JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Effect of Transgenes on Global Gene Expression in Soybean Is within the Natural Range of Variation of Conventional Cultivars

K. C. CHENG,[†] J. BEAULIEU,[†] E. IQUIRA,[§] F. J. BELZILE,[§] M. G. FORTIN,[†] AND M. V. STRÖMVIK*,[†]

Department of Plant Science, McGill University, 21, 111 Lakeshore Road, Sainte Anne de Bellevue, Québec H9X 3V9, Canada, and 1243 Marchand Building, Département de Phytologie, Université Laval, Québec G1K 7P4, Canada

Current safety assessment for novel crops, including transgenic crops, uses a targeted approach, which relies on compositional analysis. The possibility that transgene expression could lead to unintended effects remains a debated issue. This study used transcriptome profiling as a nontargeted approach to evaluate overall molecular changes in transgenic soybean cultivars. Global gene expression was measured in the first trifoliate leaves of two transgenic and three conventional soybean cultivars using the soybean Affymetrix GeneChip. It was found that gene expression differs more between the two conventional cultivars than between the transgenics and their closest conventional cultivar investigated and that the magnitudes of differences measured in gene expression and genotype (determined by SSR analysis) do not necessarily correlate. A MySQL database coupled with a CGI Web interface was developed to store and present the results (http://soyxpress.agrenv.mcgill.ca/). By integrating the microarray data with gene annotations and other soybean data, a comprehensive view of differences in gene expression is explored between cultivars.

KEYWORDS: Pleiotropic effects; transgenic plants; gene expression

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is one of the economically most important crops in North America and worldwide, and to improve performance, many new cultivars are developed every year using traditional breeding and/or genetic engineering. These plants with novel traits have, for instance, improved seed quality and cold tolerance, as well as disease, pest, and herbicide resistance (*1*).

The safety of transgenic crops has been a debated topic since the mid 1900s when the first genetically modified (GM) crops were released on the market (*2*). The typical compositional analysis performed in the context of regulatory review of transgenic crops focuses on main constituents and does not normally allow ruling out of other pleiotropic changes. Unintended effects of modifications at the DNA level could ultimately cause modifications to the compounds or compound composition normally found in plants (*3*).

Current safety assessment of new crops is based on the concept of "substantial equivalence". If the chemical composition of new crops is substantially similar to that of existing crops, it is not considered to pose a health risk (*4*). In addition, compounds that could be generated through expression of an introduced gene are also analyzed. However, this is a targeted approach, which analyzes only a certain number of compounds for which the amount is speculated to vary as a result of transgene expression. Unknown and unintended effects on metabolism outside those specific compounds are typically not assessed. Because the current methods of plant transformation do not offer control over the insertion site, the number of copies transferred, or the integrity of the gene cassette, unintended effects may result from disruption of a functional gene at the point of insertion, rearrangements of the gene cassette, or ectopic co- expression of neighboring genes (*1, 5*). Modifications such as overexpression of transcription factors, introduction or alterations of biosynthetic pathways, and expression of transgenes to increase tolerance to biotic or abiotic challenges all carry the potential to give rise to unexpected interactions between gene products as well as increases or decreases in the availability and activity of other plant biochemicals. Therefore, nontargeted profiling approaches have been considered to be a better approach to assess changes in new crops and, hence, provide better tools to assess the safety of newly introduced traits. A number of recent studies in wheat and *Arabidopsis* have shown that DNA microarray technology, as a nontargeted and unbiased approach, is a promising tool to detect unintended effects (*6*−*8*).

* Author to whom correspondence should be addressed [telephone (514) 398-8627; fax (514) 398-7897; e-mail martina.stromvik@mcgill.ca].

[†] McGill University.

[§] Université Laval.

10.1021/jf073505i CCC: $40.75 © 2008 American Chemical Society
Published on Web 04/23/2008

A few studies have compared gene expression profiles of transgenic plants with their nontransgenic counterparts, but to date no study has presented comparisons of gene expression profiles in transgenic soybean. One of the most remarkable studies using microarray technology was done on transgenic *Arabidopsis* plants generated with simple T-DNA constructs containing the marker gene *nptII* and the reporter *uidA* (GUS), which were subjected to various environmental stresses (*7*). Only a small number of differentially expressed genes were found (varying between 39 and 180) between transformed and nontransformed control lines. These represented a very small portion (0.17−0.8%) of the genes screened using the Affymetrix ATH1 *Arabidopsis* GeneChip (22500 probe sets, representing the *Arabidopsis* transcriptome). The results showed that the insertion of the commonly used marker genes *nptII* and *uidA* has minimal effect on the global gene expression levels in transgenic *Arabidopsis* under optimal growth conditions and that the T-DNA insertion of the transgenes leads to very little functional disturbance to the genomes of transgenic plants. More importantly, the number of genes affected by the insertion of transgenes was significantly lower than the number of genes affected by common abiotic stresses such as heat, cold, salt, and drought (varying between 1080 and 4406). Also, when the gene expression profiles of transgenic lines were compared with the profile of the control line under abiotic stresses, the stress response was not different, meaning that the transgenes did not affect the stress response. The conclusion was that transgenic plants generated with simple T-DNA constructs containing common marker genes are equivalent to nontransgenic plants (*7*).

Two studies on wheat compared substantial equivalence of transgenic and nontransgenic crops at the transcriptome level (*6, 8*). The first study employed a 9K cDNA microarray to compare the gene expression profiles of the developing seeds in transgenic wheat transgenic for an *Aspergillus fumigatus* phytase gene with wild-type wheat (*6*). The results showed very slight variations for the three sampling time points (8, 16, and 32 days after pollination) but the differentially expressed genes could not be confirmed by real time RT-PCR. The authors concluded that the expression of the *A. fumigatus* phytase gene had no significant effects on the global gene expression pattern in the developing seeds of transgenic wheat (*6*).

The second study reported the comparison of gene expression profiles of transgenic and conventionally bred wheat lines that overexpress genes encoding high molecular weight subunits of glutenin (*8*). The same 9K cDNA microarray was used for pairwise comparisons between the transgenic wheat line, the conventionally bred wheat sister line, and the nontransgenic control line. The numbers of differentially expressed genes in the comparison between transgenic line and the nontransgenic line at 8, 14, and 28 days postanthesis were 6, 5, and 2, respectively. It represented only a small proportion (0.06, 0.05, and 0.02%) of the genes spotted on the microarray. In the comparison between the conventionally bred line and the nontransgenic line, the number of differentially expressed genes varied from 26 to 527 (0.27−5.59%). In the comparison between the transgenic line and conventionally bred lines, the number of differentially expressed genes varied from 4 to 154 (0.04−1.63%). The results showed that transgenic manipulation led to very small changes in expression profiles. Most importantly, there were greater differences in gene expression due to conventional breeding than genetic modification in transgenic wheat. This implied that the presence of the transgene and associated T-DNA with marker and reporter genes has a smaller impact on global gene expression patterns than gene recombination through conventional breeding. As with the previous study, the conclusion is that a single transgene has minimal effects on the transcriptome and that a transgenic crop can be substantially equivalent to the control nontransformed line (*8*).

In this study, we have profiled five soybean cultivars at the genomic level using simple sequence repeats (SSR) marker analysis and at the gene expression level using DNA microarrays. The soybean lines selected for this project included both conventional and transgenic lines. To explore and interpret potential differences in gene expression between cultivars, we also developed a database to allow users to retrieve data and results of the microarray experiments with cross-referenced annotations of the expressed sequence tags (EST) and hyperlinks to external public databases. We show that consistent with previous reports in other crops (*6, 8*), soybean cultivars transgenic for glyphosate tolerance are not necessarily substantially different from conventional cultivars and that the global gene expression in the cultivars tested is minimally affected by the insertion of a transgene.

This is the first study that attempts to link "substantial equivalence" with general plant composition in transgenic soybean, and the database we report provides new tools to assist researchers and regulators in assessing the frequency and magnitude of changes in plant composition as well as to develop an understanding of the biological significance of these changes.

MATERIALS AND METHODS

Plant Material. Five mid- to late-maturing soybean cultivars were used: OAC Bayfield, developed by the University of Guelph and registered in 1993 (*9*); S03-W4, developed by Syngenta Seeds Inc. (Minneapolis, MN) and registered in 1998; 2601R, a descendant of the glyphosate-resistant 40-3-2 line and registered by First Line Seeds Ltd. (Guelph, ON, Canada) in 1998 (*9*); PS46RR, another descendant of 40-3-2 registered by First Line Seeds Ltd in 2000 (*10*); and Mandarin (Ottawa), obtained from the Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada. Plants were grown in growth chambers in ambient humidity, 16 h photoperiod, and 25/19 °C day/night temperatures. At the V2 growth stage (*11*), completely unrolled first trifoliate leaves were harvested by cutting the petiole a few millimeters below the leaflets, immediately frozen in liquid nitrogen, and stored at −80 °C. DNA was extracted, and Polymerase Chain Reaction (PCR) was performed to confirm the presence of CP4-EPSPS in the transgenic lines and its absence in nontransgenic lines using primers (*sttmf3a* and *sttmr2a*) designed by Padgette et al. (*12*).

Soybean Genotype Analysis Using SSR Markers. Twenty-nine SSR loci were used in this study on the basis of their polymorphism and distribution across the 20 soybean chromosomes (Supporting Information). Standard PCR was performed using the following primers: a forward primer carrying a 17 bp M13 tail, a reverse primer, and an IRD700-labeled M13 primer. PCR products were separated on a LI-COR Global Edition IR^2 DNA Analyzer (Lincoln, NE) and Saga Generation 2 software was used to visualize and estimate allele sizes. For each allele at each of the 29 SSR loci, data were scored as "1" for presence and "0" for absence. Genetic distances between each pair of cultivars were estimated using PAUP v. 4.0b10 (*13*). Similarity coefficients based on SSR profiles were calculated according to the method of Nei and Li (*14*).

GeneChip Expression Profiling. Total RNA was prepared from trifoliate leaves using the RNeasy Plant Mini Kit (Qiagen). RNA integrity was tested for each sample using the Agilent 2100 bioanalyzer (Palo Alto, CA). The five first-quality samples of RNA of each of the five soybean cultivars were further selected for hybridization to Affymetrix Soybean GeneChips (total of 25 chips). Target preparation, hybridization and scanning were carried out at the McGill University and Genome Quebec Innovation Center Microarray platform using the protocol recommended by Affymetrix (*15*).

Microarray Data Processing and Database Construction. All computations were performed on a Mac Power PC G5 running Mac OS X operating system version 10.4.9 equipped with 8 GB RAM. Perl (version 5.8.6) (www.perl.com) scripts were written for parsing data files and to load data into a MySQL relational database (version 5.0.18) (http://www.mysql.com). The core tables for the sequence data are based on the ESTIMA database (*16*). Perl CGI (http://search.cpan.org/dist/CGI.pm/) scripts were used to create the Web interfaces, and the Perl module DBI and DBD::mysql (http://dev.mysql.com/downloads/dbi.html) were used to connect CGI scripts to our database, and the CGIwithR package (*17*) was used for running R (*18*) statistical analysis within the CGI script. A total of 380,095 soybean EST sequences were annotated using SwissProt (*19, 20*) and BLAST (*21*), Gene Ontology (*22*), Enzyme Commission numbers (http://www.chem.qmul.ac.uk/iubmb/enzyme/), and Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways (*23*) and also linked with tentative consensus information from TIGR (*24*). The sequences and annotations were integrated with the microarray data. All raw microarray data (e.g., probe intensities), preprocessed data (e.g., normalized probe-set intensities), and results from statistical analysis (e.g., fold change, statistics score, and *p* value) were organized and stored in the database. Information about the microarray GeneChip such as probe sequences, probe location of the chip, and corresponding GenBank accession number of the probes were integrated into the soybean EST and annotation database to describe the microarray data. Probe identifiers were linked to GenBank accession numbers, which were further linked to the SwissProt protein identities and functional annotations. Information on the probes (such as soybean probe sequences, consensus sequences of the probes, probe locations on the chip) was downloaded from the Affymetrix Web site (http://www.affymetrix.com/support/technical/byproduct.affx?product=soy). The full database schema and architecture will be described elsewhere.

Microarray Data Preprocessing. R (*18*) and the BioConductor packages (*25*) such as *affy*, *limma*, *cluster*, and *made4* were used for data analysis. Quality assessment of the microarray data was carried out using the *affyRNAdeg* function from the *affy* package. All 61,170 probe sets (138,734 probes) including control probes and probes for *Glycine max*, *Phytophthora sojae*, and *Heterodera glycines* were preprocessed and normalized together. The microarray data were preprocessed using three different normalization methods: RMA (*26*), MAS5 (*27*), and dChip (*28*).

Gene Expression Data Analysis. Soybean cultivar-specific patterns of gene expression (classifying them into groups) were examined by principal component analysis (PCA) for the 25 nonprocessed chips, and unsupervised hierarchical clustering for the 25 normalized chips was carried out using Euclidean distance and average linkage. The most closely related pairs of transgenic/nontransgenic soybean cultivar were defined from the cluster analysis. Pairwise comparison for every two cultivars was done using Linear Models for Microarray Data (LIMMA) (*29*) at a p value of <0.01 and a fold change of >2. The RMA processed data are in log 2 base; MAS5 and dChip processed data are in log 10 base. Before using LIMMA, MAS5 and dChip processed data were transformed to log 2 base for statistical analysis. For calculation of the differences of intensity in two samples by fold change, RMA processed data were transformed to log 10 base. All microarray data, including raw intensities from CEL files, preprocessed data using three normalizations and summarized methods, log transformed intensities, t scores and p values from LIMMA analysis, and information about the probes, were stored and integrated into the soybean database.

RESULTS

SSR Genotyping Shows the Distance between Soybean Cultivars. Five different mid- to late-maturing soybean cultivars, OAC Bayfield, Mandarin (Ottawa), S03W4, 2601R, and PS46RR, were selected and genotyped using 29 SSR markers (see Supporting Information). As seen in **Table 1**, overall, the genetic distances between pairs of soybean cultivars ranged between 0.074 and 0.438, with the shortest distance being observed between 2601R and OAC Bayfield and the greatest distance between Mandarin (Ottawa) and S03W4. OAC Bayfield was the conventional cultivar most closely related to both transgenic cultivars (0.074 vs 2601R and 0.193 vs PS46RR).

Table 1. Distance Matrix (Nei–Li) of Five Soybean Cultivars Based on 29 SSR Markers Distributed over 20 Soybean Chromosomes

	2601RR	Mandarin (Ottawa)	P546RR	SO3W4
Mandarin (Ottawa)	0.178			
P546RR	0.183	0.282		
SO3W4	0.278	0.438	0.239	
OAC Bayfield	0.074	0.130	0.193	0.236

Transcriptome Profiling in Soybean First Trifoliate Leaves. To profile the transcriptome in the first trifoliate leaves of five different soybean cultivars under one growth condition, RNA was extracted from five individual plants of each cultivar and used for hybridization to the soybean Affymetrix GeneChip. The integrity of the RNA samples was assessed using *affy RNAdeg* functions from the R-Bioconductor *affy* package without transforming the intensities of probes to log 2 base. Five RNA samples from two cultivars (S03W4-1, S03W4-3, 2601R-2, 2601R-4, and 2601R-6) were shown to have a higher degree of RNA degradation than the rest of the samples (data not shown), but hybridization intensity remained strong. **Figure 1a**, shows the results of group classification using PCA for nonprocessed data from the 25 chips. Mandarin (Ottawa) forms a group separate from the other samples, whereas the more degraded samples grouped together, and the remaining 15 samples from transgenic and nontransgenic soybean cultivars form one heterogeneous group. It is clear that RNA quality had a large impact on hybridization results.

Using RMA-normalized data for the 25 chips in hierarchical clustering using Euclidean distance and average linkage, results similar to those obtained from the PCA are generated (**Figure 1b**). All five samples of Mandarin (Ottawa) form a separate cluster, whereas the five samples that displayed more degradation (S03W4-1, S03W4-3, 2601R-2, 2601R-4, and 2601R-6) cluster into a separate group; the other 15 samples of nontransgenic and transgenic soybeans form yet another cluster. This suggests that the variation in gene expression between four cultivars [excluding Mandarin (Ottawa)] is very small. Except for the samples with some RNA degradation, all samples of the transgenic 2601R and PS46RR cultivars grouped with OAC Bayfield and with S03W4 in both the PCA and hierarchical clustering. The two transgenic cultivars do not cluster into a separate group from the nontransgenic cultivars, and so the transgenic cultivars analyzed cannot be said to be different from the nontransgenic cultivars in these group classifications based on gene expression.

Pairwise Comparisons Define Gene Expression Variations among the Five Soybean Cultivars. Further comparison of gene expression profiles used two different approaches: (1) pairwise comparison of a transgenic cultivar with its closest related nontransgenic counterpart (as per the group clustering analysis) and (2) comparison of each transgenic cultivar with a group of nontransgenic cultivars based on the concept of "substantial equivalence". To make exhaustive intercultivar comparisons, pairwise comparisons using LIMMA were carried out between each soybean cultivar. **Figure 2** shows that data preprocessed with different methods (RMA, MAS5, and dCHIP) follow similar trends. Mandarin (Ottawa) displayed the largest number of differentially expressed genes, in relation to other cultivars, among all five cultivars. After processing with the RMA or MAS5 normalization methods, more than 1000 genes (from 37583 total soybean genes on the Affymetrix chip) are

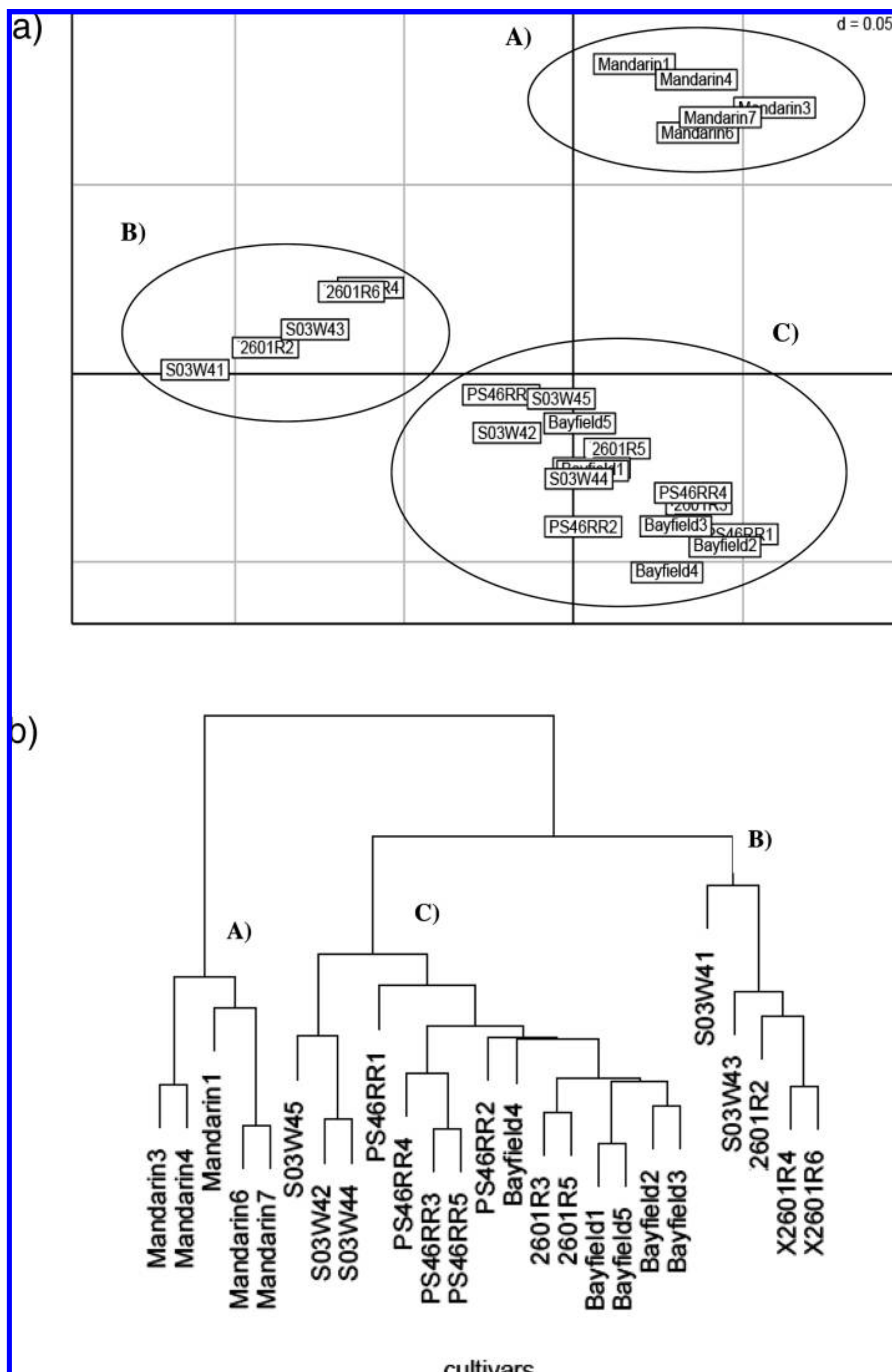

Figure 1. Group comparisons of microarray data from five soybean cultivars: (**a**) PCA of the 25 microarrays (raw data) from the five cultivars. Mandarin forms a separate group (A) from the other soybean cultivars, and the five samples (S03W4 and 2601R mixed) with lower RNA integrity form one group (B). Soybean cultivars cannot be classified into independent gene expression groups based solely on whether they are transgenic or not as most samples (15) form one group, mixed of transgenic (2601R, PS46RR) and nontransgenic cultivars (OAC Bayfield, S03W4) (C) (distance = 0.05). (**b**) Hierarchical clustering for 25 samples of five cultivars using all probes on the arrays. Three main clusters are distinguished: (A) Mandarin (Ottawa) samples; (B) poor-quality samples of S03W4 and 2601R; (C) remaining 15 transgenic and nontransgenic samples of 2601R, PS46RR, S03W4, and OAC Bayfield.

differentially expressed at p values of <0.01 and intensities of >2-fold change in the comparisons between Mandarin and any other cultivar. The other four cultivars, OAC Bayfield, S03W4, 2601R, and PS46RR, are less different from each other (with fewer than 350 differentially expressed genes of a total of 37583 genes). After RMA preprocessing, only 44 genes were differentially expressed between OAC Bayfield and 2601R, whereas 109 genes were differentially expressed when OAC Bayfield was compared to PS46RR. The number of differentially expressed genes between OAC Bayfield and each of the two transgenic cultivars was less than the number of differentially expressed genes between the two transgenic cultivars (137 differentially expressed genes). In the comparison of the other nontransgenic cultivar S03W4 to both transgenic cultivars, there were 248 genes differentially expressed when S03W4 was compared to 2601R; and 290 genes were differentially expressed

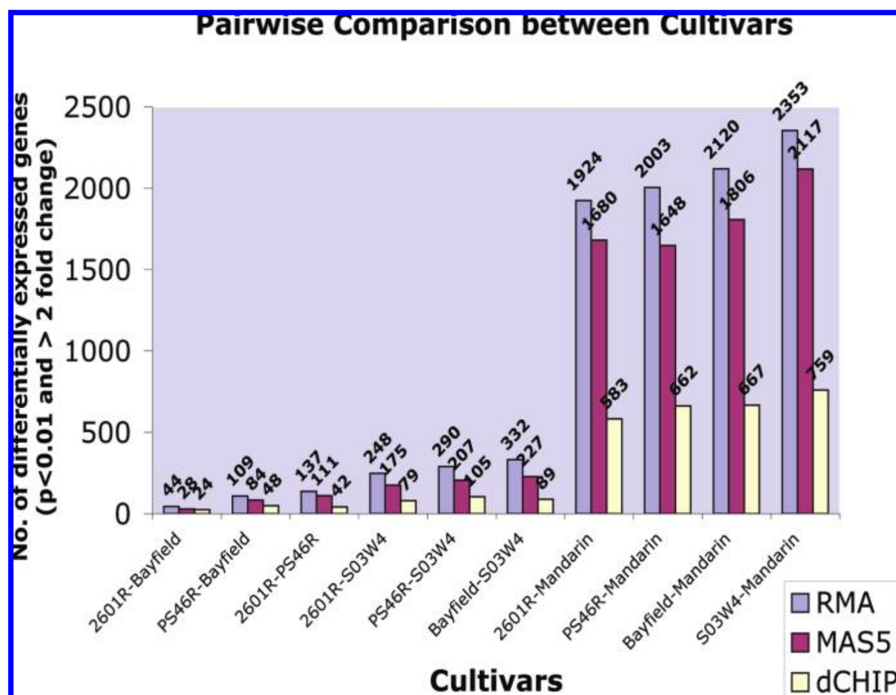

Figure 2. Pairwise comparison between five different soybean cultivars: LIMMA analysis on three sets of different preprocessed microarray data (using RMA, MAS5, or dCHIP). The numbers of differentially expressed genes (p value < 0.01, intensities > 2-fold change) are located above the bars.

when S03W4 was compared to PS46RR. The differences between transgenic and nontransgenic soybeans were smaller than the differences between two nontransgenic soybeans (332 differentially expressed genes). On the basis of the fewest differentially expressed genes, OAC Bayfield is again shown likely to be the closest related nontransgenic cultivar to each of the transgenic cultivars.

Comparison of Each Transgenic Cultivar to the OAC Bayfield (Nontransgenic) Cultivar. Because OAC Bayfield was the cultivar found to be the closest conventional relative to the two transgenic cultivars used in this study, it was compared individually with transgenic cultivars 2601R and PS46RR to test for differential gene expression. Within the 44 and 109 differentially expressed genes found in the previous pairwise comparison of 2601R and PS46RR to OAC Bayfield (using the RMA preprocessing method, cutoff at p value < 0.01 and fold change > 2), only 8 genes were differentially expressed in common in both transgenic cultivars (**Table 2**). Only three of these genes are annotated with Gene Ontology (GO) terms. Two genes that were down-regulated were in the category "cysteine protease inhibitor activity", and one gene that was down-regulated was in the category "dihydroflavonol-4-reductase activity". One of the up-regulated genes belonged to a tentative contig annotated as "cinnamoyl-CoA reductase activity". To understand the molecular function of these genes, pairwise comparisons in each of the transgenic cultivars were interpreted in terms of GO molecular function (using parent terms that describe the functions in more general annotations). In the comparison between 2601R and OAC Bayfield using the RMA preprocessing method (cut-off at p value < 0.01 and fold change > 2), 2 genes were identified as involved in "endopeptidase inhibitor activity", 5 genes were related to "transferase activities", 5 genes were involved in "binding", and 1 gene was involved in each of the functions "lyase activity", "signal transducer activity", "isomerase activity", "oxidoreductase activity", "transporter activity", and "hydrolase activity". The results were similar in the comparison of PS46RR to OAC Bayfield using the same method: 2 genes were involved in "endopeptidase inhibitor activity", 13 genes in "transferase activity", 11 genes in "binding", 5 genes in "hydrolase activity", 5 genes in "oxidoreductase activity", 3 genes in "signal transducer activity", and 1 gene each in "transporter activity and "nutrient reservoir activity".

Comparison of Each Transgenic Cultivar to a Group of Nontransgenic Cultivars. The possibility of applying the concept of substantial equivalence in microarray experiments was evaluated in our second approach by grouping the data from the three nontransgenic cultivars together as the reference group and comparing the gene expression with each of our two transgenic cultivars using LIMMA at p value < 0.01 and fold change > 2. The concept of in vitro substantial equivalence is based on the concept of defining the molecular "signature" of a species and then using this consensus "signature" as a reference against which new lines or new cultivars are compared. Existing cultivars currently on the market are recognized as "GRAS" (Generally Recognized as Safe), and new cultivars are then compared to that collective consensus composition to assess the safety of new lines. In this approach, comparison of one transgenic soybean to a reference group (the GRAS group) is used to identify deviation from the norm. As a GRAS comparator, we used the three nontransgenic soybean lines instead of pairwise comparisons. The number of differentially expressed genes was reduced from 44 to 10 genes in 2601R and from 109 to 49 genes in PS46RR. There were only five genes differentially expressed in common in both of the two transgenic cultivars (2601R and PS46RR). However, only two of these were assigned GO annotations, both involved in "cysteine protease inhibitor activity". **Table 3** shows the 10 differentially expressed genes in 2601R. Three of them (protein sequences similar to flavonol 3-*O*-glucosyltransferase, phytochrome A, and zeatin *O*-xylosyltransferase, or indole-3-acetate β-glucosyltransferase, respectively) are involved in "transferase activity". Phytochrome is also related in "binding" and "signal transducer activity". Two other genes (protein sequences similar to cysteine proteinase inhibitor A and multicystatin) are involved in "endopeptidase inhibitor activity". Nine of these 10 genes were

Table 2. Differentially Expressed Genes Common to Both Transgenic Soybean Cultivars (2601R and PS46R), Compared with Nontransgenic Soybean (OAC) Bayfield

probe set ID	protein ID/contig ID	protein/Contig description	GO terms	fold change (log 2) (2601R)	*p* value (2601R)	fold change (log 2) (PS46R)	*p* value (PS46R)
Gma.3314.1.S1_a_at	CYTM_SOLTU	(P37842) multicystatin	(GO:0004866) endopeptidase inhibitor activity (GO:0004869) cysteine protease inhibitor activity	−4.157	5.88E−11	−4.018	6.12E−12
Gma.3314.2.S1_x_at	CYTA_HELAN	(Q10992) cysteine proteinase inhibitor A	(GO:0004866) endopeptidase inhibitor activity (GO:0004869) cysteine protease inhibitor activity	−5.185	5.88E−11	−5.087	1.43E−11
GmaAffx.18584.1.S1_x_at				−2.674	1.89E−09	−2.370	1.29E−08
Gma.5206.1.A1_at	TC209225	Rev interacting protein mis3-like (partial 19%)		2.413	5.07E−08	1.922	1.38E−06
GmaAffx.78465.1.S1_s_at	TC226919	cinnamoyl CoA reductase-like protein (partial 52%) (EC 1.2.1.44)		1.686	0.0003	1.387	0.0023
GmaAffx.57421.1.S1_at	TC221352			1.103	0.0011	1.249	0.0001
GmaAffx.52672.1.S1_at	TC217896	replication factor C 110 kDa subunit (partial 18%)		2.153	0.0029	1.758	0.0039
Gma.15664.1.S1_at	DFRA_VITVI	(P51102) dihydro flavonol-4-reductase (EC 1.1.1.219)	(GO:0009813) flavonoid biosynthesis (GO:0016491) oxido reductase activity (GO:0045552) dihydro kaempferol 4-reductase activity	−1.449	0.0030	−1.355	0.0033

also differentially expressed in the comparison using only OAC Bayfield as the comparator except the probe GmaAffx.52838.1.S1_at, which has a *p* value of 0.06 in comparison with OAC Bayfield. However, it was significantly down-regulated in comparison with Mandarin (Ottawa) and S03W4 at p value < 0.0002. Unfortunately, there were no similar sequences found in the BLAST search using *e* value < 0.01, so no annotation for this gene and no information of its molecular function were provided. **Table 4** shows 12 (from 49) differentially expressed genes that have GO term annotations in comparison of PS46RR with the reference group of 3 nontransgenic soybeans [OAC Bayfield, Mandarin (Ottawa), and S03W4]. Six of the differentially expressed genes are involved in the molecular function "binding"; four genes in "transferase activity"; and two genes in "cysteine protease inhibitor activity". Some of the genes have multiple functions such as GmaAffx.55247.1.S1_at and GmaAffx.54889.1.S1_at, which are involved in both "transferase activity" and "binding". Most of these 12 genes are also differentially expressed when using only OAC Bayfield for comparison, except in the cases of Gma.16328.1.S1_at and Gma.2590.1.A1_s_at, which have fold changes between PS46RR and Bayfield that are slightly below 2 (i.e., 1:1.92 and 1:1.82, respectively) and consequently cannot be said to be differentially expressed with the same strict criteria.

DISCUSSION

In this study, we have tested the concept of using molecular data as a source of information on the composition of plant products. The current food safety approach for new plant products is based on the comparison of basic food compositional data (for example, proteins and lipids) to assess the risk of deleterious effects. The justification for the use of very few compounds for a content comparison is rooted in a risk management approach that focuses on known risk factors. In contrast, pleiotropic effects following transgene expression cannot be known a priori and, hence, can be tested for only by using a genome-wide (transcriptome, proteome, or metabolome) analysis. Compositional (targeted) analysis has previously shown that seeds from glyphosate-resistant soybean and seeds from their isogenic conventional counterparts are substantially equivalent (*30*). The first trifoliate leaf stage was chosen in our study because this is a common growth stage at which glyphosate application is made in the field. Alterations in metabolic pathways in leaves that act as source tissues would have an impact on the general state of a plant and so could have an effect on the final seed composition. Previous studies show, however, that in the case of the soybean vegetative storage proteins (VSPα/β), the plant can compensate for altered temporary N storage composition in leaves, leaving the final seed composition virtually unaltered (*31*).

The results reported in this study demonstrate that the insertion of a transgene need only have minimal effects on global gene expression. The transcriptomic divergence was more pronounced between the conventional cultivar Mandarin (Ottawa) and the other cultivars as more differentially expressed genes were found in pairwise comparisons and in cluster analysis. Mandarin (Ottawa) was released in 1934 (*32*) and has a longer history of commercialization than the other four soybean cultivars. Although Mandarin (Ottawa) is a major ancestor of North American soybean cultivars and has contributed 11−22% to the genomes of present-day northern soybean elite lines (*33, 34*), its contribution to the northern gene pool has been reduced in the past 10−15 years (*34*). Therefore, the older soybean cultivar might be more distant in terms of global gene expression [the expression levels of over 1000 genes were measured to be significantly different (p value < 0.01 and fold change > 2)] compared to the recently developed cultivars, which are more inbred and have more desirable traits in common. The other soybean cultivars tested in this study were very similar in gene expression patterns.

Table 3. Differentially Expressed Genes in Transgenic Soybean 2601R, Compared with a Reference Group of Three Nontransgenic Soybean Cultivars [OAC Bayfield, S03W4, and Mandarin (Ottawa)]

probe set ID	protein ID/contig ID	protein description	GO terms	fold change (log 2)	*p* value
GmaAffx.75645.1.A1_at	UFOG4_MANES	(Q40286) flavonol 3-*O*-glucosyltransferase 4 (EC 2.4.1.91)	(GO:0016740) transferase activity (GO:0047893) flavonol 3-*O*-glucosyltransferase activity	−3.723	1.36E−13
GmaAffx.84169.1.A1_at	ZOX_PHAVU	(P56725) zeatin *O*-xylosyltransferase (EC 2.4.2.40)	(GO:0016740) transferase activity	2.367	4.79E−11
GmaAffx.70608.1.S1_at	IAAG_MAIZE	(Q41819) indole-3-acetate β-glucoxyltransferase (EC 2.4.1.121)	(GO:0050404) zeatin *O*-β-D-xylosyltransferase activity (GO:0047215) indole-3-acetate β-glucosyltransferase activity (GO:0047215) indole-3-acetate β-glucosyltransferase activity	−3.059	1.26E−05
Gma.13345.1.S1_at				−1.260	1.38E−05
GmaAffx.57421.1.S1_at	TC221352			1.074	0.0002
Gma.17524.1.S1_at	PHYA_SOYBN	(P42500) phytochrome A	(GO:0016740) transferase activity (GO:0000155) two-component sensor activity (GO:0005524) binding (GO:0005524) ATP binding (GO:0004871) signal transducer activity (GO:0008020) G-protein coupled photoreceptor activity	−1.229	0.0002
Gma.3314.2.S1_x_at	CYTA_HELAN	(Q10992) cysteine proteinase inhibitor A (cystatin A)	(GO:0004866) endopeptidase inhibitor activity	−4.344	0.0022
Gma.13860.1.A1_at			(GO:0004869) cysteine protease inhibitor activity	1.419	0.0031
Gma.3314.1.S1_a_at	CYTM_SOLTU	(P37842) multicystatin (MC)	(GO:0004866) endopeptidase inhibitor activity	−3.385	0.0065
GmaAffx.52838.1.S1_at			(GO:0004869) cysteine protease inhibitor activity	−1.102	0.0077

In terms of genetic distance, Mandarin (Ottawa) also exhibited the two highest pairwise genetic distance values (0.438 and 0.282), suggesting that it is genetically quite distant from P546RR and S03W4. Overall, however, it is S03W4 that exhibited the greatest average genetic distance relative to the four cultivars to which it was compared. Therefore, genetic distance, as measured by SSR markers, is not highly correlated with the distance resulting from cluster analysis based on similarity in gene expression profiles.

Our hierarchical clustering analysis could not distinguish the group of transgenic soybean cultivars (2601R and PS46RR) from the other [non-Mandarin (Ottawa)] nontransgenic cultivars (OAC Bayfield and S03W4), and fewer than 332 genes (approximately 1% of the soybean genes arrayed) differed significantly (p value < 0.01) with expression levels >2-fold in any pairwise comparisons among these four cultivars. Interestingly, the number of differentially expressed genes between nontransgenic cultivars (OAC Bayfield/S03W4) was higher than the number of differentially expressed genes between transgenic and nontransgenic soybeans. It has previously been shown that a single insertion of T-DNA and common reporter genes did not appreciably affect gene expression profiles in transgenic *Arabidopsis* plants (*7*), also shown in a study of the effect of a dhurrin in *Arabidospis* (*35*). The result is similar to previous findings in crop plants as well—the expression of *A. fumigatus* phytase had minimal effect on gene expression patterns in transgenic wheat seedlings (*6*)—and also similar to a recent cDNA microarray study in wheat lines expressing genes encoding high molecular weight subunits of glutenin (*8*). Our results are in strong agreement with these reports, and taken together, these data suggest that the expression of a transgene generally has little impact on the transcriptome. Similar results were seen with RoundUp-treated sensitive and resistant soybean seedlings (Patrick Tranel, personal communication). This broad observation must of course be tempered by the limited number of transgenes that have been expressed so far in studies of pleiotropic impact of transgene expression.

The inclusion of cultivar Mandarin (Ottawa) in this study generated variation in gene expression levels among soybean cultivars. Its omission would have narrowed the substantive range of gene expression levels and therefore increased the relative variation caused by the expression of a transgene. The same observation could in fact have been made if levels of a series of plant compounds had been measured in more conventional plant composition analyses usually performed to assess the safety of transgenic crops. The reference used for comparing a plant with novel traits with GRAS cultivars determines the outcome of the comparison. Hence, the definition of what is "soybean" for the purpose of creating a reference set for the crop species is an important question. Comparisons of transgenic cultivars to sister cultivars simplify the laboratory analysis aimed at establishing the equivalence of a new transgenic cultivar, but risk overestimating differences by ignoring a large range of variation observed (such as reported here) when different soybean cultivars are used. This is illustrated by results obtained in this study where comparison between 2601R and OAC Bayfield and between PS46RR and OAC Bayfield showed more divergence in gene expression than when the comparator was the group of conventional soybean lines.

The application of the test for substantial equivalence between plants with novel traits and transgenic lines using gene expression experiments as a measure will require that a database of the natural range of variation for each crop cultivar be established, just as databases of the typical content

Table 4. Differentially Expressed Genes That Have GO Term Annotations in Transgenic Soybean PS46R, Compared with a Reference Group of Three Nontransgenic Soybean Cultivars [OAC Bayfield, S03W4, and Mandarin (Ottawa)]

probe set ID	protein ID	protein description	GO terms (molecular function category)	fold change (log 2)	*p* value
Gma.4564.1.A1_at	RGA2_SOLBU	(Q7XBQ9) disease resistance protein RGA2 (RGA2-blb) (blight resistance protein RPI)	(GO:0005524) ATP binding	5.909	1.20E−28
Gma.16328.1.S1_at	ROC1_NICSY	(Q08935) 29 kDa ribonucleoprotein A, chloroplast precursor (CP29A)	GO:0003723) RNA binding	−1.064	1.16E−08
GmaAffx.55247.1.S1_at	SUVH9_ARATH	(Q9T0G7) probable histone-lysine *N*-methyltransferase, H3 lysine-9 specific 9 (EC 2.1.1.43) (histone H3-K9 methyltransferase 9)	(GO:0018024) histone-lysine *N*-methyltransferase activity	1.031	3.54E−08
		(H3-K9-HMTase 9) (suppressor of variegation 3-9 homologue 9) [Su(var.)3-9 homologue 9]	(GO:0008270) zinc ion binding		
GmaAffx.86027.1.S1_at	HCBT3_DIACA	(O23917) anthranilate *N*-benzoyltransferase protein 2 (EC 2.3.1.144) (anthranilate *N*-hydroxycinnamoyl/benzoyltransferase 2)	(GO:0008415) acyltransferase activity (GO:0047672) anthranilate *N*-benzoyltransferase activity	1.804	3.12E−07
GmaAffx.784.1.A1_at	CWF26_SCHPO	(O94417) cell cycle control protein cwf26	(GO:0000398) nuclear mRNA splicing via spliceosome	−1.437	5.43E−07
GmaAffx.54889.1.S1_at	WRK52_ARATH	(Q9FH83) probable WRKY transcription factor 52 (WRKY DNA-binding protein 52) (disease resistance protein RRS1) (resistance to *Ralstonia solanacearum* 1 protein)	(GO:0003677) DNA binding (GO:0005524) ATP binding	2.076	5.75E−07
			(GO:0004888) transmembrane receptor activity (GO:0004674) protein serine/threonine kinase activity (GO:0017111) nucleoside-tri phosphatase activity		
Gma.15686.1.A1_at	PIP22_ARATH	(P43287) aquaporin PIP2.2 (plasma membrane intrinsic protein 2b) (PIP2b) (TMP2b)	(GO:0005215) transporter activity	−1.760	2.40E−05
Gma.2590.1.A1_s_at	RS24_ARATH	(Q9SS17) 40S ribosomal protein S24	(GO:0003735) structural constituent of ribosome	−1.123	0.0001
Gma.8137.1.S1_at	BAK1_ARATH	(Q94F62) brassinosteroid insensitive 1-associated receptor kinase 1 precursor (EC 2.7.1.37) (BRI1-associated receptor kinase 1) (somatic embryogenesis receptor-like kinase 3)	(GO:0004674) protein serine/threonine kinase activity (GO:0005524) ATP binding	1.128	0.0002
Gma.3314.2.S1_x_at	CYTM_HELAN	(Q10992) cysteine proteinase inhibitor A (cystatin A) (SCA)	(GO:0004869) cysteine protease inhibitor activity	−4.247	0.0003
Gma.9827.1.S1_at	RGA2_SOLBU	(Q7XBQ9) disease resistance protein RGA2 (RGA2-blb) (blight resistance protein RPI)	(GO:0005524) ATP binding	2.125	0.0018
Gma.3314.1.S1_a_at	CYTM_SOLTU	(P37842) multicystatin	(GO:0004869) cysteine protease inhibitor activity	−3.246	0.0026

in plant compounds were. Gene expression databases will, however, suffer from the weakness of the link between gene expression data and biological activity of each of the data point.

To obtain biological information from the gene expression data, many researchers translate a list of differentially expressed genes to relevant biological processes and pathways manually through literature and public databases searches (*36*). However, this is a tedious and time-consuming process. We have integrated nucleotide information for the soybean genes on the microarray with BLAST search results (SwissProt protein IDs), GO terms annotation, and KEGG pathways in one database, in order to interpret the differentially expressed genes based on functional annotations in terms of gene ontology molecular function category. We have developed Web interfaces to retrieve and display these data (**Figure 3**). The mapping of information with corresponding protein names and functional (GO) terms provides insight into functional differences between samples and enhances the prediction of unintended effects in transgenic soybean cultivars. Our database and pathway environment are versatile and can be applied to identify which genes and pathways have altered profiles in any tissue.

The insertion of a transgene in the soybean cultivars we examined did not generate large changes in gene expression relative to a group of conventional soybean cultivars. Our study included two transgenic cultivars, both derived from the same insertion event, and three conventional cultivars, one tissue, and one growth condition, and although in this study minimal differences were seen, every transgene insertion has the potential of unintended effects. The pairwise analysis in the comparison of transgenic soybean to the closest conventional counterparts produced a list of differentially expressed gene and revealed that, in both transgenic cultivars, genes involved in cysteine protease inhibitor activity and dihydroflavonol-4-reductase activity were downregulated. This may be an effect of the insertion event, an effect of the transgene product and, thus, a real unintended

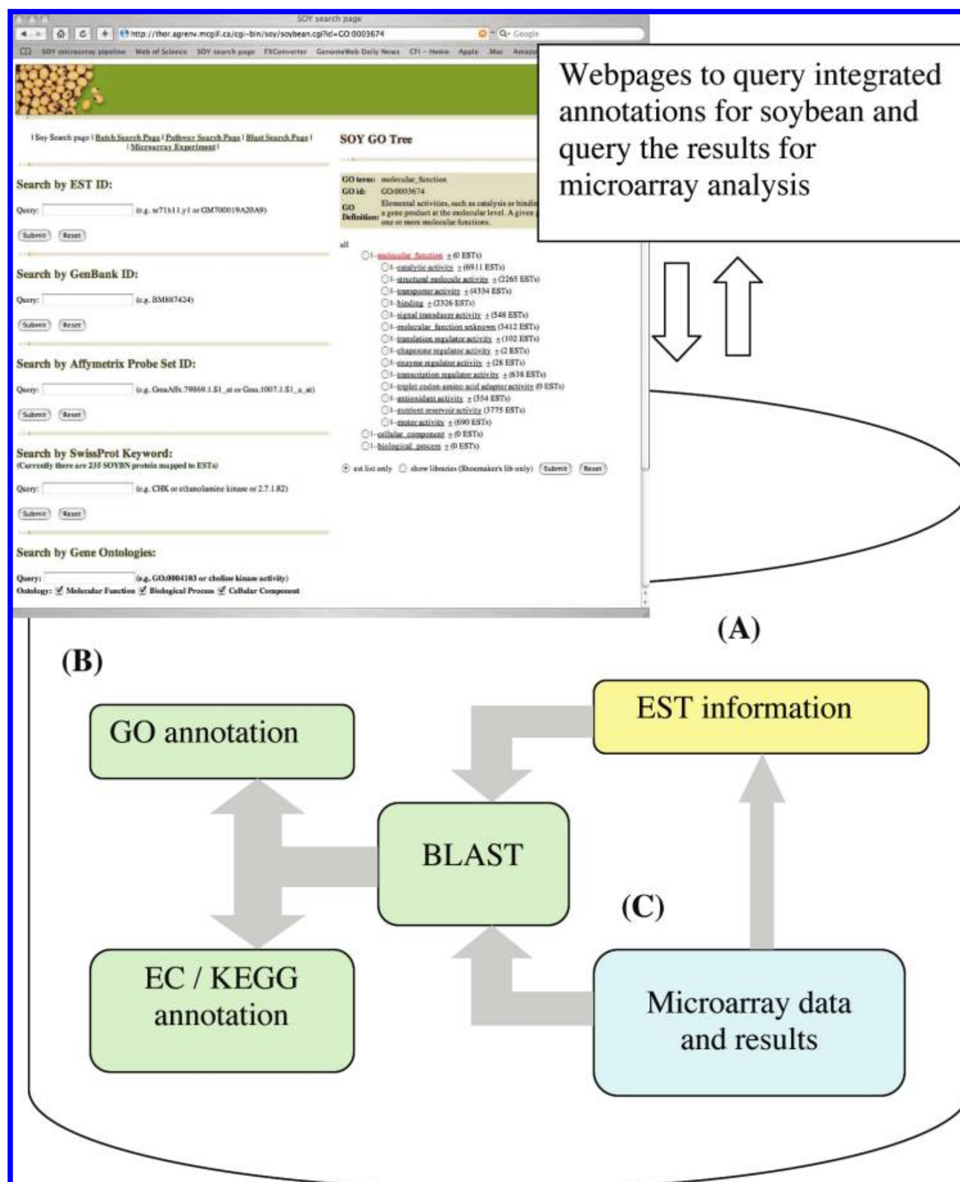

Figure 3. Overview of the soybean database structure. (**A**) Soybean EST and contig information obtained from public EST projects, GenBank, and TIGR were organized in tables modified from the ESTIMA database schema to include EST sequencing pipeline data. (**B**) Functional annotations such as GO terms, EC numbers. and KEGG molecular pathways link to the EST and microarray data through the protein names obtained from the BLASTX results. (**C**) Microarray data and results link to the annotations through BLASTX results and GenBank accession numbers of the EST sequences.

effect, or a natural variation of the parent genotype. Further investigations in the laboratory will be needed to assess effects like this. Future studies should include many more transgenes and transgene insertions, isogenic lines, and conventional cultivars, and tests should be performed on all plant tissues. The use of gene expression as an analytical tool for the assessment of differences between plants with novel traits and conventional lines will require tools for extracting biological information from the molecular functions and pathways influenced by transgene insertion. A list of genes with altered expression patterns does not constitute a safety assessment. The concept of substantial equivalence was developed as a practical approach to identify deviation from a norm, but the deviation is an indicator of risk and should not be used to define a hazard; it is rather a starting point that leads to a safety assessment (*37*).

ABBREVIATIONS USED

BLAST, Basic Local Alignment Search Tool; cDNA, complementary deoxyribonucleic acid; CEL, cell intensity file; dChip, DNA-Chip analyzer; DNA, deoxyribonucleic acid; EC, Enzyme Commission; EMBL, European Molecular Biology Laboratory; EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; EST, expressed sequence tag; ESTIMA, Expressed Sequence Tag Information Management and Annotation; GB, gigabyte; GO, gene ontology; GRAS, Generally Recognized As Safe); ID, identifier; KEGG, Kyoto Encyclopedia of Genes and Genomes; LIMMA, Linear Models for Microarray Data; MAS, Microarray Suite software; NCBI, National Center for Biotechnology Information; OS, operating system; PC, personal computer; PCA, principal component analysis; RAM, random access memory; RMA, robust multichip average; RNA, ribonucleic acid; RT-PCR,

reverse transcriptase Polymerase Chain Reaction; SQL, structured query language; T-DNA, transferred deoxyribonucleic acid; TIGR, The Institute of Genomic Research.

ACKNOWLEDGMENT

We thank Ernest Retzel, Lei Liu, and Lee Zamparo for bioinformatics assistance.

Accession Codes: The microarray data will be accessible at NCBI (Gene Expression Omnibus) GEO under accession numbers GSE9374, GSM238030, GSM238031, GSM238032, GSM238033, GSM238034, GSM238036, GSM238038, GSM238039, GSM238041, GSM238043, GSM238047, GSM238048, GSM238049, GSM238050, GSM238051, GSM238052, GSM238053, GSM238054, GSM238055, GSM238056, GSM238057, GSM238058, GSM238059, GSM238060, and GSM238061.

Supporting Information Available: Table of SSR markers. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Dunwell, J. M. Transgenic crops: the current and next generations. *Methods Mol. Biol.* **2005**, *286*, 377–398.
(2) Tabashnik, B. E. Seeking the root of insect resistance to transgenic plants. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3488–3490.
(3) Kuiper, H. A.; Kleter, G. A.; Noteborn, H. P. J. M.; Kok, E. J. Assessments of the food safety issues related to genetically modified foods. *Plant J.* **2001**, *27* (6), 503–528.
(4) Millstone, E.; Brunner, E.; Mayer, S. Beyond 'substantial equivalence'. *Nature* **1999**, *401*, 525–526.
(5) Kohli, A.; Twyman, R. M.; Abranches, R.; Wegel, E.; Stoger, E.; Christou, P. Transgene integration, organization and interaction in plants. *Plant Mol. Biol.* **2003**, *52*, 247.
(6) Gregersen, P. L.; Brinch-Pedersen, H.; Holm, P. B. A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic Res.* **2005**, *14*, 887–905.
(7) Ouakfaoui, S. E.; Miki, B. The stability of the *Arabidopsis* transcriptome in transgenic plants expressing the marker genes nptII and uidA. *Plant J.* **2005**, *41*, 791–800.
(8) Baudo, M. M.; Lyons, R.; Powers, S.; Pastori, G. M.; Edwards, K. J.; Holdsworth, M. J.; Shewry, P. R. Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnol. J.* **2006**, *4*, 369–380.
(9) Tanner, J. W.; Luzzi, B. M.; Gostovic, P.; Montminy, W.; Hume, D. J. OAC Bayfield soybean. *Can. J. Plant Sci.* **1998**, *78*, 625–626.
(10) CFIA List of varieties with novel traits and their progeny registered under the Canada Seeds Act and Regulations. Canadian Food Inspection Agency, Ottawa, Canada; 2005; http://www.inspection.gc.ca/english/plaveg/variet/pntvcne.shtml.
(11) Fehr, W. R.; Caviness, C. E.; Burmood, D. T.; Pennington, J. S. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* **1971**, *11*, 929–931.
(12) Padgette, S. R.; Kolacz, K. H.; Delannay, X.; Re, D. B.; LaVallee, B. J.; Tinius, C. N.; Rhodes, W. K.; Otero, Y. I.; Barry, G. F.; Eichholtz, D. A.; Peschke, V. M.; Nida, D. L.; Taylor, N. B.; Kishore, G. M. Development, identification, and characterization of a glyphosate-tolerant soybean line. *Crop Sci.* **1995**, *35*, 1451–1461.
(13) Swofford, D. L. PAUP*: phylogenetic analysis using pasimony (and other methods), version 4.0 beta test 10; Sinaur Associates, Sunderland, MA; 2002.
(14) Nei, M.; Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 5269–5273.
(15) Affymetrix. Data Sheet: GeneChip Soybean Genome Array. Documentation from Affymetrix website: 1-2, 2004−2005, accessed Aug 7, 2007; http://www.affymetrix.com/support/technical/datasheets/soybean_datasheet.pdf.
(16) Kumar, C. G.; LeDuc, R.; Gong, G.; Roinishivili, L.; Lewin, H. A.; Liu, L. ESTIMA, a tool for EST management in a multi-project environment. *BMC Bioinf.* **2005**, *5* (article 176), 1–10.
(17) Firth, D. CGIwithR: Facilities for processiong web forms using R. *J. Statistical Software* **2003**, *8* (10), 1–8.
(18) Hornik. The R FAQ. Documentation from R-Project website. (ISBN 3-900051-08-9), accessed Aug 7, 2007; http://www.r-project.org/.
(19) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788.
(20) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M. C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370.
(21) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
(22) The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29.
(23) Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **2006**, *34* (database issue), D354–D357.
(24) Quackenbush, J.; Lian, F.; Holt, I.; Pertea, G.; Upton, J. The TIGR Gene Indics: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res.* **2000**, *28*, 141–145.
(25) Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A. J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. Y.; Zhang, J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5* (10, article R80), 1–16.
(26) Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U.; Speed, T. P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **2003**, *4* (2), 249–264.
(27) Affymetrix. Statistical algorithms description document. Documentation from Affymetrix website. 1-28, accessed August 7, 2007; http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf.
(28) Li, C.; Wong, W. H. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 31–36.
(29) Smyth, G. K.; Michaud, J.; Scott, H. The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **2005**, *21*, 2067–2075.
(30) Padgette, S. R.; Taylor, N. B.; Nida, D. L.; Bailey, M. R.; MacDonald, J.; Holden, L. R.; Fuchs, R. L. The composition of glyphosate-tolerant soybean seeds is equivalent to that of conventional soybeans. *J. Nutr.* **1996**, *126*, 702–716.
(31) Staswick, P. E.; Zhang, Z.; Clemente, T. E.; Specht, J. E. Efficient down-regulation of the major vegetative storage protein genes in transgenic soybean does not compromise plant productivity. *Plant Physiol.* **2001**, *127*, 1819–1826.
(32) Kumudini, S.; Hume, D. J.; Chu, G. Genetic improvement in short season soybeans: I, dry matter accumulation, partitioning, and leaf area duration. *Crop Sci.* **2001**, *41*, 391–398.
(33) Kisha, T. J.; Diers, B. W.; Hoyt, J. M.; Sneller, C. H. Genetic diversity among soybean plant introductions and north American germplasm. *Crop Sci.* **1998**, *38*, 1669–1680.

(34) Sneller, C. H. Impact of transgenic genotypes and subdivision on diversity within elite North American soybean germplasm. *Crop Sci.* **2003**, *43*, 409–414.

(35) Kristensen, C.; Morant, M.; Olsen, C. E.; Ekstrom, C. T.; Galbraith, D. W.; Lindberg Meller, B.; Bak, S. Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (5), 1779–1784.

(36) Draghici, S.; Khatri, P.; Martins, R. P.; Ostermeier, C.; Krawetz, S. A. Global functional profiling of gene expression. *Genomics* **2003**, *81*, 98–104.

(37) FAO/WHO. Safety aspects of genetically modified foods of plant origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, accessed Aug 7, 2007; http://www.who.int/foodsafety/publications/biotech/en/ec_june2000_en.pdf.

Received for review November 30, 2007. Revised manuscript received February 24, 2008. Accepted February 28, 2008. This project was funded by the NCE AFMNet research network and by NSERC discovery grants to M.V.S. and M.G.F. We acknowledge support from the Centre Sève (FQRNT).

JF073505I