JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Transcriptome Analysis of Potato Tubers—Effects of Different Agricultural Practices

JEROEN P. VAN DIJK,*,†,‡ KATARINA CANKAR,†,§ STANLEY J. SCHEFFER,‡ HENRIEK G. BEENEN,‡ LOUISE V. T. SHEPHERD,# DEREK STEWART,# HOWARD V. DAVIES,# STEVE J. WILKOCKSON,⊥ CARLO LEIFERT,⊥ KRISTINA GRUDEN,§ AND ESTHER J. KOK‡

RIKILT Institute of Food Safety, Bornsesteeg 45, P.O. Box 230, 6700 AE Wageningen, The Netherlands; Department of Biotechnology and Systems Biology, National Institute of Biology, Vecna pot 111, 1000 Ljubljana, Slovenia, ; Plant Products and Food Quality Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland; and Nafferton Ecological Farming Group, Nafferton Farm, Newcastle University, Stocksfield, Northumberland NE43 7XD, United Kingdom

The use of profiling techniques such as transcriptomics, proteomics, and metabolomics has been proposed to improve the detection of side effects of plant breeding processes. This paper describes the construction of a food safety-oriented potato cDNA microarray (FSPM). Microarray analysis was performed on a well-defined set of tuber samples of two different potato varieties, grown under different, well-recorded environmental conditions. Data were analyzed to assess the potential of transcriptomics to detect differences in gene expression due to genetic differences or environmental conditions. The most pronounced differences were found between the varieties Sante and Lady Balfour, whereas differences due to growth conditions were less significant. Transcriptomics results were confirmed by quantitative PCR. Furthermore, the bandwidth of natural variation of gene expression was explored to facilitate biological and/or toxicological evaluation in future assessments.

KEYWORDS: ***Solanum tuberosum*; cDNA microarray; gene expression; food safety; organic input systems**

INTRODUCTION

The number of technological tools that are at plant breeders' disposal to introduce new desired traits into the plant genome is increasing. In the past this has only in rare cases led to food and feed safety issues in the newly bred plant varieties. Often-cited exceptions are the potatoes with high glycoalkaloid levels (*1*) and the celery variety with high psoralen levels (*2*). An assessment of the safety of new crop plant varieties is, however, not a routine part of approval procedures. With the current development of new plant varieties with more drastic changes in the plant's physiology, it can now be questioned whether the assessment of the safety of new crop varieties should not become a more standard item in the approval procedure of these new varieties (*3*).The food and feed safety assessment of plant products derived from genetically modified organisms has been the subject of debate in a number of international expert meetings (*4*–*6*). One of the recommendations was to further investigate the use of "omics" technologies as unbiased tools to assess any unintended changes in the plant's physiology (*4*, *6*). In subsequent years a number of national and European research projects have investigated the potential of transcriptomics, proteomics, and metabolomics for the detection of undesired side effects of plant breeding processes (*7*) (http://www.entransfood.nl/RTDprojects/GMOCARE/GMOCARE.html; http://www.food.gov.uk/science/research/researchinfo/foodcomponentsresearch/novelfoodsresearch/g02programme/g02projectlist/g02001/).

Recently, a number of papers describing transcriptomics of genetically modified (GM) plants have appeared. Baudo et al. (*8*) used transcriptomics to analyze GM and conventional wheat varieties. They primarily investigated the bandwidth of natural variation and found the natural variation in gene expression patterns in conventionally bred wheat varieties to be much larger than the variation between different GM lines. Another study on the bandwidth of natural variation as a basis for using transcriptomics as part of safety assessment protocols was performed in tomatoes (*9*). Batista et al. (*10*) compared an irradiated stable mutant rice line and a GM rice line with their respective parent lines, using transcriptomics. They found that transcriptome changes were more frequent in mutagenized plants, compared to GM plants. Similarly, Cheng et al. (*11*) found differences in gene expression with whole genome soy

* Corresponding author [telephone +31 (0) 317480398; fax +31 (0) 317417717; e-mail jeroen.vandijk@wur.nl].
† These authors contributed equally to this publication.
‡ RIKILT Institute of Food Safety.
§ National Institute of Biology.
# Scottish Crop Research Institute.
⊥ Newcastle University.

10.1021/jf802815d CCC: $40.75
Published on Web 01/27/2009

microarrays to be more frequent and more pronounced between conventional lines than between GM and conventional lines. However, they also found changes in *cysteine protease inhibitor* expression levels as a potential unintended effect in GM soybeans, although they also state that this could still fall within natural variation had more conventional soy lines been included in the study.

Insight into the natural variation in gene expression emerged as an important issue in these papers. With a holistic approach such as transcriptomics some changes are likely to be discovered in all cases between wild type (WT) and new plant varieties, whether or not GM. Such changes would trigger unnecessary follow-up analyses if no data were present on the natural variation of the specific genes in various conditions considered to be normal in agricultural practices. Natural variation of gene expression should be investigated in different locations, climates, years of harvest, and different farming practices. This would result in data on transciptome variation under current growing conditions and could also be used in time to assess potential health benefits or risks associated with new breeding systems. In this paper we describe transcriptomics of well-defined potato samples of two different potato varieties, grown under different, well-recorded environmental conditions. Potato was chosen as a model crop as it is the world's fourth most widely grown crop and a staple food in many countries. Data on its potential food safety risks are well recorded. Additionally, potatoes are known to produce natural toxins (glycoalkaloids) in certain conditions, which makes them interesting as a food safety model. The conditions here are represented by foliar late blight challenge and different agricultural practices. Other environmental factors were not investigated in this study. The data were analyzed to assess the potential of transcriptomics to detect differences in gene expression as a result of genetic differences or environmental conditions. Quantitative PCR was used for confirmation of the transcriptomics findings. The results are discussed in light of the food safety assessment of new plant varieties.

MATERIALS AND METHODS

Field Experimental Design. Potatoes were grown at Nafferton Experimental Farm in a field that was converted to organic farming practices in 2001 and cropped with a grass clover ley between 2001 and 2003. The field trial was established as a completely randomized block design with four replicate blocks, with all combinations of cultivar, fertilization, and crop protection being established on one plot in each block. The size of individual plots was 15 × 6 m. Organic seed tubers (supplied by Greenvale plc, Dundee, U.K.) were used for both varieties included in the trial, Sante (moderately resistant to foliar blight) and Lady Balfour (highly resistant to foliar blight), which are both relatively late-maturing main potato varieties. Potatoes were planted into soils to which either chicken manure pellets or dairy manure compost (prepared using a Sandberger windrow compost turner) was applied at a rate equivalent to 170 kg of N/ha. Fertilizers were applied immediately prior to plowing of soils in March 2004. The foliage of developing potato crops was then sprayed with either (a) copper oxychloride (product Headland Copper, two applications of 7.5 L/ha, equivalent to 3 kg of Cu/ha), (b) an extract of *Lychnis viscaria* L. (German catch fly) that contains brassinosteroids known to induce plant resistance (product COMCAT, two applications of 200 g/ha), or (c) water (at the same volume as used for the copper fungicide and alternative treatment sprays) as a control treatment. All plants were challenged by foliar blight (caused by exposure to *Phytophthora infestans*). Each of the 6 genotype/treatment combinations was performed in 4 biological replicates, leading to a total of 48 potato tuber samples. Nonchitted potato seed tubers were planted using a semiautomatic two-row potato planter (Checci & Magli, Budrio Bologna, Italy). Defoliation took place 2−3 weeks before harvest. Tops were removed by hand using sickles. The Blight MOP trial was harvested in September 2005 using a single-row potato harvester (Ransomes, Ipswitch, U.K.). In each plot, the middle row was harvested for tuber selection and sampling. All tubers from that row were bulked together and placed in a 25 kg bag. The bags were stored in a large shed at ambient temperature with no additional heating. After around 3 weeks the potatoes from each bag were graded and put back in the bag. Tubers were sampled after 7 weeks of storage. About 800 g (fresh weight) from four to five tubers was combined for each replicate. Two opposite eights (to minimize gradient effects within the tuber) were taken from each tuber within a replicate and bulked together. The resulting ca. 200 g of fresh weight was chopped, frozen in liquid nitrogen, freeze-dried, and ground in a laboratory mill fitted with a 1 mm screen. Freeze-dried and milled potato powders were stored at −80 °C until RNA extraction.

Construction of the cDNA Potato Microarray. Potato cDNA libraries were prepared by suppression subtractive hybridization using the Clontech PCR-Select cDNA Subtraction Kit (Clontech Laboratories Inc., Palo Alto, CA). Subtractions included tuber tissue versus nonedible plant parts (stem, leaf, stolon, immature tuber), light-exposed stressed tuber tissue versus nonstressed tuber, PVY-infected versus noninfected cDNA, and developing tubers versus mature tubers. Additionally, a selection of randomly sequenced clones from *Solanum phureja* was added. Clones were sequenced and sequences were consequently assembled into contigs using Gap4 software. Sequence analysis was performed by Greenomics, Plant Research International, Wageningen, The Netherlands. The resulting consensus sequences were blasted against "nr" and "EST-others" public sequence databases of the National Center for Biotechnology Information (NCBI) using blastx and blastn algorithms, respectively. Sequences have been submitted to the NCBI EST database with the following accession numbers: FG54827−FG551327.

Selection of clones for the microarray was made on the basis of unique sequence and sequence identity. To reduce the redundancy of the sequences selected for the microarray, different sequences from the same gene were included only when they represented distinct areas within the gene. Detailed data on the cDNA sequences included in the microarray are given in the Supporting Information. The automated annotation of chosen potato clones was performed as described previously (*12*).

Selected potato cDNA sequences were amplified by PCR using primers 20F1mod (aggcgattaagttgggtaac) and M13Rmod (agcggataacaatttcacac) that anneal to the pGEM T-easy vector multiple cloning site (Promega, Madison, WI). 5′-C6-amine linker was added to the primers to allow for binding to the aldehyde groups present on the silylated slides. Two microliters of glycerol stock of bacterial clones was used directly for a PCR reaction in a 100 μL reaction volume. End concentrations of PCR components were as follows: 1× PCR buffer (Invitrogen, Breda, The Netherlands); 1.5 mM $MgCl_2$ (Invitrogen); 200 μM dNTPs; 0.4 μM each primer and 5 units of Taq DNA polymerase (Invitrogen). Reactions were run on DNA Engine PCR cycler (Bio-Rad, Richmond, CA) with initial DNA denaturation for 2 min at 94 °C, then 40 cycles of 40 s at 94 °C, 60 s at 55 °C, and 150 s at 72 °C, with a final extension of 10 min at 72 °C. For quality control the PCR products were electrophoresed on a 1.5% agarose gel. PCR fragments were purified through Sephadex G-50 columns (Sigma-Aldrich, St. Louis, MO) and precipitated by overnight incubation with a $^1/_{10}$ volume of 3 M sodium acetate (pH 5.2) and 1 volume of isopropanol. The pellets were washed with 70% ethanol, air-dried, redissolved in 12 μL of 5× SSC, and used for spotting. Microarrays were spotted on silylated slides (CELAssociates) at 19 °C and 55% relative humidity using a Microgrid instrument (Apogent Technologies) in six blocks containing 26 × 26 spots. As positive controls three different luciferase fragments corresponding to the 5′, middle, and 3′ part of the cDNA, as well as the entire luciferase cDNA, were spotted seven times in each block of the array as well as a *Salmonella* gene fragment as a negative control (*13*). Microarrays were dried and stored at room temperature until use. Spotting quality was assessed for two of the spotted arrays in the series by staining the slides with 0.2 μM cyanine nucleic acid dye POPO-3 (Molecular Probes, Eugene, OR). Prior to hybridization free aldehyde groups were blocked with $NaBH_4$ (*14*).

Total RNA Isolation from Potato Tubers. Tubers were preprocessed prior to RNA extraction as described previously (*15*). RNA was isolated from 0.4 g of freeze-dried tuber material. The protocol for RNA extraction from polysaccharide-rich tissues based on using CTAB and consecutive chloroform/isoamyl alcohol extractions with an overnight LiCl precipitation (*16*) was adopted for potato tuber samples with the following modifications: the extraction buffer was warmed to 60 °C before use, the chloroform/isoamyl alcohol extraction was repeated three times before LiCl precipitation, and the final precipitation of RNA in 96% ethanol was performed by cooling the tubes on ice and centrifuging at 4 °C for 15 min at 14000*g*. The RNA was dissolved in 100 μL of 10 mM Tris (pH 7) by heating to 65 °C for 10 min.

RNA concentration and purity was assessed from the absorbance measurements by a Nanodrop 1000 instrument (NanoDrop Technologies). To assess RNA degradation 1 μg of RNA was electrophoresed on a 1% agarose gel with 5% formamide. Gel images were analyzed using Quantity One 1-D (Bio-Rad). Samples with the ratio of the quantity of 18S rRNA to total rRNA above 40% and the ratio of non-rRNA to rRNA below 30% were considered to be of adequate quality for microarray analysis. Adequate RNA quality was confirmed for all 48 RNA samples.

Fluorescent Labeling of cDNA and Hybridizations of the Microarray. RNA (25 μg) was labeled by incorporation of Cy5-dCTP during a cDNA synthesis reaction using 21-mer oligo-dT primers (*14, 17, 18*). One microliter of luciferase RNA was added to each sample. Labeled cDNA was dissolved in 40 μL of hybridization buffer containing 5× SSC, 0.2% SDS, 5× Denhardt's, 50% (v/v) formamide, and 0.2 mg/mL denatured herring sperm DNA.

A mixture of all PCR products spotted on the microarray was transcribed into cRNA in vitro using a Riboprobe SP6/T7 in vitro transcription kit (Promega) and was used as a universal control for all hybridizations. After in vitro transcription, the DNA was removed by addition of RQ-1 RNase-free DNase (Promega) and the cRNA was purified by two consecutive extractions with equal volumes of phenol/chloroform/isoamyl alcohol (25:24:1); the aqueous phase was precipitated with a $^1/_{10}$ volume of sodium acetate (pH 5.2) and 2 volumes of ethanol. RNA concentration was determined by spectrophotometric measurements using the Nanodrop 1000 instrument (NanoDrop Technologies). Universal control RNA was labeled with Cy3-dCTP during a cDNA synthesis reaction diluted 100 times to prevent competition with RNA samples at hybridization and dissolved in hybridization buffer.

Microarray slides were prehybridized in the hybridization buffer at 42 °C overnight. Prior to hybridization the slides were washed twice with Milli-Q water and once with isopropanol and quickly dried by centrifugation (1 min, 800*g*). The hybridization was performed in a hybridization frame (Geneframe, Implen, Munich, Germany) in 80 μL of hybridization buffer containing Cy5-labeled potato tuber RNA and Cy3-labeled universal control cDNA, in equal volumes. Slides were hybridized overnight at 42 °C in a hybridization chamber in darkness. After hybridization, the slides were washed twice with 1× SSC/0.1% SDS (5 min), twice with 0.1× SSC/0.1% SDS (5 min), and twice with 0.1× SSC (5 and 1 min) and dried by centrifugation (1 min, 800*g*). The slides were stored at room temperature in darkness until scanning.

Scanning and Image Analysis. Microarrays were scanned after excitation of Cy5 dye with a 633 nm laser and Cy3 dye with a 543 nm laser using scanner ScanArray Express HT (Perkin-Elmer). The microarrays were scanned at constant laser power (90%) and 10 μm resolution settings. Tiff images were imported into the ArrayVision software (Imaging Research, Waalwijk, The Netherlands) and the fluorescent intensity, background, and signal-to-noise ratio (S/N) were determined for each spot. The background signal was defined as the average signal in the four corners surrounding each spot. The S/N was defined as the spot signal minus the background signal, divided by the standard deviation of the background signal.

Microarray Data Analysis. Data were imported into statistical software R (*19*). Signal intensity, background, and S/N ratio were first examined on individual arrays to determine array quality. The $\log_2$ transformed ratio between Cy5 and Cy3 signals was calculated and used in further analysis. The arrays were divided into 12 groups according to potato cultivar, fertilization, and crop protection treatments, and 853 spots having at least one positive value for each of the treatments were selected for subsequent statistical analysis. Data were median normalized with the median being calculated for each array using a set of 568 spots for which at least 35 of 48 spots on all arrays showed a S/N ratio above 3.

Two independent data analyses were performed. Principal component analysis (PCA) was performed with Genemaths XT software (Applied Maths, St-Martens-Latem, Belgium) to determine whether the treatments influence the potato transcriptome and whether distinct differences can be seen among various treatment groups (*20*). A second analysis of variance (ANOVA) was performed with the R statistical software (*19*) employing a split plot design to determine significantly differentially expressed genes among treatment variables and their interaction. The biological replication was considered to be one of the factors in the ANOVA model to account for the variability within treatment groups. The false discovery rate was determined using Bonferonni, Holms, and Benjamini and Hochberg corrections (*21–23*). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (*24*) and are accessible through GEO Series accession number GSE11875 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11875).

Quantitative Reverse Transcriptase PCR (Q-RT-PCR). Q-PCR was used for the confirmation of the microarray results. Primers specific for different types of protease inhibitors and for housekeeping genes were designed in Beacon Designer software v. 5.10 (PREMIER Biosoft

Table 1. Primer Sequences for Q-RT-PCR

target gene	FSMP clone	primer name	nucleotide sequence (5′−3′)
PI Kunitz type A/B	FSPM2425, FSPM2948, FSPM0710, FSPM2456	PIAB-For	TGTTGTTGGAGACKGGAGGAACC
		PIAB-Rev	SCTKGAKGAAGACATCAAGAGGATTK
PI Kunitz type C	FSPM2334, FSPM4347, FSMP4349	PIC-For	CGTGTTCATTCGTAAGTCGGAGTC
		PIC-Rev	GTTCATTATTAACTTTCCAAACAGTTTCGTC
PI-I	FSMP2356, FSMP2856	PI-I-For	GTTTGCTCACATCATTGTTTTCTTTCTTC
		PI-I-Rev	GTTCTATGACTTCTGGTCCATCACTTTC
PI-II	FSPM1524, FSPM2919	PI-II-For	GCCTATTCAAAATGTCCCCGTTCAG
		PI-II-Rev	GGGCTCATCACTCTCTCCTTCAC
S. tuberosum eukaryotic elongation factor 5A-3	FSPM2812	EF5A3-For	GGAGGAGGTGGCTGAAGATTGG
		EF5A3-Rev	GGCTGATTGTGGTTCTGGTCTTATAC
potato ribosomal protein L10	FSPM2814	PRPL10-For	GTTACCGCCAGATTAAGAACAAGCC
		PRPL10-Rev	TCCCTTTCTCTTCATACCCACATCG
potato starch phosphorylase	FSPM2263	PSP-For	TGTCAACCGTACCGACTATCAGC
		PSP-Rev	CATCGCAGACATAACAGACACCAC

International). Four assays were designed to detect expression of genes from the *protease inhibitor (PI)* families *I* and *II* and *Kunitz type A, B,* and *C* (see **Table 1**). *PI* of *Kunitz type A* and *B* could not be differentiated. Therefore, a primer set employing ambiguous nucleotides was designed to detect the expression of both classes with a single assay. Three candidate housekeeping genes with high expression level and a small overall variation between samples were selected after the analysis of microarray data: *Solanum tuberosum eukaryotic elongation factor 5A-3*, the gene for *potato ribosomal protein L10*, and *potato starch phosphorylase*. Real-time PCR assays were designed for these genes (see **Table 1**) and compared with regard to overall variation in expression level and the Ct difference between Sante and Lady Balfour cultivars (data not shown). The *starch phosphorylase* gene was selected as the housekeeping gene for subsequent real-time PCR experiments. For PCR experiments, samples were chosen for the two cultivars in such a way that each cultivar was represented by five comparable samples.

Total RNA (10 μg) was treated with RQ-1 RNase-free DNase (Promega). One microgram of RNA was reverse transcribed using random hexamers and iScript cDNA Synthesis Kit (Bio-Rad), and 1 μL of cDNA (equivalent to 10 ng of reverse transcribed RNA) was added to a 25 μL real-time PCR reaction containing 2× iQ SYBR Green Supermix (Bio-Rad) and amplified on a MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad). The thermal cycling conditions were as follows: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C and 45 s at 55 °C followed by a melting curve analysis. Real-time PCR products for each assay were visualized on agarose gel, and only single bands of expected size were observed.

Each sample was assayed on two independent PCR plates, each containing the *starch phosphorylase* assay and two *PI* assays. A standard curve composed of 5 serial 10× dilutions of the most positive sample and a negative control were included on each plate. PCRs had previously been screened on DNase-treated RNA to amplify no trace DNA amounts. Transcript quantity was determined by interpolation of Ct values in a standard regression curve and normalized for the quantity of the starch phosphorylase. The data were log transformed, and the ratio between expression levels in Lady Balfour and Sante cultivars was calculated to enable direct comparison to microarray data. The difference between the Lady Balfour and Sante samples was assayed with a two-tailed *t* test on log-transformed data.

Natural Variation. For the factor cultivar, a variation factor was calculated on the basis of the SD of the $\log_2$-transformed expression values. The SD was calculated for the median values of the 12 individual treatments to reduce the influence of outliers. The variation factor was expressed as the factorial difference between the average$_{(\log\ values)}$ + SD and average$_{(\log\ values)}$ − SD, or 2× SD, as follows: variation factor = $2^{(2\times\ SD)}$. The upper and lower limit between which ∼95% of the data are expected to occur can be calculated directly from the SD (upper = $2^{(average\ +\ SD)}$; lower = 2(average − SD)), or from the NVF [upper = average(antilog) × ($\sqrt{}$NVF), and lower = average(antilog)/($\sqrt{}$NVF)].

RESULTS

Construction of the Food Safety-Oriented Potato Microarray (FSPM). For the safety assessment of novel plant varieties, including GM plants, relevant metabolic routes may be nutrient-related pathways as well as metabolic pathways that are part of the plants' basal physiology. Other metabolic routes of interest may be stress-related metabolic pathways as it is known that up-regulation of these pathways may lead to an increased production of antinutrients, including natural toxins (*25*). A dedicated potato microarray, the FSPM, was constructed comprising these three different categories of probes. Several cDNA libraries were constructed by cDNA subtraction (**Table 2**). To obtain specific genes for tubers, representing the edible plant part, cDNA subtraction was performed on tuber cDNAs minus a pool of cDNAs from other plant parts (stem, leaf, stolon, and immature tuber). The reverse subtraction was performed to obtain potato plant-specific sequences that may be unintentionally expressed in tubers as a result of applied breeding and growing schemes. Also, a cDNA subtraction was performed for tuber samples exposed to light to enrich for stress-related genes, including genes linked to the biosynthesis of glycoalkaloid potato natural toxins, such as *UDP-galactose:solanidine galactosyltransferase* (*sgt1*). Prior to the subtraction, *sgt1* expression was confirmed to be increased in the light-exposed potato tubers with real-time PCR. Expression of *sgt1* relative to *patatin* was 250 times higher in the light- versus the dark-stored tubers. Additionally, to include pathways related to biotic stress, cDNA sequences from a potato virus Y (PVY) infected versus noninfected potato leaf library were added. A number of protease inhibitor clones were also added to the array (*26*, *27*). Finally, cDNA sequences were included from a library enriched for genes involved in potato tuber development and sequences from the diploid potato variety *S. phureja*, also referred to as creole, or native, potato. The final FSPM microarray contained 4006 spots, including control spots for positive hybridization and negative hybridization. Positive controls were four different parts of the *luciferase* gene to hybridize with luciferase mRNA spiked in the labeling reaction. Negative control spots consisted of a *Salmonella* cDNA sequence (see Materials and Methods).

Table 2. Composition of the 4K Food Safety-Oriented Potato cDNA Microarray (FSPM)

library	no. of sequences
plant-specific cDNA sequences[a]	772
tuber-specific cDNA sequences[a]	301
stressed tuber-specific cDNA sequences[a]	1155
PVY infection-related cDNA sequences[a]	757
developing tuber-specific cDNA sequences[b]	404
Solanum phureja-derived cDNA sequences[b]	365
positive control	209
negative control	43
total	4006

[a] Combined, nonredundant set of cDNA sequences. [b] cDNA sequences may show limited overlap with the above libraries.

The chosen set of genes was annotated (*12*), and the genes were divided into 13 functional groups: photosynthesis (5.4% of clones), carbohydrate metabolism (4.6%), amino acid metabolism (2.8%), secondary metabolism (2.4%), signaling (4.4%), transport (4.2%), structure and development (6.5%), hormone related (3.0%), minor metabolism (5.8%), protein (16.1%), DNA/RNA/nucleotide metabolism (10.5%), stress (4.2%), and a group of genes for which the function was not successfully assigned (29.9%). Representation of functional groups differed for the potato cDNA libraries (**Figure 1**). As expected, the potato tuber library was highly enriched for genes related to carbohydrate metabolism, whereas photosynthesis genes prevailed among plant-specific cDNA sequences. Many stress-related genes were obtained in the developing tuber library. The presentation of functional groups was more equal in the *S. phureja*, light stress-related, and PVY stress-related libraries. Slight enrichment for genes involved in secondary metabolism and hormone-related genes was seen for the stress-related libraries, whereas transport-related genes were highly represented in the cDNA library connected to PVY infection.

Furthermore, for identified potato antinutrients and potential allergens (*28*) a limited set of genes is known, and these were all included in the FSPM. For antinutrients these consist of genes coding for protease inhibitors in general and genes involved in the biosynthesis of the glycoalkaloid potato natural toxins: *sgt1* (*UDP-galactose:solanidine galactosyltransferase*), *sgt2* (*UDP-glucose:solanidine glucosyltransferase*), *hmg1* (*3-hydroxy-3-*

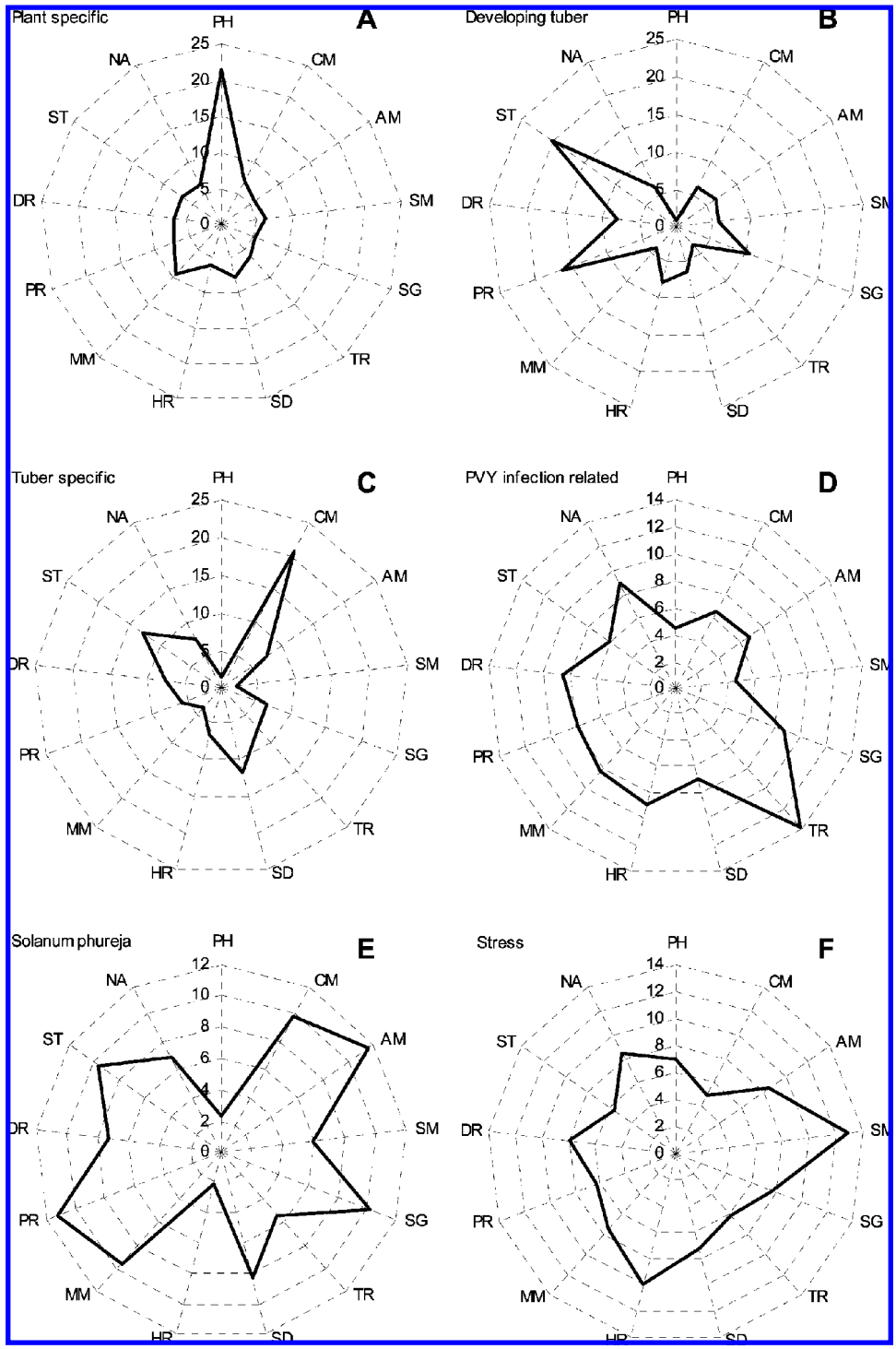

Figure 1. Functional classification for sequences from potato DNA. cDNA libraries are shown for (**A**) nonedible potato parts library, (**B**) developing tuber library, (**C**) tuber-specific library, (**D**) PVY infection-related library, (**E**) *Solanum phureja* sequences, and (**F**) light stress-related cDNA library. (PH, photosynthesis; CM, carbohydrate metabolism; AA, amino acid metabolism; SM, secondary metabolism; SG, signaling; TR, transport; SD, structure and development; HR, hormone related; MM, minor metabolism; PR, protein; DR, DNA/RNA/nucleotide metabolism; ST, stress; NA, functional group not assigned.) The numbers on the axes represent percentages of clones.

methylglutaryl coenzyme A reductase 1), and *pss1* (*squalene synthase 1*). For potential allergens these comprise genes coding for patatin and Kunitz type protease inhibitors A, B, and C (*29, 30*).

Microarray Analysis of Potato Tubers. We used a model study (**Figure 2**) to investigate the possibilities for microarray analysis to identify differences in gene expression within different agricultural settings. The combination of the different factors in the study led to 12 different treatment groups, all of which were replicated in quadruplicate in the field, yielding a total of 48 samples. After hybridization and scanning of the microarrays for all samples, a general data assessment was performed. After removal of the control spots, on average 959 of the 3754 spots (26%) showed a signal-to-noise ratio (S/N) larger than 3, ranging from 481 to 2139 spots, indicating the dormant state of potato tubers. When arrays were compared,

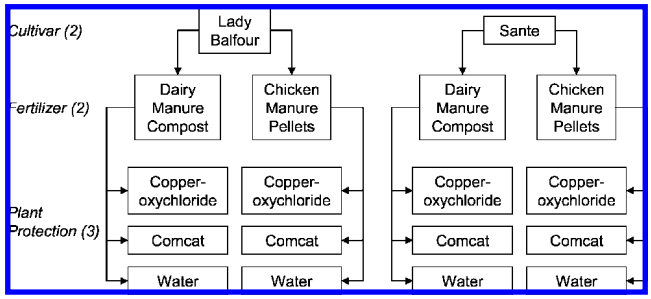

Figure 2. Overview of the BlightMOP study. All plants were challenged by foliar blight (caused by exposure to *Phytophthora infestans*). Two types of cultivar were used: Sante (moderately resistant to foliar blight) and Lady Balfour (highly resistant to foliar blight). Both cultivars were grown in two differently fertilized soils. All four treatments were exposed to two different antifungal regimens plus water as control, yielding a total of 12 different sample groups. The field trial was established as a completely randomized block design with four replicate blocks, with all combinations of cultivar, fertilization, and crop protection being established on one plot in each block.

499 spots showed a S/N < 3 for all arrays, whereas 74 spots showed a S/N > 3 for all arrays. We used an array-based cRNA pool as a universal reference to normalize spot quality differences between arrays. For this reference, 3528 of the 4006 spots showed a S/N > 3 on average (88%), ranging from 2323 to 3677. To reduce the noise in the data set, a selection of spots was made that showed consistent good signals throughout the experiments. We chose those spots that showed a S/N > 3 in at least one of the four arrays per treatment for both Cy3 and Cy5. This led to a data set of 853 spots used in further statistical analysis.

After normalization, principal component analysis (PCA) was used to visualize different sources of variation in the data set. The combination of the first and second components showed a complete separation of the two cultivars, together explaining 43% of the total variation in the data set (**Figure 3**). Further downstream (combinations of) components showed separation relating to neither type of fertilizer nor type of plant protection treatment.

A three-way, split plot analysis of variance (ANOVA) model was applied to identify the spots most significantly differentially expressed between the two cultivars, the different fertilizers, and the plant protection treatments. "Cultivar", "fertilizer", and "plant protection treatment" were included as factors, and the biological replication was included as error. Again, the most significant differential gene expression was observed for the factor cultivar, with 10 cDNA sequences having a p value of 2×10^{-16}, the smallest value that is given by the ANOVA function "aov" in the R software. The 15 most up- and the 15 most down-regulated cDNAs for the factor cultivar are listed in **Table 3**. The most significantly different genes identified by ANOVA corresponded well with the genes having the highest loadings for the components showing the separation between the cultivars in the PCA (**Figure 4**). A general overview of the number of genes that show statistically significant changes based on the ANOVA is given in **Table 4**. After cultivar, the factor fertilizer showed the most significant changes, the lowest p value being 1.4×10^{-3}. The factor antifungal treatment and the interactions between the different factors showed even less significant differences. A complete list of the ANOVA outcome can be found in the Supporting Information. Among the most differentially up-regulated spots for the Sante cultivar were clones representing *patatin*, *metallocarboxypeptidase inhibitor*, a *tomato elongation factor homologue*, and a *protease inhibitor*.

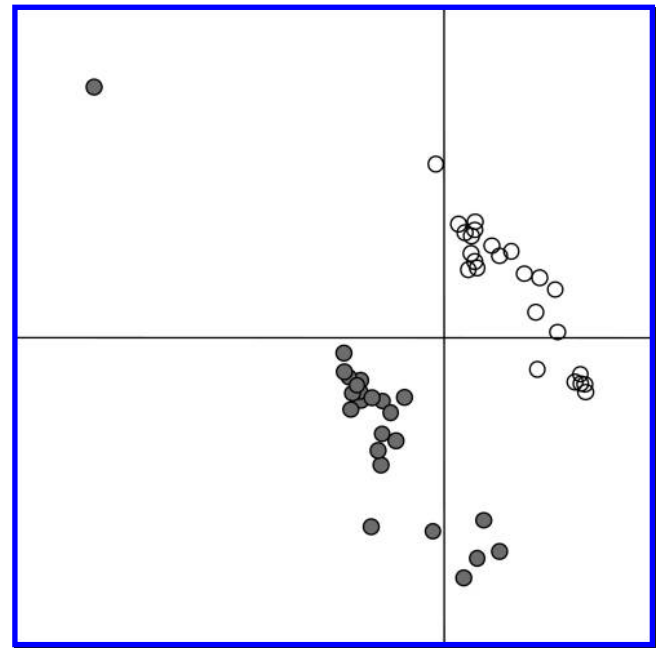

Figure 3. Complete separation of the Lady Balfour and Sante cultivars in principle component analysis (PCA). Representation of individual arrays in the combination of the first two components: gray circles represent arrays hybridized with Lady Balfour samples, open circles represent those with Sante samples. *X*-axis, first component, explaining 23% of the variation; *Y*-axis, second component, explaining 20% of the variation. PCA was performed on a selection of 853 spots on all 48 arrays.

The most up-regulated spots for the Lady Balfour cultivar included clones for another *protease inhibitor*, *heat shock proteins*, and *leucine amino peptidase*.

To estimate the percentage of false positives at a certain p value, different methods for estimating the false discovery rate (FDR), or type I error, were applied. We used the classic, rather strict, Bonferroni method (*22*), the Holms adaptation thereof (*23*), and the less conservative Benjamini and Hochberg (BH) method (*21*). Within the factor cultivar, when applying the Bonferroni method, an FDR of 1% corresponded to a cutoff value for the ANOVA p values of 1.2×10^{-5} and of 1.3×10^{-5} and 2.8×10^{-3} for the Holm and BH methods, respectively. This corresponded to 93, 127, and 242 differentially expressed spots with <1% being false positives according to the Bonferroni, Holm, and BH methods for FDR estimation, respectively, within the data set of 853 spots. For the factor fertilizer, the FDR was already 36% for the most significant spot (p value $= 1.4 \times 10^{-3}$) applying the BH method. This indicated that at least 36% of the spots were estimated to be false positives, even with p values slightly higher than 1.4×10^{-3}. Moreover, the Holm and Bonferroni methods yielded an FDR of 100% for all spots for the factor fertilizer, implying none of the spots being true positives, consistent with the lack of separation for this factor in the PCA. FDR analyses were not performed for factor plant protection treatment or for the interactions between factors, which were even less statistically significant.

Q-RT-PCR as Verification of Microarray Results. Quantitative reverse transcription Polymerase Chain Reaction (Q-RT-PCR) was used for verification of the microarray results. Analysis of the most prominent differences in the microarray data (**Table 3**) suggested a differential regulation of two different *protease inhibitors* between the two cultivars. Further analysis showed that all five different *protease inhibitors* present on the array showed differential regulation. For the cultivar Sante,

Table 3. Significantly Up-regulated Clones

(**A**) 15 Most Significantly Up-regulated Clones in Sante (S) Compared to Lady Balfour (LB)

clone ID	*p* ANOVA	S/LB[a]	closest homology	species
FSPM2639	2.0E−16	4.8	*ubiquitin-conjugating enzyme 18*	*Arabidopsis*
FSPM2941	2.0E−16	3.8	*60S ribosomal protein L21-like protein*	potato
FSPM2610	2.0E−16	3.7	*EF-1-alpha-related GTP-binding protein*	tobacco
FSPM2789	2.0E−16	3.3	*malate dehydrogenase*	potato
FSPM2536	2.0E−16	3.1	*patatin*[b]	potato
FSPM2412	2.0E−16	2.4	*metallocarboxypeptidase inhibitor*[c]	potato
FSPM0816	1.7E−15	2.9	*glucanase*	*Camellia sinensis*
FSPM2961	2.7E−15	2.0	*ADP-glucose pyrophosphorylase small subunit*	potato
FSPM3039	4.1E−15	2.1	*adenylate kinase*[d]	potato
FSPM2358	6.6E−15	1.8	*EST344548 potato stolon (est)*	potato
FSPM1594	7.7E−14	1.9	*cell wall protein*[e]	tomato
FSPM2115	2.9E−13	1.9	*ferredoxin-NADP reductase*	tobacco
FSPM2686	5.3E−13	1.7	*β tubulin*	*Arabidopsis*
FSPM2957	6.2E−11	1.5	*Drm3-like protein*[f]	potato
FSPM3283	2.0E−10	1.6	no sequence information	

(**B**) 15 Most Significantly Up-regulated Clones in Lady Balfour Compared to Sante

clone ID	*p* ANOVA	LB/S[a]	closest homology	species
FSPM2919	2.0E−16	3.00	*protease inhibitor II*[g]	potato
FSPM2771	2.0E−16	2.67	*cDNA clone 97821 5′ (est)*	potato
FSPM3019	2.0E−16	1.99	*mitochondrial small heat shock protein*	tomato
FSPM2732	2.0E−16	1.63	*mixed leaf Solanum tuberosum cDNA clone 64682 5′ (est)*	potato
FSPM2946	1.6E−15	1.72	*annexin p34*	potato
FSPM1618	6.7E−14	1.73	*H1 histone-like protein*	tomato
FSPM0120	1.1E−13	1.88	*DnaJ-like protein*[h]	potato
FSPM3165	1.5E−12	1.58	*human P23 tumor protein-like mRNA*[i]	potato
FSPM3305	2.5E−12	1.98	*cDNA clone STDB003K01 (est)*	potato
FSPM2221	2.8E−12	1.50	unknown sequence	
FSPM2945	3.0E−12	1.58	*cytosolic class I small heat shock protein 1B*	tobacco
FSPM0868	6.4E−12	2.18	*17.6 kD class I small heat shock protein*	tomato
FSPM0292	9.1E−12	1.46	*leucine aminopeptidase*	tomato
FSPM2922	5.8E−11	1.70	*similar to aquaporin*	tomato
FSPM2956	5.9E−11	1.55	*ribosomal protein L27a-like protein*	potato

[a] Ratio of mean expression values. [b] Similar *patatin* expression was observed for four clones, of which one represented the 3′ end of the gene, whereas three overlapping clones represented the 5′ end. [c] Similar expression for metallocarboxypeptidase inhibitor was observed for four clones, of which one represented the 5′ end of the gene, whereas three overlapping clones represented the 3′ end. [d] Similar expression for *adenylate cyclase* was observed for two almost identical clones representing the 3′ end of the gene. [e] Similar expression for *cell wall protein* was observed for two almost identical clones representing the entire cDNA. [f] Similar expression for *Drm3-like protein* was observed for three clones representing the 5′ end, the middle part, and the 3′ end of the gene, covering the entire CDS. [g] Similar expression for *protease inhibitor I* was observed for four almost identical clones representing nearly the entire coding sequence. [h] Similar expression for *DnaJ-like protein* was observed for three clones of which one showed homology to the second quarter of the gene, and two overlapping clones represented the 3′ end. [i] Similar expression for *human P23 tumor protein-like mRNA* was observed for two almost identical clones representing the entire cDNA.

higher expression was observed for *protease inhibitors type I* and *Kunitz type A* and *B*. For the cultivar Lady Balfour, higher expression was observed for *protease inhibitors type II* and *Kunitz type C*. The differential expression observed showed differences both in magnitude and in significance, ranging from a factor 3.0 with a p value of 2×10^{-16} to a factor 0.74 with a p value of 1.2×10^{-6}. As the latter p value was close to the cutoff value for 1% FDR using the Bonferroni and BH methods, this series of genes was selected for gene expression verification in the range from small to large differences. Q-RT-PCRs were developed for all *protease inhibitors* on the array. For the *Kunitz type A* and *B* a PCR was developed that recognized both targets, as the sequences showed too much homology to develop a discriminating PCR (**Table 1**). For normalization, PCRs were developed for three potential reference genes, based on microarray results: *S. tuberosum eukaryotic elongation factor 5A-3*, *potato ribosomal protein L10*, and *potato starch phosphorylase*. *Potato starch phosphorylase* was selected as reference gene as it showed the least variability in gene expression levels between cultivars and among the samples. Of the four different *protease inhibitors*, *type II* and *Kunitz type C* were clearly confirmed to have differential expression between the cultivars by two-tailed *t* test on Q-RT-PCR results. Expression differences for the *type I protease inhibitor* were on the verge of significant confirmation ($p = 0.048$), and the difference found showed a ratio of expression between the two cultivars very similar to the ratio based on the microarray data. The *Kunitz type A/B protease inhibitors* could not be confirmed to be differentially expressed. Although the magnitude of the difference observed with Q-RT-PCR in this case was very similar to the microarray results, the p value of the t test for Q-RT-PCR results was only 0.48 (**Figure 5**).

Natural Variation of Gene Expression. To gain insight into the natural variation in our data set, we expressed a natural variation factor (NVF) as $2^{(2\times\ SD)}$, as the data were $\log_2$ transformed, for all genes with significant levels of expression. For most genes this NVF was very small. Eighty-nine percent of the genes (755) had a NVF of <1.5, 68 genes showed a NVF between 1.5 and 2.0, and 30 genes showed a NVF of >2.0, the highest being 5.7 (annex 1, **Figure 6**). This was mainly due to the variation in expression between the two cultivars in our study, as this was the main source of variation. In fact, of the 30 spots with an NVF > 2.0, 26 were in the top 15 of either up- or down-regulated genes between the two cultivars (**Table 3**). Of the known food safety-related genes in potato, two gene families showed a large variation in the two cultivars. For

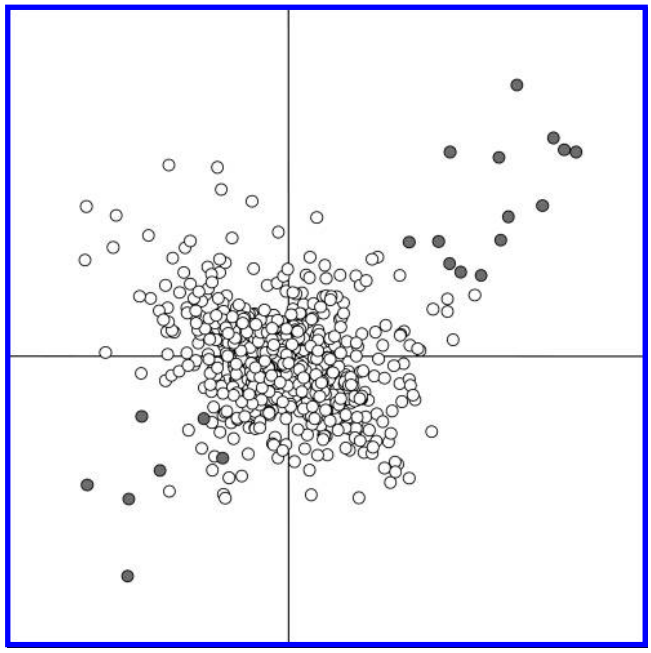

Figure 4. Concordance between ANOVA and PCA results. Shown is a loadings plot for the individual spots for the combination of the first two principal components of the PCA. Spots in the upper right panel correspond with genes having a higher expression in the Sante samples than in the Lady Balfour samples, and vice versa for the lower left panel. Indicated in gray are the spots that showed a p value of $<5.0 \times 10^{-15}$ in the ANOVA for the factor cultivar, showing a good concordance between the two methods of data analysis. *X*-axis, first component, explaining 23% of the variation; *Y*-axis, second component, explaining 20% of the variation. PCA and ANOVA were performed on a selection of 853 spots on all 48 arrays.

Table 4. Summary of ANOVA Analysis[a]

p value	C	F	T	C × F	C × T	F × T	C × F × T
≥ 0.01	550	842	850	850	848	849	853
$<0.01 \geq 1 \times 10^{-5}$	180	11	3	3	5	4	
$<1 \times 10^{-5} \geq 1 \times 10^{-8}$	63						
$<1 \times 10^{-8} \geq 1 \times 10^{-11}$	23						
$<1 \times 10^{-11} \geq 1 \times 10^{-14}$	15						
$<1 \times 10^{-14} \geq 2 \times 10^{-16}$	22						

[a] C, cultivar; F, fertilizer; T, antifungal treatment; ×, interaction.

patatin, identified as coding for a potential allergen, the average $\log_2$ expression was 0.44 with a NVF of 5.73, based on the *patatin* clone FSPM2981, showing in fact the largest NVF in the entire data set. This number indicates that the average (detransformed) expression value for *patatin* was 1.35 and that the majority of expression values would be expected between 0.57 and 3.23, as $1.35/(\sqrt{5.73})$ equals 0.57 and $1.35 \times (\sqrt{5.73})$ equals 3.23. Again, the main cause for this variation was the variation in patatin expression between the two cultivars, with the means differing by a factor 4.6. For the various *protease inhibitors* (also coding for potential allergens and antinutrients), especially *type II* showed a large NVF (3.24), whereas other *protease inhibitors* showed moderate NVFs (*PI-I*, 1.69; *PI Kunitz type A*, 1.51; *PI Kunitz type B*, 1.54; *PI Kunitz type C*, 1.43). Other genes with relation to the food safety of potatoes concern biosynthesis of glycoalkaloids (*sgt1*, *sgt2*, *hmg1*, and *pss1*). For all of these, the level of expression was below the detection limit of our microarray system in the samples investigated. This meant that a variation ratio could not be calculated for these genes.

DISCUSSION

Current safety assessment procedures for new plant varieties with known significant compositional alterations and for new GM plant varieties require the compositional analysis of the new plants or relevant plant parts. These compositional studies form an important starting point of the comparative safety assessment (CSA) (*3*, *25*, *31*) of new plant products. At present, transcriptomics may provide the best coverage over the plant's metabolic networks, as for some species "whole genome" arrays are already available. In the future, analysis of changes in the proteome and metabolome may provide more direct information on changes in the plant's physiology.

We constructed a food safety-oriented potato cDNA microarray (FSPM), through various cDNA subtractive libraries, to include basal plant and tuber metabolic pathways as well as light-induced or biotic stress-related pathways. Rather than the construction of an array with only known food safety-related genes, we chose to include all genes from subtractive cDNA libraries enriched for genes to be expressed in stress situations. Although not all sequences on the array could be related to known genes, this approach resulted in the inclusion of genes known so far to be involved in the food safety of potato tubers. These genes were identified by comparing literature on gene expression with the OECD consensus document on key nutrients, antinutrients, and toxicants in potato tubers (*28*). We focused on the pathways for antinutrients and allergens as these may be the most relevant for food safety. The metabolic routes, and underlying genes, for the glycoalkaloid potato natural toxins present in cultivated potato have not yet been entirely elucidated (*32*). However, both *steroidal alkaloid glycosyltransferase* genes *sgt1* and *sgt2* that are involved in the biosynthesis of α-solanine and β-chaconine, respectively (*33*, *34*), are present on the FSPM, as well as the genes *hmg1* and *pss1*, the expressions of which are correlated with high levels of steroidal glycoalkaloids (*32*). Furthermore, genes encoding for the antinutrients (protease inhibitors in general) and allergens (patatin, Kunitz type protease inhibitors), which were identified as such (*28*), were also included in the FSPM. The potato was until recently not considered to be a source of allergens, but patatin, the main storage protein in potatoes, was reported to induce allergic reactions in sensitive children and to bind IgE, which may indicate allergenic potential (*29*). Likewise, potential allergenicity was established for the cysteine and aspartic protease inhibitors, including the cathepsin D-protease inhibitors, belonging to the family of soybean trypsin inhibitors (Kunitz type) (*30*). On the basis of DNA homology, the potato Kunitz type protease inhibitors are grouped in three major families (A, B, and C), with members showing distinct but overlapping profiles with regard to type of protease inhibiting potential (*35*–*38*). Members of all three groups of Kunitz type protease inhibitors were included on the FSPM array, including the identified allergens. Members of families of wound-inducible protease inhibitors types I and II (*39*) were also included on the array.

In the present study two cultivars, Sante and Lady Balfour, were exposed to foliar blight infection and were grown using different types of organic fertilizer and under different types of plant protection regimens. Major differences in gene expression were observed between the two potato cultivars included in the study. When analyzed with PCA, 43% of the variation within the data set could be attributed to cultivar differences. No separation was seen for the other factors in the study on the basis of PCA. The large differences in gene

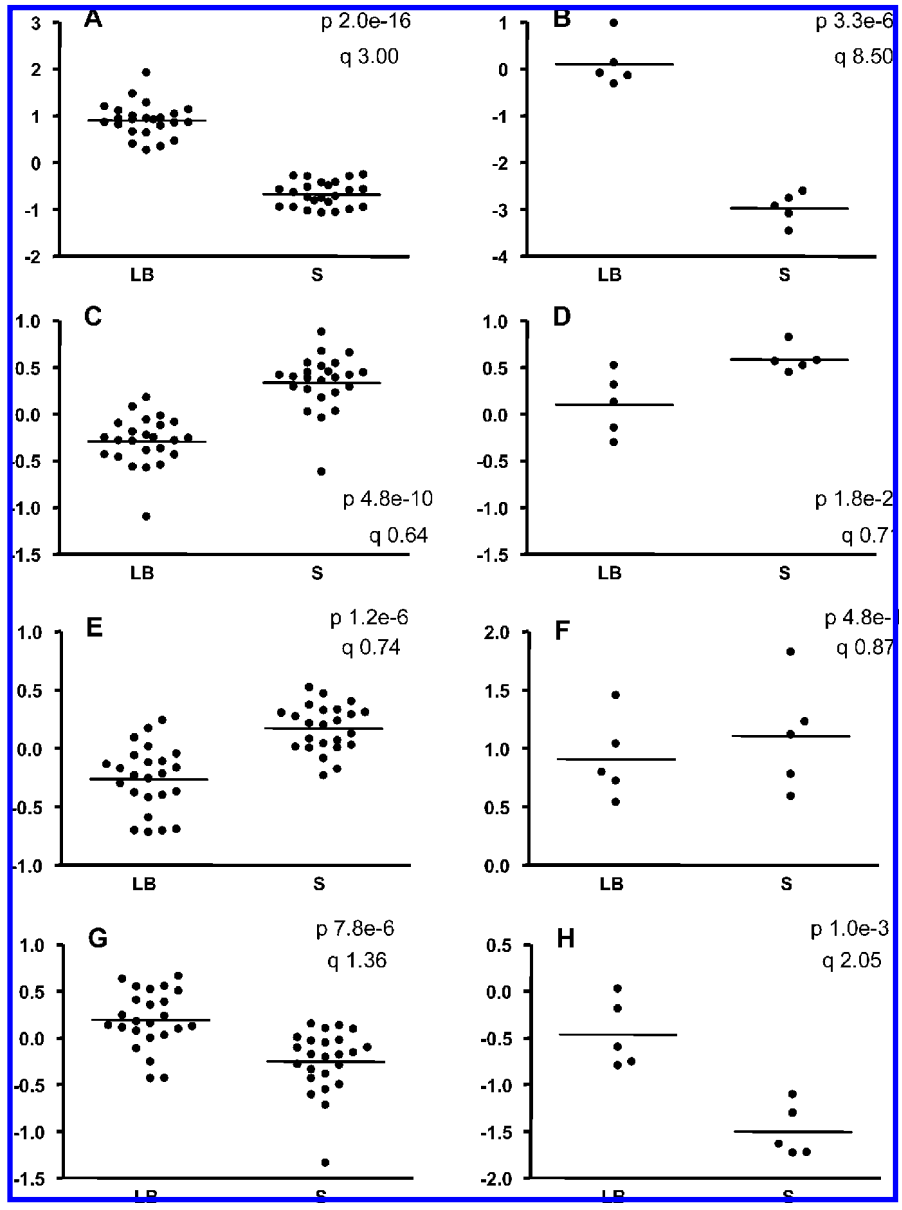

Figure 5. Microarray verification by Q-RT-PCR. Left-hand panels show microarray results; right-hand panels show Q-RT-PCR results. In all cases the significance (*p*) and the magnitude (*q*) of the difference in expression between cultivars is given (LB, Lady Balfour; S, Sante). Panels **A** and **B** show confirmation of microarray results with Q-RT-PCR results for protease inhibitor (PI) type II. Panels **C** and **D** show a trend toward confirmation of microarray results for PI type I. Panels **E** and **F** show no confirmation of microarrays results for PI Kunitz type A/B. Panels **G** and **H** show confirmation of microarray results for PI Kunitz type C. All data are normalized $\log_2$ transformed expression values. For each gene the *y*-axis spans the same magnitude for microarray and Q-RT-PCR data, although this is different for different genes.

expression between cultivars were confirmed by ANOVA, also after a subsequent FDR correction, and by quantitative PCR of selected cDNA sequences. Smaller differences for applied fertilizer and plant protection regimen were only observed with ANOVA analysis, but were annihilated when FDR correction was applied. The observation that the combination of ANOVA with FDR correlated more closely with PCA than ANOVA alone confirmed the necessity of applying FDR to correct for the type I error in multiple ANOVAs.

In the present study, we did not investigate corresponding protein levels. However, protein expression is at least partially transcriptionally regulated for two of the major differentially expressed genes, *patatin* and *protease inhibitors*. For *patatin*, accumulation of both mRNA and protein occurs during the onset of tuber formation, and an antisense *patatin1* gene was shown to inhibit both mRNA and protein expression of patatin (*40*). For *protease inhibitors*, a clear correlation between induction of both gene and protein expression is observed after wounding (reviewed in ref *41*), although in some specific cases also posttranslational regulation has been described (*42*).

As the cultivars and agricultural practices in this study are already commercially available for consumers, the observed gene expression variation can be considered as part of the natural variation in commercial potato varieties with a history of safe use. These data can form part of a data set for future reference when assessing the safety of new potato varieties. Noteworthy is that a large variation was observed for two gene families identified as potential allergens, protease inhibitors, and patatin, the first being already known for their antinutritional characteristics as well (*28*). This information is of relevance for future comparisons between new potato plant varieties and comparator varieties already on the market. On the other hand, genes involved in the glycoalkaloid biosynthesis pathway showed

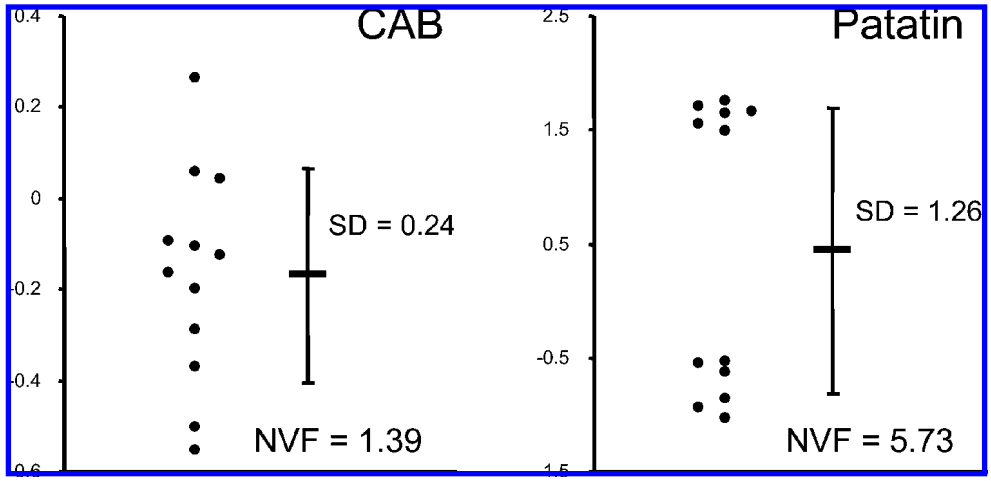

Figure 6. Natural variation factor for a normal and bimodal data distribution. Shown are the $\log_2$ transformed data (dots) for a normal (left) and bimodal (right) distribution of data. The bars indicate the average ($\log_2$) expression and the standard deviation, from which the NVF can be calculated (NVF = $2^{(2\times\ SD)}$). CAB, chlorophyll *a*/*b* binding protein, FSPM1778; patatin, FSPM2981.

expression levels beyond the sensitivity of our system, implying that significant expression of these genes in new potato varieties, if observed, may be a cause for further investigation.

A first attempt to assess the biological relevance of observed differences in future assessments can be performed by establishing a so-called natural variation factor (NVF), which provides an indication of the observed expression range for genes of interest. With the NVF, the SD of log-transformed data is translated to a factor of variation on a linear scale. Although data analysis of microarray data usually makes use of log transformation to obtain a normal distribution of the entire data set, there is no reason to assume that individual gene expression would not fit a normal distribution on a linear scale in different varieties and under different agricultural conditions. For a normal distribution the NVF and the upper and lower limits of expression approximate the 95% confidence interval. However, the NVF is not limited to normally distributed data; it can also be determined for genes in this study, such as patatin, where clearly two groups of data can be observed for the two cultivars. In these cases the NVF provides a better insight in the variation than the 95% confidence interval because the actual standard deviation is used instead of a modeled one. In these cases the 95% confidence interval would indicate a narrower margin of expected gene expression than is actually the case. The data from the present study may form a starting point for the compilation of a larger set of data on natural variation, enabling the comparison of new varieties with a window of observed gene expression values in varieties on the market with a history of safe use. An obstacle in transcriptomics analysis could be that the nature of the probes on different platforms, such as cDNA probes at different locations in the gene sequence, or oligonucleotides of different lengths, or other procedural differences, might result in a different relative expression of the same gene. However, also in these cases the variation factors, being independent of the actual gene expression, could still be used as a basis for comparison. Nevertheless, the future use of uniform microarray platforms and SOPs would greatly enhance the comparability of data and the usefulness of such a database of natural variation of gene expression in potato tubers as part of the safety assessment of new potato plant varieties. Additionally, careful recording of storage time before sampling is necessary to allow an equal comparison between tubers. This is of importance because, even as the tubers are dormant, marked metabolic and transcriptomic changes occur during this phase, prior to sprouting (*43*). For instance, expression of two members of the potato *NCED (9-cis-epoxycarotenoid dioxygenase)* gene family was found to correlate with changes in abscisic acid content in meristems (*StNCED2*) and cortex (*StNCED1*) during tuber storage (*44*). Likewise, factors such as time between defoliation and harvest and weather/climate conditions need to be recorded as metadata for each data set to come to a useful database for natural variation in potato tubers and a valid comparison of new varieties to the existing set of potato cultivars.

In conclusion, a new microarray (FSPM) was constructed and used to assay a set of potato tubers of two varieties grown under different field practices. The results of the gene expression analysis were confirmed by Q-RT-PCR and discussed in light of food safety assessment. The present study underscores the value of the application of transcriptomics analysis to assess the safety of new potato varieties. It was shown that observed differences can be confirmed by real-time PCR. Furthermore, the data set in this study provides a basis to further assess the bandwidth of natural variation of gene expression that should in safety studies be taken into account to assess the biological and/or toxicological relevance of observed differences between the new potato plant variety and a suitable comparator already on the market.

ACKNOWLEDGMENT

We thank Polona Kogovsek and Dr. Maruša Pompe-Novak for the construction of PVY subtraction libraries and Prof. Dr. Borut Štrukelj (Faculty of Pharmacy, University of Ljubljana, Slovenia) for contributing the proteinase inhibitor clones for the construction of the array.

Supporting Information Available: A spreadsheet (Appendix 1) is available with a number of characteristics for the spots on the array that were selected for analysis including average gene expressions, standard deviations, natural variation factors, and *p* values for the two-way ANOVA. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Van Gelder, W. M. J.; Vinke, J. H.; Scheffer, J. J. C. Steroidal glycoalkaloids in tubers and leaves of *Solanum* species used in potato breeding. *Euphytica* **1988**, 147–158 (Suppl.).

(2) Seligman, P. J.; Mathias, C. G.; O'Malley, M. A.; Beier, R. C.; Fehrs, L. J.; Serrill, W. S.; Halperin, W. E. Phytophotodermatitis

from celery among grocery store workers. *Arch. Dermatol.* **1987**, *123* (11), 1478–1482.
(3) Kok, E. J.; Keijer, J.; Kleter, G. A.; Kuiper, H. A. Comparative safety assessment of plant-derived foods. *Regul. Toxicol. Pharmacol.* **2008**, *50* (1), 98–113.
(4) *FAO/WHO Biotechnology and Food Safety. Report of a Joint FAO/WHO Consultation*; Food and Agriculture Organisation of the United Nations: Rome, Italy, 1996.
(5) Konig, A.; Cockburn, A.; Crevel, R. W.; Debruyne, E.; Grafstroem, R.; Hammerling, U.; Kimber, I.; Knudsen, I.; Kuiper, H. A.; Peijnenburg, A. A.; Penninks, A. H.; Poulsen, M.; Schauzu, M.; Wal, J. M. Assessment of the safety of foods derived from genetically modified (GM) crops. *Food Chem. Toxicol.* **2004**, *42* (7), 1047–1088.
(6) *OECD Food Safety Evaluation*; Organisation for Economic Cooperation and Development: Paris, France, 1996.
(7) Cellini, F.; Chesson, A.; Colquhoun, I.; Constable, A.; Davies, H. V.; Engel, K. H.; Gatehouse, A. M.; Karenlampi, S.; Kok, E. J.; Leguay, J. J.; Lehesranta, S.; Noteborn, H. P.; Pedersen, J.; Smith, M. Unintended effects and their detection in genetically modified crops. *Food Chem. Toxicol.* **2004**, *42* (7), 1089–1125.
(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* (4), 369–380.
(9) Kok, E. J.; Lehesranta, S. J.; van Dijk, J. P.; Helsdingen, J. R.; Dijksma, W. T. P.; Van Hoef, A. M. A.; Koistinen, K. M.; Kärenlampi, S. O.; Kuiper, H. A.; Keijer, J. Changes in gene and protein expression during tomato ripening: consequences for the safety assessment of new crop plant varieties. *Food Sci. Technol. Int.* **2009**, (in press).
(10) Batista, R.; Saibo, N.; Lourenco, T.; Oliveira, M. M. Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (9), 3640–3645.
(11) Cheng, K. C.; Beaulieu, J.; Iquira, E.; Belzile, F. J.; Fortin, M. G.; Strömvik, M. V. Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. *J. Agric. Food Chem.* **2008**, *56*, 3057–3067.
(12) Rotter, A.; Usadel, B.; Baebler, S.; Stitt, M.; Gruden, K. Adaptation of the MapMan ontology to biotic stress responses: application in solanaceous species. *Plant Methods* **2007**, *3*, 10.
(13) Kok, E. J.; Franssen-van Hal, N. L.; Winnubst, L. N.; Kramer, E. H.; Dijksma, W. T.; Kuiper, H. A.; Keijer, J. Assessment of representational difference analysis (RDA) to construct informative cDNA microarrays for gene expression analysis of species with limited transcriptome information, using red and green tomatoes as a model. *J. Plant Physiol.* **2007**, *164* (3), 337–349.
(14) Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**, *270* (5235), 467–470.
(15) Lehesranta, S. J.; Koistinen, K. M.; Massat, N.; Davies, H. V.; Shepherd, L. V.; McNicol, J. W.; Cakmak, I.; Cooper, J.; Luck, L.; Karenlampi, S. O.; Leifert, C. Effects of agricultural production systems and their components on protein profiles of potato tubers. *Proteomics* **2007**, *7* (4), 597–604.
(16) Chang, S.; Puryear, J.; Cairney, J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **1993**, *11*, 113–116.
(17) Boeuf, S.; Klingenspor, M.; Van Hal, N. L.; Schneider, T.; Keijer, J.; Klaus, S. Differential gene expression in white and brown preadipocytes. *Physiol. Genomics* **2001**, *7* (1), 15–25.
(18) Franssen-van Hal, N. L.; Vorst, O.; Kramer, E.; Hall, R. D.; Keijer, J. Factors influencing cDNA microarray hybridization on silylated glass slides. *Anal. Biochem.* **2002**, *308* (1), 5–17.
(19) *R-Development-Core-Team R: A Language and Environment for Statistical Computing (Manual)*; R Foundation for Statistical Computing: Vienna, Austria, 2005; ISBN 3-900051-07-0.
(20) Yeung, K. Y.; Ruzzo, W. L. Principal component analysis for clustering gene expression data. *Bioinformatics* **2001**, *17* (9), 763–774.
(21) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B—Method.* **1995**, *57* (1), 289–300.
(22) Bonferroni, C. E. Teoria statistica delle classi e calcolo delle probabilità. *Pubbl. R. Ist. Sup. Sci. Econ. Commer. Firenze* **1936**, *8*, 3–62.
(23) Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6* (2), 65–70.
(24) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30* (1), 207–210.
(25) Kok, E. J.; Kleter, G. A.; van Dijk, J. P. *Use of the cDNA Microarray Technology in the Safety Assessment of GM Food Plants*; 558 TemaNord 2003; Nordic Council of Ministers: Copenhagen, Denmark, 2003; pp 5—84.
(26) Gruden, K.; Strukelj, B.; Ravnikar, M.; PoljsakPrijatelj, M.; Mavric, I.; Brzin, J.; Pungercar, J.; Kregar, I. Potato cysteine proteinase inhibitor gene family: molecular cloning, characterisation and immunocytochemical localisation studies. *Plant Mol. Biol.* **1997**, *34* (2), 317–323.
(27) Strukelj, B.; Ravnikar, M.; Mesko, P.; Poljsak-Prijatelj, M.; Pungercar, J.; Kopitar, G.; Kregar, I.; Turk, V. Molecular cloning and immunocytochemical localization of jasmonic acid inducible cathepsin D inhibitors from potato. *Adv. Exp. Med. Biol.* **1995**, *362*, 293–298.
(28) *OECD Consensus Document on Compositional Considerations for New Varieties of Potatoes: Key Food and Feed Nutrients, Anti-Nutrients and Toxicants*; ENV/JM/MONO(2002)5; Organisation for Economic Cooperation and Development: Paris, France, 2002.
(29) Seppala, U.; Alenius, H.; Turjanmaa, K.; Reunala, T.; Palosuo, T.; Kalkkinen, N. Identification of patatin as a novel allergen for children with positive skin prick test responses to raw potato. *J. Allergy Clin. Immunol.* **1999**, *103* (1 Pt 1), 165–171.
(30) Seppala, U.; Majamaa, H.; Turjanmaa, K.; Helin, J.; Reunala, T.; Kalkkinen, N.; Palosuo, T. Identification of four novel potato (*Solanum tuberosum*) allergens belonging to the family of soybean trypsin inhibitors. *Allergy* **2001**, *56* (7), 619–626.
(31) Kuiper, H. A.; Kleter, G. A.; Noteborn, H. P.; Kok, E. J. Assessment of the food safety issues related to genetically modified foods. *Plant J.* **2001**, *27* (6), 503–528.
(32) Krits, P.; Fogelman, E.; Ginzberg, I., Potato steroidal glycoalkaloid levels and the expression of key isoprenoid metabolic genes. *Planta* **2007**, *227* (1), 143—150.
(33) McCue, K. F.; Allen, P. V.; Shepherd, L. V.; Blake, A.; Whitworth, J.; Maccree, M. M.; Rockhold, D. R.; Stewart, D.; Davies, H. V.; Belknap, W. R. The primary in vivo steroidal alkaloid glucosyltransferase from potato. *Phytochemistry* **2006**, *67* (15), 1590–1597.
(34) Moehs, C. P.; Allen, P. V.; Friedman, M.; Belknap, W. R. Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. *Plant J.* **1997**, *11* (2), 227–236.
(35) Heibges, A.; Glaczinski, H.; Ballvora, A.; Salamini, F.; Gebhardt, C. Structural diversity and organization of three gene families for Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Mol. Genet. Genomics* **2003**, *269* (4), 526–534.
(36) Heibges, A.; Salamini, F.; Gebhardt, C. Functional comparison of homologous members of three groups of Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Mol. Genet. Genomics* **2003**, *269* (4), 535–541.
(37) Ishikawa, A.; Ohta, S.; Matsuoka, K.; Hattori, T.; Nakamura, K. A family of potato genes that encode Kunitz-type proteinase inhibitors: structural comparisons and differential expression. *Plant Cell Physiol.* **1994**, *35* (2), 303–312.
(38) Walsh, T. A.; Twitchell, W. P. Two Kunitz-type proteinase inhibitors from potato tubers. *Plant Physiol.* **1991**, *97* (1), 15–18.
(39) Mosolov, V. V.; Valueva, T. A. Proteinase inhibitors and their function in plants: a review. *Prikl. Biokhim. Mikrobiol.* **2005**, *41* (3), 261–282.
(40) Huai-Jun, S.; Jun, L.; Cong-Hua, X. Transformation of potato using an antisense class I patatin gene and its effect on microtuber formation. *Chin. J. Agric. Biotechnol.* **2005**, *2* (1), 7–11.

(41) Lawrence, P. K.; Koundal, K. R. Plant protease inhibitors in control of phytophagous insects. *Electronic J. Biotechnol.* **2002**, *5* (1).

(42) Nelson, C. E.; Walker-Simmons, M.; Ryan, C. A. Regulation of proteinase inhibitor synthesis in tomato leaves: in vitro synthesis of inhibitors I and II with mRNA from excised leaves induced with PIIF (proteinase inhibitor inducing factor). *Plant Physiol.* **1981**, *67* (4), 841–844.

(43) Sonnewald, U. Control of potato tuber sprouting. *Trends Plant Sci.* **2001**, *6* (8), 333–335.

(44) Destefano-Beltran, L.; Knauber, D.; Huckle, L.; Suttle, J. C. Effects of postharvest storage and dormancy status on ABA content, metabolism, and expression of genes involved in ABA biosynthesis and metabolism in potato tuber tissues. *Plant Mol. Biol.* **2006**, *61* (4−5), 687–697.

Received for review September 11, 2008. Revised manuscript received November 18, 2008. Accepted December 17, 2008. Collaboration of the National Institute of Biology was supported by the Slovenian Ministry of Science, Education and Sport (Ph.D. Grant 3311-02-831032 and Programme Plant Physiology and Biotechnology P4-0165). Support from the EU under the FP5 RTD Project Blight-MOP (QLK5-CT-2000-01065) and the FP6 Project Safefoods (Food-CT-2004-506446) is gratefully acknowledged. Also, support from the Dutch Ministry of Agriculture, Nature and Food Quality is gratefully acknowledged.

JF802815D