DEVELOPING TRANSCRIPTIONAL PROFILES AND GENE EXPRESSION ANALYSIS OF GRAPE PLANT RESPONSE TO *XYLELLA FASTIDIOSA*

Project Leader:
Hong Lin
Crop Diseases, Pests, and Genetics
USDA, ARS
Parlier, CA 93648

Cooperators:
Harsha Doddapaneni and Andrew Walker
Department of Viticulture and Enology
University of California
Davis, CA, 95616

Reporting Period: The results reported here are from work conducted October 2004 to September 2005.

ABSTRACT
In the present study, a set of 1,942 non-redundant SSH ESTs in response to *Xylella fastidiosa* (*Xf*) infection were cloned from susceptible and resistant sibling genotypes from a *Vitis rupestris* x *Vitis arizonica* genetic mapping population. The majorities (54 %) of these ESTs were novel and the rest included genes known to be involved in defense responses. Transcriptional profiling using a custom high-density (382,900 probes) microarray chip of 20,020 *Vitis* transcripts showed a significant variation in response between the susceptible and resistant genotypes to *Xf* infection. Out of the 793 transcripts that showed significant response to *Xf* infection, 28.1% (223 ESTs) were derived from this project.

INTRODUCTION
The impact of Pierce's Disease (PD) on the California grape industry has been significant since the introduction and establishment of a more effective vector, *Homalodisca coagulata* (*H. coagulata*), the glassy-winged sharpshooter (GWSS) (Almeida and Purcell 2003). Development of resistance in grape is stymied by the relatively limited amount of genetic and molecular information regarding genotype specific resistance to PD infection (Davis et al. 1978). Breeding efforts confirm that resistance is inheritable and molecular mapping has linked DNA markers to *Xf* resistance (Krivanek et al. 2005). Once cloned, the next step is to incorporate PD resistance genes into traditional grape cultivars. This objective can be achieved in fewer generations by elucidating the molecular basis of resistance and pathogenicity, which prompted us to develop a functional genomic approach for PD research.

SSH (Suppression Subtractive Hybridization) DNA cloning is one of the most efficient and comprehensive methods used for identifying genes involved in differentially regulated conditions. This is particularly important in host-pathogen interactions where many pathogen-related genes are expressed at low abundance and limited to particular tissues or cell types at certain times (Caturla et al., 2002). Some of these genes are less likely to be cloned by standard EST cloning methods. Here, we explored the utility of subtractive DNA cloning to characterize differentially expressed genes in response to *Xf* infection between highly susceptible and resistant sibling genotypes segregating from the *Vitis rupestris x Vitis arizonica* population. To maximize the chances of cloning expression profiles associated with the tissues and at various stages of host-pathogen interactions, we designed a time course sampling scheme and constructed tissue specific cDNA libraries.

Plants respond to pathogen attack through a variety of signaling pathways consisting of a large number of regulatory as well as effector genes. Microarrays facilitate automated analysis of transcriptional profiling data to enable complete understanding of such gene function and interactions. A custom 60-mer high-density oligoarray chip was designed using the generated EST collection as well as incorporating the entire *Vitis* transcriptome information in the public domain to understand the *Xf* / *Vitis* interaction.

OBJECTIVES
1. Sequence analyses of SSH cDNA libraries.
2. Design of high-density expression array.
3. Microarray gene expression analysis.

RESULTS
Objective 1: Sequence analyses of SSH cDNA libraries
Sequencing, EST assembling and dbEST submission
A total of 5,794 ESTs, with ~500 ESTs from the libraries, were sequenced from 12 constructed SSH libraries (Figure 1). The average EST size was 282 bp with more than 5,400 sequences of at least 100 bp in length. In order to reduce the transcript redundancy, the EST collection was subjected to PHRED, PHRAP and BLAST analysis to do clustering, first within each library and then among all 12 libraries. The final assembling resulted in 1,942 unique sequences including 993 clusters (contigs) and 949 singleton ESTs. All the sequenced ESTs that are at least 100 bp in length (5421 ESTs) were submitted to the NCBI's ESTdb under the accession numbers DN942225 to DN947645.

EST similarity search and functional assignments
Comparison of the 1,942 non-redundant sequences against the non-redundant protein database of the NCBI revealed that 716 sequences have significant similarity ($\leq 1E^{-5}$) and the remaining had no hits. A list of EST hits against the known disease resistance related proteins are presented in Table 1. Functional annotation was carried out using an ontology database available in the USDA-ARS system that is based on the functional classification schemes such as Gene Ontology (GO), Enzyme Commission numbers (EC), BioCarta Pathways, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. All the non-redundant sequences were searched against this database using default blast parameters and a cut off E value of 10^{-4}. Based on the generated GO information, 906 sequences were divided into the three principal GO categories: molecular function (30%), cellular component (9%) and biological process (7%) (Figure 2).

Under the molecular function category, ligand binding and carrier contributed for 25% of the total contigs followed by the ribosomal coding transcripts. Similarly, transport followed by signal transduction and defense response formed majority of the total numbers in the biological process category, while chloroplast, membrane and nucleus subsections had major contributions in the cellular component category. More than half of the sequences (54 %) could not be annotated due to lack of significant similarly with the known proteins.

EST cluster analysis
Co-expressivity of the transcripts was accessed across the 12 SSH libraries, using hirearchial clustering based on the transcript abundance. A matrix file for 437 contigs that are represented by 5 or more ESTs across the 12 libraries were selected out of the total 993 contigs similar to Ewing et al. (1999). A total of 11 clusters were generated using partitional clustering (Qtclust algorithm) to divide the data into clusters of coexpressed contigs with the maximal inter-contig Pearson correlation of 0.99 and with the minimal cluster size of 5. There were 5 to 73 member contigs per cluster. To reduce the bias resulting from outliners, the distances used for computing clusters were jackknifed. Hierarchical tress was next generated from these data sets by calculating the Euclidean distance with complete linkage option. Based on the above parameters, 37% (174 contigs) could not be assigned to any cluster.

In general, across all the clusters, the level of expression for a particular contig was significantly different between the tissue types that were used as tester and driver populations suggesting that the preparation of the subtractive libraries was optimal. For instance, Cluster-A had 73 contigs which were abundantly expressed in four libraries, Lib 2, Lib 5, Lib 4 and Lib 6, out of which, 12 contigs (top left, including 11 with 'no hit" and 1 metallothionin like protein) showed abundant expression only in the Lib 2 (Stem library from resistant genotype infected with *Xylella*) and are potential candidates from the PD point of view. Similarly, a set of 11 contigs enriched with pathogen responsive P-rich protein genes showed hyper expression in the non-infected stem library from the same genotype. Cluster-B has groups of contigs abundantly expressed in the infected stem tissue library from susceptible genotype (Lib 3), all of which interesting are known to play a role in the defense mechanism. This cluster also groups contigs over expressed in non-infected leaf tissue from the resistant genotype (Lib 4) and all of them appear to be novel.

Objective 2: Design of high-density expression array
EST assembly
To maximize gene discovery, we have designed a custom microarray chip based on our ESTs and the publicly available EST sequences from all *Vitis* (*V.*) varieties and species. From the public domain, a total of 33,933 ESTs -12,593 unigene set from the *V. vinifera*, 10,704 accessions of *V. shuttleworthii*, 6,533 of *V. rupestris* 'A. de Serres' x *V.* sp, 2,117 of *V. aestivalis* and 1,986 sequences of *V. riparia* were collected. Redundancy in each of the non-*vinifera* EST sets was reduced based on BLAST analysis and the longest EST from each cluster was selected. Next, repeat ESTs between the sets were removed and at this step between matching ESTs, and the *vinifera* ESTs were discarded to facilitate enrichment of the final set with non-*vinifera* ESTs.

These efforts resulted in a total of 20,020 ESTs with 1,947 from the SSH libraries, including 40 from the cDNA-AFLP experiments, 10,014 from *V. vinifera*, 5,470 from *V. shuttleworthii*, 1,219 from *V. aestivalis*, 780 from *V. rupestris* x *V.* sp and 588 from *V. riparia.*

Probe design
Probe design was carried out in collaboration with Nimblegen Inc., with active input from our group. Nine individual 60-mer probes were designed for each EST and for 1,634 ESTs, probes were designed from both the strands. A total of 191,450 probes were selected for the entire set and there were two spots for each probe on the slide totaling 382,900 spots per slide.

Objective 3: Microarray gene expression analysis
Experimental set-up
A total of 18 slides (9 of genotype 9621-67 (resistant) and 9 of genotype 9621-94 (susceptible)) were hybridized in a two color hybridization experiment using the monochromatic dyes Cy 5 and Cy 3. Total RNA from stem tissues at three stages of

disease development- week-1, week-6 and week-10 from both control (non-infected) and experimental (infected) was used. There were three biological replications for each stage and this included a dye flip.

Data analysis

Following normalization, data from the hybridization experiments were statistically analyzed using Perl scripts and Excel package. For each gene there were 54 data points per each stage (18 per slide x 3 biological replications) of disease development. Only signal ratios that had less than 20% SE across the measured values were included for results interpretation. A cut-off value of 2-fold response was considered significant and as evident from Table 2, the response was more surprising for the susceptible genotype both in terms of numbers and magnitude in the later stages of disease development. Further experiments based on RT-PCR are in progress to confirm the observed microarray results.

CONCLUSIONS

Characterizing the molecular basis of the grape response to *Xf* is critical to understanding the mechanisms of PD resistance and pathogenesis. The generated EST pool and subsequent microarray based genome-wide transcription profiles have identified, for the first time, a pool of ESTs expressed under defined conditions and might be the candidates in determining resistant and susceptible interactions. Efforts are underway to generate transcription profiles in leaf tissue. Currently, we are developing a relational database incorporating the generated transcriptional data that will be open to public access.

REFERENCES

Almeida, R.P. and A.H. Purcell. 2003. Transmission of *Xylella fastidiosa* to grapevines by *Homalodisca coagulata* (Hemiptera: Cicadellidae). J Econ Entomol. 96:264-271.

Caturla, M., C. Chaparro, K. Schroeyers, and M. Holsters. 2002. Suppression subtractive hybridization to enrich low-abundance and submergence-enhanced transcripts of adventitious root primordia of *Sesbania rostrata*. Plant Science 162:915-921.

Davis, M. J., A. H. Purcell, and S. V. Thomson.1978. Pierce's disease of grapevines: isolation of the causal bacterium. Science 199: 75-77.

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

Krivanek, A.F., T. R. Famula, A. Tenscher, and M.A. Walker. 2005. Inheritance of resistance to *Xylella fastidiosa* within a *Vitis rupestris x Vitis arizonica* hybrid population. Theor Appl Genet. Apr 28; [Epub ahead of print]

FUNDING AGENCIES

Funding for this project was provided by the CDFA Pierce's Disease and Glassy-winged Sharpshooter Board.

Table 1. Blast hits of a sub-set of the ESTs showing high similarity to known disease resistant genes.

EST I.D.	Length	BlastX- top Hit	E-value
EST-1	737	emb|CAA95857.1| S-adenosyl-L-methionine synthetase 2 [*Catharanthus roseus*]	1.00E-122
EST-2	610	emb|CAC14015.1| chitinase [*Vitis vinifera*]	5.00E-98
EST-3	535	gb|AAR13304.1| phytochelatin synthetase-like protein [*Phaseolus vulgaris*]	4.00E-87
EST-4	468	emb|CAB91554.1| beta 1-3 glucanase [*Vitis vinifera*]	6.00E-64
EST-5	448	emb|CAC16165.1| pathogenesis-related protein 10 [*Vitis vinifera*]	4.00E-53
EST-6	302	dbj|BAA76424.1| rac-type small GTP-binding protein [*Cicer arietinum*]	5.00E-48
EST-7	842	ref|NP_850638.1| zinc finger (DHHC type) family protein [*Arabidopsis thaliana*]	3.00E-32
EST-8	348	pir||T07139 cysteine proteinase inhibitor - soybean dbj|BAA19608.1|	2.00E-30
EST-9	456	gb|AAN71733.1| WRKY transcription factor IId-4 [*Lycopersicon esculentum*]	2.00E-21
EST-10	265	gb|AAM21199.1| pathogenesis-related protein 5-1 [*Helianthus annuus*]	3.00E-21
EST-11	171	gb|AAD55090.1| thaumatin [*Vitis riparia*]	1.00E-19
EST-12	372	gb|AAN75467.1| mitogen-activated protein kinase [*Lycopersicon esculentum*]	1.00E-14
EST-13	418	gb|AAN15621.1| O-methyltransferase-like protein [*Arabidopsis thaliana*]	4.00E-14
EST- 14	116	gb|AAP23944.1| leucine-rich repeat protein [x *Citrofortunella mitis*]	1.00E-13

Table 2. Microarray hybridization response of *Vitis* stem tissue to *Xf* infection in both susceptible and resistant genotypes. Numbers in parenthesis represent the ESTs generated from SSH experiments.

Stage	Response	Genotype			
		9621-67		9621-94	
		# Of ESTs	Fold-Change	# Of ESTs	Fold-Change
Week-1	Up-regulated	38 (4)	2.0 - 3.1	2 (2)	2.1 - 2.36
	Down-regulated	24 (11)	0.49 - 0.33	1 (1)	0.49
Week-6	Up-regulated	19 (11)	2.0 -2.43	152 (48)	2.0 - 37
	Down-regulated	11 (2)	0.49 - 0.33	71 (15)	0.49 - 0.12
Week-10	Up-regulated	81 (5)	2.0 -3.57	99 (24)	2.0 - 22.38
	Down-regulated	61 (33)	0.49 - 0.15	234 (67)	0.49 - 0.11

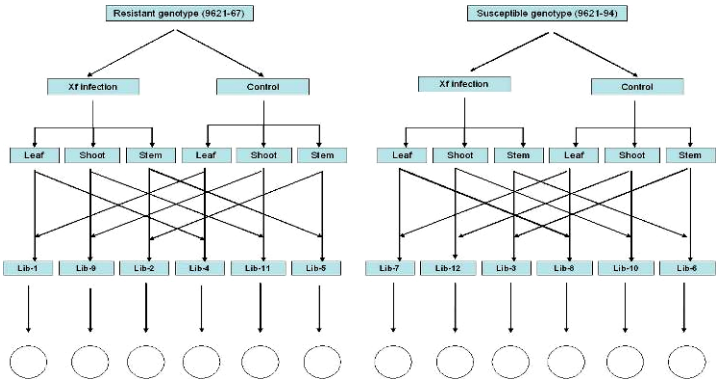


Figure 1. Strategy for SSH library construction.

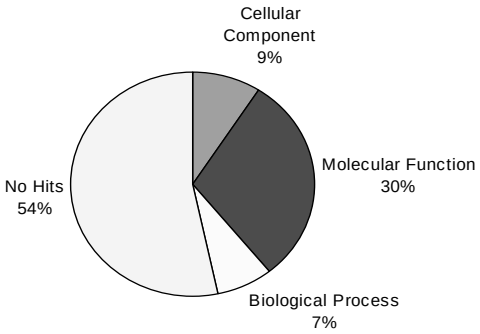
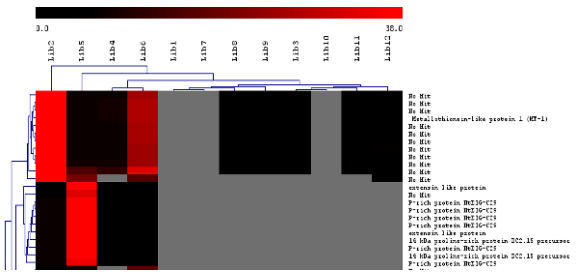


Figure 2. EST distribution among the three GO principles.

Cluster A



Cluster B