Lehrstuhl für Pflanzenbau und Pflanzenzüchtung
der Technischen Universität München
in Freising-Weihenstephan

Identification of differentially expressed genes associated with sugarcane mosaic virus resistance in maize (*Zea mays L.*)

Chun Shi

Vollständiger Abdruck der von der Fakultät Wissenschaftzentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ-Prof. Dr. rer. nat. G. Forkmann
Prüfer der Dissertation:	1. Univ.-Prof. Dr. rer. nat. G. Wenzel
	2. Univ.-Prof. Dr. agr. F. J. Zeller
	3. Priv.-Doz. Dr. rer. nat. T. Lübberstedt

Die Dissertation wurde am bei 02.12.2004 der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftzentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 27.01.2005 angenommen.

Contents

Introduction

Resistance to sugarcane mosaic virus (SCMV)

SCMV causes mosaic diseases in sugarcane, maize, sorghum and other Poaceous species worldwide. It has resulted in considerable economic losses in sugarcane and failure of commercial clones in several countries. Yield losses of 30 - 40% and sometimes 60 – 80% have been recorded in the western hemisphere (King 1955-56, Forbes and Steib 1964, Koike and Gillaspie 1989). SCMV is also responsible for yield losses of 10 – 30% and 10 – 50% in China and South Africa, respectively (Chiu 1988, Fauquet and Wechmar 1988). So far, it is one of the most important virus diseases of maize in Europe and causes serious yield losses in susceptible cultivars (Fuchs and Gruntzig 1995) (Figure 1).

Figure 1. SCMV infected maize leafs with different levels of mosaic symptoms. Infection level increases from left to right.

SCMV particles are flexuous, rods of 730 – 755 nm long and 13 nm wide and composed of a single polypeptide species of 28,500 – 35,000 Daltons consisting of 264 – 328 amino acid residues surrounding a single stranded, positive sense RNA species (Koike and Gillaspie 1989, Teakle et al. 1989). It is readily transmitted by grafting, mechanical inoculation and a number of aphids in a non-persistent manner (Koike and Gillaspie 1989). SCMV was formerly denoted as a MDMV isolate, MDMV-B (Shukla et al. 1989). Together with wheat streak mosaic virus (WSMV), Johnson grass mosaic virus (JGMV), Sorghum mosaic virus (SrMV), and MDMV, it belongs to the same taxonomic group of related pathogenic potyviruses in maize. Since the 1980s, SCMV and the closely related maize dwarf mosaic virus (MDMV) have been found in Germany (Fuchs and Kozelska 1984). In Germany, SCMV is more prevalent than MDMV and causes increasing damage to maize (Fuchs et al. 1996), while MDMV is a widespread viral disease in the southern US Corn Belt (Louie et al. 1991).

Though chemical control of vectors is commonly practiced for the management of viral diseases, it has not found its' place in SCMV management due to the non-persistent transmission of aphids. Cultivation of resistant maize varieties is the most efficient and environmentally sound approach to manage SCMV. In a study with 122 early-maturing European maize inbreds, three lines (FAP1360A, D21, and D32) displayed complete resistance and four lines displayed partial resistance (FAP1396A, D06, D09, and R2306) against SCMV and maize dwarf mosaic virus (MDMV) (Kuntze et al. 1997). In field trials, resistance of all three European lines D21, D32, and FAP1360A seemed to be controlled by one to three genes (Melchinger et al. 1998). Two major QTL regions, *Scmv1* and *Scmv2*, conferring resistance to SCMV were mapped to chromosome arms 6S

and 3L. In cross D145 × D32 quantitative trait locus (QTL) analysis (Xia et al. 1999) and in cross F7 × FAP1360A bulked segregant analysis (BSA) (Xia et al. 1999) and QTL analysis (Dussle et al. 2000) were applied. Minor QTLs affecting SCMV resistance were identified on chromosomes 1, 5, and 10 (Xia et al. 1999). For complete resistance to SCMV, presence of both *Scmv1* and *Scmv2* is essential. *Scmv1* suppresses symptom expression throughout all developmental growth stages at a high level, whereas *Scmv2* was mainly expressed at later stages of infection (Xia et al. 1999, Dussle et al. 2000).

Selection of candidate genes (CGs)

Positional cloning is the major approach used to characterize genes underlying QTL, but it is very laborious and time consuming. The candidate-gene approach provides an alternative for pinpointing genes underlying SCMV resistance, especially in view of the planned sequencing of major parts of the genome (Martienssen et al. 2004). CGs are proposed from two classes: functional CGs based on molecular and physiological studies, and positional CGs based on linkage data of the locus being characterized.

Maize resistance gene analogues (RGA) involved in initial pathogen recognition, were chosen as starting point for isolation of genes conferring SCMV resistance (Collins et al. 1998). Mapping of RGAs in relation to *Scmv1* and *Scmv2* suggested that RGA *pic19* is a candidate for *Scmv1* and *pic13* for *Scmv2* (Quint et al. 2002). *pic19* and *pic13* were used to screen a BAC library of B73 and three paralogues clustering in the *Scmv1* region were isolated from the maize genome (Quint et al. 2003), currently analyzed in more detail (Xu and Lübberstedt, unpublished results).

Construction of specific cDNA libraries corresponding to different organs, developmental stages or stress responses coupled to differential screening of these

libraries fosters the isolation of CGs. For instance, Mazeyrat et al. (1998) identified genes specifically induced during plant defense by screening cDNA libraries corresponding to fungi-infected and healthy sunflowers. Near isogenic lines (NILs) are excellent materials to construct subtractive libraries (Borevitz and Chory 2004). Because these lines are almost identical, the background noise due to variable genome regions is eliminated. In this study, five SSH (suppression subtractive hybridization) libraries constructed from the NILs F7 (SCMV susceptible) and $F7^+$ (SCMV resistant, carrying *Scmv1* and *Scmv2* regions from FAP1360A) were screened to identify candidate genes for the previously mapped QTL, but also genes from other chromosomal locations involved in subsequent steps leading to resistance or susceptibility after the initial recognition of SCMV.

cDNA- and oligonucleotide microarray technologies hold great promise for identifying CGs and for monitoring the expression of mRNAs or the occurrence of polymorphisms in genomic DNA (Pflieger et al. 2001) as already shown in strawberry (Aharoni et al. 2000) and tomato (Giovanonni 2000). We investigated the NILs F7 and $F7^+$ to conduct microarray experiments. Differentially expressed genes might be derived from the *Scmv1* or *Scmv2* genome regions, and thus, be candidate genes for the previously mapped QTL. If located in other genome regions, these genes might be further downstream in the signal transduction pathway and induced by genes located in the *Scmv1* and / or *Scmv2* regions.

Once genes responsible for quantitative variation of SCMV resistance become available, information can be passed on to plant breeders in the form of functional markers (Andersen and Lubberstedt 2003). Functional markers are superior to random DNA markers such as RFLPs, SSRs and AFLPs owing to complete linkage with trait

locus alleles. Due to polygenic trait of SCMV resistance, marker-assisted selection (MAS) programs with functional markers would increase breeding efficiency.

A mechanistic view of maize-SCMV interactions

Except the identification of *Scmv* candidate genes, gene expression studies also provide a strong tool to reveal the defense mechanisms of SCMV resistance. An unusually high frequency of genes conferring recessive resistance has been observed in relation to potyviruses (40% versus 20% for resistance against other viruses), in which the plant lacks one or more factors required for virus replication or movement (Provvidenti and Hampton 1992). However, resistance genes *Scmv1* (*Scmv1a*, *Scmv1b*), and *Scmv2* displayed at least partial dominance in different studies (Xia et al. 1999, Dussle et al. 2000, Yuan et al. 2003). Moreover, no hypersensitive response (HR) symptoms are observed for maize leaves infected with SCMV. The defense mechanism without HR applying to SCMV resistance is poorly understood.

Due to the widespread application of global transcript profiling technology in the field of plant–pathogen interactions, it's now clear that the plant response to pathogen infection is associated with massive changes in gene expression (Katagiri 2004). In an array representing about 8,000 Arabidopsis genes, more than 2,000 genes changed expression level within nine hours of inoculation with the bacterial pathogen *Pseudomonas syringae* (Tao et al. 2003). Recent opinion about plant-pathogen interaction is that the plant defense response is probably not highly specialized. When a plant detects a pathogen, it does not tailor its response to the pathogen at hand. Instead, it turns on many of the defense mechanisms it has, among which some may be effective against a particular pathogen (Katagiri 2004). It is difficult to define the difference between genes

that are part of the defense response and genes that play other roles during infection. For example, turning on defense mechanisms is energy intensive, and some genes might be induced or repressed to promote efficient energy utilization during defense (Katagiri 2004). Although the importance of low false-positive rates in expression profiles is often emphasized for gene discovery studies, low false-negative rates are also important for global analysis. The statistical criteria chosen for defining genes with significant changes in expression level should provide a balance between false-positive and false-negative rates that is appropriate for the purpose of the analysis (Katagiri 2004).

When the resistance of a plant to a particular pathogen is controlled by gene-for-gene relationships (Dangl and Jones 2001), there is usually a very clear phenotypic difference between the resistant and susceptible responses. For this reason, the idea that resistance is associated with resistance-specific responses has been emphasized. Although resistance-specific responses certainly exist, large sections of the global changes revealed by expression profiles are qualitatively similar in resistant and susceptible responses (Katagiri 2004). The major differences between resistant and susceptible responses are quantitative and/or kinetic. That is, the shapes of the expression profiles from resistant and susceptible interactions are similar at early stages of the interactions, but the amplitude of the profile from the susceptible interaction is lower than that from the resistant interaction (Katagiri 2004). This quantitative/kinetic notion of resistance and susceptibility was proposed long ago (Lamb et al. 1992), but global expression profiles have revealed that it is the rule rather than an exception (Katagiri 2004). Thus, global transcript profiling, as a broad-spectrum phenotyping method, has begun to reveal large-scale behaviors of the signaling network that were previously difficult to study. The

application of transcript profiling technologies to SCMV resistance study will advance our understanding of maize-SCMV interactions to a higher level.

Methods of transcript profiling

High-throughput transcript profiling methods can be divided into two classes: (1) direct analysis, including procedures involving nucleotide sequencing (EST sequencing, SSH, SAGE) and fragment sizing (e.g., cDNA-AFLP); and (2) indirect analysis (macro- or microarray based expression profiling), involving nucleic acid hybridization of mRNA or cDNA fragments (Donson et al. 2002).

Large-scale EST sequencing is attractive because they do not rely on established sequence data from the organism under study, and they also fit well with labs already equipped to carry out high-throughput DNA sequencing (Adams et al. 1991). However, even at a few dollars per sequence the process can be expensive if one desires to progress beyond cursory screening of abundant mRNAs to in depth analysis (Ohlrogge and Benning 2000). Auxiliary techniques are available that reduce the amount of sequencing. These include subtraction hybridization (Sargent 1987) and related methods, SSH (Diatchenko et al. 1996). Except for lower set-up costs, SSH procedure enriched the library for low-abundant and differentially expressed mRNAs by normalization (Diatchenko et al. 1996). Otherwise, abundant pathogenesis-related transcripts (e.g., genes coding for PR proteins) would very likely have masked important SCMV-specific transcripts expressed at much lower levels. (Birch et al. 1999) have used SSH to isolate potato genes induced during an early stage of the HR to *Phytophthora infestans*.

Though a similar sequence-based method to EST analysis, serial analysis of gene expression (SAGE) achieves a cost-efficiency, by the concatenation and punctuation of

multiple sequence tags of 10 – 14 bp, prior to cloning (Donson et al. 2002). By the size selection of inserts containing 25 – 50 tags, a comparable reduction of cost or increase in depth of analysis can be achieved over the sequencing of single ESTs. However, this increased efficiency comes at the price of more extensive sequence reads. Consequently, this technology is best applied to organisms whose genomic sequences are known or that have a substantial cDNA sequence database. Even with a reference database, because the tags are so short, there can be a redundancy of matches (Donson et al. 2002).

Fragment sizing involves the discrimination of mRNAs by differential separation of representative cDNA fragments on matrices (Donson et al. 2002). Amplified restriction fragment length polymorphism (AFLP) of cDNA is the most popular used method of this approach and easily and inexpensively performed as SSH. It doesn't rely on EST databases or existing cDNA libraries, allow detection of rare transcripts, and require relatively small amounts of mRNA. The main disadvantages include heterogeneity of final products, the need to clone and sequence the product for identification and the need to isolate a full-length cDNA after obtaining the PCR product (Baldwin et al. 1999).

Recently, cDNA microarraying has had substantial impact on molecular biology, invited by the availability of genomic sequences. It has become the predominant method for the parallel analysis of gene expression in phytopathology research (Wan et al. 2002), under different defense-related treatments and over different time points. These technologies open up tremendous opportunities to identify new pathogenesis-related genes, to identify co-regulated genes and the associated regulatory systems, and to reveal interactions between different signaling pathways (Wan et al. 2002). Baldwin et al.

(1999) identified 117 genes that consistently showed altered mRNA expression in maize 6 h after various treatments with the fungal pathogen *Cochliobolus carbonum*, using a maize DNA microarray representing 1,500 maize genes. Using a similar approach, Nadimpalli et al. (2000) identified nearly 70 genes having a more than twofold change in mRNA abundance in the lesion mimic maize mutant, *Les9* compared to wild-type plants. In contrast to comprehensive microarrays in *Arabidopsis* and rice (Tao et al. 2003), publicly available maize unigene-microarrays (Nakazono et al. 2003) actually contain 9,841 different unigenes, which account for only 20% of the about 50,000 maize genes (Martienssen et al. 2004). Thus, another expression profiling method, combining SSH and macroarray hybridization, was conducted in a companion study to detect SCMV-related transcripts in maize. Combining both results will provide a more comprehensive understanding of SCMV-maize interaction.

Objectives

In this thesis, I used NILs to identify differentially expressed genes associated with SCMV resistance in maize by two transcript profiling methods, SSH-macroarray and unigene-microarray. The objectives were to (i) identify CGs for *Scmv* QTLs and downstream genes in the signal transduction pathway, (ii) reveal the defense mechanism of maize-SCMV interactions, (iii) investigate expression-level polymorphisms among different maize inbreds.

Materials and methods

Plant materials

$F7^+$, F7, D21, D32, FAP1360A, Pa405, D145, and D408 were grown and maintained in growth chambers under a 12 -h photoperiod at 23 ^oC and 50% relative humidity. Two-week-old plants were mechanically inoculated by an air brush technique using a tractor-mounted air compressor at constant pressure of 799 kPa (Fuchs and Gruntzig 1995). Non-infected plants and infected plants were kept in separate growth chambers after inoculation. Non-infected and infected leaves were harvested 24 hours after inoculation in parallel. For biological replicates, two independent sets of leaf materials were harvested. To confirm resistance or susceptibility of infected plants used for leaf harvest, plants were grown for additional two weeks.

RNA extraction

Total RNA from maize leaves was extracted using TRIzol reagent (Invitrogen, Carlsbad, California, USA). Poly (A)+ RNA was separated from total RNA using Oligotex mRNA Midi Kit (Qiagen GmbH, Hilden, Germany).

Suppression subtractive hybridization (SSH) and cDNA library construction

cDNA synthesis and SSH were carried out using a PCR-Select cDNA Subtraction Kit (BD Biosciences, San Jose, CA, USA) according to manufacturer's protocol. The subtracted tester cDNA was cloned into pCRII-TOPO TA cloning vector (Invitrogen, Carlsbad, CA, USA) and transformed into One Shot® TOP10 electrocompetent *E. coli* cells (Invitrogen, Carlsbad, CA, USA). Finally, five SSH libraries were constructed. For two tester/driver cDNA pairs (infected $F7^+$ versus infected F7; non-infected $F7^+$ versus

non-infected F7) subtractions were conducted in both directions. For the tester/driver cDNA pair infected $F7^+$ versus non-infected $F7^+$ only forward direction was conducted (Table 1).

Table 1 Overview over five cDNA libraries generated by suppression subtractive hybridization

No.	Tester	Driver	Genes expected to be enriched
1	Infected $F7^+$	Not infected $F7^+$	SCMV induced genes in resistant genotype
2	Infected $F7^+$	Infected F7	Genetic discrepancy between NILs including induced resistance factors
3	Infected F7	Infected $F7^+$	Genetic discrepancy between NILs including repressed resistance factors
4	Not infected $F7^+$	Not infected F7	Genetic discrepancy between NILs including preformed
5	Not infected F7	Not infected $F7^+$	resistance or susceptibility factors

Construction of macroarrays

From each subtracted cDNA library, more than 384 colonies were randomly picked and PCR-amplified in 25 µl reactions in 96-well plates. Specific primers were designed (forward primer: 5'-ATGCTTCCGGCTCGTT-3'; reverse primer: 5'-CAGGGTTTTCCCAGTC-3'). After gel eletrophoresis, clones having inserts were collected in 96-well plates. Thereafter, two independent re-amplification 100 µl reactions were performed and pooled to reduce variation in PCR efficiency. Pooled PCR products were concentrated from 200 µl to approximately 25 µl by MultiScreen-PCR plates (Millipore, Billerica, Mass, USA) before transfer to 384-well plates. We used the BioGrid robotic system (BioRobotics Ltd. Cambridge, U.K.) with a 384 gridding tool (radius 0.4 mm, 5 transfers per spot) to spot cDNA clones onto H-bond nylon transfer membranes (Amersham, Piscataway, NJ, USA). The spotting scheme followed a 4x4 secondary grid pattern, with each secondary grid containing seven clones spotted in duplicate, plus two empty local background spots. DNA on membranes were crosslinked in a GS Gene Linker chamber (Bio-Rad Laboratories, Hercules, CA ,USA). In total, 2688 clones were

spotted on each macroarray, including internal controls and two RGAs (*pic13* and *pic19*) (Quint et al. 2003). Microarray hybridization data were evaluated by the SpotReport™ Alien™ cDNA Array Validation System (Stratagene, La Jolla, CA, USA), containing positive, negative, and ten spiking controls. Equal amounts of PCR products from ten different alien genes were spotted on the macroarray as spiking controls. Hybridization signals detected from control spots on the macroarray were evaluated to determine (i) quality of both the macroarray and the mRNA, (ii) the macroarray orientation, and (iii) the sensitivity, specificity, signal linearity, and consistency of the assay.

Macroarray hybridizations

In each RT reaction, 5, 2.5, 1.25, 0.625, 0.312, 0.156, 0.078, 0.04, 0.02 and 0.01 ng of the ten different alien mRNA spikes were added to 5 µg of total RNA. Using the Strip-EZ RT kit (Ambion, Austin, TX, USA), cDNA synthesis was primed using oligo(dT) and [α-^{32}P]dATP (MP Biomedicals, Irvine, CA,USA). cDNA was separated from unincorporated nucleotides using Micro Bio-Spin chromatography columns (Bio-Rad Laboratories, Hercules, CA ,USA) filled with Sephadex G-50 (Amersham, Piscataway, NJ, USA) equilibrated in water. tRNA and oligo(dA) was added to the hybridization probe to suppress cross-hybridization. The prehybridization and hybridization steps were conducted as described in the manual of the Strip-EZ RT kit. Labeled cDNA probes were stripped from the arrays using the Strip-EZ system (Ambion, Austin, TX, USA), and the process checked by phosphorimaging. For technical replicates, every RNA sample was used in two independent labeling and hybridization experiments. Hybridization signals were detected using the Storm 860 phosphorimager (Amersham, Piscataway, NJ, USA) with a resolution of 50 µm.

Unigene-microarray hybridizations

For NILs $F7^{+}$ and F7, the same total RNA extracted from SSH-macroarray approach was also used for unigene-microarray hybridization. Poly (A)+ RNA was isolated from Total RNA via Dynabeads® Oligo (dT) 25 (Dynal biotech, Oslo, Norway). According to TIGR Microarray Protocols (Hegde et al. 2000), each mRNA sample was indirectly labeled with Cy3 or Cy5 (Amersham Pharmacia, Piscataway, NJ, USA) and hybridized with maize unigene-microarrays. Fluorescence signals were detected using the arrayWoRx® Biochip Reader (Applied Precision, Issaquah, Washington, USA).

Maize unigene-microarrays were generated by the laboratory of Prof. Schnable (Iowa University, USA) and contain 11,827s maize ESTs (http://www.plantgenomics.iastate.edu/maizechip/). Among them, 11,027 ESTs were spotted once, 391 ESTs duplicate and 6 ESTs triple. Thus, 11, 424 unique ESTs, clustered into 9841 unigenes, are on the maize unigene-microarray, and 8.3% (949 of 11,424) of them have been mapped. The EST collection at the maize unigene-microarrray was derived from fifteen EST libraries, including 486 (immature leaf), 605 (endosperm), 606 (ear tissue), 614 (root), 618 (tassel primordia), 660 (mixed stages of anther and pollen), 683 (14 day immature embryo), 687 (mixed stages of embryo development), 707/945 (mixed adult tissues) and ISUM3/4/5/6/7 (seedling and silk), made from plants grown under normal environmental conditions and two stress-induced EST libraries, including 496 (stressed shoot) and 603 (stressed root).

For each comparison, four replications, including two biological replications and dye swap replications in each biological replication, were conducted. Thus, four maize gene chips were used in each comparison.

Raw data acquisition

The image data obtained were imported into the software program ArrayVision 7.0 (Imaging Research, St. Catharines, Ontario, Canada) for spot detection and quantification of hybridization signals. For the macroarrays, local background calculated from empty spots in each secondary grid, were subtracted using ArrayVision 7.0 to obtain raw signal intensities, whereas local background calculated from the corners between spots for the unigene-microarrays.

Data Analysis

Raw data were exported from ArrayVision 7.0 (Imaging Research, St. Catharines, Ontario, Canada) into Excel. Duplicate spots at macroarrays were averaged. According to spiking controls, data of different macroarrays were normalized and converted to TIGR Array Viewer (TAV) format files, whereas Excel files from the unigene-microarrays were directly converted to TAV files.

Using the TIGR Microarray Data Analysis System (MIDAS) (Saeed et al. 2003), first signals were filtered to exclude low intensity signals, and then " lowess (locally weighted linear regression) normalization" was employed to adjust intensity-dependent effects in log2 (ratio) values. "Replicate consistency checking" removed poorly reproducible genes, and finally "slice analysis" was utilized to identify differentially expressed genes, which were induced/repressed more than 1.96 standard deviations from the local mean in each comparison (Quackenbush 2002).

Afterwards, the dataset of all inbreds (F7, D21, D32, FAP1360A, Pa405, D145, and D408) were imported in Multiexperiment Viewer (MeV) (Saeed et al. 2003). "Hierarchical cluster analysis" (Eisen et al. 1998) were conducted to discover similar

expression pattern, and then "between-subject t-tests with adjusted Bonferroni correction" (Pan 2002) were utilized to identify candidate genes differentially expressed between SCMV resistant and susceptible inbreds. Finally, the pair-wise correlation within the inbreds was obtained by "Gene distance matrix" (Saeed et al. 2003).

Sequence Analysis

Differentially expressed SSH clones were sequenced by MWG (Ebersberg, Munich, Germany). All sequences were compared with the EST database in Maize GDB (http://www.maizegdb.org) by BLASTN analysis with a threshold E value of 10^{-5}. Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). Each gene was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. Gene mapping information came from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/). If an EST was assigned to a mapped gene cluster, we assumed identical chromosome location of this EST and the gene cluster. The distribution of mapped genes was drawn by MapChart (Voorrips 2002).

Results

Identification of differentially expressed genes

After SCMV inoculation, mosaic symptoms were observed on each infected plant of susceptible lines (D145, D408, and F7) within two weeks, whereas infected plants of resistant lines ($F7^+$, D21, D32, FAP1360A, and Pa405) remained without symptoms after five weeks.

In order to capture a wide spectrum of differentially expressed genes, five SSH libraries were constructed (Table 1). The macroarrays involving clones from all SSH libraries were hybridized with cDNA preparations from non-infected F7 / $F7^+$ and infected F7 / $F7^+$ (Figure 2). Comparisons A and B included different genotypes with the same treatment. Comparison A: F7 Non-infected versus $F7^+$ Non-infected, which is constitutive genetic discrepancy. Structural and chemical barriers of the plant effectively exclude the majority of organisms. The genotypic difference between NILs might include constitutive resistance or susceptibility factors. B: F7 infected versus $F7^+$ infected, which is inducible genetic discrepancy. If constitutive defense of a plant is overcome, a sensitive surveillance system can detect foreign pathogens and trigger a rapid response to injury or virus attack. Genetic discrepancy after SCMV inoculation might include induced or repressed resistance factors between NILs. Comparison C and D include identical genotypes (F7 or $F7^+$) with different treatments. Comparison C: F7 infected versus F7 non-infected, (compatible interaction): virus replicates and moves systemically in cells of intact susceptible plants. Comparison D: $F7^+$ infected versus $F7^+$ non-infected, (incompatible interaction): virus multiplication is limited to initially infected cells of resistant plants.

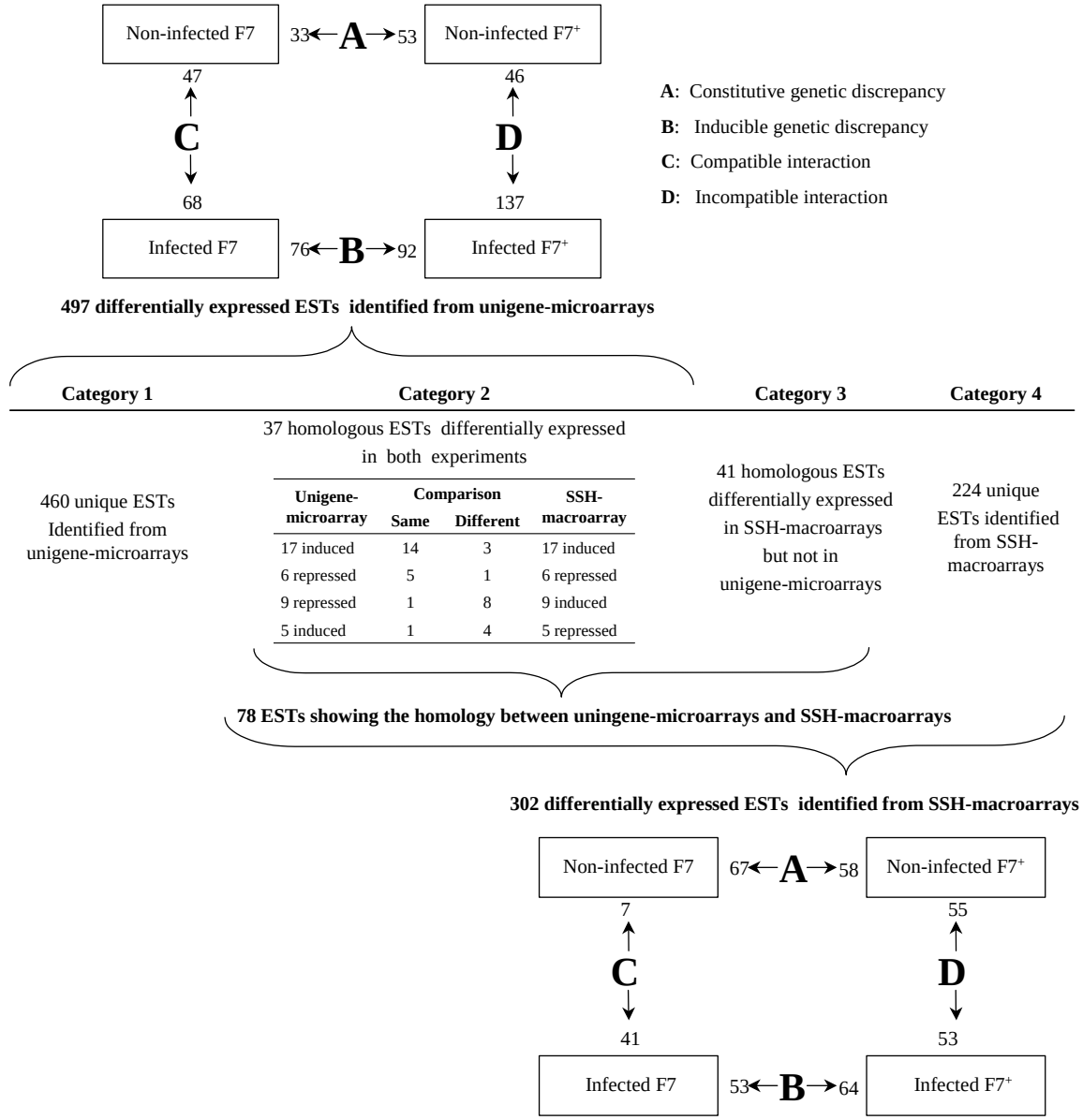


Figure 2. Result comparison of unigene-microarray and SSH-macroarray experiments.

In the approach of SSH-macroarrays, the number of at least 2-fold induced or 2-fold repressed genes, was 67 and 58 for Comparison A, 53 and 64 for Comparison B, 7 and 41 for Comparison C, as well as 53 and 55 for Comparison D, respectively (Figure 2). In addition, 24 differentially expressed genes were found when comparing different genotypes with the same treatment (Comparison A and B), and 12 were identified when

comparing identical genotypes (F7 or $F7^+$) with different treatments (Comparison C and D). RGAs *pic19* and *pic13* were not differentially expressed in all four comparisons.

The unigene-microarrrays were also hybridized with the probes from non-infected F7 / $F7^+$ and infected F7 / $F7^+$ (Figure 2). In total, 497 ESTs were differentially expressed in one or more comparisons, which accounted for 4.1% of 11, 827 ESTs deposited on the unigene-microarray. The number of at least 2-fold induced ESTs, was 33 for non-infected F7 and 53 for non-infected $F7^+$ in Comparison A, 76 for infected F7 and 92 for infected $F7^+$ in Comparison B, 47 for non-infected F7 and 68 for infected F7 in Comparison C, as well as 46 for non-infected $F7^+$ and 137 for infected $F7^+$ in Comparison D (Figure 2). In total 50.4% of these ESTs were induced more than 4-fold up to 25-fold.

We assigned to each differentially expressed SSH-EST to a Genbank accession number (GA) to identify respective microarray-ESTs based on BlastN hits (E-value$< 10^{-5}$) against the MaizeGDB EST database. The comparison of unigene-microarray and SSH-ESTs is summarized in Figure 2. 460 differentially expressed ESTs were exclusively present on microarrays (Category 1), and 224 on SSH-based macroarrays (Category 4). 78 differentially expressed ESTs were present both on micro- and macroarrays. A Bland-Altman plot (Bland and Altman 1986) revealed no significant difference between both experiments. Among those, 37 homologous ESTs were differentially expressed both in unigene-microarray and SSH-macroarray experiments (Category 2): 17 ESTs were induced in both approaches with 14 in the same comparison and 3 in different comparisons; 6 ESTs were repressed in both approaches with 5 in the same comparison and 1 in different comparisons; 9 ESTs were repressed in unigene-microarray but induced in SSH-macroarray experiments with 1 in the same and 8 in different comparisons; 5

ESTs were repressed in SSH-macroarray but induced in unigene-microarray experiments with 1 in the same and 4 in different comparisons. 41 homologous ESTs differentially expressed in SSH-macroarray were not differentially expressed in unigene-microarray experiments (Category 3). If all 78 ESTs in Category 2 and 3 are taken into account expression patterns of unigene-microarray and SSH-macroarray experiments from the same comparisons were consistent (Fisher's exact test: $P = 0.0117$).

Although more differentially expressed ESTs were identified based on microarrays (497) than SSH-macroarrays (302), the efficiency of gene discovery, determined as the ratio between the number of differentially expressed to all spotted cDNAs, was much higher in SSH-macroarray (59%) (Shi et al., submitted) than in unigene-microarray experiments (4.1%). However, due to five-time redundancy in SSH clones revealed by sequencing (Shi et al. submitted), the actual efficiency of gene discovery by SSH-macroarray experiments was approximately 10%.

Functional classification of differentially expressed genes

In total, more differentially expressed ESTs from the unigene-microarray experiment (72%) were unclassified than in the SSH-macroarray experiment (44%). Among classified ESTs, the largest category was “metabolism” (25.3%) in unigene-microarray experiments (Figure 3), and “cell rescue, defense, cell death and ageing” (32.0%) in SSH-macroarray experiments. Further ranking of classification categories was “cell rescue, defense, cell death and ageing” (19.0%), “signal transduction” (15.5%) and “transcription” (11.3%) for unigene-microarray experiments, and “transcription” (17.4%), “metabolism” (16.3%) and “signal transduction” (11.1%) for SSH-macroarray experiments. In spite of different ranks between both experiments, the top four categories

were the same, and three of them ("cell rescue, defense, cell death and ageing", "signal transduction", and "transcription"), are pathogenesis-related. In contrast to 60.5% in SSH-macroarray experiments, 45.8% of differentially expressed ESTs in unigene-microarray experiments were classified into pathogenesis-related categories.

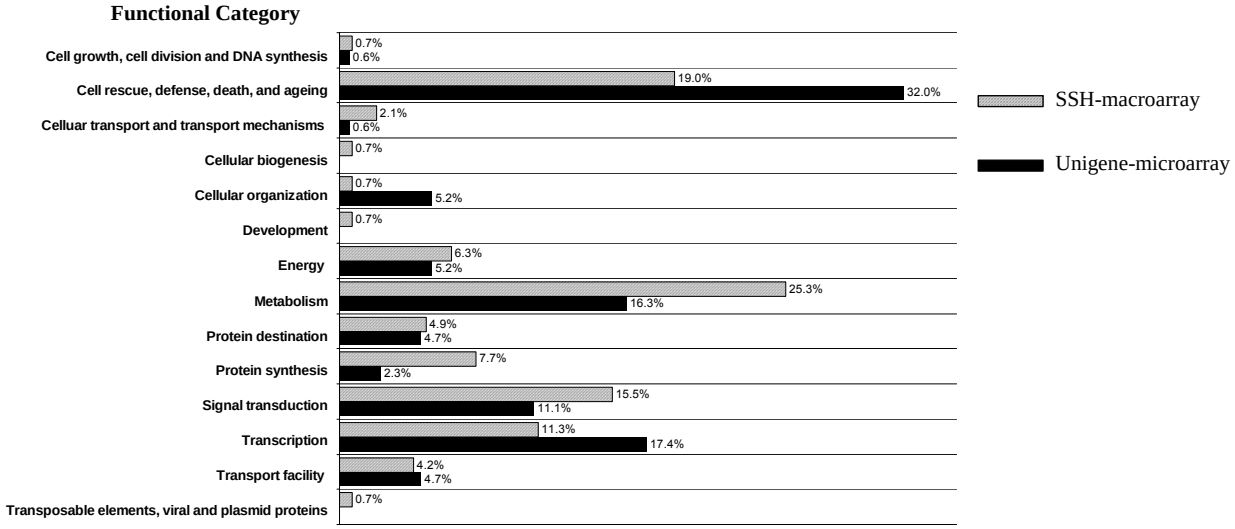


Figure 3. Comparison of gene distribution in functional classes in unigene-microarray and SSH-macroarray experiments. Unclassified microarray-ESTs (72%) and SSH-ESTs (42%) didn't been taken into account.

Map position of differentially expressed ESTs

Altogether 20% (100 of 497) of ESTs identified from unigene-microarray experiments have been previously assigned to 51 bins (Gardiner et al., 1993) distributed over all 10 maize chromosomes (Figure 4), whereas the same proportion of ESTs (in total 60) identified from SSH-macroarray were assigned to fewer genome regions (29 bins). ESTs were randomly distributed to chromosomes in both unigene-microarray (P = 0.3979, χ^2 = 9.438, df = 9) and SSH-macroarray experiment (P = 0.1806, χ^2 = 12.62, df = 9).

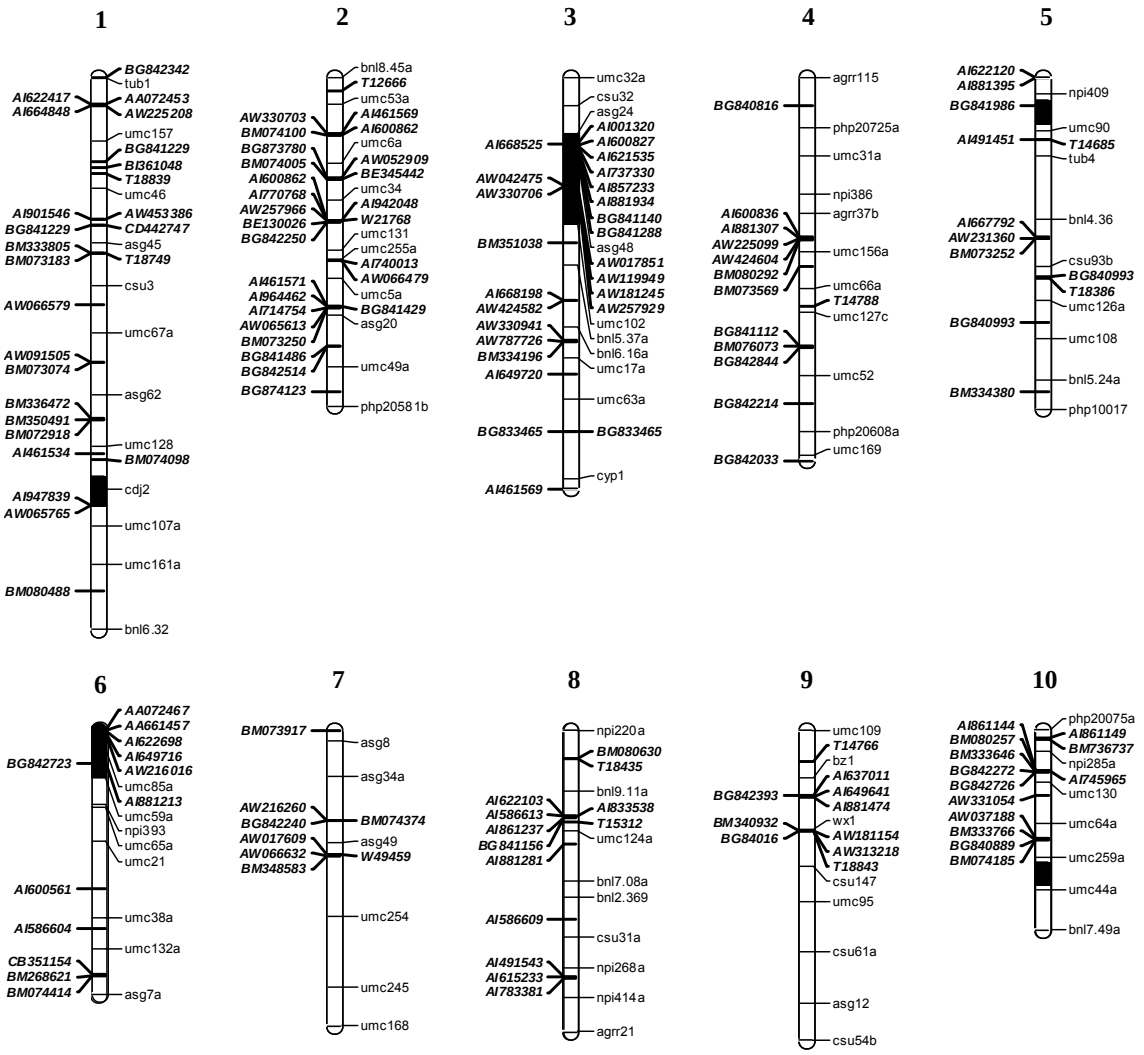


Figure 4. Distribution of differentially expressed genes on maize chromosomes with respect to SCMV resistance identified in unigene-microarray and SSH-macroarray experiments. Loci in *bold* and *italics* placed on the *left* side of each chromosome were mapped ESTs identified from unigene-microarray experiments. Loci placed on the *right* side of each chromosome are a set of core markers that defines a bin boundary (Gardiner *et al*., 1993), while the loci in *bold* and *italics* were mapped ESTs identified in SSH-macroarray experiments. EST mapping information was from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins. *Scmv1* is

highlighted on chromosome 6, *Scmv2* on chromosome 3 and three minor *QSCMs* on chromosomes 1, 5, and 10 (Xia *et al.*, 1999).

In contrast to the 30% (18 of 60) of mapped ESTs located in bin 3.04-3.05 (12) and bin 6.00-6.02 (6) in SSH-macroarray experiments, only 4% (4 of 100) of the mapped ESTs from unigene-microarray experiments in bin 3.04-3.05 (3) and bin 6.00-6.02 (1) (Table 2). The proportion of ESTs mapped in vicinity of *Scmv* QTLs was significantly higher (P = 0.0013) in SSH-macroarray than in unigene-microarray experiments. Among the ESTs mapped in bins 3.04-3.05 and 6.00-6.02, no homology was found between microarray- and SSH-ESTs. None of these microarray-ESTs, but at least 6 SSH-ESTs, showed homology to defense-related genes, such as, AI881934 (17.5 kDa class II heat shock protein), AI649716 (Cytochrome P450 71C2). Finally, AI947839, AW065765 and BG841986 identified from unigene-microarray experiments were co-localized with QSCM on chromosomes 1, 1, and 5 (Figure 4).

Association between gene expression patterns and 'SCMV resistance'

302 differentially expressed genes between NILs $F7^{+}$ and F7 were further investigated among four resistant (FAP1360A, D21, D32, and Pa405) and three susceptible inbreds (D145, D408, and F7). Generally, more genes were differentially expressed in incompatible reactions than in compatible reactions (Figure 5). Among incompatible reactions, the greatest overall response was observed in FAP1360A (177 (112 up regulated, 65 down regulated)). Slightly fewer genes were identified in D21 (163 (87 up, 76 down)) and D32 (165 (92 up, 73 down)), and only one third of genes in Pa405 (62 (47 up, 15 down)). Among compatible reactions, the overall response was highest in

D408 (93 (57 up, 36 down)), whereas substantially fewer genes showed differential expression in F7 (47 (40 up, 7 down)) and D145 (37 (15 up, 22 down)). The number of genes differentially expressed for pairs of resistant inbreds were 153 (between FAP1360A and D21), 36 (between D21 and Pa405), 38 (between Pa405 and D32) and 155 (between D32 and FAP1360A), respectively. In contrast, the commonly differentially expressed genes between pairs of susceptible inbreds were 10 (between F7 and D145), 9 (between D145 and D408) and 24 (between D408 and F7), respectively. 34 genes were differentially expressed in all resistant inbreds, as compared to only 3 genes among all susceptible inbreds. The proportion of commonly differentially expressed genes was significantly higher ($P = 0.0203$) among incompatible compared to compatible reactions.

Based on hierarchical cluster analysis of expression patterns (Eisen et al. 1998). The inbreds were divided into two groups: D32, D21, FAP1360A, and D408 formed one group, Pa405, D145 and F7 the second group (Figure 6A). Using t-tests with adjusted Bonferroni correction, the expression patterns of 22 genes were significantly ($P < 0.05$) different between resistant (D21, D32, FAP1360A, and Pa405) and susceptible inbreds (D408, D145, and F7) (Figure 6A, cluster A). According to marker-based haplotype analysis for the *Scmv1* and *Scmv2* regions (Xu et al. 2000), the origin of the resistant U.S. inbred Pa405 was largely different from the resistant European inbreds (D21, D32, and FAP1360A). A t-test without Pa405 revealed a substantially higher number of genes (112) with group-specific expression patterns for resistant (D21, D32, and FAP1360A) versus susceptible inbreds (D408, D145, and F7) (Figure 6A, cluster B).

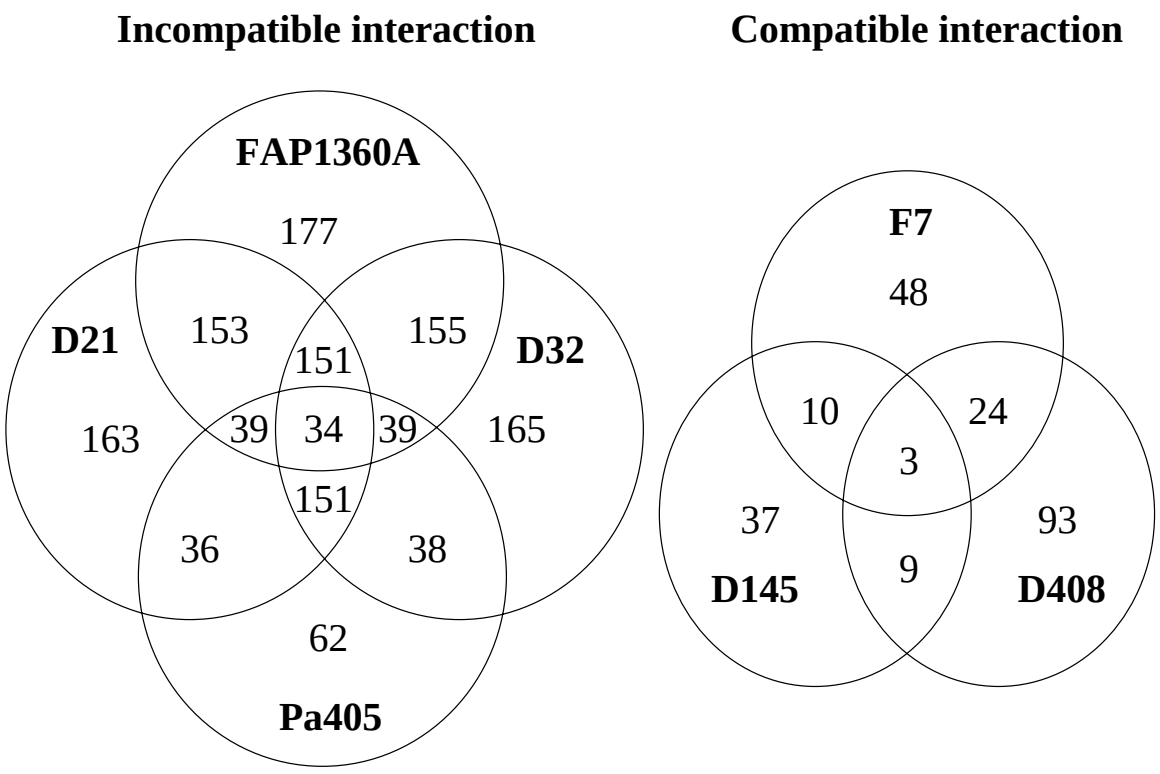


Figure 5. Venn diagrams for comparison of the numbers of differentially expressed genes within resistant or susceptible inbreds. Total numbers of genes differentially expressed in individual inbreds (FAP1360A, D21, D32, Pa405, F7, D145, or D408) are indicated in respective circles. In parentheses, the first number indicates up-regulated genes, whereas the last number in *italics* down-regulated genes. (a) Intersection of genes identified in resistant inbreds (incompatible interaction), (b) Intersection of genes identified in susceptible inbreds (compatible interaction).

Pair-wise correlations (Figure 6B) were obtained by GDM (gene distance matrix) (Saeed et al. 2003). Maximum similarity (scaled distance = 1.0) was found between FAP1360A and D408. Expression patterns of all inbreds are similar (scaled distance ≥ 0.48). The expression pattern of D408 (average scaled distance = 0.84) showed the highest correlation with the other six inbreds: Pa405 (0.77), FAP1360A (0.70), D32

(0.69), F7 (0.69), D21 (0.67), and D145 (0.58). The average similarity within the resistant group (D21, D32, FAP1360A, and Pa405) (average scaled distance = 0.69) was higher than between susceptible inbreds (D408, D145, and F7) (Average scaled distance = 0.65).

Figure 6. Cluster analysis based on expression patterns of SCMV resistant and susceptible lines with and without SCMV infection. (A) Hierarchical cluster analysis (Eisen et al. 1998) of differential gene expression with and without SCMV infection for SCMV resistant (D32, D21, FAP1360A, and Pa405) and susceptible (D408, D145, and F7) inbreds. The color saturation reflects the magnitude of the log2 expression ratio (Cy5/Cy3) for each transcript. Red color means higher transcript levels than the reference, whereas green means lower transcript levels than the reference. The color log2 scale is provided at the bottom of the figure. The vertical bars on the left and right side of the tree indicate cluster A including 22 genes and Cluster B including 112 genes, respectively. (B) A table of pair-wise correlations among the samples shown in (A).

B

	D21	**D32**	**FAP1360A**	**Pa405**	**D408**	**D145**
D32	0.56					
FAP1360A	0.53	0.55				
Pa405	0.83	0.81	0.83			
D408	0.88	0.94	1.00	0.85		
D145	0.48	0.51	0.48	0.68	0.75	
F7	0.75	0.78	0.81	0.62	0.61	0.58

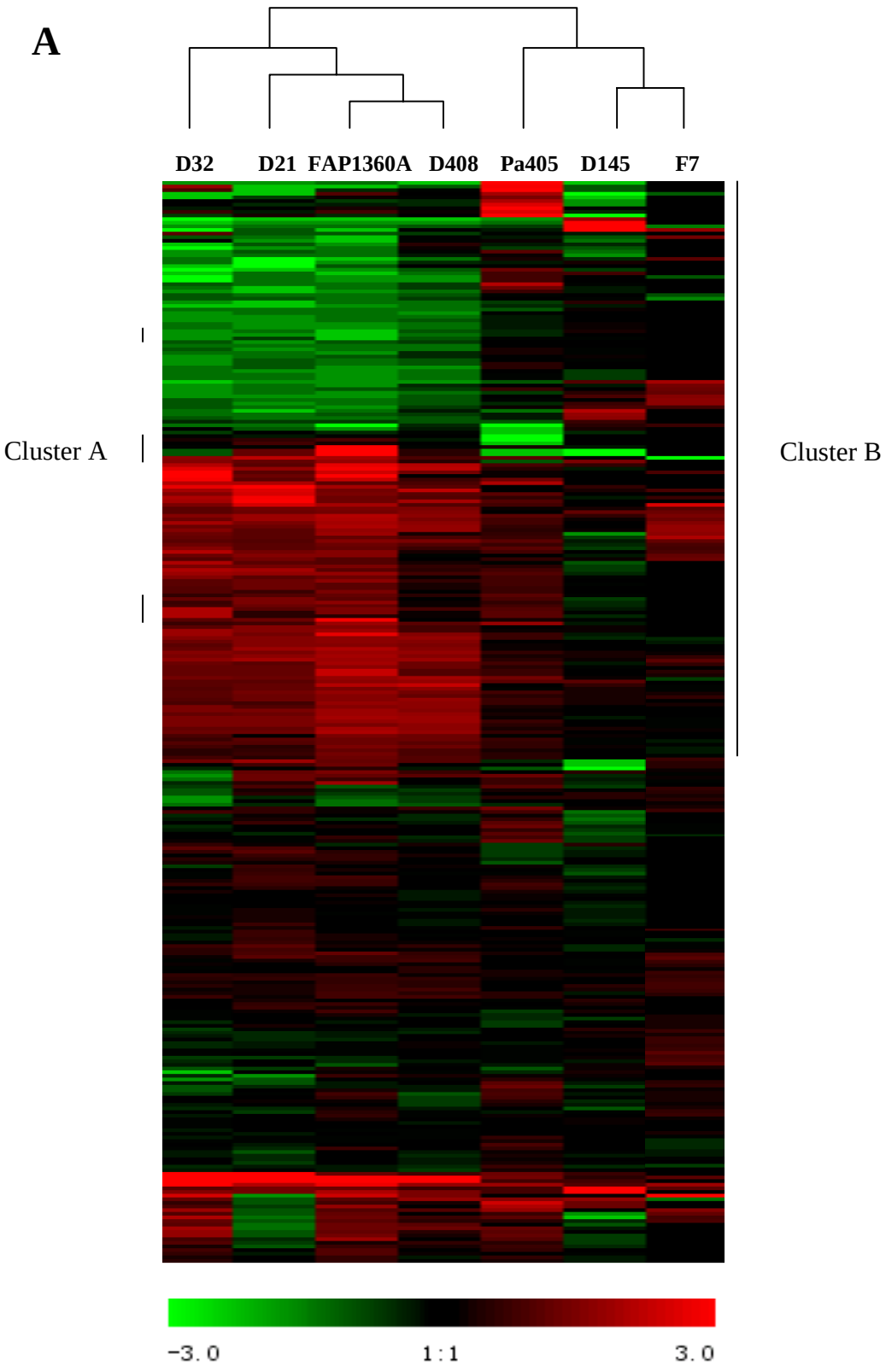

A
D32
D21
FAP1360A
D408
Pa405
D145
F7
Cluster A
Cluster B
−3.0
1:1
3.0

Table 2. Mapped ESTs co-localized with major *Scmv* QTLs

GA[a]	Annotation[b]	Similarity[b]	Ratio[c]	Bin[d]	Microarray				SSH+Macroarray			
					A[e]	B	C	D	A	B	C	D
***Scmv 1* region**												
AA661457	gb|AAK58690.1|receptor-like kinase Xa21-binding protein 3 {Oryza sativa}	24%	3.3, 5.0	6.01					+[f]	+		
AA072467	PIR|T06273|benzothiadiazole-induced protein {Triticum aestivum}	10%	2.5, 5.0	6.01					+	+		
AI622698	GP|4514655|dbj|BAA75493.1||AB024058 IDS3 {Hordeum vulgare}	43%	2.3, 2.1	6.01							+	+
AI649716	SP|Q43255|Cytochrome P450 71C2{Zea mays}	100%	2.1	6.01							+	
AW216016	dbj|BAB92879.2|{Oryza sativa}	16%	-2.5	6.01								+
AI881213	SP|Q9NVW2|RING finger protein 12 {Homo sapiens}	5%	-2.5	6.02						+		
BG842723	GP|17385700|dbj|BAB78651.1||AP003022 P0681B11.18 {Oryza sativa}	7%	5.8, 2.7	6.02	+			+				
***Scmv 2* region**												
AI881934	SP|P24631|17.5 kDa class II heat shock protein {Zea mays}	100%	8.7	3.04								+
AI600827	dbj|BAB44108.1|{Oryza sativa}	71%	6.0	3.04								+
AI857233	dbj|BAB21196.1|{Oryza sativa}	26%	3.7	3.04						+		
BG841140	emb|CAB40376.1|adenosine kinase {Zea mays}	50%	2.5	3.04						+		
AI621535	gb|AAC98962.1|nucleic acid binding protein {Oryza sativa}	61%	2.4	3.04						+		
AI737330	EGAD|65434|troponin I (TNI) (wings apart-a protein) {Drosophila melanogaster}	7%	2.1	3.04						+		
BG841288	dbj|BAB62599.1|{Oryza sativa}	98%	-2.0	3.04						+		
AI668525	GP|7299996|gb|AAF55168.1||AE003708 CG4913-PA {Drosophila melanogaster}	4%	-3.0	3.04			+					
AI001320	Unknown		-10.0	3.04								+
AW181245	dbj|BAB13743.1|pseudo-response regulator 4 {Arabidopsis thaliana}	4%	8.5	3.05						+		
AW257929	gb|AAL62060.1|RAD21-3 {Arabidopsis thaliana}	7%	7.2	3.05								+
AW017851	emb|CAA11391.1| phytase {Zea mays}	44%	5.5	3.05								+
AW330706	Unknown		2.8	3.05				+				
AW119949	dbj|BAB90212.1|{Oryza sativa}	95%	2.1	3.05						+		
AW042475	GP|15528797|dbj|BAB64692.1||AP003683 P0431G06.3 {Oryza sativa}	10%	-4.5	3.05			+					
Sum					1	0	2	2	2	10	2	7

[a] Genbank accession number. [b] Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). [c] If the ratio is less than one, the negative reciprocal is listed. [d] Mapping information is from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins (Gardiner *et al*., 1993). [e] Four experiments on differential gene expression were conducted: Comparison A (constitutive genetic discrepancy), Comparison B (inducible genetic discrepancy), Comparison C (compatible interaction) and Comparison D (incompatible interaction). [f] The mapped EST was identified from this comparison.

Discussion

Technical comparison between SSH-macroarray and unigene-microarray

Recently, microarrays were widely recognized as a significant technological advance providing genome-scale information on gene expression patterns (Richmond and Somerville 2000). Complete transcriptome arrays are allowed to assay traits without preconceived ideas. Although a comprehensive microarray is not available in maize yet, the microarrays used in this study, contained 9,841 different unigenes, accounting for only 20% of the about 50,000 maize genes (Martienssen et al. 2004). In contrast to the unigene-microarray, the macroarray used in a companion study contained only a limited number of SSH clones specifically developed for studying SCMV resistance. One major limitation of SSH and similar methods is the difficulty to cover multiple comparisons when comparing a series of RNA samples, since SSH libraries are produced from pairwise comparisons (Donson et al. 2002). In our study, only five SSH libraries were constructed instead of twelve covering all combinations (there are only six comparisons, but there are 12 possibilities for subtraction) between four RNA samples (infected F7, infected $F7^{+}$, not-infected F7 and not-infected $F7^{+}$). This might be one explanation for 460 differentially expressed ESTs exclusively present on microarrays, which have not been recovered by the SSH-macroarray procedure (Figure 3).

Only 8.8% (1045 of 11827) of the EST collection in the unigene-microarray was derived from two stress-induced EST libraries, including 496 (stressed shoot) and 603 (stressed root), whereas 91.2% (10771 of 11827) from fifteen EST libraries made from plants grown under normal environmental conditions. The number of differentially expressed genes discovered from library 496 (9%) was substantially higher than from

other EST libraries (average: 4%). Since more ESTs differentially expressed in SSH-macroarray experiments (60.5%) were classified into pathogenesis-related categories than in unigene-microarray experiments (45.8%), the SSH cDNAs complemented the ESTs printed on microarrays. In addition, SSH-macroarray procedure enriches for low-abundant and differentially expressed mRNAs by normalization (Diatchenko et al. 1996). The normalization step is particularly important because abundant pathogenesis-related transcripts (e.g., genes coding for PR proteins) very likely mask important SCMV-specific transcripts expressed at much lower levels.

Regardless of the procedure, the reproducibility was high both in unigene-microarray and SSH-macroarray experiments. Fisher's exact test ($P = 0.0117$) showed consistent expression patterns of microarray-ESTs and SSH-ESTs from the same comparison (Figure 3). However, the degree of consistency was limited, such as AI691482 was induced in unigene-microarray experiment but repressed in SSH-macroarray experiment. Discrepancies between both approaches can be explained by i) different targets spotted on the arrays; ii) different labeling procedures; iii) different ratio measurements. Because different cDNAs of different length, different parts of the genes were deposited on micro- or macroarrays, it might mask changes in transcript levels due to cross-hybridization to gene family members (Girke et al. 2000). Furthermore, probe labeling was different. The same total RNA samples were indirectly labeled with fluorescent dyes Cy3 or Cy5 using random primers in the unigene-microarray experiment, whereas direct labeling with radioactive P^{32} using oligo(dT) was employed in the SSH-macroarray experiment. In another study comparing array-based results with northern blots, arrays were less sensitive in measuring a subset of the genes (Taniguchi et

al. 2001). Moreover, different ratio measurements were implemented. In unigene-microarray experiments two cDNA samples were hybridized on the same glass slide in parallel, allowing the direct measurement of ratios on the unigene-microarray. In contrast, internal controls were included on nylon membranes employed in SSH–macroarray experiments. Only one probe was hybridized per membrane and ratios were obtained by indirect comparisons. Among 21 homologous ESTs from the same comparison in Category 2, the correlation ($r = 0.88$) of the expression level between two approaches was highly significant by Pearson correlation calculations ($P < 0.0001$). The finding of 41 ESTs in Category 3 corroborates a higher risk of smaller fragments cross-hybridizing with other gene family members (Finkelstein et al. 2002).

Molecular mechanisms of maize-SCMV interaction

Typical mosaic symptoms have been observed in leaves of susceptible F7 after systemic movement and replication of SCMV. Thus, the F7 - SCMV interaction is a compatible interaction (Comparison C) (Figure 7). In compatible interactions, the distribution of genes among functional classes looked similar to the incompatible reaction, regardless of unigene-microarray or SSH-macroarray. It is consistent with the hypothesis that viruses induce defense response both in susceptible and resistant plants at early stages (Matthews and Hull 2002). Whitham et al. (Whitham et al. 2003) reported that diverse RNA viruses, including cucumber mosaic cucumovirus, oil seed rape tobamovirus, turnip vein clearing tobamovirus, potato virus X potexvirus and turnip mosaic potyvirus, elicited the expression of common sets of genes in susceptible *Arabidopsis*. Totally 62% of the differentially expressed genes in Comparison C, 72% (33 of 47) for SSH-macroarray (Shi et al, submitted) and 58% (67 of 115) for unigene-

microarray, can be found in this common set of genes, while the remaining genes without annotation could be maize-specific.

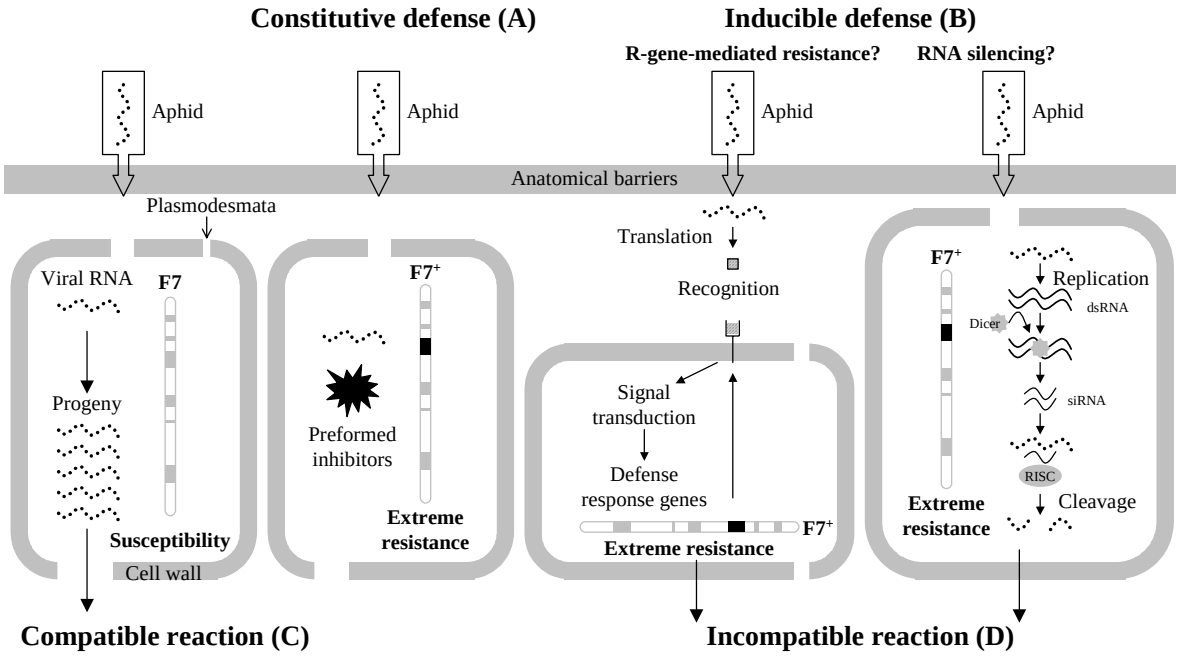


Figure 7. Diagrammatic view of SCMV-maize interaction

SCMV enters maize symplasm by non-persistent transmission through aphids. In susceptible F7, SCMV rapidly replicates and spreads from cell to cell through plasmodesmata. Comparison C revealed differentially expressed ESTs in this compatible interaction. In contrast, extreme resistance was detected in resistant $F7^{+}$. Preformed inhibitors in the cells play a major role in constitutive defense, revealed in Comparison A, while inducible defense (Comparison B) can be found together with R-gene-mediated resistance or RNA silencing. Comparison D revealed differentially expressed ESTs in this incompatible interaction. The models of R-gene-mediated resistance and RNA silencing were adapted from Lucas and Dickinson (1998), Waterhouse and Helliwell (2003), respectively.

Plants of NIL $F7^+$ displayed no SCMV symptoms in the infection trial, thus $F7^+$ is completely resistant to SCMV and the $F7^+$ - SCMV interaction is an incompatible interaction (Comparison D) (Figure 7). Gene expression profiles of incompatible reactions, including TMV in tomato and *Chenopodium* (Cooper 2001, Golem and Culver 2003), revealed similarities at the gene level with Comparison D. It corroborates most likely resistant plants utilize a common mechanism for defense against virus attack (Matthews and Hull 2002).

Several putative preformed inhibitors, which could be involved in constitutive defense, were revealed in Comparison A (Figure 7). Such as, AW011679 show homology to genes encoding UMP/CMP kinase, and BM335333 is the homolog of an ankyrin-kinase. It corroborates previous finding that SCMV can be detected and, thus, replicates in primary infected leaves of resistant genotypes (Louie 1995).

So far, two types of inducible defence are defined: hypersensitivity response (HR) and extreme response (ER) (Matthews and Hull 2002). HR limits virus infection to a zone of cells around the initially infected cell of the resistant host, usually with the formation of visible necrotic local lesions (Matthews and Hull 2002). ER limits virus multiplication to initially infected cells because of an ineffective virus-coded movement protein, giving rise to latent infection. No HR symptoms are observed for maize leaves infected with SCMV, thus maize resistance to SCMV might be extreme resistance. Further experiments conducted at single-cell level, usually in protoplast, are warranted (Matthews and Hull 2002).

ER is most often triggered by dominant or semi-dominant resistance (R) genes and occurring in a strain-specific or “gene-for-gene” manner (Figure 7) (Matthews and

Hull 2002). In potato, two extreme resistance genes (*Rx1* and *Rx2*) to PVX have been cloned, which belong to the nucleotide binding, leucine-rich repeat (NBS-LRR) super-family of R-genes (Bendahmane et al. 1999, Bendahmane et al. 2000). In addition, ER might be triggered by RNA silencing (Figure 7). In contrast to resistance triggered by the NBS-LRR–type R genes, resistance through silencing appears not to depend on a gene-for-gene recognition event (Whitham et al. 2000). Differentially expressed pathogen-related genes identified from Comparison B, such as AI664862 (stress-induced protein OZI1 precursor) and AI795699 (peroxidase), have been found together with R-gene-mediated resistance or RNA silencing. So far, little is known about the genes involved in signal transduction of HR and ER, it is even possible that they use the same genes for signaling. Therefore, both mechanisms might be involved in SCMV resistance.

Association between SCMV resistance and expression-level polymorphisms

Analysis of expression profiling data across a collection of lines well characterized with respect to SCMV resistance has the potential to associate ELPs with our trait of interest. This is comparable to association studies at the levels of DNA polymorphisms, were also no experimental populations such as segregating populations are required. It also avoids the limitation of positional cloning by NILs. This is due to the difficulty of developing NILs for loci that explain less than 20% of the variance and to constraints created by only using two alleles. Recent work (Potokina et al. 2004) confirmed that the variation of the complex trait “malting quality” in a set of 10 barley genotypes was reflected at the RNA level by using a cDNA array with 1400 ESTs. Between 17 and 30 candidate genes were identified for each of the six malting parameters analyzed.

Of 112 genes differentially expressed between European resistant (FAP1360A, D21, and D32) and susceptible lines (D405, D148 and F7), 42, 40, 19 and 58 genes were common with the genes identified from constitutive genetic discrepancy (125), inducible genetic discrepancy (117), compatible interaction (48) and incompatible interaction (108) between NILs $F7^{+}$ and F7, respectively. It shows at least 40% genes (In case of compatible interaction: 39.6%) identified from the NILs can be used to characterize the expression pattern of European inbred lines from the same comparison, especially for incompatible reaction (53.7%).

Cluster analysis based on expression patterns of SCMV resistant and susceptible lines with and without SCMV infection confirmed the close relationship of the European resistant lines (D21, D32, and FAP1360A) and separated them from the U.S. line Pa405. It is consistent with pedigree records (Kuntze et al. 1995) and shows that chromosome segments in common among the three European resistant genotypes are leading to more similar expression patterns. The strong reduction of commonly differentially expressed candidate genes from 112 to 22 after including Pa405 indicates the presence of different resistance genes in the three resistant European lines compared to Pa405. Both results corroborate the conclusion of haplotype analysis that Pa405 had unique haplotypes both in the *Scmv1* and *Scmv2* region comparing to European resistant lines (Xu et al. 2000). The standard Bonferroni correction is very stringent and may exclude many genes that are really significant, whereas the adjusted Bonferroni correction is less conservative, and more likely to include significant genes while still controlling the error rate (Pan 2002). This method should provide a reasonable balance between false-positive and false-negative rates for our analysis.

After inclusion of susceptible lines D408, D145, and F7, the seven inbreds were divided into two groups based on expression profiling data: D32, D21, FAP1360A, and D408, and Pa405, D145, and F7 (Figure 3A). As described before (Melchinger et al. 1998), D32, D21, FAP1360A, and D408 are early-maturing European Dent inbreds, whereas D145 and F7 are Flint lines. Pa405 is a Dent line, but from a different pool as the European Dent lines. Thus the grouping of expression patterns among the inbreds was more according to Dent-Flint than resistant-susceptible. The presence of population stratification can result in nonfunctional, spurious associations (Flint-Garcia et al. 2003), as described for studies on association mapping of maize. It would, therefore, be more meaningful to compare resistant Dent with susceptible Dent lines to avoid confounding with heterotic grouping. However, an expansion of this study is difficult due to the low number of available SCMV resistant inbred lines (three) identified in a large collection of European inbreds (Kuntze et al. 1995). Therefore, the most interesting genes are those differentially expressed according to resistance / susceptibility, including the 22 genes shared between the three resistant European inbreds and Pa405, but we cannot rule out especially for the 90 remaining genes only shared between the three resistant European inbreds but not Pa405, that they include some genes showing up due to the dent – flint grouping. Among those 90 genes, those ones not shared with D408 might more likely be associated with SCMV resistance than those shared with D408. Because the expression pattern of D408 was closer to the resistant lines than F7 and D145, it supposes *Scmv2* region might be already present in D408. Previous field trials also showed fewer susceptible plants were found in F2 populations with D408 as compared to F7 and D145 (Melchinger et al. 1998).

Candidate gene (CG) selection

The CG approach consists of three subsequent steps: the choice, screening and validation of CGs (22). In this study, good candidates associated with SCMV resistance can be chosen from at least three classes: i) positional CGs mapping to bins 3.04 – 3.05 and 6.00 – 6.02, ii) functional CGs showing the homology to pathogenesis-related genes, or iii) the ESTs in Category 2 showing consistent expression pattern in both approaches.

So far, 18.6% (696 of 3737) of all mapped maize ESTs are located in bins 3.04 – 3.05 (426) and 6.00 – 6.02 (270) (http://www.plantgenomics.iastate.edu/maizechip/). In contrast to the 30% (18 of 60) of the mapped ESTs from SSH-macroarray located in bin 3.04-3.05 (12) and bin 6.00-6.02 (6), only 4% (4 of 100) of the mapped microarray-ESTs were located in bins 3.04-3.05 (3) and bin 6.00-6.02 (1) (Table 2). While no homology was found between SSH-ESTs (18) and microarray-ESTs (4), 50% (9 of 18) mapped SSH-ESTs belong to Category 3 (Figure 2). One possible explanation is that *Scmv*-specific ESTs were under represented in genome-wide unigene-microarray, whereas SSH libraries enriched them after normalization step. Identification of a larger number of differentially expressed genes mapping to these two regions can be explained by (i) genes differentially expressed due to the polymorphism between F7 and $F7^{+}$ in these two regions but without relation to SCMV resistance, or (ii) clustering of genes involved in SCMV resistance in these two regions. Except for positional CGs, many CGs (pathogenesis-related genes) were revealed from the 80% non-mapped ESTs, such as AI621822 (Avr9 elicitor response protein), AI999974 (maize catalase isozyme 3).

Comparing to other ESTs identified from either unigene-microarray or SSH-macorarray experiments, 19 consistent ESTs in Category 2 (Figure 2) are the most

promising candidates for being differentially expressed in the context of SCMV resistance. However, due to uncompleted annotation, four ESTs (AI665633, AI855243, AW330660 and AI974914) have no tentative annotation from TIGR Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml), and the annotation of six genes (AI649641, AI861142, AI942105, AW257966, AI714860 and AW438364) was based on proteins from the *Arabidopsis* or rice genome. Of all nine remaining genes, AI461569, AI621758, AI942048, AW052909 and AI941971 were related to RNA binding, while AI738263, BG842726, BG840993 and BM334179 were homologous of catalytic proteins

Evaluation of transcript profiling data

In total, 762 distinctly differentially expressed genes associated with SCMV resistance in maize were identified by both SSH-macroarray and unigene-microarray. An urgent question, we should immediately confront with, is how to validate the large data sets that are generated. Chuaqui et al. (2002) divided the validation process into three areas: experimental quality control, independent confirmation of data and universality of results.

Great care has been taken when we conducted all experiments. Four replications were utilized through all comparisons, including two biological replicates and two technical replicates in each biological replicate (In case of unigene-microarray: dye swap replication). In addition, the clones were spotted on the SSH-macroarray in duplicate patterns. Internal controls were also included in SSH-macroaray, containing ten spiking controls and negative controls. Our results showed the reproducibility of the experiments was high across all comparisons within both SSH-macroarrays and unigene-microarrays.

There are two approaches for the confirmation of array data: *in silico* analysis and laboratory-based analysis (Chuaqui et al. 2002). The *in silico* method compares array results with information available in the literature and in public or private expression databases, and provides the opportunity to validate data without further experimentation. A handful of genes in our candidate gene list are pathogenesis-related genes, which validates the general performance of our system and provides confidence in the overall data. Laboratory-based validation of data provides independent, experimental verification of gene-expression levels, and typically begins with the same samples that were studied in the initial array experiments. The methodology used varies depending on the scientific question, but commonly used techniques include semi-quantitative reverse transcription PCR (RT–PCR), real time RT–PCR and northern blot. Before choosing either method to validate our CGs, the following questions need to be considered. First, precisely how well do array results correlate with measurements using other techniques? Second, what is the cost and effort involved in carrying out follow-up studies on a large scale? Third, are there differences in sensitivity among the methods? So far, no accepted standards are existed in array community that enabling meaningful comparison of array data between different research groups. Therefore, the development of uniform validation methods and a more complete understanding of how to compare and contrast results derived by different techniques will be important for the future of microarray technology (Chuaqui et al. 2002).

Once array data have been analyzed and independently verified, investigators must determine whether the expression profiles are a universal feature of the biological phenomenon under study—in other words, are the data an essential descriptor of the

biological state? The CGs isolated by SSH were employed to investigate their association with SCMV resistance across seven resistant or susceptible inbred lines. Due to the genetic structure among the seven inbred lines, genetic background and resistance response are confounded. With or without the resistant U.S. inbred line Pa405, 7.3% (22 of 302) and 37.1% (112 of 302) genes showed different expression patterns between resistant (D21, D32, and FAP1360A) and susceptible (D145, D408, and F7) inbred lines by *t* tests, respectively. Due to SCMV resistance is complex trait, the next step will correlate the CGs with multi-environmental factors, such as, temperature and plant development stages. 'Tissue lysate array', where lysates from cell populations collected by laser capture microdissection (LCM) are arrayed on a nyloncoated slide, could be chosen as experimental platform to conduct such experiments (Schnable et al. 2004).

Future perspectives

Recently, plant breeding has been transformed from selection of phenotypes toward selection of genes, either directly or indirectly. Plant breeders try to optimize the use of the genetic variation in nature by bringing good alleles together that maximize yield, resistance to stress, etc. This strong demands abundant knowledge of genes with assigned functions. This thesis shows high-throughput assay, expression profiling, can be used as the starting point to identify candidate genes on a large scale. Further experiments, such as TILLING (targeting-induced local lesions in genomes) (McCallum et al. 2000), or RNA interference (RNAi) (Waterhouse and Helliwell 2003), can be used to determine gene function.

TILLING is emerging as a standard reverse-genetic strategy for plant functional genomics. In this strategy, chemical mutagenesis is followed by screening for point

mutations in pooled DNAs (McCallum et al. 2000). Chemical mutagens, such as ethylmethanesulfonate (EMS), can produce a relatively high density of irreversible mutations and allow for saturation mutagenesis to be achieved using relatively few individuals. Therefore, TILLING can provide allelic series of mutations, including knock-outs (Henikoff and Comai 2003). In contrast to insertional mutagenesis, the high density of chemically induced point mutations makes TILLING suitable for targeting the genes underlying the complex traits, such as, SCMV resistance. In addition, TILLING is completely general, as chemical mutagenesis has been successfully applied to most major taxa. Currently, this method is being established in crop plants (http://www.evry.inra.fr/public/projects/tilling/tilling.html).

In plant functional genomics, RNAi is another hot method to identify the gene functions in the growth and development of the plants. The essence of RNA interference is the delivery of double-stranded RNA (dsRNA) into an organism, or cell, to induce a sequence-specific RNA degradation mechanism that effectively silences a targeted gene (Waterhouse and Helliwell 2003). This approach can circumvent several limitations of insertional mutagenesis. For example, insertional mutagenesis cannot be used to investigate the functions of duplicated genes, and many mutant phenotypes in these lines are caused by disruptions to genes other than those into which the DNA tag is inserted. An important aspect of using RNAi in plant genomics research is the delivery of the silence-inducing dsRNA. This RNA can be delivered by stably transforming plants with transgenes that encode dsRNA. It can also be transiently delivered by bombarding plants with nucleic-acid-coated beads, by infiltrating plant cells with transgene-carrying *Agrobacterium tumefaciens* or by infecting plants with a virus, either on its own or

together with a satellite virus (Waterhouse and Helliwell 2003). SCMV belongs to potyvirus, the largest plant virus genus (Provvidenti and Hampton 1992), thus building up virus-induced gene silencing system in crop is promising using SCMV as the vector.

Once genes responsible for quantitative variation of SCMV resistance become available, information can be passed on to plant breeders in form of functional markers, which are derived from polymorphic sites within genes causally affecting phenotypic trait variation (Andersen and Lubberstedt 2003). Functional markers are superior to random DNA markers such as RFLPs, SSRs and AFLPs owing to complete linkage with trait locus alleles. More general, it allows reliable application of markers in populations without prior mapping, the use of markers in mapped populations without risk of information loss owing to recombination and better representation of genetic variation in natural or breeding populations. Recently, projects on structural and functional genomics have been established in maize (http://www.maizegenetics.net/). The knowledge generated from this project will help systematic development of functional markers for SCMV resistance. Due to polygenic trait of SCMV resistance, marker-assisted selection (MAS) programs with functional markers would significantly increase breeding efficiency.

At present, SCMV also causes substantial yield losses in susceptible maize cultivars in China. In contrast to developed countries, the majority of Chinese farmers still produces most of their own food and depends on small-scale farming for their incomes and livelihoods. In the future, the genes underlying SCMV resistant can be used to develop transgenic line to benefit small farmers in China. As we known, insect-protected cotton containing a natural insecticide protein from *Bacillus thuringiensis* (Bt

cotton) is providing millions of farmers with increased yields, reduced insecticide costs and fewer health risks (Toenniessen et al. 2003). From a human welfare standpoint, the greatest benefits of the research described in this thesis will surely be derived from the doption of improved crop varieties in the developing countries to raise the next wave of ‘Green Revolution”.

Literature Cited

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252(5013): 1651-6.

Aharoni A, Keizer LC, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AM, De Vos RC, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12(5): 647-62.

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8(11): 554-60.

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2(2): 96-103.

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr. Opin. Plant Biol. 2(2): 96-103.

Bendahmane A, Kanyuka K, Baulcombe DC (1999) The Rx gene from potato controls separate virus resistance and cell death responses. Plant Cell 11(5): 781-92.

Bendahmane A, Querci M, Kanyuka K, Baulcombe DC (2000) *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: application to the *Rx2* locus in potato. Plant J. 21(1): 73-81.

Birch PRJ, Avrova AO, Duncan JM, Lyon GD, Toth RL (1999) Isolation of potato genes that are induced during an early stage of the hypersensitive response to Phytophthora infestans. Mol Plant Microbe Interact 12(4): 356-361.

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476): 307-10.

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7(2): 132-6.

Chiu WF (1988). The economic impact of filamentous plant virus - China. The filamentous plant viruses. Milne RG. New York, Plenum Press**:** 387-391.

Chuaqui RF, Bonner RF, Best CJM, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR (2002) Post-analysis follow-up and validation of microarray experiments. Nature Genetics 32: 509-514.

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 11(10): 968-78.

Cooper B (2001) Collateral gene expression changes induced by distinct plant viruses during the hypersensitive resistance reaction in Chenopodium amaranticolor. Plant J. 26(3): 339-349.

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 411(6839): 826-33.

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA 93(12): 6025-30.

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol 48(1-2): 75-97.

Dussle CM, Melchinger AE, Kuntze L, Stork A, Lubberstedt T (2000) Molecular mapping and gene action of Scm1 and Scm2, two major QTL contributing to SCMV resistance in maize. Plant Breeding 119(4): 299-303.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-8.

Fauquet C, Wechmar MBV (1988). The economic impact of filamentous plant virus - Africa. The filamentous plant viruses. Milne RG. New York, Plenum Press**:** 379-384.

Finkelstein D, Ewing R, Gollub J, Sterky F, Cherry JM, Somerville S (2002) Microarray data quality analysis: lessons from the AFGC project. Arabidopsis Functional Genomics Consortium. Plant Mol. Biol. 48(1-2): 119-31.

Flint-Garcia SA, Thornsberry JM, Buckler ESt (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54: 357-74.

Forbes IL, Steib RJ (1964) Controlling sugarcane mosaic. Sugar J. 27: 51-52.

Fuchs E, Gruntzig M (1995) Influence of Sugarcane Mosaic-Virus (Scmv) and Maize-Dwarf Mosaic-Virus (Mdmv) on the Growth and Yield of 2 Maize Varieties. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-

Journal of Plant Diseases and Protection 102(1): 44-50.

Fuchs, E., Grüntzig, M., Kuntze, L. and Hohmann, F. (1996) Zur Epidemiologie der Potyviren des Maises in Deutschland. In Bericht über die 46. Arbeitstagung der Arbeitsgemeinschaft der Saatzuchtleiter im Rahmen der Vereinigung österreichischer Pflanzenzüchter 46: 43-49.

Fuchs, E. and Kozelska, S. (1984) Zum Auftreten und zur Schadwirkung des Europäischen Mais Mosaik Virus (EMMV) in der DDR sowie Ergebnisse der Prüfung der Anfälligkeit von Inzuchtlinien und Hybriden. Nachrichtenblatt Pflanzenschutz DDR 38: 162-164.

Giovanonni J (2000). Tools for tomato genome analysis: application for isolation of fruit ripening loci and expression profiling. International Conference on the Status of Plant and Animal Genome Research VIII. Abstracts, San Diego, CA.

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol 124(4): 1570-81.

Golem S, Culver JN (2003) Tobacco mosaic virus induced alterations in the gene expression profile of Arabidopsis thaliana. Mol. Plant Microbe Interact. 16(8): 681-688.

Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J (2000) A concise guide to cDNA microarray analysis. Biotechniques 29(3): 548-50, 552-4, 556 passim.

Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Biol 54: 375-401.

Katagiri F (2004) A global view of defense gene expression regulation ?a highly interconnected signaling network. Current Opinion in Plant Biology 7: 1-6.

King NC (1955-56) Cane diseases and weed control. Ann. Rept. Expt. Stat. S. Afr. Sugar Assn.: 25-30.

Koike H, Gillaspie AG (1989). Mosaic. Diseases of Sugarcane - Major Diseases. Ricaud et al. Amsterdam, Elsevier Science Publishers: 301-322.

Kuntze L, Fuchs E, Gruntzig M, Schulz B, Henning U, Hohmann F, Melchinger AE (1995) Evaluation of maize inbred lines for resistance to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Agronomie 15(7-8): 463-467.

Kuntze L, Fuchs E, Gruntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early-maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breeding 116(5): 499-501.

Lamb CJ, Ryals JA, Ward ER, Dixon RA (1992) Emerging Strategies for Enhancing Crop Resistance to Microbial Pathogens. Bio-Technology 10(11): 1436-1445.

Louie R (1995) Vascular Puncture of Maize Kernels for the Mechanical Transmission of Maize White Line Mosaic-Virus and Other Viruses of Maize. Phytopathology 85(2): 139-143.

Louie R, Findley WR, Knoke JK, Mcmullen MD (1991) Genetic-Basis of Resistance in Maize to 5 Maize-Dwarf Mosaic-Virus Strains. Crop Science 31(1): 14-18.

Martienssen RA, Rabinowicz PD, OShaughnessy A, McCombie WR (2004) Sequencing the maize genome. Curr Opin Plant Biol 7(2): 102-7.

Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie WR (2004) Sequencing the maize genome. Curr. Opin. Plant Biol. 7(2): 102-7.

Matthews REF, Hull R (2002) Matthews' plant virology. Academic Press, San Diego.

Mazeyrat F, Mouzeyar S, Nicolas P, Tourvieille de Labrouhe D, Ledoigt G (1998) Cloning, sequence and characterization of a sunflower (Helianthus annuus L.) pathogen-induced gene showing sequence homology with auxin-induced genes from plants. Plant Mol Biol 38(5): 899-903.

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123(2): 439-442.

Melchinger AE, Kreps R, Spath R, Klein D, Schulz B (1998) Evaluation of early-maturing European maize inbreds for resistance to the European corn borer. Euphytica 99(2): 115-125.

Melchinger AE, Kuntze L, Gumber RK, Lubberstedt T, Fuchs E (1998) Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm. Theor Appl Genet 96(8): 1151-1161.

Nadimpalli R, Yalpani N, Johal GS, Simmons CR (2000) Prohibitins, stomatins, and plant disease response genes compose a protein superfamily that controls cell proliferation, ion channel regulation, and death. J. Biol. Chem. 275(38): 29579-86.

Nakazono M, Qiu F, Borsuk LA, Schnable PS (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15(3): 583-96.

Ohlrogge J, Benning C (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol 3(3): 224-8.

Pan W (2002) A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 18(4): 546-554.

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Molecular Breeding 7(4): 275-291.

Potokina E, Caspers M, Prasad M, Kota R, Zhang H, Sreenivasulu N, Wang M, Graner A (2004) Functional association between malting quality trait components and cDNA array based expression patterns in barley (Hordeum vulgare L.). Molecular Breeding 14(2): 153-170.

Provvidenti R, Hampton RO (1992) Sources of resistance to viruses in the Potyviridae. Arch. Virol. Suppl. 5: 189-211.

Provvidenti R, Hampton RO (1992) Sources of resistance to viruses in the Potyviridae. Arch Virol Suppl 5: 189-211.

Quackenbush J (2002) Microarray data normalization and transformation. Nat. Genet. 32: 496-501.

Quint M, Dussle CM, Melchinger AE, Lubberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS. Theor Appl Genet 106(7): 1171-7.

Quint M, Dussle CM, Melchinger AE, Lubberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS. Theor. Appl. Genet. 106(7): 1171-7.

Quint M, Mihaljevic R, Dussle M, Xu L, Melchinger E, Lubberstedt T (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor Appl Genet 105(2-3): 355-363.

Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3(2): 108-16.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2): 374-8.

Sargent TD (1987) Isolation of Differentially Expressed Genes. Methods in Enzymology 152: 423-432.

Schnable PS, Hochholdinger F, Nakazono M (2004) Global expression profiling applied to plant development. Curr Opin Plant Biol 7(1): 50-6.

Shukla DD, Tosic M, Jilka J, Ford RE, Toler RW, Langham MAC (1989) Taxonomy of Potyviruses Infecting Maize, Sorghum, and Sugarcane in Australia and the United-States as Determined by Reactivities of Polyclonal Antibodies Directed Towards Virus-Specific N-Termini of Coat Proteins. Phytopathology 79(2): 223-229.

Taniguchi M, Miura K, Iwao H, Yamanaka S (2001) Quantitative assessment of DNA microarrays--comparison with Northern blot analyses. Genomics 71(1): 34-9.

Tao Y, Xie Z, Chen W, Glazebrook J, Chang HS, Han B, Zhu T, Zou G, Katagiri F (2003) Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae. Plant Cell 15(2): 317-30.

Teakle DS, Shukla DD, Ford RE (1989) Sugarcane Mosaic Virus. AAB Description of Plant Viruses. No. 342.

Toenniessen GH, O'Toole JC, DeVries J (2003) Advances in plant biotechnology and its adoption in developing countries. Curr Opin Plant Biol 6(2): 191-8.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93(1): 77-8.

Wan J, Dunning FM, Bent AF (2002) Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. Funct. Integr. Genomics 2(6): 259-73.

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4(1): 29-38.

Whitham SA, Anderberg RJ, Chisholm ST, Carrington JC (2000) Arabidopsis RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. Plant Cell 12(4): 569-82.

Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang X, Hou YM (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible Arabidops*is thaliana plants.* Plant J. 33(2): 271-83.

Xia XC, Melchinger AE, Kuntze L, Lubberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89(8): 660-667.

Xu ML, Melchinger AE, Lubberstedt T (2000) Origin of Scm1 and Scm2 - two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor Appl Genet 100(6): 934-941.

Yuan L, Dussle CM, Melchinger AE, Utz HF, Lubberstedt T (2003) Clustering of QTL conferring SCMV resistance in maize. Maydica 48(1): 55-62.

Summary

SCMV (Sugarcane mosaic virus) is one of the most important virus diseases of maize in Europe and causes serious yield losses in susceptible cultivars. It is readily transmitted by aphids in a non-persistent manner. Thus, chemical control is not efficient for control of SCMV. Cultivation of resistant maize varieties is the most efficient and environmentally sound approach to limit yield-loss caused by SCMV.

So far, the molecular mechanisms underlying the development and progression of SCMV infection in maize are poorly understood. Furthermore, the characterization of genes underlying QTL by positional cloning is very laborious and time consuming. Recently, the methods becoming available within the functional genomics "toolbox" provide an alternative for pinpointing genes underlying SCMV resistance, especially in view of the planned sequencing of major parts of maize genome. Two complementary approaches, SSH (Suppression subtractive hybridization) and microarray-based expression profiling, were used to isolate and identify candidate genes associated with SCMV resistance (1st and 2nd paper, respectively). Since current maize microarrays include less than 30% of all maize genes, SSH was conducted to identify rare transcripts associated with SCMV resistance. Expression profiling has become the predominant high-throughput transcript profiling method in understanding host-pathogen interaction.

In the 1st paper, SSH was combined with macroarray hybridization to identify genes differently expressed in NILs (Near isogenic lines) F7+ (SCMV resistant, carrying *Scmv1* and *Scmv2* regions from FAP1360A) and F7 (SCMV susceptible). Altogether, 302 differentially expressed SSH-ESTs were identified in four comparisons addressing constitutive genetic discrepancy, inducible genetic discrepancy, compatible reaction, and

incompatible reaction. Except for genes related to “metabolism”, most of the classified genes belonged to the three pathogenesis-related categories, “cell rescue, defense, cell death and ageing”, “signal transduction” and “transcription”, which accounted for 56-66% of the classified genes. In total, 19% (60 of 302) of the identified SSH-ESTs have previously been assigned to 29 bins distributed over all the 10 maize chromosomes. Among the mapped SSH-ESTs, 30% (18 of 60) were located within the *Scmv2* and *Scmv1* genome regions on chromosomes 3 and 6, respectively, conferring resistance to SCMV. Promising candidate genes have been identified, such as AA661457 (receptor-like kinase Xa21-binding protein 3) for *Scmv1*.

In the 2nd paper, genes associated with SCMV resistance in the NIL pair $F7^{+}$ and F7 were identified by transcript profiling based on maize unigene-microarrays. Altogether, 497 differentially expressed genes were identified in the same four comparisons as in the SSH approach, addressing constitutive genetic discrepancy, inducible discrepancy, compatible reaction, and incompatible reaction. Compared to the SSH approach, expression patterns of microarray-ESTs and SSH-ESTs were consistent for the same comparisons despite of technical discrepancies. Since pathogen-induced transcripts were underrepresented on the unigene-microarray, fewer microarray-ESTs (45.8%) were classified into pathogenesis-related categories than by using SSH-ESTs (60.5%). Moreover, fewer microarray-ESTs (4) co-segregated with *Scmv* QTL than SSH-ESTs (18). Therefore, our results demonstrate that SSH-macroarray complements incomprehensive microarrays.

The candidate genes isolated and described in the 1st paper were employed to investigate their association with SCMV resistance across seven resistant or susceptible

inbreds in the 3rd paper. The number of differentially expressed genes (SCMV infected versus non-infected) in individual lines was 177, 163, 165, 62, 47, 37, and 93, for FAP1360A, D21, D32, Pa405, F7, D145, and D408, respectively. All inbreds were divided into two groups by hierarchical cluster analysis: D32, D21, FAP1360A and D408 formed one group; Pa405, D145 and F7 another group. Due to the genetic structure among the seven inbreds, genetic background and resistance response are confounded. With or without the resistant U.S. inbred line Pa405, 22 and 112 genes were identified by *t* tests between resistant (D21, D32, and FAP1360A) and susceptible (D145, D408, and F7) inbred lines, respectively. The 112 candidate genes were divided into three clusters by K-means clustering and analyzed in more detail. These candidate genes identified from present analysis can be further investigated in a segregating population by "genetical genomics" approach.

In conclusion, this thesis demonstrates the usefulness of expression profiling to study SCMV resistance and to identify candidate genes potentially affecting the signal transduction pathway or even for previously identified SCMV QTL. This information is relevant for plant breeders in view of development of functional markers. Due to oligogenic inheritance of SCMV resistance, marker-assisted selection (MAS) programs with functional markers would increase the breeding efficiency.

Zusammenfassung

SCMV (Sugarcane mosaic virus) ist eine der wichtigsten Viruserkrankungen beim Mais in Europa und verursacht erhebliche Ernteverluste in anfälligen Sorten. Er wird leicht von Aphiden auf eine nicht-persistente Weise übertragen. Daher sind chemische Bekämpfungsmassnahmen des SCMV nicht effektiv. Der Anbau resistenter Maissorten ist die effektivste und umweltfreundlichste Massnahme zur Begrenzung von Ernteverlusten.

Bisher sind die molekularen Mechanismen, denen die Entwicklung und das Fortschreiten einer Infektion mit SCMV unterliegen, wenig verstanden. Zudem ist die Charakterisierung von Genen, die für QTL verantwortlich sind, durch positionelles Klonieren sehr arbeits- und zeitintensiv. Seit kurzem bieten Methoden der funktionalen Genomik eine Alternative für die Identifikation von Genen, die an der Ausprägung der SCMV-Resistenz beteiligt sind. Zwei sich gegenseitig ergänzende Ansätze, SSH (suppressive subtraktive Hybridisierung) und Microarray-basiertes Expressionsprofiling, wurden verwendet, um mit SCMV-Resistenz assoziierte Kandidatengene zu isolieren und zu identifizieren (erste und zweite Veröffentlichung). Da aktuelle Mais Microarrays weniger als 30% aller Maisgene enthalten, wurde die SSH zur Identifizierung seltener Transkripte eingesetzt. Expressionsprofilierung ist mittlerweile eine wichtige high-throughput (Hochdurchsatz-) Transkript-Profilierungsmethode zum Verständnis von Wirt-Pathogen Interaktion.

In der ersten Veröffentlichung wurde die SSH mit Microarray-Hybridisierung kombiniert, um unterschiedlich exprimierte Gene in den NILs (nah-isogene Linien) F7+ (SCMV-resistent, Träger der *Scmv1* und *Scmv2* Regionen von FAP1360A) und F7 (SCMV-anfällig) zu identifizieren. Insgesamt wurden in vier Vergleichen 302 differentiell exprimierte SSH-ESTs identifiziert, die konstitutive Abwehr, induzierbare Abwehr, kompatible und inkompatible Reaktion ansprechen.

Mit Ausnahme der mit dem „Stoffwechsel“ zusammenhängenden Gene gehörten die meisten klassifizierten Gene (56-66%) zu den drei Klassen, „Zellrettung, Abwehr, Zelltod und Altern“, „Signaltransduktion“ und „Transkription“. Insgesamt waren bereits vorher 19% der identifizierten SSH-ESTs (60 von 302) 29 Chromosomenregionen (bins) zugeordnet worden, die über alle 10 Maischromosomen verteilt sind. 30% der kartierten SSH-ESTs (18 von 60) waren innerhalb der *Scmv2* und *Scmv1* Genomregionen auf den Chromosomen 3 und 6 lokalisiert, die Resistenz gegen SCMV vermitteln. Vielversprechende Kandidatengene wie z.B. AA661457 (receptor-like kinase Xa21-binding protein 3) für *Scmv1*, wurden identifiziert.

In der zweiten Veröffentlichung wurden Gene, die mit SCMV-Resistenz assoziiert sind, in dem NIL-Paar $F7^{+}$ und F7 durch Transkriptprofilierung auf der Grundlage von Mais „unigene microarrays“ identifiziert. Insgesamt wurden 497 unterschiedlich exprimierte Gene in den gleichen vier Vergleichen wie im SSH-Ansatz identifiziert, die konstitutiver und induzierbarer Abwehr, sowie kompatibler und inkompatibler Reaktion entsprechen. Verglichen mit der SSH stimmten die Expressionsmuster der Microarray-ESTs und der SSH-ESTs trotz technischer Diskrepanzen grundsätzlich für die gleichen Vergleiche überein. Da Pathogen-induzierte Transkripte auf den „unigene microarrays“ unterrepräsentiert waren, wurden weniger Microarray-ESTs (45,8%) in Pathogenese-bezogene Kategorien eingeteilt als bei der Verwendung von SSH-ESTs (60,5%). Darüberhinaus kosegregierten weniger Microarray-ESTs (4) mit *Scmv* QTL als SSH-ESTs (18). Daher demonstrieren unsere Ergebnisse, dass SSH-Microarrays unvollständige Microarrays koplementieren.

Die isolierten Kandidatengene, die in der ersten Veröffentlichung beschrieben wurden, wurden herangezogen, um ihre Verbindung mit SCMV-Resistenz über sieben resistente oder anfällige Inzuchtlininien in einer dritten Veröffentlichung zu untersuchen. Die Anzahl der unterschiedlich exprimierten Gene (SCMV infiziert versus nicht-infiziert) war 177, 163, 165, 62, 47, 37 und 93 für die einzelnen Linien FAP1360A, D21, D32, Pa405 (alle resistent), F7, D145, and D408 (anfällig). Alle Inzuchtlinien wurden durch hierarchische Clusteranalyse in

zwei Gruppen eingeteilt: D32, D21, FAP1360A und D408 bildeten die eine Gruppe; Pa405, D145 und F7 bildeteten die andere Gruppe. In der ersten Gruppe handelt es sich um europäische Dent-Linien, bei der zeiten um Flint-Linien bzw. eine U.S.-amerikanische Linie (Pa405).Mit und ohne die U.S. Inzuchtlinie Pa405 wurden 22 bzw. 112 Gene durch t-Tests als gemeinsam differentiell exprimiert zwischen resistenten (D21, D32 und FAP1360A) und anfälligen Inzuchtlinien (D145, D408 und F7) identifiziert. Die 112 Kandidatengene wurden durch „K-Mittelwert-Clusterung“ in drei Cluster eingeteilt und detaillierter analysiert. Diese identifizierten Kandidatengene der aktuellen Analysen können in einer spaltenden Population durch „genetical genomics“ weiter untersucht werden.

Zusammenfassend zeigt diese Arbeit die Zweckmäßigkeit des Expressionsprofiling für für die Identifikation potentieller Kandidatengene für SCMV Resistenz, die Genen der Signaltransduktionskette oder sogar früher identifizierten SCMV QTL entsprechen. Diese Information ist für Pflanzenzüchter im Hinblick auf die Entwicklung funktionaler Marker relevant. Da SCMV-Resistenz oligogen vererbt wird, würden markergestützte Selektionsprogramme mit funktionalen Markern die Züchtungseffizienz deutlich erhöhen.

Appendix: List of publications

Contributions to the manuscripts presented in this thesis are noted in brackets.

1. **Chun Shi, Christina Ingvardsen, Fritz Thümmler, Albrecht E. Melchinger, Gerhard Wenzel, Thomas Lübberstedt** (2004) Identification of differentially expressed genes between maize near-isogenic lines in association with SCMV resistance using suppression subtractive hybridization. *Molecular Genetics and Genomics.* Revised.

(Concept by T. Lübberstedt and A. E. Melchinger; Experiments conducted by C. Shi except for SSR analysis by C. Ingvardsen; Experimental support by F. Thümmler and G. Wenzel; Writing by C. Shi and T. Lübberstedt)

2. **Chun Shi, Fritz Thümmler, Albrecht E. Melchinger, Gerhard Wenzel, Thomas Lübberstedt** (2004) Comparison of transcript profiles between near-isogenic maize lines in association with SCMV resistance based on unigene-microarrays. *Molecular plant pathology.* Submitted.

(Concept by T. Lübberstedt and A. E. Melchinger; Experiments conducted by C. Shi except for SSR analysis by C. Ingvardsen; Experimental support by F. Thümmler and G. Wenzel; Writing by C. Shi and T. Lübberstedt)

3. **Chun Shi, Fritz Thümmler, Albrecht E. Melchinger, Gerhard Wenzel, Thomas Lübberstedt** (2004) Association between SCMV resistance and macroarray-based expression patterns in different maize inbreds. *Molecular breeding*. Submitted.

(Concept by T. Lübberstedt and A. E. Melchinger; Experiments conducted by C. Shi except for SSR analysis by C. Ingvardsen; Experimental support by F. Thümmler and G. Wenzel; Writing by C. Shi and T. Lübberstedt)

Identification of differentially expressed genes between maize near-isogenic lines in association with SCMV resistance using suppression subtractive hybridization

Chun Shi[1] , Christina Ingvardsen[4], Fritz Thümmler[2], Albrecht E. Melchinger[3], Gerhard Wenzel[1], Thomas Lübberstedt[4,*]

[1]Chair of Plant Breeding, Technical University of Munich, Am Hochanger 2, 85350 Freising, Germany
[2]Vertis Biotechnologie AG, Lise-Meitner-Straße 30, 85354 Freising, Germany
[3]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
[4]Department of Plant Biology, Research Centre Flakkebjerg, 4200 Slagelse, Denmark

* For correspondence (tel. +45 5811 3484; fax +45 5811 3301; e-mail Thomas.Luebberstedt@agrsci.dk)

The molecular mechanisms underlying the development and progression of sugarcane mosaic virus (SCMV) infection in maize are poorly understood. A study on differential expression was conducted to identify genes involved in SCMV resistance. In this study, we combined suppression subtractive hybridization (SSH) and macroarray hybridization to identify genes differently expressed in the near isogenic line (NIL) pair $F7^+$ (SCMV resistant) and F7 (susceptible). Altogether, 302 differentially expressed genes were identified in four comparisons addressing constitutive genetic discrepancy, inducible genetic discrepancy, compatible interaction, and incompatible interaction. Except for genes related to "metabolism", most of the classified genes belonged to the three pathogenesis-related categories, "cell rescue, defense, cell death and ageing", "signal transduction" and "transcription", which accounted for 56-66% of the classified genes. In total, 19% (60 of 302) of the identified genes have previously been assigned to 29 bins distributed over all the 10 maize chromosomes. Among the mapped genes, 31% (18 of 58) were located within the *Scmv2* and *Scmv1* genome regions on chromosomes 3 and 6, respectively, conferring resistance to SCMV. Promising candidate genes have been identified, such as AA661457 (receptor-like kinase Xa21-binding protein 3) for *Scmv1*. Implications of the genomic distribution of differentially expressed genes from this isogenic comparison are discussed in view of resistance breeding.

Keywords: Maize, SCMV resistance, NILs, SSH, macroarray, expression profiling

Introduction

Sugarcane mosaic virus (SCMV) is an important pathogen of maize (Zea mays L.) in Europe and China, causing substantial yield losses in susceptible cultivars (Fuchs and Gruntzig 1995). In previous studies, Kuntze et al. (1997) screened 122 early-maturing European inbred lines for resistance to SCMV and MDMV (maize dwarf mosaic virus) and identified only three inbreds (D21, D32, and FAP1360A) displaying complete resistance under both field and greenhouse conditions. Two major QTL regions, Scmv1 and Scmv2, conferring resistance to SCMV were mapped to chromosome arms 6S and 3L. In cross D145 (susceptible) × D32 (resistant) quantitative trait locus (QTL) analysis (Xia et al. 1999) and in cross F7 (susceptible) × FAP1360A (resistant) bulked segregant analysis (BSA) (Xu et al. 1999) and QTL analysis (Dussle et al. 2000) were applied. For full resistance to SCMV, both Scmv regions are required. Scmv1 is sufficient for resistance in earlier stages of plant development, whereas Scmv2 is expressed only at a later stage of plant development and contributed a higher degree of resistance in both populations D145 × D32 and F7 × FAP1360A (Dussle et al. 2000). Epistatic effects were found between Scmv1 and Scmv2 in population D145 × D32 (Xia et al. 1999) but not in population F7 × FAP1360A (Dussle et al. 2000). The Scmv1 region contains a minimum of two QTL (Scmv1a and Scmv1b) (Yuan et al. 2003). Together with three additional minor QTL identified on chromosomes 1, 5, and 10 (Xia et al. 1999), a minimum of six genes are involved in the oligogenic inherited SCMV resistance.

Recent work indicates that the candidate gene approach is an efficient way to establish the association between resistance gene candidates and both qualitative and quantitative resistances in rice (Ramalingam et al. 2003). Maize resistance gene analogues (RGA) involved in initial pathogen recognition, were chosen as starting point for isolation of genes conferring SCMV resistance (Collins et al. 1998). Mapping of RGAs in relation to *Scmv1* and *Scmv2* suggested that RGA *pic19* is a candidate for *Scmv1* and *pic13* for *Scmv2* (Quint et al. 2002). *pic19* and *pic13* were used to screen a BAC library of inbred line B73 and three paralogues clustering in the *Scmv1* region were isolated from the maize genome (Quint et al. 2003), currently analyzed in more detail (Xu and Lübberstedt, unpublished results). However, candidate genes involved in subsequent steps leading to resistance or susceptibility after the initial recognition of SCMV are not available so far.

Besides stunting, typical SCMV symptoms are dark green islands parallel to the mid vein next to leaf tissue lighter in color. A complex of factors is involved in the development of mosaic symptoms including systemic and local movement of the virus, the ability of the virus to invade meristematic tissues, the strain of the virus and its propensity to mutate and, probably above all, the conflict between the invasiveness of the virus and the response of the host defense system (Matthews and Hull 2002).

Microarray-based expression profiling methods, together with genomic and/or EST (expressed sequence tag) sequence data resulted in significant progress in characterization of plant pathogenesis-related responses (Wan et al. 2002). Diverse RNA viruses (cucumber mosaic cucumovirus, oil seed rape tobamovirus, turnip vein clearing tobamovirus, potato virus X potexvirus, and turnip mosaic potyvirus) elicited expression of common sets of genes in susceptible *Arabidopsis thaliana* plants using Arabidopsis GeneChip microarrays (Whitham et al. 2003). Baldwin et al. (1999) identified 117 genes that consistently showed altered mRNA expression in maize 6 h after various treatments with the fungal pathogen *Cochliobolus carbonum*, using a maize DNA microarray representing 1,500 maize genes. Using a similar approach, Nadimpalli et al. (2000) identified nearly 70 genes having a more than twofold change in mRNA abundance in the lesion mimic maize mutant, *Les9* compared to wild-type plants. In the Kansas State University Defense Gene Collection (http://www.ksu.edu/ksudgc/), 31 defense genes contributed by different researchers were either reported to be involved in maize defense responses or are known resistance genes. These genes were classified into seven functional groups: resistance gene homologues, amino acid metabolism, lipid metabolism, detoxification, phenylpropanoid pathway, PR proteins, DNA binding.

In contrast to comprehensive microarrays in Arabidopsis and rice (Zhu et al. 2003), publicly available maize microarrays contain 9,841 different unigenes(http://www.plantgenomics.iastate.edu/maizechip/), which account for 20% - 39% of 25,000 - 50,000 total maize genes (Martienssen et al. 2004). Other expression profiling methods, such as SSH (suppression subtractive hybridization) (Diatchenko et al. 1996) and cDNA-AFLP (cDNA amplified fragment length polymorphism) (Vos et al. 1995), might in this situation be more powerful to detect differentially expressed rare transcripts. For instance, by SSH and cDNA-AFLP Birch et al. (1999) isolated potato genes induced during an early stage of hypersensitive response to *Phytophthora infestans*. The particular strength of SSH is identification of rare transcripts (Diatchenko et al. 1996), which can be combined with macroarray technology for high-throughput screening, without need to obtain previously cloned cDNAs (Yang et al. 1999).

We investigated the NILs F7 (SCMV susceptible) and $F7^+$ (SCMV resistant, carrying *Scmv1* and *Scmv2* regions from FAP1360A) to conduct expression profiling experiments. Because these lines are almost identical, the background noise due to variable genome regions is eliminated. Differentially expressed genes might be derived from the *Scmv1* or *Scmv2* genome regions, and thus, be candidate genes for the previously mapped QTL. If located in other genome regions, these genes may function further downstream in the signal transduction pathway and their induction may be mediated by genes located in the *Scmv1* and / or *Scmv2* regions. To date, over 3,000 maize ESTs have mapping information from the Maize GDB and the Maize IDP project. Establishing a transcriptom map will help us to select positional candidate genes, mapped inside QTL intervals, and functional candidate genes.

The objectives of our study were to 1) phenotypically and genotypically evaluate NILs F7 and $F7^+$ by infection trials and SSR markers, and 2) identify genes associated with SCMV resistance in maize combining SSH and macroarray techniques.

Materials and Methods

Development and evaluation of NILs F7 and F7+

NIL F7+ was developed using phenotypic and marker-assisted selection. The early maturing European maize inbreds, FAP1360A, resistant to SCMV, and F7, highly susceptible to SCMV (Kuntze et al. 1997), were crossed to produce an F1 generation, and backcrossed seven times to F7 with two generations per year from 1995 to 1998 (Dussle et al. 2003). Seed of the homozygous line F7+ was produced by three subsequent selfing steps starting from one SCMV resistant BC7 plant carrying the donor regions from FAP1360A at *Scmv1* and *Scmv2*.

In 2001-2002, FAP1360A, F7 and F7+ were evaluated for resistance to SCMV in two replicated field trials and one greenhouse trial at TU Munich - Weihenstephan. Both trials included two replications with 25 plants per row. Plants at the three to four leaf stage were mechanically inoculated twice at a weekly interval by an air brush technique using a tractor-mounted air compressor at constant pressure of 799 kPa (Fuchs and Gruntzig 1995). Resistance to SCMV was visually scored in 2-weekly intervals.

A total of 25 simple-sequence repeat (SSR) markers mapping to the short arm of chromosome 6 (Bin 6.00 and 6.01) and near to the centromere region of chromosome 3 (Bin 3.04 and 3.05) (Dussle et al. 2003) were chosen to evaluate NIL F7+. Sequences of all SSR markers were obtained from the Maize GDB (http://www.maizegdb.org) and primers synthesized by Metabion (Munich, Germany). PCR amplification and MetaPhor gel-electrophoresis were performed as described by Lübberstedt et al. (1998).

Plant materials for RNA extraction

NILs F7 and F7+ were grown and maintained in growth chambers under a 12 -h photoperiod at 23 oC and 50% relative humidity. Two-week-old plants were used for virus inoculations. Non-infected plants and infected plants were kept in separate growth chambers after inoculation. Non-infected and infected leaves were harvested 24 hours after inoculation in parallel. For biological replicates, two independent sets (five plants per each set) of leaf materials were harvested. To confirm resistance or susceptibility of infected plants used for leaf harvest, plants were grown for additional two weeks. After this period, mosaic symptoms were observed on each infected F7 plant, whereas infected F7+ remained without symptoms.

RNA extraction

Total RNA from maize leaves was extracted using TRIzol reagent (Invitrogen, Carlsbad, California, USA). Poly (A)+ RNA was separated from total RNA using Oligotex mRNA Midi Kit (Qiagen GmbH, Hilden, Germany).

Suppression subtractive hybridization (SSH) and cDNA library construction

cDNA synthesis and SSH were carried out using a PCR-Select cDNA Subtraction Kit (BD Biosciences, San Jose, CA, USA) according to manufacturer's protocol. In brief, 2 µg of poly (A)+ RNA from the tester and the driver, defined in Table 1, were used for cDNA synthesis. After digestion with RsaI, the tester cDNA preparation was divided into two subpopulations, which were ligated with different adaptors. The two subpopulations were then hybridized with an excess amount of driver cDNA, after which they were combined and hybridized again in the presence of driver cDNA, without denaturing the DNA before the second hybridization. Following the second hybridization, two PCR rounds were performed to enrich and amplify the differentially expressed sequences. The subtracted tester cDNA was cloned into pCRII-TOPO TA cloning vector (Invitrogen, Carlsbad, CA, USA) and transformed into One Shot® TOP10 electrocompetent E. coli cells (Invitrogen, Carlsbad, CA, USA). Finally, five SSH libraries were constructed. For two tester/driver cDNA pairs (infected F7+ versus infected F7; non-infected F7+ versus non-infected F7) subtractions were conducted in both directions. For the tester/driver cDNA pair infected F7+ versus non-infected F7+ only forward direction was conducted (Table 1).

Construction of macroarrays

From each subtracted cDNA library, more than 96 colonies were randomly picked and PCR-amplified in 25 µl reactions in 96-well plates. Specific primers were designed (forward primer: 5'-ATGCTTCCGGCTCGTT-3'; reverse primer: 5'-CAGGGTTTTCCCAGTC-3'). After gel eletrophoresis, the PCR products of clones having inserts were collected in a 96-well plate. We used

Table 1. Overview over five cDNA libraries generated by suppression subtractive hybridization

No.	Tester	Driver	Genes expected to be enriched
1	Infected F7$^+$	Not infected F7$^+$	SCMV induced genes in resistant genotype
2	Infected F7$^+$	Infected F7	Genetic discrepancy between NILs including induced resistance factors
3	Infected F7	Infected F7$^+$	Genetic discrepancy between NILs including repressed resistance factors
4	Not infected F7$^+$	Not infected F7	Genetic discrepancy between NILs including preformed
5	Not infected F7	Not infected F7$^+$	resistance or susceptibility factors

Nano-plotter™ (Gesim, Großerkmannsdorf, Germany) to spot cDNA clones onto Biodyne B transfer membrane (Pall Europe Ltd., Portsmouth, England). The preliminary macroarrays containing 96 clones were constructed and hybridized with unsubtracted tester and driver probes labeled with [α-^{32}P] dATP (MP Biomedicals, Irvine, CA, USA).

After checking subtraction efficiency of each SSH library, more clones were prepared in the same way as the clones for preliminary macroarrays. Thereafter, using the PCR product of each clone as the template, two independent re-amplification 100 µl reactions were performed and pooled to reduce variation in PCR efficiency. Pooled PCR products were concentrated from 200 µl to approximately 25 µl by MultiScreen-PCR plates (Millipore, Billerica, Mass, USA) before transfer to 384-well plates. We used the BioGrid robotic system (BioRobotics Ltd. Cambridge, U.K.) with a 384 gridding tool (radius 0.4 mm, 5 transfers per spot) to spot cDNA clones onto H-bond nylon transfer membranes (Amersham, Piscataway, NJ, USA). The spotting scheme followed a 4x4 secondary grid pattern, with each secondary grid containing seven clones spotted in duplicate, plus two empty local background spots. DNA on membranes were crosslinked in a GS Gene Linker chamber (Bio-Rad Laboratories, Hercules, CA, USA). In total, 2688 clones were spotted on each macroarray, including Library 1, 2 and 3 contributing 576 clones each, Library 4 and 5 contributing 384 clones each, internal controls and two RGAs (*pic13* and *pic19*) (Quint et al. 2003). The SpotReport™ Alien™ cDNA Array Validation System (Stratagene, La Jolla, CA, USA) was used for internal controls, which contain ten spiking controls and negative controls including poly $(dA)_{40\text{-}60}$ and water. Equal amounts of PCR products from ten different alien genes with no significant homology to any known nucleic acid, were spotted on the macroarray as spiking controls. As described in the manual, hybridization signals detected from control spots on the macroarray were evaluated to determine (i) quality of both the macroarray and the mRNA, (ii) the macroarray orientation, and (iii) the sensitivity, specificity, signal linearity, and consistency of the assay.

Macroarray hybridizations

In each labeling reaction, 5, 2.5, 1.25, 0.625, 0.312, 0.156, 0.078, 0.04, 0.02 and 0.01 ng of the ten different alien mRNA spikes were added to 5 µg of total RNA. Using the Strip-EZ RT kit (Ambion, Austin, TX, USA), cDNA synthesis was primed using oligo(dT) and [α-^{32}P]dATP (MP Biomedicals, Irvine, CA,USA). cDNA was separated from unincorporated nucleotides using Micro Bio-Spin chromatography columns (Bio-Rad Laboratories, Hercules, CA ,USA) filled with Sephadex G-50 (Amersham, Piscataway, NJ, USA) equilibrated in water. tRNA and oligo(dA) was added to the hybridization probe to suppress cross-hybridization. The prehybridization and Hybridization steps were conducted as described in the manual of the Strip-EZ RT kit. After scanning, labeled cDNA probes were stripped from the arrays using the Strip-EZ system (Ambion, Austin, TX, USA), and the process checked by phosphorimaging. For technical replicates, every RNA sample was used in two independent labeling and hybridization experiments.

Raw data acquisition

Hybridization signals were detected using the Storm 860 phosphorimager (Amersham, Piscataway, NJ, USA) with a resolution of 50 µm. The image data obtained were imported into the software program ArrayVision 7.0 (Imaging Research, St. Catharines, Ontario, Canada) for spot detection and quantification of hybridization signals. Local background calculated from empty spots in each secondary grid, were subtracted using ArrayVision 7.0 to obtain raw signal intensities.

Macroarray Data Analysis

Raw data were exported from ArrayVision 7.0 into Excel. Duplicate spots at macroarrays were averaged. According to spiking controls, data of different macroarrays were normalized and converted to TIGR Array Viewer (TAV) format files.

Using the TIGR Microarray Data Analysis System (MIDAS) (Saeed et al. 2003), Signals were first filtered to exclude low intensity signals, and then "lowess (locally weighted linear regression) normalization" was employed to adjust intensity-dependent effects in log2 (ratio) values. "Replicate consistency checking" removed genes giving poorly reproducible signals, and finally "slice analysis" was utilized to identify differentially expressed genes, which were induced/repressed more than 1.96 standard deviations from the local mean in each comparison (Quackenbush 2002). Cluster analysis of differentially expressed genes was conducted by Multiexperiment Viewer (MeV) (Saeed et al. 2003).

Sequence Analysis

Differentially expressed clones were sequenced by MWG (Ebersberg, Munich, Germany). All sequences were compared with the EST database in Maize GDB (http://www.maizegdb.org) by BLASTN analysis with a threshold E value of 10^{-5}. The EST homologs with highest scores were used to represent our ESTs. Of them, several ESTs have mapping information from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/). Putative mapping position of our ESTs were deduced from those mapped ESTs according to the criteria, at least 70% identity over at least 60% the length (Salse et al. 2004). Annotation of each EST was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). Each EST was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS)(http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. The Graphical genotype of $F7^{+}$ was displayed by GTT software (van Berloo 1999). The distribution of mapped ESTs was drawn by MapChart (Voorrips 2002).

Results

Evaluation of NIL $F7^{+}$ by infection trials and SSR markers

In two field trials, the infection level of the susceptible parent F7 was 90% in the first and

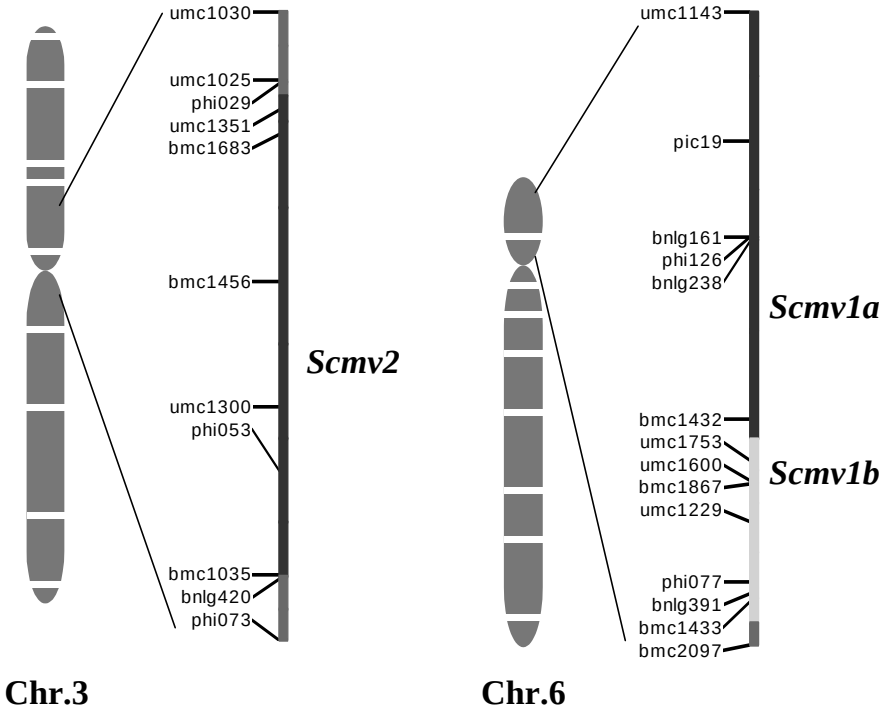


Figure 1. Graphical genotype of near isogenic line $F7^{+}$.
The chromosome segments fixed in F7+ for the FAP1360A and F7 allele are displayed in black and dark gray, respectively, and light gray otherwise. The approximate positions of *Scmv1a*, *Scmv1b*, and *Scmv2* are included according to Yuan et al. (2002).

100% in the second trial four weeks after initial SCMV inoculation. No plants of the resistant line FAP1360A and F7$^+$ displayed SCMV symptoms in both trials. In the greenhouse, FAP1360A and F7$^+$ were completely resistant to SCMV, whereas F7 was 100% susceptible.

A set of SSR markers employed in a previous study (Dussle et al. 2003) were used to evaluate the *Scmv1* and *Scmv2* genome regions of NIL F7$^+$. Among them, 14 SSR markers cover 68.7 cM of the *Scmv1* region on chromosome 6, and 11 SSRs cover 57.6 cM of the *Scmv2* region on chromosome 3. On chromosome 3, all SSRs in the interval between umc1351 and bnlg420 revealed fixation of the FAP1360A allele in F7$^+$, whereas the F7 allele was fixed outside this interval (Fig. 1). On chromosome 6, the FAP1360A allele was fixed in F7$^+$ for the SSR interval between umc1143 and bmc1432, and for the F7 allele at bmc2097, whereas the segment between umc1753 and bmc1433 harboring *Scmv1b* was not fixed (Fig. 1).

Experiment quality control

Subtraction efficiency of each SSH library was tested by a preliminary macroarray containing 96 clones. It was hybridized with unsubtracted tester and driver probes. On average, hybridization signals of tester probes were approximately two - three folds stronger than driver probes, indicating a successful subtraction.

Sensitivity of the assay was tested by ten spiking controls. It showed linear hybridization signals from 20 pg to 2500 pg. The lowest amount of mRNA reliably distinguished from the background was 20 pg, corresponding to 0.01% of total mRNA used for probe preparation. The differences in signal intensity among the macroarrays, due to the differences in isotope incorporation and quantum yield, were normalized according to the mean intensity of spiking controls in each macroarray.

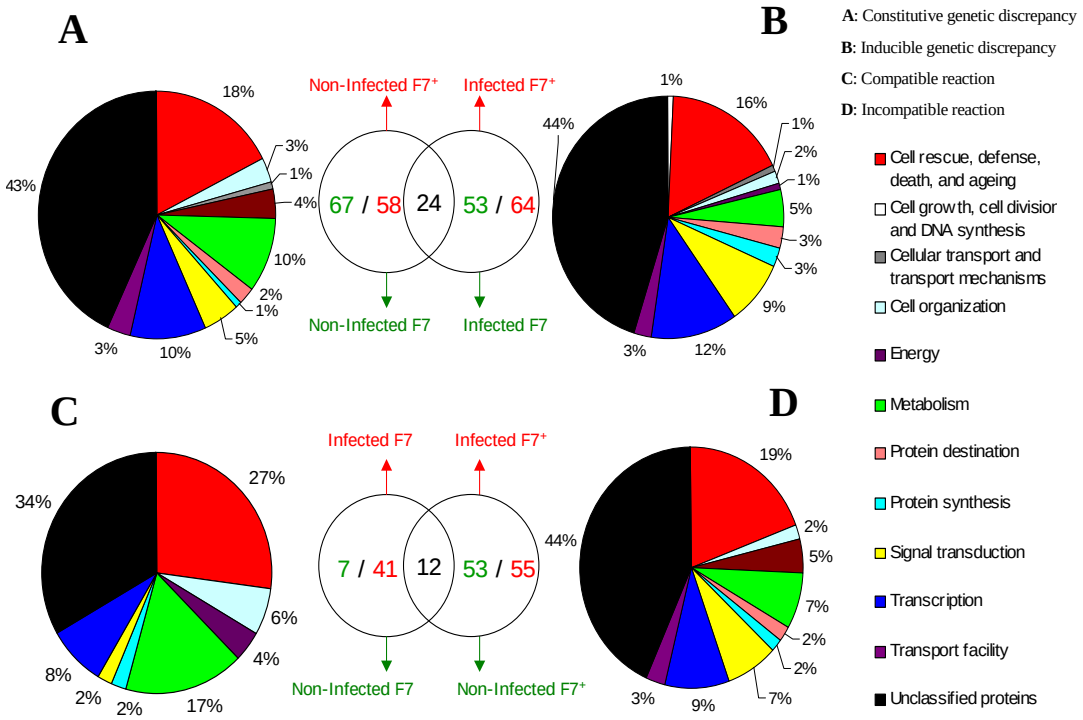


Figure 2. Distribution of differentially expressed genes in functional classes. Four experiments on differential gene expression were conducted: Comparison A (constitutive genetic discrepancy), Comparison B (inducible genetic discrepancy), Comparison C (compatible interaction) and Comparison D (incompatible interaction). The numbers of at least 2-fold down-regulated or 2-fold up-regulated genes are given in green or red, respectively, within circles. The number of differentially expressed genes in common between comparisons is noted in the overlap of circles. Each gene was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. Unclassified indicates no significant similarity to genes of known function.

The reproducibility within the same macroarray was high, with a Pearson correlation coefficient exceeding 0.99 between duplicate spots across all tests. When ^{32}P labeled cDNA probes prepared from the same mRNA were hybridized to the same macroarray, over 95 % of the ratios varied less than 1.5-fold (Pearson correlation coefficient 0.95±0.02). When probes were prepared from biological replicates and hybridized to the same macroarray, over 90% of the ratios calculated from these two data sets varied by less than 2-fold (Pearson correlation coefficient 0.97± 0.02).

Identification and classification of differentially expressed genes

In order to capture a wide spectrum of differentially expressed genes, five SSH libraries were constructed (Table 1). The macroarrays involving clones from all SSH libraries were hybridized with cDNA preparations from non-infected F7, non-infected F7$^+$, infected F7 and infected F7$^+$ (Fig. 2). Of all 2688 clones analyzed, 1603 (59%) clones were identified as differentially expressed in one or more comparisons.

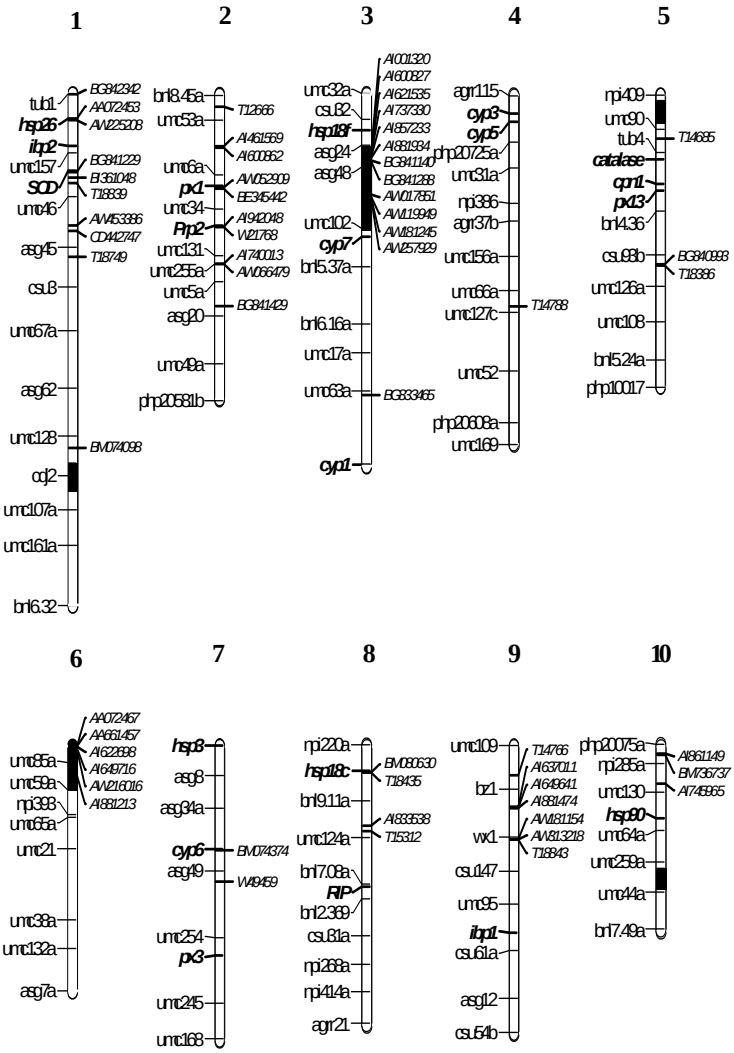


Figure 3. Distribution of genes differentially expressed with respect to SCMV resistance on maize chromosomes.The loci placed on the *left* side of each chromosome were a set of core markers that defines a bin boundary (Gardiner et al. 1993), while the loci in *bold* and *italics* are mapped loci related with pathogenesis in maize (Ramalingam et al. 2003). The loci placed on the *right* side of each chromosome are putative mapped ESTs, from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins. *Scmv1* is highlighted on chromosome 6, *Scmv2* on chromosome 3 and three minor *QSCMs* on chromosomes 1, 5, and 10 (Xia et al. 1999). Hsp: heat shock protein; Ibp: initiator binding protein; Prp: pathogenesis-related protein; px: peroxidase; cyp: cytochrome P450; RIP: ribosome-inactivating protein; cpn: chaperonin.

Sequencing was done in seven subsequent batches of 96 clones. The seventh sequencing batch revealed only 8 additional new genes, for which reason only 672 and not all 1603 clones were sequenced. The 672 sequenced clones, ranging in length from 96 to 843 bp, clustered into 302 ESTs. Among them, 147 were singletons and 155 were contigs made up of 2-16 overlapping clones.

In this study, four comparisons addressing biological questions were made (Fig. 2). Comparisons A and B included different genotypes with the same treatment. Comparison A: F7 Non-infected versus $F7^+$ Non-infected, which is constitutive genetic discrepancy. Structural and chemical barriers of the plant effectively exclude the majority of organisms. The genotypic difference between NILs might include constitutive resistance or susceptibility factors. B: F7 infected versus $F7^+$ infected, which is inducible genetic discrepancy. If constitutive defense of a plant is overcome, a sensitive surveillance system can detect foreign pathogens and trigger a rapid response to injury or virus attack. Genetic discrepancy after SCMV inoculation might include induced or repressed resistance factors between NILs. Comparison C and D include identical genotypes (F7 or $F7^+$) with different treatments. Comparison C: F7 infected versus F7 non-infected, (compatible interaction): virus replicates and moves systemically in cells of intact susceptible plants. Comparison D: $F7^+$ infected versus $F7^+$ non-infected, (incompatible interaction): virus multiplication is limited to initially infected cells of resistant plants.

The number of at least 2-fold induced or 2-fold repressed genes, was 67 and 58 for Comparison A, 53 and 64 for Comparison B, 7 and 41 for Comparison C, as well as 53 and 55 for Comparison D, respectively (Fig. 2). In addition, 24 differentially expressed genes were found when comparing different genotypes with the same treatment (Comparison A and B), and 12 were identified when comparing identical genotypes (F7 or $F7^+$) with different treatments (Comparison C and D). RGAs *pic19* and *pic13* were not differentially expressed in all four comparisons.

Based on automatic genome annotation (Schoof and Karlowski 2003), 34-44% of the differentially expressed genes within the four comparisons were not classified, such as homologues of disease resistance protein AIG1, zinc finger transcription factors ZF1, and ZFP2. The second largest category (16-27% of differentially expressed genes) was "cell rescue, defense, cell death and ageing", including homologues of PR1, Cytochrome P450, a thaumatin-like gene, and several heat shock protein (HSP) genes. Further ranking of classification categories was "metabolism"(10%), "transcription"(10%), "signal transduction"(5%) for Comparison A; "transcription"(12%), "signal transduction"(9%), "metabolism"(5%) for Comparison B; "metabolism"(17%),

Table 2. Putative mapped SSH-ESTs co-localized with major *Scmv* QTLs

GA[a]	Annotation[b]	Similarity[b]	Ratio[c]	Bin[d]	Comparison[e]			
					A	B	C	D
***Scmv 1* region**								
AA661457	gb|AAK58690.1|receptor-like kinase Xa21-binding protein 3 {Oryza sativa}	24%	3.3, 5.0	6.01	+[f]	+		
AA072467	PIR|T06273|benzothiadiazole-induced protein {Triticum aestivum}	10%	2.5, 5.0	6.01	+	+		
AI622698	GP|4514655|dbj|BAA75493.1||AB024058 IDS3 {Hordeum vulgare}	43%	2.3, 2.1	6.01			+	+
AI649716	SP|Q43255|Cytochrome P450 71C2{Zea mays}	100%	2.1	6.01			+	
AW216016	dbj|BAB92879.2|{Oryza sativa}	16%	-2.5	6.01				+
AI881213	SP|Q9NVW2|RING finger protein 12 {Homo sapiens}	5%	-2.5	6.02		+		
***Scmv 2* region**								
AI881934	SP|P24631|17.5 kDa class II heat shock protein{Zea mays}	100%	8.7	3.04				+
AW181245	dbj|BAB13743.1|pseudo-response regulator 4 {Arabidopsis thaliana}	4%	8.5	3.05		+		
AW257929	gb|AAL62060.1|RAD21-3 {Arabidopsis thaliana}	7%	7.2	3.05				+
AI600827	dbj|BAB44108.1|{Oryza sativa}	71%	6.0	3.04				+
AW017851	emb|CAA11391.1| phytase {Zea mays}	44%	5.5	3.05				+
AI857233	dbj|BAB21196.1|{Oryza sativa}	26%	3.7	3.04		+		
BG841140	emb|CAB40376.1|adenosine kinase {Zea mays}	50%	2.5	3.04		+		
AI621535	gb|AAC98962.1|nucleic acid binding protein {Oryza sativa}	61%	2.4	3.04		+		
AI737330	EGAD|65434|troponin I (TNI) (wings apart-a protein) {Drosophila melanogaster}	7%	2.1	3.04		+		
AW119949	dbj|BAB90212.1|{Oryza sativa}	95%	2.1	3.05		+		
BG841288	dbj|BAB62599.1|{Oryza sativa}	98%	-2.0	3.04		+		
AI001320	Unknown		-10.0	3.04				+
	Sum				2	10	2	7

[a] Genbank accession number.[b] Annotation of each SSH-EST sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml).[c] If the ratio is less than one, the negative reciprocal is listed. [d] Mapping information is from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins (Gardiner et al. 1993).[e] Four experiments on differential gene expression were conducted: Comparison A (constitutive genetic discrepancy), Comparison B (inducible genetic discrepancy), Comparison C (compatible interaction) and Comparison D (incompatible interaction).[f] The mapped SSH-EST was identified from this comparison.

"transcription"(8%), "cell organization"(6%) for Comparison C and "transcription"(9%), "signal transduction" (7%), "metabolism"(7%) for Comparison D. Most genes classified under signal transduction are known or suspected protein kinases, and most genes classified under transcription are known or suspected DNA binding proteins and transcription factors.

Digital northern analysis

Based on the assumption that the number of EST clones is proportional to the abundance of the mRNA (Audic and Claverie 1997) and most maize EST collections are derived from non-normalized cDNA libraries, the abundance of identified genes were analyzed by "digital northern analysis" (Ohlrogge and Benning 2000). Fernandes et al. (2002) considered maize genes with fewer than five ESTs to represent rarely transcribed genes. 143 (47%) differentially expressed genes are in the rarely transcribed category. Only 4 (1 %) of the genes (AA143906, T18435, AW216194, W49435) are represented by more than 100 ESTs and can be classified as abundantly transcribed. The remaining 154 (51 %) genes correspond to genes transcribed at moderate rate (6-88 ESTs).

Putative map position and annotation of differentially expressed genes

Altogether 19% (60 of 302) of the identified ESTs have been previously assigned to 29 bins (Gardiner et al. 1993) distributed over all 10 maize chromosomes (Fig. 3).

A total of 31% (18 of 58) of the mapped genes were located in bin 3.04-3.05 (12) and bin 6.00-6.02 (6) (Table 2). Ten mapped genes (seven in bin 3.04-3.05, three in bin 6.00-6.02) were revealed in Comparison B, two genes (bin 6.00-6.02) in Comparison A, two genes (bin 6.00-6.02) in Comparison C, and seven genes (five in bin 3.04-3.05; two in bin 6.00-6.02) in comparison D. Only 7 of these 18 genes were more than 50% similar to a protein sequence based on tentative annotation from TIGR Gene Index. For four genes, annotation (AI600827, AI621535, BG841288, and AW119949) was based on proteins from the *Arabidopsis* or rice genome. Of all three remaining genes, AI881934 (17.5 kDa class II heat shock protein) and AI649716 (Cytochrome P450 71C2) were defense-related genes while BG841140 (adenosine kinase) was an unclassified protein.

In addition, BM074098 co-localized with QSCM on chromosome 1 (Fig. 3). Twelve genes co-localized with 7 maize pathogenesis-related genes (Ramalingam et al. 2003) on chromosomes 1, 2, 7, 8, such as AA072453 and AW225208 (hsp26, Ibp2); BG841229, BI361048 and T18839 (SOD); AW052909 and BE345442 (px1); AI942048 and W21768 (prp2); BM074374 (cyp6); BM080630 and T18435 (hsp18c).

Hierarchical clustering of genes in association with SCMV

Hierarchical clustering (Eisen et al. 1998) of 302 genes revealed at least 13 distinct clusters (Fig. S1). Clusters 3, 10, and 13 displayed unique expression patterns in Comparison A, clusters 1, 5, and 9 in Comparison B, cluster 12 in Comparison C, and clusters 7 and 11 in Comparison D. Clusters 2 and 4 showed increased expression level in Comparisons A to D, while clusters 6 and 8 showed reduced expression in Comparisons A to D. No obvious functional relationship was found among genes within each of the 13 clusters.

Discussion

High throughput identification of low-abundance pathogenesis-related genes combining SSH and macroarray

According to Maize EST statistics in the Genbank (31/12/2003 assembly), the distribution of high-, medium-, and low-abundant ESTs was 3%, 64%, and 33%, respectively. In contrast, "digital northern" analysis revealed a shift towards medium- and low-abundant ESTs within our SSH libraries with 51% and 47%, respectively. Thus, the SSH procedure enriched the library for low-abundant and differentially expressed mRNAs by normalization (Diatchenko et al. 1996). Otherwise, abundant pathogenesis-related transcripts (e.g., genes coding for PR proteins) would very likely have masked important SCMV-specific transcripts expressed at much lower levels.

Transcript profiling methods can be divided into two types of analysis: (1) direct analysis, including procedures involving nucleotide sequencing (EST sequencing, SAGE, SSH) and fragment sizing (e.g., cDNA-AFLP); and (2) indirect analysis (macro- or microarray based expression profiling), involving nucleic acid hybridization of mRNA or cDNA fragments (Donson et al. 2002). EST sequencing is compatible with ongoing large-scale sequencing projects, but statistically significant coverage can

be expensive. SAGE is less expensive than EST sequencing, but it requires a comprehensive reference database and short tags can be redundant in databases. cDNA-AFLP methodology has lower set-up costs as SSH. However, extensive band isolation and sequencing are time-consuming. Alternatively, maize cDNA microarrays are publicly available but cost more than $200 per slide and cover less than 30% of the maize genome (Martienssen et al. 2004). In comparison to the 7,500 genes spotted to the publicly available microarray (http://www.plantgenomics.iastate.edu/maizechip/) 74% (224 of 302) genes identified in our study by SSH were lacking. In addition, 12 (4%) identified genes had a poor match (E value>10^{-5}) to the 306,218 maize EST entries in the Genbank (07/04/2004 assembly). In conclusion, the SSH procedure revealed genes of particular interest and offered a low cost alternative based on macroarrays for studying the maize–SCMV pathosystem compensating for a so far not available comprehensive maize microarray.

Microarray technology has undergone a rapid development in the last few years (Holloway et al. 2002) and in the present study allowed screening of cDNA clones from five SSH libraries reliably and at high speed. Three layers of replications were utilized in this study (Churchill 2002), involving biological and technical replicates as well as duplicated spotting of clones on the macroarray. Averaging duplicated spots and combining data from two technical replicates resulted in high reproducibility (>94%) between biological replicates, obligatory for drawing conclusions from expression profiling experiments (Churchill 2002).

Among the four comparisons, the number of differentially expressed genes was 125 (Comparison A), 117 (Comparison B), 108 (Comparison D), 48 (Comparison C). Fewer differentially expressed genes in comparison C were in agreement with the lack of a SSH library constructed especially for this comparison, whereas SSH library 1 was constructed for Comparison D, libraries 2 and 3 for Comparison B, and libraries 4 and 5 for Comparison A (Table 1). Thus, the SSH procedure successfully enriched for clones of genes expressed in the respective tester mRNA populations (Diatchenko et al. 1996).

The distribution of genes among functional classes was similar for all four comparisons (Fig. 2). Except for the category “metabolism”, most of the classified genes belonged to three pathogenesis-related categories, “cell rescue, defense, cell death and ageing”, “signal transduction”, and “transcription”, which accounted for 56-66% of classified genes. Recently, 78% of classified genes elicited by diverse RNA viruses were classified into these three categories in susceptible Arabidopsis (Whitham et al. 2003), whereas without biotic stresses 37% for Arabidopsis guard cell and 31% for mesophyll cell (Leonhardt et al. 2004).

Molecular mechanisms of maize-SCMV interaction

An unusually high frequency of genes conferring recessive resistance has been observed in relation to potyviruses (40% versus 20% for resistance against other viruses), in which the plant lacks one or more factors required for virus replication or movement (Provvidenti and Hampton 1992). Susceptibility to some potyviruses is associated with the host translation initiation factors eIF4E and eIF(iso) 4E, components of eIF4F and eIF(iso)4F, which may interact with the VPg protein that is covalently linked to the 5’ end of potyviral RNA genomes (Whitham and Wang 2004). However, resistance genes *Scmv1* (*Scmv1a*, *Scmv1b*), and *Scmv2* displayed at least partial dominance in different studies (Xia et al. 1999, Dussle et al. 2000, Yuan et al. 2003).

So far, two major molecular mechanisms for active resistance response to viruses are known, hypersensitive response (HR) and RNA silencing (Carrington and Whitham 1998). HR or extreme resistance (ER) is triggered by dominant or semi-dominant resistance (R) genes at initial infection sites, occurring most often in a strain-specific or “gene-for-gene” manner. Four cloned virus R-genes (*N*, *Rx1*, *Rx2* and *Sw-5*) belong to the nucleotide binding, leucine-rich repeat (NBS-LRR) super-family of R-genes (Whitham et al. 1994, Bendahmane et al. 1999, Bendahmane et al. 2000, Brommonschenkel et al. 2000). However, no HR symptoms are observed on maize leaves infected with SCMV. Moreover, no differential expression occurred in the two RGAs *pic19* and *pic13*, supposed to be respective candidate genes of *Scmv2* and *Scmv1*, as it was found for *Rx2* four days after inoculation (Bendahmane *et al.*, 2000). Some resistant host plants can recognize viral nucleic acids via RNA silencing and customize a sequence-specific response (Vance and Vaucheret 2001). Thus, virus-induced silencing can limit virus accumulation, promote recovery (in some cases) from a systemic infection, and confer resistance

to secondary infections with the same or homologous viruses (Voinnet 2001). In contrast to resistance triggered by the NBS-LRR–type R genes, resistance through silencing appears not to depend on a gene-for-gene recognition event and occurs without HR (Whitham et al. 2000). Infection of *Nicotiana clevelandi* by the tomato blackring nepovirus (TBRV) strain W22 resulted in an initial symptomatic phase in which the virus moves systemically, followed by a recovery state in which new tissue developing post-inoculation is asymptomatic and largely devoid of virus (Ratcliff et al. 1997). In addition, resistance of *Arabidopsis* to tobacco etch potyvirus (TEV), involving two dominant genes *RTM1* and *RTM2*, are not associated with HR or ER (Chisholm et al. 2001). In this case, resistance is not caused by activation of known defense pathways but appears to be due to interference with long-distance movement of TEV through sieve elements. In conclusion, since SCMV resistance is conferred by (partially) dominant resistant genes not leading to HR, more than one of the above described mechanisms might be involved at the level of long-distance transport of virus (Lei and Agrios 1986).

Plant defense systems can be classified as either constitutive or inducible, depending on whether they are pre-existing features of the plant, or are switched on after challenge. Passive protection against plant virus vectors (in case of SCMV: aphids) is provided by anatomical barriers (such as the cuticle and cell wall) and, more importantly, virus infectivity is reduced by preformed antiviral compounds within cells (Lucas and Dickinson 1998). Increasing evidence shows that constitutively expressed genes encoding enzymes associated with normal plant metabolism play critical roles in the induction of plant defenses against viruses (Eckardt 2004). Recently, it was shown that plant SNF1 kinase, otherwise involved in regulating the carbon metabolism, is associated with viral defense (Hao et al. 2003). Furthermore, a plant adenosine kinase, controlling the flux through the S-adenosyl-L-Met–dependent methylation cycle, also played a crucial role in viral defense (Hao et al. 2003, Wang et al. 2003). Comparison A revealed several putative preformed inhibitors. CD437477, BQ486978, and BM074098 are the homologs encoding chitinases (Datta and Muthukrishnan 1999), which can hydrolyze the cell wall of many fungi and the exoskeleton of invertebrates. AA661457 and AI600506 show homology to genes encoding receptor-like protein kinases (Morris and Walker 2003) and AI820401 to a protein kinase C inhibitor. Ribosome-inactivating proteins (RIPs) can block virus replication via affecting ribosomes in host cells (Lucas and Dickinson 1998). This type of protein or other known antiviral proteins were not found in Comparison A. This is, however, in agreement with the finding that SCMV can be detected and, thus, replicate in primary infected leaves of resistant genotypes (Louie 1995). Differentially expressed pathogen-related genes, identified from Comparison B (inducible genetic discrepancy), such as BM337818 (pathogenesis-related protein 1, PR1), BG841140 (adenosine kinase) and BM660017 (Dnaj protein homolog ZMDJ1), have been found together with HR or RNA silencing (Whitham et al. 2003).

Gene expression profiles of incompatible interactions, including TMV in tomato and *Chenopodium* (Cooper 2001, Golem and Culver 2003), revealed similarities at the gene level with Comparison D, such as for AI943646 (cytochrome b245) and BM379389 (thioredoxin H-type). Most likely resistant plants utilize a common mechanism for defense against virus attack (Matthews and Hull 2002). In the compatible interaction (Comparison C), the distribution of genes among functional classes looked similar to the incompatible interaction. It is consistent with the hypothesis that viruses induce defense response both in susceptible and resistant plants at early stages (Matthews and Hull 2002). Whitham et al. (2003) reported that diverse RNA viruses, including cucumber mosaic cucumovirus, oil seed rape tobamovirus, turnip vein clearing tobamovirus, potato virus X potexvirus and turnip mosaic potyvirus, elicited the expression of common sets of genes in susceptible *Arabidopsis*. 72% (33 of 47) of the differentially expressed genes in Comparison C can be found in this common set of genes, such as BM501006 (F5M15.13), AI820398 (AIG2 protein homolog F8F16.130), while the remaining genes without annotation could be maize-specific. Technical advances in sampling tissue in individual cell types at specific stages of viral infection, such as LCM (laser-capture microdissection), will enable us to better understand molecular mechanisms by increasing the spatial and temporal resolution of expression studies (Schnable et al. 2004).

Co-localization of differentially expressed genes and *Scmv* QTL

Two RGAs (*pic13* and *pic19*) (Quint et al. 2003), supposed to be candidate genes of *Scmv2* and *Scmv1*, respectively were not differentially expressed in all four comparisons. In addition,

none of the identified genes showed homology to NBS-LRR resistance genes. This is in agreement with several recent DNA microarray experiments, showing no dramatic transcriptional regulation of most known R genes during plant-pathogen interactions. This may be due to very low transcript levels of many R genes, and a limited sensitivity of current DNA microarray techniques for detecting low-abundance transcripts and their changes (Wan et al. 2002). Alternatively, these genes might be not pathogen-induced but expressed constitutively.

Positional cloning is the major approach used to characterize genes underlying QTL, but it is very laborious and time consuming. Consequently, there are so far no reports about positional cloned maize QTL. Transposon mutagenesis was used to isolate *Tb1* cosegregating with a QTL affecting apical dominance in maize. Thus, the candidate-gene approach may provide an alternative for pinpointing genes underlying resistance QTL in maize, especially in view of the planned sequencing of major parts of the genome (Martienssen et al. 2004). In addition, gene expression studies comparing NILs differing for short chromosome segments offer an alternative to identify candidate genes for QTL located within such segments (Borevitz and Chory 2004).

Despite the limitations of *in silico* mapping, such as duplicated genomes of maize, multi-copy genes, it provides the opportunity to validate data without further experimentation and facilitate the identification of candidate genes. In this study, we investigated NILs F7 and $F7^+$ to conduct expression profiling experiments. Ten differentially expressed genes are co-localized with *Scmv* QTL in Comparison B, whereas only two genes in Comparison A. This was consistent with the hypothesis that differentially expressed genes derived from the *Scmv1* or *Scmv2* genome regions might be candidate genes for the previously mapped QTL or at least be involved in SCMV resistance. Two mapped genes (AA661457 and AA072467) in Comparison A were located within the *Scmv1* region and were stronger induced in Comparison B (5.0) than A (3.3). AA661457 is the homolog of gene encoding receptor-like kinase Xa21-binding protein 3, which is a member of the leucine- rich repeat, kinase (LRR-TM-Kinase) super-familiy of R-genes (Hulbert et al. 2001), and thus, a promising candidate gene for *Scmv1*. In the vicinity of the *Scmv2* region, six mapped genes (AW181245, AI857233, BG841140, AI621535, AI737330 and AW119949) were up-regulated in Comparison B. AW181245 (pseudo-response regulator 4) was expressed over eight fold, whereas BG841140 (adenosine kinase) and AI621535 (nucleic acid binding protein), known to be pathogenesis-related genes, were induced at a low level.

Other mapped genes were co-localized with *Scmv* QTL on chromosomes 3 and 6, including two down-regulated genes (AI881213, BG841288) in Comparison B, eight genes from Comparison C and D. Those genes and 69% of the mapped genes, located in other genome regions, may act further downstream in the signal transduction pathway and may be induced by resistance genes located in the *Scmv1* and / or *Scmv2* regions. This has implications in view of candidate gene identification and application of results from differential expression studies in plant breeding. The majority of differentially expressed genes originated from monomorphic regions between both isogenic lines F7 and F7+. Therefore, these genes can't be employed in conventional breeding for SCMV resistance, since this requires "positive" and "negative" alleles for either phenotypic or marker-assisted selection. However, some of the identified differential expression of genes outside the *Scmv1* and *Scmv2* regions may be due to undetected residual heterozygosity, although F7+ has been derived from F7 after seven backcross generations. This is most likely for clusters of differentially expressed genes outside the *Scmv1* and *Scmv2* regions.

With regard to the *Scmv1* and *Scmv2* regions, the expectation so far was the presence of 1-2 resistance genes within each of these chromosome segments. The identification of a larger number of differentially expressed genes mapping to these two regions can be explained by (i) the presence of polymorphisms between F7 and F7+ unrelated to SCMV resistance, or (ii) clustering of genes involved in SCMV resistance in these two regions. Since a substantially larger number of these genes showed differential expression only after SCMV infection, both the *Scmv1* and the *Scmv2* regions appear to harbor clusters of genes responding to SCMV infection. This would complicate map-based gene isolation of respective *Scmv* resistance genes and optimization of complex loci for plant breeding. The question of *Scmv* resistance gene clusters in both regions is currently addressed by studying $F7^+$ derived recombinant sub-lines (Lübberstedt, unpublished).

Acknowledgements

We thank Olaf Neuschaefer-Rube and Michael Goetz for spotting macroarrays at GSF (Neuherberg). Critical comments on the manuscript from Volker Mohler and Gerhard Schwarz are appreciated. This work was supported by the German research foundation (DFG LU601/6-1).

Reference

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7: 986-95.

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2: 96-103.

Bendahmane A, Kanyuka K, Baulcombe DC (1999) The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell 11: 781-92.

Bendahmane A, Querci M, Kanyuka K, Baulcombe DC (2000) *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: application to the *Rx2* locus in potato. Plant J 21: 73-81.

Birch PRJ, Avrova AO, Duncan JM, Lyon GD, Toth RL (1999) Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. Mol Plant Microbe Interact 12: 356-361.

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7: 132-6.

Brommonschenkel SH, Frary A, Tanksley SD (2000) The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene Mi. Mol Plant Microbe Interact 13: 1130-8.

Carrington JC, Whitham SA (1998) Viral invasion and host defense: strategies and counter-strategies. Curr Opin Plant Biol 1: 336-341.

Chisholm ST, Parra MA, Anderberg RJ, Carrington JC (2001) Arabidopsis *RTM1* and *RTM2* genes function in phloem to restrict long-distance movement of tobacco etch virus. Plant Physiol 127: 1667-75.

Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet 32: 490-495.

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 11: 968-78.

Cooper B (2001) Collateral gene expression changes induced by distinct plant viruses during the hypersensitive resistance reaction in *Chenopodium amaranticolor*. Plant J 26: 339-349.

Datta SK, Muthukrishnan S (1999) Pathogenesis-related proteins in plants. CRC Press, Boca Raton.

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93: 6025-30.

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol 48: 75-97.

Dussle CM, Melchinger AE, Kuntze L, Stork A, Lübberstedt T (2000) Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTL contributing to SCMV resistance in maize. Plant Breeding 119: 299-303.

Dussle CM, Quint M, Melchinger AE, Xu ML, Lübberstedt T (2003) Saturation of two chromosome regions conferring resistance to SCMV with SSR and AFLP markers by targeted BSA. Theor Appl Genet 106: 485-93.

Eckardt NA (2004) Aminotransferases confer "enzymatic resistance" to downy mildew in melon. Plant Cell 16: 1-3.Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863-8.

Fernandes J, Brendel V, Gai X, Lal S, Chandler VL, Elumalai RP, Galbraith DW, Pierson EA, Walbot V (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. Plant Physiol 128: 896-910.

Fuchs E, Gruntzig M (1995) Influence of Sugarcane Mosaic-Virus (Scmv) and Maize-Dwarf Mosaic-Virus (Mdmv) on the Growth and Yield of 2 Maize Varieties. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection 102: 44-50.

Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S (1993) Development of a core RFLP map in maize using an immortalized F2 population. Genetics 134: 917-30.

Golem S, Culver JN (2003) Tobacco mosaic virus induced alterations in the gene expression profile of *Arabidopsis thaliana*. Mol Plant Microbe Interact 16: 681-688.

Hao L, Wang H, Sunter G, Bisaro DM (2003) Geminivirus AL2 and L2 proteins interact with and inactivate SNF1 kinase. Plant Cell 15: 1034-48.

Holloway AJ, van Laar RK, Tothill RW, Bowtell DDL (2002) Options available - from start to finish - for obtaining data from DNA microarrays. Nat Genet 32: 481-489.

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39: 285-312.

Kuntze L, Fuchs E, Gruntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early-maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breeding 116: 499-501.

Lei JD, Agrios GN (1986) Mechanisms of Resistance in Corn to Maize-Dwarf Mosaic-Virus. Phytopathology 76: 1034-1040.

Leonhardt N, Kwak JM, Robert N, Waner D, Leonhardt G, Schroeder JI (2004) Microarray expression analyses of *Arabidopsis* guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. Plant Cell 16: 596-615.

Louie R (1995) Vascular Puncture of Maize Kernels for the Mechanical Transmission of Maize White Line Mosaic-Virus and Other Viruses of Maize. Phytopathology 85: 139-143.

Lübberstedt T, Dussle C, Melchinger AE (1998) Application of microsatellites from maize to teosinte and other relatives of maize. Plant Breeding 117: 447-450.

Lucas JA, Dickinson CH (1998) Plant pathology and plant pathogens. Blackwell Science, Malden, MA.

Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie WR (2004) Sequencing the maize genome. Curr Opin Plant Biol 7: 102-7.

Matthews REF, Hull R (2002) Matthews' plant virology. Academic Press, San Diego.

Morris ER, Walker JC (2003) Receptor-like protein kinases: the keys to response. Curr Opin Plant Biol 6: 339-42.

Nadimpalli R, Yalpani N, Johal GS, Simmons CR (2000) Prohibitins, stomatins, and plant disease response genes compose a protein superfamily that controls cell proliferation, ion channel regulation, and death. J Biol Chem 275: 29579-86.

Ohlrogge J, Benning C (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol 3: 224-8.

Provvidenti R, Hampton RO (1992) Sources of resistance to viruses in the Potyviridae. Arch Virol Suppl 5: 189-211.

Quackenbush J (2002) Microarray data normalization and transformation. Nat Genet 32: 496-501.

Quint M, Dussle CM, Melchinger AE, Lübberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS. Theor Appl Genet 106: 1171-7.

Quint M, Mihaljevic R, Dussle M, Xu L, Melchinger E, Lübberstedt T (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor Appl Genet 105: 355-363.

Ramalingam J, Cruz CMV, Kukreja K, Chittoor JM, Wu JL, Lee SW, Baraoidan M, George ML, Cohen MB, Hulbert SH, Leach JE, Leung H (2003) Candidate Defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. Mol Plant Microbe Interact 16: 14-24.

Ratcliff F, Harrison BD, Baulcombe DC (1997) A similarity between viral defense and gene silencing in plants. Science 276: 1558-1560.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374-8.

Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (Oryza sativa L.) and maize (Zea mays L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J 38: 396-409.

Schnable PS, Hochholdinger F, Nakazono M (2004) Global expression profiling applied to plant development. Curr Opin Plant Biol 7: 50-6.

Schoof H, Karlowski WM (2003) Comparison of rice and *Arabidopsis* annotation. Curr Opin Plant Biol 6: 106-12.

van Berloo R (1999) GGT: Software for the display of graphical genotypes. J Hered 90: 328-329.

Vance V, Vaucheret H (2001) RNA silencing in plants--defense and counterdefense. Science 292: 2277-80.

Voinnet O (2001) RNA silencing as a plant immune system against viruses. Trends Genet 17: 449-59.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93: 77-8.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-14.

Wan J, Dunning FM, Bent AF (2002) Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. Funct Integr Genomics 2: 259-73.

Wang H, Hao L, Shung CY, Sunter G, Bisaro DM (2003) Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. Plant Cell 15: 3020-32.

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78: 1101-15.

Whitham SA, Anderberg RJ, Chisholm ST, Carrington JC (2000) Arabidopsis *RTM2* gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. Plant Cell 12: 569-82.

Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang X, Hou YM (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. Plant J 33: 271-83.

Whitham SA, Wang Yz (2004) Roles for host factors in plant viral pathogenicity. Curr Opin Plant Biol 7: 1-7.

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89: 660-667.

Xu ML, Melchinger AE, Xia XC, Lübberstedt T (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. Mol Gen Genet 261: 574-81.

Yang GP, Ross DT, Kuang WW, Brown PO, Weigel RJ (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucleic Acids Res 27: 1517-23.

Yuan L, Dussle CM, Melchinger AE, Utz HF, Lübberstedt T (2003) Clustering of QTL conferring SCMV resistance in maize. Maydica 48: 55-62.Zhu T, Budworth P, Chen WQ, Provart N, Chang HS, Guimil S, Su WP, Estes B, Zou GZ, Wang X (2003) Transcriptional control of nutrient partitioning during rice grain filling. Plant Biotechnology J 1: 59-70.

Comparison of Transcript Profiles between Near-Isogenic Maize Lines in Association with SCMV Resistance Based on Unigene-Microarrays

Chun Shi[1], Fritz Thümmler[2], Albrecht E. Melchinger[3], Gerhard Wenzel[1], Thomas Lübberstedt[4,*]

[1]Chair of Plant Breeding, Technical University of Munich, Am Hochanger 2, 85350 Freising, Germany
[2]Vertis Biotechnologie AG, Lise-Meitner-Straße 30, 85354 Freising, Germany
[3]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
[4]Department of Plant Biology, Research Centre Flakkebjerg, 4200 Slagelse, Denmark

* For correspondence (tel. +45 5811 3484; fax +45 5811 3301; e-mail Thomas.Luebberstedt@agrsci.dk).

Summary

The molecular mechanisms underlying the development and progression of sugarcane mosaic virus (SCMV) infection in maize are poorly understood. A transcript profiling study based on maize unigene-microarrays was conducted to identify genes associated with SCMV resistance in the near isogenic line (NIL) pair $F7^+$ (SCMV resistant) and F7 (susceptible). Altogether, 497 differentially expressed genes were identified in four comparisons addressing constitutive genetic discrepancy, inducible genetic discrepancy, compatible reaction, and incompatible reaction. Compared to a suppression subtractive hybridization (SSH) approach on the same materials, expression patterns of microarray-ESTs and SSH-ESTs were consistent for the same comparisons despite of technical discrepancies. Since pathogen-induced transcripts were underrepresented on the unigene-microarray, fewer microarray-ESTs (45.8%) were classified into pathogenesis-related categories than SSH-ESTs (60.5%). Moreover, fewer microarray-ESTs (4) co-segregated with *Scmv* QTL than SSH-ESTs (18). Therefore, our results demonstrate that SSH-macroarray complements incomprehensive microarrays. Good candidates genes (CGs) associated with SCMV resistance can be chosen from three classes: i) positional CGs co-localized with major *Scmv* QTL, ii) functional CGs showing the homology to pathogenesis-related genes, or iii) differentially expressed ESTs showing consistent expression pattern in both approaches.

Introduction

Sugarcane mosaic virus (SCMV) is an important pathogen of maize (*Zea mays L.*) in Europe and China, causing substantial yield losses in susceptible cultivars (Fuchs and Gruntzig, 1995). In previous studies, Kuntze et al. (1997) screened 122 early-maturing European inbred lines for resistance to potyviruses SCMV and MDMV (maize dwarf mosaic virus) and identified only three inbreds (D21, D32, and FAP1360A) displaying complete resistance under both field and greenhouse conditions. Two major QTL regions, *Scmv1* and *Scmv2*, conferring resistance to SCMV were mapped to chromosome arms 6S and 3L. In cross D145 × D32 quantitative trait locus (QTL) analysis (Xia et al., 1999) and in cross F7 × FAP1360A bulked segregant analysis (BSA) (Xu et al., 1999) and QTL analysis (Dussle et al., 2000) were applied. The *Scmv1* region contains a minimum of two QTL (*Scmv1a* and *Scmv1b*) (Yuan et al., 2003). Together with three additional minor QTL identified on chromosomes 1, 5, and 10 (Xia et al., 1999), a minimum of six genes are involved in the oligogenic inherited SCMV resistance.

An unusually high frequency of genes conferring recessive resistance has been observed in relation to potyviruses (40% versus 20% for resistance against other viruses), in which the plant lacks one or more factors required for virus replication or movement (Provvidenti and Hampton, 1992). However, resistance genes *Scmv1* (*Scmv1a*, *Scmv1b*), and *Scmv2* displayed at least partial dominance in different studies (Dussle et al., 2000, Xia et al., 1999, Yuan et al., 2003). Moreover, no HR symptoms are observed for maize leaves infected with SCMV. The defense mechanism without HR applying to SCMV resistance is poorly understood.

Positional cloning is the major approach used to characterize genes underlying QTL, but it is very laborious and time consuming. The candidate-gene approach provides an alternative for pinpointing genes underlying SCMV resistance, especially in view of the planned sequencing of major parts of the

genome (Martienssen et al., 2004). Maize resistance gene analogues (RGA) involved in initial pathogen recognition, were chosen as starting point for isolation of genes conferring SCMV resistance (Collins et al., 1998). Mapping of RGAs in relation to *Scmv1* and *Scmv2* suggested that RGA *pic19* is a candidate for *Scmv1* and *pic13* for *Scmv2* (Quint et al., 2002), which is currently investigated in more detail (Lübberstedt et al., unpublished results). Gene expression studies comparing NILs differing for short chromosome segments offer an alternative to identify candidate genes for QTL located within such segments (Borevitz and Chory, 2004), but also genes from other chromosomal locations involved in subsequent steps leading to resistance or susceptibility after the initial recognition of SCMV.

Microarray-based expression profiling methods, together with genomic and/or EST (expressed sequence tag) sequence data resulted in significant progress in characterization of plant pathogenesis-related responses (Wan et al., 2002). Baldwin et al. (Baldwin et al., 1999) identified 117 genes that consistently showed altered mRNA expression in maize 6 h after various treatments with the fungal pathogen *Cochliobolus carbonum*, using a maize DNA microarray representing 1,500 maize genes. Using a similar approach, Nadimpalli et al. (Nadimpalli et al., 2000) identified nearly 70 genes having a more than twofold change in mRNA abundance in the lesion mimic maize mutant, *Les9* compared to wild-type plants.

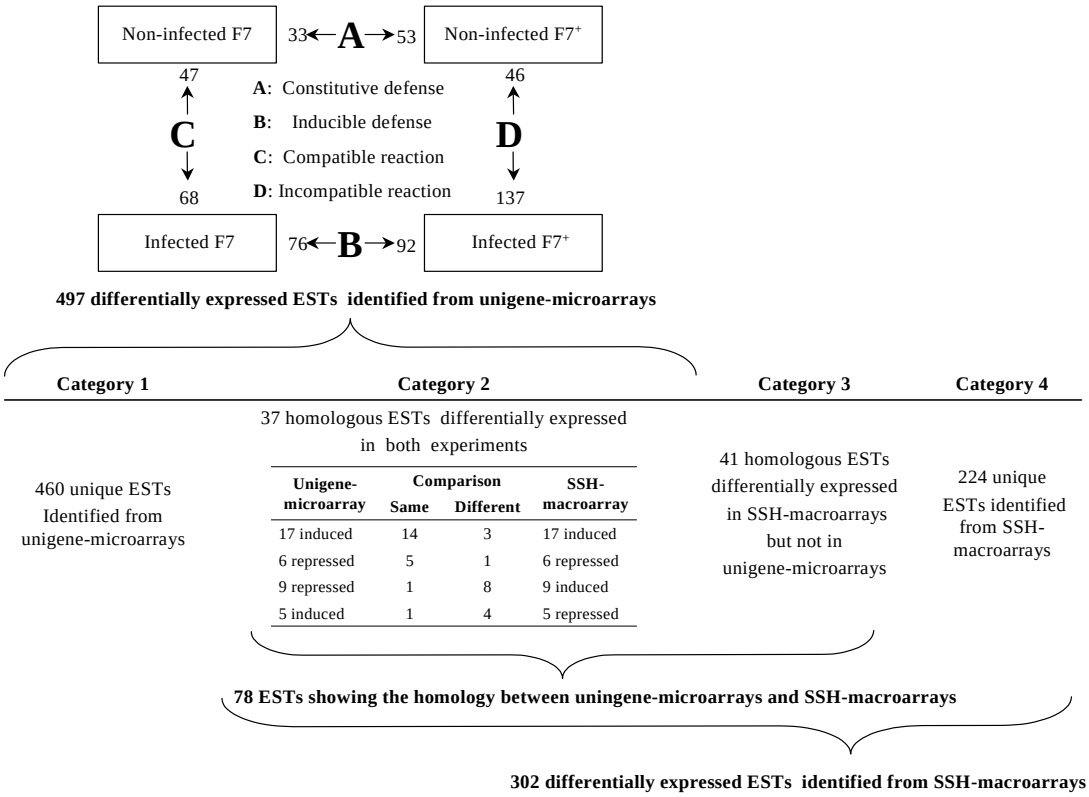


Fig. 1. Result comparison of unigene-microarray and SSH-macroarray experiments.
In total, differentially expressed ESTs, identified from unigene-microarrays and SSH-macroarrays, were classified into four categories. Category 1 and 2 shows differentially expressed ESTs identified from unigene-microarray experiments in four comparisons, Comparison A (constitutive defense), Comparison B (inducible defense), Comparison C (compatible reaction) and Comparison D (incompatible reaction). The number indicates up-regulated gene numbers identified in each probe (arrow head) from each experiment (arrow tail – there are no numbers at arrow tails). Category 2 and 3 contains the ESTs showing the homology in both approaches. The differentially expressed ESTs identified from SSH-macroarrays were showed in Category 2, 3 and 4.

In this study, we used publicly available maize unigene-microarrays (Nakazono et al., 2003) containing 11,424 distinct ESTs, including 949 mapped sequences. In contrast to comprehensive microarrays in *Arabidopsis* and rice (Zhu, 2003), it actually contains 9,841 different unigenes, which account for only 20% of the about 50,000 maize genes (Martienssen et al., 2004). Thus, another expression profiling method, combining suppression subtractive hybridization (SSH) (Diatchenko et al., 1996) and macroarray hybridization, was conducted in a companion study to detect SCMV-related transcripts in maize (Shi et al., submitted). Combining both results will provide a more comprehensive understanding of SCMV-maize interaction.

The objectives of our study were to 1) identify genes associated with SCMV resistance in maize using unigene-microarray hybridization; 2) propose the molecular mechanisms underlying the development and progression of SCMV infection combining the results of SSH-macroarray and unigene-microarray experiments; 3) identify candidate genes underlying major *Scmv* QTL through comparing SSH-macroarray and unigene-microarray results.

Results

Identification of differentially expressed genes by unigene-microarrays

In this study, four comparisons addressing relevant biological questions were made (Fig. 1). Comparisons A and B included different genotypes with the same treatment, while Comparisons C and D included different treatments on the same genotype (either F7 or $F7^+$). Comparison A: non-infected F7 versus $F7^+$ (constitutive genetic discrepancy): structural and chemical barriers of the plant effectively exclude the majority of organisms. Comparison B: infected F7 versus $F7^+$ (inducible genetic discrepancy): if inducible genetic discrepancy of a plant is overcome, a sensitive surveillance system can detect foreign pathogens and trigger a rapid response to injury or virus attack. Comparison C: infected versus non-infected F7 (compatible reaction): virus replicates and moves systemically in cells of intact susceptible plants. Comparison D: infected versus non-infected $F7^+$ (incompatible reaction): virus multiplication is limited to initially infected cells of resistant plants.

The reproducibility of unigene-microarray experiments was high across all comparisons. In dye swap replications prepared from the same mRNA, over 90% of the ratios calculated from

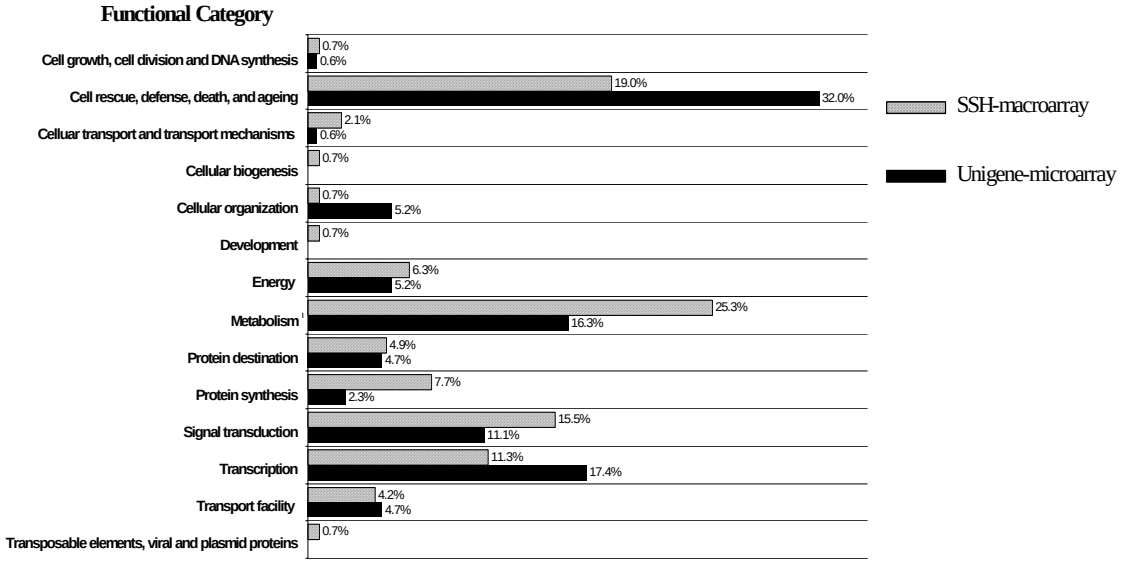


Fig. 2. Comparison of gene distribution in functional classes in unigene-microarray and SSH-macroarray experiments. Each EST was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. And unclassified microarray-ESTs (72%) and SSH-ESTs (42%) didn't been taken into account.

technical replications varied by less than two-fold (Pearson correlation coefficient 0.87±0.03). However, when the Cy3 signal of slide 1 and the Cy5 signal of slide 2 were averaged (data set 1), and the Cy5 signal of slide 1 and the Cy3 signal of slide 2 were averaged (data set 2), more than 95% of the ratios varied by less than 1.5-fold (Pearson correlation coefficient 0.94±0.02). After averaging technical replications, more than 88% of all ratios varied by less than 1.5-fold (Pearson correlation coefficient 0.90±0.01 between biological replications).

In total, 497 ESTs were differentially expressed in one or more comparisons, which accounted for 4.1% of 11, 827 ESTs deposited on the unigene-microarray. The number of at least 2-fold induced ESTs, was 33 for non-infected F7 and 53 for non-infected $F7^+$ in Comparison A, 76 for infected F7 and 92 for infected $F7^+$ in Comparison B, 47 for non-infected F7 and 68 for infected F7 in Comparison C, as well as 46 for non-infected $F7^+$ and 137 for infected $F7^+$ in Comparison D (Fig. 1). In total 50.4% of these ESTs were induced more than 4-fold up to 25-fold.

The EST collection printed on the maize unigene-microarrray was derived from 17 EST libraries (see "Materials and Methods" for details). The discovery ratio of differentially expressed genes from EST libraries was 8.0% (111 of 1380), 9.0% (34 of 379), 4.4% (29 of 666), 4.8% (14 of 291), 3.4% (54 of 1598), 3.0% (71 of 2371), 2.6% (15 of 574), 0% of (0 of 15), 4.7% (12 of 258), 4.2% (16 of 383), 6.0% (64 of 1060), and 4.6% (132 of 2852) for 486, 496, 603, 605, 606, 614, 618, 683, 687, 707/945 and ISUM3/4/5/6/7, respectively.

In "digital northern analysis" (Audic and Claverie, 1997), 151 (30%) differentially expressed microarray-ESTs were in the rare transcript category with fewer than five sequencess in the public EST collection (Fernandes et al., 2002). 43 (9 %) of differentially expressed ESTs were represented by more than 100 ESTs and can be classified as abundantly transcribed, whereas the remaining 303 (61 %) corresponded to genes transcribed at a moderate rate (6 - 98 ESTs).

Comparison of SSH-macroarray and unigene-microarray results

We assigned to each differentially expressed SSH-EST to a Genbank accession number (GA) to identify respective microarray-ESTs based on BlastN hits (E-value < 0.0001) against the MaizeGDB EST database (Shi et al., submitted). In average, PCR fragments printed on microarrays were significantly ($P < 0.0001$) longer (499 bp) than SSH fragments (245 bp).

The comparison of unigene-microarray and SSH-ESTs is summarized in Fig. 1. 460 differentially expressed ESTs were exclusively present on microarrays (Category 1), and 224 on SSH-based macroarrays (Category 4). 78 differentially expressed ESTs were present both on micro- and macroarrays. A Bland-Altman plot (Bland and Altman, 1986) revealed no significant difference between both experiments. Among those, 37 homologous ESTs were differentially expressed both in unigene-microarray and SSH-macroarray experiments (Category 2): 17 ESTs were induced in both approaches with 14 in the same comparison and 3 in different comparisons; 6 ESTs were repressed in both approaches with 5 in the same comparison and 1 in different comparisons; 9 ESTs were repressed in unigene-microarray but induced in SSH-macroarray experiments with 1 in the same and 8 in different comparisons; 5 ESTs were repressed in SSH-macroarray but induced in unigene-microarray experiments with 1 in the same and 4 in different comparisons. 41 homologous ESTs differentially expressed in SSH-macroarray were not differentially expressed in unigene-microarray experiments (Category 3). If all 78 ESTs in Category 2 and 3 are taken into account expression patterns of unigene-microarray and SSH-macroarray experiments from the same comparisons were consistent (Fisher's exact test: $P = 0.0117$).

Although more differentially expressed ESTs were identified based on microarrays (497) than SSH-macroarrays (302), the efficiency of gene discovery, determined as the ratio between the number of differentially expressed to all spotted cDNAs, was much higher in SSH-macroarray (59%) (Shi et al., submitted) than in unigene-microarray experiments (4.1%). However, due to five-time redundancy in SSH clones revealed by sequencing (Shi et al. submitted), the actual efficiency of gene discovery by SSH-macroarray experiments was approximately 10%.

Classification of differentially expressed genes identified by unigene-microarrays and SSH-macroarrays

In total, more differentially expressed ESTs from the unigene-microarray experiment (72%) were unclassified than in the SSH-macroarray experiment (44%). Among classified ESTs, the largest category was "metabolism" (25.3%) in unigene-microarray

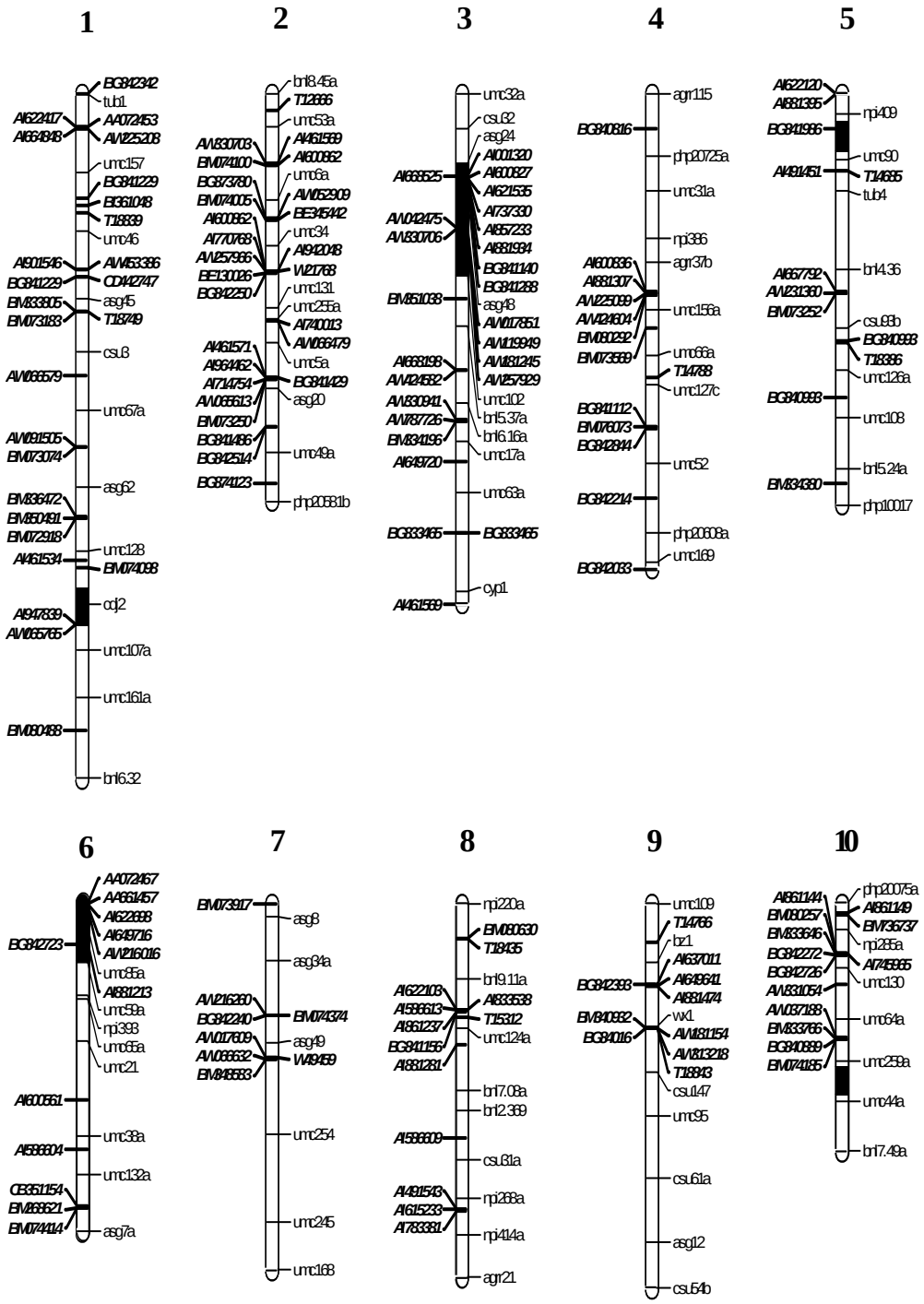


Fig. 3. Distribution of differentially expressed genes on maize chromosomes with respect to SCMV resistance identified in unigene-microarray and SSH-macroarray experiments.
Loci in *bold* and *italics* placed on the *left* side of each chromosome were mapped ESTs identified from unigene-microarray experiments. Loci placed on the *right* side of each chromosome are a set of core markers that defines a bin boundary (Gardiner et al., 1993), while the loci in *bold* and *italics* were mapped ESTs identified in SSH-macroarray experiments. EST mapping information was from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins. *Scmv1* is highlighted on chromosome 6, *Scmv2* on chromosome 3 and three minor *QSCMs* on chromosomes 1, 5, and 10 (Xia et al., 1999).

Table 1. The list of candidate genes (CGs) associated with SCMV resistance

GA[a]	Annotation[b]	Similarity[b]	Bin[c]	Comparison			
				A	B	C	D
Positional CGs mapping to bins 3.04 – 3.05 and 6.00 – 6.02							
AI668525	gb|AAF55168.1||AE003708 CG4913-PA {Drosophila melanogaster}	4%	3.04			-3.0[d]	
AW330706	Unknown		3.05				2.8
AW042475	dbj|BAB64692.1||AP003683 P0431G06.3 {Oryza sativa}	10%	3.05			-4.5	
BG842723	dbj|BAB78651.1||AP003022 P0681B11.18 {Oryza sativa}	7%	6.02	5.8			2.7
Functional CGs showing the homology to pathogenesis-related genes							
AI491543	gb|AAF68389.1 hypersensitive-induced response protein {Zea mays}	100%	8.07				2.8
AI621822	emb|CAA06925.1 Avr9 elicitor response protein {Nicotiana tabacum}	69%				6.0	
AI619128	dbj|BAB89081.1 dnaJ-like protein {Oryza sativa}	84%					-3.3
AI615100	SP|P33890 Cold shock induced protein TIR2 precursor – yeast	5%		3.0			
AI664862	PIR|S59544 stress-induced protein OZI1 precursor {Arabidopsis thaliana}	92%			3.5		
AI795699	emb|CAC21392.1 peroxidase {Zea mays}	100%			5.3		
AI999974	SP|P18123 Catalase isozyme 3 {Zea mays}	42%		6.9	7.2		
BM073434	PIR|T02055 pathogenesis related protein-5 {Zea mays}	98%		3.3			
Consistent ESTs in Category 2							
AI461569	PIR|S65781|S54179 acidic ribosomal protein P2 {Zea mays}	100%					2.5, 3.4[e]
AI600862	dbj|BAB93128.1||AP003196 beta-1,3-glucanase-like protein {Oryza sativa}	94%	2.04	-9.7, -10.0			
AI621758	dbj|BAB09296.1||AB011476 RNA-binding protein-like {Arabidopsis thaliana}	64%			4.8, 2.0		
AI649641	dbj|BAB16858.1||AP002537 P0001B06.11 {Oryza sativa}	70%					3.0, 2.9
AI665633	Unknown					9.0, 5.0	
AI714860	gb|AAM98096.1||AY139778 AT3g13690/MMM17_12 {Arabidopsis thaliana}	15%		-3.5, -10.0			
AI738263	gb|AAB49338.1 delta-24-sterol methyltransferase {Triticum aestivum}	40%		8.3, 5.8			
AI855243	Unknown				7.0, 5.3		
AI941971	PIR|S33633|S33633 ubiquitin / ribosomal protein CEP52 {Oryza sativa}	100%				5.8, 2.0	
AI942048	gb|AAK67147.1 nucleosome/chromatin assembly factor C {Zea mays}	25%				8.5, 5.3	
AI942105	dbj|BAC55693.1||AP004275 P0453E05.3 {Oryza sativa}	14%		9.7, 6.1			
AI974914	Unknown				-2.3, -1.9		
AW052909	gb|AAN08216.1||AC090874 ribosomal protein L15 {Oryza sativa}	100%				6.7, 5.0	
AW330660	Unknown						2.7, 2.1
AW331161	gb|AAO74140.1||AY228468 ORF64c {Pinus koraiensis}	100%					2.7, 3.1
AW438364	gb|AAL08230.1||AY056374 AT4g22990/F7H19_170 {Arabidopsis thaliana}	25%			-3.8, -2.3		
BG840993	SP|Q8W425 proteasome non-ATPase regulatory subunit 6 {Oryza sativa}	90%	5.06				2.3, 6.3
BG841229	SP|P48489 Serine/threonine protein phosphatase PP1 {Oryza sativa}	98%	1.03	-4.3, -10.0			
BG842726	dbj|BAB93128.1||AP003196 beta-1,3-glucanase-like protein {Oryza sativa}	84%	10.02				24.7, 5.6

[a] Genbank accession number.
[b] Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml).
[c] Mapping information is from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins (16).
[d] If the ratio is less than one, the negative reciprocal is listed.
[e] The first ratio is from unigene-microarray experiment, whereas the second one from SSH-macroarray experiment.

experiments (Fig. 2), and "cell rescue, defense, cell death and ageing" (32.0%) in SSH-macroarray experiments. Further ranking of classification categories was "cell rescue, defense, cell death and ageing" (19.0%), "signal transduction" (15.5%) and "transcription" (11.3%) for unigene-microarray experiments, and "transcription" (17.4%), "metabolism" (16.3%) and "signal transduction" (11.1%) for SSH-macroarray experiments. In spite of different ranks between both experiments, the top four

categories were the same, and three of them ("cell rescue, defense, cell death and ageing", "signal transduction", and "transcription"), are pathogenesis-related. In contrast to 60.5% in SSH-macroarray experiments, 45.8% of differentially expressed ESTs in unigene-microarray experiments were classified into pathogenesis-related categories, such as AI664862 (stress-induced protein OZI1 precursor), AI795699 (peroxidase), AI491543 (hypersensitive-induced response protein) and BM073434 (pathogenesis related protein-5).

Comparing map position of differentially expressed ESTs between SSH-macroarray and unigene-microarray

Altogether 20% (100 of 497) of ESTs identified from unigene-microarray experiments have been previously assigned to 51 bins (Gardiner et al., 1993) distributed over all 10 maize chromosomes (Fig. 3), whereas the same proportion of ESTs (in total 60) identified from SSH-macroarray were assigned to fewer genome regions (29 bins). ESTs were randomly distributed to chromosomes in both unigene-microarray ($P = 0.3979$, $\chi^2 = 9.438$, $df = 9$) and SSH-macroarray experiment ($P = 0.1806$, $\chi^2 = 12.62$, $df = 9$).

In contrast to the 30% (18 of 60) of mapped ESTs located in bin 3.04-3.05 (12) and bin 6.00-6.02 (6) in SSH-macroarray experiments, only 4% (4 of 100) of the mapped ESTs from unigene-microarray experiments in bin 3.04-3.05 (3) and bin 6.00-6.02 (1) (Table 1). The proportion of ESTs mapped in vicinity of *Scmv* QTLs was significantly higher ($P = 0.0013$) in SSH-macroarray than in unigene-microarray experiments. Among the ESTs mapped in bins 3.04-3.05 and 6.00-6.02, no homology was found between microarray- and SSH-ESTs. None of these microarray-ESTs, but at least 6 SSH-ESTs, showed homology to defense-related genes (Shi et al., submitted). Finally, AI947839, AW065765 and BG841986 identified from unigene-microarray experiments were co-localized with QSCM on chromosomes 1, 1, and 5 (Fig. 3).

Discussion

Comparison of SSH-macroarray and unigene-microarray experiments

Recently, microarrays were widely recognized as a significant technological advance providing genome-scale information on gene expression patterns (Richmond and Somerville, 2000). Complete transcriptome arrays are allowed to assay traits without preconceived ideas. Although a comprehensive microarray is not available in maize yet, the microarrays used in this study, contained 9,841 different unigenes, accounting for only 20% of the about 50,000 maize genes (Martienssen et al., 2004). In contrast to the unigene-microarray, the macroarray used in a companion study contained only a limited number of SSH clones specifically developed for studying SCMV resistance. One major limitation of SSH and similar methods is the difficulty to cover multiple comparisons when comparing a series of RNA samples, since SSH libraries are produced from pairwise comparisons (Donson et al., 2002). In our study (Shi et al., submitted), only five SSH libraries were constructed instead of twelve covering all combinations (there are only six comparisons, but there are 12 possibilities for subtraction) between four RNA samples (infected F7, infected $F7^+$, not-infected F7 and not-infected $F7^+$). This might be one explanation for 460 differentially expressed ESTs exclusively present on microarrays, which have not been recovered by the SSH-macroarray procedure (Fig. 1).

Only 8.8% (1045 of 11827) of the EST collection in the unigene-microarray was derived from two stress-induced EST libraries, including 496 (stressed shoot) and 603 (stressed root), whereas 91.2% (10771 of 11827) from fifteen EST libraries made from plants grown under normal environmental conditions. The number of differentially expressed genes discovered from library 496 (9%) was substantially higher than from other EST libraries (average: 4%). Since more ESTs differentially expressed in SSH-macroarray experiments (60.5%) were classified into pathogenesis-related categories than in unigene-microarray experiments (45.8%), the SSH cDNAs complemented the ESTs printed on microarrays.

In addition, SSH-macroarray procedure enriches for low-abundant and differentially expressed mRNAs by normalization (Diatchenko et al., 1996). The normalization step is particularly important because abundant pathogenesis-related transcripts (e.g., genes coding for PR proteins) very likely mask important SCMV-specific transcripts expressed at much lower levels. According to Maize EST statistics in the Genbank (31/12/2003 assembly), the distribution of high-, medium-, and low-abundant ESTs was 3%, 64%, and 33%, respectively. "Digital northern" analysis revealed a shift towards medium- and low-abundant ESTs within our SSH libraries with 51% and 47%, respectively (Shi et al., submitted), in contrast to a shift towards high-abundant ESTs (9%) on the unigene-microarray, most of them associated with metabolism, within differentially expressed microarray-ESTs. Therefore, the 224 unique

SSH-ESTs (Fig. 1) demonstrate the usefulness of the SSH-macroarray procedure to isolate target trait – specific genes.

Regardless of the procedure, the reproducibility was high both in unigene-microarray and SSH-macroarray experiments. Fisher's exact test (P = 0.0117) showed consistent expression patterns of microarray-ESTs and SSH-ESTs from the same comparison (Fig. 2). However, the degree of consistency was limited, such as AI691482 was induced in unigene-microarray experiment but repressed in SSH-macroarray experiment. Discrepancies between both approaches can be explained by i) different targets spotted on the arrays; ii) different labeling procedures; iii) different ratio measurements.

Because different cDNAs of different length, different parts of the genes were deposited on micro- or macroarrays, it might mask changes in transcript levels due to cross-hybridization to gene family members (Girke et al., 2000). Furthermore, probe labeling was different. The same total RNA samples were indirectly labeled with fluorescent dyes Cy3 or Cy5 using random primers in the unigene-microarray experiment, whereas direct labeling with radioactive P^{32} using oligo(dT) was employed in the SSH-macroarray experiment. In another study comparing array-based results with northern blots, arrays were less sensitive in measuring a subset of the genes (Taniguchi et al., 2001). Moreover, different ratio measurements were implemented. In unigene-microarray experiments two cDNA samples were hybridized on the same glass slide in parallel, allowing the direct measurement of ratios on the unigene-microarray. In contrast, internal controls were included on nylon membranes employed in SSH–macroarray experiments. Only one probe was hybridized per membrane and ratios were obtained by indirect comparisons. Among 21 homologous ESTs from the same comparison in Category 2, the correlation (r = 0.88) of the expression level between two approaches was highly significant by Pearson correlation calculations (P < 0.0001). The finding of 41 ESTs in Category 3 corroborates a higher risk of smaller fragments cross-hybridizing with other gene family members (Finkelstein et al., 2002).

Molecular mechanisms of maize-SCMV interaction

Typical mosaic symptoms have been observed in leaves of susceptible F7 after systemic movement and replication of SCMV. Thus, the F7 - SCMV interaction is a compatible interaction (Comparison C) (Fig. 4). In compatible interactions, the distribution of genes among functional classes looked similar to the incompatible reaction, regardless of unigene-microarray or SSH-macroarray. It is consistent with the hypothesis that viruses induce defense response both in susceptible and resistant plants at early stages (Matthews and Hull, 2002). Whitham et al. (Whitham et al., 2003) reported that diverse RNA viruses, including cucumber mosaic cucumovirus, oil seed rape tobamovirus, turnip vein clearing tobamovirus, potato virus X potexvirus and turnip mosaic potyvirus, elicited the expression of common sets of genes in susceptible *Arabidopsis*. Totally 62% of the differentially expressed genes in Comparison C, 72% (33 of 47) for SSH-macroarray (Shi et al, submitted) and 58% (67 of 115) for unigene-microarray, can be found in this common set of genes, such as BM501006 (F5M15.13), AI820398 (AIG2 protein homolog F8F16.130), while the remaining genes without annotation could be maize-specific.

Plants of NIL $F7^+$ displayed no SCMV symptoms in all infection trials (Shi et al., submitted), thus $F7^+$ is completely resistant to SCMV and the $F7^+$ - SCMV interaction is an incompatible interaction (Comparison D) (Fig. 4). Gene expression profiles of incompatible reactions, including TMV in tomato and *Chenopodium* (Cooper, 2001, Golem and Culver, 2003), revealed similarities at the gene level with Comparison D, such as AI491543 (hypersensitive-induced response protein) and BM073434 (pathogenesis related protein-5) induced after infection. It is consistent with SSH-macroarray experiments and corroborates most likely resistant plants utilize a common mechanism for defense against virus attack (Matthews and Hull, 2002).

In SSH-macroarray experiments, both constitutive and inducible genetic discrepancy were discussed as resistance mechanisms of $F7^+$ against SCMV (Shi et al., submitted). In unigene-microarray experiments, several putative preformed inhibitors were also revealed in Comparison A (Fig. 4, Constitutive defense). AW011679 show homology to genes encoding UMP/CMP kinase, and BM335333 is the homolog of an ankyrin-kinase. It corroborates previous finding that SCMV can be detected and, thus, replicates in primary infected leaves of

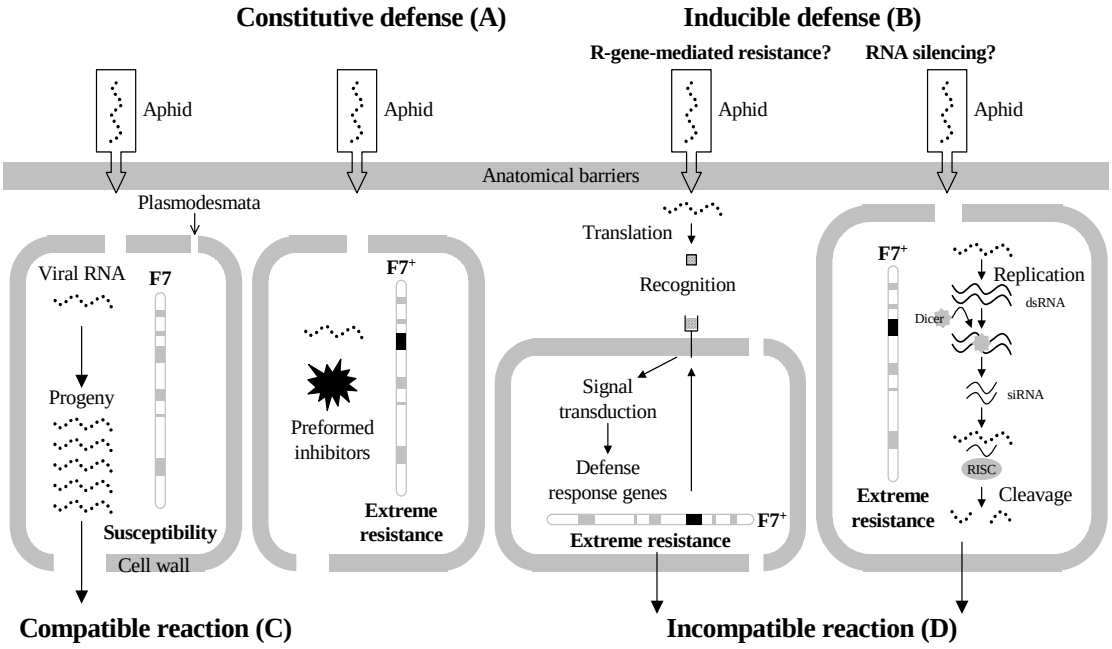


Fig. 4. Diagrammatic view of SCMV-maize interaction
SCMV enters maize symplasm by non-persistent transmission through aphids. In susceptible F7, SCMV rapidly replicates and spreads from cell to cell through plasmodesmata. Comparison C revealed differentially expressed ESTs in this compatible reaction. In contrast, extreme resistance was detected in resistant $F7^+$. Preformed inhibitors in the cells play a major role in constitutive defense, revealed in Comparison A, while inducible defense (Comparison B) can be found together with R-gene-mediated resistance or RNA silencing. Comparison D revealed differentially expressed ESTs in this incompatible reaction. The models of R-gene-mediated resistance and RNA silencing were adapted from Lucas and Dickinson (1998), Waterhouse and Helliwell (2003), respectively.

resistant genotypes (Louie, 1995). So far, two types of inducible defence are defined: hypersensitivity response (HR) and extreme response (ER) (Matthews and Hull, 2002). HR limits virus infection to a zone of cells around the initially infected cell of the resistant host, usually with the formation of visible necrotic local lesions (Matthews and Hull, 2002). ER limits virus multiplication to initially infected cells because of an ineffective virus-coded movement protein, giving rise to latent infection. No HR symptoms are observed for maize leaves infected with SCMV, thus maize resistance to SCMV might be extreme resistance. Further experiments conducted at single-cell level, usually in protoplast, are warranted (Matthews and Hull, 2002).

ER is most often triggered by dominant or semi-dominant resistance (R) genes and occurring in a strain-specific or "gene-for-gene" manner (Fig. 4) (Matthews and Hull, 2002). In potato, two extreme resistance genes (*Rx1* and *Rx2*) to PVX have been cloned, which belong to the nucleotide binding, leucine- rich repeat (NBS-LRR) super-family of R-genes (Bendahmane et al., 1999, Bendahmane et al., 2000). In addition, ER might be triggered by RNA silencing (Fig. 4). In contrast to resistance triggered by the NBS-LRR–type R genes, resistance through silencing appears not to depend on a gene-for-gene recognition event (Whitham et al., 2000). In unigene-microarray experiments, differentially expressed pathogen-related genes, identified from Comparison B (inducible genetic discrepancy), such as AI664862 (stress-induced protein OZI1 precursor) and AI795699 (peroxidase), have been found together with R-gene-mediated resistance or RNA silencing. This is in agreement with the finding from SSH-macroarray experiments (Shi et al., submitted). So far, little is known about the genes involved in signal transduction of HR and ER, it is even possible that they use the same genes for signaling. Therefore, both mechanisms might be involved in SCMV resistance.

Candidate gene (CG) selection for *Scmv* QTLs

The CG approach consists of three subsequent steps: the choice, screening and validation of CGs (Pflieger et al., 2001). In this study, good candidates associated with SCMV resistance can be chosen from at least three classes: i) positional CGs mapping to bins 3.04 – 3.05 and 6.00 – 6.02, ii) functional CGs showing the homology to pathogenesis-related genes, or iii) the ESTs in Category 2 showing consistent expression pattern in both approaches. Although these three classes contain candidates for genes affecting SCMV resistance, only the first class contains candidate genes for *Scmv1* or *Scmv2*.

So far, 18.6% (696 of 3737) of all mapped maize ESTs are located in bins 3.04 – 3.05 (426) and 6.00 – 6.02 (270) (http://www.plantgenomics.iastate.edu/maizechip/). In contrast to the 30% (18 of 60) of the mapped ESTs from SSH-macroarray located in bin 3.04-3.05 (12) and bin 6.00-6.02 (6) (Shi et al. submitted), only 4% (4 of 100) of the mapped microarray-ESTs were located in bins 3.04-3.05 (3) and bin 6.00-6.02 (1) (Table 1). While no homology was found between SSH-ESTs (18) and microarray-ESTs (4), 50% (9 of 18) mapped SSH-ESTs belong to Category 3 (Fig. 1). One possible explanation is that *Scmv*-specific ESTs were under represented in genome-wide unigene-microarray, whereas SSH libraries enriched them after normalization step. Identification of a larger number of differentially expressed genes mapping to these two regions can be explained by i) genes differentially expressed due to the polymorphism between F7 and $F7^+$ in these two regions but without relation to SCMV resistance, or ii) clustering of genes involved in SCMV resistance in these two regions. Except for positional CGs, several CGs (pathogenesis-related genes) were revealed from the 80% non-mapped ESTs, such as AI621822 (Avr9 elicitor response protein), AI999974 (maize catalase isozyme 3).

Comparing to other ESTs identified from either unigene-microarray or SSH-macorarray experiments, 19 consistent ESTs in Category 2 (Table 1) are the most promising candidates for being differentially expressed in the context of SCMV resistance. However, due to uncompleted annotation, four ESTs (AI665633, AI855243, AW330660 and AI974914) have no tentative annotation from TIGR Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml), and the annotation of four genes (AI649641, AI714860, AI942105 and AW438364) was based on proteins from the *Arabidopsis* or rice genome. Of all eleven remaining ESTs, AI461569, AI621758, AI941971, AI942048 and AW052909 were related to RNA binding, while AI600862, AI738263, AW331161, BG842726, BG841229 and BG840993 were homologous of catalytic proteins involving in defense response. In addition, AI600862, BG840993, BG841229 and BG842726 were already mapped in bin 2.04, 5.06, 1.03 and 10.02, respectively. These genes, located outside of *Scmv* QTL regions, might be further downstream in the signal transduction pathway and induced by genes located in the *Scmv1* and / or *Scmv2* regions.

Once genes responsible for quantitative variation of SCMV resistance become available, information can be passed on to plant breeders in the form of functional markers (Andersen and Lübberstedt, 2003). Functional markers are superior to random DNA markers such as RFLPs, SSRs and AFLPs owing to complete linkage with trait locus alleles. Due to polygenic trait of SCMV resistance, marker-assisted selection (MAS) programs with functional markers would increase breeding efficiency.

Experimental procedures

Plant materials

NILs F7 and $F7^+$ were grown and maintained in growth chambers under a 12 -h photoperiod at 23°C and 50% relative humidity. Two-week-old plants were mechanically inoculated by an air brush technique using a tractor-mounted air compressor at constant pressure of 799 kPa (Fuchs and Gruntzig, 1995). Non-infected plants and infected plants were kept in separate growth chambers after inoculation. Non-infected and infected leaves were harvested 24 hours after inoculation in parallel. For biological replicates, two independent sets of leaf materials were harvested. To confirm resistance or susceptibility of infected plants used for leaf harvest, plants were grown for additional two weeks. After this period, mosaic symptoms were observed on each infected F7 plant, whereas infected $F7^+$ remained without symptoms.

Probe preparation and unigene-microarray hybridization

Total RNA from maize leaves was extracted using TRIzol reagent (Invitrogen, Carlsbad, California, USA). The same total RNA extracted was also used for probe preparation in SSH-macroarray approach. Poly (A)+ RNA was isolated from Total RNA via Dynabeads® Oligo (dT) 25 (Dynal biotech, Oslo, Norway). According to TIGR

Microarray Protocols (Hegde et al., 2000), each mRNA sample was indirectly labeled with Cy3 or Cy5 (Amersham Pharmacia, Piscataway, NJ, USA) and hybridized with maize unigene-microarrays.

Maize unigene-microarrays were generated by the laboratory of Prof. Schnable (Iowa University, USA) and contain 11,827s maize ESTs (http://www.plantgenomics.iastate.edu/maizechip/). Among them, 11,027 ESTs were spotted once, 391 ESTs duplicate and 6 ESTs triple. Thus, 11, 424 unique ESTs, clustered into 9841 unigenes, are on the maize unigene-microarray, and 8.3% (949 of 11,424) of them have been mapped. The EST collection at the maize unigene-microarrray was derived from fifteen EST libraries, including 486 (immature leaf), 605 (endosperm), 606 (ear tissue), 614 (root), 618 (tassel primordia), 660 (mixed stages of anther and pollen), 683 (14 day immature embryo), 687 (mixed stages of embryo development), 707/945 (mixed adult tissues) and ISUM3/4/5/6/7 (seedling and silk), made from plants grown under normal environmental conditions and two stress-induced EST libraries, including 496 (stressed shoot) and 603 (stressed root).

For each comparison, four replications, including two biological replications and dye swap replications in each biological replication, were conducted. Thus, four maize gene chips were used in each comparison.

Raw data acquisition

Fluorescence signals were detected using the arrayWoRx® Biochip Reader (Applied Precision, Issaquah, Washington, USA). The image data obtained were imported into the software program ArrayVision 7.0 (Imaging Research, St. Catharines, Ontario, Canada) for spot detection and quantification of hybridization signals. Local background calculated from the corners between spots, were subtracted using ArrayVision 7.0 to obtain raw signal intensities.

Unigene-microarray Data Analysis

Raw data were exported from ArrayVision 7.0 into Excel and converted to TIGR Array Viewer (TAV) format files. Using the TIGR Microarray Data Analysis System (MIDAS) (Saeed et al., 2003), first signals were filtered to exclude low intensity signals, and then "lowess (locally weighted linear regression) normalization" was employed to adjust intensity-dependent effects in log2 (ratio) values. "Replicate consistency checking" removed poorly reproducible genes, and finally "slice analysis" was utilized to identify differentially expressed genes, which were induced/repressed more than 1.96 standard deviations from the local mean in each comparison (Quackenbush, 2002).

Sequence Analysis

Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). Each gene was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. Gene mapping information came from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/). If an EST was assigned to a mapped gene cluster, we assumed identical chromosome location of this EST and the gene cluster. The distribution of mapped genes was drawn by MapChart (Voorrips, 2002).

Acknowledgements

We are grateful to Patrick S. Schnable for offering maize unigene-microarrays. We thank Preben B. Holm, Per Gregersen and Christina Ingvardsen for assistance with unigene-microarray experiments in Research Centre Flakkebjerg. We also thank Olaf Neuschaefer-Rube for technical assistance at GSF (Neuherberg). This work was supported by the German research foundation (DFG LU601/6-1).

References

Andersen, J. R. and Lübberstedt, T. (2003) Functional markers in plants. Trends Plant Sci. 8(11): 554-60.

Audic, S. and Claverie, J. M. (1997) The significance of digital gene expression profiles. Genome Res 7(10): 986-95.

Baldwin, D., Crane, V. and Rice, D. (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr. Opin. Plant Biol. 2(2): 96-103.

Bendahmane, A., Kanyuka, K. and Baulcombe, D. C. (1999) The Rx gene from potato controls separate virus resistance and cell death responses. Plant Cell 11(5): 781-92.

Bendahmane, A., Querci, M., Kanyuka, K. and Baulcombe, D. C. (2000) *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: application to the Rx2 locus in potato. Plant J. 21(1): 73-81.

Bland, J. M. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476): 307-10.

Borevitz, J. O. and Chory, J. (2004) Genomics tools for QTL analysis and gene discovery. Curr. Opin. Plant Biol. 7(2): 132-6.

Collins, N. C., Webb, C. A., Seah, S., Ellis, J. G., Hulbert, S. H. and Pryor, A. (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol. Plant Microbe Interact. 11(10): 968-78.

Cooper, B. (2001) Collateral gene expression changes induced by distinct plant viruses during the hypersensitive resistance reaction in Chenopodium amaranticolor. Plant J. 26(3): 339-349.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. and Siebert, P. D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA 93(12): 6025-30.

Donson, J., Fang, Y., Espiritu-Santo, G., Xing, W., Salazar, A., Miyamoto, S., Armendarez, V. and Volkmuth, W. (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol 48(1-2): 75-97.

Dussle, C. M., Melchinger, A. E., Kuntze, L., Stork, A. and Lübberstedt, T. (2000) Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTL contributing to SCMV resistance in maize. Plant Breeding 119(4): 299-303.

Fernandes, J., Brendel, V., Gai, X., Lal, S., Chandler, V. L., Elumalai, R. P., Galbraith, D. W., Pierson, E. A. and Walbot, V. (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. Plant Physiol 128(3): 896-910.

Finkelstein, D., Ewing, R., Gollub, J., Sterky, F., Cherry, J. M. and Somerville, S. (2002) Microarray data quality analysis: lessons from the AFGC project. Arabidopsis Functional Genomics Consortium. Plant Mol. Biol. 48(1-2): 119-31.

Fuchs, E. and Gruntzig, M. (1995) Influence of Sugarcane Mosaic-Virus (Scmv) and Maize-Dwarf Mosaic-Virus (Mdmv) on the Growth and Yield of 2 Maize Varieties. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection 102(1): 44-50.

Gardiner, J. M., Coe, E. H., Melia-Hancock, S., Hoisington, D. A. and Chao, S. (1993) Development of a core RFLP map in maize using an immortalized F2 population. Genetics 134(3): 917-30.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol 124(4): 1570-81.

Golem, S. and Culver, J. N. (2003) Tobacco mosaic virus induced alterations in the gene expression profile of Arabidopsis thaliana. Mol. Plant Microbe Interact. 16(8): 681-688.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. and Quackenbush, J. (2000) A concise guide to cDNA microarray analysis. Biotechniques 29(3): 548-50, 552-4, 556 passim.

Kuntze, L., Fuchs, E., Gruntzig, M., Schulz, B., Klein, D. and Melchinger, A. E. (1997) Resistance of early-maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breeding 116(5): 499-501.

Louie, R. (1995) Vascular Puncture of Maize Kernels for the Mechanical Transmission of Maize White Line Mosaic-Virus and Other Viruses of Maize. Phytopathology 85(2): 139-143.

Lucas, J. A. and Dickinson, C. H. (1998) Plant pathology and plant pathogens. Malden, Mass., Blackwell Science.

Martienssen, R. A., Rabinowicz, P. D., O'Shaughnessy, A. and McCombie, W. R. (2004) Sequencing the maize genome. Curr. Opin. Plant Biol. 7(2): 102-7.

Matthews, R. E. F. and Hull, R. (2002) Matthews' plant virology. San Diego, Academic Press.

Nadimpalli, R., Yalpani, N., Johal, G. S. and Simmons, C. R. (2000) Prohibitins, stomatins, and plant disease response genes compose a protein superfamily that controls cell proliferation, ion channel regulation, and death. J. Biol. Chem. 275(38): 29579-86.

Nakazono, M., Qiu, F., Borsuk, L. A. and Schnable, P. S. (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15(3): 583-96.

Pflieger, S., Lefebvre, V. and Causse, M. (2001) The candidate gene approach in plant genetics: a review. Molecular Breeding 7(4): 275-291.

Provvidenti, R. and Hampton, R. O. (1992) Sources of resistance to viruses in the Potyviridae. Arch. Virol. Suppl. 5: 189-211.

Quackenbush, J. (2002) Microarray data normalization and transformation. Nat. Genet. 32: 496-501.

Quint, M., Mihaljevic, R., Dussle, M., Xu, L., Melchinger, E. and Lübberstedt, T. (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor. Appl. Genet. 105(2-3): 355-363.

Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3(2): 108-16.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2): 374-8.

Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001) Quantitative assessment of DNA microarrays--comparison with Northern blot analyses. Genomics 71(1): 34-9.

Voorrips, R. E. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93(1): 77-8.

Wan, J., Dunning, F. M. and Bent, A. F. (2002) Probing plant-pathogen interactions and downstream defense signaling using DNA microarrays. Funct. Integr. Genomics 2(6): 259-73.

Waterhouse, P. M. and Helliwell, C. A. (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4(1): 29-38.

Whitham, S. A., Anderberg, R. J., Chisholm, S. T. and Carrington, J. C. (2000) Arabidopsis RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. Plant Cell 12(4): 569-82.

Whitham, S. A., Quan, S., Chang, H. S., Cooper, B., Estes, B., Zhu, T., Wang, X. and Hou, Y. M. (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. Plant J. 33(2): 271-83.

Xia, X. C., Melchinger, A. E., Kuntze, L. and Lübberstedt, T. (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89(8): 660-667.

Xu, M. L., Melchinger, A. E., Xia, X. C. and Lübberstedt, T. (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. Mol. Gen. Genet. 261(3): 574-81.

Yuan, L., Dussle, C. M., Melchinger, A. E., Utz, H. F. and Lübberstedt, T. (2003) Clustering of QTL conferring SCMV resistance in maize. Maydica 48(1): 55-62.

Zhu, T. (2003) Global analysis of gene expression using GeneChip microarrays. Curr Opin Plant Biol 6(5): 418-25.

Association between SCMV resistance and macroarray-based expression patterns in different maize inbreds

Chun Shi[1], Fritz Thümmler[2], Albrecht E. Melchinger[3], Gerhard Wenzel[1], Thomas Lübberstedt[4,*]

[1]Chair of Plant Breeding, Technical University of Munich, Am Hochanger 2, 85350 Freising, Germany
[2]Vertis Biotechnologie AG, Lise-Meitner-Straße 30, 85354 Freising, Germany
[3]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
[4]Department of Plant Biology, Research Centre Flakkebjerg, 4200 Slagelse, Denmark

* For correspondence (tel. +45 5811 3484; fax +45 5811 3301; e-mail Thomas.Luebberstedt@agrsci.dk).

Key words: Candidate gene identification, Different maize inbred lines, Functional genomics, Gene expression, SCMV resistance

Abstract

Recent advances in RNA profiling offer an opportunity to establish functional links between genotype and phenotype for complex traits like SCMV resistance. The change of RNA profiles was monitored on a macroarray containing SSH (suppression subtractive hybridization) clones. The number of differentially expressed genes (SCMV infected versus non-infected) in individual lines was 177, 163, 165, 62, 47, 37, and 93, for FAP1360A, D21, D32, Pa405, F7, D145, and D408, respectively. All inbreds were divided into two groups by hierarchical cluster analysis: D32, D21, FAP1360A and D408 formed one group; Pa405, D145 and F7 another group. Due to the genetic structure among the seven inbreds, genetic background and resistance response are confounded. With or without the resistant U.S. inbred line Pa405, 22 and 112 genes were identified by *t* tests between resistant (D21, D32, and FAP1360A) and susceptible (D145, D408, and F7) inbred lines, respectively. The 112 candidate genes were divided into three clusters by K-means clustering and analyzed in more detail. These candidate genes identified from the present analysis can be further investigated in a segregating population by a "genetical genomics" approach.

Introduction

Sugarcane mosaic virus (SCMV) is an important pathogen of maize (*Zea mays L.*), causing substantial yield loss in susceptible cultivars (Fuchs and Gruntzig 1995). Pa405 is the most intensively studied inbred line associated with SCMV resistance in U.S.A. (Louie et al. 1991). In previous studies, Kuntze et al. (1997) screened 122 early-maturing European inbred lines for resistance to SCMV and MDMV (maize dwarf mosaic virus) and identified only three inbreds (D21, D32, and FAP1360A) displaying complete resistance under both field and greenhouse conditions. Two major quantitative trait locus (QTL) regions, *Scmv1* and *Scmv2*, conferring resistance to SCMV were mapped to chromosome arms 6S and 3L (Xia et al. 1999, Dussle et al. 2000). Based on pedigree records, D21 and D32 are closely related by descent, whereas FAP1360A and Pa405 were developed independently (Kuntze et al. 1995). Pedigree relationship analysis by molecular markers (Xu et al. 2000) indicated that FAP1360A, D21, and D32 shared the same marker haplotype in the *Scmv1* region despite of different donors of this region. The *Scmv2* region differed between FAP1360A, D21, and D32, whereas Pa405 had unique haplotypes both in the *Scmv1* and *Scmv2* region. Three F_2 populations including SCMV susceptible inbreds F7, D145, and D408 were studied in more detail: F7 × FAP1360A, D145 × D32 and D408 × D21 (Melchinger et al. 1998), including quantitative trait locus (QTL) (Xia et al. 1999, Dussle et al. 2000) and bulked segregant analysis (BSA) (Xia et al. 1999).

Recent advances of microarray-based expression profiling allows gene expression monitoring on a genome-wide scale and offers an opportunity to establish functional links between genotype and phenotype for complex traits like SCMV resistance (Cheung and Spielman 2002). If taken in its simplest setting, a change in the expression level of a particular gene when comparing different genotypes can be considered as expression-level

polymorphism (ELP). ELPs can be associated with regions of the genome, like QTL. As there are thousands of genes, some of which are functionally related, this approach seems to have great potential for dissecting the complex traits and identifying the genes underlying QTL (Doerge 2002).

So far, microarray-based expression profiling was mainly used to compare a wild-type plant with a corresponding mutant, or different treatments on or developmental stages of the same genotype, whereas the comparison of different genotypes, breeding lines, or cultivars that exhibit differences in quantitative traits, has rarely been reported. Recently, a new approach has been developed for relating differences in gene expression to the complex trait 'malting quality' among ten barley genotypes. Between 17 and 30 candidate genes were identified for each of the six malting parameters analyzed (Potokina et al. 2004).

In this study, the above-mentioned seven inbreds well-characterized with regard to SCMV resistance were chosen to analyze associations between SCMV resistance and macroarray-based expression patterns. The macroarrays containing SSH (suppression subtractive hybridization) (Diatchenko et al. 1996) clones were used to analyze ELPs among these inbreds. In a companion study, the same macroarrays have been successfully applied to detect SCMV-related transcripts between NILs (Near isogenic line) $F7^+$ and F7, resistant or susceptible to SCMV, respectively (Shi et al. 2004). The objectives of our study were to 1) compare the pedigree relationships among the maize inbreds based on marker and expression profiling data, 2) identify genes consistently differentially expressed between resistant and susceptible inbreds, and 3) study in detail the expression patterns of previously identified candidate genes for *Scmv* QTL (Shi et al. 2004) in resistant versus susceptible inbreds.

Materials and Methods

Plant materials

D21, D32, FAP1360A, Pa405, F7, D145, and D408 were grown and maintained in growth chambers under a 12 -h photoperiod at 23°C and 50% relative humidity. Two-week-old plants were used for SCMV inoculation. Artificial inoculation followed the procedure described by Fuchs and Grüntzig (1995). Non-infected plants and infected plants were kept in separate growth chambers after inoculation. Non-infected and infected leaves were harvested 24 hours after inoculation in parallel. For biological replicates, two independent sets of leaf materials including five plants in each set were harvested. To confirm resistance or susceptibility of infected plants used for leaf harvest, plants were grown for additional two weeks.

The introduction of macroarray

The macroarrays including SSH clones derived from comparison of NILs $F7^+$ (resistant to SCMV) and F7 (susceptible to SCMV) (Shi et al. 2004) were used in this study. In total, 2688 clones were spotted in duplicate on each macroarray. These clones were randomly picked from five SSH libraries. For two tester/driver cDNA pairs (infected $F7^+$ versus infected F7; non-infected $F7^+$ versus non-infected F7) subtractions were conducted in both directions. For the tester/driver cDNA pair infected $F7^+$ versus non-infected $F7^+$ only forward direction was conducted. Microarray hybridization data were evaluated by the SpotReport™ Alien™ cDNA Array Validation System (Stratagene, La Jolla, CA, USA), including positive, negative, and ten spiking controls.

Of 2688 clones, 672 clones, ranging in length from 96 to 843 bp, were sequenced by MWG (Ebersberg, Munich, Germany) and clustered into 302 genes. Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). Each gene was assigned to a functional class using the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/thal/db/tables/tables_func_frame.html) classification scheme by BLASTX with a threshold E value of 10. Gene mapping information came from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/).

Macroarray hybridizations

Total RNA from maize leaves was extracted using TRIzol reagent (Invitrogen, Carlsbad, California, USA). In each RT reaction, 5, 2.5, 1.25, 0.625, 0.312, 0.156, 0.078, 0.04, 0.02 and 0.01 ng of the ten different alien mRNA spikes were added to 5 µg of total RNA. Using the Strip-EZ RT kit (Ambion, Austin, TX, USA), cDNA synthesis was primed using oligo(dT) and [α-^{32}P]dATP (MP Biomedicals, Irvine, CA,USA). cDNA was separated from unincorporated nucleotides using Micro Bio-Spin chromatography columns (Bio-Rad Laboratories, Hercules, CA ,USA) filled with Sephadex G-50 (Amersham, Piscataway, NJ, USA) equilibrated in water. tRNA and oligo(dA) was added to the hybridization probe to suppress cross-hybridization. The prehybridization and hybridization steps were conducted as described in

the manual of the Strip-EZ RT kit. Labeled cDNA probes were stripped from the arrays using the Strip-EZ system (Ambion, Austin, TX, USA), and the process checked by phosphorimaging. For technical replicates, every RNA sample was used in two independent labeling and hybridisation experiments.

Raw data acquisition

Hybridization signals were detected using the Storm 860 phosphorimager (Amersham, Piscataway, NJ, USA) with a resolution of 50 µm. The image data obtained were imported into the software program ArrayVision 7.0 (Imaging Research, St. Catharines, Ontario, Canada) for spot detection and quantification of hybridization signals. Local background calculated from empty spots in each secondary grid, were subtracted using ArrayVision 7.0 to obtain raw signal intensities.

Macroarray Data Analysis

Raw data were exported from ArrayVision 7.0 into Excel. Duplicate spots at macroarrays were averaged. According to spiking controls, data of different macroarrays, hybridized with individual non-infected and infected inbreds, were normalized and converted to TIGR Array Viewer (TAV) format files. Using the TIGR Microarray Data Analysis System (MIDAS) (Saeed et al. 2003), first signals were filtered to exclude low intensity signals, and then " lowess (locally weighted linear regression) normalization" was employed to adjust intensity-dependent effects in log2 (ratio) values. "Replicate consistency checking" removed poorly reproducible genes, and finally "slice analysis" was utilized to identify differentially expressed genes in individual inbreds, which were induced/repressed more than 1.96 standard deviations from the local mean in each comparison (Quackenbush 2002).

Afterwards, the dataset of all inbreds were imported in Multiexperiment Viewer (MeV) (Saeed et al. 2003). "Hierarchical cluster analysis" (Eisen et al. 1998) were conducted to discover similar expression pattern, and then "between-subject t-tests with adjusted Bonferroni correction" (Pan 2002) were utilized to identify candidate genes differentially expressed between SCMV resistant and susceptible inbreds. The pair-wise correlation within the inbreds was obtained by "Gene distance matrix" (Saeed et al. 2003). Finally, 112 candidate genes were grouped by K-mean clustering (Soukas et al. 2000).

AFLP Data Analysis

Based on the 12 AFLP primer combination (E-AAC/M-CAT, E-AAC/M-CTA, E-AAC/M-CTT, E-ACA/M-CTT, E-ACT/M-CAA, E-ACT/M-CTC, E-AAG/M-CTG, E-AGG/M-CTA, E-AGG/M-CTG, E-ACC/M-CAC, E-ACC/M-CAG, and E-ACC/M-CTC) (Xu et al. 2000), 857 polymorphic AFLP markers were detected between the seven inbred lines FAP1360A, D21, D32, Pa405, D145, D408, and F7. Theses bands were scored as either 1 (present) or 0 (absent) and compiled into a binary data matrix of 7 × 857. Further analysis was carrie out by NTSYSpc 2.11a (Rohlf 1998). Similarity matrix (AFLP matrix) was produced from this matrix using SIMQUAL module. It was analyzed by UPGMA (unweighted pair group method using arithmetic averages) clustering method in the SAHN module. The dendrogram was created using the TREE module.

The dataset used in Mev analysis was also imported into NTSYSpc 2.11a and used to produce a distance matrix (Expression matrix) with the SIMINT module. Mantel test (Mantel 1967) was conducted between AFLP and Expression matrices using MXCOMP module.

Results

SCMV inoculation effects on gene expression in individual inbreds

After SCMV inoculation, mosaic symptoms (compatible reaction) were observed on each infected plant of susceptible inbreds (D145, D408, and F7) within two weeks, whereas infected plants of resistant inbreds (D21, D32, FAP1360A, and Pa405) remained without symptoms after five weeks (incompatible reaction).

The reproducibility within the same macroarray experiment was high, with a Pearson correlation coefficient exceeding 0.99 between duplicate spots across all tests. The mean intensity of spiking controls in each macroarray was used to normalize the differences in signal intensity among the macroarrays. When ^{32}P labeled cDNA probes prepared from the same mRNA were hybridized to the same macroarray, over 96 % of the ratios varied less than 1.5-fold (Pearson correlation coefficient 0.94±0.02). When probes were prepared from biological replicates and hybridized to the same macroarray, over 90% of the ratios calculated from these two data sets varied by less than 2-fold (Pearson correlation coefficient 0.86±0.03).

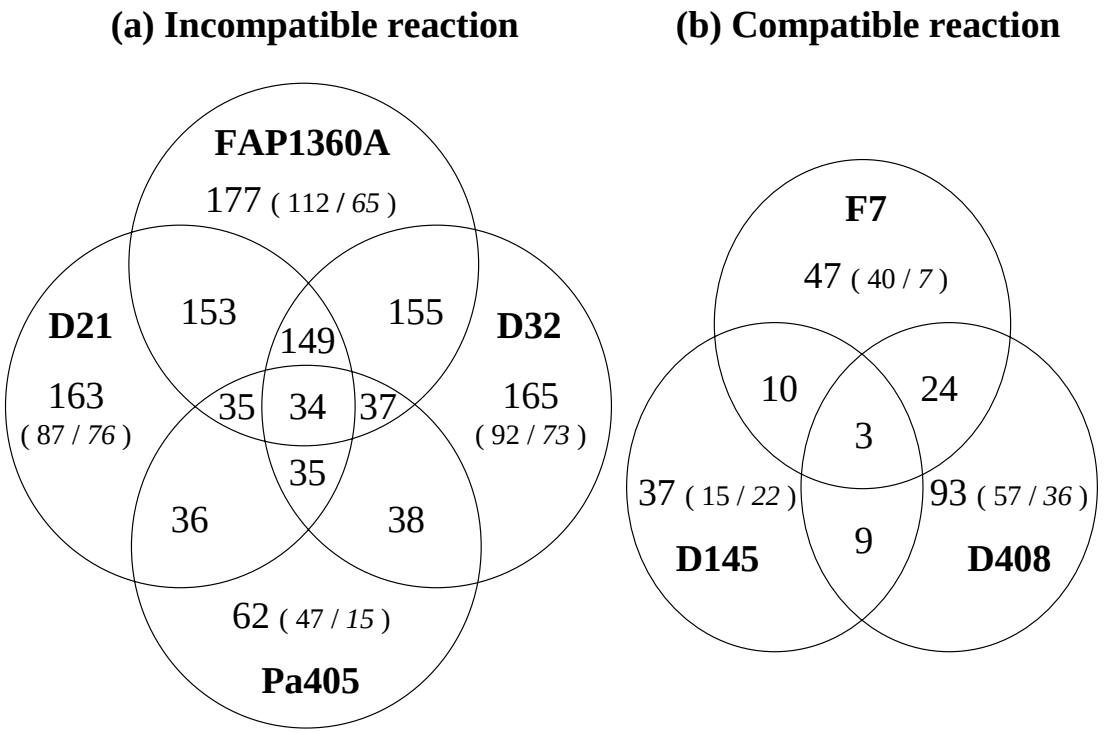


Figure 1. Venn diagrams for comparison of the numbers of differentially expressed genes within resistant or susceptible inbreds

Total numbers of genes differentially expressed in individual inbreds (FAP1360A, D21, D32, Pa405, F7, D145, or D408) are indicated in respective circles. In parentheses, the first number indicates up-regulated genes, whereas the last number in *italics* down-regulated genes. (a) Intersection of genes identified in resistant inbreds (incompatible reaction), (b) Intersection of genes identified in susceptible inbreds (compatible reaction).

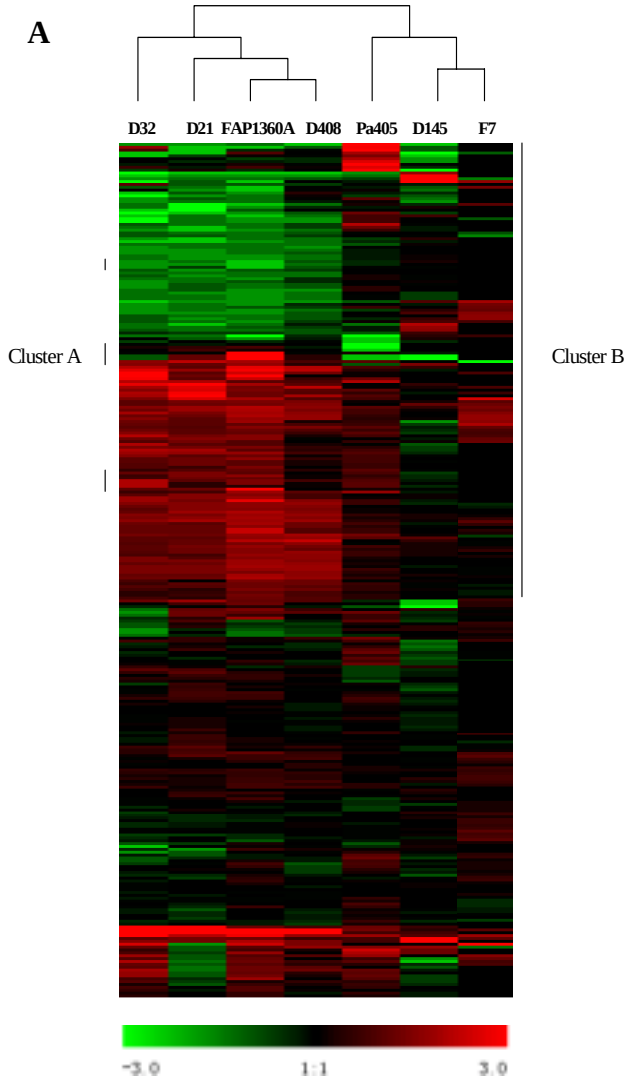
However, when first duplicate spots, and then technical replicates were averaged, close to 100% of biological replicates varied less than 1.5-fold (Pearson correlation coefficient 0.96± 0.02).

Recent work (Shi et al. 2004) identified 302 differentially expressed genes between NILs $F7^{+}$ and F7. This study focuses on these 302 genes to compare four resistant and three susceptible inbreds. Generally, more genes were differentially expressed in incompatible reactions than in compatible reactions (Figure 1). Among incompatible reactions, the greatest overall response was observed in FAP1360A (177 (112 up regulated, 65 down regulated)). Slightly fewer genes were identified in D21 (163 (87 up, 76 down)) and D32 (165 (92 up, 73 down)), and only one third of genes in Pa405 (62 (47 up, 15 down)). Among compatible reactions, the overall response was highest in D408 (93 (57 up, 36 down)), whereas substantially fewer genes showed differential expression in F7 (47 (40 up, 7 down)) and D145 (37 (15 up, 22 down)). The number of genes differentially expressed for pairs of resistant inbreds were 153 (between FAP1360A and D21), 36 (between D21 and Pa405), 38 (between Pa405 and D32) and 155 (between D32 and FAP1360A), respectively. In contrast, the commonly differentially expressed genes between pairs of susceptible inbreds were 10 (between F7 and D145), 9 (between D145 and D408) and 24 (between D408 and F7), respectively. A total of 34 genes were differentially expressed in all resistant inbreds, as compared to only 3 genes among all susceptible inbreds. The proportion of commonly differentially expressed genes was significantly higher (P = 0.0203) among incompatible compared to compatible reactions.

Association between gene expression patterns and 'SCMV resistance'

Based on hierarchical cluster analysis of expression patterns (Eisen et al. 1998), the inbreds were divided into two groups: D32, D21, FAP1360A, and D408 formed one group, Pa405, D145 and F7 the second group (Figure 2A). AFLP fingerprinting grouped the seven inbreds according to their heterotic grouping (Figure 2C). Expression patterns and genetic distances

A



B

	D21	D32	FAP1360A	Pa405	D408	D145
D32	0.56					
FAP1360A	0.53	0.55				
Pa405	0.83	0.81	0.83			
D408	0.88	0.94	1.00	0.85		
D145	0.48	0.51	0.48	0.68	0.75	
F7	0.75	0.78	0.81	0.62	0.61	0.58

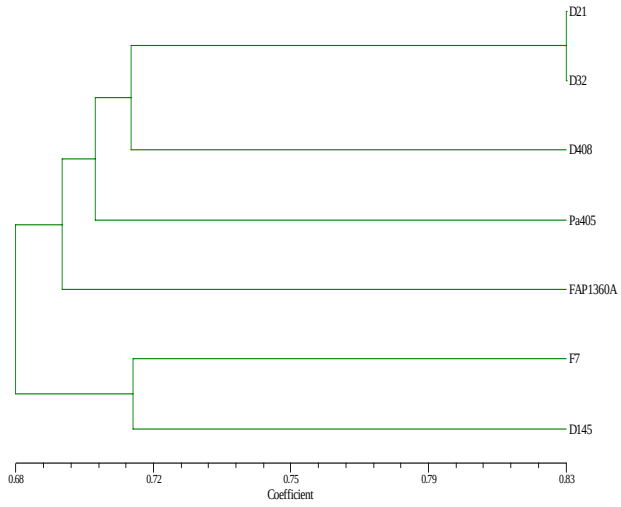
Figure 2. Cluster analysis based on expression patterns of SCMV resistant and susceptible lines with and without SCMV infection and AFLP markers

(A) Hierarchical cluster analysis (Eisen et al. 1998) of differential gene expression with and without SCMV infection for SCMV resistant (D32, D21, FAP1360A, and Pa405) and susceptible (D408, D145, and F7) inbreds. The color saturation reflects the magnitude of the log2 expression ratio (Cy5/Cy3) for each transcript. Red color means higher transcript levels than the reference, whereas green means lower transcript levels than the reference. The color log2 scale is provided at the bottom of the figure. The vertical bars on the left and right side of the tree indicate cluster A including 22 genes and Cluster B including 112 genes, respectively.

(B) A table of pair-wise correlations among the samples shown in (A).

(C) AFLP-based UPGMA dendrogram among the inbreds shown in (A).

C



assessed by AFLP fingerprinting were correlated (r = 0.8102, P=0.012) according to the Mantel test (Mantel 1967). Using t-tests with adjusted Bonferroni correction, the expression patterns of 22 genes were significantly (P < 0.05) different between resistant (D21, D32, FAP1360A, and Pa405) and susceptible inbreds (D408, D145, and F7) (Table 1, Figure 2A). According to marker-based haplotype analysis for the *Scmv1* and *Scmv2* regions (Xu et al. 2000), the origin of the resistant U.S. inbred Pa405 was largely different from the resistant European inbreds (D21, D32, and FAP1360A). A t-test without Pa405 revealed a substantially higher number of genes (112) with group-specific expression patterns for resistant (D21, D32, and FAP1360A) versus susceptible inbreds (D408, D145, and F7) (Table 1, Figure 2A, cluster B).

Pair-wise correlations (Figure 2B) were obtained by GDM (gene distance matrix) (Saeed et al. 2003). Maximum similarity (scaled distance = 1.0) was found between FAP1360A and D408. Expression patterns of all inbreds are similar (scaled distance ≥ 0.48). The expression pattern of D408 (average scaled distance = 0.84) showed the highest correlation with the other six inbreds: Pa405 (0.77), FAP1360A (0.70), D32 (0.69), F7 (0.69), D21 (0.67), and D145 (0.58). The average similarity within the resistant group (D21, D32, FAP1360A, and Pa405) (average scaled distance = 0.69) was higher than between susceptible inbreds (D408, D145, and F7) (Average scaled distance = 0.65).

Analysis of 112 candidate genes

According to expression patterns across the seven inbreds, 112 genes were clustered into three sub-clusters, showing distinct expression patterns based on K-mean clustering (Figure 3). Cluster 1 included 36 genes, twofold (on average) induced in European resistant inbreds after SCMV infection, but not differentially expressed in susceptible inbreds. In cluster 3, 43 genes were repressed at least twofold in European resistant inbreds, but not differentially expressed in susceptible inbreds. Contrasting expression patterns between European resistant and susceptible lines is characteristic for the 33 genes of cluster 2. For example, AI600292 was up-regulated in resistant but down-regulated in susceptible lines. The distribution of the 22 genes, showing distinct expression patterns between all resistant (including Pa405) and susceptible lines, is 12, 8, and 2 in clusters 1, 2, and 3, respectively.

Genes in clusters 1 and 2 were analyzed in more detail (Tables 2 and 3). Despite of low similarities in some cases, most of these genes were annotated in TIGR maize gene index (http://www.tigr.org/tdb/tgi/plant.shtml). More than half of the genes in cluster 1 were not classified (Table 2), while the other genes were assigned to 'cell rescue, defense, cell death and ageing' (10), 'cellular organization' (1), 'metabolism' (2), 'signal transduction' (3), and 'transcription' (1). In contrast, less than half of the genes in cluster 3 were not classified (Table 3), while the other genes were assigned to 'cell rescue, defense, cell death and ageing' (9), 'cellular organization' (3), 'energy' (3), 'metabolism' (3), 'protein destination' (1), 'signal transduction' (3) and 'transcription' (2). Of 19 mapped genes, five genes were located in bin 3.04-3.05.

Discussion

Association between SCMV resistance and expression-level polymorphisms

Analysis of expression profiling data across a collection of lines well characterized with respect to SCMV resistance has the potential to associate ELPs with important agronomic traits. This is comparable to association studies at the levels of DNA polymorphisms, were also no experimental populations such as segregating populations are required. It also avoids the limitation of positional cloning by NILs. This is due to the difficulty of developing NILs for loci that explain less than 20% of the variance and to constraints created by only using two alleles. Recent work (Potokina et al. 2004) confirmed that the variation of the complex

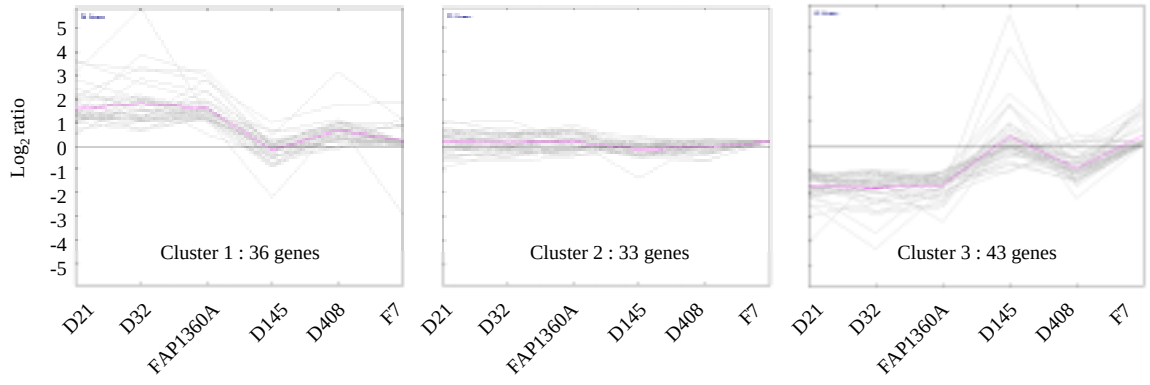


Figure 3. K-mean clustering of 112 genes in Cluster B of Fig. 2A
112 genes in Cluster B of Fig. 2A were grouped into 3 clusters using K-mean algorithm (Soukas et al. 2000). Each graph displays the mean pattern of expression of the ESTs in the cluster in pink. The number of ESTs in each cluster is at the bottom of each graph. The *y*-axis represents log2 of gene-expression levels.

Table 1. Differentially expressed genes identified by t-tests with adjusted Bonferroni correction ($P < 0.05$)

Item	Group A	Group B	Significant Genes
With Pa405	Resistant (D21, D32, FAP1360A, Pa405)	Susceptible (D408, D145, F7)	22
Without Pa405	Resistant (D21, D32, FAP1360A)	Susceptible (D408, D145, F7)	112
Without Pa405	Dent (D21, D32, FAP1360A, D408)	Flint (D145, F7)	112

trait "malting quality" in a set of 10 barley genotypes was reflected at the RNA level by using a cDNA array with 1400 ESTs. Between 17 and 30 candidate genes were identified for each of the six malting parameters analyzed.

302 differentially expressed genes were identified between NILs $F7^{+}$ and F7 in four comparisons addressing genetic discrepancy, response discrepancy, compatible interaction and incompatible interaction (Shi et al. 2004). In this study, of 112 genes differentially expressed between European resistant and susceptible lines, 42, 40, 19 and 58 genes were common with the genes identified from genetic discrepancy (125), response discrepancy (117), compatible interaction (48) and incompatible interaction (108), respectively. Although compatible and incompatible interactions concerned in this study, only a slightly higher ratio of common genes is found in compatible (39.6%) and incompatible interactions (53.7%) than genetic (33.6%) and response discrepancy (34.2%) due to overlapped genes between four comparisons (Shi et al. 2004), yet it shows at least 40% genes (In the case of compatible interaction: 39.6%) identified from the NILs can be used to characterize the expression pattern of European inbred lines from the same comparison, especially for incompatible reaction (53.7%).

Cluster analysis based on expression patterns of SCMV resistant and susceptible lines with and without SCMV infection and AFLP markers confirmed the close relationship of the European resistant lines (D21, D32, and FAP1360A) and separated them from the U.S. line Pa405. This is consistent with pedigree records (Kuntze et al. 1995) and shows that chromosome segments in common among the three European resistant genotypes are leading to more similar expression patterns. The strong reduction of commonly differentially expressed candidate genes from 112 to 22 after including Pa405 indicates the presence of different resistance genes in the three resistant European lines compared to Pa405. Both results corroborate the conclusion of haplotype analysis that Pa405 had unique haplotypes both in the *Scmv1* and *Scmv2* region comparing to European resistant lines (Xu et al. 2000). The standard Bonferroni correction is very stringent and may exclude many genes that are really significant, whereas the adjusted Bonferroni correction is less conservative, and more likely to include significant genes while still controlling the error rate (Pan 2002). This method should provide a reasonable balance between false-positive and false-negative rates for our analysis.

After inclusion of susceptible lines D408, D145, and F7, the seven inbreds were divided into two groups based on expression profiling data: D32, D21, FAP1360A, and D408, and Pa405, D145, and F7 (Figure 2A). As described before (Melchinger et al. 1998), D32, D21, FAP1360A, and D408 are early-maturing European Dent inbreds, whereas D145 and F7 are flint lines. Pa405 is a dent line, but from a different pool as the European dent lines. Thus the grouping of expression patterns among the inbreds was more according to dent-flint than resistant-susceptible. Population stratification can result in nonfunctional, spurious associations (Flint-Garcia et al. 2003), as described for studies on association mapping of maize. It would, therefore, be more meaningful meaningful to compare resistant vs. susceptible dent lines to avoid confounding with heterotic grouping. However, an expansion of this study is difficult due to the low number of available SCMV resistant inbred lines (three) identified in a large collection of European inbreds (Kuntze et al. 1995). Therefore, the most interesting genes are those differentially expressed according to resistance / susceptibility, including the 22 genes shared between the three resistant European inbreds and Pa405, but we cannot rule out especially for the 90 remaining genes only shared between the three resistant European inbreds but not Pa405, that they include some genes reflecting the dent – flint grouping. Among these 90 genes, those ones not shared with D408 might more likely be associated with SCMV resistance than those shared with D408. Because the expression pattern of D408 was closer to the resistant lines than F7 and D145, it is supposed that the *Scmv2* region might be already present in D408. Previous field trials also showed fewer susceptible plants were found in F2 populations with D408 as compared to F7 and D145 (Melchinger et al. 1998).

Table 2. List of genes in Cluster 1 in Fig. 3

GA [a]	Annotation [b]	Similarity [b]	Bin [c]	Ratio [d]					
				D21	D32	FAP 1360A	D145	D408	F7
Cell rescue, defense, cell death and ageing									
AI664965	PIR|T01412 hsp 22k precursor	100%		2.1	3.0	2.4	-1.1	1.2	1.0
AI714822	SP|Q08275 hsp 17.0k	100%		1.4	4.1	2.7	-1.2	1.2	1.2
AI738301	PIR|S22697 extensin	8%		12.0	9.5	9.3	1.4	9.1	2.1
AI770912	SP|P24631 hsp 17.5k	100%		2.9	2.1	3.0	1.2	1.7	1.0
AI795298	GP|18157649 RAD21-3	7%		2.4	2.2	3.2	1.0	1.6	1.0
AI861149	GP|2072553 salT gene product	16%	10.02	2.3	2.3	5.1	1.0	2.0	1.1
AI942105	GP|27817929 P0453E05.3	14%		2.7	1.6	3.1	-1.4	1.3	1.1
BM032385	GP|15624025 rbohA	44%		2.5	15.4	8.1	-1.1	1.2	1.1
BM337818	PIR|T02054 PR-1 protein	100%		2.5	1.7	2.3	-1.4	1.2	1.1
CD058538	PIR|G96806 thaumatin-like protein	44%		4.3	10.3	6.8	2.0	3.4	3.6
Cellular Organization									
BM429053	GP|8778400 F16A14.28	13%		1.4	4.0	1.4	-1.4	1.2	1.1
Metabolism									
AI622113	PIR|H84602 40S ribosomal protein S25	100%		1.6	2.1	2.7	1.1	-1.0	1.2
AI891183	GP|8978342 alcohol dehydrogenase-like	94%		2.1	4.0	2.7	-1.2	1.8	1.0
Signal transduction									
AW585276	GP|20161273 ankyrin-kinase -like	42%		2.5	2.2	7.2	-1.1	-1.0	1.1
BM277028	SP|P35182 phosphatase 2C homolog 1	6%		2.5	2.2	2.2	-1.1	1.3	1.1
BQ034241	GP|12060388 response regulator 6	57%		2.5	2.0	2.2	1.1	1.3	1.2
Transcription									
BM660017	PIR|T01643 DnaJ protein homolog	100%		2.7	2.3	2.2	-1.7	1.0	1.2
Unclassified proteins									
AI600827	dbj|BAB44108.1|{Oryza sativa}	71%	3.04	2.3	2.3	5.1	1.0	2.0	1.0
AI374523	GP|3955073 phytase	100%		2.0	1.5	2.3	-1.1	1.4	1.2
AI621767	GP|20186 calmodulin	100%		3.8	4.0	2.2	-1.6	1.4	1.0
AI795726	PIR|A30900 calmodulin	100%		2.6	2.8	2.7	-1.7	1.5	1.0
AI820398	PIR|T04498 protein homolog F8F16.130	47%		2.9	2.5	3.0	-1.3	-1.1	2.3
AI881638	GP|4097585 NTGP4	29%		3.7	2.4	2.3	-4.6	2.0	1.0
AI920449	Unknown			2.1	4.0	2.0	-1.9	1.3	1.1
AI941971	PIR|S33633 ribosomal protein CEP52	100%		3.0	3.6	3.6	1.2	1.7	1.1
AW017851	emb|CAA11391.1| phytase	44%	3.05	3.3	3.5	3.3	-1.5	1.7	1.1
AW257966	GP|21740628 OSJNBb0012E08.10	96%		2.2	2.5	2.5	-1.5	2.0	1.9
AW313218	Unknown		9.03	5.5	2.9	2.8	1.0	2.1	1.1
AW324587	GP|6635236 elicitor-inducible LRR	14%		8.8	59.6	2.5	1.6	1.7	1.1
BE345442	GP|22121720 SET102	100%	2.04	2.6	4.5	2.9	-1.0	1.3	1.9
BM074005	PIR|JC5445 glutaredoxin	95%		4.8	3.2	3.8	1.1	2.0	-7.7
BM334179	GP|10177015 ubiquitin-like protein	21%		7.2	4.0	2.5	0.8	2.0	1.6
BM500257	PIR|T51593 GTP-binding protein	25%		4.2	4.4	2.2	1.0	2.0	1.2
CA400607	GP|24413982 dioxygenase-like	78%		12.8	6.5	3.8	1.5	2.2	1.2
CA404973	GP|4099914 ethylene-responsive	29%		2.1	7.3	4.9	1.1	1.6	1.2
CD442188	Unknown			2.2	3.0	2.3	-1.8	1.6	1.8

[a] Genbank accession number. [b] Annotation of each gene sequence was taken from the TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). [c] Mapping information is from the Maize GDB (http://www.maizegdb.org) and the IDP mapping project (http://www.plantgenomics.iastate.edu/maizechip/), according to map bins. [d] If the ratio is less than one, the negative reciprocal is listed.

Candidate genes

Studies of expression profiling patterns across phenotypically distinct genotypes have the potential to identify candidate genes underlying QTL (Borevitz and Chory 2004). Of 19 mapped genes, five genes were located in bin 3.04 - 3.05, indicating that the chromosome regions surrounding *Scmv2* significantly contributes to SCMV resistance, despite of different haplotypes of the four resistant inbreds (Xu et al. 2000). Since these genes also showed different expression patterns

Table 3. List of genes in Cluster 3 in Fig. 3

GA [a]	Annotation [b]	Similarity [b]	Bin [c]	Ratio [d]					
				D21	D32	FAP 1360A	D145	D408	F7
Cell rescue, defense, cell death and ageing									
AI637093	GP|21741683 oj000126_13.7	15%		-5.5	-6.9	-5.8	3.3	-4.5	1.0
AI666166	PIR|S22697|S22697 extensin	8%		-3.3	-2.1	-2.1	1.1	-1.8	1.0
BG842726	GP|21104533 beta-1, 3-glucanase-like	84%		-2.2	-2.9	-2.2	1.0	-2.0	1.1
AW056039	SP|Q43250 Cytochrome P450 71C1	100%		-2.0	-5.1	-4.0	-1.4	1.4	1.1
AW261292	PIR|B42424 chitinaseprecursor	100%		-2.8	-2.8	-2.4	1.0	-2.6	1.1
AW574496	PIR|A42424 chitinaseA - maize	100%		-3.4	-3.9	-3.5	1.9	-2.1	2.9
BG835847	PIR|B42424 chitinaseprecursor	100%		-2.4	-2.1	-4.2	-2.0	-1.1	2.6
BG841429	PIR|S53051 glycine rich protein	56%	2.07	-2.3	-3.1	-2.1	1.0	-1.7	2.5
BI361048	SP|P24631 hsp 17.5k	100%	1.03	-4.2	-2.3	-2.5	-1.3	-1.7	1.1
Cellular organization									
AI770787	SP|Q42443 TRX-H	100%		-2.6	-2.7	-9.5	1.4	-1.4	1.6
AW461165	PIR|T04359 pectin methylesterase-like	56%		-2.7	-2.7	-2.2	3.3	-1.6	1.1
BM379389	SP|Q42443TRX-H	89%		-2.6	-2.8	-3.0	-1.7	-2.7	1.1
Energy									
AI677504	PIR|A00049 cytochrome c	100%		-5.1	-18.8	-3.8	-1.7	-1.9	1.0
AI947525	GP|17065918 aldehyde dehydrogenase	27%		-2.8	-2.9	-3.3	1.4	-2.6	2.3
AW330667	GP|19401698 uncoupling protein	100%		-4.5	-2.5	-2.4	4.5	-1.2	1.1
Metabolism									
AW927689	GP|15983466 At1g33990	19%		-2.4	-2.1	-2.7	3.2	-2.1	1.1
BM736737	PIR|T04567 T12H17.110	78%	10.02	-4.4	-3.0	-3.3	-1.2	-2.7	1.2
T18753	GP|4514655 DS3	43%		-2.1	-6.9	-5.4	16.8	1.0	3.3
Protein destination									
BM736555	GP|28140231 disulphide isomerase	90%		-3.4	-3.7	-4.8	1.1	-2.7	1.2
Signal transduction									
AI001320	Unknown		3.04	-2.6	-2.8	-3.0	-1.7	-2.7	1.0
AI745965	Unknown		10.03	-3.6	-2.7	-3.0	-1.6	-2.8	1.0
AW066479	GP|20977604 chromatin assembly factor 104	100%	2.07	-2.4	-2.1	-2.8	-1.1	-1.7	1.1
Transcription									
BM499990	PIR|I39161 dystonin isoform 2	7%		-16.8	-2.3	-2.9	1.3	-1.5	1.2
CD437477	GP|16930693 AT3g07810/F17A17_15	37%		-7.8	-5.1	-2.1	1.0	1.1	1.2
Unclassified proteins									
AI586715	GP|21740681 OSJNBb0011N17.1	56%		-3.8	-2.4	-2.9	1.2	-2.3	1.0
AI677406	GP|170396 calmodulin	95%		-2.5	-4.9	-6.1	1.7	-2.4	1.0
AI861106	GP|13873004 B1085F09.4	71%		-3.1	-3.2	-2.8	1.0	-2.8	1.0
AI881474	PIR|T04159 histone H1 homolog	53%	9.03	-2.3	-4.0	-2.9	-1.3	-1.4	2.6
AI920581	GP|20453106 At2g04030/F3C11.14	80%		-2.5	-3.5	-2.1	1.2	-1.4	1.1
AI973529	GP|29028836 At2g37790	27%		-2.4	-2.4	-2.9	-1.1	-1.3	1.1
AW147100	PIR|JQ1060 glycine-rich protein 1	25%		-4.3	-3.5	-2.5	1.4	-2.2	1.1
AW179540	Unknown			-3.8	-2.4	-2.9	1.2	-2.3	1.1
AW324638	Unknown			-2.5	-3.1	-4.9	1.3	-1.9	1.1
AW574438	PIR|T04731 cytochrome P450 homolog	32%		-4.6	-6.5	-4.0	-1.2	-2.2	1.1
AW585282	PIR|T03026|T03026 chitinase, acidic	96%		-2.4	-2.4	-2.6	1.2	-2.4	1.1
BE012214	PIR|T03026|T03026 chitinase, acidic	96%		-2.7	-2.0	-2.8	1.0	-2.2	3.1
BE056921	PIR|A42424 chitinaseA - maize	65%		-3.7	-2.1	-2.5	1.6	-2.1	3.1
BE186517	GP|500656 Yhr143wp	5%		-2.1	-3.3	-2.5	-1.1	-2.1	1.1
BE509606	Unknown			-3.3	-5.2	-3.5	2.0	-3.1	3.8
BE575027	GP|13160397 aldose reductase	88%		-3.8	-12.0	-2.6	-2.1	1.2	1.1
BG842342	GP|21741210 OSJNBb0086G13.5	72%	1.01	-4.0	-3.4	-2.8	1.1	-2.0	1.1
CD961191	PIR|T04498 AIG2 homolog F8F16.130	38%		-3.3	-4.4	-3.6	44.2	-3.2	1.2
CF047798	Unknown			-2.2	-2.0	-2.9	1.9	-1.6	1.8

[a b c d] The same footnote as Table 2.

between NILs $F7^{+}$ and F7 (Shi et al. 2004), they are candidate genes for *Scmv2* QTL. However, the genes co-localized with *Scmv1* QTLs (Shi et al. 2004) only differentially expressed in one or two of three European resistant lines, thus were excluded from the list of 112 genes. In addition, none of 19 mapped genes were mapped in bin 6.00 – 6.02. The possible explanation is the pleiotropy of resistant genes, despite of in that region the same haplotype shared within three resistant lines (Xu et al. 2000). Among five candidate genes co-localized with *Scmv2*, AW017851 and AI737330 show homology to phytase and troponin 1, respectively, whereas AI600827 and AI857233 are the homologs of predicted rice proteins and AI001320 has no annotation from TIGR Maize Gene Index (http://www.tigr.org/tdb/tgi/plant.shtml). Proximity of a candidate gene to a QTL provides only circumstantial evidence, because of the large physical size of QTL confidence

intervals on genetic maps. Therefore, to validate candidate genes, further fine mapping is in progress in our lab.

22 genes were most interesting, since they were differentially expressed between all resistant and susceptible inbreds lines. For instance, AW324587 is the homolog of EILP (elicitor inducible LRR protein) gene coding for a leucine-rich repeat (LRR) protein in tobacco. EILP is involved in both preexisting and inducible surveillance systems. And the product of EILP may be involved in non-host disease resistance in tobacco (Takemoto et al. 2000). BQ034241 and BM277028 show homology to response regulator 6 in maize and protein phosphatase 2C homolog 1, respectively. Both proteins are involved in signal transduction. BM660017 is DnaJ protein homolog ZMDJ1, which is a maize promoter driving GUS expression at moderate levels in a variety of seedling and mature plant tissues (Baszczynski et al. 1997). Among the 90 genes commonly differentially expressed between the 3 European resistant inbreds, many pathogenesis-related genes were identified. AW261292, AW574496, and BG835847 are homologs of chitinases (Datta and Muthukrishnan 1999), which can hydrolyze the cell wall of many fungi and the exoskeleton of invertebrates. CD058538 shows homology to a gene encoding thaumatin, induced by the presence of pathogenic molds and fungi and referred as pathogenesis-related proteins 5 (PR-5) (Hu and Reddy 1997). In addition, this list also includes a lot of genes without obvious relation to SCMV resistance, e.g., AI677504 (Cytochrome C) and AI947525 (Aldehyde dehydrogenase). However, defense mechanisms are energy intensive, and those genes might be induced or repressed to promote efficient energy utilization during defense reaction. More generally, it is consistent with massive changes in gene expression observed in other studies of plant response to pathogen infection (Katagiri 2004).

The information, generated from association analysis of expression profiling data across a collection of lines, can be passed on plant breeders in view of development of functional markers (Andersen and Lübberstedt 2003). Although morphological markers, e.g. mosaic symptoms for susceptible lines, are easily monitored, it can be affected by the environment and restricted by their limited numbers. With the advent of DNA markers, these limitations have been overcome. Thousands of phenotypically neutral, random DNA markers, such as RFLPs, SSRs and AFLPs, can be generated for any species and have been successfully used in many studies to represent genomes in biodiversity studies or to map trait genes. However, genetic linkage between a specific DNA marker and a target locus allele, established by QTL studies, for example, can be broken by genetic recombination. Functional markers based on expression profiling data are superior to random DNA markers owing to complete linkage with trait locus alleles, despite of high cost of marker development. Due to oligogenic inheritance of SCMV resistance, marker-assisted selection (MAS) programs with functional markers would increase the breeding efficiency.

Acknowledgements

We thank Olaf Neuschaefer-Rube and Michael Goetz for spotting macroarrays at GSF (Neuherberg). This work was supported by the German research foundation (DFG LU601/6-1).

References

Andersen, J. R. and Lübberstedt, T. 2003. Functional markers in plants. Trends Plant Sci 8(11): 554-60.

Baszczynski, C. L., Barbour, E., Zeka, B. L., Maddock, S. E. and Swenson, J. L. 1997. Characterization of a genomic clone for a maize DnaJ-related gene, ZmdJ1, and expression analysis of its promoter in transgenic plants. Maydica 42(2): 189-201.

Borevitz, J. O. and Chory, J. 2004. Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7(2): 132-6.

Cheung, V. G. and Spielman, R. S. 2002. The genetics of variation in gene expression. Nature Genetics 32: 522-525.

Datta, S. K. and Muthukrishnan, S. 1999. Pathogenesis-related proteins in plants. Boca Raton, CRC Press.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. and Siebert, P. D. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA 93(12): 6025-30.

Doerge, R. W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3(1): 43-52.

Dussle, C. M., Melchinger, A. E., Kuntze, L., Stork, A. and Lübberstedt, T. 2000. Molecular mapping and gene action of Scm1 and Scm2, two major QTL contributing to SCMV resistance in maize. Plant Breeding 119(4): 299-303.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-8.

Flint-Garcia, S. A., Thornsberry, J. M. and Buckler, E. S. t. 2003. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54: 357-74.

Fuchs, E. and Gruntzig, M. 1995. Influence of Sugarcane Mosaic-Virus (Scmv) and Maize-Dwarf Mosaic-Virus (Mdmv) on the Growth and Yield of 2 Maize Varieties. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection 102(1): 44-50.

Hu, X. and Reddy, A. S. 1997. Cloning and expression of a PR5-like protein from Arabidopsis: inhibition of fungal growth by bacterially expressed protein. Plant Mol Biol 34(6): 949-59.

Katagiri, F. 2004. A global view of defense gene expression regulation ?a highly interconnected signaling network. Current Opinion in Plant Biology 7: 1-6.

Kuntze, L., Fuchs, E., Gruntzig, M., Schulz, B., Henning, U., Hohmann, F. and Melchinger, A. E. 1995. Evaluation of maize inbred lines for resistance to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Agronomie 15(7-8): 463-467.

Kuntze, L., Fuchs, E., Gruntzig, M., Schulz, B., Klein, D. and Melchinger, A. E. 1997. Resistance of early-maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breeding 116(5): 499-501.

Louie, R., Findley, W. R., Knoke, J. K. and Mcmullen, M. D. 1991. Genetic-Basis of Resistance in Maize to 5 Maize-Dwarf Mosaic-Virus Strains. Crop Science 31(1): 14-18.

Mantel, N. 1967. Detection of Disease Clustering and a Generalized Regression Approach. Cancer Research 27(2P1): 209-&.

Melchinger, A. E., Kreps, R., Spath, R., Klein, D. and Schulz, B. 1998. Evaluation of early-maturing European maize inbreds for resistance to the European corn borer. Euphytica 99(2): 115-125.

Melchinger, A. E., Kuntze, L., Gumber, R. K., Lübberstedt, T. and Fuchs, E. 1998. Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm. Theor Appl Genet 96(8): 1151-1161.

Pan, W. 2002. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 18(4): 546-554.

Potokina, E., Caspers, M., Prasad, M., Kota, R., Zhang, H., Sreenivasulu, N., Wang, M. and Graner, A. 2004. Functional association between malting quality trait components and cDNA array based expression patterns in barley (Hordeum vulgare L.). Molecular Breeding 14(2): 153-170.

Quackenbush, J. 2002. Microarray data normalization and transformation. Nat Genet 32: 496-501.

Rohlf, F. J. 1998. NTSYSpc Numerical Taxonomy and Multivariate Analysis System. Setauket, New York, ExeterSoftware.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2): 374-8.

Shi, C., Ingvardsen, C., Thuemmler, F., Melchinger, A. E., Wenzel, G. and Lübberstedt, T. 2004. Identification of differentially expressed genes between maize near-isogenic lines in association with SCMV resistance using suppression subtractive hybridization. Mol Genet Genomics Revised.

Soukas, A., Cohen, P., Socci, N. D. and Friedman, J. M. 2000. Leptin-specific patterns of gene expression in white adipose tissue. Genes Dev 14(8): 963-80.

Takemoto, D., Hayashi, M., Doke, N., Nishimura, M. and Kawakita, K. 2000. Isolation of the gene for EILP, an elicitor-inducible LRR receptor-like protein, from tobacco by differential display. Plant and Cell Physiology 41(4): 458-464.

Xia, X. C., Melchinger, A. E., Kuntze, L. and Lübberstedt, T. 1999. Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89(8): 660-667.

Xu, M. L., Melchinger, A. E. and Lübberstedt, T. 2000. Origin of Scm1 and Scm2 - two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor Appl Genet 100(6): 934-941.

Acknowledgement

I would like to take this opportunity to thank the following persons for their special contributions, without many of which it would have been impossible to write this thesis. First and foremost, I would like to express my deep gratitude to my supervisor Dr. Thomas Lübberstedt for his excellent supervision, as well as outstanding organization and creative advices.

Sincere thanks to Prof. Dr. Gerhard Wenzel for giving me a precious chance to conduct research in his laboratory. I appreciate his generosity in every concern.

My regards to Dr. Fritz Thümmler for initiating the experiments and his continuous advice and scientific support.

A special tribute is addressed to all colleagues and technicians in the Chair of Plant Breeding at the Technical University of Munich, Freising, Germany. My special thank goes to Dr. Gerhard Schwarz and Dr. Volker Mohler, with whom I have shared the difficulties and problems of working. I particularly enjoy the friendship with Imad Zein and Anna Uzarowska in the workgroup of Quantitative Genomics.

I also want to thank the colleagues in Eureka project, especially Dr. Milena Ouzunova for her valuable advices.

Finally, I need to express my deepest gratitude and respect to my parents for their constant support and encouragement.

This list is far from exhaustive. I pray for forgiveness from those I did not mention by name and include them in my heart-felt gratitude.

Curriculum Vitae

Name	Chun Shi
Date of Birth	22.05.1976 in Shanghai
Nationality	China
Marital Status	Single

Education

09/1994-07/1998	Shanghai Fisheries University, China Awarded the degree of Bachelor
09/1998-03/2001	Fudan University, China Awarded the degree of Master
04/2001-09/2004	Technical University of Munich, Germany Doctorate candidate

Practical Experience

01/1998-06/1998	Research assistant at the School of Fisheries, Shanghai Fisheries University, China
09/1998-01/2001	Research assistant at the School of Life Sciences, Fudan University, China
02/2004-03/2004	Visiting scientist at the Danish Institute of Agricultural Sciences, Denmark