

RESEARCH

A. Major research and education activities and findings from these activities

1) Development of core tissue cDNA libraries and the associated EST database for tomato (J. Giovannoni, S. Tanksley, G. Martin in collaboration with Cathy Ronning, TIGR)

Below is a list of the tissues/libraries targeted for library construction as well as the actual number of ESTs generated from each library. Sequencing is being done by TIGR through a separate contract. Thus far more than 62,000 ESTs have been generated with a target of 152,000 by the end of the project. We are on schedule for completing all ESTs by the end of the project.

Library	# Successful ESTs	Target
callus	10,072	10,000
ovary (5d pre-a to 5d post-a)	10,222	10,000
P.syr. R and S plant	11,365	10,000
mixed elicitor	11,912	10,000
seed (germinating and dormant)	1,145	10,000
tomato fruit (4 stages)	5,395	20,000
roots (3 stages and mineral stress)	5,955	10,000
Trichomes (L.hir and L.pen)	3,183	5,000
Flowers (buds-open flowers, staged)	2,062	10,000
shoot	1,348	5,000
selected clones (via screens)		5000
mapped markers (3' and 5')		2000
suspension culture (subtracted, various-induced)(in prep.)		15000
suspension culture (control)(in prep.)		5000
pollen tubes (germinating)(in prep.)		5000
shoots meristems (staged)	1348	7000
callus (agro-induced)		5000
demethylated-genomic		8,000
Total	62,659	152,000

2) Generation and analysis of genome-wide expression patterns during fruit development and in response to mutations of key regulatory loci and in response to hormone and light signal transduction (Jim Giovannoni, Texas A&M)

Implementation of microarray technology in the laboratory.

Specific activities have included testing of three different probe labeling strategies combined with various methods for RNA template preparation to develop a cost-effective

and reliable protocol. We have also used a total of six different fluorescent labels and yet seem to have the best sensitivity with probes labeled with Cy3 and Cy5 as in our initial experiments. We have recently changed from a GeneMachines arrayer to a Genetic Microsystems arrayer which in preliminary tests seems to give much more reproducible spotting. Our current arrays are comprised of either 650 non-redundant breaker fruit ESTs of which a sub-set of 50 ESTs are spotted onto nylon membranes as controls (for hybridization with radio-labeled probes derived from the same RNA used on the chip experiments), or a collection of 288 non-redundant ESTs mined from all sequences available in the February build of the TIGR database with particular emphasis on those which may impact ripening, ethylene response, or fruit quality/nutritional characteristics. **Figure 1** shows hybridization of duplicate ripe fruit derived probes from inbred lines developed at the University of Florida and differing in quality characteristics, to a chip containing the 288 EST set of non-redundant sequences.

Efforts toward optimization of microarray production, probe synthesis, scanning and subsequent software manipulations for image capture and analysis have resulted in our being at the stage where the majority of hybridizations give clean and reproducible data. Large amounts of staged and treated fruit tissues have been generated and we are now poised for initiation of extensive expression profiling.

Generation of tissues for expression profiling during fruit development and ripening.

Normal tomatoes were tagged at anthesis and harvested at daily intervals from 7 dpa through ripe/senescent. As a result, daily samples were taken daily for 43 days with a minimum of 4 fruit per stage. Tomato fruit harboring single gene mutations in genes influencing developmental (*rin*, *nor*), ethylene (*Nr*), and light (*hp-1*) signal transduction have also been isolated from the mature green (33 dpa) through 50 dpa in parallel with normal fruit. Additional normal fruit treated with heat (42C), cold (8C), and exogenous ethylene have also been generated. Total RNA (for cDNA probe synthesis and control RNA gel-blot) has been extracted from nearly all of this tissues and initial hybridizations will begin by early June. The general plan for expression analysis will involve simultaneous analyses of adjacent time point tissues. Each tissue will be compared to the previous and subsequent day fruit tissues so that a continuum of data is established permitting any time point to be compared to any other (**Figure 2**).

Development of collaboration with Ed Dougherty (Texas A&M Electrical Engineering) for analysis of regulatory networks via microarray expression analysis.

Dr. Dougherty is in his fourth year of collaboration with Dr. Jeffery Trent's lab (NHGRI) working on image analysis, statistical analysis, and informatics relating to microarrays. The computing environment is being developed at NIH and, with new equipment soon coming on line, will provide a high performance multi-processor system dedicated to the analysis of genomic control using cDNA microarray data. NHGRI is planning to use expression data derived from our tomato fruit development microarrays as a means of testing and tuning the software applications with accurately staged and genetically identical tissues.

Our collaboration with Ed Dougherty's group will provide a unique opportunity to extend interpretation of expression array results beyond the level of expression pattern clustering. This collaboration will provide us with the unique opportunity to analyze our expression with prototype next-generation software for computationally intensive analysis of regulatory relationships based solely on expression data. Such analysis requires utilization of treatments and genetic mutations which will alter gene expression (ie. provide variation in gene expression) in otherwise genetically and developmentally identical material.

Data-mining of the Tomato Gene Index and additional screening of arrayed "deep" EST libraries including 300,000 clones that will not be sequenced under this project.

We have identified nearly 3000 putative fruit development/ripening-related clones via this approach and will use a combination of sequencing and microarrays to begin classifying these sequences in terms of potential relationship to fruit ripening and/or fruit development, in addition to assessing redundancy (**Figure 3, Figure 4**).

Preliminary analysis of sequences recovered via mining the Tomato Gene Index and low stringency screening of the "deep" tomato cDNA libraries has yielded a large number of genes related to the RIN and NOR putative transcriptional regulators in addition to genes potentially associated with a range of fruit ripening and quality characteristics. The RIN gene is a member of the MADS family of DNA binding proteins (J. Vrebalov and J. Giovannoni unpublished) and we have identified 5 additional MADS genes expressed in ripening fruit with only minimal analysis of the clones recovered via our screens. NOR represents a distinct family of transcriptional regulators of which we have identified 4 additional family members - none of which have been previously described nor associated with fruit development.

Giovannoni laboratory will move to Ithaca, New York.

Our laboratory will move from Texas A&M to the ARS Plant Soil and Nutrition Lab in Ithaca NY in September 2000. All personnel except undergraduates listed above will remain in the lab. Paxton Payton and Shanna Moore will concentrate on taking expression profiling during fruit development and of mutants from testing and technical improvement to experimental analysis. Zhangjun Fei and Ruth White will focus on identification of fruit development and ripening-related genes from the collection of nearly 3000 candidate genes mentioned above. Microarrays will be used to associate expression with fruit development and DNA sequencing (Diana Medrano) will additionally be performed to narrow/hypothesize involvement in fruit biology. Diana Medrano has applied to graduate school at Cornell to earn a M.S. in Plant Biology while continuing to work in the lab. Her research project will involve becoming trained with Payton and Moore for expression profiling of inbred lines varying in fruit quality and nutrition characters with the eventual objective of identifying candidates for manipulation in transgenic plants.

3) Gene expression profiling of plant responses to pathogens (Gregory Martin, Boyce Thompson Institute)

Progress in setting up and using cDNA microarrays for expression profiling of plant defense responses.

We are developing tomato cDNA microarrays for gene expression profiling of plant responses to various pathogens and elicitors. This effort involves: 1) development of cDNA libraries derived from tomato leaves undergoing defense responses; 2) sequencing of large numbers of these cDNAs and the establishment of an EST database (by TIGR, see above); 3) derivation of a "unigene" set consisting of one representative cDNA of each unique tomato gene; and 4) development of cDNA microarrays for expression profiling.

To date, 5,000 of unigenes have been re-arrayed into 384-well microtiter plates and their inserts are being PCR amplified for microarraying onto glass slides (**Figure 5**). The microarraying is being done in the new *Center for Gene Expression Profiling* that was established in the past year at the Boyce Thompson Institute (see infrastructure section below). PCR products (approx. 4 nl each) are spotted at a 200 micron pitch onto silanated glass slides with a Genetic Microsystems arrayer. As controls for hybridization and for standardizing among slides the following genes are replicated several times on each slide: rubisco, osmotin, ubiquitin, actin, several non-plant genes, and the pBlueScript polylinker (**Figure 6**). In close collaboration with the Giovannoni lab we are continuing to optimize our microarraying protocols further in preparation for spotting the nonredundant clone set of 8,000 elements. An example of an array containing a selected

set of cDNAs from the tomato EST database and control clones is shown in **Figure 7**. The array was probed with cDNA derived from *Pseudomonas*-infected tomato leaf tissue harvested 4 hours post inoculation. cDNA was labeled with Cy3 using Genisphere's (Oakland, NJ) dendrimer technology.

Development of cDNA microarrays containing genes encoding signaling components.

Our research group is especially interested in genes involved in signal transduction pathways that might play a role in disease resistance. As one step towards targeting these genes in microarray experiments we are using automated BLAST searching algorithms to search the tomato EST database for all protein kinases, protein phosphatases, transcription factors, and pathogenesis-related genes. These efforts have identified a set of approximately 1,000 genes encoding signaling components and these are being used to develop a special "signaling array". Another approach to enrich for genes that might be involved in plant disease resistance is to search for ESTs that appear only in pathogen-infected tissues. To accomplish this search we have used an algorithm to extract ESTs from the TPR and TPS libraries that are present in: a) both libraries, b) only the resistant library, or c) only the susceptible library (**Figure 9**). A similar procedure is being used to extract ESTs that are unique to the mixed-elicitor treated leaf tissue library.

Expression profiling of plant defense responses.

We have many projects underway to use cDNA microarrays for examining tomato responses to various pathogens. The first phase of the ongoing work is focussing on gene expression profiling of tomato leaves over a time course in response to *P. syringae* pv. tomato with (or without) the *avrPto* gene. The second phase will center on analyses of tomato leaf responses to other pathogens (e.g. *Phytophthora infestans*, *Xanthomonas vesicatoria*, tomato spotted wilt virus) and to various elicitors (e.g. salicylic acid, jasmonic acid, ethylene, fenthion). A summary of ongoing expression profiling projects follows:

Bacterial pathogens:	<i>Pseudomonas</i> , incompatible/compatible
Fenthion response:	Comparison with <i>Pseudomonas</i>
Nonhost pathogens:	<i>Pseudomonas tabaci</i> , <i>P. phaseolicola</i> etc.
Fungal pathogens:	<i>Alternaria alternata</i> toxin
Specific elicitors:	SA, JA, C ₂ H ₄ , H ₂ O ₂ , Nitric oxide (using suspension cells)
TF overexpression:	Arabidopsis +/- Pti4 transcription factor
Transient assays:	<i>N. benthamiana</i> , Pto mutants: (S198A, Y207D)
Non-HR expression:	Examine non-cell death related gene expression

Collaborations:	
<i>Phytophthora infestans</i> :	Bill Fry and Chris Smart
Bacterial effectors:	Alan Collmer, Sandy Lazarowitz, Terry Delaney
Insect resistance:	Gregg Howe, Michigan State University
Elicitors in potato:	Barbara Baker, UC-Berkeley

Development of serial analysis of gene expression for analyzing plant defense responses.

We are setting up to use serial analysis of gene expression technique (SAGE) to identify and compare gene expression between different pathogen-treated tissues. To optimize this approach in our lab we are comparing gene expression differences between a wild type *Arabidopsis thaliana* genotype and a transformant overexpressing the tomato Pti4 transcription factor gene. The SAGE protocol entails the production of short ~9 bp 'tags' from defined location within transcripts, the cloning and concatemerization of these tags, and then sequencing to identify genes that are expressed in a particular tissue. The 9 bp tag is sufficient to assign the tag to a particular gene, provided it comes from a defined location within the gene. In addition to being a qualitative approach to genome-wide gene expression, the SAGE technique also has been shown as being quantitative provided that enough tags are sequenced. We have used the original SAGE protocol and incorporated a modification that involves different reaction conditions for the amplification of ditags. The modified PCR conditions resulted in excellent amplification of ditags from all our SAGE libraries. Currently, we are in the process of mass amplifying ditags from the two *Arabidopsis* genotypes and preparing them for sequencing.

4). Development of a tomato-Arabidopsis synteny map (S. Tanksley, Cornell University)

The purpose of this part of the project is to identify a common set of conserved orthologs between the tomato and *Arabidopsis* genomes and, based on these common orthologs, determine the level of microsynteny and macrosynteny between the two genomes. By accomplishing these goals we wish to use the *Arabidopsis* genome sequence to predict gene content and marker position in tomato as well as to shed light on the mechanisms by which plant genomes evolve over long periods of time.

Determining tomato:Arabidopsis synteny.

The strategy for identifying conserved orthologs, using the newly-developed tomato EST database, is depicted graphically in **Figure 9** and has the following steps:

- 1) Produce a large pool of tomato ESTs, representing as many genes as possible
- 2) Individually tBLASTx each tomato EST against the *Arabidopsis* genome tiling path
- 3) Identify tomato ESTs that fit these criteria:
 - Match a single *Arabidopsis* BAC at expect value less than -10
 - If there is a second matching BAC, must be at an expect value different by $\Delta 10$
- 4) Survey (via southern) the ESTs meeting these criteria
- 5) Map ones in tomato that are truly single copy

The "Arabidopsis tiling path" is a file constructed from the published sequence of *Arabidopsis* (approximately 85% of the genome is currently represented) (**Figure 10**) and allows each tomato EST to be tested for a potential match to a conserved ortholog on a specific BAC in the *Arabidopsis* genome (**Figure 11**). Approximately 20% of the tomato

ESTs are able to be assigned to a unique Arabidopsis BAC using the criteria described above. Approximately 50% of these ESTs are determined to be single copy after genomic southern blots against tomato DNA and pass through to the mapping phase.

At this time, 620 single copy orthologs have been identified computationally, 570 were surveyed on genomic southern blots, and 35 were found to be single/low copy and mappable (the rest were multiple copy or monomorphic; **Figure 12**). Mapping on new F2 population has been initiated (population being maintained in sterile culture). This same F2 is being imaged for floral structures (and other characters) in order to identify loci involved in carpel/fruit size and shape. The goals are to identify a total 500-1000 single copy orthologs by August 2000 and to have 500 mapped within the next year. The new mapping population will have an average bin size of approximately 1 cM and we are targeting to assign each ortholog to a BIN and to determine the BAC tile match in Arabidopsis.

In parallel with the whole genome approach for determining macrosynteny, we have also begun to examine, in finer detail, specific regions of the tomato genome to determine the level of microsynteny with Arabidopsis. Most advanced is the mapping of a section of the long arm of tomato chromosome 2 that has a match with a segment of Arabidopsis chromosome 4 (**Figure 13**). In this instance, conserved orthologs map in the same linear order in both genomes, indicating that microsynteny does exist. We plan to expand the detailed microsynteny study to include four other regions of the tomato genome. This will be in addition to the global, macrosynteny study described above.

cDNA libraries and EST database.

My lab is coordinating cDNA library construction and scheduling for sequencing among the four labs (Ronning, Martin, Giovannoni, Tanksley). cDNA library construction will continue throughout this year (see previous table of targeted tissues/elicitors for library construction). It is anticipated that all library construction will be completed by March 2001 and that all sequencing will be completed by the end of the funded period. It is projected that 152,000 ESTs will have been generated by the end of the project. This number is greater than that originally proposed (100,000) and covers a much broader range of tissues/elicitors (see previous table) in order to maximize gene discovery. All sequences will be immediately deposited in GenBank and the TIGR website.

Tomato:Arabidopsis synteny project.

We will continue to screen tomato ESTs to identify a set of 500-1000 single copy orthologs (between tomato and Arabidopsis genomes). We plan to map at least 500 of these in tomato within the next year. Map locations in tomato will be assigned using the new mapping population (average bin size of approximately 1 cM -- see earlier section

for details). Microsynteny will also be determined in 3-5 selected regions of tomato based on BAC sequences. We expect a general picture of the extent and nature of both macro and micro synteny, between the tomato and Arabidopsis genomes, will emerge later this year. We are also coordinating the tomato:Arabidopsis synteny research with parallel research in potato in order to tie together all three genomes.

SGN database development.

We will continue with development of the Solanaceae Genomics Network database, providing the first on-line service by August 2000, with continued upgrades throughout the year. We also hope to begin interfacing this database, and the synteny results, with other dicot synteny research ongoing in other labs.

NOTE: Figures are placed in the Findings Section.