Recent Advances In Tobacco Science

Volume 36

Tobacco Research in the Era of Biotechnology and Genomics

Symposium Proceedings
64th Meeting
TOBACCO SCIENCE RESEARCH CONFERENCE

October 3-6, 2010
Hilton Head Island, South Carolina USA

Symposium of the 64th Tobacco Science Research Conference

Tobacco Research in the Era of Biotechnology and Genomics

– Symposium Chair –
Balazs Siminszky

– Editors –
Ray Robertson
Edward Robinson
Balazs Siminszky

CONTENTS

Symposium of the
64th Tobacco Science Research Conference

Tobacco Research in the Era of Biotechnology
and Genomics

– CONTRIBUTORS –

Marta Bokowiec

Joe Chappell

Ralph Dewey

Wanda Fields

Satrio Husodo

Scott Kinison

Ramsey Lewis

Tom Nieuhaus

Paul Rushton

Michael Timko

Robert Williams

Shuiqin Wu

Hongbo Zhang

PREFACE

The Program Editorial Committee of the 64th Tobacco Science Research Conference is pleased to present the 36th volume of Recent Advances in Tobacco Science publication. Each year the Program Editorial Committee of the Conference selects a theme that highlights a scientific or regulatory issue that is relevant, engaging and thought-provoking for the tobacco community. To review the recent advances made in molecular biology and its related disciplines and to assess the impact of these developments on the research of the tobacco plant and in the toxicological testing of tobacco products, the main topic chosen for this symposium is "Tobacco Research in the Era of Biotechnology and Genomics". Four speakers were invited to share their knowledge and discuss the latest developments in these subjects, and this publication contains the synopses of the symposium presentations and some introductory remarks that include a brief biographical sketch of the symposium speakers. Members of the Program Editorial, Ray Robertson, Ed Robinson and I wish to express our sincere appreciation to the speakers: Drs. Joe Chappell, Wanda Fields, Ramsey Lewis, Mike Timko and their colleagues for the significant time and effort spent preparing the publications and presentations. Information on ordering additional copies of this booklet, as well as previous volumes, is included at the back of this book.

Balazs Siminszky, Chair
Program Editorial Committee
64th Tobacco Science Research Conference

INTRODUCTION TO THE SYMPOSIUM

Balazs Siminszky

Philip Morris International
Neuchatel, Switzerland

The application of molecular biology and genomics has made a profound impact in the advancement of biological knowledge and tobacco-related sciences are no exceptions to this trend. Several important discoveries have been recently made in tobacco harm reduction, toxicology, agronomic performance and in the alternative use of tobacco with the application of molecular techniques. The symposium of the 64th TSRC was organized to provide an overview of the most important achievements and the directions of ongoing research in these fields.

We are fortunate to have four prominent scientists as symposium speakers who are actively involved in plant and human molecular biology, genomics, toxicology and some of the related disciplines.

Dr. Ramsey Lewis will review the history and current goals and challenges of traditional tobacco breeding, genomics, and biotechnology in the area of harm reduction. He will focus on research efforts to alter the alkaloid content and lower the levels of some of the harmful constituents in tobacco. He will also discuss his research in developing ultra low NNN tobacco varieties.

Ramsey Lewis, an Associate Professor at North Carolina State University, has received his BS at the University of Nebraska, and his MS and Ph.D. degrees at North Carolina State University. A plant breeder and molecular geneticist, Dr. Lewis leads an internationally renown research program to develop tobacco cultivars that possess desirable characteristics for the U.S. growers and the tobacco industry. His research also aims at increasing our understanding about the genetic control of economically important traits that benefit a large number of plant biologists. He has been serving on the tobacco and plant breeding community in a number of capacities such as the member of the Flue-Cured Tobacco Variety Standards Committee, the North Carolina Seed Board, the Flue-Cured Tobacco Variety Evaluation Committee, the Regional Burley Tobacco Variety Evaluation/ Standards Subcommittee and the N.C. State University Breeder's Release Board. He is actively involved in undergraduate and graduate student education and has authored more than 20 scientific publications.

Dr. Michael P. Timko will discuss his research on developing reduced harm tobacco by defining the signaling components and transcription factors

controlling the formation of nicotine and various other minor alkaloids in tobacco. He will explain how his laboratory is using genomic, bioinformatic, and transgenic technologies to unravel the cellular regulation of alkaloid formation in response to various developmental, phytohormonal, and environmental cues.

Michael P. Timko is currently a Professor of Biology at the University of Virginia and Director of the Distinguished Major Program in Human Biology. He received a BS from the Rutgers College of Agriculture and Environmental Science in 1975, studied genetics and plant breeding at Michigan State University, and earned his PhD from Rutgers University in 1980. He held postdoctoral positions at Brandeis University and Rockefeller University, where he was involved in some of the seminal studies of gene expression in plants and the construction of transgenic crops. He joined the faculty at the University of Virginia in 1986. He has authored or co-authored over 100 research papers, book chapters and review articles and holds multiple US and world-wide patents in agricultural and nutritional biotechnology. In 2009 he was recipient of the Hartwell Foundation Individual Biomedical Research Award for his work on probiotic-based therapeutics. He was recently appointed Guest Professor at Zhejiang University in China, and Visiting Professor at the Joint Sino-US Tobacco Molecular Breeding Laboratory at the Yunnan Academy of Tobacco Agricultural Science.

Dr. Joseph Chappell will describe his research in using tobacco for the production of long, branched-chain hydrocarbons, compounds that are amenable to the conversion into biofuels. He will highlight various experimental strategies designed to maximize the plants' capacity to produce and accumulate these value-added products.

Dr. Chappell has been on the faculty at the University of Kentucky since April 1985, where he has developed an internationally recognized research program pioneering the molecular genetics and biochemistry of natural products in plants. Dr. Chappell earned his B.A. degree in Biology from the University of California, San Diego in 1977, his Ph.D. in Biology in 1981 from the University of California, Santa Cruz, and pursued postdoctoral studies at the University of Freiburg, Germany and the Max Planck Institute – Cologne, Germany. At the University of Kentucky, Dr. Chappell's research has focused on the mechanisms plants use to defend themselves against microbial pathogens, especially the biosynthesis of anti-microbial terpene-type compounds, as well as natural product biosynthesis and production in plants. Recent work from his laboratory has been published (approximately 70 publications) in Plant Physiology, Journal of Biological Chemistry, Archives of Biochemistry & Biophysics, Proceedings of the National Academy of Sciences, USA and Nature Biotechnology. He currently serves on the editorial boards of the Journal of Biological Chemistry and Plant Physiology.

Dr. Wanda R. Fields, R. J. Reynolds Tobacco, Co., will provide an overview on the application of genomics, transcriptomics, proteomics and metabolomics to assess the toxicological responses of chemicals using *in vitro* and *in vivo* models. She will discuss the role of molecular biology and "omics" technology in biomarker evaluation and marker discovery in the toxicological testing of tobacco products.

Dr. Fields is a Senior Staff Scientist at R. J. Reynolds Tobacco Company in Winston-Salem, N.C. She has 13 years of Research and Development experience with R. J. Reynolds Tobacco Company. Dr. Fields earned her bachelors degree in Biochemistry at N.C. State and her Ph.D. in Biochemistry at Wake Forest University. She received a Leon Golberg Toxicology Postdoctoral Fellowship through R.J. Reynolds' postdoctoral training program through Duke University before joining R.J. Reynolds as a regular employee in 1997. Dr. Fields' work at R&D is concerned with using molecular genetic techniques to understand lung response to tobacco smoke components and genetic predisposition to lung disease. She focuses on modifications to DNA and RNA, specifically key genes involved in the development of cancer and other lung diseases. She is a member of the Genetics and Environmental Mutagenesis Society and the American Association for Cancer Research, and has authored or co-authored 43 external publications and presentations. In 2006, she received an appointment as an Adjunct Assistant Professor for the Department of Physiology and Pharmacology of Wake Forest University School of Medicine and has also been appointed as an Adjunct Assistant Professor with the Integrated Toxicology Program of Duke University, through which she supports R.J. Reynolds' postdoctoral training programs. She was also a *2007 Minorities in Research Science Emerald Honor's Award Recipient* of the "Community Outreach" award from Science Spectrum Magazine in recognition of her hard work and dedication to helping young people with an interest in science, especially those at risk, to move toward careers in science and achieve their goals.

The Editorial Committee wishes that these symposium presentations will stimulate your interest and provide new insights into the latest molecular technologies used in modern tobacco science and the transformational impact of these methodologies in various aspects of tobacco and cigarette research.

Balazs Siminszky
64th TSRC Symposium Chair

GENETICS-BASED MODIFICATION OF TOBACCO CHEMISTRY IN A REGULATORY ENVIRONMENT

Ramsey S. Lewis and Ralph E. Dewey
Crop Science Department
North Carolina State University, Raleigh, NC

Introduction

Tobacco products contain a number of constituents classified as carcinogens by the International Agency for Research on Cancer (IARC) (Hecht, 1998, 2003; National Toxicology Program, 2005). Some of these compounds are found in cured tobacco leaves, the primary raw ingredient of tobacco products. The majority of the most potent carcinogens, however, are generated by pyrolysis, the most prevalent method of consumption. To date, over sixty compounds found in cigarette smoke and up to twenty-five contained in unburned tobacco products have been associated with carcinogenesis in laboratory animals (Hoffman and Djordjevic, 1997; Hecht, 2003, 2006). Although avoidance/cessation remains the best way to avert health risks associated with tobacco use, potential reduced-exposure products (PREPs) might be used within complementary strategies to reduce tobacco-related harm at the population level (Stratton *et al*., 2001; Gartner *et al*., 2007). Good product stewardship would dictate that tobacco manufacturers venture to reduce or eliminate harmful constituents of their products.

During 2009, the "Family Smoking Prevention and Tobacco Control Act" was signed into law in the United States. This legislation gives the U.S. Food and Drug Administration (FDA) regulatory authority to mandate the lowering or elimination of certain chemical constituents from tobacco products marketed in the U.S. Specific possible modifications will be the subject of discussion as the FDA begins its regulatory role during 2010. A list of toxicants suggested for lowering in cigarette smoke has already been outlined by the IARC and the World Health Organization (WHO) (Burns *et al*., 2008), and includes tobacco specific nitrosamines (TSNAs), acetaldehyde, acrolein, benzene, benzo[a] pyrene, 1,3-butadiene, carbon monoxide, and formaldehyde. Nicotine *per se* is not viewed as a carcinogen, but some groups have also suggested regulation of this compound as part of a tobacco control strategy (Russell, 2000; Benowitz *et al*., 2007).

Manufacturers will need to determine the best combination of methods to modify their products for various chemistries to be in compliance with upcoming regulations. Levels of some constituents may be altered through agronomic practices, growing conditions, curing methods, and product design. Improved genetics has contributed greatly to improvements in yields and disease resistance

of tobacco varieties over the last eighty years (Bowman *et al.*, 1984; Wernsman and Rufty, 1987). What role might tobacco plant genetics play in lowering levels of toxicants in tobacco products? The objectives of this paper are to describe, with examples, how traditional plant breeding, genomics, and biotechnology might be used to develop products with 'reduced harm' characteristics and help comply with future regulations. Some methods require knowledge of DNA sequence information and/or foreign gene transfer to *N. tabacum* via genetic engineering. Others exploit naturally occurring genetic variation with no understanding of specific gene-trait associations. Seamless integration of all approaches is needed to best capitalize on opportunities.

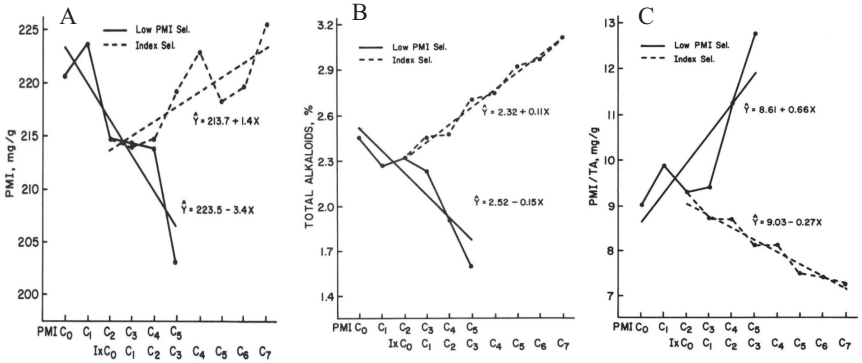
Some Previous History

The prospect of using plant genetics to reduce levels of harmful constituents in tobacco products is not new. A prerequisite for conventional selection designed to affect levels of a chemical constituent in *N. tabacum* or its smoke is heritable variation within the species or its close relatives. Abundant phenotypic variation exists among *N. tabacum* accessions for a range of morphological traits and disease resistance characteristics. Quantitative and qualitative genetic variation is also present for the potential of tobacco leaves to accumulate a number of natural products, or to produce toxic substances upon pyrolysis. For example, Rathkamp *et al.* (1973) observed differences in mainstream smoke from four flue-cured tobacco cultivars for total particulate matter (TPM), CO, CO_2, acetaldehyde, acrolein, gas phase hydrogen cyanide, phenols, benzo[a]pyrene (BaP), and benz[a]anthracene. In addition, Julio *et al.* (2006) found smoke produced by parental lines of a mapping population to be significantly different for TPM, tar, CO, and BaP contents. Transgressive segregation was observed amongst derived lines for CO, tar, and total particulate matter. This leads to the suggestion that levels of certain health-related constituents might be genetically manipulated.

Initial genetic selection investigations with respect to harm reduction were directed towards reducing TPM, defined as solid materials trapped on a Cambridge filter pad after smoking under specified conditions. TPM is related to tar in that tar is the weight of TPM minus nicotine and water. Several papers reported significant differences among flue-cured tobacco cultivars for TPM when leaves were subjected to combustion (Tso and Chaplin, 1977; Wernsman *et al.*, 1977; Chaplin and Spurr, 1982). These apparent genetic differences prompted Matzinger *et al.* (1978) to explore the potential of using recurrent mass selection to reduce particulate matter index (PMI), an efficient predictor of TPM in cigarette smoke. Recurrent selection methods are designed to increase the frequency of favorable alleles affecting the trait(s) of interest in a population under selection. Five cycles of selection for lower PMI in a synthetic flue-cured tobacco population reduced PMI by an average 1.5% per cycle (Matzinger *et al.*, 1984) (Figure 1). A correlated decrease in total alkaloids (TA) of 6.1% per cycle

was also observed, however. The more rapid decrease for percent total alkaloids relative to PMI led to a 7.6% average per cycle increase of the PMI/TA ratio, a potentially undesirable outcome from a harm reduction point of view. A selection index was later applied in the same population to select against the correlation between PMI and TA. The objective was to decrease PMI while maintaining TA at the mean of the population from which selection was initiated. After seven cycles of restricted index selection, the PMI/TA ratio was reduced by an average of 3% per cycle (Matzinger *et al.*, 1984) (Figure 1). This work did not result in commercial varieties, but demonstrated the effect that genetic selection in the absence of specific DNA information can have on levels of declared toxicants in tobacco smoke.

Figure 1. Responses to selection for lower particulate matter index (PMI) (indicated by solid lines) and a selection index (indicated by dashed lines) designed to lower PMI while keeping % total alkaloids constant in a flue-cured tobacco synthetic population. Results are presented for (A) PMI, (B) % total alkaloids, and (C) PMI/TA ratio. Results are from Matzinger et al. (1984).



Modern Genetic Information

The above-mentioned 'conventional' plant breeding approach led to directed change in the potential of flue-cured tobacco leaves to produce TPM upon pyrolysis. This was accomplished in the absence of gene information, except for knowledge that tobacco genes reside on one of twenty-four linkage groups. Technologies that utilize DNA sequence information can now be combined with traditional plant breeding methodologies to increase the potential for impacting tobacco product chemistry and human health. The complete plastid and mitochondrial sequences have been published for *N. tabacum* and several related species (Shinozake *et al.*, 1986; Sugiyama *et al.*, 2005; Yukawa *et al.*, 2006), and a much larger amount of nuclear DNA sequence information has recently been generated. The largest effort was that of the Tobacco Genome

Initiative (TGI) carried out by researchers at N.C. State University (Opperman *et al.*, 2007). This project sequenced > 85,000 EST sequences for *N. tabacum* and > 38,000 ESTs from *N. benthamiana* using mRNA collected from an array of different tissues from both species. A second major objective of the TGI was sequencing of methyl-filtered *N. tabacum* genomic clones in a strategy designed to selectively sequence gene-rich regions. Results suggested a nearly 10-fold increase in gene discovery in methyl-filtered vs. non-filtered libraries (Gadani *et al.*, 2003). More than 1.4 million genomic sequence reads in addition to the EST data are now available in public databases. A second major tobacco sequencing effort was that of the 'European Sequencing of Tobacco' (ESTobacco) project carried out as a collaborative arrangement between Advanced Technologies LTD (Cambridge) and the Institute du Tabac de Bergerac (Altadis) (Dorlhac de Borne *et al.*, 2006). More than 46,000 EST sequences from this effort have been made publicly available. It should be pointed out that, with the advent of more modern DNA sequencing platforms, it is now even much more cost-efficient to quickly generate large amounts of sequence information, and to conduct transcript comparisons between different types or lines of tobacco. With large amounts of sequence information for *N. tabacum* now available, it is relatively straightforward to identify tobacco sequences that share sequence identity to those characterized in other species (*Arabidopsis thaliana*, rice, and tomato, for example) where greater amounts of genomics research are conducted. Because tobacco is easily transformed using a number of different methods (Birch, 1997), and also because some *Nicotiana* species are very amenable to virus induced gene silencing techniques (Baulcombe, 1999), it is not difficult to investigate gene function for this species.

To date, genetic mapping research in *Nicotiana* has been limited compared to other major crops. This is mostly due to relatively low amounts of DNA polymorphism that were revealed among genotypes of *N. tabacum* using early marker systems such as RFLPs, RAPDs, or AFLPs (Brandle and Bai, 1999; del Piano *et al.*, 2000; Ren and Timko, 2001; Rossi *et al.*, 2001). The situation has greatly improved, however, with the development of a large number of microsatellite markers partially based on the above-mentioned DNA sequence information. These codominant markers are highly reproducible, amplify only one or few bands, exhibit greater levels of polymorphism, and are transferable across multiple populations. The most complete genetic linkage map published to date for *N. tabacum* is that of Bindler *et al.* (2007) and is based on approximately 300 microsatellite markers. It is expected that maps with much greater marker density will be published soon. Mapping experiments have thus far been useful for locating genomic regions affecting quantitative levels of disease resistance (Nishi *et al.*, 2003; Vontimitta, 2010) and genes affecting leaf surface chemistry (Vontimitta *et al.*, 2010). Julio *et al.* (2006) used DNA markers to identify tobacco genomic regions putatively associated with TPM, tar, carbon monoxide, and benzo[a]pyrene in smoke.

TSNA Reduction: Combined Application of Genomics, Biotechnology, & Plant Breeding

Tobacco specific nitrosamines (TSNAs) are a class of compounds that have been recommended for regulation by the IARC and WHO (Burns *et al.*, 2008), and it is likely that the FDA will set maximum tolerable limits for these constituents in U.S. products. Research into development of genetic technologies for reducing TSNAs in tobacco highlights the combined use of traditional and cutting edge methods to influence levels of a toxicant in tobacco products. Activities in this area also point to the issue of technology deployment in the face of industry attitudes towards genetic engineering.

TSNAs are most prevalent in air-cured tobacco types, and are generated through the nitrosation of naturally-occurring pyridine alkaloids during the curing and storage of tobacco (Bush *et al.*, 2001). The TSNAs *N*-nitrosoanatabine (NAT) and *N*-nitrosoanabasine (NAB) are formed via nitrosation of anatabine and anabasine, respectively, and are reported to have little to no biological activity (Hecht and Hoffmann, 1989; Hoffman *et al.*, 1994; Hecht, 1998). In contrast, *N*-nitrosonornicotine (NNN) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) derived from nornicotine and nicotine, respectively, represent IARC Group I carcinogens (Hecht, 1998, 2003; Hoffmann *et al.*, 1994). NNN and NNK are present in tobacco smoke, and are reported to be the most important carcinogens in smokeless tobacco products (Hoffmann and Djordjevic, 1997). The impact on overall toxicity that might be created by reducing or eliminating these compounds in smoke is unknown. It is expected, however, that the reduction of these compounds in smokeless tobacco would have significant influence on lowering health risks associated with the use of these products.

To reduce or eliminate TSNAs, one could: (1) reduce or eliminate alkaloid precursors, (2) reduce availability of nitrosating agents, or (3) cure and store tobacco under conditions not conducive for TSNA formation. Genetics can be used to influence the first and second possibilities. In air-cured tobaccos, most effort to date has focused on reduction of NNN through reduction of its nornicotine precursor.

Nornicotine is produced almost entirely via the oxidative *N*-demethylation of nicotine by nicotine *N*-demethylase enzymes (Figure 2). One major (*CYP82E4*) and two minor nicotine demethylase genes (*CYP82E5v2* and *CYP82E10*) have been identified based on gene expression studies and/or use of gene sequence databases (Siminszky *et al.*, 2005; Gavilano and Siminszky, 2007; Lewis *et al.*, 2010). Knowledge of sequence information for *CYP82E4* permitted the development and testing of transgenic tobacco lines possessing an RNA interference (RNAi) construct designed to dramatically reduce or stop expression of this gene (Gavilano *et al.*, 2006). Although targeted specifically to *CYP82E4*,

the two minor nicotine demethylase genes were likely silenced as well, given that these genes all share over 90% DNA sequence identity. The effect is dependent upon genetic background, but selected lines exhibited a six-fold reduction in nornicotine content compared to untransformed controls, and analysis of cured leaves revealed commensurate reductions in NNN (Lewis *et al.*, 2008; Figure 3).

Figure 2. Primary biochemical mechanism for formation of NNN in tobacco.

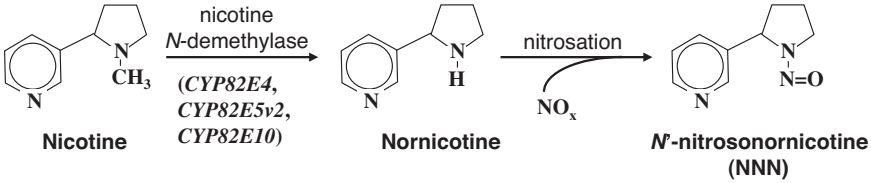
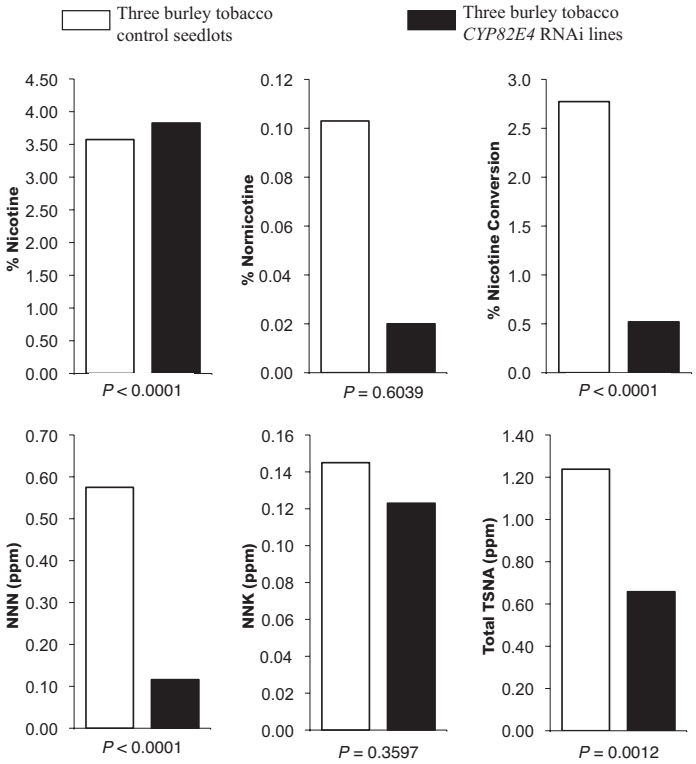
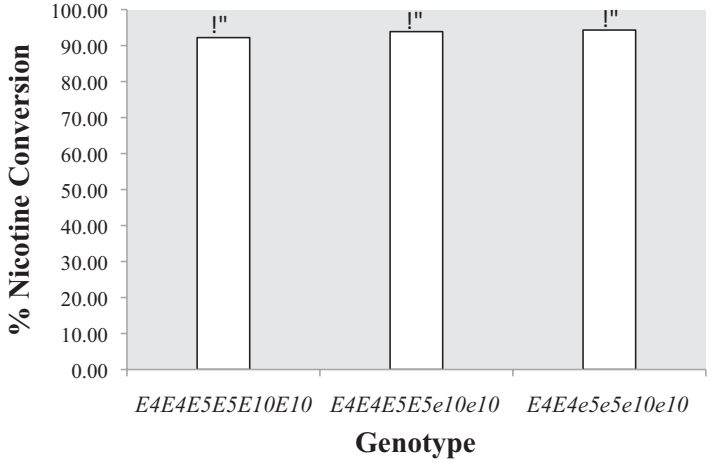


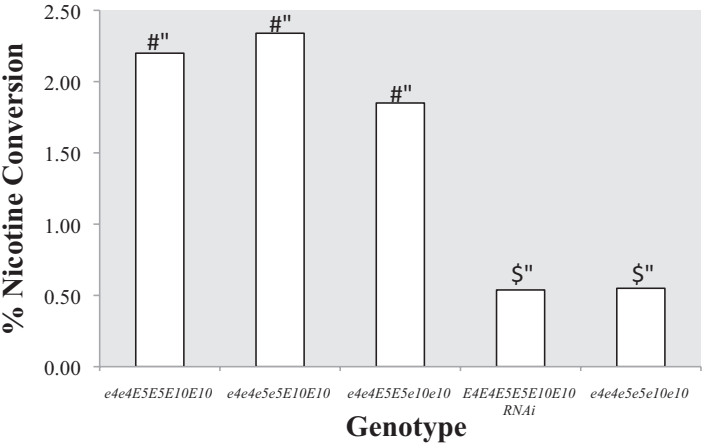
Figure 3. Burley tobacco non-transformed controls (mean of three independently-derived seed lots) versus burley tobacco CYP82E4 RNAi lines (mean of three transgenic lines). P-values were generated from single degree of freedom contrasts of group means. Data are from Lewis et al. (2008).



The above-mentioned transgenic, or 'GMO,' technology was demonstrated to be a simple and effective means of significantly lowering the level of a key defined carcinogen present in tobacco products. From a plant breeding point of view, this method is attractive because a single transgene insertion can stop the expression of three independently segregating genes with probably little adverse effect on tobacco yields or cured leaf quality. The mainstream tobacco industry is not prepared to utilize cured leaf derived from GMO tobacco cultivars, however. We have since pursued a mutation breeding approach to suppress nicotine demethylase activity to levels equivalent to that achieved using the transgenic method. To accomplish this, seed was treated with a chemical mutagen, ethane methyl sulfonate (EMS), to introduce primarily single base pair changes across the entire tobacco genome. High throughput gene sequencing was then used to identify cases in which deleterious mutations were introduced into each of the three characterized nicotine demethylase genes (Lewis *et al.*, 2010). Mutations encoding for premature stop codons that prevent production of functional gene products were identified for *CYP82E4* and *CYP82E5v2*. A similar truncation mutation was not identified for *CYP82E10*, but several substitution mutations yielding the same null phenotypic effect (*i.e.* no measurable nicotine demethylase activity) were found for this gene as well. Different combinations of these mutations permitted production of lines with a range of different levels of nicotine conversion, and lines homozygous for mutations in all three genes exhibit very low levels of nicotine conversion that are comparable to what can be achieved using RNA interference (Lewis *et al.*, 2010; Figure 4). Materials derived from mutation breeding are accepted by the tobacco industry, and we are currently in the process of transferring these mutations to existing air-cured tobacco cultivars using conventional breeding methods. The outcome of this research will demonstrate the result of collaboration between molecular biologists involved in gene discovery research and plant breeders working to develop and commercialize improved tobacco cultivars carrying reduced harm characteristics.

Figure 4. Mean percent nicotine conversion for burley tobacco plants with varying mutant combinations at CYP82E4 (E4), CYP82E5v2 (E5), and CYP82E10 (E10) loci. Means with different letters are significantly different at the P < 0.05 level.





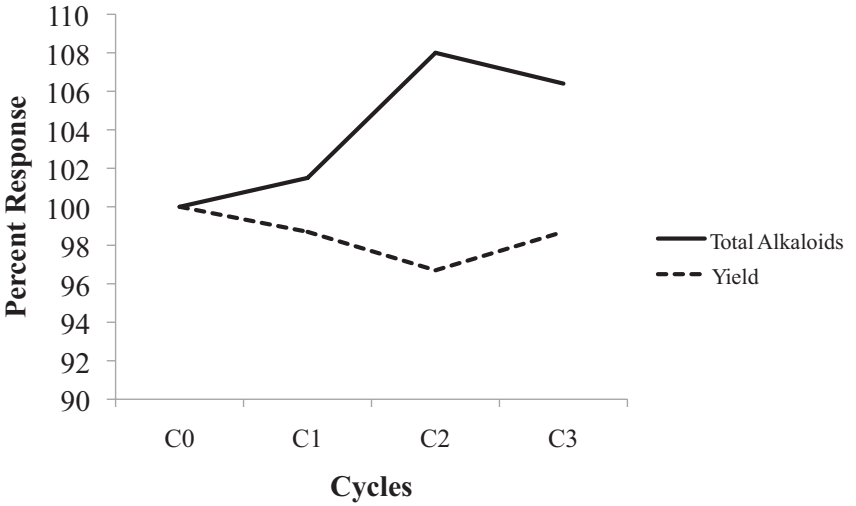
Genetics-Based Alteration of Tobacco Alkaloids

Nicotine is the major alkaloid in tobacco products, while the minor alkaloids include nornicotine, anabasine, and anatabine. Under the Family Smoking Prevention and Tobacco Control Act, the FDA has authority to regulate alkaloid levels in tobacco products, but cannot require the reduction of nicotine yields to zero. Legitimate arguments can be made regarding whether public health would be better served by increasing or decreasing nicotine levels in tobacco products.

Russell (2000) suggested that less harmful cigarettes might be those with lowered toxicant yields combined with either medium or increased nicotine yields. In contrast, others have argued for regulatory strategies where nicotine levels would be gradually reduced or immediately lowered to subaddictive threshold levels (Benowitz and Henningfield, 1994; Benowitz *et al.*, 2007). Even in the absence of directives on nicotine levels per se, this alkaloid will be an important component of future regulations because toxicant amounts will likely be reported per mg of nicotine in smoke. Specific alkaloid levels are also important because of their correlation with accumulation of corresponding TSNAs derived through nitrosation reactions. The possibility of future regulation warrants consideration of genetic options for modifying alkaloid levels in tobacco products.

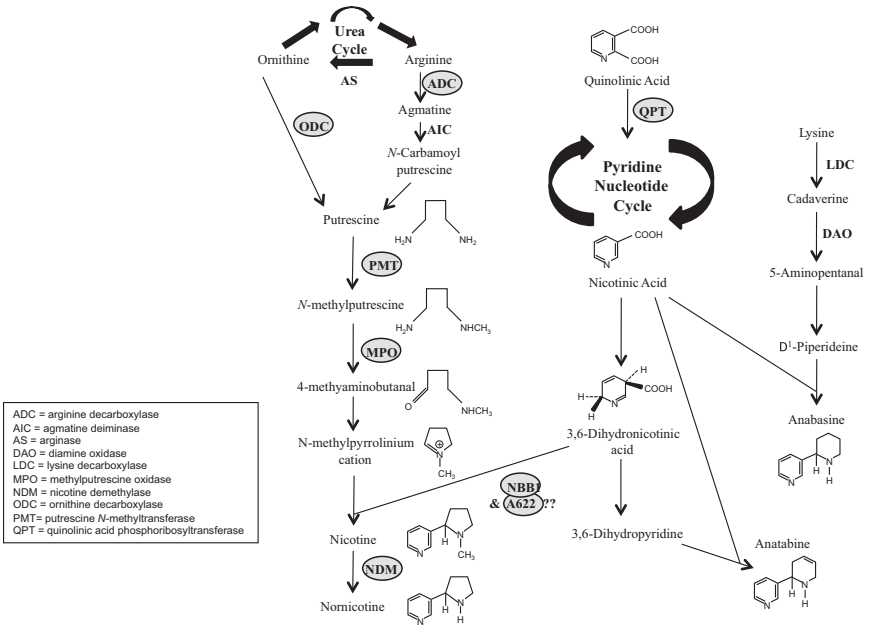
Alkaloid levels are highly influenced by the environment, cultural practices, and plant genetics. Genes at two loci, *Nic1* and *Nic2* (also called *A* and *B* loci), are known to have large effects and likely encode for transcription factors that globally regulate expression of a number of genes involved in alkaloid biosynthesis (Hibi *et al.*, 1994; Riechers and Timko, 1999; Reed and Jelesko, 2004). Recessive alleles at both of these loci can reduce alkaloid levels from between 1.5% and 4.5% to approximately 0.2% (Legg *et al.*, 1969; Legg and Collins, 1971; Chaplin and Weeks, 1976). An unknown number of additional minor genes also influence alkaloid levels as evidenced by the successful application of recurrent selection to continually increase nicotine levels in populations fixed for dominant alleles at the *Nic1* and *Nic2* loci (Figure 5) (Matzinger *et al.*, 1972, 1989). Interspecific genetic variability might also be used to affect alkaloid accumulation as Dr. James Chaplin worked to transfer genes from the high-nicotine species *N. rustica* to *N. tabacum* in an attempt to develop tobacco cultivars with elevated alkaloid contents as part of a strategy to contribute to the development of cigarettes with a lower tar/nicotine ratio (Chaplin and Spurr, 1982; Chaplin, 1987).

Figure 5. Percentage response from three cycles of index selection to increase total alkaloids and hold yield at the population mean in a 'McNair 13' x 'McNair 135' population. Data are from Matzinger et al. (1989).



Considerable knowledge has been gained in the last fifteen years regarding the biochemical pathways involved in tobacco alkaloid biosynthesis, and a number of participating genes have been characterized (Figure 6). In addition to the nicotine demethylase genes mentioned previously (Siminszky *et al.*, 2005; Gavilano and Siminszky, 2007; Lewis *et al.*, 2010), isolated genes directly involved in alkaloid synthesis include those encoding for ornithine decarboxylase (ODC) (Wang *et al.*, 2000), arginine decarboxlyase (ACD) (Wang *et al.*, 2000), putrescine *N*-methyltransferase (PMT) (Hibi *et al.*, 1994; Riechers and Timko, 1999), quinolinic acid phosphoribosyltransferase (QPT) (Sinclair *et al.*, 2000), and methylputrescine oxidase (MPO) (Heim *et al.* 2006; Katoh *et al.*, 2007). Genes designated as *A622* and *NBB1* have also been described (Hibi *et al.*, 1994; Hashimoto and Kato, 2007, Hashimoto and Kajikawa, 2008; Kajikawa *et al.*, 2009). The precise role of these two gene products in alkaloid synthesis is not currently known, but they may be involved in the final condensation reaction of nicotine formation (Shoji *et al.*, 2002; Hashimoto and Kato, 2007; Hashimoto and Kajikawa, 2008; DeBoer *et al.*, 2009; Kajikawa *et al.*, 2009) (Figure 6). Transcriptional regulators of alkaloid biosynthetic genes (De Sutter *et al.*, 2005; Todd *et al.*, 2010) and genes involved in sequestration of nicotine in leaf vacuoles have also been reported (Morita *et al.*, 2009; Shoji *et al.*, 2009).

Figure 6. Primary biosynthetic pathways for major alkaloids of tobacco mostly as outlined by Chintapakorn & Hamill (2003) and Cane et al. (2005). Enzymes for which genes have been characterized are circled and shaded in grey.



Gene sequence information can permit detection of naturally occurring allelic variants affecting alkaloid levels or allow identification of induced mutations in relevant genes. Gene information also allows modification of alkaloid profiles via genetic engineering. The first demonstrated use of gene information to alter alkaloid accumulation in *Nicotiana* was by Sato *et al.* (2001) who both overexpressed and silenced *PMT* in *N. sylvestris* to increase and decrease nicotine content, respectively. In addition, Xie *et al.* (2004) used silencing of QPT to generate the transgenic commercial burley tobacco cultivar, 'Vector 21-41,' which exhibits dramatically reduced total alkaloid levels (~0.23% vs 2.3% for wild-type). Other examples of transgene-based modification of alkaloid profiles have involved the use of up- or down-regulation of *NBB1* or *A622* using overexpression or RNA interference (Hashimoto and Kato, 2007; Hashimoto and Kajikawa, 2008; Kajikawa *et al.*, 2009). We are currently in the process of investigating the phenotypic effects of induced mutations in several selected genes of the alkaloid biosynthetic pathway. Most of these genes are present in multiple copies in the *N. tabacum* genome. It is expected that one could have quantitative impacts on alkaloid formation by introducing deleterious mutations into one or several copies of a single gene, or by combining deleterious mutations in multiple genes. Some genetics-based modifications of alkaloid levels have been achieved without

obvious undesired alterations in plant productivity (Xie *et al*, 2004). One must be cautious in using these directed approaches for modifying plant alkaloid levels, however, because of a possible buildup of upstream intermediates that may be deleterious to normal plant growth and development (Sato *et al*., 2001; Kajikawa *et al*., 2009).

Additional Possibilities

Cadmium (Cd) is a heavy metal that can be toxic to humans at high levels and that is taken up by the plant from the soil where the element exists in variable amounts as a contaminant. Because of its presence in both smoke and non-burning tobacco products, there has been significant interest in developing technologies that might reduce its accumulation in above-ground plant parts. Two papers have reported on the impact of transgenic expression of mammalian metallothionein genes (*mMTI* or *hMTII*) to entrap Cd in tobacco root tissues through cytosolic chelation and reduce translocation to leaves (Yeargan *et al*., 1992; Dorlhac de Borne *et al*., 1998). Reductions in Cd accumulation in the lamina of transgenic field grown plants ranged from 14% to 73% relative to controls. Metallothioneins can also strongly fix metal ions such as copper and zinc, however. Descriptions of corresponding negative effects on plant productivity that might be predicted based on possible deficiencies in these elements uptake were variable.

Expression of vacuole cation exchange genes has also been evaluated for the potential to sequester Cd in root vacuoles. Transgenic tobacco field-grown plants expressing CAX2 or CAX4 Cd antiporter genes from *Arabidopsis thaliana* under the control of root-specific promoters accumulated 15% to 25% less Cd in the lamina as compared to control plants (Korenkov *et al*., 2009). Combining the two transgenic mechanisms (*mMT* and *CAX2*) into single tobacco genotypes indicates that additive reductions in lamina Cd contents might be achieved (Wagner *et al*., 2009).

A third documented strategy for reducing leaf Cd accumulation involves the use of RNA interference to knockdown the expression of *N. tabacum* heavy metal ATP-ase (*HMA*) genes (Hayes *et al*., 2009) that play a role in transport of heavy metal cations across cellular membranes in plants. Introduction of mutations into these genes might contribute to comparable Cd reductions in leaves (Hussain *et al*., 2004).

In all of these methods, one might need to be concerned about zinc deficiencies since the proposed genetic manipulations may not effectively discriminate between this element and Cd. Potential problems in this area might be counteracted by adding supplemental zinc to soils.

Prospects for Transgenic Tobacco?

Genetic variability is a general requirement to effect desired change for a trait through conventional selection. For certain tobacco chemistry traits, such variability may not exist within the *Nicotiana* genus and alternative strategies are needed. The use of mutation breeding to dramatically reduce expression of target genes was outlined previously. In order to increase expression of a particular gene or to add a new trait, however, genetic engineering is usually required. Several examples of this approach were presented here, as well.

A curious relationship exists between the tobacco industry and transgenic tobacco. Tobacco was the first plant species to be genetically engineered (Barton *et al.*, 1983; De Block *et al.*, 1984; Horsch *et al.*, 1985; An *et al.*, 1986), and dozens of traits have been documented for this commodity that could reduce crop production expenses for growers and supply costs for the industry. Genetically engineered varieties of other major crop species have been widely planted in many parts of the world. To date, the mainstream tobacco industry has been resistant to the use of cured leaf from transgenic tobacco cultivars, however. The only transgenic tobacco cultivar to be commercialized to this point is burley tobacco variety 'Vector 21-41,' which has been grown on an extremely limited basis in the U.S. to produce cured tobacco with ultra-low nicotine content (Xie *et al.*, 2004).

The scientific risks associated with the use of transgenic tobacco appear to be minimal. The potential for transgenic escape is very small because modern tobacco cultivars are typically male-sterile, and flowers are generally removed as a part of routine tobacco production. While a cautious approach seems warranted anytime alien genes are introduced into the genome of any commodity, it is difficult to foresee how the presence of most transgenes of commercial interest could increase risk to users of cigarettes.

One must assume that current objection to transgenic tobacco involves concerns associated with public perception. Tobacco and genetic engineering are socially and politically charged words, and the industry has already received negative attention over the development of flue-cured tobacco variety 'Y1' with reported elevated nicotine levels achieved through plant breeding. Under increasingly stringent tolerance levels for harmful constituents that are likely to develop over time, it may be become exceedingly difficult to comply with regulations without using GMO technologies. If this occurs, business may then become the driving force behind rapid adoption of transgenic tobaccos. Indeed, perception could be damaged if the public becomes aware of the existence of transgenic technologies that might reduce risk, but that were never put into use. To date, it appears that regulatory suggestions for harmful constituents have at least been partially based upon what is technologically achievable (Burns *et al.*, 2008; O'Connor *et al.*, 2008). Existence of related transgenic technologies may ultimately contribute to future standards that must be met by the industry.

Conclusions

Regulation of tobacco product chemistry will begin in many countries in the next few years. Although it may be a formidable task to dramatically reduce levels of certain constituents produced during pyrolysis, product design as well as plant genetics will play a role in reducing consumer risk and in complying with regulations. Several examples for combining the use of gene-specific information and conventional plant breeding approaches to modify tobacco product chemistry were presented here. No single research group possesses all the necessary expertise, infrastructure, or germplasm needed to commercialize a new tobacco cultivar carrying a harm reduction trait. Ultimate deployment will require a high degree of cooperation between those involved in gene discovery, molecular biology, and plant breeding. The issue of intellectual property protection was not discussed here, but patents associated with genes or technologies could delay inclusion of cured leaf of new tobacco cultivars in future products.

REFERENCES

1. An, G., B.D. Watson, and C.C. Cheng. 1986. Transformation of tobacco, tomato, potato, and *Arabidopsis thaliana* using a binary Ti vector system. Plant Physiol. 81:301-305.
2. Barton, K.A., A.N. Binns, A.J.M. Matzke, and M.D. Chilton. 1983. Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. Cell 32:1033-1043.
3. Baulcombe, D.C. 1999. Fast forward genetics based on virus-induced gene silencing. Curr. Opin. Plant Biol. 2:109-113.
4. Benowitz, N.L., S.M. Hall, S. Stewart, W. Wilson, D. Dempsey, and P. Jacob. 2007. Nicotine and carcinogen exposure with smoking of progressively reduced nicotine content cigarette. Cancer Epidemiol. Biomarkers Prev. 16:2479-2485.
5. Benowitz, N.L., and J.E. Henningfield. 1994. Establishing a nicotine threshold for addiction. New England J. Med. 331:123-125.
6. Bindler, G., R. van der Hoeven, I. Gunduz, J. Plieske, M. Ganal, L. Rossi, F. Gadani, and P. Donini. 2007. A microsatellite marker based linkage map of tobacco. Theor. Appl. Genet. 114:341-349.
7. Birch, R.G. 1997. Plant transformation: problems and strategies for practical application. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:297-326.
8. Bowman, D.T., E.A Wernsman, T.C. Corbin, and A.G. Tart. 1984. Contribution of genetics and production technology to long-term yield and quality gains in flue-cured tobacco. Tob. Sci. 28:30-35.
9. Brandle, J.D., and D. Bai. 1999. Biotechnology: Uses and applications in tobacco improvement. *In*: Davis DL, Nielsen MT (eds) Tobacco Production, Chemistry, and Technology. Blackwell Science Ltd., Oxford, pp 49-65.

10. Burns, D.M., E. Dybing, N. Gray, S. Hecht, C. Anderson, T. Sanner, R. O'Connor, M. Djordjevic, C. Dresler, P. Hainaut, M. Jarvis, A. Opperhuizen, and K. Straif. 2008. Mandated lowering of toxicants in tobacco smoke: a description of the World Health Organization TobReg proposal. Tobacco Control 17:132-141.
11. Bush, L.P., Cui, M., Shi, H. and Burton, H.R. 2001. Formation of tobacco specific nitrosamines in air-cured tobacco. Rec. Adv. Tob. Sci. 27:23-46.
12. Cane, K.A., M. Mayer, A.J. Lidgett, A.J. Michael, and J.D. Hamill. 2005. Molecular analysis of alkaloid metabolism in AABB v. aabb genotype *Nicotiana tabacum* in response to wounding of aerial tissues and methyl jasmonate treatment of cultured roots. Func. Plant Biol. 32:305-320.
13. Chaplin, J.F. 1987. Die Züchtung von Tabak im Hinblick auf verschieden hohe Alkaloidgehalte. Beitrage zur Tabakforschung International 14:1-9.
14. Chaplin, J.F., and H.W. Spurr. 1982. Altering condensate levels in tobacco smoke by genetic techniques. Beitr. Tabakforschung 11:151-160.
15. Chaplin, J.F., and W.W. Weeks. 1976. Association between percent total alkaloids and other traits in flue-cured tobacco. Crop Sci. 16:416-418.
16. Chintapakorn, Y., and J.D. Hamill. 2003. Antisense-mediated down-regulation of putrescine *N*-methyltransferase activity in transgenic *Nicotiana tabacum* L. can lead to elevated levels of anatabine at the expense of nicotine. Plant Mol. Biol. 53:87-105.
17. De Block, M., L. Herrera-Estrella, M. van Montagu, J. Schell, and P. Zambryski. 1984. Expression of foreign genes in regenerated plants and their progeny. EMBO J. 3:1681-1689.
18. DeBoer, K., J.C. Lye, C.D. Aitken, A.K.-K. Su, and J.D. Hamill. 2009. The *A622* gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Mol. Biol. 69:299-312.
19. del Piano, L., M. Abet, C. Sorrentino, F. Acanfora, E. Cozzolino, and A. Di Muro. 2000. Genetic variability in *Nicotiana tabacum* and *Nicotiana* species as revealed by RAPD markers: 1. Development of the RAPD procedure. Beiträge zur tabakforshung International Contribution to Tobacco Research 19:1–15.
20. De Sutter, V., R. Vanderhaeghen, S. Tilleman, F. Lammertyn, I. Vanhoutte, M. Karimi, D. Inze, A. Goossens, and P. Hilson. 2005. Exploration of jasmonate signaling via automated and standardized transient expression assays in tobacco cells. Plant J. 454:1065-1076.
21. Dorlhac de Borne, F., T. Elmayan, C. de Roton, L. de Hys, and M. Tepfer. 1998. Cadmium portioning in transgenic tobacco plants expressing a mammalian metallothionein gene. Mol. Breeding 4:83-90.
22. Dorlhac de Borne, F., S. Coates, J. Ross, J.L. Verrier, L. Jones, E. Julio and R. Delon. 2006. A comprehensive survey of the *N. tabacum* transcriptome (The European Sequencing of Tobacco project, ESTobacco). 2006 CORESTA Congress, Oct. 15-20, Paris, France.

23. Gadani, F., A. Hayes, C.H. Opperman, S.A. Lommel, B.R. Sosinski, M. Burke, L. Hi, R. Brierly, A. Salstead, J. Heer, G. Fuelner, and N. Lakey. 2003. Large scale sequencing and analysis of *Nicotiana tabacum* : the Tobacco Genome Initiative. 5th Bergerac Tobacco Scienctific Meeting, 4-5 September, 2005, Bergerac, France.
24. Gartner, C.E., W.D. Hall, T. Vos, M.Y. Bertram, A.L. Wallace, and S.S. Lim. 2007. Assessment of Swedish snus for tobacco harm reduction: an epidemiological study. The Lancet 369:2010-2014.
25. Gavilano, L.B., N.P. Coleman, L.-E. Burnley, M.L. Bowman, N.E. Kalengamaliro, A. Hayes, L. Bush, and B. Siminszky. 2006. Genetic engineering of *Nicotiana tabacum* for reduced nornicotine content. J. Agric. Food Chem. 54:9071-9078.
26. Gavilano, L.B., and B. Siminszky. 2007. Isolation and characterization of the cytochrome P450 gene *CYP82E5v2* that mediates nicotine to nornicotine conversion in the green leaves of tobacco. Plant Cell Physiol. 48:1567-1574.
27. Hashimoto, T., and M. Kajikawa. 2008. Increasing levels of nicotinic alkaloids in plants. United States Patent Application US 2008/0120737 A1.
28. Hashimoto, T., and A. Kato. 2007. Reducing levels of nicotinic alkaloids in plants. United States Patent Application US 2007/0240728 A1.
29. Hayes, A.J., C. Kudithipudi, and R.S. van der Hoeven. 2009. Transgenic plants modified for reduced cadmium transport, derivative products, and related methods. Unites States Patent Application US 2009/0183280 A1.
30. Hecht, S.S. 1998. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. Chem. Res. Toxicol. 11:559-603.
31. Hecht, S.S. 2003. Tobacco carcinogens, their biomarkers and tobacco-induced cancers. Nature Rev. Cancer 3:733-744.
32. Hecht, S.S. 2006. Cigarette smoking: cancer risks, carcinogens, and mechanisms. Langenbecks Arch. Surg. 391:603-613.
33. Hecht, S.S., and D. Hoffmann, D. 1989. The relevance of tobacco-specific nitrosamines to human cancer. Cancer Surv. 8:273-294.
34. Heim, W.G., K.A. Sykes, S.B. Hildreth, J. Sun, R.-H. Lu, and J.G. Jelesko. 2006. Cloning and characterization of a *Nicotiana tabacum* methylputrescine oxidase transcript. Phytochemistry 68:454-463.
35. Hibi, N., S. Higashiguchi, T. Hashimoto, and Y. Yamada. 1994. Gene expression in tobacco low-nicotine mutants. Plant Cell 6:723-735.
36. Hoffmann, D., and M.V. Djordjevic. 1997. Chemical composition and carcinogenicity of smokeless tobacco. Adv. Dent. Res. 11:322-329.
37. Hoffmann, D., K.D. Brunnemann, B. Prokopczyk, and M.V. Djordjevic. 1994. Tobacco-specific *N*-nitrosamines and areca-derived *N*-nitrosamines – Chemistry, biochemistry, carcinogenicity, and relevance to humans. J. Toxicol. Environ. Health 41:1-52.

38. Horsch, R.B., J.E. Fry, N.L. Hoffman, D. Eichholtz, S.G. Rogers, and R.T. Fraley. 1985. A simple and general method for transferring genes into plants. Science 227:1229-1231.
39. Hussain, D., M.J. Haydon, Y. Wang, E. Wong, S.M. Sherson, J. Young, J. Camakaris, J.F. Harper, and C.S. Cobbett. 2004. P-Type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. Plant Cell 16:1327-1339.
40. Julio E, B. Denoyes-Rothan, J.-L.Verrier, and F. Dorlhac de Borne. 2006. Detection of QTLs linked to leaf and smoke properties in *Nicotiana tabacum* based on a study of 114 recombinant inbred lines. Mol. Breed. 18:69–91.
41. Kajikawa, M., N. Hirai, and T. Hashimoto. 2009. A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Mol. Biol. 69:287-298.
42. Katoh, A., T. Shoji, and T. Hashimoto. 2007. Molecular cloning of *N*-methylputrescine oxidase from tobacco. Plant Cell Physiol. 48:550-554.
43. Korenkov, V., B. King, K. Hirschi, and G.J. Wagner. 2009. Root-selective expression of *AtCAX4* and *AtCAX2* results in reduced lamina cadmium in field-grown *Nicotiana tabacum* L. Plant Biotech. J. 7:219-226.
44. Legg, P.D., J.F. Chaplin, and G.B. Collins. 1969. Inheritance of percent total alkaloids in *Nicotiana tabacum* L. J. Hered. 60:213-217.
45. Legg, P.D., and G.B. Collins. 1971. Inheritance of percent total alkaloids in *Nicotiana tabacum* L. II. Genetic effects of two loci in Burley 21 x LA Burley 21 populations. Can. J. Genet. Cytol. 13:287-291.
46. Lewis, R.S., S.W. Bowen, M.R. Keogh, and R.E. Dewey. 2010. Three nicotine demethylase genes mediate nornicotine accumulation in tobacco: functional characterization of the *CYP82E10* gene. (In Review)
47. Lewis, R.S., A.M. Jack, J.W. Morris, V.J.M. Robert, L.B. Gavilano, B. Siminszky, L.P. Bush, A.J. Hayes, and R.E. Dewey. 2008. RNA interference (RNAi)-induced suppression of nicotine demethylase activity reduces levels of a key carcinogen in cured tobacco leaves. Plant Biotech. J. 6:346-354.
48. Matzinger, D.F., T.J. Mann, and C.C. Cockerham. 1972. Recurrent family selection and correlated response in *Nicotiana tabacum* L. I. 'Dixie Bright 244' x 'Coker 139.' Crop Sci. 12:40-43.
49. Matzinger, D.F., W.W. Weeks, and E.A. Wernsman. 1984. Genetic modification of total particulate matter. Rec. Adv. Tob. Sci. 10:15-51.
50. Matzinger, D.F., E.A. Wernsman, and W.W. Weeks. 1978. Genetic modification of total particulate matter in tobacco smoke. Tob. Sci. 22:138-140.
51. Matzinger, D.F., E.A. Wernsman, and W.W. Weeks. 1989. Restricted index selection for total alkaloids and yield in tobacco. Crop Sci. 29:74-77.

52. Morita, M., N. Shitan, K. Sawada, M.C.E. Van Montagu, D. Inze, H. Rischer, A. Goossens, K. Oksman-Caldentey, Y. Moriyama, and K. Yazaki. 2009. Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. Proc. Natl. Acad. Sci. 106:2447-2452.
53. National Toxicology Program. 2005. Report on Carcinogens, Eleventh Edition, U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program.
54. Nishi, T., T. Tajima, S. Noguchi, H. Ajisaka, and H. Negishi. 2003. Identification of DNA markers of tobacco linked to bacterial wilt resistance. Theor. Appl. Genet. 106:765-770.
55. O'Connor, R.J., and P.J. Hurley. 2008. Existing technologies to reduce specific toxicant emissions in cigarette smoke. Tobacco Control 17 (Suppl. 1):i39-i48.
56. Opperman, C.H., S.A. Lommel, M. Burke. 2007. Sequencing and analysis of the *Nicotiana tabacum* genome. Rec. Adv. Tob. Sci. 33:5-14.
57. Rathkamp, G., T.C. Tso, and D. Hoffmann. 1973. Chemical studies on tobacco smoke, XX: smoke analysis of cigarettes made from bright tobaccos differing in variety and stalk positions. Beitr. Tabakforsch 7:179-189.
58. Reed, D.G., and J.G. Jelesko. 2004. The A and B loci of *Nicotiana tabacum* have non-equivalent effects on the mRNA levels of four alkaloid biosynthetic genes. Plant Science 167:1123-1130.
59. Ren, N., and M.P. Timko. 2001. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome 44:559-571.
60. Riechers, D.E., and M.P. Timko. 1999. Structure and expression of the gene family encoding putrescine *N*-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. Plant Mol. Biol. 41:387-401.
61. Rossi, L., G. Bindler, H. Pijnenburg, P.G. Isaac, I. Giraud-Henry, M. Mahe, C. Orvain, and F. Gadani. 2001. Potential of molecular marker analysis for variety identification in processed tobacco. Plants Varieties and Seeds 14:89-101.
62. Russell M.A.H. 2000. Public health and levels of nicotine: should nicotine levels in cigarettes be minimized or maximized? *In*: R. Ferrence, J. Slade, R. Room, *et al* (eds). *Nicotine and Public Health*. Washington DC: American Public Health Association.
63. Sato, F., T. Hashimoto, A. Hachiya, K. Tamura, K.-B. Choi, T. Morishige, H. Fujimoto, and Y. Yamada. 2001. Metabolic engineering of plant alkaloid biosynthesis. Proc. Nat. Acad. Sci. 98:367-372.

64. Shinozaki, K., M. Ohme, M. Tanaka, T. Wagasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and exression. The EMBO J. 5:2043-2049.
65. Shoji, T., K. Inai, Y. Yazaki, Y. Sato, H. Takase, N. Shitan, K. Yazaki, Y. Goto, K. Toyooka, K. Matsuoka, and T. Hashimoto. 2009. Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. Plant Physiol. 149:708-718.
66. Shoji, T., R. Winz, T. Iwase, K. Nakajima, Y. Yamada, and T. Hashimoto. 2002. Expression patterns of two isoflavone reductase-like genes and their possible roles in secondary metabolism in tobacco. Plant Mol. Biol. 50:427-440.
67. Siminszky, B., L. Gavilano, S.W. Bowen, and R.E. Dewey. 2005. Conversion of nicotine to nornicotine in *Nicotiana tabacum* is mediated by *CYP82E4*, a cytochrome P450 monooxygenase. Proc. Natl. Acad. Sci. USA 102:14919-14924.
68. Sinclair, S.J., K.J. Murphy, C.D. Birch, and J.D. Hamill. 2000. Molecular characterization of quinolinate phosphoribosyltransferase (QPRTase) in *Nicotiana*. Plant Mol. Biol. 44:603-617.
69. Stratton, K.R., P. Shetty, R. Wallace, and S. Bondurant. 2001. Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction. National Academy Press, Washington, D.C.
70. Sugiyama, Y., Y. Watase, N. Nagase, N. Makita, S. Yagura, A. Hirai, and M. Sugiura. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol. Gen. Genomics 272:603-615.
71. Todd, A.T., E. Liu, S.L. Polvi, R.T. Pammett, and J.E. Page. 2010. A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J. 62:589-600.
72. Tso, T.C., and J.F. Chaplin. 1977. Simple correlation and multiple regression among leaf characteristics, smoke components, and biological responses to bright tobacco. U.S.D.A., A.R.S. Tech. Bull. 1551.
73. Vontimitta, V. 2010. Genetic mapping of genes controlling two leaf surface chemistry traits and identification of quantitative trait loci (QTL) associated with resistance to *Phytophthora nicotianae* in tobacco (*Nicotiana tabacum* L). Ph.D. Dissertation, N.C. State University, Raleigh, NC.

74. Vontimitta, V., D.A. Danehower, T. Steede, H.S. Moon, and R.S. Lewis. 2010. Analysis of a *Nicotiana tabacum* L. genomic region controlling two leaf surface chemistry traits. J. Agric. Food Chem. 58:294-300.
75. Wagner, G.J., V. Korenkov, and B. King. 2009. Impacts of expressing *AtCAX* and metallothionein genes on Cd accumulation in field grown tobacco. 2009 CORESTA Congress, Oct. 4-8, Rovinj, Croatia.
76. Wang, J., M. Sheehan, H. Brookman, and M.P. Timko. 2000. Characterization of cDNAs differentially expressed in roots of tobacco (*Nicotiana tabacum* cv Burley 21) during the early stages of alkaloid biosynthesis. Plant Science 158:19-32.
77. Wernsman, E.A., D.F. Matzinger, and W.W. Weeks. 1977. Potential for particulate matter reduction through tobacco genetic and cultural modifications. Proc. Amer. Chem. Soci. Symp. Recent Advances in the Chemical Composition of Tobacco and Tobacco Smoke. pp. 340-364.
78. Wernsman, E.A., and R.C. Rufty. 1987. Tobacco. p. 669-698. *In* W.R. Fehr W.R. (ed.) Principles of Cultivar Development, Vol. 2 - Crop Species. Macmillian Publishing Company, New York.
79 Xie, J., W. Song, W. Maksymowicz, W. Jin, K. Cheah, W. Chen, C. Carnes, J. Ke, and M.A. Conkling. 2004. Biotechnology: A tool for reduced-risk tobacco products – the nicotine experience from test tube to cigarette pack. Rec. Adv. Tob. Sci. 30:17-38.
80. Yeargan, R., I.B. Maiti, M.T. Nielsen, A.G. Hunt, and G.J. Wagner. 1992. Tissue portioning of cadmium in transgenic tobacco seedlings and field grown plants expressing the mouse metallothionein I gene. Transgenic Research 1:261-267.
81. Yukawa, M., T. Tsudzuki, and M. Sugiura. 2006. The chloroplast genome of Nicotiana sylvestris and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. Mol. Gen. Genomics 275:367-373.

FUNCTIONAL GENOMIC APPROACHES TO HARM REDUCTION IN TOBACCO PRODUCTS

Michael P. Timko, Paul J. Rushton*, Marta T. Bokowiec and Hongbo Zhang

Department of Biology
University of Virginia
Charlottesville, Virginia 22904 USA

*Current address is:
Department of Biology and Microbiology
South Dakota State University
Brookings, South Dakota 57007 USA

Abstract

There is overwhelming evidence that the use of tobacco products plays a major role in the pathogenesis of lung and oral cancers, cardiovascular and chronic obstructive pulmonary diseases, and a variety of other human maladies. In the absence of complete elimination of their use, a viable alternative is the development of tobacco products that deliver reduced levels of constituents harmful to the human body. Among the constituents implicated in increased risks for various pathologies are a variety of tobacco-specific nitrosamines (TSNAs), polycyclic aromatic hydrocarbons (PAHs), and aromatic amines. Nicotine, the most prevalent alkaloid found in cultivated tobacco (*Nicotiana tabacum* L.), nornicotine, and various minor alkaloids (*e.g.*, anatabine, anabasine, anatalline) contribute significantly to the production of TSNAs during post-harvest curing and fermentation and are considered key targets for harm reduction. Recent advances in our knowledge of the metabolic pathways controlling the biosynthesis and accumulation of tobacco alkaloids and their derivatives, and the development of new genomic, proteomic, and metabolomic tools for analysis and manipulation of leaf chemistry and composition offer new opportunities to reduce the risks associated with tobacco use. Using oligonucleotide microarrays and transcriptomic profiling to examine global changes in gene expression and transgenic approaches (ectopic overexpression / RNAi knockdowns) in cultured cells and whole plants, we are working to uncover the transcriptional circuitry that controls alkaloid formation in response to various developmental, phytohormonal, and environmental cues and identify novel cellular targets controlling alkaloid biosynthesis, accumulation and derivation. Our progress towards understanding the control of tobacco alkaloid formation and its use in the generation of harm reduced products is discussed.

Introduction

Since the mid 1950's, there has been a growing concern about the detrimental effects of the use of tobacco products on human health (Wynder and Hoffmann, 1979). Overwhelming evidence now exists that tobacco use, and in particular exposure to tobacco smoke, plays a major role in cardiovascular disease, chronic obstructive pulmonary disease (COPD), and the pathogenesis of various lung, oral cavity, throat/esophageal and organ cancers (Bartecchi *et al.*, 1995; Hecht 2003; van der Vaart *et al.*, 2004). Similar concerns of increased risk for a variety of oral and throat cancers and other human diseases exist over the use of smokeless tobacco products (Stepanov *et al.*, 2008). While the surest way to reduce tobacco use–related disease is to promote complete cessation, many members of society are unwilling to abandon the use of tobacco products or unable to quit due to the addictive properties of nicotine (Djordjevic *et al.*, 1997). As an alternative the development of tobacco products with lower levels of potential harmful constituents delivered to human body and the manufacture of Potential Reduced Exposure Products (PREPs) such as cigarette-like devices (Bombick *et al.*, 1997; Hoffman *et al.* 1997; Breland *et al.*, 2002; Hatsukami *et al.*, 2007; Pederson and Nelson, 2007) are being promoted by various sectors (Stratton *et al.*, 2001). Both strategies remain controversial, owing mainly to the lack of clear data showing positive health effects from their use in comparison to traditional smoked and smokeless tobacco products (Hecht *et al.*, 2007; Prignot *et al.*, 2008).

In recent years there has been considerable advancement in our understanding of the metabolic pathways that control the accumulation of tobacco alkaloids and their conversion into constituents harmful to human health during pre- and post-harvest processing. Coupled with the significant expansion of plant genomics (including transcriptomics, metabolomics and bioinformatics) and enhanced technologies for plant gene manipulation (*e.g.*, gene transfer, ectopic overexpression, RNA interference (RNAi)-mediated gene silencing), new research opportunities exist to reduce the risks associated with tobacco use though directed alteration of leaf chemistry and composition. In this paper, we describe the results of studies in tobacco based on the application of several different emerging technologies for understanding global gene expression as a means of uncovering signal pathway components involved in the control of alkaloid biosynthesis and identification of potential targets for selective expression alteration to bring about changes leading to harm reduction.

Transcriptional circuits controlling alkaloid biosynthesis gene expression

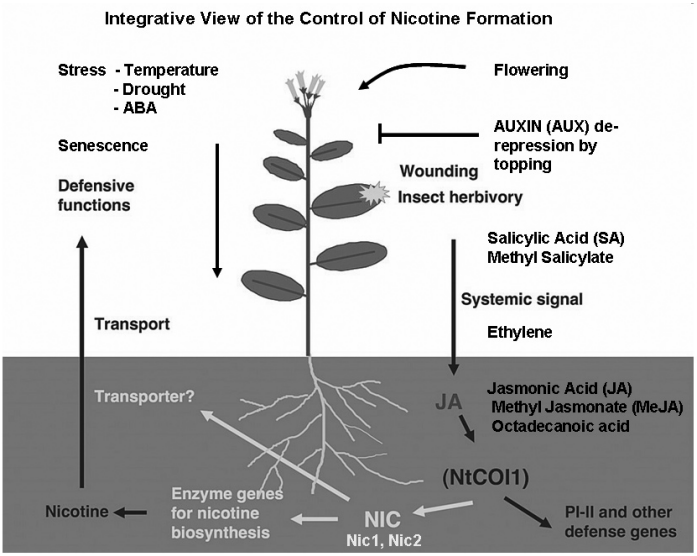
Wild and cultivated tobaccos (*Nicotiana* spp.) produce an array of alkaloids that play an important role in the plant's defense against herbivore and insect attack (Sisson and Severson, 1990; Steppuhn *et al.*, 2004). The main alkaloid found in cultivated tobacco (*N. tabacum* L.) is nicotine which constitutes approximately 0.6–3.0% of leaf dry weight in undamaged tobacco plants. Real or simulated

herbivory (wounding) of tobacco leaves may cause an increase of up to 1–4%, which is sufficient to deter even nicotine-adapted insects (Steppuhn *et al.*, 2004).

Nicotine is synthesized by condensation of two transitory compounds, a 1-methyl-Δ^1-pyrrolium cation and nicotinic acid. The 1-methyl-Δ^1-pyrrolium cation is derived from *N*-methylputrescine, synthesized from the polyamine putrescine by the *S*-adenosylmethionine-dependent enzyme putrescine *N*-methyltransferase (PMT) (Facchini, 2001). PMT is the key regulatory enzyme in the biosynthesis of nicotine and others pyridine alkaloids. Nicotinic acid is formed from quinolinic acid through the pyridine nucleotide cycle, and this process is controlled by quinolinate phosphoribosyltransferase (QPT) (Sinclair *et al.*, 2000). Condensation of the pyrrolidine and pyridine rings is mediated by the presumed nicotine synthesis enzyme known simply as A622 (De Boer *et al.*, 2009; Kajikawa *et al.*, 2009).

Nicotine biosynthesis occurs predominantly in the roots of tobacco and the molecule is then translocated through the xylem to aerial parts of the plant where it accumulates in the leaves and is exuded by trichomes to protect against insect defoliation (Thurston *et al.*, 1966: Hashimoto and Yamata, 2003; Morita *et al.*, 2009). A variety of factors, including plant developmental age, biotic and abiotic stresses and phytohormone levels control nicotine biosynthesis and accumulation (Figure 1) (Facchini, 2001; Kessler and Baldwin, 2002; Goossens *et al.*, 2003; Goossens and Rischer, 2007). Jasmonic acid (JA, 3-oxo-2-(cis-2-pentenyl)-1-cyclopentaneacetic acid) and its volatile methyl ester (MeJA) and auxins (AUX) are well documented phytohormone regulators of nicotine formation (Pauw and Memlink, 2004). Increased endogenous levels of JA and exogenous application of JA /MeJA have been shown to rapidly induce the transcription of genes encoding the enzymes of nicotine biosynthesis (Shoji *et al.*, 2000, 2008; Goossens *et al.*, 2003; Katoh *et al.*, 2005). Nicotine accumulation can also be induced by wounding (Baldwin, 1988) and this response is tied into the JA signaling pathway (Baldwin *et al.*, 1994). On the other hand, exogenous auxin application can inhibit these wound-induced responses and significantly reduces MeJA-inducible accumulation transcripts for many of the enzymes involved in nicotine synthesis (Imanishi *et al.*, 1998; Xu *et al.*, 2004). Removal of the flowering head and young leaves leads to increased nicotine concentration in tobacco leaves (Hashimoto and Yamada, 1994; Hibi *et al.*, 1994). The factors controlling the increase in alkaloid biosynthesis induced by removal of the apex are not known (Xu *et al.*, 2004), but application of an AUX transport inhibitor around the stem directly under the shoot apex of intact plants caused an increase in nicotine concentration in the whole plant. The results strongly suggest that AUX serves as a negative signal to regulate nicotine synthesis in roots of tobacco plants (Shi *et al.*, 2006).

Figure 1. Integrative view of the control of nicotine formation. (Adapted from Katoh et al. (2005).



Recent studies have shown that the perception of a bioactive jasmonate derivative (jasmonoyl-L-isoleucine; JA-Ile) by the receptor F-box protein CORONATINE INSENSITIVE1 (COI1) (Yan *et al.*, 2009) leads to the formation of a stable COI1/JA-Ile complex that binds to a JA- inducible zinc finger-like ZIM motif-containing transcriptional repressor protein known as JAZ (Chini *et al.*, 2007) resulting in ubiquitination of JAZ proteins by the SCF^{COI1} complex and their degradation by the 26S proteasome (Xu et al., 2002). In the absence of JA-Ile, JAZ proteins repress the activity of the basic helix-loop-helix (bHLH) transcription factor (TF) MYC2 required for activation of the downstream JA responses (Chini *et al.*, 2007; Thines *et al.*, 2007; Katsir *et al.*, 2008). When MYC2 is liberated, it alone or in cooperation with other TFs activates transcription of target genes.

We previously reported that five PMT genes are present in *N. tabacum* (*NtPMT1a, NtPMT1b, NtPMT2, NtPMT3* and *NtPMT4*) whose kinetics of expression and levels of transcript accumulation differ after topping and in response to phytohormone treatment (*e.g.*, MeJA, ethylene (ETH) and AUX) (Riechers and Timko, 1999; Xu and Timko, 2004). Using *N. tabacum* cv. Bright Yellow 2 (BY-2) cell cultures, a well-defined model for the analysis of transcription circuitry and signal transduction pathways in alkaloid and isoprenoid formation (Goossens *et al.*, 2003; Nagata *et al.*, 2004), we previously demonstrated that MeJA treatment rapidly induces *NtPMT1a* transcription and that as little as 111 bp of promoter upstream of the transcriptional start site of the *NtPMT1a* gene are sufficient to

confer MeJA-responsiveness (Xu and Timko, 2004). Within this region are three recognizable elements: a G box-like element at -103 bp to -96 bp, an AT-rich region (92.8% A and T residues) that extends from -80 bp to -69 bp and a GCC box-like element located at -62 bp to -56 bp. We have designated this tripartite group of elements as the GAG motif.

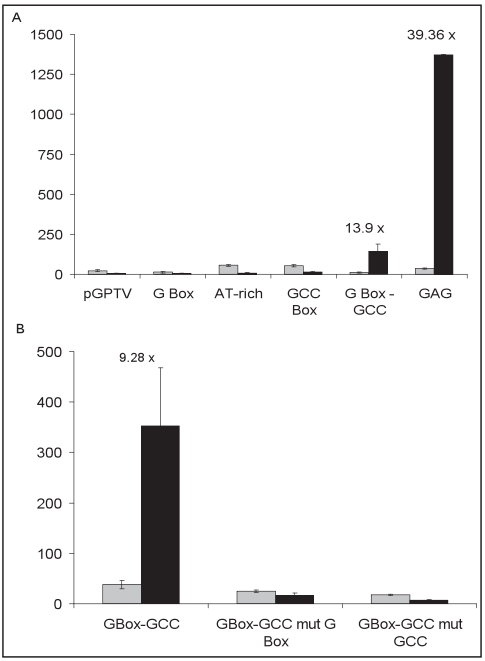
Using a gain-of-function approach, in which various synthetic promoters containing the individual elements (G-box, AT rich region, and GCC-box) or mutated versions of the elements, individually and in combination, were fused to a CaMV 35S -46 minimal promoter-GUS reporter gene, we have further demonstrated that tetramers of the G box like element alone showed no MeJA responsiveness (Figure 2). Similarly, tetramers of the AT-rich region and GCC box-like element also show no MeJA responsiveness. In contrast, tetramers consisting of a combination of the G box-like element and the GCC box-like element show significant MeJA responsiveness with a ~10-15 fold induction. However, the strength of the tetramer of the G box – GCC combination is lower than tetramers of the complete GAG motif, suggesting that the AT-rich region plays a role in the function of the tripartite GAG motif. Mutation of the core sequence in the G box-like element from ACGT to TTAA, or the GCC box-like element from GCGCCC to TTAATT abolished MeJA-induced expression. These finding suggest that more than one element is required for MeJA-induced *NtPMT1a* gene expression.

Examination of the promoter regions of available *NtPMT* genes from *N. tabacum* shows that they all contain a GAG motif at the same approximate location, ~110 bp upstream of their respective transcription start sites and that a small number of sequence differences distinguish the GAG motifs in each gene promoter. Gain-of-function experiments designed to test the regulatory properties and MeJA responsiveness of the GAG motifs from the different genes showed that all give high level MeJA-inducible expression (~25-80-fold induction) (data not shown). BLAST searches of available public databases failed to identify any other gene promoters with this novel motif, suggesting that the GAG motifs are a fundamental component of coordinated MeJA-inducible *NtPMT* gene expression in nicotine biosynthesis.

Two members of the AP2/ERF-domain TF family in tobacco called NtORC1 and NtJAP1 were previously shown to independently up-regulate the activity of the tobacco *NtPMT2* promoter in transient assays in tobacco protoplasts and function in a synergistic manner (De Sutter *et al.*, 2005). Both *NtORC1* and *NtJAP1* gene expression is induced by MeJA (Goossens *et al.*, 2003). To identify TFs that bind to the GAG motif, we carried out yeast one-hybrid screens using a 4 x GAG motif containing fragment as bait and a cDNA library constructed from MeJA-treated BY-2 cell suspensions. Several ERF family members were

recovered which specifically bound to the GCC-box in the *NtPMT1a* promoter in electrophoretic mobility shift assays. These ERFs were members of the Group IX subfamily (Rushton *et al.*, 2008) that includes *NtORC1* and *NtJAP1*, are MeJA-responsive (Rushton, *et al.*, 2008) and have been previously implicated in JA regulated defense responses (Kessler and Baldwin, 2002; Howe and Jander, 2008).

Figure 2. Functional analysis of the elements in the GAG motif. (A) Activity of tetramers of individual elements and a combination of the G box-like and GCC box-like elements. (B) Functional analysis of the G box – GCC combination. GUS activity is presented on the Y axis and is measured in pmoles/min/mg protein. Black bars represent the level of GUS activity 48 h after the addition of 100µM MeJA and grey bars the level after 48 h in control samples. The fold induction (+/- SE) of the constructs is shown.



Recently, six TFs from three different TF families were identified that alter constitutive and MeJA induced leaf nicotine levels in *N. benthamiana* (Todd *et al.*, 2010). Among these were two bHLH TFs that served as positive regulators of *NbPMT* gene expression. We have recently characterized three genes (*NtMYC2a/b/c*) encoding basic helix-loop-helix (bHLH) TFs with homology to MYC2 a key regulator of JA response in other plant systems. Expression of *NtMYC2a/b*

is rapidly induced by MeJA. Moreover, NtMYC2a/b specifically activate JA-inducible *NtPMT1a* expression by binding a G-box motif within the *NtPMT1a* promoter under both *in vivo* and *in vitro* assay conditions (Zhang H, Bokowiec MT, Rushton PJ, Han S and Timko, MP, unpublished). Using split-YFP assays we further showed that in the absence of JA, NtMYC2a and NtMYC2b are present as nuclear complexes with the NtJAZ1 repressor. RNAi-mediated knockdown of *NtMYC2a* and *NtMYC2b* expression results in significant decreases in MeJA-inducible *NtPMT1a* transcript levels, as well as reduced levels of transcripts encoding multiple other enzymes in nicotine and minor alkaloid biosynthesis, including an 80-90% reduction in the level of transcripts encoding the putative nicotine synthase gene *NtA662*. In contrast, ectopic overexpression of *NtMYC2a* and *NtMYC2b* had no effect on *NtPMT1a* expression in the presence or absence of exogenously added JA. Our studies suggest that NtMYC2a/b/c are key regulators of multiple genes in the nicotine biosynthetic pathway of tobacco.

Our current understanding of the regulation of nicotine biosynthetic gene expression by NtMYC2a/b is as follows. In the presence of MeJA, NtMYC2a/b/c are released from a repressor complex, allowing them to bind to the G-box-like element for the potential activation of *NtPMT1a*. However, as noted above, availability of NtMYC2a/b is insufficient alone to direct transcriptional activation since JA-induced *NtPMT1a* transcription also requires the recruitment of an AP2/ERF type TF that binds to the GCC-like element (Xu and Timko, 2004; Bokowiec MT, Han S, Zhang H, Rushton PJ and Timko MP, unpublished). This additional level of complexity may be a reflection of tight regulation of the first committed step in nicotine biosynthesis in tobacco.

Harm reduction through selective changes in alkaloid biosynthetic gene expression

Two biotechnological approaches have been employed to selectively alter alkaloid compositions and levels of tobacco specific nitrosamines (TSNAs) derived from them: transgenic manipulation of candidate genes (by ecotpic over expression or gene silencing) or systematic and automated identification of mutations in TILLING (Targeting Induced Local Lesions in Genomes) populations created by chemical or radiation induced mutagenesis (Colbert *et al.*, 2001; Comai and Henikoff, 2005).

The first successful gene knockdowns experiments aimed at altering nicotine biosynthesis used antisense constructs directed against PMT gene expression necessary for production of the 1-methyl-Δ^1-pyrrolium cation. Decrease in PMT transcript levels led to the generation of tobacco plants with varying reductions in nicotine levels with the perhaps unexpected consequence of an elevation in anatabine (Sato *et al.*, 2001; Voelckel *et al.*, 2001; Chintapakorn and Hamill, 2003; Steppuhn *et al.*, 2004; Wang *et al.*, 2008, 2009). Antisense-mediated reduction

in ADC activity (a step prior to PMT) led to minor alterations of alkaloids such as nicotine and anatabine in tobacco hairy roots and transgenic plants (Chintapakorn and Hamill, 2007) whereas RNAi mediated knockdown of the MPO expression (a step after PMT) resulted in transgenic lines with the contents of nicotine and nornicotine markedly decreased (1–7% and 3–7%, respectively) compared to controls plants, while the respective contents of anatabine, anabasine and anatalline increased 5- to 10-fold, 1.5- to 6-fold and 5- to 7-fold compared with the controls, respectively (Shoji *et al*., 2008).

Antisense suppression of QPT gene expression, involved in production of nicotinic acid, also provided significant nicotine reductions in transgenic tobacco plants (Xie *et al*., 2004). Interestingly, RNAi-mediated reduction of A622 transcripts encoding an isoflavone reductase (IFR)-like protein thought to be required in nicotine synthesis (Shoji *et al*., 2002; Goossens *et al*., 2003; Kidd *et al*. 2006) markedly reduces the capacity of *N. glauca* to produce anabasine resulting in plants with scarcely any pyridine alkaloids in leaf tissues, even after damage to apical tissues (De Boer *et al*., 2008). In addition, analysis of hairy roots containing the NgA622-RNAi construct shows a substantial reduction in both anabasine and nicotine levels within these tissues, even if stimulated with MeJA, indicating a role for the A622 enzyme in the synthesis of both alkaloids in roots of *N. glauca*. Kajikawa *et al*. (2009) identified two orthologous genes of *N. sylvestris* and *N. tomentosiformis* origins 622 and A622L that are 95.5% identical and 98.4%. RNAi suppression of A622/A622L inhibited synthesis of the non-pyrrolidine-type alkaloids but did not lead to accumulation of the 1-methyl-Δ^1-pyrrolium cation. These results indicate that A622 is required for the coupling reactions between a pyrrolidine precursor and the 1-methyl-Δ^1-pyrrolium cation to give nicotine, and between two pyrrolidine precursors (to yield anatabine). Knockdown of NtabCYP82E4 a cytochrome P450 encoding nicotine N-demethylase reduced the level of nornicotine (Lewis *et al*., 2008).

Alkaloid biosynthesis can also be controlled by targeting of transcription factors known to control genes at various steps in the biosynthetic pathway or involved in signal transduction leading to gene activation. For example, silencing of *COI1* in tobacco plants, as well as over-expressing of C-terminal deletion derivatives of NtJAZ1 or NtJAZ3 abolished the MeJA-responsive expression of nicotine biosynthesis genes including *NtPMT*, as well as MeJA-responsive nicotine accumulation (Shoji *et al*., 2008).

As noted above, we recently showed that RNAi-mediated knock down of NtMYC2a and NtMYC2b expression results in significant decreases in MeJA-inducible *NtPMT1a* transcript levels and alkaloid levels. Similarly, RNAi silencing of the bHLH TF genes *NbbHLH1* or *NbbHLH2* that regulate the *N. sylvestris NsPMT* in

response to MeJA treatment, resulted in the reduced expression of genes encoding nicotine biosynthetic enzymes, and reduced alkaloid levels (Todd *et al.*, 2010).

As an alternative to transgenic approaches for harm reduction, Julio et al (2008) used a mutant-based TILLING strategy to identify mutated tobacco lines with no or low levels of nornicotine. A similar strategy has also been used to identify tobacco lines containing mutations in NtERF2 and NtERF5, key TF genes necessary for NtPMT regulation identified in our studies described above, and these lines show reduced levels of nicotine and nornicotine (Kudithipudi C and Hayes A, personal communication).

Global changes in gene expression during MeJA-induced alkaloid formation
Over one million gene-space sequence reads (GSRs) (representing ~1.2 Gb of nuclear genomic sequence) were generated from the *N. tabacum* cultivar 'Hicks Broadleaf', a highly inbred and fairly homozygous tobacco cultivar using Gene Thresher® methylation filtering technology by the Tobacco Genome Initiative (TGI; http://www.tobaccogenome.org/), The availability of these genomic sequences greatly facilitated genome-wide analysis, large scale functional genomics, and gene discovery for the identification of important biosynthetic pathways related to secondary metabolism and down-stream targets for harm reduction.

Using the TGI sequence data, we carried out an *in silico* analysis and identified approximately 2,500 TFs present in the tobacco genome (Rushton *et al.*, 2008). These data are available to the public at TOBFAC: The Database of Tobacco Transcription Factors (http://compsysbio.achs.virginia.edu/tobfac/) and serve as a significant knowledge base for use by the broader plant community.

From our analysis we determined that at least one gene family member could be identified from each of the 64 TF families currently reported to exist in flowering plants. The largest gene family is the ERF family that has at least 239 members. There are also at least 35 of the related AP2 genes. Among the other large TF families in tobacco are the R2R3MYB, bHLH, C2H2 zinc finger, NAC, homeodomain, MADS box, WRKY, bZIP and AS2 gene families. At the other end of the spectrum are the LEAFY, HRT-like, NOZZLE, VOX and whirly families which have only two members each, and the S1Fa-like family that is only represented by a single gene in our dataset. Fifty-one of the 64 identified TF families in tobacco were of similar relative size to the families present in three sequenced plant genomes (Arabidopsis, rice and poplar). Among 8 of the largest TF families in tobacco (the ERF, R2R3MYB, bHLH, NAC, homeodomain, MADS box, WRKY and bZIP families) that account for over a third of the total number of tobacco TFs, 7 families appear similar in composition to those from other vascular plants, the one notable exception being the NAC family.Thirteen

families, however, appeared to be significantly different in size when comparing our tobacco dataset to the TFs from the three complete plant genomes. Tobacco appears to have proportionally more ERF/AP2, C2H2 zinc finger, homeodomain, GRF, TCP, zinc finger homeodomain, BES and SAP genes. With some families (*e.g.*, BES and SAP) this increase is associated with the acquisition of a novel subfamily. Other TFs including the ARF, CCAAT HAP5, CPP, ZIM and PcG gene families may be under-represented in tobacco. Taken together, our data suggest that evolution of the Solanaceae was not associated with the wholesale gain or loss of TF families but rather that Solanaceae-specific, or even tobacco-specific, expansion of TF subfamilies occurred (Rushton *et al.*, 2008).

To facilitate functional genomic studies in tobacco, we developed an ultra high-density long-oligonucleotide microarray (NimbleGen Systems, Inc., Madison, WI) that contained 385,000 oligonucleotides capable of interrogating ~40,642 individual tobacco genes of known and unknown function, including the dastaset of TFs listed in TOBFAC (Rushton *et al.*, 2008) and *N. tabacum* sequences available in public databases at the time. Each predicted gene coding region was represented by six to eight 60-mer oligonucleotide probes.

The utility of this microarray for monitoring global gene expression changes in tobacco was examined using tobacco BY-2 cell suspensions, a well defined model system for the analysis of transcription circuitry and signal transduction pathways in alkaloid and isoprenoid formation (Goossens *et al.*, 2003; Hakkinen *et al.*, 2004, 2007; Matsuoka and Galis, 2006; Nagata *et al.* 2004). BY-2 cells retain their ability for phytohormonal induced "metabolic differentiation" and, in particular, are responsive to treatment with JA, a well-characterized inducer of tobacco alkaloid biosynthesis (Goossens *et al.*, 2003; Swiatek *et al.*, 2004). To investigate global gene expression in response to MeJA treatment, BY-2 cell suspension cultures were treated with 50 mM MeJA and RNA isolated at 0, 0.5, 2 , 6, 12 and 24 hours post-treatment. Using stringent criteria (triplicate samples, 95% confidence as determined using student's t-test and Benjamini and Hochberg methods) it was possible to detect the first changes in gene expression as early as 30 min post-treatment (Table 1). The uniformity of the BY-2 cells and their rapid perception and response to MeJA, resulted in very high levels of induction in our experiments and therefore a high cut off value of 8-fold induction/repression was chosen for comparison. A total of 85 genes showed 8-fold changes in expression at 30 min and 790 genes were either induced or repressed by 24 h. This wide and reproducible change in signal intensity makes these results more robust than most microarray experiments. The microarray results were validated using quantitative real-time PCR (qRT-PCR). In over 95% of the cases the level of induction or repression of gene expression observed on the microarrays was confirmed by qRT-PCR.

Table 1. The number of MeJA-induced genes in BY-2 cells using a custom microarray.

	8-fold up	8-fold 90% confidence	8-fold 95% confidence	8-fold 99% confidence
30 minutes	85	70	48	6
2 hours	352	352	347	272
6 hours	459	458	453	337
12 hours	440	439	430	305
24 hours	790	790	789	719

By comparison, Goossens *et al.* (2003) observed 376 MeJA up-regulated genes out of a total of ~20,000 transcript tags analyzed. They also observed only 83 MeJA repressed gene over the 48 h period of the experiment. Matsuoka and Galis (2006) similarly reported that 828 genes showed a 2.5-fold or higher level of induction in MeJA-treated cells relative to control cells over any of the time points used in the analysis (*i.e.*, 3, 6, 24 and 48 h post treatment with MeJA) and a total of 938 genes that were down-regulated by at least 2.5-fold in the presence of MeJA. They described six groups of genes with similar MeJA-inducible expression characteristics. Groups 1–3 contained genes with different amplitudes of induction over the 48 h of the experiment, Group 4 contained early but transiently expressed genes (3–24 h), group 5 included genes with a late response to MeJA (24–48 h), and most of the genes in Group 6 had less regular patterns of weaker induction over the later period of the experiment (24–48 h). In this analysis, Groups 1–4 mostly represent primary response genes, while Groups 5 and 6 contain late(r)-responding genes.

In our experiments, several different patterns of MeJA-inducible gene expression could be recognized. Some genes show an immediate and sustained increase in mRNA levels, whereas others show a rapid and transient increase in mRNA. These genes represent the immediate early/primary response MeJA-inducible genes and are perhaps the most interesting of the inducible genes because many are signalling molecules (*e.g.*, kinases, phosphatases, TFs). It is these genes that most likely regulate MeJA responses in tobacco.

Our detailed analysis of the genes that are significantly up-regulated by MeJA gave novel insights both into the pathways that are involved and the regulatory networks that control these pathways. In BY-2 cells, two major metabolic groups, nicotine and related alkaloids and phenylpropanoids, are known to be induced by MeJA (Goossens *et al.*, 2003; Hakkinen *et al.*, 2004; Matsuoka and Galis 2006; Galis *et al.*, 2006). The former pathway starts with PMT, the latter

with phenylalanine ammonia-lyase (PAL). We observed that treatment of BY-2 cells with MeJA induced accumulation of transcripts for genes involved in both pathways.

Genes in Nicotine Biosynthesis: Induction of nicotine and related alkaloids in BY-2 cells treated by MeJA has been previously reported although the signaling pathways that coordinate the expression of these genes remain largely unknown. We observed significant up-regulation of Nt*PMT*, Nt*QPT*, Nt*A622* and lysine decarboxylase (Nt*LDC*) genes. This is consistent with a coordinated up-regulation of nicotine biosynthetic genes as a response to MeJA treatment. By contrast, upstream of the first committed step in nicotine biosynthesis (PMT), only Nt*ODC* genes were up-regulated.

Genes in other metabolic processes: Similarly, several of the genes involved in phenylpropanoid metabolism, including *PAL*, *C4H* and *4CL-2*, were induced early and their expression remained high. Some of these genes encode metabolic enzymes involved in advanced/late phenolic metabolism (flavonoids, lignin) that are substrate-driven by gradually increasing amounts of PAL-derived phenolic substrates. In support of this, several other phenylpropanoid pathway genes, such as those encoding caffeic acid 3-O-methyltransferases, are strongly up-regulated. Our data suggest that these genes may be under common transcriptional control in MeJA-treated BY-2 cells.

Other processes that appear to up-regulated as part of the response of BY-2 cells to MeJA are flavonoid biosynthesis, terpenoid biosynthesis and wax biosynthesis. It is possible that the regulation of these pathways may share some common regulatory nodes with both the nicotine and phenylpropanoid pathways. A number of defense-related transcripts are also up-regulated by MeJA in BY-2 cells, including cysteine and serine proteases, proteinase inhibitors, elicitor-induced genes, PR proteins, ABC transporters and other genes reportedly involved in plant defense.

Phytohormones and the response to MeJA: One of the most notable aspects of the response to MeJA is the up-regulation not only of all the genes involved in JA biosynthesis itself (*e.g.*, allene oxide synthase, lipase, 12-oxophytodienoate reductase, and omega-3 fatty acid desaturase) but also of genes involved in ethylene biosynthesis and perception (ACC synthase, ACC oxidase and ethylene receptor genes). The clear message is that JA and ETH responses are being turned on. By contrast, a number of AUX induced proteins and AUX transport proteins are down-regulated. This is consistent with the previously mentioned role of AUX in antagonizing JA responses.

Transcription Factors: Our microarray analyses revealed that many TFs are strongly up-regulated by MeJA (manuscript in preparation). As part of this analysis, we have looked at the expression profiles of the group IX ERF genes because they appear to play key roles in the regulation of MeJA-inducible processes. Of the 56 Group IX ERF TF genes examined, 25 showed at least 3 x fold induction during the first 24 h of response to MeJA in BY-2 cells at some point during the time course. This demonstrates that many of the group IX ERF genes in tobacco are up-regulated as a part of the transcriptional reprogramming associated with the response to MeJA. A number of the genes show extremely high levels of induction. The five most strongly induced genes are *ERF91* (230 x fold), *ERF210* (171 x fold), *ERF29* (133 x fold), *JAP1* (96 x fold) and *ERF66* (75 x fold). It is also apparent that there is a correlation between phylogenetic position and the kinetics of induction of the genes. One clade, containing *ERF66, JAP1, ERF34* and *ERF1*, contains members that show an early maximum in mRNA levels (30 min or 2 h). In comparison, a second clade containing *ERF29, ERF91, ERF210* and *ORC1* contains members that show a late maximum in mRNA levels. All of the 25 MeJA-inducible ERF genes, except for *ERF57*, fall into these two clades with the genes in the remaining part of the ERF phylogenetic tree showing no inducibility. These data suggest that the molecular mechanisms associated with MeJA induction are different for the early maximum group compared to the late maximum group

Conclusions

The studies described above demonstrate that effective strategies exist for the successful manipulation of alkaloid levels in tobacco making the prospect of reducing some of the harm components in tobacco products a reality. The strategy of manipulation of candidate gene expression though RNAi silencing has led to the selective reduction of nicotine, nornicotine and modifications of minor alkaloid compositions, all of which contribute to the formation of TSNAs, well known to be associated with the elicitation of human disease. A microarray based approach has provided comprehensive data on changes in gene expression in various metabolic pathways associated with alkaloid and isoprenoid formation and, together with progressing metabolomic approaches, will allow efficient identification of additional gene targets. This will allow selective modification of metabolic processes underlying accumulation of harm components. Once the candidate genes are identified both transgenic and non-transgenic approaches are available to allow the rapid generation and selection of tobacco lines for inclusion in breeding programs for the development and manufacture of tobacco-leaf based harm-reduced products.

Acknowledgements

We thank various members of the Timko laboratory for their input on various aspects of this work, in particular Shengcheng Han, Ryan Thompson, Tom

Laudeman, Jim Roberts, and Xianfeng Chen. Research described in this article was supported in part by Altria Client Services Inc.

REFERENCES

1. Baldwin IT (1988) Short-term damage-induced increases in tobacco alkaloids protect plants. Oecologia 75: 367-370.
2. Baldwin IT, Schmelz EA, and Ohnmeiss TE (1994) Wound-induced changes in root and shoot jasmonic acid pools correlate with induced nicotine synthesis in Nicotiana-sylvestris Spegazzini and Comes. J Chem Ecol 20: 2139-2157.
3. Bartecchi CE, MacKenzie TD and Schrier RW (1995) The global tobacco epidemic. Sci Am 272: 44–51.
4. Bombick DW, Bombick BR, Ayres PH, Putnam K, Avalos J, Borgerding MF, and Doolittle DJ (1997) Evaluation of the genotoxic and cytotoxic potential of mainstream whole smoke and smoke condensate from a cigarette containing a novel carbon filter. Fundam Appl Toxicol 39: 11-17.
5. Breland AB, Buchhalter AR, Evans SE, and Eissenberg T (2002) Evaluating acute effects of potential reduced-exposure products for smokers: Clinical laboratory methodology. Nicotine & Tobacco Res 4: S131-S140.
6. Chintapakorn Y, and Hamill JD (2003) Antisense-mediated downregulation of putrescine N-methyltransferase activity in transgenic *Nicotiana tabacum* L. can lead to elevated levels of anatabine at the expense of nicotine. Plant Mol Biol 53: 87-105.
7. Chintapakorn Y, and Hamill JD (2007) Antisense-mediated reduction in ADC activity causes minor alterations in the alkaloid profile of cultured hairy roots and regenerated transgenic plants of *Nicotiana tabacum*. Phytochemistry 68: 2465-2479.
8. Chini A, Fonseca S, Fernández G, Adie B, Chico JM Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, Micol JL, and Solano R (2007) The JAZ family of repressors is the missing link in jasmonate signalling . Nature 448: 666-671.
9. Colbert T, Till BJ, Tompa R et al (2001) High-throughput screening for induced point mutations. Plant Physiol 126: 480-484.
10. Comai L, and Henikoff S (2006) TILLING: practical single nucleotide mutation discovery. Plant J 45: 684-694.
11. De Boer KD, Lye JC, Aitken CD, Su AK, and Hamill JD (2009) The A622 gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Mol Biol 69: 299-312.
12. Djordjevic M, Hoffmann D, and Hoffmann I (1997) Nicotine regulates smoking patterns. Prev Med 26: 435-440.

13. De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inze D, Goossens A, and Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J 44: 1065-1076.
14. Facchini PJ (2001) Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu Rev Plant Physiol Plant Molec Biol 52: 29-66.
15. Galis I, Simek P, Narisawa T, Sasaki M, Horiguchi T, Fukuda H, and Matsuoka K (2006) A novel R2R3 MYB transcription factor NtMYBJS1 is a methyl jasmonate-dependent regulator ofphenylpropanoid-conjugate biosynthesis in tobacco. Plant J 46: 573-592.
16. Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Soderlund H, Zabeau M, Inze D, and Oksman-Caldentey KM (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci(USA) 100: 8595–8600.
17. Goossens A, and Rischer H (2007) Implementation of functional genomics for gene discovery in alkaloid producing plants. Phytochem Rev 6: 35–49.
18. Hakkinen ST, Rischer H, Laakso I, Maaheimo H, Seppanen-Laakso T, and Oksman-Cladentey KM (2004) Anatalline and other methyl jasmonate-inducible nicotine alkaloids from Nicotiana tabacum cv. BY-2 cell cultures. Planta Med 70:936–941.
19. Hakkinen ST, Tilleman S, Swiatek A, De Sutter V, Rischer H, Vanhoutte I, Van Onckelen H, Hilson P, Inze D, Oksman-Caldentey K, and Gossens A (2007) Functional characterization of genes involved in pyridine alkaloid biosynthesis in tobacco. Phytochem 68: 2773–2785.
20. Hashimoto T, and Yamada Y (2003) New genes in alkaloid metabolism and transport. Curr Opin Biotechnol 14:163–168.
21. Hatsukami DK, Ebbert JO, Feuer RM, Stepanov I and Hecht SS (2007) Changing Smokeless Tobacco Products New Tobacco-Delivery Systems. Am J Prev Med 33: S368–S378.
22. Hecht SS (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer 3: 733–744.
23. Hecht SS, Carmella SG, Murphy SE, Riley WT, Le C, Luo XH, Mooney M, and Hatsukami DK (2007) Similar exposure to a tobacco-specific carcinogen in smokeless tobacco users and cigarette smokers. Cancer Epidemiol Biomarkers Prev 16: 1567–1572.
24. Hoffmann D, Djordjevic MV, and Hoffmann I (1997) The changing cigarette. Prev Med 26: 427–434.
25. Hibi N, Higashiguchi S, Hashimoto T, and Yamada Y (1994) Gene expression in tobacco low-nicotine mutants. Plant Cell 6:723–735.

26. Howe GA, and Jander G (2008) Plant immunity to insect herbivores. Annu Rev Plant Biol 59: 41-66.
27. Imanishi S, Hashizume K, Nakakita M, Kojima H, Matsubayashi Y, Hashimoto T, Sakagami Y, Yamada Y, and Nakamura K (1998) Differential induction by methyl jasmonate of genes encoding ornithine decarboxylase and other enzymes involved in nicotine biosynthesis in tobacco cell cultures. Plant Mol Biol 38: 1101–1111.
28. Julio E, Laporte F, Reis S, Rothan C, and Dorlhac de Borne F (2008) Reducing the content of nornicotine in tobacco via targeted mutation breeding. Mol Breeding 21: 369–381.
29. Kajikawa M, Hirai N, and Hashimoto T (2009) A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Mol Biol 69: 287–298
30. Katoh A, Ohki H, Inai K, and Hashimoto T (2005) Molecular regulation of nicotine biosynthesis. Plant Biotechnol 22: 389–392.
31. Katsir L, Chung HS, Koo AJ, and Howe GA (2008) Jasmonate signaling: a conserved mechanism of hormone sensing. Curr Opin Plant Biol 11:428-435.
32. Kessler A, and Baldwin IT (2002) Plant responses to insect herbivory: the emerging molecular analysis. Annu Rev Plant Biol 53: 299–328.
33. Kidd SK, Melillo AA, Lu R-H, Reed DG, Kuno N, Uchida K, Furuya M, and Jelesko JG (2006) The A and B loci in tobacco regulate a network of stress response genes, few of which are associated with nicotine biosynthesis. Plant Mol Biol 60: 699-716.
34. Lewis RS, Jack AM, Morris JW, Robert VJM, Gavilano LB, Siminszky B, Bush LP, Hayes AJ, and Dewey RE (2008) RNA interference (RNAi)-induced suppression of nicotine demethylase activity reduces levels of a key carcinogen in cured tobacco leaves. Plant Biotech Jour 6: 346-354.
35. Matsuoka K, and Galis I, (2006) VI.2 EST and Microarray Analysis of Tobacco BY-2 Cells. In Biotechnology in Agriculture and Forestry, Vol. 5 Tobacco BY-2 Cells: From Cellular Dynamics to Omics (ed. by T. Nagata, K. Matsuoka, and D. Inzé), Springer-Verlag Berlin Heidelberg pp. 293-311.
36. Morita M, Shitana N, Sawada K, Van Montagu MCE, Inze D, Rischer H, Goossens A, Oksman-Caldentey K-M, Moriyama Y, and Yazaki K (2009) Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in Nicotiana tabacum. Proc Natl Acad Sci (USA) 106: 2447-2452.
37. Nagata T, Hasezawa S, and Inze D, eds. (2004) Tobacco BY-2 cells. Biotechnology in Agriculture and Forestry, vol. 53. Springer, Berlin Heidelberg New York.
38. Pauw B, and Memelink J (2004) Jasmonate-responsive gene expression. J Plant Growth Regul 23: 200-210.

39. Pederson LL, and Nelson DE. (2007). Literature review and summary of perceptions, attitudes, beliefs, and marketing of potentially reduced exposure products: communication implications. Nicotine Tob Res 9: 525-534.
40. Prignot JJ, Sasco AJ, Poulet E, Gupta PC, and Aditama TY (2008) Alternative forms of tobacco use. Int J Tuberc Lung Dis 12: 718-772.
41. Riechers DE, and Timko MP (1999) Structure and expression of the gene family encoding putrescine N-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. Plant Molecular Biology 41: 387-401.
42. Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, Laudeman TW, and Timko MP (2008) Tobacco transcription factors: Novel insights into transcriptional regulation in the Solanaceae. Plant Physiol 147: 280-295.
43. Sato F, Hashimoto T, Hachiya A, Tamura K, Choi KB, Morishige T, Fujimoto H, and Yamada Y (2001) Metabolic engineering of plant alkaloid biosynthesis. Proc Natl Acad Sci (USA) 98: 367-372.
44. Shi Q, Li C, and Zhang F (2006) Nicotine synthesis in *Nicotiana tabacum* L. induced by mechanical wounding is regulated by auxin. J Exp Bot 57: 2899-2907.
45. Shoji T, Yamada Y, and Hashimoto T (2000) Jasmonate induction of putrescine N-methyltransferase genes in the roots of *Nicotiana sylvestris*. Plant Cell Physiol 41: 831-839.
46. Shoji T, Winz R, Iwase T, Nakajima K, Yamada Y, and Hashimoto T (2002) Expression patterns of two tobacco isoflavone reductase like genes and their possible roles in secondary metabolism in tobacco. Plant Mol Biol 50: 427-440.
47. Shoji T, Ogawa T, and Hashimoto T (2008) Jasmonate-induced nicotine formation in tobacco is mediated by tobacco *COI1* and *JAZ* genes. Plant Cell Physiol 49: 1003-1012.
48. Sinclair SJ, Murphy KJ, Birch CD, and Hamill JD (2000) Molecular characterization of quinolinate phosphoribosyltransferase (QPRTase) in Nicotiana. Plant Mol Biol 44: 603–617.
49. Sisson VA, and Severson RF (1990) Alkaloid composition of the *Nicotiana* species. Beitr Tabakforsch 14: 327-339.
50. Stepanov I, Jensen J, Hatsukami D, and Hecht SS (2008) New and traditional smokeless tobacco: Comparison of toxicant and carcinogen levels. Nicotine & Tobacco Res 10: 1773-1782.
51. Steppuhn A, Gase K, Krock B, Halitschke R, and Baldwin IT (2004) Nicotine's defensive function in nature. PLoS Biol 2:1074-1080.
52. Stratton K, Shetty P, Wallace R, and Bondurant S, eds. (2001) Clearing the smoke: assessing the science base for tobacco harm reduction. Institute of Medicine (IOM). Washington DC: National Academies Press.

53. Swiatek A, Van Dongen W, Esmans EL, and Van Onckelen H (2004) Metabolic fate of jasmonates in tobacco bright yellow-2 cells. Plant Physiol 135:161-172.
54. Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, and Browse J (2007) JAZ repressor proteins are targets of the SCF^{COI1} complex during jasmonate signalling. Nature 448: 661-665.
55. Thurston R, Smith WT, and Cooper BP (1966) Alkaloid secretion by trichomes of Nicotiana species and resistance to aphids. Entomologia Experimentalis et Applicata 9: 428-432.
56. Todd AT, Liu E, Polvi SL, Pammett RT, and Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J 62: 589-600.
57. van der Vaart H, Postma DS, Timens W, and Ten Hacken NH (2004) Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax 59: 713–721.
58. Voelckel C, Krugel T, Gase K, Heidrich N, van Dam NM, Winz R, and Baldwin IT (2001) Anti-sense expression of putrescine N-methyltransferase confirms defensive role of nicotine in *Nicotiana sylvestris* against *Manduca sexta*. Chemoecology 11: 121-126.
59. Wang P, Liang Z, Zeng J, Li W, Sun X, Miao Z and Tang K 2008 Generation of tobacco lines with widely different reduction in nicotine levels via RNA silencing approaches. J Biosci 33: 177-184.
60. Wang P, Zeng J, Liang Z, Miao Z, Sun X, and Tang K (2009) Silencing of PMT expression caused a surge of anatabine accumulation in tobacco. Mol Biol Rep 36: 2285-2289.
61. Wynder EL, and Hoffmann D (1979) Tobacco and health: a societal challenge. N Engl J Med 300: 894-903.
62. Xie J, Song W, Maksymowicz W, Jin W, Cheah K, Chen W, Carnes C, Ke J, and Conkling MA. (2004) Biotechnology: A tool for reduced risk tobacco products - The nicotine experience from test tube to cigarette pack. Recent Adv Tobacco Sci 30: 17-37.
63. Xie DX, Feys BF, James S, Nieto-Rostro M, and Turner JG (1998) *COI1*: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science 280: 1091-1094.
64. Xu B, Timko M (2004) Methyl jasmonate induced expression of the tobacco putrescine N -methyltransferase genes requires both G-box and GCC-motif elements. Plant Mol Biol 55: 743-761.
65. Xu BF, Sheehan MJ, and Timko MP (2004) Differential induction of ornithine decarboxylase (ODC) gene family members in transgenic tobacco (*Nicotiana tabacum* L. cv. Bright Yellow 2) cell suspensions by methyl-jasmonate treatment. Plant Growth Regul 44:101-116.

66. Xu LH, Liu FQ, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang DF, and Xie DX (2002) The SCF^{COI1} ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis. Plant Cell 14: 1919-1935.
67. Yan J, Zhang C, Gu M, Bai Z, Zhang W, Qi T, Cheng Z, Peng W, Luo H, Nan F, Wang Z, Xie D (2009) The Arabidopsis CORONATINE INSENSITIVE1 protein is a jasmonate receptor. Plant Cell 21: 2220-2236.

ENGINEERING HIGH VALUE OIL PRODUCTION IN TOBACCO

Shuiqin Wu[1], Tom Niehaus, Scott Kinison, Satrio Husodo[2],
Robert Williams[3] and Joe Chappell
Department of Plant & Soil Sciences
University of Kentucky
Lexington, KY 40546-0312

Current Addresses:
[1]Sapphire Energy, Inc. San Diego, CA
[2]IBS Graduate Program, University of Kentucky, Lexington, KY
[3]Medical School, University of Kentucky, Lexington, KY

Abstract

Assuming biofuels generated via the fermentation of sugars derived from cellulosic and non-cellulosic constituents of biofuels crops will provide a substantial contribution to our future energy needs, augmenting and amending the productivity of these biofuel crops is now a major research thrust worldwide. One way of enhancing these biofuels crops will be to engineer them for value-added components such as oils that can be used for efficient fuel production and the manufacturing of other high-value products currently derived from petroleum oils. Towards this end, we have developed an engineering strategy for optimized production of long, branched-chain hydrocarbon biosynthesis in plants. Branched chain hydrocarbons, like methylated triterpenes, are readily cracked into paraffins and naphthenes that can either be distilled to combustible fuels (gasoline, JP-8 and diesel), or can be used directly for the synthesis of plastics, nylons, paints and other oil-derived products manufactured by diverse chemical industries. Our working hypothesis has been that success in generating high level production platforms for triterpene oils in plants can be accomplished by targeting this metabolism to the chloroplast compartment of cells, thus eliminating the regulatory mechanisms that normally operate to control this metabolism occurring in the cytoplasm, and providing a means for the direct channeling of photosynthetically fixed CO_2 to the biosynthesis of novel, value-added products. Preliminary experiments suggests that engineering this metabolism into chloroplasts of photosynthetically active mesophyll cells can result in plants yielding up to 0.1% of their dry weight as triterpenes. In comparison, plants having this metabolism engineered into secretory trichomes can yield up to 1% of their dry weight as triterpenes. These results have important ramifications for our understanding of basic metabolism in plants, as well as the development of novel chemical production platforms in plants.

Backdrop

One intention of this chapter is to suggest that the molecular genetic tools of today are sufficient to recapitulate the biochemistry that existed some 600 million years ago at the bottom of the then ocean floor. But why would we think it is important to understand how natural products were biosynthesized so long ago? Well, if we understand the biochemical transformations that occurred in progenitor enzymes and pathways in comparison to extant ones, then we might be able to infer what changes have occurred in fashioning evolutionary advantages to these pathways, enzymes and the organisms harboring them. Such information might also allow us to craft analogous means for evolving new functionalities for particular desired outcomes. The second intention is to address a more practical problem in the field of metabolic engineering. Why can't we deregulate metabolic pathways in plants like that done with microbes to create a much greater production capacity for a desired chemical constituent? Plant production systems would seem to have several important advantages over microbial platforms. Plants do not require an exogenous feedstock for growth. Plant systems are easily scaled from 0.1 acre to 1,000s of acres. And, the infrastructure costs for agricultural production are generally much lower than that for fermentation. Our contention is that plant production platforms will not be considered seriously until production capacity for a chemical target is equivalent to 1% of the biomass weight. Production at such levels alleviate all kinds of downstream processing concerns and allow one to think that a 30,000 Kg yield of biomass per acre leads to 300 Kg of a desired chemical product. Such a yield would obviously be attractive for the production of specialty, high value compounds, but it also allows one to start thinking about using plants to produce more commodity type compounds rather than just specialty, high price ones.

With these two seemingly disparate notions swirling in the readers mind, the remainder of this chapter will walk the reader through a rationale for our interest in an ancient algal species, our efforts to capture the genetic blueprints for the biosynthesis of a unique oil produced by this algae, and then a brief summary of our progress in engineering this unique capacity into tobacco.

Introducing *Botryococcus braunii*

Botryococcus braunii is a colony-forming, freshwater, green microalgae with individual cells of the colony held together by an extracellular matrix consisting of liquid hydrocarbons embedded within a cross-linked hydrocarbon polymer core (Fig. 1) [Maxwell *et al.*, 1968; Knights *et al.*, 1970]. Previous studies have classified *B. braunii* into three races (A, B, and L) depending on the type of hydrocarbon associated with the extracellular matrix [Metzger *et al.*, 1985, 1990]. Triterpene hydrocarbons, 30 carbon polymers decorated with up to 4 additional methyl substituents and collectively referred to as botryococcenes and methylated squalenes, are the major matrix components associated with the B

race (Fig. 2). Race A accumulates alkadienes and alkatrienes derived from fatty acids, and race L accumulates tetraterpenes like lycopadiene [Metzger *et al.* 1990, Templier *et al.*, 1984, 1991].

Fig. 1. Light micrograph of a Botryococcus braunii Race B colony. Cells containing single, cup-shaped chloroplasts are embedded in an oil-laden matrix (fluorescence from nile red stain). Micrograph kindly provided by Taylor Weiss and Tim Devarenne from TAMU.

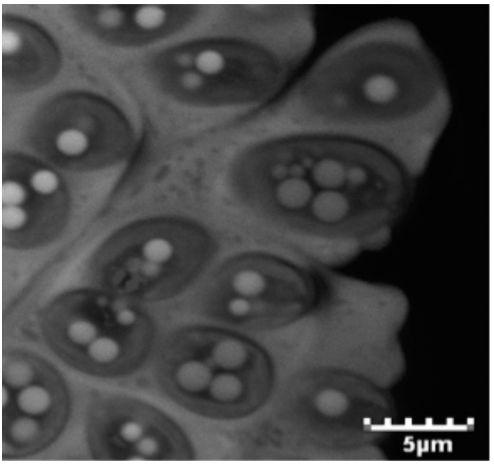
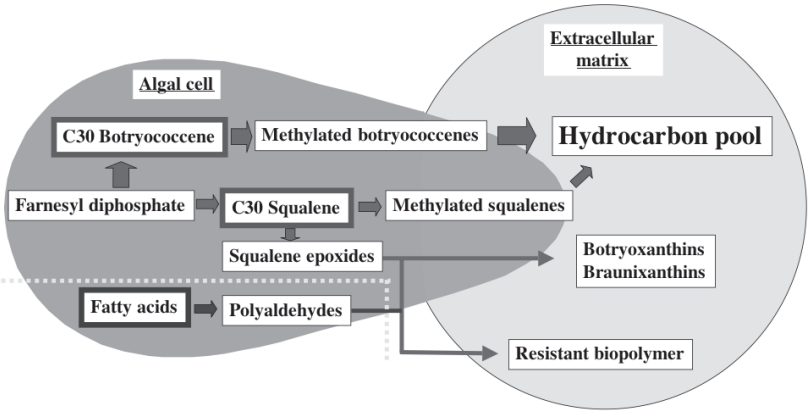


Fig. 2. Depiction of the production and deposition hydrocarbon related metabolites by Botryococcus braunii race B.



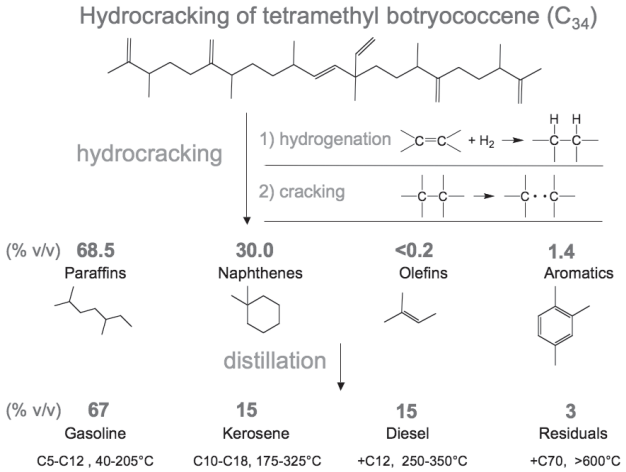
The B race has attracted special attention for several reasons:
First, this race accumulates botryococcene and squalene derivatives up to 30–40% of their dry weight [Metzger *et al.*, 1985; Okada *et al.*, 1995]. Engineering production platforms for natural products in terrestrial plants can certainly benefit from a better understanding of the mechanisms underlying the accumulation of particular substances to such high levels as observed with *Botryococcus.*

Second, catalytic hydrolysis or cracking (as performed in standard chemical and petroleum refineries) of these highly branched, poly-unsaturated triterpenes results in the generation of aliphatic and aromatic molecules (illustrated in Fig. 3). The alkanes and alkenes are key substrates for many chemical industrial applications. For instance, these are the essential monomeric units utilized in polymer chemistry for plastics and nylon production. The alkenes also lend themselves to further chemical modifications, such a halogenation and introduction of other substituent groups. Alcohols formed from oxygenation of the alkenes are the basis for many household cleaning products. The uses for these hydrolytic products are very diverse and extensive. However, the greatest utility of these branched-chain hydrocarbons are as fuels. Hillen *et al.* [1982] previously reported on the catalytic cracking of the triterpene oils extracted from the green alga *Botryococcus* (primarily the botryococcenes and squalene derivatives) and observed an overall conversion of 97% of the oil to combustible grade fuels under standard cracking conditions. Overall, 67% of the converted oil was to gasoline grade fuel, 15% to aviation turbine fuel, 15% to diesel fuel with a residue of only 3%.

Third, *B. braunii* represents an evolutionarily ancient organism (originating >600 million years ago), and the botryococcenes and methylated squalenes of *B. braunii* have been found in many fossil petroleum deposits [Summons *et al.*, 2002]. This information has lead to speculations that ancient hydrocarbon pools of *B. braunii* were major progenitors to the current day oil shale and coal deposits [Walters *et al.*, 2005].

Fourth, *B. braunii* grows very slowly with doubling times on the order of days relative to hours or fractions of a day for other algae like *Chlamydomonas reinhardtii* and *Chloroella* species. Whether this is a consequence of its large allocation of energy and resources to the accumulation of prodigious levels of triterpene oils (greater than 30% of its dry weight), its growth as a colony rather than single cells, or whether this arises from a lack of adaptation to changing environmental conditions over evolutionary time remains unknown. Nonetheless, *B. braunii* presumably flourished during the Ordovician time period (433 to 488 million years ago) when CO_2 levels were 15 times higher than they are today, overall atmospheric conditions much more reducing, and water temperatures considerably higher than they are today.

Fig. 3. Catalytic processing ("cracking") of tetramethylated botryococcene isolated from Botryococcus braunii race B yields petroleum distillate-like products that can be used directly as fuels, for the manufacturing of industrial polymers (i.e. nylons, plastics), and for the generation of alcohols and other chemicals used in a wide range of industrial products (cleaning and manufacturing fluids). Adapted from Hillen et al. [1982].



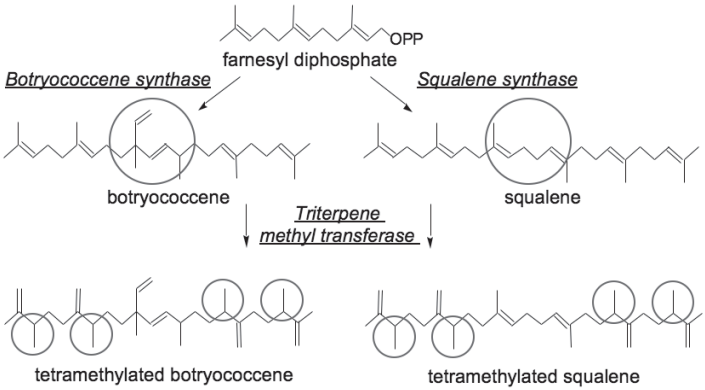
Taking all this information together, though the hydrocarbon oils of *Brotyococcus braunii* race B are highly valued as a direct feedstock for bio-fuels and bio-based manufacturing, large-scale cultivation of this alga does not seem realistic. However, capturing the genetic blueprints for the biosynthesis of this unique oil and deploying it via genetic engineering methodologies into other organisms capable of rapid growth and high yields represents an obvious and logical alternative for creating a sustainable production platform for these triterpene oils.

Biochemical Studies with *B. braunii* – *Characterization of the triterpene synthases*

The potential use of the botryococcene and methylated squalene derivatives as a renewable hydrocarbon source has led to many studies on the chemistry and biosynthesis of these compounds. Because of the similarity between the structure of squalene and botryococcene, our particular studies have focused on mechanistic and molecular comparisons between the biosynthesis of these two families of compounds (Fig. 4). Both botryococcene and squalene are 30 carbon compounds with a common backbone of two 15-carbon farnesyl residues. However, they differ in the linkage between the farnesyl residues; squalene has a 1'-1 linkage and botryococcene has a 1'-3 linkage. Squalene is synthesized by

the enzyme squalene synthase (SS) in a two-step reaction in which the first step condenses two molecules of farnesyl diphosphate (FPP) to form presqualene diphosphate (PSPP) [Robinson *et al.*, 1993; Gu *et al.*, 1998]. The second step cleaves the cyclopropane ring in PSPP to form the 1'-1 linkage coincident with a NADPH dependent reduction. From our own results, the formation of the 1'-3 linkage found in C30 botryococcene can be rationalized by a similar two-step reaction carried out by a SS-like catalyst or enzyme [Okada *et al.*, 2004].

Fig. 4. The biosynthetic pathways for botryococcenes and tetramethylsqualene. Both classes of methylated triterpenes are derived from an initial condensation of 2 FPPs, which then undergo a unique modification by a novel methyltransferase activity.



We previously prepared a high quality cDNA library against mRNA isolated from *B. braunii* race B cultures in rapid growth phase and have used this important tool for the isolation and functional characterization of genes involved in the unique triterpene metabolism of B. braunii. For example, utilizing PCR strategies we have isolated and characterized a full-length squalene synthase (BSS) [Okada *et al.*, 2000, 2004] and more recently a putative *Botryococcus* botryococcene-synthase cDNA (BBS) (Fig. 5). Functional characterization of the BBS cDNA included expression of this gene in bacteria, but we only observed a modest level of apparent BBS enzyme activity in these bacterial lysates (Fig. 6). In comparison, lysates of *Botryococcus* algal cells exhibit a slightly greater level of NADPH dependent botryococcene synthase activity. However, when equal aliquots of the *E. coli* and *B. braunii* lysates were mixed, a very significant stimulation of NADPH dependent botryococcene synthase activity was observed (Fig. 6).

Fig. 5. Sequence comparison between B. braunii squalene synthase (BSS) and botryococcene synthase (BBS) relative to other SS enzymes with an emphasis on those domains correlated with catalytic/functional activities (upper panel). Note that domains III and IV, those known to be associated with the formation of the pre-squalene intermediate and reduction to the final product exhibit the greatest sequence diversity.

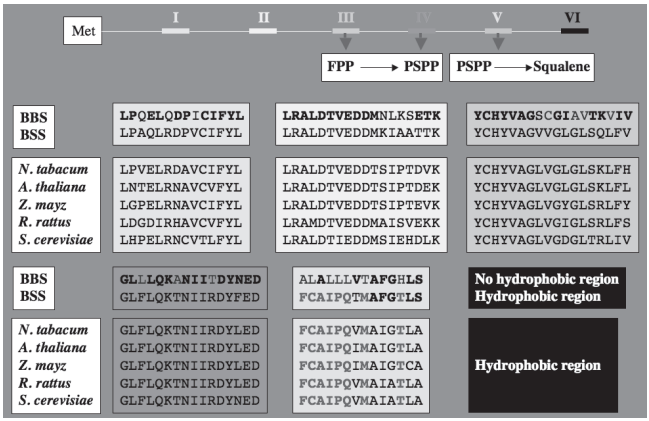
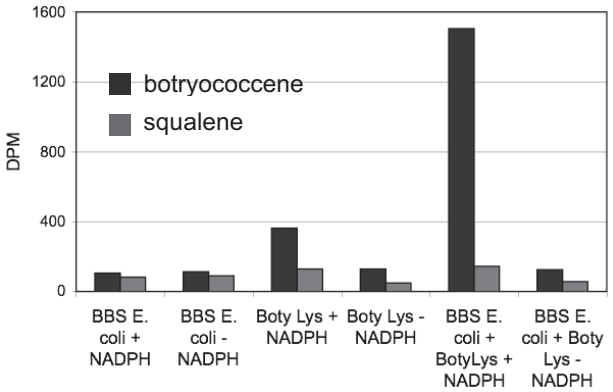


Fig. 6. The cDNA encoding for the BBS enzyme illustrated in Fig. 5 was expressed in bacteria, a cell lysate prepared and used as the BBS enzyme source (BBS E. coli). Botryococcus algal cells were also homogenized and a 2,000xg supernatant prepared (Boty Lys). Samples were incubated with 3H-FPP, +/- NADPH, and the amount of radioactivity incorporated into squalene and botryococcene determined by TLC separation/scintillation counting.



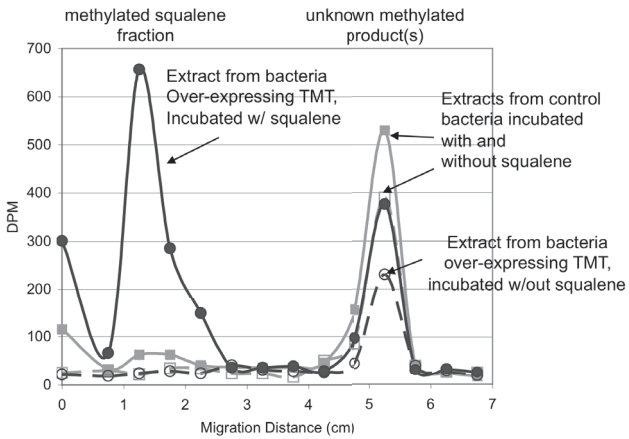
More recent work has focused on identifying what the stimulatory factor for the *Botryococcus* BBS enzyme activity might be. The enhanced activity is dependent

on the integrity of both the purified BBS enzyme and the algal lysate. Proteinase K and heat treatment of either the purified protein or algal lysate abolished the restoration of botryococcene synthase activity. We thus suspect that the *Botryococcus* lysate is providing a unique proteinaceous factor to complete the botryococcene synthase enzyme activity. That co-factor may serve to supply reducing equivalents to the enzyme, or it may actively participate in the catalytic conversion of PSPP to botryococcene, or both.

Characterization of a Triterpene Methyltransferase

Unlike their tetramethylated derivatives, catalytic hydrolysis of squalene and botryococcene does not yield high-rated (combustible) octanes. Instead, smaller, less combustible molecules are generated [Hillen *et al*., 1982]. "Cracking" the tetramethylated forms of squalene and botryococcene, in contrast, yield a much higher quality petroleum-like product. Hence, a renewable production platform for petroleum-like oils not only requires engineering a high level production platform for triterpenes, but also their biosynthetic conversion to their tetramethylated forms. To date, methyltransferase enzymes responsible for the methylation of proteins, phenolic derivatives and sterols have been well characterized [Grillo & Colombetto, 2005; Willits *et al*., 2004; Bouvier-Nave *et al*., 1998], but a squalene or botryococcene methyltransferase has not.

Fig. 7. TLC separation of reaction products generated upon incubation of extracts from bacteria over-expressing the TMT gene with (solid black symbols) and without squalene (open black symbols). Extracts of bacteria transformed with empty vector (no TMT gene) were also incubated under identical conditions with (solid grey symbols) and without (open grey symbols) squalene addition. Reactions products extracted from the respective reactions were separated by reverse phase TLC and the radioactivity associated with the indicated zones determined by scintillation counting.



From the approximate 600 randomly selected and sequenced cDNAs from the existing *B. braunii* cDNA library, many of the sequences exhibited high sequence similarities to enzymes involved in photosynthesis and other housekeeping functions. However, a significant number of the cDNAs exhibited sequence similarities to enzymes potentially involved in triterpene metabolism. One cDNA possessed sequence similarity to known sterol methyl-transferases, but also contained differences associated with the putative substrate specificity domains. A full-length clone of this cDNA was subsequently isolated and has been demonstrated to encode for the squalene/botryococcene methyltransferase (triterpene methyltransferase or TMT) activity (Fig. 7).

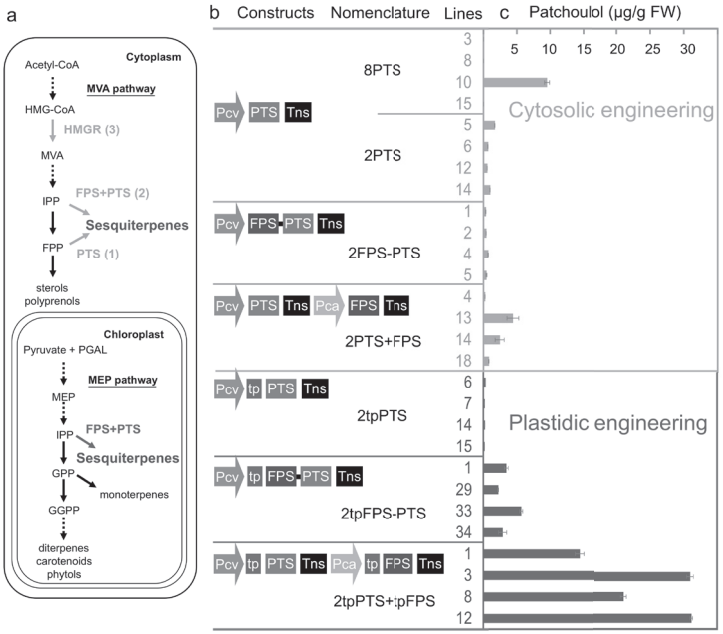
Genetic Engineering Update

Terpene biosynthesis is a complicated process associated with two independent biosynthetic pathways (Fig. 8a). In all eukaryotes, the mevalonate (MVA) pathway operates in the cytoplasm and is responsible for the biosynthesis of the universal C_5 building blocks isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are used for the biosynthesis of larger compounds like sesquiterpenes (C_{15}), sterols (C_{30}) and dolichols ($C_{40\text{-}50}$). In prokaryotes, hopanes (C_{30} equivalents to sterols), sesquiterpenes and other terpenes are derived from the methyl-erythritol phosphate (MEP) pathway. Interestingly, both pathways exist in plants, the MVA pathway in the cytoplasm and the MEP pathway in the chloroplast, and each has evolved an intriguing division of labor providing for the complex array of molecules essential for plant growth and development and those responsible for mediating interactions between plants and their environment. The cytosolic pathway (MVA) is predominately responsible for the generation of terpenes like sterols and sesquiterpenes, compounds consisting of 30 (C_{30}) or 15 (C_{15}) carbons, respectively. Monoterpenes (C_{10}), diterpenes (C_{20}) and carotenoids (C_{40}) are synthesized via the plastidic pathway (MEP) (Fig. 8a).

There have been many efforts to manipulate terpene metabolism in microbes [Martin *et al.*, 2003] and plants [Zook *et al.*, 1996, Wallaart *et al.*, 2001] in part, to evaluate the contribution of specific classes of molecules to the organisms' life cycle, and as a means for functional characterization of putative biosynthetic genes. Our particular interest in engineering these pathways has been to develop production platforms providing sufficient materials for industrial and commercial uses. In 2006, we reported on the development of a strategy to engineer high-level production of sesquiterpenes in transgenic plants [Wu *et al.*, 2006]. We assumed that regulation might be imposed by feedback mechanisms operating in the downstream portions of these pathways, and especially by FPP itself [Gardner and Hampton, 1991]. We therefore developed a strategy to divert the first common intermediates in both pathways, IPP and DMAPP, intermediates which serve as the precursors for FPP biosynthesis in the cytosol, and geranyl diphosphate (GPP) and geranylgeranyl diphosphate (GGPP) biosynthesis in the

chloroplast. We assumed that perhaps previous efforts to engineer robust cyclic terpene production by directly diverting FPP from the MVA pathway, and GPP or GGPP from the MEP pathway in chloroplasts were stymied in part because these endogenous pools are small and highly regulated.

Fig. 8. Strategies for diverting carbon from the mevalonate (cytosolic) or the methyl-erythritol phosphate (plastidic) pathways for novel sesquiterpene biosynthesis as reported by Wu et al. [2006]. A schematic outline of the two terpene biosynthetic pathways operating in plants, their intracellular locations, and the genetic manipulations used to engineer in the MVA and MEP pathways (a). Gene constructs used to engineer sesquiterpene metabolism in the cytoplasm consisted of the patchoulol synthase (PTS) gene, the farnesyl diphosphate synthase (FPS) gene or a gene fusion of PTS and FPS (PTS-FPS) inserted downstream of strong, constitutive promoters (Pca, 35S cauliflower mosaic viral promoter; Pcv, cassava vein mosaic viral promoter), and were identical for plastid engineering except a plastid targeting signal sequence was fused to the 5' end of the respective genes (b). Wild type tobacco (Nicotiana tabacum cv. Xanthi, 14-2) and a line (14-8) previously engineered for over-expression of a truncated 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) gene were transformed with the respective gene constructs and antibiotic selected R1 lines were assessed from patchoulol accumulation by GC-MS (c).



Gene constructs consisting of patchoulol synthase (PTS), a sesquiterpene synthase, directed by a strong, constitutive promoter, by itself and in combination with an avian FPP synthase gene (FPS) [Tarshis *et al.*, 1994], either as separate genes or as a gene fusion with PTS were prepared (Fig. 8b). These initial gene constructs were devoid of any obvious intracellular targeting signals, hence the encoded enzymes were targeted to the cytoplasmic compartment (*i.e.* PTS). The engineered genes were also further modified by appending a chloroplast targeting signal sequence (tp) to the amino terminus of the proteins (*i.e.* tpPTS) to target the encoded proteins to the MEP pathway. The importance of carbon flux within the MVA pathway was also determined by utilizing plant lines harboring a truncated form of 3-hydroxy-3-methylglutaryl CoA reductase, HMGR (line 14-8), which we had previously shown to accumulate upwards of a 10-fold higher level of sterols and sterol biosynthetic intermediates than wildtype (14-2) plants [Chappell *et al.*, 1995].

Multiple independent transgenic lines were generated per construct transformation, and the resulting transgenic lines evaluated for terpene accumulation by GC-MS analysis. Simply targeting the PTS enzyme to the cytosolic compartment resulted in transgenic lines capable of accumulating low, but readily detectable levels of patchoulol (average 0.5 μg/g F.W.) (Fig. 8c). Coupled over-expression of PTS with the avian FPP synthase increased patchoulol accumulation 2- to 5-fold over those plants harboring only PTS, and this was increased another 2- to 6-fold in a single transgenic line over-expressing of a truncate form of HMGR. Co-expression of un-regulated forms of PTS, FPS and HMGR thus appear to relieve important constraints on carbon flux through the MVA pathway to sesquiterpene biosynthesis.

Targeting the PTS enzyme to the plastid compartment also yielded plants accumulating patchoulol, which was somewhat unexpected since FPP biosynthesis in chloroplasts has not been documented. More significantly, however, co-expression of PTS and FPP synthase targeted to the chloroplast, either as a protein fusion or as independent enzymes, resulted in transgenic lines accumulating 5 to 30 μg of patchoulol per gm fresh weight of plant material, representing 100 to 1,000 times more sesquiterpene accumulation than previously reported for a genetic engineered plant [Aharoni *et al.*, 2003].

Patchoulol is a volatile sesquiterpene and we determined in head-space gas analyses that upwards of 75% of the patchoulol synthesized per day was lost to volatile emission. Therefore, the actual levels of patchoulol accumulating reflects the minimal biosynthetic rate of the engineered sesquiterpene, which could be as much as 4 times greater than as reported in Fig. 8.

Evidence for the Engineering of Triterpene Metabolism

Given the success of engineering sesquiterpene metabolism, we have begun to extend this strategy and approach to other terpene classes, including linear triterpenes. The target for these initial studies was squalene, and our aim was to determine if an analogous strategy as used for sesquiterpenes was applicable to squalene. The first construct iterations were designed to over-express squalene synthase in combination with the avian farnesyl diphosphate synthase, with the encoded enzymes targeted either to the MVA pathway (cytosol) or to the MEP pathway (plastid) (as illustrated in Fig. 9). Because squalene synthase is natively tethered to the endoplasmic reticulum via a carboxy-terminal hydrophobic domain, we first screened truncated versions of squalene synthase for soluble catalytic activity. This was accomplished by inserting 3' truncated versions of the rat, yeast, tobacco, *Arabidopsis* and *Botryococcus* squalene synthase genes into a bacterial expression vector, then evaluating the level of soluble squalene synthase activity in the bacterial lysates after gene induction treatment. The truncated yeast squalene synthase gene (ySQS) yielded the highest level of soluble squalene synthase activity in comparison to the others and hence was chosen for our subsequent vector construction (Fig. 9). Expression of the cytosolic and plastid targeted (tp, transit peptide of the Arabidopsis RuBisCO small-subunit protein) genes were driven by either strong constitutive promoters (cauliflower/cassava mosaic viral promoters [Wu *et al.* 2006]) or trichome specific promoters (cbts, cembratrienol synthase [Ennajdaoui *et al.*, 2010] and cyt16, diterpene hydroxylase [Wang *et al.*, 2001]). The respective expression cassettes were then used to generate independent tobacco transgenic lines and the initial Ro plants were screened for their ability to accumulate squalene by GC-MS analysis. Example data for this preliminary screen are presented in Fig. 9.

Squalene does not accumulate to any appreciable level in wild type plants and borders on being within our technical detection limits of GC-MS. While plants engineered with the squalene synthase and FPS enzymes targeted to the cytosolic MVA pathway appeared normal in terms of growth habit and stature, 16 of 30 lines surveyed accumulated upwards of 10-times higher levels of squalene than the wild type controls. In comparison, about 20% of the plants engineered for plastid targeting of the SS and FPS proteins exhibited some growth abnormalities, mostly stunted growth. Seven of the R0 transgenic lines tested up to this point accumulated much higher levels (≥100x) of squalene than the control plants or those having their cytosolic MVA pathway engineered (Fig. 9). Interestingly, no correlation between growth characteristics and squalene accumulation was observed. For instance, plant line #15 grew comparable to control plant while plant line #7 was stunted. Nonetheless, plant #15 accumulated 1,200 times more squalene than control plants while #7 accumulated about a 300-fold greater amount. A modest, yet positive correlation between leaf developmental age and squalene accumulation was also evident for the lines engineered for squalene

biosynthesis in the chloroplasts, but much less so for the cytosolic engineered lines.

Fig. 9. Preliminary evidence for diverting carbon flux from the mevalonate (MVA, cytosolic) or the methyl-erythritol phosphate (MEP, plastidic) pathways for novel squalene biosynthesis and accumulation. Gene constructs used to engineer squalene metabolism in the cytoplasm consist of a truncated form of the yeast squalene synthase (ySQS) gene and the avain farnesyl diphosphate synthase (FPS) gene inserted downstream of strong, constitutive promoters (Pca, 35S cauliflower mosaic viral promoter; Pcv, cassava vein mosaic viral promoter) or enhanced, trichome specific promoters (e, 35S enhancer; cbts, cyp16, cembriene synthase and hydroxylase promoters, respectively), and were identical for plastid engineering except a plastid targeting signal sequence (tp) was fused to the 5' end of the respective genes. Wild type tobacco (Nicotiana tabacum accession 1068) was transformed with the respective gene constructs and antibiotic selected R0 lines propagated in the greenhouse were assessed from squalene accumulation by GC-MS.

construct	Plant line	Leaf development	Squalene (µg/gm fr.wt.)
Wild type	wt	intermediate	0.4
Constitutive, cytosolic (MVA) targeted Pcv – ySQS – Pca – FPS	# 16	young	6.4
		intermediate	5.5
		mature	7.4
Constitutive, plastidic (MEP) targeted Pcv – tp ySQS – Pca – tp FPS	# 7	young	30.1
		intermediate	121.2
		mature	147.4
	# 15	young	329.3
		intermediate	450.4
		mature	667.5
Trichome, plastidic (MEP) targeted e Pcbts1 – tp ySQS – e Pcyp16 – tp FPS	# 31	young	527.6
		intermediate	594.7
		mature	1,760.2

By far the greatest amount of squalene accumulation was documented for plants engineered for trichome specific expression of the squalene synthase and FPS proteins targeted to the chloroplast compartment. The intent of this expression vector design was to facilitate trichome specific expression in hopes that whatever squalene might be produced, it would be secreted and/or sequestered similarly to the diterpenes and sucrose esters that normally accumulate as leaf exudates [Wagner *et al.*, 2004]. While we currently do not know where the squalene is accumulating (inside the trichomes, inside the plastid compartment, or secreted to the leaf surface), the observed levels of approximately 1 mg squalene per g fresh weight of leaf tissue are upwards of several 1,000-fold higher than in the wild type

control plants. The accumulation pattern also exhibits a modest correlation with leaf development and some of the high and low squalene accumulating plants do exhibit various degrees of stunted growth.

If one assumes water weight accounts for about 90% of the leaf fresh weight, then the levels of squalene accumulating (0.5-1.7 mg/ g fresh weight) approach 5 to 10 mg per g dry weight, which corresponds to approximately 0.5 to 1% of the plant material dry weight. Such a calculation is undoubtedly preliminary at this stage. Nonetheless, one milestone often discussed as being necessary for the commercialization of chemical targets produced in plants is accumulation to 1 to 4% of the plant dry weight [Horsch, 1993; Snell and Peoples, 2009].

Final Remarks

The aim of this chapter was to introduce the rationale and means to engineering terrestrial plants with the ability to accumulate high levels of oils. Not fatty acid derived oils, but triterpene oils because of their demonstrated potential as direct drop-in biofuels, and because of their potential as renewable feedstocks for chemical, manufacturing, and food industries [Knights *et al.*, 1970; Hillen *et al.*, 1982; Sharma *et al.*, 1994]. We also hope our genetic engineering approach is taken as illustrative of how new value-added products can be added to a crop plant. That is, engineering triterpene oils as a possible trichome exudate should lend itself to an extraction process leaving relatively intact plant biomass behind for cell wall/sugar and other processing, and thus enhancing the value of the crop. Lastly, while the results presented here are for squalene accumulation in transgenic plants, our ultimate objective is in generating transgenic lines accumulating methylated triterpenes and especially methylated botryococcenes. Having these genes now in hand provides just such an opportunity.

Acknowledgements

The work described here in the Chappell laboratory was supported by grants from the NSF.

REFERENCES CITED:

1. Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel WJ, Verstappen FWA, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ. (2003) Terpenoid metabolism in wild-type and transgenic arabidopsis plants. Plant Cell 15:2866-2884.
2. Bouvier-Nave P, Husselstein T, Benveniste P. (1998) Two families of sterol methyltransferases are involved in the first and the second methylation steps of plant sterol biosynthesis. European Journal Of Biochemistry 256:88-96.
3. Chappell J, Wolf F, Proulx J, Cuellar R, Saunders C. (1995) Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl coenzyme-a reductase a rate-limiting step for isoprenoid biosynthesis in plants. Plant Physiology 109:1337-1343.

4. Ennajdaoui H, Vachon G, Giacalone C, Besse I, Sallaud C, Herzog M, Tissier A. (2010) Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. Plant Mol Biol (in press)
5. Gardner RG, Shan H, Matsuda SPT, Hampton RY (2001) An oxysterol-derived positive signal for 3-hydroxy-3-methylglutaryl-CoA reductase degradation in yeast. J Biol Chem 276: 8681-8694
6. Grillo MA, Colombatto S. (2005) S-adenosylmethionine and protein methylation. Amino Acids 28:357-362.
7. Gu PD, Ishii Y, Spencer TA, Shechter I. (1998) Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. J Biol Chem 273:12515-12525.
8. Hillen LW, Pollard G, Wake LV, White N. (1982) Hydrocracking of the oils of botryococcus-braunii to transport fuels. Biotechnol Bioeng 24:193-205.
9. Horsch RB. (1993) Commercialization of genetically-engineered crops. Philos Trans R Soc Lond Ser B-Biol Sci 342:287-291.
10. Knights BA, Brown AC, Conway E, Middledi.Bs. (1970) Hydrocarbons from green form of freshwater alga botryococcus-braunii. Phytochemistry 9:1317-&.
11. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. (2003) Engineering a mevalonate pathway in escherichia coli for production of terpenoids. Nat Biotechnol 21:796-802.
12. Maxwell JR, Douglas AG, Eglinton G, McCormic.A. (1968) Botryococcenes-hydrocarbons of novel structure from alga botryococcus braunii kutzing. Phytochemistry 7:2157-2171.
13. Metzger P, Allard B, Casadevall E, Berkaloff C, Coute A. (1990) Structure and chemistry of a new chemical race of botryococcus-braunii (chlorophyceae) that produces lycopadiene, a tetraterpenoid hydrocarbon. Journal Of Phycology 26:258-266.
14. Metzger P, Berkaloff C, Casadevall E, Coute A. (1985) Alkadiene-producing and botryococcene-producing races of wild strains of *botryococcus-braunii*. Phytochemistry 24:2305-2312.
15. Okada S, Devarenne TP, Chappell J. (2000) Molecular characterization of squalene synthase from the green microalga *botryococcus braunii*, race b. Arch Biochem Biophys 373:307-317.
16. Okada S, Devarenne TP, Murakami M, Abe H, Chappell J. (2004) Characterization of botryococcene synthase enzyme activity, a squalene synthase-like activity from the green microalga *botryococcus braunii*, race b. Arch Biochem Biophys 422:110-118.
17. Okada S, Murakami M, Yamaguchi K. (1995) Hydrocarbon composition of newly isolated strains of the green microalga *botryococcus braunii*. Journal Of Applied Phycology 7:555-559.

18. Robinson GW, Tsay YH, Kienzle BK, Smithmonroy CA, Bishop RW. (1993) Conservation between human and fungal squalene synthetases - similarities in structure, function, and regulation. Molecular And Cellular Biology 13:2706-2717.
19. Sharma DK, Behera BK, Arora M. (1994) Hydrogen-Transfer Hydrocracking of C-Procera Latex under Ambient-Pressure Conditions to Get Value-Added Chemicals and Fuels. Fuel Sci. Technol. Int. 12: 1131-1155
20. Snell KD, Peoples OP. (2009) Pha bioplastic: A value-added coproduct for biomass biorefineries. Biofuels Bioprod Biorefining 3:456-467.
21. Summons RE, Metzger P, Largeau C, Murray AP, Hope JM. (2002) Polymethylsqualanes from *botryococcus braunii* in lacustrine sediments and crude oils. Organic Geochemistry 33:99-109.
22. Tarshis LC, Yan MJ, Poulter CD, Sacchettini JC. (1994) Crystal-structure of recombinant farnesyl diphosphate synthase at 2.6-angstrom resolution. Biochemistry 33:10871-10877.
23. Templier J, Largeau C, Casadevall E. (1984) Hydrocarbon formation in the green-alga *botryococcus-braunii* .4. Mechanism of non-isoprenoid hydrocarbon biosynthesis in *botryococcus-braunii*. Phytochemistry 23:1017-1028.
24. Templier J, Largeau C, Casadevall E. (1991) Biosynthesis of normal-alkatrienes in *botryococcus-braunii*. Phytochemistry 30:2209-2215.
25. Wagner GJ, Wang E, Shepherd RW. (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot 93:3-11.
26. Wallaart TE, Bouwmeester HJ, Hille J, Poppinga L, Maijers NCA. (2001) Amorpha-4,11-diene synthase: Cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. Planta 212:460-465.
27. Walters CC, Curry DJ, Isaksen GH. (2005) Modification of the petroleum system concept: Origins of alkanes and isoprenoids in crude oils: Discussion. Aapg Bulletin 89:1239-1244.
28. Wang EM, Gan SS, Wagner GJ. (2002) Isolation and characterization of the cyp71d16 trichome-specific promoter from *Nicotiana tabacum* l. J Exp Bot 53:1891-1897.
29. Willits MG, Giovanni M, Prata RTN, Kramer CM, De Luca V, Steffens JC, Graser G. (2004) Bio-fermentation of modified flavonoids: An example of in vivo diversification of secondary metabolites. Phytochemistry 65:31-41.
30. Wu SQ, Schalk M, Clark A, Miles RB, Coates R, Chappell J. (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat Biotechnol 24:1441-1447.
31. Zook M, Hohn T, Bonnen A, Tsuji J, Hammerschmidt R. (1996) Characterization of novel sesquiterpenoid biosynthesis in tobacco expressing a fungal sesquiterpene synthase. Plant Physiology 112:311-318.

TOXICOGENOMICS AND EMERGING TECHNOLOGIES IN TOXICOLOGICAL RESEARCH

Wanda R. Fields
R. J. Reynolds Tobacco, Co., Winston-Salem, NC 27105

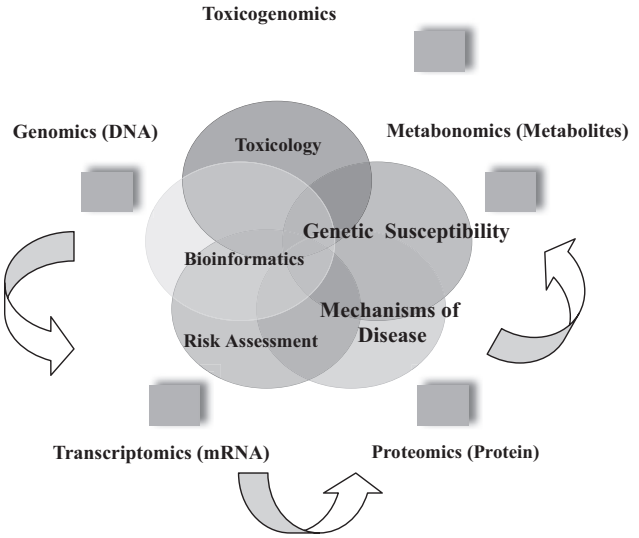
Abstract
Research efforts over the last several decades have significantly impacted and accelerated the study of biology. In particular, the early study of macromolecules spawned a revolution in the field of molecular biology. Molecular biology in simple terms is the study of life at the molecular level. Fundamental to molecular biology is the understanding that genetic material is maintained in the form of DNA and can be transcribed to RNA and finally translated into protein, providing a flow of biological information from the genetic code to cellular function. Toxicological testing and biomarker discovery as well as drug development and therapy have been influenced by advances in molecular biological techniques such as "–omics" technologies. These technologies utilize a broad range of molecular tools to support the study of various macromolecules or subcellular components in response to exposures and treatments. The leading -omic disciplines are genomics (DNA), transcriptomics (RNA), proteomics (protein) and metabolomics (metabolites).

Toxicogenomics combines genomics and bioinformatics to characterize and identify mechanisms of toxicity induced by various exposures. Within this scientific discipline, genomics, transcriptomics, proteomics and metabolomics may be applied to assess the toxicological responses of chemicals using *in vitro* and *in vivo* models. Data from such studies can be analyzed by tools that combine biology, computer software and statistics to generate biological information that support efforts in predictive toxicology, mechanistic toxicology, biomarker discovery, risk assessment, identification of mechanisms of disease and drug development. This presentation will provide a summary of toxicogenomics in toxicology and tobacco research. Standard testing measures, challenges and current advances will be discussed.

Introduction
Toxicological testing, biomarker discovery, drug development and therapy have been influenced by advances in molecular biological techniques such as "–omics" technologies. The four major –omics disciplines are defined as follows: 1) **genomics:** the study of gene sequences and genetic variability (humans possess ~30,000 genes), 2) **transcriptomics:** the quantitative measurement of gene expression (mRNA) in a cell or tissue by various measures, 3) **proteomics:** the measurement of protein production and levels in a cell, tissue or biologic fluid, and 4) **metabonomics:** the multiparametric measurement of metabolites.

The Human Genome Project which led to the sequencing of the entire human genome has fostered the expansion of toxicological testing to include **toxicogenomics.** Toxicogenomics combines genomics and bioinformatics to characterize and identify mechanisms of toxicity induced by various chemicals including drugs. Within this scientific discipline, genomics, transcriptomics, proteomics and metabonomics have been applied to assess the toxicological responses of chemicals using *in vitro* and *in vivo* models and to assist with predictive toxicology, mechanistic toxicology, and risk assessment (Figure 1; Burcyzynski *et al.*, 2000; Farr and Dunn, 1999; Burczynsky 2004; Moggs *et al.*, 2004; Kramer *et al.*, 2004; Inoue *et al.*, 2002; Medlin *et al.*, 2002; Waters *et al.*, 2003; Pennie *et al.*, 2002; Darvas *et al.*, 2004).

Figure 1: Interdisciplinary aspects of Toxicological Analyses.



The development of technologies used in toxicology and toxicogenomics has evolved significantly over the last several decades. Technologies that could only be dreamed of only 20 to 30 years ago are standard practice in today's laboratories.

This review will describe the current technologies used in toxicogenomics and provide examples as they relate to tobacco research where appropriate. Key technologies to be discussed include polymerase chain reaction (PCR) and quantitative reverse transcriptase/polymerase (qRT-PCR), microarrays (transcriptomics), methylation-specific PCR (DNA methylation), protein arrays (proteomics), and metabolic profiling (metabonomics) (Table I). Extensions and applications of these technologies will be discussed as they relate to bioinfomatics, transgenics, genomics, emerging technologies and systems biology.

Table I - Technologies used in Toxicogenomic Research

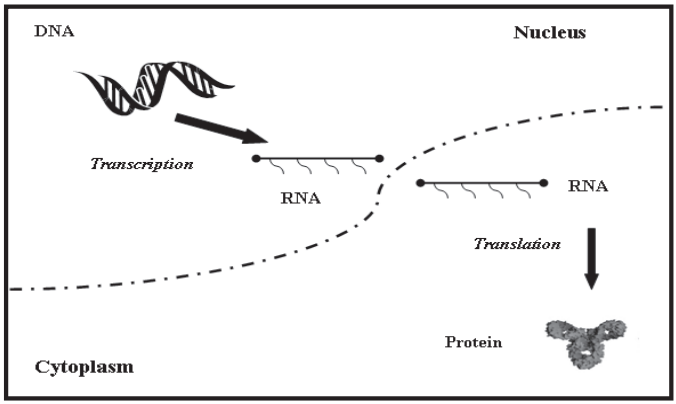
Technique	Application	Advantages	Limitations
Quantitative RT/PCR	Gene Expression	Screen Multiple Samples; High-throughput screening	Limited gene targets per assay – even with multiplexing and gene cards; PCR array is improving this issue
SiRNA/RNAi	Gene Silencing	*In vitro* assessment can be performed to test efficacy prior to *in vivo* assessments; drug discovery	Transfection efficiencies; toxicity in certain cases
Microarray (filter, chip)	Genomics (Transcripts, single nucleotide polymorphisms (SNPS), DNA polymorphisms)	High-throughput screening of gene targets - thousands of genes	High-throughput analyses (*i.e.* many samples) are expensive; Potentially large data set generation with limited understanding of biological impact.
Protein Array (filter, chip, solution)	Proteomics	Profiling of multiple proteins	Some difficulties in correlating with mRNA changes; variations in time requirements for regulation of mRNA (transcription/stability) versus protein (translation, modifications)
Metabolic Profiling (NMR, Mass Spectrometry)	Metabonomics	Profiling of multiple metabolites; provides understanding of gene functions	Requires expensive/ specialized equipment
Methylation-specific PCR	Epigenetic DNA modification (*i.e.* methylation)	Provides gene specific analysis	Multistep process leading to potential for sample lost; limited to gene specific characterization - global methylation patterns must be assessed by other measures – Advances are in progress

Polymerase Chain Reaction and Quantitative RT/PCR

The polymerase chain reaction (PCR) is a technique used to amplify specific DNA sequences with the aid of gene specific primers (small sequences of DNA specifically designed for a gene target), heat-stable DNA polymerase enzymes, nucleotides and other reagents. The process is often performed as a precursor to DNA isolation, forensic analysis and disease diagnosis. It has been an integral part of *in vitro* toxicology and other scientific disciplines as applied to: 1) gene expression analysis from *in vitro* and *in vivo* samples, 2) *in situ* procedures of paraffin-embedded tissue, 3) gene cloning and 4) genotyping since the nineties (Vanden Heuvel, 1997).

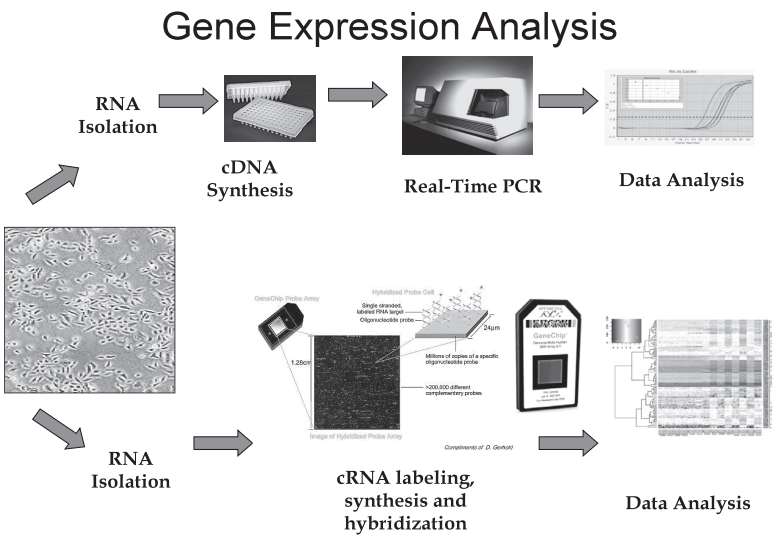
Gene expression allows for genes which are coded as DNA to be expressed as RNA and ultimately protein which serves as the functional product of the DNA code (Figure 2). Gene expression analysis requires the measurement of messenger RNA (mRNA) or protein. Since the late 90's, quantitative reverse transcription polymerase chain reaction (qRT-PCR) has dominated the field in assessing steady state mRNA levels for individual gene analyses. In this method, RNA is reverse transcribed (RT) to a complementary DNA copy which is then amplified by PCR. Advances in chemical components developed to modify primers, probes and reaction materials have resulted in significant improvements in quantitative and qualitative assessment of gene expression, thereby overcoming the prior limitations associated with detecting changes in mRNA levels (Ferre *et al.*, 1994; Heid *et al.*, 1996; Figure 3).

Figure 2. Cellular Context of Gene Expression. RNA is produced in the nucleus through a process termed transcription in which the DNA code is transcribed to RNA. The RNA is subsequently transported to the cytoplasm where it is translated to protein which completes the chain for functional expression of the DNA (gene) following post-translational processing.



Advances in quantitative qRT-PCR facilitated the adoption of gene expression technology into toxicological testing and led to advanced methods of large scale gene expression profiling via microarrays and gene chips.

Figure 3. Advances in Gene Expression Technologies. Upper Panel. Real-time quantitative RT/PCR. RNA is reversed transcribed to a complimentary DNA (cDNA) copy and subsequently amplified in the presence of buffers, enzymes, DNA bases and fluorescent reagents vial real-time. Representative data analysis is depicted by the quantitation of mRNA facilitated by amplification of a relative standard curve prepared with positive control cDNA. ***Lower Panel.*** *Affymetrix gene chip expression analysis is conducted with RNA that is labeled, subjected to synthesis and hybridized to a specific gene chip. Various normalization, statistical and bioinformatic data analysis steps are conducted to simultaneously generate gene expression data from thousands of genes which can be subsequently represented via hierarchical data comparisons (as depicted) for visual representation of differentially expressed genes.*



Microarray Technology

Microarrays are grids of DNA or DNA fragments aligned on a chip to evaluate large amounts of biological information. Microarray technology has progressed from the use of nylon membranes formed to the size of an index card to gene chips designed as small as a letter-sized return address label. The arrays allow for the simultaneous assessment of genes/transcripts (*i.e.* 1,000 – 30,000, entire genomes) involved in specific focused categories such as signal transduction (process that transfers the external cellular impact of chemical exposure to internal compartments and effects), stress response, cell cycle regulation, DNA

synthesis/repair, inflammatory responses and metabolism. Arrays are designed for different species (*i.e.* human, mouse, rat) and for disease and stress states (*i.e.* cancer, toxicology, inflammation). Array designs are aided by genome sequences from public databases and unique probe designs, and are analyzed using bioinformatics techniques. Photolithographic techniques (adapted from the semiconductor industry) are used in the manufacturing process of microarray chips (Figure 3; Kreiner *et al.*, 2005). Aided by various data analysis methods including cluster and hierarchical analyses, researchers can generate details on interactions in gene pathways.

Validation and standardization are important aspects of microarray research that are still under discussion. These issues addressed by Rockett and Hellmann (2004) include measures for validation (quantitative RT/PCR, western or northern blots, in silico analyses), uniform formats for RNA amplification, normalization techniques, data analysis and representation, journal requirements for publication of array data and assay design and standardizations for across platform comparisons. The Minimum Information About a Microarray Experiment (MIAME) checklist (www.mged.org) provides a good reference framework for microarray investigations.

DNA Methylation

Epigenetic modifications are changes in DNA that do not modify the gene sequence but may alter gene expression or phenotypes (characteristics). DNA methylation is a process by which a methyl group is covalently added to the 5-carbon position of cytosine in a cytosine-guanine (CpG) sequence of base pairs. The methylation alters gene expression by preventing transcription of the gene. This process is referred to as gene "silencing". This process is catalyzed by DNA methyltransferase (Bird, 1992; Jones and Takai, 2001) and may be reversible. Since specific ratios of methylation in GC-regions of DNA are required for normal maintenance of gene expression, global changes observed by hyper- and hypo-methylation of genomic DNA can indicate epigenetic alterations induced by chemicals during disease progression (Watson *et al.*, 2004). Hyper- and hypo-methylation patterns can lead to phenotypic changes that affect toxicological response to chemical exposure and may also define disease pathogenesis (*i.e.* carcinogenesis, aging, etc.) (Watson and Goodman, 2002; Watson *et al.*, 2004, Jones and Baylin 2002; Issa, 2000; Richardson, 2003).

DNA methylation changes may be either global or gene specific (*i.e.* p16, GST π, HOXA5). Global methylation---the overall level of methyl cytosines in a genome--- may be determined via methyl-accepting capacity assays (SssI DNA methyltransferase) and chromatographic assays (high-performance liquid chromatography, thin-layer chromatography or liquid chromatography/

mass spectroscopy). The latter assay requires digestion of the DNA into single nucleotides.

Numerous methods have been developed for analyzing gene-specific methylation including bisulfite based methods. Bisulfite can selectively deaminate cytosine but not 5-methylcytosine to uracil. This leads to a sequence change in the DNA that discriminates cytosine from 5-methylcytosine. After the conversion has taken place, sequence differences between a methylated and unmethylated cytosine can be assessed by several methods including direct sequencing, restriction enzyme digestion, nucleotide extension assays, and methylation-specific PCR (MSP).

Cells can be exposed *in vitro* to chemicals causing demethylation and then analyzed for reversal patterns. Such investigations have led to the understanding of the efficacy of certain drug therapies as well as the level of methylation required to alter cellular functions. *In vivo* investigations have studied methylation (either permanent or reversible) in genes controlling cell proliferation, cell cycling, and in disease states such as lung cancer and skin tumorigenesis.

Proteomics

Proteomics is the measurement of protein production and levels in a cell, tissue or biological fluid. These measurements are important since proteins are the expressed characteristics of genes and function as effecter molecules in cells. A variety of protein alterations result from cellular responses to stimuli and stress. For example, cells may secrete cytokines or chemokines, proteins that play a role in immune response and inflammation. Alternatively, proteins may be modified by phosphorylation, a process that controls protein activity by the addition of phosphate molecules and supports the transfer of signaling through a cell.

Tissue profiling by proteomic techniques has led to the detection of biomarkers as well as establishment of specific signal pathway profiling (Bichsel *et al.*, 2001; Caldwell and Caprioli, 2005). Early proteomic technologies consisted of 2-D gel electrophoresis followed by liquid chromatography/mass spectrometry [LC/MS] and subsequent biomarker validation with western blot (protein detection via antibodies on protein binding membranes) or enzyme-linked immunoassay (ELISA; protein detection in plastic multi-welled plates coated with protein binding matrix layered with antibodies). Recent technologies provide time-saving, higher-throughput and increased reproducibility (for example, matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) and surface enhanced laser desorption/ionization (SELDI)-based protein chip technologies). These technologies allow for the identification of proteins in biological matrixes by analysis of proteins digested into smaller units (peptides) which can be distinguished on a mass (molecular weight) basis.

Protein arrays (filter, chip and suspension) facilitate screening protein profiles of tissue samples and biological fluids on a smaller scale. Protein arrays are used to assess biomarkers of diseases and pathogenic states, including alterations in expression, cytokines/chemokines and post-translational protein modifications (e.g., phosphorylation, transcription factors). They may be used to assess the effects of candidate drugs and environmental toxicants/irritants on the biology of cells, research animals and human subjects, and may be applied to functional genomics, validation requirements in drug discovery and toxicology. Protein filter arrays are similar to microarrays with the screening process performed by antibody probes instead of labeled DNA probes. However, compared to microarrays, the number of targets is dramatically reduced due to difficulties related to protein chemistry of each antibody protein interaction.

Proteomics has sub-disciplines or categories as well. For example, "kinomics" combines genomics and proteomics for the study of kinases, proteins that function as enzymes to support phosphorylation of macromolecules. Kinases promote the activation of proteins involved in transducing signals from external to internal cellular environments. This sub-discipline provides an important complement to genomic assessments since some gene products exhibit modified expression and activity past the mRNA state; hence, complete understanding of the cellular responses to treatments or chemical exposures requires a combination of technologies that measure gene and gene product changes across a continuum.

Metabonomics

The need to advance the understanding of the biological relevance of genomic data is ever present. Metabonomics is a seminal technology in functional genomics – a discipline which involves assessing gene function and its control mechanisms. As the genome represents all of the gene sequences of an organism, the metabolome represents all the low-molecular weight molecules in a cell at a given time. Analysis of the metabolome was born from two similar yet distinct methods that have evolved from the need to assess animal and microbial/plant biochemistry. Metabonomics and metabolomics involve characterization of metabolites through multiparametric measurements. Metabonomics deals with biological systems, including extracellular environments in an integrated and multicellular approach allowing for quantitative measurements of metabolic responses following a stimulus. Metabolomics predominately deals with concentrations of intracellular metabolites in simple cell systems. In recent years, the terms have been used interchangeably and most often referred to as metabonomics.

In concert with chemometric and bioinformatics tools, metabonomic data can generate biochemically-based fingerprints (biomarkers) of drugs and toxicants

and hence support biologically relevant endpoints to data obtained from various "-omics" technologies. Metabonomics offers the promise for obtaining drug-effect and disease endpoints to support drug development, risk and toxicological assessments in living systems (Nicholson *et al.*, 1999, 2002) via analysis of biofluids, tissues and cell lysates. This technology has been applied in ecotoxicological assessment of environmental contaminants (Bundy *et al.*, 2004), physiological influences on biofluids (Bollard *et al.*, 2004), systems biology in pharmaceutical research (Lindon *et al.*, 2004a), physiological monitoring, drug safety assessments and disease diagnosis (Lindon *et al.*, 2004b), toxicological assessments (Keun, *et al.*, 2004; Holmes *et al.*, 2001), and cigarette smoke effects in cultured lung cells (Vulmiri *et al.*, 2009).

Applications and Extensions of Toxicogenomic Technologies

Gene Expression and Genomics

Differential gene expression profiling and genomic evaluations via PCR, qRT/PCR, microarrays have been conducted in *in vitro*, *in vivo* and human samples to support chemical evaluations, risk assessment, drug discovery, genetic susceptibility and analysis of tissue for disease pathogenesis (Table II).

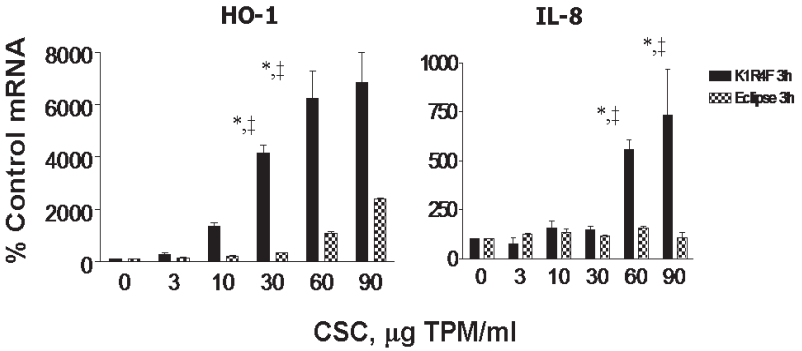
Complex mixtures of chemicals such as cigarette smoke condensate (CSC) and diesel exhaust extracts can alter gene expression in lung cells as detected by qRT/PCR. For example, human bronchial epithelial cells exposed to CSC exhibited changes in interleukin (IL-8) mRNA and secreted cytokine levels, changes associated with inflammation (Hellermann *et al.*, 2002). Individual smoke constituents may also cause alterations. *c-myc* (a gene that promotes uncontrollable cellular growth) was evaluated in normal human bronchial epithelial cell cultures exposed to benzo[a]pyrene and benzo[a]pyrene diol epoxide. Differential mRNA responses were observed between the chemicals and correlated with distinctions in DNA adduct accumulation and cell cycle regulation (Fields *et al.*, 2004).

Table II – Toxicological gene expression and methods of detection

Agent/Disease	Model	Method	Reference
Diesel Exhaust	*In vitro* (human airway cells; rat alveolar macrophages)	Semi qRT-PCR	Baulig *et al.*, 2003; Koike *et al.*, 2002
Bromobenzene	*In vivo* (rats)	Transcriptomics; Proteomics	Heijne *et al.*, 2003
Cytotoxic Anti-inflammatory Drugs; DNA damaging agents	*In vitro* (HepG2)	cDNA microarray	Burczynski *et al.*, 2000
Arsenic	*In vivo* (Tg.AC mice)	DNA methylation	Xie Y *et al.*, 2004
Benzo[a] pyrene	*In vitro*	qRT-PCR	Fields *et al.*, 2004
Lung Cancer	*In vitro*	cDNA microarray	Hellmann *et al.*, 2001
Lung Cancer	*In vivo*	cDNA microarray; qRT-PCR	Powell *et al.*, 2003; Dressman *et al.*, 2006; Potti *et al.*, 2006;Gustafson *et al.*, 2010; Spira *et al.*, 2010
COPD	*In vivo*	cDNA microarray	Steiling *et al.*, 2009; Silverman *et al.* 2009
Cigarette Smoke	*In vivo*	cDNA microarray; Gene Chip	Hackett *et al.*, 2003; Shah *et al.*, 2005; Gebel *et al.*, 2004; Beane *et al.*, 2007
Cigarette Smoke Condensate/ Extracts; Cigarette Smoke	*In vitro* (human bronchial cells; SWISS 3T3 cells)	real-time qRT-PCR; PCR array; cDNA microarray	Fields *et al.*, (2005; 2009); Maunders *et al.*, (2007); Parsanejad *et al.*, (2008a and 2008b); Bosio *et al.*, 2002
Asbestos	*In vitro*	cDNA microarray, real-time PCR	Ramos-Nino *et al.*, 2003
Oxidative Stress	*In vitro* (Hep-G2)	cDNA microarray, real-time PCR	Morgan *et al.*, 2002

Fukano *et al.* (2006) evaluated the effect of different cigarette filters on gene expression in in vitro lung cultures. Cells exposed to smoke from cigarettes with a carbon-filter exhibited reduced activation of heme oxygenase 1 (HO-1), a gene responsive to oxidative stress, compared to cells exposed to smoke from cigarettes with cellulose acetate filters. In studies conducted at our laboratory, changes were observed in HO-1 as well as genes associated with inflammation (COX-2, IL-8) and stress (heat shock protein 70, HSP-70) in lung cells exposed to burn versus primarily-heat cigarettes. (Figure 4; Fields *et al.*, 2005a, Fields *et al.*, 2005b).

Figure 4: Differential expression of HO-1 and IL-8 in NHBE cells following CSC exposure. Cells were exposed to 3, 10, 30, 60 or 90 μg TPM/ml or DMSO (0.9%) in growth media continuously for 3 hours. Percent of control values were determined for each gene by comparing the normalized fluorescence values of the solvent control sample to each treated sample. Asterisks (*) denote statistical significance compared to the solvent control, and superscript symbol (‡) denotes statistical significance between K1R4F and Eclipse, using one-way ANOVA with the Bonferoni adjustment. The level of statistical significance was expressed at p< 0.05. (Data are represented as the mean + SEM.)

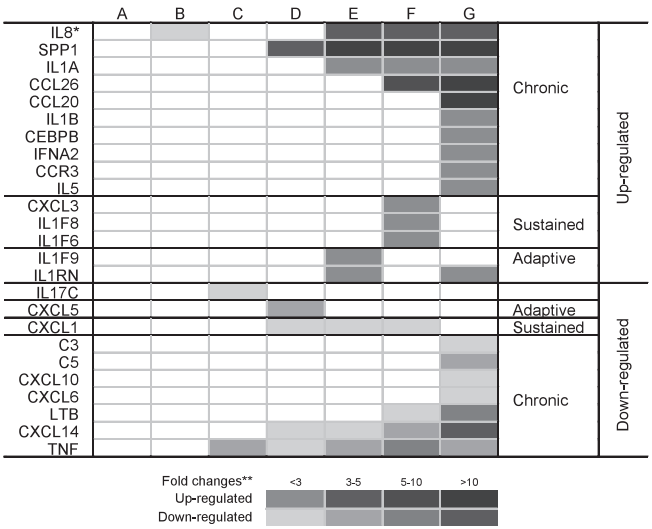


Taken together, qRT-PCR of selected tobacco responsive genes can be an effective assessment for the evaluation of potential reduced exposure products (PREPs).

Pathway specific analyses have recently been enhanced by PCR array technologies and may be used in tobacco-related research. Typical platforms such as the PCR arrays from SA Biosciences offer 96 and 384-well formats to allow for quantitative and high-throughput screening for specific disease states, cellular conditions, signaling pathways, biological functions, gene regulation and DNA methylation. The Signal Transduction Pathway Finder PCR array and Inflammatory Cytokine PCR Array were used by Parsanejad *et al.* (2008a and 2008b) to determine the

effect of tobacco smoke on selected molecular pathways. Assessment of whole smoke in NHBE cells yielded responses in pathways associated with inflammation, apoptosis and wound healing. Our group subsequently evaluated a repetitive smoke exposure regimen to mimic a smoker's frequency of exposure (Figure 5; Parsanejad *et al.*, 2008b). The resulting cytokine responses exhibited adaptive, sustained and chronic expression patterns over the time-course of exposure.

*Figure 5: Inflammatory Cytokine PCR Array Responses to CSC exposure in NHBE Cells. Gene expression patterns in NHBE cells displaying adaptive, sustained and chronic expression in response to 60 µg TPM/ml media. Groups A to G represent different treatment methodologies from 15 minutes to alternating repeated exposure and recovery times over a 24 hour period. In this study, adaptive gene regulation refers to genes that adapted to the CSC exposure and short-term recovery (3 hours) by returning to the baseline; sustained gene regulation refers to genes that remained up- or down-regulated only during CSC exposure and short-term recovery; and chronic gene regulation refers to genes initiating or maintaining elevation or suppression patterns after a 12-hour (long-term) recovery. * IL8 was the exception in expression pattern by exhibiting both up- and down-regulation in the 7 groups. ** The fold-changes are scaled by color intensity. Upregulation indicates gene expression is increased and is represented on the upper panel. Downregulation indicates that gene expression is deceased and is represented on the lower panel. The color scale shows enhanced color intensity as the fold change in expression as compared to the control samples gets larger either positively or negatively. The categories are <3, 3 - 5, 5 - 10 and > 10.*



Gene expression profiling associated with tobacco research is leading to the development of relevant models of smoking-related disease and large amounts of data to support the understanding of the mechanisms and pathogenic changes in smoke-exposed tissues. For example, Maunders *et al.* (2007) have also observed that primary human bronchial epithelial cells grown in a 3-D air-liquid interface exposed to air or whole mainstream cigarette smoke for 1 hour exhibited differential gene expression for several cellular processes including oxidant/antioxidant balance, xenobiotic metabolism and DNA damage and repair.

Beane *et al.* (2007) have also used a combination of microarray and qRT-PCR to identify reversible, slowly reversible and non-reversible gene expression changes in lungs of former smokers as compared to non-smokers. To support collective understanding of the vast amount of gene changes associated with smoking or lung disease, the smoking induced epithelial gene expression database (SIEGE) has been compiled as a signature profile for smoke-induced changes in vivo and compares lung responses in disease-free current, former and non-smokers (Shah *et al.*, 2005).

Spira (2010) examined gene expression profiles of buccal cells and compared the gene expression profile to that found in bronchial epithelial cells from smokers. He identified correlations between the two sources of cells from smokers indicating that the buccal cell gene expression response may serve as a surrogate for responses observed in respiratory tract (*i.e.* lung) of smokers. Additional research may lead to buccal cell collection and samplings as a measure for a minimally invasive source of tissue that can provide clinical diagnostics. However, a confounding issue for some of the current non-invasive techniques for buccal sampling is RNA integrity and quantity. Recent studies however are supporting the use of buccal cells for the assessment of epigenetic changes (*i.e.* DNA methylation).

DNA Methylation

Recently, Liu *et al.* (2010) reported the effect of CSC on DNA methylation in normal small airway epithelial cells and immortalized bronchial epithelial cells following chronic exposure (9 months). Changes in expression profiles of genes associated with DNA methylation, altered hypo- and hyper-DNA methylation patterns along with growth of cells in soft agar (preneoplastic characteristic) were observed. The authors conclude that CSC induced cancer-related epigenetic changes along with the development of preneoplastic morphological transitions in the cells. The resulting model may support *in vitro* studies of lung cancer pathogenesis.

Proteomics

Steiling *et al.* (2009) evaluated proteomic and transcriptomic profiles in bronchial brushings from current and never smokers. Twenty-three proteins were observed

to be differentially regulated in current as compared to never smokers. Many of the corresponding mRNAs (86%) for these proteins also exhibited differential regulation.

Individual and multiplex cytokine analysis has supported the evaluation of inflammatory response in smokers as well as *in vitro* cultures exposed to smoke and tobacco products. Differential regulation of IL-8 has been observed in lung cells exposed to smoke from cigarettes that burn as compared to primarily heat tobacco (Fields *et al.*, 2005) and in bronchial lavage samples from smokers that switched to cigarettes that primarily heat tobacco. In both cases, the subjects and cells exposed to smoke from cigarettes that primarily heat tobacco presented with reduced inflammatory effects (Rennard *et al.*, 2002).

Bioinformatics

Bioinformatics was developed out of a need to maintain, correlate and evaluate large volumes of biological data generated from molecular and toxicogenomic studies. Bioinformatics encompasses biological sciences, computer science and statistics. Integration of data from toxicogenomic studies into databases supports investigations for risk assessment, mechanistic and predictive toxicology and pattern recognitions in disease progression. Such efforts have been initiated by various agencies and research teams (Table III), and provide valuable means for understanding toxicological responses at the molecular level. However, there are some challenges with database design and maintenance involving data integration (standardized data storage and exchange), uniform nomenclature and standardized assay design (Mattes *et al.*, 2004).

Table III – Toxicogenomic Databases

Agency	**Database Link**
NIH Center for Bioinformatics	www.discover.nci.nih.gov/tools.jsp
NIEHS: National Center for Toxicogenomics	www.niehs.nih.gov/nct/home.htm
FDA's National Center for Toxicological Research (NCTR)	www.fda.gov/ScienceResearch/ BioinformaticsTools/Arraytrack/
EMBL-EBI The European Bioinformatics Institute	www.ebi.ac.uk/microarray-as/ae/
Enviromental Genome Project	www.genome.utah.edu/genesnps/
Biocarta	www.biocarta.com

Additional web-based databases and software platforms are also useful resources. The Biocarta system provides information on gene function and specific gene

pathways as well as information on reagent/assay resources and links to pertinent scientific citations (PubMed). Software platforms such as Ingenuity, SpotFire, GeneSpring, GeneSifter and Pathway Studio are also valuable tools for analyzing and supporting the interpretation of genomic data.

Transgenics

In vitro gene reporter assays serve as measures to detect toxicological affects of chemicals on promoters of genes involved in cellular regulation (*i.e.* transcription factors). Gene reporters consist of a gene of interest promoter (controlling elements) linked to DNA for luciferase or green fluorescence protein (proteins that liberate fluorescent light when induced or turned-on). Gene reporter assays are facilitated by transfection (insertion) of gene reporter constructs into a variety of cell types, and subsequent quantitation of signal or protein generation in response to chemical treatment (Figure 6). Our laboratory is investigating the use of luciferase models to evaluate the nucleus-related factor 2 (Nrf2) pathway. Nrf2 is a transcription factor (protein) that plays a key role in regulating the responses of genes involved in oxidative stress via control of DNA transcription. CSC has been observed to affect Nrf2 responsive genes and dose-dependently activates the luciferase gene through the Nrf2 promoter (Fields *et al.*, 2010; Figure 7).

Figure 6. Gene Reporter Assay. Chemical exposures, processes that generate reactive oxygen species, DNA damage or inflammatory mediators can direct transcription factors (such as Nrf2) to complimentary DNA binding elements (bases) of a gene. After binding, the transcription factor can activate the expression of multiple target genes.

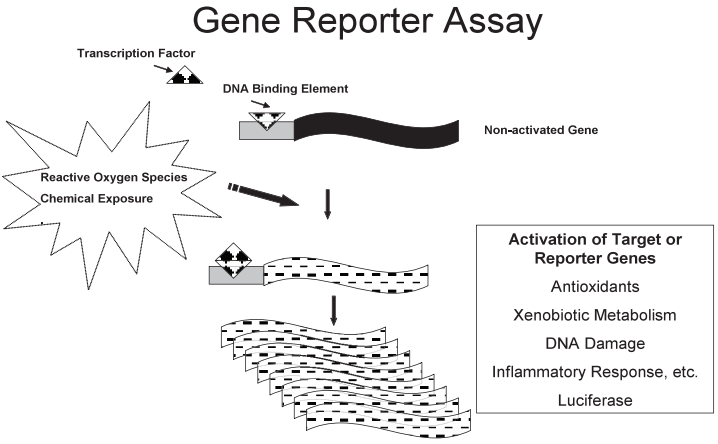


Figure 7. BEAS-2B cells were transiently transfected with a Nrf2 luciferase reporter plasmid. Time- and dose-dependent inductions in the Nrf2 promoter activity were assessed following CSC exposures. Specificity of the Nrf2 activation was evaluated by co-treatment with all-trans retinoic acid (ALTRA), a Nrf2 inhibitor.

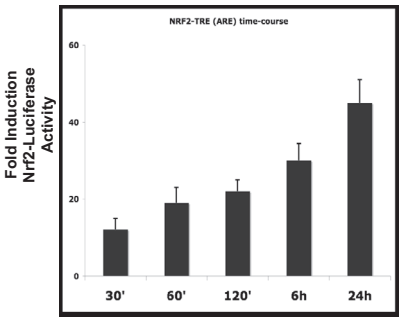


Figure 7a: Time-dependent activation of Nrf2 by CSC (50µg/ml); ARE: Antioxidant Response Element

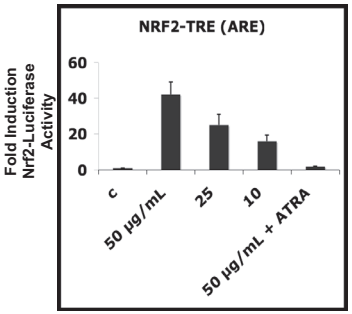


Figure 7b: Dose-dependent activation of Nrf2 by CSC; ARE: Antioxidant Response

Knockout (gene capability removed) and transgenically-enhanced (gene capability increased) models are also being investigated to advance knowledge on the response of specific gene modifications on toxic responses and disease pathogenesis. Mouse models have been developed with enhanced tumorigenic potential due to genetic alterations. The Tg.AC mouse harbors increased ras gene activity and enhanced growth via transgene expression of oncogenic viral Harvey ras (v-Ha-ras) within the skin. UL53-3 X A/J mice harbor a dominant-negative p53 mutation which inhibits cell cycle control. Such genetic alterations may support studies designed to evaluate tumorigenic potentials of chemical exposures in previously unresponsive models or models requiring long-term exposure regimens. These models have been used to assess the multistage aspects of tumorigenesis via exposures to 7,12-dimethylbenz(a)anthracene (DMBA), 12-O-tetradecanoylphorbol-13-acetate (TPA) or cigarette smoke (Owens *et al.*, 1995, 1999; De Flora *et al.*, 2003). Lung tumorigenicity has also been assessed to compare the A/J and rasH2 transgenic mouse models in response to mainstream tobacco-smoke whole-body and nose-only exposure regimens (Curtin *et al.*, 2004).

Further efforts for assessing genotype variations in relation to toxicology, susceptibility to environmental exposures and disease models have been undertaken by the Environmental Genome Project and commercial laboratories. Commercially available animal models include transgenic models for several research interests including apoptosis, cancer, diabetes and obesity, immunology and inflammation, cardiovascular diseases and metabolism. The ApoE-/- mouse is under investigation by researchers at RJRT as a model of cardiovascular disease. ApoE is a gene involved in lipoprotein clearance. Inactivation of the

apoE gene results in poor LDL-cholesterol clearance culminating in cholesterol accumulation and the development of atherosclerosis. Current studies are showing dose-dependent effects of cigarette smoke exposures on atherosclerotic lesions and gene expression profiles in mice fed a high-fat diet but little responses in mice consuming a normal chow diet (Curtin *et al.*, 2008; Nordskog *et al.*, 2010).

Genomics: Drug Discovery, Therapeutics and Perspectives of Governmental and Regulatory Agencies

The pharmaceutical industry and clinical research teams have expanded the use of genomics to drug development and clinical investigations to deliver new therapeutics for advanced treatment options, for the enhancement of diagnosis of disease and for monitoring response to therapies (Ito and Demers, 2004). Two terms, pharmacogenomics and pharmacogenetics, have emerged. These terms have been used interchangeably in recent years, although they are different by definition. "Pharmacogenomics" is the study of genes that determine drug behavior, while "pharmacogenetics" is the study of inherited differences that define drug metabolism and response.

The US Food and Drug Administration (FDA) recommends that pharmaceutical companies evaluate the pharmacogenomics of their respective drugs. The recommendation has been followed with guidelines detailed in a document prepared by the FDA (www.fda.gov/RegulatoryInformation/Guidances). The guidelines further define genomic biomarkers as "a measurable DNA and/or RNA characteristic that is an indicator of normal biological processes, pathogenic processes, and/or response to therapeutic or other interventions", since the FDA believes that such is key to understanding pharmacogenomics and pharmacogenetics. Pharmaceutical companies have initiated studies to identify biomarkers and additional information that may be used to predict patient drug responses (responsive, non-responsive, adverse effects). Subsequent goals included the development of clinical assays that may differentiate patients.

Gene signatures are being employed in current cancer research investigations to predict outcome of chemotherapeutics and to develop custom therapies. Gene profiles and the current knowledge of molecular changes in lung tumors may identify patients as low-risk or high-risk for certain disease. Current theories and trials suggest that those with low disease risk have a lower percentage of positive response to a specific therapy, while patients with a high risk may have a higher percentage of positive response. This theory must be taken with caution as the complexity of the molecular events associated with the high risk patient can complicate the therapy and potential outcome.

The Committee on Emerging Issues and Data on Environment Contaminants was established in the mid 2000's to allow for a network of individuals representing academia, governmental agencies, industry, environmental as well as public interest groups to discuss new technologies such as toxicogenomics. Governmental research agencies such as the National Institute of Environmental Health Sciences (NIEHS) are active in characterizing toxicogenomic applications in toxicological experiments. The Toxicogenomics Research Consortium (TRC) of the National Center for Toxicogenomics (NCT), a division of the NIEHS, has goals to evaluate toxicant-specific patterns of gene expression as related to dose-response, molecular mechanisms and biomarkers of human exposure. The consortium also plans to integrate gene expression profiling with proteomics, metabonomics and phenotypic anchoring, study toxicological effects of chemical mixtures and contribute gene expression and proteomics data to the Chemical Effects in Biological Systems Database (CEBS; www.niehs.nih.gov). Given the efforts of the TRC as well as individual laboratories in toxicogenomic research, the application of genomics-based techniques has become common place in recent years and is expected to expand. Key learnings from these consortiums and agency directives are anticipated to extend to toxicogenomics of tobacco research.

Emerging Technologies

In the post-genomic era, new techniques are still emerging for characterizing gene expression and mechanisms of disease. These include micro RNA (miRNA) expression analysis, applications with activity-based probes (ABPs) to support functional proteomics, enhanced cell-based assays for high-throughput screening, in vivo surrogates (*i.e.* organotypic, 3D cell cultures), micro-/nano-biotechnologies and laser capture microdissection (LCM; Sioud and Rosok, 2004; Spira *et al*., Viravaidya and Shuler, 2004; www.Asterand.com). These technologies along with micro total analysis systems (microTAS) or "lab-on-a-chip" can be used for toxicological assessment of cellular structures and functions, such as cellular adhesion, signal transduction, motility, membrane elasticity, chemotaxis, phagocytosis, protein expression and metabolism (Zieziulewicz, *et al*., 2003). As these techniques emerge, the promise for their application to toxicology are yet to be fully realized. Therefore, only laser capture microdissection (LCM) will be covered here as it relates to tissue-specific analysis as well as cellular and molecular interactions, since it is currently being employed in toxicogenomic applications.

LCM provides for a homogenous cell population and allows gene expression analysis of pooled single cells, cell subpopulations and cell populations, thereby eliminating the obscurity in gene expression data that may occur as whole-tissue analyses are performed. (Todd *et al*., 2002; Player *et al*., 2004; Elkahloun *et al*., 2002; Best *et al*., 2001). The LCM system consists of an inverted microscope

equipped with low-power near-infrared laser. Following preparation of tissue sections on glass slides, which includes placement of a thermal plastic (ethylene-vinyl acetate) film, specific cell populations or single cells are extractable from the tissue as energy liberated from the laser melts the thermoplastic film at designated locations. The ability to dissect single cells or clusters of cells is accomplished by adjusting of the laser diameter between 7.5 – 30 micron. The thermal plastic film provides a means for acquiring the cells with minimal or no detectable damage to the underlying macromolecules. The dissected cells can be transferred to a transparent cap that fits into a microcentrifuge tube and are subsequently subjected to designated extraction methods for molecular analyses (Bonner *et al*., 1997; Simone *et al*., 1998; Banks *et al*., 1999).

LCM has been applied to studies conducted to evaluate the effect of cigarette smoke in animals and humans. Such studies were performed to address the impact of aging on the susceptibility to smoke-induced inflammation and antioxidant capacity, differential role of cigarette smoke in expression of profibrotic mediators in small airways versus parenchyma in relation to the development of emphysema, and chemokine expression after single versus repeated smoke exposure (Moriyama *et al*., 2010; Churg *et al*., 2009; Suzuki *et al*., 2008). Suzuki *et al*., (2008) examined the levels of vascular endothelial growth factor (VEGF) expression in bronchial epithelium harvested from mice and humans via LCM or bronchial brushing. VEGF is required for normal lung function and deficiencies of the protein are associated with the development of emphysema in mice. Down-regulation of VEGF was exhibited by LCM captured tissues from smokers versus non-smokers as well as mice exposed to cigarette smoke for short and extended periods. Specifically, mice exposed to repeated exposures of smoke expressed decreased levels of VEGF compared to control animals and prior to the onset of emphysema. In humans, smokers with COPD had decreased expression of VEGF as opposed to smokers without COPD and life-long non-smokers.

Systems Biology

Systems biology is the current philosophy that is being applied to aid in combining genomics, proteomics and metabolomic data. The development of systems biology has been supported by the enhanced understanding of genes, their respective proteins and biological functions via advances in genetic information from the human genome project and new bioinformatic and molecular technologies. The aim of systems biology is to analyze and compare the large data sets from -omics and as well as other molecular tests into information that can inform the researcher of the total impact to the biological system. As stated by the Institute for Systems Biology,

> Systems biology is the study of an organism, viewed as an *integrated* and *interacting network* of genes, proteins and biochemical reactions

> which give rise to life. Instead of analyzing individual components or aspects of the organism, such as sugar metabolism or a cell nucleus, systems biologists focus on all the components and the interactions among them, all as part of one system. These interactions are ultimately responsible for an organism's form and functions (www.systemsbiology.org; Accessed May 24, 2010).

Sullivan *et al.*, (2010) are currently working on methods to combine two multi-year studies that have incorporated the use of genomics and proteomics data into the study of infectious disease. Challenges such as data integration, quality control and synchronization of data are under review by this team. Resolution of these changes may support other multi-omic toxicological research efforts.

Conclusions

Chemical exposure modulated by inherited familial genetic polymorphisms may drive molecular genetic changes during the pathogenesis of disease. For example, lung cancers possess modifications of genes that regulate DNA damage, growth, metabolism, oxidative stress and inflammation. The IOM report "Clearing the Smoke" describes the importance of establishing gene expression fingerprints for tissue-specific chemical exposure. The report further defines the need to use various genomic tools including gene expression assays in the battery of genetic toxicological analyses for determining the effectiveness of potential reduced exposure products (PREPS). Since the issuance of the IOM report, increased application of genomic technologies has been observed in academia, clinical, industrial and governmental organizations as well as tobacco research.

In 2007, the National Research Council of the National Academy of Sciences issued a document, "Toxicity Testing in the 21st Century: A Vision and a Strategy". In this document, the authors suggest that the advances in fields such as molecular biology and biotechnology are offering enhanced methods for researchers to evaluate the health risks of potentially toxic chemicals in cellular systems of human origin. Johnson *et al.* (2009) recently provided an overview of *in vitro* assays used in tobacco testing. Traditional and emerging technologies were evaluated for their merit and potential to relate to cancer for assessing tobacco products. The team concluded that in vitro toxicology methods are useful for screening toxicity; however, additional methods are required in light of the current regulatory environment. The application of toxicogenomics to tobacco research was also assessed and identified as a research gap. However, it was suggested that –omic technologies may serve as potential tools for future tobacco testing once potential issues (*i.e.* consensus methods, experimental designs, validation, criteria for fold-change and statistical significance, bioinformatic analysis platforms) are resolved.

Collectively, application of toxicogenomics in exposure assessment, hazard screening, risk assessment, human susceptibility and mechanisms of action is showing promise and expected to become an integral part of tobacco research. Current advances in drug discovery, personalized medicine and by the National Center for Toxicogenomics support this assertion and should provide key information to support the integration of toxicogenomics into toxicological test batteries.

REFERENCES

1. Banks, R. E. *et al* (1999).The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis: Preliminary findings. Electrophoresis. 20:689–700.
2. Baulig, A., Garlatti, M., Bonvallot, V., Marchand, A, Barouki, R., Marano, F., and Baeza-Squiban, A. (2003). Involvement of reactive oxygen species in the metabolic pathways triggered by diesel exhaust particles in human airway epithelial cells. Am. J. Physiol Lung Cell. Mol. Physiol. 285:671-679.
3. Beane, J., Sebastiani, P., Liu, G., Brody, J.S., Lenburg, M.E. and Spira, A. (2007). Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol. 8:201.
4. Best, C.J. and Emmert-Buck, M.R. (2001). Molecular profiling of tissue samples using laser capture microdissection. Expert Rev Mol Diagn.1:53-60.
5. Bird, A. (1992).The essentials of DNA methylation. Cell, 70, 5–8.
6. Bichsel, V.E., Liotta, L.A., and Petricoin, E.F. 3rd. (2001). Cancer proteomics: from biomarker discovery to signal pathway profiling. Cancer J. 2001 Jan-Feb 7(1):69-78.
7. Bollard, M.E., Stanley, E.G., Lindon, J.C., Nicholson, J.K., and Holmes, E. (2005). NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed. 18:143-162.
8. Bonner, R. F. *et al.* (1997). Laser capture microdissection: Molecular analysis of tissue. Science. 278:1481–1483.
9. Bosio, A., Knorr, C., Janssen, U., Gebel, S., Haussmann H. and Muller, T. (2002). Kinetics of gene expression profiling in Swiss 3T3 cells exposed to aqueous extracts of cigarette smoke. Carcinogenesis. 23:741-748.
10. Bundy, J.G., Spurgeon, D.J., Svendsen ,C., Hankard, P.K., Weeks, J.M., Osborn, D., Lindon J.C. and Nicholson, J.K. (2004). Environmental metabonomics: applying combination biomarker analysis in earthworms at a metal contaminated site. Ecotoxicology. 13:797-806.
11. Burczynski, M.E., McMillian, M. Ciervo, J., Li, L, Parker, J.B., Dunn, R.T., Hicken, S., Farr, S. and Johnson, M.D. (2000). Toxicogenomic-based discrimination of toxic mechanism in HepG2 human hepatoma cells. Toxicol. Sci. 58.399-415.

12. Caldwell, R. L. and Caprioli, R. M. (2005). Tissue profiling by mass spectrometry: a review of methodology and applications. Mol Cell Proteomics. 4:394-401.
13. Churg, A., Zhou, S., Preobrazhenska, O., Tai, H., Wang, R.. and Wright, J.L. (2009). Expression of profibrotic mediators in small airways versus parenchyma after cigarette smoke exposure. Am J Respir Cell Mol Biol. 40(3):268-76.
14. Curtin, G.M., Higuchi, M.A., Ayres, P.H., Swauger, J.E. and Mosberg, A.T. (2004). Lung tumorigenicity in A/J and rasH2 transgenic mice following mainstream tobacco smoke inhalation. Toxicol Sci. 81:26-34.
15. Curtin, G.M., Boudyguina, E., Meckley, D.R., Ayres, P.H., Doolittle, D.J. and Parks, J.S. (2008). Mainstream tobacco smoke inhalation exposure induces time- and dose-dependent increases for aortic cholesterol content within ApoE-/- knockout mice maintained on an atherogenic diet. Arteriosclerosis, Thrombosis and Vascular Biology 2008 Scientific Sessions.
16. Darva, F., Dorman, G. Krajcsi, P., Puskas, L.G. Kovari, Z., Lorincz, Z. and Urge, L. (2004). Recent advances in chemical genomics. 11:3119-3145.
17. De Flora, S., Balansky, R.M., D'Agostini, F., Izzotti, A., Camoirano, A., Bennicelli, C., Zhang, Z., Wang, Y., Lubet, R.A., and You, M. (2003). Molecular alterations and lung tumors in p53 mutant mice exposed to cigarette smoke. Cancer Res. 63:793-800.
18. Elkahloun, A.G., Gaudet, J., Robinson, G.S. and Sgroi, D.C. (2002). In situ gene expression analysis of cancer using laser capture microdissection, microarrays and real time quantitative PCR. Cancer Biol Ther. 1:354-358.
19. Farr, S. and Dunn, R.T. (1999). Concise review: gene expression applied to toxicology. Toxicol. Sci. 50:1-9.
20. Fields, W.R., Desiderio, J.G., Leonard, R.M., Burger, E.E., Brown, B.G. and Doolittle, D. J. (2004). Differential *c-myc* expression profiles in normal human bronchial epithelial cells following treatment with benzo[a] pyrene, benzo[a]pyrene-4,5 epoxide and benzo[a]pyrene-7,8-9,10 diol epoxide. Mol Carcinog. 40:79-89.
21. Fields, W.R., Leonard, R.M., Odom, P.S., Nordskog, B.K., Ogden, M.W. and Doolittle, D.J. (2005a). Gene expression in normal human bronchial epithelial (NHBE) cells following *in vitro* exposure to cigarette smoke condensate. Toxicol. Sci. 86:84-91.
22. Fields, W.R., Bombick, B.R. and Doolittle, D.J Gene expression profiles in normal human bronchial epithelial (NHBE) cells following exposure to smoke condensate from cigarettes that burn or primarily heat tobacco. Proc. Am. Association Cancer Res. 46:5439 (2005b).

23. Fields, W.R., Liedtke, W., Li, J. and Bombick, B. R. Assessment of Nrf2 luciferase reporter assay and heme oxygenase expression as models of oxidative stress in *in vitro* cultures of lung cells exposed to cigarette smoked condensate. Proc. Am. Association Cancer Res. 51:1693 (April 2010).
24. Fukano, Y., Yoshimura, H. and Yoshida, T. (2006). Heme oxygenase-1 gene expression in human alveolar epithelial cells (A549) following exposure to whole cigarette smoke on a direct in vitro exposure system. Exp Toxicol Pathol. 57:411-418.
25. Ferre, F., Marchese, A., Pexxoli, P., Griffin, S., Buxton, E. and Boyer, V. Quantitative PCR: an overview. *In*: Mullis KB, Gibbs RA, eds. The Polymerase Chain Reaction. Birkhauser, Boston, 1994, pp. 67-88.
26. Gebel, S., Gerstmayer, B., Bosio, A., Haussmann H., Miert, E.V. and Muller, T. (2004). Gene expression profiling in respiratory tissue from rats exposed to mainstream cigarette smoke. Carcinogenesis. 25:169-178.
27. Gustafson, A.M., Soldi, R., Anderlind, C., Scholand, M.B., Qian, J., Zhan, X., Cooper, K., Walker, D., McWilliams, A., Liu, G., Szabo, E., Brody, J., Massion, P.P., Lenburg, M.E., Lam, S., Bild, A.H. and Spira, A. (2010). Airway P13K pathway activation is an early and reversible event in lung cancer development. Sci Transl Med. 2:26.
28. Hackett, N.R., Heguy, A., Harvey, B.G., O'Conner, T.P., Luettich, K., Flieder, D.B., Kaplan, R., and Crystal, R.G. (2003). Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am. J. Respir. Cell Mol. Biol. 29:331-343.
29. Heid, C.A., Stevens, J., Livak, K. J, Williams P.M. (1996). Real time quantitative PCR. Genome Meth. 6:986-994.
30. Heijne, W.H., Stierum, R.H., Slijper, M., van Bladeren, P.J. and van Ommen, B. (2003) Toxicogenomics of bromobenzene hepatotoxicity: a combined transcriptomics and proteomics approach. Biochem Pharmacol. 65:857-875.
31. Hellmann, G.M., Fields, W.R. and Doolittle, D.J. (2001) Gene Expression Profiling of Cultured Human Bronchial Epithelial and Lung Carcinoma Cells. Toxicol. Sci. 61: 154-163.
32. Holmes, E., Nicholson, J.K. and Tranter, G.. (2001). Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks. Chem. Res. Toxicol. 14:182-191.
33. Inoue, T. and Pennie W.D. (2003) Toxicogenomics. Publisher: Tokyo; New York: Springer.
34. Issa, J.P. (2000). CpG-island methylation in aging and cancer. Curr. Top. Microbiol. Immunol., 249, 101–118.

35. Ito, R.K. and L.M. Demers. (2004). Pharmacogenomics and pharmacogenetics: future role of molecular diagnostics in the clinical diagnostic laboratory. Clinical Chemistry 50: 1526-1527.
36. Johnson, M.D., Schilz, J., Djordjevic, M.V., Rice, J.R. and Shields, P.G. (2009). Evaluation of in vitro assays for assessing the toxicity of cigarette smoke and smokeless tobacco. Cancer Epidemiol Biomarkers Prev. 18(12):3263-304.
37. Jones, P. A. and Takai, D. (2001). The role of DNA methylation in mammalian epigenetics. Science, 293, 1068–1070.
38. Jones, P. A. and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. Nature Rev. Genet., 3, 415–428.
39. Keun, H. C., Ebbels, T. M., Bollard, M. E., Beckonert, O., Antti, H., Holmes, E., Lindon, J. C., and Nicholson, J. K. (2004). Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. Chem Res Toxicol. 17:579-587.
40. Koike, E., Hirano, S., Shimojo, N., and Kobayashi, T. (2002). cDNA microarray analysis of gene expression in rat alveolar macrophages in response to organic extract of diesel exhaust particles. Toxicol Sci. 67:241-246.
41. Kreiner, T. and Buck, K.T. (2005). Moving toward whole-genome analysis: a technology perspective. Am. J. Health Syst. Pharm. 62:296-305.
42. Lindon, J. C., Holmes, E., Nicholson, J. K. (2004a). Metabonomics: systems biology in pharmaceutical research and development. Curr Opin Mol Ther. 6:265-72.
43. Lindon, J.C., Holmes, E., Bollard, M. E., Stanley, E. G. and Nicholson, J. K. (2004b). Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. Biomarkers 9:1-31.
44. Liu, F., Killian, J.K., Yan, M., Walker, R.L., Hong, J.A., Zhang, M., Davis, S., Zhang, Y., Hussain, M. XI, S., Rao, M., Meltzer, P.A., Schrump, D.S. (2010). Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. Oncogene. (Epub ahead of pint)
45. Maunders, H., Patwardham, S., Phillips, J., Clack, A. and Ritcher, A. (2007). Human bronchial epithelial cells: gene expression changes following acute exposure to whole cigarette smoke in vitro. Am. J. Physiol. Lung Cell Mol Physiol. 292:1248-1256.
46. Mattes, W. B., Pettit, S.D., Sansone, S. A., Bushel, P. R., and Waters, M. D. (2004). Database development in toxicogenomics: issues and efforts. Environ Health Perspect. 112:495-505.
47. Medlin, J. (2002). Two committees tackle toxicogenomics. Environmen Health Perspect. 110):A746-747.

48. Moggs, J.G. (2004). Phenotypic anchoring of gene expression changes during estrogen-induced uterine growth. Environ Health Perspect. 112: 1589.
49. Morgan, K.T., Ni, H. Brown, H.R., Yoon, L., Qualls, C.W., Crosby, L.M., Reynolds, R., Gaskill, B., Anderson, S.P., Kepler, T.B., Brainard, T., Liv, N., Easton, M., Merrill, C., Creech, D., Sprenger, D., Conner, G., Johnson, P.R., Fox, T., Sartor, M., Richard, E., Kuruvilla, S., Casey, W. and Benavudes, G. (2002). Application of cDNA microarray technology to in vitro toxicology and the selection of genes for a real-time RT-PCR-based screen for oxidative stress in Hep-G2 cells. Toxicol. Pathol. 30:435-451.
50. Moriyama C, Betsuyaku T, Ito Y, Hamamura I, Hata J, Takahashi H, Nasuhara Y, Nishimura M. (2010). Aging enhances susceptibility to cigarette smoke-induced inflammation through bronchiolar chemokines. Am J Respir Cell Mol Biol. 42(3):304-11.
51. Nicholson, J. K, Lindon, J. C. and Holmes, E. (1999)."Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181-1189.
52. Nicholson, J.K., Connelly, J., Lindon, J.C. and Holmes, E. (2002). Metabonomics: A platform for studying drug toxicity and gene function. Nature Reviews Drug Discovery. 1, 153-161.
53. Nordskog , B.K., Curtin, G.M., Brown, J.E. and Bombick, B.R. (2010). Gene expression analyses of the thoracic aorta from ApoE-/- mice exposed to mainstream cigarette smoke. Arteriosclerosis, Thrombosis and Vascular Biology 2010 Scientific Sessions.
54. Owens, D. M., Spalding, J. W., Tennant, R. W., and Smart, R. C. (2005). Genetic alterations cooperate with v-Ha-ras to accelerate multistage carcinogenesis in Tg.AC transgenic mouse skin. Cancer Res. 55:3171-3178.
55. Owens, D. M. , Wei, S-JC, and Smart, R. C. (1999). A multihit, multistage model of chemical carcinogenesis. Carcinogenesis. 20:1837-1844.
56. Parsanejad R., Fields, W.R., Bombick, B.R., Morgan, W.P. and Doolittle, D.J. (2008a). The time course for expression of genes involved in specific pathways in human bronchial epithelial cells following exposure to cigarette smoke. Exp. Lung Res 34:513-530.
57. Parsanejad R., Fields, W.R., Steichen T.J., Bombick, B.R. and Doolittle, D.J. (2008b). Distinct regulatory profiles of interleukins and chemokines in response to cigarette smoke condensate normal human bronchial epithelial (NHBE) cells. J. Cytokine & Interferon Research. 28:703- 12.
58. Pennie, W., Pettit, S. D., and Lord, P.G. (2004). Toxicogenomics in risk assessment: an overview of an HESI collaborative research program. Environ Health Perspect. 2:417-419.

59. Player, A., Barrett, J. C., and Kawasaki, E. S. (2004). Laser capture microdissection, microarrays and the precise definition of a cancer cell. Expert Rev Mol Diagn. 4:831-840.
60. Ramos-Nino, M.E., Heintz, N. Scappoli, L., Martinelli, M., Land, S., Nowak, N., Haegens, A., Manning, B., Manning, N., MacPherson, M., Stern, M. and Mossman, B. (2003). Gene profiling and kinase screening in asbestos-exposed epithelial cells and lungs. Am. J. Respir. Cell Mol. Biol. 29:S51-58.
61. Richardson, B. (2003). Impact of aging on DNA methylation. Ageing Res. Rev., 2, 245–261.
62. Rockett, J.C. and Hellmann, G.M. (2004).Confirming microarray data—is it really necessary? Genomics. 83:541-549.
63. Shah, V., Sridhar, S.,Beane, J., Brody, J.S., and Spira, A. (2005). SIEGE: smoking induced epithelial gene expression database. Nucleic Acids Res. 33:573-579.
64. Silverman, F.K., Spira, A., and Pare', P.D. (2009) Genetics and genomics of chromic obstructive pulmonary disease. Proc. Am Thorac. Soc. 6:539-542.
65. Simone, N. L. *et al.* (1998). Laser-capture microdissection: Opening the microscopic frontier to molecular analysis. Trends In Genetics. 4:272–276.
66. Sioud, M. and Rosok, O. (2004). Profiling microRNA expression using sensitive cDNA probes and filter arrays. Biotechniques. 37:574-576.
67. Spira A. (2010). Upper airway gene expression in smokers: the mouth as a "window to the soul" of lung carcinogenesis? Cancer Prev Res. 3:255-258.
68. Steiling, K., Lenburg, M.E. and Spira A. (2009). Airway gene expression in chronic obstructive pulmonary disease. Proc. Am. Thorac. Soc. 6:697-700.
69. Steiling, K., Kadar, A.Y., Bergerat, A., Flanigon, J., Sridhar, S., Shah, V., Ahmad, Q.R., Brody, J.S., Lenburg, M.E., Steffen, M. and Spira A. (2009). Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. PLoS One. 4: (Epub ahead of print).
70. Sullivan, D.E, Gabbard, J.L Jr, Shukla, M. and Sobral. B. (2010). Data integration for dynamic and sustainable systems biology resources: challenges and lessons learned. Chem Biodivers. 7(5):1124-41.
71. Suzuki M, Betsuyaku T, Nagai K, Fuke S, Nasuhara Y, Kaga K, Kondo S, Hamamura I, Hata J, Takahashi H and Nishimura M. (2008). Decreased airway expression of vascular endothelial growth factor in cigarette smoke-induced emphysema in mice and COPD patients. Inhal Toxicol. 20(3):349-59.

72. Todd R, Lingen MW, Kuo WP. (2002). Gene expression profiling using laser capture microdissection. Expert Rev Mol Diagn. 2:497-507.
73. Vanden Heuvel, J.P. (1997) PCR Protocols in Molecular Toxicology. Publisher: Informa Healthcare.
74. Viravaidya, K. and Shuler, M.L. (2004). Incorporation of 3T3-L1 cells to mimic bioaccumulation in a microscale cell culture analog device for toxicity studies. Biotechnol. Prog. 20:590-597.
75. Vulmiri, S.V., Misra, M., Hamm, J.T., Mitchell, M. And Berger, A. (2009). Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. Chem. Res. Toxicolo. 22:492-503.
76. Waters, M.D., Olden, K., and Tennant, R.W. (2003). Toxicogenomic approach for assessing toxicant-related disease. Mutat Res. 544:415-424.
77. Watson, R.E., and Goodman, J.I. (2002). Epigenetics and DNA methylation come of age in toxicology. Toxicol Sci. 67:11-6.
78. Watson R. E., McKim, J.M., Cockerel.l G.L. and Goodman J. I. (2004). The value of DNA methylation analysis in basic, initial toxicity assessments. Toxicol Sci. 79:178-188.
79. Zieziulewicz ,T. J., Unfricht, D.W., Hadjout, N., Lynes, M. A., and Lawrence D. A. (2003). Shrinking the biologic world--nanobiotechnologies for toxicology. Toxicol Sci. 74:235-244.

INFORMATION FOR ORDERING PREVIOUS SYMPOSIA

FREQUENCY: Annual. One volume per year.

PUBLISHER: The Tobacco Science Research Conference (TSRC) [formerly Tobacco Chemists' Research Conference, TCRC]

ORDER FROM THE DISTRIBUTOR:

Jeanne Marie Wallace
Communication Services
Butler Communications Services Building
Campus Box 7603, North Carolina State University
Raleigh, NC 27695-7603
(919) 513-3152 / fax (919) 515-8602
jeannemarie_wallace@ncsu.edu
http://www.cals.ncsu.edu/agcomm/tobacco_literature.html

TERMS:

1. All orders MUST be prepaid
2. Invoice sent upon request.
3. Payment accepted ONLY in U.S. currency or with a check drawn on an American bank with the bank's nine-digit American Banking Association (ABA) routing transit number. Overseas purchasers: Please instruct the bank to include the following information with payment:
    a. Publication being ordered
    b. Number of copies ordered
    c. Correct shipping address.

Checks without this information will be returned to the bank for instructions, delaying the order. Overseas subscribers may find it helpful for currency exchange reasons to deal with a serials vendor. See our web site for a link to a vendor list.

SUBSCRIPTIONS: Subscriptions are not accepted, since the number of copies available varies by year. However, we will hold copies upon request, pending payment; and a purchase order may serve to reserve orders, which are otherwise filled on a first-come, first-served basis.

BACK NUMBERS: Inquire for availability or see our web site.

REFUNDS/RETURNS: Not allowed.

DISCOUNTS: No multiple copy discount to subscription agents or vendors.

PRICES: **2006 and Subsequent Meetings:** Shipments for these meetings include a CD containing Powerpoint slides for presentations from this meeting.
$100 per volume for U.S. mailing address
$140 per volume for overseas - surface
$160 per volume for airmail
Previous Meetings: See web site above for pricing

SHIPMENT: Surface, unless airmail or FedEx shipment is purchased; cost includes postage.

CONTENTS: Each volume concerns a particular topic in tobacco science. A contents list appears at: http://www.cals.ncsu.edu/agcomm/rats_2006_contents.html
All *Recent Advances* orders are shipped with the accompanying meeting program booklet, when available.