

Minireview

***Arabidopsis* genome sequence as a tool for functional genomics in tomato**

Kirankumar S Mysore*, Robert P Tuori* and Gregory B Martin*,[†]

Addresses: *Boyce Thompson Institute for Plant Research and [†]Department of Plant Pathology, Cornell University, Tower Road, Ithaca, NY 14853, USA.

Correspondence: Gregory B Martin. E-mail: gbm7@cornell.edu

Published: 12 January 2001

Genome Biology 2001, **2**(1):reviews1003.1–1003.4

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/1/reviews/1003>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Tomato is a well-established model organism for studying many biological processes including resistance and susceptibility to pathogens and the development and ripening of fleshy fruits. The availability of the complete *Arabidopsis* genome sequence will expedite map-based cloning in tomato on the basis of chromosomal synteny between the two species, and will facilitate the functional analysis of tomato genes.

Arabidopsis thaliana is widely used as a model for the study of many aspects of plant biology. Because of its small genome size (125 Mb) it was chosen as the subject of the first plant genome sequencing project, an effort that was recently completed [1,2]. The availability of the *Arabidopsis* genome sequence will greatly enhance our knowledge of the entire complement of genes expressed by a typical flowering plant and will lead to a thorough analysis of the function of these genes. A comprehensive molecular-marker-based linkage map exists for *Arabidopsis*, and the map-based cloning of genes conferring specific phenotypes will become even easier with the availability of genomic sequence information [3]. In addition to map-based cloning, several other approaches for isolating and determining the function of specific genes are available in *Arabidopsis* including T-DNA and activation tagging, transposon tagging, and various gene silencing methods [4-9]. The many resources available for the *Arabidopsis* experimental system also make it an ideal springboard for understanding the function of genes in economically important plants, such as tomato.

Tomato is a member of the Solanaceae family, a widely distributed group of plants to which many other economically important species also belong, including potato, pepper, eggplant, tobacco, and Petunia. In addition to its economic

importance, several features make tomato an excellent species for understanding basic plant biology. Tomato is diploid, has a relatively small genome (950 Mb), has a moderately short life cycle, tolerates inbreeding yet is easily cross-hybridized, and is easy to grow and maintain. Recently, a miniature tomato variety with an accelerated life cycle has been developed and is gaining favor among tomato researchers [10,11]. As a result of intensive breeding and genetic research over the past 50 years, many resources are available for tomato studies, including extensive germplasm collections of wild and cultivated species, numerous natural, induced, and transgenic mutants and genetic variants [12], routine transformation [11], a dense restriction-fragment length polymorphism (RFLP) map [13], numerous cDNA, yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC) and other genomic libraries [14,15], a transposon tagging system [16], a virus-induced gene silencing system in the related plant *Nicotiana benthamiana* [17,18], and recently an extensive database of over 100,000 expressed sequence tags (ESTs) [19]. Because of its experimental tractability, tomato has been the subject of many fundamental discoveries in plant biology including, for example, transgenic analysis of genes that impact fruit development and ripening [20,21], the discovery of a peptide hormone (systemin) [22], and the

mapping and molecular cloning of many disease-resistance (R) genes [23,24].

Although many resources are available to study tomato genes, it remains a daunting prospect, when compared with *Arabidopsis*, to assign a function to each tomato gene. The tomato genome is approximately seven times the size of the *Arabidopsis* genome, and that makes gene mapping and map-based cloning more difficult. In addition, the highly efficient floral-dip method used for *Arabidopsis* transformation [25] has not yet been successfully adapted to tomato and thus the development of tomato transgenics remains labor-intensive and inefficient. In this article we discuss some new approaches being used by labs including our own that might facilitate the characterization of tomato genes by using the resources developed for the study of *Arabidopsis*.

Synteny among *Arabidopsis* and tomato chromosomes

The availability of extensive *Arabidopsis* genome sequence has stimulated investigations of synteny among the chromosomes of *Arabidopsis* and several other plant species, including tomato. The exploitation of chromosomal colinearity with *Arabidopsis* can aid high-resolution linkage mapping and map-based cloning experiments in many plant species and will be especially helpful in crop plants that have large genomes. There are several reports of the conservation of gene repertoire between *Arabidopsis* and other members of the Brassica family [26,27]. Remarkably, conservation of gene order within small chromosomal regions has also been found between *Arabidopsis* and distantly related plants such as rice, sorghum and soybean [28-30].

Recently, Ku *et al.* [31] reported that colinearity over small chromosomal segments ('microsynteny') exists between *Arabidopsis* and tomato. Using sequencing and computational analysis, they compared the gene content and gene order of a 105 kb segment of tomato chromosome 2 (TC2) to its homologs in the *Arabidopsis* genome. They identified 17 open reading frames (ORFs) in the TC2 segment, of which four (24%) had no matches with any *Arabidopsis* BAC at the established statistical threshold. The remaining 13 ORFs had significant matches (at the amino-acid level) with ORFs from the *Arabidopsis* genome; 12 of these had matches to one or more of four different *Arabidopsis* BAC/P1 clones that correspond to four different chromosomal regions. The authors suggest that this network of microsynteny between the TC2 segment and different regions of the *Arabidopsis* genome must be due to multiple rounds of duplication in the *Arabidopsis* lineage. Although the order of the studied genes is the same within the *Arabidopsis* segments and TC2, none of the *Arabidopsis* regions contained the full set of 12 matching ORFs. These observations indicate that establishing syntenic relationships between chromosomes of species that belong to families as divergent as those of *Arabidopsis* and

tomato is likely to be challenging. Nevertheless, when the microsynteny that is identified can be coupled with high-resolution mapping, it might greatly expedite the cloning of tomato genes on the basis of their map position.

Expression of tomato genes in *Arabidopsis* for functional studies

Arabidopsis has a very efficient, fast, and high-throughput transformation system when compared with that available for tomato [25]. This advantage has prompted some researchers to express tomato genes in *Arabidopsis* as a rapid method of analyzing their potential function. Tieman and Klee [32] used *Arabidopsis* to study two tomato genes, *LeETR4* and *LeETR5*, which encode members of the ethylene-receptor family. To assess the potential role of *LeETR4* and *LeETR5* in ethylene perception, they introduced a putative dominant-negative mutation (Cys65 to serine) into the membrane-spanning domain of these proteins. They introduced these mutated gene constructs along with the wild-type genes into *Arabidopsis* and showed that, as expected, the mutated tomato genes conferred ethylene insensitivity in *Arabidopsis* [32].

Heterologous expression of tomato genes in *Arabidopsis* might also facilitate the investigation of plant defense responses against pathogens that attack both species. For example, the pathogen *Pseudomonas syringae* pv. tomato causes disease in both tomato and *Arabidopsis*. The tomato gene *Pto* confers resistance to strains of *Pseudomonas syringae* pv. tomato expressing the corresponding avirulence gene *avrPto* [23]. *Pto* encodes a protein kinase that has been intensively studied [33]. Several genes encoding *Pto*-interacting (*Pti*) proteins have been previously identified using yeast two-hybrid screens [34,35]: *Pti1* encodes a protein kinase that is phosphorylated by the *Pto* kinase and appears to play a role in the hypersensitive response. Although we have been unable to show that expression of *Pto* or *Pti1* alone in *Arabidopsis* confers elevated resistance to *Pseudomonas*, we have preliminary evidence that when *Pto* and *Pti1* are co-expressed in *Arabidopsis*, they do confer increased resistance to *Pseudomonas* strain DC3000 when compared to wild-type *Arabidopsis* (K.S.M. and G.B.M, unpublished observations). We are currently investigating whether this phenotype requires the *Pto*-*AvrPto* interaction or is due to generally enhanced resistance.

Other genes encoding *Pto*-interacting proteins are *Pti4*, *Pti5* and *Pti6*, whose products are transcription factors that bind the GCC-box element present in the promoter of many pathogenesis-related (PR) genes. We have previously reported evidence that, in tomato, *Pti4*, *Pti5* and *Pti6* form a direct link between the *Pto*-mediated pathogen recognition and activation of several PR genes [36]. We overexpressed *Pti4* and *Pti6* in *Arabidopsis* to test whether this heterologous expression might be used to efficiently identify possible

target genes of these transcription factors. Of the PR genes we examined in these lines, we found constitutive expression of *PR1*, *PR2*, *PR4*, *PDF1.2* and *Thi2.1* in plants overexpressing *Pti4* and of *PR1*, *PR2* and *Thi2.1* in plants overexpressing *Pti6* (Y. Gu and G.B.M., unpublished observations).

Heterologous gene expression profiling

We have used serial analysis of gene expression (SAGE) [37] to investigate differential gene expression patterns between wild-type *Arabidopsis thaliana* Col-0 and an isogenic transformant overexpressing a heterologous gene, tomato *Pti4* (R.P.T. and G.B.M., unpublished observations). After construction of SAGE libraries from leaves of both genotypes and subsequent large-scale sequencing of concatenated clones, SAGE tags corresponding to 27,927 transcripts were extracted. When we limited our study to tags identified at least twice (to reduce the potential for sequencing errors affecting the analysis), 20,215 tags remained. These tags included 9,978 from the wild-type *Arabidopsis* library and 10,237 from the *Pti4*-overexpressing library. The two projects when combined represent 3,312 different SAGE tags and probably represent that number of different genes. Figure 1 shows the distribution of the extent of induction or repression for all the genes identified in this project. Although 236 out of the 3,312 total genes (7.1%) were either induced or repressed fourfold or more, overall comparison of the two libraries revealed surprisingly few dramatic differences. There were only 25 genes whose tag abundance varied sevenfold or greater and six genes whose tag abundance varied tenfold or greater. The great majority (3,078 or 92.9%) had essentially similar levels of expression in the two libraries analyzed. Interestingly, this study identified almost as many genes that are repressed fourfold or greater (113) as those induced fourfold or greater (123) in the *Pti4* transformant.

We then performed extensive database similarity searches with the 1,139 most abundant transcripts (those that had an abundance of at least four SAGE tags in the combined project). Database searches (National Center for Biotechnology Information database searches done in September 2000) yielded four general classes: 340 (29%) matched full-length genes annotated in GenBank, 506 (44%) matched ESTs, 256 (23%) matched only genomic sequence, and 37 (3%) had no match in the *Arabidopsis* sequence databases. In total, 846 of the 1,139 different genes, or 74%, matched an *Arabidopsis* gene or EST sequence in the database. A complete description of the *Pti4*-overexpressing lines and our SAGE analysis will be published elsewhere. We plan to use the *Arabidopsis* ESTs identified in this study to isolate orthologs from the tomato EST collection. The corresponding tomato cDNA clones will then be used to develop a targeted microarray for a detailed analysis of changes in PR gene expression that occur in tomato in response to inoculation by virulent or avirulent *Pseudomonas* strains.

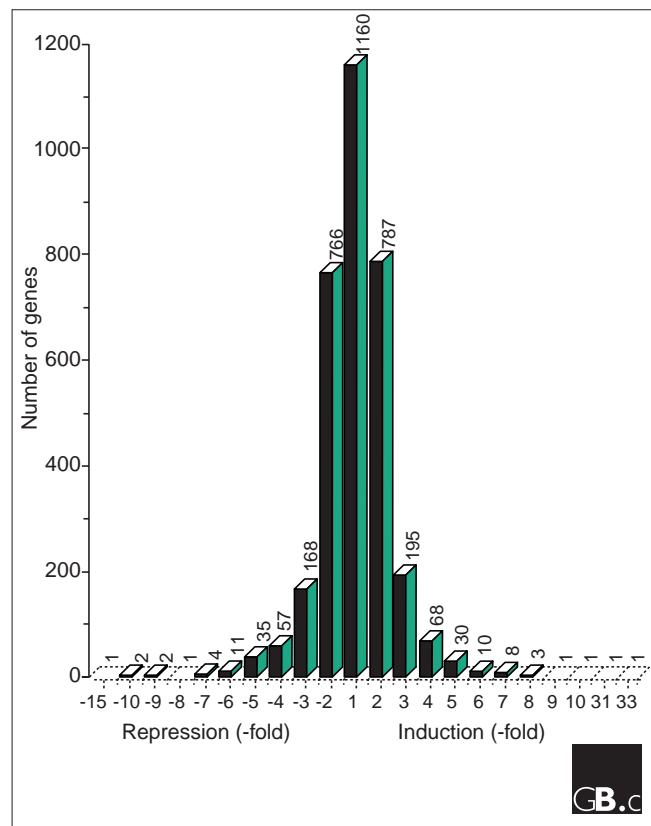


Figure 1
Histogram showing distribution of genes based on their 'fold' induction or repression in *Arabidopsis* plants overexpressing *Pti4* compared with wild-type plants.

We have discussed several approaches for functional genomics of tomato that will benefit from the availability of the completed genome sequence for *Arabidopsis*. It is also possible that resources available in the Solanaceae such as 'fast-forward' genetics using virus-induced gene silencing might benefit functional studies of *Arabidopsis* genes [17]. Clearly, as the number of tomato ESTs continues to increase and additional resources for genomics approaches become available for tomato, there will be further opportunities for cross-feeding of knowledge between tomato and *Arabidopsis*.

References

1. The Arabidopsis Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature* 2000, **408**:796-815.
2. **Arabidopsis Genome Initiative** [<http://www.arabidopsis.org/agi.html>]
3. **The Arabidopsis Information Resource** [<http://www.arabidopsis.org>]
4. Feldmann KA: **T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum**. *Plant J* 1991, **1**:71-82.
5. Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JDG: **Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: A tool for functional genomics**. *Plant Cell* 1999, **11**:1841-1852.
6. Winkler RG, Frank MR, Galbraith DW, Feyereisen R, Feldmann KA: **Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis***. *Plant Physiol* 1998, **118**:743-750.

7. Dalmay T, Hamilton A, Mueller E, Baulcombe DC: **Potato virus X amplicons in *Arabidopsis* mediate genetic and epigenetic gene silencing.** *Plant Cell* 2000, **12**:369-379.
8. Chuang CF, Meyerowitz EM: **Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*.** *Proc Natl Acad Sci USA* 2000, **97**:4985-4990.
9. Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM: **Total splicing by intron-spliced hairpin RNAs.** *Nature* 2000, **407**:319-320.
10. Scott JW, Harbaugh BK: **Micro-Tom: A miniature dwarf tomato.** *Florida Agr Expt Sta Circ* 1989, **370**:1-6.
11. Meissner R, Jacobson Y, Melamed S, Levyatuv S, Shalev G, Ashri A, Elkind Y, Levy A: **A new model system for tomato genetics.** *Plant J* 1997, **12**:1465-1472.
12. Rick CM, Yoder JT: **Classical and molecular genetics of tomato: highlights and perspectives.** *Annu Rev Genet* 1988, **22**:281-300.
13. Tanksley SD, Ganai MW, Prince JP, de Vicente MC, Bonierbale PB, Broun P, Fulton TM, Giovanonni JJ, Grandillo S, Martin GB, *et al.*: **High density molecular linkage maps of the tomato and potato genomes.** *Genetics* 1992, **132**:1141-1160.
14. Martin GB, Ganai MW, Tanksley SD: **Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci.** *Mol Gen Genet* 1992, **233**:25-32.
15. Budiman MA, Mao L, Wood TC, Wing RA: **A deep-coverage tomato BAC library and prospectus toward development of an STC framework for genome sequencing.** *Genome Res* 2000, **10**:129-136.
16. Briza J, Carroll BJ, Klimyuk VI, Thomas CM, Jones DA, Jones JDG: **Distribution of unlinked transposition of a Ds element from a T-DNA locus on tomato chromosome 4.** *Genetics* 1995, **141**:383-390.
17. Baulcombe DC: **Fast forward genetics based on virus-induced gene silencing.** *Curr Opin Plant Biol* 1999, **2**:109-113.
18. Burton RA, Gibeaut DM, Bacic A, Findlay K, Roberts K, Hamilton A, Baulcombe DC, Fincher GB: **Virus-induced silencing of a plant cellulose synthase gene.** *Plant Cell* 2000, **12**:691-706.
19. **The Institute for Genomic Research** [<http://www.tigr.org>]
20. Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap EV, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD: **fw2.2: a quantitative trait locus key to the evolution of tomato fruit size.** *Science* 2000, **289**:85-88.
21. Wilkinson JQ, Lanahan MB, Yen H, Giovannoni JJ, Klee HJ: **An ethylene-inducible component of signal transduction encoded by *Never-ripe*.** *Science* 1995, **270**:1807-1809.
22. Pearce G, Strydom D, Johnson S, Ryan CA: **A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins.** *Science* 1991, **253**:895-898.
23. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganai MW, Spivey R, Wu T, Earle ED, Tanksley SD: **Map-based cloning of a protein kinase gene conferring disease resistance in tomato.** *Science* 1993, **262**:1432-1436.
24. Grube RC, Radwanski ER, Jahn M: **Comparative genetics of disease resistance within the Solanaceae.** *Genetics* 2000, **155**:873-887.
25. Clough SJ, Bent AF: **Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*.** *Plant J* 1998, **16**:735-743.
26. Sadowski J, Quiros CF: **Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the *RPS2* gene in the *Brassica nigra* genome.** *Theor Appl Genet* 1998, **96**:468-474.
27. Acarkan A, Rossberg M, Koch M, Schmidt R: **Comparative genome analysis reveals extensive conservation of genome organization for *Arabidopsis thaliana* and *Capsella rubella*.** *Plant J* 2000, **23**:55-62.
28. Dodweerd AV, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I: **Identification and analysis of homeologous segments of the genomes of rice and *Arabidopsis thaliana*.** *Genome* 1999, **42**:887-892.
29. Paterson AH, Lan TH, Rieschmann KP, Chang C, Lin YR, Liu SC, Burrow MD, Kowalski SP, Katsar CS, DelMonte TA, *et al.*: **Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence.** *Nat Genet* 1996, **14**:380-382.
30. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*.** *Proc Natl Acad Sci USA* 2000, **97**:4168-4173.
31. Ku H, Vision T, Liu J, Tanksley SD: **Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny.** *Proc Natl Acad Sci USA* 2000, **97**:9121-9126.
32. Tieman DM, Klee HJ: **Differential expression of two novel members of the tomato ethylene-receptor family.** *Plant Physiol* 1999, **120**:165-172.
33. Sessa G, Martin GB: **Signal recognition and transduction mediated by the tomato Pto kinase: a paradigm of innate immunity in plants.** *Microbes Infect* 2000, **2**:1591-1597.
34. Zhou J, Loh Y, Bressan RA, Martin GB: **The tomato gene *PtiI* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response.** *Cell* 1995, **83**:925-935.
35. Zhou J, Tang X, Martin GB: **The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes.** *EMBO J* 1997, **16**:3207-3218.
36. Gu YQ, Yang C, Thara VK, Zhou J, Martin GB: ***Pti4* is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase.** *Plant Cell* 2000, **12**:771-786.
37. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.