 sites to each end. A single step recombination reaction (BP reaction; utilizing attP1/2 sites) replaces the 5' ccdB unit with a sense PCR product and the 3' ccdB unit with an antisense PCR product. Red arrows represent PCR products of LeERE BP1 which replaced the 5' and 3' ccdB sequences by site-specific recombination.

pHELLSGATE 2 were confirmed by Xho I and Xba I digestion to release the size verified inserts and by sequencing both inserts using insert-specific vector primers.

### ***Confirmation of transgenic plants***

Transgenic plants were confirmed by PCR of genomic DNA and DNA gel-blot analysis. For sense transgenic plants, the primers used for genomic DNA PCR are from the 35S promoter (5'-GGAAAAAGAAGACGTTCCAACCAC-3') and LeEREBP1 (5'-GAATCGACGATCCCTGAAAA-3'). For RNAi transgenic plants, one pair of primers were used to confirm the inserted transgene sequence residing between the Xba I sites (Figure 7). They are from the vector Pdk intron (5'-CGATAAGCTTGGATCCTCTAG-3') and the gene specific LeEREBP1 primer including the attB1 site (5'-GGGACAAGTTTGTACAAAAAAGCAGGCTACCACCAGCTAAGCGACAGT-3'). A separate pair of primers were used to confirm the inserted transgene sequence between the Xho I sites (Figure 7). They are from the vector Pdk intron (5'-GCTGCAGCTGGATGGCAAAT-3') and the LeEREBP1 gene specific primer including the attB1 site (5'-GGGACAAGTTTGTACAAAAAAGCAGGCTACCACCAGCTAAGCGACAGT-3'). DNA gel-blot analysis to verify transgene insertion was as described above except that the NPTII gene was used as the probe.

## **Results**

### ***Sequence analysis of LeEREBP1***

Through queries of both the TIGR tomato gene index and the SGN tomato unigene build, a single cDNA clone containing the full ORF of LeEREBP1 was obtained and

chosen for further characterization. The complete sequence of LeEREBP1 is shown in Figure 8. LeEREBP1 is 1165 bp in length and contains an open reading frame (ORF) starting with the first in-frame ATG at nucleotide 64 and ending with a TGA stop codon at nucleotide 826. This ORF codes for a polypeptide of 254 amino acid residues with a calculated molecular weight of 28.29 kDa (Figure 8). The region from amino acids 91 to 148 (highlighted in black in Figure 8) has high similarity to the conserved AP2/ERF domain, a DNA-binding domain found in APETALA2 and EREBP family transcription factors (Figure 9; Riechmann and Meyerowitz, 1998). Protein sequence comparisons indicate that LeEREBP1 is most similar (80% identity) to a protein encoded by the chilling induced potato (*Solanum tuberosum*) cDNA CIP353 (Mine et al., 2003) and to a partial tobacco (*Nicotiana tabacum*) cDNA (GB#: CAC12822; Shen, unpublished data) having 60% identity to LeEREBP1 (Figure 10).

#### ***Expression of LeEREBP1 in normal and ripening-impaired mutant tomato plants***

RNA gel-blot analysis of tomato leaves, flowers, and fruits at different developmental stages was performed to examine the expression pattern of LeEREBP1 (Figure 11A). LeEREBP1 mRNA was not detectable at the level of resolution of this analysis in leaf tissue. The expression level of LeEREBP1 was low in flower, immature green fruit, mature green fruit, and red ripe fruit (15 days after breaker), but highly induced in breaker stage (early ripening) fruit (Figure 11A).

The expression level of LeEREBP1 in mature green and ripening fruits in normal and ripening mutant backgrounds (*rin*, *nor*, *Nr*) was also investigated (Figure 11B). As in Figure 11A, in wild type tomato fruit the expression of LeEREBP1 was induced at the

```

1   CTCGAGACTA GTTCTCTCTC TTCTACTCAG AAAACAAAAA GAAAAAAAAA AGAGTTAGAA
61  AAGATGTGTG GAGGTGCCAT AATCTCCGAT TATGATCCCG CCGGAAGCTT CTACCCGAAA
    M C G G A I I S D Y D P A G S F Y R K 19
121 CTTTCTGCTC GTGACCTCTG GGCTGAGCTG GACCCTATCT CCGACTACTG GTCCTCTTCT
    L S A R D L W A E L D P I S D Y W S S S 39
181 TCCTCATCCT CAACCGTCGG AAAACCTGAT TCCGCTCTGT CGCCGGTGAC TCACTCCGTC
    S S S S T V G K P D S A L S P V T H S V 59
241 GATAAGCCTA ATAAATCAGA TTCCGGCAAA AAAGGTAATA AGACTGTGAA GGTTGAGAAG
    D K P N K S D S G K K G N K T V K V E K 79
301 GAGAAGAGTA GTGGACCAAG GCCAAGGAAG AACAAGTACA GAGGAATAAG ACAGAGGCCA
    E K S S G P R P R K N K Y R G I R Q R P 99
361 TGGGGAAAAA GGGCTGCTGA GATTCGCGAT CCACAGAAGG GTGTACGCGT TTGGCTTGGT
    W G K W A A E I R D P Q K G V R V W L G 119
421 ACATTCAACA CAGCAGAAGA TGCTGCTAGA GCCTATGATG AGGCTGCTAA GCGCATTTCGT
    T F N T A E D A A R A Y D E A A K R I R 139
481 GGTGATAAGG CTAAACTCAA CTTTCCAGCC CCATCACCAC CAGCTAAGCG ACAGTGCACCT
    G D K A K L N F P A P S P P A K R Q C T 159
541 AGCACTGTGC CTGCTGCTGA TACACCACCA GCACTACTCC TTGAGAGTTC TGACAACCTCT
    S T V A A A D T P P A L L L E S S D N S 179
601 CCTTTGATGA ACTTTGGATA TGATGTCCAG TATCAGAGCC AACTCCCTA CTACCCCATG
    P L M N F G Y D V Q Y Q S Q T P Y Y P M 199
661 GAAATGCCCA TAGTTAGTGA AGATTATGAA CTGAAGGAAC AGATTTCCAA TTTGGAATCG
    V M P I V S E D Y E L K E Q I S N L E S 219
721 TTCCTGGAAT TGGAGCCATC TGATCAATTT TCAGGGATCG TCGATTCTGA TCCTCTTAAT
    F L E L E P S D Q F S G I V D S D P L N 239
781 GTTTTTCTGA TGGAGGACTT TGCTTCAACT CATCATCAGT TCTACTGAAC ATGATCAAAG
    V F L M E D F A S T H H Q F Y * 254
841 TGTTATTTTC GGAGTATTTT ATGGCTGAGT AGTTTATTAA TTAGCTTTTG GGTGTAGTAG
901 TAGTAATAGT AGTAATAGAG ATTAGAGAGT ACATATGATG ATAATAATAA TAAGTTGCGT
961 GCCTTAGCAT GCAATTGTAA TAGTATTAGT GTTTGTGTC TTGTGCTGTT GTTTATGCTT
1021 TTCTAAATCT TGGATTTACC TTATAATGTT TGGTGCATGT ATTTAACTA TATATATATA
1081 TATATATATA TATATATATA TATATATATG GAGTACTTTG TTTAATATT ATCTTTGTTC
1141 ATTCTTAAAA AAAAAAAAAA AAAA 1165

```

**Figure 8. cDNA and predicted amino acid sequence of LeEREBP1.** The predicted amino acid sequence for the open reading frame is shown directly below the nucleotide sequence. Numbers at left indicate nucleotide position while numbers at right indicates amino acid position. The AP2/ERF domain is highlighted in black.



LeEREBP1	91	KYRGIRQRPWGKWAAEIRDPOK-GVRVWLGTFNTAEDAARAYDEAAKRIRGDKAKLNFP	148
CIP353	96	KYRGIRQRPWGKWAAEIRDPOK-GVRVWLGTFNTAEDAARAYDEAAKRIRGNKAKLNFP	153
RAP2.3	101	VYRGIRKRPWGKWAAEIRDPRK-GVRVWLGTFNTAEEAAMAYDVAAKQIRGDKAKLNFP	158
EREBP1	104	HYRGVRRRPWGEFAAEIRDPKNGARVWHRTYETDEEAAIAYDKAAYRMRGSKAHLNFP	162
Pti4	105	HYRGVRQRPWGKFAAEIRDPKNGARVWLGTYETAEEAAIAYDKAAYRMRGSKAHLNFP	163

**Figure 9. LeEREBP1 has homology to AP2/ERF domain plant DNA-binding proteins.**

The deduced amino acid sequence of the AP2/ERP-domain of LeEREBP1 along with those of other AP2/ERF-domain proteins: potato CIP353; Arabidopsis RAP2.3; tobacco EREBP1; and tomato Pti4 were aligned. Amino acids shaded in black represent identical matches while gray-shaded sequences represent conservative changes. Dashes indicate gaps introduced to maximize alignment. ClustalX was used to perform multiple protein sequence alignments and BoxShade was used for background coloring.

```

LeEREBP1      1  MCGGALLISDYDYPAGSFYRKLSARDLWAEILDPIISDYWSSSSSSSTVCKPDSALSPTVTHSVD
CIP353        1  MCGGALLISDYDFAGNFYRKLSARDLWAEILDPIISDYWSSSSSSSTVENFYSAQSPVTHSVD
CAC12822     1  -----ISDFWSSSSSSSIAGKSDSVQSITHSYN

LeEREBP1     61  KPNKSDSGK-----KGNKTVKVEKEKSSGPRPRKNNKYRGIQRFPWGKWAAEIRDPOKGV
CIP353       61  KPKKSDSGKSNQIKKGNKTVKVEKEKSGPRKGNKTVKVEKEKSGPRKNNKYRGIQRFPWGKWAAEIRDPOKGV
CAC12822    30  KPKKSDSGKLNQIEKCTISVKKVEKESSGPRKGNKTVKVEKEKSGPRKNNKYRGIQRFPWGKWAAEIRDPOKGV

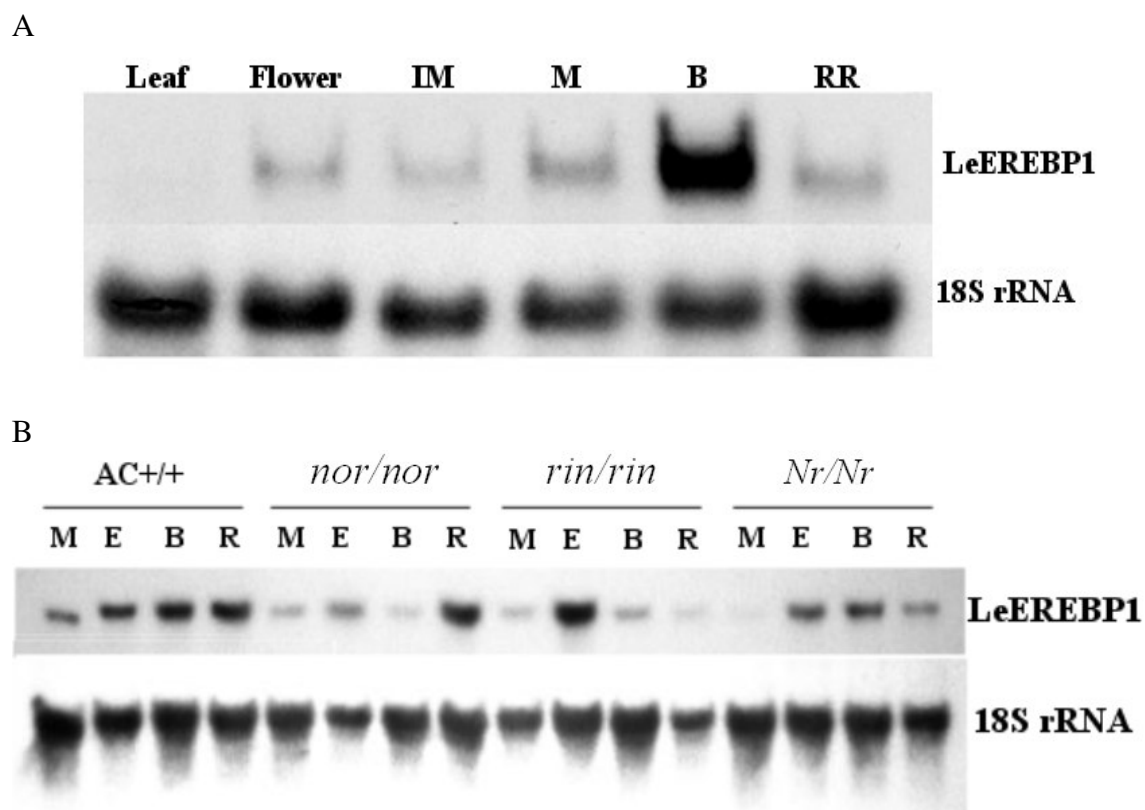
LeEREBP1     116 VWLGTFTNTAEDAARAYDEAAKIRIGDKAKLNFPPAP--SPPAKROCTSTVAAADTTPPALLLE
CIP353      121  VWLGTFTNTAEDAARAYDEAAKIRIGDKAKLNFPPAP--SPPAKROCTSTVAAADTTPPALLLE
CAC12822    89  VWLGTFTNTAEAAARAYDEAAKIRIGDKAKLNFPPAP--SPPAKRHCCTSTPDPQPTRSDLNLSQ

LeEREBP1     175  SS-----DNSPLMNFGYDVQYQSQTPLYPMEMPVYSE-----DYELKEQISNLESTLELEP
CIP353      179  SSNIISYNNSPLMNFGYDVQ--SQTFYPMEMPVASD-----DYELKEQISNLESTLELEP
CAC12822   149  KP-----ASTMLNNGYENQ-----TPYPMEMPVAAEDPQHHDYELKEQISNLESTLELEP

LeEREBP1     226  SD---QFSGIVDSDPLNNVFLMEDFASVTHQ---FY
CIP353      233  ADSDDQFSGIVDPPDPLNNVFLMEDFASVTHQ---FY
CAC12822   199  DS-----GIVDSDPLNIFLMDDFAAVQQQLFY

```

**Figure 10. Alignment of the amino acid sequences of the LeEREBP1 ORF with those of closely related proteins:** CIP353 from potato, CAC12822 from tobacco. Amino acid shaded in black represents identical matches and gray-shaded represent conservative changes. Dashes indicate gaps introduced to maximize alignment. ClustalX was used to perform multiple protein sequence alignments and BoxShade was used for sequence background coloring.



**Figure 11. RNA gel-blot analysis of LeERE BP1.** A, Expression of LeERE BP1 in leaf, flower, immature green (IM), mature green (M), breaker (B), and B + 15 days (RR) fruit. B, Expression of LeERE BP1 in M, M after 12 hours exposure to 10 ppm ethylene (E), B, and B + 7 days (R) fruit from wild type and ripening mutant lines. Total RNA (10  $\mu$ g per lane) was electrophoresed under denaturing conditions, blotted onto a nylon membrane, and probed with a  $^{32}$ P-labeled probe derived from the 3'-untranslated region of LeERE BP1. 18s rRNA probe was used as a loading control

breaker stage and persists through at least 7 days post breaker (note that 15 days post breaker was used in 11A). Combining the data for LeEREBP1 expression in Figures 11A and 11B it can be concluded that expression of LeEREBP1 is induced at breaker stage and is maintained for at least seven days post breaker stage before dropping precipitously by 15 days post breaker.

In the homozygous *nor* mutant, expression of LeEREBP1 was very low in mature green and breaker stage fruits, yet was induced at seven days after breaker stage. While in the homozygous *rin* mutant, the expression level of LeEREBP1 was constitutively low during fruit development through seven days after the age where breaker stage would have occurred. In the homozygous *Nr* mutant, expression of LeEREBP1 was undetectable in mature green fruits, and moderately induced in breaker fruit and in fruits seven days older than normal breaker stage. These results suggest that LeEREBP1 may be regulated in part by ethylene, while elevated expression in the non-climacteric *nor* mutant suggests a developmental component of regulation of this gene as well.

To clarify whether or not LeEREBP1 is influenced by ethylene during ripening, and because some AP2/ERF domain containing proteins are known to be induced by ethylene (Buttner and Singh, 1997; Fujimoto et al., 2000; Gu et al., 2000), I exposed pre-climacteric mature green fruit from wild type, *rin*, *nor*, and *Nr* genotypes to 10 ppm ethylene for 12 hours. The expression of LeEREBP1 was induced to varying degrees in mature green fruits following 12-hour ethylene treatment in all four genotypes.

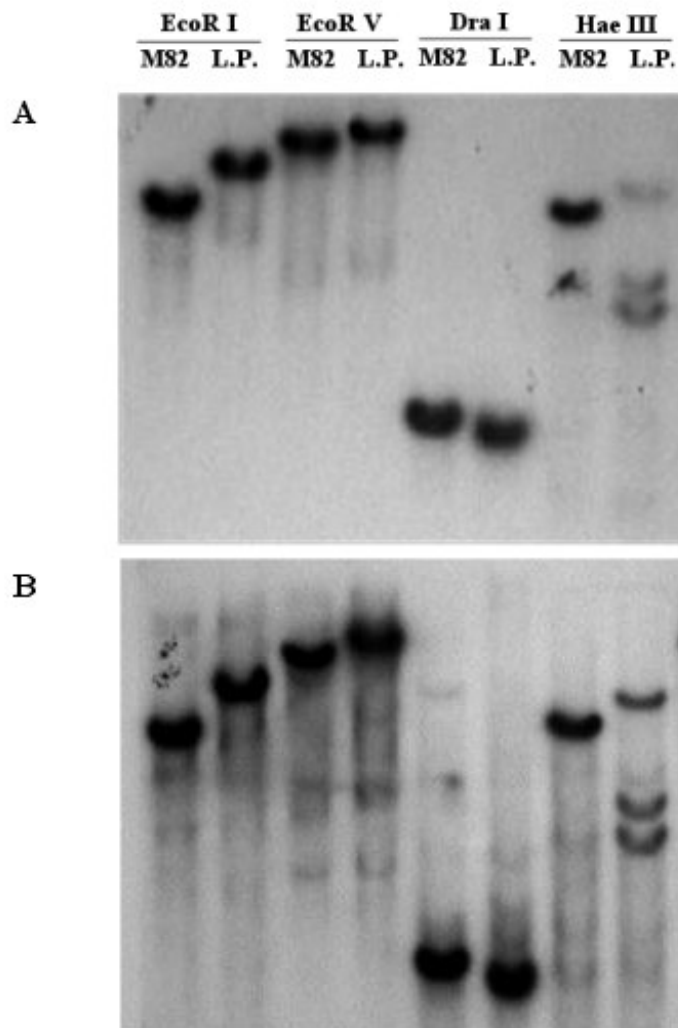
### ***DNA gel-blot analysis and genetic mapping of LeEREBP1***

Because plant AP2/EREBP genes exist as a multigene family, DNA gel blots were performed to investigate the copy number of LeEREBP1 in tomato (*L. esculentum*, cv M82) and the wild relative *L. pennellii*. Using the 3'-untranslated region of LeEREBP1 as probe, a single strong band was observed in M82, indicating that LeEREBP1 is a low copy gene. Analysis of *L. pennellii* suggested one to three copies of LeEREBP1 could be found in its genome (Figure 12A). Interestingly, although LeEREBP1 belongs to a multigene family, a similar result was obtained using the full length LeEREBP1 cDNA as probe (Figure 12B). Reducing stringency did not alter this result (hybridized and washed at 50°C; data not shown) indicating the DNA sequence of LeEREBP1 is considerably diverged from other members of the tomato AP2/ERF family.

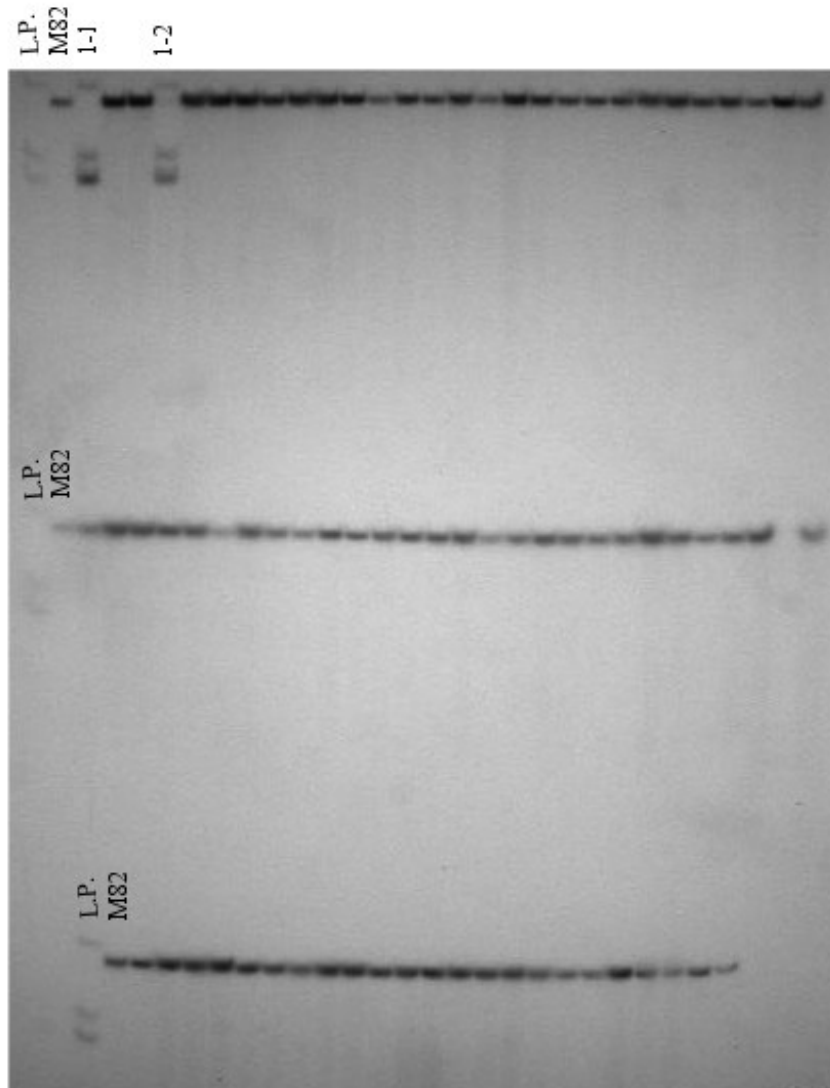
Figure 12 shows that a restriction fragment length polymorphism for LeEREBP1 occurs between M82 and *L. pennellii* when the genomic DNA was digested with Hae III. This RFLP was used to map LeEREBP1 in the population of 76 ILs whose genomic DNA was also hybridized with LeEREBP1 (3' UTR region). IL line 1-1 and 1-2 show RFLPs (Figure 13), indicating that chromosome segments from *L. pennellii* in both IL 1-1 and 1-2 contain the LeEREBP1 locus specifically at the region designated 1-D ([http://www.sgn.cornell.edu/maps/pennellii\\_il/chr1.html](http://www.sgn.cornell.edu/maps/pennellii_il/chr1.html)).

### ***Confirmation of transgenic plants***

LeEREBP1 sense and RNAi constructs were developed as described in the Materials and Methods section. Following plant transformation and kanamycin selection,



**Figure 12. DNA gel-blot analysis of LeERE BP1.** Each lane contains 15  $\mu\text{g}$  genomic DNA extracted from leaves of tomato M82 and *L. pennellii* following digestion with the indicated enzyme. Genomic DNA were size-fractionated on an agarose gel, blotted and hybridized with the  $^{32}\text{P}$ -labeled 3' UTR region of LeEBERP1 (A) and the full length LeERE BP1 cDNA (B).



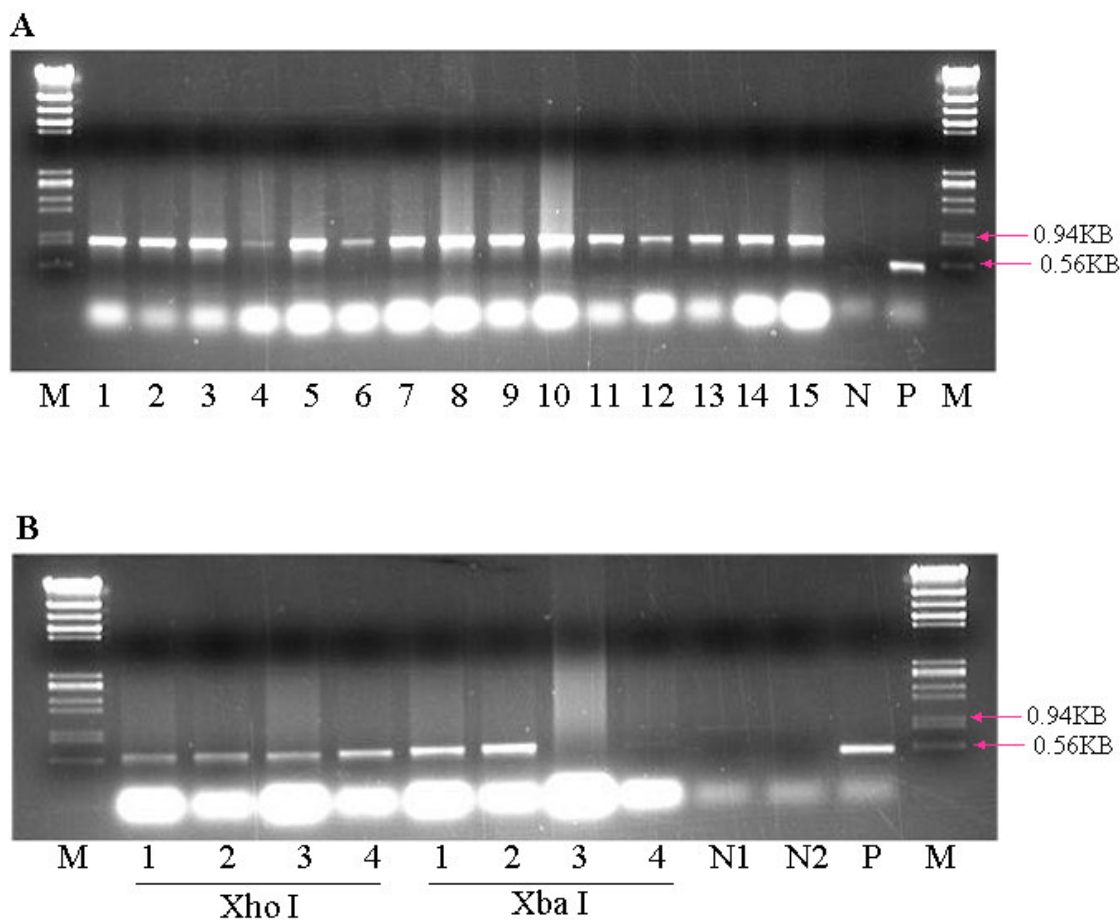
**Figure 13. Genetic mapping of LeEREBP1.** Genomic DNA from 76 *L. pennellii*-derived ILs and their parents, M82 and *L. pennellii*, were digested with Hae III, size-fractionated on an agarose gel, blotted and hybridized with the  $^{32}\text{P}$ -labeled 3' UTR region of LeEBERP1.

seventeen independent transformants harboring the LeEREBP1 sense construct and nine transformed with the RNAi construct were obtained. To confirm whether or not these plants contained the corresponding transgenes, genomic DNA was extracted from expanding young leaves of each putative transformant followed by PCR and DNA gel-blot verification of transgene integration. At present, fifteen sense transgenic plants and four RNAi plants are mature enough for PCR confirmation, while 10 sense transgenic lines and one RNAi plant are mature enough for DNA gel-blot analysis.

To check for transgene integration in the putative sense transgenic plants by PCR, one primer from the CaMV 35S promoter region of vector pBI121 and one LeEREBP1 specific primer, were employed. This primer pair should generate a PCR product with a predicted size of 909 bps. In addition, a positive control was used which amplifies from non-transformed lines via a pair of LeEREBP1 specific primers to yield a PCR product with a predicted size of 560 bps. All fifteen sense transgenic plants tested positive by this screen for transgene integration (Figure 14A).

To test for transgene integration in RNAi transgenic plants, a strategy was employed to confirm whether or not both copies were integrated into the target tomato genome. To check the insert at the Xho I site, a primer from the Pdk intron of the vector and a LeEREBP1 specific primer were used to generate a PCR product predicted to be 643 bps. To check the insert at the Xba I site, a different primer from the Pdk intron of the vector and the same LeEREBP1 specific primer were used to generate a PCR product predicted to be 670 bps. All four independent RNAi transgenic plants yielded





**Figure 14. PCR confirmation of transgenic plants.** Genomic DNA was extracted from fifteen independent sense lines (A) and four independent RNAi lines (B). PCR was performed using primers described in the material and methods section. (A) 1 to 15 denotes the fifteen sense lines, N denotes the negative control, P denotes the positive control, and M denotes the HHR MW marker. (B) 1 to 4 denotes the four RNAi lines, N1 and N2 denote negative controls, P denotes the positive control, and M denotes the HHR MW marker. Xho I and Xba I denote the amplification target regions.

the expected PCR product from the Xho I site, while only two lines also yielded the corresponding Xba I site PCR product of expected size (Figure 14B).

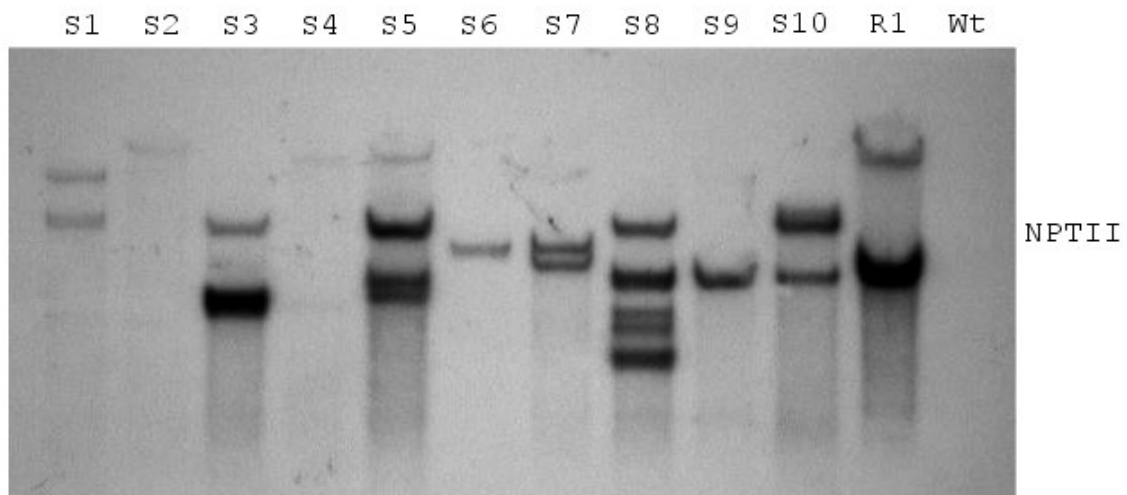
To further confirm transgene integration in both sense and RNAi lines, DNA gel-blot analysis was performed. The NPTII gene was used as probe for both sense and RNAi plants (Figure 15). All ten sense lines and one RNAi line were confirmed for transgene integration to date.

### ***RNA gel-blot analysis of transgenic plants***

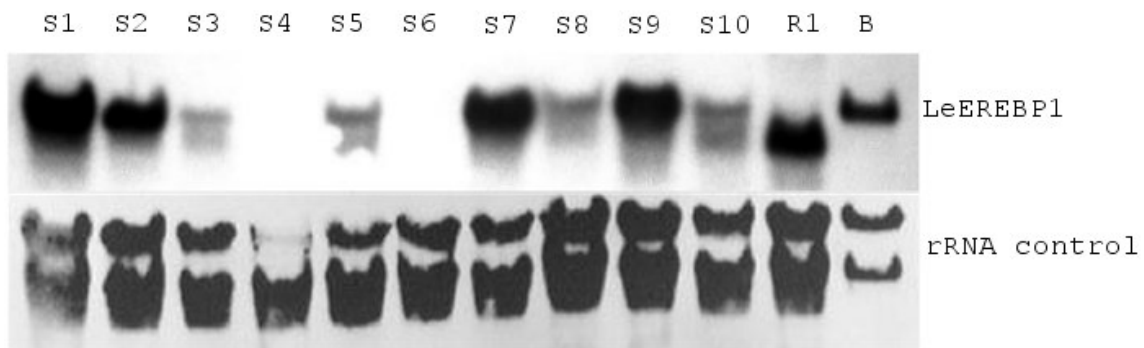
Since the expression of LeEREBP1 is undetectable in leaf tissue from wild type plants by RNA gel blot analysis (Figure 11), sense overexpression transgenic plants can be easily determined by analyzing the LeEREBP1 level in leaves from each transgenic line. In addition, exogenous transcripts from the RNAi construct, which are smaller than endogenous LeEREBP1 transcripts, can also be distinguished in leaf tissue. Ten sense transgenic lines and one RNAi transgenic line were assayed for expression via RNA gel-blot analysis (Figure 16). Among the ten sense transgenic lines, four (line S1, S2, S7, and S9) have significantly increased expression of LeEREBP1, while four others (line S3, S5, S8, and S10) show only slightly increased expression of LeEREBP1, and the remaining two (S4, S6) show no expression of LeEREBP1. In the one tested RNAi line (R1), as anticipated the expression of the endogenous LeEREBP1 is not detected, while the exogenous transcript from the RNAi construct is highly prevalent (Figure 16).

### ***Future studies***

The remaining putative transgenic plants need to be confirmed by PCR and DNA/RNA gel-blot analysis. The expression level of LeEREBP1 in fruit (mainly at



**Figure 15. DNA gel-blot confirmation of transgenic plants.** Genomic DNA was extracted from ten independent sense transformation lines (S1 to S10), one RNAi transgenic line (R1) and a wild type non-transformed tomato plant (Wt). Genomic DNA was digested with EcoR V and size-fractionated on a 1% agarose gel, blotted and hybridized with a  $^{32}\text{P}$ -labeled NPTII gene as probe.



**Figure 16. RNA gel-blot analysis of LeEREBP1 expression in transgenic plants.** Total RNA was extracted from leaves of ten independent sense transgenic lines (S1 to S10), one RNAi transgenic line (R1), and from fruits at breaker stage of a wild type non-transformed line (B). RNA was size fractionated by electrophoresis under denaturing conditions, blotted to a nylon membrane and hybridized with a  $^{32}\text{P}$  labeled-LeEREBP1 probe.

breaker stage) will initially be investigated for all the transgenic plants to assay transgene effects. The number of inserts will also be determined for each transgenic line by DNA gel-blot hybridization. Currently all transgenic plants are at the T<sub>0</sub> generation. After obtaining homozygous progeny (T<sub>1</sub>) with increased or decreased LeEREBP1 expression, physiological and biochemical analysis will be performed to determine the impact of LeEREBP1 on fruit development and ripening. Examples of parameters monitored will include but not be limited to time to ripening, time for ripening to reach completion, average fruit diameter and mass, ethylene evolution, seedling triple response, carotenoid accumulation, and expression of ripening-related genes such as polygalacturonase, ACC oxidase, and phytoene synthase. Lines that demonstrate reproducible phenotypes will be further characterized via microarray expression analysis to more fully characterize the function and effects of LeEREBP1.

## **Discussion**

A ripening-induced putative ethylene signaling transcription factor, LeEREBP1, was isolated as described in Chapter II and characterized here. As shown in Figure 9, the central part of the deduced amino acid sequence (from residues 91 to 148) has high similarity to the conserved AP2/EREBP DNA binding domain (Riechmann and Meyerowitz, 1998). Proteins from the AP2/EREBP family represent transcription factors unique to plants. They are a group of novel DNA-binding proteins, neither showing homology to known DNA-binding proteins and transcription factors nor containing leucine zipper or zinc finger motifs (Hao et al., 1998). AP2/EREBP genes form a large

family and can be divided into two subfamilies based on whether the proteins contain one or two AP2 domains, the EREBP subfamily and the AP2 subfamily, respectively. LeEREBP1 contains one AP2 domain (Figure 8) so by definition belongs to the EREBP subfamily. Several EREBP gene products have been shown to bind to a cis-regulatory sequence widely conserved among ethylene-responsive pathogenesis-related (PR) genes, suggesting a role for EREBPs in plant disease resistance (Ohme-Takagi and Shinshi, 1995; Buttner and Singh, 1997; Zhou et al., 1997; Gu et al., 2000; Gu et al., 2002). The Arabidopsis ERF1 gene, which is an EREBP, can directly interact with EIN3, suggesting a role in the ethylene signaling pathway (Solano et al., 1998).

While ethylene plays a significant role in tomato fruit ripening, an EREBP associated with this process remains to be identified. Expression analysis of LeEREBP1 suggests a possible role in ripening and regulation by both ethylene and developmental factors. This mode of regulation is found in several key ripening-related genes including E8 (Lincoln and Fischer, 1988) and the NR ethylene receptor (Wilkinson et al., 1995).

In wild type plants, the expression of LeEREBP1 is undetectable in leaf, and very low in flower, immature green fruit, and mature green fruit. LeEREBP1 is highly induced at breaker stage and continues to persist at later stages of fruit ripening (seven days after breaker) and then drops to a low level in late ripening (fifteen days after breaker) (Figure 11). The expression level of LeEREBP1 in mature green fruit is highly induced by 12-hours of exogenous ethylene treatment (Figure 11B). The expression profile of LeEREBP1 is similar to that of previously described ripening-related genes

including E4, E8, and NR (Biggs et al., 1986; Lincoln et al., 1987; Lincoln and Fischer, 1988; Wilkinson et al., 1995).

Homozygous mutant *rin* fruit do not ripen in response to exogenous ethylene, yet display induction of at least some ethylene responsive genes, indicating retention of fruit ethylene sensitivity (Lincoln and Fischer, 1988). As expected, LeEREBP1 mRNA accumulation is markedly reduced in *rin*, but is also highly induced by ethylene (Figure 11B).

Like *rin*, *nor* is also a single locus mutation and the mutant displays inhibition of all measurable ripening phenomena. LeEREBP1 mRNA accumulation is very low in mature green and breaker age *nor* fruit, but induced in red ripe age fruit (still unripe) suggesting contribution to ripening expression by factors other than ethylene. LeEREBP1 expression is slightly induced in *nor* by exogenous ethylene (Figure 11B).

Ripening-associated gene expression has been reported to be ~10 to 30% of normal levels in *Nr* fruit (Dellapenna et al., 1989; Knapp et al., 1989). The *Nr* mutant is ethylene insensitive and *Nr* fruit does not ripen fully, even in the presence of exogenous ethylene (Lanahan et al., 1994). The NR gene was cloned and encodes a member of the ethylene receptor family (Wilkinson et al., 1995; Tieman et al., 2000). As expected, the expression of LeEREBP1 in *Nr* mutant fruit is reduced during fruit development and in response to exogenous ethylene treatment as compared to wild type (Figure 11B).

In order to elucidate how LeEREBP1 might participate in fruit ripening, I developed constructs and generated transgenic plants containing LeEREBP1 sense or RNAi constructs. Seventeen independent sense lines and nine independent RNAi lines

have been obtained from tissue culture to date and most were PCR and DNA gel-blot confirmed (Figures 14 and 15). Among putative sense lines investigated, four show significant overexpression of LeEREBP1 transcript in leaf tissue. Currently all transgenic plants are in the T<sub>0</sub> generation. Homozygous progeny (T<sub>1</sub>) of transgenic plants with increased and reduced expression of LeEREBP1 will be obtained. Parameters related to tomato fruit ripening such as time to ripening, expression of known ripening-related genes, and carotenoid accumulation during fruit development will be analyzed. In addition, since it is reported that EREBPs participate in plant disease resistance (Ohme-Takagi and Shinshi, 1995; Buttner and Singh, 1997; Zhou et al., 1997; Gu et al., 2000; Gu et al., 2002), analysis of susceptibility of these transgenic plants to plant pathogens will be pursued through future collaborations.

An EREBP gene, ERF1, which can directly interact with EIN3 and is directly involved in the ethylene signal transduction pathway, was cloned from Arabidopsis (Solano et al., 1998). To test biochemically whether LeEREBP1 is a component of ethylene signaling pathway, it would be valuable to determine whether or not interaction of LeEREBP1 and LeEILs (Tieman et al., 2001) can occur via two-hybrid analysis.



## CHAPTER IV

### DIGITAL EXPRESSION ANALYSIS OF TOMATO EST LIBRARIES

#### Summary

A large tomato expressed sequence tag (EST) dataset was mined to gain insights into differential gene expression among diverse plant tissues representing a range of developmental programs and biological responses. The EST collection generated from tomato contains high quality sequences of 152,635 distinct ESTs from 27 different cDNA libraries. These ESTs were clustered and assembled to a total of 31,012 unique genes with 15,925 tentative consensus (TCs) sequences and 15,087 singletons. To better understand tomato gene expression at a global level and to identify differentially expressed and tissue specific genes, we developed and implemented a digital expression analysis protocol. By clustering genes according to their relative abundance in the various EST libraries, expression patterns of genes across various tissues were generated and genes with similar patterns were grouped. In addition, tissues themselves were clustered for relatedness based on relative gene expression as a means of validating the integrity of the EST data as representative of relative gene expression. Statistical analysis of EST prevalence in the fruit and pathogenesis-related libraries resulted in 333 genes being classified as fruit ripening-induced, 185 as fruit ripening-repressed, and 169 as pathogenesis-related ( $P < 0.05$ ). By choosing a significance level of  $P < 0.01$ , we estimated that approximately 1379 and 370 loci in the tomato genome are differentially expressed during the stages analyzed for fruit and flower development, respectively.

Tomato fruit digital expression data was also compared to publicly available grape EST data to gain insight into molecular manifestation of ripening processes governing ripening regulation across species and of climacteric versus non-climacteric fruits. In addition, an online database was created to allow researchers to mine tomato EST expression data interactively.

## **Introduction**

Expressed sequence tags (ESTs) are created by partially sequencing randomly isolated gene transcripts that have been converted into cDNA and cloned (Adams et al., 1991). ESTs have played significant roles in accelerating gene discovery including gene family expansion (Papadopoulos et al., 1994; Rogaev et al., 1995; Bourdon et al., 2002), facilitating large-scale gene expression analysis (Ewing et al., 1999; Ewing and Claverie, 2000; Ogihara et al., 2003), and elucidating phylogenetic relationships (Nishiyama et al., 2003). Recent developments in high volume biotechnology combined with advanced DNA sequencing technology have made it feasible to perform large-scale EST projects. Currently there are approximately 17 million ESTs in the NCBI public collection – dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST/>). With many large-scale EST sequencing projects in progress and new projects being initiated, the number of ESTs in the public domain will continue likely to increase substantially for the foreseeable future.

One obvious outcome of large-scale EST projects is that ESTs provide a comprehensive window into the transcriptome as contrasted to whole genome sequences which yield information about genome content and organization but not expression. It

has been shown previously that EST databases are a valid and reliable source of gene expression data (Ewing et al., 1999; Ewing and Claverie, 2000; Ogihara et al., 2003). With the rapid expansion of available EST data, opportunities for digital gene expression analysis will continue to expand. As a result of advances in computational molecular biology and biostatistics, it is possible to mine and analyze large-scale EST datasets efficiently and exhaustively (Ewing et al., 1999; Ewing and Claverie, 2000; Ogihara et al., 2003).

Tomato is a member of the family Solanaceae which includes several additional economically important crops such as potato, pepper, and eggplant and as such represents the most valuable family in terms of vegetable crops. Tomato has long served as a model system for plant genetics, development, physiology, pathology and fruit ripening resulting in the accumulation of substantial information regarding the biology of this economically important organism. Many experimental tools and features of tomato make it an excellent model system including: extensive germplasm collections, numerous natural, induced, and transgenic mutants, routine transformation technology, a dense and expanding RFLP map, relatively short life-cycle and ease of growth and maintenance. A large, publicly available tomato EST database has been generated with support from the National Science Foundation Plant Genome Program (<http://www.tigr.org/tdb/tgi/lgi>; <http://www.sgn.cornell.edu>) (Quackenbush et al., 2001; Van der Hoeven et al., 2002). Several recently developed bioinformatics and statistical tools have allowed us to perform global expression analysis using EST data generated under this and other projects (Audic and Claverie, 1997; Ewing et al., 1999; Greller and

Tobin, 1999; Stekel et al., 2000). Here we present correlated expression profile for 6,758 genes across 25 different tissues. In addition, we identified sets of differentially expressed genes associated with pathogen response and plant development with an emphasis on fruit ripening. We also performed similar analysis with available grape ESTs derived from fruit and compared fruit ripening-induced genes between tomato and grape to gain insights into common and distinctive features of climacteric and non-climacteric fruits at the molecular level. A public database that allows researchers to mine the tomato EST expression data interactively was created as a result of this study.

## **Methods**

### ***Datasets acquisition and data analyses***

The EST dataset and the Unigene dataset of the TIGR Tomato Gene Index (LeGI; <http://www.tigr.org/tdb/tgi/lgi/>) release 9.0 and the TIGR Grape Gene Index release 2.0 (VvGI; <http://www.tigr.org/tdb/tgi/vvgi/>) were provided through the TIGR ftp server. GenBank accession numbers of all the ESTs in VvGI were used to query against the GenBank database to obtain tissue information for each clone. For the purpose of digital expression analysis, only the 5' sequence for each EST was used as this is the predominant sequence information for both the tomato and grape EST collections. All sequence data were loaded into a local MySQL relational database for manipulation.

The unigene dataset from both LeGI and VvGI were electronically annotated by performing sequence similarity searches against the NCBI nr database using the

BLASTX function with a cutoff E value at  $1e-10$ . BLASTX analysis was performed on the Cornell CBSU (Computational Biology Service Unit) parallel blast driver (pblast, <http://ser-loopp.tc.cornell.edu/cbsu/pblast.htm>) on a MS Windows PC computer cluster (128 Pentium III 1GHz CPUs) at the Cornell Theory Center.

A set of in-house perl scripts were written and used for data management and extraction from the MySQL database, for parsing the blast results and for statistical calculations.

### ***Coordinated gene expression analysis of tomato***

cDNA libraries with more than 1000 ESTs and TCs (EST contigs) with at least five members were selected for digital gene expression analysis. A two-way matrix table of raw EST counts was created with rows corresponding to contigs (genes) and columns corresponding to cDNA libraries. Raw data were normalized by: 1) transforming the raw EST counts into relative expression values as defined by Schmitt et al. (1999) as the ratio of homologous ESTs to the total number of ESTs in the corresponding pool. 2) The expression values for each gene was then normalized to itself by dividing all gene expression values for a given gene by the median of its expression over all the 25 tissues. If the median of any given gene was less than one then the number one was used for this normalization.

All genes were clustered using the Gene Tree classification and all conditions (cDNA libraries) were clustered using the Experiment Tree classification of GeneSpring software (v5.0; Silicon Genetics, Redwood City, CA, USA). In both classifications,

Pearson correlation was used as measurement of similarity with the minimum distance set at 0.001.

### ***Identification of differentially expressed genes***

AC statistics (Audic and Claverie, 1997) and general  $\chi^2$  test were used to identify differentially expressed genes for pair-wise and multiple cDNA library comparisons, respectively as described in Romualdi et al. (2001). P values for AC statistics and general  $\chi^2$  test were calculated using the IDEG6 program (<http://telethon.bio.unipd.it/bioinfo/IDEG6/>) or an in-house perl script.

Ripening-induced, ripening-repressed, and pathogenesis-related genes identified in this study were subdivided into functional categories. Functional classification was based on the detailed information of each TC obtained by querying the Swiss-Prot and TrEMBL protein knowledgebases (<http://us.expasy.org/sprot/>) using the putative function/product/gene name of each TC as the query term. The putative function/product/gene name was obtained by blasting each gene against the GenBank nr database with a cutoff e value of 1e-10.

### ***Tomato digital expression database***

The tomato digital expression database was implemented as a relational database using a MySQL database management system. Perl/CGI was used to implement client-server communication, in conjunction with the DBI module allowing scripts to connect to the back-end database. The tomato digital expression database operates on a Windows 2000 advanced server and can be accessed over the Internet using a web browser (<http://ted.bti.cornell.edu/digital>).

## **Results**

### ***EST foundation for tomato digital expression analysis***

154,054 high quality sequences representing 152,635 distinct ESTs from 27 different cDNA libraries (Table 7) were generated under two NSF-funded tomato functional genomics projects (see acknowledgements). cDNA clones were sequenced at TIGR and cleaned to remove untrimmed vector, linkers, low quality regions, and poly A/T tails as well as vector contaminated sequences. These ESTs were then used to generate the TIGR Tomato Gene Index (LeGI; <http://www.tigr.org/tdb/tgi/lgi>), which incorporates both new ESTs and tomato sequences downloaded from dbEST and GenBank. Individual sequences were clustered using stringent criteria (minimum 94% identity over a 40 nt or longer region with fewer than 30 bases of mismatched sequences at either end). Each cluster was then assembled separately at high stringency using the Paracel Transcript Assembler version 2.6.2 to produce TCs (<http://www.paracel.com>; Huang and Madan, 1999). After clustering and assembly, these ESTs were reduced to 31,012 unique genes, of which 15,925 are TCs and 15,087 are singletons. The TIGR Tomato Gene Index/unigene set is available at TIGR (<http://www.tigr.org/tdb/tgi/lgi>) and the Solanaceae Genome Network (<http://www.sgn.cornell.edu>).

### ***Identification of highly abundant and tissue specific genes***

In digital expression analysis, the relative abundance of a gene is defined simply as the ratio of homologous ESTs to the total number of ESTs in the corresponding pool (Schmitt et al., 1999). Singletons are likely to represent relatively low abundance transcripts, but due to their rarity are difficult to characterize further via digital

**Table 7.** Tomato cDNA library statistics

Cat#	Library description	No. of ESTs	No. of TCs	No. of singletons (% <sup>a</sup> )
T1045	ovary	9878	3072	623 (16.9)
T10018	developing/immature green fruit	4240	2287	376 (14.1)
T1356	mature green fruit	5317	2540	658 (20.6)
T1775	breaker fruit	15207	4022	815 (16.8)
T1391	red ripe fruit	3895	1605	238 (12.9)
T1526	flower, 0-3mm buds	6259	3018	802 (21.0)
T1527	flower, 3-8mm buds	5524	2597	523 (16.8)
T1528	flower, 8mm-preanthesis buds	5759	2665	575 (17.7)
T1529	flower, anthesis (open flower)	5643	2630	659 (20.0)
T10393	flower library from a mixture of developmental stages (mixed flower)	7009	2799	884 (24.0)
T10227	wild tomato pollen	5426	796	440 (35.6)
T1079	pseudomonas susceptible	5243	2469	485 (16.4)
T1080	pseudomonas resistant	5127	2289	438 (16.1)
T1297	mixed elicitors	9135	3138	599 (16.0)
T10304	shoot/meristem	9122	3930	901 (18.7)
T1005	shoot	898	627	115 (15.5)
T1481	root, plant at pre-anthesis	3259	1775	434 (19.6)
T1480	root, plant at fruit set	3142	1782	405 (18.5)
T1450	nutrient deficient root	3175	1788	506 (22.1)
T1482	root, etiolated radicle	2373	1436	205 (12.5)
T1437	germinating seed	3927	2004	331 (14.2)
T1048	quiescent seed	547	256	183 (41.7)
T1451	L. pennellii trichome	2729	1448	579 (28.6)
T1452	L. hirsutum trichome	2457	1236	463 (27.3)
T1207	callus tissue	14114	4440	1288 (22.5)
T10600	suspension culture, untreated	8026	2821	579 (17.0)
T10284	crown gall	5204	2749	624 (18.5)
<b>TOTAL</b>		<b>152635</b>		

<sup>a</sup> Calculated as the ratio of the number of singletons to the number of total unique sequences in the corresponding library.



expression analysis. In each of the 27 tomato cDNA libraries analyzed, singletons represented from 12.5% to 41.7% of the unique genes in the library, while in most libraries, singletons represented less than 20% of the unique genes (Table 7). In contrast, Table 8 lists the ten most abundant transcripts in the combined tomato EST database including genes known to be highly expressed, for example, in photosynthetic tissues (TC123709, TC123753, TC113710, encoding RuBisCO subunits), or ripening fruit (TC123821, beta-fructosidase, Elliott et al., 1993).

The 27 cDNA libraries (Table 7) were classified into nine different tissue categories: leaf, shoot, fruit, root, flower, seed, trichome, callus, suspension culture, and crown gall. The most abundant transcript in each tissue is shown in Table 9. Many of the most abundant genes represent those that would be expected to be highly expressed in the respective tissue based on previous reports (Table 9), thus supporting the underlying assumption of digital gene expression that relative abundance in EST libraries is indeed reflective of relative gene expression.

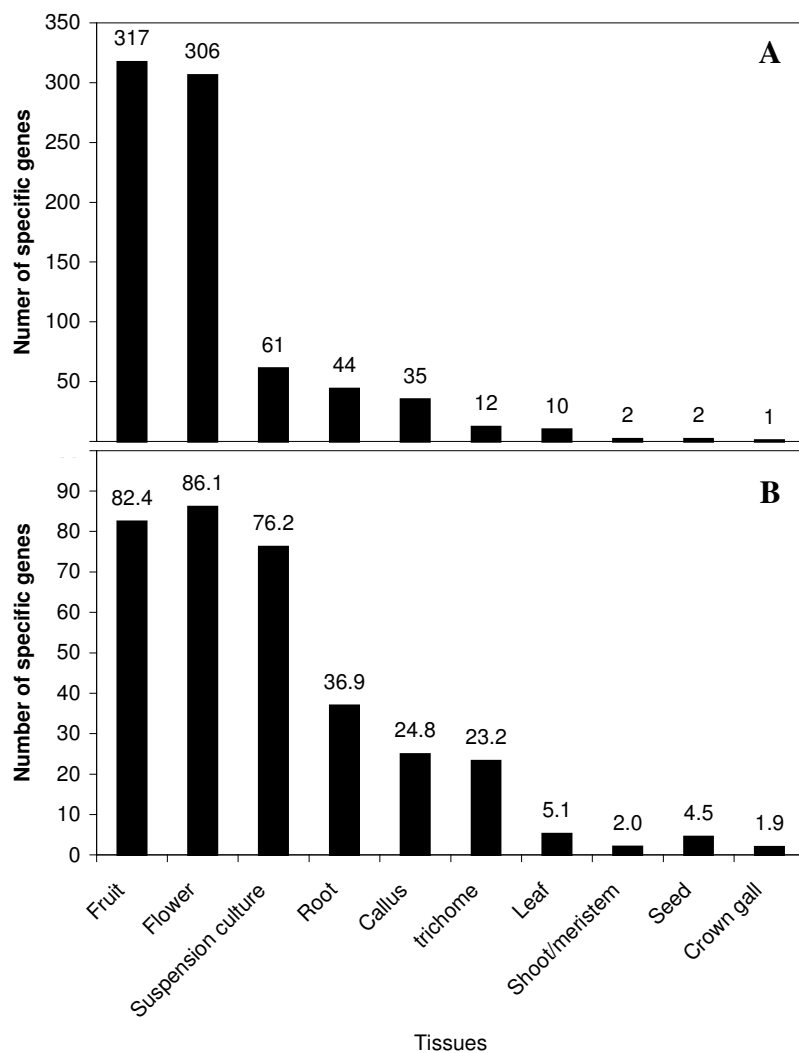
Genes that are primarily expressed in one tissue or in response to a particular stimulus are of special interest relative to the unique biology of said tissue/response and in terms of relationship to potential tissue/response-specific promoters. In order to more accurately identify genes expressed in a tissue-specific manner, we focused on TCs with a minimum of 4 EST members from a single tissue ( $P < 0.05$ ) (Audic and Claverie, 1997). Figure 17 depicts the number of genes uniquely expressed in each tissue based on this criteria. Fruit and flower have far more specific genes than other tissues, with 317 and 306, respectively (Figure 17A), in part due to the large numbers of ESTs recovered

**Table 8.** The ten most abundant genes in tomato EST database

<b>TC #</b>	<b>No. of ESTs</b>	<b>Putative function</b>	<b>E value</b>
TC123709	887	RuBisCO small subunit 3A/3C	6e-097
TC123753	766	RuBisCO small subunit 2A	1e-096
TC123710	567	RuBisCO small subunit 3B	e-100
TC123803	510	Glucan endo-1,3-beta-glucosidase	0.0
TC115845	447	Translationally controlled tumor protein homolog	1e-082
TC115811	364	unknown	
TC115865	326	Catalase isozyme 1	0.0
TC123867	311	heat shock protein 80	0.0
TC123821	302	beta-fructosidase	0.0
TC123778	296	elongation factor 1-alpha	0.0

**Table 9.** The most abundant genes in various tissues in the EST collection

<b>Tissue</b>	<b>TC</b>	<b>No. of ESTs</b>	<b>Putative function</b>	<b>E value</b>	<b>Reference</b>
Fruit	TC123821	284	beta-fructosidase	0.0	Elliott et al. (1993)
Flower	TC115920	250	Pectinesterase	e-139	Mu, et al. (1994)
Leaf	TC123709	676	RuBisCO small subunit 3A/3C	6e-097	
Shoot	TC123709	115	RuBisCO small subunit 3A/3C	6e-097	
Root	TC123896	64	S-adenosylmethionine synthetase	0.0	Espartero et al. (1994)
Seed	TC124040	108	lipase	e-119	
Trichome	TC116136	69	sesquiterpene synthase 1	e-136	van der Hoeven et al.(2000)
Callus	TC123803	457	Glucan endo-1,3-beta-glucosidase	0.0	
Suspension culture	TC115994	111	P-rich protein Nt-SubC29	2e-020	Hoeberichts et al. (2001)
Crown Gall	TC115774	36	rRNA intron-encoded homing endonuclease	5e-021	



**Figure 17. Number of specific genes in each tissue.** TCs with at least four members and only expressed in one of the ten tissue classes were defined as tissue specific genes. (A) Total number of specific genes in each tissue; (B) Relative number of tissue specific genes. Total number of specific genes from (A) was first normalized for the proportion of total ESTs represented in each tissue. The resulting ratio was multiplied by 10020 (median number of ESTs across all tissues) in order to generate values easily utilized for comparative analysis.

from these tissues. Nevertheless, normalization for the number of ESTs sequenced from each library (Figure 17B) suggests these tissues indeed require relatively more complex expression regimes for their respective developmental programs to be completed.

Within floral and fruit development the richest sources of tissue specific genes are pollen and ripe fruit. Ripening specific genes were defined as those that were only expressed in breaker and red ripe fruit, while pollen specific genes were only expressed in pollen, open flower, and mixed flower (and with higher relative abundance in pollen than both open flower and mixed flower). We identified 113 ripening specific (35.6% of total fruit specific) and 169 pollen specific (55.2% of total flower specific) genes. In a recent study using Affymetrix Arabidopsis 8K GeneChip arrays, Honys and Twell (2003) also found Arabidopsis pollen had significantly more uniquely expressed genes (387) than flower buds (107) or open flower (67). Lists of all the tissue specific tomato genes and their putative functions based on sequence homology are available through the Appendix B section and at <http://ted.bti.cornell.edu/digital>.

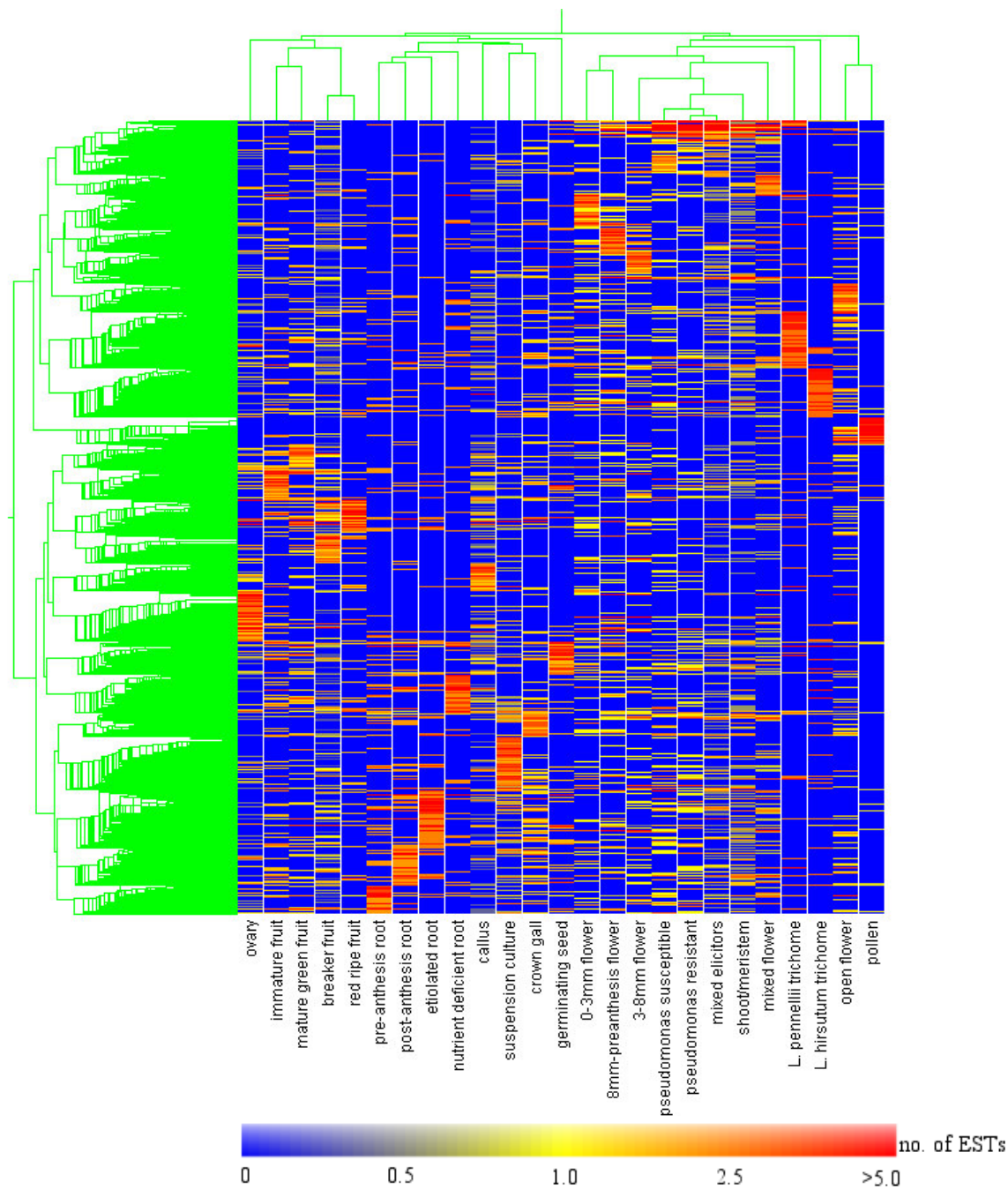
### ***Coordinated gene expression analysis of tomato***

To generate expression patterns of genes across various tissues as represented by cDNA libraries and to access the overall similarities and differences between transcriptomes of different tissues/organs, we performed coordinated gene expression analysis. Only cDNA libraries with more than 1000 independent EST sequences and TCs represented by at least five individual ESTs were selected for analysis as described previously (Ewing et al., 1999). 6,758 genes and 25 cDNA libraries met these criteria. Instead of using raw EST counts as in prior studies (Ewing et al., 1999; Ewing and

Claverie, 2000; Ogihara et al., 2003), we first normalized the data on the basis of the number of ESTs per tissue for more accurate relative expression values (see Materials and Methods). Following normalization, data was loaded into GeneSpring software for hierarchical clustering on the basis of both genes and tissues using Gene Tree and Experiment Tree programs, respectively.

Figure 18 is the complete cluster of all 6,758 genes across 25 tissues. Correlations of genes with development or biology of specific tissues was observed with two examples shown in Figure 19. Figure 19A depicts a cluster of genes highly expressed in leaf and shoot tissues, many of which have homology to genes involved in photosynthesis. Figure 19B shows genes highly expressed in ripening fruit. Most of the genes in this cluster have previously documented roles in fruit ripening {e.g. E8 (Deikman and Fischer, 1988), beta-fructosidase (Elliott et al., 1993), phytoene synthase (Bartley et al., 1992)}. Genes of unknown or minimally defined functions residing in such clusters (i.e. displaying coordinate expression) may represent additional participants in the same or associated biological processes.

Red regions of Figure 18 specific to each tissue represent tissue/treatment-specific gene expression in the context of the 25 tissues analyzed. The tissue specific region for pollen in Figure 18 represents 226 individual TCs. The list of these TCs and their putative functions are available at Appendix A Table V. Many have been previously documented to be pollen specific or pollen related genes including LePRK1 and LePRK2 (TC118503 and TC125301; Muschietti et al., 1998), LAT52 (TC115959, Tang et al., 2002), and LeFRK4 (TC116279; German et al., 2002).



**Figure 18. Correlated expression analysis of tomato EST data.** 6,758 TCs with at least five members and 25 cDNA libraries with at least 1000 ESTs were used for clustering. Both gene tree (left) and cDNA library tree (top) were created using gene tree and experiment tree programs in GeneSpring software. Differences in gene expression are shown in color as per the lower scale.



**Figure 19. Expanded region of two sub-clusters derived from Figure 18.** Color representation of expression data and the order of cDNA libraries in the cluster are same as in Figure 2. The expression profile and putative function is shown for each gene. (A) Cluster of genes that are highly expressed in leaf and shoot tissues. (B) Cluster of genes that are highly expressed in ripening fruit (breaker and red ripe fruit). Pearson correlated coefficients of all the gene expression profiles within both clusters are  $> 0.9$ .



Gene expression profiles for each tissue/treatment were also clustered relative to each other (Figure 18). As predicted, EST collections derived from the most similar tissues or representing a developmental continuum typically clustered together. All five fruit, four root, four leaf/shoot and two trichome libraries clustered most tightly together, respectively. Furthermore, most libraries from the two sets representing developmental time courses (fruit and flower development) clustered in developmental order further validating the analysis methods employed. The only exception was for the open flower library that clustered most tightly and separately with the pollen EST collection, most likely due to the large number of pollen related genes predominant in both EST collections.

#### ***Identification of fruit ripening-related genes***

Fruit ripening is a biologically and agriculturally important process unique to plants and for which tomato has been widely studied as a model system (Giovannoni, 2001). We initially focused on ripening for a more detailed analysis of differential gene expression, especially the transition to ripening as demonstrated in the "breaker" stage where the fruit display the first signs of lycopene accumulation (i.e. breaking color), ethylene evolution, and tissue softening (Seymour et al., 1993 and references therein).

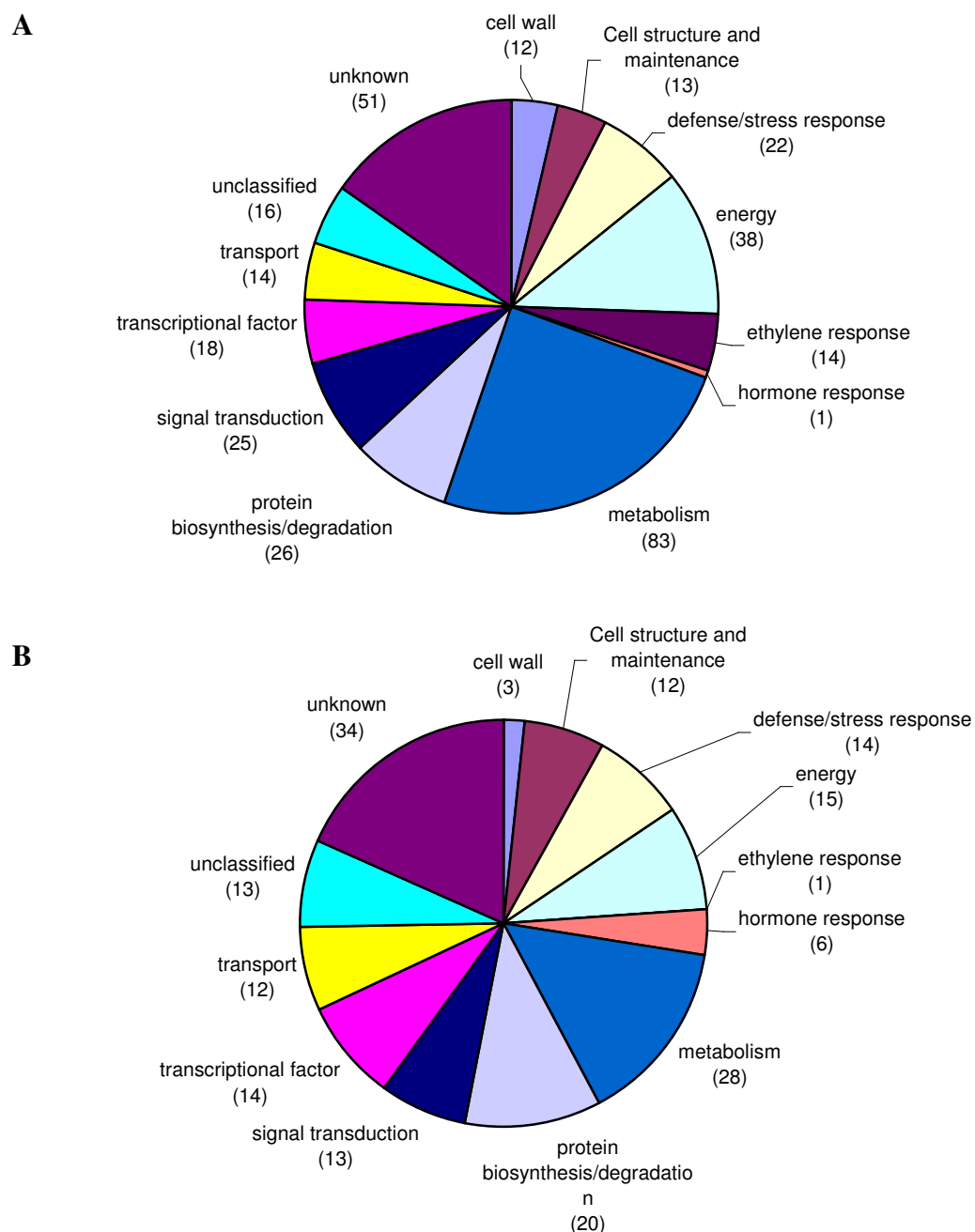
A number of robust and independent statistical approaches for digital gene expression analysis (also referred to as *in-silico* differential display or electronic Northern analysis) have been reported (Audic and Claverie, 1997; Grellier and Tobin, 1999; Stekel et al., 2000; Strausberg et al., 2001). Romualdi et al. (2001) analyzed the efficiency of these statistical methods both by a simulation approach and by analysis of

datasets of human ESTs from the UniGene database and concluded that when dealing with pair-wise comparisons, the Audic and Claverie (1997) (AC) statistical tool was the most efficient method for detecting differences in gene expression, while the general  $\chi^2$  test was the most efficient in multi-tag sampling experiments. In this study we used the AC statistical tool for pair-wise comparisons and the general  $\chi^2$  test for multiple condition comparisons.

We compared tomato expression in mature green fruit (2 to 3 days prior to the onset of ripening) to that in breaker fruit (fruit showing first visible signs of ripening). 333 ripening-induced genes (TCs showing greater expression in breaker,  $P < 0.05$ ) and 185 ripening-repressed genes (TCs showing greater expression in mature green fruit,  $P < 0.05$ ) were identified.

The combined 518 ripening-related genes were subdivided into thirteen categories according to putative function based on DNA sequence homology (Figure 20 and Appendix A Table VI and VII). As a point of clarification, ethylene response element binding proteins (EREBPs) were placed in the “ethylene response” category although they are also technically transcription factors. Such limitations in annotation should be considered in any report of large-scale genomics data.

Among the ripening-induced genes identified through this approach, a number had been previously shown to be associated with fruit ripening, including ACC oxidase 1 (TC123932; Barry et al., 1996), E8 (TC115848; Deikman and Fischer, 1988), *Never-ripe* ethylene receptor (TC124584; Wilkinson et al., 1995), lipoxygenase (TC124020; Ferrie et al., 1994), polygalacturonase 2A precursor (TC124082; Bird et al., 1988),



**Figure 20. Distribution of ripening-related genes by functional category.** 333 ripening-induced genes (A) and 185 ripening-repressed genes (B) were annotated by blast against the GenBank nr database and then classified into corresponding categories according to their putative functions. Number in parenthesis in each category represents the number of ripening-induced genes (A) and ripening repressed genes (B) classified into the corresponding category.

histidine decarboxylase (TC123915; Picton et al., 1993), and phytoene synthase (TC115970, TC116252; Bartley et al., 1992; Ray et al., 1992). Several additional genes previously associated with ripening via differential display were also identified here (TC116368, TC116059; Giovannoni et al., 1999, Cordes et al., 1989).

Among the ripening-induced genes, fourteen share homology with sequences involved in ethylene biosynthesis or signal transduction. Within this group two previously uncharacterized EREBP genes (TC116548 and TC116831) were identified in this study as ripening-induced genes.

Twelve ripening-induced genes putatively encode loci involved in cell wall metabolism, twenty-two are related to defense or stress response sequences, and eleven are homologous to genes involved in carotenoid metabolism (Appendix A Table VII).

Of the 185 ripening-repressed genes, only three are associated with cell wall metabolism, one with ethylene response and none were identified having clear relationships to carotenoid metabolism. In addition, fourteen genes are related to defense/stress response in contrast to 22 members of this classification group being induced during ripening. Particularly prevalent among the ripening-induced stress genes are members of the heat shock protein family suggesting a need for protein stabilization during ripening. Among the three ripening-repressed genes associated with cell wall metabolism, one (TC116388) corresponds to the  $\beta$  subunit of polygalacturonase isoenzyme 1 (PG1). The expression of the  $\beta$  subunit of PG1 was previously reported to be down regulated by fruit ripening (Zheng et al., 1992).

Few transcription factors other than the *RIPENING-INHIBITOR* MADS gene *LeMADS-RIN* (Vrebalov et al., 2002) have been functionally associated with the ripening process. Interestingly, eighteen putative transcription factors in addition to three EREBPs were induced during ripening and fourteen transcription factors in addition to an EREBP displayed expression negatively correlated with ripening, suggesting a large pool of putative positively and negatively acting regulatory elements for future functional analysis. Furthermore, fifty-one genes with no known or putative function are ripening-induced while thirty-four are ripening-repressed.

#### ***Comparison of tomato and grape ripening-induced genes***

Fruit can be physiologically classified as climacteric or non-climacteric depending on the presence or absence of a burst in respiration at the onset of ripening. Climacteric fruit also typically increases ethylene production at the onset of ripening and requires this hormone for completion of the ripening process (Biale and Young, 1981). Common ripening regulatory mechanisms operating in both climacteric and non-climacteric fruit remain elusive (Vrebalov et al., 2002). Tomato is a climacteric fruit that produces and requires ethylene for ripening while grape is a non-climacteric fruit for which a significant number of fruit derived EST sequences are publicly available. In order to gain insight into common regulatory mechanisms for both climacteric and non-climacteric fruit, we compared EST collections from ripening grape and tomato fruits.

The TIGR Grape Gene Index (VvGI, <http://www.tigr.org/tdb/tgi/vvgi/>) includes 83,675 sequences representing 66,501 ESTs and 17,091 unique genes (of which 9,571 are represented by contigs and 7,520 are singletons). 4,126 ESTs in VvGI are from green

fruit (stage II) and 5,167 are from veraison stage fruit. Veraison represents the onset of grape ripening and thus is approximately comparable to breaker stage of tomato.

In comparing ESTs from green versus veraison fruit and employing the same criteria as with tomato, ninety-five grape ripening-induced genes were identified at  $P < 0.05$  (Appendix A Table VIII). Nine of the ninety-five genes are homologous to genes involved in cell wall metabolism and five represent previously characterized ripening-related genes (Davies and Robinson, 2000; Nunan et al., 2001). While fourteen ethylene synthesis/signaling genes were identified via digital expression analysis of tomato, none was revealed by analysis of grape ESTs.

The cDNA sequences of all ninety-five grape and 333 tomato ripening-induced genes were compared at the translated amino acid level using the TBLASTX program (cutoff  $e$  value of  $1e-10$ ). Fifteen homologues of ripening-induced genes from both tomato and grape were identified (Table 10). Interestingly, three transcriptional factors, including members of the MADS box, zinc finger, and bZIP transcription factor families, are among those identified.

During the onset of fruit ripening, both climacteric and non-climacteric fruits typically undergo tissue softening mediated by cell wall modification, changes in the levels of flavor metabolites, and become more susceptible to pathogen attack. Table 10 was the result of independent digital expression analysis of unripe versus ripening fruit from both tomato and grape, followed by DNA sequence homology analysis between the two resulting lists to identify homologous ripening-induced genes from both tomato and grape. A number of genes associated with the ripening traits mentioned above were

**Table 10.** Genes induced both by tomato and grape ripening

<b>Tomato TC</b>	<b>Grape TC</b>	<b>Annotation</b>
TC125305/TC125359	TC4377	MADS box protein
TC124244/TC124112	TC4730	bZIP transcription factor
TC124196/TC125034	TC9044	zinc finger transcription factor
TC116030	TC4282	xyloglucan endo-1,4-beta-D-glucanase
TC123883/TC124274	TC4394	alcohol dehydrogenase
TC115998	TC4249	Pathogenesis-related protein
TC125239	TC4910	calcineurin B like protein
TC123982	TC9086	Calmodulin
TC116962/TC116318/TC116319	TC4046/TC4559/TC4193/ TC4064/TC4181/TC4034	heat shock protein
TC124903/TC126297/TC126413	TC4348	heat shock protein
TC124001	TC9134/TC4581	heat shock protein
TC115895	TC4236/TC4822	ubiquitin
TC123771	TC4209	elongation factor 1-alpha
TC124929	TC9284	heavy metal ion transport protein
TC124731	TC9962	endoplasmatic reticulum retrieval protein

identified including xyloglucan endo-1,4-beta-D-glucanase, which contributes to cell wall disassembly and is associated with fruit softening (Ishimaru and Kobayash, 2002), alcohol dehydrogenase, which could impact fruit flavor (Speirs et al., 1998) and a gene encoding a putative PR protein.

A number of additional genes not previously associated with ripening, yet with putative roles in signal transduction, stress response and transport, were also common ripening-induced genes in both tomato and grape (Table 10).

### ***Identification of pathogenesis-related genes***

Digital expression analysis was also performed on available tomato EST collections to identify pathogenesis-related genes via analysis of leaves from infected *Pseudomonas* susceptible plants, infected *Pseudomonas* resistant plants (harboring a Pto transgene), and leaves of plants treated with a number of elicitors of plant defense responses (Table 7). We identified 169 differentially expressed genes (Appendix A Table IX). Because all tissues were pathogen or elicitor challenged this was not the optimal screen for pathogenesis-related expression and demonstrates one of the limitations of digital expression analysis (i.e. availability of suitable EST collections for optimal comparative analysis). A number of pathogenesis-related proteins identified through this analysis were previously characterized in terms of pathogen response and include TC117463 (van Kan et al., 1992), TC115911 (Tornerio et al., 1997), TC124733 (Linthorst et al., 1991), and TC116258 (Gadea et al., 1996)

The 169 putative pathogenesis-related genes were subdivided into twelve functional categories based on sequence homology (Figure 21 and Appendix Table IX).



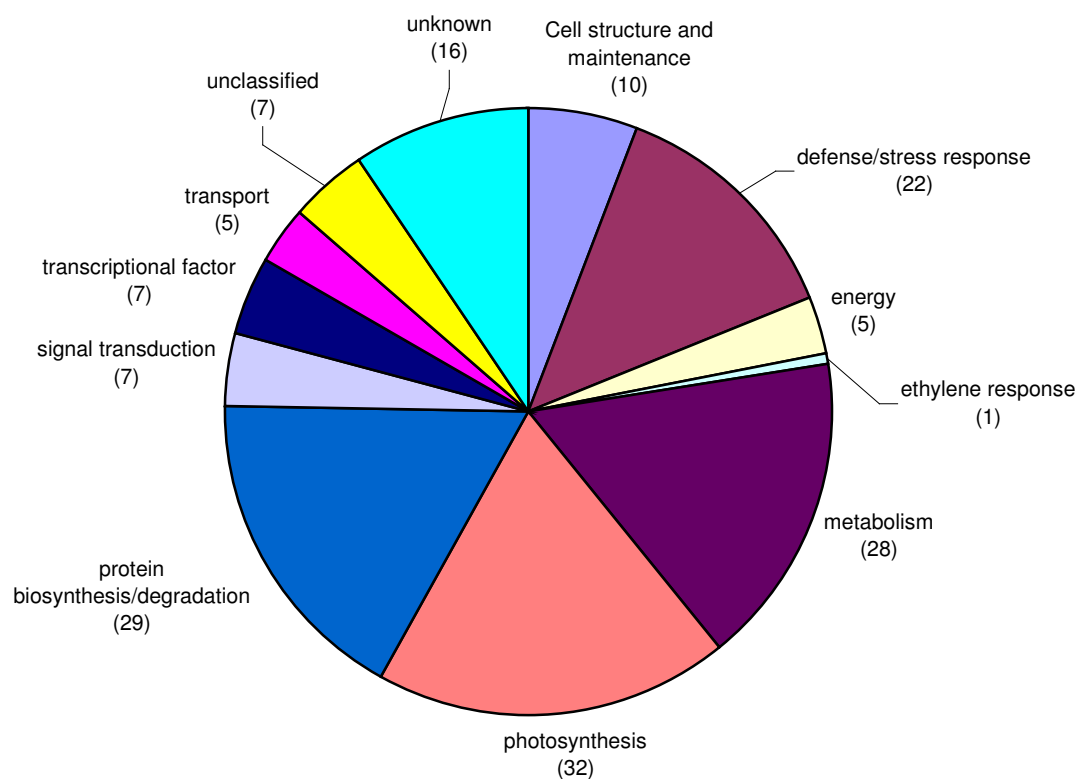
Most were classified as being involved in photosynthesis, protein biosynthesis/degradation, or defense/stress response. In addition, we identified six transcription factors not previously associated with pathogenesis, including two EREBPs (TC116750 and TC116368).

***An estimate of the number of loci that are differentially expressed during tomato flower and fruit development***

Both tomato fruit and flower development are complex processes during which the expression level of many genes changes significantly. Using the available EST information, we attempted to estimate the number of differentially expressed genes during flower and fruit development, respectively.

Five fruit (ovary, immature green, mature green, breaker, and red ripe) and four flower development stages (0-3mm, 3-8mm, 8mm-preanthesis, and open flower) were characterized by digital expression analysis. TCs differentially expressed during the five development stages and the four flower development stages were identified. 931 and 250 differentially expressed genes during fruit and flower development were identified, respectively, at a significance level of  $P < 0.01$  (Appendix Table X and XI).

The tomato unigene set represents an approximate 35% overestimation of the true number of independent ESTs and has been hypothesized to represent approximately half of the predicted 35,000 genes residing in the tomato genome (Van der Hoeven et al., 2002). Because we have sampled only four and five stages of the multi-day to multi-week flower and fruit developmental programs, respectively, any estimate of differential expression is by nature a lower limit. Nevertheless, we can set this lower limit at



**Figure 21. Distribution of pathogenesis-related genes by functional category.** 169 pathogenesis-related TCs were annotated by blast against the GenBank nr database and then classified into corresponding categories according to their putative functions. Number in parenthesis in each category represents the number of pathogenesis-related genes classified into the corresponding category.

approximately 1379 and 370 genes in tomato that are differentially expressed during fruit and flower development, respectively (number of differentially expressed genes  $\times 2 / 1.35$ ).

### ***Tomato digital expression database***

In order to mine the tomato EST digital expression data efficiently and allow public use of this resource, we constructed an on-line digital expression database (<http://ted.bti.cornell.edu/digital>). The information presented includes the annotation and sequence of each TC and singleton, raw expression data of all the TCs, and normalized expression data for the 6,758 TCs used in digital expression analysis.

The database also contains several interactive tools for expression data manipulation and analysis including tools for identifying differentially expressed genes between any two tissues and a function which identifies gene expression patterns similar to any selected gene of interest.

## **Discussion**

### ***Utility of digital expression profiling***

In this study, we performed large-scale transcriptome analysis and identified differentially expressed genes among diverse plant tissues by using available tomato EST information as a source of expression data for digital expression profiling. Digital expression analysis has several advantages over conventional microarray approaches in that the reliability of the later can be reduced by cross-hybridization of closely related sequences (Kachalo et al., 2002) and development of stable probe secondary structures

(Southern et al., 1999). The main limitation of digital expression analysis is the availability of large unbiased cDNA libraries from tissues of interest. In this regard, the resolution of digital expression profiling can be low if a given EST collection is small. The overall effort and cost of EST sequencing clearly precludes digital expression analysis as a general tool for expression analysis, nevertheless, as more EST collections are created, this approach becomes a valuable means of extracting expression information from available EST collections and for performing cross-species comparisons.

More than 150,000 tomato ESTs from different tissues and/or treatments were analyzed. These ESTs were reduced to a total of 31,012 unique genes through clustering and assembly. By analyzing EST data following normalization for library size, we were able to identify highly abundant (Tables 8 and 9) and tissue specific genes (Figure 17).

### ***Approach and analysis***

Relative expression patterns of 6,758 tomato genes across 25 different tissues/treatments were developed (Figure 18). Unlike other studies which employ raw EST counts (Ewing et al., 1999; Ewing and Claverie, 2000; Ogihara et al., 2003), we normalized the number of ESTs in two ways. First, the raw EST counts for genes represented by sufficient number of ESTs for further analysis were transformed into relative expression values (see Materials and Methods) to correct for the different sampling sizes of the various EST collections (cDNA libraries). The relative expression value of every gene was then divided by the median expression of said gene across all 25 libraries. This step of normalization puts all genes on the same relative scale for

comparison among all tissues. As validation of this approach we found that genes with similar functions or in the same pathways (e.g. photosynthesis and fruit ripening, respectively), often have correlated expression patterns (Figure 19). Genes of previously unknown function were also found in these clusters and may now be candidates for associated functions.

Gene expression patterns of 25 different tissues/treatments were clustered according to their digital profiles (Figure 18). EST collections derived from similar tissues or adjacent developmental stages typically clustered together providing further validation of the analysis pipeline.

Interestingly, open flower and pollen, which clustered together, did not cluster with other flower tissues. Open flower contains a large amount of pollen while other flower tissues (0-3mm flower buds, 3-8mm flower buds, and 8mm-preanthesis flower) do not. When re-analyzing the data excluding the subset of pollen specific genes, open flower clustered with other flower tissues, while pollen remained distant (data not shown). This result combined with the fact that pollen harbored the largest number of tissue-specific genes provides an explanation for why these two tissues represent an out-group in this analysis.

### ***Digital gene expression analysis of fruit ripening***

Fruit is a major component of the human diet contributing a large portion of the vitamins, minerals and fiber. Fruit ripening is a complex process influenced by numerous factors including light, hormones, temperature and genotype. Considerable attention has been directed toward elucidating the molecular basis of fruit ripening in

tomato (Giovannoni, 2001). In this study, we identified genes that are differentially expressed during fruit ripening by comparing mature green and breaker stage fruit EST collections. While a number of ripening-related genes identified in this study were reported previously, more than 300 ripening-induced and 180 ripening-repressed genes that had heretofore not been characterized in terms of ripening were identified in this study. One novel ripening-induced gene (TC115942) encodes a putative acyltransferase. This gene is highly expressed in ripening fruit (152 ESTs in breaker, 34 in red ripe and none in green fruit or any other tissues) and shows high similarity to acetyl-CoA:benzyl alcohol acetyl transferase from *Clarkia breweri* responsible for the production of the floral volatile benzylacetate (D'Auria et al. 2002). Alcohol acetyl transferases capable of catalyzing the formation of volatile esters which contribute to fruit flavor and aroma have been identified in strawberry (Aharoni et al., 2000) and this highly regulated gene in ripening tomato fruit is a likely candidate for a role in tomato fruit aroma.

TC116718 encodes a putative flavonol synthase, an enzyme involved in flavonoid biosynthesis. Flavonoids are a group of plant secondary metabolites associated with health-promoting properties (Muir et al., 2001). Overexpression of a petunia chalcone isomerase, another enzyme involved in flavonoid biosynthesis, in tomato significantly increased flavonoid concentration (Muir et al., 2001). TC116718 may also represent a potential candidate for engineering fruit nutrient value.

Another interesting gene in terms of ripening expression is TC115905, which encodes pectate lyase and is highly expressed in ripening fruit (131 ESTs in breaker, 105 from red ripe library and only one in green fruit). The relationship between pectate lyase

and fruit softening is well characterized in other species including banana, strawberry, and grape (Marín-Rodríguez et al., 2002). In tomato, polygalacturonase was thoroughly studied as a key mediator of fruit softening. Nevertheless, transgenic tomato plants with suppressed polygalacturonase expression showed no effects on fruit softening (Smith et al., 1989), implying that this process is more complicated than originally thought. Indeed, additional cell wall hydrolases and expansins have been more recently associated with tomato fruit softening (Rose et al., 1997; Smith and Gross, 2000). The prevalence and tight ripening regulation of pectate lyase during fruit development suggests the possibility of an additional contributor to tomato fruit softening.

Ten ripening-induced genes encoding putative heat shock proteins (HSPs) were identified through digital expression analysis. Recently it was reported that *vis1*, which is fruit ripening-induced and encodes a small heat shock protein, contributes to pectin depolymerization and juice viscosity (Ramakrishna et al. 2003). It was proposed that HSPs play a role in facilitating fruit ripening by protecting cellular machinery against heat stress during the daytime rise in field temperature (Ramakrishna et al. 2003). HSPs, which were originally thought to be protective factors induced specifically by heat stress, were also found to be developmentally regulated in the absence of stress (Vierling, 1991).

Thirty-six ripening-related genes identified by digital expression analysis encode putative transcription factors. Twenty-one (including three EREBPs) are ripening-induced while fifteen (including one EREBP) are ripening-repressed. Most of these genes were first associated with ripening in this study. These transcription factors may

play significant roles in fruit ripening. Among this group of transcription factors, four belong to the AP2/EREBP family (TC116548, TC116831, TC116368, TC116320). While EREBP/AP2 transcription factors represent a large gene family in tomato, to date their role in fruit ripening remains unknown. Three putative MADS box genes (TC125359, TC125305, and TC124719) were also among this group. The association of MADS box genes with fruit ripening was reported recently by cloning the tomato *LeMADS-RIN* gene (Vrebalov et al., 2002). MADS-box genes are known to act as multimers (Riechmann et al., 1996), thus MADS box genes in addition to *LeMADS-RIN* could be anticipated to impact ripening. Other putative ripening-related transcription factors identified in this study, such as homeodomain proteins, basic leucine zipper proteins, and zinc finger proteins (Appendix Tables VI and VII), have been found to regulate various processes of plant and animal development, while a relationship between these transcription factors and fruit ripening remains undocumented.

Included among the ripening-related genes identified in this study are many whose sequences show no similarity to any previously described gene of known function. Association of expression of said genes with fruit development and ripening provides initial insights pertaining to general areas of possible function.

#### ***Digital expression analysis facilitates cross-species comparisons***

Tomato is a climacteric fruit requiring ethylene for ripening, while grape is non-climacteric. Nevertheless, the ripening processes of both share several common features including fruit softening, changes in pigmentation, increased vulnerability to pathogen infection, and accumulation of sugars responsible for at least part of the ripening-



associated change in flavor quality. Through digital expression profiling of the public grape EST collection, we identified 95 ripening-induced genes in grape at  $P < 0.05$ . In order to gain insight into common regulatory mechanisms of ripening between climacteric and non-climacteric fruit, we compared grape and tomato ripening-induced genes at the translated amino acid level. Homologous genes induced by both grape and tomato fruit ripening are shown in Table 10. Of note are three transcription factors including a MADS box gene, a zinc finger protein, and member of the bZIP transcription factor family. The necessary role of a MADS box gene acting upstream of ethylene in climacteric fruit ripening was recently described and proposed to represent a common regulator of ripening in climacteric and non-climacteric fruit (Vrebalov et al., 2002). Grape MADS box gene TC4377 was previously shown to be induced by ripening (Boss et al., 2002). This MADS box gene is most similar (at the amino acid level) to TDR5 of tomato (e-113) which mediates organ differentiation in the three inner whorls of tomato flowers and is not associated with fruit ripening (Pnueli et al., 1994). Digital expression analysis indicates that the expression of TDR5 is not induced by tomato ripening suggesting the possibility that different MADS box genes might participate in fruit ripening in different species.

A relationship of either zinc finger proteins or bZIP transcription factors to ripening has not been demonstrated to date, though this analysis suggests their possible roles in a conserved mechanism of ripening control transcending climacteric and non-climacteric distinctions.

### ***Pathogenesis-related gene identification***

Plant diseases caused by bacteria, fungi, viruses and nematodes cause substantial crop losses and require expensive and potentially harmful applications of pesticides. A deeper understanding of plant defense mechanisms will facilitate development of disease resistant crops.

We have documented 169 pathogenesis-related genes through digital expression analysis. Experimentally identified differentially expressed genes in transgenic tomatoes over-expressing the Pto kinase were reported in two previous studies (Xiao et al., 2001; Mysore et al., 2003) and include approximately one-third of those observed in this analysis. However, more than one hundred additional putative pathogenesis-related genes were identified here. Most of these additional genes can be categorized as defense/stress response, photosynthesis, and protein biosynthesis or degradation. The roles of these gene categories in plant defense are discussed in detail by Mysore et al. (2003).

### ***Limitations of digital expression analysis***

Employing digital expression analysis of available EST collections from tomato, we were able to develop crude estimates of the number of different loci that are differentially expressed during fruit and flower development. Employing a significance level of  $P < 0.01$ , we estimate that approximately 1379 and 370 loci are differentially expressed during tomato fruit and flower development, respectively, corresponding to 3.94% and 1.06%, respectively, of all tomato genes. However, caution should be taken in interpreting such estimates. Specifically, differentially expressed genes identified via

digital expression analysis are biased toward moderate or highly expressed genes to a degree dependent upon the total number of ESTs generated per tissue/treatment. For example, some well-documented differentially expressed genes during fruit ripening (e.g. lycopene  $\beta$ -cyclase and lycopene epsilon-cyclase, Ronen et al., 1999) may not be revealed by this approach due to the limited availability of corresponding ESTs from tissues analyzed. In addition, genes that are differentially expressed in tissues or stages for which no ESTs are available obviously cannot be identified. Nevertheless, by setting appropriate statistical criteria, a substantive amount of expression information can be derived from even modest sized EST collections. Here we have clearly identified numerous genes that had not been previously associated with ripening, floral development, or pathogen responses and which may prove valuable in future efforts toward elucidating the genetic mechanisms underlying these processes. The online database developed in conjunction with this effort will allow researchers to take advantage of the expression data recovered through digital expression analysis of the tomato EST collection.

## CHAPTER V

### IMPLEMENTATION OF A TOMATO EXPRESSION DATABASE

#### Summary

Members of the lab are embarking on large-scale functional genomics using cDNA microarrays to obtain expression profiles of tomato genes during fruit development and ripening. Expression data for approximately 11,049 ESTs over a time course of wild type tomato fruit development were generated (R. Alba, P. Payton, S. Moore and J. Giovannoni, unpublished). The ESTs employed represent sequences unique to specific tissues (e.g. fruit, flower), as well as defined genetic, biochemical and physiological functions (e.g. protein kinases, transcriptional factors). In order to provide the research community access to our normalized and replicated microarray data as a tool to assess relative expression of genes of interest, I developed a publicly accessible online database - Tomato Expression Database (TED: <http://ted.bti.cornell.edu>). Through this database, I provide multiple approaches to pursue analysis of specific genes of interest and/or access the larger microarray dataset to identify sets of genes that may behave in a pattern of interest to the user. To achieve this goal, a set of useful data mining and data visualization tools were developed with the anticipation of continuing expansion according to users' requirements. Developed initially as a data mining and analysis resource, TED also contains comprehensive annotation of each EST including homology derived from sequence similarity searches of GenBank and GO terms assigned manually according to putative functions. Future expansions will include

integration of expression data from tomato fruit ripening mutants and from normal tissues in response to treatments that influence ripening, integration of tomato metabolite data, and addition of interfaces for expanded data analyses including clustering.

## **Introduction**

In the past few years the rapid progress of large-scale EST sequencing projects, cDNA microarrays, and computational biology has provided us with enormous potential for novel gene discovery and gene expression analysis on a large-scale. In tomato, a publicly available EST database was created which includes more than 150,000 ESTs from more than 27 different tissues (<http://www.tigr.org/tdb/tgi/lgi>; <http://www.sgn.cornell.edu/>; Van der Hoeven et al., 2002). To take full advantage of this valuable EST data, the Giovannoni and Martin labs at BTI developed a publicly available tomato cDNA microarray – TOM1 (<http://bti.cornell.edu/CGEP/CGEP.html>). The array contains ~11,049 ESTs representing ~8,800 distinct genes, an estimated one fourth of the entire tomato gene complement (Van der Hoeven et al., 2002).

In order to gain insights into the molecular basis of fruit development and dissect the regulatory network governing this process, others in the lab performed large-scale transcriptome analysis using the TOM1 cDNA microarray. Gene expression data over ten time points spanning tomato fruit development from one week post anthesis through late ripening was recovered.

The use of cDNA microarrays, especially of the complexity of the TOM1 array, and when capturing replicated data for 10 time points, generates enormous amounts of

data. Effective data management and storage is critical to successful cDNA microarray experimentation and interpretation. To address these needs, the Microarray Gene Expression Data group (MGED; <http://www.mged.org>) developed a standard definition of minimum information required to unambiguously interpret microarray data and to subsequently allow independent verification of this data at a later stage. The resulting protocol developed by this group has been given the acronym MIAME (Minimum Information About a Microarray Experiment; Brazma et al., 2001).

In order to disseminate our data to the research community and provide access to expression data for genes of interest, I built two online databases – Tomato Expression Database (TED; <http://ted.bti.cornell.edu>) and Tomato Microarray Database (TMD; <http://tmd.bti.cornell.edu>). TMD is MIAME compliant and only serves as a data warehouse to store raw microarray data generated by anyone using the TOM1 array. It is not implemented with any data analysis, or visualization tools and will not be described further in this chapter. The database is completed but not yet public. Passwords will be provided to committee members who may wish to view TMD. The remainder of the chapter will focus on the development of the TED database.

## **Methods**

### ***Microarray data acquisition***

The ten time points analyzed during fruit development are 7 DAP (days after pollination/anthesis), 17DAP, 27 DAP, 39 DAP, 41 DAP, 42 DAP, 43 DAP, 47 DAP, 52 DAP, and 57 DAP, with 39 DAP corresponding to the mature green stage and 42 DAP

corresponding to the breaker stage. Two-color cDNA microarrays were employed to compare consecutive time points, which is termed a direct sequential design (Yang and Speed, 2002). Each comparison includes six replicates - three biological replicates each with a dye-swap control (Kerr et al., 2002). Raw data [image files and data files generated by image process program - ImaGene (BioDiscovery, CA)] were generated using a Scan Array 5000 laser scanner. Raw data was processed and normalized using GeneSpring software (SiliconGenetics, CA) as outlined in the following text. ESTs with median signal values less than the local background median plus three standard deviations were omitted. As were microarray spots with less than three replicates. Data was then normalized by using the print-tip LOWESS (intensity dependent) normalization strategy (Yang et al., 2002). Omitted ESTs, negative/positive control ESTs, and “blanks” were not used during normalization. For each EST comparison, the mean of ratio, standard deviation, and P value (calculated using the standard student t test) were generated. The relative expression levels of all the ESTs in ten different developmental stages were obtained by arbitrarily setting the expression at 7 DAP as 1.0 for every gene. ESTs lacking data for more than two of the nine pair-wise comparisons in the time course were discarded for the relative expression calculation. Expression profiles for ESTs lacking data for less than two pair-wise comparisons in the time course were subjected to the KNNimpute algorithm (Troyanskaya et al., 2001) to estimate the missing value(s).

### ***Functional annotation of ESTs on the TOM1 array***

The TOM1 cDNA array contains approximately 11,049 distinct tomato EST elements (and approximately 700 control spots). All the tomato ESTs were initially sequenced by TIGR (<http://www.tigr.org>). All of the tomato ESTs on the TOM1 array are in the process of being re-sequenced to confirm their identities in a collaboration with INRA, France as EST projects by the nature of their size and complexity of data tracking tend to introduce large numbers of tracking errors. At the writing of this thesis, approximately half of the cDNA clones have been re-sequenced from both ends while the remainder should be completed by the Fall of 2003.

All of the TOM1 ESTs were functionally annotated. Putative sequence homologies were obtained by performing sequence similarity searches against the GenBank nr database using the BLASTX program (Altschul et al., 1997) with a cutoff  $e$  value of  $1e-10$ . Clones showing homology to known proteins were further annotated manually by assigning Gene Ontology (GO; Ashburner et al., 2000; Ashburner et al., 2001) terms by other members of the lab.

### ***Implementation of TED***

TED is implemented on the World Wide Web and can be easily accessed over the Internet using standard web browsers. TED runs on a Windows 2000 advanced server and the data is stored in a relational database (MySQL database management system). Perl/CGI is used to implement client-server communication in conjunction with the Perl/DBI module that allows scripts to connect to the back-end database. When Perl/CGI scripts receive a client request, they create and send SQL-commands to the



database server, and then use the resulting data to create a HTML response which is returned to the client.

## **Results**

### ***Functional annotation of ESTs on the TOM1 array***

A key part of mining microarray data is the ability to efficiently link the data to available information/annotation about specific genes of interest. In the analysis of microarray expression data, especially time course data, clustering of genes according to their expression similarities is a typical informative output. It is hypothesized that genes in the same clusters have a higher probability of being related functionally (Quackenbush, 2001). In this way, genes with unknown functions can be given tentative assignments to involvement in particular processes or responses based on the known function of other genes in the same cluster. Consequently, accurate functional annotation information for a subset of known genes is indispensable in attempting to develop new knowledge based on gene expression data.

The TOM1 tomato cDNA array contains approximately 11,049 tomato ESTs representing approximately 8,800 independent genes. The tomato ESTs are derived from divergent tomato tissues (Table 11). While more than 90% come from fruits, leaves and callus (with almost 50% from fruit, Table 11). This should not be interpreted to mean that these genes are specific to these tissues. While the TOM1 array was intentionally skewed to contain a disproportionate number of genes from fruit and related to fruit development, many of the ESTs recovered from a given library had contig members in

libraries derived from other tissues as well. In short, the specific list of clones to be picked was developed to minimize the number of libraries handled during the re-array process used to develop the TOM1 collection.

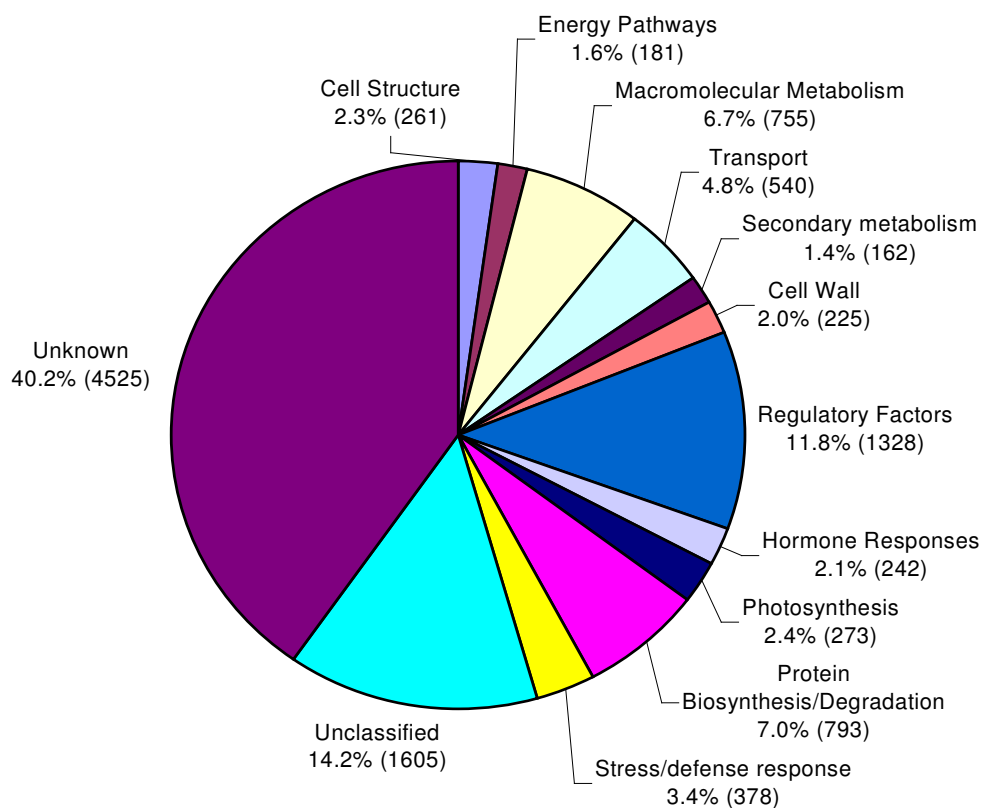
All the ESTs were initially sequenced by TIGR (Van der Hoeven et al., 2002) and approximately half of the TOM1 ESTs have been re-sequenced by Genoscope (<http://www.genoscope.cns.fr>) (see methods). Putative sequence homologies of all the ESTs were derived by performing sequence similarity searches using the BLASTX algorithm (Altschul et al., 1997) at a cutoff  $e$  value of  $1e-10$ . Nearly 7,000 TOM1 ESTs show homology to known proteins. These ESTs were further annotated manually by others in the lab using terminology and a hierarchy developed by the Gene Ontology Consortium (Ashburner et al., 2000; Ashburner et al., 2001). ESTs on the TOM1 cDNA array were classified into different functional categories (Figure 22).

### ***TED content***

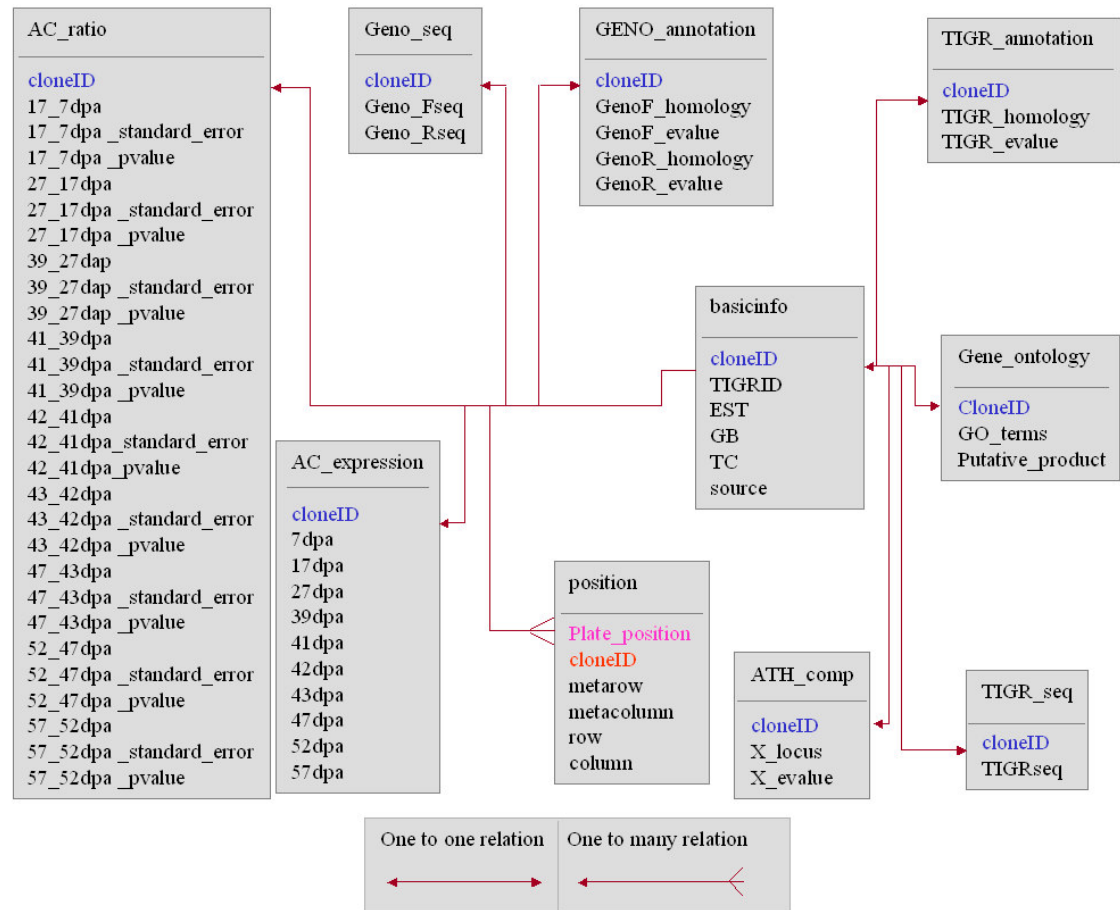
Currently TED contains two major data components: TOM1 EST clone information and normalized microarray data from fruit development experiments (Figure 23). Each clone has basic clone identifier information including the Cornell clone ID, a dbEST EST number, and the GenBank accession number. In addition, each clone was assigned a TIGR tentative consensus (TC) number, if available, by comparing its sequence to the TIGR tomato unigene set. TED also contains information on clone physical location on the TOM1 array and in the original stock collection. All sequence information (including TIGR original sequence and Genoscope re-sequences in addition

**Table 11.** Tissue source of ESTs on the TOM1 cDNA array

Tissue	Number of ESTs	Percentage (%)
Fruit	5427	49.12
Leaf	2531	22.91
Callus	2207	19.97
Flower	404	3.66
Seed	156	1.41
Shoot	135	1.22
Root	129	1.17
Trichome	30	0.27
Suspension culture	18	0.16
Crown gall	12	0.11
Total	11049	100



**Figure 22. Functional classification of ESTs on the TOM1 array.** ESTs showing homology to known genes were sorted into 11 different categories of biological function according to their putative functions. ESTs showing no homology to previously described genes or with homology to predicted proteins of unknown function were classified into the “Unknown” category. The remaining ESTs were placed in the “Unclassified” category. Number in parenthesis in each category represents the number of ESTs from the TOM1 collection classified into the corresponding category.



**Figure 23. Entity relation (ER) diagram of TED.** Column name in blue serves as both primary key and foreign key; Column name in pink serves only as primary key; Column name in red serves only as foreign key.

to smaller contributions of re-sequencing performed by other collaborators) for each clone and the EST annotation information described above were also included in TED.

Currently TED contains normalized and processed microarray data for ten time points with nine pair-wise comparisons during wild type fruit development, including the ratio data for each comparison and the relative expression data for each time point.

### ***Search functions and analysis tools in TED***

The maximal value of stored microarray data and EST annotation can only be extracted when appropriate database query and datamining tools are implemented. TED was implemented with several useful query interfaces for data retrieval and several tools for data visualization and analysis.

#### ***Clone identifier search***

Single clone information can be retrieved by searching TED using several distinct clone identifiers, including the Cornell clone ID (e.g. cLEF43N7), an EST number (e.g. EST356658), and the GenBank accession number (e.g. AW930278). Clone information including normalized expression ratio, standard deviation, and P value from each of the nine pair-wise comparisons during fruit development, GenBank best hits, and GO annotation can be retrieved and displayed in the "result" page returned from a clone identifier search (Figure 24). The queried clone is further linked to external databases such as GenBank, TIGR, and SGN (<http://www.sgn.cornell.edu>) where more information about the clone may be available. TIGR and re-sequence information, if available, will be provided through linked pages. Sequences can be translated and blasted through the link to the SGN web site.

**Information about clone:cLET21F22**

General Information

EST#	Clone ID	GenBank#	TC	Link to SGN	Link to TIGR
EST285870	cLET21F22	AW092690	TC120273	<a href="#">Click here</a>	<a href="#">Click here</a>

Relative Expression Data for TOM1 Array

17dap/7dap	27dap/17dap	39dap/27dap
0.93±0.097 p=0.5183	0.98±0.134 p=0.8653	1.33±0.127 p=0.0148

41dap/39dap	42dap/41dap	43dap/42dap
0.93±0.12 p=0.619	2.14±0.266 p=0.0001	1.06±0.068 p=0.4068

47dap/43dap	52dap/47dap	57dap/52dap
0.88±0.161 p=0.5015	0.89±0.078 p=0.2546	1.37±0.158 p=0.034

[View Expression Pattern](#)

GenBank BLAST information

Sequence Info

Genoscope Forward Sequence GenBank Best Hit	E value
<a href="#">sp P37273 PSY2_LYCES</a> Phytoene synthase 2, chloroplast precursor (Fragment)	1e-083

Genoscope Reverse Sequence GenBank Best Hit	E value
<a href="#">sp P37273 PSY2_LYCES</a> Phytoene synthase 2, chloroplast precursor (Fragment)	1e-083

TIGR Sequence GenBank Best Hit	E value
<a href="#">sp P37273 PSY2_LYCES</a> Phytoene synthase 2, chloroplast precursor	2e-083

[View Sequence Information](#)

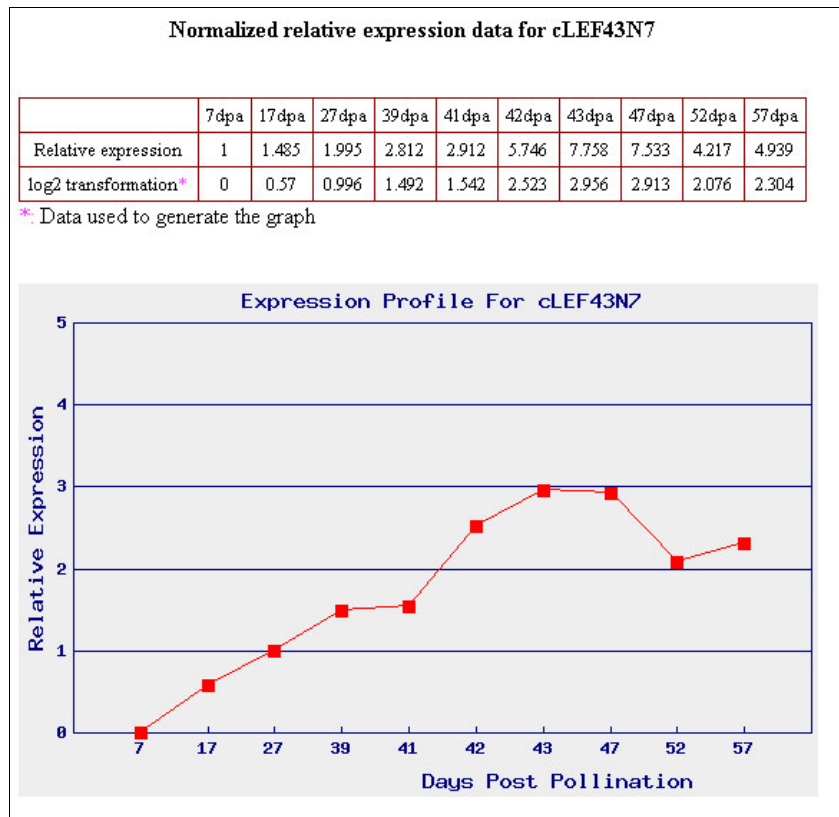
**Figure 24.** The TED result page obtained through a clone identifier search. Clone information including expression data and annotation is displayed. While other information including sequence is linked.

The result page from a clone identifier search also provides a link for the user to view the expression pattern of the queried EST during fruit development. The linked page displays the normalized relative expression data for each of the ten time points during fruit development in log<sub>2</sub> transformed form (Figure 25). The graph of the expression profile is generated by an incorporated computer script dynamically using the log<sub>2</sub> transformed data (Figure 25).

*Finding clones with similar or inverse expression pattern*

In the laboratory's large-scale transcriptome analysis during fruit development, one major goal is to find novel genes related to fruit development through identification of genes with similar expression profiles to those documented to be involved in ripening. In the TOM1 cDNA array collection, more than 40% of the included ESTs have no homology to previously described genes (Figure 22). These genes can now be given expression annotation based on their profiles and known genes which have similar expression patterns. In addition, genes with inverse expression patterns may also have target process related functions. For example, in human, tumor-suppressor gene p53 acts as a transrepressor of several other genes (Rocha et al., 2003) and as the level of p53 mRNA is elevated, the expression of its targets are repressed in an inverse relationship. I implemented a data analysis tool that can identify ESTs with similar or inverse expression patterns to a specified clone (Figure 26). The Pearson correlation coefficient is used to measure gene expression pattern similarity. A Pearson correlation coefficient greater than 0.9 indicates strong similarity while a Pearson correlation coefficient less





**Figure 25. TED normalized relative expression data.** Normalized relative expression data was obtained by setting the expression for all genes at 7 DPA as 1.0. The data was log<sub>2</sub> transformed and the transformed data used to generate the graph dynamically in TED.

Clones have similar or inverse expression pattern with clone cLEN14C20*			
* Correlation coefficient is calculated based on Pearson Correlation. The cutoff value is 0.9 or -0.9 (strong similarity or strong inverse pattern)			
Clone ID	Correlation Coefficient	Annotation	Graph Pattern
cLEN13P22	0.999	PABA synthase, para-aminobenzoate synthase/glutamine amidotransferase	<a href="#">View the graph</a>
cLEG10I11	0.999	hypersensitive response, hsr201 protein	<a href="#">View the graph</a>
cLEC33C10	0.996	glucanase	<a href="#">View the graph</a>
cLEG42I1	0.994	hypersensitive response protein, hsr20	<a href="#">View the graph</a>
cLEC10G10	0.994		<a href="#">View the graph</a>
cLEG36J15	0.993		<a href="#">View the graph</a>
cLED6O21	0.993		<a href="#">View the graph</a>
cLEN14I14	0.992	eRF2	<a href="#">View the graph</a>
cLEN10E18	0.992	phytoene synthase	<a href="#">View the graph</a>
cLEG42M9	0.99	E8 homolog	<a href="#">View the graph</a>
cLEG41K22	0.99		<a href="#">View the graph</a>
cLEG30K11	0.99	E8 homolog	<a href="#">View the graph</a>
cLER20K22	0.989		<a href="#">View the graph</a>
cLEG41P17	0.989		<a href="#">View the graph</a>
cLEG30N5	0.989	phytoene synthase	<a href="#">View the graph</a>
cLEG30J18	0.989	histidine decarboxylase	<a href="#">View the graph</a>
cLED4G24	0.989		<a href="#">View the graph</a>
cLEC14P2	0.989	NADH dehydrogenase	<a href="#">View the graph</a>
cLES6L11	0.988		<a href="#">View the graph</a>
cLEG30G13	0.988		<a href="#">View the graph</a>
cLEC18A5	0.988	putative protein	<a href="#">View the graph</a>
cLHT11N5	0.987		<a href="#">View the graph</a>
cLEM5L7	0.987	CDK/cyclin inhibitor	<a href="#">View the graph</a>
cLEG42G5	0.987	histidine decarboxylase	<a href="#">View the graph</a>
cLEG18E1	0.987	phytoene synthase	<a href="#">View the graph</a>
cLEC21C10	0.987	tocopherol methyltransferase, gamma-tocopherol methyltransferase	<a href="#">View the graph</a>
cLEC14L7	0.987	phosphatase, protein phosphatase 2C, serine/threonine specific	<a href="#">View the graph</a>
cLEG37L17	0.986	ethylene receptor	<a href="#">View the graph</a>
cLEI16B17	0.985	nuclease, exonuclease RRP41	<a href="#">View the graph</a>
cLEF16H17	0.985	glutamate-ammonia ligase	<a href="#">View the graph</a>

**Figure 26. Result page returned by TED for finding similar or inverse expression patterns with a specific query EST. The similarity in expression profile is based on the Pearson's correlation coefficient and the process is run dynamically by TED.**

than -0.9 indicates strong inverse expression. As for graph representations of expression profiles, log<sub>2</sub> transformed data was used for measuring gene expression similarity. The advantage of using the logarithm of relative expression levels for graph representation and finding clones with similar expression profiles is straightforward. Genes that are upregulated by a factor of 2 have an expression of 2, whereas those downregulated by the same factor have an expression of 0.5. As a result, downregulated genes are compressed between the values of 1 and 0. After log transformation, a gene upregulated by a factor of 2 has a value of 1, while a gene downregulated by a factor of 2 has a value of -1 so that both down regulated and up regulated genes can be described in the same relative terms of scale.

#### *Keyword search*

Many researchers will be interested in identifying genes based on keywords related to biochemical function, pathway or process. To meet this requirement, we developed a query interface that allows users to retrieve clones using keywords as query terms. Keywords will retrieve EST information if clone annotation information described above including Genbank best hits for TIGR sequences or Genoscope re-sequences, GO terms or manual annotation, contains the keywords employed. As with all such search systems, keywords will only elicit returns if the corresponding terminology resides somewhere in the annotation for said gene. It is thus important for the person performing the search to be exhaustive in performing keyword searches if desired hits are not returned. Several keywords combined together by using Boolean language terms “and” or “or” can be used to query the database simultaneously. Figure

27A shows the resulting page using the keyword “phytoene synthase” as an example. The expression profiles of all ESTs retrieved by a given keyword search can be displayed in a single graph (Figure 27B). The expression profile of each EST (each line in the graph) in Figure 27B is linked to the corresponding EST information page (Figure 24).

#### *Search and sort by expression data*

TED can also be searched by expression ratios of interest for the various pair-wise comparisons to retrieve ESTs displaying expression patterns determined by the searcher to be of interest. Up to five comparisons can be queried at the same time (Figure 28).

In addition, the expression ratios of all the ESTs in any single pair-wise comparison can be sorted and displayed in descending order of relative expression. Genes with ratios greater than 2.0 (upregulated) are displayed in red, genes with ratios less than 0.5 (downregulated) are displayed in green, while genes with ratios between 0.5 and 2.0 (no to minor change in expression) are displayed in yellow in the resulting data file (not shown).

#### ***Data mining tools in TED***

In addition to the search functions and analysis tools described above, several data mining tools that allow users to retrieve information pertaining to clone annotation and relative expression data for a subset or all of the ESTs on TOM1 were developed and implemented in TED (Figure 29). These tools are referred to as batch search tools

A

**TOMATO EXPRESSION DATABASE**

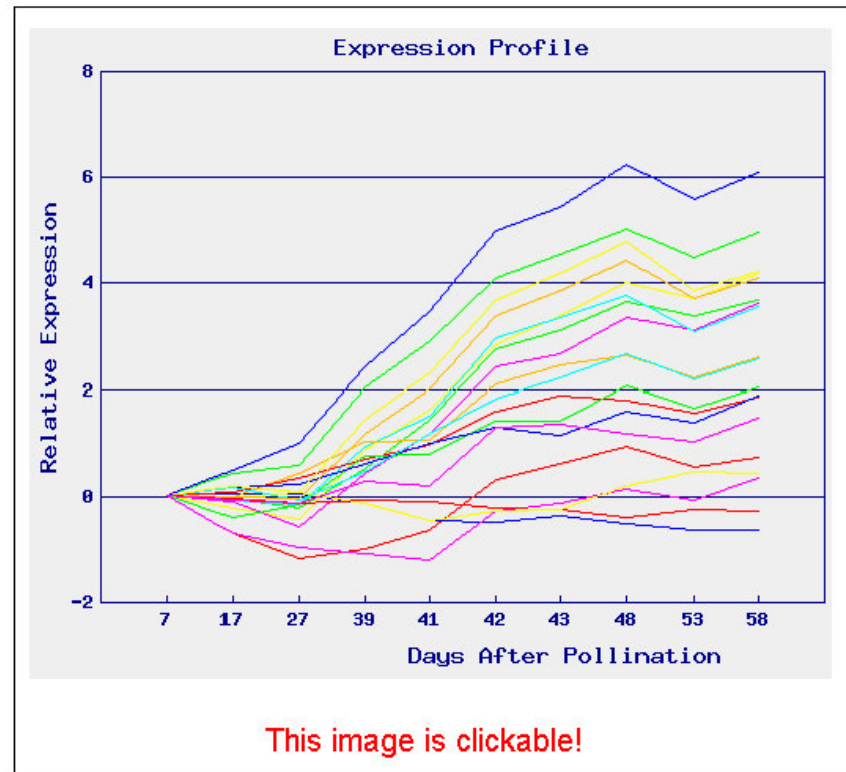
[HOME](#) | [SEARCH](#) | [DATAMINING](#) | [CONTACT US](#) | [LINKS](#)

21 records match your search criteria!

EST#	Clone ID	GenBank#	TC	Link to SGN	Link to TIGR
EST440339	cLEG42E8	BF112749	TC115970	<a href="#">Click here</a>	<a href="#">Click here</a>
EST440538	cLEG42L18	BF113028	TC115968	<a href="#">Click here</a>	<a href="#">Click here</a>
EST398872	cLEG7H3	BE432343	TC115970	<a href="#">Click here</a>	<a href="#">Click here</a>
EST400088	cLEG18E2	BE433559	TC120339	<a href="#">Click here</a>	<a href="#">Click here</a>
EST400132	cLEG18M14	BE433603	TC115970	<a href="#">Click here</a>	<a href="#">Click here</a>
EST405656	cLEG18E1	BE434578	TC115970	<a href="#">Click here</a>	<a href="#">Click here</a>
EST407219	cLEG30J18	BE436141	TC123915	<a href="#">Click here</a>	<a href="#">Click here</a>
EST407161	cLEG30N5	BE436083	TC115971	<a href="#">Click here</a>	<a href="#">Click here</a>
EST407273	cLEG31C21	BE436195	TC124239	<a href="#">Click here</a>	<a href="#">Click here</a>
EST407274	cLEG31C23	BE436196	TC115972	<a href="#">Click here</a>	<a href="#">Click here</a>
EST412247	cLEG36J15	BE460828	TC128108	<a href="#">Click here</a>	<a href="#">Click here</a>
EST412593	cLEG37B16	BE461174	TC124680	<a href="#">Click here</a>	<a href="#">Click here</a>
EST412510	cLEG37B5	BE461091	TC115968	<a href="#">Click here</a>	<a href="#">Click here</a>
EST413626	cLEM1D16	BE458334	TC122589	<a href="#">Click here</a>	<a href="#">Click here</a>
EST413605	cLEM1P1	BE458313	TC115968	<a href="#">Click here</a>	<a href="#">Click here</a>
EST299858	cLEN10E18	AW223047	TC115968	<a href="#">Click here</a>	<a href="#">Click here</a>
EST300339	cLEN12A10	AW223528	TC115970	<a href="#">Click here</a>	<a href="#">Click here</a>
EST300692	cLEN13P22	AW223881	TC128944	<a href="#">Click here</a>	<a href="#">Click here</a>
EST300773	cLEN14C20	AW223962	TC128754	<a href="#">Click here</a>	<a href="#">Click here</a>
EST310960	cLEN17L5	AW441564	TC115968	<a href="#">Click here</a>	<a href="#">Click here</a>
EST285870	cLET21F22	AW092690	TC120273	<a href="#">Click here</a>	<a href="#">Click here</a>

[View Expression Pattern](#)

B



**Figure 27. TED result page from keyword search.** (A) 21 clones with annotation including “phytoene synthase” were retrieved using this term as a keyword to query TED. (B) The expression profiles of all 21 clones from (A) during fruit development. The expression profile (line in the graph) of every clone is clickable to display the corresponding clone information (example shown in Figure 24).

### Search by Relative Expression Ratio

Relative expression level > 2 In experiment 41dap/39dap

AND

Relative expression level > 2 In experiment 42dap/41dap

AND

Relative expression level > In experiment 17dap/7dap

AND

Relative expression level > In experiment 17dap/7dap

AND

Relative expression level > In experiment 17dap/7dap

**23 records match your search criteria!**

EST#	Clone ID	GenBank#	TC	Link to SGN	Link to TIGR	41dap/39dap	42dap/41dap
<a href="#">EST398810</a>	cLEG7E12	<a href="#">BE432281</a>	TC124903	<a href="#">Click here</a>	<a href="#">Click here</a>	3.26	7.4
<a href="#">EST406129</a>	cLEG24J2	<a href="#">BE435051</a>	TC128182	<a href="#">Click here</a>	<a href="#">Click here</a>	2.59	2.5
<a href="#">EST252409</a>	cLED29K9	<a href="#">AI771213</a>	TC117542	<a href="#">Click here</a>	<a href="#">Click here</a>	2.5	2.31
<a href="#">EST440379</a>	cLEG42K20	<a href="#">BF112789</a>	TC115849	<a href="#">Click here</a>	<a href="#">Click here</a>	2.5	3.05
<a href="#">EST329046</a>	cLEI13B13	<a href="#">AW650592</a>	TC124043	<a href="#">Click here</a>	<a href="#">Click here</a>	2.46	2.44
<a href="#">EST412608</a>	cLEG37F10	<a href="#">BE461189</a>	TC126287	<a href="#">Click here</a>	<a href="#">Click here</a>	2.36	3.24
<a href="#">EST440301</a>	cLEG42M9	<a href="#">BF112711</a>	TC115849	<a href="#">Click here</a>	<a href="#">Click here</a>	2.36	2.42
<a href="#">EST300455</a>	cLEN12J21	<a href="#">AW223644</a>	TC118251	<a href="#">Click here</a>	<a href="#">Click here</a>	2.29	5.73
<a href="#">EST336481</a>	cLEG30I15	<a href="#">BE431665</a>	TC115848	<a href="#">Click here</a>	<a href="#">Click here</a>	2.27	2.74
<a href="#">EST399917</a>	cLEG13M4	<a href="#">BE433388</a>	TC116650	<a href="#">Click here</a>	<a href="#">Click here</a>	2.25	3.26
<a href="#">EST298731</a>	cLEN6I7	<a href="#">AW221920</a>	TC130652	<a href="#">Click here</a>	<a href="#">Click here</a>	2.25	6.75
<a href="#">EST406865</a>	cLEG29C17	<a href="#">BE435787</a>	TC119160	<a href="#">Click here</a>	<a href="#">Click here</a>	2.24	2.87
<a href="#">EST311524</a>	cLEN21O6	<a href="#">AW442128</a>	TC129207	<a href="#">Click here</a>	<a href="#">Click here</a>	2.18	3.04
<a href="#">EST407181</a>	cLEG30B12	<a href="#">BE436103</a>		<a href="#">Click here</a>	<a href="#">Click here</a>	2.16	3.16
<a href="#">EST310708</a>	cLEN15I15	<a href="#">AW441312</a>	TC127753	<a href="#">Click here</a>	<a href="#">Click here</a>	2.07	3.35
<a href="#">EST336459</a>	cLEG30E11	<a href="#">BE431643</a>	TC117720	<a href="#">Click here</a>	<a href="#">Click here</a>	2.06	2.02
<a href="#">EST300773</a>	cLEN14C20	<a href="#">AW223962</a>	TC128754	<a href="#">Click here</a>	<a href="#">Click here</a>	2.06	2.84
<a href="#">EST407006</a>	cLEG29J18	<a href="#">BE436012</a>	TC128098	<a href="#">Click here</a>	<a href="#">Click here</a>	2.05	2.68
<a href="#">EST326686</a>	cLEI3L6	<a href="#">AW648232</a>	TC125440	<a href="#">Click here</a>	<a href="#">Click here</a>	2.05	2.23
<a href="#">EST244017</a>	cLED4G2	<a href="#">AI485696</a>	TC115853	<a href="#">Click here</a>	<a href="#">Click here</a>	2.04	3.16
<a href="#">EST406722</a>	cLEG28M21	<a href="#">BE435644</a>		<a href="#">Click here</a>	<a href="#">Click here</a>	2.04	7.59
<a href="#">EST299765</a>	cLEN9P14	<a href="#">AW222954</a>	TC123921	<a href="#">Click here</a>	<a href="#">Click here</a>	2.04	3.5
<a href="#">EST414483</a>	cLEM5D12	<a href="#">BE459191</a>	TC116841	<a href="#">Click here</a>	<a href="#">Click here</a>	2.01	2.65

**Figure 28. Sample TED expression ratio search form and result.** The expression ratio from up to five comparisons can be queried in one search form.

**TOMATO EXPRESSION DATABASE**

[HOME](#)   [SEARCH](#)   [DATAMINING](#)   [CONTACT US](#)   [LINKS](#)

[TOM1 clone GenBank annotation](#)  
[TOM1 clone GO annotation](#)  
[TOM1 expression data retrieval](#)  
[TOM1 blast search](#)  
[TOM1 TC search](#)  
[TOM1 contiguous clone search](#)  
[TOM1 chip position search](#)  
[TOM1 clone position search](#)



---

[Home](#) / [Search](#) / [Datamining](#) / [Contact](#) / [Links](#)

© 2001 Jim Giovannoni's lab. All rights reserved.

Any comments or questions to : [Webmaster](#)

**Figure 29. Data mining tools in TED.**

because information about more than one EST can be retrieved in a single query. Figure 30 shows a batch query interface that accepts a set of TCs and recovers all the clones on the TOM1 array belonging to the corresponding TCs. These data mining tools are extremely useful as users can recover defined subsets or entire collections of deposited expression data and analyze them locally.

It is worth noting that one can do sequence homology searches through the “TOM1 blast search” tool by submitting a set of query sequences (nucleotide or protein) in FASTA format which can be blasted against sequences of all the TOM1 array clones via BLASTN, TBLASTN or TBLASTX programs of the stand alone BLAST package (Altschul et al., 1997). The result page provides link to corresponding clone information as shown in Figure 24.

For most tools, in addition to the results displayed in web browser in HTML format, a tab delimited text file is also provided for users to download. The text file can be easily imported into other software applications such as MS Excel spreadsheets for easy manipulation and further analysis.

### ***Future directions***

Once TOM1 re-sequencing is completed, effort will be directed toward re-annotating the ESTs computationally followed by further manual curation.

Others in the lab are expected to obtain microarray data for the same ten time points during fruit development from tomato ripening mutants including *rin*, *nor*, and *Nr*. These data will be integrated into TED and a set of tools that allow users to compare



### TOM1 TC Search

Only the TC No. from the most recently updated TIGR tomato gene index (currently version 8.0) can be searchable. These TC No. fall into range from TC115712 to TC131636. If you TC No. are out of this range, please update them from [here](#). The TC number must be separated by spaces, carriage returns (new lines) or tabs

Please input or paste the TC list:

TC118459  
TC131378  
TC125527  
TC118006  
TC125381

Or upload the file containing the TC list:

Here is a sample file: [sample.txt](#)

### TOM1 TC Search

[Right click here to get tab-delimited txt file](#)

TC #	Clone ID	GenBank Best hit	E-value	Graph Pattern
TC118459	cLED24H19	dbj BAB89702.1  P0504E02.2 [Oryza sativa (japonica cultivar-group)]	3e-021	<a href="#">graph pattern</a>
TC131378	cLED24H3	gb AAN03626.1 AF406702_1 BEL1-related homeotic protein 29 [Solanum tuberosum]	5e-035	<a href="#">graph pattern</a>
	cLED19C14	gb AAN03626.1 AF406702_1 BEL1-related homeotic protein 29 [Solanum tuberosum]	5e-035	<a href="#">graph pattern</a>
TC125527	cLED11B4	ref NP_194258.1  UMP/CMP kinase like protein; protein id: At4g25280.1 [Arabidopsis thaliana]	4e-058	<a href="#">graph pattern</a>
	cLED24H4	sp O24464 KAD_PRUAR ADENYLATE KINASE (ATP-AMP TRANSPHOSPHORYLASE)	7e-020	<a href="#">graph pattern</a>
TC118006	cLED24H7	dbj BAB44155.1  hydroxypyruvate reductase [Bruguiera gymnorrhiza]	7e-037	<a href="#">graph pattern</a>
TC125381	cLEC6M6	ref NP_195272.1  putative protein; protein id: At4g35470.1, supported by cDNA: gi_18252196 [Arabidopsis thaliana]	5e-033	<a href="#">graph pattern</a>
	cLED24I1	ref NP_179336.1  unknown protein; protein id: At2g17440.1 [Arabidopsis thaliana]	5e-033	<a href="#">graph pattern</a>

**Figure 30. TED TC mining query form and its results.** The query form accepts a set of TCs or a file containing a set of TCs. All of the clones on TOM1 belonging to the queried TCs will be displayed (right).

different datasets will be developed. Additional data analysis tools including clustering will be implemented in TED as the database develops.

## **Conclusion**

An online interactive database, currently containing tomato expression profiling data for fruit development, was created in this study. Microarray data from ten time points with nine pair-wise comparisons during fruit development was generated in our lab using tomato TOM1 cDNA arrays (<http://bti.cornell.edu/CGEP/CGEP.html>). All the data in TED was normalized and processed using methodology as described in this study. The raw microarray data can be obtained from the tomato microarray database (TMD; <http://tmd.bti.cornell.edu>) by users who want to analyze the data using their own methods.

Several useful data mining, visualization, and analysis tools were implemented in TED and new tools are under continuous development based on users' requests and needs. TED is a valuable data mining and analysis resource for researchers who are interested in tomato fruit development and ripening-related studies.

## CHAPTER VI

### SUMMARY

The first part of this research was focused on isolation and characterization of fruit ripening related genes. Four tomato cDNA libraries were screened at low stringency with 10 fruit development and ripening-related genes resulting in ~3000 primary positives. Expression analysis of half of these positives in fruits at mature green and breaker stage using cDNA microarray technology identified eight ripening induced genes. Northern blot analysis and previously published data confirmed seven of them as being ripening-related. One novel gene identified in this analysis, LeEREBP1, was chosen for further characterization. LeEREBP1 encodes an AP2/ERF-domain transcription factor and is ethylene inducible. The expression profiles of LeEREBP1 comply with other typical ripening related genes. Transgenic plants with increased and decreased expression of LeEREBP1 were generated.

In the second part of this research, a large tomato EST dataset was mined for transcriptome analysis by using raw EST counts as a measure of relative expression value. By clustering genes according to their tissue expression profiles, correlated expression patterns of genes were generated and genes with similar functions were usually grouped together. In addition, tissues were clustered for relatedness according to their profiles. Statistical analysis of EST prevalence in the fruit and pathogenesis-related libraries resulted in 343 genes being classified as fruit ripening induced, 185 genes were identified as fruit ripening repressed, and 169 genes were identified as pathogenesis

related. I further compared ripening induced genes between tomato and grape and identified homologous genes induced by both tomato and grape fruit ripening.

The third part of this research was to implement an online interactive database – Tomato Expression Database (TED). TED contains normalized and processed expression data for approximately 11,049 ESTs over ten time points during fruit development. It also contains comprehensive annotation information of each EST. Through TED, we provide multiple approaches to pursue analysis of specific genes of interest and/or access the larger microarray data set to identify sets of genes that may behave in a pattern of interest to the user. In addition, a set of useful data mining and data visualization tools were developed.

## LITERATURE CITED

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC** (1991) Complementary-DNA sequencing - expressed sequence tags and human genome project. *Science* **252**: 1651–1656
- Aharoni A, Keizer LCP, Bouwmeester HJ, Sun ZK, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AMML, De Vos RCH et al.** (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**: 647–661
- Ahmad M, Cashmore AR** (1996) Seeing blue: the discovery of cryptochrome. *Plant Mol Biol* **30**: 851–861
- Alba R, Cordonnier-Pratt MM, Pratt LH** (2000) Fruit-localized phytochromes regulate lycopene accumulation independently of ethylene production in tomato. *Plant Physiol* **123**: 363–370
- Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR** (1999) EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* **284**: 2148–2152
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Ampomah-Dwamena C, Morris BA, Sutherland P, Veit B, Yao JL** (2002) Down-regulation of TM29, a tomato SEPALLATA homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol* **130**: 605–617
- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815
- Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP** (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**: 2270–2275
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K et al.** (2000) Gene ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29
- Ashburner M, Ball CA, Blake JA, Butler H, Cherry JM, Corradi J, Dolinski K, Eppig JT et al.** (2001) Creating the gene ontology resource: design and implementation. *Genome Res* **11**: 1425–1433
- Audic S, Claverie JM** (1997) The significance of digital gene expression profiles. *Genome Res* **7**: 986–995
- Barry CS, Blume B, Bouzayen M, Cooper W, Hamilton AJ, Grierson D** (1996) Differential expression of the 1-aminocyclopropane-1-carboxylate oxidase gene family of tomato. *Plant J* **9**: 525–535
- Bartley GE, Viitanen PV, Bacot KO, Scolnik PA** (1992) A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthesis pathway. *J Biol Chem* **267**: 5036–5039

- Biale J, Young R** (1981) Respiration and ripening in fruits-retrospect and prospect. *In* J Friend, M Rhodes, eds, Recent Advance in the Biochemistry of Fruit and Vegetables. Academic Press, London, pp 1–39
- Biggs MS, Harriman RW, Handa AK** (1986) Changes in gene-expression during tomato fruit ripening. *Plant Physiol* **81**: 395–403
- Bird CR, Smith CJS, Ray JA, Moureau P, Bevan MW, Bird AS, Hughes S, Morris PC, Grierson D, Schuch W** (1988) The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. *Plant Mol Biol* **11**: 651–662
- Bishop JO, Morton JG, Rosbash M, Richards M** (1974) 3 abundance classes in heLa-cell messenger-RNA. *Nature* **250**: 199–204
- Bleecker AB, Estelle MA, Somerville C, Kende H** (1988) Insensitivity to ethylene conferred by a dominant mutation in *Arabidopsis thaliana*. *Science* **241**: 1086–1089
- Bortoluzzi S, d'Alessi F, Romualdi C, Danieli GA** (2000) The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach. *Genome Res* **10**: 344–349
- Boss PK, Sensi E, Hua C, Davies C, Thomas MR** (2002) Cloning and characterization of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci* **162**: 887–895
- Bourdon V, Naef M, Rao PH, Reuter V, Mok SC, Bosl GJ, Koul S, Murty VVVS, Kucherlapati RS, Chaganti RSK** (2002) Genomic and expression analysis of the 12p11-p12 amplicon using EST Arrays identifies two novel amplified and overexpressed genes. *Cancer Res* **62**: 6218–6223
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J et al.** (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet* **29**: 365–371
- Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N et al.** (2003) ArrayExpress - a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* **31**: 68–71
- Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D** (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* **97**: 262–267
- Buttner M, Singh KB** (1997) *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proc Natl Acad Sci USA* **94**: 5961–5966
- Chang C, Kwok SF, Bleecker AB, Meyerowitz EM** (1993) *Arabidopsis* ethylene-response gene ETR1 - similarity of product to 2-component regulators. *Science* **262**: 539–544
- Chao QM, Rothenberg M, Solano R, Roman G, Terzaghi W, Ecker JR** (1997) Activation of the ethylene gas response pathway in *Arabidopsis* by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell* **89**: 1133–1144
- Chou HH, Holmes MH** (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* **17**: 1093–1104

- Clark KL, Larsen PB, Wang XX, Chang C** (1998) Association of the Arabidopsis CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors. *Proc Natl Acad Sci USA* **95**: 5401–5406
- Claverie JM** (1999) Computational methods for the identification of differential and coordinated gene expression. *Hum Mol Genet* **8**: 1821–1832
- Coomes KR, Baggerly KA, Stivers DN, Wang J, Gold D, Sung H, Lee S** (2002) Biology-driven clustering of microarray data. *In* S Lin, K Johnson, eds, *Methods of Microarray Data Analysis II*, pp 65–79
- Cordes S, Deikman J, Margossian LJ, Fischer RL** (1989) Interaction of a developmentally regulated DNA-binding factor with sites flanking two different fruit-ripening genes from tomato. *Plant Cell* **1**: 1025–1034
- D'Auria JC, Chen F, Pichersky E** (2002) Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiol* **130**: 466–476
- Davies C, Robinson SP** (2000) Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. *Plant Physiol* **122**: 803–812
- Deikman J, Fischer RL** (1988) Interaction of a DNA-binding factor with the 5'-flanking region of an ethylene-responsive fruit ripening gene from tomato. *Embo J* **7**: 3315–3320
- Dellapenna D, Lincoln JE, Fischer RL, Bennett AB** (1989) Transcriptional analysis of polygalacturonase and other ripening associated genes in rutgers, *rin*, *nor*, and *Nr* tomato fruit. *Plant Physiol* **90**: 1372–1377
- Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P et al.** (1998) A physical map of 30,000 human genes. *Science* **282**: 744–746
- Deng XW, Caspar T, Quail PH** (1991) COP1 - a regulatory locus involved in light-controlled development and gene-expression in Arabidopsis. *Gene Deve* **5**: 1172–1182
- DeRisi JL, Iyer VR, Brown PO** (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686
- Diatchenko L, Lau YFC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD** (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93**: 6025–6030
- Duggan DJ, Bittner M, Chen YD, Meltzer P, Trent JM** (1999) Expression profiling using cDNA microarrays. *Nat Genet* **21**: 10–14
- Edgar R, Domrachev M, Lash AE** (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207–210
- Eisen MB, Spellman PT, Brown PO, Botstein D** (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868

- Elliott KJ, Butler WO, Dickinson CD, Konno Y, Vedvick TS, Fitzmaurice L, Mirkov TE** (1993) Isolation and characterization of fruit vacuolar invertase genes from 2 tomato species and temporal differences in messenger-RNA levels during fruit ripening. *Plant Mol Biol* **21**: 515–524
- Espartero J, Pintortoro JA, Pardo JM** (1994) Differential accumulation of s-adenosylmethionine synthetase transcripts in response to salt stress. *Plant Mol Biol* **25**: 217–227
- Ewing B, Green P** (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186–194
- Ewing B, Hillier L, Wendl MC, Green P** (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175–185
- Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM** (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* **9**: 950–959
- Ewing RM, Claverie JM** (2000) EST databases as multi-conditional gene expression datasets. *Pac Symp Biocomput*: 430–432
- Eyal Y, Meller Y, Levyadun S, Fluhr R** (1993) A basic-type PR-1 promoter directs ethylene responsiveness, vascular and abscission zone-specific expression. *Plant J* **4**: 225–234
- Fankhauser C, Chory J** (2000) RSF1, an Arabidopsis locus implicated in phytochrome A signaling. *Plant Physiol* **124**: 39–45
- Ferrie BJ, Beaudoin N, Burkhart W, Bowsher CG, Rothstein SJ** (1994) The cloning of 2 tomato lipoxygenase genes and their differential expression during fruit ripening. *Plant Physiol* **106**: 109–118
- Fillatti J, Kiser J, Rose B, Comai L** (1987). Tomato transformation. *In* D Nevins, R Jones, eds, *Tomato Biotechnology*. Alan R. Liss, New York, pp 199–210
- Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M** (2000) Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* **12**: 393–404
- Gadea J, Mayda ME, Conejero V, Vera P** (1996) Characterization of defense-related genes ectopically expressed in viroid-infected tomato plants. *Mol Plant Microbe Interac* **9**: 409–415
- Gardiner-Garden M, Littlejohn TG** (2001) A comparison of microarray databases. *Briefings in Bioinformatics* **2**: 143–158
- German MA, Dai N, Chmelnitsky I, Sobolev I, Salts Y, Barg R, Schaffer AA, Granot D** (2002) LeFRK4, a novel tomato (*Lycopersicon esculentum* Mill.) fructokinase specifically expressed in stamens. *Plant Sci* **163**: 607–613
- Giovannoni J** (2001) Molecular biology of fruit maturation and ripening. *Annu Rev Plant Physiol Plant Mol Biol* **52**: 725–749
- Giovannoni J, Kannan P, Lee S, Yen HC** (1998) Genetic approaches to manipulation of fruit development and quality in tomato. *In* K Cockshull, ed, *Genetic and Environmental Manipulation of Horticultural Crops*. CAB International Publishers, pp 1–15



- Giovannoni J, Yen H, Shelton B, Miller S, Vrebalov J, Kannan P, Tieman D, Hackett R, Grierson D, Klee H** (1999) Genetic mapping of ripening and ethylene-related loci in tomato. *Theor Appl Genet* **98**: 1005–1013
- Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al.** (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100
- Gray JE, Picton S, Giovannoni JJ, Grierson D** (1994) The use of transgenic and naturally-occurring mutants to understand and manipulate tomato fruit ripening. *Plant Cell and Environment* **17**: 557–571
- Greller LD, Tobin FL** (1999) Detecting selective expression of genes and proteins. *Genome Res* **9**: 282–296
- Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang CM, He XH, Han Y, Martin GB** (2002) Tomato transcription factors Pti4, Pti5, and Pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell* **14**: 817–831
- Gu YQ, Yang C, Thara VK, Zhou J, Martin GB** (2000) Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. *Plant Cell* **12**: 771–785
- Guzman P, Ecker JR** (1990) Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant Cell* **2**: 513–523
- Hacia J** (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet* **21**: 42–47
- Hamilton AJ, Lycett GW, Grierson D** (1990) Antisense gene that inhibits synthesis of the hormone ethylene in transgenic plants. *Nature* **346**: 284–287
- Hao DY, Ohme-Takagi M, Sarai A** (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *J Biol Chem* **273**: 26857–26861
- Hart CM, Nagy F, Meins F** (1993) A 61 bp enhancer element of the tobacco beta-1,3-glucanase b-gene interacts with one or more regulated nuclear proteins. *Plant Mol Biol* **21**: 121–131
- Herrero J, Valencia A, Dopazo J** (2001) A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* **17**: 126–136
- Hobson G, Grierson D** (1993) Tomato. *In* GB Seymour, JE Taylor, GA Tucker, eds, *Biochemistry of Fruit Ripening*, pp 405–422
- Hoerberichts FA, Orzaez D, van der Plas LHW, Woltering EJ** (2001) Changes in gene expression during programmed cell death in tomato cell suspensions. *Plant Mol Biol* **45**: 641–654
- Honys D, Twell D** (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol* **132**: 640–652
- Hua J, Chang C, Sun Q, Meyerowitz EM** (1995) Ethylene insensitivity conferred by Arabidopsis ERS gene. *Science* **269**: 1712–1714
- Hua J, Meyerowitz EM** (1998) Ethylene responses are negatively regulated by a receptor gene family in Arabidopsis thaliana. *Cell* **94**: 261–271

- Hua J, Sakai H, Nourizadeh S, Chen QHG, Bleecker AB, Ecker JR, Meyerowitz EM** (1998) EIN4 and ERS2 are members of the putative ethylene receptor gene family in Arabidopsis. *Plant Cell* **10**: 1321–1332
- Huang XQ, Madan A** (1999) CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868–877
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA et al.** (2000) Functional discovery via a compendium of expression profiles. *Cell* **102**: 109–126
- Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, Wang HY et al.** (1997) A genome-based resource for molecular cardiovascular medicine - Toward a compendium of cardiovascular genes. *Circulation* **96**: 4146–4203
- Ishimaru M, Kobayashi S** (2002) Expression of a xyloglucan endo-transglycosylase gene is closely related to grape berry softening. *Plant Sci* **162**: 621–628
- Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO** (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**: 83–87
- Johnson PR, Ecker JR** (1998) The ethylene gas signal transduction pathway: A molecular perspective. *Annu Rev Genet* **32**: 227–254
- Kachalo S, Arbieva Z, Liang J** (2002) Assessing effect of crosshybridization on oligonucleotide microarrays. *In Critical Assessment of Microarray Data Analysis (CAMDA02)*, **In press**
- Kende H, Zeevaart JAD** (1997) The five "classical" plant hormones. *Plant Cell* **9**: 1197–1210
- Kerr MK, Afshari CA, Bennett L, Bushel P, Martinez J, Walker NJ, Churchill GA** (2002) Statistical analysis of a gene expression microarray experiment with replication. *Stat Sinica* **12**: 203–217
- Khan J, Bittner ML, Saal LH, Teichmann U, Azorsa DO, Gooden GC, Pavan WJ, Trent JM, Meltzer PS** (1999) cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proc Natl Acad Sci USA* **96**: 13264–13269
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS** (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* **7**: 673–679
- Kieber JJ, Rothenberg M, Roman G, Feldmann KA, Ecker JR** (1993) CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein-kinases. *Cell* **72**: 427–441
- Klee HJ** (2002) Control of ethylene-mediated processes in tomato at the level of receptors. *J Exp Bot* **53**: 2057–2063
- Knapp J, Moureau P, Schuch W, Grierson D** (1989) Organization and expression of polygalacturonase and other ripening-related genes in Ailsa Craig, Never ripe and Ripening inhibitor tomato mutants. *Plant Mol Biol* **12**: 105–116

- Lanahan MB, Yen HC, Giovannoni JJ, Klee HJ** (1994) The never ripe mutation blocks ethylene perception in tomato. *Plant Cell* **6**: 521–530
- Leclercq J, Adams-Phillips LC, Zegzouti H, Jones B, Latche A, Giovannoni JJ, Pech JC, Bouzayen M** (2002) LeCTR1, a tomato CTR1-like gene, demonstrates ethylene signaling ability in Arabidopsis and novel expression patterns in tomato. *Plant Physiol* **130**: 1132–1142
- Leerkes MR, Caballero OL, Mackay A, Torloni H, O'Hare MJ, Simpson AJG, de Souza SJ** (2002) *In silico* comparison of the transcriptome derived from purified normal breast cells and breast tumor cell lines reveals candidate upregulated genes in breast tumor cells. *Genomics* **79**: 257–265
- Liang P, Pardee AB** (1992) Differential display of eukaryotic messenger-RNA by means of the polymerase chain-reaction. *Science* **257**: 967–971
- Lin Z, Hackett RM, Payton S, Grierson D** (1998) A tomato sequence (AJ005077) encoding an Arabidopsis CTR1 homologue. *Plant Physiol* **117**: 1125–1125
- Lincoln JE, Cordes S, Read E, Fischer RL** (1987) Regulation of gene-expression by ethylene during lycopersicon-esculentum (tomato) fruit-development. *Proc Natl Acad Sci USA* **84**: 2793–2797
- Lincoln JE, Fischer RL** (1988) Regulation of gene-expression by ethylene in wild-type and *rin* tomato (*lycopersicon esculentum*) fruit. *Plant Physiol* **88**: 370–374
- Linthorst HJM, Danhash N, Brederode FT, Vankan JAL, Dewit PJGM, Bol JF** (1991) Tobacco and tomato PR proteins homologous to win and pro-hevein lack the hevein domain. *Mol Plant Microbe Interac* **4**: 586–592
- Marín-Rodríguez MC, Orchard J, Seymour GB** (2002) Pectate lyases, cell wall degradation and fruit softening. *J Exp Bot* **53**: 2115–2119
- Mateos A, Herrero J, Tamames J, Dopazo J** (2002) Supervised neural networks for clustering conditions in DNA array data after reducing noise by clustering gene expression profiles. In S Lin, K Johnson, eds, *Methods of Microarray Data Analysis II*, pp 91–103
- Maunder MJ, Holdsworth MJ, Slater A, Knapp JE, Bird CR, Schuch W, Grierson D** (1987) Ethylene stimulates the accumulation of ripening-related messenger-RNAs in tomatoes. *Plant Cell Environ* **10**: 177–184
- Mine T, Hiyoshi T, Kasaoka K, Ohyama A** (2003) CIP353 encodes an AP2/ERF-domain protein in potato (*Solanum tuberosum* L.) and responds slowly to cold stress. *Plant Cell Physiol* **44**: 10–15
- Mu JH, Stains JP, Kao TH** (1994) Characterization of a pollen-expressed gene encoding a putative pectin esterase of *petunia inflata*. *Plant Mol Biol* **25**: 539–544
- Muir SR, Collins GJ, Robinson S, Hughes S, Bovy A, De Vos CHR, van Tunen AJ, Verhoeyen ME** (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat Biotech* **19**: 470–474
- Muschietti J, Eyal Y, McCormick S** (1998) Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* **10**: 319–330

- Mustilli AC, Fenzi F, Ciliento R, Alfano F, Bowler C** (1999) Phenotype of the tomato high pigment-2 mutant is caused by a mutation in the tomato homolog of DEETIOLATED1. *Plant Cell* **11**: 145–157
- Mysore KS, D'Ascenzo M, He X, Martin G** (2003) Overexpression of the disease resistant gene Pto in tomato induces gene expression changes similar to immune responses in human and Drosophila. *Plant Physiol* **132**: 1901–1912
- Neff MM, Fankhauser C, Chory J** (2000) Light: an indicator of time and place. *Gene Dev* **14**: 257–271
- Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M** (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. *Proc Natl Acad Sci USA*. **100**: 8007–8012
- Nunan KJ, Davies C, Robinson SP, Fincher GB** (2001) Expression patterns of cell wall-modifying enzymes during grape berry development. *Planta* **214**: 257–264
- O'Donnell PJ, Calvert C, Atzorn R, Wasternack C, Leyser HMO, Bowles DJ** (1996) Ethylene as a signal mediating the wound response of tomato plants. *Science* **274**: 1914–1917
- Oeller PW, Wong LM, Taylor LP, Pike DA, Theologis A** (1991) Reversible inhibition of tomato fruit senescence by antisense RNA. *Science* **254**: 437–439
- Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y** (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J* **33**: 1001–1011
- Ohme-Takagi M, Shinshi H** (1995) Ethylene-inducible DNA-binding proteins that interact with an ethylene-responsive element. *Plant Cell* **7**: 173–182
- Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K** (1992) Large-scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene-expression. *Nat Genet* **2**: 173–179
- Ouaked F, Rozhon W, Lecourieux D, Hirt H** (2003) A MAPK pathway mediates ethylene signaling in plants. *Embo J* **22**: 1282–1288
- Pan QL, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R** (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: Tomato and *Arabidopsis*. *Genetics* **155**: 309–322
- Papadopoulos N, Nicolaidis NC, Wei YF, Ruben SM, Carter KC, Rosen CA et al.** (1994) Mutation of a MUTL homolog in hereditary colon-cancer. *Science* **263**: 1625–1629
- Pelech SL, Sanghera JS** (1992) Mitogen-activated protein-kinases - versatile transducers for cell signaling. *Trends Biochem Sci* **17**: 233–238
- Penmetza RV, Cook DR** (1997) A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* **275**: 527–530

- Peters JL, Schreuder MEL, Verduin SJW, Kendrick RE** (1992) Physiological characterization of a high-pigment mutant of tomato. *Photochem Photobiol* **56**: 75–82
- Picton S, Gray JE, Payton S, Barton SL, Lowe A, Grierson D** (1993) A histidine decarboxylase-like messenger-RNA is involved in tomato fruit ripening. *Plant Mol Biol* **23**: 627–631
- Pirrung MC** (1999) Histidine kinases and two-component signal transduction systems. *Chem Biol* **6**: R167–R175
- Pnueli L, Hareven D, Broday L, Hurwitz C, Lifschitz E** (1994) The TM5 MADS box gene mediates organ differentiation in the 3 inner whorls of tomato flowers. *Plant Cell* **6**: 175–186
- Quackenbush J** (2001) Computational analysis of microarray data. *Nat Rev Genet* **2**: 418–427
- Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J** (2001) The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res* **29**: 159–164
- Ramakrishna W, Deng ZP, Ding CK, Handa AK, Ozminkowski RH** (2003) A novel small heat shock protein gene, *vis1*, contributes to pectin depolymerization and juice viscosity in tomato fruit. *Plant Physiol* **131**: 725–735
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL** (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**: 1742–1745
- Ray J, Moureau P, Bird C, Bird A, Grierson D, Maunders M, Truesdale M, Bramley P, Schuch W** (1992) Cloning and characterization of a gene involved in phytoene synthesis from tomato. *Plant Mol Biol* **19**: 401–404
- Riechmann JL, Meyerowitz EM** (1998) The AP2/EREBP family of plant transcription factors. *Biol Chem* **379**: 633–646
- Riechmann JL, Wang M, Meyerowitz EM** (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETELA1, APETELA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res* **24**: 3134–3141
- Ringwald M, Eppig JT, Begley DA, Corradi JP, McCright IJ, Hayamizu TF, Hill DP, Kadin JA, Richardson JE** (2001) The Mouse Gene Expression Database (GXD). *Nucleic Acids Res* **29**: 98–101
- Rocha S, Martin A, Meek D, Perkins N** (2003) p53 represses cyclin D1 transcription through down regulation of Bcl-3 and inducing increased association of the p52 NF- $\kappa$ B subunit with histone deacetylase. *Mol Cell Biol* **23**: 4713–4727
- Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y et al.** (1995) Familial alzheimers-disease in kindreds with missense mutations in a gene on chromosome-1 related to the alzheimers-disease type-3 gene. *Nature* **376**: 775–778
- Roman G, Lubarsky B, Kieber JJ, Rothenberg M, Ecker JR** (1995) Genetic-analysis of ethylene signal-transduction in *Arabidopsis thaliana* - 5 novel mutant loci integrated into a stress-response pathway. *Genetics* **139**: 1393–1409

- Romualdi C, Bortoluzzi S, Danieli GA** (2001) Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests. *Hum Mol Genet* **10**: 2133–2141
- Ronen G, Cohen M, Zamir D, Hirschberg J** (1999) Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J* **17**: 341–351
- Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR et al.** (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol* **131**: 419–429
- Rose JKC, Lee HH, Bennett AB** (1997) Expression of a divergent expansion gene is fruit-specific and ripening-regulated. *Proc Natl Acad Sci USA* **94**: 5955–5960
- Sakai H, Hua J, Chen QHG, Chang CR, Medrano LJ, Bleecker AB, Meyerowitz EM** (1998) ETR2 is an ETR1-like gene involved in ethylene signaling in Arabidopsis. *Proc Natl Acad Sci USA* **95**: 5812–5817
- Sato F, Kitajima S, Koyama T, Yamada Y** (1996) Ethylene-induced gene expression of osmotic-like protein, a neutral isoform of tobacco PR-5, is mediated by the AGCCGCC cis-sequence. *Plant Cell Physiol* **37**: 249–255
- Schaller GE, Ladd AN, Lanahan MB, Spanbauer JM, Bleecker AB** (1995) The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer. *J Biol Chem* **270**: 12526–12530
- Schena M, Shalon D, Davis RW, Brown PO** (1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* **270**: 467–470
- Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW** (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* **93**: 10614–1061
- Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM** (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proc Natl Acad Sci USA* **97**: 11655–11660
- Schmitt AO, Specht T, Beckmann G, Dahl E, Pilarsky CP, Hinzmann B, Rosenthal A** (1999) Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucleic Acids Res* **27**: 4251–4260
- Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE et al.** (1996) A gene map of the human genome. *Science* **274**: 540–546
- Semple CAM, Morris SW, Porteous DJ, Evans KL** (2000) *In silico* identification of transcripts and SNPs from a region of 4p linked with bipolar affective disorder. *Bioinformatics* **16**: 735–738
- Seymour GB, Fray RG, Hill P, Tucker GA** (1993) Down-regulation of two non-homologous endogenous tomato genes with a single chimaeric sense gene construct. *Plant Mol Biol* **23**: 1–9
- Shalon D, Smith SJ, Brown PO** (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**: 639–645

- Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC et al.** (2001) The Stanford Microarray Database. *Nucleic Acids Res* **29**: 152–155
- Smith DL, Gross KC** (2000) A family of at least seven beta-galactosidase genes is expressed during tomato fruit development. *Plant Physiol* **123**: 1173–1183
- Smith CJS, Watson CF, Morris PC, Bird CR, Seymour GB, Gray JE, Harding SE, Tucker GA, Schuch W, Grierson D** (1989) Inheritance and effect on ripening of antisense polygalacturonase genes in transgenic tomatoes. *Plant Mol Bio* **14**: 369–379
- Soh MS, Hong SH, Hanzawa H, Furuya M, Nam HG** (1998) Genetic identification of FIN2, a far red light-specific signaling component of *Arabidopsis thaliana*. *Plant J* **16**: 411–419
- Solano R, Stepanova A, Chao QM, Ecker JR** (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Gene Dev* **12**: 3703–3714
- Southern E, Mir K, Shchepinov M** (1999) Molecular interactions on microarrays. *Nat Genet* **21**: 5–9
- Southern EM** (1975) Detection of specific sequences among DNA fragments separated by gel-electrophoresis. *J Mol Biol* **98**: 503–517
- Speirs J, Lee E, Holt K, Yong-Duk K, Scott NS, Loveys B, Schuch W** (1998) Genetic manipulation of alcohol dehydrogenase levels in ripening tomato fruit affects the balance of some flavor aldehydes and alcohols. *Plant Physiol* **117**: 1047–1058
- Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S et al.** (2002) Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol* **3**: research0046.0041–0046.0049
- Staub JM, Wei N, Deng XW** (1996) Evidence for FUS6 as a component of the nuclear-localized COP9 complex in *Arabidopsis*. *Plant Cell* **8**: 2047–2056
- Stekel DJ, Git Y, Falciani F** (2000) The comparison of gene expression from multiple cDNA libraries. *Genome Res* **10**: 2055–2061
- Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, Weitz CJ** (2002) Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**: 78–83
- Strausberg RL, Greenhut SF, Grouse LH, Schaefer CF, Buetow KH** (2001) *In silico* analysis of cancer through the cancer genome anatomy project. *Trends Cell Biol* **11**: S66–S71
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR** (1999) Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* **96**: 2907–2912
- Tang WH, Ezcurra I, Muschietti J, McCormick S** (2002) A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell* **14**: 2277–2287
- Tanimoto M, Roberts K, Dolan L** (1995) Ethylene is a positive regulator of root hair development in *Arabidopsis thaliana*. *Plant J* **8**: 943–948

- Tanksley SD, Ganai MW, Prince JP, Devicente MC, Bonierbale MW et al.** (1992) High-density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM** (1999) Systematic determination of genetic network architecture. *Nat Genet* **22**: 281–285
- Tieman DM, Ciardi JA, Taylor MG, Klee HJ** (2001) Members of the tomato LeEIL (EIN3-like) gene family are functionally redundant and regulate ethylene responses throughout plant development. *Plant J* **26**: 47–58
- Tieman DV, Taylor MG, Ciardi JA, Klee HJ** (2000) The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. *Proc Natl Acad Sci USA* **97**: 5663–5668
- Tornero P, Gadea J, Conejero V, Vera P** (1997) Two PR-1 genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development. *Mol Plant Microbe Interac* **10**: 624–634
- Tournier B, Sanchez-Ballesta MT, Jones B, Pesquet E, Regad F, Latche A, Pech JC, Bouzayen M** (2003) New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett* **550**: 149–154
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB** (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**: 520–525
- Van der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC** (2000) Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. *Plant Cell* **12**: 2283–2294
- Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S** (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* **14**: 1441–1456
- Van Kan JAL, Joosten MHJ, Wagemakers CAM, Vandenbergvelthuis GCM, Dewit PJGM** (1992) Differential accumulation of messenger-RNAs encoding extracellular and intracellular PR proteins in tomato induced by virulent and avirulent races of *cladosporium-fulvum*. *Plant Mol Biol* **20**: 513–527
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW** (1995) Serial analysis of gene expression. *Science* **270**: 484–487
- Vierling E** (1991) The roles of heat-shock proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* **42**: 579–620
- Vingron M, Hoheisel J** (1999) Computation aspects of expression data. *J Mol Med* **77**: 3–7
- Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J** (2002) A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*Rin*) locus. *Science* **296**: 343–346



- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Gandour G, Perkins N, Winchester E, Lander ES** (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**: 1077–1082
- Wang KLC, Li H, Ecker JR** (2002) Ethylene biosynthesis and signaling networks. *Plant Cell* **14**: S131–S151
- Waterhouse PM, Helliwell CA** (2003) Exploring plant genomes by RNA-induced gene silencing. *Nat Rev Genet* **4**: 29–38
- Wei N, Chamovitz DA, Deng XW** (1994) Arabidopsis COP9 is a component of a novel signaling complex mediating light control of development. *Cell* **78**: 117–124
- Wei N, Deng XW** (1999) Making sense of the COP9 signalosome - a regulatory protein complex conserved from Arabidopsis to human. *Trends Genet* **15**: 98–103
- Weigel D** (1995) The APETALA2 domain is related to a novel type of DNA-binding domain. *Plant Cell* **7**: 388–389
- Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM** (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* **27**: 581–590
- Wilkinson JQ, Lanahan MB, Clark DG, Bleecker AB, Chang C, Meyerowitz EM, Klee HJ** (1997) A dominant mutant receptor from Arabidopsis confers ethylene insensitivity in heterologous plants. *Nat Biotech* **15**: 444–447
- Wilkinson JQ, Lanahan MB, Yen HC, Giovannoni JJ, Klee HJ** (1995) An ethylene-inducible component of signal-transduction encoded by Never-Ripe. *Science* **270**: 1807–1809
- Xiao FM, Tang XY, Zhou JM** (2001) Expression of 35S::Pto globally activates defense-related genes in tomato plants. *Plant Physiol* **126**: 1637–1645
- Yang SF, Hoffman NE** (1984) Ethylene biosynthesis and its regulation in higher plants. *Annu Rev Plant Physiol Plant Mol Biol* **35**: 155–189
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP** (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: art. no.–e15
- Yang YH, Speed T** (2002) Design issues for cDNA microarray experiments. *Nat Rev Genet* **3**: 579–588
- Yen HC, Lee SY, Tanksley SD, Lanahan MB, Klee HJ, Giovannoni JJ** (1995) The tomato Never-Ripe locus regulates ethylene-inducible gene-expression and is linked to a homolog of the arabidopsis ETR1 gene. *Plant Physiol* **107**: 1343–1353
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X et al.** (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79–92
- Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latche A, Pech JC, Bouzayen M** (1999) Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J* **18**: 589–600

- Zheng LS, Heupel RC, Dellapenna D** (1992) The beta-subunit of tomato fruit polygalacturonase isoenzyme-1 - isolation, characterization, and identification of unique structural features. *Plant Cell* **4**: 1147–1156
- Zhou JM, Tang XY, Martin GB** (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *Embo J* **16**: 3207–3218
- Zhou XH, Kao MCJ, Wong WH** (2002) Transitive functional annotation by shortest-path analysis of gene expression data. *Proc Natl Acad Sci USA* **99**: 12783–12788

## APPENDIX A

**Table I.** Ripening-induced clones identified by cDNA microarray analysis and homologous to their corresponding probes

CloneID	Probe	BB/MG	Corresponding TC number and annotation
cLEG139L10	NOR	3.307	TC100002 AtNAC2
cLEG140I14	NOR	7.452	TC100002 AtNAC2
cLEG114O24	NOR	44.421	TC100002 AtNAC2
cLEG027H14	NOR	3.423	TC100002 AtNAC2
cLED023E3	NOR	10.26	TC100002 AtNAC2
cLEG069E17	NOR	2.144	TC100002 AtNAC2
cLEG125E12	NOR	3.357	TC100002 AtNAC2
cLEG086B17	NOR	3.641	TC100002 AtNAC2
cLEG007K12	NOR	3.223	TC100002 AtNAC2
cLEG107J16	NOR	3.626	TC100002 AtNAC2
cLEC102O3	NOR	4.921	TC100002 AtNAC2
cLEG099P1	NOR	3.759	TC100002 AtNAC2
cLEG107H6	NOR	2.405	TC100002 AtNAC2
cLEG091F18	NOR	3.189	TC100002 AtNAC2
cLEG013F8	NOR	5.717	TC100002 AtNAC2
cLEG007J4	NOR	5.414	TC100002 AtNAC2
cLEG022L18	NOR	16.007	TC101914 AtNAC2
cLEG077D19	NOR	3.961	TC101914 AtNAC2
cLEG105C4	NOR	7.796	TC101914 AtNAC2
cLEG014G2	NOR	4.397	TC101914 AtNAC2
cLEG080G17	NOR	4.651	TC101914 AtNAC2
cLEG004K4	NOR	4.846	TC101914 AtNAC2
cLEG013M7	NOR	2.58	TC101914 AtNAC2
cLEG088I14	NOR	4.133	TC101914 AtNAC2
cLEG107H4	NOR	8.37	TC101914 AtNAC2
cLEG028E2	NOR	5.129	TC101914 AtNAC2
cLEG041P4	NOR	4.417	TC101914 AtNAC2
cLEG005B5	NOR	5.261	TC101914 AtNAC2
cLEG129G2	NOR	9.543	TC101914 AtNAC2
cLEG114P11	NOR	7.242	TC101914 AtNAC2
cLEG079D24	NOR	5.618	TC101914 AtNAC2
cLEG019L21	NOR	2.294	TC101914 AtNAC2
cLEG123J15	NOR	3.546	TC101914 AtNAC2
cLED071K19	NOR	3.049	TC99988 OsNAC6
cLEG144I4	NOR	3.215	TC99988 OsNAC6
cLEG024A11	NOR	4.9	TC99988 OsNAC6
cLEG062E12	NOR	2.192	TC99988 OsNAC6
cLEG063N14	NOR	2.941	TC99988 OsNAC6

cLEG089O18	NOR	3.843	TC99988 OsNAC6
cLEG132B11	NOR	3.122	TC99988 OsNAC6
cLED075F16	NOR	2.008	TC99988 OsNAC6
cLEG030H17	NOR	3.2	TC99988 OsNAC6
cLEG070G14	NOR	2.736	TC99988 OsNAC6
cLED068I21	NOR	3.087	TC99988 OsNAC6
cLEG124A5	NOR	5.203	TC99988 OsNAC6
cLEG109L19	NOR	2.203	TC99988 OsNAC6
cLEG002J3	NOR	1.916	TC99988 OsNAC6
cLEG141K19	NOR	3.359	TC99988 OsNAC6
cLEG091K3	RIN/MC	2.424	TC100261 TDR4
cLEG034A15	RIN/MC	2.596	TC100261 TDR4
cLEG081A1	RIN/MC	3.202	TC100261 TDR4
cLED158G16	RIN/MC	2.194	TC100261 TDR4
cLEG113I23	RIN/MC	2.673	TC100261 TDR4
cLEG137O1	RIN/MC	3.911	TC100261 TDR4
cLEG127C4	RIN/MC	2.298	TC100261 TDR4
cLEG134P9	RIN/MC	2.226	TC100261 TDR4
cLEG082O24	RIN/MC	2.335	TC100261 TDR4
cLEG041F9	RIN/MC	3.534	TC100261 TDR4
cLEG084H12	RIN/MC	2.322	TC100261 TDR4
cLEG095N8	RIN/MC	2.55	TC100261 TDR4
cLEG037M6	RIN/MC	2.809	TC100261 TDR4
cLEG087A10	RIN/MC	5.196	TC100261 TDR4
cLEG074D15	RIN/MC	3.133	TC100261 TDR4
cLEG096C23	RIN/MC	2.707	TC100261 TDR4
cLEG107D20	RIN/MC	2.718	TC100261 TDR4
cLEG134H10	RIN/MC	2.566	TC100261 TDR4
cLED158P16	RIN/MC	3.039	TC100261 TDR4
cLED160L20	RIN/MC	2.985	TC100261 TDR4
cLED167B24	RIN/MC	3.041	TC100261 TDR4
cLEG118K5	RIN/MC	3.09	TC100261 TDR4
cLEG128J5	RIN/MC	2.078	TC100261 TDR4
cLEG113C10	RIN/MC	4.223	TC98600 MADS-RIN
cLEG105B6	RIN/MC	4.641	TC98600 MADS-RIN
cLEG017O1	DDTFR10/A	2.214	TC99621 AP2
cLEG014G19	DDTFR10/A	2.204	TC99621 AP2
cLEG010F21	DDTFR10/A	3.326	TC99621 AP2
cLEG038G22	DDTFR10/A	3.236	TC99621 AP2
cLEG055J16	DDTFR10/A	4.044	TC99621 AP2
cLED038C3	DDTFR10/A	2.167	TC99621 AP2
cLEG041N6	DDTFR10/A	3.615	TC99621 AP2
cLEG038D8	DDTFR10/A	2.267	TC99621 AP2
cLEG005L19	DDTFR10/A	3.959	TC99621 AP2
cLED040I20	DDTFR10/A	2.552	TC99621 AP2

cLEG057C13	DDTFR10/A	2.398	TC99621 AP2
cLEC010A3	DDTFR10/A	3.886	TC99621 AP2
cLEC046F9	DDTFR10/A	2.523	TC99346 DDTFR10/A
cLED012I5	DDTFR10/A	2.341	TC99346 DDTFR10/A
cLED035J3	DDTFR10/A	1.963	TC99346 DDTFR10/A
cLEG029B22	DDTFR10/A	7.554	TC103299 ethylene-responsive heat shock protein
cLEG094L8	NR	1.961	TC99771 ethylene receptor

**Table II.** Ripening-induced clones identified by cDNA microarray analysis and not homologous to their corresponding probes (false positives)

CloneID	Probe	BB/MG	Corresponding TC number and annotation
cLEG010J5	COP9	9.173	TC98539 phytoene synthase
cLEG011I21	COP9	2.288	TC98447 RuBisCO subunit binding-protein
cLED059L7	TCTR1	1.897	TC98313 chaperonin 60 beta chain precursor
cLEG005F14	DDTFR10/A	5.421	TC101691 heat shock protein 18p
cLEC054C23	DDTFR10/A	3.436	TC100946 probable triacylglycerol
cLEG062N24	DDTFR10/A	17.245	TC99248 cysteine protease component of protease-inhibitor complex
cLEG038D9	DDTFR10/A	2.474	TC101032 unknown
cLEC026H3	DDTFR10/A	5.33	TC101234 hypothetical protein
cLEG064P24	DDTFR10/A	2.131	TC106432
cLEG058K2	DDTFR10/A	2.179	AW223176 ferritin
cLEG006N8	DDTFR10/A	2.654	AW626011
cLEG029D14	DDTFR10/A	2.75	No hits found
cLEC062L9	DDTFR10/A	2.973	Bad sequence
cLEC030N3	NOR	3.759	TC99010 cold-induced glucosyl transferase
cLEC014C12	NOR	3.91	TC98766 alcohol dehydrogenase
cLEG074N11	NOR	3.895	TC98771 pectate lyase
cLEC078L12	NOR	2.305	TC98737 hypothetical protein AAG29221.1
cLED069H1	NOR	2.387	TC99218 unknown protein
cLED002D9	NOR	2.788	TC101733
cLEG057J21	NOR	3.634	TC104873
cLEL172F18	NOR	4.112	TC106617 Serine protease inhibitor
cLEG099N11	NOR	3.72	AW442272 putative pectate-lyase
cLEG077P3	NOR	2.278	Bad sequence
cLEC036E24	RIN/MC	1.882	TC99030 bZIP transcription factor
cLEG115M20	RIN/MC	3.077	TC98683 vacuolar invertase precursor
cLEL346A24	RIN/MC	11.582	BI933036 wound-induced proteinase inhibitor
cLEG021G7	RIN/MC	4.037	Bad sequence
cLEC028B23	RIN/MC	3.774	Bad sequence

**Table III.** Ripening-repressed clones identified by cDNA microarray analysis and homologous to their corresponding probes

CloneID	Probe	BB/MG	Corresponding TC number and annotation
cLEC049H19	RIN	0.404	TC99375 MADSbox protein
cLEG010F16	RIN	0.189	TC99375 MADSbox protein
cLEG020A3	RIN	0.082	TC99375 MADSbox protein
cLEG046K7	RIN	0.309	TC99375 MADSbox protein
cLEL130P8	RIN	0.489	TC99375 MADSbox protein
cLEL338H14	RIN	0.006	TC99375 MADSbox protein
cLEC087H23	RIN	0.428	TC99375 MADS-box protein
cLEG016L17	RIN	0.296	TC99375 MADS-box protein
cLED094F17	RIN	0.098	TC99375 MADS-box protein
cLEL119H20	RIN	0.403	TC99978 MADS-box protein
cLED023E12	RIN	0.437	TC99978 MADS-box protein MADS1
cLEG032F5	RIN	0.33	TC101162 MADS box transcription factor MADS1
cLEL169I17	RIN	0.462	TC101989 TDR5
cLEG029C3	RIN	0.455	TC99899 APETALA3 homolog LeAP3
cLEC002A12	NOR	0.377	TC106246 NAM (no apical meristem)-like protein
cLED010N16	NOR	0.147	TC106246 NAM (no apical meristem)-like protein
cLED014C15	NOR	0.426	TC106246 NAM (no apical meristem)-like protein
cLED039J4	NOR	0.458	TC106246 NAM (no apical meristem)-like protein
cLEG087M20	NOR	0.281	TC106246 NAM (no apical meristem)-like protein
cLEG128O14	NOR	0.248	TC102618 putative NAC domain protein
cLED079K11	NOR	0.421	TC99219 NAC domain protein NAC2
cLEC001B15	10A	0.119	TC100180 ethylene-responsive element binding protein
cLEC011E20	10A	0.159	TC100180 ethylene-responsive element binding protein
cLEC011H15	10A	0.336	TC100180 ethylene-responsive element binding protein
cLEC020L6	10A	0.412	TC100180 ethylene-responsive element binding protein
cLEC053G1	10A	0.011	TC100180 ethylene-responsive element binding protein
cLEC053I2	10A	0.184	TC100180 ethylene-responsive element binding protein
cLEC056P13	10A	0.436	TC100180 ethylene-responsive element binding protein
cLEC070G11	10A	0.107	TC100180 ethylene-responsive element binding protein
cLEC079I17	10A	0.398	TC100180 ethylene-responsive element binding protein
cLEC080M7	10A	0.207	TC100180 ethylene-responsive element binding protein
cLEC082I23	10A	0.078	TC100180 ethylene-responsive element binding protein
cLED014D22	10A	0.153	TC100180 ethylene-responsive element binding protein
cLED021I2	10A	0.482	TC100180 ethylene-responsive element binding protein
cLED024B9	10A	0.209	TC100180 ethylene-responsive element binding protein
cLED033M18	10A	0.095	TC100180 ethylene-responsive element binding protein
cLED045J12	10A	0.434	TC100180 ethylene-responsive element binding protein
cLEG010A6	10A	0.316	TC100180 ethylene-responsive element binding protein

cLEL124I12	10A	0.299	TC100180 ethylene-responsive element binding protein
cLEL255B15	10A	0.404	TC100180 ethylene-responsive element binding protein
cLEL279M9	10A	0.065	TC100180 ethylene-responsive element binding protein
cLEC029C9	10A	0.497	TC99660 contain AP2 domain transcription factor
cLEC057F19	10A	0.293	TC99660 contain AP2 domain transcription factor
cLED082P13	10A	0.064	TC99660 contain AP2 domain transcription factor
cLED153D19	10A	0.078	TC99660 contain AP2 domain transcription factor
cLEG038C13	10A	0.082	TC99660 contain AP2 domain transcription factor
cLEG057C9	10A	0.209	TC99660 contain AP2 domain transcription factor
cLEG096A21	10A	0.289	TC99660 contain AP2 domain transcription factor
cLEG097I3	10A	0.227	TC99660 contain AP2 domain transcription factor
cLEC032J7	10A	0.203	TC99989 AP2 domain protein
cLEC038G19	10A	0.109	TC99989 AP2 domain protein
cLEC061M20	10A	0.118	TC99989 AP2 domain protein
cLEC061N21	10A	0.086	TC99989 AP2 domain protein
cLED011I13	10A	0.488	TC99989 AP2 domain protein
cLED036D22	10A	0.048	TC99989 AP2 domain protein
cLED038A6	10A	0.056	TC99989 AP2 domain protein
cLED043J13	10A	0.356	TC99989 AP2 domain protein
cLED075O9	10A	0.299	TC99989 AP2 domain protein
cLEG074I21	10A	0.033	TC99989 AP2 domain protein
cLEG091A22	10A	0.186	TC99677 ethylene responsive element binding factor
cLEC042L7	10A	0.086	TC99677 ethylene responsive element binding factor
cLEL156J17	10A	0.327	AW220395 ethylene-responsive element binding factor
cLEC036L21	10A	0.163	TC100620 EREBP3
cLED168F20	10A	0.323	TC103863 Pti5



**Table IV.** Ripening-repressed clones identified by cDNA microarray analysis and not homologous to their corresponding probes (false positives)

CloneID	Probe	BB/MG	Corresponding TC number and annotation
cLEC075J4	RIN	0.073	TC100255 KED
cLED010G3	RIN	0.426	TC100255 KED
cLED039C23	RIN	0.173	TC100255 KED
cLED083D10	RIN	0.095	TC100255 KED
cLED095C12	RIN	0.075	TC100255 KED
cLED096F17	RIN	0.075	TC100255 KED
cLEG046I6	RIN	0.386	TC100255 KED
cLEG094D6	RIN	0.497	TC100255 KED
cLEG120M7	RIN	0.46	TC100255 KED
cLEG143L19	RIN	0.399	TC100255 KED
cLEG038L21	RIN	0.36	Bad sequence
cLEL202F15	RIN	0.087	Bad sequence
cLEG048E14	RIN	0.334	TC100119 ubiquitin-specific protease
cLEG044M20	RIN	0.412	TC101152 translation initiation factor 5
cLEG134K5	RIN	0.367	TC104461 protease-like protein
cLEG125K2	RIN	0.441	TC107590
cLEG111L18	RIN	0.039	TC108039 Protoporphyrinogen oxidase
cLED154J6	RIN	0.345	TC110131
cLEG008H3	RIN	0.376	TC99067 embryo-abundant protein
cLEG042C5	RIN	0.388	TC99259 Hypothetical protein
cLEG048K19	RIN	0.29	TC99556 leucine-rich repeat resistance protein
cLEC092K19	RIN	0.239	TC99944 AMP-binding protein
cLED083J6	NR	0.381	TC101050 WIZZ
cLED043G20	NR	0.459	TC99522 unknown
cLEL310E21	NOR	0.404	Bad sequence
cLEC079K9	NOR	0.181	BI930993
cLEG058B8	NOR	0.141	TC101151 transcription factor SF3
cLEL351E23	NOR	0.372	TC102975 60S ribosomal protein
cLEL158I23	NOR	0.456	TC98835 14-3-3 protein
cLEL213F7	NOR	0.047	TC99465 probable amine oxidase
cLEC008C12	NOR	0.39	TC99468 arbutin synthase
cLEC033C9	NOR	0.413	TC99706 dihydroipoamide dehydrogenase
cLEL205L2	EIN3	0.15	TC100143 EIL1 [ <i>Lycopersicon esculentum</i> ]
cLED013I21	COP9	0.484	AI781809
cLEC044O7	COP9	0.137	Bad sequence
cLEC045D13	COP9	0.252	TC100094 mutator-like transposase
cLED041D15	COP9	0.471	TC100094 mutator-like transposase
cLEL227D22	COP9	0.256	TC100094 mutator-like transposase

cLED077P19	COP9	0.074	TC100094 mutator-like transposase
cLEC033J2	COP9	0.201	TC101016 hypothetical protein
cLED081H11	COP9	0.089	TC101525 unknown protein
cLED083L15	COP9	0.053	TC102141 armadillo repeat-containing protein
cLEC035B22	10A	0.339	AI773054
cLEG021O18	10A	0.291	Bad sequence
cLED003D1	10A	0.362	TC102931
cLED037M14	10A	0.35	TC108528 glycoprotein specific UDP-glucuronyltransferase
cLEC082I21	10A	0.14	TC108805
cLEC023D10	10A	0.122	TC110446 cytochrome c oxidase
cLEC028B6	10A	0.185	TC99317 vacuolar ATP synthase
cLED004L22	10A	0.245	TC99912 small GTP-binding protein
cLEL140G17	10A	0.22	Bad sequence
cLEL102N17	10A	0.021	TC100434 auxin response
cLEC017G9	TCOP11	0.321	TC100022 calcium binding protein

**Table V.** Tomato pollen specific/related genes

TC	Putative function	E value
TC115714	l-ascorbate oxidase [Petunia x hybrida]	0.0
TC115719	Pollen-specific protein NTP303 precursor	0.0
TC115749	Pollen-specific protein NTP303 precursor	e-115
TC115750	putative beta-galactosidase [Nicotiana tabacum]	0.0
TC115920	Pectinesterase precursor (Pectin methylesterase) (PE)	e-139
TC115958	Anther specific LAT52 protein precursor	2e-079
TC115959	Anther specific LAT52 protein precursor	1e-085
TC115985	cell-wall invertase [Lycopersicon esculentum]	0.0
TC115986	cell-wall invertase [Lycopersicon esculentum]	e-156
TC115991	PGPS/D4 [Petunia x hybrida]	1e-094
TC115993	PGPS/D4 [Petunia x hybrida]	3e-041
TC116077	Probable pectate lyase P56 precursor	0.0
TC116113	tumor-related protein [Nicotiana glauca x Nicotiana langsdorffii]	3e-057
TC116143	Probable pectinesterase precursor (Pectin methylesterase) (PE)	2e-065
TC116144	pectin methylesterase-like protein [Oryza sativa (japonica cultivar-group)]	3e-012
TC116146	pectinesterase family [Arabidopsis thaliana]	e-106
TC116147	pectinesterase - Arabidopsis thaliana	2e-065
TC116184	putative protein [Arabidopsis thaliana]	e-130
TC116195	endo-1,3-beta-glucanase-like protein [Pyrus pyrifolia]	e-105
TC116202	cysteine-rich protein [Nicotiana tabacum]	1e-012
TC116243	ubiquitin conjugating enzyme E2 [Lycopersicon esculentum]	1e-061
TC116279	fructokinase [Lycopersicon esculentum]	0.0
TC116300	contains similarity to MtN3 protein [Arabidopsis thaliana]	2e-045
TC116304	nodulin MtN3 family protein [Arabidopsis thaliana]	9e-069
TC116306	PGPS/D7 [Petunia x hybrida]	2e-031
TC116321	actin	e-115
TC116336	Anther specific LAT52 protein precursor	2e-049
TC116374	fasciclin-like arabinogalactan-protein [Arabidopsis thaliana]	3e-031
TC116375	putative glyoxal oxidase [Oryza sativa (japonica cultivar-group)]	e-143
TC116389	unknown [Arabidopsis thaliana]	5e-031
TC116415	No hits found	
TC116437	LIM domain protein PLIM-2 [Nicotiana tabacum]	e-102
TC116449	Probable pectate lyase P59 precursor	0.0
TC116450	Probable pectate lyase P59 precursor	0.0
TC116474	Anther specific LAT52 protein precursor	2e-053
TC116520	putative glutamate-/aspartate-binding peptide [Arabidopsis thaliana]	4e-049

TC116541	putative monosaccharide transporter 1 [Petunia x hybrida]	0.0
TC116576	succinate dehydrogenase iron-protein subunit -like [Arabidopsis thaliana]	e-126
TC116632	Omega-3 fatty acid desaturase, endoplasmic reticulum	1e-084
TC116684	putative allantoinase [Solanum tuberosum]	e-147
TC116692	alternative oxidase 1a [Lycopersicon esculentum]	0.0
TC116795	unknown protein [Arabidopsis thaliana]	8e-027
TC116797	LIM domain protein PLIM1 [Nicotiana tabacum]	e-102
TC116826	No hits found	
TC116851	putative phospholipid cytidyltransferase [Arabidopsis thaliana]	e-138
TC116905	delta-12 fatty acid desaturase [Borago officinalis]	e-120
TC116915	cysteine-rich protein [Nicotiana tabacum]	3e-013
TC116926	protein kinase [Raphanus sativus]	3e-075
TC116969	putative pectinesterase [Arabidopsis thaliana]	8e-033
TC117023	putative phospholipase [Arabidopsis thaliana]	5e-066
TC117028	zinc finger protein [imported] - garden pea	e-164
TC117036	zinc finger domain-containing protein [Arabidopsis thaliana]	7e-057
TC117074	No hits found	
TC117094	Probable pectate lyase P59 precursor	0.0
TC117249	putative protein [Arabidopsis thaliana]	e-109
TC117344	amino acid transporter family [Arabidopsis thaliana]	1e-053
TC117351	No hits found	
TC117517	pectinesterase (EC 3.1.1.11) [imported] - Salix gilgiana	3e-015
TC117518	hexose transporter - like protein [Arabidopsis thaliana]	e-158
TC117543	unknown protein [Arabidopsis thaliana]	2e-013
TC117577	synaptobrevin-like protein [Oryza sativa (japonica cultivar-group)]	4e-085
TC117643	CBL-interacting protein kinase 11 [Arabidopsis thaliana]	6e-072
TC117668	No hits found	
TC117674	contains similarity to ABC transporter [Arabidopsis thaliana]	1e-044
TC117756	small GTP-binding protein	e-109
TC117771	No hits found	
TC117894	auxin-regulated protein [Phaseolus vulgaris]	2e-025
TC117940	hydrolase, alpha/beta fold family [Arabidopsis thaliana]	4e-057
TC117943	serine-rich protein [Arabidopsis thaliana]	9e-041
TC117964	auxin-regulated protein [Phaseolus vulgaris]	3e-035
TC117972	putative integral membrane protein [Arabidopsis thaliana]	3e-057
TC118016	unknown [Solanum chacoense]	4e-015
TC118049	No hits found	
TC118076	No hits found	
TC118201	hypothetical protein [Arabidopsis thaliana]	3e-035

TC118237	phosphatidic acid phosphatase-like protein [Arabidopsis thaliana]	3e-053
TC118241	peroxidase, putative [Arabidopsis thaliana]	2e-057
TC118298	NTS1 protein (Sperm Cell Expressed Gene) [Nicotiana tabacum]	5e-071
TC118303	cationic amino acid transporter -related protein [Arabidopsis thaliana]	0.0
TC118320	small basic membrane integral protein ZmSIP2-1 [Arabidopsis thaliana]	3e-071
TC118398	No hits found	
TC118473	putative peroxidase [Oryza sativa (japonica cultivar-group)]	1e-012
TC118503	receptor-like protein kinase PRK1 - tomato	0.0
TC118513	aluminium tolerance associated - like protein [Arabidopsis thaliana]	5e-036
TC118530	No hits found	
TC118571	No hits found	
TC118648	No hits found	
TC118745	zinc finger - like protein [Arabidopsis thaliana]	6e-032
TC118746	auxin-induced protein family [Arabidopsis thaliana]	2e-070
TC118757	hypothetical protein MMB12.11 [imported] - Arabidopsis thaliana	1e-082
TC118759	syntaxin SYP124 [Arabidopsis thaliana]	1e-051
TC118785	No hits found	
TC118887	senescence-associated protein-like protein [Oryza sativa (japonica cultivar-group)]	7e-037
TC119008	hypothetical protein [Arabidopsis thaliana]	3e-061
TC119024	eukaryotic initiation factor 4, eIF4-like protein [Arabidopsis thaliana]	e-146
TC119071	putative RHO GDP-dissociation inhibitor 1 [Arabidopsis thaliana]	1e-017
TC119101	peroxidase, putative [Arabidopsis thaliana]	2e-051
TC119107	GTPase activating protein-like [Arabidopsis thaliana]	5e-048
TC119186	No hits found	
TC119315	Serine/threonine Kinase [Persea americana]	3e-013
TC119368	No hits found	
TC119370	putative pathogenesis-related protein [Arabidopsis thaliana]	2e-029
TC119372	putative high-affinity potassium transporter [Oryza sativa (japonica cultivar-group)]	5e-085
TC119399	ubiquitin family [Arabidopsis thaliana]	1e-029
TC119464	GTPase activating protein-like [Arabidopsis thaliana]	7e-070
TC119514	No hits found	
TC119590	omega-6 fatty acid desaturase [Sesamum indicum]	2e-053
TC119617	annexin-like protein [Medicago sativa]	2e-062
TC119632	OJ1117_G01.5 [Oryza sativa (japonica cultivar-group)]	6e-032
TC119683	syntaxin SYP72 [Arabidopsis thaliana]	5e-041
TC119746	similar to ARF GAP-like zinc finger-containing protein [Oryza sativa (japonica cultivar-group)]	2e-052
TC123908	glutamate decarboxylase isozyme 1 [Nicotiana tabacum]	0.0
TC123935	Tubulin beta-1 chain	0.0
TC123975	L-ascorbate oxidase homolog precursor (Ascorbase)	e-161

TC123978	Calmodulin	7e-080
TC124087	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	2e-071
TC124088	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	3e-079
TC124102	putative purple acid phosphatase precursor [Ipomoea batatas]	e-117
TC124211	Pollen-specific protein NTP303 precursor	3e-077
TC124212	pectinesterase (pectin methylesterase) family [Arabidopsis thaliana]	e-159
TC124219	actin 97	0.0
TC124285	elicitor inducible gene Nt-SA1-1 [Nicotiana tabacum]	3e-033
TC124292	beta-tubulin [Naegleria gruberi]	0.0
TC124307	Elicitor inducible gene product Nt-SubE80 [Nicotiana tabacum]	1e-049
TC124346	pectin methyl-esterase - like protein [Arabidopsis thaliana]	e-145
TC124368	late-embryogenesis protein lea5 - common tobacco	9e-021
TC124405	No hits found	
TC124442	peptidyl-prolyl isomerase protein [Arabidopsis thaliana]	1e-077
TC124456	S-adenosylmethionine decarboxylase proenzyme	e-136
TC124504	nucleotide sugar epimerase-like protein [Arabidopsis thaliana]	0.0
TC124517	putative pectinesterase [Arabidopsis thaliana]	3e-032
TC124600	putative protein [Arabidopsis thaliana]	2e-029
TC124618	Pectate lyase precursor	e-104
TC124625	pollen specific actin-depolymerizing factor 1 [Nicotiana tabacum]	4e-064
TC124636	P-type H <sup>+</sup> -ATPase [Vicia faba]	0.0
TC124646	No hits found	
TC124648	S-adenosyl-methionine-sterol-C- methyltransferase [Nicotiana tabacum]	0.0
TC124666	lysine and histidine specific transporter, putative [Arabidopsis thaliana]	e-111
TC124695	plasma membrane H <sup>+</sup> -ATPase [Oryza sativa (japonica cultivar-group)]	3e-039
TC124697	plasma membrane ATPase 3 [Arabidopsis thaliana]	e-134
TC124722	phosphate/phosphoenolpyruvate translocator-like protein [Arabidopsis thaliana]	e-128
TC124750	14-3-3 protein 2	e-140
TC124779	No hits found	
TC124790	G-protein beta family [Arabidopsis thaliana]	1e-093
TC124791	predicted GPI-anchored protein [Arabidopsis thaliana]	0.0
TC124820	No hits found	
TC124826	polygalacturonase (EC 3.2.1.15) precursor (clone G27.1, G27.2) - common tobacco	0.0
TC124869	putative heterogeneous nuclear ribonucleoprotein A1 [Oryza sativa (japonica cultivar-group)]	e-118
TC124947	P0492F05.25 [Oryza sativa (japonica cultivar-group)]	1e-065
TC124963	unknown protein [Arabidopsis thaliana]	e-155
TC125018	expressed protein [Arabidopsis thaliana]	2e-086
TC125079	putative C-4 sterol methyl oxidase [Arabidopsis thaliana]	e-114
TC125085	shock protein SRC2 homolog [imported] - Arabidopsis thaliana	3e-022

TC125121	calcineurin-like phosphoesterase family [Arabidopsis thaliana]	3e-066
TC125133	Pollen-specific membrane integral protein	2e-090
TC125146	PGPS/NH15 [Petunia x hybrida]	4e-037
TC125280	bZIP transcription factor BZI-4 [Nicotiana tabacum]	6e-054
TC125301	receptor-like protein kinase PRK2 - tomato	0.0
TC125324	putative fatty acid desaturase [Arabidopsis thaliana]	e-150
TC125363	unknown protein [Arabidopsis thaliana]	9e-023
TC125387	auxin-repressed protein-like protein [Nicotiana tabacum]	6e-033
TC125392	proline transport protein 1 [imported] - tomato	0.0
TC125396	pectinacylesterase family [Arabidopsis thaliana]	6e-028
TC125445	No hits found	
TC125458	No hits found	
TC125484	serine/threonine protein kinase [Arabidopsis thaliana]	e-145
TC125488	putative protein [Arabidopsis thaliana]	1e-023
TC125529	Ki1 protein [Beta vulgaris]	e-131
TC125530	Serine/threonine Kinase [Persea americana]	e-111
TC125549	LIM domain protein PLIM-2 [Nicotiana tabacum]	3e-089
TC125621	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-114
TC125661	calcineurin B [Pisum sativum]	e-112
TC125702	No hits found	
TC125705	cysteine-rich protein [Nicotiana tabacum]	3e-015
TC125707	putative protein [Arabidopsis thaliana]	2e-034
TC125715	potassium transporter - like protein [Arabidopsis thaliana]	7e-063
TC125735	hypothetical protein [Arabidopsis thaliana]	3e-014
TC125760	putative protein kinase [Arabidopsis thaliana]	3e-023
TC125763	putative GPI-anchored protein [Arabidopsis thaliana]	8e-019
TC125777	putative protein destination factor [Arabidopsis thaliana]	2e-054
TC125881	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	2e-070
TC125964	No hits found	
TC126012	peroxidase, putative [Arabidopsis thaliana]	3e-025
TC126044	carbonic anhydrase -like protein [Arabidopsis thaliana]	1e-068
TC126079	actin	e-124
TC126101	hypothetical protein [Arabidopsis thaliana]	3e-016
TC126126	senescence-associated protein -related [Arabidopsis thaliana]	2e-014
TC126161	No hits found	
TC126222	probable transcription factor SF3 - common tobacco	4e-060
TC126262	bZIP transcriptional activator	4e-24
TC126268	receptor-like protein kinase 3 [Lycopersicon esculentum]	5e-046
TC126283	unknown protein [Arabidopsis thaliana]	1e-050

TC126311	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-110
TC126313	SINA1 protein [imported] - Vitis vinifera	e-116
TC126324	Probable pectate lyase P59 precursor	5e-025
TC126373	pectin methylesterase-like protein [Arabidopsis thaliana]	3e-063
TC126374	No hits found	
TC126394	putative glycosyl transferase [Arabidopsis thaliana]	5e-053
TC126449	Expressed protein [Arabidopsis thaliana]	7e-059
TC126497	histidine-rich glycoprotein precursor	6e-018
TC126568	purine permease - related [Arabidopsis thaliana]	3e-021
TC126596	putative protein [Arabidopsis thaliana]	8e-046
TC126645	GTP-binding protein - garden pea	4e-084
TC126689	heat shock protein DNAJ homolog Pfj4 [Plasmodium falciparum 3D7]	3e-012
TC126759	cyclic nucleotide-regulated ion channel, putative [Arabidopsis thaliana]	1e-053
TC126763	putative receptor - like kinase [Arabidopsis thaliana]	1e-031
TC126788	putative protein [Arabidopsis thaliana]	1e-022
TC126833	DNAJ-like heatshock protein [Arabidopsis thaliana]	e-112
TC126916	PGPS/D7 [Petunia x hybrida]	5e-019
TC126959	Serine/threonine Kinase [Persea americana]	2e-077
TC127002	β-glucanase [Sorghum bicolor]	3e-018
TC127056	No hits found	
TC127117	No hits found	
TC127131	actin depolymerizing factor-like [Arabidopsis thaliana]	3e-050
TC127293	putative protein [Arabidopsis thaliana]	7e-044
TC127328	No hits found	
TC127363	fasciclin-like arabinogalactan-protein [Arabidopsis thaliana]	7e-028
TC127401	putative sterol dehydrogenase [Arabidopsis thaliana]	1e-044
TC127423	phosphoinositide-specific phospholipase C (EC 3.1.4.-) plc3 - potato	e-112
TC127463	putative fatty acid hydroxylase [Oryza sativa (japonica cultivar-group)]	2e-058
TC127523	No hits found	
TC127524	No hits found	
TC127550	expressed protein [Arabidopsis thaliana]	3e-020
TC127574	mitochondrial carrier protein family [Arabidopsis thaliana]	1e-028
TC127642	No hits found	
TC127692	Na <sup>+</sup> /H <sup>+</sup> antiporter-like protein [Arabidopsis thaliana]	e-126



**Table VI.** Tomato ripening-induced genes

<b>TC</b>	<b>BB<sup>a</sup></b>	<b>MG<sup>b</sup></b>	<b>P value</b>	<b>Putative function</b>	<b>E value</b>	<b>Functional category</b>
TC124731	6	0	0.042864	AtRer1A [Arabidopsis thaliana]	2e-024	cell structure/maintenance
TC118249	9	0	0.017436	C2 domain-containing protein [Arabidopsis thaliana]	2e-056	cell structure/maintenance
TC118143	7	0	0.03176	Centromere/kinetochore protein zw10 homolog	e-105	cell structure/maintenance
TC116127	20	1	0.003505	CIG1 [Nicotiana tabacum]	0.0	cell structure/maintenance
TC118221	6	0	0.042864	cullin-like protein1 [Pisum sativum]	0.0	cell structure/maintenance
TC117977	8	0	0.023532	fimbrin [Arabidopsis thaliana]	3e-087	cell structure/maintenance
TC116741	11	0	0.009572	GT-related trihelix DNA-binding protein [Arabidopsis thaliana]	e-104	cell structure/maintenance
TC116672	7	0	0.03176	histone H2B [Cicer arietinum]	1e-041	cell structure/maintenance
TC115798	82	23	0.015426	induced stolon tip protein	2e-092	cell structure/maintenance
TC125123	6	0	0.042864	microtubule-associated protein [Arabidopsis thaliana]	e-118	cell structure/maintenance
TC116029	27	3	0.005575	poly(A)-binding protein [Nicotiana tabacum]	0.0	cell structure/maintenance
TC123940	7	0	0.03176	Putative beta tubulin [Oryza sativa]	0.0	cell structure/maintenance
TC124135	33	6	0.012888	YABBY2 [Arabidopsis thaliana]	1e-054	cell structure/maintenance
TC123890	203	4	3.19E-22	acid beta-fructofuranosidase precursor	0.0	cell wall
TC123825	59	1	8.36E-08	acid beta-fructofuranosidase precursor	0.0	cell wall
TC123821	209	1	8.58E-27	beta-fructosidase [Lycopersicon esculentum]	0.0	cell wall
TC126105	7	0	0.03176	beta-galactosidase [Arabidopsis thaliana]	e-127	cell wall
TC124240	10	0	0.012919	endo-beta-mannanase [Lycopersicon esculentum]	1e-059	cell wall
TC117251	6	0	0.042864	glycosyl hydrolase family 5/cellulase [Arabidopsis thaliana]	2e-072	cell wall
TC124239	42	1	9.8E-06	inactive endo-beta-mannanase [Lycopersicon esculentum]	0.0	cell wall
TC123895	10	0	0.012919	minor allergen beta-fructofuranosidase precursor [Lycopersicon esculentum]	2e-095	cell wall
TC115905	131	1	7.74E-17	pectate lyase [Arabidopsis thaliana]	0.0	cell wall
TC116135	60	22	0.025193	Pectinesterase 1 precursor	0.0	cell wall
TC124082	55	0	1.78E-08	Polygalacturonase 2A precursor	0.0	cell wall
TC116030	18	4	0.038664	xyloglucan endo-1,4-beta-D-glucanase	e-118	cell wall
TC116318	24	0	0.000194	17.7 kD class I small heat shock protein [Lycopersicon esculentum]	2e-064	defense/stress response

TC124878	10	0	0.012919	Avr9 elicitor response protein [Nicotiana tabacum]	e-158	defense/stress response
TC124566	9	1	0.045169	ceo protein [Arabidopsis thaliana]	e-103	defense/stress response
TC124090	20	6	0.044836	chaperonin 21 precursor [Lycopersicon esculentum]	e-136	defense/stress response
TC115816	11	0	0.009572	chitinase [Euonymus europaeus]	1e-016	defense/stress response
TC124179	7	0	0.03176	Chorismate synthase 1	0.0	defense/stress response
TC124101	23	7	0.041778	cold-induced glucosyl transferase [Solanum soganandinum]	0.0	defense/stress response
TC116013	40	7	0.00789	dehydrin homolog CI7 - potato	9e-039	defense/stress response
TC124917	10	0	0.012919	florfenicol resistance protein [Clostridium tetani E88]	9e-052	defense/stress response
TC126297	7	0	0.03176	heat shock cognate 70 kd protein	0.0	defense/stress response
TC123868	6	0	0.042864	heat shock cognate protein 80	0.0	defense/stress response
TC116962	8	0	0.023532	heat shock protein 17.6 - tomato	6e-068	defense/stress response
TC126413	7	0	0.03176	heat shock protein hsp70t-2 [Arabidopsis thaliana]	e-157	defense/stress response
TC124001	20	0	0.000644	heat shock protein MTSHP precursor - tomato	2e-078	defense/stress response
TC124903	13	0	0.005255	heat shock protein, 70K, chloroplast - cucumber	0.0	defense/stress response
TC116593	26	2	0.002704	HSP100/ClpB [Arabidopsis thaliana]	0.0	defense/stress response
TC116319	18	3	0.027136	Hsp20.0 protein [Lycopersicon peruvianum]	4e-065	defense/stress response
TC115998	28	3	0.004573	Pathogenesis-related protein PR P23	e-154	defense/stress response
TC116910	12	2	0.043315	putative Hsp70 binding protein [Oryza sativa]	e-140	defense/stress response
TC116269	12	0	0.007092	stress inducible protein (sti) [Arabidopsis thaliana]	0.0	defense/stress response
TC126972	6	0	0.042864	susceptibility antioxidant protein [Oryza sativa]	1e-022	defense/stress response
TC115989	14	3	0.046033	systemic acquired resistance-related protein SRE1b - potato	2e-021	defense/stress response
TC116692	7	0	0.03176	alternative oxidase 1a [Lycopersicon esculentum]	0.0	energy
TC117099	10	0	0.012919	ATP synthase alpha chain, mitochondrial	0.0	energy
TC117268	7	0	0.03176	ATP synthase delta chain	1e-018	energy
TC116313	18	6	0.047746	bifunctional lysine-ketoglutarate reductase/saccharopine dehydrogenase [Gossypium hirsutum]	0.0	energy
TC124071	11	1	0.029757	chlorophyll a/b-binding protein (cab-11) - tomato	e-145	energy
TC123987	8	0	0.023532	Cytochrome P450	e-140	energy
TC126100	7	0	0.03176	cytochrome P450 [Arabidopsis thaliana]	e-119	energy

TC116032	14	0	0.003894	cytochrome P450 family [Arabidopsis thaliana]	0.0	energy
TC116883	15	2	0.026332	cytosolic ascorbate peroxidase [Fragaria x ananassa]	e-110	energy
TC123931	46	9	0.008858	Enolase	0.0	energy
TC117210	6	0	0.042864	Enolase	0.0	energy
TC124099	14	2	0.031357	Ferredoxin--NADP reductase	0.0	energy
TC117431	11	0	0.009572	ferric reductase-like transmembrane component family [Arabidopsis thaliana]	e-130	energy
TC115872	28	4	0.009478	fructose-bisphosphate aldolase [Arabidopsis thaliana]	0.0	energy
TC116094	17	2	0.018176	fructose-bisphosphate aldolase [Persea americana]	e-173	energy
TC115908	9	1	0.045169	glyceraldehyde 3-phosphate dehydrogenase [Solanum tuberosum]	e-170	energy
TC123860	18	5	0.046075	Glyceraldehyde 3-phosphate dehydrogenase A	0.0	energy
TC124020	53	0	3.25E-08	lipoxygenase - tomato	0.0	energy
TC116986	6	0	0.042864	lipoxygenase [Lycopersicon esculentum]	0.0	energy
TC124825	8	0	0.023532	lipoxygenase loxC - tomato	0.0	energy
TC116115	30	8	0.031873	malate dehydrogenase precursor - tomato	0.0	energy
TC116554	21	1	0.00272	monooxygenase [Solanum tuberosum]	0.0	energy
TC116633	9	1	0.045169	NADH-ubiquinone oxireductase -related [Arabidopsis thaliana]	e-154	energy
TC116702	12	1	0.023885	NADPH-cytochrome P450 oxidoreductase - common tobacco	0.0	energy
TC124140	6	0	0.042864	oxidoreductase, 2OG-Fe(II) oxygenase family [Arabidopsis thaliana]	3e-055	energy
TC116738	12	2	0.043315	oxidoreductase, 2OG-Fe(II) oxygenase family [Arabidopsis thaliana]	e-103	energy
TC115873	26	3	0.006772	oxidoreductase, zinc-binding dehydrogenase family [Arabidopsis thaliana]	e-133	energy
TC115875	12	1	0.023885	oxidoreductase, zinc-binding dehydrogenase family [Arabidopsis thaliana]	e-145	energy
TC115874	6	0	0.042864	oxidoreductase, zinc-binding dehydrogenase family [Arabidopsis thaliana]	e-137	energy
TC116538	15	1	0.011958	pfkB type carbohydrate kinase protein family [Arabidopsis thaliana]	4e-044	energy
TC116182	35	0	7.17E-06	Photosystem II 22 kDa protein	2e-083	energy
TC116358	21	3	0.016797	Probable vacuolar ATP synthase subunit d 2	0.0	energy
TC116071	42	2	5.58E-05	pulmonary flavin-containing monooxygenase 2 [Arabidopsis thaliana]	2e-099	energy
TC118103	6	0	0.042864	pyrophosphate-dependent phosphofructo-1-kinase [Arabidopsis thaliana]	e-155	energy
TC116324	21	0	0.000477	pyruvate decarboxylase [Solanum tuberosum]	0.0	energy
TC123850	6	0	0.042864	pyruvate kinase-like [Deschampsia antarctica]	e-150	energy

TC123859	27	8	0.037711	RuBisCO subunit binding-protein alpha subunit	0.0	energy
TC116205	10	0	0.012919	Triosephosphate isomerase, cytosolic (TIM)	e-116	energy
TC116296	7	0	0.03176	ACC oxidase [Solanum tuberosum]	e-162	ethylene response
TC123932	104	9	2.33E-07	ACC oxidase 1	e-169	ethylene response
TC123934	10	0	0.012919	ACC oxidase 1	e-117	ethylene response
TC115848	149	11	1.12E-10	ACC oxidase homolog (protein E8)	0.0	ethylene response
TC115852	17	0	0.001584	ACC oxidase homolog (protein E8)	e-175	ethylene response
TC115849	37	7	0.011819	ACC oxidase homolog (protein E8)	6e-098	ethylene response
TC115853	31	1	0.000197	ACC oxidase homolog (protein E8)	e-100	ethylene response
TC116548	16	1	0.009414	AP2/ERF-domain protein [Solanum tuberosum]	e-116	ethylene response
TC124584	15	1	0.011958	ethylene receptor - tomato (strain UC82-B)	0.0	ethylene response
TC116831	7	0	0.03176	ethylene response factor 3 [Lycopersicon esculentum]	4e-097	ethylene response
TC116574	20	2	0.009987	ethylene-responsive protein [Arabidopsis thaliana]	5e-047	ethylene response
TC116326	32	0	1.76E-05	ripening protein E8 homolog - tomato	0.0	ethylene response
TC116368	17	0	0.001584	ripening regulated protein DDTFR10/A [Lycopersicon esculentum]	e-107	ethylene response
TC116700	15	0	0.002885	similar to ACC deaminase [Arabidopsis thaliana]	e-114	ethylene response
TC116203	9	1	0.045169	auxin-regulated protein [Arabidopsis thaliana]	3e-026	hormone response
TC124221	6	0	0.042864	spermidine synthase [Arabidopsis thaliana]	e-155	metabolism
TC124472	6	0	0.042864	2-oxoglutarate dehydrogenase E2 subunit [Arabidopsis thaliana]	e-175	metabolism
TC126030	6	0	0.042864	3(2),5-bisphosphate nucleotidase [Arabidopsis thaliana]	3e-070	metabolism
TC126635	7	0	0.03176	3,4-dihydroxy-5-hexaprenylbenzoate methyltransferase - Arabidopsis thaliana	8e-096	metabolism
TC117340	6	0	0.042864	3-desoxy-D-manno octulosonic acid-8-phosphate synthase [Lycopersicon esculentum]	e-158	metabolism
TC124154	12	0	0.007092	4-hydroxyphenylpyruvate dioxygenase	e-180	metabolism
TC115961	71	5	3.17E-06	acetyl-CoA C-acyltransferase[Arabidopsis thaliana]	0.0	metabolism
TC124728	7	0	0.03176	acyl carrier protein [Capsicum chinense]	1e-052	metabolism
TC115942	152	0	4.17E-21	acyltransferase 2 [Capsicum chinense]	0.0	metabolism
TC115943	14	0	0.003894	acyltransferase 2 [Capsicum chinense]	2e-063	metabolism
TC116044	24	1	0.001257	Adenosylhomocysteinase	0.0	metabolism
TC123883	25	2	0.003389	Alcohol dehydrogenase 2	0.0	metabolism

TC124274	9	0	0.017436	alcohol dehydrogenase class III [Arabidopsis thaliana]	0.0	metabolism
TC117676	6	0	0.042864	amidase-like protein [Nicotiana tabacum]	0.0	metabolism
TC116466	15	0	0.002885	amine oxidase family [Arabidopsis thaliana]	0.0	metabolism
TC124056	9	0	0.017436	arginine decarboxylase 1 [Datura stramonium]	0.0	metabolism
TC116807	7	0	0.03176	arginine/serine-rich splicing factor (atSRp34) [Arabidopsis thaliana]	7e-074	metabolism
TC117269	6	0	0.042864	beta-alanine synthase [Lycopersicon esculentum]	e-175	metabolism
TC116638	17	0	0.001584	beta-cyanoalanine synthase [Solanum tuberosum]	e-174	metabolism
TC116652	17	0	0.001584	branched-chain amino acid aminotransferase [Capsicum annuum]	e-147	metabolism
TC123953	16	4	0.046647	Carbonic anhydrase	4e-098	metabolism
TC116434	31	1	0.000197	carboxyphosphoenolpyruvate mutase [Arabidopsis thaliana]	e-156	metabolism
TC124283	23	0	0.000262	CER1 protein [Arabidopsis thaliana]	0.0	metabolism
TC125310	10	1	0.036814	cytochrome P450-dependent fatty acid hydroxylase [Nicotiana tabacum]	e-179	metabolism
TC117055	7	0	0.03176	DEAD box protein (RNA helicase) [Homo sapiens]	e-142	metabolism
TC116440	12	0	0.007092	DEAD/DEAH box helicase [Arabidopsis thaliana]	e-163	metabolism
TC124535	8	0	0.023532	GCN5-related N-acetyltransferase (GNAT) family [Arabidopsis thaliana]	3e-068	metabolism
TC116558	6	0	0.042864	GcpE (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase) [Lycopersicon esculentum]	0.0	metabolism
TC116406	10	1	0.036814	Glucose-6-phosphate 1-dehydrogenase	0.0	metabolism
TC125793	6	0	0.042864	glucose-6-phosphate 1-dehydrogenase [Solanum tuberosum]	0.0	metabolism
TC117044	14	0	0.003894	glucuronosyl transferase homolog, ripening-related - tomato	0.0	metabolism
TC116414	6	0	0.042864	Glutamate decarboxylase	8e-057	metabolism
TC116293	21	2	0.008105	glutathione S-transferase [Capsicum annuum]	e-112	metabolism
TC116034	15	2	0.026332	glutathione transferase - potato	e-116	metabolism
TC123915	100	0	2.46E-14	Histidine decarboxylase	0.0	metabolism
TC123916	58	0	7.26E-09	Histidine decarboxylase	e-146	metabolism
TC123922	25	0	0.000144	Histidine decarboxylase	e-118	metabolism
TC123921	8	0	0.023532	Histidine decarboxylase	1e-056	metabolism
TC124859	14	0	0.003894	lipase (class 3) family [Arabidopsis thaliana]	e-174	metabolism
TC125633	7	0	0.03176	lipase [Arabidopsis thaliana]	7e-046	metabolism

TC126615	6	0	0.042864	lipase [Arabidopsis thaliana]	8e-077	metabolism
TC116755	10	0	0.012919	long-chain acyl-CoA synthetase [Arabidopsis thaliana]	0.0	metabolism
TC117352	7	0	0.03176	MutT/nudix family protein (hydrolase) [Arabidopsis thaliana]	2e-082	metabolism
TC116497	9	1	0.045169	NAD-dependent epimerase/dehydratase family [Arabidopsis thaliana]	e-177	metabolism
TC125232	6	0	0.042864	nucleotide sugar epimerase family [Arabidopsis thaliana]	e-164	metabolism
TC116516	13	0	0.005255	peroxiredoxin - like protein [Arabidopsis thaliana]	6e-066	metabolism
TC124567	14	0	0.003894	Probable adenylate kinase 2	6e-080	metabolism
TC116222	21	1	0.00272	probable short chain alcohol dehydrogenase - common tobacco	2e-091	metabolism
TC124491	10	0	0.012919	putative 6-phosphogluconolactonase [Elaeis guineensis]	e-101	metabolism
TC124305	19	2	0.012254	putative acyl-CoA synthetase [Capsicum annuum]	0.0	metabolism
TC116199	38	1	2.95E-05	putative dehydrogenase [Arabidopsis thaliana]	7e-088	metabolism
TC125703	9	0	0.017436	putative DNA cytosine methyltransferase Zmet3 [Zea mays]	e-118	metabolism
TC125884	8	0	0.023532	putative esterase [Arabidopsis thaliana]	1e-063	metabolism
TC126727	6	0	0.042864	putative RNA helicase [Arabidopsis thaliana]	6e-019	metabolism
TC124282	41	0	1.19E-06	putative short-chain type alcohol dehydrogenase [Solanum tuberosum]	e-117	metabolism
TC116522	25	4	0.01539	Putative Transcription initiation factor IIE, beta subunit [Oryza sativa]	6e-093	metabolism
TC125841	6	0	0.042864	RNA helicase -related [Arabidopsis thaliana]	2e-085	metabolism
TC116947	7	0	0.03176	RNase NGR2 [Nicotiana glutinosa]	4e-091	metabolism
TC123879	23	4	0.020715	S-adenosylmethionine decarboxylase proenzyme	0.0	metabolism
TC116875	6	0	0.042864	serine acetyltransferase 4 [Nicotiana tabacum]	e-158	metabolism
TC125273	7	0	0.03176	serine hydroxymethyltransferase	0.0	metabolism
TC117579	6	0	0.042864	Shikimate kinase	e-120	metabolism
TC116890	9	0	0.017436	short-chain dehydrogenase/reductase family protein [Arabidopsis thaliana]	6e-073	metabolism
TC117842	7	0	0.03176	short-chain dehydrogenase/reductase family protein [Arabidopsis thaliana]	8e-087	metabolism
TC124537	23	0	0.000262	similar to limonene cyclase [Arabidopsis thaliana]	e-123	metabolism
TC125903	7	0	0.03176	Soluble glycogen [starch] synthase	0.0	metabolism
TC125629	11	0	0.009572	steroleosin (Sterol-binding dehydrogenase/reductase) [Sesamum indicum]	e-108	metabolism
TC124762	6	0	0.042864	thioesterase family [Arabidopsis thaliana]	e-173	metabolism
TC115829	6	0	0.042864	thioredoxin peroxidase [Nicotiana tabacum]	e-128	metabolism

TC119047	6	0	0.042864	trehalose phosphatase family [Arabidopsis thaliana]	1e-093	metabolism
TC124977	7	0	0.03176	UDP-glycosyltransferase family [Arabidopsis thaliana]	e-158	metabolism
TC116630	19	2	0.012254	xanthine dehydrogenase [Arabidopsis thaliana]	0.0	metabolism
TC119001	6	0	0.042864	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	8e-074	metabolism, carotenoid
TC124119	22	2	0.006551	cinnamic acid 4-hydroxylase [Capsicum annuum]	0.0	metabolism, carotenoid
TC124613	6	0	0.042864	farnesyl-pyrophosphate synthetase FPS1 - tomato	0.0	metabolism, carotenoid
TC116718	12	0	0.007092	Flavonol synthase	e-171	metabolism, carotenoid
TC116882	13	1	0.019059	geranyl diphosphate synthase [Arabidopsis thaliana]	e-148	metabolism, carotenoid
TC117185	10	0	0.012919	isoflavone reductase homolog	e-144	metabolism, carotenoid
TC115968	11	1	0.029757	mutant phytoene synthase [Lycopersicon esculentum]	2e-042	metabolism, carotenoid
TC115971	7	0	0.03176	mutant phytoene synthase [Lycopersicon esculentum]	1e-033	metabolism, carotenoid
TC125667	11	0	0.009572	orcinol O-methyltransferase [Rosa hybrid cultivar]	5e-091	metabolism, carotenoid
TC116252	34	0	9.68E-06	tomato fruit ripening specific URF, phytoene synthase [Lycopersicon esculentum]	0.0	Metabolism, carotenoid
TC115970	55	1	2.59E-07	Phytoene synthase 1	0.0	metabolism, carotenoid
TC123990	8	0	0.023532	60S ribosomal protein L10A [Arabidopsis thaliana]	2e-095	protein biosynthesis/degradation
TC123924	6	0	0.042864	60S ribosomal protein L13 [Arabidopsis thaliana]	4e-085	protein biosynthesis/degradation
TC125192	8	0	0.023532	asparaginyl-tRNA synthetase [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC125544	9	0	0.017436	ATP-dependent clp protease	0.0	protein biosynthesis/degradation
TC124060	16	1	0.009414	cathepsin B-like cysteine proteinase [Nicotiana rustica]	e-174	protein biosynthesis/degradation
TC116078	23	3	0.011846	elongation factor [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC123771	42	12	0.027567	elongation factor 1-alpha	0.0	protein biosynthesis/degradation
TC116059	19	0	0.000869	Peptide methionine sulfoxide reductase (Fruit-ripening protein E4)	e-114	protein biosynthesis/degradation
TC124171	32	3	0.002007	peptidylprolyl isomerase [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC116893	18	0	0.001173	pol polyprotein [Citrus x paradisi]	e-129	protein biosynthesis/degradation
TC125212	8	0	0.023532	presenilin-like protein 4 [Mus musculus]	3e-056	protein biosynthesis/degradation
TC124361	7	0	0.03176	probable protein disulfide-isomerase - common tobacco	0.0	protein biosynthesis/degradation
TC123951	13	0	0.005255	protease inhibitor [Arabidopsis thaliana]	2e-024	protein biosynthesis/degradation
TC116931	8	0	0.023532	Proteasome subunit alpha type 6	e-135	protein biosynthesis/degradation
TC124039	23	4	0.020715	Protein disulfide isomerase precursor	e-178	protein biosynthesis/degradation

TC124796	15	2	0.026332	putative prolyl 4-hydroxylase, alpha subunit [Arabidopsis thaliana]	4e-084	protein biosynthesis/degradation
TC124163	7	0	0.03176	Ribosomal protein L18a	1e-091	protein biosynthesis/degradation
TC116868	7	0	0.03176	ribosome-like protein [Cucumis sativus]	1e-034	protein biosynthesis/degradation
TC124255	25	1	0.000969	serine carboxypeptidase [Arabidopsis thaliana]	e-133	protein biosynthesis/degradation
TC125764	6	0	0.042864	Similar to Secale chloroplast ribosomal protein L12 [Arabidopsis thaliana]	2e-017	protein biosynthesis/degradation
TC116315	37	3	0.000677	thiol protease [Matricaria chamomilla]	e-120	protein biosynthesis/degradation
TC117405	9	0	0.017436	translation initiation factor IF1 [Lycopersicon esculentum]	3e-051	protein biosynthesis/degradation
TC117001	8	0	0.023532	tripeptidyl peptidase II [Homo sapiens]	9e-075	protein biosynthesis/degradation
TC115895	27	8	0.037711	ubiquitin [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC123950	20	6	0.044836	ubiquitin-conjugating enzyme UBC2 [Mesembryanthemum crystallinum]	7e-083	protein biosynthesis/degradation
TC116777	10	0	0.012919	ubiquitin-specific protease 6 (UBP6), putative [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC116260	12	2	0.043315	14-3-3 protein 5	e-137	signal transduction
TC124245	13	1	0.019059	14-3-3 protein 7	e-135	signal transduction
TC125042	8	0	0.023532	ankyrin-kinase [Medicago truncatula]	0.0	signal transduction
TC123941	21	7	0.04426	ankyrin-repeat protein HBPI [Nicotiana tabacum]	e-162	signal transduction
TC118596	7	0	0.03176	ARF GTPase-activating domain-containing protein [Arabidopsis thaliana]	4e-025	signal transduction
TC125239	6	0	0.042864	calcineurin B-like protein 1 (CBL1) [Arabidopsis thaliana]	6e-099	signal transduction
TC117720	7	0	0.03176	calcium-binding EF-hand family protein [Arabidopsis thaliana]	e-136	signal transduction
TC115834	7	0	0.03176	calcyclin binding protein-like [Oryza sativa]	9e-069	signal transduction
TC123982	6	0	0.042864	Calmodulin	8e-080	signal transduction
TC125519	7	0	0.03176	calmodulin-binding protein [Arabidopsis thaliana]	0.0	signal transduction
TC124839	7	0	0.03176	CBL-interacting protein kinase 14 [Arabidopsis thaliana]	e-141	signal transduction
TC124569	6	0	0.042864	GTP-binding nuclear protein RAN2	e-116	signal transduction
TC115894	33	2	0.000522	guanylate kinase [Nicotiana tabacum]	e-150	signal transduction
TC115887	6	0	0.042864	guanylate kinase [Nicotiana tabacum]	1e-038	signal transduction
TC117113	6	0	0.042864	phragmoplastin (GTPase) [Nicotiana tabacum]	0.0	signal transduction
TC116467	22	3	0.014143	pinhead-like protein [Arabidopsis thaliana]	e-138	signal transduction
TC115882	17	0	0.001584	probable calcium-binding protein - potato	0.0	signal transduction
TC115884	10	0	0.012919	probable calcium-binding protein - potato	1e-049	signal transduction



TC116229	13	1	0.019059	probable serine/threonine kinase SNFL2 - sorghum	0.0	signal transduction
TC125331	7	0	0.03176	protein kinase 5 [Arabidopsis thaliana]	e-136	signal transduction
TC118048	6	0	0.042864	protein kinase family [Arabidopsis thaliana]	e-103	signal transduction
TC125164	7	0	0.03176	putative tubby protein [Oryza sativa]	e-149	signal transduction
TC125502	7	0	0.03176	Rac-like GTP-binding protein [Arabidopsis thaliana]	e-102	signal transduction
TC116439	6	0	0.042864	shaggy protein kinase 4 - garden petunia	0.0	signal transduction
TC125349	6	0	0.042864	two-component phosphorelay mediator [Arabidopsis thaliana]	3e-038	signal transduction
TC124244	17	2	0.018176	bZIP DNA-binding protein [Lycopersicon esculentum]	7e-077	transcription factor
TC124112	8	0	0.023532	bZIP transcription factor ATB2 [Glycine max]	2e-034	transcription factor
TC118144	7	0	0.03176	CONSTANS B-box zinc finger family protein [Arabidopsis thaliana]	3e-045	transcription factor
TC117443	6	0	0.042864	FLN29 gene product (zinc finger) [Mus musculus]	1e-023	transcription factor
TC124192	27	3	0.005575	homeotic protein VAHOX1 - tomato	e-170	transcription factor
TC125359	9	0	0.017436	MADS box transcription factor MADS1 [Capsicum annuum]	8e-097	transcription factor
TC118434	6	0	0.042864	scarecrow transcription factor family [Arabidopsis thaliana]	2e-044	transcription factor
TC124194	19	0	0.000869	PHAP2A protein [Petunia x hybrida]	e-116	transcription factor
TC125037	7	0	0.03176	putative heat shock transcription factor [Oryza sativa]	1e-010	transcription factor
TC116904	6	0	0.042864	similar to CHP-rich zinc finger protein [Arabidopsis thaliana]	1e-066	transcription factor
TC125034	7	0	0.03176	similar to zinc finger - like protein [Arabidopsis thaliana]	2e-038	transcription factor
TC116186	8	0	0.023532	SPF1 protein - sweet potato	e-158	transcription factor
TC125032	8	0	0.023532	SQUAMOSA-promoter binding protein 1	2e-025	transcription factor
TC125305	9	0	0.017436	TDR4 transcription factor [Lycopersicon esculentum]	e-136	transcription factor
TC116929	10	0	0.012919	transcription factor Hap5a, putative [Arabidopsis thaliana]	1e-051	transcription factor
TC117192	6	0	0.042864	VIP2 protein [Arabidopsis thaliana]	7e-071	transcription factor
TC117853	7	0	0.03176	WRKY family transcription factor [Arabidopsis thaliana]	2e-082	transcription factor
TC124196	8	0	0.023532	putative zinc finger transcription factor ZFP33 [Oryza sativa]	7e-026	transcription factor
TC125379	9	0	0.017436	ABC transporter family protein [Arabidopsis thaliana]	3e-082	transport
TC116352	6	0	0.042864	ABC transporter-related [Arabidopsis thaliana]	0.0	transport
TC125341	8	0	0.023532	clathrin adaptor medium chain protein MU1B [Arabidopsis thaliana]	0.0	transport
TC117654	6	0	0.042864	coatmer beta subunit [Arabidopsis thaliana]	0.0	transport

TC124929	14	2	0.031357	copper-binding protein family [Arabidopsis thaliana]	2e-063	transport
TC126210	6	0	0.042864	metal transporter family [Arabidopsis thaliana]	1e-031	transport
TC124700	14	0	0.003894	NTGP4 (heavy metal ion transport) [Nicotiana tabacum]	e-158	transport
TC123738	6	0	0.042864	Preprotein translocase secY subunit	e-168	transport
TC124836	9	0	0.017436	putative ABC transporter ATPase [Arabidopsis thaliana]	e-115	transport
TC118806	6	0	0.042864	putative Na <sup>+</sup> -dependen inorganic phosphate cotransporter [Oryza sativa]	3e-056	transport
TC117647	6	0	0.042864	syntaxin SNAP33 [Arabidopsis thaliana]	4e-091	transport
TC124811	7	0	0.03176	translocon Tic40 [Pisum sativum]	6e-061	transport
TC116726	8	0	0.023532	Vacuolar ATP synthase 16 kDa proteolipid subunit	3e-058	transport
TC124706	7	0	0.03176	v-ATPase subunit D (vATPD) [Arabidopsis thaliana]	e-112	transport
TC116311	24	1	0.001257	PGPD14 (pollen germination related protein) [Oryza sativa]	e-122	unclassified
TC125220	6	0	0.042864	putative plastid protein [Oryza sativa]	9e-072	unclassified
TC124284	32	0	1.76E-05	aluminum-induced protein-like protein [Thellungiella halophila]	5e-058	unclassified
TC123760	9	1	0.045169	B2 protein [Daucus carota]	1e-078	unclassified
TC123970	11	0	0.009572	early nodulin ENOD18 [Vicia faba]	3e-034	unclassified
TC115890	18	0	0.001173	Fruit-specific protein	4e-049	unclassified
TC117511	8	0	0.023532	KH domain protein [Arabidopsis thaliana]	e-126	unclassified
TC116209	13	1	0.019059	leucine rich repeat protein [Arabidopsis thaliana]	e-111	unclassified
TC124203	11	1	0.029757	membrane protein, 37K, precursor - common tobacco	e-170	unclassified
TC125213	12	1	0.023885	PBng143 [Vigna radiata]	1e-070	unclassified
TC116636	13	1	0.019059	putative early light induced protein [Arachis hypogaea]	9e-046	unclassified
TC117433	13	0	0.005255	putative exosome component [Oryza sativa]	5e-074	unclassified
TC117486	9	0	0.017436	regulatory protein -related [Arabidopsis thaliana]	0.0	unclassified
TC116478	9	0	0.017436	ripening regulated protein DDTFR18 [Lycopersicon esculentum]	e-170	unclassified
TC116477	7	0	0.03176	ripening regulated protein DDTFR18 [Lycopersicon esculentum]	4e-056	unclassified
TC117472	9	0	0.017436	WD repeat protein; periodic tryptophan protein 1 homolog [Schizosaccharomyces pombe]	1e-065	unclassified
TC123761	45	1	4.26E-06	expressed protein [Arabidopsis thaliana]	6e-091	unknown
TC116373	28	2	0.001708	expressed protein [Arabidopsis thaliana]	4e-097	unknown
TC124373	21	1	0.00272	expressed protein [Arabidopsis thaliana]	1e-058	unknown

TC125390	13	0	0.005255	expressed protein [Arabidopsis thaliana]	4e-033	unknown
TC124873	11	0	0.009572	expressed protein [Arabidopsis thaliana]	4e-071	unknown
TC124948	11	0	0.009572	expressed protein [Arabidopsis thaliana]	3e-058	unknown
TC125000	11	0	0.009572	expressed protein [Arabidopsis thaliana]	4e-050	unknown
TC116975	10	0	0.012919	expressed protein [Arabidopsis thaliana]	e-127	unknown
TC124461	10	0	0.012919	expressed protein [Arabidopsis thaliana]	6e-041	unknown
TC125167	9	0	0.017436	expressed protein [Arabidopsis thaliana]	5e-017	unknown
TC125205	8	0	0.023532	expressed protein [Arabidopsis thaliana]	2e-028	unknown
TC116884	7	0	0.03176	expressed protein [Arabidopsis thaliana]	e-163	unknown
TC125462	7	0	0.03176	expressed protein [Arabidopsis thaliana]	e-110	unknown
TC115812	10	1	0.036814	expressed protein [Arabidopsis thaliana]	2e-053	unknown
TC116521	6	0	0.042864	expressed protein [Arabidopsis thaliana]	e-156	unknown
TC117500	6	0	0.042864	expressed protein [Arabidopsis thaliana]	6e-058	unknown
TC118046	6	0	0.042864	expressed protein [Arabidopsis thaliana]	1e-055	unknown
TC118500	6	0	0.042864	expressed protein [Arabidopsis thaliana]	1e-059	unknown
TC118508	6	0	0.042864	expressed protein [Arabidopsis thaliana]	e-129	unknown
TC119035	6	0	0.042864	expressed protein [Arabidopsis thaliana]	6e-023	unknown
TC119113	6	0	0.042864	expressed protein [Arabidopsis thaliana]	2e-046	unknown
TC124470	6	0	0.042864	expressed protein [Arabidopsis thaliana]	2e-090	unknown
TC124915	6	0	0.042864	expressed protein [Arabidopsis thaliana]	6e-083	unknown
TC124998	6	0	0.042864	expressed protein [Arabidopsis thaliana]	e-156	unknown
TC125227	6	0	0.042864	expressed protein [Arabidopsis thaliana]	4e-083	unknown
TC125400	6	0	0.042864	expressed protein [Arabidopsis thaliana]	1e-077	unknown
TC125423	6	0	0.042864	expressed protein [Arabidopsis thaliana]	2e-039	unknown
TC126467	6	0	0.042864	expressed protein [Arabidopsis thaliana]	e-120	unknown
TC125384	8	0	0.023532	hypothetical protein [Arabidopsis thaliana]	9e-077	unknown
TC124904	6	0	0.042864	hypothetical protein [Oryza sativa]	5e-091	unknown
TC118108	7	0	0.03176	hypothetical protein F12M12.190 - Arabidopsis thaliana	7e-094	unknown
TC116801	6	0	0.042864	hypothetical protein F8J2.120 - Arabidopsis thaliana	0.0	unknown

TC125570	11	0	0.009572	No hits found		unknown
TC124708	10	0	0.012919	No hits found		unknown
TC116911	8	0	0.023532	No hits found		unknown
TC118447	8	0	0.023532	No hits found		unknown
TC116190	7	0	0.03176	No hits found		unknown
TC117075	7	0	0.03176	No hits found		unknown
TC124661	7	0	0.03176	No hits found		unknown
TC126074	7	0	0.03176	No hits found		unknown
TC126287	7	0	0.03176	No hits found		unknown
TC126382	7	0	0.03176	No hits found		unknown
TC126678	7	0	0.03176	No hits found		unknown
TC116853	10	1	0.036814	No hits found		unknown
TC124758	10	1	0.036814	No hits found		unknown
TC119067	6	0	0.042864	No hits found		unknown
TC126451	6	0	0.042864	No hits found		unknown
TC126522	6	0	0.042864	No hits found		unknown
TC124704	11	0	0.009572	unknown [Arabidopsis thaliana]	8e-054	unknown
TC125113	10	1	0.036814	unknown protein [Arabidopsis thaliana]	e-176	unknown
TC116376	6	0	0.042864	unknown protein [Arabidopsis thaliana]	2e-048	unknown

<sup>a</sup> This column represents the number of ESTs from breaker fruit library with total of 15207 ESTs

<sup>b</sup> This column represents the number of ESTs from mature green fruit library with total of 5317 ESTs

**Table VII.** Tomato ripening-repressed genes

TC	MG <sup>a</sup>	BB <sup>b</sup>	P value	Putative function	E value	Functional category
TC116116	8	2	0.000371	actin [imported] - Malva pusilla	0.0	cell structure/maintenance
TC116117	6	1	0.001162	actin [imported] - Malva pusilla	0.0	cell structure/maintenance
TC115716	6	3	0.007653	alpha-tubulin [Nicotiana tabacum]	0.0	cell structure/maintenance
TC115782	6	4	0.014176	alpha-tubulin [Nicotiana tabacum]	0.0	cell structure/maintenance
TC116390	3	1	0.03818	cell division protein [Arabidopsis thaliana]	e-174	cell structure/maintenance
TC125090	4	0	0.003337	chloroplast nucleoid DNA-binding protein[Arabidopsis thaliana]	e-134	cell structure/maintenance
TC124326	5	3	0.019694	DNA-binding protein 4 - common tobacco	2e-058	cell structure/maintenance
TC121307	3	0	0.012882	dynammin related protein involved in chloroplast division [Cyanidioschyzon merolae]	3e-034	cell structure/maintenance
TC116713	3	0	0.012882	purine-rich single-stranded DNA-binding protein [Arabidopsis thaliana]	2e-061	cell structure/maintenance
TC126546	3	0	0.012882	putative Sec31p [Oryza sativa]	2e-059	cell structure/maintenance
TC118475	3	0	0.012882	PWWP domain protein [Arabidopsis thaliana]	4e-020	cell structure/maintenance
TC124273	4	3	0.047513	SKP1-like protein [Nicotiana clevelandii]	1e-069	cell structure/maintenance
TC125565	3	1	0.03818	Beta-galactosidase precursor	0.0	cell wall
TC116388	8	3	0.001009	polygalacturonase 1 beta chain precursor - tomato	0.0	cell wall
TC116031	6	2	0.003443	xyloglucan endo-1,4-beta-D-glucanase - Arabidopsis thaliana	4e-069	cell wall
TC124106	10	12	0.01786	abscisic stress ripening protein 1	4e-021	defense/stress response
TC117728	4	2	0.027482	CARD15-like protein [Homo sapiens]	2e-021	defense/stress response
TC115865	30	38	0.000377	Catalase isozyme 1	0.0	defense/stress response
TC116265	8	8	0.017573	dehydration-induced protein family [Arabidopsis thaliana]	0.0	defense/stress response
TC124673	3	1	0.03818	dehydration-induced protein RD22-like protein [Gossypium hirsutum]	2e-086	defense/stress response
TC124430	4	1	0.012364	disease resistance protein family (LRR) [Arabidopsis thaliana]	e-143	defense/stress response
TC123960	12	15	0.013107	dnaJ protein homolog - potato	0.0	defense/stress response
TC124842	3	0	0.012882	DS2 protein [Solanum tuberosum]	1e-064	defense/stress response
TC124714	5	1	0.003844	glutathione S-transferase domain protein [Brucella suis 1330]	3e-085	defense/stress response
TC123867	26	10	5.26E-09	heat shock cognate protein 80	0.0	defense/stress response

TC124487	7	5	0.010262	iron-stress related protein [Citrus junos]	5e-021	defense/stress response
TC124012	11	3	3.87E-05	TSI-1 protein - tomato	5e-066	defense/stress response
TC119023	4	0	0.003337	wound-induced protein 1	2e-031	defense/stress response
TC119597	3	1	0.03818	wound-induced protein Sn-1, vacuolar membrane - pepper	6e-046	defense/stress response
TC124444	3	0	0.012882	aldehyde dehydrogenase [Arabidopsis thaliana]	e-179	energy
TC116402	3	1	0.03818	ATP synthase gamma chain	e-154	energy
TC124182	8	9	0.024594	cytochrome c [Arabidopsis thaliana]	e-144	energy
TC118187	3	0	0.012882	flavin-containing monooxygenase family [Arabidopsis thaliana]	1e-063	energy
TC123985	10	14	0.029737	Formate dehydrogenase	0.0	energy
TC116377	5	2	0.009968	fructokinase [Lycopersicon esculentum]	0.0	energy
TC116003	55	23	1.41E-16	glycolate oxidase [Arabidopsis thaliana]	e-175	energy
TC116076	4	2	0.027482	glycolate oxidase [Arabidopsis thaliana]	0.0	energy
TC124943	5	4	0.032832	Lipoxygenase A	0.0	energy
TC124391	5	3	0.019694	NADPH-ferrihemoprotein reductase - mung bean	0.0	energy
TC124810	15	1	1.4E-08	phosphoenolpyruvate carboxylase 2 [Lycopersicon esculentum]	0.0	energy
TC115712	7	1	0.000344	plastidic aldolase [Nicotiana paniculata]	0.0	energy
TC123871	22	20	0.000118	plastidic aldolase NPALDP1 [Nicotiana paniculata]	0.0	energy
TC116725	3	1	0.03818	Rubisco subunit binding-protein beta subunit - Arabidopsis thaliana	3e-076	energy
TC124785	9	4	0.000839	uroporphyrinogen III synthase [Arabidopsis thaliana]	1e-080	energy
TC116320	4	0	0.003337	dehydration-responsive element binding protein 3 [Lycopersicon esculentum]	6e-046	ethylene response
TC116549	5	2	0.009968	gda-1, putative [Arabidopsis thaliana]	e-113	hormone response
TC116615	6	1	0.001162	IAA-Ala hydrolase (IAR3) [Arabidopsis thaliana]	e-170	hormone response
TC119522	3	0	0.012882	IAA-Ala hydrolase (IAR3) [Arabidopsis thaliana]	e-113	hormone response
TC117228	4	1	0.012364	jasmonic acid 2 [Lycopersicon esculentum]	0.0	hormone response
TC116481	3	0	0.012882	putative auxin growth promotor protein [Lycopersicon esculentum]	e-113	hormone response
TC116559	5	3	0.019694	putative auxin-repressed protein [Prunus armeniaca]	6e-027	hormone response
TC115787	40	49	3.53E-05	2-oxoglutarate-dependent dioxygenase [Arabidopsis thaliana]	3e-073	metabolism
TC116382	5	2	0.009968	Acetolactate synthase I	0.0	metabolism

TC117284	3	0	0.012882	adenylate kinase b	e-121	metabolism
TC125279	4	0	0.003337	arabinosidase ARA-1 [ <i>Lycopersicon esculentum</i> ]	0.0	metabolism
TC116397	6	3	0.007653	beta-amylase [ <i>Prunus armeniaca</i> ]	e-178	metabolism
TC117741	5	1	0.003844	copper amine oxidase-like protein - <i>Arabidopsis thaliana</i>	0.0	metabolism
TC117186	4	0	0.003337	dihydrolipoamide S-acetyltransferase precursor - <i>Arabidopsis thaliana</i>	e-122	metabolism
TC116394	3	0	0.012882	DNA repair protein RAD23 [ <i>Arabidopsis thaliana</i> ]	1e-090	metabolism
TC116752	4	1	0.012364	esterase D [ <i>Arabidopsis thaliana</i> ]	e-129	metabolism
TC124655	3	0	0.012882	Glycine dehydrogenase	0.0	metabolism
TC124323	5	2	0.009968	histidine decarboxylase [ <i>Arabidopsis thaliana</i> ]	0.0	metabolism
TC116745	3	1	0.03818	hydrolase, alpha/beta fold family [ <i>Arabidopsis thaliana</i> ]	e-118	metabolism
TC116475	3	0	0.012882	Hydroquinone glucosyltransferase (Arbutin synthase)	0.0	metabolism
TC116869	7	3	0.002832	Inositol-3-phosphate synthase	0.0	metabolism
TC117406	4	0	0.003337	invertase inhibitor precursor - tomato	3e-083	metabolism
TC124188	9	12	0.031336	LYTB-like protein precursor [ <i>Adonis palaestina</i> ]	0.0	metabolism
TC125067	3	1	0.03818	MutT/nudix family protein [ <i>Arabidopsis thaliana</i> ]	0.0	metabolism
TC124178	3	0	0.012882	neutral invertase [ <i>Arabidopsis thaliana</i> ]	e-180	metabolism
TC126057	3	1	0.03818	phosphoethanolamine cytidyltransferase [ <i>Hordeum vulgare</i> ]	e-109	metabolism
TC125496	4	3	0.047513	putative heme A farnesyltransferase homolog [ <i>Oryza sativa</i> ]	5e-090	metabolism
TC121503	3	0	0.012882	putative non-LTR retroelement reverse transcriptase [ <i>Oryza sativa</i> ]	6e-014	metabolism
TC117132	3	1	0.03818	putative nucleoside hydrolase protein [ <i>sinorhizobium meliloti</i> ]	5e-055	metabolism
TC126463	3	0	0.012882	RNA helicase-like protein DB10	e-172	metabolism
TC123896	41	13	1.48E-14	S-adenosylmethionine synthetase 1	0.0	metabolism
TC116337	5	5	0.048653	Spermidine synthase	0.0	metabolism
TC116527	3	0	0.012882	serine hydroxymethyltransferase	0.0	metabolism
TC116583	4	3	0.047513	topoisomerase I [ <i>Nicotiana tabacum</i> ]	0.0	metabolism
TC124513	7	5	0.010262	UDP-glucose 4-epimerase [ <i>Arabidopsis thaliana</i> ]	e-169	metabolism
TC124743	3	0	0.012882	26S proteasome regulatory subunit (RPN6) [ <i>Arabidopsis thaliana</i> ]	0.0	protein biosynthesis/degradation
TC124242	3	1	0.03818	40S ribosomal protein S3A	e-120	protein biosynthesis/degradation

TC116128	4	1	0.012364	40S ribosomal protein S5	5e-091	protein biosynthesis/degradation
TC124317	4	0	0.003337	40S ribosomal protein S8	1e-075	protein biosynthesis/degradation
TC124213	5	3	0.019694	40S ribosomal protein S9 [Arabidopsis thaliana]	1e-091	protein biosynthesis/degradation
TC115858	3	0	0.012882	60S ribosomal protein L21 [Oryza sativa]	2e-078	protein biosynthesis/degradation
TC123995	3	1	0.03818	60S ribosomal protein L5	e-114	protein biosynthesis/degradation
TC124145	13	11	0.001618	ATP-dependent clp protease	0.0	protein biosynthesis/degradation
TC123779	3	0	0.012882	elongation factor 1-ALPHA	5e-068	protein biosynthesis/degradation
TC125228	3	0	0.012882	nhp2-like protein - Arabidopsis thaliana	2e-060	protein biosynthesis/degradation
TC116068	4	3	0.047513	polyubiquitin - tomato	0.0	protein biosynthesis/degradation
TC116189	10	7	0.002405	probable 60S acidic ribosomal protein PO - Arabidopsis thaliana	1e-080	protein biosynthesis/degradation
TC124311	3	1	0.03818	protein disulfide isomerase-related protein [Arabidopsis thaliana]	7e-073	protein biosynthesis/degradation
TC124354	3	0	0.012882	ribosomal protein L1 protein [imported] - Arabidopsis thaliana	e-113	protein biosynthesis/degradation
TC116081	8	6	0.007469	RUB1 conjugating enzyme [Olea europaea]	3e-097	protein biosynthesis/degradation
TC125214	6	0	0.000224	serine protease-like protein [Nicotiana tabacum]	0.0	protein biosynthesis/degradation
TC123998	37	31	2.95E-07	ubiquitin conjugating protein [Avicennia marina]	9e-086	protein biosynthesis/degradation
TC124000	3	1	0.03818	ubiquitin-conjugating enzyme (E2) [Nicotiana tabacum]	2e-028	protein biosynthesis/degradation
TC116024	13	13	0.003701	Ubiquitin-conjugating enzyme E2-17 kDa (Ubiquitin-protein ligase) (Ubiquitin carrier protein)	6e-084	protein biosynthesis/degradation
TC115969	4	1	0.012364	ubiquitin-conjugating enzyme family [Arabidopsis thaliana]	2e-067	protein biosynthesis/degradation
TC127174	5	0	0.000865	Ca2+-binding EF-hand common family protein, putative [Arabidopsis thaliana]	8e-057	signal transduction
TC124117	6	7	0.047119	Calreticulin precursor	0.0	signal transduction
TC126098	3	1	0.03818	cryptochrome 1b [Lycopersicon esculentum]	0.0	signal transduction
TC116496	4	2	0.027482	GTP-binding protein [Capsicum annum]	0.0	signal transduction
TC125731	4	0	0.003337	leucine-rich repeat transmembrane protein kinase 1 [Arabidopsis thaliana]	0.0	signal transduction
TC124217	6	3	0.007653	phosphoprotein phosphatase 2A regulatory chain - common tobacco	0.0	signal transduction
TC118990	3	0	0.012882	protein kinase 5 [Arabidopsis thaliana]	2e-097	signal transduction
TC116492	6	7	0.047119	putative serine/threonine protein kinase [Nicotiana tabacum]	e-162	signal transduction
TC125517	3	0	0.012882	putative signal sequence receptor, alpha subunit [Arabidopsis thaliana]	6e-060	signal transduction



TC118148	5	0	0.000865	serine/threonine protein kinase - spinach	0.0	signal transduction
TC116971	7	7	0.024414	serine/threonine protein kinase-like protein [Arabidopsis thaliana]	e-114	signal transduction
TC117693	3	1	0.03818	Serine/threonine protein phosphatase PP1 isozyme 1	e-156	signal transduction
TC119785	3	0	0.012882	transducin / WD-40 repeat protein family [Arabidopsis thaliana]	0.0	signal transduction
TC116278	8	7	0.011859	BEL1-related homeotic protein 11 [Solanum tuberosum]	0.0	transcription factor
TC124643	3	1	0.03818	CONSTANS-like protein 1 [Malus x domestica]	5e-083	transcription factor
TC124189	16	6	3.76E-06	Homeobox protein knotted-1 like LET12	0.0	transcription factor
TC124719	7	5	0.010262	MADS box protein [Capsicum annuum]	e-108	transcription factor
TC117705	3	1	0.03818	Mob1-like protein [Arabidopsis thaliana]	e-121	transcription factor
TC126685	3	1	0.03818	putative VIP2 protein [Oryza sativa]	3e-011	transcription factor
TC116647	3	1	0.03818	putative zinc finger protein [Oryza sativa]	8e-081	transcription factor
TC118002	4	0	0.003337	RING zinc finger ankyrin protein -related [Arabidopsis thaliana]	0.0	transcription factor
TC116039	7	0	0.000058	TCP family transcription factor, putative [Arabidopsis thaliana]	2e-055	transcription factor
TC119117	3	0	0.012882	TCP family transcription factor, putative [Arabidopsis thaliana]	7e-032	transcription factor
TC125054	5	3	0.019694	transcriptional coactivator-like protein [Arabidopsis thaliana]	5e-046	transcription factor
TC117006	4	3	0.047513	transcriptional regulator, putative [Arabidopsis thaliana]	2e-063	transcription factor
TC117576	3	1	0.03818	zinc finger protein [Arabidopsis thaliana]	3e-093	transcription factor
TC125348	3	0	0.012882	zinc finger-like protein [Arabidopsis thaliana]	8e-039	transcription factor
TC127852	3	0	0.012882	ADP-ribosylation factor [Arabidopsis thaliana]	1e-059	transport
TC124356	3	0	0.012882	Alpha-soluble NSF attachment protein	e-142	transport
TC117772	3	1	0.03818	auxin efflux carrier protein family [Arabidopsis thaliana]	e-139	transport
TC124296	6	0	0.000224	coatamer delta subunit [Arabidopsis thaliana]	0.0	transport
TC115715	27	4	1.02E-12	major intrinsic protein 2 [Solanum tuberosum]	e-145	transport
TC125603	3	0	0.012882	mitochondrial carrier protein family [Arabidopsis thaliana]	e-148	transport
TC117219	3	0	0.012882	Nonspecific lipid-transfer protein 1 precursor	2e-038	transport
TC116225	9	9	0.012743	putative ABC transporter [Oryza sativa]	0.0	transport
TC117451	4	2	0.027482	putative bactericidal permeability-increasing protein precursor [Arabidopsis thaliana]	2e-071	transport
TC117137	3	0	0.012882	sugar transporter [Arabidopsis thaliana]	e-103	transport

TC124128	6	3	0.007653	triose phosphate/phosphate translocator	0.0	transport
TC116357	4	1	0.012364	vacuolar processing enzyme-1b [Nicotiana tabacum]	0.0	transport
TC124343	10	4	0.000304	CBS domain containing protein [Arabidopsis thaliana]	e-134	unclassified
TC117358	3	1	0.03818	CBS domain containing protein [Arabidopsis thaliana]	1e-068	unclassified
TC119637	3	1	0.03818	dem (defective embryo and meristems gene [Arabidopsis thaliana]	e-121	unclassified
TC124638	7	3	0.002832	Kelch repeat containing F-box protein family [Arabidopsis thaliana]	5e-053	unclassified
TC123969	65	1	2.69E-37	late-embryogenesis protein homolog - tomato	7e-039	unclassified
TC125364	8	2	0.000371	leucine rich repeat protein [Arabidopsis thaliana]	e-129	unclassified
TC115918	3	0	0.012882	nodulin MtN3 family protein [Arabidopsis thaliana]	3e-068	unclassified
TC125069	3	0	0.012882	Ntdin [Nicotiana tabacum]	2e-077	unclassified
TC124928	3	0	0.012882	predicted GPI-anchored protein [Arabidopsis thaliana]	e-159	unclassified
TC124092	3	0	0.012882	putative ripening-related protein [Vitis vinifera]	e-132	unclassified
TC124585	3	0	0.012882	selenium binding protein [Medicago sativa]	e-176	unclassified
TC115845	41	40	8.67E-07	translationally controlled tumor protein homolog	1e-082	unclassified
TC116050	14	4	4.19E-06	VTC2 [Arabidopsis thaliana]	e-172	unclassified
TC124197	14	8	0.000132	expressed protein [Arabidopsis thaliana]	e-140	unknown
TC124441	6	1	0.001162	expressed protein [Arabidopsis thaliana]	3e-013	unknown
TC124647	8	4	0.002243	expressed protein [Arabidopsis thaliana]	8e-034	unknown
TC125118	5	2	0.009968	expressed protein [Arabidopsis thaliana]	e-113	unknown
TC116087	4	1	0.012364	expressed protein [Arabidopsis thaliana]	4e-033	unknown
TC116088	4	1	0.012364	expressed protein [Arabidopsis thaliana]	1e-024	unknown
TC116419	3	0	0.012882	expressed protein [Arabidopsis thaliana]	e-117	unknown
TC124334	3	0	0.012882	expressed protein [Arabidopsis thaliana]	3e-017	unknown
TC127883	3	0	0.012882	expressed protein [Arabidopsis thaliana]	1e-079	unknown
TC116181	6	5	0.023108	expressed protein [Arabidopsis thaliana]	3e-080	unknown
TC125081	4	2	0.027482	expressed protein [Arabidopsis thaliana]	e-154	unknown
TC124374	3	1	0.03818	expressed protein [Arabidopsis thaliana]	3e-059	unknown
TC116163	7	9	0.044676	expressed protein [Arabidopsis thaliana]	1e-034	unknown

TC124382	4	3	0.047513	expressed protein [Arabidopsis thaliana]	2e-037	unknown
TC116250	5	5	0.048653	expressed protein [Arabidopsis thaliana]	8e-082	unknown
TC116996	6	3	0.007653	hypothetical protein [Arabidopsis thaliana]	3e-031	unknown
TC124922	7	7	0.024414	hypothetical protein [Arabidopsis thaliana]	3e-027	unknown
TC125281	3	1	0.03818	hypothetical protein [Arabidopsis thaliana]	3e-045	unknown
TC116978	7	5	0.010262	hypothetical protein [Catharanthus roseus]	2e-041	unknown
TC125623	3	0	0.012882	hypothetical protein [Cicer arietinum]	4e-013	unknown
TC125450	4	2	0.027482	hypothetical protein [imported] - Arabidopsis thaliana	3e-073	unknown
TC116010	3	0	0.012882	No hits found		unknown
TC121532	3	0	0.012882	No hits found		unknown
TC126793	3	0	0.012882	No hits found		unknown
TC127340	3	0	0.012882	No hits found		unknown
TC129392	3	0	0.012882	No hits found		unknown
TC117221	4	2	0.027482	No hits found		unknown
TC117309	3	1	0.03818	No hits found		unknown
TC115811	30	52	0.007035	No hits found		unknown
TC118165	3	0	0.012882	unknown [Arabidopsis thaliana]	8e-020	unknown
TC119027	4	0	0.003337	unknown [Lycopersicon esculentum]	1e-039	unknown
TC117357	4	3	0.047513	unknown protein [Arabidopsis thaliana]	e-111	unknown
TC125321	4	1	0.012364	unknown protein [Arabidopsis thaliana]	1e-078	unknown
TC124289	6	4	0.014176	unknown protein [Oryza sativa]	1e-015	unknown

<sup>a</sup> This column represents the number of ESTs from mature green fruit library with total of 5317 ESTs

<sup>b</sup> This column represents the number of ESTs from breaker fruit library with total of 15207 ESTs

**Table VIII.** Grape ripening-induced genes

TC	green <sup>a</sup>	veraison <sup>b</sup>	P value	Putative function	E value	Functional category
TC10004	0	6	0.016428	BY-2 kinesin-like protein 10 [Nicotiana tabacum]	3e-049	cell structure and maintenance
TC9962	0	5	0.029546	endoplasmatic reticulum retrieval protein Rer1B [validated] - Arabidopsis thaliana	6e-072	cell structure and maintenance
TC9486	0	7	0.009134	exocyst subunit EXO70 family [Arabidopsis thaliana]	6e-085	cell structure and maintenance
TC4469	1	10	0.007668	N3 like protein [Medicago truncatula]	1e-081	cell structure and maintenance
TC10316	0	6	0.016428	putative DNA binding protein [Atriplex hortensis]	e-118	cell structure and maintenance
TC4902	1	8	0.020294	RNA recognition motif (RRM)-containing protein [Arabidopsis thaliana]	1e-027	cell structure and maintenance
TC4321	2	11	0.013422	expansin [Vitis labrusca x Vitis vinifera]	e-139	cell wall
TC9448	0	10	0.00157	proline rich protein 3 [Cicer arietinum]	7e-046	cell wall
TC4311	1	7	0.032443	putative proline-rich cell wall protein [Vitis vinifera]	4e-028	cell wall
TC8886	0	5	0.029546	putative proline-rich cell wall protein {Vitis vinifera}	0.0	cell wall
TC8883	2	74	4.27E-17	putative proline-rich cell wall protein {Vitis vinifera}	0.0	cell wall
TC8913	3	31	3.65E-06	putative proline-rich cell wall protein {Vitis vinifera}	0.0	cell wall
TC4180	5	21	0.002798	putative proline-rich cell wall protein {Vitis vinifera}	0.0	cell wall
TC9656	0	5	0.029546	reversibly glycosylated polypeptide [Gossypium hirsutum]	5e-056	cell wall
TC4282	0	6	0.016428	Xyloglucan endo-transglycosylase [Vitis labrusca x Vitis vinifera]	e-100	cell wall
TC4034	0	10	0.00157	18.2 KD class I heat shock protein	2e-057	defense/stress response
TC4064	2	10	0.020427	18.2 KD class I heat shock protein	2e-058	defense/stress response
TC4348	0	10	0.00157	cell-autonomous heat shock cognate protein 70 [Cucurbita maxima]	0.0	defense/stress response
TC4046	0	5	0.029546	cytosolic class II low molecular weight heat shock protein [Prunus dulcis]	4e-053	defense/stress response
TC9273	0	5	0.029546	DnaJ homolog [Salix gilgiana]	0.0	defense/stress response
TC4181	0	10	0.00157	heat shock protein 18p - common tobacco	3e-054	defense/stress response
TC4193	0	6	0.016428	heat shock protein 18p - common tobacco	1e-054	defense/stress response
TC9134	0	5	0.029546	LMW heat shock protein [Euphorbia esula]	5e-037	defense/stress response
TC4559	0	13	0.00027	low-molecular-weight heat shock protein [Cuscuta japonica]	1e-037	defense/stress response
TC8995	3	12	0.019329	probable glutathione transferase - soybean	4e-092	defense/stress response

TC8874	13	67	3.80E-08	putative metallothionein-like protein [Vitis vinifera]	1e-033	defense/stress response
TC8853	18	39	0.008868	putative metallothionein-like protein [Vitis vinifera]	5e-034	defense/stress response
TC4581	0	7	0.009134	small heat shock protein [Lycopersicon esculentum]	8e-067	defense/stress response
TC10077	0	5	0.029546	UVB-resistance protein-related [Arabidopsis thaliana]	e-180	defense/stress response
TC4249	0	11	0.000873	VVTL1 (fruit-specific thaumatin-like protein) [Vitis vinifera]	e-124	defense/stress response
TC9208	0	7	0.009134	ATP synthase [Arabidopsis thaliana]	5e-085	energy
TC9453	0	5	0.029546	dicyanin [Lycopersicon esculentum]	7e-066	energy
TC8961	0	6	0.016428	acyl-coA-binding protein	2e-031	metabolism
TC4394	1	8	0.020294	alcohol dehydrogenase 2 [Vitis vinifera]	0.0	metabolism
TC4300	1	12	0.002801	alcohol dehydrogenase 7 [Vitis vinifera]	3e-040	metabolism
TC4084	0	19	7.97E-06	DAD1 (lipase) [Petunia x hybrida]	9e-053	metabolism
TC9404	0	5	0.029546	GDSL-motif lipase/hydrolase protein [Arabidopsis thaliana]	e-112	metabolism
TC4592	0	9	0.002824	hydrogenase protein [Arabidopsis thaliana]	2e-090	metabolism
TC9155	2	11	0.013422	omega-6 fatty acid desaturase [Sesamum indicum]	0.0	metabolism
TC9915	0	5	0.029546	PII protein [Ricinus communis]	3e-054	metabolism
TC5599	0	6	0.016428	putative mitochondrial glyoxalase II [Cicer arietinum]	1e-088	metabolism
TC8933	1	38	1.98E-09	rRNA intron-encoded homing endonuclease [Oryza sativa]	2e-025	metabolism
TC4404	1	8	0.020294	S-adenosylmethionine synthetase 1	0.0	metabolism
TC9777	0	9	0.002824	small nuclear ribonucleoprotein U2B - potato	2e-074	metabolism
TC8975	3	13	0.013227	Thioredoxin H-type (TRX-H)	2e-043	metabolism
TC4413	0	7	0.009134	60S ribosomal protein L24 [Prunus avium]	5e-058	protein biosynthesis/degradation
TC4655	0	7	0.009134	60S ribosomal protein L9	2e-090	protein biosynthesis/degradation
TC4209	0	5	0.029546	elongation factor 1-alpha [Lilium longiflorum]	0.0	protein biosynthesis/degradation
TC4822	0	5	0.029546	polyubiquitin [Arabidopsis thaliana]	e-179	protein biosynthesis/degradation
TC9955	0	5	0.029546	protein phosphatase PP2A catalytic subunit [Arabidopsis thaliana]	e-100	protein biosynthesis/degradation
TC4520	0	5	0.029546	ribosomal protein S3a [Cicer arietinum]	e-110	protein biosynthesis/degradation
TC4236	2	8	0.045051	ubiquitin precursor - common sunflower	0.0	protein biosynthesis/degradation
TC4910	0	6	0.016428	calcineurin B [Pisum sativum]	e-105	signal transduction

TC9086	0	5	0.029546	calmodulin [Arabidopsis thaliana]	1e-079	signal transduction
TC9412	0	6	0.016428	putative JUN kinase activation domain binding protein [Medicago sativa]	0.0	transcription factor
TC4730	0	5	0.029546	bZIP transcription factor ATB2 [Glycine max]	3e-039	transcription factor
TC4377	1	7	0.032443	MADS-box protein 4 [Vitis vinifera]	e-135	transcription factor
TC4658	0	6	0.016428	probable CCCH-type zinc finger protein - Arabidopsis thaliana	2e-029	transcription factor
TC4163	34	48	0.042923	putative transcription factor [Vitis vinifera]	5e-021	transcription factor
TC9044	0	7	0.009134	zinc finger - like protein [Arabidopsis thaliana]	5e-039	transcription factor
TC4095	32	69	0.001558	2S albumin precursor [Vitis vinifera]	4e-020	transport
TC9284	1	7	0.032443	ATFP3 (heavy metal ion transport) [Arabidopsis thaliana]	3e-058	transport
TC4153	77	137	0.001384	nonspecific lipid-transfer protein A	1e-028	transport
TC4223	0	12	0.000485	plasma intrinsic protein 2,2 [Juglans regia]	e-120	transport
TC4297	0	11	0.000873	putative aquaporin PIP1-3 [Vitis berlandieri x Vitis rupestris]	e-111	transport
TC4218	3	10	0.0393	putative aquaporin PIP2-1 [Vitis berlandieri x Vitis rupestris]	e-161	transport
TC4481	0	8	0.005079	Vacuolar processing enzyme precursor (VPE)	0.0	transport
TC4495	1	7	0.032443	Bet v I allergen family [Arabidopsis thaliana]	1e-040	unclassified
TC5050	0	5	0.029546	gb protein [Sorghum bicolor]	2e-053	unclassified
TC4259	1	29	3.00E-07	putative ripening-related protein [Vitis vinifera]	1e-096	unclassified
TC9018	0	21	2.46E-06	putative ripening-related protein [Vitis vinifera]	e-119	unclassified
TC8964	2	13	0.005586	putative ripening-related protein [Vitis vinifera]	1e-083	unclassified
TC9052	9	22	0.018525	putative ripening-related protein {Vitis vinifera}	0.0	unclassified
TC4256	12	30	0.008123	putative ripening-related protein {Vitis vinifera}	0.0	unclassified
TC4056	0	9	0.002824	putative senescence-associated protein [Pisum sativum]	2e-077	unclassified
TC10107	0	5	0.029546	specific tissue protein 2 [Cicer arietinum]	3e-022	unclassified
TC9112	1	24	4.70E-06	expressed protein [Arabidopsis thaliana]	3e-050	unknown
TC9200	0	15	8.34E-05	expressed protein [Arabidopsis thaliana]	4e-067	unknown
TC4426	0	13	0.00027	expressed protein [Arabidopsis thaliana]	5e-045	unknown
TC4340	7	16	0.038678	expressed protein [Arabidopsis thaliana]	e-147	unknown
TC4374	7	17	0.03036	hypothetical protein - castor bean	0.0	unknown

TC9399	0	7	0.009134	hypothetical protein [Arabidopsis thaliana]	4e-035	unknown
TC10419	0	6	0.016428	No hits found		unknown
TC10208	0	5	0.029546	No hits found		unknown
TC4049	0	5	0.029546	No hits found		unknown
TC4642	0	5	0.029546	No hits found		unknown
TC5241	0	5	0.029546	No hits found		unknown
TC5867	0	5	0.029546	No hits found		unknown
TC5896	0	5	0.029546	No hits found		unknown
TC9766	0	5	0.029546	No hits found		unknown
TC8968	0	5	0.029546	orf107a [Arabidopsis thaliana]	2e-028	unknown
TC9778	0	7	0.009134	P0434A03.35 [Oryza sativa]	5e-012	unknown
TC9587	0	5	0.029546	unknown [Arabidopsis thaliana]	4e-043	unknown
TC10569	0	5	0.029546	unknown protein [Arabidopsis thaliana]	5e-027	unknown
TC9335	0	5	0.029546	unnamed protein product [Oryza sativa ]	2e-012	unknown

<sup>a</sup> This column represents the number of ESTs from green berry library with total of 4126 ESTs

<sup>b</sup> This column represents the number of ESTs from veraison stage berry library with total of 5167 ESTs

**Table IX.** Tomato pathogenesis-related genes

TC	A <sup>a</sup>	B <sup>b</sup>	C <sup>c</sup>	P value	Putative function	E value	Functional category
TC124475	7	1	0	0.00048	1,3-beta-glucanase - tomato	e-172	cell structure/maintenance
TC115782	2	2	21	0.00097	alpha-tubulin [Nicotiana tabacum]	0.0	cell structure/maintenance
TC116507	0	4	0	0.00366	Cyanogenic Beta-Glucosidase	e-159	cell structure/maintenance
TC116536	6	0	3	0.01814	cystatin [Lycopersicon esculentum]	e-130	cell structure/maintenance
TC124852	4	0	0	0.00433	hydroxyproline-rich glycoprotein family [Arabidopsis thaliana]	e-165	cell structure/maintenance
TC116800	0	1	10	0.01252	profilin [Lycopersicon esculentum]	1e-061	cell structure/maintenance
TC125012	0	3	0	0.01489	Profilin 1	5e-054	cell structure/maintenance
TC124886	0	0	6	0.03315	RNA binding protein homolog - common tobacco	e-164	cell structure/maintenance
TC127763	3	0	0	0.01689	T1N6.18 protein (nucleic acid binding activity) - Arabidopsis thaliana	6e-024	cell structure/maintenance
TC116804	0	0	7	0.01879	xyloglucan endo-1,4-beta-D-glucanase - tomato	e-179	cell structure/maintenance
TC116619	4	1	15	0.0274	Acidic 26 kDa endochitinase precursor	e-149	defense/stress response
TC124673	0	0	6	0.03315	dehydration-induced protein RD22 [Gossypium hirsutum]	2e-086	defense/stress response
TC126072	0	3	0	0.01489	DnaJ-like protein [Arabidopsis thaliana]	3e-044	defense/stress response
TC124001	7	1	1	0.00258	heat shock protein - tomato	2e-078	defense/stress response
TC116767	3	4	0	0.03914	In2-1 protein (glutathione transferase activity) [Glycine max]	4e-089	defense/stress response
TC116105	5	13	9	0.03558	Metallothionein-like protein type 2	1e-016	defense/stress response
TC125209	5	0	2	0.02341	Metallothionein-like protein type 2 A	6e-024	defense/stress response
TC117215	4	0	1	0.02556	outer membrane lipo protein [Arabidopsis thaliana]	2e-075	defense/stress response
TC117463	4	0	0	0.00433	Pathogenesis-related leaf protein 4 precursor (P4)	2e-094	defense/stress response
TC115911	72	8	12	0	Pathogenesis-related leaf protein 6 precursor	9e-094	defense/stress response
TC124773	10	2	5	0.01153	Pathogenesis-related protein P2 precursor	7e-083	defense/stress response
TC124667	3	0	0	0.01689	Pathogenesis-related protein STH-2	1e-083	defense/stress response
TC116258	2	4	0	0.03639	peroxidase, defense-related - tomato	e-165	defense/stress response
TC116130	2	0	9	0.04786	P-rich protein EIG-I30 [Nicotiana tabacum]	1e-034	defense/stress response
TC115713	5	6	1	0.0254	probable chaperonin 60 beta chain precursor - potato	0.0	defense/stress response



TC124153	3	0	0	0.01689	Probable glutathione S-transferase	5e-076	defense/stress response
TC124148	12	5	3	0.00192	Probable glutathione S-transferase	e-105	defense/stress response
TC124150	4	0	0	0.00433	Probable glutathione S-transferase	3e-079	defense/stress response
TC117270	4	1	0	0.02163	similar to hsr203J [ <i>Lycopersicon esculentum</i> ]	0.0	defense/stress response
TC123751	0	3	19	0.00067	snakin2 [ <i>Solanum tuberosum</i> ]	8e-049	defense/stress response
TC125684	0	5	2	0.02014	submergence induced protein 2A [ <i>Arabidopsis thaliana</i> ]	8e-085	defense/stress response
TC124012	5	1	1	0.02819	TSI-1 protein (pathogenesis-related protein) - tomato	5e-066	defense/stress response
TC116380	5	1	0	0.00629	Cytochrome c	1e-059	energy
TC123973	24	19	66	0.01332	Ferredoxin I	6e-066	energy
TC116344	1	0	10	0.01256	inorganic pyrophosphatase [ <i>Arabidopsis thaliana</i> ]	e-122	energy
TC115873	5	1	1	0.02819	oxidoreductase, zinc-binding dehydrogenase family [ <i>Arabidopsis thaliana</i> ]	e-133	energy
TC123957	2	12	11	0.01971	putative chloroplast thiazole biosynthetic protein [ <i>Nicotiana tabacum</i> ]	e-173	energy
TC115851	4	1	0	0.02163	ripening protein E8 homolog - tomato	1e-086	ethylene response
TC124345	1	5	1	0.0244	28 kDa ribonucleoprotein	e-104	metabolism
TC116469	0	4	1	0.02234	4-coumarate--CoA ligase 1	0.0	metabolism
TC124155	5	2	17	0.04442	4-hydroxyphenylpyruvate dioxygenase	8e-036	metabolism
TC124286	4	11	7	0.04092	acetyl-CoA C-acyltransferaseB precursor - mango	1e-089	metabolism
TC116721	1	4	0	0.01906	acyltransferase family [ <i>Arabidopsis thaliana</i> ]	2e-039	metabolism
TC116006	5	2	0	0.01453	Caffeoyl-CoA O-methyltransferase	2e-033	metabolism
TC124080	7	22	33	0.01684	Carbonic anhydrase	e-145	metabolism
TC124119	0	4	1	0.02234	cinnamic acid 4-hydroxylase [ <i>Capsicum annuum</i> ]	0.0	metabolism
TC124409	3	0	0	0.01689	DEAD/DEAH box helicase [ <i>Arabidopsis thaliana</i> ]	e-163	metabolism
TC115766	6	1	0	0.00176	dihydroflavonol reductase [ <i>Arabidopsis thaliana</i> ]	0.0	metabolism
TC125638	0	3	0	0.01489	fatty acid hydroxylase [ <i>Arabidopsis thaliana</i> ]	e-110	metabolism
TC124747	3	0	0	0.01689	glutamate-tRNA ligase	0.0	metabolism
TC125721	3	0	0	0.01689	glutamine synthetase	e-155	metabolism
TC116076	6	12	31	0.03327	glycolate oxidase [ <i>Arabidopsis thaliana</i> ]	0.0	metabolism
TC124202	5	5	36	0.00011	H-Protein precursor (glycine catabolism) [ <i>Flaveria pringlei</i> ]	1e-073	metabolism

TC116542	0	6	0	0.00022	Inositol-3-phosphate synthase	0.0	metabolism
TC117373	3	0	0	0.01689	prohibitin (inhibits DNA synthesis) - common tobacco	e-123	metabolism
TC124306	0	3	0	0.01489	putative acyl-CoA synthetase [Capsicum annuum]	0.0	metabolism
TC116287	2	4	17	0.02828	putative nitrilase-associated protein [Arabidopsis thaliana]	3e-030	metabolism
TC116556	3	0	0	0.01689	pyruvate dehydrogenase E1 alpha subunit [Beta vulgaris]	e-165	metabolism
TC117923	3	0	0	0.01689	ribokinase [Arabidopsis thaliana]	e-123	metabolism
TC123879	12	12	4	0.00256	S-adenosylmethionine decarboxylase proenzyme	0.0	metabolism
TC125552	3	0	0	0.01689	shikimate kinase family [Arabidopsis thaliana]	9e-062	metabolism
TC123761	10	0	4	0.00055	similar to PHZF [Arabidopsis thaliana]	6e-091	metabolism
TC116836	5	0	0	0.00111	small nuclear ribonucleo protein D2 [Arabidopsis thaliana]	6e-044	metabolism
TC124975	0	4	1	0.02234	transcription initiation factor IIB [Arabidopsis thaliana]	e-127	metabolism
TC116152	0	3	0	0.01489	tyramine hydroxycinnamoyl transferase [Solanum tuberosum]	2e-085	metabolism
TC124546	4	0	1	0.02556	udp-glucose 6-dehydrogenase	0.0	metabolism
TC117379	0	7	3	0.00507	ATP synthase delta chain	e-102	photosynthesis
TC124294	6	10	4	0.02431	ATP synthase gamma chain	e-171	photosynthesis
TC116049	9	9	37	0.00978	chlorophyll a/b binding protein CP29 [Vigna radiata]	e-138	photosynthesis
TC124113	12	3	23	0.03446	chlorophyll a/b-binding protein type III precursor - tomato	e-136	photosynthesis
TC124015	2	5	17	0.04281	Chlorophyll A-B binding protein 13	e-146	photosynthesis
TC123782	11	28	37	0.02172	Chlorophyll A-B binding protein 1B	e-146	photosynthesis
TC123794	8	14	5	0.0033	Chlorophyll A-B binding protein 1B	e-141	photosynthesis
TC123857	5	11	30	0.02001	Chlorophyll A-B binding protein 3C	e-148	photosynthesis
TC115982	10	17	46	0.01066	Chlorophyll A-B binding protein 4	e-155	photosynthesis
TC124093	6	19	16	0.01049	Chlorophyll A-B binding protein 7	e-160	photosynthesis
TC115908	20	7	6	4.3E-05	glyceraldehyde 3-phosphate dehydrogenase [Solanum tuberosum]	e-170	photosynthesis
TC116141	5	19	20	0.01254	oxygen-evolving complex protein 16 [Arabidopsis thaliana]	2e-058	photosynthesis
TC115938	5	13	35	0.0059	Oxygen-evolving enhancer protein 1	0.0	photosynthesis
TC123967	8	10	34	0.02495	Oxygen-evolving enhancer protein 2	e-127	photosynthesis
TC123968	2	8	20	0.02884	Oxygen-evolving enhancer protein 2	3e-077	photosynthesis

TC124266	2	5	18	0.029	photosystem I psaH precursor - wood tobacco	2e-051	photosynthesis
TC123728	9	4	30	0.0063	Photosystem I reaction center subunit IV B	1e-038	photosynthesis
TC115785	31	35	123	3E-06	Photosystem II 10 kDa polypeptide	9e-066	photosynthesis
TC116182	12	7	6	0.03068	Photosystem II 22 kDa protein	2e-083	photosynthesis
TC115748	11	12	44	0.00813	Photosystem II core complex proteins psbY	4e-037	photosynthesis
TC116084	1	10	38	1.8E-05	photosystem II reaction center 6.1KD protein [Arabidopsis thaliana]	4e-027	photosynthesis
TC124109	16	28	20	0.00437	Plastocyanin	6e-060	photosynthesis
TC115855	32	35	26	0.00108	Ribulose bisphosphate carboxylase small chain 1	5e-095	photosynthesis
TC115862	4	1	0	0.02163	Ribulose bisphosphate carboxylase small chain 1	3e-075	photosynthesis
TC115856	12	20	16	0.04357	Ribulose bisphosphate carboxylase small chain 1	5e-095	photosynthesis
TC123753	58	97	296	0	Ribulose bisphosphate carboxylase small chain 2A	1e-096	photosynthesis
TC123733	12	19	75	3E-06	Ribulose bisphosphate carboxylase small chain 3A/3C	2e-096	photosynthesis
TC123709	139	182	355	0.00047	Ribulose bisphosphate carboxylase small chain 3A/3C	6e-097	photosynthesis
TC123710	91	135	213	0.00673	Ribulose bisphosphate carboxylase small chain 3B	e-100	photosynthesis
TC115976	1	5	1	0.0244	Ribulose bisphosphate carboxylase/oxygenase activase 1	0.0	photosynthesis
TC121438	3	0	0	0.01689	Ribulose-1,5 bisphosphate carboxylase/oxygenase large subunit N-methyltransferase	9e-071	photosynthesis
TC123791	3	3	17	0.03368	Ribulose-phosphate 3-epimerase	e-140	photosynthesis
TC117109	5	0	0	0.00111	40S ribosomal protein S17	4e-068	protein biosynthesis/degradation
TC125478	0	3	0	0.01489	40S ribosomal protein S18 [Arabidopsis thaliana]	8e-069	protein biosynthesis/degradation
TC123757	3	0	0	0.01689	40S ribosomal protein S23	4e-076	protein biosynthesis/degradation
TC123807	8	3	1	0.00434	40S ribosomal protein S4	e-144	protein biosynthesis/degradation
TC116156	8	3	1	0.00434	40S ribosomal protein S7 [Arabidopsis thaliana]	4e-085	protein biosynthesis/degradation
TC116939	0	7	2	0.00178	50S ribosomal protein L15 [Arabidopsis thaliana]	4e-082	protein biosynthesis/degradation
TC125685	0	5	1	0.00607	50S ribosomal protein L28	1e-060	protein biosynthesis/degradation
TC124290	4	0	1	0.02556	60S ribosomal protein L26 [Arabidopsis thaliana]	6e-043	protein biosynthesis/degradation
TC123831	6	0	2	0.0074	60S ribosomal protein L37A	1e-046	protein biosynthesis/degradation
TC124709	7	0	1	0.00054	60S ribosomal protein L4	e-170	protein biosynthesis/degradation

TC116223	2	4	0	0.03639	CmE8 [Cucumis melo]	e-109	protein biosynthesis/degradation
TC116913	4	2	0	0.03958	Csf-1 [Cucumis sativus]	2e-060	protein biosynthesis/degradation
TC116219	0	3	0	0.01489	cytoplasmic ribosomal protein S14 [Brassica napus]	9e-060	protein biosynthesis/degradation
TC116078	6	5	1	0.0264	elongation factor [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC123773	15	1	0	0	elongation factor-1 alpha [Nicotiana tabacum]	0.0	protein biosynthesis/degradation
TC120932	0	3	0	0.01489	Eukaryotic peptide chain release factor subunit 1[Arabidopsis thaliana]	e-143	protein biosynthesis/degradation
TC116555	0	4	1	0.02234	miraculin homolog - tomato	e-117	protein biosynthesis/degradation
TC124442	7	0	1	0.00054	peptidyl-prolyl isomerase protein [Arabidopsis thaliana]	1e-077	protein biosynthesis/degradation
TC124061	1	5	0	0.00531	plastid ribosomal protein S10 precursor [Mesembryanthemum crystallinum]	2e-063	protein biosynthesis/degradation
TC126123	3	0	0	0.01689	putative glycyl tRNA synthetase [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC124532	1	4	0	0.01906	ribosomal protein L13p family [Arabidopsis thaliana]	e-100	protein biosynthesis/degradation
TC117610	3	0	0	0.01689	ribosomal protein L2 [Atropa belladonna]	e-150	protein biosynthesis/degradation
TC115864	8	2	3	0.0186	ribosomal protein L27a [Petunia x hybrida]	6e-079	protein biosynthesis/degradation
TC116329	4	2	0	0.03958	ribosomal protein S19	4e-062	protein biosynthesis/degradation
TC116227	0	0	6	0.03315	SUI1 protein [Coffea arabica]	1e-055	protein biosynthesis/degradation
TC116149	0	1	13	0.00241	translation-inhibitor protein [Gentiana triflora]	1e-050	protein biosynthesis/degradation
TC117134	0	0	9	0.00603	translation-inhibitor protein [Gentiana triflora]	2e-074	protein biosynthesis/degradation
TC124028	1	1	10	0.04027	ubiquinol-cytochrome-c reductase - potato	e-128	protein biosynthesis/degradation
TC116777	3	0	0	0.01689	ubiquitin-specific protease 6 [Arabidopsis thaliana]	0.0	protein biosynthesis/degradation
TC124332	0	2	9	0.04713	14-3-3 protein [Solanum tuberosum]	e-135	signal transduction
TC125878	3	0	0	0.01689	calcium/calmodulin-dependent protein kinase CaMK2 [Nicotiana tabacum]	0.0	signal transduction
TC116805	5	1	1	0.02819	calnexin 1 [Arabidopsis thaliana]	0.0	signal transduction
TC124565	0	3	0	0.01489	F-box containing tubby family protein [Arabidopsis thaliana]	e-178	signal transduction
TC123741	0	5	2	0.02014	photolyase/blue-light receptor [Arabidopsis thaliana]	e-150	signal transduction
TC125399	3	0	0	0.01689	protein kinase [Arabidopsis thaliana]	e-156	signal transduction
TC116291	1	5	1	0.0244	putative steroid membrane binding protein [Oryza sativa]	5e-077	signal transduction
TC126049	3	0	0	0.01689	CCAAT-binding transcription factor subunit A [Arabidopsis thaliana]	8e-054	transcription factor
TC116368	3	0	0	0.01689	ripening regulated protein DDTFR10/A [Lycopersicon esculentum]	e-107	transcription factor

TC116750	6	1	0	0.00176	ethylene response factor 2 (AP2 domain protein) [ <i>Lycopersicon esculentum</i> ]	e-137	transcription factor
TC124958	2	7	2	0.01759	SENU1, senescence up-regulated - tomato	9e-056	transcription factor
TC125827	0	3	0	0.01489	small zinc finger-like protein [ <i>Lycopersicon esculentum</i> ]	4e-042	transcription factor
TC116053	5	1	1	0.02819	transcription factor BTF3 [ <i>Arabidopsis thaliana</i> ]	2e-047	transcription factor
TC118067	3	0	0	0.01689	zinc finger (C3HC4-type RING finger) protein family [ <i>Arabidopsis thaliana</i> ]	6e-038	transcription factor
TC118444	0	3	0	0.01489	Ammonium transporter 1, member 2 (LeAMT1;2)	0.0	transport
TC125614	0	3	0	0.01489	anion exchange protein family [ <i>Arabidopsis thaliana</i> ]	6e-063	transport
TC125673	8	0	1	0.00014	lipid transfer protein LTP1 precursor [ <i>Capsicum annuum</i> ]	3e-053	transport
TC116611	1	0	11	0.00728	rhodanese-like domain protein [ <i>Arabidopsis thaliana</i> ]	2e-073	transport
TC115869	1	11	17	0.01583	synaptobrevin-like protein [ <i>Oryza sativa</i> ]	e-111	transport
TC124159	0	1	9	0.0215	allergen-like protein BRSn20 [ <i>Sambucus nigra</i> ]	6e-044	unclassified
TC124368	7	1	0	0.00048	late-embryogenesis protein lea5 - common tobacco	9e-021	unclassified
TC125069	0	1	11	0.00725	Ntdin [ <i>Nicotiana tabacum</i> ]	2e-077	unclassified
TC117978	0	3	0	0.01489	seed maturation-like protein [ <i>Arabidopsis thaliana</i> ]	9e-037	unclassified
TC116757	1	9	7	0.02344	SEND32, senescence down-regulated - tomato	3e-029	unclassified
TC115845	20	3	35	0.00124	translationally controlled tumor protein homolog	1e-082	unclassified
TC116498	2	5	1	0.04934	tuberisation-related protein - potato	e-105	unclassified
TC127027	4	0	0	0.00433	expressed protein [ <i>Arabidopsis thaliana</i> ]	2e-080	unknown
TC125226	0	1	10	0.01252	expressed protein [ <i>Arabidopsis thaliana</i> ]	5e-037	unknown
TC125740	0	3	0	0.01489	expressed protein [ <i>Arabidopsis thaliana</i> ]	8e-032	unknown
TC117064	0	0	7	0.01879	expressed protein [ <i>Arabidopsis thaliana</i> ]	0.0	unknown
TC118511	4	0	1	0.02556	expressed protein [ <i>Arabidopsis thaliana</i> ]	1e-028	unknown
TC124991	0	0	6	0.03315	expressed protein [ <i>Arabidopsis thaliana</i> ]	1e-053	unknown
TC117214	3	0	0	0.01689	hypothetical protein [ <i>Arabidopsis thaliana</i> ]	e-111	unknown
TC116280	4	5	27	0.0031	No hits found		unknown
TC127230	0	4	0	0.00366	No hits found		unknown
TC115798	5	0	1	0.00723	No hits found		unknown
TC127589	0	3	0	0.01489	No hits found		unknown

TC125892	4	0	1	0.02556	No hits found		unknown
TC117061	1	0	8	0.03688	No hits found		unknown
TC123898	26	33	21	0.00054	No hits found		unknown
TC124704	1	6	3	0.04997	unknown [Arabidopsis thaliana]	8e-054	unknown
TC116050	9	7	3	0.02139	unknown protein [Arabidopsis thaliana]	e-179	unknown

<sup>a</sup> This column represents the number of ESTs from pseudomonas susceptible library with total of 5243 ESTs

<sup>b</sup> This column represents the number of ESTs from pseudomonas resistant library with total of 5127 ESTs

<sup>c</sup> This column represents the number of ESTs from mixed elicitor library with total of 9135 ESTs

**Table X.** Differentially expressed genes during tomato fruit development

TC	A <sup>a</sup>	B <sup>b</sup>	C <sup>c</sup>	D <sup>d</sup>	E <sup>e</sup>	P value	Putative function	E value
TC116318	0	0	0	24	0	0	17.7 kD class I small heat shock protein [ <i>Lycopersicon esculentum</i> ]	2e-064
TC123932	31	1	9	104	15	0	l-aminocyclopropane-1-carboxylate oxidase 1	e-169
TC115848	7	0	11	149	71	0	l-aminocyclopropane-1-carboxylate oxidase homolog (protein E8)	0.0
TC115849	1	0	7	37	47	0	l-aminocyclopropane-1-carboxylate oxidase homolog (protein E8)	6e-098
TC115850	0	0	2	10	19	0	l-aminocyclopropane-1-carboxylate oxidase homolog (protein E8)	e-131
TC115853	1	1	1	31	11	0	l-aminocyclopropane-1-carboxylate oxidase homolog (protein E8)	e-100
TC116729	16	1	0	0	0	0	l-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) - tomato	0.0
TC115961	4	1	5	71	12	0	3-ketoacyl-CoA thiolase	0.0
TC123825	0	1	1	59	20	0	acid beta-fructofuranosidase precursor	0.0
TC123890	0	2	4	203	51	0	acid beta-fructofuranosidase precursor	0.0
TC115942	0	0	0	152	34	0	acyltransferase 2 [ <i>Capsicum chinense</i> ]	0.0
TC116721	13	0	0	0	0	0	acyltransferase family [ <i>Arabidopsis thaliana</i> ]	2e-039
TC123883	13	9	2	25	90	0	Alcohol dehydrogenase 2	0.0
TC123885	1	0	0	3	12	0	Alcohol dehydrogenase 2	e-135
TC115716	1	26	6	3	0	0	alpha-tubulin [ <i>Nicotiana tabacum</i> ]	0.0
TC124284	0	0	0	32	11	0	aluminum-induced protein-like protein [ <i>Theillungiella halophila</i> ]	5e-058
TC124952	16	0	0	0	0	0	ankyrin like protein [ <i>Arabidopsis thaliana</i> ]	0.0
TC116592	25	0	0	3	0	0	AP2 domain transcription factor, putative [ <i>Arabidopsis thaliana</i> ]	4e-021
TC115728	20	0	0	0	0	0	apyrase precursor	e-145
TC123948	22	0	0	0	0	0	Aspartic protease inhibitor 1 precursor	1e-072
TC116850	20	0	0	0	0	0	ATP synthase CF0 C chain [ <i>Nicotiana tabacum</i> ]	1e-024
TC117244	13	0	0	0	0	0	BCS1 protein-like protein [ <i>Arabidopsis thaliana</i> ]	4e-075
TC123821	0	1	1	209	73	0	beta-fructosidase [ <i>Lycopersicon esculentum</i> ]	0.0
TC125991	0	0	0	0	9	0	calcineurin-like phosphoesterase family [ <i>Arabidopsis thaliana</i> ]	6e-041

TC115865	10	1	30	38	15	0	Catalase isozyme 1	0.0
TC123946	20	0	0	0	0	0	Cathepsin D Inhibitor [ <i>Lycopersicon esculentum</i> ]	e-118
TC117667	0	0	0	0	5	0	Cell wall-associated hydrolase [ <i>Vibrio vulnificus</i> CMCP6]	7e-021f
TC117059	0	0	0	2	10	0	cellulase (EC 3.2.1.4) Cel2 precursor - tomato	0.0
TC124179	23	0	0	7	3	0	Chorismate synthase 1, chloroplast precursor	0.0
TC116126	14	8	0	0	0	0	CIG1 [ <i>Nicotiana tabacum</i> ]	0.0
TC117297	0	0	0	5	9	0	cytochrome p450, putative [ <i>Arabidopsis thaliana</i> ]	e-152
TC118293	0	0	0	0	5	0	DHHC-type zinc finger domain-containing protein [ <i>Arabidopsis thaliana</i> ]	e-119
TC116347	14	0	0	1	0	0	DIP-1 [ <i>Citrullus lanatus</i> ]	0.0
TC116394	0	3	3	0	9	0	DNA repair protein RAD23 homolog [ <i>Arabidopsis thaliana</i> ]	1e-090
TC123931	3	28	9	46	4	0	Enolase	0.0
TC124105	46	2	0	0	0	0	ent-kaurenoic acid oxidase [ <i>Pisum sativum</i> ]	2e-097
TC125086	0	0	0	4	7	0	expansin 1 - tomato	e-152
TC117481	0	0	0	1	5	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	5e-037
TC116748	0	7	1	1	0	0	fasciclin-like arabinogalactan-protein (FLA10) [ <i>Arabidopsis thaliana</i> ]	e-109
TC119588	0	0	0	0	4	0	F-box protein (lectin-related) [ <i>Arabidopsis thaliana</i> ]	3e-060
TC116071	2	2	2	42	4	0	flavin-containing monooxygenase (FMO) family [ <i>Arabidopsis thaliana</i> ]	4e-090
TC116435	20	0	0	0	0	0	gamma-thionin-like protein precursor - tomato	5e-048
TC124169	1	18	3	3	0	0	GDP-mannose pyrophosphorylase [ <i>Solanum tuberosum</i> ]	0.0
TC124329	0	7	0	0	0	0	GDSL-motif lipase/hydrolase protein [ <i>Arabidopsis thaliana</i> ]	e-150
TC116564	0	7	0	0	0	0	glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) small chain - potato	0.0
TC123860	0	15	5	18	0	0	Glyceraldehyde 3-phosphate dehydrogenase A, chloroplast precursor	0.0
TC116003	13	29	55	23	0	0	glycolate oxidase, putative [ <i>Arabidopsis thaliana</i> ]	e-175
TC123774	5	13	3	8	0	0	glycosyl hydrolase family 19 (chitinase) [ <i>Arabidopsis thaliana</i> ]	e-134
TC124185	15	2	0	0	0	0	glyoxalase family protein [ <i>Arabidopsis thaliana</i> ]	8e-040
TC124186	15	0	0	0	0	0	glyoxalase family protein [ <i>Arabidopsis thaliana</i> ]	1e-044
TC123867	7	13	26	10	1	0	heat shock cognate protein 80	0.0



TC123915	0	0	0	100	12	0	Histidine decarboxylase	0.0
TC123916	0	0	0	58	4	0	Histidine decarboxylase	e-146
TC123922	0	0	0	25	15	0	Histidine decarboxylase	e-118
TC116529	29	0	1	0	0	0	homeodomain leucine zipper protein HDZ2 [Phaseolus vulgaris]	1e-073
TC117125	15	0	1	1	0	0	homeodomain protein [Arabidopsis thaliana]	1e-056
TC124451	33	0	0	0	0	0	homeodomain protein [Lycopersicon esculentum]	7e-098
TC124640	20	1	1	1	0	0	hypothetical protein SENU5, senescence up-regulated - tomato	e-134
TC124239	0	0	1	42	5	0	inactive endo-beta-mannanase [Lycopersicon esculentum]	0.0
TC116249	18	0	0	4	0	0	isocitrate lyase	0.0
TC124383	14	1	0	0	0	0	jasmonic acid 1 [Lycopersicon esculentum]	e-113
TC124494	14	0	1	0	0	0	late embryogenesis (Lea)-like protein - tomato	3e-082
TC116653	24	0	1	0	0	0	late embryogenesis (Lea)-like protein ER5, ethylene-responsive - tomato	2e-087
TC123969	2	4	65	1	4	0	late-embryogenesis protein homolog - tomato	7e-039
TC124834	0	1	3	4	9	0	lateral organ boundaries (LOB) domain protein 41 [Arabidopsis thaliana]	7e-058
TC125491	13	0	0	0	0	0	lipase (class 3) family [Arabidopsis thaliana]	5e-066
TC116728	21	2	0	0	0	0	lipase, putative [Arabidopsis thaliana]	3e-077
TC124020	0	0	0	53	15	0	lipoygenase (EC 1.13.11.12) - tomato	0.0
TC115715	0	3	27	4	2	0	major intrinsic protein 2 [Solanum tuberosum]	e-145
TC124174	47	0	0	0	0	0	miraculin precursor [Richadella dulcifica]	3e-018
TC115968	0	1	1	11	11	0	mutant phytoene synthase [Lycopersicon esculentum]	2e-042
TC116288	43	0	1	0	0	0	NAC domain protein NAC2 [Phaseolus vulgaris]	5e-084
TC115798	1	36	23	82	6	0	No hits found	
TC120361	0	0	0	0	4	0	No hits found	
TC124534	20	0	0	0	0	0	No hits found	
TC125395	0	0	0	0	4	0	No hits found	
TC125679	0	0	0	0	6	0	No hits found	
TC115922	15	0	1	0	0	0	nonspecific lipid-transfer protein 2 precursor	1e-051

TC115998	0	0	3	28	17	0	NP24 protein precursor (Pathogenesis-related protein PR P23) (Salt-induced protein)	e-154
TC125601	0	0	0	0	6	0	P0034C11.13 [ <i>Oryza sativa</i> (japonica cultivar-group)]	8e-081
TC116135	0	9	22	60	4	0	Pectinesterase 1 precursor (Pectin methylesterase 1) (PE 1)	0.0
TC124172	14	0	1	0	0	0	Phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplast precursor	0.0
TC124810	0	5	15	1	0	0	phosphoenolpyruvate carboxylase 2 [ <i>Lycopersicon esculentum</i> ]	0.0
TC116182	0	2	0	35	0	0	Photosystem II 22 kDa protein, chloroplast precursor	2e-083
TC115970	0	1	1	55	19	0	Phytoene synthase 1, chloroplast precursor	0.0
TC123871	0	4	22	20	3	0	plastidic aldolase NPALDP1 [ <i>Nicotiana paniculata</i> ]	0.0
TC124337	24	1	0	0	0	0	PnFL-2 [ <i>Ipomoea nil</i> ]	3e-033
TC116388	1	18	8	3	0	0	polygalacturonase (EC 3.2.1.15) 1 beta chain precursor - tomato	0.0
TC124082	0	0	0	55	49	0	Polygalacturonase 2A precursor (PG-2A) (Pectinase)	0.0
TC116904	0	0	0	6	8	0	prenylated rab acceptor (PRA1) family [ <i>Arabidopsis thaliana</i> ]	2e-067
TC125175	14	0	0	1	0	0	prephenate dehydratase family [ <i>Arabidopsis thaliana</i> ]	e-112
TC124735	16	0	0	0	0	0	probable inositol-1,4,5-trisphosphate 5-phosphatase - <i>Arabidopsis thaliana</i>	e-159
TC116198	19	5	1	0	0	0	proline rich protein [ <i>Lycopersicon esculentum</i> ]	2e-028
TC124098	17	0	0	0	0	0	proteinase inhibitor type II TR8 precursor	e-129
TC116861	16	0	0	0	0	0	Pto-responsive gene 1 protein [ <i>Lycopersicon esculentum</i> ]	e-111
TC123957	0	66	9	25	1	0	putative chloroplast thiazole biosynthetic protein [ <i>Nicotiana tabacum</i> ]	e-173
TC116199	3	0	1	38	11	0	putative dehydrogenase [ <i>Arabidopsis thaliana</i> ]	7e-088
TC116192	32	0	3	4	4	0	putative embryo-abundant protein [ <i>Arabidopsis thaliana</i> ]	6e-089
TC125993	0	0	0	0	4	0	putative ethylene-responsive protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	4e-078
TC117039	15	0	0	0	0	0	putative glucosyl transferase [ <i>Arabidopsis thaliana</i> ]	8e-086
TC124553	0	7	1	1	0	0	putative inosine-5-monophosphate dehydrogenase [ <i>Arabidopsis thaliana</i> ]	e-132
TC115787	43	0	40	49	31	0	putative oxidoreductase [ <i>Arabidopsis thaliana</i> ]	3e-073
TC115905	0	0	1	131	105	0	putative pectate lyase [ <i>Arabidopsis thaliana</i> ]	0.0
TC116912	2	7	0	0	0	0	putative pollen surface protein [ <i>Arabidopsis thaliana</i> ]	e-108
TC124870	0	5	0	0	0	0	putative pollen surface protein [ <i>Gossypium raimondii</i> ]	4e-034

TC117053	14	0	0	0	0	0	putative protein [Arabidopsis thaliana]	1e-015
TC123761	0	3	1	45	6	0	putative protein [Arabidopsis thaliana]	6e-091
TC125000	0	0	0	11	9	0	putative protein [Arabidopsis thaliana]	4e-050
TC124092	73	4	3	0	0	0	putative ripening-related protein [Vitis vinifera]	e-132
TC124282	2	0	0	41	6	0	putative short-chain type alcohol dehydrogenase [Solanum tuberosum]	e-117
TC119941	0	0	0	0	4	0	retroelement pol polyprotein-like [Arabidopsis thaliana]	1e-027
TC116326	0	0	0	32	20	0	ripening protein E8 homolog - tomato	0.0
TC123896	4	6	41	13	0	0	S-adenosylmethionine synthetase 1	0.0
TC123976	0	12	3	4	0	0	S-adenosylmethionine synthetase 3	0.0
TC116186	38	0	0	8	0	0	SPF1 protein - sweet potato	e-158
TC124049	54	1	2	10	0	0	syringolide-induced protein 19-1-5 [Glycine max]	e-124
TC115826	143	1	0	0	0	0	Threonine dehydratase biosynthetic, chloroplast precursor	0.0
TC124888	16	2	0	0	0	0	TMV response-related gene product [Nicotiana tabacum]	e-100
TC116252	3	1	0	34	18	0	tomato fruit ripening specific URF [Lycopersicon esculentum]	0.0
TC115845	18	12	41	40	14	0	translationally controlled tumor protein homolog	1e-082
TC116149	36	0	0	0	0	0	translation-inhibitor protein [Gentiana triflora]	1e-050
TC123998	7	3	37	31	9	0	ubiquitin conjugating protein [Avicennia marina]	9e-086
TC124052	1	17	9	16	4	0	UDP-glucose pyrophosphorylase [Solanum tuberosum]	0.0
TC124703	0	6	0	0	0	0	unknown [Arabidopsis thaliana]	2e-090
TC126688	0	0	0	0	4	0	unknown [Oryza sativa (japonica cultivar-group)]	1e-062
TC116050	5	6	14	4	0	0	unknown protein [Arabidopsis thaliana]	e-179
TC126118	0	6	0	0	0	0	Uracil phosphoribosyltransferase	e-116
TC124785	0	0	9	4	0	0	uroporphyrinogen III synthase [Arabidopsis thaliana]	1e-080
TC116099	23	2	0	1	0	0	viroid RNA-binding protein [Lycopersicon esculentum]	0.0
TC117558	13	0	0	0	0	0	wall-associated kinase 3 [Arabidopsis thaliana]	1e-055
TC124156	55	0	0	0	0	0	WIZZ [Nicotiana tabacum]	e-101
TC116040	19	0	1	0	0	0	wound-induced proteinase inhibitor ii precursor	3e-073

TC116257	43	0	1	2	0	0	WRKY family transcription factor [Arabidopsis thaliana]	3e-029
TC116804	0	9	0	0	0	0	xyloglucan endo-1,4-beta-D-glucanase (EC 3.2.1.-) - tomato	e-179
TC125541	0	6	0	0	0	0	zinc-finger protein, putative [Arabidopsis thaliana]	3e-052
TC124656	15	1	2	0	0	0.000001	12-oxophytodienoate reductase 3 [Lycopersicon esculentum]	0.0
TC115782	2	10	6	4	0	0.000001	alpha-tubulin [Nicotiana tabacum]	0.0
TC124855	12	0	0	0	0	0.000001	Asparagine synthetase	0.0
TC124885	12	0	0	0	0	0.000001	CHP-rich zinc finger protein, putative [Arabidopsis thaliana]	2e-022
TC116032	26	2	0	14	0	0.000001	cytochrome p450 family [Arabidopsis thaliana]	0.0
TC125480	0	3	0	0	5	0.000001	expressed protein [Arabidopsis thaliana]	2e-034
TC124825	0	0	0	8	8	0.000001	lipoxygenase (EC 1.13.11.12) loxC, chloroplast - tomato	0.0
TC124313	16	0	0	2	7	0.000001	Mutator-related transposase [Arabidopsis thaliana]	1e-031
TC124171	2	1	3	32	0	0.000001	peptidylprolyl isomerase (EC 5.2.1.8) ROF1 - Arabidopsis thaliana	0.0
TC116231	0	5	1	0	0	0.000001	probable protein kinase - Arabidopsis thaliana	e-103
TC117205	13	0	1	0	0	0.000001	protein kinase family [Arabidopsis thaliana]	e-109
TC116214	0	8	1	3	1	0.000001	proton pump interactor [Arabidopsis thaliana]	e-116
TC116434	0	4	1	31	1	0.000001	putative carboxyphosphoenolpyruvate mutase [Arabidopsis thaliana]	e-156
TC124255	0	4	1	25	12	0.000001	serine carboxypeptidase, putative [Arabidopsis thaliana]	e-133
TC116315	0	2	3	37	7	0.000001	thiol protease [Matricaria chamomilla]	e-120
TC125410	12	0	0	0	0	0.000001	transcription factor CRC [Arabidopsis thaliana]	8e-050
TC115902	2	0	0	1	6	0.000001	ubiquitin extension protein - potato	2e-068
TC125512	11	0	0	0	0	0.000002	ATP synthase CF0 A chain [Nicotiana tabacum]	e-122
TC118567	0	4	0	0	0	0.000002	auxin-regulated protein [Zinnia elegans]	1e-050
TC123813	11	0	0	0	0	0.000002	caffeoyl-CoA O-methyltransferase (EC 2.1.1.104) 5 - common tobacco	e-124
TC124538	11	0	0	0	0	0.000002	cytochrome P450, putative [Arabidopsis thaliana]	e-176
TC125584	11	0	0	0	0	0.000002	expressed protein [Arabidopsis thaliana]	1e-026
TC119444	0	4	0	0	0	0.000002	glycosyltransferase family 2 [Arabidopsis thaliana]	4e-068
TC115949	9	1	2	4	11	0.000002	histone H3.3 [Arabidopsis thaliana]	3e-070

TC124385	11	0	0	0	0	0.000002	homeotic protein - common sunflower	9e-078
TC116866	11	0	0	0	0	0.000002	mitochondrial carrier protein family [Arabidopsis thaliana]	e-113
TC116670	0	4	0	0	0	0.000002	phenylalanyl-trna synthetase - like protein [Arabidopsis thaliana]	e-142
TC125112	11	0	0	0	0	0.000002	photosystem I P700 apoprotein A1 [Nicotiana tabacum]	0.0
TC118795	0	4	0	0	0	0.000002	phytochrome B2 [Lycopersicon esculentum]	0.0
TC117212	0	4	0	0	0	0.000002	probable proteinase inhibitor precursor - tomato	4e-079
TC116039	3	7	7	0	0	0.000002	TCP family transcription factor, putative [Arabidopsis thaliana]	2e-055
TC127174	0	0	5	0	0	0.000003	Ca <sup>2+</sup> -binding EF-hand common family protein, putative [Arabidopsis thaliana]	8e-057
TC124025	12	1	0	0	0	0.000003	No hits found	
TC125327	12	1	0	0	0	0.000003	No hits found	
TC117177	12	0	1	0	0	0.000003	putative protein [Arabidopsis thaliana]	8e-024
TC118148	0	0	5	0	0	0.000003	serine/threonine protein kinase - spinach	0.0
TC116031	0	0	6	2	6	0.000003	xyloglucan endotransglycosylase (XTR4), putative [Arabidopsis thaliana]	4e-069
TC116117	0	0	6	1	0	0.000004	actin - Malva pusilla	0.0
TC124001	0	0	0	20	0	0.000004	heat shock protein MTSHP precursor, mitochondrial - tomato	2e-078
TC125364	0	2	8	2	0	0.000004	leucine rich repeat protein-related [Arabidopsis thaliana]	e-129
TC116059	1	0	0	19	9	0.000004	Peptide methionine sulfoxide reductase	e-114
TC117022	1	1	0	0	5	0.000004	putative methylmalonate semi-aldehyde dehydrogenase [Arabidopsis thaliana]	0.0
TC116397	2	0	6	3	8	0.000005	beta-amylase [Prunus armeniaca]	e-178
TC116874	13	0	0	2	0	0.000006	auxin-regulated protein [Arabidopsis thaliana]	3e-069
TC124296	0	5	6	0	1	0.000006	coatomer delta subunit [Arabidopsis thaliana]	0.0
TC115890	1	7	0	18	11	0.000006	Fruit-specific protein	4e-049
TC124537	2	0	0	23	9	0.000006	similarity to limonene cyclase [Arabidopsis thaliana]	e-123
TC117081	10	0	0	0	0	0.000008	l-aminocyclopropane-1-carboxylate oxidase 2	e-169
TC118036	10	0	0	0	0	0.000008	bHLH protein [Arabidopsis thaliana]	9e-027
TC117227	4	6	0	0	0	0.000008	epoxide hydrolase [Nicotiana tabacum]	2e-097
TC124809	10	0	0	0	0	0.000008	glycosyl hydrolase family 19 (class IV chitinase) [Arabidopsis thaliana]	e-102

TC124407	10	0	0	0	0	0.000008	putative protein [Arabidopsis thaliana]	e-135
TC116067	14	0	4	7	12	0.000009	21K protein precursor - alfalfa	2e-053
TC116116	1	1	8	2	0	0.000009	actin - Malva pusilla	0.0
TC116311	0	0	1	24	2	0.000009	putative PGPD14 protein [Oryza sativa (japonica cultivar-group)]	e-122
TC124012	7	8	11	3	0	0.000009	TSI-1 protein - tomato	5e-066
TC124097	15	3	0	2	0	0.00001	arginine decarboxylase 1 [Datura stramonium]	0.0
TC124189	10	3	16	6	3	0.000012	Homeobox protein knotted-1 like LET12	0.0
TC116444	13	2	1	0	0	0.000013	hypothetical protein - tomato	e-136
TC116869	0	0	7	3	0	0.000014	Inositol-3-phosphate synthase	0.0
TC116914	13	2	0	1	0	0.000015	KED [Nicotiana tabacum]	2e-011
TC124490	11	0	0	1	0	0.000015	phosphoprotein phosphatase 1a catalytic chain - Madagascar periwinkle	e-164
TC124788	11	0	0	1	0	0.000015	xyloglucan endo-1,4-beta-D-glucanase (EC 3.2.1.-) precursor - tomato	e-118
TC124106	0	1	10	12	9	0.000016	abscisic stress ripening protein 1	3e-021
TC116695	0	7	0	4	2	0.000016	hypothetical protein [Arabidopsis thaliana]	e-111
TC116081	16	2	8	6	13	0.000016	RUB1 conjugating enzyme [Olea europaea]	3e-097
TC115712	2	0	7	1	0	0.000017	plastidic aldolase [Nicotiana paniculata]	0.0
TC116749	3	0	0	0	5	0.000017	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx)	1e-093
TC127464	0	0	0	1	4	0.000018	No hits found	
TC116413	0	0	10	23	1	0.000019	Glutamate decarboxylase	0.0
TC124777	12	1	0	1	0	0.00002	mutator-like transposase [Arabidopsis thaliana]	4e-025
TC124701	0	5	1	1	0	0.000021	cellulase (EC 3.2.1.4) Cel3, membrane-anchored - tomato	0.0
TC124918	12	0	0	2	0	0.000021	glycosyltransferase family 8 [Arabidopsis thaliana]	0.0
TC128754	0	0	0	0	3	0.000023	anti-silencing function 1a protein [Arabidopsis thaliana]	5e-070
TC115788	0	0	0	0	3	0.000023	AX110P -like protein [Arabidopsis thaliana]	e-129
TC125018	0	0	0	0	3	0.000023	expressed protein [Arabidopsis thaliana]	2e-086
TC126340	0	0	0	0	3	0.000023	expressed protein [Arabidopsis thaliana]	1e-035
TC121206	0	0	0	0	3	0.000023	heterogeneous nuclear ribonucleoprotein (hnRNP), putative [Arabidopsis thaliana]	6e-093

TC127430	0	0	0	0	3	0.000023	hypothetical protein [Spinacia oleracea]	9e-029
TC127476	0	0	0	0	3	0.000023	Lignin forming anionic peroxidase precursor	9e-086
TC120982	0	0	0	0	3	0.000023	No hits found	
TC125574	0	0	0	0	3	0.000023	No hits found	
TC125986	0	0	0	0	3	0.000023	No hits found	
TC123716	0	0	0	0	3	0.000023	protein phosphatase 2C (PP2C) [Arabidopsis thaliana]	e-115
TC128172	0	0	0	0	3	0.000023	putative cytochrome b-561 [Oryza sativa (japonica cultivar-group)]	3e-048
TC124452	0	0	0	0	3	0.000023	putative membrane protein [Oryza sativa (japonica cultivar-group)]	e-130
TC128944	0	0	0	0	3	0.000023	putative para-aminobenzoate synthase and glutamine amidotransferase, a bifunctional enzyme [Arabidopsis thaliana]	2e-087
TC115907	0	0	0	0	3	0.000023	putative pectate-lyase [Capsicum annuum]	1e-096
TC126242	0	0	0	0	3	0.000023	putative RING zinc finger protein [Arabidopsis thaliana]	1e-040
TC118749	0	0	0	0	3	0.000023	putative SCARECROW gene regulator [Oryza sativa]	4e-013
TC126276	0	0	0	0	3	0.000023	putative translation releasing factor2 [Oryza sativa (japonica cultivar-group)]	e-143
TC126626	0	0	0	0	3	0.000023	unknown [Arabidopsis thaliana]	3e-011
TC120826	0	0	0	0	3	0.000023	unknown protein [Arabidopsis thaliana]	6e-028
TC125438	0	0	0	0	3	0.000023	unknown protein; 69440-65718 [Arabidopsis thaliana]	3e-014
TC120248	0	0	0	0	3	0.000023	wound inducive gene [Nicotiana tabacum]	e-121
TC117397	0	0	0	0	3	0.000023	Yippee-like protein	2e-028
TC125981	0	0	0	0	3	0.000023	zinc finger (C3HC4-type RING finger) protein family [Arabidopsis thaliana]	5e-053
TC116593	3	0	2	26	0	0.000027	HSP100/ClpB, putative [Arabidopsis thaliana]	0.0
TC115811	13	7	30	52	10	0.000029	No hits found	
TC118137	9	0	0	0	0	0.00003	18.1 KD class i heat shock protein	7e-023
TC124046	9	0	0	0	0	0.00003	expressed protein [Arabidopsis thaliana]	4e-014
TC116923	9	0	0	0	0	0.00003	flavanone 3-hydroxylase-like protein [Arabidopsis thaliana]	e-139
TC125919	9	0	0	0	0	0.00003	glutamic acid-rich protein [Arabidopsis thaliana]	2e-047
TC126045	9	0	0	0	0	0.00003	No hits found	
TC126184	9	0	0	0	0	0.00003	No hits found	

TC126198	9	0	0	0	0	0.00003	No hits found	
TC117529	9	0	0	0	0	0.00003	pectinesterase family [Arabidopsis thaliana]	6e-054
TC116018	9	0	0	0	0	0.00003	Phenylalanine ammonia-lyase (PAL)	e-148
TC125906	9	0	0	0	0	0.00003	photoperiod responsive protein [Solanum tuberosum subsp. andigena]	3e-075
TC124609	9	0	0	0	0	0.00003	WRKY DNA-binding protein [Nicotiana tabacum]	0.0
TC116615	0	1	6	1	0	0.000031	IAA-Ala hydrolase (IAR3) [Arabidopsis thaliana]	e-170
TC126452	0	4	0	0	1	0.000031	iron superoxide dismutase, putative [Arabidopsis thaliana]	2e-062
TC124145	0	4	13	11	3	0.000032	ATP-dependent clp protease ATP-binding subunit clpA homolog CD4B, chloroplast precursor	0.0
TC123939	0	4	1	0	0	0.000039	beta tubulin [Arabidopsis thaliana]	0.0
TC125674	10	0	0	0	1	0.000042	Putative SKP1-like protein [Oryza sativa (japonica cultivar-group)]	2e-016
TC125214	1	0	6	0	3	0.000042	serine protease-like protein [Nicotiana tabacum]	0.0
TC124289	10	2	6	4	11	0.000043	unknown protein [Oryza sativa]	9e-016
TC124725	10	1	0	0	0	0.000044	Kunitz-type enzyme inhibitor S9C11 [Solanum tuberosum]	2e-019
TC115894	2	3	2	33	4	0.000045	guanylate kinase [Nicotiana tabacum]	e-150
TC124194	0	1	0	19	7	0.000046	PHAP2A protein [Petunia x hybrida]	e-116
TC116522	0	0	4	25	2	0.000047	Putative Transcription initiation factor IIE, beta subunit [Oryza sativa (japonica cultivar-group)]	6e-093
TC125279	0	0	4	0	0	0.00005	arabinosidase ARA-1 [Lycopersicon esculentum]	0.0
TC119027	0	0	4	0	0	0.00005	unknown [Lycopersicon esculentum]	1e-039
TC119023	0	0	4	0	0	0.00005	wound-induced protein 1	2e-031
TC123771	1	12	12	42	7	0.000054	elongation factor 1-ALPHA (EF-1-ALPHA)	0.0
TC125191	10	0	0	1	0	0.000058	AAA-type ATPase family [Arabidopsis thaliana]	2e-069
TC125260	10	0	0	1	0	0.000058	beta-fructofuranosidase [Lycopersicon esculentum]	0.0
TC116506	10	0	0	1	0	0.000058	katanin p60 subunit A 1-like [Oryza sativa (japonica cultivar-group)]	e-144
TC117741	0	0	5	1	0	0.00006	copper amine oxidase -like protein [Arabidopsis thaliana]	0.0
TC117323	0	4	0	1	0	0.000061	putative beta-mannosidase [Gossypium hirsutum]	0.0
TC125665	1	5	2	0	0	0.000063	No hits found	
TC124441	0	2	6	1	0	0.000069	expressed protein [Arabidopsis thaliana]	3e-013



TC116297	0	3	0	0	0	0.000071	110 kDa 4Snc-Tudor domain protein [Pisum sativum]	0.0
TC117904	0	3	0	0	0	0.000071	ACT domain-containing protein [Arabidopsis thaliana]	e-116
TC116765	0	3	0	0	0	0.000071	acyltransferase homolog [Petunia x hybrida]	0.0
TC117089	0	3	0	0	0	0.000071	alcohol NADP+ oxidoreductase [Solanum tuberosum]	e-158
TC123925	0	3	0	0	0	0.000071	disease resistance protein family (LRR) [Arabidopsis thaliana]	e-120
TC125604	0	3	0	0	0	0.000071	elicitor-inducible cytochrome P450 [Nicotiana tabacum]	e-110
TC120297	0	3	0	0	0	0.000071	expressed protein [Arabidopsis thaliana]	5e-038
TC128577	0	3	0	0	0	0.000071	expressed protein [Arabidopsis thaliana]	8e-073
TC118950	0	3	0	0	0	0.000071	glycosyl hydrolase family 9 [Arabidopsis thaliana]	5e-081
TC125822	0	3	0	0	0	0.000071	mitochondrial carrier protein family [Arabidopsis thaliana]	2e-063
TC128512	0	3	0	0	0	0.000071	No hits found	
TC127074	0	3	0	0	0	0.000071	phantastica [Lycopersicon esculentum]	e-173
TC125929	0	3	0	0	0	0.000071	putative anthocyanidine rhamnosyl-transferase [Capsicum annuum]	e-147
TC125259	0	3	0	0	0	0.000071	putative beta-galactosidase [Lycopersicon esculentum]	0.0
TC119028	0	3	0	0	0	0.000071	putative protein [Arabidopsis thaliana]	7e-025
TC117054	0	3	0	0	0	0.000071	putative replication protein A1 [Arabidopsis thaliana]	3e-030
TC117398	0	3	0	0	0	0.000071	putative translation initiation factor [Arabidopsis thaliana]	e-109
TC128945	0	3	0	0	0	0.000071	putative vacuolar protein sorting-associated protein [Arabidopsis thaliana]	2e-082
TC118008	0	3	0	0	0	0.000071	unknown protein [Arabidopsis thaliana]	2e-081
TC124913	0	2	0	4	6	0.000076	glucan 1,3-beta-glucosidase (EC 3.2.1.58) - common tobacco	0.0
TC124287	11	0	0	2	0	0.000076	pectinesterase (pectin methylesterase) family [Arabidopsis thaliana]	0.0
TC116554	0	0	1	21	5	0.000086	monooxygenase [Solanum tuberosum]	0.0
TC124638	1	0	7	3	0	0.000091	Kelch repeat containing F-box protein family [Arabidopsis thaliana]	5e-053
TC124197	6	5	14	8	1	0.000093	expressed protein [Arabidopsis thaliana]	e-140
TC123985	32	2	10	14	6	0.000093	Formate dehydrogenase, mitochondrial precursor	0.0
TC124343	2	3	10	4	1	0.000094	CBS domain containing protein [Arabidopsis thaliana]	e-134
TC124806	0	6	2	3	0	0.000102	transaldolase ToTAL2 [Lycopersicon esculentum]	0.0

TC116841	0	6	0	4	1	0.000108	2-oxoglutarate dehydrogenase E2 subunit [Arabidopsis thaliana]	1e-090
TC123962	2	5	1	0	0	0.000112	40S ribosomal protein SA	e-112
TC116530	8	0	0	0	0	0.000115	homeodomain leucine zipper protein HDZ2 [Phaseolus vulgaris]	1e-022
TC115913	8	0	0	0	0	0.000115	N3 like protein [Medicago truncatula]	7e-075
TC118357	8	0	0	0	0	0.000115	No hits found	
TC125848	8	0	0	0	0	0.000115	putative protein [Arabidopsis thaliana]	9e-032
TC126246	8	0	0	0	0	0.000115	Serine carboxypeptidase II-3 precursor (CP-MII.3)	1e-063
TC125825	8	0	0	0	0	0.000115	SYNC1 protein [Arabidopsis thaliana]	4e-057
TC126106	8	0	0	0	0	0.000115	syringolide-induced protein 14-1-1 [Glycine max]	1e-018
TC126335	0	1	1	0	4	0.00012	putative protein [Arabidopsis thaliana]	1e-017
TC125091	0	4	2	0	0	0.000122	dihydroxyacid dehydratase, putative [Arabidopsis thaliana]	0.0
TC124619	0	4	2	0	0	0.000122	nucleic acid binding protein [Nicotiana tabacum]	e-133
TC116893	0	0	0	18	2	0.000124	pol polyprotein [Citrus x paradisi]	e-129
TC124513	0	0	7	5	0	0.000134	uridine diphosphate glucose epimerase [Arabidopsis thaliana]	e-169
TC116324	1	4	0	21	0	0.000143	pyruvate decarboxylase [Solanum tuberosum]	0.0
TC116389	6	1	1	0	6	0.000149	unknown [Arabidopsis thaliana]	5e-031
TC116357	1	0	4	1	5	0.000151	vacuolar processing enzyme-1b [Nicotiana tabacum]	0.0
TC116029	4	14	3	27	5	0.000154	poly(A)-binding protein [Nicotiana tabacum]	0.0
TC116024	2	2	13	13	2	0.000156	Ubiquitin-conjugating enzyme E2-17 kDa	6e-084
TC124100	9	1	0	0	0	0.000158	Aminopeptidase 1, chloroplast precursor	0.0
TC117133	11	2	0	1	0	0.000159	hypersensitive response assisting protein [Capsicum annuum]	7e-022
TC116293	0	0	2	21	2	0.000162	glutathione S-transferase [Capsicum annuum]	e-112
TC125731	0	0	4	0	3	0.000175	leucine-rich repeat transmembrane protein kinase 1, putative [Arabidopsis thaliana]	0.0
TC123778	0	8	13	30	4	0.000181	elongation factor 1-ALPHA (EF-1-ALPHA)	0.0
TC124099	0	8	2	14	0	0.000187	Ferredoxin--NADP reductase, leaf-type isozyme, chloroplast precursor	0.0
TC124193	14	0	2	3	6	0.000208	expressed protein [Arabidopsis thaliana]	2e-099
TC123901	10	1	1	0	0	0.000212	glutamate decarboxylase isozyme 1 [Nicotiana tabacum]	0.0

TC116896	10	1	1	0	0	0.000212	membrane related protein CP5, putative [Arabidopsis thaliana]	4e-095
TC116834	9	0	0	1	0	0.000214	l-amino-cyclopropane-1-carboxylate synthase [Lycopersicon esculentum]	0.0
TC126132	0	2	1	0	4	0.000215	FH protein interacting protein FIP1	2e-083
TC124188	0	0	9	12	1	0.000226	LYTB-like protein precursor [Adonis palaestina]	0.0
TC116483	12	0	2	2	0	0.000244	whitefly-induced gp91-phox [Lycopersicon esculentum]	0.0
TC125583	10	0	0	1	1	0.000246	inter-alpha-trypsin inhibitor heavy chain-related [Arabidopsis thaliana]	7e-044
TC124700	0	0	0	14	0	0.000254	NTGP4 [Nicotiana tabacum]	e-158
TC117003	10	0	0	2	0	0.000268	GHMP kinase family [Arabidopsis thaliana]	e-160
TC124305	0	0	2	19	1	0.000268	putative acyl-CoA synthetase [Capsicum annuum]	0.0
TC124688	9	0	0	0	2	0.000272	S-adenosylmethionine:2-demethylmenaquinone methyltransferase-like protein [Oryza sativa (japonica cultivar-group)]	3e-081
TC115852	0	0	0	17	3	0.000305	l-aminocyclopropane-1-carboxylate oxidase homolog (PROTEIN E8)	e-175
TC124283	7	0	0	23	0	0.000306	maize gl1 homolog [Arabidopsis thaliana]	0.0
TC123734	1	1	0	2	5	0.000308	ABA-responsive protein - like [Arabidopsis thaliana]	3e-096
TC116044	2	1	1	24	2	0.000309	Adenosylhomocysteinase	0.0
TC124192	0	4	3	27	3	0.00031	homeotic protein VAHOX1 - tomato	e-170
TC117493	9	2	0	0	0	0.000311	Avr9/Cf-9 rapidly elicited protein 146 [Nicotiana tabacum]	2e-055
TC124512	11	0	1	2	0	0.000321	WRKY family transcription factor [Arabidopsis thaliana]	1e-055
TC115895	0	2	8	27	3	0.000326	ubiquitin [Arabidopsis thaliana]	0.0
TC116078	1	2	3	23	0	0.000335	elongation factor, putative [Arabidopsis thaliana]	0.0
TC116391	0	5	0	4	0	0.000354	Argonaute (AGO1)-like protein [Arabidopsis thaliana]	0.0
TC126447	3	0	0	0	4	0.000374	F-box protein family [Arabidopsis thaliana]	2e-017
TC125233	10	0	1	0	2	0.000377	expressed protein [Arabidopsis thaliana]	8e-034
TC115882	0	1	0	17	1	0.000388	probable calcium-binding protein (clone Y8) - potato	0.0
TC120097	0	1	0	0	3	0.000391	No hits found	
TC118737	0	1	0	0	3	0.000391	phi-2 [Nicotiana tabacum]	e-145
TC124584	1	0	1	15	7	0.00041	ethylene receptor - tomato	0.0
TC118094	7	0	0	0	0	0.000433	bHLH protein SPATULA (SPT) [Arabidopsis thaliana]	6e-041

TC125077	7	0	0	0	0	0.000433	ENOD18 protein [ <i>Vicia faba</i> ]	6e-055
TC118061	7	0	0	0	0	0.000433	expressed protein [ <i>Arabidopsis thaliana</i> ]	8e-034
TC118688	7	0	0	0	0	0.000433	formin-like protein [ <i>Arabidopsis thaliana</i> ]	2e-068
TC118721	7	0	0	0	0	0.000433	No hits found	
TC126129	7	0	0	0	0	0.000433	No hits found	
TC126461	7	0	0	0	0	0.000433	No hits found	
TC126505	7	0	0	0	0	0.000433	No hits found	
TC116349	7	0	0	0	0	0.000433	ovule/fiber cell elongation protein Ghfe1 [ <i>Gossypium hirsutum</i> ]	0.0
TC126426	7	0	0	0	0	0.000433	pentatricopeptide (PPR) repeat-containing protein [ <i>Arabidopsis thaliana</i> ]	1e-024
TC124027	7	0	0	0	0	0.000433	Polyphenol oxidase F, chloroplast precursor	0.0
TC115794	7	0	0	0	0	0.000433	putative oxidoreductase [ <i>Arabidopsis thaliana</i> ]	4e-034
TC124761	7	0	0	0	0	0.000433	thioesterase family [ <i>Arabidopsis thaliana</i> ]	e-106
TC125797	7	0	0	0	0	0.000433	transfactor [ <i>Nicotiana tabacum</i> ]	2e-055
TC124601	7	0	0	0	0	0.000433	unknown protein [ <i>Arabidopsis thaliana</i> ]	1e-072
TC117489	7	0	0	0	0	0.000433	zinc finger protein (PMZ), putative [ <i>Arabidopsis thaliana</i> ]	1e-042
TC116115	0	4	8	30	4	0.000437	malate dehydrogenase precursor - tomato	0.0
TC116574	0	0	2	20	4	0.000455	ethylene-responsive protein, putative [ <i>Arabidopsis thaliana</i> ]	5e-047
TC117044	0	0	0	14	5	0.000462	glucuronosyl transferase homolog, ripening-related - tomato	0.0
TC116538	0	0	1	15	0	0.000478	fructokinase, putative [ <i>Arabidopsis thaliana</i> ]	4e-044
TC125267	0	2	0	0	3	0.000481	beta 1,3-glycosyltransferase-like protein I [ <i>Lycopersicon esculentum</i> ]	0.0
TC125090	0	1	4	0	0	0.000496	putative chloroplast nucleoid DNA-binding protein [ <i>Arabidopsis thaliana</i> ]	e-134
TC118002	0	1	4	0	0	0.000496	putative RING zinc finger ankyrin protein [ <i>Arabidopsis thaliana</i> ]	0.0
TC116978	0	0	7	5	1	0.000509	hypothetical protein [ <i>Catharanthus roseus</i> ]	2e-041
TC117433	0	0	0	13	0	0.000511	putative exosome component [ <i>Oryza sativa</i> (japonica cultivar-group)]	5e-074
TC125390	0	0	0	13	0	0.000511	unknown protein [ <i>Arabidopsis thaliana</i> ]	4e-033
TC124671	0	4	0	2	0	0.000516	prohibitin 1-like protein [ <i>Brassica napus</i> ]	e-133
TC124641	0	4	0	2	0	0.000516	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase alpha subunit	0.0

TC117795	6	0	0	0	4	0.000527	BAC19.10 [ <i>Lycopersicon esculentum</i> ]	4e-085
TC123807	3	11	5	10	3	0.000536	40S ribosomal protein S4	e-144
TC125421	8	1	0	0	0	0.000558	putative ripening-related protein - like [ <i>Arabidopsis thaliana</i> ]	2e-091
TC116278	21	1	8	7	7	0.000569	BEL1-related homeotic protein 11 [ <i>Solanum tuberosum</i> ]	0.0
TC124224	1	6	1	4	0	0.000604	aluminum-induced protein-related [ <i>Arabidopsis thaliana</i> ]	e-102
TC125644	8	0	1	0	0	0.000612	No hits found	
TC124922	0	0	7	7	5	0.0007	glycine-rich protein [ <i>Arabidopsis thaliana</i> ]	3e-022
TC115835	1	0	0	0	3	0.000705	putative mitochondrial NAD-dependent malate dehydrogenase [ <i>Solanum tuberosum</i> ]	2e-053
TC116753	11	1	2	1	0	0.000723	putative sensory transduction histidine kinase [ <i>Arabidopsis thaliana</i> ]	0.0
TC124658	1	0	2	2	5	0.000771	GTP-binding protein, ras-related - common tobacco	e-110
TC118066	8	0	0	1	0	0.000776	No hits found	
TC118149	8	0	0	1	0	0.000776	similar to AP2 domain transcription factor, putative [ <i>Arabidopsis thaliana</i> ]	2e-033
TC125633	0	0	0	7	5	0.000798	lipase, putative [ <i>Arabidopsis thaliana</i> ]	7e-046
TC117900	0	3	0	0	1	0.000813	cochlorine N-methyltransferase [ <i>Arabidopsis thaliana</i> ]	e-154
TC126654	0	3	0	0	1	0.000813	putative myosin heavy chain [ <i>Arabidopsis thaliana</i> ]	2e-049
TC123879	0	1	4	23	5	0.000813	S-adenosylmethionine decarboxylase proenzyme	0.0
TC119167	0	0	0	1	3	0.000823	expressed protein [ <i>Arabidopsis thaliana</i> ]	1e-089
TC127657	0	0	0	1	3	0.000823	Mob1-like protein [ <i>Arabidopsis thaliana</i> ]	e-115
TC126853	0	0	0	1	3	0.000823	UDP-GlcNAc:dolichol phosphate N-acetylglucosamine-1-phosphate transferase, putative [ <i>Arabidopsis thaliana</i> ]	e-112
TC126025	2	0	1	0	4	0.000829	unknown protein [ <i>Arabidopsis thaliana</i> ]	2e-011
TC125618	0	0	0	3	4	0.000831	aldose 1-epimerase - like protein [ <i>Arabidopsis thaliana</i> ]	e-132
TC117570	0	0	0	3	4	0.000831	glucosyltransferase IS5a (EC 2.4.1.-), salicylate-induced - common tobacco	e-133
TC125211	0	0	0	3	4	0.000831	glutathione S-transferase-like protein [ <i>Lycopersicon esculentum</i> ]	e-127
TC117401	0	0	0	3	4	0.000831	zinc finger (C3HC4-type RING finger) protein family [ <i>Arabidopsis thaliana</i> ]	2e-083
TC124611	1	7	1	6	1	0.000837	protein F9L1.16 - <i>Arabidopsis thaliana</i>	2e-052
TC117284	0	0	3	0	0	0.000882	adenylate kinase	e-121
TC127852	0	0	3	0	0	0.000882	ADP-ribosylation factor -like protein [ <i>Arabidopsis thaliana</i> ]	1e-059

TC119117	0	0	3	0	0	0.000882	auxin-induced basic helix-loop-helix transcription factor, putative [Arabidopsis thaliana]	7e-032
TC124842	0	0	3	0	0	0.000882	DS2 protein [Solanum tuberosum]	1e-064
TC121307	0	0	3	0	0	0.000882	dynamamin-like protein [Arabidopsis thaliana]	e-101
TC123779	0	0	3	0	0	0.000882	elongation factor 1-ALPHA (EF-1-ALPHA)	5e-068
TC127883	0	0	3	0	0	0.000882	expressed protein [Arabidopsis thaliana]	1e-079
TC116475	0	0	3	0	0	0.000882	Hydroquinone glucosyltransferase (Arbutin synthase)	0.0
TC125623	0	0	3	0	0	0.000882	hypothetical protein [Cicer arietinum]	4e-013
TC119522	0	0	3	0	0	0.000882	IAA-Ala hydrolase [Arabidopsis thaliana]	e-113
TC117137	0	0	3	0	0	0.000882	integral membrane protein, putative [Arabidopsis thaliana]	e-103
TC125228	0	0	3	0	0	0.000882	nhp2-like protein [Oryza sativa (japonica cultivar-group)]	3e-061
TC116010	0	0	3	0	0	0.000882	No hits found	
TC121532	0	0	3	0	0	0.000882	No hits found	
TC126793	0	0	3	0	0	0.000882	No hits found	
TC127340	0	0	3	0	0	0.000882	No hits found	
TC129392	0	0	3	0	0	0.000882	No hits found	
TC118990	0	0	3	0	0	0.000882	protein kinase 5 [Arabidopsis thaliana]	2e-097
TC116713	0	0	3	0	0	0.000882	purine-rich single-stranded DNA-binding protein -related [Arabidopsis thaliana]	2e-061
TC121503	0	0	3	0	0	0.000882	putative non-LTR retroelement reverse transcriptase [Oryza sativa (japonica cultivar-group)]	6e-014
TC125517	0	0	3	0	0	0.000882	putative signal sequence receptor, alpha subunit [Arabidopsis thaliana]	6e-060
TC119785	0	0	3	0	0	0.000882	putative WD-repeat protein [Arabidopsis thaliana]	0.0
TC118165	0	0	3	0	0	0.000882	unknown [Arabidopsis thaliana]	8e-020
TC116419	0	0	3	0	0	0.000882	unknown protein [Arabidopsis thaliana]	e-117
TC116700	0	2	0	15	0	0.000905	putative 1-aminocyclopropane-1-carboxylate deaminase [Betula pendula]	3e-094
TC124448	10	1	0	1	1	0.000912	GTP-binding protein - curled-leaved tobacco	3e-084
TC117110	9	0	0	2	0	0.000919	xyloglucan endo-1,4-beta-D-glucanase precursor - tomato	e-154
TC116936	0	0	2	0	3	0.000937	similarity to lysophosphatidic acid acyltransferase [Arabidopsis thaliana]	e-150
TC124721	11	1	0	3	0	0.00094	putative protein [Arabidopsis thaliana]	e-153

TC124743	0	3	3	0	0	0.000982	26S proteasome regulatory subunit, putative [Arabidopsis thaliana]	0.0
TC124570	0	2	9	18	1	0.000988	AMP-binding protein, putative [Arabidopsis thaliana]	0.0
TC124119	3	0	2	22	1	0.001003	cinnamic acid 4-hydroxylase [Capsicum annuum]	0.0
TC124039	0	2	4	23	2	0.001008	Protein disulfide isomerase precursor	e-178
TC116178	0	5	1	2	2	0.00102	ketol-acid reductoisomerase	0.0
TC124154	0	0	0	12	0	0.001023	4-hydroxyphenylpyruvate dioxygenase	e-180
TC116440	0	0	0	12	0	0.001023	DEAD/DEAH box helicase carpel factory-related [Arabidopsis thaliana]	e-163
TC124723	10	0	1	2	0	0.001052	auxin-regulated protein [Lycopersicon esculentum]	0.0
TC125305	0	0	0	9	5	0.001061	TDR4 transcription factor [Lycopersicon esculentum]	e-136
TC116047	0	1	1	1	4	0.001071	putative phloem transcription factor M1 [Apium graveolens]	3e-073
TC125075	0	3	1	0	0	0.001096	BAG domain containing protein [Arabidopsis thaliana]	3e-072
TC116006	0	3	1	0	0	0.001096	caffeoyl-CoA O-methyltransferase [Solanum tuberosum]	2e-033
TC127562	0	3	1	0	0	0.001096	expressed protein [Arabidopsis thaliana]	7e-021
TC123834	0	3	1	0	0	0.001096	GTP-binding protein Rab6 - common tobacco	e-102
TC116857	0	3	1	0	0	0.001096	No hits found	
TC119244	0	3	1	0	0	0.001096	Notchless protein homolog [Arabidopsis thaliana]	4e-061
TC116354	0	3	1	0	0	0.001096	Pectinesterase U1 precursor (Pectin methylesterase) (PE)	0.0
TC124442	0	3	1	0	0	0.001096	peptidyl-prolyl isomerase protein [Arabidopsis thaliana]	1e-077
TC125804	0	3	1	0	0	0.001096	putative protein [Arabidopsis thaliana]	4e-021
TC120494	0	3	1	0	0	0.001096	SOL1 protein [Arabidopsis thaliana]	4e-066
TC115723	19	3	3	11	11	0.001174	senescence-associated protein, putative [Arabidopsis thaliana]	2e-078
TC124128	0	2	6	3	0	0.001178	triose phosphate/phosphate translocator, chloroplast precursor	0.0
TC117567	0	0	1	2	4	0.001179	PGPD14 protein [Arabidopsis thaliana]	1e-092
TC116225	1	3	9	9	0	0.001278	putative ABC transporter [Oryza sativa (japonica cultivar-group)]	0.0
TC124933	0	0	2	4	5	0.001342	No hits found	
TC116327	0	0	0	0	2	0.001357	l-aminocyclopropane-1-carboxylate oxidase homolog (protein E8)	e-105
TC118824	0	0	0	0	2	0.001357	20S proteasome alpha subunit B [Arabidopsis thaliana]	1e-087

TC121886	0	0	0	0	2	0.001357	36kDa-peroxisomal membrane protein [Arabidopsis thaliana]	4e-061
TC124489	0	0	0	0	2	0.001357	36kDa-peroxisomal membrane protein [Arabidopsis thaliana]	1e-060
TC118563	0	0	0	0	2	0.001357	ADP-ribosylation factor 3 [Arabidopsis thaliana]	2e-087
TC121211	0	0	0	0	2	0.001357	amidase [Arabidopsis thaliana]	2e-050
TC121507	0	0	0	0	2	0.001357	ATPase [Solanum tuberosum]	2e-095
TC126510	0	0	0	0	2	0.001357	AtRer1A [Arabidopsis thaliana]	1e-069
TC128713	0	0	0	0	2	0.001357	auxin response transcription factor 3; ARF3 [Arabidopsis thaliana]	5e-014
TC131305	0	0	0	0	2	0.001357	Catalase	e-105
TC118918	0	0	0	0	2	0.001357	CCR4-associated factor 1-like protein [Arabidopsis thaliana]	7e-090
TC117446	0	0	0	0	2	0.001357	chloroplast import-associated channel protein homolog [Arabidopsis thaliana]	3e-047
TC120620	0	0	0	0	2	0.001357	copia-like retrotransposable element [Arabidopsis thaliana]	1e-036
TC120139	0	0	0	0	2	0.001357	Cystinosin homolog	4e-050
TC125092	0	0	0	0	2	0.001357	dihydroxyacid dehydratase, putative [Arabidopsis thaliana]	5e-061
TC125378	0	0	0	0	2	0.001357	elongation factor TU, chloroplast precursor	e-119
TC129477	0	0	0	0	2	0.001357	esterase/lipase/thioesterase family [Arabidopsis thaliana]	1e-053
TC120357	0	0	0	0	2	0.001357	ethylene-responsive protein -related [Arabidopsis thaliana]	1e-022
TC118098	0	0	0	0	2	0.001357	expressed protein [Arabidopsis thaliana]	6e-055
TC118982	0	0	0	0	2	0.001357	FKBP-like [Arabidopsis thaliana]	e-102
TC126084	0	0	0	0	2	0.001357	glutaredoxin, putative [Arabidopsis thaliana]	5e-023
TC122862	0	0	0	0	2	0.001357	glycotransferase family 19 [Arabidopsis thaliana]	3e-045
TC127299	0	0	0	0	2	0.001357	inhibitor of apoptosis-like protein [Arabidopsis thaliana]	6e-026
TC118538	0	0	0	0	2	0.001357	membrane associated protein [Arabidopsis thaliana]	7e-076
TC131072	0	0	0	0	2	0.001357	multidrug resistance P-glycoprotein, putative [Arabidopsis thaliana]	6e-034
TC122153	0	0	0	0	2	0.001357	NADP-specific glutamate dehydrogenase, putative [Arabidopsis thaliana]	e-114
TC118655	0	0	0	0	2	0.001357	No hits found	
TC118788	0	0	0	0	2	0.001357	No hits found	
TC120045	0	0	0	0	2	0.001357	No hits found	



TC120326	0	0	0	0	2	0.001357	No hits found	
TC120338	0	0	0	0	2	0.001357	No hits found	
TC120749	0	0	0	0	2	0.001357	No hits found	
TC122071	0	0	0	0	2	0.001357	No hits found	
TC122293	0	0	0	0	2	0.001357	No hits found	
TC122560	0	0	0	0	2	0.001357	No hits found	
TC122676	0	0	0	0	2	0.001357	No hits found	
TC122824	0	0	0	0	2	0.001357	No hits found	
TC122959	0	0	0	0	2	0.001357	No hits found	
TC123044	0	0	0	0	2	0.001357	No hits found	
TC123151	0	0	0	0	2	0.001357	No hits found	
TC123351	0	0	0	0	2	0.001357	No hits found	
TC126553	0	0	0	0	2	0.001357	No hits found	
TC130519	0	0	0	0	2	0.001357	No hits found	
TC130813	0	0	0	0	2	0.001357	No hits found	
TC131383	0	0	0	0	2	0.001357	No hits found	
TC131415	0	0	0	0	2	0.001357	No hits found	
TC118837	0	0	0	0	2	0.001357	P0638D12.11 [Oryza sativa (japonica cultivar-group)]	2e-011
TC125995	0	0	0	0	2	0.001357	P0678F11.22 [Oryza sativa (japonica cultivar-group)]	4e-015
TC121563	0	0	0	0	2	0.001357	Photosystem II 10 kDa polypeptide, chloroplast precursor	2e-032
TC122986	0	0	0	0	2	0.001357	POZ domain protein family [Arabidopsis thaliana]	1e-025
TC122987	0	0	0	0	2	0.001357	protein F14N23.29 - Arabidopsis thaliana	2e-015
TC122640	0	0	0	0	2	0.001357	protein phosphatase 2C [Arabidopsis thaliana]	2e-064
TC122867	0	0	0	0	2	0.001357	putative actin-related complex protein [Oryza sativa (japonica cultivar-group)]	2e-015
TC128185	0	0	0	0	2	0.001357	putative ankyrin [Arabidopsis thaliana]	2e-043
TC122544	0	0	0	0	2	0.001357	putative mitochondrial carrier protein [Arabidopsis thaliana]	3e-016
TC117080	0	0	0	0	2	0.001357	putative protein [Arabidopsis thaliana]	6e-087

TC118682	0	0	0	0	2	0.001357	putative protein [Arabidopsis thaliana]	3e-083
TC119914	0	0	0	0	2	0.001357	putative protein [Arabidopsis thaliana]	2e-050
TC123522	0	0	0	0	2	0.001357	putative protein [Arabidopsis thaliana]	5e-070
TC123437	0	0	0	0	2	0.001357	putative ribosomal protein [Oryza sativa (japonica cultivar-group)]	1e-011
TC119582	0	0	0	0	2	0.001357	putative salt-inducible protein [Arabidopsis thaliana]	1e-074
TC116343	0	0	0	0	2	0.001357	putative steroid binding protein [Arabidopsis thaliana]	2e-037
TC121620	0	0	0	0	2	0.001357	putative uridine kinase [Oryza sativa (japonica cultivar-group)]	2e-021
TC126699	0	0	0	0	2	0.001357	putative WD-repeat membrane protein [Arabidopsis thaliana]	9e-099
TC120490	0	0	0	0	2	0.001357	RPS6-related protein [Arabidopsis thaliana]	5e-053
TC120758	0	0	0	0	2	0.001357	serine C-palmitoyltransferase like protein [Arabidopsis thaliana]	e-118
TC117943	0	0	0	0	2	0.001357	serine-rich protein [Arabidopsis thaliana]	9e-041
TC117391	0	0	0	0	2	0.001357	similar to non-phototropic hypocotyl-like protein [Arabidopsis thaliana]	e-121
TC118472	0	0	0	0	2	0.001357	similar to PIG-L [Arabidopsis thaliana]	9e-029
TC117662	0	0	0	0	2	0.001357	small nuclear ribonucleoprotein, putative [Arabidopsis thaliana]	1e-048
TC130652	0	0	0	0	2	0.001357	somatic embryogenesis receptor-like kinase -like protein [Arabidopsis thaliana]	3e-026
TC119875	0	0	0	0	2	0.001357	somatic embryogenesis receptor-like kinase, putative [Arabidopsis thaliana]	3e-080
TC120560	0	0	0	0	2	0.001357	thylakoid lumenal 17.9 kDa protein, chloroplast precursor [Arabidopsis thaliana]	5e-052
TC117794	0	0	0	0	2	0.001357	unknown [Arabidopsis thaliana]	2e-089
TC130025	0	0	0	0	2	0.001357	unknown [Arabidopsis thaliana]	6e-033
TC118873	0	0	0	0	2	0.001357	WRKY family transcription factor [Arabidopsis thaliana]	3e-039
TC127188	0	0	0	0	2	0.001357	zinc finger protein homolog - Arabidopsis thaliana	2e-074
TC116373	5	1	2	28	3	0.001399	unknown protein [Arabidopsis thaliana]	4e-097
TC116358	0	1	3	21	4	0.001496	Probable vacuolar ATP synthase subunit d 2	0.0
TC124931	0	2	1	1	4	0.001505	bHLH protein [Arabidopsis thaliana]	5e-074
TC124182	0	3	8	9	0	0.001507	cytochrome c, putative [Arabidopsis thaliana]	e-144
TC115943	0	0	0	14	2	0.001528	acyltransferase 2 [Capsicum chinense]	2e-063
TC125054	0	0	5	3	0	0.001542	transcriptional coactivator-like protein [Arabidopsis thaliana]	5e-046

TC116886	2	5	2	1	0	0.001587	Kelch repeat-containing protein [Arabidopsis thaliana]	e-165
TC124647	1	3	8	4	4	0.001594	unknown protein [Arabidopsis thaliana]	8e-034
TC116984	6	0	0	0	0	0.001608	4-coumarate--CoA ligase 2	e-145
TC125659	6	0	0	0	0	0.001608	50S ribosomal protein L4 - Arabidopsis thaliana	2e-033
TC118651	6	0	0	0	0	0.001608	acetyl-CoA:benzylalcohol acetyltransferase [Clarkia concinna]	4e-029
TC118390	6	0	0	0	0	0.001608	C2 domain-containing protein [Arabidopsis thaliana]	3e-080
TC117207	6	0	0	0	0	0.001608	cysteine protease [Solanum tuberosum]	3e-077
TC118943	6	0	0	0	0	0.001608	Cytochrome P450 71D7	4e-047
TC123952	6	0	0	0	0	0.001608	E2, ubiquitin-conjugating enzyme, putative [Arabidopsis thaliana]	1e-082
TC127115	6	0	0	0	0	0.001608	equilibrative nucleoside transporter ENT3 -related [Arabidopsis thaliana]	7e-087
TC126786	6	0	0	0	0	0.001608	F16A14.19 [Arabidopsis thaliana]	8e-039
TC126875	6	0	0	0	0	0.001608	gibberellin 3-oxidase 1 [Nicotiana sylvestris]	4e-062
TC127166	6	0	0	0	0	0.001608	hypothetical protein [Arabidopsis thaliana]	1e-030
TC117225	6	0	0	0	0	0.001608	lipase, putative [Arabidopsis thaliana]	e-110
TC116723	6	0	0	0	0	0.001608	MADS-box protein homolog DEF4 - potato	e-120
TC118562	6	0	0	0	0	0.001608	membrane protein PTM1-like [Oryza sativa (japonica cultivar-group)]	2e-047
TC119138	6	0	0	0	0	0.001608	Mutator-like transposase [Arabidopsis thaliana]	1e-023
TC116289	6	0	0	0	0	0.001608	No apical meristem (NAM) protein family [Arabidopsis thaliana]	1e-055
TC124876	6	0	0	0	0	0.001608	No apical meristem (NAM) protein family [Arabidopsis thaliana]	1e-052
TC119131	6	0	0	0	0	0.001608	No hits found	
TC119139	6	0	0	0	0	0.001608	No hits found	
TC119211	6	0	0	0	0	0.001608	No hits found	
TC126487	6	0	0	0	0	0.001608	No hits found	
TC126495	6	0	0	0	0	0.001608	No hits found	
TC126542	6	0	0	0	0	0.001608	No hits found	
TC126891	6	0	0	0	0	0.001608	No hits found	
TC116614	6	0	0	0	0	0.001608	peroxidase [Arabidopsis thaliana]	3e-011

TC118835	6	0	0	0	0	0.001608	phosphoglycerate mutase-like protein [Arabidopsis thaliana]	5e-033
TC118941	6	0	0	0	0	0.001608	potassium transport protein-like [Arabidopsis thaliana]	2e-040
TC119149	6	0	0	0	0	0.001608	probable mitotic control protein dis3 - Arabidopsis thaliana	3e-032
TC126818	6	0	0	0	0	0.001608	probable UDP-glucuronosyltransferase - garden pea	5e-076
TC119039	6	0	0	0	0	0.001608	protein kinase-like protein [Arabidopsis thaliana]	1e-022
TC126588	6	0	0	0	0	0.001608	putative cytoskeletal protein [Arabidopsis thaliana]	1e-034
TC126180	6	0	0	0	0	0.001608	putative inorganic polyphosphate/ATP-NAD kinase [Oryza sativa (japonica cultivar-group)]	e-100
TC125355	6	0	0	0	0	0.001608	putative phosphatidylinositol-4-phosphate 5-kinase [Arabidopsis thaliana]	2e-073
TC118618	6	0	0	0	0	0.001608	putative protein [Arabidopsis thaliana]	6e-015
TC119181	6	0	0	0	0	0.001608	putative receptor-like serine-threonine protein kinase [Solanum tuberosum]	3e-076
TC126878	6	0	0	0	0	0.001608	RNA polymerase beta	0.0
TC126127	6	0	0	0	0	0.001608	tetratricopeptide repeat (TPR)-containing protein [Arabidopsis thaliana]	1e-037
TC126731	6	0	0	0	0	0.001608	Tryptophan synthase beta chain 2, chloroplast precursor	1e-092
TC127169	6	0	0	0	0	0.001608	unknown [Arabidopsis thaliana]	8e-037
TC118532	6	0	0	0	0	0.001608	zinc finger (C3HC4-type RING finger) protein family [Arabidopsis thaliana]	4e-030
TC116996	0	4	6	3	1	0.00164	unknown protein [Arabidopsis thaliana]	3e-031
TC117293	7	0	0	1	4	0.001661	putative DNA-binding protein [Arabidopsis thaliana]	1e-051
TC124859	0	2	0	14	0	0.001673	lipase (class 3) family [Arabidopsis thaliana]	e-174
TC117149	7	3	0	0	0	0.001689	putative chaperonin [Arabidopsis thaliana]	0.0
TC116516	0	1	0	13	0	0.001731	peroxiredoxin - like protein [Arabidopsis thaliana]	6e-066
TC116971	1	0	7	7	0	0.00175	CBL-interacting protein kinase 25 [Arabidopsis thaliana]	e-114
TC116432	0	3	2	0	0	0.001799	14-3-3 protein 6	e-142
TC125002	0	3	2	0	0	0.001799	heterogeneous nuclear ribonucleoprotein (hnRNP), putative [Arabidopsis thaliana]	0.0
TC117275	0	3	2	0	0	0.001799	KH domain protein [Arabidopsis thaliana]	e-125
TC125052	0	3	2	0	0	0.001799	putative serine carboxypeptidase II [Arabidopsis thaliana]	e-121
TC124985	0	6	1	6	1	0.001833	GTP-binding protein - related [Arabidopsis thaliana]	0.0
TC116823	0	3	0	1	0	0.00186	26S proteasome AAA-ATPase subunit RPT1a [Arabidopsis thaliana]	0.0

TC127480	0	3	0	1	0	0.00186	CTD phosphatase-like 3 [Arabidopsis thaliana]	e-105
TC127051	0	3	0	1	0	0.00186	hypothetical protein [Oryza sativa (japonica cultivar-group)]	6e-035
TC118004	0	3	0	1	0	0.00186	putative RING protein [Populus x canescens]	5e-036
TC116402	1	5	3	1	1	0.001868	ATP synthase gamma chain, mitochondrial precursor	e-154
TC125118	0	1	5	2	0	0.001885	expressed protein [Arabidopsis thaliana]	e-113
TC117067	7	1	0	0	0	0.001903	OSJNBa0086B14.2 [Oryza sativa (japonica cultivar-group)]	1e-085
TC117638	7	1	0	0	0	0.001903	type 5 serine/threonine phosphatase 55 kDa isoform [Lycopersicon esculentum]	0.0
TC118218	7	1	0	0	0	0.001903	ZPT2-13 [Petunia x hybrida]	2e-042
TC124271	1	2	5	6	7	0.001946	GTP-binding nuclear protein RAN1	e-127
TC124565	9	0	2	1	0	0.001962	F-box containing tubby family protein [Arabidopsis thaliana]	e-178
TC117431	0	0	0	11	0	0.002039	FRO2-like protein; NADPH oxidase-like [Arabidopsis thaliana]	e-130
TC124704	0	0	0	11	0	0.002039	unknown [Arabidopsis thaliana]	8e-054
TC124669	8	1	0	0	1	0.002048	small GTP-binding protein	e-111
TC116994	0	0	0	4	4	0.002058	auxin response factor-like protein [Mangifera indica]	e-174
TC125904	0	0	0	4	4	0.002058	MRP-like ABC transporter [Oryza sativa (japonica cultivar-group)]	0.0
TC117557	0	0	0	4	4	0.002058	O-diphenol-O-methyl transferase, putative [Arabidopsis thaliana]	e-138
TC124936	0	5	2	2	2	0.002098	DegP protease, putative [Arabidopsis thaliana]	0.0
TC117152	8	0	1	0	1	0.002223	DNA-directed RNA polymerase II, third largest subunit [Arabidopsis thaliana]	e-149
TC124323	0	0	5	2	2	0.00225	histidine decarboxylase, putative [Arabidopsis thaliana]	0.0
TC125227	1	0	0	6	5	0.002368	putative protein [Arabidopsis thaliana]	4e-083
TC123783	2	5	0	2	3	0.00241	putative ATP synthase [Arabidopsis thaliana]	5e-082
TC116883	0	0	2	15	5	0.00243	cytosolic ascorbate peroxidase [Fragaria x ananassa]	e-110
TC124644	4	0	2	1	5	0.002475	putative protein [Arabidopsis thaliana]	e-113
TC124429	9	4	0	2	0	0.002478	unknown protein [Arabidopsis thaliana]	1e-087
TC124698	0	1	2	7	6	0.002517	phosphatidylinositol 3- and 4-kinase family [Arabidopsis thaliana]	0.0
TC124496	1	5	0	3	1	0.002542	Ferritin 2, chloroplast precursor (NtFer2)	e-107
TC116467	2	3	3	22	0	0.00262	pinhead-related protein [Arabidopsis thaliana]	e-138

TC124827	0	3	0	3	4	0.002647	proline-rich protein family [Arabidopsis thaliana]	7e-047
TC115797	0	6	3	7	0	0.002648	No hits found	
TC116466	0	2	0	15	1	0.002697	polyamine oxidase, putative [Arabidopsis thaliana]	0.0
TC117135	9	0	0	3	0	0.002714	No hits found	
TC116237	7	0	0	1	0	0.002743	obtusifoliol-14-demethylase [Nicotiana tabacum]	e-149
TC117072	7	0	0	1	0	0.002743	purine permease family [Arabidopsis thaliana]	4e-041
TC125374	7	0	0	1	0	0.002743	Zinc finger protein constans-like 15	7e-079
TC124903	0	0	0	13	2	0.002767	heat shock protein, 70K, chloroplast - cucumber	0.0
TC124219	0	2	0	0	0	0.002789	ACTIN 97	0.0
TC125251	0	2	0	0	0	0.002789	ATP-dependent Clp proteinase catalytic chain P4 - Arabidopsis thaliana	2e-086
TC122459	0	2	0	0	0	0.002789	BAC19.10 [Lycopersicon esculentum]	6e-037
TC118322	0	2	0	0	0	0.002789	calmodulin-like protein [Arabidopsis thaliana]	1e-029
TC126604	0	2	0	0	0	0.002789	contains similarity to vacuolar protein sorting-associated protein VPS28 [Arabidopsis thaliana]	7e-087
TC131037	0	2	0	0	0	0.002789	copine-related [Arabidopsis thaliana]	4e-021
TC129124	0	2	0	0	0	0.002789	cullin-like protein 1 [Pisum sativum]	5e-036
TC118649	0	2	0	0	0	0.002789	Cytochrome P450 72A1	e-108
TC125510	0	2	0	0	0	0.002789	cytosolic tRNA-Ala synthetase - Arabidopsis thaliana	e-179
TC126956	0	2	0	0	0	0.002789	Dihydrodipicolinate synthase, chloroplast precursor (DHDPS)	e-117
TC129054	0	2	0	0	0	0.002789	dihydrofolate synthetase /folylpolyglutamate synthetase [Arabidopsis thaliana]	2e-041
TC118631	0	2	0	0	0	0.002789	expressed protein [Arabidopsis thaliana]	1e-054
TC124907	0	2	0	0	0	0.002789	expressed protein [Arabidopsis thaliana]	e-144
TC126095	0	2	0	0	0	0.002789	F-box protein family [Arabidopsis thaliana]	1e-076
TC116410	0	2	0	0	0	0.002789	Ferredoxin--NADP reductase, root-type isozyme, chloroplast precursor (FNR)	9e-011
TC124955	0	2	0	0	0	0.002789	fructose-1,6-bisphosphatase precursor [Solanum tuberosum]	e-128
TC126237	0	2	0	0	0	0.002789	galactose-1-phosphate uridyl transferase-like protein [Arabidopsis thaliana]	e-109
TC119991	0	2	0	0	0	0.002789	G-box binding factor 4, GBF4 [Arabidopsis thaliana]	2e-031
TC127697	0	2	0	0	0	0.002789	geranylgeranyl transferase type II beta subunit, putative [Arabidopsis thaliana]	1e-084

TC123477	0	2	0	0	0	0.002789	glycosyl hydrolase family 3 [Arabidopsis thaliana]	8e-061
TC127284	0	2	0	0	0	0.002789	H+/Ca2+ exchanger 2 [Ipomoea nil]	e-112
TC117998	0	2	0	0	0	0.002789	high mobility group protein 2-like [Arabidopsis thaliana]	2e-019
TC116470	0	2	0	0	0	0.002789	hypersensitive-induced response protein [Arabidopsis thaliana]	e-141
TC119334	0	2	0	0	0	0.002789	hypothetical protein [Arabidopsis thaliana]	e-122
TC128506	0	2	0	0	0	0.002789	hypothetical protein [Arabidopsis thaliana]	2e-041
TC120121	0	2	0	0	0	0.002789	hypothetical protein F10N7.130 - Arabidopsis thaliana	e-112
TC120365	0	2	0	0	0	0.002789	inositol phosphate kinase [Zea mays]	4e-035
TC117316	0	2	0	0	0	0.002789	JAB [Lycopersicon esculentum]	8e-074
TC131043	0	2	0	0	0	0.002789	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	2e-053
TC116242	0	2	0	0	0	0.002789	light-harvesting chlorophyll a/b binding protein [Arabidopsis thaliana]	7e-072
TC127931	0	2	0	0	0	0.002789	lipoic acid synthase-like protein [Arabidopsis thaliana]	e-102
TC122568	0	2	0	0	0	0.002789	MATE efflux protein family [Arabidopsis thaliana]	5e-030
TC119415	0	2	0	0	0	0.002789	membrane protein PTM1-like [Oryza sativa (japonica cultivar-group)]	6e-073
TC116254	0	2	0	0	0	0.002789	mitochondrial ATPase beta subunit [Nicotiana sylvestris]	0.0
TC126429	0	2	0	0	0	0.002789	MTD1 [Medicago truncatula]	4e-025
TC123762	0	2	0	0	0	0.002789	NADH-ubiquinone oxidoreductase 23 kDa subunit, mitochondrial precursor	e-114
TC116952	0	2	0	0	0	0.002789	No hits found	
TC117102	0	2	0	0	0	0.002789	No hits found	
TC118629	0	2	0	0	0	0.002789	No hits found	
TC119116	0	2	0	0	0	0.002789	No hits found	
TC120875	0	2	0	0	0	0.002789	No hits found	
TC121060	0	2	0	0	0	0.002789	No hits found	
TC122555	0	2	0	0	0	0.002789	No hits found	
TC122570	0	2	0	0	0	0.002789	No hits found	
TC122686	0	2	0	0	0	0.002789	No hits found	
TC123490	0	2	0	0	0	0.002789	No hits found	

TC128013	0	2	0	0	0	0.002789	No hits found	
TC128440	0	2	0	0	0	0.002789	No hits found	
TC128743	0	2	0	0	0	0.002789	No hits found	
TC129285	0	2	0	0	0	0.002789	No hits found	
TC129956	0	2	0	0	0	0.002789	No hits found	
TC130412	0	2	0	0	0	0.002789	No hits found	
TC131105	0	2	0	0	0	0.002789	No hits found	
TC131276	0	2	0	0	0	0.002789	No hits found	
TC131587	0	2	0	0	0	0.002789	No hits found	
TC125446	0	2	0	0	0	0.002789	ORF; able to induce HR-like lesions [ <i>Nicotiana tabacum</i> ]	2e-057
TC116021	0	2	0	0	0	0.002789	Oxygen-evolving enhancer protein 1, chloroplast precursor	2e-076
TC126075	0	2	0	0	0	0.002789	pectate lyase [ <i>Salix gilgiana</i> ]	e-112
TC117377	0	2	0	0	0	0.002789	pectinesterase [ <i>Citrus sinensis</i> ]	e-122
TC129521	0	2	0	0	0	0.002789	PHAP2A protein [ <i>Petunia x hybrida</i> ]	2e-040
TC124218	0	2	0	0	0	0.002789	phosphoprotein phosphatase 2A regulatory chain - common tobacco	1e-088
TC126931	0	2	0	0	0	0.002789	phosphoribosyl pyrophosphate synthase isozyme 3 [ <i>Spinacia oleracea</i> ]	e-134
TC124235	0	2	0	0	0	0.002789	pollen-specific protein homolog - tomato	0.0
TC116533	0	2	0	0	0	0.002789	porin-like protein [ <i>Arabidopsis thaliana</i> ]	5e-057
TC118696	0	2	0	0	0	0.002789	probable polygalacturonase (EC 3.2.1.15) 1 - tomato	0.0
TC129297	0	2	0	0	0	0.002789	probable transcription factor F12K22.14 - <i>Arabidopsis thaliana</i>	5e-091
TC125687	0	2	0	0	0	0.002789	profilin [ <i>Lycopersicon esculentum</i> ]	6e-072
TC117144	0	2	0	0	0	0.002789	proline-rich protein [ <i>Arabidopsis thaliana</i> ]	3e-038
TC125107	0	2	0	0	0	0.002789	Proteasome subunit alpha type 4	e-122
TC116734	0	2	0	0	0	0.002789	Proteasome subunit beta type 5 precursor	e-100
TC120748	0	2	0	0	0	0.002789	putative acyl-CoA synthetase [ <i>Arabidopsis thaliana</i> ]	2e-071
TC120501	0	2	0	0	0	0.002789	putative arginine methyltransferase [ <i>Arabidopsis thaliana</i> ]	e-129
TC127100	0	2	0	0	0	0.002789	putative BSD domain containing protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	6e-031



TC129198	0	2	0	0	0	0.002789	putative CAAX prenyl protease [Arabidopsis thaliana]	e-135
TC129986	0	2	0	0	0	0.002789	putative cellulose synthase [Arabidopsis thaliana]	1e-083
TC123958	0	2	0	0	0	0.002789	putative chloroplast thiazole biosynthetic protein [Nicotiana tabacum]	1e-052
TC118457	0	2	0	0	0	0.002789	putative elongation factor P (EF-P) [Arabidopsis thaliana]	2e-080
TC119291	0	2	0	0	0	0.002789	putative polyprotein [Oryza sativa (japonica cultivar-group)]	1e-014
TC120095	0	2	0	0	0	0.002789	putative protein [Arabidopsis thaliana]	8e-057
TC125954	0	2	0	0	0	0.002789	putative protein [Arabidopsis thaliana]	e-100
TC130604	0	2	0	0	0	0.002789	putative ternary complex factor MIP1 [Antirrhinum majus]	2e-017
TC123847	0	2	0	0	0	0.002789	Putative transporter [Oryza sativa (japonica cultivar-group)]	0.0
TC128410	0	2	0	0	0	0.002789	putative zinc finger protein [Oryza sativa (japonica cultivar-group)]	3e-095
TC127690	0	2	0	0	0	0.002789	putative zinc finger protein MA23 [Mus musculus]	1e-034
TC118505	0	2	0	0	0	0.002789	Ras-related protein Rab11D	e-116
TC115851	0	2	0	0	0	0.002789	ripening protein E8 homolog - tomato	1e-086
TC128249	0	2	0	0	0	0.002789	SET-domain transcriptional regulator family [Arabidopsis thaliana]	2e-095
TC128656	0	2	0	0	0	0.002789	similar to glycosyl hydrolase family 3 [Arabidopsis thaliana]	e-147
TC119260	0	2	0	0	0	0.002789	storage protein, putative [Arabidopsis thaliana]	9e-024
TC117424	0	2	0	0	0	0.002789	strictosidine synthase-related [Arabidopsis thaliana]	e-114
TC127421	0	2	0	0	0	0.002789	succinic semialdehyde dehydrogenase (gabD), putative [Arabidopsis thaliana]	9e-061
TC117898	0	2	0	0	0	0.002789	sulfate adenylyltransferase (EC 2.7.7.4) met3-2 - potato	e-179
TC127772	0	2	0	0	0	0.002789	T27G7.21 [Arabidopsis thaliana]	9e-048
TC116392	0	2	0	0	0	0.002789	Thioredoxin H-type 1 (TRX-H1)	2e-052
TC123732	0	2	0	0	0	0.002789	Tic62 protein [Pisum sativum]	e-127
TC130464	0	2	0	0	0	0.002789	tospovirus resistance protein B [Lycopersicon esculentum]	1e-026
TC130409	0	2	0	0	0	0.002789	transcription factor bZIP61, pelota-related protein [Arabidopsis thaliana]	4e-051
TC119145	0	2	0	0	0	0.002789	transcription factor JERF2 [Lycopersicon esculentum]	2e-036
TC128951	0	2	0	0	0	0.002789	transducin / WD-40 repeat protein family [Arabidopsis thaliana]	2e-069
TC126713	0	2	0	0	0	0.002789	transducin / WD-40 repeat protein family [Arabidopsis thaliana]	1e-084

TC128309	0	2	0	0	0	0.002789	transducin / WD-40 repeat protein family [Arabidopsis thaliana]	e-121
TC121980	0	2	0	0	0	0.002789	transducin / WD-40 repeat protein family [Arabidopsis thaliana]	2e-081
TC118850	0	2	0	0	0	0.002789	translin-like protein [Arabidopsis thaliana]	2e-098
TC116498	0	2	0	0	0	0.002789	tuberisation-related protein - potato	e-105
TC120964	0	2	0	0	0	0.002789	unknown [Arabidopsis thaliana]	1e-055
TC121151	0	2	0	0	0	0.002789	unknown [Arabidopsis thaliana]	3e-045
TC126124	0	2	0	0	0	0.002789	unknown [Arabidopsis thaliana]	4e-042
TC125249	0	2	0	0	0	0.002789	unknown protein [Arabidopsis thaliana]	1e-059
TC126733	0	2	0	0	0	0.002789	unknown protein [Arabidopsis thaliana]	1e-029
TC128052	0	2	0	0	0	0.002789	unknown protein [Arabidopsis thaliana]	9e-021
TC130534	0	2	0	0	0	0.002789	unknown protein [Arabidopsis thaliana]	8e-035
TC115967	0	2	0	0	0	0.002789	unknown protein [Oryza sativa]	4e-027
TC117147	0	2	0	0	0	0.002789	vacuolar H <sup>+</sup> -ATPase A1 subunit isoform [Lycopersicon esculentum]	0.0
TC119671	0	2	0	0	0	0.002789	zinc finger (C3HC4-type RING finger) protein family [Arabidopsis thaliana]	5e-011
TC128943	0	2	0	0	0	0.002789	zinc finger protein [Arabidopsis thaliana]	3e-062
TC125328	8	1	0	1	0	0.002844	glucosyltransferase NTGT2 [Nicotiana tabacum]	2e-099
TC116072	0	7	10	22	5	0.002875	T29M8.12 protein - Arabidopsis thaliana	5e-072
TC125663	0	1	1	0	3	0.002916	putative protein [Arabidopsis thaliana]	1e-093
TC118070	8	0	0	2	0	0.003038	No hits found	
TC125069	0	0	3	0	2	0.003069	Ntdin [Nicotiana tabacum]	2e-077
TC118027	8	0	1	1	0	0.003086	UDP-glucose 4-epimerase GEPI48	e-130
TC125547	2	0	0	0	3	0.003111	metacaspase 1 [Lycopersicon esculentum]	0.0
TC117406	2	0	4	0	0	0.003113	invertase inhibitor precursor - tomato	3e-083
TC124317	3	5	4	0	2	0.003221	40S ribosomal protein S8	1e-075
TC124454	4	0	0	1	4	0.003285	membrane related protein CP5, putative [Arabidopsis thaliana]	e-140
TC124742	1	4	0	2	0	0.003325	putative protein [Arabidopsis thaliana]	6e-068
TC124749	0	1	0	3	4	0.003332	protease-related [Arabidopsis thaliana]	e-176

TC124620	1	1	2	5	6	0.003338	Xaa-Pro aminopeptidase 2 [ <i>Lycopersicon esculentum</i> ]	0.0
TC123861	3	5	2	1	0	0.00342	Phospho-2-dehydro-3-deoxyheptonate aldolase 1, chloroplast precursor	0.0
TC125213	0	0	1	12	0	0.003552	thioredoxin-like protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	3e-062
TC117299	0	0	1	3	4	0.00369	coded for by <i>A. thaliana</i> cDNA Z18000 [ <i>Arabidopsis thaliana</i> ]	1e-072
TC124393	0	0	1	3	4	0.00369	unknown [ <i>Arabidopsis thaliana</i> ]	9e-065
TC123995	3	5	3	1	0	0.003747	60S RIBOSOMAL PROTEIN L5	e-114
TC117186	1	1	4	0	0	0.003777	dihydrolipoamide S-acetyltransferase precursor - <i>Arabidopsis thaliana</i>	e-122
TC124739	3	4	1	0	0	0.003809	pectinesterase (pectin methylesterase) family [ <i>Arabidopsis thaliana</i> ]	0.0
TC116573	0	3	1	0	2	0.003871	probable ubiquitin-protein ligase (EC 6.3.2.19) - common tobacco	0.0
TC124444	0	3	3	0	1	0.003904	aldehyde dehydrogenase, putative [ <i>Arabidopsis thaliana</i> ]	e-179
TC124240	0	0	0	10	0	0.004036	endo-beta-mannanase [ <i>Lycopersicon esculentum</i> ]	1e-059
TC117185	0	0	0	10	0	0.004036	isoflavone reductase homolog	e-144
TC124708	0	0	0	10	0	0.004036	No hits found	
TC115858	0	2	3	0	0	0.004083	60S ribosomal protein L21 [ <i>Oryza sativa</i> (japonica cultivar-group)]	2e-078
TC124334	0	2	3	0	0	0.004083	expressed protein [ <i>Arabidopsis thaliana</i> ]	3e-017
TC117219	0	2	3	0	0	0.004083	Nonspecific lipid-transfer protein 1 precursor (LTP 1)	2e-038
TC116481	0	2	3	0	0	0.004083	putative auxin growth promotor protein [ <i>Lycopersicon esculentum</i> ]	e-113
TC124083	9	1	0	1	2	0.004207	peroxidase [ <i>Nicotiana tabacum</i> ]	e-128
TC124894	10	0	0	5	0	0.004232	putative protein phosphatase [ <i>Arabidopsis thaliana</i> ]	e-141
TC116911	0	5	0	8	4	0.004256	No hits found	
TC116001	0	4	1	2	3	0.004324	mitochondrial processing peptidase [ <i>Solanum tuberosum</i> ]	0.0
TC124028	0	3	9	20	5	0.004354	ubiquinol-cytochrome-c reductase Rieske iron-sulfur protein - potato	e-128
TC125352	0	1	0	1	3	0.004428	Ferrochelatase II, chloroplast precursor	e-167
TC125252	0	1	0	1	3	0.004428	putative protein kinase [ <i>Oryza sativa</i> (japonica cultivar-group)]	0.0
TC125405	0	1	1	2	4	0.004498	F-box containing tubby family protein [ <i>Arabidopsis thaliana</i> ]	e-173
TC125157	1	0	1	0	3	0.004617	unknown [ <i>Arabidopsis thaliana</i> ]	3e-068
TC116076	0	0	4	2	0	0.004661	glycolate oxidase, putative [ <i>Arabidopsis</i> ]	0.0

TC126675	0	2	0	0	2	0.004734	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	e-103
TC120254	0	2	0	0	2	0.004734	SABRE [Arabidopsis thaliana]	1e-075
TC116319	0	1	3	18	2	0.004772	Hsp20.0 protein [Lycopersicon peruvianum]	4e-065
TC116339	0	5	1	8	5	0.004908	GTP-binding nuclear protein RAN2	e-116
TC118238	0	0	0	2	3	0.005048	expressed protein [Arabidopsis thaliana]	2e-080
TC117239	0	0	0	2	3	0.005048	phosphoenolpyruvate carboxylase (EC 4.1.1.31) 1 - upland cotton	0.0
TC117287	0	0	0	2	3	0.005048	phosphoenolpyruvate carboxylase kinase 2 [Lycopersicon esculentum]	e-159
TC118163	0	0	0	2	3	0.005048	Probable methionyl-tRNA synthetase	e-180
TC116200	0	0	0	2	3	0.005048	putative dehydrogenase [Arabidopsis thaliana]	8e-036
TC117564	0	0	0	2	3	0.005048	RNase NGR2 [Nicotiana glutinosa]	e-123
TC118475	0	0	3	0	1	0.005174	PWWP domain protein [Arabidopsis thaliana]	4e-020
TC125348	0	0	3	0	1	0.005174	zinc finger - like protein [Arabidopsis thaliana]	8e-039
TC118969	0	0	1	1	3	0.005209	DnaJ protein homolog - Salix gilgiana	1e-096
TC118786	0	0	1	1	3	0.005209	MORF-related gene 15 [Homo sapiens]	9e-038
TC123921	0	0	0	8	4	0.005337	Histidine decarboxylase (HDC) (TOM92)	1e-056
TC124909	3	0	0	0	3	0.005377	26S proteasome regulatory subunit (RPN11), putative [Arabidopsis thaliana]	e-150
TC116638	3	0	0	17	2	0.005382	beta-cyanoalanine synthase [Solanum tuberosum]	e-174
TC124330	18	1	3	11	1	0.0054	MADS-box protein [Lycopersicon esculentum]	e-123
TC125300	0	3	0	1	2	0.005475	unknown protein [Arabidopsis thaliana]	6e-035
TC126496	4	0	0	0	3	0.005507	No hits found	
TC118187	0	1	3	0	0	0.005653	flavin-containing monooxygenase (FMO) family [Arabidopsis thaliana]	1e-063
TC124655	0	1	3	0	0	0.005653	Glycine dehydrogenase [decarboxylating], mitochondrial precursor	0.0
TC125603	0	1	3	0	0	0.005653	mitochondrial carrier protein-like [Arabidopsis thaliana]	e-148
TC124354	0	1	3	0	0	0.005653	ribosomal protein L1 protein - Arabidopsis thaliana	e-113
TC126463	0	1	3	0	0	0.005653	RNA helicase-like protein DB10	e-172
TC124585	0	1	3	0	0	0.005653	selenium binding protein [Medicago sativa]	e-176
TC116527	0	1	3	0	0	0.005653	serine hydroxymethyltransferase, mitochondrial precursor	0.0

TC124928	1	3	3	0	0	0.005695	predicted GPI-anchored protein [Arabidopsis thaliana]	e-159
TC124217	1	4	6	3	1	0.005768	phosphoprotein phosphatase 2A regulatory chain - common tobacco	0.0
TC124201	5	0	0	0	0	0.005837	aldo/keto reductase family [Arabidopsis thaliana]	e-165
TC125277	5	0	0	0	0	0.005837	aux/IAA protein [Populus tremula x Populus tremuloides]	2e-054
TC119275	5	0	0	0	0	0.005837	auxin-induced protein family [Arabidopsis thaliana]	2e-025
TC127471	5	0	0	0	0	0.005837	auxin-induced protein family [Arabidopsis thaliana]	4e-015
TC116008	5	0	0	0	0	0.005837	caffeoyl-CoA O-methyltransferase [Solanum tuberosum]	e-134
TC118460	5	0	0	0	0	0.005837	expansin-related protein 1-like [Oryza sativa (japonica cultivar-group)]	3e-078
TC116812	5	0	0	0	0	0.005837	extensin homolog - potato	1e-070
TC125805	5	0	0	0	0	0.005837	Extracellular ribonuclease LE precursor (RNase LE)	e-135
TC117359	5	0	0	0	0	0.005837	glutathione peroxidase, putative [Arabidopsis thaliana]	1e-069
TC127636	5	0	0	0	0	0.005837	glycine-rich protein [Arabidopsis thaliana]	3e-034
TC119068	5	0	0	0	0	0.005837	glycosyltransferase family 1 [Arabidopsis thaliana]	1e-084
TC115824	5	0	0	0	0	0.005837	hydroxymethylglutaryl coenzyme A synthase [Hevea brasiliensis]	0.0
TC126312	5	0	0	0	0	0.005837	hypothetical protein [Arabidopsis thaliana]	3e-071
TC126163	5	0	0	0	0	0.005837	MutT/nudix family protein [Arabidopsis thaliana]	5e-099
TC126368	5	0	0	0	0	0.005837	myb-related protein 305 - garden snapdragon	7e-081
TC118787	5	0	0	0	0	0.005837	myb-related protein Ph2 - garden petunia	6e-059
TC117264	5	0	0	0	0	0.005837	MybSt1 [Solanum tuberosum]	0.0
TC116173	5	0	0	0	0	0.005837	No hits found	
TC119285	5	0	0	0	0	0.005837	No hits found	
TC119312	5	0	0	0	0	0.005837	No hits found	
TC119352	5	0	0	0	0	0.005837	No hits found	
TC119462	5	0	0	0	0	0.005837	No hits found	
TC119500	5	0	0	0	0	0.005837	No hits found	
TC119521	5	0	0	0	0	0.005837	No hits found	
TC124369	5	0	0	0	0	0.005837	No hits found	

TC126979	5	0	0	0	0	0.005837	No hits found	
TC127388	5	0	0	0	0	0.005837	No hits found	
TC127442	5	0	0	0	0	0.005837	No hits found	
TC127597	5	0	0	0	0	0.005837	No hits found	
TC127632	5	0	0	0	0	0.005837	No hits found	
TC117691	5	0	0	0	0	0.005837	nonclathrin coat protein zeta1-COP [ <i>Lycopersicon esculentum</i> ]	4e-090
TC119715	5	0	0	0	0	0.005837	nucellin-like aspartic protease [ <i>Zea mays</i> ]	8e-037
TC127472	5	0	0	0	0	0.005837	oj991113_30.18 [ <i>Oryza sativa</i> (japonica cultivar-group)]	3e-050
TC117458	5	0	0	0	0	0.005837	ornithine carbamoyltransferase; OCTase [ <i>Canavalia lineata</i> ]	8e-042
TC118609	5	0	0	0	0	0.005837	PHAP2A protein [ <i>Petunia x hybrida</i> ]	2e-089
TC126789	5	0	0	0	0	0.005837	phosphatidic acid phosphatase (PAP2)-related [ <i>Arabidopsis thaliana</i> ]	e-110
TC116472	5	0	0	0	0	0.005837	probable cytochrome P-450 - curled-leaved tobacco	e-171
TC125457	5	0	0	0	0	0.005837	protein kinase family [ <i>Arabidopsis thaliana</i> ]	e-151
TC126113	5	0	0	0	0	0.005837	protein kinase, putative [ <i>Arabidopsis thaliana</i> ]	2e-075
TC117732	5	0	0	0	0	0.005837	putative 7-transmembrane G-protein-coupled receptor [ <i>Solanum chacoense</i> ]	0.0
TC118007	5	0	0	0	0	0.005837	putative bHLH transcription factor bHLH027 [ <i>Arabidopsis thaliana</i> ]	1e-035
TC118131	5	0	0	0	0	0.005837	putative NAC domain protein [ <i>Solanum tuberosum</i> ]	e-149
TC117172	5	0	0	0	0	0.005837	putative peroxidase [ <i>Solanum tuberosum</i> ]	e-130
TC127351	5	0	0	0	0	0.005837	putative phosphatidylcholine acyltransferase [ <i>Medicago truncatula</i> ]	7e-071
TC118203	5	0	0	0	0	0.005837	putative protein [ <i>Arabidopsis thaliana</i> ]	4e-029
TC118662	5	0	0	0	0	0.005837	putative protein [ <i>Arabidopsis thaliana</i> ]	8e-025
TC127312	5	0	0	0	0	0.005837	putative protein [ <i>Arabidopsis thaliana</i> ]	8e-061
TC125809	5	0	0	0	0	0.005837	putative protein kinase [ <i>Arabidopsis thaliana</i> ]	e-163
TC127335	5	0	0	0	0	0.005837	Similar to exopolyphosphatase [ <i>Arabidopsis thaliana</i> ]	2e-019
TC119543	5	0	0	0	0	0.005837	transducin / WD-40 repeat protein family [ <i>Arabidopsis thaliana</i> ]	2e-028
TC125125	5	0	0	0	0	0.005837	tropinone reductase homolog	1e-089
TC125299	5	0	0	0	0	0.005837	Ubiquinol-cytochrome C reductase complex 7.8 kDa protein	6e-037

TC116124	5	0	0	0	0	0.005837	unknown [Arabidopsis thaliana]	3e-071
TC118938	5	0	0	0	0	0.005837	unknown protein [Arabidopsis thaliana]	7e-044
TC119232	5	0	0	0	0	0.005837	ZF-HD homeobox protein [Flaveria bidentis]	2e-021
TC126877	2	3	0	0	0	0.005903	GDP-L-fucose synthetase, putative [Arabidopsis thaliana]	5e-093
TC125029	2	3	0	0	0	0.005903	tuberisation-related protein - potato	9e-057
TC126612	2	3	0	0	0	0.005903	unknown [Solanum tuberosum]	7e-083
TC118554	6	0	0	0	1	0.005911	No hits found	
TC116601	6	0	0	0	1	0.005911	probable glucosyltransferase twi1 (EC 2.4.1.-) - tomato	0.0
TC116179	9	1	2	1	0	0.006023	cytoplasmic ribosomal protein S13 [Panax ginseng]	5e-077
TC125570	0	0	0	11	1	0.006171	No hits found	
TC125667	0	0	0	11	1	0.006171	orcinol O-methyltransferase [Rosa hybrid cultivar]	4e-091
TC124948	0	0	0	11	1	0.006171	putative protein [Arabidopsis thaliana]	3e-058
TC125629	0	0	0	11	1	0.006171	steroleosin [Sesamum indicum]	e-108
TC124673	0	3	3	1	0	0.006206	dehydration-induced protein RD22-like protein [Gossypium hirsutum]	2e-086
TC117825	6	1	0	0	0	0.006217	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	e-120
TC125189	6	1	0	0	0	0.006217	metal transporter family [Arabidopsis thaliana]	1e-056
TC117915	6	1	0	0	0	0.006217	S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase [Petunia x hybrida]	1e-082
TC125540	6	1	0	0	0	0.006217	trehalose-6-phosphate synthase, putative [Arabidopsis thaliana]	e-153
TC125767	6	1	0	0	0	0.006217	unknown protein [Oryza sativa (japonica cultivar-group)]	4e-019
TC116559	0	0	5	3	2	0.006226	putative auxin-repressed protein [Prunus armeniaca]	6e-027
TC116752	0	3	4	1	1	0.006235	putative esterase D [Arabidopsis thaliana]	e-130
TC125321	0	3	4	1	1	0.006235	unknown protein [Arabidopsis thaliana]	1e-078
TC116701	0	0	2	3	4	0.006263	putative SF16 protein [Arabidopsis thaliana]	6e-087
TC115873	2	5	3	26	5	0.006273	oxidoreductase, zinc-binding dehydrogenase family [Arabidopsis thaliana]	e-133
TC116203	0	5	1	9	0	0.006337	ACT domain-containing protein [Arabidopsis thaliana]	5e-088
TC116618	3	4	2	0	0	0.006442	14-3-3 protein [Solanum tuberosum]	e-134
TC123974	0	4	0	5	0	0.006681	No hits found	

TC115881	0	0	2	10	5	0.006776	probable calcium-binding protein - potato	7e-090
TC124071	0	0	1	11	0	0.006798	chlorophyll a/b-binding protein - tomato	e-145
TC124391	1	0	5	3	0	0.006971	NADPH-ferrihemoprotein reductase (EC 1.6.2.4) - mung bean	0.0
TC116273	6	0	1	0	0	0.006977	aquaporin 2 [Samanea saman]	6e-099
TC117120	6	0	1	0	0	0.006977	Avr9/Cf-9 rapidly elicited protein 216 [Nicotiana tabacum]	e-175
TC123927	6	0	1	0	0	0.006977	oxidase like protein - tomato	0.0
TC125582	6	0	1	0	0	0.006977	pectinesterase [Arabidopsis thaliana]	3e-040
TC117937	6	0	1	0	0	0.006977	putative protein kinase [Arabidopsis thaliana]	8e-046
TC117604	1	0	0	1	3	0.006989	metalloprotease, putative [Arabidopsis thaliana]	0.0
TC115755	9	4	0	2	1	0.006992	bZIP protein [Arabidopsis thaliana]	e-111
TC124017	3	4	9	21	0	0.00741	cysteine proteinase (EC 3.4.22.-) precursor - tomato	0.0
TC115713	4	9	14	29	6	0.007503	probable chaperonin 60 beta chain precursor, chloroplast - potato	0.0
TC117875	0	0	1	4	4	0.007611	putative jasmonic acid regulatory protein [Arabidopsis thaliana]	2e-088
TC119337	0	3	0	1	1	0.007798	steroid 22-alpha-hydroxylase (DWF4) [Arabidopsis thaliana]	e-121
TC118249	0	0	0	9	0	0.007924	C2 domain-containing protein [Arabidopsis thaliana]	2e-056
TC125167	0	0	0	9	0	0.007924	hypothetical protein [Arabidopsis thaliana]	5e-017
TC125359	0	0	0	9	0	0.007924	MADS box transcription factor MADS1 [Capsicum annuum]	8e-097
TC125703	0	0	0	9	0	0.007924	putative DNA cytosine methyltransferase Zmet3 [Zea mays]	e-118
TC116478	0	0	0	9	0	0.007924	ripening regulated protein DDTR18 [Lycopersicon esculentum]	e-170
TC124242	10	1	3	1	1	0.007925	40S ribosomal protein S3A	e-120
TC125381	8	0	0	3	0	0.008046	PSR9 [Brassica nigra]	e-165
TC125554	9	1	0	3	0	0.00808	stomatal cytokinesis defective protein [Arabidopsis thaliana]	0.0
TC124567	1	1	0	14	1	0.008804	Probable adenylate kinase 2, chloroplast precursor	6e-080
TC124262	2	3	0	4	5	0.00897	plasma membrane protein - common tobacco	2e-051
TC123970	1	1	0	11	5	0.009055	early nodulin ENOD18 [Vicia faba]	3e-034
TC125672	3	3	0	0	0	0.009124	putative protein [Arabidopsis thaliana]	1e-089
TC116880	7	2	1	0	0	0.009153	unknown [Lycopersicon esculentum]	e-119



TC116368	12	1	0	17	0	0.009266	ripening regulated protein DDTFR10/A [ <i>Lycopersicon esculentum</i> ]	e-107
TC118177	6	0	0	1	0	0.009362	AP2 domain protein RAP2.8 (RAV2) [ <i>Arabidopsis thaliana</i> ]	7e-051
TC126572	6	0	0	1	0	0.009362	hypothetical protein [ <i>Arabidopsis thaliana</i> ]	1e-021
TC118018	6	0	0	1	0	0.009362	similar to carboxylesterase [ <i>Oryza sativa</i> (japonica cultivar-group)]	3e-051
TC117117	6	0	0	1	0	0.009362	T complex protein	e-167
TC124479	0	0	5	8	0	0.009401	Translocase of chloroplast 34	e-115
TC125082	0	3	0	2	0	0.009523	putative carnitine/acylcarnitine translocase [ <i>Oryza sativa</i> ]	e-128
TC126280	0	3	0	2	0	0.009523	putative TPR repeat nuclear phosphoprotein [ <i>Arabidopsis thaliana</i> ]	e-131
TC124830	0	3	0	2	0	0.009523	remorin 1 [ <i>Lycopersicon esculentum</i> ]	3e-057
TC124368	7	0	0	2	0	0.009603	late-embryogenesis protein lea5 - common tobacco	9e-021
TC124796	2	0	2	15	0	0.009629	putative prolyl 4-hydroxylase, alpha subunit [ <i>Arabidopsis thaliana</i> ]	4e-084
TC116975	0	0	0	10	3	0.009696	expressed protein [ <i>Arabidopsis thaliana</i> ]	e-127
TC116229	1	5	1	13	0	0.009732	probable serine/threonine kinase (EC 2.7.1.-) SNFL2 - sorghum	0.0
TC124159	7	0	1	1	0	0.009769	allergen-like protein BRSn20 [ <i>Sambucus nigra</i> ]	5e-044
TC116094	1	1	2	17	5	0.009877	fructose-bisphosphate aldolase [ <i>Persea americana</i> ]	e-173
TC115918	1	0	3	0	0	0.009966	MTN3-like protein [ <i>Arabidopsis thaliana</i> ]	3e-068

<sup>a</sup> This column represents the number of ESTs from overy library with total of 9878 ESTs

<sup>b</sup> This column represents the number of ESTs from immature green fruit library with total of 4240 ESTs

<sup>c</sup> This column represents the number of ESTs from mature green library with total of 5317 ESTs

<sup>d</sup> This column represents the number of ESTs from breaker fruit library with total of 15207 ESTs

<sup>e</sup> This column represents the number of ESTs from red ripe fruit library with total of 3895 ESTs

**Table XI.** Differentially expressed genes during tomato flower development

TC	A <sup>a</sup>	B <sup>b</sup>	C <sup>c</sup>	D <sup>d</sup>	P value	Annotation	E value
TC124100	3	7	44	1	0	Aminopeptidase 1, chloroplast precursor	0.0
TC124960	0	0	0	17	0	anthranilate N-hydroxycinnamoyl/benzoyltransferase-like protein [Arabidopsis thaliana]	1e-071
TC115985	0	0	0	15	0	cell-wall invertase [Lycopersicon esculentum]	0.0
TC116195	0	0	0	12	0	endo-1,3-beta-glucanase-like protein [Pyrus pyrifolia]	e-105
TC124201	0	1	1	40	0	aldo/keto reductase family [Arabidopsis thaliana]	e-165
TC116052	0	2	3	30	0	Light-induced protein, chloroplast precursor	e-175
TC124618	0	0	0	11	0	Pectate lyase precursor	e-104
TC116146	0	0	0	13	0	pectinesterase family [Arabidopsis thaliana]	e-106
TC115920	0	0	0	47	0	Pectinesterase precursor (Pectin methylesterase)	e-139
TC115991	0	0	0	12	0	PGPS/D4 [Petunia x hybrida]	1e-094
TC124027	51	15	32	7	0	Polyphenol oxidase F, chloroplast precursor	0.0
TC116077	0	0	0	17	0	Probable pectate lyase P56 precursor	0.0
TC116143	0	0	0	18	0	Probable pectinesterase precursor (Pectin methylesterase)	2e-065
TC115750	0	0	0	18	0	putative beta-galactosidase [Nicotiana tabacum]	0.0
TC123753	20	58	24	0	0	RuBisCO small subunit 2A	1e-096
TC123709	8	49	11	2	0	RuBisCO small subunit 3A/3C	6e-097
TC123710	2	21	10	1	0	RuBisCO small subunit 3B	e-100
TC117559	0	0	0	12	0	SAM:benzoic acid carboxyl methyltransferase [Antirrhinum majus]	1e-039
TC117334	0	0	0	11	0	SRG1-like protein [Arabidopsis thaliana]	7e-079
TC115826	1	5	22	4	0	Threonine dehydratase biosynthetic, chloroplast precursor	0.0
TC116149	1	3	29	0	0	translation-inhibitor protein [Gentiana triflora]	1e-050
TC116040	2	3	24	3	0	Wound-induced proteinase inhibitor ii precursor	3e-073
TC124953	0	11	1	0	0.000001	anther-specific protein - wood tobacco	0.0
TC123943	3	19	7	0	0.000001	Elongation factor 1-BETA	3e-060

TC118069	0	0	0	10	0.000001	lectin receptor kinase [Arabidopsis thaliana]	1e-030
TC124645	0	0	0	10	0.000001	malonyl CoA:anthocyanin 5-O-glucoside-6"-O-malonyltransferase [Perilla frutescens]	1e-091
TC124110	7	2	1	20	0.000001	methionine synthase [Solanum tuberosum]	0.0
TC116969	0	0	0	10	0.000001	putative pectinesterase [Arabidopsis thaliana]	8e-033
TC124102	0	0	0	10	0.000001	putative purple acid phosphatase precursor [Ipomoea batatas]	e-117
TC118189	0	9	0	0	0.000002	gamma-thionin-like protein precursor - tomato	4e-034
TC115845	10	18	32	4	0.000003	Translationally controlled tumor protein homolog	1e-082
TC124346	0	0	0	9	0.000004	pectin methyl-esterase - like protein [Arabidopsis thaliana]	e-145
TC116305	0	0	0	9	0.000004	PGPS/D7 [Petunia x hybrida]	3e-031
TC117094	0	0	0	9	0.000004	Probable pectate lyase P59 precursor	0.0
TC124636	0	0	0	9	0.000004	P-type H+-ATPase [Vicia faba]	0.0
TC116182	1	10	0	0	0.000005	Photosystem II 22 kDa protein, chloroplast precursor (CP22)	2e-083
TC125342	0	8	0	0	0.000012	glycine-rich protein - tomato	8e-025
TC124292	0	0	0	8	0.000016	beta-tubulin [Naegleria gruberi]	0.0
TC124179	0	0	0	8	0.000016	Chorismate synthase 1, chloroplast precursor	0.0
TC125621	0	0	0	8	0.000016	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-114
TC116375	0	0	0	8	0.000016	putative glyoxal oxidase [Oryza sativa (japonica cultivar-group)]	e-143
TC123778	36	7	11	14	0.000018	Elongation factor 1-ALPHA (EF-1-ALPHA)	0.0
TC116347	2	0	15	5	0.000019	DIP-1 [Citrullus lanatus]	0.0
TC115921	5	18	15	0	0.000024	lipid transfer protein 2 [Lycopersicon pennellii]	3e-056
TC116957	1	0	9	0	0.000041	Mitogen-activated protein kinase homolog NTF3 (P43)	e-130
TC123754	0	7	0	0	0.000054	40S ribosomal protein S23 (S12)	1e-075
TC126393	0	7	0	0	0.000054	putative protein [Arabidopsis thaliana]	e-129
TC125757	0	8	1	0	0.000091	NTM19 [Nicotiana tabacum]	3e-014
TC118779	0	0	7	0	0.000096	No hits found	
TC115925	0	0	7	0	0.000096	putative nascent polypeptide associated complex alpha chain [Oryza sativa (japonica cultivar-group)]	4e-058
TC123975	0	0	1	8	0.000123	L-ascorbate oxidase homolog precursor (Ascorbase)	e-161

TC124456	0	0	1	8	0.000123	S-adenosylmethionine decarboxylase proenzyme	e-136
TC115898	1	0	0	8	0.000125	polyubiquitin 6 - parsley	0.0
TC116632	0	6	1	12	0.000146	Omega-3 fatty acid desaturase, endoplasmic reticulum	1e-084
TC123967	0	11	4	1	0.000167	Oxygen-evolving enhancer protein 2, chloroplast precursor	e-127
TC116523	0	6	0	0	0.00025	ribosomal protein L33 [Castanea sativa]	6e-056
TC125245	0	6	0	0	0.00025	senescence-associated protein -related [Arabidopsis thaliana]	3e-028
TC124113	2	1	11	1	0.00027	chlorophyll a/b-binding protein type III precursor - tomato	e-136
TC124056	9	2	9	21	0.000286	arginine decarboxylase 1 [Datura stramonium]	0.0
TC124015	3	10	1	0	0.000321	Chlorophyll A-B binding protein 13, chloroplast precursor	e-146
TC116099	4	5	14	0	0.000353	viroid RNA-binding protein [Lycopersicon esculentum]	0.0
TC123815	0	3	9	0	0.000376	putative proline-rich protein APG isolog [Cicer arietinum]	e-146
TC118366	0	7	1	0	0.000397	chalcone synthase - like protein [Arabidopsis thaliana]	e-112
TC126185	0	7	1	0	0.000397	No hits found	
TC123786	5	13	2	1	0.000407	Chlorophyll A-B binding protein 1B, chloroplast precursor	e-145
TC119088	0	0	6	0	0.000408	Ac-like transposase [Arabidopsis thaliana]	5e-021
TC117703	0	0	6	0	0.000408	Floral homeotic protein PMADS 2	9e-099
TC118104	0	0	6	0	0.000408	hypothetical protein F27H5.90 - Arabidopsis thaliana	6e-034
TC116753	0	0	6	0	0.000408	putative sensory transduction histidine kinase [Arabidopsis thaliana]	0.0
TC116051	0	0	6	0	0.000408	putative transcription factor BTF3 [Arabidopsis thaliana]	1e-047
TC124304	0	0	6	0	0.000408	Serine/threonine protein phosphatase PP2A catalytic subunit	e-174
TC116269	0	2	0	8	0.000444	TPR-repeat protein [Arabidopsis thaliana]	0.0
TC124141	0	0	2	8	0.000458	putative gibberellin beta-hydroxylase [Arabidopsis thaliana]	2e-040
TC117100	8	0	0	1	0.000483	pyruvate kinase [Arabidopsis thaliana]	0.0
TC116519	1	8	0	1	0.000519	histone H2A, putative [Arabidopsis thaliana]	1e-044
TC123946	3	10	13	0	0.000529	Cathepsin D Inhibitor [Lycopersicon esculentum]	e-118
TC116391	10	0	1	2	0.000681	Argonaute (AGO1)-like protein [Arabidopsis thaliana]	0.0
TC123859	24	7	6	7	0.000684	RuBisCO subunit binding-protein alpha subunit, chloroplast precursor	0.0

TC116050	15	1	2	9	0.000718	unknown protein [Arabidopsis thaliana]	e-179
TC116028	19	5	3	6	0.000982	Phosphoglycerate kinase, cytosolic	0.0
TC123989	0	9	3	1	0.001001	60S ribosomal protein L10a-1	7e-095
TC115931	6	0	0	0	0.001017	Elongation factor 1-gamma	e-166
TC124479	6	0	0	0	0.001017	Translocase of chloroplast 34	e-115
TC124122	0	5	0	0	0.00114	40S ribosomal protein S11	2e-071
TC119054	0	5	0	0	0.00114	Avr9/Cf-9 rapidly elicited protein 146 [Nicotiana tabacum]	3e-037
TC126363	0	5	0	0	0.00114	bHLH protein [Arabidopsis thaliana]	2e-028
TC116913	0	5	0	0	0.00114	Csf-1 [Cucumis sativus]	2e-060
TC124320	0	5	0	0	0.00114	CycD3;3 [Lycopersicon esculentum]	2e-049
TC125872	0	5	0	0	0.00114	transferase - related [Arabidopsis thaliana]	4e-063
TC126630	0	5	0	0	0.00114	unknown [Arabidopsis thaliana]	3e-053
TC119156	0	5	0	0	0.00114	GDSL-motif lipase/hydrolase protein [Arabidopsis thaliana]	1e-055
TC119528	0	5	0	0	0.00114	unknown protein [Arabidopsis thaliana]	3e-026
TC116412	0	7	0	2	0.001317	neutral leucine aminopeptidase preprotein [Lycopersicon esculentum]	0.0
TC116723	0	9	2	2	0.001329	MADS-box protein homolog DEF4 - potato	e-120
TC115774	37	27	24	8	0.001334	rRNA intron-encoded homing endonuclease [Oryza sativa]	5e-021
TC125042	0	0	0	5	0.001404	ankyrin-kinase [Medicago truncatula]	0.0
TC116302	0	0	0	5	0.001404	contains similarity to MtN3 protein [Arabidopsis thaliana]	3e-070
TC119009	0	0	0	5	0.001404	GTPase activator protein of Rab-like small GTPases-like protein [Arabidopsis thaliana]	1e-034
TC124087	0	0	0	5	0.001404	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	2e-071
TC124696	0	0	0	5	0.001404	plasma membrane H <sup>+</sup> -ATPase [Oryza sativa (japonica cultivar-group)]	4e-042
TC115719	0	0	0	5	0.001404	Pollen-specific protein NTP303 precursor	0.0
TC125760	0	0	0	5	0.001404	putative protein kinase [Arabidopsis thaliana]	3e-023
TC126240	0	0	0	5	0.001404	SINA1 protein [imported] - Vitis vinifera	7e-083
TC119562	0	0	0	5	0.001404	unknown protein [Arabidopsis thaliana]	4e-024
TC116926	0	0	0	5	0.001404	protein kinase [Raphanus sativus]	3e-075

TC124785	0	0	0	5	0.001404	uroporphyrinogen III synthase [Arabidopsis thaliana]	1e-080
TC116772	0	3	10	2	0.001461	GDSL-motif lipase/hydrolase protein [Arabidopsis thaliana]	e-159
TC125546	7	0	1	0	0.001656	putative rubisco subunit binding-protein alpha subunit [Arabidopsis thaliana]	e-173
TC118159	0	6	1	0	0.001681	glycine-rich protein - tomato	7e-019
TC123973	1	9	4	0	0.001686	Ferredoxin I, chloroplast precursor	6e-066
TC116818	2	8	1	0	0.001701	Systemin precursor	1e-079
TC126749	0	0	5	0	0.001707	FH protein NFH1 [Nicotiana tabacum]	2e-071
TC127246	0	0	5	0	0.001707	gamma glutamyl hydrolase, putative [Arabidopsis thaliana]	8e-026
TC119336	0	0	5	0	0.001707	No hits found	
TC118730	0	0	5	0	0.001707	ethylene-insensitive 2 (EIN2) [Arabidopsis thaliana]	2e-042
TC118964	0	0	5	0	0.001707	ZF-HD homeobox protein [Flaveria bidentis]	1e-018
TC125123	0	0	5	0	0.001707	microtubule-associated protein [Arabidopsis thaliana]	e-118
TC117730	0	0	5	0	0.001707	root cap 1 (RCP1) [Arabidopsis thaliana]	8e-012
TC115751	2	0	2	9	0.001828	Catalase isozyme 2	0.0
TC124163	0	2	7	0	0.002005	Ribosomal protein L18a	1e-091
TC125514	0	1	0	6	0.002059	DEAD/DEAH box RNA helicase, putative [Arabidopsis thaliana]	0.0
TC125048	1	7	1	0	0.002076	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, chloroplast precursor	0.0
TC116338	1	7	1	0	0.002076	Vacuolar ATP synthase subunit G 1	7e-034
TC116295	0	0	1	6	0.002082	1-aminocyclopropane-1-carboxylate oxidase [Lycopersicon esculentum]	e-169
TC124034	0	0	1	6	0.002082	Water-stress induced tonoplast intrinsic protein	e-165
TC124725	2	2	9	0	0.002272	Kunitz-type enzyme inhibitor S9C11 [Solanum tuberosum]	2e-019
TC116514	0	1	6	0	0.002514	gene C-7 protein - common tobacco	e-107
TC126715	0	1	6	0	0.002514	male sterility protein 2 - Arabidopsis thaliana	4e-034
TC117653	1	0	6	0	0.002598	unknown [Arabidopsis thaliana]	9e-038
TC117295	1	0	1	7	0.002604	mitochondrial processing peptidase [Solanum tuberosum]	0.0
TC123871	11	3	3	0	0.003298	plastidic aldolase NPALDP1 [Nicotiana paniculata]	0.0
TC118093	5	0	0	0	0.00363	dem protein - tomato	0.0

TC116029	13	0	4	6	0.003735	poly(A)-binding protein [Nicotiana tabacum]	0.0
TC116417	3	3	12	2	0.003892	OSJNBa0084K20.16 [Oryza sativa (japonica cultivar-group)]	9e-091
TC116482	1	3	2	10	0.004055	catechol O-methyltransferase (EC 2.1.1.6) III - common tobacco	e-149
TC123948	14	16	23	4	0.004351	Aspartic protease inhibitor 1 precursor	1e-072
TC117170	7	2	0	0	0.004555	ZF-HD homeobox protein [Flaveria bidentis]	5e-040
TC123879	35	11	17	16	0.004738	S-adenosylmethionine decarboxylase proenzyme]	0.0
TC115976	1	1	3	9	0.004787	RuBisCO activase 1	0.0
TC116939	0	4	0	0	0.005112	50S ribosomal protein L15, chloroplast precursor [Arabidopsis thaliana]	4e-082
TC116060	0	4	0	0	0.005112	60S ribosomal protein L30	4e-056
TC116930	0	4	0	0	0.005112	acyl carrier protein, putative [Arabidopsis thaliana]	3e-044
TC116411	0	4	0	0	0.005112	Chalcone--flavonone isomerase A (Chalcone isomerase A)	5e-094
TC127390	0	4	0	0	0.005112	contains similarity to heparanase [Arabidopsis thaliana]	2e-078
TC126323	0	4	0	0	0.005112	cytochrome P450, putative [Arabidopsis thaliana]	5e-063
TC118169	0	4	0	0	0.005112	expressed protein [Arabidopsis thaliana]	2e-094
TC126366	0	4	0	0	0.005112	emp24/gp25L/p24 family [Arabidopsis thaliana]	3e-074
TC126468	0	4	0	0	0.005112	vacuolar ATPase subunit e-like [Mesembryanthemum crystallinum]	8e-029
TC119205	0	4	0	0	0.005112	gamma-thionin [Lycopersicon esculentum]	2e-015
TC124821	0	4	0	0	0.005112	histone H2B [Cicer arietinum]	3e-043
TC116873	0	4	0	0	0.005112	histone H3, putative [Arabidopsis thaliana]	5e-060
TC127455	0	4	0	0	0.005112	ubiquitin family [Arabidopsis thaliana]	2e-030
TC116710	0	4	0	0	0.005112	hypothetical protein [Arabidopsis thaliana]	e-120
TC118198	0	4	0	0	0.005112	mitotic checkpoint protein, putative [Arabidopsis thaliana]	e-158
TC126744	0	4	0	0	0.005112	myb-related transcription factor THM18 - tomato	e-129
TC120308	0	4	0	0	0.005112	No hits found	
TC120498	0	4	0	0	0.005112	No hits found	
TC128203	0	4	0	0	0.005112	No hits found	
TC128245	0	4	0	0	0.005112	No hits found	

TC116141	0	4	0	0	0.005112	oxygen-evolving complex protein 16, chloroplast precursor [Arabidopsis thaliana]	2e-058
TC118064	0	4	0	0	0.005112	peptidylprolyl isomerase (EC 5.2.1.8) Cyp1 - rice	4e-067
TC123833	0	4	0	0	0.005112	Proteasome subunit beta type 1	e-122
TC116763	0	4	0	0	0.005112	putative protein [Arabidopsis thaliana]	1e-044
TC118481	0	4	0	0	0.005112	putative protein [Arabidopsis thaliana]	2e-067
TC118734	0	4	0	0	0.005112	Ribosomal protein L7Ae-like [Arabidopsis thaliana]	2e-051
TC125970	0	4	0	0	0.005112	ripening regulated protein DDTFR8 [Lycopersicon esculentum]	2e-060
TC124727	0	4	0	0	0.005112	thioredoxin-like protein CDSP32, chloroplast - potato	e-166
TC125572	6	1	0	0	0.005448	replication factor A - like protein [Arabidopsis thaliana]	0.0
TC116961	6	0	0	1	0.00548	enoyl-ACP reductase [Petunia x hybrida]	0.0
TC117810	6	0	1	0	0.00551	Peptide deformylase, mitochondrial precursor	e-157
TC116531	3	8	10	0	0.005689	unnamed protein product [Oryza sativa (japonica cultivar-group)]	1e-088
TC125600	0	0	2	6	0.005836	phosphatidylglycerol specific phospholipase C [Arabidopsis thaliana]	0.0
TC115927	2	0	2	8	0.005864	Elongation factor 1-gamma	3e-096
TC119747	0	0	0	4	0.006028	anthranilate N-hydroxycinnamoyl/benzoyltransferase - like protein [Arabidopsis thaliana]	2e-034
TC120402	0	0	0	4	0.006028	auxin-responsive-related protein [Arabidopsis thaliana]	4e-091
TC123821	0	0	0	4	0.006028	beta-fructosidase [Lycopersicon esculentum]	0.0
TC117132	0	0	0	4	0.006028	putative inosine-uridine preferring nucleoside hydrolase [Oryza sativa (japonica cultivar-group)]	3e-088
TC120341	0	0	0	4	0.006028	pentatricopeptide (PPR) repeat-containing protein [Arabidopsis thaliana]	5e-037
TC119649	0	0	0	4	0.006028	feebly-like protein [Arabidopsis thaliana]	7e-082
TC119294	0	0	0	4	0.006028	formamidase - like protein [Arabidopsis thaliana]	e-101
TC120291	0	0	0	4	0.006028	globulin-like protein [Daucus carota]	6e-079
TC120237	0	0	0	4	0.006028	AMP-dependent synthetase and ligase family [Arabidopsis thaliana]	1e-040
TC125924	0	0	0	4	0.006028	hypothetical protein [Arabidopsis thaliana]	e-125
TC124937	0	0	0	4	0.006028	transcription factor Hap5a, putative [Arabidopsis thaliana]	9e-061
TC124666	0	0	0	4	0.006028	lysine and histidine specific transporter, putative [Arabidopsis thaliana]	e-111
TC118050	0	0	0	4	0.006028	No hits found	



TC126836	0	0	0	4	0.006028	No hits found	
TC120178	0	0	0	4	0.006028	Phosphatidylinositol Transfer Protein, putative [Arabidopsis thaliana]	e-110
TC117060	0	0	0	4	0.006028	PRL1 protein [Arabidopsis thaliana]	3e-056
TC116450	0	0	0	4	0.006028	Probable pectate lyase P59 precursor	0.0
TC127251	0	0	0	4	0.006028	putative cap-binding protein [Arabidopsis thaliana]	2e-074
TC117389	0	0	0	4	0.006028	putative gigantea protein [Arabidopsis thaliana]	e-143
TC124517	0	0	0	4	0.006028	putative pectinesterase [Arabidopsis thaliana]	3e-032
TC125707	0	0	0	4	0.006028	putative protein [Arabidopsis thaliana]	2e-034
TC117860	0	0	0	4	0.006028	putative t-SNARE SED5 [Oryza sativa (japonica cultivar-group)]	2e-089
TC120247	0	0	0	4	0.006028	putative uridine kinase [Oryza sativa (japonica cultivar-group)]	9e-013
TC125301	0	0	0	4	0.006028	receptor-like protein kinase PRK2 - tomato	0.0
TC125530	0	0	0	4	0.006028	Serine/threonine Kinase [Persea americana]	e-111
TC117491	0	0	0	4	0.006028	Suberization-associated anionic peroxidase 1 precursor	0.0
TC127566	0	0	0	4	0.006028	sulfate transporter ATST1 [Arabidopsis thaliana]	2e-052
TC128250	0	0	0	4	0.006028	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Clostridium acetobutylicum]	4e-031
TC128408	0	0	0	4	0.006028	transporter-related [Arabidopsis thaliana]	1e-044
TC116571	0	0	0	4	0.006028	unknown protein [Arabidopsis thaliana]	7e-058
TC127248	0	0	0	4	0.006028	Putative myosin heavy chain-like [Oryza sativa]	3e-012
TC116859	8	1	0	2	0.006036	No hits found	
TC124715	8	0	1	2	0.00608	similar to carboxylesterase [Oryza sativa (japonica cultivar-group)]	3e-087
TC124619	3	9	3	0	0.006391	nucleic acid binding protein [Nicotiana tabacum]	e-133
TC118183	0	5	1	0	0.006848	5B protein - tomato	6e-048
TC123857	0	5	1	0	0.006848	Chlorophyll A-B binding protein 3C, chloroplast precursor	e-148
TC127031	0	5	1	0	0.006848	b-keto acyl reductase, putative [Arabidopsis thaliana]	2e-053
TC116427	0	5	1	0	0.006848	Malate dehydrogenase, glyoxysomal precursor	e-174
TC116489	0	5	1	0	0.006848	Photosystem II 5 kDa protein, chloroplast precursor	2e-012
TC116604	1	5	0	0	0.00702	Ribosomal protein L18a	5e-093

TC126171	0	0	4	0	0.00703	ATP-dependent Clp protease proteolytic subunit, putative [Arabidopsis thaliana]	4e-085
TC120464	0	0	4	0	0.00703	gibberellin 20-oxidase - common tobacco	e-104
TC116191	0	0	4	0	0.00703	GTP-binding protein, ras-related - common tobacco	2e-092
TC120500	0	0	4	0	0.00703	No hits found	
TC128398	0	0	4	0	0.00703	No hits found	
TC117782	0	0	4	0	0.00703	protein phosphatase 2C (PP2C) [Fagus sylvatica]	0.0
TC120311	0	0	4	0	0.00703	putative anthranilate N-hydroxycinnamoyl/benzoyltransferase [Arabidopsis thaliana]	4e-052
TC125293	0	0	4	0	0.00703	putative glucosyltransferase [Lycopersicon esculentum]	1e-047
TC128295	0	0	4	0	0.00703	putative phytochrome-associated protein 3 [Arabidopsis thaliana]	3e-011
TC128354	0	0	4	0	0.00703	putative protein kinase [Arabidopsis thaliana]	3e-034
TC126656	0	0	4	0	0.00703	unknown protein [Arabidopsis thaliana]	5e-034
TC124171	7	1	0	8	0.007216	peptidylprolyl isomerase (EC 5.2.1.8) ROF1 - Arabidopsis thaliana	0.0
TC115982	6	7	16	3	0.007475	Chlorophyll A-B binding protein 4, chloroplast precursor	e-155
TC116718	1	6	0	1	0.00785	Flavonol synthase (FLS)	e-171
TC116602	1	6	0	1	0.00785	PAPS-reductase-like protein [Catharanthus roseus]	0.0
TC124421	1	6	1	0	0.007888	40S ribosomal protein S8	1e-092
TC124278	1	6	1	0	0.007888	ribosomal protein L33 [Castanea sativa]	3e-055
TC116925	0	1	0	5	0.008019	AER [Nicotiana tabacum]	0.0
TC116595	0	1	0	5	0.008019	probable transcription factor - fava bean	8e-036
TC126040	0	6	3	0	0.00803	flavonoid 1-2 rhamnosyltransferase [Citrus maxima]	5e-029
TC116917	0	6	3	0	0.00803	putative protein [Arabidopsis thaliana]	1e-097
TC115800	24	13	9	6	0.008046	unspecific monooxygenase (EC 1.14.14.1) - common tobacco	e-100
TC125815	0	0	1	5	0.008126	3-methyl-2-oxobutanoate dehydrogenase E1-alpha chain precursor, mitochondrial - tomato	0.0
TC117643	0	0	1	5	0.008126	CBL-interacting protein kinase 11 [Arabidopsis thaliana]	5e-072
TC116610	0	0	1	5	0.008126	unknown protein [Oryza sativa (japonica cultivar-group)]	e-123
TC123851	12	17	18	3	0.008175	Chlorophyll A-B binding protein 3C, chloroplast precursor	e-145
TC116428	1	0	0	5	0.00833	putative chorismate mutase/prephenate dehydratase [Arabidopsis thaliana]	e-167

TC125109	1	0	0	5	0.00833	RNA helicase-like protein DB10	2e-084
TC126175	1	0	0	5	0.00833	SDL-1 protein [Nicotiana glumbaginifolia]	e-114
TC124672	1	0	0	5	0.00833	TOM (target of myb1) -like protein [Arabidopsis thaliana]	e-159
TC124114	3	6	0	0	0.009319	photosystem I subunit XI [Nicotiana attenuata]	e-102
TC125376	0	1	5	0	0.009408	expansin [Lycopersicon esculentum]	e-105
TC117305	0	1	5	0	0.009408	putative protein [Arabidopsis thaliana]	3e-076
TC117910	0	1	5	0	0.009408	unknown protein [Arabidopsis thaliana]	1e-071
TC125168	0	0	5	1	0.009472	splicing factor U2AF small subunit, putative [Arabidopsis thaliana]	3e-092
TC124464	1	0	1	6	0.009486	transaldolase [Lycopersicon esculentum]	0.0
TC125666	1	0	5	0	0.009772	serine/threonine kinase-like protein [Arabidopsis thaliana]	6e-095

<sup>a</sup> This column represents the number of ESTs from 0-3mm flower bud library with total of 6259 ESTs

<sup>b</sup> This column represents the number of ESTs from 3-8mm flower bud library with total of 5524 ESTs

<sup>c</sup> This column represents the number of ESTs from 8mm-preanthesis flower bud library with total of 5759 ESTs

<sup>d</sup> This column represents the number of ESTs from open flower library with total of 15207 ESTs

## APPENDIX B

### Fruit specific genes

TC	T1045	T10018	T1356	T1775	T1391	Annotation	E value
TC115942	0	0	0	152	34	acyltransferase 2 [Capsicum chinense]	0.0
TC123915	0	0	0	100	12	Histidine decarboxylase (HDC) (TOM92)	0.0
TC124082	0	0	0	55	49	Polygalacturonase 2A precursor (PG-2A) (Pectinase)	0.0
TC116135	0	9	22	60	4	Pectinesterase 1 precursor (Pectin methylesterase 1) (PE 1)	0.0
TC124020	0	0	0	53	15	lipoxygenase (EC 1.13.11.12) - tomato	0.0
TC123916	0	0	0	58	4	Histidine decarboxylase (HDC) (TOM92)	e-146
TC116252	3	1	0	34	18	tomato fruit ripening specific URF [Lycopersicon esculentum]	0.0
TC116326	0	0	0	32	20	ripening protein E8 homolog - tomato	0.0
TC116315	0	2	3	37	7	thiol protease [Matricaria chamomilla]	e-120
TC124239	0	0	1	42	5	inactive endo-beta-mannanase [Lycopersicon esculentum]	0.0
TC124537	2	0	0	23	9	putative protein [Arabidopsis thaliana]	e-144
TC124373	6	3	1	21	2	putative protein [Arabidopsis thaliana]	1e-058
TC124194	0	1	0	19	7	PHAP2A protein [Petunia x hybrida]	e-116
TC116319	0	1	3	18	2	Hsp20.0 protein [Lycopersicon peruvianum]	4e-065
TC116893	0	0	0	18	2	pol polyprotein [Citrus x paradisi]	e-129
TC125000	0	0	0	11	9	putative protein [Arabidopsis thaliana]	4e-050
TC117044	0	0	0	14	5	glucuronosyl transferase homolog, ripening-related - tomato	0.0
TC115797	0	6	3	7	0	No hits found	
TC115943	0	0	0	14	2	acyltransferase 2 [Capsicum chinense]	2e-063
TC115760	7	0	1	7	0	hypothetical protein [Arabidopsis thaliana]	4e-020
TC125310	0	0	1	10	4	cytochrome P450-dependent fatty acid hydroxylase [Nicotiana tabacum]	e-179
TC123792	0	0	4	6	4	No hits found	
TC125305	0	0	0	9	5	TDR4 transcription factor [Lycopersicon esculentum]	e-136
TC116479	0	0	3	10	0	ripening regulated protein DDTFR18 [Lycopersicon esculentum]	2e-095
TC117558	13	0	0	0	0	wall-associated kinase 3 [Arabidopsis thaliana]	1e-055
TC123987	2	0	0	8	3	Cytochrome P450 71A4 (CYPLXXIA4) (P-450EG2)	e-140
TC125491	13	0	0	0	0	hypothetical protein F25P12.93 [imported] - Arabidopsis thaliana	4e-070
TC125554	9	1	0	3	0	B1112D09.9 [Oryza sativa (japonica cultivar-group)]	0.0
TC123895	0	1	0	10	1	minor allergen beta-fructofuranosidase precursor [Lycopersicon esculentum]	2e-095
TC123921	0	0	0	8	4	Histidine decarboxylase (HDC) (TOM92)	1e-056
TC125570	0	0	0	11	1	No hits found	
TC125583	10	0	0	1	1	F14D16.26 [Arabidopsis thaliana]	7e-044
TC125629	0	0	0	11	1	steroleosin [Sesamum indicum]	e-108

TC125633	0	0	0	7	5	unnamed protein product [ <i>Oryza sativa</i> (japonica cultivar-group)]	5e-049
TC125667	0	0	0	11	1	orcinol O-methyltransferase [ <i>Rosa hybrid cultivar</i> ]	4e-091
TC115878	0	0	4	4	2	putative protein [ <i>Arabidopsis thaliana</i> ]	5e-034
TC116890	0	1	0	9	0	carbonyl reductase-like protein [ <i>Arabidopsis thaliana</i> ]	6e-073
TC117071	0	0	1	6	3	putative protein [ <i>Arabidopsis thaliana</i> ]	4e-064
TC117939	0	1	1	8	0	putative protein [ <i>Arabidopsis thaliana</i> ]	2e-021
TC118027	8	0	1	1	0	UDP-glucose 4-epimerase GEPI48	e-130
TC118036	10	0	0	0	0	bHLH protein [ <i>Arabidopsis thaliana</i> ]	9e-027
TC118070	8	0	0	2	0	No hits found	
TC124240	0	0	0	10	0	endo-beta-mannanase [ <i>Lycopersicon esculentum</i> ]	1e-059
TC116478	0	0	0	9	0	ripening regulated protein DDTFR18 [ <i>Lycopersicon esculentum</i> ]	e-170
TC117529	9	0	0	0	0	pectinesterase family [ <i>Arabidopsis thaliana</i> ]	6e-054
TC118137	9	0	0	0	0	18.1 KD CLASS I HEAT SHOCK PROTEIN (HSP 18.1)	7e-023
TC118149	8	0	0	1	0	similar to AP2 domain transcription factor, putative [ <i>Arabidopsis thaliana</i> ]	2e-033
TC118172	3	0	2	1	3	Serine/threonine Kinase [ <i>Persea americana</i> ]	e-121
TC118249	0	0	0	9	0	putative protein [ <i>Arabidopsis thaliana</i> ]	2e-056
TC124046	9	0	0	0	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	4e-014
TC125991	0	0	0	0	9	protein F21D18.16 [imported] - <i>Arabidopsis thaliana</i>	4e-043
TC126045	9	0	0	0	0	No hits found	
TC126184	9	0	0	0	0	No hits found	
TC126198	9	0	0	0	0	No hits found	
TC116530	8	0	0	0	0	homeodomain leucine zipper protein HDZ2 [ <i>Phaseolus vulgaris</i> ]	1e-022
TC118332	0	0	1	7	0	acyl-ACP thioesterase [ <i>Capsicum chinense</i> ]	e-142
TC118357	8	0	0	0	0	No hits found	
TC118424	6	0	0	2	0	unknown protein [ <i>Arabidopsis thaliana</i> ]	4e-016
TC118434	0	0	0	6	2	putative short-root protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	1e-045
TC118447	0	0	0	8	0	No hits found	
TC124972	0	0	1	7	0	unknown protein [ <i>Arabidopsis thaliana</i> ]	9e-048
TC126034	0	0	2	5	1	glutathione S-transferase GST 14 [ <i>Glycine max</i> ]	4e-042
TC126125	0	0	1	7	0	TDR4 transcription factor [ <i>Lycopersicon esculentum</i> ]	e-134
TC126246	8	0	0	0	0	Serine carboxypeptidase II-3 precursor (CP-MII.3)	1e-063
TC126257	0	0	4	4	0	guanylate kinase [ <i>Lilium longiflorum</i> ]	e-114
TC126297	0	1	0	7	0	hsp70 (AA 6 - 651) [ <i>Petunia x hybrida</i> ]	0.0
TC126361	4	0	0	3	1	protein F3F19.18 [imported] - <i>Arabidopsis thaliana</i>	1e-048
TC115971	0	0	0	7	0	mutant phytoene synthase [ <i>Lycopersicon esculentum</i> ]	1e-033
TC116074	0	0	0	5	2	OSJNBb0072N21.4 [ <i>Oryza sativa</i> (japonica cultivar-group)]	1e-078
TC116477	0	0	0	7	0	ripening regulated protein DDTFR18 [ <i>Lycopersicon esculentum</i> ]	4e-056
TC117075	0	0	0	7	0	No hits found	
TC118554	6	0	0	0	1	No hits found	

TC118596	0	0	0	7	0	putative protein [Arabidopsis thaliana]	4e-025
TC118688	7	0	0	0	0	formin-like protein [Arabidopsis thaliana]	2e-068
TC118721	7	0	0	0	0	No hits found	
TC118806	1	0	0	6	0	P0529E05.3 [Oryza sativa (japonica cultivar-group)]	2e-093
TC123829	0	0	0	5	2	beta-fructosidase [Lycopersicon esculentum]	4e-076
TC126129	7	0	0	0	0	No hits found	
TC126496	4	0	0	0	3	No hits found	
TC126505	7	0	0	0	0	No hits found	
TC126572	6	0	0	1	0	hypothetical protein [Arabidopsis thaliana]	1e-021
TC126631	1	2	0	4	0	starch synthase, isoform V [Vigna unguiculata]	e-161
TC126635	0	0	0	7	0	3,4-dihydroxy-5-hexaprenylbenzoate methyltransferase - Arabidopsis thaliana	8e-096
TC126678	0	0	0	7	0	No hits found	
TC115854	0	0	1	2	3	ripening protein E8 homolog - tomato	2e-079
TC115887	0	0	0	6	0	guanylate kinase [Nicotiana tabacum]	1e-038
TC116194	5	0	0	0	1	embryo-abundant protein EMB [Pisum sativum]	2e-051
TC116414	0	0	0	6	0	Glutamate decarboxylase (GAD) (ERT D1)	8e-057
TC117579	0	0	0	6	0	Shikimate kinase, chloroplast precursor	e-120
TC117928	5	0	0	0	1	unknown protein [Arabidopsis thaliana]	1e-061
TC118436	0	0	0	5	1	putative protein [Arabidopsis thaliana]	7e-083
TC118835	6	0	0	0	0	putative protein [Arabidopsis thaliana]	2e-046
TC118863	2	0	0	4	0	No hits found	
TC118865	1	0	0	5	0	putative protein [Arabidopsis thaliana]	5e-035
TC118941	6	0	0	0	0	potassium transport protein-like [Arabidopsis thaliana]	2e-040
TC118943	6	0	0	0	0	Cytochrome P450 71D7	4e-047
TC118980	0	0	0	6	0	exclusion [Bacteriophage lambda]	3e-092
TC119001	0	0	0	6	0	4-coumarate-CoA ligase-like protein [Arabidopsis thaliana]	8e-074
TC119035	0	0	0	6	0	expressed protein [Arabidopsis thaliana]	6e-023
TC119039	6	0	0	0	0	protein kinase-like protein [Arabidopsis thaliana]	1e-022
TC119047	0	0	0	6	0	trehalose-6-phosphate synthase, putative [Arabidopsis thaliana]	1e-093
TC119067	0	0	0	6	0	No hits found	
TC119113	0	0	0	6	0	expressed protein [Arabidopsis thaliana]	2e-046
TC119131	6	0	0	0	0	No hits found	
TC119138	6	0	0	0	0	Mutator-like transposase [Arabidopsis thaliana]	1e-023
TC119139	6	0	0	0	0	No hits found	
TC119149	6	0	0	0	0	probable mitotic control protein dis3 [imported] - Arabidopsis thaliana	3e-032
TC119181	6	0	0	0	0	putative receptor-like serine-threonine protein kinase [Solanum tuberosum]	3e-076
TC119190	0	0	0	5	1	expressed protein [Arabidopsis thaliana]	1e-056
TC119211	6	0	0	0	0	No hits found	

TC123892	0	0	0	4	2	acid invertase; AI [ <i>Lycopersicon esculentum</i> ]	0.0
TC124502	1	0	0	5	0	protein-tyrosine-phosphatase (EC 3.1.3.48) 1 [validated] - <i>Arabidopsis thaliana</i>	9e-046
TC125033	0	0	1	5	0	SQUAMOSA-PROMOTER BINDING PROTEIN 1	4e-042
TC125349	0	0	0	6	0	two-component phosphorelay mediator	3e-038
TC125355	6	0	0	0	0	putative phosphatidylinositol-4-phosphate 5-kinase [ <i>Arabidopsis thaliana</i> ]	2e-073
TC125423	0	0	0	6	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	2e-039
TC125659	6	0	0	0	0	50S ribosomal protein L4 [imported] - <i>Arabidopsis thaliana</i>	2e-033
TC126495	6	0	0	0	0	No hits found	
TC126571	3	0	0	3	0	No hits found	
TC126738	0	0	2	4	0	calcium-dependent protein kinase 3 [ <i>Nicotiana tabacum</i> ]	9e-079
TC126786	6	0	0	0	0	F16A14.19 [ <i>Arabidopsis thaliana</i> ]	8e-039
TC126818	6	0	0	0	0	probable UDP-glucuronosyltransferase (EC 2.4.1.-) - garden pea	5e-076
TC126866	0	0	0	3	3	gene C-7 protein - common tobacco	1e-097
TC126875	6	0	0	0	0	gibberellin 3-oxidase 1 [ <i>Nicotiana sylvestris</i> ]	4e-062
TC126878	6	0	0	0	0	RNA polymerase beta	0.0
TC126880	2	0	0	4	0	glycosyl hydrolase family 85 [ <i>Arabidopsis thaliana</i> ]	3e-025
TC126891	6	0	0	0	0	No hits found	
TC126972	0	0	0	6	0	putative protein [ <i>Arabidopsis thaliana</i> ]	4e-028
TC127115	6	0	0	0	0	putative protein [ <i>Arabidopsis thaliana</i> ]	5e-088
TC127132	0	0	0	5	1	putative protein [ <i>Arabidopsis thaliana</i> ]	1e-055
TC127166	6	0	0	0	0	hypothetical protein [ <i>Arabidopsis thaliana</i> ]	1e-030
TC127169	6	0	0	0	0	unknown [ <i>Arabidopsis thaliana</i> ]	8e-037
TC116737	0	0	3	2	0	hypothetical protein T13L16.2 - <i>Arabidopsis thaliana</i>	1e-018
TC116751	3	0	0	2	0	ethylene response factor 2 [ <i>Lycopersicon esculentum</i> ]	5e-054
TC116837	0	0	0	5	0	putative protein [ <i>Arabidopsis thaliana</i> ]	1e-012
TC117309	0	0	3	1	1	No hits found	
TC117983	0	0	0	5	0	armadillo repeat-containing protein [ <i>Nicotiana tabacum</i> ]	6e-011
TC118116	0	0	0	5	0	No hits found	
TC118792	0	0	0	5	0	dnaK-type molecular chaperone hsc70-3 - tomato	0.0
TC118962	2	1	0	2	0	No hits found	
TC118969	0	0	1	1	3	DnaJ protein homolog [imported] - <i>Salix gilgiana</i>	1e-096
TC119045	1	0	0	4	0	hypothetical protein [ <i>Arabidopsis thaliana</i> ]	3e-048
TC119183	1	0	0	4	0	hypothetical protein [ <i>Arabidopsis thaliana</i> ]	5e-099
TC119232	5	0	0	0	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	1e-032
TC119275	5	0	0	0	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	2e-025
TC119285	5	0	0	0	0	No hits found	
TC119286	0	0	1	4	0	putative protein [ <i>Arabidopsis thaliana</i> ]	e-128
TC119312	5	0	0	0	0	No hits found	

TC119331	1	0	0	4	0	putative protein [Arabidopsis thaliana]	3e-030
TC119337	0	3	0	1	1	steroid 22-alpha-hydroxylase (DWF4) [Arabidopsis thaliana]	e-121
TC119352	5	0	0	0	0	No hits found	
TC119435	0	0	0	5	0	No hits found	
TC119462	5	0	0	0	0	No hits found	
TC119468	1	0	0	4	0	unknown protein [Arabidopsis thaliana]	8e-067
TC119500	5	0	0	0	0	No hits found	
TC119509	0	0	0	5	0	ORF_14; similar to NAD dependent epimerase/dehydratase family [Pseudomonas aeruginosa]	2e-048
TC119521	5	0	0	0	0	No hits found	
TC119543	5	0	0	0	0	gblAAF53066.1~gene_id:MDC11.13~similar to unknown protein [Arabidopsis thaliana]	2e-028
TC119597	0	1	3	1	0	wound-induced protein Sn-1, vacuolar membrane - pepper	6e-046
TC119602	0	0	0	5	0	unknown protein [Arabidopsis thaliana]	e-109
TC119626	0	0	0	4	1	No hits found	
TC119637	0	1	3	1	0	dem-like protein [Arabidopsis thaliana]	e-121
TC119665	0	0	1	4	0	AP47/50p (gblAAB88283.1) [Arabidopsis thaliana]	e-141
TC119693	0	0	0	5	0	similar to zinc protease PQQL [Oryza sativa (japonica cultivar-group)]	4e-069
TC119715	5	0	0	0	0	expressed protein [Arabidopsis thaliana]	6e-066
TC123919	0	0	0	4	1	Histidine decarboxylase (HDC) (TOM92)	e-125
TC124344	4	0	0	1	0	putative protein [Arabidopsis thaliana]	2e-013
TC124369	5	0	0	0	0	No hits found	
TC124540	0	1	2	2	0	cytochrome P450, putative [Arabidopsis thaliana]	e-128
TC126789	5	0	0	0	0	putative protein [Arabidopsis thaliana]	e-110
TC126911	0	0	1	4	0	expressed protein [Arabidopsis thaliana]	e-105
TC127174	0	0	5	0	0	Ca2+-binding EF-hand common family protein, putative [Arabidopsis thaliana]	8e-057
TC127189	0	0	3	2	0	putative RING zinc finger protein-like protein [Thellungiella halophila]	4e-037
TC127224	0	0	0	5	0	unknown protein [Arabidopsis thaliana]	3e-048
TC127225	0	2	2	1	0	pectinesterase family [Arabidopsis thaliana]	3e-091
TC127312	5	0	0	0	0	putative protein [Arabidopsis thaliana]	8e-061
TC127335	5	0	0	0	0	hypothetical protein [imported] - Arabidopsis thaliana	2e-019
TC127344	0	0	0	5	0	expressed protein [Arabidopsis thaliana]	1e-020
TC127351	5	0	0	0	0	putative phosphatidylcholine acyltransferase [Medicago truncatula]	7e-071
TC127366	0	0	0	5	0	hypothetical protein [Arabidopsis thaliana]	1e-038
TC127379	0	0	0	5	0	Unknown protein [Oryza sativa (japonica cultivar-group)]	4e-023
TC127387	0	0	0	5	0	Similar to ORF3 protein [Oryza sativa (japonica cultivar-group)]	2e-018
TC127388	5	0	0	0	0	No hits found	
TC127391	0	1	1	3	0	ankyrin-like protein [Arabidopsis thaliana]	e-127
TC127442	5	0	0	0	0	No hits found	



TC127457	0	0	0	5	0	ABC transporter family protein [Arabidopsis thaliana]	3e-096
TC127464	0	0	0	1	4	No hits found	
TC127466	1	1	0	3	0	putative protein [Arabidopsis thaliana]	1e-062
TC127471	5	0	0	0	0	Expressed protein [Arabidopsis thaliana]	4e-015
TC127472	5	0	0	0	0	oj991113_30.18 [Oryza sativa (japonica cultivar-group)]	3e-050
TC127491	0	0	1	4	0	unknown protein [Arabidopsis thaliana]	1e-061
TC127538	0	0	0	5	0	orf, conserved hypothetical protein [Shigella flexneri 2a str. 301]	2e-027
TC127597	5	0	0	0	0	No hits found	
TC127618	0	0	0	5	0	hypothetical protein [Arabidopsis thaliana]	3e-052
TC127632	5	0	0	0	0	No hits found	
TC127636	5	0	0	0	0	putative protein [Arabidopsis thaliana]	3e-034
TC127637	4	1	0	0	0	bHLH transcription factor GBOF-1 [Tulipa gesneriana]	2e-043
TC127653	0	0	2	3	0	No hits found	
TC127668	0	0	0	5	0	No hits found	
TC115885	3	0	0	1	0	No hits found	
TC115972	0	0	0	4	0	Phytoene synthase 1, chloroplast precursor (Fruit ripening specific protein pTOM5)	3e-032
TC116070	0	0	0	3	1	21K protein precursor - alfalfa	9e-023
TC116445	4	0	0	0	0	hypothetical protein - tomato	6e-017
TC117187	4	0	0	0	0	dihydrolipoamide S-acetyltransferase [Arabidopsis thaliana]	6e-034
TC117234	0	0	0	4	0	unknown protein [Arabidopsis thaliana]	3e-024
TC117595	0	0	0	4	0	unknown protein [Arabidopsis thaliana]	2e-031
TC117618	0	0	0	4	0	ADP-RIBOSYLATION FACTOR -like protein [Arabidopsis thaliana]	5e-040
TC117646	3	0	0	1	0	expressed protein [Arabidopsis thaliana]	3e-011
TC117929	1	0	0	3	0	ReMembr-H2 protein JR702 [Arabidopsis thaliana]	3e-045
TC118406	4	0	0	0	0	expressed protein [Arabidopsis thaliana]	1e-025
TC118795	0	4	0	0	0	phytochrome B2 [Lycopersicon esculentum]	0.0
TC119142	4	0	0	0	0	No hits found	
TC119398	4	0	0	0	0	TAGL11 transcription factor [Lycopersicon esculentum]	e-121
TC119454	0	0	0	4	0	TPA: putative phytosulfokine peptide precursor [Solanum tuberosum]	2e-017
TC119478	0	0	0	4	0	No hits found	
TC119512	0	0	0	4	0	DEMETER protein [Arabidopsis thaliana]	1e-079
TC119757	3	0	0	0	1	No hits found	
TC119765	0	2	2	0	0	cyclin-dependent kinase inhibitor [Nicotiana tabacum]	2e-012
TC119801	0	0	1	1	2	protein F21B7.29 [imported] - Arabidopsis thaliana	3e-023
TC119818	0	0	0	4	0	No hits found	
TC119858	2	0	0	2	0	putative protein [Arabidopsis thaliana]	e-122
TC119885	4	0	0	0	0	No hits found	
TC119937	4	0	0	0	0	unknown protein [Arabidopsis thaliana]	2e-013

TC119941	0	0	0	0	4	hypothetical protein [Arabidopsis thaliana]	1e-027
TC119942	4	0	0	0	0	expressed protein [Arabidopsis thaliana]	9e-013
TC119956	4	0	0	0	0	RNA polymerase beta II subunit [Atropa belladonna]	e-105
TC119974	0	0	0	2	2	unknown protein [Arabidopsis thaliana]	5e-036
TC119978	0	1	0	1	2	hypothetical protein [Arabidopsis thaliana]	1e-072
TC120008	4	0	0	0	0	No hits found	
TC120024	4	0	0	0	0	No hits found	
TC120037	3	0	0	1	0	unknown protein [Arabidopsis thaliana]	7e-079
TC120043	0	1	0	3	0	hypothetical protein AT4g03420 [imported] - Arabidopsis thaliana	1e-035
TC120097	0	1	0	0	3	No hits found	
TC120104	0	0	0	4	0	No hits found	
TC120122	4	0	0	0	0	No hits found	
TC120125	0	0	0	2	2	VPS13 - like protein [Arabidopsis thaliana]	1e-079
TC120132	0	0	0	4	0	No hits found	
TC120137	4	0	0	0	0	Avr9/Cf-9 rapidly elicited protein 231 [Nicotiana tabacum]	2e-051
TC120158	1	0	0	3	0	CTV.2 [Poncirus trifoliata]	1e-049
TC120175	1	0	0	3	0	contains similarity to 5'-nucleotidase [Arabidopsis thaliana]	e-161
TC120188	0	0	0	4	0	No hits found	
TC120225	0	0	0	4	0	unknown protein [Arabidopsis thaliana]	9e-015
TC120234	4	0	0	0	0	ornithine decarboxylase [Phaseolus vulgaris]	1e-047
TC120235	0	0	1	2	1	putative 3-hydroxybutyryl-CoA dehydrogenase [Arabidopsis thaliana]	4e-088
TC120254	0	2	0	0	2	SABRE [Arabidopsis thaliana]	1e-075
TC120261	0	0	2	2	0	F5M15.7 [Arabidopsis thaliana]	2e-016
TC120263	0	0	0	4	0	No hits found	
TC120277	4	0	0	0	0	No hits found	
TC120288	0	0	1	3	0	No hits found	
TC120289	0	0	0	4	0	putative lipase [Arabidopsis thaliana]	2e-067
TC120293	0	0	1	3	0	hypothetical protein LOC55954 [Homo sapiens]	7e-012
TC120321	4	0	0	0	0	protein T7N9.16 [imported] - Arabidopsis thaliana	1e-042
TC120361	0	0	0	0	4	No hits found	
TC120367	4	0	0	0	0	No hits found	
TC120386	0	0	0	4	0	No hits found	
TC120401	0	1	0	3	0	AT5g51140/MWD22_8 [Arabidopsis thaliana]	5e-028
TC120414	0	0	0	4	0	hypothetical protein [Arabidopsis thaliana]	5e-062
TC120419	0	0	0	4	0	No hits found	
TC120424	0	0	0	4	0	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	3e-022
TC120425	4	0	0	0	0	No hits found	
TC120438	0	0	0	4	0	acid phosphatase, putative [Arabidopsis thaliana]	5e-024

TC120453	0	0	0	4	0	hypothetical protein F24B22.200 - Arabidopsis thaliana	2e-029
TC120461	1	1	0	2	0	unknown [Arabidopsis thaliana]	2e-084
TC120468	4	0	0	0	0	No hits found	
TC120480	0	0	0	4	0	No hits found	
TC120484	0	0	0	2	2	putative protein [Arabidopsis thaliana]	8e-052
TC120491	4	0	0	0	0	hypothetical protein At2g44640 [imported] - Arabidopsis thaliana	1e-040
TC120494	0	3	1	0	0	hypothetical protein F14O23.7 [imported] - Arabidopsis thaliana	4e-066
TC123918	0	0	0	4	0	Histidine decarboxylase (HDC) (TOM92)	2e-042
TC124157	4	0	0	0	0	WIZZ [Nicotiana tabacum]	1e-013
TC125444	4	0	0	0	0	ZPT2-13 [Petunia x hybrida]	8e-049
TC125537	4	0	0	0	0	protein T23J18.3 [imported] - Arabidopsis thaliana	6e-020
TC126386	0	0	0	4	0	putative protein [Arabidopsis thaliana]	1e-015
TC127469	0	1	0	1	2	OSJNBa0094O15.17 [Oryza sativa (japonica cultivar-group)]	3e-099
TC127733	0	1	0	3	0	putative vacuolar protein sorting-associated protein [Oryza sativa (japonica cultivar-group)]	e-161
TC127789	0	0	1	3	0	unknown protein [Arabidopsis thaliana]	7e-019
TC127824	4	0	0	0	0	L-ASCORBATE OXIDASE PRECURSOR (ASCORBASE) (ASO)	4e-045
TC127827	4	0	0	0	0	No hits found	
TC127888	0	0	0	4	0	putative protein [Arabidopsis thaliana]	4e-032
TC127899	0	1	2	0	1	En/Spm-like transposon protein [Arabidopsis thaliana]	5e-060
TC127920	0	0	1	2	1	unknown protein [Arabidopsis thaliana]	3e-093
TC127941	0	0	0	4	0	No hits found	
TC128023	2	0	0	2	0	No hits found	
TC128025	2	0	2	0	0	unknown protein [Arabidopsis thaliana]	2e-064
TC128066	4	0	0	0	0	symbiosis-related protein [Daucus carota]	5e-049
TC128071	4	0	0	0	0	expressed protein [Arabidopsis thaliana]	2e-026
TC128072	4	0	0	0	0	No hits found	
TC128077	3	1	0	0	0	No hits found	
TC128093	0	0	0	4	0	No hits found	
TC128100	0	0	0	4	0	putative receptor ser/thr protein kinase [Arabidopsis thaliana]	4e-033
TC128108	0	0	0	4	0	hypothetical protein At2g15270 [imported] - Arabidopsis thaliana	9e-023
TC128138	0	0	0	4	0	No hits found	
TC128140	4	0	0	0	0	hypothetical protein [Arabidopsis thaliana]	9e-046
TC128155	4	0	0	0	0	similar to c-myc binding protein [Arabidopsis thaliana]	4e-013
TC128156	0	0	0	3	1	putative casein kinase I [Arabidopsis thaliana]	e-128
TC128167	4	0	0	0	0	ABC transporter family protein [Arabidopsis thaliana]	3e-026
TC128170	4	0	0	0	0	F27J15.3 [Arabidopsis thaliana]	2e-031
TC128178	4	0	0	0	0	No hits found	
TC128179	0	0	0	4	0	putative clathrin-associated protein [Oryza sativa (japonica cultivar-group)]	1e-087

TC128184	0	0	1	3	0	No hits found	
TC128186	0	0	0	4	0	transposon related protein [Arabidopsis thaliana]	6e-038
TC128189	0	0	0	4	0	COP9 complex subunit, FUS4 [Arabidopsis thaliana]	4e-073
TC128263	0	0	0	4	0	Putative pyridoxamine 5-phosphate oxidase [Oryza sativa]	2e-018
TC128298	0	0	0	4	0	No hits found	
TC128300	0	0	0	4	0	double-stranded RNA-specific editase, putative [Arabidopsis thaliana]	3e-047
TC128320	2	1	1	0	0	nucleolar protein-like [Arabidopsis thaliana]	9e-086
TC128327	0	0	0	2	2	No hits found	
TC128334	3	0	0	1	0	No hits found	
TC128336	0	0	0	4	0	putative DNA polymerase I [Oryza sativa (japonica cultivar-group)]	6e-054
TC128344	4	0	0	0	0	No hits found	
TC128356	0	0	0	4	0	No hits found	
TC128360	0	1	0	3	0	cytochrome P-450LXXIA1 [similarity] - avocado	5e-060
TC128367	0	0	0	4	0	No hits found	
TC128370	0	0	0	4	0	expressed protein [Arabidopsis thaliana]	2e-065
TC128373	4	0	0	0	0	No hits found	
TC128394	4	0	0	0	0	No hits found	
TC128422	0	0	1	3	0	Unknown protein [Arabidopsis thaliana]	3e-075

- \* T1045: tomato ovary  
T10018: tomato developing/immature green fruit  
T1356: tomato mature green fruit  
T1775: tomato breaker fruit  
T1391: tomato red ripe fruit

### Ripening fruit specific genes

TC	T1775	T1391	Annotation	E value
TC115942	152	34	acyltransferase 2 [Capsicum chinense]	0.0
TC123915	100	12	Histidine decarboxylase (HDC) (TOM92)	0.0
TC124082	55	49	Polygalacturonase 2A precursor (PG-2A) (Pectinase)	0.0
TC124020	53	15	lipoxygenase (EC 1.13.11.12) - tomato	0.0
TC123916	58	4	Histidine decarboxylase (HDC) (TOM92)	e-146
TC116326	32	20	ripening protein E8 homolog - tomato	0.0
TC116893	18	2	pol polyprotein [Citrus x paradisi]	e-129
TC125000	11	9	putative protein [Arabidopsis thaliana]	4e-050
TC117044	14	5	glucuronosyl transferase homolog, ripening-related - tomato	0.0
TC115943	14	2	acyltransferase 2 [Capsicum chinense]	2e-063
TC125305	9	5	TDR4 transcription factor [Lycopersicon esculentum]	e-136

TC123921	8	4	Histidine decarboxylase (HDC) (TOM92)	1e-056
TC125570	11	1	No hits found	
TC125629	11	1	steroleosin [ <i>Sesamum indicum</i> ]	e-108
TC125633	7	5	unnamed protein product [ <i>Oryza sativa</i> (japonica cultivar-group)]	5e-049
TC125667	11	1	orcinol O-methyltransferase [ <i>Rosa hybrid cultivar</i> ]	4e-091
TC124240	10	0	endo-beta-mannanase [ <i>Lycopersicon esculentum</i> ]	1e-059
TC116478	9	0	ripening regulated protein DDTFR18 [ <i>Lycopersicon esculentum</i> ]	e-170
TC118249	9	0	putative protein [ <i>Arabidopsis thaliana</i> ]	2e-056
TC125991	0	9	protein F21D18.16 [imported] - <i>Arabidopsis thaliana</i>	4e-043
TC118434	6	2	putative short-root protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	1e-045
TC118447	8	0	No hits found	
TC115971	7	0	mutant phytoene synthase [ <i>Lycopersicon esculentum</i> ]	1e-033
TC116074	5	2	OSJNBb0072N21.4 [ <i>Oryza sativa</i> (japonica cultivar-group)]	1e-078
TC116477	7	0	ripening regulated protein DDTFR18 [ <i>Lycopersicon esculentum</i> ]	4e-056
TC117075	7	0	No hits found	
TC118596	7	0	putative protein [ <i>Arabidopsis thaliana</i> ]	4e-025
TC123829	5	2	beta-fructosidase [ <i>Lycopersicon esculentum</i> ]	4e-076
TC126635	7	0	3,4-dihydroxy-5-hexaprenylbenzoate methyltransferase (EC 2.1.1.-) precursor [validated] - <i>Arabidopsis thaliana</i>	8e-096
TC126678	7	0	No hits found	
TC115887	6	0	guanylate kinase [ <i>Nicotiana tabacum</i> ]	1e-038
TC116414	6	0	Glutamate decarboxylase (GAD) (ERT D1)	8e-057
TC117579	6	0	Shikimate kinase, chloroplast precursor	e-120
TC118436	5	1	putative protein [ <i>Arabidopsis thaliana</i> ]	7e-083
TC119001	6	0	4-coumarate-CoA ligase-like protein [ <i>Arabidopsis thaliana</i> ]	8e-074
TC119035	6	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	6e-023
TC119047	6	0	trehalose-6-phosphate synthase, putative [ <i>Arabidopsis thaliana</i> ]	1e-093
TC119067	6	0	No hits found	
TC119113	6	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	2e-046
TC119190	5	1	expressed protein [ <i>Arabidopsis thaliana</i> ]	1e-056
TC123892	4	2	acid invertase; AI [ <i>Lycopersicon esculentum</i> ]	0.0
TC125349	6	0	two-component phosphorelay mediator, putative	3e-038
TC125423	6	0	expressed protein [ <i>Arabidopsis thaliana</i> ]	2e-039
TC126866	3	3	gene C-7 protein - common tobacco	1e-097
TC126972	6	0	putative protein [ <i>Arabidopsis thaliana</i> ]	4e-028
TC127132	5	1	putative protein [ <i>Arabidopsis thaliana</i> ]	1e-055
TC116837	5	0	putative protein [ <i>Arabidopsis thaliana</i> ]	1e-012
TC117983	5	0	armadillo repeat-containing protein [ <i>Nicotiana tabacum</i> ]	6e-011
TC118116	5	0	No hits found	
TC118792	5	0	dnaK-type molecular chaperone hsc70-3 - tomato	0.0

TC119435	5	0	No hits found	
TC119509	5	0	ORF_14; similar to NAD dependent epimerase/dehydratase family [Pseudomonas aeruginosa]	2e-048
TC119602	5	0	unknown protein [Arabidopsis thaliana]	e-109
TC119626	4	1	No hits found	
TC119693	5	0	similar to zinc protease PQQL [Oryza sativa (japonica cultivar-group)]	4e-069
TC123919	4	1	Histidine decarboxylase (HDC) (TOM92)	e-125
TC127224	5	0	unknown protein [Arabidopsis thaliana]	3e-048
TC127344	5	0	expressed protein [Arabidopsis thaliana]	1e-020
TC127366	5	0	hypothetical protein [Arabidopsis thaliana]	1e-038
TC127379	5	0	Unknown protein [Oryza sativa (japonica cultivar-group)]	4e-023
TC127387	5	0	Similar to ORF3 protein [Oryza sativa (japonica cultivar-group)]	2e-018
TC127457	5	0	ABC transporter family protein [Arabidopsis thaliana]	3e-096
TC127464	1	4	No hits found	
TC127538	5	0	orf, conserved hypothetical protein [Shigella flexneri 2a str. 301]	2e-027
TC127618	5	0	hypothetical protein [Arabidopsis thaliana]	3e-052
TC127668	5	0	No hits found	
TC115972	4	0	Phytoene synthase 1, chloroplast precursor (Fruit ripening specific protein pTOM5)	3e-032
TC116070	3	1	21K protein precursor - alfalfa	9e-023
TC117234	4	0	unknown protein [Arabidopsis thaliana]	3e-024
TC117595	4	0	unknown protein [Arabidopsis thaliana]	2e-031
TC117618	4	0	ADP-RIBOSYLATION FACTOR -like protein [Arabidopsis thaliana]	5e-040
TC119454	4	0	TPA: putative phytosulfokine peptide precursor [Solanum tuberosum]	2e-017
TC119478	4	0	No hits found	
TC119512	4	0	DEMETER protein [Arabidopsis thaliana]	1e-079
TC119818	4	0	No hits found	
TC119941	0	4	hypothetical protein [Arabidopsis thaliana]	1e-027
TC119974	2	2	unknown protein [Arabidopsis thaliana]	5e-036
TC120104	4	0	No hits found	
TC120125	2	2	VPS13 - like protein [Arabidopsis thaliana]	1e-079
TC120132	4	0	No hits found	
TC120188	4	0	No hits found	
TC120225	4	0	unknown protein [Arabidopsis thaliana]	9e-015
TC120263	4	0	No hits found	
TC120289	4	0	putative lipase [Arabidopsis thaliana]	2e-067
TC120361	0	4	No hits found	
TC120386	4	0	No hits found	
TC120414	4	0	hypothetical protein [Arabidopsis thaliana]	5e-062
TC120419	4	0	No hits found	

TC120424	4	0	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	3e-022
TC120438	4	0	acid phosphatase, putative [Arabidopsis thaliana]	5e-024
TC120453	4	0	hypothetical protein F24B22.200 - Arabidopsis thaliana	2e-029
TC120480	4	0	No hits found	
TC120484	2	2	putative protein [Arabidopsis thaliana]	8e-052
TC123918	4	0	Histidine decarboxylase (HDC) (TOM92)	2e-042
TC126386	4	0	putative protein [Arabidopsis thaliana]	1e-015
TC127888	4	0	putative protein [Arabidopsis thaliana]	4e-032
TC127941	4	0	No hits found	
TC128093	4	0	No hits found	
TC128100	4	0	putative receptor ser/thr protein kinase [Arabidopsis thaliana]	4e-033
TC128108	4	0	hypothetical protein At2g15270 [imported] - Arabidopsis thaliana	9e-023
TC128138	4	0	No hits found	
TC128156	3	1	putative casein kinase I [Arabidopsis thaliana]	e-128
TC128179	4	0	putative clathrin-associated protein [Oryza sativa (japonica cultivar-group)]	1e-087
TC128186	4	0	transposon related protein [Arabidopsis thaliana]	6e-038
TC128189	4	0	COP9 complex subunit, FUS4 [Arabidopsis thaliana]	4e-073
TC128263	4	0	Putative pyridoxamine 5-phosphate oxidase [Oryza sativa]	2e-018
TC128298	4	0	No hits found	
TC128300	4	0	double-stranded RNA-specific editase, putative [Arabidopsis thaliana]	3e-047
TC128327	2	2	No hits found	
TC128336	4	0	putative DNA polymerase I [Oryza sativa (japonica cultivar-group)]	6e-054
TC128356	4	0	No hits found	
TC128367	4	0	No hits found	
TC128370	4	0	expressed protein [Arabidopsis thaliana]	2e-065

- \* T1775: tomato breaker fruit  
T1391: tomato red ripe fruit

### Flower specific genes

TC	T1526	T1527	T1528	T10393	T1529	T10227	Annotation	E value
TC115920	0	0	0	4	47	203	Pectinesterase precursor (Pectin methylesterase) (PE)	e-139
TC123975	0	0	1	0	8	165	L-ascorbate oxidase homolog precursor (Ascorbase)	e-161
TC115991	0	0	0	3	12	143	PGPS/D4 [Petunia x hybrida]	1e-094
TC115985	0	0	0	4	15	130	cell-wall invertase [Lycopersicon esculentum]	0.0
TC115959	0	0	0	5	3	137	Anther specific LAT52 protein precursor	1e-085
TC115714	0	0	0	6	1	101	l-ascorbate oxidase [Petunia x hybrida]	0.0

TC116113	0	0	0	0	0	104	tumor-related protein [ <i>Nicotiana glauca</i> x <i>Nicotiana langsdorffii</i> ]	3e-057
TC124102	0	0	0	4	10	83	putative purple acid phosphatase precursor [ <i>Ipomoea batatas</i> ]	e-117
TC116184	0	0	0	1	0	90	putative protein [ <i>Arabidopsis thaliana</i> ]	e-130
TC116195	0	0	0	5	12	68	endo-1,3-beta-glucanase-like protein [ <i>Pyrus pyrifolia</i> ]	e-105
TC115719	0	0	0	0	5	78	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	0.0
TC116202	0	0	0	0	0	83	cysteine-rich protein [ <i>Nicotiana tabacum</i> ]	1e-012
TC124087	0	0	0	0	5	69	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	2e-071
TC116279	0	0	0	0	0	64	fructokinase [ <i>Lycopersicon esculentum</i> ]	0.0
TC115750	0	0	0	2	18	42	putative beta-galactosidase [ <i>Nicotiana tabacum</i> ]	0.0
TC116336	0	0	0	1	1	54	Anther specific LAT52 protein precursor	2e-049
TC124211	0	0	0	0	1	55	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	3e-077
TC124292	0	0	0	1	8	46	beta-tubulin [ <i>Naegleria gruberi</i> ]	0.0
TC116306	0	0	0	0	0	53	PGPS/D7 [ <i>Petunia x hybrida</i> ]	2e-031
TC116374	0	0	0	2	3	45	fasciclin-like arabinogalactan-protein, putative (FLA3) [ <i>Arabidopsis thaliana</i> ]	3e-031
TC116375	0	0	0	1	8	41	putative glyoxal oxidase [ <i>Oryza sativa</i> (japonica cultivar-group)]	e-143
TC124307	0	0	0	1	0	46	Elicitor inducible gene product Nt-SubE80 [ <i>Nicotiana tabacum</i> ]	1e-049
TC116415	0	0	0	0	0	46	No hits found	
TC124346	0	0	0	1	9	35	pectin methyl-esterase - like protein [ <i>Arabidopsis thaliana</i> ]	e-145
TC116437	0	0	0	0	2	41	LIM domain protein PLIM-2 [ <i>Nicotiana tabacum</i> ]	e-102
TC116304	0	0	0	1	0	40	Expressed protein [ <i>Arabidopsis thaliana</i> ]	9e-069
TC116474	0	0	0	0	0	41	Anther specific LAT52 protein precursor	2e-053
TC124405	0	0	0	1	0	40	No hits found	
TC116541	0	0	0	1	3	33	putative monosaccharide transporter 1 [ <i>Petunia x hybrida</i> ]	0.0
TC115958	0	0	0	0	0	36	Anther specific LAT52 protein precursor	2e-079
TC116143	0	0	0	1	18	16	Probable pectinesterase precursor (Pectin methylesterase) (PE)	2e-065
TC116321	0	0	0	0	3	32	ACTIN	e-115
TC124517	0	0	0	0	4	30	putative pectinesterase [ <i>Arabidopsis thaliana</i> ]	3e-032
TC124625	0	0	0	1	0	30	pollen specific actin-depolymerizing factor 1 [ <i>Nicotiana tabacum</i> ]	4e-064
TC124088	0	0	1	0	1	27	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	3e-079
TC124666	0	0	0	0	4	23	hypothetical protein F14O23.2 [imported] - <i>Arabidopsis thaliana</i>	e-111
TC116797	0	0	0	0	1	25	LIM domain protein PLIM1 [ <i>Nicotiana tabacum</i> ]	e-102
TC124722	0	0	0	4	1	21	phosphate/phosphoenolpyruvate translocator-like protein [ <i>Arabidopsis thaliana</i> ]	e-128
TC116449	0	0	0	1	0	24	Probable pectate lyase P59 precursor	0.0
TC124636	0	0	0	0	9	16	P-type H <sup>+</sup> -ATPase [ <i>Vicia faba</i> ]	0.0
TC116826	0	0	0	0	0	24	No hits found	
TC124646	0	0	0	2	0	22	No hits found	
TC116926	0	0	0	0	5	18	unnamed protein product [ <i>Oryza sativa</i> (japonica cultivar-group)]	1e-074



TC124791	0	0	0	0	2	20	unknown protein [Arabidopsis thaliana]	0.0
TC116915	0	0	0	0	0	21	cysteine-rich protein [Nicotiana tabacum]	3e-013
TC124285	0	0	0	1	0	20	elicitor inducible gene Nt-SA1-1 [Nicotiana tabacum]	3e-033
TC116146	0	0	0	0	13	7	pectinesterase family [Arabidopsis thaliana]	e-106
TC124826	0	0	0	0	3	17	polygalacturonase (EC 3.2.1.15) precursor - common tobacco	0.0
TC116969	0	0	0	0	10	9	putative pectinesterase [Arabidopsis thaliana]	8e-033
TC117094	0	0	0	1	9	9	Probable pectate lyase P59 precursor	0.0
TC116144	0	0	0	0	0	18	pectin methylesterase-like protein [Oryza sativa (japonica cultivar-group)]	3e-012
TC125085	0	0	0	0	0	18	shock protein SRC2 homolog [imported] - Arabidopsis thaliana	3e-022
TC125121	3	0	0	0	0	15	hypothetical protein [Arabidopsis thaliana]	3e-066
TC117074	0	0	0	0	0	17	No hits found	
TC124697	0	0	0	0	1	15	plasma membrane ATPase 3 [Arabidopsis thaliana]	e-134
TC125133	0	0	0	0	0	16	POLLEN-SPECIFIC MEMBRANE INTEGRAL PROTEIN	2e-090
TC124953	0	11	1	4	0	0	anther-specific protein - wood tobacco	0.0
TC117351	0	0	0	0	0	15	No hits found	
TC125301	0	0	0	1	4	9	receptor-like protein kinase PRK2 - tomato	0.0
TC125387	0	0	0	0	0	14	auxin-repressed protein-like protein [Nicotiana tabacum]	6e-033
TC125396	0	0	0	0	1	13	pectinacetyl esterase family [Arabidopsis thaliana]	6e-028
TC115986	0	0	0	0	0	13	cell-wall invertase [Lycopersicon esculentum]	e-156
TC116147	0	0	0	0	0	13	pectinesterase (EC 3.1.1.11) [imported] - Arabidopsis thaliana	2e-065
TC117517	0	0	0	0	3	10	pectinesterase (EC 3.1.1.11) [imported] - Salix gilgiana	3e-015
TC117577	0	0	0	0	0	13	synaptobrevin-like protein [Oryza sativa (japonica cultivar-group)]	4e-085
TC123978	0	0	0	0	0	13	CALMODULIN	7e-080
TC125146	0	0	0	0	0	13	PGPS/NH15 [Petunia x hybrida]	4e-037
TC125363	0	0	0	0	0	13	unknown protein [Arabidopsis thaliana]	9e-023
TC125445	0	0	0	0	0	13	No hits found	
TC125458	0	0	0	0	0	13	No hits found	
TC125549	0	0	0	0	3	10	LIM domain protein PLIM-2 [Nicotiana tabacum]	3e-089
TC125342	0	8	0	5	0	0	glycine-rich protein - tomato	8e-025
TC115749	0	0	0	6	0	6	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	e-115
TC117559	0	0	0	0	12	0	SAM:benzoic acid carboxyl methyltransferase [Antirrhinum majus]	1e-039
TC117668	0	0	0	1	0	11	No hits found	
TC117674	0	0	0	0	0	12	contains similarity to ABC transporter [Arabidopsis thaliana]	1e-044
TC124212	0	0	0	0	0	12	pectinesterase (pectin methylesterase) family [Arabidopsis thaliana]	e-159
TC125621	0	0	0	0	8	4	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-114
TC117643	0	0	1	0	5	6	At2g30360/T9D9.17 [Arabidopsis thaliana]	5e-072
TC116450	0	0	0	1	4	6	Probable pectate lyase P59 precursor	0.0

TC117334	0	0	0	0	11	0	SRG1-like protein [Arabidopsis thaliana]	7e-079
TC117771	0	0	0	0	0	11	No hits found	
TC117894	0	0	0	0	1	10	auxin-regulated protein [Phaseolus vulgaris]	2e-025
TC125702	0	0	0	0	0	11	No hits found	
TC125707	0	0	0	0	4	7	putative protein [Arabidopsis thaliana]	2e-034
TC125715	0	0	0	0	0	11	potassium transporter - like protein [Arabidopsis thaliana]	7e-063
TC125735	0	0	0	0	0	11	hypothetical protein [Arabidopsis thaliana]	3e-014
TC125760	0	0	0	0	5	6	putative protein kinase [Arabidopsis thaliana]	3e-023
TC125777	0	0	0	0	1	10	putative protein destination factor [Arabidopsis thaliana]	2e-054
TC125757	0	8	1	2	0	0	NTM19 [Nicotiana tabacum]	3e-014
TC117964	0	0	0	0	0	10	auxin-regulated protein [Phaseolus vulgaris]	3e-035
TC118016	0	0	0	0	0	10	unknown [Solanum chacoense]	4e-015
TC118069	0	0	0	0	10	0	lectin receptor kinase [Arabidopsis thaliana]	1e-030
TC118076	0	0	0	0	0	10	No hits found	
TC125705	0	0	0	0	0	10	cysteine-rich protein [Nicotiana tabacum]	3e-015
TC126012	0	0	0	0	1	9	peroxidase, putative [Arabidopsis thaliana]	3e-025
TC125964	1	0	0	0	0	9	No hits found	
TC117703	0	0	6	4	0	0	Floral homeotic protein PMADS 2	9e-099
TC116305	0	0	0	0	9	0	PGPS/D7 [Petunia x hybrida]	3e-031
TC117518	0	0	0	1	3	5	hexose transporter - like protein [Arabidopsis thaliana]	e-158
TC118201	0	0	0	0	0	9	hypothetical protein [Arabidopsis thaliana]	3e-035
TC124582	0	0	0	9	0	0	putative alcohol dehydrogenase [Lycopersicon esculentum]	3e-063
TC125763	0	0	0	1	0	8	putative GPI-anchored protein [Arabidopsis thaliana]	8e-019
TC126044	0	0	0	0	1	8	carbonic anhydrase -like protein [Arabidopsis thaliana]	1e-068
TC126079	0	0	0	0	1	8	ACTIN	e-124
TC126101	0	0	0	0	0	9	hypothetical protein [Arabidopsis thaliana]	3e-016
TC126161	0	0	0	0	2	7	No hits found	
TC126222	0	0	0	3	0	6	probable transcription factor SF3 - common tobacco	4e-060
TC125079	0	1	0	0	0	8	putative C-4 sterol methyl oxidase [Arabidopsis thaliana]	e-114
TC118196	0	4	2	3	0	0	ANTHER-SPECIFIC PROTEIN TA-29	6e-050
TC125872	0	5	0	4	0	0	Expressed protein [Arabidopsis thaliana]	4e-063
TC126194	0	5	4	0	0	0	cytochrome P450, putative [Arabidopsis thaliana]	e-101
TC126040	0	6	3	0	0	0	flavonoid 1-2 rhamnosyltransferase [Citrus maxima]	5e-029
TC126185	0	7	1	1	0	0	No hits found	
TC118189	0	9	0	0	0	0	gamma-thionin-like protein precursor - tomato	4e-034
TC118241	0	0	0	0	3	5	peroxidase, putative [Arabidopsis thaliana]	2e-057
TC118298	0	0	0	0	1	7	NTS1 protein [Nicotiana tabacum]	5e-071
TC118398	0	0	0	0	0	8	No hits found	
TC118473	0	0	0	0	1	7	putative peroxidase [Oryza sativa (japonica cultivar-group)]	1e-012

TC118513	0	0	0	0	2	6	aluminium tolerance associated - like protein [Arabidopsis thaliana]	5e-036
TC126268	0	0	0	0	3	5	receptor-like protein kinase 3 [Lycopersicon esculentum]	5e-046
TC126311	0	0	0	0	1	7	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-110
TC126262	0	0	0	0	1	7	bZIP transcriptional activator	4e-24
TC126324	0	0	0	0	0	8	Probable pectate lyase P59 precursor	5e-025
TC126374	0	0	0	0	0	8	No hits found	
TC126394	0	0	0	0	0	8	putative glycosyl transferase [Arabidopsis thaliana]	5e-053
TC118125	0	3	1	4	0	0	PROTEIN 108 PRECURSOR	4e-036
TC118183	0	5	1	2	0	0	5B protein - tomato	6e-048
TC118366	0	7	1	0	0	0	chalcone synthase - like protein [Arabidopsis thaliana]	e-112
TC118530	0	0	0	0	0	7	No hits found	
TC118745	0	0	0	0	1	6	zinc finger - like protein [Arabidopsis thaliana]	6e-032
TC118759	0	0	0	0	1	6	syntaxin SYP124 [Arabidopsis thaliana]	1e-051
TC118785	0	0	0	0	2	5	No hits found	
TC124695	0	0	0	0	2	5	plasma membrane H <sup>+</sup> -ATPase [Oryza sativa (japonica cultivar-group)]	3e-039
TC126373	0	0	0	0	1	6	pectin methylesterase-like protein [Arabidopsis thaliana]	3e-063
TC126497	0	0	0	0	1	6	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR	6e-018
TC126568	0	0	0	0	2	5	expressed protein [Arabidopsis thaliana]	5e-023
TC126596	0	0	0	0	2	5	putative protein [Arabidopsis thaliana]	8e-046
TC126645	0	0	0	0	0	7	GTP-binding protein - garden pea	4e-084
TC118779	0	0	7	0	0	0	No hits found	
TC126623	0	1	4	0	2	0	cytochrome P450, putative [Arabidopsis thaliana]	5e-055
TC126715	0	1	6	0	0	0	male sterility protein 2 - Arabidopsis thaliana	4e-034
TC118159	0	6	1	0	0	0	glycine-rich protein - tomato	7e-019
TC116300	0	0	0	0	0	6	contains similarity to MtN3 protein [Arabidopsis thaliana]	2e-045
TC118503	0	0	0	0	0	6	receptor-like protein kinase PRK1 - tomato	0.0
TC118887	0	0	0	0	3	3	senescence-associated protein-like protein [Oryza sativa (japonica cultivar-group)]	7e-037
TC119008	0	0	0	0	0	6	hypothetical protein [Arabidopsis thaliana]	3e-061
TC119071	0	0	0	0	0	6	putative RHO GDP-dissociation inhibitor 1 [Arabidopsis thaliana]	1e-017
TC119101	0	0	0	0	0	6	peroxidase, putative [Arabidopsis thaliana]	2e-051
TC119107	0	0	0	0	0	6	GTPase activating protein-like [Arabidopsis thaliana]	5e-048
TC119186	0	0	0	0	0	6	No hits found	
TC124600	0	0	0	0	1	5	putative protein [Arabidopsis thaliana]	2e-029
TC126689	0	0	0	0	0	6	heat shock protein DNAJ homolog Pfj4 [Plasmodium falciparum 3D7]	3e-012
TC126763	0	0	0	0	3	3	putative receptor - like kinase [Arabidopsis thaliana]	1e-031
TC127002	0	0	0	0	0	6	β-glucanase [Sorghum bicolor]	3e-018
TC127056	0	0	0	0	1	5	No hits found	
TC127117	0	0	0	0	0	6	No hits found	

TC127131	0	0	0	3	0	3	actin depolymerizing factor-like [Arabidopsis thaliana]	3e-050
TC118869	3	0	0	0	3	0	isoamylase isoform 1 [Solanum tuberosum]	0.0
TC126814	1	0	2	3	0	0	unknown protein [Arabidopsis thaliana]	1e-023
TC126573	3	0	2	0	0	1	Zwille protein [imported] - Arabidopsis thaliana	8e-094
TC119088	0	0	6	0	0	0	Ac-like transposase [Arabidopsis thaliana]	5e-021
TC127144	5	1	0	0	0	0	expressed protein [Arabidopsis thaliana]	6e-026
TC118963	2	1	1	2	0	0	OSJNBa0029H02.4 [Oryza sativa (japonica cultivar-group)]	6e-024
TC127049	3	1	1	0	1	0	putative DNA primase [Oryza sativa (japonica cultivar-group)]	1e-083
TC118664	0	2	0	4	0	0	metallocarboxypeptidase inhibitor IIa precursor [validated] - potato	1e-026
TC119048	0	2	4	0	0	0	putative chloroplast nucleoid DNA-binding protein [Arabidopsis thaliana]	2e-079
TC119161	0	3	3	0	0	0	dihydro-flavanoid reductase-like protein [Zea mays]	e-101
TC119127	0	4	2	0	0	0	putative DNA-binding protein [Arabidopsis thaliana]	1e-029
TC127031	0	5	1	0	0	0	Expressed protein [Arabidopsis thaliana]	2e-053
TC115993	0	0	0	0	0	5	PGPS/D4 [Petunia x hybrida]	3e-041
TC116302	0	0	0	0	5	0	contains similarity to MtN3 protein [Arabidopsis thaliana]	3e-070
TC118049	0	0	0	0	0	5	No hits found	
TC118237	0	0	0	0	0	5	phosphatidic acid phosphatase-like protein [Arabidopsis thaliana]	3e-053
TC119315	0	0	0	0	1	4	Serine/threonine Kinase [Persea americana]	3e-013
TC119368	0	0	0	0	0	5	No hits found	
TC119372	0	0	0	0	0	5	putative high-affinity potassium transporter [Oryza sativa (japonica cultivar-group)]	5e-085
TC119399	0	0	0	0	3	2	expressed protein [Arabidopsis thaliana]	1e-029
TC119514	0	0	0	0	0	5	No hits found	
TC119520	0	0	0	5	0	0	No hits found	
TC119562	0	0	0	0	5	0	unknown protein [Arabidopsis thaliana]	4e-024
TC119617	0	0	0	0	0	5	annexin-like protein [Medicago sativa]	2e-062
TC119632	0	0	0	0	1	4	OJ1117_G01.5 [Oryza sativa (japonica cultivar-group)]	6e-032
TC119683	0	0	0	0	0	5	hypothetical protein F18N11.40 - Arabidopsis thaliana	5e-041
TC119701	0	0	0	5	0	0	No hits found	
TC119746	0	0	0	0	0	5	similar to ARF GAP-like zinc finger-containing protein [Oryza sativa (japonica cultivar-group)]	2e-052
TC124696	0	0	0	0	5	0	plasma membrane H <sup>+</sup> -ATPase [Oryza sativa (japonica cultivar-group)]	4e-042
TC126240	0	0	0	0	5	0	SINA1 protein [imported] - Vitis vinifera	7e-083
TC126916	0	0	0	0	2	3	PGPS/D7 [Petunia x hybrida]	5e-019
TC127363	0	0	0	0	0	5	fasciclin-like arabinogalactan-protein, family (FLA14) [Arabidopsis thaliana]	7e-028
TC127463	0	0	0	0	0	5	putative fatty acid hydroxylase [Oryza sativa (japonica cultivar-group)]	2e-058
TC127523	0	0	0	0	0	5	No hits found	
TC127524	0	0	0	0	1	4	No hits found	
TC127531	0	0	0	4	1	0	Omega-3 fatty acid desaturase, endoplasmic reticulum	1e-056

TC127550	0	0	0	0	0	5	expressed protein [Arabidopsis thaliana]	3e-020
TC127566	0	0	0	0	4	1	sulfate transporter ATST1 [Arabidopsis thaliana]	2e-052
TC127642	0	0	0	0	0	5	No hits found	
TC127692	0	0	0	0	0	5	Na <sup>+</sup> /H <sup>+</sup> antiporter-like protein [Arabidopsis thaliana]	e-126
TC127286	1	0	0	0	4	0	thylakoid lumen 15.0-kDa protein [Arabidopsis thaliana]	4e-077
TC127423	1	0	0	0	0	4	phosphoinositide-specific phospholipase C (EC 3.1.4.-) plc3 - potato	e-112
TC119289	3	0	0	0	2	0	RNA helicase - like protein [Arabidopsis thaliana]	1e-096
TC127670	1	0	3	1	0	0	hypothetical protein [Arabidopsis thaliana]	2e-026
TC119481	2	0	3	0	0	0	Expressed protein [Arabidopsis thaliana]	2e-062
TC119714	2	0	3	0	0	0	Putative ABC transporter [Oryza sativa (japonica cultivar-group)]	e-114
TC127488	2	0	3	0	0	0	No hits found	
TC119336	0	0	5	0	0	0	No hits found	
TC127246	0	0	5	0	0	0	gamma glutamyl hydrolase, putative [Arabidopsis thaliana]	8e-026
TC119487	0	1	0	4	0	0	strictosidine synthase-related [Arabidopsis thaliana]	5e-071
TC119685	0	1	2	0	2	0	No hits found	
TC127297	0	1	4	0	0	0	putative protein [Arabidopsis thaliana]	2e-020
TC119266	3	2	0	0	0	0	No hits found	
TC118686	2	3	0	0	0	0	phosphoribosyl pyrophosphate synthase isozyme 4 [Spinacia oleracea]	e-100
TC124954	0	3	2	0	0	0	anther-specific protein - wood tobacco	9e-097
TC117281	0	4	1	0	0	0	RNA binding protein [Elaeis oleifera]	1e-069
TC119528	0	5	0	0	0	0	gene_id:MTG10.9~pir T10744~similar to unknown protein [Arabidopsis thaliana]	3e-026
TC124320	0	5	0	0	0	0	CycD3;3 [Lycopersicon esculentum]	2e-049
TC116301	0	0	0	0	0	4	No hits found	
TC116303	0	0	0	0	0	4	contains similarity to MtN3 protein [Arabidopsis thaliana]	1e-036
TC117954	0	0	0	0	0	4	3-hydroxy-3-methylglutaryl CoA reductase 2 [Lycopersicon esculentum]	1e-081
TC118050	0	0	0	0	4	0	No hits found	
TC118236	0	0	0	0	1	3	phosphatidic acid phosphatase-like protein [Arabidopsis thaliana]	1e-064
TC119657	0	0	0	0	0	4	dormancy-associated protein, putative [Arabidopsis thaliana]	2e-020
TC119797	0	0	0	0	0	4	No hits found	
TC119918	0	0	0	0	1	3	unknown [Arabidopsis thaliana]	1e-057
TC119921	0	0	0	0	0	4	ABC transporter family protein [Arabidopsis thaliana]	e-135
TC119947	0	0	0	2	0	2	Pectinesterase PPE8B precursor (Pectin methylesterase) (PE)	3e-017
TC119962	0	0	0	0	1	3	No hits found	
TC120035	0	0	0	4	0	0	No hits found	
TC120066	0	0	0	2	0	2	Pectinesterase precursor (Pectin methylesterase) (PE)	3e-064
TC120142	0	0	0	4	0	0	alkylation repair-like protein [Arabidopsis thaliana]	4e-036
TC120145	0	0	0	0	1	3	ABC transporter family protein [Arabidopsis thaliana]	3e-029

TC120168	0	0	0	0	0	4	No hits found	
TC120173	0	0	0	0	0	4	hypothetical protein [Arabidopsis thaliana]	6e-047
TC120178	0	0	0	0	4	0	Phosphatidylinositol Transfer Protein, putative [Arabidopsis thaliana]	e-110
TC120237	0	0	0	0	4	0	hypothetical protein [Arabidopsis thaliana]	1e-040
TC120247	0	0	0	0	4	0	putative uridine kinase [Oryza sativa (japonica cultivar-group)]	9e-013
TC120291	0	0	0	0	4	0	hypothetical protein - carrot	6e-079
TC120313	0	0	0	0	0	4	No hits found	
TC120340	0	0	0	0	0	4	No hits found	
TC120341	0	0	0	0	4	0	Expressed protein [Arabidopsis thaliana]	5e-037
TC120402	0	0	0	0	4	0	AT5g54510/F24B18_13 [Arabidopsis thaliana]	4e-091
TC120462	0	0	0	0	2	2	serine palmitoyltransferase [Solanum tuberosum]	2e-078
TC124308	0	0	0	0	0	4	Elicitor inducible gene product Nt-SubE80 [Nicotiana tabacum]	1e-049
TC125407	0	0	0	4	0	0	NONSPECIFIC LIPID-TRANSFER PROTEIN 2 PRECURSOR (LTP 2)	2e-017
TC125602	0	0	0	0	0	4	putative mitochondrial carrier protein [Oryza sativa (japonica cultivar-group)]	2e-028
TC127427	0	0	0	0	0	4	No hits found	
TC127593	0	0	0	0	0	4	hypothetical protein [Arabidopsis thaliana]	6e-055
TC127710	0	0	0	4	0	0	No hits found	
TC127784	0	0	0	4	0	0	No hits found	
TC127830	0	0	0	2	2	0	aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2A precursor, mitochondrial - common tobacco	1e-054
TC127859	0	0	0	0	0	4	OJ1117_G01.5 [Oryza sativa (japonica cultivar-group)]	5e-012
TC127869	0	0	0	0	0	4	putative Septum-promoting GTP-binding protein 1 (GTPase spg1) [Oryza sativa (japonica cultivar-group)]	8e-028
TC127878	0	0	0	0	0	4	No hits found	
TC127897	0	0	0	3	1	0	hypothetical protein F15K9.23 - Arabidopsis thaliana	1e-014
TC127934	0	0	0	0	0	4	putative protein [Arabidopsis thaliana]	1e-018
TC127955	0	0	0	0	0	4	polygalacturonase, putative [Arabidopsis thaliana]	2e-071
TC127958	0	0	0	1	0	3	expressed protein [Arabidopsis thaliana]	9e-014
TC128063	0	0	0	0	1	3	No hits found	
TC128087	0	0	0	0	2	2	No hits found	
TC128090	0	0	0	4	0	0	kinesin like protein [Arabidopsis thaliana]	2e-019
TC128134	0	0	0	0	0	4	expressed protein [Arabidopsis thaliana]	2e-015
TC128147	0	0	0	4	0	0	globulin-like protein [Arabidopsis thaliana]	1e-049
TC128250	0	0	0	0	4	0	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Clostridium acetobutylicum]	4e-031
TC128259	0	0	0	4	0	0	dihydro-flavanoid reductase-like protein [Zea mays]	3e-038
TC128381	0	0	0	0	0	4	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	6e-073
TC128408	0	0	0	0	4	0	unknown [Arabidopsis thaliana]	1e-044
TC128417	0	0	0	0	0	4	No hits found	
TC128269	1	0	0	0	3	0	putative DnaJ protein [Arabidopsis thaliana]	2e-057

TC128353	1	0	0	0	3	0	hypothetical protein F1B16.12 [imported] - Arabidopsis thaliana	4e-057
TC118428	2	0	0	0	2	0	putative dynamin-like protein ADL2 [Oryza sativa (japonica cultivar-group)]	2e-072
TC128193	3	0	0	1	0	0	Dof zinc finger protein [Arabidopsis thaliana]	1e-026
TC120224	4	0	0	0	0	0	No hits found	
TC118407	0	0	1	3	0	0	putative protein [Arabidopsis thaliana]	5e-079
TC128280	0	0	2	2	0	0	No hits found	
TC119922	0	0	3	0	1	0	T4O12.25 [Arabidopsis thaliana]	4e-076
TC128095	0	0	3	0	1	0	No hits found	
TC128221	0	0	3	1	0	0	Serine protease inhibitor 5 precursor (gCDI-B1)	5e-042
TC120311	0	0	4	0	0	0	putative anthranilate N-hydroxycinnamoyl/benzoyltransferase [Arabidopsis thaliana]	4e-052
TC120464	0	0	4	0	0	0	gibberellin 20-oxidase - common tobacco	e-104
TC120500	0	0	4	0	0	0	No hits found	
TC128295	0	0	4	0	0	0	putative phytochrome-associated protein 3 [Arabidopsis thaliana]	3e-011
TC128354	0	0	4	0	0	0	putative protein kinase [Arabidopsis thaliana]	3e-034
TC128398	0	0	4	0	0	0	No hits found	
TC119866	2	1	0	0	1	0	putative protein [Arabidopsis thaliana]	2e-066
TC128230	3	1	0	0	0	0	No hits found	
TC119040	0	1	3	0	0	0	putative strictosidine synthase [Arabidopsis thaliana]	5e-074
TC128228	0	1	3	0	0	0	cytochrome P450 CYP703A1 - garden petunia	e-114
TC120410	0	2	0	0	2	0	No hits found	
TC115988	1	2	0	0	1	0	No hits found	
TC119919	1	2	0	1	0	0	Floral homeotic protein GLOBOSA	2e-078
TC126543	1	2	0	0	1	0	peptide transporter, putative [Arabidopsis thaliana]	3e-048
TC127774	1	2	0	0	1	0	hypothetical protein T32A16.30 - Arabidopsis thaliana	3e-040
TC126983	2	2	0	0	0	0	myb-related protein - upland cotton	5e-059
TC128199	1	2	1	0	0	0	mitochondrial carrier protein family [Arabidopsis thaliana]	e-104
TC127803	0	2	2	0	0	0	unknown protein [Arabidopsis thaliana]	2e-056
TC120471	0	3	0	0	1	0	No hits found	
TC120316	1	3	0	0	0	0	pectate lyase-like protein [Arabidopsis thaliana]	2e-046
TC127894	1	3	0	0	0	0	hypothetical protein [Arabidopsis thaliana]	1e-012
TC120385	0	3	1	0	0	0	No hits found	
TC128364	0	3	1	0	0	0	No hits found	
TC119205	0	4	0	0	0	0	gamma-thionin [Lycopersicon esculentum]	2e-015
TC120308	0	4	0	0	0	0	No hits found	
TC120498	0	4	0	0	0	0	No hits found	
TC126744	0	4	0	0	0	0	myb-related transcription factor THM18 - tomato	e-129
TC128203	0	4	0	0	0	0	No hits found	
TC128245	0	4	0	0	0	0	No hits found	

- \* T1526: tomato flower, 0-3mm buds  
 T1527: tomato flower, 3-8mm buds  
 T1528: tomato flower, 8mm-preanthesis buds  
 T10393: tomato mixed flower  
 T1529: tomato open flower  
 T10227: wild tomato pollen

### Pollen specific genes

TC	T10393	T1529	T10227	Annotation	E value
TC115920	4	47	203	Pectinesterase precursor (Pectin methylesterase) (PE)	e-139
TC115991	3	12	143	PGPS/D4 [Petunia x hybrida]	1e-094
TC115959	5	3	137	Anther specific LAT52 protein precursor	1e-085
TC115985	4	15	130	cell-wall invertase [Lycopersicon esculentum]	0.0
TC116113	0	0	104	tumor-related protein [Nicotiana glauca x Nicotiana langsdorffii]	3e-057
TC115714	6	1	101	l-ascorbate oxidase [Petunia x hybrida]	0.0
TC116184	1	0	90	putative protein [Arabidopsis thaliana]	e-130
TC124102	4	10	83	putative purple acid phosphatase precursor [Ipomoea batatas]	e-117
TC116202	0	0	83	cysteine-rich protein [Nicotiana tabacum]	1e-012
TC115719	0	5	78	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	0.0
TC124087	0	5	69	pectin acetylerase (EC 3.1.1.-) precursor - mung bean	2e-071
TC116195	5	12	68	endo-1,3-beta-glucanase-like protein [Pyrus pyrifolia]	e-105
TC116279	0	0	64	fructokinase [Lycopersicon esculentum]	0.0
TC124211	0	1	55	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	3e-077
TC116336	1	1	54	Anther specific LAT52 protein precursor	2e-049
TC116306	0	0	53	PGPS/D7 [Petunia x hybrida]	2e-031
TC124292	1	8	46	beta-tubulin [Naegleria gruberi]	0.0
TC124307	1	0	46	Elicitor inducible gene product Nt-SubE80 [Nicotiana tabacum]	1e-049
TC116415	0	0	46	No hits found	
TC116374	2	3	45	fasciclin-like arabinogalactan-protein, putative (FLA3) [Arabidopsis thaliana]	3e-031
TC115750	2	18	42	putative beta-galactosidase [Nicotiana tabacum]	0.0
TC116375	1	8	41	putative glyoxal oxidase [Oryza sativa (japonica cultivar-group)]	e-143
TC116437	0	2	41	LIM domain protein PLIM-2 [Nicotiana tabacum]	e-102
TC116474	0	0	41	Anther specific LAT52 protein precursor	2e-053
TC116304	1	0	40	Expressed protein [Arabidopsis thaliana]	9e-069
TC124405	1	0	40	No hits found	
TC115958	0	0	36	Anther specific LAT52 protein precursor	2e-079
TC124346	1	9	35	pectin methyl-esterase - like protein [Arabidopsis thaliana]	e-145



TC116541	1	3	33	putative monosaccharide transporter 1 [Petunia x hybrida]	0.0
TC116321	0	3	32	ACTIN	e-115
TC124517	0	4	30	putative pectinesterase [Arabidopsis thaliana]	3e-032
TC124625	1	0	30	pollen specific actin-depolymerizing factor 1 [Nicotiana tabacum]	4e-064
TC116797	0	1	25	LIM domain protein PLIM1 [Nicotiana tabacum]	e-102
TC116449	1	0	24	Probable pectate lyase P59 precursor	0.0
TC116826	0	0	24	No hits found	
TC124666	0	4	23	hypothetical protein F14O23.2 [imported] - Arabidopsis thaliana	e-111
TC124646	2	0	22	No hits found	
TC124722	4	1	21	phosphate/phosphoenolpyruvate translocator-like protein [Arabidopsis thaliana]	e-128
TC116915	0	0	21	cysteine-rich protein [Nicotiana tabacum]	3e-013
TC124791	0	2	20	emb/CAB38841.1~gene_id:K10D20.12~similar to unknown protein [Arabidopsis thaliana]	0.0
TC124285	1	0	20	elicitor inducible gene Nt-SA1-1 [Nicotiana tabacum]	3e-033
TC116926	0	5	18	unnamed protein product [Oryza sativa (japonica cultivar-group)]	1e-074
TC116144	0	0	18	contains EST C73169(E3252)-pectin methylesterase-like protein [Oryza sativa (japonica cultivar-group)]	3e-012
TC125085	0	0	18	shock protein SRC2 homolog [imported] - Arabidopsis thaliana	3e-022
TC124826	0	3	17	polygalacturonase (EC 3.2.1.15) precursor (clone G27.1, G27.2) - common tobacco	0.0
TC117074	0	0	17	No hits found	
TC124636	0	9	16	P-type H <sup>+</sup> -ATPase [Vicia faba]	0.0
TC125133	0	0	16	POLLEN-SPECIFIC MEMBRANE INTEGRAL PROTEIN	2e-090
TC124697	0	1	15	plasma membrane ATPase 3 (proton pump) (spIP20431) [Arabidopsis thaliana]	e-134
TC117351	0	0	15	No hits found	
TC125387	0	0	14	auxin-repressed protein-like protein [Nicotiana tabacum]	6e-033
TC125396	0	1	13	pectinacetylerase family [Arabidopsis thaliana]	6e-028
TC115986	0	0	13	cell-wall invertase [Lycopersicon esculentum]	e-156
TC116147	0	0	13	pectinesterase (EC 3.1.1.11) [imported] - Arabidopsis thaliana	2e-065
TC117577	0	0	13	synaptobrevin-like protein [Oryza sativa (japonica cultivar-group)]	4e-085
TC123978	0	0	13	CALMODULIN	7e-080
TC125146	0	0	13	PGPS/NH15 [Petunia x hybrida]	4e-037
TC125363	0	0	13	unknown protein [Arabidopsis thaliana]	9e-023
TC125445	0	0	13	No hits found	
TC125458	0	0	13	No hits found	
TC117674	0	0	12	contains similarity to ABC transporter [Arabidopsis thaliana]	1e-044
TC124212	0	0	12	pectinesterase (pectin methylesterase) family [Arabidopsis thaliana]	e-159
TC117668	1	0	11	No hits found	
TC117771	0	0	11	No hits found	
TC125702	0	0	11	No hits found	
TC125715	0	0	11	potassium transporter - like protein [Arabidopsis thaliana]	7e-063

TC125735	0	0	11	hypothetical protein [Arabidopsis thaliana]	3e-014
TC117517	0	3	10	pectinesterase (EC 3.1.1.11) [imported] - Salix gilgiana	3e-015
TC125549	0	3	10	LIM domain protein PLIM-2 [Nicotiana tabacum]	3e-089
TC117894	0	1	10	auxin-regulated protein [Phaseolus vulgaris]	2e-025
TC125777	0	1	10	putative protein destination factor [Arabidopsis thaliana]	2e-054
TC117964	0	0	10	auxin-regulated protein [Phaseolus vulgaris]	3e-035
TC118016	0	0	10	unknown [Solanum chacoense]	4e-015
TC118076	0	0	10	No hits found	
TC125705	0	0	10	cysteine-rich protein [Nicotiana tabacum]	3e-015
TC125301	1	4	9	receptor-like protein kinase PRK2 - tomato	0.0
TC126012	0	1	9	peroxidase, putative [Arabidopsis thaliana]	3e-025
TC118201	0	0	9	hypothetical protein [Arabidopsis thaliana]	3e-035
TC126101	0	0	9	hypothetical protein [Arabidopsis thaliana]	3e-016
TC125763	1	0	8	putative GPI-anchored protein [Arabidopsis thaliana]	8e-019
TC126044	0	1	8	carbonic anhydrase -like protein [Arabidopsis thaliana]	1e-068
TC126079	0	1	8	ACTIN	e-124
TC118398	0	0	8	No hits found	
TC126324	0	0	8	Probable pectate lyase P59 precursor	5e-025
TC126374	0	0	8	No hits found	
TC126394	0	0	8	putative glycosyl transferase [Arabidopsis thaliana]	5e-053
TC125707	0	4	7	putative protein [Arabidopsis thaliana]	2e-034
TC126161	0	2	7	No hits found	
TC118298	0	1	7	NTS1 protein [Nicotiana tabacum]	5e-071
TC118473	0	1	7	putative peroxidase [Oryza sativa (japonica cultivar-group)]	1e-012
TC126311	0	1	7	glycosyl hydrolase family 17 [Arabidopsis thaliana]	e-110
TC126262	0	1	7	bZIP transcriptional activator	4e-24
TC118530	0	0	7	No hits found	
TC126645	0	0	7	GTP-binding protein - garden pea	4e-084
TC115749	6	0	6	POLLEN-SPECIFIC PROTEIN NTP303 PRECURSOR	e-115
TC116450	1	4	6	Probable pectate lyase P59 precursor	0.0
TC125760	0	5	6	putative protein kinase [Arabidopsis thaliana]	3e-023
TC126222	3	0	6	probable transcription factor SF3 - common tobacco	4e-060
TC118513	0	2	6	aluminium tolerance associated - like protein [Arabidopsis thaliana]	5e-036
TC118745	0	1	6	zinc finger - like protein [Arabidopsis thaliana]	6e-032
TC118759	0	1	6	syntaxin SYP124 [Arabidopsis thaliana]	1e-051
TC126373	0	1	6	pectin methylesterase-like protein [Arabidopsis thaliana]	3e-063
TC126497	0	1	6	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR	6e-018
TC116300	0	0	6	contains similarity to MtN3 protein~gene_id:MQB2.17 [Arabidopsis thaliana]	2e-045
TC118503	0	0	6	receptor-like protein kinase PRK1 - tomato	0.0

TC119008	0	0	6	hypothetical protein [Arabidopsis thaliana]	3e-061
TC119071	0	0	6	putative RHO GDP-dissociation inhibitor 1 [Arabidopsis thaliana]	1e-017
TC119101	0	0	6	peroxidase, putative [Arabidopsis thaliana]	2e-051
TC119107	0	0	6	GTPase activating protein-like [Arabidopsis thaliana]	5e-048
TC119186	0	0	6	No hits found	
TC126689	0	0	6	heat shock protein DNAJ homolog Pfj4 [Plasmodium falciparum 3D7]	3e-012
TC127002	0	0	6	3-glucanase [Sorghum bicolor]	3e-018
TC127117	0	0	6	No hits found	
TC117518	1	3	5	hexose transporter - like protein [Arabidopsis thaliana]	e-158
TC118241	0	3	5	peroxidase, putative [Arabidopsis thaliana]	2e-057
TC126268	0	3	5	receptor-like protein kinase 3 [Lycopersicon esculentum]	5e-046
TC118785	0	2	5	No hits found	
TC124695	0	2	5	plasma membrane H <sup>+</sup> -ATPase [Oryza sativa (japonica cultivar-group)]	3e-039
TC126568	0	2	5	expressed protein [Arabidopsis thaliana]	5e-023
TC126596	0	2	5	putative protein [Arabidopsis thaliana]	8e-046
TC124600	0	1	5	putative protein [Arabidopsis thaliana]	2e-029
TC127056	0	1	5	No hits found	
TC115993	0	0	5	PGPS/D4 [Petunia x hybrida]	3e-041
TC118049	0	0	5	No hits found	
TC118237	0	0	5	phosphatidic acid phosphatase-like protein [Arabidopsis thaliana]	3e-053
TC119368	0	0	5	No hits found	
TC119372	0	0	5	putative high-affinity potassium transporter [Oryza sativa (japonica cultivar-group)]	5e-085
TC119514	0	0	5	No hits found	
TC119617	0	0	5	annexin-like protein [Medicago sativa]	2e-062
TC119683	0	0	5	hypothetical protein F18N11.40 - Arabidopsis thaliana	5e-041
TC119746	0	0	5	similar to ARF GAP-like zinc finger-containing protein [Oryza sativa (japonica cultivar-group)]	2e-052
TC127363	0	0	5	fasciclin-like arabinogalactan-protein, family (FLA14) [Arabidopsis thaliana]	7e-028
TC127463	0	0	5	putative fatty acid hydroxylase [Oryza sativa (japonica cultivar-group)]	2e-058
TC127523	0	0	5	No hits found	
TC127550	0	0	5	expressed protein [Arabidopsis thaliana]	3e-020
TC127642	0	0	5	No hits found	
TC127692	0	0	5	Na <sup>+</sup> /H <sup>+</sup> antiporter-like protein [Arabidopsis thaliana]	e-126
TC119315	0	1	4	Serine/threonine Kinase [Persea americana]	3e-013
TC119632	0	1	4	OJ1117_G01.5 [Oryza sativa (japonica cultivar-group)]	6e-032
TC127524	0	1	4	No hits found	
TC116301	0	0	4	No hits found	
TC116303	0	0	4	contains similarity to MtN3 protein [Arabidopsis thaliana]	1e-036
TC117954	0	0	4	3-hydroxy-3-methylglutaryl CoA reductase 2 [Lycopersicon esculentum]	1e-081

TC119657	0	0	4	dormancy-associated protein, putative [Arabidopsis thaliana]	2e-020
TC119797	0	0	4	No hits found	
TC119921	0	0	4	ABC transporter family protein [Arabidopsis thaliana]	e-135
TC120168	0	0	4	No hits found	
TC120173	0	0	4	hypothetical protein [Arabidopsis thaliana]	6e-047
TC120313	0	0	4	No hits found	
TC120340	0	0	4	No hits found	
TC124308	0	0	4	Elicitor inducible gene product Nt-SubE80 [Nicotiana tabacum]	1e-049
TC125602	0	0	4	putative mitochondrial carrier protein [Oryza sativa (japonica cultivar-group)]	2e-028
TC127427	0	0	4	No hits found	
TC127593	0	0	4	hypothetical protein [Arabidopsis thaliana]	6e-055
TC127859	0	0	4	OJ1117_G01.5 [Oryza sativa (japonica cultivar-group)]	5e-012
TC127869	0	0	4	putative Septum-promoting GTP-binding protein 1 (GTPase spg1) [Oryza sativa (japonica cultivar-group)]	8e-028
TC127878	0	0	4	No hits found	
TC127934	0	0	4	putative protein [Arabidopsis thaliana]	1e-018
TC127955	0	0	4	polygalacturonase, putative [Arabidopsis thaliana]	2e-071
TC128134	0	0	4	expressed protein [Arabidopsis thaliana]	2e-015
TC128381	0	0	4	pectin acetyltransferase (EC 3.1.1.-) precursor - mung bean	6e-073
TC128417	0	0	4	No hits found	
TC126916	0	2	3	PGPS/D7 [Petunia x hybrida]	5e-019
TC118236	0	1	3	phosphatidic acid phosphatase-like protein [Arabidopsis thaliana]	1e-064
TC119918	0	1	3	unknown [Arabidopsis thaliana]	1e-057
TC119962	0	1	3	No hits found	
TC120145	0	1	3	ABC transporter family protein [Arabidopsis thaliana]	3e-029
TC127958	1	0	3	expressed protein [Arabidopsis thaliana]	9e-014
TC128063	0	1	3	No hits found	

\* T10393: tomato mixed flower

T1529: tomato open flower

T10227: wild tomato pollen

### Suspension culture specific genes

TC	T10060	Annotation	E value
TC125357	9	cim1 protein - soybean	2e-042
TC126302	8	No hits found	
TC118654	7	No hits found	
TC118700	7	No hits found	

TC126796	6	No hits found	
TC126835	6	Expressed protein [Arabidopsis thaliana]	1e-015
TC126904	6	unknown protein [Arabidopsis thaliana]	8e-023
TC127155	6	No hits found	
TC118904	5	F21F23.21 protein - Arabidopsis thaliana	3e-022
TC119328	5	No hits found	
TC119340	5	predicted by genscan and genefinder [Arabidopsis thaliana]	9e-028
TC119342	5	No hits found	
TC119367	5	S-locus protein 8 [Brassica rapa]	3e-015
TC119384	5	protein kinase precursor - like [Arabidopsis thaliana]	2e-040
TC119422	5	SEC14 cytosolic factor, putative [Arabidopsis thaliana]	1e-061
TC119531	5	No hits found	
TC119664	5	probable beta-fructofuranosidase, [imported] - Arabidopsis thaliana	3e-073
TC126812	5	hypothetical protein [Oryza sativa (japonica cultivar-group)]	5e-011
TC127321	5	No hits found	
TC127519	5	No hits found	
TC127643	5	cytochrome p450 family [Arabidopsis thaliana]	4e-054
TC115738	4	No hits found	
TC116036	4	glutathione transferase (EC 2.5.1.18), class-phi - Commerson's wild potato	5e-046
TC119758	4	No hits found	
TC119775	4	No hits found	
TC119826	4	putative protein [Arabidopsis thaliana]	9e-029
TC119859	4	No hits found	
TC119889	4	No hits found	
TC119898	4	No hits found	
TC119905	4	putative protein [Arabidopsis thaliana]	3e-018
TC119960	4	No hits found	
TC119965	4	putative carnitine/acylcarnitine translocase [Oryza sativa]	3e-041
TC120082	4	No hits found	
TC120212	4	No hits found	
TC120236	4	putative polyprotein [Oryza sativa (japonica cultivar-group)]	3e-014
TC120266	4	hypothetical protein [Arabidopsis thaliana]	1e-012
TC120269	4	orf105b [Beta vulgaris]	4e-016
TC120347	4	putative leucine rich repeat containing protein kinase [Oryza sativa (japonica cultivar-group)]	6e-020
TC120411	4	No hits found	
TC120417	4	D2 protein, photosystem II subunit [Antirrhinum majus]	1e-031
TC120429	4	CDK5 activator-binding protein-like [Arabidopsis thaliana]	1e-017
TC120433	4	putative receptor like kinase [Arabidopsis thaliana]	2e-020
TC120443	4	No hits found	

TC120478	4	AT5g39790/MKM21_80 [Arabidopsis thaliana]	2e-011
TC127935	4	No hits found	
TC127969	4	No hits found	
TC128035	4	expressed protein [Arabidopsis thaliana]	6e-043
TC128124	4	No hits found	
TC128146	4	RRM-containing protein [Arabidopsis thaliana]	6e-034
TC128152	4	No hits found	
TC128164	4	retrotransposon del1-46	1e-020
TC128197	4	short chain alcohol dehydrogenase-like [Arabidopsis thaliana]	4e-040
TC128200	4	Cc11 [Brassica napus]	e-117
TC128209	4	unknown [Prunus armeniaca]	1e-032
TC128235	4	No hits found	
TC128241	4	No hits found	
TC128301	4	unknown [Arabidopsis thaliana]	8e-055
TC128407	4	putative protein [Arabidopsis thaliana]	4e-059
TC128423	4	expressed protein [Arabidopsis thaliana]	2e-023
TC128428	4	No hits found	
TC128430	4	No hits found	

\* T10600: tomato suspension culture

### Root specific genes

TC	T1481	T1480	T1450	T1482	Annotation	E value
TC124822	4	10	4	4	putative oxidoreductase [Arabidopsis thaliana]	9e-078
TC124865	8	4	3	3	nitrate transporter [Nicotiana tabacum]	0.0
TC125532	13	0	0	0	unknown protein [Arabidopsis thaliana]	2e-021
TC117718	6	5	0	0	probable peroxidase (EC 1.11.1.7) (clone PC36) - spinach	e-115
TC125947	1	2	1	1	specific tissue protein 2 [Cicer arietinum]	2e-036
TC118229	3	0	0	0	unknown [Capsicum annuum]	5e-076
TC118389	1	3	0	0	putative protein [Arabidopsis thaliana]	2e-038
TC126033	4	3	0	0	No hits found	
TC126334	1	3	2	2	peroxidase (EC 1.11.1.7) TPX2 precursor - tomato	0.0
TC126658	0	4	0	0	peroxidase (EC 1.11.1.7) TPX2 precursor - tomato	5e-097
TC117091	3	3	0	0	No hits found	
TC118796	2	3	1	1	probable germin protein - tomato	e-121
TC118925	3	1	0	0	glutamate-ammonia ligase (EC 6.3.1.2) 1-5, cytosolic - common tobacco	e-116
TC118956	0	2	1	1	peroxidase; peroxidase ATP19a [Arabidopsis thaliana]	1e-067
TC126059	4	1	0	0	Metallothionein-like protein type 2	7e-024

TC126928	0	3	3	3	putative serine carboxypeptidase I [Arabidopsis thaliana]	1e-085
TC119170	2	1	0	0	putative glutathione S-transferase T4 [Lycopersicon esculentum]	e-130
TC119269	1	1	3	3	bZIP transcription factor [Arabidopsis thaliana]	2e-015
TC119314	1	0	2	2	peroxidase, putative [Arabidopsis thaliana]	7e-062
TC119641	1	1	0	0	expressed protein [Arabidopsis thaliana]	2e-028
TC126091	0	0	0	0	extensin-like protein Ext1 [Lycopersicon esculentum]	0.0
TC126256	0	2	0	0	unknown [Petroselinum crispum]	9e-057
TC127422	0	0	0	0	major latex-like protein [Prunus persica]	9e-035
TC127437	4	0	1	1	putative cytochrome P450 [Nicotiana tabacum]	1e-042
TC127546	5	0	0	0	Hyoscyamine 6-dioxygenase (Hyoscyamine 6-beta-hydroxylase)	2e-048
TC127645	0	3	1	1	putative protein [Arabidopsis thaliana]	1e-049
TC119049	2	2	0	0	specific tissue protein 2 [Cicer arietinum]	1e-036
TC119654	2	2	0	0	subtilisin-like protease [Lycopersicon esculentum]	0.0
TC119793	3	0	1	1	potassium transporter - like protein [Arabidopsis thaliana]	3e-037
TC119796	1	2	0	0	dehydration-induced protein RD22-like protein [Gossypium hirsutum]	2e-049
TC120172	2	1	1	1	peroxidase (EC 1.11.1.7) TPX2 precursor - tomato	3e-071
TC120205	0	0	0	0	No hits found	
TC120320	3	0	0	0	alcohol NADP+ oxidoreductase [Solanum tuberosum]	1e-063
TC124037	1	1	0	0	water channel protein [Nicotiana excelsior]	e-159
TC127872	0	0	0	0	putative ring-H2 zinc finger protein [Oryza sativa (japonica cultivar-group)]	3e-012
TC127902	0	1	3	3	cytochrome P450 monooxygenase CYP72A26 [Zea mays subsp. mays]	4e-030
TC127951	0	1	0	0	expressed protein [Arabidopsis thaliana]	1e-043
TC127979	4	0	0	0	oxygenase, pathogen-induced - common tobacco	e-106
TC128085	0	4	0	0	protein BYJ15 - common tobacco	2e-037
TC128191	2	2	0	0	wound-induced protein Sn-1, vacuolar membrane - pepper	7e-050
TC128229	2	0	1	1	No hits found	
TC128271	0	4	0	0	putative nitrate transporter [Oryza sativa (japonica cultivar-group)]	5e-052
TC128405	1	1	1	1	OSJNBa0072F16.9 [Oryza sativa (japonica cultivar-group)]	4e-054
TC128433	3	1	0	0	putative flavanone 3-hydroxylase [Saussurea medusa]	4e-053

- \* T1481: tomato root, plant at pre-anthesis  
T1480: tomato root, plant at fruit set  
T1450: tomato nutrient deficient root  
T1482: tomato root, etiolated radicle

### Callus specific genes

TC	T1207	Annotation	E value
TC124683	28	Probable serine protease inhibitor 6 precursor (AM66)	8e-057

TC124726	26	unknown protein [Arabidopsis thaliana]	e-100
TC125084	17	glucosyltransferase NTGT3 [Nicotiana tabacum]	e-139
TC116197	15	SIEP1L protein precursor - beet	2e-076
TC118240	9	expressed protein [Arabidopsis thaliana]	7e-075
TC118491	8	gibberellin 2-oxidase-like protein [Pisum sativum]	9e-073
TC126469	8	patatin-like protein 1 [Nicotiana tabacum]	e-160
TC118608	7	myb-related transcription factor LBM1 [Nicotiana tabacum]	e-120
TC116224	6	unknown [Prunus armeniaca]	5e-067
TC117925	6	subtilisin-like proteinase (EC 3.4.21.-) - tomato	0.0
TC118979	6	similar to axi 1 protein from Nicotiana tabacum [Arabidopsis thaliana]	2e-016
TC119056	5	Phenylalanine ammonia-lyase (PAL)	0.0
TC119441	5	WRKY family transcription factor [Arabidopsis thaliana]	2e-052
TC119624	5	cytochrome P450, putative [Arabidopsis thaliana]	e-108
TC124716	5	Histidine decarboxylase (HDC) (TOM92)	3e-043
TC127658	5	symbiotic ammonium transport protein SAT1 - soybean	3e-065
TC116539	4	putative chloroplast 50S ribosomal protein L6 [Oryza sativa (japonica cultivar-group)]	2e-013
TC119895	4	helix-loop-helix-like protein [Cucumis melo]	2e-014
TC119933	4	putative reverse transcriptase [Arabidopsis thaliana]	3e-019
TC119985	4	ankyrin-like protein [Arabidopsis thaliana]	e-158
TC120081	4	putative protein kinase [Oryza sativa]	2e-015
TC120086	4	putative protein [Arabidopsis thaliana]	8e-036
TC120174	4	glycosyl hydrolase family 29 (alpha-L-fucosidase) [Arabidopsis thaliana]	6e-044
TC120299	4	putative receptor serine/threonine kinase [Oryza sativa (japonica cultivar-group)]	6e-045
TC120336	4	(+)-delta-cadinene synthase isozyme A (D-cadinene synthase)	2e-096
TC120473	4	Probable WRKY transcription factor 61 (WRKY DNA-binding protein 61)	3e-040
TC126747	4	germin -like protein [Arabidopsis thaliana]	9e-080
TC127994	4	DRE binding protein [Arabidopsis thaliana]	4e-024
TC128198	4	putative protein [Arabidopsis thaliana]	4e-042
TC128211	4	No hits found	
TC128264	4	hypothetical protein [Arabidopsis thaliana]	2e-048
TC128270	4	NAM (no apical meristem)-like protein [Arabidopsis thaliana]	1e-046
TC128339	4	No hits found	
TC128391	4	No hits found	
TC128453	4	expressed protein [Arabidopsis thaliana]	1e-028

\* T1027: tomato callus tissue



## Trichome specific genes

TC	T1451	T1452	Annotation	E value
TC116136	0	69	sesquiterpene synthase 1 [ <i>Lycopersicon hirsutum</i> ]	e-136
TC116138	0	10	sesquiterpene synthase 1 [ <i>Lycopersicon hirsutum</i> ]	e-103
TC123854	6	0	Chlorophyll A-B binding protein 3C, chloroplast precursor	3e-087
TC116140	0	5	sesquiterpene synthase 2 [ <i>Lycopersicon hirsutum</i> ]	e-111
TC127295	3	2	cytochrome p450, putative [ <i>Arabidopsis thaliana</i> ]	2e-056
TC115990	2	2	No hits found	
TC116221	1	3	No hits found	
TC117335	0	4	SRG1-like protein [ <i>Arabidopsis thaliana</i> ]	2e-041
TC120092	4	0	extra sporogenous cells [ <i>Arabidopsis thaliana</i> ]	e-103
TC120309	0	4	cytochrome P450 [ <i>Petunia x hybrida</i> ]	2e-062
TC124550	0	4	dehydrodolichyl diphosphate synthase, putative [ <i>Arabidopsis thaliana</i> ]	5e-031
TC127760	1	3	putative protein [ <i>Arabidopsis thaliana</i> ]	2e-017

\* T1451: *L. pennellii* trichome  
T1452: *L. hirsutum* trichome

## Leaf specific genes

TC	T1079	T1080	T1297	Annotation	E value
TC116065	3	2	3	chlorophyll a/b-binding protein type I precursor (cab-6A) - tomato	7e-059
TC126286	4	2	2	photoassimilate-responsive protein PAR-1c precursor - common tobacco	2e-076
TC118638	1	4	1	unknown protein [ <i>Arabidopsis thaliana</i> ]	4e-018
TC119090	1	1	4	unknown protein [ <i>Arabidopsis thaliana</i> ]	5e-056
TC123882	2	1	3	PsbQ domain protein family [ <i>Arabidopsis thaliana</i> ]	1e-043
TC127230	0	4	0	No hits found	
TC127553	0	2	2	photosystem II protein D1 precursor - common tobacco chloroplast	e-112
TC127729	3	0	1	heat shock protein, putative [ <i>Arabidopsis thaliana</i> ]	2e-040
TC127822	0	1	3	putative protein [ <i>Arabidopsis thaliana</i> ]	1e-041
TC127892	0	2	2	ferredoxin [ <i>Arabidopsis thaliana</i> ]	1e-046

\* T1079: tomato *pseudomonas* susceptible  
T1080: tomato *pseudomonas* resistant  
T1297: tomato mixed elicitor

## Shoot specific genes

TC	T1005	T10304	Annotation	E value
TC118360	1	6	pathogenesis-related protein 5-1 [ <i>Helianthus annuus</i> ]	2e-092
TC119456	1	3	unknown [ <i>Arabidopsis thaliana</i> ]	2e-093

\* T1005: tomato shoot

T10304: tomato shoot/meristem

## Seed specific genes

TC	T1437	T1048	Annotation	E value
TC118369	7	0	endo-beta-mannanase [ <i>Lycopersicon esculentum</i> ]	0
TC127599	4	0	mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - tomato	e-115

\* T1437: tomato germinating seed

T1048: tomato quiescent seed

## Crown gall specific genes

TC	T10284	Annotation	E value
TC124438	21	antifungal protein [ <i>Capsicum annum</i> ]	2e-020

\* T10284: tomato crown gall

**VITA**

NAME: Zhangjun Fei  
PERMANENT ADDRESS: 134 Graham Road #3B1  
Ithaca, NY14853

**EDUCATION BACKGROUND**

Degree: Master of Science  
Major subject: Plant Physiology  
University: China Agricultural University  
Date of graduation: July 1998

Degree: Bachelor of Science  
Major subject: Horticulture  
University: China Agricultural University  
Date of graduation: July 1993

**PROFESSIONAL EXPERIENCE**

Research Assistant, Texas A&M University, with Dr. Jim Giovannoni, Sept. 1998 – August 2003

Research Assistant, China Agricultural University, with Dr. Dapeng Zhang, Sept. 1995 – July 1998

Research Assistant, Nanjing Agricultural Institute, Jiangsu, China, July 1993 – August 1995