The Plant Journal (2004) **40**, 47–59 doi: 10.1111/j.1365-313X.2004.02188.x

Comprehensive EST analysis of tomato and comparative genomics of fruit ripening

Zhangjun Fei[1,2], Xuemei Tang[1], Rob M. Alba[1], Joseph A. White[3], Catherine M. Ronning[3], Gregory B. Martin[1,4], Steven D. Tanksley[5] and James J. Giovannoni[1,6,*]

[1]*Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA,*
[2]*Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843, USA,*
[3]*The Institute for Genome Research, Rockville, MD 20850, USA,*
[4]*Department of Plant Pathology, Cornell University, Ithaca, NY 14853, USA,*
[5]*Department of Plant Breeding, Cornell University, Ithaca, NY 14853, USA, and*
[6]*United States Department of Agriculture, Plant, Soil, and Nutrition Laboratory, Ithaca, NY 14853, USA*

Received 19 March 2004; revised 26 June 2004; accepted 2 July 2004.
*For correspondence (fax +1 607 254 2958; e-mail jjg33@cornell.edu).

Summary

A large tomato expressed sequence tag (EST) dataset (152 635 total) was analyzed to gain insights into differential gene expression among diverse plant tissues representing a range of developmental programs and biological responses. These ESTs were clustered and assembled to a total of 31 012 unique gene sequences. To better understand tomato gene expression at a plant system 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. *Arabidopsis* and grape EST collections were also characterized to facilitate cross-species comparisons where possible. Tomato fruit digital expression data was specifically compared with publicly available grape EST data to gain insight into molecular manifestation of ripening processes across diverse taxa and resulted in identification of common transcription factors not previously associated with ripening.

Keywords: tomato, gene expression, expressed sequence tag, fruit ripening, comparative genomics.

Introduction

Expressed sequence tags (ESTs) are created by partially sequencing randomly isolated gene transcripts that have been converted into cDNA (Adams *et al.*, 1992). ESTs have played significant roles in accelerating gene discovery including gene family expansion (Bourdon *et al.*, 2002; Rogaev *et al.*, 1995), large-scale expression analysis (Ewing *et al.*, 1999; Ogihara *et al.*, 2003; Ronning *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 sequencing projects. Currently there are near 20 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 likely increase substantially providing additional opportunities for intra and inter-specific expression comparisons on a genomics scale.

It has been shown previously that EST databases are a valid and reliable source of gene expression data (Ewing *et al.*, 1999; Ogihara *et al.*, 2003; Ronning *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; Ogihara *et al.*, 2003; Ronning *et al.*, 2003).

Tomato is a member of the family Solanaceae that includes several additional economically important crops such as potato, pepper, and eggplant and as such represents the most valuable plant family in terms of vegetable crops with important contributions to human health and nutrition. 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. Recently 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; Strausberg *et al.*, 2001). Here we present correlated expression profile for 6758 genes across 25 different tomato tissues. We highlight highly expressed genes in various tomato tissues and compare with corresponding *Arabidopsis* and grape tissues where possible. In addition, we identified sets of differentially expressed genes associated with plant development with an emphasis on fruit ripening. We also performed comparative analysis with available grape ESTs derived from fruit to gain insights into common and distinctive features of these diverse fruit species at the molecular level and identified a set of transcription factors induced during both tomato and grape ripening that had not previously been associated with ripening.

Results

EST foundation for tomato digital expression analysis

A total of 154 054 high quality sequences representing 152 635 distinct ESTs from 27 different cDNA libraries (Table 1) were generated under two NSF-funded tomato functional genomics projects (see Acknowledgements). After clustering and assembly, these ESTs were reduced to 31 012 unique genes, of which 15 925 are Tentative Consensus sequences (TCs) and 15 087 are singletons. The tomato unigene set is available through both TIGR (http://www.tigr.org/tdb/tgi/lgi) and the Solanaceae Genome Network (http://www.sgn.cornell.edu).

Table 1 Tomato cDNA library statistics

Category no.	Library description	No. of ESTs	No. of TCs	No. of singletons (%[a])
T1045	Fruit, ovary	9878	3072	623 (16.9)
T10018	Fruit, developing/immature green	4240	2287	376 (14.1)
T1356	Fruit, mature green	5317	2540	658 (20.6)
T1775	Fruit, breaker	15207	4022	815 (16.8)
T1391	Fruit, red ripe	3895	1605	238 (12.9)
T1526	Flower, 0–3 mm buds	6259	3018	802 (21.0)
T1527	Flower, 3–8 mm buds	5524	2597	523 (16.8)
T1528	Flower, 8 mm preanthesis buds	5759	2665	575 (17.7)
T1529	Flower, anthesis (open flower)	5643	2630	659 (20.0)
T10393	Flower, mixed stages	7009	2799	884 (24.0)
T10227	Flower, wild tomato pollen	5426	796	440 (35.6)
T1079	Leaf, *Pseudomonas* susceptible	5243	2469	485 (16.4)
T1080	Leaf, *Pseudomonas* resistant	5127	2289	438 (16.1)
T1297	Leaf, mixed elicitor-induced	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	Root, nutrient deficient	3175	1788	506 (22.1)
T1482	Root, etiolated radicle	2373	1436	205 (12.5)
T1437	Seed, germinating	3927	2004	331 (14.2)
T1048	Seed, quiescent	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	14114	4440	1288 (22.5)
T10600	Suspension culture, untreated	8026	2821	579 (17.0)
T10284	Crown gall	5204	2749	624 (18.5)
Total		152 635		

[a]Calculated as the ratio of the number of singletons to the number of total unique sequences in the corresponding library.

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 are difficult to characterize further via digital expression analysis in a statistically meaningful manner because of their rarity. 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 1). In contrast, we identified the most abundant transcripts in the combined tomato EST database and in each tissue, in addition to tissue-specific genes [defined as genes with a minimum of four EST members from a single tissue ($P < 0.05$; Audic and Claverie, 1997)]. The list of all these genes is available at the tomato digital expression database (http://ted.bti.cornell.edu).

Comparative analysis of tomato and Arabidopsis digital expression data

Digital expression analysis allows for genomics-scale cross-species comparisons that cannot be conducted via conventional wet-lab genomics approaches. For example, digital expression analysis can provide gene expression information that can be used to facilitate functional comparison of gene family members across species and on a genomics scale.

To demonstrate the utility of digital expression analysis for cross-species comparisons, we compared tomato EST collections to EST collections from equivalent *Arabidopsis* tissues. The TIGR *Arabidopsis* Gene Index contains EST sequence derived from five tissues that we compared with similar tomato EST collections. It should be noted that this comparison can be made because the compared libraries have not been normalized, subtracted or otherwise altered in their ability to represent steady-state gene expression of the tissues examined. The *Arabidopsis* EST collections that we compared with tomato include leaf (5452 ESTs), root (16 646 ESTs), flower buds (7679 ESTs), green silique (fruit) (13 045 ESTs), and seed (9936 ESTs). During comparative digital expression analysis, the raw EST count was normalized for the proportion of total ESTs. We initiated our analysis by (i) identifying tomato TCs with at least five EST members in the five comparable tissues and (ii) determined their corresponding AtGI top hits (comparing translated amino acids with an e-value of 1e-80 or less). The expression profile of tomato TCs and those of their corresponding AtGI hits across the five tissues were compared using the Pearson correlation coefficient (r) as a measurement of similarity. We identified 74 tomato TCs displaying a statistically similar expression profile in the five tissues analyzed as compared with *Arabidopsis* homologues ($r > 0.9$). In this highest quality data subset, the 74 corresponding *Arabidopsis* best hits also contained at least five EST members. The list of these genes is available at http://ted.bti.cornell.edu. The following examples demonstrate the value of large-scale comparative sequence and expression analysis in assessing functionally conserved homologs across species. In Figure 1(a) tomato TC116371 (likely 14-3-3 protein) has high-sequence similarity to all of its top six AGI hits (e-value less than e-130) making it difficult to identify the corresponding *Arabidopsis* gene likely to be most functionally similar based on sequence alone. However, TC116371 expression in the five comparable tissues is only similar to that of *Arabidopsis* TC183911 ($r = 0.9689$). Combined sequence and digital expression information can thus assist in identification of functionally conserved homologs across species.

Figure 1(b) shows that tomato TC117718 (likely peroxidase) is much more conserved at the sequence level to *Arabidopsis* TC180382 (e-136) than TC183774 (2e-91), yet its expression profile is more similar to TC183774, suggesting that while the sequence of TC183774 is more diverged, it is likely a better candidate for the homolog of TC117718 with biologically conserved function. Finally, Figure 1(c) shows that tomato TC124329 (likely GDSL-motif lipase/hydrolase) is highly conserved in both sequence and expression profile with both of its top *Arabidopsis* hits. While this particular result does not further facilitate elucidation of corresponding orthologs, it does suggest that additional homologous gene(s) may remain to be identified in tomato (whose genome remains to be sequenced), a hypothesis that would not have been a logical extension of examination of the tomato data alone. In addition to the 74 genes for which high-quality data is available, lower significance correlation can be viewed for additional genes (2889 from tomato and 2340 from *Arabidopsis*) at http://ted.bti.cornell.edu.

Coordinated gene expression analysis of tomato

To generate expression patterns of numerous genes across various tissues and to assess the overall similarities and differences between transcriptomes of different tissues/organs, we performed coordinated gene expression analysis of the tomato EST collection. Only cDNA libraries with more than 1000 independent ESTs and TCs represented by at least five individual ESTs were selected for analysis, as described previously (Ewing *et al.*, 1999). A total of 6758 genes and 25 cDNA libraries, respectively, met these criteria.

Figure 2(a) is the complete cluster of all 6758 genes across 25 tissues. Red regions specific to each tissue represent tissue/treatment-specific gene expression in the context of the 25 tissues analyzed. For example, the red region for pollen in Figure 2(a) represents 226 individual unigenes.

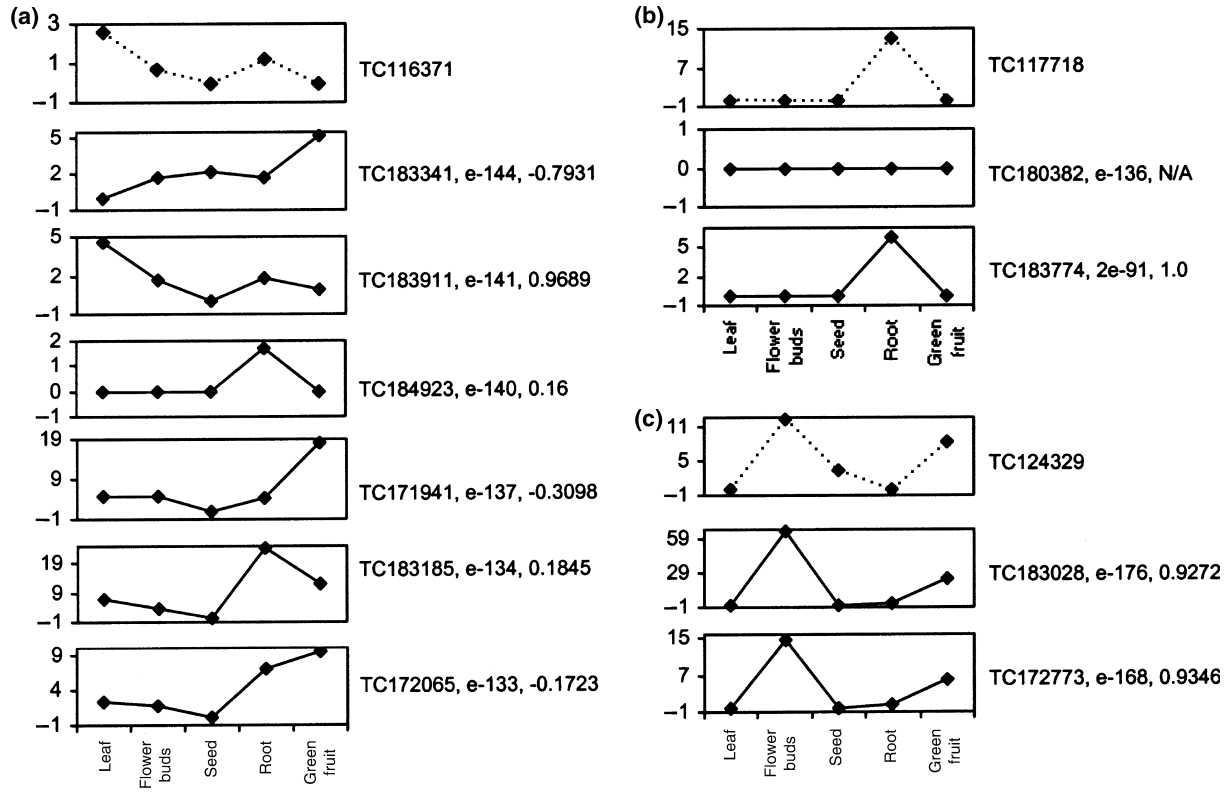


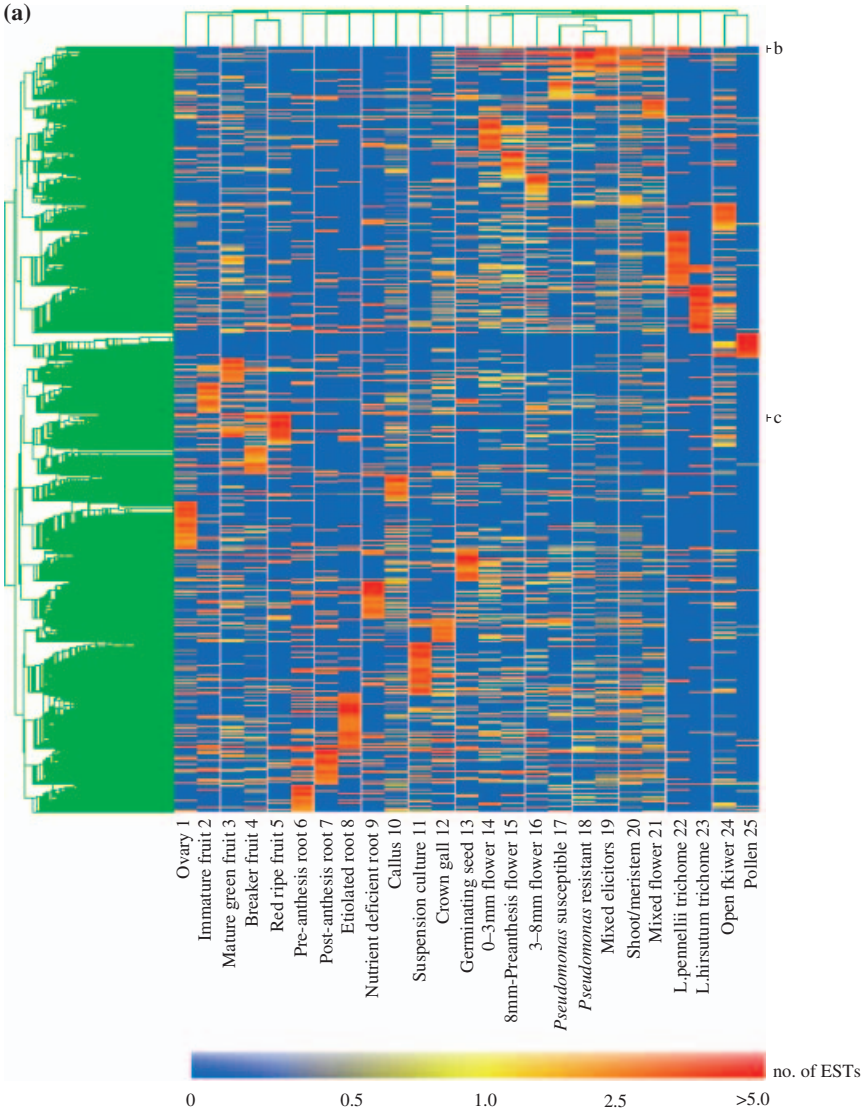
Figure 1. Comparative digital expression analysis between tomato and *Arabidopsis* gene family members.
Expression profiles of tomato TC116371 (14-3-3 protein) and its top six AGI hits (a), TC117718 (peroxidase) and its top two AGI hits (b) and TC124329 (GDSL-motil lipase/hydrolyase) and its top two AGI hits (c) are shown. For each AGI hit, the e-value (obtained by comparing its sequence with the corresponding tomato TC sequence at translated sequence level) and the Pearson correlation coefficient (measuring the degree of expression profile similarity) are listed.

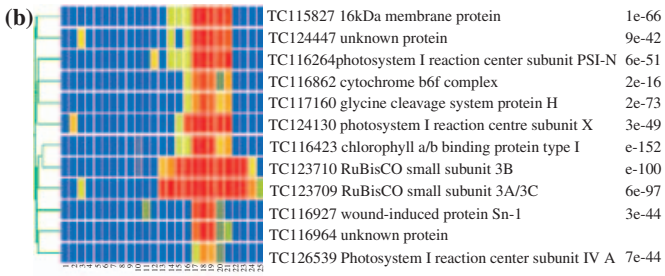
Many have been previously documented to be pollen-specific or pollen-related genes including LePRK1 and LePRK2 (TC118503 and TC125301; Muschietti *et al.*, 1998), LAT52 (TC115959; Tang *et al.*, 2002), and LeFRK4 (TC116279; German *et al.*, 2002).

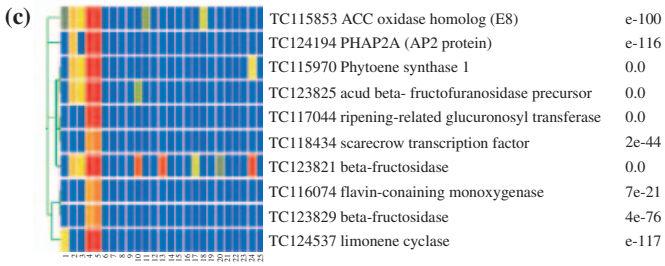
Correlations of genes with development or biology of specific tissues was observed and two examples are shown in Figure 2. Figure 2(b) depicts a cluster of genes highly expressed in leaf and shoot tissues, many of which have homology to genes involved in photosynthesis. Figure 2(c) shows genes highly expressed in ripening fruit. Most genes in this cluster have previously documented roles in fruit ripening [e.g. E8 (Deikman and Fischer, 1988), β-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 and thus may represent candidates for further analysis.

Gene expression profiles from each tissue/treatment were also clustered (Figure 2a). EST collections derived from the most similar tissues or representing a developmental continuum typically clustered together. All five-fruit, four-root, three-leaf and one shoot and two trichome libraries clustered most tightly together, respectively. Furthermore, most libraries from the two sets of EST libraries representing developmental time courses (i.e. fruit and flower development) clustered in developmental order, further validating the analysis methods employed. The *pseudomonas*-resistant leaf library and mixed elicitors-treated leaf library also clustered most tightly. The only exception was for the open flower library that clustered most tightly and separately with the pollen EST collection (as opposed to the other floral libraries).

Figure 2. Correlated expression analysis of tomato EST data.
(a) 6758 TCs with at least five members and 25 cDNA libraries with at least 1000 ESTs were used for clustering. Both TCs and cDNA libraries were clustered using Gene Tree and Experiment Tree programs in GeneSpring software (gene tree not shown). Differences in gene expression are shown in color as per the lower scale.
(b) Subcluster of genes that are highly expressed in leaf and shoot tissues.
(c) Subcluster of genes that are highly expressed in ripening fruit (breaker and red ripe fruit). In (b) and (c), the expression profile, putative function and e-value are shown for each gene. Pearson correlation coefficients of all the gene expression profiles within both subclusters are >0.9.


(a)
b
c
Ovary 1
Immature fruit 2
Mature green fruit 3
Breaker fruit 4
Red ripe fruit 5
Pre-anthesis root 6
Post-anthesis root 7
Etiolated root 8
Nutrient deficient root 9
Callus 10
Suspension culture 11
Crown gall 12
Germinating seed 13
0–3mm flower 14
8mm-Preanthesis flower 15
3–8mm flower 16
Pseudomonas susceptible 17
Pseudomonas resistant 18
Mixed elicitors 19
Shoot/meristem 20
Mixed flower 21
L.pennellii trichome 22
L.hirsutum trichome 23
Open fkiwer 24
Pollen 25
no. of ESTs
0
0.5
1.0
2.5
>5.0


(b)
TC115827 16kDa membrane protein 1e-66
TC124447 unknown protein 9e-42
TC116264photosystem I reaction center subunit PSI-N 6e-51
TC116862 cytochrome b6f complex 2e-16
TC117160 glycine cleavage system protein H 2e-73
TC124130 photosystem I reaction centre subunit X 3e-49
TC116423 chlorophyll a/b binding protein type I e-152
TC123710 RuBisCO small subunit 3B e-100
TC123709 RuBisCO small subunit 3A/3C 6e-97
TC116927 wound-induced protein Sn-1 3e-44
TC116964 unknown protein
TC126539 Photosystem I reaction center subunit IV A 7e-44


(c)
TC115853 ACC oxidase homolog (E8) e-100
TC124194 PHAP2A (AP2 protein) e-116
TC115970 Phytoene synthase 1 0.0
TC123825 acud beta- fructofuranosidase precursor 0.0
TC117044 ripening-related glucuronosyl transferase 0.0
TC118434 scarecrow transcription factor 2e-44
TC123821 beta-fructosidase 0.0
TC116074 flavin-conaining monoxygenase 7e-21
TC123829 beta-fructosidase 4e-76
TC124537 limonene cyclase e-117

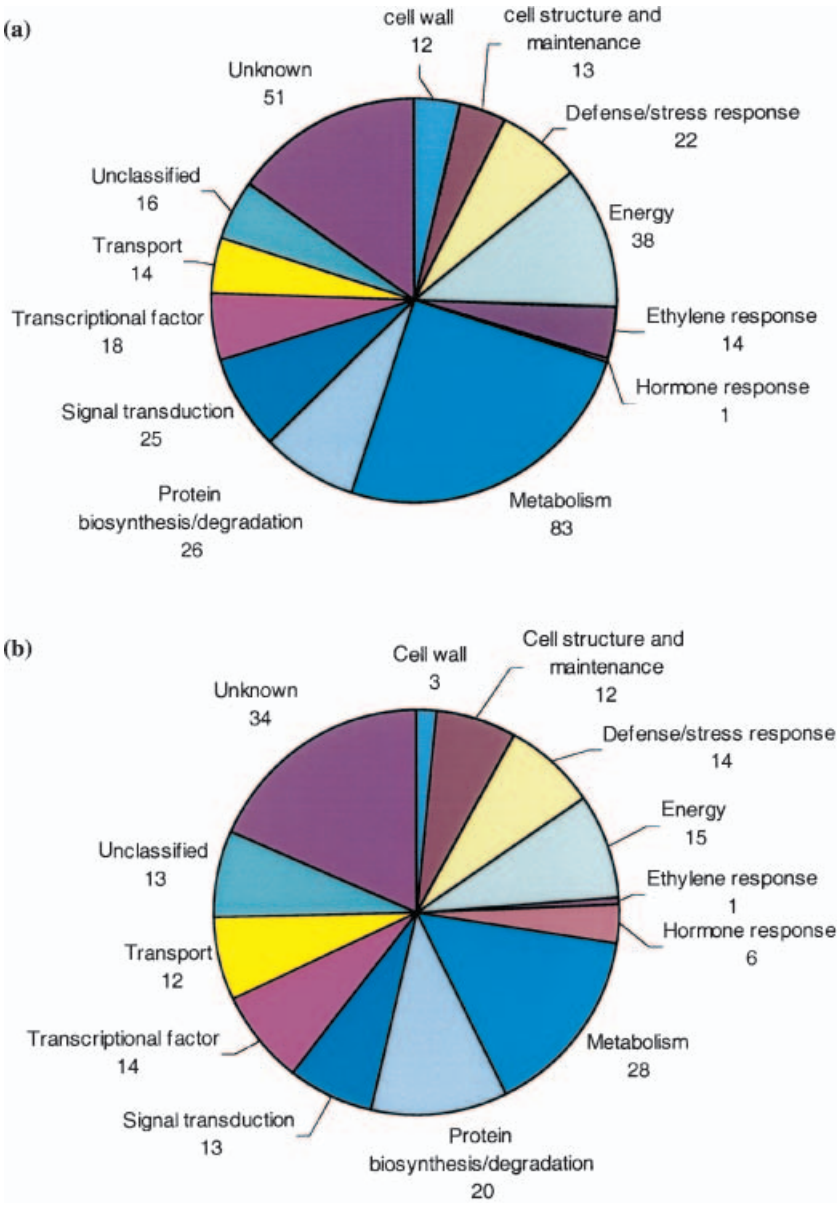


Figure 3. Distribution of tomato ripening-related TCs by different functional categories.
A total of 333 ripening-induced TCs (a) and 185 ripening-repressed TCs (b) were annotated by blast against the GenBank nr database and then classified into corresponding categories according to their putative functions. Number in each category represents the number of ripening-induced (a) and ripening-repressed (b) genes classified into the corresponding category.

Identification of tomato fruit ripening-related genes

We initially focused on ripening for a more detailed analysis of differential gene expression, with emphasis on 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 (Giovannoni, 2001; Seymour *et al.*, 1993 and references therein).

We compared tomato expression in mature green fruit (2–3 days prior to the onset of ripening) to that in breaker fruit (fruit showing the first visible signs of ripening). Whether or not expression was detected in other tissues was not a factor in this analysis. A total of 333 ripening-induced genes (greater expression in breaker, $P < 0.05$) and 185 ripening-repressed genes (greater expression in mature green fruit, $P < 0.05$) were identified.

The combined 518 ripening-related genes were subdivided into 13 categories according to putative function based on DNA sequence homology (Figure 3 and Tables S1 and S2). 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, including the possibility of error propagation because of electronic 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), 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; Cordes *et al.*, 1989; Giovannoni *et al.*, 1999).

Among the ripening-induced genes, 14 share homology with sequences involved in ethylene biosynthesis or signal transduction. Within this group two are previously uncharacterized EREBP genes (TC116548 and TC116831). Twelve ripening-induced genes putatively encode loci involved in cell wall metabolism, 22 are related to defense or stress response sequences, and 11 are homologous to genes involved in fruit pigmentation.

Of the 185 ripening-repressed genes, only three are associated with cell wall metabolism, one with ethylene response (TC116320, an EREBP gene) and none were identified having clear relationships to fruit pigmentation, as would be anticipated. In addition, 14 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 (HSP) 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 β subunit of polygalacturonase isoenzyme 1 (PG1), previously reported to be downregulated during fruit ripening and consistent with its role in limiting pectin depolymerization (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, 18 putative transcription factors in addition to three EREBPs were induced during ripening and 14 transcription factors in addition to one EREBP displayed expression negatively correlated with ripening, suggesting a large pool of putative regulatory elements for future functional analysis. Furthermore, 51 genes with no known or putative function are ripening-induced while 34 are ripening-repressed, suggesting an additional pool of genes for further analysis.

Digital expression profiling permits inter-species comparisons on a genomics scale: 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 require this hormone for completion of the ripening process (Biale and Young, 1981). Tomato is a climacteric fruit while grape is considered to be a non-climacteric fruit (Davies and Robinson, 2000). Common ripening regulatory mechanisms operating in both climacteric and non-climacteric fruit remain elusive (Vrebalov *et al.*, 2002). In order to gain insight into common regulatory mechanisms among diverse fruit species and ripening physiologies, we compared EST collections from ripening grape and tomato fruits.

The TIGR Grape Gene Index includes 83 675 sequences representing 66 501 ESTs and 17 091 unique genes. A total of 4126 ESTs in VvGI are from green fruit (stage II) and 5167 are from veraison stage fruit. Veraison represents the onset of grape ripening and thus is approximately comparable with breaker stage of tomato.

In comparing ESTs from green versus veraison fruit, 95 grape ripening-induced genes were identified at $P < 0.05$ (Table S3). Nine of the 95 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). Fourteen ethylene synthesis/signaling genes were identified via digital expression analysis of tomato, while none was revealed by analysis of grape ESTs.

The cDNA sequences of all 95 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). Twenty-three tomato ripening-induced genes had homologues in the set of grape-induced genes (Table 2). This number is an overestimate as several pairs of related tomato ripening genes were homologous to a single grape ripening gene. Interestingly, three transcription factors, including members of the MADS box, zinc finger, and bZIP transcription factor families, are among those identified in both species.

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. A number of genes associated with these ripening traits were identified in both species (Table 2), including xyloglucan endo-1,4-β-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 2). We additionally identified 181 grape ripening-repressed genes at $P < 0.05$ (Table S4) and 35 ripening-repressed genes from

Table 2 Homologous genes induced by both tomato and grape ripening

Tomato TC	Grape TC	E-value[a]	Putative function of tomato TC
TC125305	TC4377	4e-055	MADS box transcription factor
TC125359	TC4377	3e-072	MADS box transcription factor
TC124244	TC4730	9e-025	bZIP transcription factor
TC124112	TC4730	2e-042	bZIP transcription factor
TC124196	TC9044	1e-025	Putative zinc finger transcription factor
TC125034	TC9044	2e-025	Similar to zinc finger-like protein
TC116030	TC4282	5e-033	Xyloglucan endo-1,4-β-D-glucanase
TC115998	TC4249	2e-080	Pathogenesis-related protein
TC123883	TC4394	0.0	Alcohol dehydrogenase 2
TC124274	TC4394	e-160	Alcohol dehydrogenase class III
TC116962	TC4046	6e-060	Heat shock protein 17.6
TC116318	TC4181	8e-052	Class I small heat shock protein
TC116319	TC4193	2e-053	HSP20.0 protein
TC124903	TC4348	9e-075	HSP70
TC126297	TC4348	0.0	HSP70
TC126413	TC4348	1e-017	HSP70
TC124001	TC9134	9e-041	Heat shock protein MTSHP precursor
TC123771	TC4209	0.0	Elongation factor 1-α
TC115895	TC4236	0.0	Ubiquitin
TC125239	TC4910	4e-090	Calcineurin B-like protein 1
TC123982	TC9086	e-102	Calmodulin
TC124929	TC9284	2e-020	Copper-binding protein family
TC124731	TC9962	8e-034	Endoplasmatic reticulum retrieval protein

[a]This column represents the e-values obtained by comparing translated amino acid sequences of tomato TCs with the corresponding grape TCs using TBLASTX program.

tomato with homology to ripening-repressed genes from grape (Table S5).

Online database – tomato digital expression database

The tomato EST digital expression data can be accessed online through the tomato digital expression database at http://ted.bti.cornell.edu. A relational database has been implemented, so users can query the data in multiple ways. Several interactive tools for expression data manipulation and analysis have been developed including tools for identifying differentially expressed genes between any two tissues, identifying additional genes with expression patterns similar to any selected gene of interest, and comparative gene expression of tomato and *Arabidopsis* homologues.

Discussion

Utility of digital expression profiling

In this study, we performed large-scale transcriptome analysis and identified differentially expressed genes among diverse plant tissues and across species by using available EST information as a source of expression data for digital expression profiling. Available EST data allowed us to focus on fruit development and ripening. Fruit are major contributors of vitamins, fiber, carbohydrates, and phytonutrient compounds in the diet. The nutritional importance of fruit and vegetables is reflected in current USDA recommendations of five or more servings of fruit or vegetables a day for a healthy diet. The World Health Organization and the United Nations Food and Agriculture Organization (FAO) recently launched an effort to enhance fruit and vegetable consumption worldwide as low consumption is considered one of the top 10 contributing factors to human mortality (http://www.fao.org/english/newsroom/news/2003/24439-en.html).

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 (Kuo *et al.*, 2002; Lipshutz *et al.*, 1999) 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. While the data resulting from cross-species comparisons is limited to the number of tissues that can be compared, it can be highly useful, for example, in building the case for functional homology across species

by providing a second criteria of comparison in addition to sequence homology.

Approach and analysis

Relative expression patterns of 6758 tomato genes across 25 different tissues/treatments were developed (Figure 2a). Unlike other studies which employ raw EST counts (Ewing *et al.*, 1999; Ogihara *et al.*, 2003; Ronning *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 Experimental procedures) to correct for the different sampling sizes of the various EST collections. 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. The normalization strategy employed in this study significantly improved the clustering (data not shown). 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 2b,c).

Gene expression patterns of 25 different tissues/treatments were clustered according to their digital profiles (Figure 2a). 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 (0–3 mm flower buds, 3–8 mm flower buds, and 8 mm 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 (Honys and Twell, 2003) provides an explanation for why these two tissues represented an outgroup in this analysis.

Digital gene expression analysis of fruit ripening

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, 2004). 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.

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.

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 several 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 expansions 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 HSPs were identified through digital expression analysis. It has recently been proposed that HSPs play roles 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. Most of these genes were first associated with ripening in this study. Among this group of transcription factors, four belong to the EREBP family, of which three are ripening-induced (TC116548, TC116831, and TC116368) and one is ripening-repressed (TC116320). It has been reported that EREBPs act both as transcriptional activators and repressors in plants (Fujimoto *et al.*, 2000). The fact that we identified both ripening-induced and repressed EREBPs fits such a model in which EREBPs dynamically regulate fruit ripening utilizing antagonistic mechanisms. While EREBP

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 the group of tomato ripening-induced genes. The association of a MADS box gene 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 (Tables S1 and S2), have been found to regulate various processes of plant and animal development, while a relationship between these transcription factors and fruit ripening has not been previously documented.

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 (proof will, of course, require further study).

Additional expression analyses

We also identified 931 and 250 differentially expressed genes during the five fruit development stages (ovary, immature green, mature green, breaker, and red ripe) and the four flower development stages (0–3, 3–8, 8 mm preanthesis, and open flower), respectively, all at a significance level of $P < 0.01$. 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 1). A total of 169 differentially expressed genes ($P < 0.05$) were identified. Information for all these genes and their putative functions based on sequence homology is listed in http://ted.bti.cornell.edu. In addition, we performed digital expression analysis on tomato flavonoid biosynthetic pathway genes (Figure S1 and Table S6). Digital expression data is consistent with that for several previously reported flavonoid pathway genes (Muir *et al*., 2001; Verhoeyen *et al*., 2002) and preliminary data for several additional genes that have not been previously characterized in ripening fruit is now available through digital expression analysis. This information may be useful for further characterization of flavonoid pathway genes and eventual manipulation of flavonoid levels in ripening fruit.

Digital expression analysis facilitates cross-specifies comparisons

Tomato is a climacteric fruit requiring ethylene for ripening, while grape is non-climacteric and the role of ethylene in its maturation remains uncertain. Nevertheless, the ripening processes of both share several common features, including tissue 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 analysis of the public grape EST collection, we identified 95 grape ripening-induced and 181 ripening-repressed genes and compared them with tomato ripening-related genes at the translated amino acid level. Homologous genes induced by both grape and tomato fruit ripening are shown in Table 2. Of note are three transcription factors, including MADS box, zinc finger, and bZIP transcription factor representatives. 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). Although TC4377 is homologous to tomato TC125305 (TDR4) and TC125359 (Table 2), 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 may have evolved functions 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, although this analysis suggests their possible roles in a conserved mechanism of ripening control transcending climacteric and non-climacteric distinctions.

Limitations of digital expression analysis

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 tomato fruit ripening (e.g. lycopene β-cyclase and lycopene epsilon-cyclase; Ronen *et al*., 1999) may not be revealed by this approach because of the limited availability of corresponding ESTs from tissues analyzed. In addition, genes that are induced by both tomato and grape fruit ripening and expressed at low levels or in specific tissue or cell types not emphasized in EST sequencing efforts may be

missed. It is also important to note that in the analysis described here we focused on only those genes with highly conserved predicted peptide sequences. Most certainly there are additional functionally equivalent genes that were missed because of higher level of sequence divergence. 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 allows researchers to take advantage of the expression data recovered through digital expression analysis of the tomato EST collection. Continued collection of ESTs and their analysis for expression within and among species holds the promise of becoming a more widely used component of annotation in assessing gene function.

Experimental procedures

cDNA library construction

Twenty-seven non-normalized nor subtracted tomato cDNA libraries were constructed as described previously (Van der Hoeven *et al.*, 2002). Sequencing of cDNA clones and construction of the tomato unigene build were as described in Van der Hoeven *et al.* (2002).

Datasets acquisition and data analyses

The EST datasets and the Unigene datasets of the TIGR Tomato Gene Index release 9.0 (LeGI; http://www.tigr.org/tdb/tgi/lgi), the TIGR *Arabidopsis* Gene Index release 10.0 (AtGI; http://www.tigr.org/tdb/tgi/agi), and the TIGR Grape Gene Index release 2.0 (VvGI; http://www.tigr.org/tdb/tgi/vvgi/) were provided through the TIGR ftp server. ESTs from subtracted and normalized libraries were eliminated for further digital expression analysis. For clones sequenced from both ends, only the 5′ sequence for each EST in LeGI and VvGI was used, while only the 3′ sequence for each EST in AtGI was used, as these are the predominant sequence information for the corresponding EST collections.

The unigene dataset from LeGI, AtGI, and VvGI were electronically annotated by performing sequence similarity searches against the NCBI nr database using the BLASTX program with a cutoff e-value at 1e-10. BLASTX analysis was performed on the Cornell Computational Biology Service Unit (CBSU) parallel blast driver (pblast; http://www.ser-loopp.tc.cornell.edu/cbsu/pblast.htm) on a MS Windows PC computer cluster (128 Pentium III 1 GHz CPUs) at the Cornell Theory Center.

Coordinated gene expression analysis of tomato

cDNA libraries with more than 1000 ESTs and TCs 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 TCs (genes) and columns corresponding to cDNA libraries. Raw data were normalized by: (i) 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. (ii) 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 (v6.0; Silicon Genetics, Redwood City, CA, USA). In both classifications, the Pearson correlation was used as a measurement of similarity.

Identification of differentially expressed genes

AC statistics (Audic and Claverie, 1997) and general chi-square 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-square test were calculated using the IDEG6 program (http://telethon.bio.unipd.it/bioinfo/IDEG6/).

Acknowledgements

The authors are grateful to Dr Cornelius Barry and Dr Suma Chakravarthy for useful discussion, and to Dr Qi Sun for providing access to the pblast driver. This work was supported by grants from the National Science Foundation (DBI-9872617 and DBI-0116076 to ST, GM, JG and DBI-9813392 to TIGR).

Supplementary material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ2188/TPJ2188sm.htm

Figure S1. Schematic illustration of the flavonoid biosynthesis pathway.
CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; F3H, flavanone 3-hydroxylase; F3′H, flavanone 3′-hydroxylase; FLS, flavonol synthase.

Table S1. Tomato ripening-induced genes

Table S2. Tomato ripening-repressed genes

Table S3. Grape ripening-induced genes

Table S4. Grape ripening-repressed genes

Table S5. Genes repressed by both tomato and grape ripening

Table S6. Expression of the flavonoid biosynthetic pathway genes in tomato fruits

References

Adams, M.D., Kelley, J.M., Gocayne, J.D. *et al.* (1992) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651–1656.

Aharoni, A., Keizer, L., Bouwmeester, H. *et al.* (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell*, **12**, 647–662.

Audic, S. and Claverie, J.M. (1997) The significance of digital gene expression profiles. *Genome Res.* **7**, 986–995.

Barry, C.S., Blume, B., Bouzayen, M., Cooper, W., Hamilton, A.J. and Grierson, D. (1996) Differential expression of the 1-aminocyclo-

propane-1-carboxylate oxidase gene family of tomato. *Plant J.* **9**, 525–535.

Bartley, G.E., Viitanen, P.V., Bacot, K.O. and Scolnik, P.A. (1992) A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthesis pathway. *J. Biol. Chem.* **267**, 5036–5039.

Biale, J. and Young, R. (1981) Respiration and ripening in fruits-retrospect and prospect. In *Recent Advance in the Biochemistry of Fruit and Vegetables* (**Friend, J. and Rhodes, M.**, eds). London: Academic Press, pp. 1–39.

Bird, C.R., Smith, C.J., Ray, J.A., Moureau, P., Bevan, M.W., Bird, A.S., Hughes, S., Morris, P.C., Grierson, D. and Schuch, W. (1988) The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. *Plant Mol. Biol.* **11**, 651–662.

Boss, P.K., Sensi, E., Hua, C., Davies, C. and Thoma, M.R. (2002) Cloning and characterisation of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci.* **162**, 887–895.

Bourdon, V., Naef, F., Rao, P.H., Reuter, V., Mok, S.C., Bosl, G.J., Koul, S., Murty, V.V., Kucherlapati, R.S. and Chaganti, R.S. (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.

Cordes, S., Deikman, J., Margossian, L.J. and Fischer, R.L. (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, J.C., Chen, F. and 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. and Robinson, S.P. (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. and Fischer, R.L. (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.

Elliott, K.J., Butler, W.O., Dickinson, C.D., Konno, Y., Vedvick, T.S., Fitzmaurice, L. and Mirkov, T.E. (1993) Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. *Plant Mol. Biol.* **21**, 515–524.

Ewing, R.M., Kahla, A.B., Poirot, O., Lopez, F., Audic, S. and Claverie, J.M. (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.* **9**, 950–959.

Ferrie, B.J., Beaudoin, N., Burkhart, W., Bowsher, C.G. and Rothstein, S.J. (1994) The cloning of two tomato lipoxygenase genes and their differential expression during fruit ripening. *Plant Physiol.* **106**, 109–118.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H. and 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.

German, M.A., Dai, N., Chmelnitsky, I., Sobolev, I., Salts, Y., Barg, R., Schaffer, A.A. and Granot, D. (2002) LeFRK4, a novel tomato (*Lycopersicon esculentum* Mill) fructokinase specifically expressed in stamens. *Plant Sci.* **163**, 607–613.

Giovannoni, J.J. (2001) Molecular biology of fruit maturation and ripening. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 725–749.

Giovannoni, J.J. (2004) Genetic regulation of fruit development and ripening. *Plant Cell*, **16**, S170–S180.

Giovannoni, J., Yen, H., Shelton, B., Miller, S., Vrebalov, J., Kannan, P., Tieman, D., Hackett, R., Grierson, D. and Klee, H. (1999) Genetic mapping of ripening and ethylene-related loci in tomato. *Theor. Appl. Genet.* **98**, 1005–1013.

Greller, L.D. and Tobin, F.L. (1999) Detecting selective expression of genes and proteins. *Genome Res.* **9**, 282–296.

Honys, D. and Twell, D. (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**, 640–652.

Ishimaru, M. and Kobayash, S. (2002) Expression of a xyloglucan endo-transglycosylase gene is closely related to grape berry softening. *Plant Sci.* **162**, 621–628.

Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. and Kohane, I.S. (2002) Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics*, **18**, 405–412.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**(Suppl. 1), 20–24.

Marín-Rodríguez, M.C., Orchard, J. and Seymour, G.B. (2002) Pectate lyases, cell wall degradation and fruit softening. *J. Exp. Bot.* **53**, 2115–2119.

Muir, S.R., Collins, G.J., Robinson, S., Hughes, S., Bovy, A., Ric De Vos, C.H., van Tunen, A.J. and Verhoeyen, M.E. (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. and 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.

Nishiyama, T., Fujita, T., Shin-I, T. *et al.* (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, K.J., Davies, C., Robinson, S.P. and Fincher, G.B. (2001) Expression patterns of cell wall-modifying enzymes during grape berry development. *Planta*, **214**, 257–264.

Ogihara, Y., Mochida, K., Nemoto, Y., Murai, K., Yamazaki, Y., Shin-I, T. and 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.

Picton, S., Gray, J.E., Payton, S., Barton, S.L., Lowe, A. and Grierson, D. (1993) A histidine decarboxylase-like mRNA is involved in tomato fruit ripening. *Plant Mol. Biol.* **23**, 627–631.

Pnueli, L., Hareven, D., Broday, L., Hurwitz, C. and 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., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R. and White, J. (2001) The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucl. Acids Res.* **29**, 159–164.

Ramakrishna, W., Deng, Z., Ding, C.K., Handa, A.K. and Ozminkowski, R.H. Jr (2003) A novel small heat shock protein gene, *vis1*, contributes to pectin depolymerization and juice viscosity in tomato fruit. *Plant Physiol.* **131**, 725–735.

Ray, J., Moureau, P., Bird, C., Bird, A., Grierson, D., Maunders, M., Truesdale, M., Bramley, P. and Schuch, W. (1992) Cloning and characterization of a gene involved in phytoene synthesis from tomato. *Plant Mol. Biol.* **19**, 401–404.

Riechmann, J.L., Wang, M. and Meyerowitz, E.M. (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETELA1, APETELA3, PISTILLATA and AGAMOUS. *Nucl. Acids Res.* **24**, 3134–3141.

Rogaev, E., Sherrington, R., Rogaeva, E. *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.

Romualdi, C., Bortoluzzi, S. and Danieli, G.A. (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. and 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, C., Stegalkina, S., Ascenzi, R. *et al.* (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* **131**, 419–429.

Rose, J.K.C., Lee, H.H. and Bennett, A.B. (1997) Expression of a divergent expansion gene is fruit-specific and ripening-regulated. *Proc. Natl Acad. Sci. USA*, **94**, 5955–5960.

Schmitt, A.O., Specht, T., Beckmann, G., Dahl, E., Pilarsky, C.P., Hinzmann, B. and Rosenthal, A. (1999) Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucl. Acids Res.* **27**, 4251–4260.

Seymour, G.B., Fray, R.G., Hill, P. and Tucker, G.A. (1993) Down-regulation of two non-homologous endogenous tomato genes with a single chimaeric sense gene construct. *Plant Mol. Biol.* **23**, 1–9.

Smith, D.L. and Gross, K.C. (2000) A family of at least seven beta-galactosidase genes is expressed during tomato fruit development. *Plant Physiol.* **123**, 1173–1183.

Smith, C.J.S., Watson, C.F., Morris, P.C., Bird, C.R., Seymour, G.B., Gray, J.E., Harding, S.E., Tucker, G.A., Schuch, W. and Grierson, D. (1989) Inheritance and effect on ripening of antisense polygalacturonase genes in transgenic tomatoes. *Plant Mol. Biol.* **14**, 369–379.

Southern, E., Mir, K. and Shchepinov, M. (1999) Molecular interactions on microarrays. *Nat. Genet.* **21**(Suppl. 1), 5–9.

Speirs, J., Lee, E., Holt, K., Yong-Duk, K., Scott, N.S., Loveys, B. and 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.

Stekel, D.J., Git, Y. and Falciani, F. (2000) The comparison of gene expression from multiple cDNA libraries. *Genome Res.* **10**, 2055–2061.

Strausberg, R.L., Greenhut, S.F., Grouse, L.H., Schaefer, C.F. and Buetow, K.H. (2001) *In silico* analysis of cancer through the Cancer Genome Anatomy Project. *Trends Cell Biol.* **11**, 66–71.

Tang, W., Ezcurra, I., Muschietti, J. and 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.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and 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.

Verhoeyen, M.E., Bovy, A., Collins, G., Muir, S., Robinson, S., de Vos, C.H.R. and Colliver, S. (2002) Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *J. Exp. Bot.* **53**, 2099–2106.

Vierling, E. (1991) The role of heat shock proteins in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **42**, 579–620.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W. and Giovannoni, J. (2002) A MADS-box gene necessary for fruit ripening at tomato *Ripening-inhibitor (Rin)* locus. *Science*, **296**, 343–346.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.-C., Giovannoni, J. and Klee, H. (1995) An ethylene-inducible component of signal transduction encoded by Never-ripe. *Science*, **270**, 1807–1809.

Zheng, L., Heupel, R. and 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.