D1156–D1163 *Nucleic Acids Research, 2011, Vol. 39, Database issue* *Published online 21 October 2010*
doi:10.1093/nar/gkq991

Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics

Zhangjun Fei[1,2,*], Je-Gun Joung[1], Xuemei Tang[1], Yi Zheng[1], Mingyun Huang[1], Je Min Lee[1], Ryan McQuinn[1], Denise M. Tieman[3], Rob Alba[1], Harry J. Klee[3] and James J. Giovannoni[1,2]

[1]Boyce Thompson Institute for Plant Research, Cornell University, [2]USDA Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853 and [3]Plant Molecular and Cellular Biology Program, Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA

Received September 1, 2010; Accepted October 5, 2010

ABSTRACT

Tomato Functional Genomics Database (TFGD) provides a comprehensive resource to store, query, mine, analyze, visualize and integrate large-scale tomato functional genomics data sets. The database is functionally expanded from the previously described Tomato Expression Database by including metabolite profiles as well as large-scale tomato small RNA (sRNA) data sets. Computational pipelines have been developed to process microarray, metabolite and sRNA data sets archived in the database, respectively, and TFGD provides downloads of all the analyzed results. TFGD is also designed to enable users to easily retrieve biologically important information through a set of efficient query interfaces and analysis tools, including improved array probe annotations as well as tools to identify co-expressed genes, significantly affected biological processes and biochemical pathways from gene expression data sets and miRNA targets, and to integrate transcript and metabolite profiles, and sRNA and mRNA sequences. The suite of tools and interfaces in TFGD allow intelligent data mining of recently released and continually expanding large-scale tomato functional genomics data sets. TFGD is available at http://ted.bti.cornell.edu.

INTRODUCTION

Tomato (*Solanum lycopersicum*) is an economically important vegetable/fruit crop throughout the world with significant importance for human health and nutrition. It has long served as a model system for fleshy fruit development, plant genetics, pathology and physiology. Currently the entire genome of tomato is being sequenced by an international consortium, resulting in a wealth of genomics resources including BAC and fosmid libraries and their end sequences, high density genetic and physical maps, a large set of molecular markers and a number of powerful computational pipelines for sequence analysis and genome annotation (1). Meanwhile, numerous large-scale functional genomics resources for tomato have been developed over the past several years with new resources accumulating at a rapid rate. A large collection of tomato Expressed Sequence Tags (ESTs) that currently represents approximately 40 000 unigenes derived from more than 300 000 ESTs has been generated (http://www.sgn.cornell.edu). In addition a collection of more than 11 000 tomato full-length cDNA sequences has been released (2). Several publicly available microarray platforms have been created based on the tomato EST collection and some have been used extensively by the community to investigate the dynamics of the tomato transcriptome in different biological processes, resulting in large amounts of gene expression data. Recently, comprehensive profiles of numerous tomato metabolites have been generated and integrated with phenotypic trait data and transcript

*To whom correspondence should be addressed. Tel: +1 607 254 3234; Fax: +1607 254 1242; Email: zf25@cornell.edu
Present addresses:
Xuemei Tang, Nature Source Genetics, 33 Thornwood Dr, Ithaca, NY 14850, USA
Rob Alba, Monsanto Company, Product Safety Center, 800 N. Lindbergh Blvd, St Louis, MO 63167, USA.

profiles to identify metabolic regulatory networks for the purpose of improving fruit quality (3,4). Profiles of fruit flavor and nutrition-related metabolites from well-defined tomato introgression lines (ILs) have also been generated in order to identify loci or genes affecting flavor and nutrition through systems analysis of genotype, metabolite and gene expression data (5,6). In the past several years, large numbers of tomato small RNA (sRNA) sequences have been accumulated and are exponentially expanding due to rapid advances in sequencing technologies (7–9). Understanding functions of these tomato sRNAs represents an emerging and relatively unexploited opportunity to provide novel insights into the regulatory mechanisms of biologically and agriculturally important processes amenable to the tomato system including fruit development and ripening.

We previously described the Tomato Expression Database (TED), which serves as a central repository of tomato expression data and contains a suite of data presentation and analysis tools to assist in the development and testing of biological hypotheses (10). With the rapid accumulation of large-scale metabolite profile and sRNA data in tomato, we have expanded the database into what we now term the Tomato Functional Genomics Database (TFGD). Besides the newly added tomato metabolite and sRNA data sets, TFGD has been significantly improved with a number of new features and analysis tools. Furthermore, we have developed computational pipelines to process and analyze raw tomato microarray expression, metabolite profile and sRNA data sets, respectively, to ensure uniformity of the analyzed results for cross-experiment comparisons.

DATABASE CONTENTS AND FUNCTIONS

TFGD is a comprehensive collection of tomato functional genomics data. Currently, the database contains three major data components: gene expression, metabolite profiles and sRNAs. All the data were collected from the public domain with regular updates and retrieval of newly released data. The database is tightly linked to the Solanaceae Genomics Network (SGN, http://www.sgn.cornell.edu), a community database containing comprehensive genomics information for solanaceous species including the tomato genome sequence (11).

Annotations of microarray probes

Currently three microarray platforms are publicly available for tomato: TOM1 cDNA array, TOM2 oligonucleotide array and Affymetrix genome array. Probes on these three array platforms were annotated by comparing their corresponding consensus or SGN unigene sequences against GenBank nr, SwissProt/TrEMBL and *Arabidopsis* protein databases. Gene Ontology (GO) terms were then assigned to array probes using the Gene Ontology Annotation Database (12) based on their top Swiss-Prot/TrEMBL hits. GO terms assigned to each probe were then mapped to a set of plant specific GO slims using a Perl script, map2slim.pl, available at the Gene Ontology website (http://www.geneontology.org/GO.slims.shtml). Array probes were further assigned to tomato metabolic pathways based on the LycoCyc database (tomato metabolic pathway database) available at SGN.

Updates on tomato gene expression data

Tomato microarray data sets have been collected from public repositories including NCBI Gene Expression Omnibus (13) and EBI ArrayExpress (14), in addition to those directly submitted to TFGD. All array data sets are archived in TFGD following MIAME guidelines (15). Currently TFGD contains a total of 1308 hybridizations from 43 experiments, of which 773 248 and 287 are from TOM1 cDNA, TOM2 oligonucleotide and Affymetrix genome arrays, respectively (Table 1).

To ensure uniformity of the analyzed results for cross-experiment comparisons, we have implemented computational pipelines to process the microarray data sets archived in our database. Briefly, for data sets generated using spotted arrays (TOM1 and TOM2), raw data were normalized using the print-tip LOWESS normalization strategy (16). Spots flagged by image quantification programs as poor quality and spots that were not expressed in both channels were filtered out. Probes with at least two replicated data points were included in downstream statistical analysis. Significance of differential gene expression was determined using Patterns from Gene Expression (PaGE) (17). For data sets generated using the Affymetrix array, raw array data (CEL files) were normalized at the probe level using the gcRMA algorithm (18) and significance of differential gene expression was determined with the LIMMA package (19). All raw and analyzed array data can be downloaded from the database without restriction.

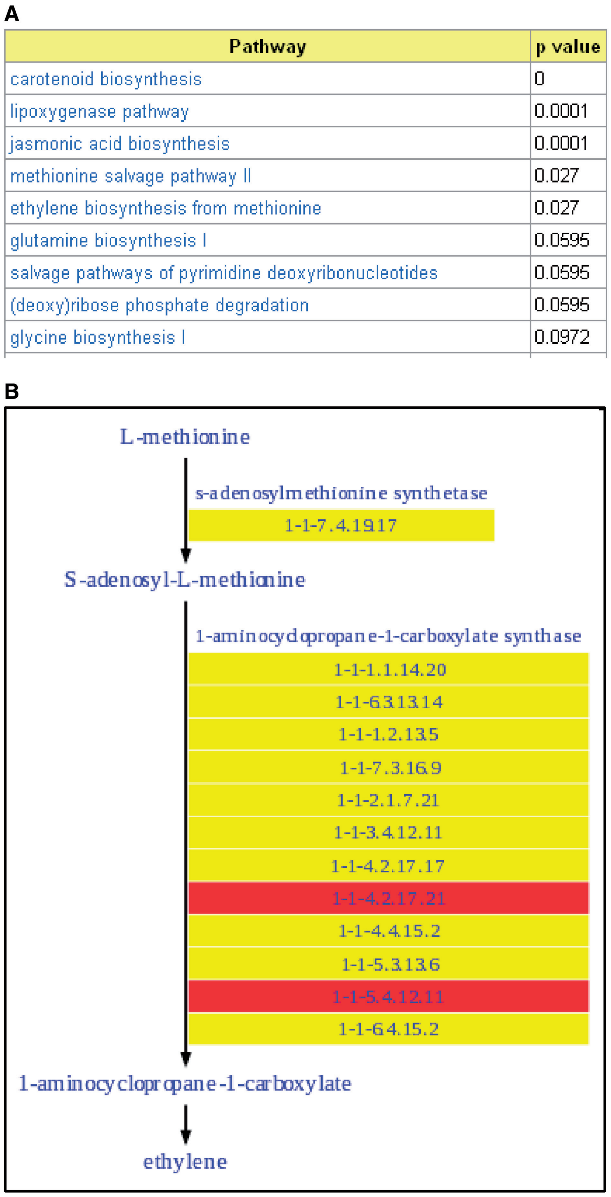
In addition to query interfaces and tools described in our previous report (10), a number of new tools to facilitate mining and analyzing the array results have been implemented. A co-expression analysis tool that can identify genes whose expression profiles are highly positively or negatively correlated with that of a given gene was implemented in TFGD. This tool can help to identify genes with similar functions since co-expressed genes are often involved in same or related pathways and biological processes. Microarray experiments typically produce a list of hundreds or thousands of interesting genes based on defined statistical criteria. Condensing and translating such a list into biologically meaningful and manageable

Table 1. Statistics of tomato microarray experiments in TFGD

Array platform	No. of experiments	No. of hybridizations[a]	No. of distinct hybridizations
TOM1 cDNA array	20	773	132
TOM2 oligonucleotide array	8	248	38
Affymetrix genome array	15	287	100
Total	43	1308	270

[a] Including biological and technical replicates.

information is required to better understand the underlying biological phenomena of interest. To achieve this goal, we implemented GO term enrichment analysis and biochemical pathway analysis tools in the database. Both tools were adopted from Plant MetGenMAP (20). The GO term enrichment tool, which was implemented based on the GO::TermFinder Perl module (21), can identify a set of over-represented GO terms reflecting highly affected biological processes from a list of user input genes or a microarray data set archived in the database. The pathway analysis tool can rapidly retrieve a list of significantly altered biochemical pathways (Figure 1A) and provide intuitive visualization of transcriptional events within a pathway with genes highlighted in different colors to reflect their expression level changes (Figure 1B). The results obtained from these analyses can provide insight into the mechanisms that underlie targeted biological phenomena or biochemical changes associated with them at the molecular level.



A

Pathway	p value
carotenoid biosynthesis	0
lipoxygenase pathway	0.0001
jasmonic acid biosynthesis	0.0001
methionine salvage pathway II	0.027
ethylene biosynthesis from methionine	0.027
glutamine biosynthesis I	0.0595
salvage pathways of pyrimidine deoxyribonucleotides	0.0595
(deoxy)ribose phosphate degradation	0.0595
glycine biosynthesis I	0.0972

Figure 1. Pathway analysis in the database. (**A**) Screenshot of an example result returned by the pathway analysis tool in TFGD which lists altered pathways identified from a gene expression data set. (**B**) Visualization of detailed transcript expression changes in a pathway.

One of the major tasks in gene expression data analysis is to sort a list of genes into different functional categories as a means of furthering downstream analysis. In TFGD, we implemented a tool that uses plant specific GO slims, which are a list of high level GO terms providing a broad overview of the ontology content (http://www.geneontology.org/GO.slims.shtml), to functionally classify a list of user input genes.

Tomato metabolite data

During the last decade, analyses of mRNA at the whole-genome level have proven central to most functional genomics initiatives. Recently, metabolite profiling has emerged as an additional layer of phenotypic information to more fully inform gene functional interpretation and has the potential not only to provide deeper insight into complex regulatory processes but also to determine biochemical and downstream phenotypes directly (22). Currently TFGD contains profiles of numerous flavor and nutrition-related metabolites. Fruit flavor and nutrition composition have clear positive human benefit. However flavor and nutrition are difficult traits to modify via either traditional breeding or transgenic approaches due to their generally complex biosynthetic and regulatory pathways. Recent advances in genomics, bioinformatics and high-throughput technologies provide an opportunity to dissect the regulatory mechanisms of fruit nutrition and flavor through systems biology to reveal key regulatory steps and thus putative targets for breeding or engineering. Profiles of a total of more than 60 flavor and nutrition-related metabolites from multiple seasons in the ripe fruit tissues of a collection of 76 *S. pennellii*-derived ILs (23) and a collection of 89 *S. habrochaites*-derived ILs (24), in addition to their corresponding parental control lines, have been generated (5,6) and archived in TFGD. These ILs represent overlapping single introgressions of the *S. pennellii* and *S. habrochaites* genomes, respectively, into the *S. lycopersicum* genome. Metabolite profiles from multiple seasons were analyzed using two-way ANOVA tests followed by *post hoc* Dunnett's tests to identify lines with significant metabolite content changes compared to parental controls. Based on these metabolite profiles, multiple loci affecting fruit nutrition and flavor have been identified (5,6).

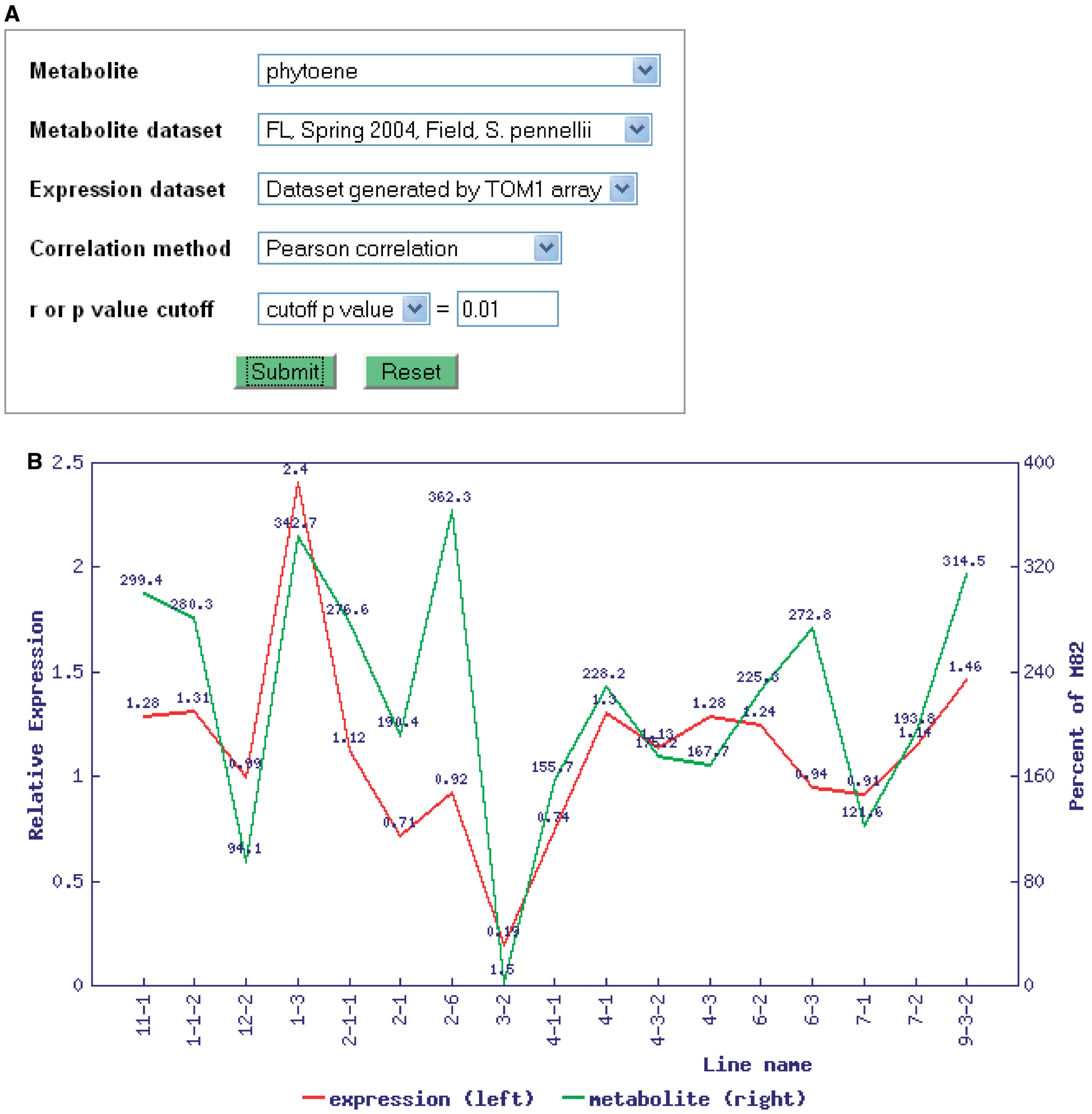
All the analyzed metabolite profile data were included in the database. Interfaces which allow users to efficiently retrieve profiles of a specific metabolite across all ILs in addition to metabolite profiles of a specific IL were implemented. Tools to identify ILs that display significant changes in a specific metabolite as well as ILs with specific metabolite properties were also implemented in the database.

Transcriptome profiles of *S. pennellii*-derived ILs from the same tissue samples used for flavor and nutrition-related metabolite profile generation are all available in

TFGD. Inclusion of both data types with tools that link them allows identification of novel genes involved in or regulating specific metabolic pathways using an integrated systems approach. To this end, a tool to correlate metabolite and transcript profiles by employing the Pearson or Spearman rank correlation coefficient to measure the similarity of profiles was implemented (Figure 2A). Using this tool, several meaningful and significant correlations between metabolite and gene expression profiles were identified (Figure 2B). In addition, a number of novel correlations have been observed. Based on these correlations, we have successfully identified a number of transcription factors associated with fruit metabolite levels and have functionally verified at least one which influences fruit ripening and carotenoid levels when repressed in transgenic tomato fruits (Lee and Giovannoni, submitted).

Tomato sRNA data

In the past few years, small RNAs (sRNAs) have been found to act as key regulators of cellular processes. They regulate gene expression by acting either on DNA to guide



Metabolite: phytoene

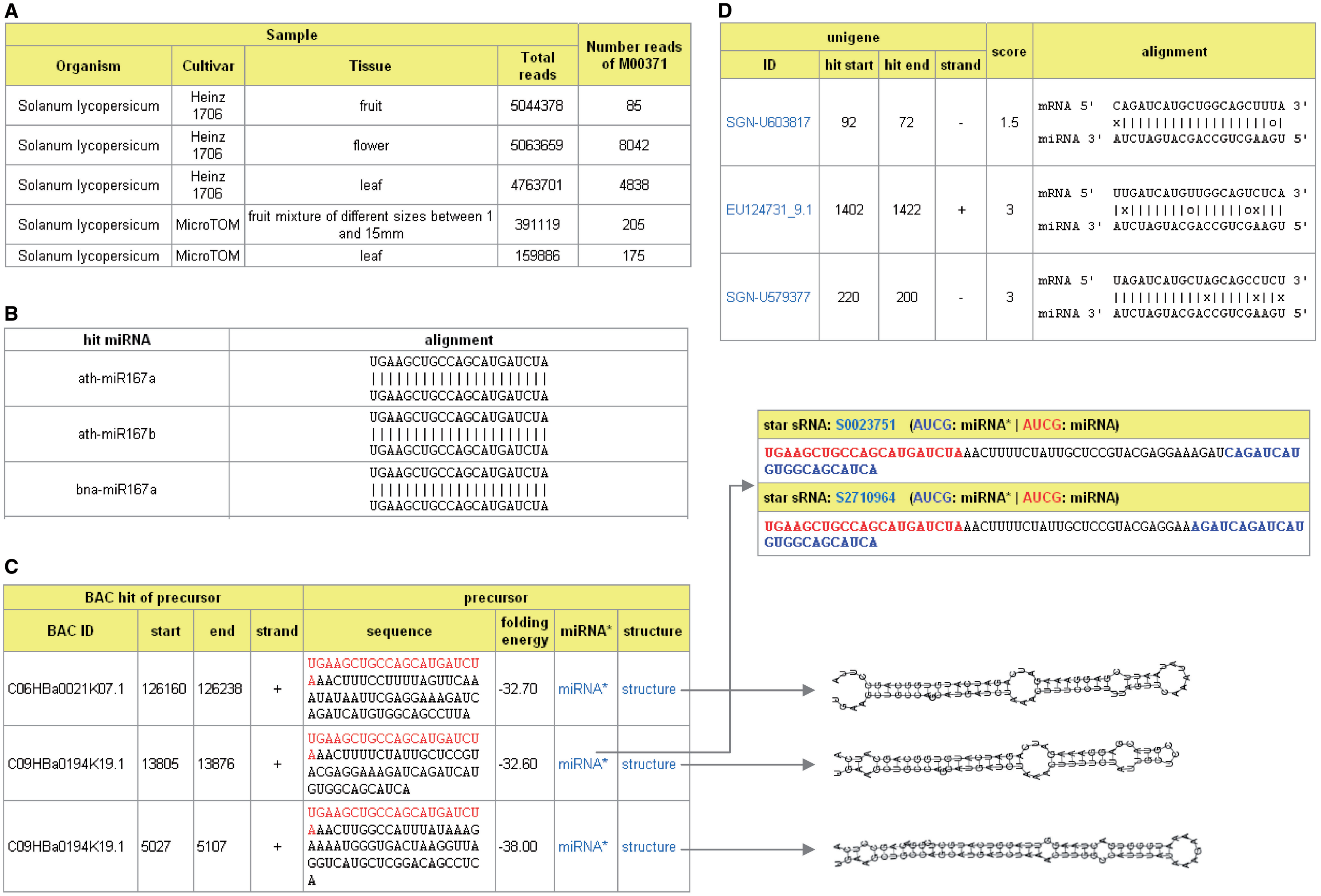
Metabolite dataset: FL, Spring 2004, Field, S. pennellii

Expression dataset: Dataset generated by TOM1 array

Correlation method: Pearson correlation

r or p value cutoff: cutoff p value = 0.01

Figure 2. Correlation analysis between gene expression and metabolite profiles in TFGD. (**A**) Interface of the correlation analysis. (**B**) An example of known correlations identified in the database: correlation between profiles of phytoene (green) and phytoene synthase (red) across 17 ILs ($r = 0.655$, $P = 0.00432$).



A

Sample				Number reads of M00371
Organism	Cultivar	Tissue	Total reads	
Solanum lycopersicum	Heinz 1706	fruit	5044378	85
Solanum lycopersicum	Heinz 1706	flower	5063659	8042
Solanum lycopersicum	Heinz 1706	leaf	4763701	4838
Solanum lycopersicum	MicroTOM	fruit mixture of different sizes between 1 and 15mm	391119	205
Solanum lycopersicum	MicroTOM	leaf	159886	175

B

<table>
<tr><th>hit miRNA</th><th>alignment</th></tr>
<tr><td>ath-miR167a</td><td>UGAAGCUGCCAGCAUGAUCUA
|||||||||||||||||||||
UGAAGCUGCCAGCAUGAUCUA</td></tr>
<tr><td>ath-miR167b</td><td>UGAAGCUGCCAGCAUGAUCUA
|||||||||||||||||||||
UGAAGCUGCCAGCAUGAUCUA</td></tr>
<tr><td>bna-miR167a</td><td>UGAAGCUGCCAGCAUGAUCUA
|||||||||||||||||||||
UGAAGCUGCCAGCAUGAUCUA</td></tr>
</table>

C

BAC hit of precursor				precursor			
BAC ID	start	end	strand	sequence	folding energy	miRNA*	structure
C06HBa0021K07.1	126160	126238	+	UGAAGCUGCCAGCAUGAUCU AAACUUUCCUUUUAGUUCAA AUAUAAUUCGAGGAAAGAUC AGAUCAUGUGGCAGCCUUA	-32.70	miRNA*	structure
C09HBa0194K19.1	13805	13876	+	UGAAGCUGCCAGCAUGAUCU AAACUUUUCUAUUGCUCCGU ACGAGGAAAGAUCAGAUCAU GUGGCAGCAUCA	-32.60	miRNA*	structure
C09HBa0194K19.1	5027	5107	+	UGAAGCUGCCAGCAUGAUCU AAACUUGGCCAUUUAUAAAG AAAAUGGGUGACUAAGGUUA GGUCAUGCUCGGACAGCCUC A	-38.00	miRNA*	structure

star sRNA: S0023751 (AUCG: miRNA* | AUCG: miRNA)

UGAAGCUGCCAGCAUGAUCUAAACUUUUCUAUUGCUCCGUACGAGGAAAGAUCAGAUCAUGUGGCAGCAUCA

star sRNA: S2710964 (AUCG: miRNA* | AUCG: miRNA)

UGAAGCUGCCAGCAUGAUCUAAACUUUUCUAUUGCUCCGUACGAGGAAAGAUCAGAUCAUGUGGCAGCAUCA

D

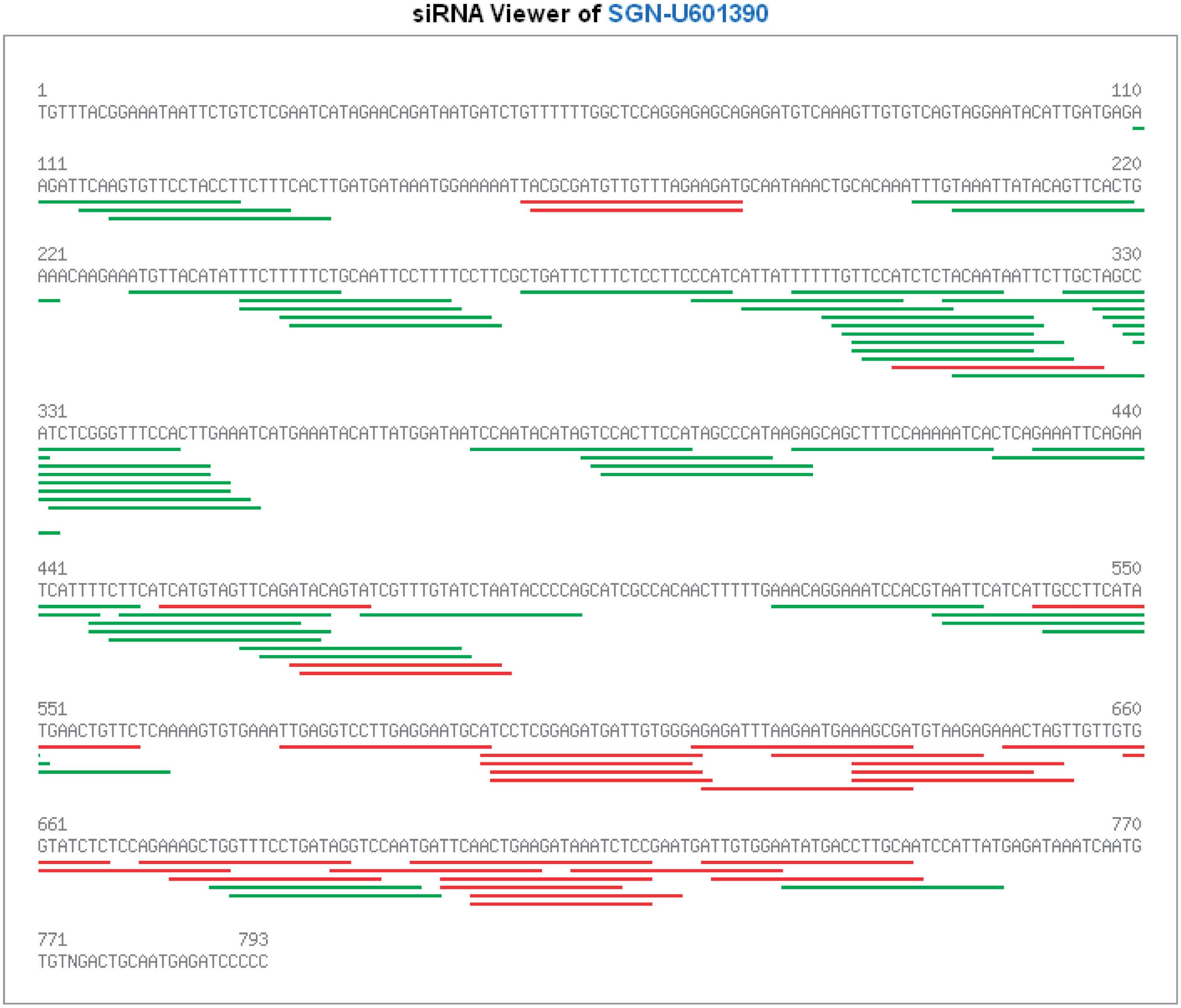
<table>
<tr><th colspan="4">unigene</th><th rowspan="2">score</th><th rowspan="2">alignment</th></tr>
<tr><th>ID</th><th>hit start</th><th>hit end</th><th>strand</th></tr>
<tr><td>SGN-U603817</td><td>92</td><td>72</td><td>-</td><td>1.5</td><td>mRNA 5' CAGAUCAUGCUGGCAGCUUUA 3'
x||||||||||||||||||o|
miRNA 3' AUCUAGUACGACCGUCGAAGU 5'</td></tr>
<tr><td>EU124731_9.1</td><td>1402</td><td>1422</td><td>+</td><td>3</td><td>mRNA 5' UUGAUCAUGUUGGCAGUCUCA 3'
|x|||||||o||||||ox|||
miRNA 3' AUCUAGUACGACCGUCGAAGU 5'</td></tr>
<tr><td>SGN-U579377</td><td>220</td><td>200</td><td>-</td><td>3</td><td>mRNA 5' UAGAUCAUGCUAGCAGCCUCU 3'
|||||||||||x|||||x||x
miRNA 3' AUCUAGUACGACCGUCGAAGU 5'</td></tr>
</table>

Figure 3. Tomato miRNA candidate information in TFGD. (**A**) Abundances of a miRNA candidate in each sample. (**B**) Conservation between the miRNA candidate and known miRNAs. (**C**) miRNA candidate precursors and their secondary structures and corresponding miRNA star sequences. (**D**) Predicted miRNA targets.

sequence elimination and chromatin remodeling or on RNA to guide cleavage and translation repression (25). Advances in high-throughput sequencing technologies have greatly accelerated the discovery and characterization of new classes of sRNAs including miRNA, ta-siRNA and nat-siRNA, as well as identification of their novel regulatory roles in diverse biological processes.

Recently several large-scale sRNA data sets have been generated for tomato (7–9; http://smallrna.udel.edu). TFGD provides a central repository with tools to disseminate these sRNAs and assist in their analysis. Currently, the database contains approximately 15.4 million sRNA sequences that are mainly derived from fruit, leaf and flower tissues, representing more than 5.3 million unique sRNAs. The sRNAs were first annotated by comparing them to rRNA, tRNA and tomato repeat sequence databases. miRNA candidates were then identified using an in-house pipeline. In short, highly abundant sRNAs were first aligned to tomato genome sequences and the flanking sequences (200 bp on each side) of sRNAs were extracted and folded *in silico* using the RNAfold program (26). Resulting folded structures were then checked with miRcheck (27) to identify potential miRNA candidates. These miRNA candidates were further compared to miRBase (28) to identify conserved miRNAs. In addition, potential miRNA star sequences for each miRNA candidate were also identified from the tomato sRNA data set. Finally, miRNA targets were identified using a program we developed according to the scoring matrix described in Jones-Rhoades and Bartel (29). Expression profiles of the target genes, if available, were provided in the database by linking to the gene expression module.

TFGD provides sRNA sequence data, annotations, their digital expression in each sample, and candidate miRNA information. For each candidate miRNA, the



siRNA Viewer of SGN-U601390

```
1                                                                                                          110
TGTTTACGGAAATAATTCTGTCTCGAATCATAGAACAGATAATGATCTGTTTTTTGGCTCCAGGAGAGCAGAGATGTCAAAGTTGTGTCAGTAGGAATACATTGATGAGA

111                                                                                                        220
AGATTCAAGTGTTCCTACCTTCTTTCACTTGATGATAAATGGAAAAATTACGCGATGTTGTTTAGAAGATGCAATAAACTGCACAAATTTGTAAATTATACAGTTCACTG

221                                                                                                        330
AAACAAGAAATGTTACATATTTCTTTTTCTGCAATTCCTTTTCCTTCGCTGATTCTTTCTCCTTCCCATCATTATTTTTTGTTCCATCTCTACAATAATTCTTGCTAGCC

331                                                                                                        440
ATCTCGGGTTTCCACTTGAAATCATGAAATACATTATGGATAATCCAATACATAGTCCACTTCCATAGCCCATAAGAGCAGCTTTCCAAAAATCACTCAGAAATTCAGAA

441                                                                                                        550
TCATTTTCTTCATCATGTAGTTCAGATACAGTATCGTTTGTATCTAATACCCCAGCATCGCCACAACTTTTTGAAACAGGAAATCCACGTAATTCATCATTGCCTTCATA

551                                                                                                        660
TGAACTGTTCTCAAAAGTGTGAAATTGAGGTCCTTGAGGAATGCATCCTCGGAGATGATTGTGGGAGAGATTTAAGAATGAAAGCGATGTAAGAGAAACTAGTTGTTGTG

661                                                                                                        770
GTATCTCTCCAGAAAGCTGGTTTCCTGATAGGTCCAATGATTCAACTGAAGATAAATCTCCGAATGATTGTGGAATATGACCTTGCAATCCATTATGAGATAAATCAATG

771                   793
TGTNGACTGCAATGAGATCCCCC
```

Figure 4. siRNA viewer in TFGD. siRNAs in red were aligned to mRNA in the forward direction while those in green were in the reverse direction.

database provides several lines of evidence to determine the confidence that a given miRNA candidate is a true miRNA. Potential evidence includes the abundance of the candidate, whether the candidate is conserved with known miRNAs, and whether the miRNA candidate has corresponding miRNA star sequences, and predicted targets (Figure 3). Several query interfaces and tools have been developed to assist in exploring and analyzing the tomato sRNAs and miRNA candidates. Users can retrieve a specific family of miRNA candidates, as well as the most abundant sRNAs in each tissue. The database also allows users to compare their own sRNA sequences against the sRNAs archived in the database. Finally, a very useful tool to identify potential miRNA targets of user-supplied miRNA sequences and to identify tomato miRNAs that potentially target specific transcript sequences was developed and added to the database.

Tomato sRNAs were further aligned to EST/mRNA sequences, in order to identify small interference RNAs (siRNAs). A siRNA viewer was developed in the database which shows the distribution of siRNAs in each tomato gene (Figure 4). Expression profiles of siRNAs and their corresponding genes can be compared, which also helps in designing siRNAs to efficiently silence their target genes. In short, we have merged our sRNA and EST/gene expression functions to facilitate prediction and likelihood analysis of sRNA involvement in regulation of specific genes via a user driven interface.

FUTURE DIRECTIONS

The complex functions of a living cell are carried out through the concerted activity of many genes and gene products. This activity is often coordinated by the organization of the genome into regulatory modules, or sets of co-regulated genes that share a common function. Identifying these regulatory modules is crucial for understanding important cellular processes. For this purpose, we are identifying tomato regulatory modules using large-scale expression data sets archived in the database. In addition, with the recent completion of the tomato genome sequence, we are in the process of mapping probes on tomato arrays to the genome and extracting promoter sequences for every probe. Tools are being developed in the database to assist in the identification of regulatory motifs from sets of co-regulated genes. We will continue to collect and archive publicly available tomato microarray, metabolite profile and sRNA data sets, as well as incorporating emerging RNA-seq, proteomics and phenotypic data sets to insure capture and utilization of the full complement of public tomato genomics resources developed and released in the plant science community.

FUNDING

National Science Foundation (IOS-0501778 and IOS-0923312). Funding for open access charge: National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Mueller,L., Lankhorst,R.K., Tanksley,S.D., Giovannoni,J.J., White,R., Vrebalov,J., Fei,Z., van Eck,J., Buels,R., Mills,A.A. *et al.* (2009) A snapshot of the emerging tomato genome sequence. *Plant Genome*, **2**, 78–92.
2. Aoki,K., Yano,K., Suzuki,A., Kawamura,S., Sakurai,N., Suda,K., Kurabayashi,A., Suzuki,T., Tsugane,T., Watanabe,M. *et al.* (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the *Solanaceae* genomics. *BMC Genomics*, **11**, 210.
3. Carrari,F., Baxter,C., Usadel,B., Urbanczyk-Wochniak,E., Zanor,M.I., Nunes-Nesi,A., Nikiforova,V., Centero,D., Ratzka,A., Pauly,M. *et al.* (2006) Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.*, **142**, 1380–1396.
4. Schauer,N., Semel,Y., Roessner,U., Gur,A., Balbo,I., Carrari,F., Pleban,T., Perez-Melis,A., Bruedigam,C., Kopka,J. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol.*, **24**, 447–454.
5. Tieman,D.M., Zeigler,M., Schmelz,E.A., Taylor,M.G., Bliss,P., Kirst,M. and Klee,H.J. (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. *J. Exp. Bot.*, **54**, 887–896.
6. Mathieu,S., Dal Cin,V., Fei,Z., Li,H., Bliss,P., Taylor,M., Klee,H. and Tieman,D. (2009) Flavor compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *J. Exp. Bot.*, **60**, 325–337.
7. Pilcher,R.L., Moxon,S., Pakseresht,N., Moulton,V., Manning,K., Seymour,G. and Dalmay,T. (2007) Identification of novel small RNAs in tomato (*Solanum lycopersicum*). *Planta*, **226**, 709–717.
8. Itaya,A., Bundschuh,R., Archual,A.J., Joung,J.G., Fei,Z., Dai,X., Zhao,P.X., Tang,Y., Nelson,R.S. and Ding,B. (2008) Small RNAs in tomato fruit and leaf development. *Biochim Biophys Acta.*, **1779**, 99–107.
9. Moxon,S., Jing,R., Szittya,G., Schwach,F., Pilcher,R.L.R., Moulton,V. and Dalmay,T. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.*, **18**, 1602–1609.
10. Fei,Z., Tang,X., Alba,R. and Giovannoni,J. (2006) Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Res.*, **34**, D766–D770.
11. Mueller,L.A., Solow,T.H., Taylor,N., Skwarecki,B., Buels,R., Binns,J., Lin,C., Wright,M.H., Ahrens,R., Wang,Y. *et al.* (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.*, **138**, 1310–1317.
12. Camon,E., Magrane,M., Barrell,D., Lee,V., Dimmer,E., Maslen,J., Binns,D., Harte,N., Lopez,R. and Apweiler,R. (2004) The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.*, **32**, D262–D266.
13. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Rudnev,D., Evangelista,C., Kim,I.F., Soboleva,A., Tomashevsky,M., Marshall,K.A. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
14. Parkinson,H., Kapushesky,M., Kolesnikov,N., Rustici,G., Shojatalab,M., Abeygunawardena,N., Berube,H., Dylag,M., Emam,I., Farne,A. *et al.* (2009) ArrayExpress update-from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.*, **37**, D868–D872.
15. Brazma,A., Hingamp,P., Quackenbush,J., Sherlock,G., Spellman,P., Stoeckert,C., Aach,J., Ansorge,W., Ball,C.A., Causton,H.C. *et al.* (2001) Minimum information about a microarray experiment (MIAME)–toward standards for microarray data. *Nat. Genet.*, **29**, 365–371.
16. Yang,Y.H., Dudoit,S., Luu,P., Lin,D.M., Peng,V., Ngai,J. and Speed,T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, **30**, e15.

17. Grant,G.R., Liu,J. and Stoeckert,C.J. Jr (2005) A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics*, **11**, 2684–2690.
18. Wu,Z., Irizarry,R.A., Gentleman,R., Martinez Murillo,F. and Spencer,F. (2004) A model based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.*, **99**, 909–917.
19. Smyth,G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, 1–29.
20. Joung,J.G., Corbett,A.M., Fellman,S.M., Tieman,D.M., Klee,H.J., Giovannoni,J.J. and Fei,Z. (2009) Plant MetGenMAP: an integrative analysis system for plant systems biology. *Plant Physiol.*, **151**, 1758–1768.
21. Boyle,E.I., Weng,S., Gollub,J., Jin,H., Botstein,D., Cherry,J.M. and Sherlock,G. (2004) GO::TermFinder-open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, **20**, 3710–3715.
22. Fiehn,O., Kopka,J., Dörmann,P., Altmann,T., Trethewey,R.N. and Willmitzer,L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.*, **18**, 1157–1161.
23. Eshed,Y. and Zamir,D. (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica*, **79**, 175–179.
24. Monforte,A. and Tanksley,S. (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome*, **43**, 803–813.
25. Vaucheret,H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.*, **20**, 759–771.
26. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
27. Rajagopalan,R., Vaucheret,H., Trejo,J. and Bartel,D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **20**, 3407–3425.
28. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
29. Jones-Rhoades,M.W. and Bartel,D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.