12-2008

Plant transcriptional responses to explosives as revealed by *Arabidopsis thaliana* microarrays and its application in phytoremediation and phytosensing

Murali Malavalli Keerthi Narayana Raghavendra Rao
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Murali Malavalli Keerthi Narayana Raghavendra Rao entitled "Plant transcriptional responses to explosives as revealed by *Arabidopsis thaliana* microarrays and its application in phytoremediation and phytosensing." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Plants, Soils, and Insects.

C. Neal Stewart Jr., Major Professor

We have read this dissertation and recommend its acceptance:

Carl Sams, Mark Radosevich, Arnold Saxton, Hamparsum Bozdogan

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Murali Malavalli Keerthi Narayana Raghavendra Rao entitled "Plant transcriptional responses to explosives as revealed by *Arabidopsis thaliana* microarrays and its application in phytoremediation and phytosensing." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Plants, Soils and Insects.

C. Neal Stewart, Jr., Major Professor

We have read this dissertation
and recommend its acceptance:

Dr. Carl Sams

Dr. Mark Radosevich

Dr. Arnold Saxton

Dr. Hamparsum Bozdogan

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Plant transcriptional responses to explosives as revealed by *Arabidopsis thaliana* microarrays and its application in phytoremediation and phytosensing

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Murali Malavalli Keerthi Narayana Raghavendra Rao

December 2008

Dedication

This dissertation is dedicated to my loving wife Nandini, my parents Keerthi Narayana Rao and Sarojamma, my aunt Usha and uncle Krishna, my very best friends Thirumalesh Chigateri and Dr. Santosh Kumar, and the rest of the family for their love, support and encouragement.

Acknowledgements

I am indebted to Dr. Neal Stewart for the financial assistance, his continuous support and guidance, and for giving me the opportunity to work with him. I am grateful to my committee members Dr. Carl Sams, Dr. Mark Radosevich, Dr. Arnold Saxton, Dr. Hamparsum Bozdogan, and Dr. Janice Zale for their support, guidance and for serving on my graduate committee. I deeply appreciate the support from all the members of the Stewart Lab. I would like to acknowledge the Armed Forces Medical Intelligence Center (AFMIC), the University of Tennessee Agricultural Experiment Station, Knoxville, and the United States Department of Agriculture for funding support. Finally, I would also like to thank all my friends for their assistance.

Abstract

This research focused on understanding genetic responses of plants to explosives, which is necessary to produce plants to detect and clean soil and water contaminated with toxic explosive compounds. The first study used microarray technology to reveal transcriptional changes in the model plant *Arabidopsis thaliana* exposed to the explosive compounds RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine; Royal Demolition Explosive or Research Department Explosive) and TNT (2,4,6-trinitrotoluene). This study yielded a list of genes up- and down-regulated by explosive compounds, which can be potentially used for phytoremediation (remediation using plants) or phytosensing (detection using plants) of explosive compounds. The second study presented biotechnology tools to enhance phytosensing that might have application in not only explosives phytosensing but also sensing of other contaminants or important biological agents. This study addressed the problem of low detectable levels of reporter gene signal from a phytosensor and the results suggest the potential use of a site-specific recombination system to amplify the reporter gene signal. The final study addressed microarray data analysis and best practices for statistical analysis of microarray data. Standard parametric approaches for microarray analysis can be very conservative, indicating no unusable information from expensive microarray experiments. A nonparametric method of analysis on a variety of microarray datasets proved to be effective in providing reliable and useful information, when the standard parametric approach used was too conservative.

Table of contents

List of Tables

List of Figures

Introduction

Contamination and plants

Human activities such as manufacturing, mining, and industrialization have contributed to widespread soil and water contamination (Cunningham et al. 1995). The subsequent necessity to remediate soils has led to the use of a variety of physical, chemical and biological technologies (Cunningham et al. 1995). Current remediation technologies available to remove contaminants from the environment comprise incineration, land filling and composting, all of which are inefficient, expensive and physically challenging (Hannink et al. 2002; Nishino et al. 2000; Peterson et al. 1998). Incineration destroys the soil structure, disturbs ecology, and costs between US $523 and $785 per cubic meter of soil, while landfilling results in displacement of contamination to another site, and composting possibly will result in partial breakdown of the contaminants with costs between $528 and $611 per cubic meter of soil (Hannink et al. 2002; Nishino et al. 2000; Peterson et al. 1998). The drawbacks of the existing technologies have resulted in efforts to search for more cost-effective technologies that are biology-based. In this regard, plant-based systems have received wide attention.

Plants are known to modify physical, chemical and biological processes that occur in their immediate surroundings (Cunningham and Ow, 1996). Plants are recognized for tolerating soil contaminants such as herbicides at levels that are significantly higher compared to the regulatory limits (Cunningham and Ow, 1996). When grown in a contaminated area, plants potentially play a role in the modification and removal of contaminants (Cunningham and Ow, 1996). Plants have already been used in the

remediation of several environmental systems (Cunningham et al. 1995). They have been used over many years to treat certain kinds of waste waters in constructed wetlands, reed beds and floating-plant based systems (Cunningham et al. 1995). Current efforts in plant-based systems have been extended to address soil and water contamination (Cunningham et al. 1995). This abatement concept of using plants to concentrate and metabolize environmental contaminants is called "phytoremediation". The two greatest advantages of phytoremediation compared with traditional abatement methods are 1. cost-effectiveness, and 2. soils remain in place thereby causing less ecosystem disruption. Cropping systems with costing between $200 and $10,000 per hectare would correspond to a remediation cost of $0.02-1.00 per cubic meter of soil; a three to four orders of magnitude savings over existing physico-chemical methods (Cunningham et al. 1995).

Phytoremediation is comprised of several processes: phytoextraction, phytodegradation, phytostabilization, rhizodegradation, and phytovolatilization (Salt et al. 1998; Burken et al. 2000; Pilon-Smits 2005). Phytoextraction is a process where the contaminant (metals or organics) is taken up by plant and stored in the harvestable parts of the plants. During phytodegradation, the contaminant is degraded in the plant to a less toxic compound. In the case of phytostabilization the contaminant is reduced to a less bioavailable compound, and rhizodegradation involves degradation of the contaminant by the microbes in the rhizosphere, which is enhanced by the plant root exudates. Finally, phytovolatilization is a process in which the contaminant is taken up by the plant and then released into the atmosphere as volatiles (Salt et al. 1998; Burken et al. 2000; Pilon-Smits 2005).

Plants may also be used to monitor and report the presence of chemical contamination: "phytosensing". In this scenario, when plants are grown in the presence of a contaminant they produce a detectable phenotypic response. Plant based monitoring systems would be a cost-effective alternative for current monitoring systems which are expensive and labor-intensive. Here we discuss how phytosensing and phytoremediation could be applied to detect and remediate explosive chemicals contamination in the environment.

Explosives as contaminants

Explosives such as TNT (2,4,6-trinitrotoluene), RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine; Royal Demolition Explosive or Research Department Explosive), and HMX (octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine; High Melting Explosive) are widely used in military ammunition (Best et al. 2001; Hannink et al. 2002; Halasz et al. 2002). Explosives in general, can be broken down into three major chemical categories, which comprise nitroaromatics, nitramines and nitrate esters (Hannink et al. 2002). Nitroaromatics include the widely used explosive, TNT, and is distinguished by an aromatic ring with three nitro groups (Hannink et al. 2002). Nitramines include RDX and HMX, in which RDX is presently the most extensively used explosive and this class of explosives is characterized by the presence of N-nitro groups (Hannink et al. 2002). Nitrate esters are the esters of nitric acid and consists of PETN (pentaerythritol tetranitrate), GTN (glycerol trinitrate or nitroglycerin), and nitrocellulose (Hannink et al. 2002). Nitrate esters usually contain many O-nitro groups (Hannink et al. 2002). All these explosives are generally recalcitrant to degradation and remain in the biosphere in ecological time, where they constitute a source of pollution resulting in toxic, mutagenic

and carcinogenic effects on humans and other biota. In humans, high and prolonged exposures to TNT cause hyperplasia of the bone marrow leading to aplastic anemia; and a drastic loss of blood platelets (Rosenblatt, 1980). Toxic hepatitis is also reported in humans from TNT exposure and RDX toxicity includes gastrointestinal, central nervous system (generalized convulsions), and renal effects (Rosenblatt, 1980). The means of exposure is inhalation or ingestion and high melting and lipid insolubility properties of RDX make skin absorption unlikely (Rosenblatt, 1980). Obviously, all of the aforementioned human health risks pale to being blown up!

Explosives and their breakdown products are the major contaminants in the environment derived exclusively from human activity; i.e., explosives are xenobiotics (compounds that are foreign to living organisms). Activities such as manufacturing, testing, field usage and improper disposal can contribute to soil and water contamination with the explosive compounds and its breakdown products (Best et al. 1997; Best et al. 2001; Halasz et al. 2002; Rosenblatt et al. 1991). There is also unexploded ordnance (UXO) on many military and non-military sites worldwide. In addition to the risk of being injured upon detonation, landmines leak explosives from inexpensive plastic cases causing uncontrolled spread of toxins. There are over 100 million landmines deployed in over 70 countries and more than 20,000 people are killed each year according to an UN estimate (http://www.un.org/Photos/mines/MINES.html). More than 50 million acres in the United Sates is contaminated with UXO and with the existing detection and remediation technologies, the projected cost for clearing is over $500 billion (Zhang et al. 2003). Clearing landmines from the civilian areas can be very difficult and dangerous

and expensive with estimates being more than US $1000 per mine cleared (Hussein and Waller, 2000).

A recent newsfocus article in *Science* reported that the U.S. National Science Foundation and the Department of Homeland Security have been actively funding research on detecting explosives and the development of an effective method for explosive detection still requires wide-spread improvements in many areas of research (Bhattacharjee, 2008). Currently used 'local' small-scale methods include visual inspection, hand-held metal detectors, mine prodders, and explosive-detecting dogs (Hussein and Waller, 2000). These are all severely limited in scope and are tedious and require disciplined, well-trained personnel (Hussein and Waller, 2000). One of the major challenges in demining is distinguishing between an anomaly and a landmine or in other words, specificity remains a big problem with current technologies (Hussein and Waller, 2000). Most of the existing and emerging landmine detection technologies focus on the detection of anomalies (Hussein and Waller, 2000). Therefore, this problem of specificity in demining remains unaddressed, and as a result, each passing day finds more and more deployed landmines (Hussein and Waller, 2000). All these current challenges in landmine and UXO clearing make plant-based wide-area sensing of landmines and UXO a novel and lucrative approach by being cost-effective, safer and more specific than the current methods. Thus, the problem of detection and removal of explosives is huge and phytoremediation and phytosensing are attractive options.

Plants and Explosives

We have very little knowledge about uptake and transport of explosives in plants (Hannink et al. 2002). In general, phytodegradation is similar to human metabolism of

xenobiotics (Ishikawa 1992; Sandermann, 1992; Ishikawa et al. 1997). Explosive contaminants such as TNT (nitroaromatics) have been largely reported to undergo phytodegradation, whereas nitramines such as RDX and HMX are reported to undergo phytoextraction (Hannink et al. 2002).

Phytotoxicity of the explosive compounds impacts the utility of plants to remediate contaminated sites; phytoremediation is predicated on tolerance of a plant species to the contaminant of interest (Hannink et al. 2002). High concentrations of TNT cause chlorosis, whereas RDX and HMX have lower toxicities (Hannink et al. 2002).

Effective application of phytoremediation also requires the knowledge of the uptake and fate of these compounds in plants. TNT is readily taken up by the plants and reduction is the most commonly observed transformation reaction in plants (Burken et al. 2000). Studies so far have indicated that plants are capable of transforming TNT and are generally accumulated in the roots (Harvey et al. 1990; Hughes et al. 1997; Larson et al. 1999). Studies on RDX uptake indicate that RDX is extensively translocated and sequestered in the leaf tissues of the plants (Harvey et al. 1991). Uptake and transformation studies were conducted by Larson et al. (1999) using ^{14}C labeled RDX in agricultural crops. They found accumulation of unknown high-molecular-weight RDX transformation products. Similar experiments conducted by Best et al. (1999) in three submersed and four emergent wetland species also indicated accumulation of RDX as unknown transformation products at places, where new plant material was produced. In poplar trees, Thompson et al. (1999) observed that RDX was readily translocated and accumulated in leaf tissues. HMX was highly recalcitrant to deposition by plants (Bhadra et al. 2001). Goel et al. (1997) showed that the nitrate ester (GTN, nitroglycerin) was

degraded by sugar beet (*Beta vulgaris* L.) cell cultures, and very little to no reduced, conjugated or cell-bound carbonaceous metabolites were formed. Williams et al. (2004) showed that five enzymes belonging to the yeast old yellow enzyme (OYE) family are capable of catalyzing reduction of TNT. The results from the research to date thus imply that plants interact with explosives and have the potential to degrade and/or sequester them.

However, plants in their natural state may not be able to sufficiently accumulate and degrade explosives and endogenous processes may simply be prohibitively slow and inefficient. Therefore, genetic engineering might be necessary to increase phytoremediation capacity, and certainly required for phytosensing applications. The first step towards enhancement of phytoremediation would be to gain a better understanding of the molecular biology, especially genomics of plants. The most important aspect is to study transcriptional responses of plants exposed to explosives (transcriptomics). This would reveal the genes potentially involved in the metabolism of explosives, which is necessary for developing phytosensors or phytoremediators.

References

Best, E.P.H., Zappi, M.E., Fredrickson, H.L., Sprecher, S.L., Larson, S.L. and Ochman, M. 1997. Screening of aquatic and wetland plant species for phytoremediation of explosives-contaminated groundwater from the Iowa Army Ammunition Plant. Ann. N. Y. Acad. Sci. 829:179-194.

Best E.P.H., Sprecher S.L., Larson S.L., Fredrickson H.L., Bader D.F. 1999. Environmental behavior of explosives in groundwater from the Milan Army Ammunition plant in aquatic and wetland plant treatments: Uptake and fate of TNT and RDX in plants. Chemosphere 39:2057-2072.

Best E.P.H., Miller J.L., Larson S.L. 2001. Tolerance towards explosives, and explosives removal from groundwater in treatment wetland mesocosms. Water Sci. Technol. 44:515-521.

Bhadra R., Wayment D.G., Williams R.K., Barman S.N., Stone M.B., Hughes J.B., Shanks J.V. 2001. Studies on plant-mediated fate of the explosives RDX and HMX. Chemosphere 44:1259-1264.

Bhattacharjee, Y. 2008. New efforts to detect explosives require advances on many fronts. Science 320:1416-1417.

Burken J.G., Shanks J.V., Thompson P.L. 2000. Phytoremediation and plant metabolism of explosives and nitroaromatic compounds. In: Biodegradation of Nitroaromatic compounds and explosives, Spain J.C., Hughes J. B., Knackmuss H. (eds.), CRC Press, Boca Raton, Florida, pp. 239-276.

Cunningham S.D., Berti W.R., Huang J.W.W. 1995. Phytoremediation of contaminated soils. Trends Biotechnol. 13:393-397.

Cunningham S.D., Ow D.W. 1996. Promises and prospects of phytoremediation. Plant Physiol. 110:715-719.

Goel A., Kumar G., Payne G.F., Dube S.K. 1997. Plant cell biodegradation of a xenobiotic nitrate ester, nitroglycerin. Nat. Biotechnol. 15:174-177.

Halasz A., Groom C., Zhou E., Paquet L., Beaulieu C., Deschamps S., Corriveau A., Thiboutot S., Ampleman G., Dubois C., Hawari J. 2002. Detection of explosive and their degradation products in soil environments. J. Chromatogr. 963:411-418.

Hannink N.K., Rosser S.J., Bruce N.C. 2002. Phytoremediation of explosives. Crit. Rev. Plant Sci. 21:511-538.

Harvey S.D., Fellows R.J., Cataldo D.A., Bean R.M. 1990. Analysis of 2,4,6-trinitrotoluene and its transformation products in soils and plant-tissues by high-performance liquid chromatography. J. Chromatogr. 518:361-374.

Harvey S.D., Fellows R.J., Cataldo D.A., Bean R.M. 1991. Fate of the explosive hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) in soil and bioaccumulation in bush bean hydroponic plants. Environ. Toxicol. Chem. 10:845-855.

Hughes J.B., Shanks J.V., Vanderford M., Lauritzen J., Bhadra R. 1997. Transformation of TNT by aquatic plants and plant tissue cultures. Environ. Sci. Technol. 31:266-271.

Hussein, E.M.A., Waller, E.J. 2000. Landmine detection: the problem and the challenge. Appl. Radiat. Isot. 53:557-563.

Ishikawa, T. 1992. The ATP-dependent glutathione s-conjugate export pump. Trends Biochem. Sci. 17:463-469.

Ishikawa, T., Li, Z.S., Lu, Y.P., Rea, P.A. 1997. The GS-X pump in plant, yeast, and animal cells: Structure, function, and gene expression. Biosci. Rep. 17:189-207.

Larson S.L., Jones R.P., Escalon L., Parker D. 1999. Classification of explosives transformation products in plant tissue. Environ. Toxicol. Chem. 18:1270-1276.

Nishino S.F., Spain J.C., He Z. 2000. Strategies for aerobic degradation of Nitroaromatic compounds by bacteria: process discovery to field application. In: Biodegradation of Nitroaromatic compounds and explosives, Spain J.C., Hughes J.B., Knackmuss H. (eds.), CRC Press, Boca Raton, Florida, pp. 7-61.

Peterson M.M., Horst G.L., Shea P.J., Comfort, S.D. 1998. Germination and seedling development of switchgrass and smooth bromegrass exposed to 2,4,6-trinitrotoluene. Environ. Pollut. 99:53-59.

Pilon-Smits E. 2005. Phytoremediation. Annu. Rev. Plant Biol. 56:15-39.

Rosenblatt D.H. 1980. Toxicology of explosives and propellants. In: Encyclopedia of Explosives and Related Items, Kaye S.M. (ed.), US Army Armament Research Development Committee, Dover, New Jersey.

Rosenblatt D.H., Burrows S.P., Mitchell W.R., Parmer D.L. 1991. Organic explosives and related compounds. In: The Handbook of Evironmental Chemistry, Hutzinger O. (ed.), Springer-Verlag, New York.

Salt D.E., Smith R.D., Raskin I. 1998. Phytoremediation. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:643-668.

Sandermann H. 1992. Plant-metabolism of xenobiotics. Trends Biochem. Sci. 17:82-84.

Thompson P.L., Ramer L.A., Schnoor J.L. 1999. Hexahydro-1,3,5-trinitro-1,3,5-triazine translocation in poplar trees. Environ. Toxicol. Chem. 18:279-284.

Williams R.E., Rathbone D.A., Scrutton N.S., Bruce N.C. 2004. Biotransformation of explosives by the old yellow enzyme family of flavoproteins. Appl. Environ. Microbiol. 70:3566-3574.

Zhang, Y., Collins, L.M., Carin, L. 2003. Unexploded ordnance detection using Bayesian physics-based data fusion. Integrated Computer-Aided Engineering 10: 231.

Transcriptional responses to explosives RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine) and TNT (2,4,6-trinitrotoluene) as revealed by *Arabidopsis thaliana* microarrays

Abstract

High explosives such as RDX (hexahydro – 1,3,5 – trinitro – 1,3,5 – triazine, Royal Demolition Explosive or Research Department Explosive), and TNT (2,4,6 – trinitrotoluene) are important contaminants in the environment and phytoremediation has been viewed as a cost-effective abatement. There remains, however, an insufficient knowledge-base about how plants respond to explosives. In this context a comprehensive microarray analysis was conducted in *Arabidopsis thaliana* to study the effect of these compounds on the transcriptional profile. Our results for both RDX and TNT were consistent with the existing theory for xenobiotic detoxification in plants. Among the genes that were differentially expressed included oxidoreductases, cytochrome P450's, transferases, transporters, and several unknown expressed proteins. We discuss the suggestive role of some of these up-regulated genes in the context of explosive metabolism in plants. This study reports the genes affected by the explosive compounds RDX and TNT and is useful not only in finding potential target genes for use in the phytoremediation of RDX and TNT, but also for phytosensing (detecting the presence of contaminants using plants) of these explosives.

Introduction

Explosives such as RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine, Royal Demolition Explosive or Research Department Explosive), and TNT (2,4,6-trinitrotoluene) are widely used in military munitions (Best et al. 2001; Hannink et al. 2002; Halasz et al. 2002). These explosives and their breakdown products are among the major human-produced contaminants in the environment; manufacturing, deployment and improper disposal contribute to contamination (Best et al. 1997; Best et al. 2001; Halasz et al. 2002; Rosenblatt et al. 1991). RDX and TNT are important constituents of unexploded ordnance (UXO) on many military and non-military sites. Landmines leak explosives from inexpensive plastic cases. There are over 100 million landmines deployed worldwide (UN estimate http://www.un.org/Photos/mines/MINES.html; http://www.unicef.org/graca/mines.htm; http://www.cyberschoolbus.un.org/sds/introduction/slideshow_print.html). These explosives are generally recalcitrant to degradation and remain in the biosphere in ecological time, where they constitute a source of pollution resulting in toxic, mutagenic and carcinogenic effects on humans and other biota. Thus, RDX and TNT require widespread environmental abatement.

Remediation technologies commonly available for environmental abatement comprise incineration, land filling and composting, which are expensive and physically challenging (Hannink et al. 2002; Nishino et al. 2000; Peterson et al. 1998). The drawbacks of these existing technologies have led to use of plant-based systems (phytoremediation), which are cost-effective and eco-friendly.

There are several studies showing that plants, in general, readily take up RDX and TNT. For example, recently Vila and others reported that crop plants (maize, soybean, wheat and rice) could grow on soils containing RDX and TNT and were able to uptake these compounds (Vila et al. 2007). In another recent study, it was reported that maize (*Zea mays* L.) and broad beans (*Vicia faba* L.) were able to remove TNT (Van Dillewijn et al. 2007). Also, *Catharanthus roseus* (Vinca) hairy root cultures, *Myriophyllum aquaticum* (parrot feather) plants, and hybrid poplars have been reported to take up RDX (Bhadra et al. 2001; Thompson et al. 1999). Harvey and others have reported bioaccumulation of RDX in bush bean (*Phaseolus vulgaris*) hydroponic plants (Harvey et al. 1991). However, plants are typically inefficient to accumulate and degrade explosives. Plants in their natural state may not be able to sufficiently accumulate and degrade explosives or endogenous processes may simply be prohibitively slow and inefficient. Therefore, genetic engineering might be necessary to increase phytoremediation capacity, and certainly required for phytosensing, i.e., using plants to report the presence of contamination. Understanding plant transcriptional responses to these compounds is thus necessary and useful for developing phytosensors or phytoremediators. The first step towards enhancement of phytoremediation would be to gain a better understanding of the molecular biology, especially functional genomics of plants. Mentewab et al. (2005) used cDNA microarrays to determine the transcriptional response of *Arabidopsis thaliana* to TNT, but the microarrays represented only about half the genome. In another study, Patel and others (2004), used microarrays to study differential gene expression of *Chlamydomonas reinhardtii* exposed to TNT. Ekman et al. (2003) used serial analysis of gene expression (SAGE) to study transcriptome responses in Arabidopsis roots

exposed to TNT. Ekman et al. (2005) also used SAGE to study the gene expression changes in Arabidopsis seedling roots exposed to RDX. Mezzari et al. (2005) analyzed expression of only few selected genes in Arabidopsis exposed to explosive compounds and chloroacetanilide herbicides. Most recently, Tanaka et al. (2007) analyzed expression of only a few selected genes in poplar exposed to RDX.

In an attempt to better understand the full complement of transcriptional expression patterns in response to RDX and TNT exposure, we conducted a comprehensive Arabidopsis oligonucleotide microarray analysis of whole Arabidopsis seedlings exposed to steady state doses of RDX and TNT. The gene expression patterns in response to RDX exposure was of specific interest since there have been no prior gene expression studies in response to RDX involving whole plants. Since RDX is extensively translocated and known to be accumulated in the leaf tissues (Best et al. 1999; Harvey et al. 1991; Thompson et al. 1999), a whole-plant investigation was warranted. Also, the transcriptional profile in response to RDX exposure was analyzed on two different microarray platforms (Affymetrix and two-color long-oligo printed glass slides), while the response to TNT was analyzed on only the Affymetrix microarray platform.

Results

Growth of plants on RDX and TNT media

The optimal concentrations for the microarray experiment were determined by analyzing the growth responses and phytotoxicity tolerance threshold of *Arabidopsis thaliana* (ecotype Columbia) plants to a range of RDX and TNT in MS media. Based on the primary root growth, 0.5 mM of RDX and 2.0 μM of TNT were considered as sub-

lethal concentrations and were used for the subsequent microarray experiments (Fig. 2.1, Fig. 2.2 and Fig. 2.3)

Microarray analysis of RDX treated plants

Two-color platform

A false discovery rate (FDR) value cut-off of 10 percent and 1.5-fold change criteria resulted in 173 genes that were differentially regulated. The top 20 up-regulated and down-regulated genes based on fold change (linear scale) are presented in Table 2.1. The most up-regulated gene in this experiment was a leucine-rich repeat family protein (*At4g33970*) with a two-fold change. Genes from this family are cell wall constituents and known to be involved in protein-protein interactions in plants, as well as transducing pathogen recognition signals (Baumberger et al. 2003; Kobe and Deisenhofer, 1994; Li and Chory, 1997). Other genes that were up-regulated included a protease inhibitor/lipid transfer protein (*At4g12500*) which is involved in lipid transport and lipid binding (Rhee et al. 2003), a putative mannitol transporter (*At4g36670*), which is located in the membrane and involved in carbohydrate transporter activity, a putative xyloglucan:xyloglucosyl transferase (*At4g14130*), multi-copper oxidase type I family protein (*At1g21850*) which has oxido-reductase activity, lipoxygenase (*LOX2*; *At3g45140*), which is targeted to chloroplast and is known to be involved in wound induced jasmonic acid accumulation in Arabidopsis (Bell et al. 1995) and several genes with unknown biological function were also up-regulated.

The down-regulated genes included a putative cysteine protease (*At4g11320*), a putative protease inhibitor (*At1g73330*) which is responsive to drought (Rhee et al. 2003), phosphoribulokinase (*At1g32060*), an ABC transporter family protein (*At5g64840*), and a

xyloglucan:xyloglucosyl transferase (*At5g57560*) that is linked to cold tolerance (Purugganan et al. 1997).

Affymetrix platform

An FDR cut-off value of 10 percent and fold change of 2.0 yielded 217 differentially expressed genes in this experiment. The top 20 up-regulated and down-regulated genes based on fold change (linear scale) from this experiment are presented in Table 2.2. Among the up-regulated genes, lipoxygenase (*LOX2*; *At3g45140*) was the most up-regulated gene with around seven fold change in expression compared to control. Genes that were also up-regulated included an ABC transporter (*At2g39350*) which was up-regulated 4.8-fold is expressed in roots and is responsive to nematodes (Rhee et al. 2003), a UDP-glucoronosyl/UDP-glucosyl transferase family protein (*At5g49690*), a putative peroxidase (*At5g39580*), a glutaredoxin family protein (*At1g03020*) which has arsenate reductase (glutaredoxin) activity (Rhee et al. 2003), a sugar transporter family protein (*At1g73220*), and several genes with unknown biological function. The most down-regulated gene (19-fold) was an unknown expressed protein (AT1g13650). Other genes exhibiting repressed transcription in response to RDX stress included genes encoding for a, neurofilament protein-related (*At3g05900*), a cytochrome p450 family protein (*At5g47990*), a putative myrcene/ocimene synthase (*At3g25820*), a putative pathogen-responsive alpha-dioxygenase (*At3g01420*), and several expressed proteins with unknown biological function.

Correlation between the Affymetrix and the two-color platforms

A simple correlation analysis indicated a positive moderate relationship between Affymetrix and two-color microarray platforms using $\log_2$ ratios of the signal intensities

for all the genes (Fig. 2.5). The Pearson correlation coefficient (r) between Affymetrix and two-color $\log_2$ ratios was 0.38 (p-value < 0.0001).

Functional characterization of genes differentially regulated in response to RDX

The GO tool used categorized genes by three categories: cellular component, biological process, and molecular function. Here discussed briefly is the categorization by molecular function. The pie charts showing functional categorization of differentially regulated genes by molecular function for both two-color, as well as Affymetrix are shown in Figure 2.4.

Two-color

Functional categorization by molecular function revealed that most of the genes (38%) were involved in other molecular functions, followed by other binding (12.2%), and transcription factor activity (10.6%) categories. In the case of down-regulated genes, other molecular functions formed the largest category with 27.8% of genes, followed by hydrolase activity (13.9%), and protein binding (11.1%).

Affymetrix

Categorization of genes up-regulated in this experiment by molecular function revealed that other molecular functions, other binding and other enzyme activity were the largest categories similar to results from two-color and consisted of 28.9%, 14.9%, and 14.0% of the up-regulated genes respectively. This was followed by transcription factor activity (8.8%) and transporter activity (7.9%) categories. Categorization of the down-regulated genes by molecular function also indicated that other molecular functions, other

binding and other enzyme activity were the largest categories with 38.2%, 12.2%, and 8.4% of the genes respectively.

Real-time RT-PCR analysis

Real-time RT-PCR analysis was carried out for six genes, three of which were up-regulated and three of which were down-regulated in the RDX microarrays. The results obtained from this analysis corresponded well with the microarray analysis (Table 2.4).

Affymetrix microarray analysis of TNT treated plants

Analysis revealed that 297 genes were differentially expressed at an FDR cut-off value of 10 percent and fold change cut-off of 2.0. The top 20 up-regulated and down-regulated genes from this experiment are presented in Table 2.3. In this experiment, the most up-regulated gene was an expressed protein with unknown biological function (*At3g15310*) with a fold-change of 17. Other up-regulated genes included an O-methyltransferase N-terminus domain containing protein (*At5g42760*), a putative pathogenesis-related protein (*At4g33720*), a putative cysteine proteinase (*At2g27420*), a myb family transcription factor (*At1g01520*), and many other expressed proteins with unknown function. Among the down-regulated genes, a male sterility MS5 family protein (*At5g48850*) was the most down-regulated gene, and other down-regulated ones included a putative CTP synthase (*At1g30820*), a putative glycine hydroxymethyltransferase (*At1g36370*), a glycosyl transferase family 20 protein (*At2g18700*), a glutaredoxin family protein (*At3g62950*), and a protease inhibitor (*AT4g12500*).

Functional characterization of genes differentially regulated in response to TNT

Functional categorization by molecular function of differentially-expressed genes was apparently different between the TNT and RDX experiments (Fig. 2.4). Categorization of the up-regulated genes revealed 27.1% of them being involved in other molecular functions followed by 17.5% of the genes being associated with transcription factor activity. Transferase activity with 7.3% of the genes came next followed by transporter activity (5.0 %), and hydrolase activity (4.5 %). With respect to categorization of the down-regulated genes by molecular function, the largest category was of the genes (25.2%) involved in other molecular functions, followed by categories transcription factor activity (12.1%), and hydrolase activity (11.2%).

Meta-analysis to identify genes unique to RDX and TNT

To ensure that the gene list obtained is specific to RDX and TNT and not a general response to similar nitrogenous compounds, different Arabidopsis microarray databases were searched for microarray experiments involving nitrogenous compounds in Arabidopsis. Only three relevant Affymetrix datasets involving nitrate treatment in Arabidopsis were found, which were downloaded. After comparing our up-regulated gene lists from the Affymetrix experiments with the gene lists from the downloaded datasets, only one gene from the Affymetrix RDX up-regulated gene list (*At4g36010*) and none from the TNT gene list were filtered.

Expression profile analysis using Genevestigator

The expression profile for the top 20 up-regulated genes from all the three microarray experiments, under different general stress conditions were studied using Genevestigator, a reference expression database and a meta-analysis system

(Zimmermann et al. 2004; Zimmermann et al. 2005). The expression profile for the top 20 up-regulated genes from RDX two-color, RDX-Affymetrix, and TNT-Affymetrix microarray experiments are presented in Figure 2.6, Figure 2.7, and Figure 2.8 respectively.

Discussion

Phytotoxicity of the explosive compounds impacts the utility of plants to remediate contaminated sites; phytoremediation is predicated on tolerance of a plant species to the contaminant of interest (Hannink et al. 2002). High concentrations of TNT cause chlorosis, whereas RDX and HMX have lower toxicities (Hannink et al. 2002). Lucero et al. (1999) conducted cell culture experiments in angel's trumpet (*Datura innoxia*) to determine phytotoxicity of explosives and found cytotoxity at 131 μM (30 mg/L) of TNT, while RDX concentrations of 173 to 270 μM (38 to 60 mg/L) did not affect cell growth. In a hydroponic study, hybrid poplars exposed to concentrations of up to 21 mg/L of RDX for 14 days did not show any symptoms of toxicity (Thompson et al. 1999). Here we confirm that TNT is toxic to *Arabidopsis thaliana* at much lower concentrations than RDX. RDX concentration of up to 0.1 mM did affect primary root growth, but TNT concentration of 2.0 μM significantly stunted root growth. The availability, uptake, and accumulation of explosives in plants are also dependent on soil properties. Poplars grown in TNT contaminated soils had reduced uptake of TNT than plants in hydroponic studies; TNT adsorbs to soil particles (Burken et al. 2000). Also, Haderlein et al. (1996) reported much lower adsorption of RDX to clay compared to TNT. The concentration of RDX and TNT in contaminated soils can range from 0.7 to 74,000

mg/kg and 0.08 to 87000 mg/kg respectively (Best et al. 2006; Talmage et al. 1999). Phytoremediation of high level contamination levels will require genetic engineering of plants to increase their tolerance and phytoremediation capacities.

RDX microarrays

The differentially expressed genes from both two-color and Affymetrix were identified using the non-parametric rank-product statistics (Breitling et al. 2004) approach. This non-parametric rank-product statistics approach offer advantages such as fewer assumptions about the data and is especially powerful when there is small number of replicates that are typical of microarray experiments (Breitling et al. 2004).

There is no earlier report on whole genome expression studies in response to RDX except for the study conducted by Ekman et al. (2005) where they studied gene expression in Arabidopsis roots, but since RDX is readily translocated and accumulated in leaf tissues (Best et al. 1999; Harvey et al. 1991; Thompson et al. 1999), studying gene expression in whole plants is more appropriate and consistent with the objective of phytosensor engineering. In another study, Mezzari et al. (2005) studied expression of only five selected genes in response to RDX and other xenobiotics using semi-quantitative reverse-transcription PCR technique. Recently, Tanaka et al. (2007) studied expression of few selected genes in poplar exposed to RDX using real time RT-PCR. Here I studied responses to RDX in Arabidopsis using two different microarray platforms aiming towards more comprehensive insight into RDX metabolism in plants.

The metabolic detoxification of xenobiotics

Plant metabolism of xenobiotics involves three phases: activation (transformation), conjugation and compartmentation (elimination) (Sandermann 1992; Ishikawa 1992;

Ishikawa 1997; Rea et al 1998; Coleman et al. 1997; Schaffner et al. 2002). Activation generally involves oxidation or hydrolysis or reduction type of reactions, where functional groups such as hydroxyl (-OH), carboxyl (-COOH) are added to the contaminant with enzymatic involvement of cytochrome P450 monooxygenases, esterases, reductases, dehalogenases, and dehydrogenases. The products of phase I (activation) are more hydrophilic and sometimes more toxic than the parent compound. In the phase II (conjugation) the activated contaminant undergoes deactivation by the formation of covalent linkages with endogenous hydrophilic molecules such as glucose, malonate, glutathione or carboxylic acids using glucosyltransferase-, glutathione-S-transferase-, and acyltransferase-mediated reactions that result in water soluble conjugates that are less toxic compared to the parent compound. Phase III (compartmentation) involves exporting conjugates to either the vacuole or apoplast using ABC transporters or multidrug and toxic compound extrusion (MATE) transporters (Sandermann 1992; Ishikawa 1992; Ishikawa 1997; Rea et al 1998; Coleman et al. 1997; Schaffner et al. 2002).

Our gene regulation results are consistent with previous RDX physiological accumulation results that might indicate detoxification of RDX in plants. For example, Best et al. (1999) conducted experiments in three submersed and four emergent wetland species and indicated accumulation of ^{14}C-RDX and unknown transformation products mostly in the shoots. Similarly, Larson et al. (1999) examined the uptake of RDX in plants utilizing ^{14}C-labelled RDX and reported accumulation of high concentrations of unknown high-molecular-weight RDX transformation products with only a small portion remaining as RDX. They also reported low level accumulation of hexahydro-1-nitroso-

3,5-dinitro-1,3,5-triazine (MNX), which is a degradation product of RDX and suggested that low-level accumulation is indicative of MNX being either a minor transformation product of RDX or MNX being further transformed to unknown products. More recently, Best et al. (2006) reported significant levels of RDX and MNX in plant tissues when exposed to RDX in soil. Best and co-workers (2005) suggested that accumulation of low levels of MNX in plants is indicative of the possibility that RDX might be metabolized via the earlier mentioned three phases of detoxification. Several other studies have also suggested partial or complete mineralization of RDX (Bhadra et al. 2001; Just and Schnoor 2004; Van et al. 2004).

Several genes induced by RDX treatment suggest RDX detoxification via the three phases (Table 2.5 & Table 2.6). The potential phase I (transformation) genes involved in RDX metabolism from both microarray platforms included cytochrome P450s, esterases, and oxidoreductases, while the putative phase II (conjugation) genes included UDP-glucosyl transferases, transferase family proteins, and amino transferases. Phase III (compartmentation) up-regulated genes consisted of ABC transporters, sugar transporters, mannitol transporters and MATE transporters.

Lipoxygenase and RDX metabolism

Lipoxygenase (*LOX2*; *At3g45140*) was strongly up-regulated by RDX as shown by both microarray platforms and confirmed by real-time RT-PCR. The *LOX2* gene product is targeted to chloroplasts and is required for wound induced accumulation of plant growth regulator jasmonic acid (Bell et al. 1995). Lipoxygenases are nonheme iron containing fatty acid dioxygenases ubiquitously present in plants, fungi, and animals (Brash 1999; Feussner and Wasternack 2002). Lipoxygenases catalyze dioxygenation of

polyunsaturated fatty acids (PUFAs), but they are also known to oxidize substrates other than fatty acids (Feussner and Wasternack 2002; Gardner 1996). The common substrates for plant lipoxygenases are linoleate and linolenate, whereas the animal lipoxygenases prefer arachidonate (Brash 1999). Lipoxygenases are known to be versatile catalysts as they can act as a dioxygenase, hydroperoxidase, or leukotriene synthase (Feussner and Wasternack 2002).

Kulkarni (2001) suggested that another role of mammalian lipoxygenases is the metabolism of xenobiotics and endobiotics. Lipoxygenases oxidize some xenobiotics by coupling the formation of lipid hydroperoxide with oxidation (co-oxidation activity), in which the xenobiotic is the co-substrate and the oxidants required are also made by lipoxygenases themselves (Kulkarni 2001). Interestingly, co-oxidation activity has been investigated using lipoxygenase from soybean (Kulkarni and Cook 1988). Lipoxygenases are also known to catalyze glutathione conjugation of some xenobiotics (Kulkarni 2001) making them candidate genes in the RDX detoxification. To our knowledge, this possible generalized role of xenobiotic metabolism by lipoxygenases has not been reported in living plants. Indeed, the plant lipoxygenase pathway is similar to the animal arachidonate pathway in many ways; plant lipoxygenases are capable of catalyzing synthesis of compounds such as leukotrienes and lipoxins that are of mammalian origin (Gardner 1991). The well-characterized soybean lipoxygenase (L-1 form) has been used as a model in several different xenobiotic studies (Kulkarni 2001). Purified soybean lipoxygenase has been shown to metabolize xenobiotics such as thiobenzamide ex vivo (Naidu and Kulkarni 1991).

RDX metabolism in plants is not well understood, however there is considerable progress in understanding microbial degradation of RDX and several mechanisms such as reduction, denitration, hydrolysis, oxidation have been proposed (Hawari 2000). Qasim et al. (2005) suggested that RDX might undergo degradation through hydroxylation, reductive mechanisms and free radical oxidation reactions. Lipoxygenases are known to catalyze both hydroxylation, and free radical reactions (Kulkarni 2001; Gardner 1991) thereby supporting the possibility that *LOX2* might be involved in RDX metabolism.

Correlation between microarray platforms and interpretation of data:

The moderate correlation between platforms can be attributed to several factors including array design, RNA amplification, labeling (single vs. double), hybridization, array scanning, image processing, and normalization techniques (Pylatuik and Fobert 2005). In other studies, Tan et al. (2003) found that considerable differences existed across three commercially available platforms (Agilent, Amersham, and Affymetrix). Additionally, Rogojina et al. (2003) reported lack of agreement between Atlas nucleotide arrays (Clontech) and Affymetrix arrays. However, there are also studies claiming reproducibility between platforms (Larkin et al. 2005; Petersen et al. 2005; Shi et al. 2006).

Despite a moderate correlation between the two platforms in this study there were several significant genes that were commonly up-regulated and down-regulated between the two platforms (Table 2.8). In no case did one platform indicate a gene was up-regulated while the other indicated down-regulation as evidence by the lack of data points in the upper left and lower right corners of Figure 2.5. Therefore, we conclude that either

platform is adequate for searching for gene candidate, while the two-color platform might be considered to be somewhat less robust.

TNT microarrays

Plants readily take up TNT and is generally accumulated in roots (Burken et al. 2000; Harvey et al. 1990; Hughes et al. 1997; Larson et al. 1999). Two other plant genome level gene expression TNT studies have been reported. The Mentewab et al. (2005) study using only half of the Arabidopsis genome found that a total of 52 genes were up-regulated and 47 genes were down-regulated (decreased expression) when Arabidopsis was exposed to TNT concentrations of 1 μM and 10 μM. A large fraction of these genes had predicted roles in cellular detoxification and defense. Although consistent with the three phase detoxification system, they found genes such as UDP-glucose glucosyltrasferase isoforms that could potentially be involved in the transformation phase. The Ekman et al. (2003) SAGE study also was consistent with the three phase detoxification system and revealed up-regulated genes such as monodehydroascorbate reductase, glutathione (GSH)-dependent dehydroascorbate reductase, and glutathione *S*-transferase (GST). A number of cytochrome P-450s, enzymes involved in detoxification of xenobiotic compounds, were also up-regulated by TNT. In another report, Mezzari et al. (2005) focused on the TNT-induced expression of only five genes from Arabidopsis that included glutathione S-trasferases (GSTs) and 12-oxophytodienoate reductases (OPRs). These researchers performed confocal microscopy on Arabidopsis root cells showing that GST-catalyzed GSH (reduced glutathione) conjugation did not occur for RDX or TNT, thereby refuting the proposed glutathione conjugation of RDX or TNT. In our study neither the OPRs nor the GSTs were

significantly differentially regulated. One reason for this could be that both these above mentioned studies analyzed plants after short term exposure to TNT, whereas our experimental design was similar to Mentewab et al. (2005); transcriptional responses to long-term exposure to explosives, which is the most plausible scenario for phytoremediation biology. Nevertheless, consistent with the suggested three phase xenobiotic detoxification system in plants, our experiment also revealed enzymes that could potentially be involved in TNT metabolism (Table 2.7). Cytochrome P-450s induced in this study may suggest their putative involvement in the transformation phase of TNT metabolism. There were six UDP-glucosyl transferase family proteins that were up-regulated suggesting their potential involvement in the conjugation phase. Higher expression for genes encoding an ABC transporter and a transporter-related protein suggest putative candidates for compartmentation of TNT conjugates and/or TNT breakdown products.

Transcriptional response differences between RDX and TNT

Arabidopsis had apparent differences in transcriptional regulation from RDX and TNT treatments. Few significant genes were commonly up-regulated or down-regulated among RDX and TNT-treated plants suggesting that plants cope with these compounds differently. This lack of overlap was also observed by Ekman et al. (2005) who studied the transcriptional responses to RDX in Arabidopsis roots and compared it to transcriptional responses to TNT in Arabidopsis roots studied earlier by Ekman et al. (2003). Common phytoremediation or phytosensing strategies between the explosives are likely not feasible. TNT and RDX are often used together in landmines;

phytoremediation would require consideration of both compounds, but phytosensing for landmine detection might be accomplished by detection of either TNT or RDX.

Meta-analysis and expression profile under general stress conditions

RDX and TNT are nitrogenous compounds (Hannink et al. 2002) and the genes up-regulated by these compounds could be a general response to nitrogen. To ensure a gene list specific in response to RDX and TNT, a meta-analysis was performed by comparing the up-regulated genes in our experiments to up-regulated gene lists from Arabidopsis microarray experiments involving nitrate treatments. After the comparison, only one gene form the RDX list was filtered out, indicating that our candidate gene list for phytoremediation or phytosensing applications is potentially unique to compounds RDX and TNT.

An expression profile analysis of the top 20 up-regulated genes from all the three microarray experiments using the reference expression database Genevestigator (Zimmermann et al. 2004; Zimmermann et al. 2005) was also performed. These expression profiles for the up-regulated genes will allow us to further determine if any of the up-regulation is a general stress response, thereby identifying genes highly specific to RDX and TNT. These specific genes and the promoter elements can be potentially used for phytoremediation and phytosensing applications.

Materials and Methods

Plants and phytotoxicity studies

Arabidopsis thaliana (ecotype Columbia) plants were grown on MS-medium supplemented with B5 vitamins, 1% sucrose, and 2% gelrite, pH 5.8. Arabidopsis seeds were sterilized using 20% bleach, and 0.1% Tween-20. Surface sterilized seeds (around 500 seeds per petri plate) were uniformly plated on solid MS medium containing RDX concentrations of 0, 0.1, 0.25, 0.50, 0.75, 1.00 mM and TNT concentrations of 0, 1, 2, 3, 4, 5, 6, 8, 10 μM. The control plates contained MS medium with DMSO (dimethyl sulfoxide) in proportion to the concentration of RDX or TNT, since DMSO was used as a solvent for both RDX and TNT. The seeds were then cold stratified at 4°C for 3-5 days and then transferred to a growth chamber at 25°C with a photoperiod of 16 h. The growth responses and phytotoxicity tolerance threshold of wild-type Arabidopsis plants to RDX and TNT were analyzed by measuring the primary root length 6-7 days after germination when grown on vertically-oriented plates. RDX was obtained from Restek Corporation, Bellefonte, PA, USA and TNT from Chem service, Inc., West Chester, PA, USA.

RNA preparation

Total RNA was extracted from Arabidopsis seedlings germinated and grown for 8-9 d on MS media containing RDX and TNT using TRI REAGENT® (Molecular Research Center, Inc., Cincinnati, OH, USA) according to manufacturer's protocol. The total RNA isolated was purified using RNeasy ® Midi kit (Qiagen Inc., Valencia, CA, USA) and then used to extract mRNA using Oligotex mRNA mini kit (Qiagen Inc., Valencia, CA,

USA) according to manufacturer's protocol. The mRNA isolated was used for synthesis and labeling of cDNA probes using Superscript™ Plus Direct/Indirect cDNA labeling system with Alexa Fluor® dyes. The labeled cDNA probes were used for hybridizing to microarray slides.

Microarray hybridizations

Two-color microarrays

Two-color hybridization was done only for the RDX treated plants. The experiment included three biological replicates and a dye swap technical replicate (to avoid dye bias) for every set of replicates. Arabidopsis oligonucleotide microarrays spotted with Qiagen-Operon Arabidopsis Genome Array Ready Oligo Set (AROS) Version 3.0 were obtained from D. Galbraith (University of Arizona, http://www.ag.arizona.edu/microarray/). Slides were prepared and hybridized according to the instructions on the supplier's webpage (http://ag.arizona.edu/microarray/Microarraymethod1.doc). After hybridization, the slides were immediately scanned using GenePix® 4000B microarray scanner (Axon Instruments, Union City, CA, USA) and analyzed using GenePix® Pro 4.1 microarray image acquisition and analysis software for quantification of oligonucleotide spot intensities. The microarray ratio data obtained from GenePix® Pro 4.1 software were further subjected to Loess normalization and $\log_2$ transformation without background subtraction using SAS software (SAS Institute Inc., Cary, NC).

Affymetrix microarrays

This experiment involved hybridization of four slides consisting of two biological replicates. Total RNA from the same biological samples that were used for two color hybridizations were used to prepare labeled cRNA. Labeled cRNA targets were prepared

according to the instructions for Arabidopsis ATH1 genome array (Affymetrix, Santa Clara, CA, USA). Labeled cRNA that was purified and fragmented was hybridized to Arabidopsis ATH1 genome array at 45°C for 16 hrs at a setting of 60 rpm. The gene chips were further washed and stained using an Affymetrix Fluidics 450 wash station, following which the gene chips were immediately scanned with a GeneChip 7G scanner. The gene chips were processed at the University of Tennessee, Knoxville Affymetrix Core Facility. Raw CEL files were created from the DAT image file of the chip using the gene chip operating software from Affymetrix. Array Assist Software (version 3.4.2152.32776; Stratagene, La Jolla, CA) was used and the GC-RMA algorithm was applied to the CEL files for background subtraction and normalization. The GC-RMA values were then $\log_2$ transformed.

Statistical analysis of microarray data

The normalized and $\log_2$ transformed data from both two-color and Affymetrix microarrays were statistically analyzed using rank-product statistics as described by Breitling et al. (2004). Bioconductor RankProd package was used to perform the rank product analysis (Hong et al. 2006; Gentleman et al. 2004). The false discovery rate (FDR) value obtained was based on 10,000 random permutations. Since 10,000 random permutations was very computer intensive, 1000 random permutations were performed 10 different times each time starting with a different random seed number and the average FDR value thus calculated was used for further analysis. The genes that had FDR values less than or equal to 0.10 were considered as differentially expressed.

Real-time RT-PCR

Quantitative real-time RT-PCR analysis was done for six genes. This experiment was performed in three replicates and the RNA samples used were prepared as described before. RT-PCR was performed using the Superscript III mix (Invitrogen, Carlsbad, CA, USA) and SYBR Green mix (Qiagen Inc., Valencia, CA, USA) on an ABI 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Primers were designed using Primer Express v. 2.0.0 (Applied Biosystems, Foster City, CA) and are as follows: Lipoxygenase (*At3g45140*) forward, 5'-CTGACCAGCGGATTACGGTAGA-3' and reverse 5'- CCCGCCGGGTAATTTAAGCT-3'; leucine-rich repeat family protein (*At4g33970*) forward, 5'-TTGCCAGTTGCCTAATTTGGTG-3' and reverse, 5'-ACGCAATCTCCTTGCGACTACC-3'; expressed protein (*At4g35720*) forward, 5'-GGGAAGCTCGTTGTGATGATGA-3' and reverse, 5'-TTCCATGGCTGCCTCTACACC-3'; pseudo-response regulator 9 (APRR9) (*At2g46790*) forward, 5'-TGTATGCTGAGAGGTGCTGCTG and reverse, 5'-TCACGCAAAGTCAGTCTTCTCCA-3'; myb related transcription factor (CCA1) (*At2g46830*) forward, 5'-CACGGGAAGAGGGAAGTCAGAAT and reverse, 5'-TGAGCTCCCCAATGGCACTAG-3'; DNA topoisomerase-related (*At3g15950*) forward, 5'-GCCTGCAGATGGTGTATGTGGT and reverse, 5'-GATGTGGTGAGCCGAGAGGTC-3'. The Arabidopsis *β-Actin*-7 (*At5g09810*) was used as the reference gene and the primer sequences (forward – AGTGGTCGTACAACCGGTATTGT; reverse – GAGGAAGAGCATTCCCCTCGTA) for this gene were taken from Campbell et al. (2003). The amplification conditions for the RT-PCR were as follows: enzyme activation at 55°C for 2 minutes, which was

followed by denaturation step at 95°C for 15 minutes and then 40 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 30 s. A negative control without reverse transcriptase was included for all the reactions to ensure that there was no genomic DNA contamination. The PCR products were confirmed for size and sequence by agarose gel electrophoresis and by sequencing the PCR products respectively. Data were analyzed according to Pfaffl (2001). The relative expression ratio was calculated using the formula: Ratio = $(E_{target})^{\Delta CP_{target}(control - sample)} / (E_{ref})^{\Delta CP_{ref}(control - sample)}$, where E_{target} is the real-time PCR efficiency of a reference target gene transcript, E_{ref} is the real-time PCR efficiency of a reference gene transcript, ΔCP_{target} is the difference between crossing points (CP) deviation of control and sample of the target gene transcript and ΔCP_{ref} is the difference between CP deviation of control and sample of the reference gene transcript.

Gene ontology

Gene ontology annotations available on the Arabidopsis information resource (TAIR) website (http://www.arabidopsis.org/tools/bulk/go/index.jsp) were used to functionally characterize the differentially regulated genes (Rhee et al. 2003). The GO tool available on the TAIR website was used to draw the functional categorization pie charts.

Meta-analysis to identify unique genes

Databases such as Gene Expression Omnibus, ArrayExpress, and the Stanford Microarray Database were searched for microarray experiments involving nitrogenous compounds in Arabidopsis. Three relevant datasets were found of which two were from Gene Expression Omnibus database (GSE6824 and GSE9148) and one from ArrayExpress database (E-MEXP-828). These datasets were downloaded and were also

analyzed using rank-product statistics similar to the analysis of our datasets and the gene list from our experiment was compared with these five datasets for any redundant genes.

Expression profile analysis using Genevestigator

The expression profile for the top 20 up-regulated genes from all the three microarray experiments under different general stress conditions were studied using Genevestigator, a reference expression database and a meta-analysis system (Zimmermann et al., 2004; Zimmermann et al., 2005; https://www.genevestigator.ethz.ch/gv/index.jsp). The array list used was selected based on the annotation and all the stress related datasets available were selected for generating the expression profile.

References

Baumberger, N., Doesseger, B., Guyot, R., Diet, A., Parsons, R.L., Clark, M.A., Simmons, M.P., Bedinger, P., Goff, S.A., Ringli, C. and Keller, B. (2003) Whole-genome comparison of leucine-rich repeat extensins in Arabidopsis and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade. Plant Physiol. 131, 1313-1326.

Bell, E., Creelman, R.A. and Mullet, J.E. (1995) A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in Arabidopsis. Proc. Natl. Acad. Sci. USA. 92, 8675-8679.

Best, E.P.H., Zappi, M.E., Fredrickson, H.L., Sprecher, S.L., Larson, S.L. and Ochman, M. (1997) Screening of aquatic and wetland plant species for phytoremediation of explosives-contaminated groundwater from the Iowa Army Ammunition Plant. Ann. N. Y. Acad. Sci. 829, 179-194.

Best, E.P.H., Sprecher, S.L., Larson, S.L., Fredrickson, H.L. and Bader, D.F. (1999) Environmental behavior of explosives in groundwater from the Milan Army Ammunition Plant in aquatic and wetland plant treatments. Uptake and fate of TNT and RDX in plants. Chemosphere 39, 2057-2072.

Best, E.P.H., Miller, J.L. and Larson, S.L. (2001) Tolerance towards explosives, and explosives removal from groundwater in treatment wetland mesocosms. Water Sci. Technol. 44, 515-521.

Best, E.P.H., Kvesitadze, G., Khatisashvili, G. and Sadunishvili, T. (2005) Plant processes important for the transformation and degradation of explosives contaminants. Z. Naturforsch., C, J. Biosci. 60, 340-348.

Best, E.P.H., Geter, K.N., Tatem, H.E. and Lane, B.K. (2006) Effects, transfer, and fate of RDX from aged soil in plants and worms. Chemosphere 62, 616-625.

Bhadra, R., Wayment, D.G., Williams, R.K., Barman, S.N., Stone, M.B., Hughes, J.B. and Shanks, J.V. (2001) Studies on plant-mediated fate of the explosives RDX and HMX. Chemosphere 44, 1259-1264.

Brash, A.R. (1999) Lipoxygenases: occurrence, functions, catalysis, and acquisition of substrate. J. Biol. Chem. 274, 23679-23682.

Breitling, R., Armengaud, P., Amtmann, A. and Herzyk, P. (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Letters 573, 83-92.

Burken, J.G., Shanks, J.V., Thompson, P.L. (2000) Phytoremediation and plant metabolism of explosives and nitroaromatic compounds. In: *Biodegradation of Nitroaromatic Compounds and Explosives*, Spain J.C., Hughes J. B., Knackmuss H. (eds.), CRC Press, Boca Raton, Florida, pp. 239-276.

Campbell, E.J., Schenk, P.M., Kazan, K., Penninckx, I.A., Anderson, J.P., Maclean, D.J., Cammue, B.P., Ebert, P.R. and Manners, J.M. (2003) Pathogen-responsive expression of a putative ATP-binding cassette transporter gene conferring resistance to the diterpenoid sclareol is regulated by multiple defense signaling pathways in Arabidopsis. Plant Physiol. 133, 1272-1284.

Coleman, J., Blake-Kalff, M. and Davies, E. (1997) Detoxification of xenobiotics by plants: chemical modification and vacuolar compartmentation. Trends Plant Sci. 2, 144-151.

Ekman, D.R., Lorenz, W.W., Przybyla, A.E., Wolfe, N.L. and Dean, J.F.D. (2003) SAGE analysis of transcriptome responses in Arabidopsis roots exposed to 2,4,6-trinitrotoluene. Plant Physiol. 133, 1397-1406.

Ekman, D.R., Wolfe, N.L. and Dean, J.F.D. (2005) Gene expression changes in *Arabidopsis thaliana* seedling roots exposed to the munition hexahydro-1,3,5-trinitro-1,3,5-triazine. Environ. Sci. Technol. 39, 6313-6320.

Feussner, I. and Wasternack, C. (2002) The lipoxygenase pathway. Annu. Rev. Plant Biol. 53, 275-297.

Gardner, H.W. (1991) Recent investigations into the lipoxygenase pathway of plants. Biochim. Biophys. Acta 1084, 221-239.

Gardner, H.W. (1996) Lipoxygenase as a versatile biocatalyst. J. Am. Oil Chem. Soc. 73, 1347-1357.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y.C., 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.H. and Zhang, J.H. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Haderlein, S.B., Weissmahr, K.W. and Schwarzenbach, R.P. (1996) Specific adsorption of nitroaromatic explosives and pesticides to clay minerals. Environ. Sci. Technol. 30, 612-622.

Halasz, A., Groom, C., Zhou, E., Paquet, L., Beaulieu, C., Deschamps, S., Corriveau, A., Thiboutot, S., Ampleman, G., Dubois, C. and Hawari, J. (2002) Detection of explosives and their degradation products in soil environments. J. Chromatogr. A 963, 411-418.

Hannink, N.K., Rosser, S.J. and Bruce, N.C. (2002) Phytoremediation of explosives. Crit. Rev. Plant Sci. 21, 511-538.

Harvey, S.D., Fellows, R.J., Cataldo, D.A. and Bean, R.M. (1990) Analysis of 2,4,6-trinitrotoluene and its transformation products in soils and plant-tissues by high-performance liquid-chromatography. J. Chromatogr. 518, 361-374.

Harvey, S.D., Fellows, R.J., Cataldo, D.A. and Bean, R.M. (1991) Fate of the explosive hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) in soil and bioaccumulation in bush bean hydroponic plants. Environ. Toxicol. Chem. 10, 845-855.

Hawari, J. (2000) Biodegradation of RDX and HMX: from basic research to field application. In: *Biodegradation of Nitroaromatic Compounds and Explosives*, Spain J.C., Hughes J.B., Knackmuss H. (eds.), CRC Press, Boca Raton, Florida, 277-310.

Hong, F.X., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L. and Chory, J. (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22, 2825-2827.

Hughes, J.B., Shanks, J., Vanderford, M., Lauritzen, J. and Bhadra, R. (1997) Transformation of TNT by aquatic plants and plant tissue cultures. Environ. Sci. Technol. 31, 266-271.

Ishikawa, T. (1992) The ATP-dependent gutathione s-conjugate export pump. Trends Biochem. Sci. 17:463-469.

Ishikawa, T., Li, Z.S., Lu, Y.P. and Rea, P.A. (1997) The GS-X pump in plant, yeast, and animal cells: Structure, function, and gene expression. Biosci. Rep. 17:189-207.

Just, C.L. and Schnoor, J.L. (2004) Phytophotolysis of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) in leaves of reed canary grass. Environ. Sci. Technol. 38, 290-295.

Kobe, B. and Deisenhofer, J. (1994) The leucine-rich repeat: a versatile binding motif. Trends Biochem. Sci. 19, 415-421.

Kulkarni, A.P. and Cook, D.C. (1988) Hydroperoxidase activity of lipoxygenase: a potential pathway for xenobiotic metabolism in the presence of linoleic acid. Res. Commun. Chem. Pathol. Pharmacol. 61, 305-314.

Kulkarni, A.P. (2001) Lipoxygenase - a versatile biocatalyst for biotransformation of endobiotics and xenobiotics. Cell. Mol. Life Sci. 58, 1805-1825.

Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R. and Quackenbush, J. (2005) Independence and reproducibility across microarray platforms. Nat Meth 2, 337-344.

Larson, S.L., Jones, R.P., Escalon, L. and Parker, D. (1999) Classification of explosives transformation products in plant tissue. Environ. Toxicol. Chem. 18, 1270-1276.

Li, J. and Chory, J. (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90, 929-938.

Lucero, M.E., Mueller, W., Hubstenberger, J., Phillips, G.C. and O'Connell, M.A. (1999) Tolerance to nitrogenous explosives and metabolism of TNT by cell suspensions of *Datura innoxia*. In Vitro Cell. Dev. Biol., Plant 35, 480-486.

Mentewab, A., Cardoza, V. and Stewart, C.N. (2005) Genomic analysis of the response of *Arabidopsis thaliana* to trinitrotoluene as revealed by cDNA microarrays. Plant Sci. 168, 1409-1424.

Mezzari, M.P., Walters, K., Jelinkova, M., Shih, M.C., Just, C.L. and Schnoor, J.L. (2005) Gene expression and microscopic analysis of Arabidopsis exposed to chloroacetanilide herbicides and explosive compounds. A phytoremediation approach. Plant Physiol. 138, 858-869.

Naidu, A.K. and Kulkarni, A.P. (1991) Role of lipoxygenase in xenobiotic metabolism - sulfoxidation of thiobenzamide by purified soybean lipoxygenase. Res. Commun. Chem. Pathol. Pharmacol. 71, 175-188.

Nishino, S.F., Spain, J.C., He, Z. (2000) Strategies for aerobic degradation of nitroaromatic compounds by bacteria: process discovery to field application. In: *Biodegradation of Nitroaromatic Compounds and Explosives*, Spain J.C., Hughes J.B., Knackmuss H. (eds.), CRC Press, Boca Raton, Florida, pp. 7-61.

Patel, N., Cardoza, V., Christensen, E., Rekapalli, L., Ayalew, M. and Stewart, C.N. (2004) Differential gene expression of *Chlamydomonas reinhardtii* in response to 2,4,6-trinitrotoluene (TNT) using microarray analysis. Plant Sci. 167, 1109-1122.

Petersen, D., Chandramouli, G.V.R., Geoghegan, J., Hilburn, J., Paarlberg, J., Kim, C., Munroe, D., Gangi, L., Han, J., Puri, R., Staudt, L., Weinstein, J., Barrett, J.C., Green, J. and Kawasaki, E. (2005) Three microarray platforms: an analysis of their concordance in profiling gene expression. BMC Genomics 6, 63.

Peterson, M.M., Horst, G.L., Shea, P.J. and Comfort, S.D. (1998) Germination and seedling development of switchgrass and smooth bromegrass exposed to 2,4,6-trinitrotoluene. Environ. Pollut. 99, 53-59.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Purugganan, M.M., Braam, J. and Fry, S.C. (1997) The Arabidopsis TCH4 xyloglucan endotransglycosylase (substrate specificity, pH optimum, and cold tolerance). Plant Physiol. 115, 181-190.

Pylatuik, J.D. and Fobert, P.R. (2005) Comparison of transcript profiling on Arabidopsis microarray platform technologies. Plant Mol. Biol. 58, 609-624.

Qasim, M., Fredrickson, H., McGrath, C., Furey, J. and Bajpai, R. (2005) Theoretical predictions of chemical degradation reaction mechanisms of RDX and other cyclic nitramines derived from their molecular structures. SAR QSAR Environ. Res. 16, 203-218.

Rea, P.A., Li, Z.S., Lu, Y.P., Drozdowicz, Y.M. and Martinoia, E. (1998) From vacuolar GS-X pumps to multispecific ABC transporters. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:727-760.

Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D.C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res. 31, 224-228.

Rogojina, A.T., Orr, W.E., Song, B.K. and Geisert, E.E., Jr. (2003) Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. Mol. Vis. 9, 482-496.

Rosenblatt, D.H., Burrows, S.P., Mitchell, W.R., Parmer, D.L. (1991) Organic explosives and related compounds. In: *The Handbook of Environmental Chemistry*, Hutzinger O. (ed.), Springer-Verlag, New York.

Sandermann, H., Jr. (1992) Plant metabolism of xenobiotics. Trends Biochem. Sci. 17, 82-84.

Schaffner, A., Messner, B., Langebartels, C. and Sandermann, H. (2002) Genes and enzymes for in-planta phytoremediation of air, water and soil. Acta Biotechnol. 22, 141-151.

Shi, L.M., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S., Lee, K.Y., Luo, Y.L., Sun, Y.M.A., Willey, J.C., Setterquist, R.A., Fischer, G.M., Tong, W.D., Dragan, Y.P., Dix, D.J., Frueh, F.W., Goodsaid, F.M., Herman, D., Jensen, R.V., Johnson, C.D., Lobenhofer, E.K., Puri, R.K., Scherf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P.K., Zhang, L., Amur, S., Bao, W.J., Barbacioru, C.C., Lucas, A.B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X.X.M., Cebula, T.A., Chen, J.J., Cheng, J., Chu, T.M., Chudin, E., Corson, J., Corton, J.C., Croner, L.J., Davies, C., Davison, T.S., Delenstarr, G., Deng, X.T., Dorris, D., Eklund, A.C., Fan, X.H., Fang, H., Fulmer-Smentek, S., Fuscoe, J.C., Gallagher, K., Ge, W.G., Guo, L., Guo, X., Hager, J., Haje, P.K., Han, J., Han, T., Harbottle, H.C., Harris, S.C., Hatchwell, E., Hauser, C.A., Hester, S., Hong, H.X., Hurban, P., Jackson, S.A., Ji, H.L., Knight, C.R., Kuo, W.P., LeClerc, J.E., Levy, S., Li, Q.Z., Liu, C.M., Liu, Y., Lombardi, M.J., Ma, Y.Q., Magnuson, S.R.,

Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B.T., Novoradovskaya, N., Orr, M.S., Osborn, T.W., Papallo, A., Patterson, T.A., Perkins, R.G., Peters, E.H., Peterson, R., Philips, K.L., Pine, P.S., Pusztai, L., Qian, F., Ren, H.Z., Rosen, M., Rosenzweig, B.A., Samaha, R.R., Schena, M., Schroth, G.P., Shchegrova, S., Smith, D.D., Staedtler, F., Su, Z.Q., Sun, H.M., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K.L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S.J., Wang, S.J., Wang, Y.H., Wolfinger, R., Wong, A., Wu, J., Xiao, C.L., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y.P. and Slikker, W. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat. Biotechnol. 24, 1151-1161.

Talmage, S.S., Opresko, D.M., Maxwell, C.J., Welsh, C.J., Cretella, F.M., Reno, P.H. and Daniel, F.B. (1999) Nitroaromatic munition compounds: environmental effects and screening values. Rev. Environ. Contam Toxicol. 161, 1-156.

Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M. and Cam, M.C. (2003) Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 31, 5676-5684.

Tanaka, S., Brentner, L.B., Merchie, K.M., Schnoor, J.L., Jong, M.Y. and Van Aken, B. (2007) Analysis of gene expression in poplar trees (*Populus deltoides* x *nigra*, DN34) exposed to the toxic explosive Hexahydro-1,3,5-Trinitro-1,3,5-Triazine (RDX). Int. J. Phytoremediation 9, 15-30.

Thompson, P.L., Ramer, L.A. and Schnoor, J.L. (1999) Hexahydro-1,3,5-trinitro-1,3,5-triazine translocation in poplar trees. Environ. Toxicol. Chem. 18, 279-284.

Van, A.B., Yoon, J.M., Just, C.L. and Schnoor, J.L. (2004) Metabolism and mineralization of hexahydro-1,3,5-trinitro-1,3,5-triazine inside poplar tissues (*Populus deltoides* x *nigra* DN-34). Environ. Sci. Technol. 38, 4572-4579.

Van Dillewijn, P., Caballero, A., Paz, J.A., Gonzalez-Perez, M.M., Oliva, J.M. and Ramos, J.L. (2007) Bioremediation of 2,4,6-trinitrotoluene under field conditions. Environ. Sci. Technol. 41, 1378-1383.

Vila, M., Lorber-Pascal, S. and Laurent, F. (2007) Fate of RDX and TNT in agronomic plants. Environ. Pollut. 148, 148-154.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol. 136, 2621-2632.

Zimmermann, P., Hennig, L. and Gruissem, W. (2005) Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci. 10, 407-409.

Appendix

Table 2.1. Top 20 up-regulated and down-regulated genes in response to RDX in the two-color microarrays along with their linear fold change (FC), false discovery rate (FDR) and the p-values associated with all the listed genes is <0.01.

Gene ID	Up-regulated genes	FC (linear scale)	FDR
At4g33970	leucine-rich repeat family protein / extensin family protein	1.996	0.000
At4g35720	expressed protein	1.926	0.000
At5g37050	hypothetical protein	1.819	0.000
At4g12500	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1.803	0.000
At4g36670	mannitol transporter, putative	1.782	0.000
At5g50335	expressed protein	1.767	0.001
At5g03545	expressed protein	1.755	0.000
At4g14130	xyloglucan:xyloglucosyl transferase, putative	1.720	0.000
At1g21850	multi-copper oxidase type I family protein	1.719	0.001
At4g08270	hypothetical protein	1.717	0.002
At5g38940	germin-like protein, putative	1.710	0.000
At1g37080	hypothetical protein	1.709	0.001
At3g50330	basic helix-loop-helix (bHLH) family protein	1.706	0.001
At3g09922	hypothetical protein	1.701	0.001
At5g15600	expressed protein	1.692	0.002
At1g15825	hydroxyproline-rich glycoprotein family protein	1.691	0.005
At1g23050	hydroxyproline-rich glycoprotein family protein	1.686	0.002
At2g34790	FAD-binding domain-containing protein	1.677	0.001
At4g17980	no apical meristem (NAM) family protein	1.671	0.002
At2g12610	expressed protein	1.670	0.002
Gene ID	**Down-regulated genes**	**FC (linear scale)**	**FDR**
At4g11320	cysteine proteinase, putative	0.428	0.000
At1g73330	protease inhibitor, putative	0.475	0.000
At3g15950	DNA topoisomerase-related	0.506	0.000
At4g14060	major latex protein-related	0.518	0.000
At2g01520	major latex protein-related	0.525	0.000
At3g05900	neurofilament protein-related	0.533	0.000
At4g11310	cysteine proteinase, putative	0.540	0.000
At1g32060	phosphoribulokinase (PRK) / phosphopentokinase	0.541	0.000
At3g25830	myrcene/ocimene synthase, putative	0.562	0.000
At3g25820	myrcene/ocimene synthase, putative	0.563	0.000
At5g64840	ABC transporter family protein	0.578	0.000
At2g46830	myb-related transcription factor	0.590	0.000
At1g58848	disease resistance protein (CC-NBS-LRR class), putative	0.595	0.000
At2g30520	signal transducer of phototropic response (RPT2)	0.595	0.001
At4g24190	shepherd protein (SHD) / clavata formation protein, putative	0.597	0.000
At5g57560	xyloglucan:xyloglucosyl transferase	0.598	0.000
At3g55800	sedoheptulose-1,7-bisphosphatase, chloroplast	0.607	0.002
At5g45820	CBL-interacting protein kinase 20 (CIPK20)	0.611	0.001
At1g48300	expressed protein	0.613	0.000
At3g54500	expressed protein	0.614	0.000

Table 2.2. Top 20 up-regulated and down-regulated genes in response to RDX in the Affymetrix microarrays along with their linear fold change (FC), false discovery rate (FDR) and the p-values associated with all the listed genes is <0.01.

Gene ID	Up-regulated genes	FC (linear scale)	FDR
AT3G45140	lipoxygenase (*LOX2*)	6.671	0.002
AT4G35720	expressed protein	5.184	0.006
AT2G39350	ABC transporter family protein	4.822	0.004
AT2G47780	rubber elongation factor (REF) protein-related	4.726	0.005
AT5G06570	expressed protein	4.184	0.011
AT5G07010	sulfotransferase family protein	4.174	0.009
AT5G49690	UDP-glucoronosyl/UDP-glucosyl transferase family protein	4.161	0.011
AT5G39580	peroxidase, putative	3.968	0.009
AT3G21720	isocitrate lyase, putative	3.934	0.011
AT5G67480	TAZ zinc finger family protein / BTB/POZ domain-containing protein	3.929	0.010
AT1G80130	expressed protein	3.922	0.009
AT1G03020	glutaredoxin family protein	3.896	0.009
AT3G51400	expressed protein	3.896	0.009
AT1G73220	sugar transporter family protein	3.883	0.015
AT1G17810	major intrinsic family protein / MIP family protein	3.701	0.009
AT5G44440	FAD-binding domain-containing protein	3.663	0.018
AT5G09570	expressed protein	3.655	0.009
AT5G67060	basic helix-loop-helix (bHLH) family protein	3.511	0.015
AT4G23060	calmodulin-binding family protein	3.470	0.016
AT1G50060	pathogenesis-related protein, putative	3.470	0.015
Gene ID	**Down-regulated genes**	**FC (linear scale)**	**FDR**
AT1G13650	expressed protein	0.051	0.000
AT1G28400	expressed protein	0.084	0.001
AT3G05900	neurofilament protein-related	0.090	0.001
AT5G47990	cytochrome P450 family protein	0.092	0.001
AT3G25830	myrcene/ocimene synthase, putative	0.113	0.001
AT4G17090	beta-amylase (CT-BMY) / 1,4-alpha-D-glucan maltohydrolase	0.126	0.004
AT4G08310	expressed protein	0.130	0.003
AT1G17360	COP1-interacting protein-related	0.134	0.003
AT3G16000	matrix-localized MAR DNA-binding protein-related	0.137	0.003
AT2G32240	expressed protein	0.139	0.003
AT4G11280	1-aminocyclopropane-1-carboxylate synthase 6 / ACC synthase 6 (ACS6)	0.153	0.004
AT3G18480	CCAAT displacement protein-related / CDP-related	0.157	0.003
AT4G33750	expressed protein	0.161	0.004
AT4G26260	expressed protein	0.166	0.006
AT2G01520	major latex protein-related / MLP-related	0.177	0.006
AT2G18370	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	0.186	0.008
AT5G09530	hydroxyproline-rich glycoprotein family protein	0.201	0.012
AT3G11450	DNAJ heat shock N-terminal domain-containing protein / cell division protein-related	0.205	0.010
AT1G65010	expressed protein	0.205	0.009
AT3G01420	pathogen-responsive alpha-dioxygenase, putative	0.207	0.010

Table 2.3. Top 20 up-regulated and down-regulated genes in response to TNT in the Affymetrix microarrays along with their linear fold change (FC), false discovery rate (FDR) and the p-values associated with all the listed genes is <0.01.

Gene ID	Up-regulated genes	FC (linear scale)	FDR
AT3G15310/ AT5G32621	expressed protein / expressed protein	17.452	0.001
AT5G42760	O-methyltransferase N-terminus domain-containing protein	14.684	0.000
AT3G21890	zinc finger (B-box type) family protein	9.597	0.002
AT3G44450	expressed protein	8.518	0.003
AT2G19650	DC1 domain-containing protein	8.006	0.005
AT4G33720	pathogenesis-related protein, putative	7.348	0.004
AT2G46790/ AT2G46670	pseudo-response regulator 9 (APRR9) / timing of CAB expression 1-like protein (TL1)	6.868	0.005
AT1G27730	zinc finger (C2H2 type) family protein (ZAT10) / salt-tolerance zinc finger protein (STZ)	6.707	0.058
AT5G54120	expressed protein	6.039	0.005
AT2G15020	expressed protein	5.917	0.005
AT2G27420	cysteine proteinase, putative	5.814	0.007
AT1G01520	myb family transcription factor	5.580	0.006
AT2G44940	AP2 domain-containing transcription factor TINY, putative	4.833	0.011
AT5G54120/ AT5G54130	expressed protein / calcium-binding EF hand family protein	4.822	0.011
AT1G13740	expressed protein	4.771	0.012
AT4G27652	expressed protein	4.636	0.023
AT3G05800	expressed protein	4.570	0.011
AT5G17350	expressed protein	4.427	0.070
AT3G14200	DNAJ heat shock N-terminal domain-containing protein	4.401	0.013
AT5G15950	adenosylmethionine decarboxylase family protein	4.367	0.016

Gene ID	Down-regulated genes	FC (linear scale)	FDR
AT5G48850	male sterility MS5 family protein	0.051	0.001
AT3G48360	speckle-type POZ protein-related	0.054	0.001
AT5G02020	expressed protein	0.070	0.002
AT1G30820	CTP synthase, putative / UTP--ammonia ligase, putative	0.075	0.002
AT1G79700	ovule development protein, putative	0.083	0.002
AT5G22920	zinc finger (C3HC4-type RING finger) family protein	0.086	0.003
AT1G36370	glycine hydroxymethyltransferase, putative / serine hydroxymethyltransferase, putative	0.095	0.002
AT1G13650	expressed protein	0.095	0.002
AT4G26260	expressed protein	0.097	0.002
AT4G33960	expressed protein	0.099	0.002
AT2G18700	glycosyl transferase family 20 protein / trehalose-phosphatase family protein	0.100	0.003
AT3G62950	glutaredoxin family protein	0.100	0.003
AT4G20820	FAD-binding domain-containing protein	0.107	0.003
AT4G24890	calcineurin-like phosphoesterase family protein	0.123	0.006
AT4G36410	ubiquitin-conjugating enzyme 17 (UBC17)	0.127	0.006
AT4G12500	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	0.135	0.007
AT5G64190	expressed protein	0.136	0.006
AT1G15040	glutamine amidotransferase-related	0.138	0.007
AT5G59080	expressed protein	0.138	0.006
AT5G12020	17.6 kDa class II heat shock protein (HSP17.6-CII)	0.139	0.007

Table 2.4. Real-time RT-PCR confirmation of lipoxygenase and other differentially regulated selected genes in the RDX microarray experiments.

		Relative expression ratio		
Gene ID	**Gene name**	**Affymetrix**	**Two-color**	**RT-PCR***
AT3G45140	Lipoxygenase	6.67	1.52	5.02 ± 1.50
AT4G33970	Leucine-rich repeat family protein	1.00	2.00	1.19 ± 0.06
AT4G35720	Expressed protein	5.18	1.93	3.90 ± 0.78
AT2G46790	Pseudo-response regulator 9 (APRR9)	0.44	0.72	0.28 ± 0.05
AT2G46830	Myb-related transcription factor (CCA1)	0.37	0.59	0.33 ± 0.09
AT3G15950	DNA topoisomerase-related	0.27	0.51	0.31 ± 0.09
AT5G09810	Arabidopsis *β-Actin*-7	1.11	0.80	1.00 ± 0.00

*average ratio ± S.E.

Table 2.5. List of potential genes suggesting RDX metabolism via the three phases of detoxification from two-color RDX microarray experiment along with their linear fold change (FC), false discovery rate (FDR) and the p-value associated with the genes.

Potential phase I enzymes (cytochrome p450, esterases, oxido reductases)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
At4g16690	esterase/lipase/thioesterase family protein	1.54	0.01	0.000
At2g48080	oxidoreductase, 2OG-Fe(II) oxygenase family protein	1.50	0.02	0.000
At3g11180	oxidoreductase, 2OG-Fe(II) oxygenase family protein	1.45	0.03	0.000
Potential phase II enzymes (glucosyl transferase, glutathione S-transferase, acyl transferase)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
At3g29590	transferase family protein	1.56	0.005	0.000
At1g78270	UDP-glucose glucosyltransferase, putative	1.48	0.015	0.000
Potential phase III enzymes (transporters)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
At4g36670	mannitol transporter, putative	1.78	0.000	0.000
At1g73220	sugar transporter family protein	1.57	0.002	0.000

Table 2.6. List of potential genes suggesting RDX metabolism via the three phases of detoxification from Affymetrix RDX microarray experiment along with their linear fold change (FC), false discovery rate (FDR) and the p-value associated with the genes.

Potential phase I enzymes (cytochrome p450, esterases, oxido reductases)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
AT5G05600	oxidoreductase, 2OG-Fe(II) oxygenase family protein	2.80	0.032	0.000
AT5G22500	acyl CoA reductase, putative/ male-sterility protein, putative	2.15	0.097	0.000
AT5G09970	cytochrome P450 family protein	2.08	0.101	0.001
AT1G64590	short-chain dehydrogenase/ reductase (SDR) family protein	2.04	0.106	0.001
AT2G12190/ AT1G64950/ AT1G64940/ AT1G64930	cytochrome P450, putative	2.00	0.115	0.001
Potential phase II enzymes (glucosyl transferase, glutathione S-transferase, acyl transferase)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
AT5G49690	UDP-glucoronosyl/UDP-glucosyl transferase family protein	4.16	0.011	0.000
AT5G39050	transferase family protein	2.68	0.037	0.000
AT2G39980	transferase family protein	2.45	0.048	0.000
AT5G01210	transferase family protein	2.15	0.093	0.000
AT3G19710	branched-chain amino acid aminotransferase, putative	2.07	0.098	0.001
Potential phase III enzymes (transporters)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**
AT2G39350	ABC transporter family protein	4.82	0.004	0.000
AT1G73220	sugar transporter family protein	3.88	0.015	0.000
AT3G05400	sugar transporter, putative	2.46	0.047	0.000
AT1G16370	transporter-related	2.18	0.078	0.000
AT4G29140	MATE efflux protein-related	2.09	0.104	0.001

Table 2.7. List of potential genes suggesting TNT metabolism via the three phases of detoxification from Affymetrix TNT microarray experiment along with their linear fold change (FC), false discovery rate (FDR) and the p-value associated with the genes.

Potential phase I enzymes (cytochrome p450, esterases, oxido reductases)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**p-value**
AT3G15650	phospholipase/carboxylesterase family protein	2.18	0.118	0.001
AT3G30180	cytochrome P450, putative	2.18	0.112	0.001
Potential phase II enzymes (glucosyl transferase, glutathione S-transferase, acyl transferase)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**p-value**
AT3G21760	UDP-glucoronosyl/UDP- glucosyl transferase family protein	3.29	0.027	0.000
AT3G21560	UDP-glucosyltransferase, putative	3.14	0.036	0.000
AT5G55380	membrane bound O-acyl transferase (MBOAT) family protein / wax synthase-related	2.84	0.052	0.000
AT3G43190	sucrose synthase, putative/ sucrose-UDP glucosyltransferase, putative	2.47	0.069	0.000
AT2G13290	glycosyl transferase family 17 protein	2.42	0.072	0.000
AT5G17050	UDP-glucoronosyl/UDP-glucosyl transferase family protein	2.39	0.073	0.000
AT2G36800/ AT2G36790	UDP-glucoronosyl/UDP-glucosyl transferase family protein	2.39	0.078	0.000
Potential phase III enzymes (transporters)				
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**p-value**
AT1G79410	transporter-related	3.05	0.042	0.000
AT3G55110	ABC transporter family protein	2.97	0.037	0.000

Table 2.8. Significant genes commonly up-regulated and down-regulated between RDX Affymetrix and RDX two-color along with their linear fold change (FC), false discovery rate (FDR) and associated p-values.

UP-REGULATED GENES							
		TWO-COLOR			**AFFYMETRIX**		
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**	**FC (Linear scale)**	**FDR**	**P-value**
At4g35720	expressed protein	1.93	0.000	0.000	5.18	0.006	0.000
At4g12500	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1.80	0.000	0.000	2.53	0.045	0.000
At5g50335	expressed protein	1.77	0.001	0.000	2.98	0.026	0.000
At4g14130	xyloglucan:xyloglucosyl transferase, putative	1.72	0.000	0.000	2.25	0.066	0.000
At4g12490	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1.65	0.001	0.000	2.24	0.069	0.000
At4g12480	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	1.65	0.001	0.000	2.66	0.037	0.000
At3g45140	lipoxygenase (LOX2)	1.65	0.004	0.000	6.67	0.002	0.000
At5g01210	transferase family protein	1.64	0.001	0.000	2.15	0.093	0.000
At1g73220	sugar transporter family protein	1.57	0.002	0.000	3.88	0.015	0.000
At1g64660	Cys/Met metabolism pyridoxal-phosphate-dependent enzyme family protein	1.56	0.003	0.000	2.26	0.070	0.000
At2g45210	auxin-responsive protein-related	1.53	0.005	0.000	2.80	0.029	0.000
At2g39980	transferase family protein	1.51	0.004	0.000	2.45	0.048	0.000
At5g23020	2-isopropylmalate synthase 2 (IMS2)	1.51	0.005	0.000	2.51	0.038	0.000

Table 2.8, continued

DOWN-REGULATED GENES							
		TWO-COLOR			AFFYMETRIX		
AGI gene ID	**Gene name**	**FC (Linear scale)**	**FDR**	**P-value**	**FC (Linear scale)**	**FDR**	**P-value**
At4g11320	cysteine proteinase, putative	0.43	0.000	0.000	0.28	0.026	0.000
At4g14060	major latex protein-related / MLP-related	0.52	0.000	0.000	0.09	0.001	0.000
At3g25830	myrcene/ocimene synthase, putative	0.56	0.000	0.000	0.32	0.049	0.000
At3g25820	myrcene/ocimene synthase, putative	0.56	0.000	0.000	0.37	0.078	0.000
At3g61060	F-box family protein / lectin-related	0.63	0.004	0.000	0.27	0.025	0.000
At2g38800	calmodulin-binding protein-related	0.63	0.001	0.000	0.39	0.094	0.001
At2g23590	hydrolase, alpha/beta fold family protein	0.64	0.002	0.000	0.20	0.012	0.000
At4g29905	expressed protein	0.65	0.006	0.000	0.39	0.083	0.000

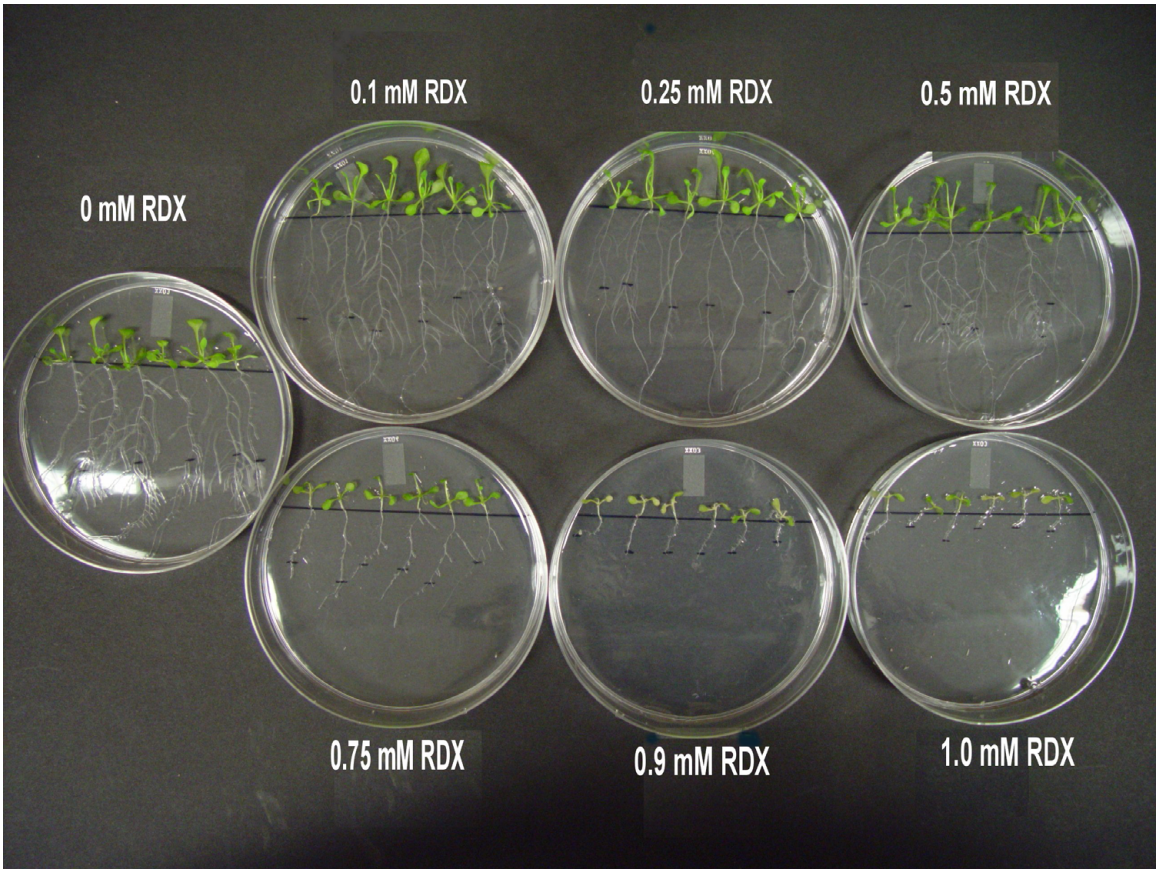

Figure 2.1. *Arabidopsis thaliana* grown on MS medium supplemented with different concentrations of RDX

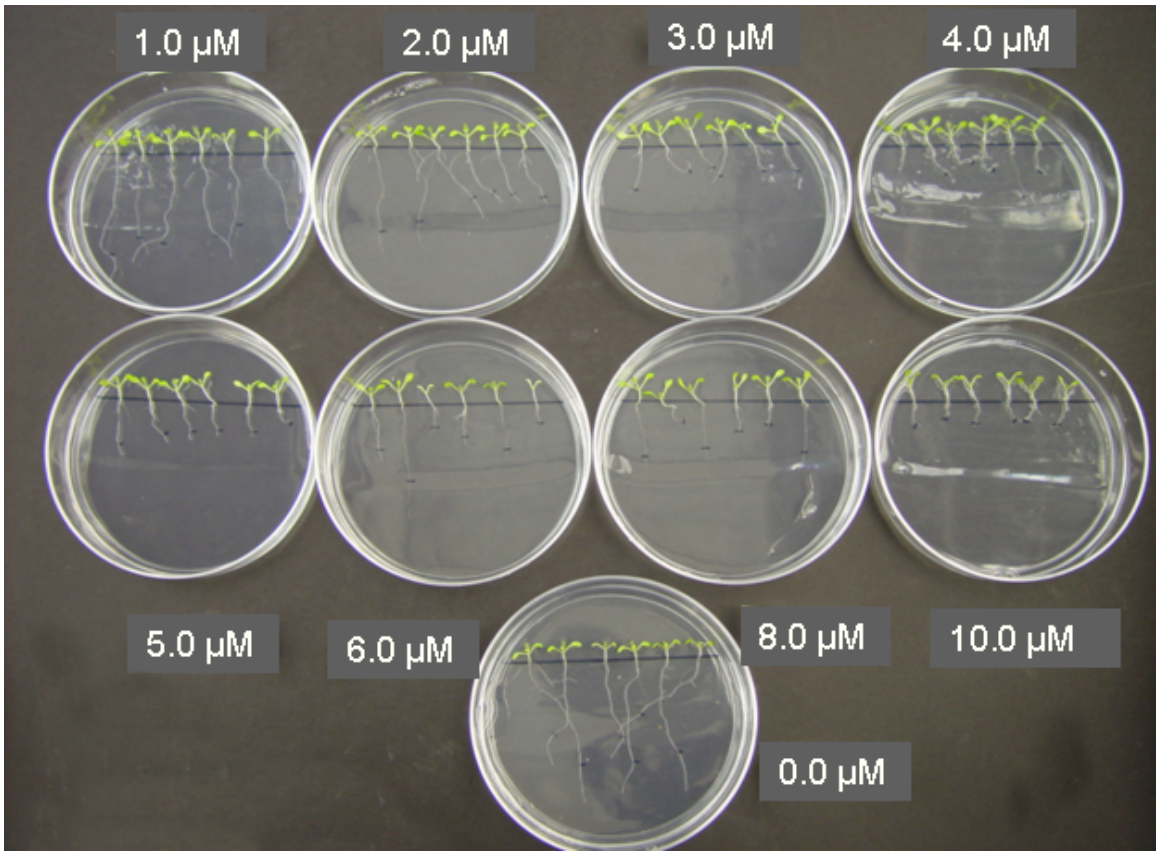

Figure 2.2. *Arabidopsis thaliana* grown on MS medium supplemented with different concentrations of TNT

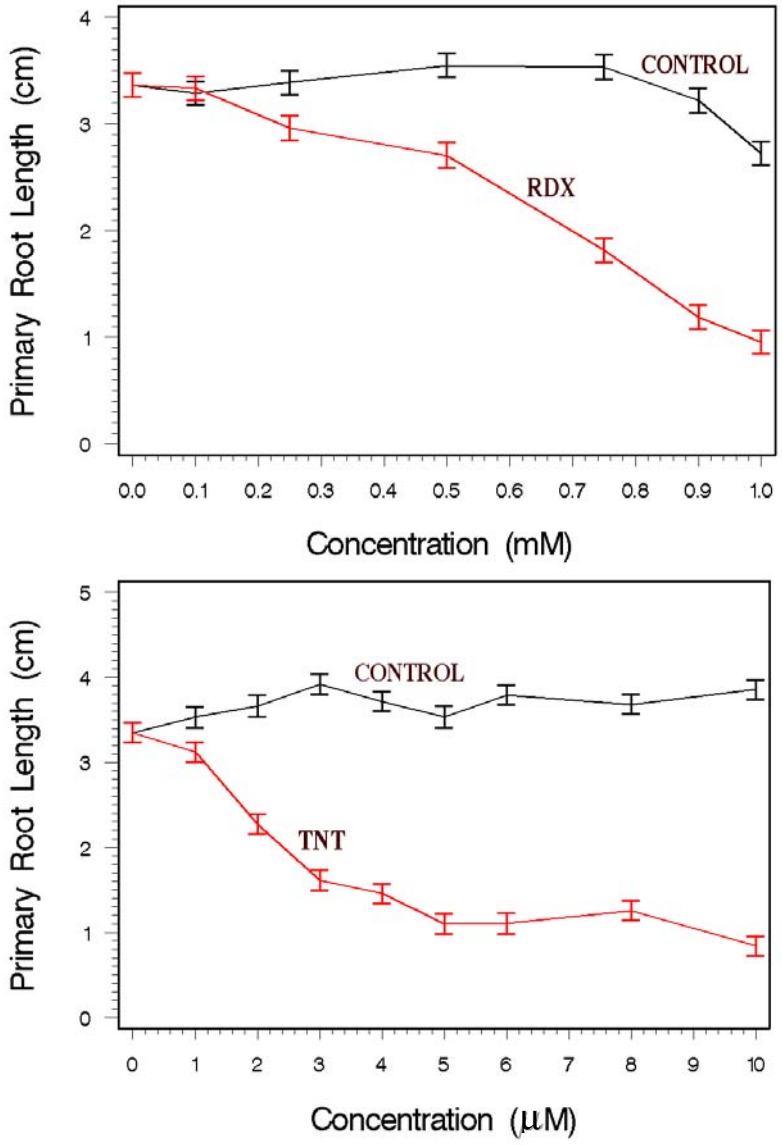

Figure 2.3. Primary root length of Arabidopsis plants exposed to RDX and TNT (6-7 days after germination). On the Y-axis is the primary root length in centimeters and on the X-axis is the concentration in millimolar for RDX and micromolar for TNT.

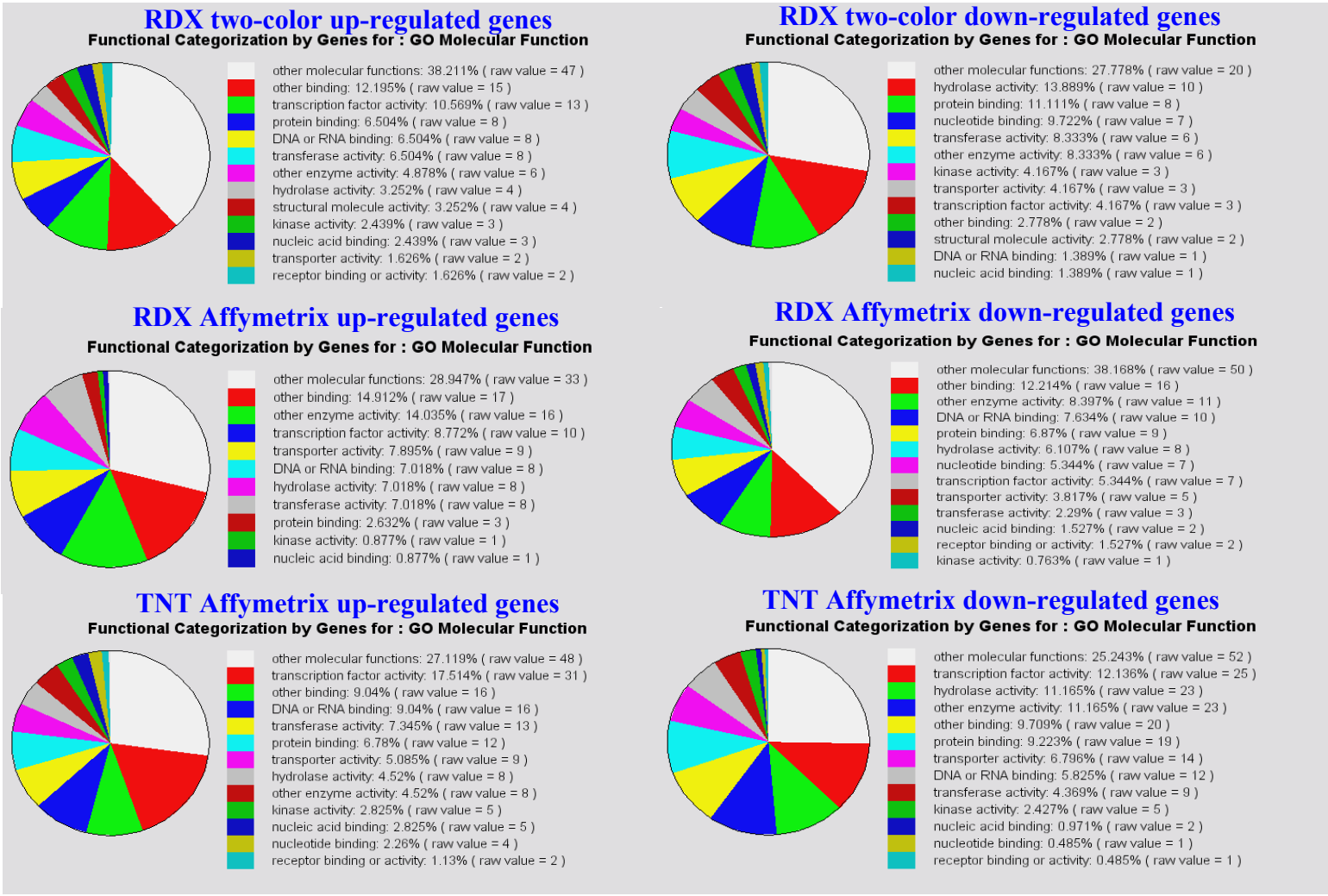

Figure 2.4. Pie charts for functional categorization by molecular function for genes differentially expressed in the RDX and TNT microarrays.

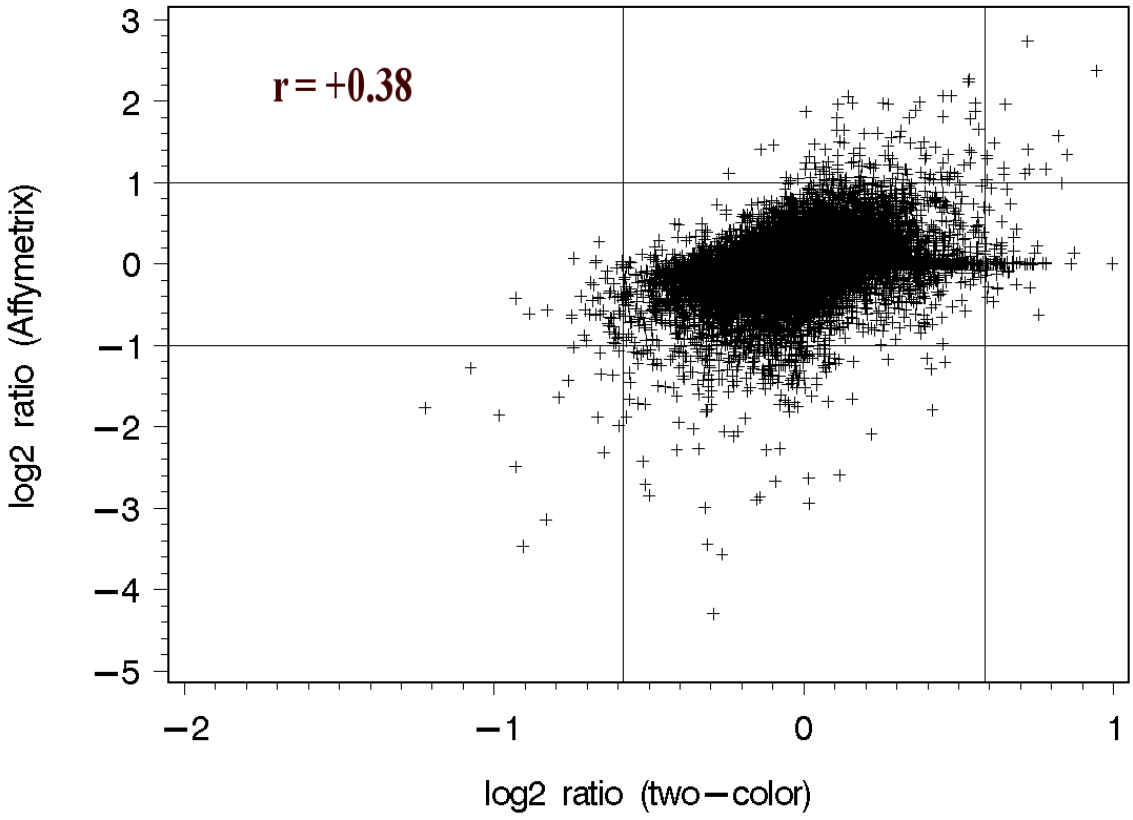

Figure 2.5. Log2 ratios of signal intensities for RDX from the two platforms (two-color and Affymetrix) plotted against each other. On the y axis are the $\log_2$ ratios from Affymetrix and on x axis are $\log_2$ ratios from two-color. The horizontal lines on the y axis at values +1 and -1 represents the cutoff value of +2.0 and -2.0 linear fold change respectively for Affymetrix and the vertical lines at values +0.58496 and -0.58496 on the x axis represents the cutoff values of +1.5 and -1.5 linear fold change respectively for two-color. The value for the Pearson correlation coefficient is represented as r on the graph.

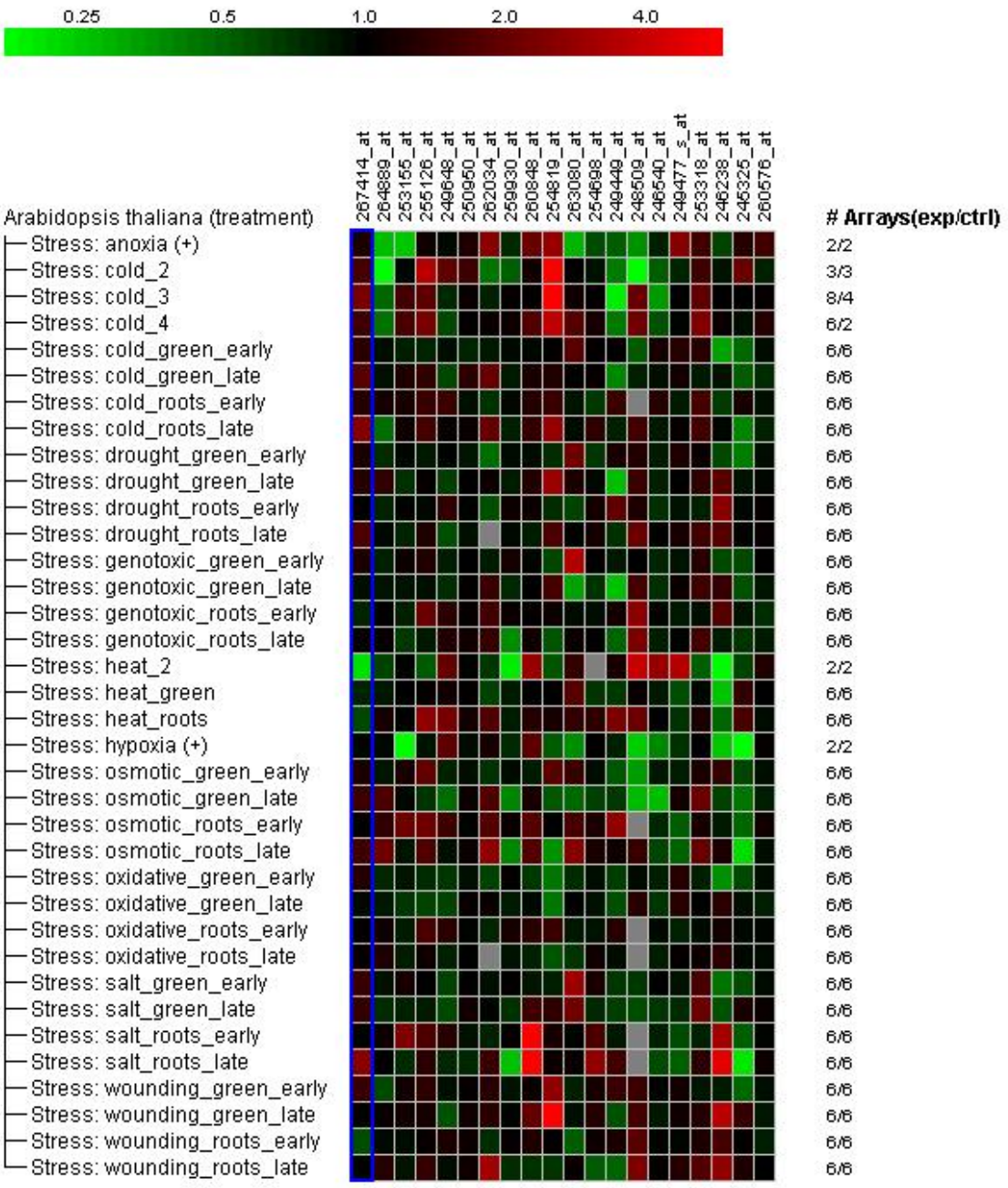

Figure 2.6. The expression profile under stress related conditions for the top 20 up-regulated genes in RDX two-color microarray experiment.

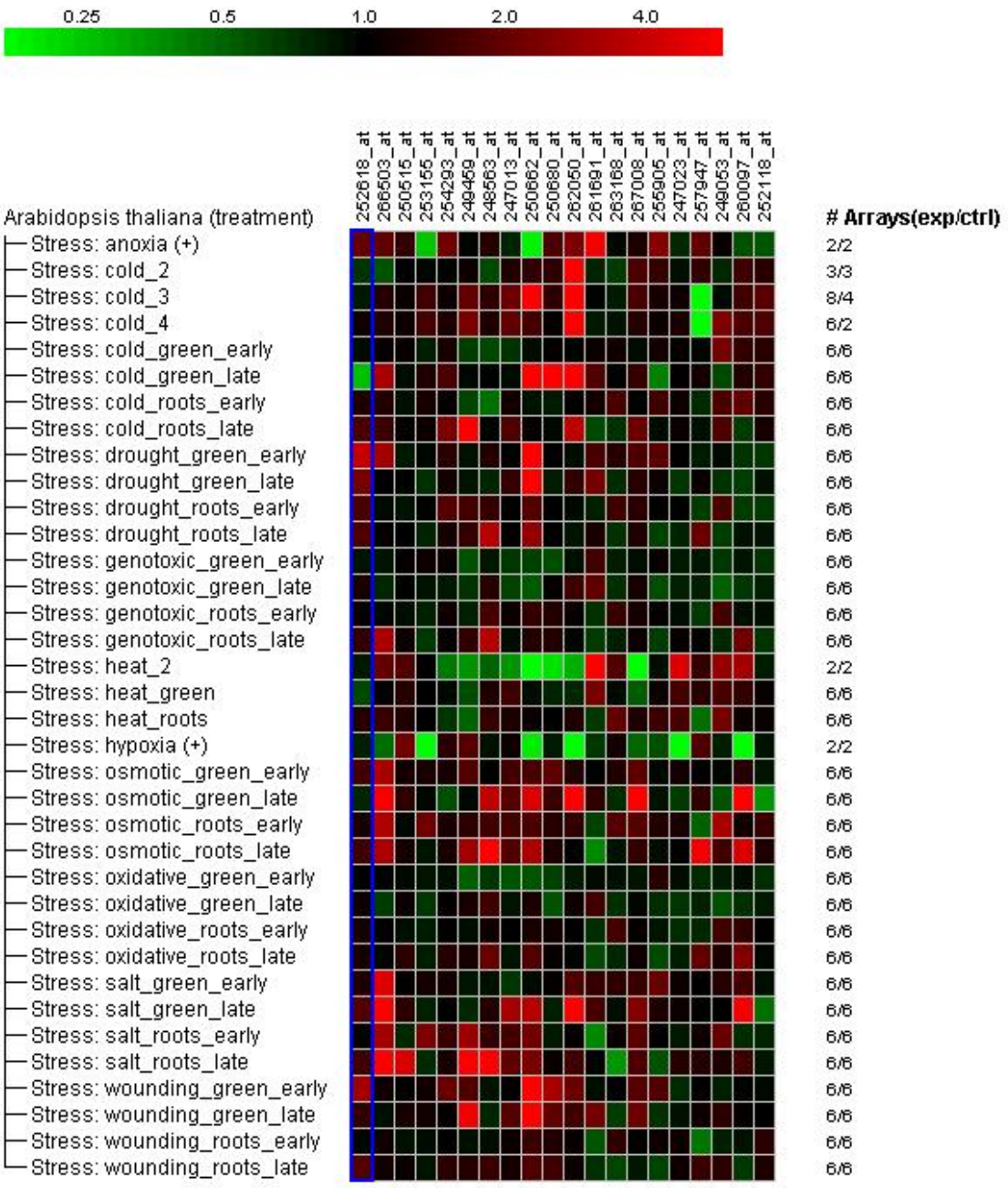

Figure 2.7. The expression profile under stress related conditions for the top 20 up-regulated genes in RDX Affymetrix microarray experiment.

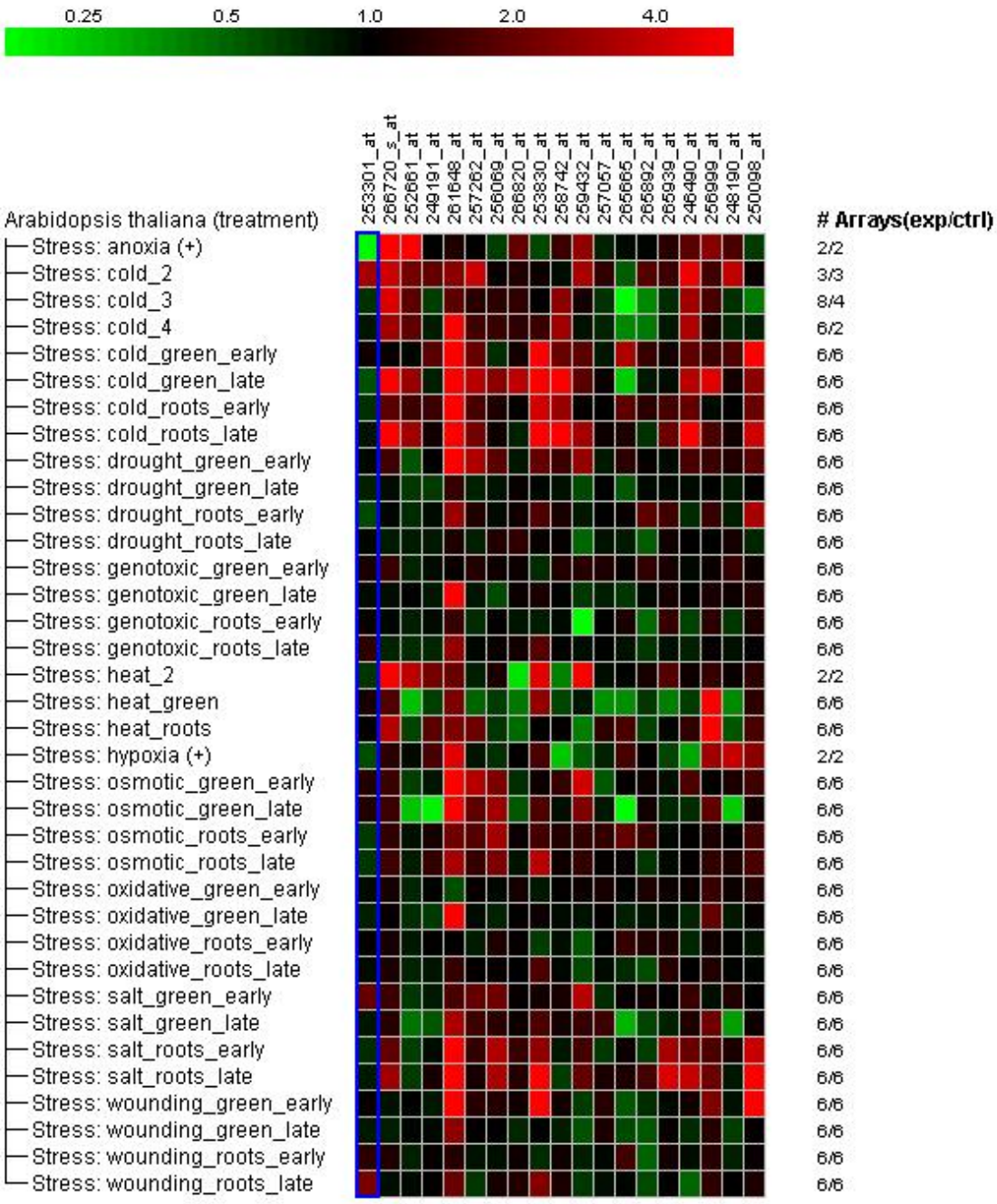

Figure 2.8. The expression profile under stress related conditions for the top 20 up-regulated genes in TNT Affymetrix microarray experiment.

Use of a site-specific recombination system to amplify inducible fluorescence in phytosensors

Abstract

A phytosensor is a plant that can detect or sense the presence of contamination including agriculturally important biological agents. Phytosensors are constructed by genetically engineering plants to contain contaminant- or pathogen-inducible promoters driving the expression of a reporter gene such as green fluorescent protein (GFP). When these phytosensors come in contact with or encounter a contaminant or a pathogen, the specific contaminant- or pathogen-inducible promoters are triggered to drive the expression of GFP. However, an inherent problem with using the native inducible promoters directly fused to reporter genes is lack of sufficient expression or in other words, the inducible promoters may not be strong enough to produce detectable levels of the reporter gene signal. In this study, a recombination strategy to amplify the signal from the reporter gene is described. In this strategy, the inducible promoter drives the expression of a recombinase gene. Upon induction and subsequent recombination event, a strong constitutive promoter such as CaMV 35S drives the expression of the reporter gene, thus amplifying the signal by several orders of magnitude.

Introduction

Homologous recombination and site-specific recombination (SSR) are broadly recognized as the two types of genetic recombination (Craig 1988). Homologous recombination is a process where two DNA segments are exchanged and can occur only when there is a high degree of homology present between the two segments (Craig 1988, Sadowski 1993). SSR, unlike homologous recombination, involves enzyme mediated rearrangement of DNA fragments that do not possess a high degree of homology (Ow and Medberry 1995; Craig 1988). Conservative site-specific recombination (CSSR) and transposition are two classes of SSR (Craig 1988, Sadowski 1993). CSSR involves exchange or recombination at highly specific regions within short stretches (recombination sites) of identical sequences in the participating DNA fragments, while transposition does not require any homology between the recombination sites (Craig 1988). CSSR can result in different DNA rearrangements depending on the relative orientation of the recombination sites, recombination sites in *cis* and oriented in the same direction results in deletion of the DNA fragment, while recombination sites in *cis* and in the opposite orientation results in inversion of the DNA fragment (Ow and Medberry 1995; Craig 1988). On the other hand, having recombination sites in *trans* on two linear DNA molecules results in exchange of DNA fragments and if one of the DNA molecules involved is circular, recombination results in a cointegration event, but this event is kinetically less favorable and less likely to occur (Ow 2002; Ow and Medberry 1995). There are several CSSR systems identified and shown to be functional in higher eukaryotes. Some of the well-characterized CSSR systems are as follows: (i) Cre-*lox* from *Escherichia coli* phage P1, where Cre (control of recombination) is the recombinase enzyme and *lox* (loci of *x*-over) is the recombination site recognized by Cre

recombinase; (ii) FLP-*FRT* from *Saccharomyces cerevisiae*, where FLP (flipping DNA) is the recombinase which recognizes *FRT* (FLP recombination target) sites; (iii) R-*RS* from *Zygosaccharomyces rouxii*, where recombination sites *RS* are recognized by the recombinase R and (iv) Gin-*gix* from bacteriophage Mu, where the Gin recombinase recognizes *gix* sites (Lyznik et al. 2003; Ow and Medberry 1995). These above mentioned recombinase systems are simple and straightforward since they require only their specific recombinase protein for recombination to take place without relying on any other accessory proteins or factors and all these four CSSR systems have been shown to work in plants (Lyznik et al. 2003; Ow and Medberry 1995; Kilby et al. 1995; Bayley et al. 1992; Dale and Ow 1991; Dale and Ow 1990). These simple and efficient CSSR systems can have a wide variety of applications in plant biotechnology. Some of these applications include excision of selectable marker genes from transgenic plants (Corneille et al. 2001; Zuo et al. 2001; Sugita et al. 2000; Gleave et al. 1999), excision of redundant copies of transgene in crop plants to reduce the extensive screening required to obtain single-copy transgenic lines (Ow 2002; Srivastava and Ow 2001; Srivastava et al. 1999), and site-specific integration of transgenes (Srivastava and Ow 2004; Lyznik et al. 2003; Ow 2002). Another novel application of these simple and efficient CSSR systems is phytosensing, where plants are used to sense environmental contaminants or agriculturally important biological agents. Phytosensors are constructed by genetically engineering plants to contain contaminant- or pathogen-inducible promoters driving the expression of a reporter gene. This phytosensor plant fluoresces when it comes in contact with or encounters the contaminant or the pathogen, thus reporting its presence. But, an inherent problem with using the native inducible promoters directly fused to reporter genes is lack of sufficient expression of the reporter gene or in other words, the inducible promoters may not be strong enough to produce

detectable levels of the reporter gene signal. In order to tackle this problem we attempt to use a proven and efficient site-specific recombination system such as FLP/*FRT* system to amplify the fluorescent signal facilitating efficient detection of the contaminant or the pathogen. A simple depiction of this system is presented in Figure 3.1. In this system, the inducible promoter (in this case, heat-shock inducible), instead of driving the expression of a green fluorescent protein (GFP), drives the expression of a recombinase gene, so that once induced, the recombination would result in the excision of the DNA fragment between the *FRT* recognition sites placing the strong constitutive CaMV 35S promoter in close proximity of GFP. This would lead to constitutive expression of GFP, amplifying the signal by several orders of magnitude and leading to efficient detection of the signal. In this study, the use of FLP/*FRT* site-specific recombination system to amplify the GFP signal in tobacco (*Nicotiana benthamiana*) using an *Agrobacterium*-mediated transient expression assay (Sparkes et al. 2006) was tested.

Results

Transient expression of GFP as measured by a handheld GFP meter

GFP expression in plants infiltrated with *Agrobacterium* carrying pBIN-HSP-FLP-GFP-Hyg (recombination construct) were significantly different when induced at both 37°C and 42°C with p-values 0.000 and 0.000 respectively, in comparison to the readings from the un-induced plants (Fig. 3.2) carrying the same construct. GFP expression in plants infiltrated with *Agrobacterium* carrying positive control construct pBI-HSP-GFP was significantly different from their corresponding un-induced counterparts at both 37°C and 42°C (Fig. 3.2) with p-values 0.000 and 0.000 respectively. Also, for the other positive control construct, pBIN-mgfp5er, the

GFP expression was significantly different at both 37°C and 42°C compared to un-induced plants with p-values 0.006 and 0.037 respectively. On the other hand, GFP expression in plants infiltrated with negative control (GV-3850; *Agrobacterium* strain alone carrying no binary constructs) showed no difference between un-induced plants and plants induced at 37°C (p-value: 0.185), although when induced at 42°C, was significantly different from the un-induced (Fig. 3.2) with a p-value of 0.000. Also, it was found that the transient expression of GFP with pBIN-HSP-FLP-GFP-Hyg (recombination construct) and pBI-HSP-GFP (direct fusion construct) was significantly lower compared to pBIN-mgfp5ER (CaMV 35S fused to GFP). These results demonstrate that the recombination took place and resulted in transient expression of GFP.

Transient expression of GFP using epifluorescence microscopy

Epifluorescence microscopy revealed the transient expression of GFP in all the infiltrated tissues except the negative control (GV-3850; *Agrobacterium* alone) (Fig. 3.3). The leaves infiltrated with GV-3850 did not show GFP either in the un-induced or induced state (37°C or 42°C). Infiltrations using the pBI-HSP-GFP construct, showed GFP only when induced at 37°C or 42°C and no GFP in the un-induced condition. The recombination construct (pBIN-HSP-FLP-GFP-Hyg) also showed GFP when induced at 37°C or 42°C with no GFP expression in the un-induced state, confirming that the recombination took place following induction, resulting in transient expression of GFP. Finally the positive control construct (pBIN-mgfp5er) showed GFP under all the three conditions (un-induced, induction at 37°C and induction at 42°C) as expected.

Discussion

Agrobacterium-mediated transient expression assay by infiltration (agroinfiltration) of tobacco epidermal cells is a fast and efficient technique to study new constructs and assess expression of transgenes (Sparkes et al. 2006). Here this technique was employed to examine the use of a well characterized recombination system to amplify fluorescent signal in a phytosensing system. This is a novel application of a recombination system in plant biotechnology. *Gmhsp*, a heat-inducible promoter from soybean (Czarnecka et al. 1989) was used to drive the recombination event and upon heat-induction, this resulted in placing CaMV 35S, a strong constitutive promoter in close proximity to GFP, thus amplifying the signal by several orders of magnitude upon induction. Once the phytosensor plant uptakes the specific contaminant or encounters the pathogen of interest, the specific inducible promoter drives the expression of the FLP recombinase enzyme. The resulting recombination event excises the DNA fragment in between CaMV 35S promoter and GFP, bringing them next to each other resulting in strong constitutive expression of GFP. The relative orientation of the recombination sites in this system are in the same direction causing the excision of the fragment in between the recognition sites, this excision process is highly efficient and reliable compared to a kinetically less favorable integration event, that occurs when the recombination sites are in the opposite orientation (Ow 2002; Ow and Medberry 1995).

This recombination system was tested via a transient *Agrobacterium*-mediated infiltration assay in tobacco. The transient expression of GFP in tobacco was confirmed using two methods: (1) measurement of the GFP signal using a GFP meter, a portable spectrofluorometer and (2) fluorescence imaging. The GFP meter readings and the epifluorescence microscopy data from

the *Agrobacterium* infiltrations using the recombination construct (pBIN-HSP-FLP-GFP-Hyg) confirmed the occurrence of recombination resulting in transient expression of GFP (Fig. 3.2; Fig. 3.3). Similarly with the pBI-HSP-GFP, the direct heat-shock promoter fusion construct, GFP was found only when induced at 37°C or 42°C as expected. The positive control construct, pBIN-mgfp5er also as anticipated showed GFP under all the conditions, i.e., induced as well as un-induced. The increase in GFP signal for the 42°C induction observed for the negative control (GV-3850; *Agrobacterium* alone) and for two of the other constructs (pBI-HSP-GFP and pBIN-mgfp5er) is likely the result of noise (signal related to stress related compounds or metabolites) leading to higher green autofluorescence and not actual GFP expression. In line with this notion, all the images from the epifluorescence microscope for the negative control (GV-3850) showed no visible GFP in either induced or un-induced state (Fig. 3.3), which may also further suggest that 37°C is better than 42°C for induction.

Another noticeable difference was the transient expression of GFP in positive control construct (pBIN-mgfp5er) in comparison to the recombination construct (Fig. 3.1). This low transient expression of GFP in the recombination construct can be partially attributed to the slow heat-induction process (Yang et al. 2000) that was followed in this experiment, resulting in lower rate of recombination and thus low level of GFP expression. Also, heat shock conditions for *Gmhsp* promoter, likely needs to be optimized for a transient assay involving recombination system. We also observed variation in the level of GFP expression for the same construct on different infiltrated leaves either on the same plant or different plants. On the other hand, the images from the epifluorescence microscope seemed convincing, confirming the event of the recombination system and demonstrating that induction at 37°C was more reliable than 42°C.

One possible solution for the problems encountered in this transient assay, is to test this system in stable transgenics.

In conclusion, the results from this study indicate that the recombination system tested for amplifying the signal from potentially weak inducible promoters is promising for explosives or pathogen phytosensing applications.

Materials and methods

Vector constructs and Agrobacterium-mediated infiltration assay

The vectors used in this study are presented in Figure 3.4. Fully expanded leaves from 4 – 6 weeks old tobacco plants were infiltrated with *Agrobacterium* containing the plant expression vector carrying FLP/FRT recombination system and GFP as the reporter gene (pBIN-HSP-FLP-GFP-Hyg). As positive controls, *Agrobacterium* containing HSP directly fused with GFP (pBI-HSP-GFP) and *Agrobacterium* containing constitutive promoter CaMV 35S directly fused with GFP (pBIN-mgfp5ER) were used. As a negative control, *Agrobacterium* alone carrying no plant expression vectors was used. At least 3 intact leaves from six independent tobacco plants were infiltrated with *Agrobacterium* carrying each of the constructs as described by Sparkes et al. (2006).

Heat-induction experiments

Heat-shock treatment was carried out as described by Yang et al. (2000). The tobacco plants were heat-shocked 48 hrs after agroinfiltration at 37°C and 42°C for 20 hrs. The plants were heat-shocked 48 hrs after agroinfiltration to allow for the integration of the T-DNA into the

plant genome. After heat-shock treatment the plants were returned to normal conditions to recover for 6 hrs before they were analyzed for GFP expression.

GFP meter readings

The transient expression of GFP was measured using a portable spectrofluorometer called GFP meter (Opti-Sciences, Tyngsboro, MA, USA). For every construct, readings from at least six spots per plant and from six independent plants were collected. The readings recorded were analyzed using two-tail two-sample t-test in Microsoft Excel.

Epifluorescence microscopy

An epifluorescence microscope under blue light excitation with a FITC filter was used to observe GFP expression and the images were captured using Q capture imaging software (Quantitative Imaging Corporation, British Columbia, Canada). The transient GFP expression was recorded under different conditions.

References

Bayley CC, Morgan M, Dale EC, Ow DW (1992) Exchange of gene activity in transgenic plants catalyzed by the Cre-Lox site-specific recombination system. Plant Molecular Biology 18: 353-361

Corneille S, Lutz K, Svab Z, Maliga P (2001) Efficient elimination of selectable marker genes from the plastid genome by the Cre-*lox* site-specific recombination system. Plant Journal 27: 171-178

Craig NL (1988) The mechanism of conservative site-specific recombination. Annual Review of Genetics 22: 77-105

Czarnecka E, Key JL, Gurley WB (1989) Regulatory domains of the *Gmhsp* 17.5-e heat-shock promoter of soybean. Molecular and Cellular Biology 9: 3457-3463

Dale EC, Ow DW (1990) Intramolecular and intermolecular site-specific recombination in plant-cells mediated by bacteriophage-P1 recombinase. Gene 91: 79-85

Dale EC, Ow DW (1991) Gene transfer with subsequent removal of the selection gene from the host genome. Proceedings of the National Academy of Sciences 88: 10558-10562

Gleave AP, Mitra DS, Mudge SR, Morris BAM (1999) Selectable marker-free transgenic plants without sexual crossing: transient expression of cre recombinase and use of a conditional lethal dominant gene. Plant Molecular Biology 40: 223-235

Kilby NJ, Davies GJ, Snaith MR, Murray JAH (1995) Flp recombinase in transgenic plants - constitutive activity in stably transformed tobacco and generation of marked cell clones in Arabidopsis. Plant Journal 8: 637-652

Lyznik LA, Gordon-Kamm WJ, Tao Y (2003) Site-specific recombination for genetic engineering in plants. Plant Cell Reports 21: 925-932

Ow DW, Medberry SL (1995) Genome manipulation through site-specific recombination. Critical Reviews in Plant Sciences 14: 239-261

Ow DW (2002) Recombinase-directed plant transformation for the post-genomic era. Plant Molecular Biology 48: 183-200

Sadowski PD (1993) Site-Specific Genetic-Recombination - Hops, Flips, and Flops. FASEB Journal 7: 760-767

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nature Protocols 1: 2019-2025

Srivastava V, Anderson OD, Ow DW (1999) Single-copy transgenic wheat generated through the resolution of complex integration patterns. Proceedings of the National Academy of Sciences of the United States of America 96: 11117-11121

Srivastava V, Ow DW (2001) Single-copy primary transformants of maize obtained through the co-introduction of a recombinase-expressing construct. Plant Molecular Biology 46: 561-566

Srivastava V, Ow DW (2004) Marker-free site-specific gene integration in plants. Trends in Biotechnology 22: 627-629

Sugita K, Kasahara T, Matsunaga E, Ebinuma H (2000) A transformation vector for the production of marker-free transgenic plants containing a single copy transgene at high frequency. Plant Journal 22: 461-469

Yang YN, Li RG, Qi M (2000) In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. Plant Journal 22: 543-551

Zuo JR, Niu QW, Moller SG, Chua NH (2001) Chemical-regulated, site-specific DNA excision in transgenic plants. Nature Biotechnology 19: 157-161

Appendix

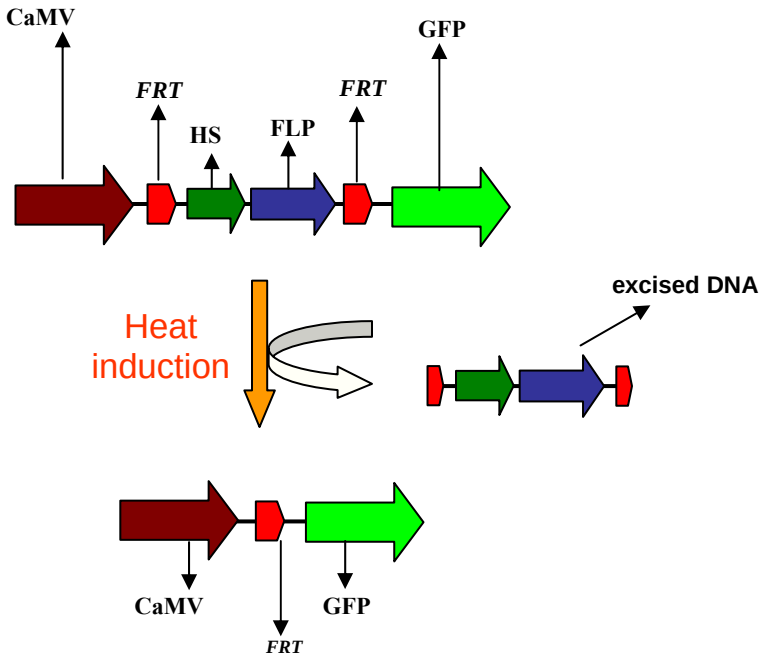

Figure 3.1. Schematic diagram of the FLP/FRT signal amplification system. Upon heat induction, the FLP recombinase protein produced will recognize the FRT sites and the region between these two FRT recognition sites will be excised and the reporter gene GFP is brought under the influence of the constitutive cauliflower mosaic virus (CaMV) 35S promoter resulting in plant-wide expression of GFP.

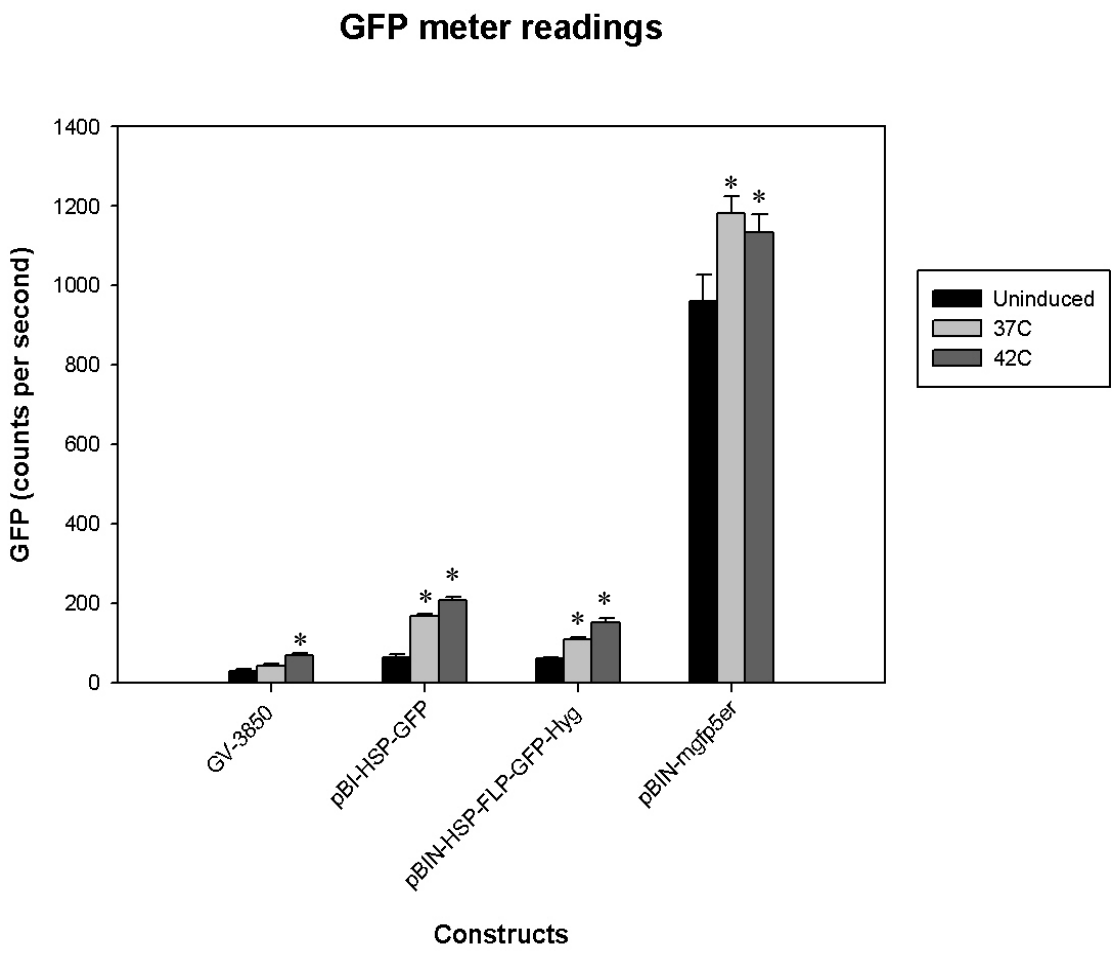

Figure 3.2. GFP meter readings measuring transient expression of GFP when un-induced and induced at 37°C and 42°C.

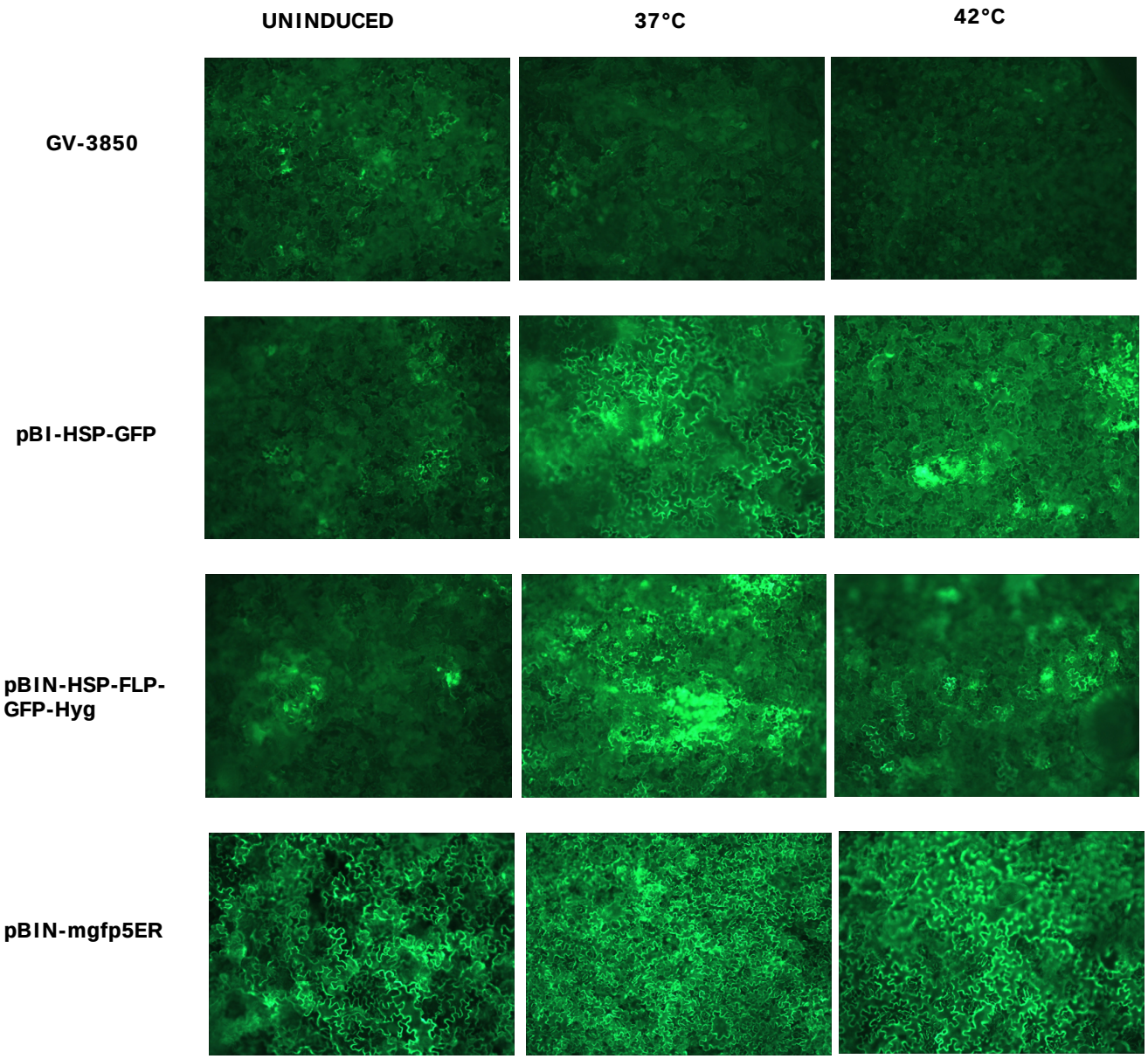

Figure 3.3. Transient GFP expression at 37°C, 42°C and un-induced conditions for different plant expression vectors.

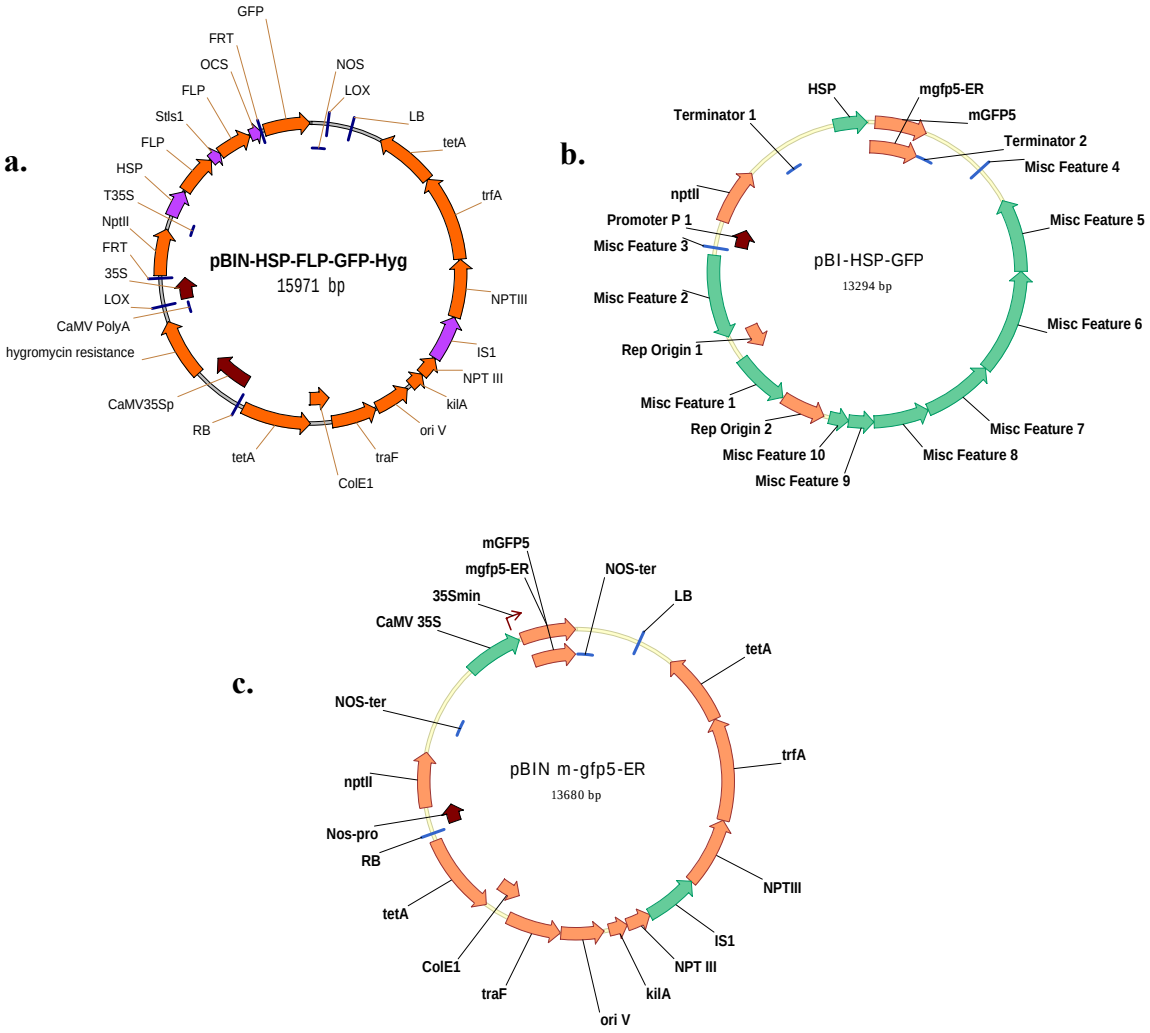

Figure 3.4. Plant expression vectors. a. pBIN-HSP-FLP-GFP-Hyg, b. pBI-HSP-GFP and c. pBIN-mgfp5ER used in the transient agroinfiltration assays.

Use of Rank Products to analyze microarray data improves false discovery rate p-values

Abstract

Analysis of microarray data involves simultaneous testing of tens of thousands of genes for significance, which is expected to result in large numbers of false positives. Q-value, a method based on false discovery rate, is widely used for controlling the number of false positives while analyzing microarray data, producing adjusted p-values that are larger, reducing the chance of false positive results. On the other hand, the behavior of p-values obtained from multiple tests in a microarray experiment can be very informative. When the null hypothesis is true, the p-values obtained follow a uniform distribution and fall in the range [0,1], but, under alternative hypothesis, the p-values tend to be smaller and group closer to zero. We observed that datasets with unusual p-value distributions had very high q-values, producing very few or no significant genes. Datasets with unusual p-value distributions and very high q-values when analyzed using the standard ANOVA method of analysis were found to give improved p-value distributions and much lower adjusted p-values when using the non-parametric Rank Products method. Empirical evidence suggests that this nonparametric method performs very well on a variety of datasets, yielding larger numbers of significant genes with an acceptable q-value. We suggest this nonparametric method to be considered for analyzing microarray data when other methods perform too conservatively.

Introduction

Microarrays have become an invaluable technique in functional genomics for scientists to study differential gene expression. With the evolution of microarray technology, an overwhelming number of analysis methods have become available to identify the significantly differentially expressed genes across two or more treatment conditions. With the availability of so many different methods, biologists with modest statistical background find it very difficult to make the right choice. Standard parametric approaches available include simple t-test, when only two treatment conditions with replicated samples are compared and ANOVA when more than two treatment conditions are being compared (Cui and Churchill 2003). Mixed model ANOVA is another choice when more than one treatment factor and different sources of variation have to be modeled (Cui and Churchill 2003). Apart from these standard approaches there are several Bayesian model based approaches such as t-test using Bayesian estimate of variance between replicates (Long et al. 2001), Bayesian framework based methods (Newton et al. 2001; Baldi and Long 2001). Nonparametric approaches available for microarray analysis include the rank products method by Breitling et al. (2004), mixture modeling method by Pan et al. (2003), and significance analysis of microarrays (SAM) by Tusher et al. (2001). Since microarray data are mostly found to have complex distributional forms and normality issues (Kim et al. 2006; Qian and Huang 2005; Zhao and Pan 2003; Troyanskaya et al. 2002; Hunter et al. 2001) nonparametric methods that do not rely on any assumptions regarding data structure are appealing.

After selecting a method of analysis, another problem that needs to be addressed while analyzing microarray data is multiple testing. A typical microarray data analysis involves testing

of tens of thousands of genes simultaneously, and is expected to result in a large number of false positives especially when typical significance levels (α-values) such as 5% or 1% are considered (Allison et al. 2006, Ge et al. 2003). In order to control the rate of false positives, initially, family-wise error rate (FWER) control methods such as Bonferroni correction had been used (Allison et al. 2006). This method proves to be very conservative and limits the number of false positives to less than the α-value, resulting in very few or no significantly differentially regulated genes (Allison et al. 2006). These conservative methods defeat the purpose of most biologists who are ready to accept some false positives provided the analysis results in some important findings. This led to the use of false discovery rate (FDR) approach to control the number of false positives instead of FWER approach (Allison et al. 2006). Benjamini and Hochberg (1995) first coined the term 'FDR', which equals the expected proportion of incorrectly rejected null hypotheses (false positives) among the genes that were found to be statistically significant (Allison et al. 2006; Pawitan et al. 2005b). Another measure of significance called the q-value (Storey 2003; Storey and Tibshirani 2003; Storey 2002), which is based on FDR has become more popular as it is less conservative and more applicable than the Benjamini and Hochberg FDR controlling method (Storey 2003; Storey 2002). Q-value method was also found to be the most powerful test when compared to several other FDR methods (Qian and Huang 2005).

It is important to note that the p-value distribution obtained from multiple tests carried out during microarray analysis contains valuable information that can be exploited to answer some of the fundamental questions biologists might have about a microarray experiment and also, the p-values contribute to the calculation of false-discovery rates (FDRs) (Allison et al. 2006; Allison et al. 2002; Delongchamp et al. 2004). P-value as defined by Hung et al. (1997) is a measure of evidence against the null hypothesis. Since p-value is a function of the random

variable in question, p-value itself is considered as a random variable and it is well known that under null hypothesis, the p-values follow a uniform distribution over the range [0,1] irrespective of sample size or the test statistic used, while under the alternative hypothesis the p-value distribution is distinctly skewed with more smaller p-values and they appear grouped more closely to zero than to one (Fig. 4.1; Hung et al. 1997; Donahue 1999; Sackrowitz and Samuel-Cahn 1999; Allison et al. 2002; Schweder and Spjøtvoll 1982; Delongchamp et al. 2004; Xiang et al. 2006). This difference in p-value distribution allows statistical testing whether the observed p-value distribution differs from the uniform p-value distribution (i.e., under null hypothesis), which in turn answers the question whether expression of any of the genes differs among treatment groups in a microarray experiment (Allison et al. 2002). The distribution of p-values from a continuous test statistic under alternative hypothesis is known to depend on sample size, effect size and the distribution of the test statistic used to compute p-values (Hung et al. 1997).

Here we present microarray datasets that display unusual p-value distributions when analyzed using a standard parametric approach, and have very high q-values thus limiting the number of statistically significant genes. We show here that a nonparametric method (Rank Products by Breitling et al. 2004) improves the behavior of the p-value distribution, and also improves q-values.

Results

Five microarray datasets were analyzed, two using Affymetrix Arabidopsis microarrays, two using Affymetrix mouse Genome 430 2.0 Array and the fifth using oligonucleotide spotted

array for *Arabidopsis thaliana*. These datasets were independently analyzed using the standard parametric approach, ANOVA, and a nonparametric approach, rank products statistics (Breitling et al. 2004). Breitling et al (2004) refer to FDR as percentage of false-positives (pfp), which is calculated based on Storey's (Storey 2003) q-value. To be consistent with the terminology, I refer to the pfp values obtained from rank products statistics as q-values. The density histograms of p-values and q-values were used to show that the nonparametric rank products statistics was a better method of analysis and resulted in many significant genes, while ANOVA gave very few or no significant genes with acceptable q-values.

In a microarray experiment since most of the genes are expected to be non-significant, the p-values are expected to be distributed uniformly between 0 and 1 and under alternative hypothesis the p-value distribution is skewed with more smaller p-values and grouped more closely to zero than one (Fig. 4.1; Yang 2004; Hung et al. 1997; Donahue 1999; Sackrowitz and Samuel-Cahn 1999; Allison et al. 2002; Schweder and Spjøtvoll 1982; Delongchamp et al. 2004; Xiang et al. 2006). A typical density histogram of p-values from a microarray experiment with statistically significant genes would look similar to Figure 4.1b (also see Yang 2004). But, irregular density histograms of p-values like the histogram in Figure 4.5a do occur. The analysis of our datasets shows that when the density histogram of p-values is irregularly shaped, the q-values values tend to be very high, thus resulting in very few or no significant genes.

The p-value and q-value distributions obtained for the two-color spotted array dataset using ANOVA and rank products statistics is presented in Figure 4.2. Figure 4.2a shows the atypical density histogram of p-values from the ANOVA analysis and resulting in corresponding very high q-values (Fig. 4.2b). Only two genes were found to be significant at a q-value cutoff of less than or equal to 10%. Figure 4.2c shows the density histogram of p-values from the rank

products method of analysis, which is markedly a big improvement on the histogram from ANOVA and appears to be close to the one from a typical microarray data. The corresponding histogram of q-values from rank products analysis (Fig. 4.2d) for this dataset also looked better than its ANOVA counterpart reporting a total of 931 significant genes differentially regulated with q-values less than or equal to 10% (Table 4.1). Similar trend of improvement with the use of nonparametric rank products statistics was observed for all the other datasets. For the two Arabidopsis Affymetrix datasets, ANOVA analysis reported no significantly differentially regulated genes with q-values less than or equal to 10% and one of the datasets had the lowest q-value of 38% and the other one had 32% as the lowest q-value. When these datasets were analyzed using rank products method, at a q-value cutoff of 10% there were a total of 217 genes reported significant for one dataset and 297 genes for the other dataset (Table 4.1). The p-value histograms for these two datasets from the ANOVA analysis were irregularly shaped, and the histograms for the q-values from ANOVA also looked unusual indicating that there were no significant genes with q-value below 10% (Fig. 4.3, Fig. 4.4, and Table 4.1). Again for these two datasets rank products statistics greatly improved the q-value histograms, while not much improvement was seen with the p-value histograms (Fig. 4.3, Fig. 4.4, and Table 4.1). The two mouse microarray datasets when analyzed using ANOVA did not have any significant genes below 10% q-value and the lowest q-values were 0.96 for one dataset and 0.45 for the other dataset again indicating that the experiment was not worth the time and money. The rank products method again performed very well here and resulted in 67 significant genes for one dataset and 368 significant genes for the other dataset at a q-value cutoff of 10% (Table 4.1). The density histograms of p-values and q-values for these two mouse microarray datasets are presented in Figure 4.5 and Figure 4.6. The density histogram of p-values from ANOVA for one

of the mouse datasets (Fig. 4.5a) shows the typical irregular shape and the corresponding histogram of q-values with very high values (Fig. 4.5b). Rank products statistics greatly improved the histogram of q-values (Fig. 4.5d), but only a marginal improvement was seen in the p-value histogram (Fig. 4.5c). The other mouse dataset was also interesting in that the p-value histogram from ANOVA (Fig. 4.6a) looked close to normal, but the corresponding q-values from ANOVA were still very high as seen in Figure 4.6b. Clearly for this dataset also, rank products method resulted in improved p-value and q-value histograms as seen in Figure 4.6c and Figure 4.6d respectively. In summary, the results indicate that the rank products statistics performed very well on our datasets, which came from different platforms and with different sample size by correcting the unusual p-value distributions and very high q-values.

Discussion

Empirical evidence suggests that the irregular shape of the p-value density histogram is an indication of obtaining very few or no significant genes below an acceptable q-value. But, the source of high q-values remains unanswered. According to Pawitan et al. (2005a), some factors that determine FDR include a) proportion of differentially regulated genes that are true b) how the true differences are distributed c) variation in the measurement and d) sample size. The two Arabidopsis Affymetrix datasets and the oligonucleotide spotted array dataset were small sample size experiments (see Methods) and are expected to be noisy, and could be one of the causes for the false discovery rates or the q-values to be very high, as sample size is one of the factors determining the characteristics of FDR (Pawitan et al. 2005a). Pawitan et al. (2005b) report that the presence of bias in the estimation of proportion of genes that are not differentially regulated

(π_0) and FDR, when using current FDR methods like q-value method by Storey (2002) could result in overall loss of power and, they also suggest an improved method to estimate π_0. Yang (2004) reported a similar problem of high FDRs resulting in very few or no significant features for some of their microarray datasets. Yang (2004) also used q-value method and argued that this method performs poorly when very few genes are expected to be significant and that other FDR control methods such as Benjamini and Hochberg (1995) performed better under this situation. Yang (2004) similar to Pawitan et al. (2005b) commented that the q-value method overestimates π_0 resulting in no significant genes, when few significant genes are expected. Yang (2004) also commented that the irregular shape of the p-value histograms is causing overestimation of π_0 and thus high q-values leading to no significant features and suggested modifications to the q-value method. Similar to Yang (2004) we observed the irregular shapes of p-value histograms which probably is resulting in not only high q-values, but also high FDRs using Benjamini and Hochberg (1995) method or family wise error rate using Bonferroni's p-value adjustment (data not shown). This problem of irregular p-value histograms and high q-values with our datasets was only observed when standard parametric ANOVA was used for analysis, which led us to the use of non-parametric rank products (Breitling et al. 2004) method for analyzing our data. Rank products method is a very attractive approach for analyzing microarray data, because of its simplicity and relative strong performance especially when the sample size is small and the data is noisy (Breitling et al. 2004). Another desirable characteristic of the rank products method is that, it is based on biological reasoning and therefore is an effective method for identifying biologically relevant changes in gene expression (Breitling et al. 2004). Breitling and co-workers (2004) in their original paper have also proved their rank products method to be more reliable and consistent than the popular nonparametric SAM method

by Tusher et al. (2001). Interestingly, the rank products method performed very well on all of our datasets and is probably indicating that the problem of high FDRs was inherent to the strong assumptions made by the parametric ANOVA method. Using the rank products method, we were able to identify several significant genes with an acceptable q-value for all of our datasets.

Owing to the inherent normality issues and complex distributional forms with most of the microarray data (Kim et al. 2006; Qian and Huang 2005; Zhao and Pan 2003; Troyanskaya et al. 2002; Hunter et al. 2001), we suggest non-parametric rank products method to be considered as the method of choice for analyzing microarray data when other methods perform too conservatively.

Methods

Microarray datasets used

The microarray datasets used are described as follows:

a. *Oligonucleotide spotted* Arabidopsis *microarray dataset:* This experiment was aimed at analyzing the transcriptional responses in *Arabidopsis thaliana* in response to explosive compound RDX (Royal Demolition explosive) and was performed at Dr. Neal Stewart's lab; Department of Plant Sciences, University of Tennessee, Knoxville and the arrays were purchased from Dr. David Galbraith at the University of Arizona. This dataset was obtained from a two-color microarray experiment consisting of six chips which included three biological replicates and a dye swap technical replicate (to avoid dye bias) for every set of replicates.

b. *Affymetrix* Arabidopsis *microarray datasets:* These experiments were also aimed at analyzing transcriptional changes in *Arabidopsis thaliana* in response to explosive compounds RDX

(Royal Demolition explosive) and TNT (2,4,6 – trinitrotoluene). The samples for both the experiments were prepared at Dr. Neal Stewart's lab and the chips were processed at the University of Tennessee, Knoxville Affymetrix Core Facility. These experiments involved hybridization of four slides consisting of two biological replicates for every explosive compound.

c. *Affymetrix mouse microarray datasets:* This experiment was aimed at studying gene expression changes in liver and adipose fat tissue of mutant and control mice fed with specific diets, to understand mechanisms underlying obesity. These experiments consisted of 10 chips for the fat tissue experiment (5 arrays for mutant mice and 5 arrays for control mice) and 10 chips for the liver tissue experiment (5 arrays for mutant mice and 5 arrays for control mice). Experiments were performed at Genome Explorations Inc. (Memphis, TN) using Affymetrix GeneChip® Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, CA) following the standard protocol.

Analysis of the microarray datasets

All the datasets were subjected to the standard parametric ANOVA (t-test) analysis and the non-parametric rank products method of analysis. ANOVA was performed using SAS® software version 9.1.3 (2000) for all the datasets and the normalized and $\log_2$ transformed data from both two-color and Affymetrix microarrays were statistically analyzed using rank products statistics as described by Breitling et al. (2004). Bioconductor RankProd package (available at http://bioconductor.org/packages/2.2/bioc/html/RankProd.html) was used to perform the rank products analysis (Hong et al. 2006, Gentleman et al. 2004). The false discovery rate (FDR) value obtained was based on 10,000 random permutations. Since 10,000 random permutations was computer intensive, 1000 random permutations were performed 10 different times each time

starting with a different random seed number and the average q-value thus calculated was used for further analysis.

Rank Product method of analysis

In a microarray experiment, the rank index for a gene in a random list of genes sorted by fold change can be calculated by r/n, where r = rank or position of the gene in the list and n = number of genes in the list. And the corresponding rank product for every gene is given by the product of the rank indexes across all replicates. This method uses rank product as a measure to identify significantly differentially regulated genes. Random permutation method is further used to determine the reference distribution of rank product values for every gene, which is in turn used to calculate a p-value i.e., to determine the probability of observing a given or more extreme rank product value in a random experiment. Genes identified with smaller rank product values and significant p-values serve as good candidates for further validation and characterization (Breitling et al., 2004).

References

Allison DB, Gadbury GL, Heo MS, Fernandez JR, Lee CK, Prolla TA, Weindruch R (2002) A mixture model approach for the analysis of microarray gene expression data. Computational Statistics & Data Analysis 39: 1-20

Allison DB, Cui XQ, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics 7: 55-65

Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics 17: 509-519

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289-300

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573: 83-92

Cui XQ, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. Genome Biology 4:

Delongchamp RR, Bowyer JF, Chen JJ, Kodell RL (2004) Multiple-testing strategy for analyzing cDNA array data on gene expression. Biometrics 60: 774-782

Donahue RMJ (1999) A note on information seldom reported via the P value. American Statistician 53: 303-306

Ge YC, Dudoit S, Speed TP (2003) Resampling-based multiple testing for microarray data analysis. Test 12: 1-77

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y.C., 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.H. and Zhang, J.H. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Hong, F.X., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L. and Chory, J. (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22, 2825-2827.

Hung HMJ, Oneill RT, Bauer P, Kohne K (1997) The behavior of the P-value when the alternative hypothesis is true. Biometrics 53: 11-22

Hunter L, Taylor RC, Leach SM, Simon R (2001) GEST: a gene expression search tool based on a novel Bayesian similarity metric. Bioinformatics 17: S115-S122

Kerr MK (2003) Design considerations for efficient and effective microarray studies. Biometrics 59: 822-828

Kim SY, Lee JW, Sohn IS (2006) Comparison of various statistical methods for identifying differential gene expression in replicated microarray data. Stat Methods Med Res 15: 3-20

Long AD, Mangalam HJ, Chan BYP, Tolleri L, Hatfield GW, Baldi P (2001) Improved Statistical Inference from DNA Microarray Data Using Analysis of Variance and A Bayesian Statistical Framework. Analysis of global gene expression in *Escherichia coli* K12. J Biol Chem 276: 19937-19944

Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW (2001) On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. Journal of Computational Biology 8: 37-52

Pan W, Lin J, Le CT (2003) A mixture model approach to detecting differentially expressed genes with microarray data. Functional & Integrative Genomics 3: 117-124

Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A (2005) False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics 21: 3017-3024

Pawitan Y, Murthy KRK, Michiels S, Ploner A (2005) Bias in the estimation of false discovery rate in microarray studies. Bioinformatics 21: 3865-3872

Qian HR, Huang S (2005) Comparison of false discovery rate methods in identifying genes with differential expression. Genomics 86: 495-503

Sackrowitz H, Samuel-Cahn E (1999) P values as random variables - Expected P values. American Statistician 53: 326-331

Schweder T, Spjøtvoll E (1982) Plots of P-Values to Evaluate Many Tests Simultaneously. Biometrika 69: 493-502

Storey JD (2002) A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology 64: 479-498

Storey JD (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. Annals of Statistics 31: 2013-2035

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America 100: 9440-9445

Troyanskaya OG, Garber ME, Brown PO, Botstein D, Altman RB (2002) Nonparametric methods for identifying differentially expressed genes in microarray data. Bioinformatics 18: 1454-1461

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America 98: 5116-5121

Xiang QF, Edwards J, Gadbury GL (2006) Interval estimation in a finite mixture model: Modeling P-values in multiple testing applications. Computational Statistics & Data Analysis 51: 570-586

Yang X (2004) Qvalue methods may not always control false discovery rate in genomic applications. Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004) (http://doi.ieeecomputersociety.org/10.1109/CSB.2004.1332493)

Zhao Y, Pan W (2003) Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. Bioinformatics 19: 1046-1054

Appendix

Table 4.1. A comparison of the number of significant genes obtained from ANOVA and rank products methods of analysis

Dataset	Microarray platform	Method of Analysis					
		ANOVA			RANKPROD		
		Number of genes up-regulated with a q-value of ≤ 0.10	Number of genes down-regulated with a q-value of ≤ 0.10	Total number of significantly differentially regulated genes	Number of genes up-regulated with a q-value of ≤ 0.10	Number of genes down-regulated with a q-value of ≤ 0.10	Total number of significantly differentially regulated genes
RDX_2C	Oligonucl-eotide spotted	1	1	2	458	473	931
RDX_AFFY	Affymetrix	0	0	0	103	114	217
TNT_AFFY	Affymetrix	0	0	0	124	173	297
Mouse_FAT	Affymetrix	0	0	0	28	39	67
Mouse_LIVER	Affymetrix	0	0	0	315	53	368

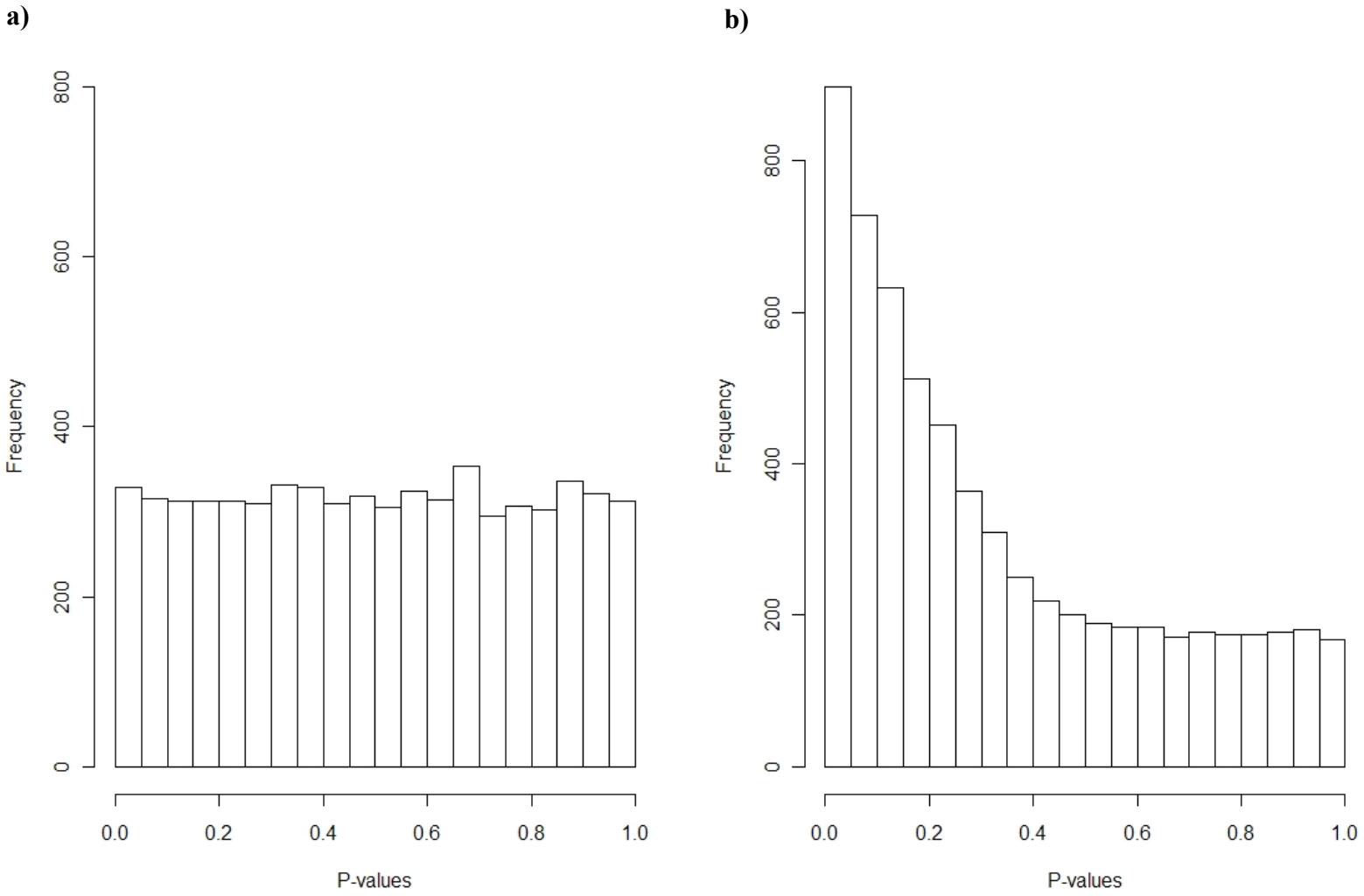

Figure 4.1. P-value density histograms under null and alternative hypotheses. a) an example of density histogram of p-values under null hypothesis b) an example of density histogram of p-values under alternative hypothesis

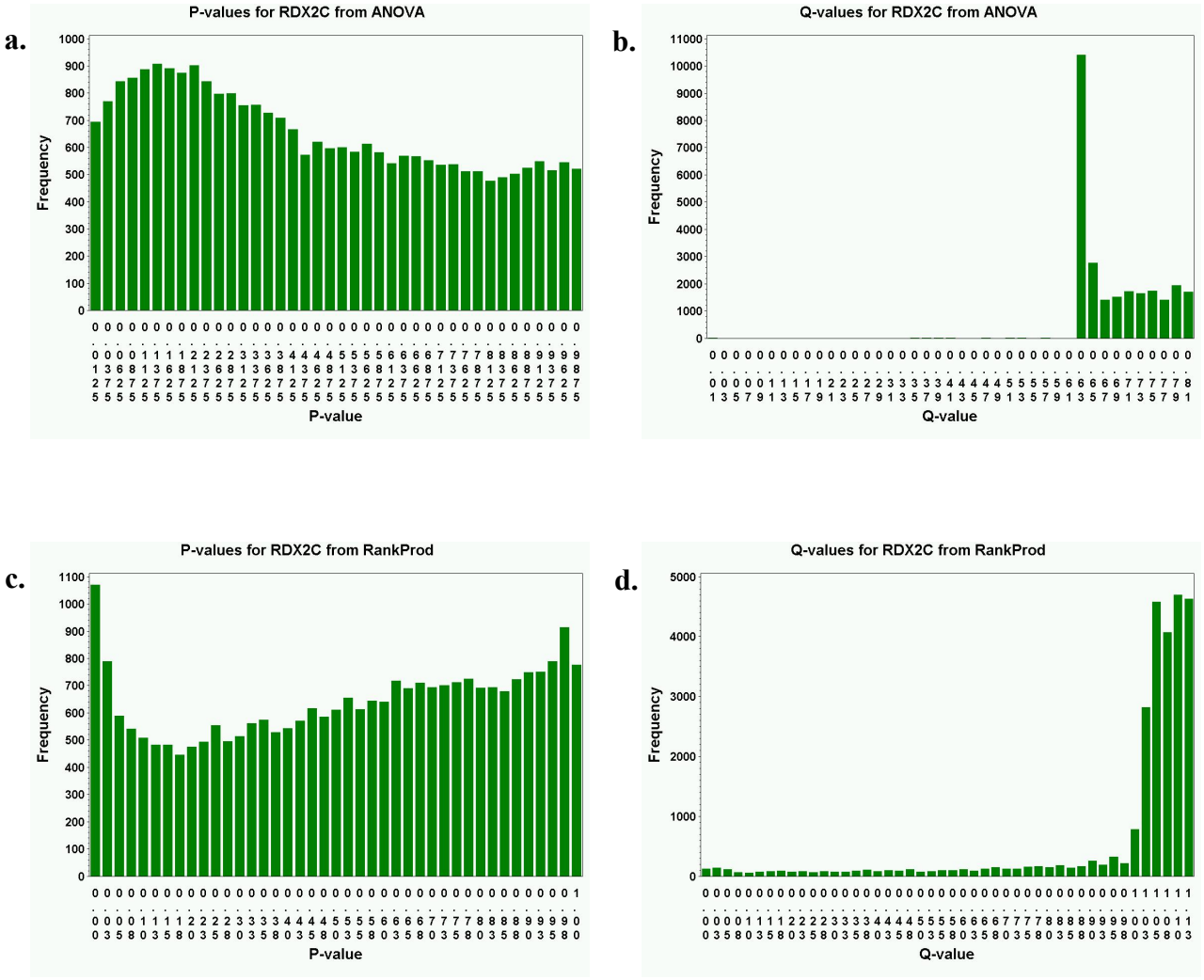

Figure 4.2. P-value and q-value density histograms for the compound RDX (Royal Demolition Explosive) treated two-color spotted Arabidopsis microarray dataset. a) distribution of p-values obtained from ANOVA analysis b) distribution of q-values obtained from ANOVA analysis c) distribution of p-values obtained from rank product analysis d) distribution of q-values obtained from rank product analysis

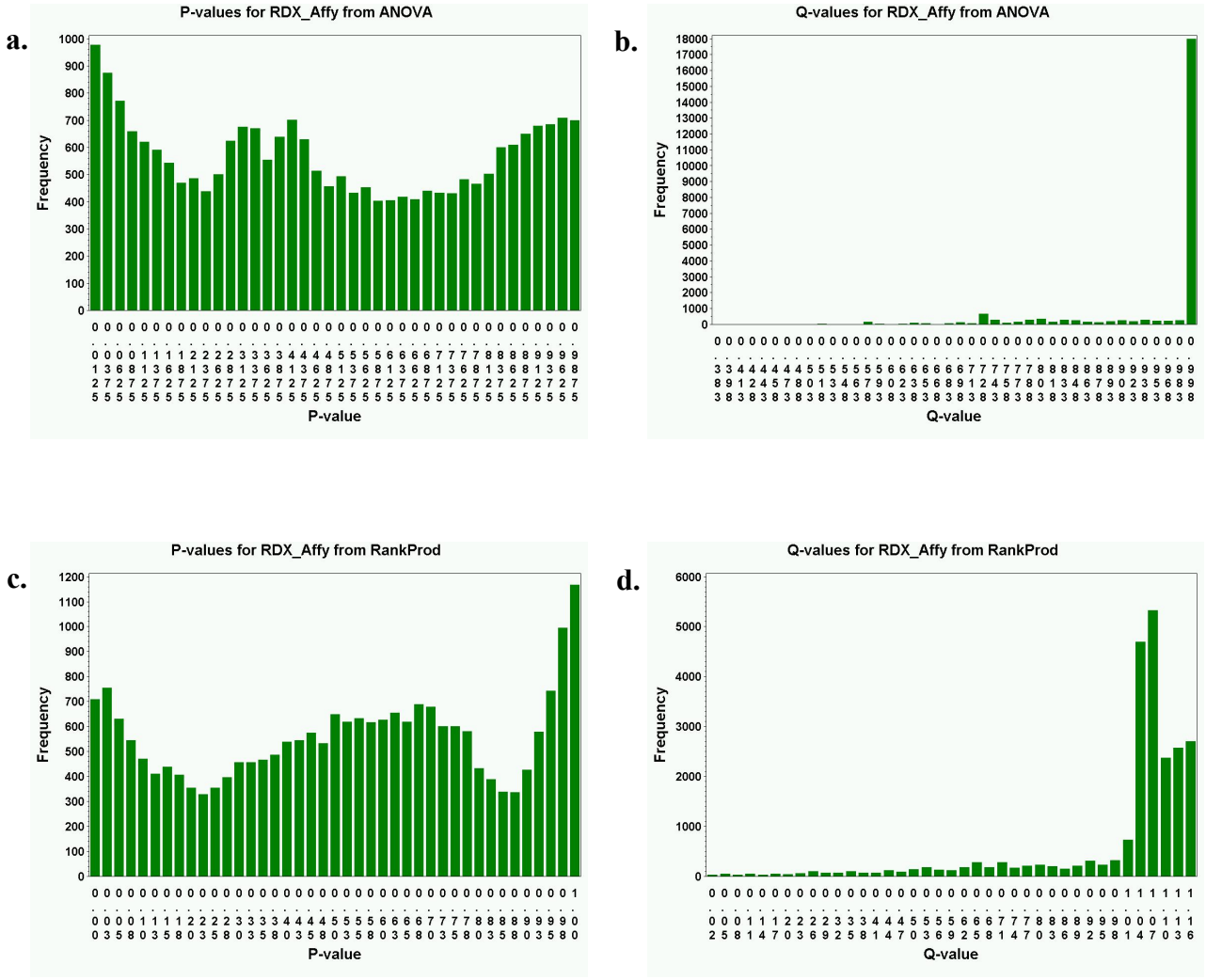

Figure 4.3. P-value and q-value density histograms for the compound RDX (Royal Demolition Explosive) treated Arabidopsis Affymetrix microarray dataset. a) distribution of p-values obtained from ANOVA analysis b) distribution of q-values obtained from ANOVA analysis c) distribution of p-values obtained from rank product analysis d) distribution of q-values obtained from rank product analysis

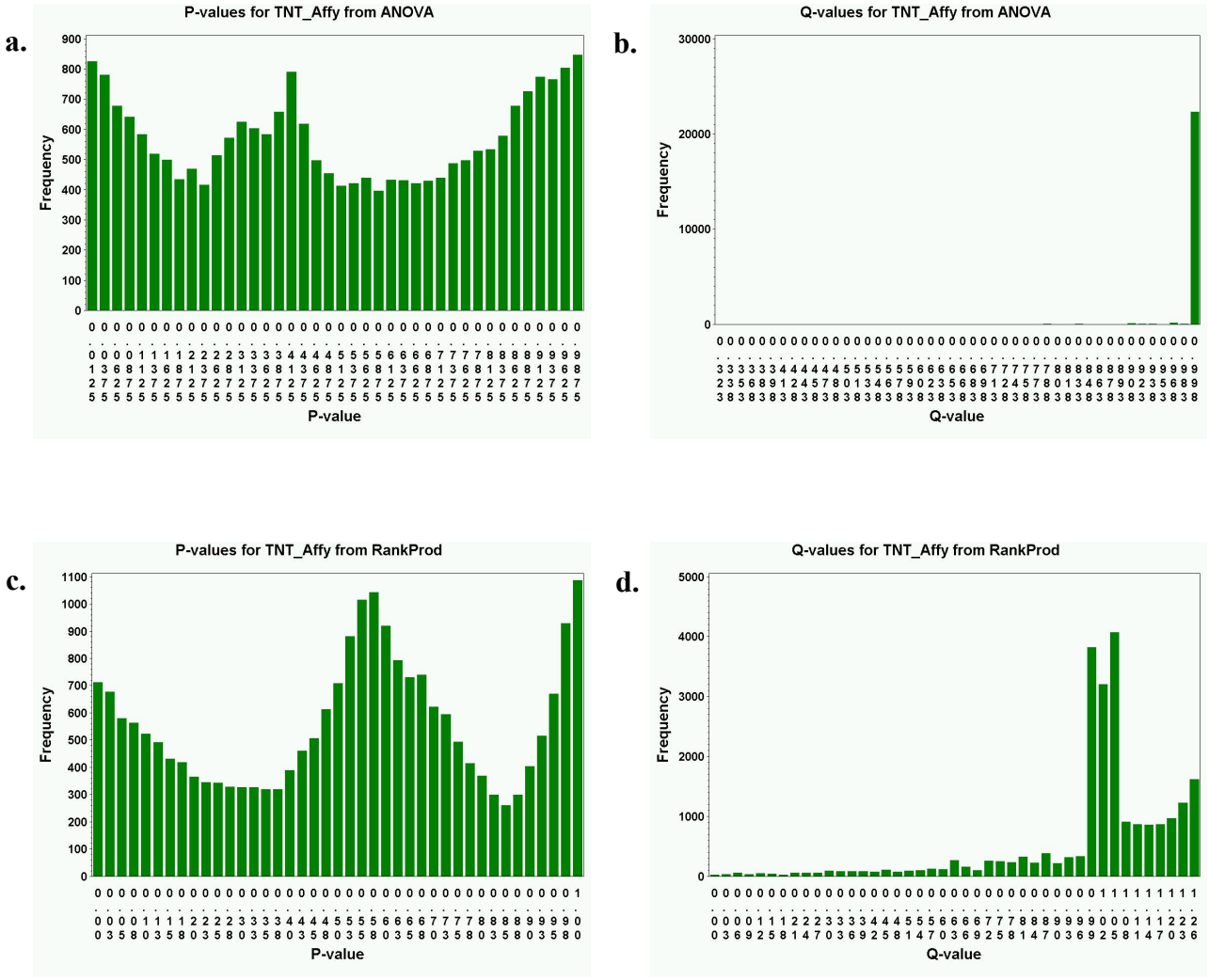

Figure 4.4. P-value and q-value density histograms for the compound TNT (2,4,6 – trinitrotoluene) treated Arabidopsis Affymetrix microarray dataset. a) distribution of p-values obtained from ANOVA analysis b) distribution of q-values obtained from ANOVA analysis c) distribution of p-values obtained from rank product analysis d) distribution of q-values obtained from rank product analysis

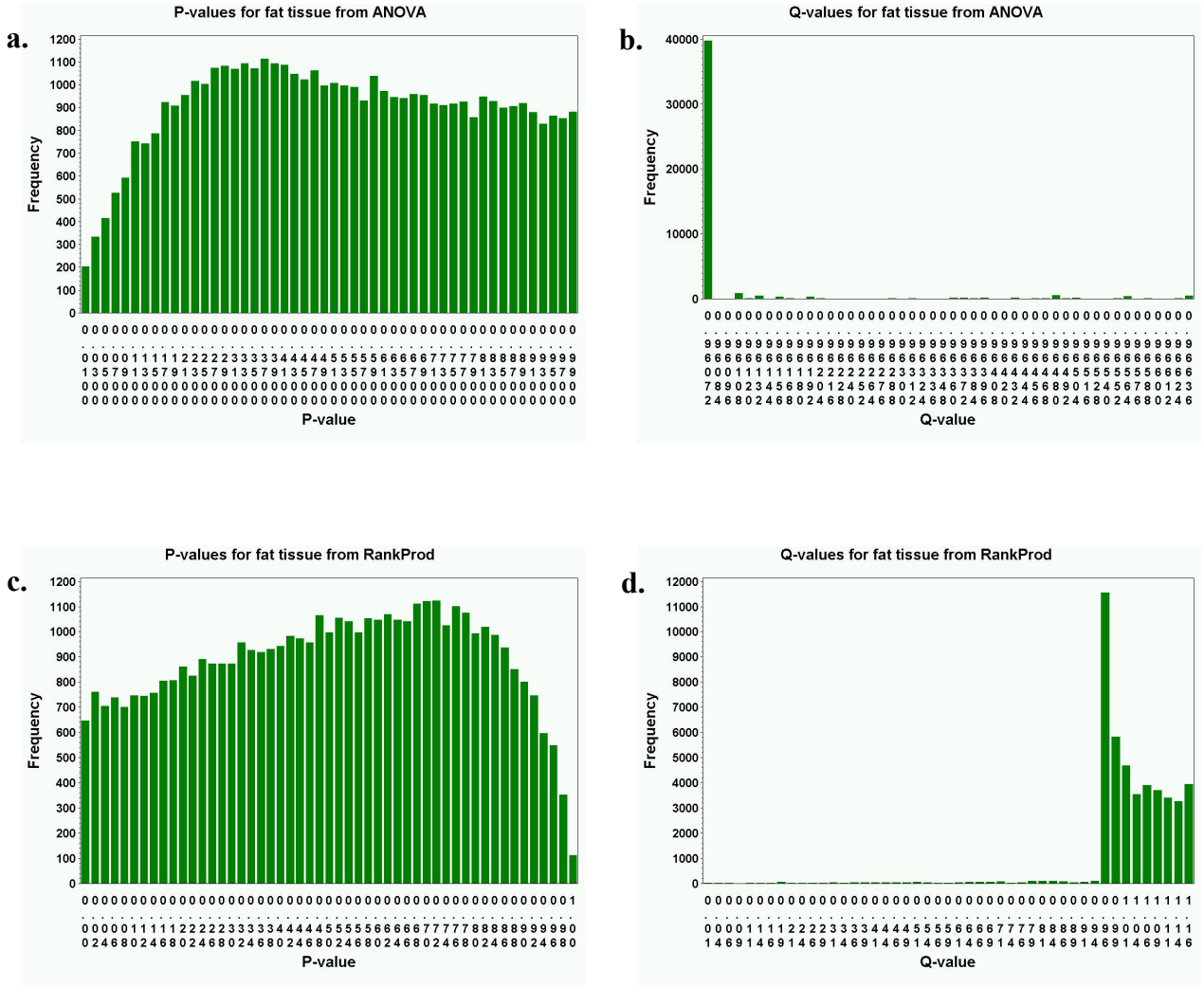

Figure 4.5. P-value and q-value density histograms for the mouse fat tissue Affymetrix microarray dataset. a) distribution of p-values obtained from ANOVA analysis b) distribution of q-values obtained from ANOVA analysis c) distribution of p-values obtained from rank product analysis d) distribution of q-values obtained from rank product analysis

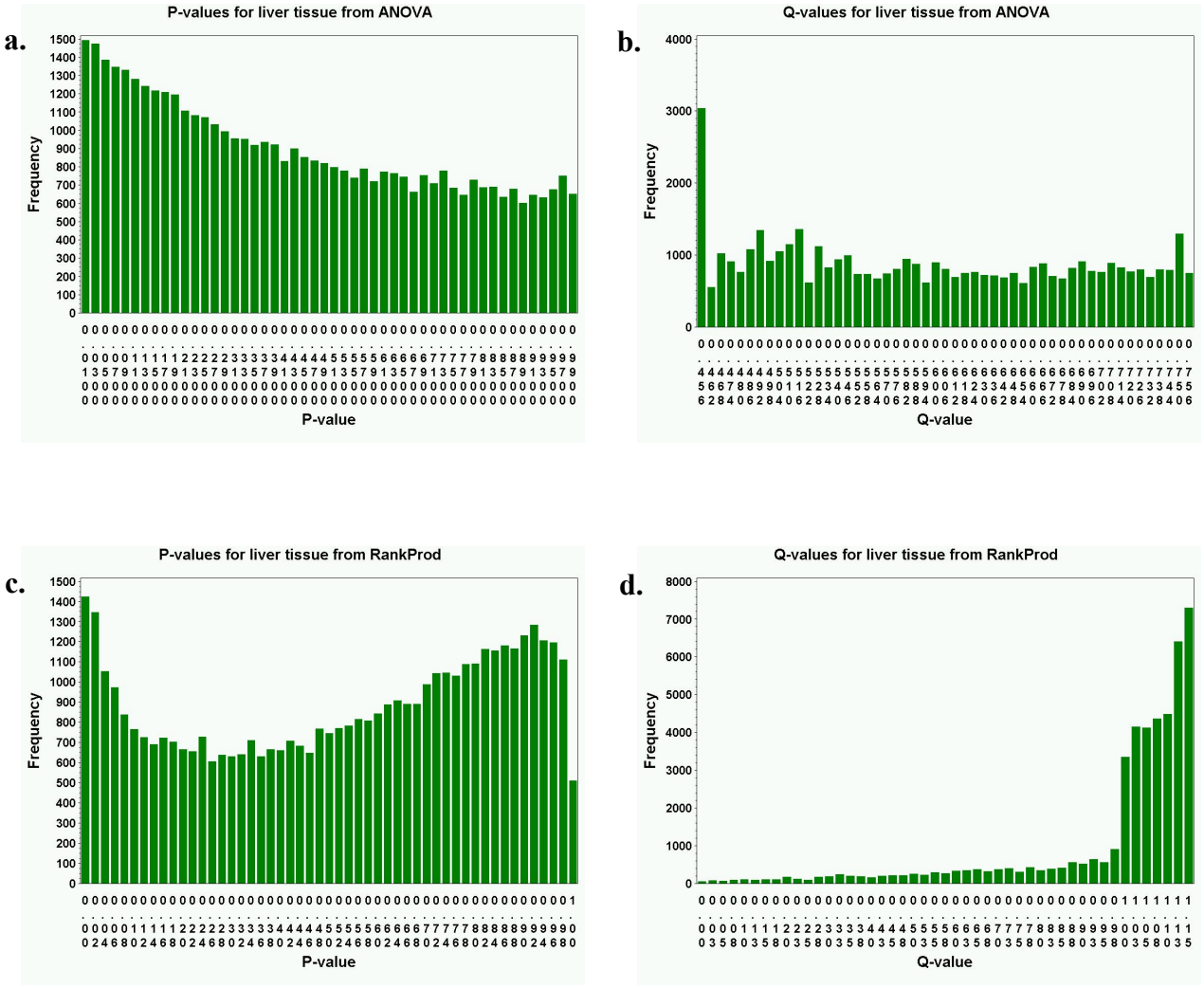

Figure 4.6. P-value and q-value density histograms for the mouse liver tissue Affymetrix microarray dataset. a) distribution of p-values obtained from ANOVA analysis b) distribution of q-values obtained from ANOVA analysis c) distribution of p-values obtained from rank product analysis d) distribution of q-values obtained from rank product analysis

Conclusion

This section briefly summarizes the contributions from each of the three studies. The first study involving microarray analysis of Arabidopsis plants exposed to explosives provided a list of candidate genes applicable in phytoremediation of explosives. And also the promoters from these candidate genes can be used in the making of phytosensors for explosives. The results from the second study indicate that the FLP/*FRT* site-specific recombination system seems promising for amplifying inducible fluorescence from phytosensors. Finally, the last study suggests the use of non-parametric rank products method for analyzing microarray data, when other methods prove to be very conservative.

Vita

Murali Raghavendra Rao was born in Bangalore, Karnataka, India on May 18th 1975. He grew up in Bangalore and graduated from Sri Aurobindo Vidya Mandir High School, Bangalore in 1990. He then pursued his pre-university studies at K.L.E College, Bangalore and graduated in 1992. From there, he went to the University of Agricultural Sciences, Bangalore to receive his Bachelor of Science in Agriculture in 1996. He further continued his studies at the University of Agricultural Sciences, Bangalore and received his Master of Science in Agriculture in 1999. He then worked for a biotechnology company in Bangalore before coming to Knoxville, TN in 2003 to pursue his doctoral studies at the University of Tennessee, Knoxville.