ORIGINAL ARTICLE

Metabolomics, metabolic diversity and genetic variation in crops

George G. Harrigan · Susan Martino-Catt ·
Kevin C. Glenn

Abstract The metabolome represents a critical aspect of a plant's physiology, growth characteristics and, ultimately, economic value. Metabolic changes underpin plant development and responses to applied stresses, whilst quality traits in many important crops and ornamental plants are dependent on metabolic composition. It is also frequently reasoned that metabolic information reflects biological endpoints more accurately than transcript or protein analysis. As such, the science of metabolomics has proven to be of increasing popularity in assessing genotypic and phenotypic diversity in plants, in defining biochemical changes associated with developmental changes during plant growth and, increasingly, in compositional comparisons. The postulated value of this metabolic information resides primarily in its potential to support breeding and selection of novel yield-enhanced and nutritionally improved crops. Plants display remarkable genetic plasticity which has served to facilitate the development of an extraordinary range of genetically distinct and metabolically diverse cultivars for any given plant species. Despite concerns regarding potential loss of genetic diversity through domestication, genetic resources still exist through wild and exotic germplasms. Additionally, emerging biotechnological approaches such as RNAi silencing and the transgenic modification of regulatory genes offer new and fascinating opportunities for enhancing diversity and eliciting trait improvements. This review is intended to promote the view that metabolomics, through comparative assessments of metabolic diversity in domesticated and non-domesticated plants, and through evaluations of the compositional impact of metabolic engineering efforts can support breeding programmes designed to elicit trait improvements.

Keywords Genetic diversity · Metabolic variation · Metabolomics

G. G. Harrigan (✉)
Product Safety Center, Monsanto Company, 800 North Lindbergh Blvd., St. Louis, MO 63167, USA
e-mail: george.g.harrigan@monsanto.com

S. Martino-Catt
Product Characterization Center, Monsanto Company, 800 North Lindbergh Blvd., St. Louis, MO 63167, USA

K. C. Glenn
Regulatory Affairs, Monsanto Company, 800 North Lindbergh Blvd., St. Louis, MO 63167, USA

1 Introduction

Most domesticated crop plants were developed within a much more abbreviated time-scale than is often perceived, many only in the past few hundred years (Gepts 2002). Yet the genetic changes induced by selective breeding are such that major domesticated crops are typically represented by hundreds, even thousands, of unique cultivars specialized for production in a wide variety of geographic and societal settings. Whilst there are concerns regarding potential loss of genetic diversity through domestication (Hoisington et al. 1999; Lee 1998; Reif et al. 2005) it is clear that the hyper-mutable and genetically fluid character of plant life has facilitated the development of an extraordinary range of genetically distinct and metabolically diverse populations for any given crop.

Extensive variation in DNA sequence is typical within plants species. Transpositions and rearrangements are also common. These natural phenomena typically arise by non-programmed sequence-independent insertions and deletions

and other modifications. Transposition activity can also be induced by stresses such as hydrostatic pressure (Long et al. 2006) or through conventional breeding (Shan et al. 2005). Individuals within a species can, through natural processes, contain different numbers of genes (Bradford et al. 2005a, b; Cellini et al. 2004). Contributing to further genetic plasticity, DNA methylation represents a layer of regulatory information beyond that encoded in the basic structure of the plant genome. This epigenetic process can greatly influence levels of gene expression. Recent studies in rice highlighted extensive changes in DNA methylation patterns (as well as gene copy number and sequence) following the introduction, through conventional breeding, of a chromosomal fragment from a related wild species (Liu et al. 2004; Long et al. 2006; Chen et al. 2006). These naturally occurring transformations and their induction through stresses and conventional breeding have impacted and will clearly continue to impact the genetic diversity of plants including domesticated crops and, potentially, the composition of foodstuffs derived from such plants.

Quality traits in important crops and ornamental plants are related to metabolic composition (Memelink 2004) whilst metabolic changes underpin plant development and responses to applied stresses. Furthermore, metabolic information is often viewed as reflecting biological endpoints more accurately than transcript or protein analysis. Indeed, the plant metabolome has been described as bridging the "genotype-phenotype gap" (Fiehn 2002; Hall 2006). As such, the science of metabolomics has proven to be of increased popularity and value in assessing genotypic and phenotypic diversity, in defining biochemical changes associated with developmental changes and organ differentiation during plant growth, and, increasingly, in comparative compositional studies. A brief illustration of such concepts is provided in the following examples prior to more extensive discussion.

Non-domesticated relatives of crop plants may retain stress and disease resistance traits that are compromised in modern cultivars developed through selective breeding for properties more directly related to yield performance and to consumer preferences. As such, Schauer et al. (2005, 2006) recently conducted a GC-MS-based profiling survey to investigate the relationship between metabolite content and stress tolerance in a number of wild species of tomato and in introgression lines. A high degree of variation in the levels of metabolites in selected wild tomato species and introgression lines was recorded, an observation that may imply a significant genetic resource for eliciting stress tolerance improvements in domesticated crops. This exemplifies the potential of metabolomics to associate metabolic diversity

with differences in desirable agronomic traits, as will be discussed in more detail later.

The metabolic profiling technologies that can be applied to investigating metabolic changes *between* different plant species and cultivars can also be applied to understanding changes *within* plants as they mature. Targeted analyses of carbohydrates (Obiadalla-Ali et al. 2004), amino acids (Boggio et al. 2000), carotenoids and isoprenoids (Fraser et al. 1994) or steryl lipids (Whitaker 1988) on ripening tomato have been conducted whilst non-targeted evaluation of developing rice (Tarpley et al. 2005) has also proven highly informative. Pursuing a parallel analysis on both flesh and seed tissues of the same tomato fruits, Mounet et al. (2007) recently highlighted compositional similarities between the different tissues at early developmental stages and a metabolic trajectory that served to increasingly differentiate the tissues during development and at maturity. Understanding patters of metabolic changes during development will clearly influence breeding strategies and a role for metabolic profiling is clearly warranted.

Current profiling technologies have been shown to be effective in identifying and describing differences and similarities in the compositions of different plant cultivars and species. Significantly, recent reports of omics studies on biotechnology-derived plants indicate that these are typically more closely related to the isogenic parental strain used in their development than the parental strain is to other cultivars of the same species. This effect has been observed at the metabolome level for wheat (Baker et al. 2006) and potato (Defernez et al. 2004; Catchpole et al. 2005). Analogous results have been observed at the transcriptome level for wheat (Baudo et al. 2006) and grape (Franks et al. 2006) and at the proteome level for *Arabidopsis thaliana* (Ruebelt et al. 2006a, b, c), potato, (Lehesranta et al. 2005), tomato (Corpillo et al. 2004) and soybean (Wei et al. 2006).

The omics sciences continue to be applied to diverse areas of plant research including, for example, transcriptomic evaluations of heterosis or hybrid vigor in maize (Auger et al. 2005; Guo et al. 2006; Swanson-Wagner et al. 2006), and proteomic evaluations of the impact of different agricultural systems on potato tubers (Lehesranta et al. 2007). These studies, as with those described above, are aimed primarily at identifying and selecting genetic and metabolic features that associate with sought-after agronomic improvements. We now describe in more detail the potential of one of the omics sciences, metabolomics, in evaluating metabolic diversity in plants, and in supporting breeding programmes based on conventional and emerging paradigms such as the use of RNAi silencing.

2 Characteristics of metabolomics

The diverse physicochemical nature of metabolites and differences in their abundances and stabilities requires a range of diverse technologies for metabolome data acquisitions. Options principally centre on different variants of nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) (see Dunn and Ellis 2005; Goodacre et al. 2004; Harrigan and Goodacre 2003; Vaidyanathan et al. 2005 for reviews). Metabolomics strategies exploiting such technologies are characterized by two different conceptual approaches (Goodacre et al. 2004) defined herein as "targeted" and "non-targeted". Non-targeted approaches aim to provide a hypothesis-free global overview of readily detectable metabolites. The range and biochemical identity of metabolites measured in a given non-targeted metabolic profiling experiment is typically dependent on the data acquisition technology adopted. Consequently, the hypothesis-generating value of non-targeted approaches can be somewhat compromised by bias and selective reporting.

Targeted approaches employ optimized measurements designed for specific classes of metabolites or pathways. As such, they offer advantages in terms of improved quantitation and interpretability of acquired data. Strategies can be developed for assessing discrete classes of metabolites including, but not restricted to, organic acids, sugars, free amino acids and lipids. Indeed, metabolomic "sub-disciplines" are now developing as technologies for assessing different metabolite classes improve. For example, lipidomics, profiling technologies focused on lipids, represents an increasingly well-developed and influential aspect of metabolomics (Mutch et al. 2006; http://www.lipidmaps.org). Analytical platforms relevant to free amino acid analysis can be designed to measure the immediate biosynthetic precursors of target amino acids as well as associated anabolites and catabolites. Thus, Dubouzet et al. (2007) recently reported methodology to assess levels of anthranilate, tryptamine, serotonin and gramine in rice producing high levels of tryptophan, that could be integrated with information on all the major free amino acids (Wakasa et al. 2006). Studies on changes induced in free amino acid metabolism in lysine-producing maize can include measurements of lysine catabolites such as saccharopine and(α-aminoadipic acid (Huang et al. 2005).

Targeted approaches can also be considered to include the use of flux-based methodology (Ratcliffe and Schachar-Hill 2006). Here, specifically designed tracer-labelled substrates can be administered to a biological test system and the fate of the substrate monitored as it is metabolized and distributed throughout the test system. At present, this approach still remains somewhat underrepresented in metabolomics but clearly has great promise in associating metabolic changes with phenotypic diversity (see Last et al. 2007 for an comprehensive review).

Many metabolic profiling studies adopt a two-tiered hierarchical approach that incorporates both non-targeted and targeted elements (Goodacre et al. 2004; Catchpole et al. 2005). Thus, for example, a rudimentary assessment of composition can be provided through "fingerprinting" analyses such as a rapid flow-injection MS (Vaidyanathan et al. 2002) as a pre-determinant to more targeted and quantitative measurements using chromatographic and higher resolution methods. Such hierarchical approaches have been applied to "differentiation" studies, for example in microbial classification (Vaidyanathan et al. 2002) or in compositional comparisons of closely related crops (Catchpole et al. 2005; Enot et al. 2007). One attractive feature of high-throughput "fingerprinting" measurements is that their short analysis time (typically < 1 min) can allow a greater number of technical and biological replicates to be incorporated into an experiment,. thus enhancing the statistical validity of generated results (Enot and Draper 2007).

Cellular and sub-cellular compartmentalization of different biochemical processes is an important aspect of metabolic regulation (Browsher and Tobin 2001; Lunn 2007). Thus, a deeper understanding of plant metabolism requires a qualitative and quantitative description of the metabolome within different compartments of the cell. One promising approach is the use of non-aqueous fractionation (NAQF) to separate, through ultra-centrifugation, sub-cellular compartments under biochemically and enzymatically inactive conditions (Benkeblia et al. 2007) followed by metabolic profiling of the compartmental fractions. This approach has not been extensively used, but aspects of spatial resolution in metabolomic studies could prove an important component in studies of developmental changes within a growing plant tissue.

Metabolomics, which shares the same etymological root as metabolism (metabol, GK. change) ultimately seeks to understand *changes* in biological systems in terms of *changes* in metabolism. To fully realize such a goal, an ideal metabolomics methodology would allow a quantitative global analysis of metabolite content and of metabolic flux, including information on spatial resolution. At present however, available metabolic profiling methods (Dunn and Ellis 2005; Goodacre et al. 2004; Harrigan and Goodacre 2003; Vaidyanathan et al. 2005) allow identification of only a subset of previously described metabolites, and due to technical limitations and the diverse physicochemical properties and relative abundances of metabolite analytes, accurate quantitation is not always achievable. However the principle that the metabolome more accurately reflects the phenotype (Fiehn 2002) [an altogether different assertion than saying that current metabolic profiling methods

actually capture the (metabolome) phenotype] continues to drive improvements in analytical technologies, informatics, standardized approaches to data reporting and dissemination (Fiehn et al. 2005; Jenkins et al. 2004).

Despite the ever-expanding range of technological options, analytical sophistication, and informatics capabilities, there remain fundamental concerns related to the acquisition, interpretation and statistical treatments of omics data, particularly non-targeted approaches (Broadhurst and Kell 2006; Lay et al. 2006). No discussion on the characteristics and applications of metabolomics would be complete without acknowledging this. Dangers include (but are not restricted to) (i) bias, unintentional or otherwise, (ii) inadequate sample size relative to the large number of measured variables in a typical non-targeted omics experiment [see Jorstad et al. (2007) for an informative discussion on sample requirements for plant transcriptomics experiments] (iii) over-fitting when utilizing supervised machine learning (iv) excessive false discovery rates due to multiple hypothesis testing, (v) artifacts and drift inherent in data acquisition technology, and (vi) artifacts introduced through improper or variable sample preparation. The reader is referred to Broadhurst and Kell (2006) and Lay et al. (2006) for highly illuminating discussions.

3 Metabolomics applications

The scope of potential applications of plant metabolomics is clearly extensive. The examples now discussed in more detail are not presented as a comprehensive survey but are intended to highlight the value of metabolomics in assessments of biological diversity, biochemical changes during development, metabolic engineering as a component of enhancing germplasm diversity, and in comparative compositional studies.

3.1 Assessment of biological diversity through population screening

As indicated earlier, the genetic and phenotypic diversity of wild relatives of important crops may prove to be an important resource in eliciting new traits of value to the consumer and environment alike. Using tomato (*Solanum lycopersicum*) as a model system, Schauer et al. (2005) sought to evaluate the contribution of metabolic diversity to differences in the trait attributes of non-domesticated species. The goal was essentially to identify biochemical markers that could be directly associated with desired traits and applied to direct progeny selection when crossed with domesticated counterparts. Using a GC-MS methodology that allowed quantitation of over 90 metabolites including organic acids, sugars, polyols, amino acids and a limited

number of secondary metabolites, profiles were generated that differentiated the fruit and leaves of tested wild species from that of a modern domesticated cultivar. A conclusion from this study was that "...the wide metabolic variance of primary metabolites in fruits of the wild species suggests that.....boosting the levels of nutritionally important metabolites such as lysine, methionine, ascorbate and tocopherol will stand a high chance of success". The higher levels of secondary metabolites in the wild species also suggested a valuable resource for flavor and color compounds such as volatiles and carotenoids.

In an analogous study designed to correlate metabolic traits with abiotic stress resistance, Semel et al. (2007) assessed the impact of water restriction on yield performance and metabolite content of the domesticated tomato variety, M82, and the F1 hybrid of its cross with the wild tomato, *S. pennellii*. The fruit yield and brix (soluble organic material) content of the F1 hybrid were less affected by water restriction than that of M82. This was also true of the generated metabolic profiles; whilst drought-induced changes in the F1 hybrid were modest, implying tight homeostatic control, water restriction induced profound metabolic changes in M82. These changes included dramatic elevations in the levels of amino acids, TCA cycle intermediates, sugars and polyols. Many of these changes were consistent with current knowledge about plant osmoprotectants and compatible solutes. Thus, increases in levels of proline, along with its biosynthetic precursor, glutamate, were consistent with its postulated stress-protective role whereas increase in levels of glycine most probably reflected accumulation of glycine betaine, a well-known compatible solute. Other less obvious findings, such as increases in the levels of branched amino acids, certain TCA metabolites, and gentobiose, may represent new opportunities for utilizing metabolite-based research to guide breeding improvements in abiotic stress tolerance. The relatively high levels of most major metabolites measured in the drought tolerant F1 hybrid fruit may imply some form of "metabolic priming" designed to withstand abiotic stress. This intriguing hypothesis is consistent with the observation that the metabolite composition of the irrigated F1 hybrid fruit is similar to that of the M82 variety grown under water restriction.

In summary, these results may identify opportunities to improve drought tolerance in a valuable crop such as tomato, and provide new metabolite biomarkers to support breeding efforts in this area. Whilst the biochemical basis of complex traits such as stress tolerance is clearly multifactorial, the importance of molecules such as proline and glycine betaine has been identified in other major crops. Studies associating stress tolerance with proline accumulation in maize seedlings (Raymond and Smirnoff 2002) and roots (Verslues and Sharp 1999) have been reported,

for example. Levels of glycine betaine in maize leaf are known to increase in growing seasons which experience drought (Saneoka et al. 1995, Yang et al. 1995).

Levels of metabolites can also be monitored in introgression lines carrying wild species alleles and metabolic differences mapped to the introgressed region and its associated phenotype. Overy et al. (2005) reported one of the first metabolite profiling studies on introgression lines. They selected six plants (based on fruit characteristics) from a library of 76 tomato introgression lines comprised of wild *S. pennellii* alleles incorporated into the domesticated *S. lycopersicum*. Analysis of polar extracts of ripe *S. lycopersicum* and *S. penellii* fruit through non-targeted direct-injection-MS indicated differences in the relative levels of numerous metabolites, primarily organic acids and sugars. Subtle differences in the metabolism of the introgression lines when compared to the parent lines were also observed but not further defined in this study.

The experiment by Overy et al. (2005) was subsequently followed by a more rigorous evaluation of the entire library of 76 introgression lines using a GC-MS-based profiling platform that allowed quantitation of 74 metabolites (Schauer et al. 2006). Since levels of the majority of these metabolites were known to be higher in *S. penellii* fruit relative to their levels in the domesticated tomato, the finding that most metabolites in the fruit of the introgression lines were also elevated was unsurprising. However, it was also shown that many metabolites, for example, certain TCA cycle intermediates, were uniquely elevated in the introgression lines with levels greater than that observed in the *S. pennelllii* and *S. lycopersicum* parents. Whilst of potential relevance to breeding for improved traits the mechanism underlying the events that induce these metabolic features is not immediately apparent.

Assessments of the relationship between the 74 quantified metabolites and yield-associated morphological traits were conducted. Correlation analysis revealed three major network modules, as well as a few minor "loosely connected" modules, characterized principally by their degree of interconnectivity with other module co-members. One of the major modules represented phosphorylated intermediates and the yield-associated morphological traits, the second included amino acids, and the third contained organic acids and sugars. In addition to the high interconnectivity *within* the three major modules, many metabolites also exhibited high interconnectivity *between* modules. Some of these observations might be anticipated for signalling molecules such as γ-aminobutyric acid and inositiol-1-phosphate; however reasons for the high interconnectivity of metabolites such as glycerol-1-phosphate, for example, appear less obvious. Further analysis demonstrated that free amino acids and sugar phosphates were "morphologically-associated", that is, changes in their levels were associated with changes in at least one yield-associated morphological trait, whereas other metabolites such as a number of fatty acids and minor sugars were "morphologically independent". A conclusion from the study was that "morphologically associated" metabolites generally comprised central pathways of intermediary metabolism, an observation that did not appear to apply to "morphologically independent" metabolites. With the caveat that the metabolite coverage in this experiment extended to only 74 analytes, Schauer et al. (2006) suggested that manipulation of less central metabolism may prove to have less impact on plant morphology than manipulation of more central pathways. The application of such an observation may depend primarily on how pathway modification is achieved, including factors such as tissue specificity. For example, increases in lycopene and β-carotene levels in fruits of tomato modified through incorporation of *Arabidopsis* DET1, a nuclear protein implicated in regulating light transduction, were accompanied by pronounced yield decreases (Davuluri et al. 2004, 2005). Association with a fruit-specific promoter, however, might be reasonably predicted to increase levels of lycopene and β-carotene levels with less impact on overall plant development.

Ultimately, as suggested by Giovannoni (2006) the metabolic genomics approach described above may "...lead to modified crop breeding strategies and to a clearer understanding of the basic plant biology underlying important fruit quality traits".

3.2 Metabolic changes during plant development

Developmental changes in the synthesis, transport, accumulation and breakdown of metabolites influence the compositional and nutritional qualities of foodstuffs. Understanding these metabolic transitions should lead to improved approaches designed to enhance the nutritional quality of major crops. Adopting a variant of hierarchical metabolic profiling in a study on early development in rice, Tarpley et al. (2005) utilized a non-targeted GC-MS based platform that allowed identification of 155 metabolites. Analysis of profiles generated on seedlings grown over 10 days allowed identification of 21 metabolites that accounted for 83% of metabolite variance associated with developmental changes. It was suggested by the authors that future developmental studies could reasonably be conducted based on assessments on only these 21 metabolites. Comprising primarily organic acids, several sugars and amino acids this "biomarker set" essentially provides the advantages of a targeted approach whilst still representing extensive metabolite coverage.

Targeted analyses have been conducted on many metabolites during fruit development, particularly in tomato. Analytes evaluated include carbohydrates (Obiada-

lla-Ali et al. 2004), amino acids (Boggio et al. 2000), carotenoids and isoprenoids (Fraser et al. 1994) and steryl lipids (Whitaker 1988). However, as explained by Mounet et al. (2007), corresponding analyses of seed and flesh tissues from the same biological material may provide novel insights into the relationships between metabolic changes and developmental patterns in seed and fruit. The use of a non-targeted approach utilizing nuclear magnetic resonance spectroscopy (NMR) in conjunction with a targeted approach utilizing HPLC-photodiode array spectroscopy (PDA) and GC-flame-ionization detection (FID) allowed quantitation of 50 metabolites, of which 44 were detected throughout a 45-day development period in both flesh and seed samples. Expectedly, there were clear differences between the metabolic composition of seeds and flesh at all developmental stages. Differences which became more pronounced during development from 8 to 45 days post-anthesis included changes in glucose and fructose, free amino acids, and isoprenoids, which greatly increased in flesh, paralleled by a concomitant decrease in seed. Most intriguingly however, a high degree of compositional similarity was found at early stages of development. Thus, at 8 days post-anthesis concentrations in mannose, choline, *N*-methylnicotinamide and the major fatty acids were similar in flesh and seeds. The high levels of choline, a precursor for phosphatidylycholine, an important membrane component and *N*-methlynicotinamide, a small molecule associated with cell cycle regulation, suggest that the observed compositional similarities is a function of their low differentiation and active cell proliferation at early development stages. Parallel trends in flesh and seeds were observed for several metabolites during development including mutual decreases in the levels of choline, chlorogenate, *N*-methlynicotinamide, and fatty acids. A parallel transitory increase in starch followed by a decline was also observed. The relevance of these to growth or differentiation in fleshy and seed tissues could provide further insights into how breeding approaches can be optimized. Indeed, in a general observation by Tarpely et al. (2005) progress in metabolic profiling in this area can address our "…lack of knowledge of the broad changes in metabolite patterns during development [that] limits our efficiency to manipulate the cellular or molecular aspects of plant development with intent to influence yield or sustainability of production".

3.3 Metabolic changes and modification of regulatory genes

Many conventionally bred improvements in desirable agricultural traits have been associated with modulations of transcription factors and other regulatory proteins. The impact of alterations in levels of regulatory proteins has been well illustrated by the classic example of the TCP transcription factors encoded by the *teosinte branched1* (*tb1*) locus that regulate the expression of morphological traits known to differentiate wild Mexican grass teosinte from its domesticated counterpart, maize (Cubas et al. 1999). Many mutations in transcription factors, as opposed to changes in their expressed levels, have also been associated with agriculturally and commercially significant trait enhancements. Many of the yield gains achieved in wheat during the "Green Revolution" are attributable to mutant alleles in the *reduced height-1* loci (Peng et al. 1999) whereas mutation of the *opaque 2* gene that encodes a bZIP transcription factor was pivotal to the development of Quality Protein Maize (QPM), the winner of the World Food Prize in 2000 (Prasanna et al. 2001; Vietmeyer 2000). A spontaneous mutation in the MADS-box transcription *ripening-inhibitor (rin)* has been widely exploited to yield "commercially ubiquitous" tomato plants that exhibit an extended shelf life (Vrebalov et al. 2002). Indeed, it could be argued that domestication, breeding, and diversification of modern crops have been driven, in large part, by positive selection in the regulatory machinery underpinning gene expression. As such, natural mutations of regulatory proteins can therefore be considered to have played a transforming role in plant biology and in agricultural and societal developments.

Eliciting trait improvement through modern biotechnological approaches to modify regulatory protein expression may prove to be an attractive and viable option, and is, in a sense, an approach that mimics natural selection. However, plants encode many nuclear protein families (Liu et al. 1999; Riechmann et al. 2000) and, to date, the function of only a small fraction of these has been determined. In attempting to provide a framework for functional classification of transcription factor families. Grotewold (2005) suggested that some families can be considered to be associated with "core" processes such as organ development, for example, whereas other families may regulate more "peripheral" processes such as plant form or secondary metabolism. Whether this binary classification of regulatory proteins will prove to be generally applicable (see earlier discussion of DET1 and its impact on both tomato plant development and secondary metabolism) remains to be determined, but clearly a fuller understanding of the structure, function, and distribution of such proteins is required to facilitate developments in the use of transcription factor biology as a tool in trait development.

Considerable attention has been paid to two related transcription factor families in maize, the MYB C1 family and the basic-helix-loop-helix (bHLH) R family. Both of these are associated with the "peripheral" process of secondary metabolism and more specifically the regulation of phenylpropanoid and flavonoid synthesis (Grotewold et al.

1998, 2000; Grotewold 2005). Expression of C1 and R appear to be required for complete anthocyanin synthesis. Another MYB protein, P1 is associated with the biosynthetically related phlobaphenes (Grotewold 2005).

Clearly, advantages in studying secondary metabolites centre on their potential agronomic and commercial significance as well as their amenability to investigation through targeted metabolic profiling. Maize transcription factor genes controlling flavonoid synthesis have now been expressed in various transgenic plant species and their activation generally monitored by resultant production of flavonoids and anthocyanin pigmentation. Heterologous expression of maize LC and C1, for example, can upregulate flavonol synthesis in tomato flesh, a tissue that does not normally accumulate these secondary metabolites (Bovy et al. 2002). At least one non-targeted metabolic profiling experiment on the transgenic expression of MYB transcription factors has been pursued. A Fourier-transform-MS analysis of *Arabidopsis thaliana* over-expressing the MYB transcription factor PAP1 Tohge et al. (2005) allowed measurement of 1,800 putative metabolites confirming accumulation of anthocyanins and quercetin-type flavonols derivatives including at least eight new metabolites. Little discussion on changes in metabolites not directly associated with flavonoid chemistry was provided though it appeared that levels of immediate intermediates of the flavonoid biosynthetic pathways were markedly decreased. Transcriptomic analyses identified a total of 38 genes whose expression was modified by MYB PAP1 expression. Most of these had putative functions, such as glycosyltransferase activity, for example, that correlated with the detected changes in the metabolome. These results suggest minimal effects on the metabolome and transcriptome through transcription factor modification, and that the influence on flavonoid chemistry was highly specific.

3.4 Metabolic engineering

Metabolic engineering seeks to modify the tissue amounts or chemical structures of specific metabolites through changes in the levels or activities of biosynthetic or catabolic enzymes, or the introduction of novel enzyme activities (see Facchini et al. 2000; Hughes and Shanks 2002; Kutchan 2005; Larkin and Harrigan 2007; Sato et al. 2001; Trethewey 2004; Verpoorte and Memelink 2002, for reviews). Predicting which enzymatic activities of a complex metabolic pathway must be modified to successfully introduce desired phenotypic changes remains a challenge despite developments in availability of modelling tools such as Metabolic Control Analysis (MCA) (Cornish-Bowden 1995; Kacser and Burns 1973; Rees and Hill 1994) and Biochemical Systems Theory (BST) (Voit and Radivoyevitch 2000). As pointed out by

Larkin and Harrigan (2007), however, transformation of most crops through transgenic technologies can usually be achieved, and current metabolic profiling technologies render it relatively straightforward to discover influential steps in a target pathway. Thus, for each enzymatic step where a corresponding gene is available, transgenic perturbations can be introduced, and metabolomics applied, to investigate control points of metabolite synthesis and accumulation. In other words, current applied transgenic and metabolomic technologies mitigate, at least partially, any current limitations in predictive analyses. This argument also implies that targeted metabolic profiling is an important tool in assessing the value and efficacy of transgenic approaches to broadening the genetic diversity of existing germplasm. Significantly, profiling analyses of metabolically engineered plants have highlighted that in some instances, pronounced secondary effects in the metabolome and transcriptome can occur in crops modified with even exquisite genetic specificity, whereas in others, the same degree of specificity effects negligible secondary changes. It has become increasingly apparent that the nature of any pathway or trait is more relevant to the potential for secondary effects than the means by which that trait is introduced. More specifically, the dynamic nature of a pathway, likelihood of metabolon channelling, localization of major intermediates and cofactors and associated transport phenomena contributes more to the potential for secondary effects than the issue of whether genetic modification is pursued through modern biotechnology, conventional breeding or other approaches such as irradiation or mutagenesis.

Recently, the entire biosynthetic pathway for dhurrin, a tyrosine-derived cyanogenic glucoside, was incorporated into *A. thaliana* (Kristensen et al. 2005; Jørgensen et al. 2005; Tattersall et al. 2001) (see Fig. 1). This required transgenic expression of two multifunctional cytochrome P450s and a glucosyl transferase from *Sorghum bicolour*. The observed dramatic increase in levels of the anti-insectant dhurrin was not accompanied by any impact on plant morphology. Focused and non-targeted profiling studies indicated only marginal effects on the metabolome and transcriptome relative to wild type. For example, the pool of free amino acids was unaffected by the channelling of tyrosine into dhurrin synthesis and changes in measured flavonoid glycosides and glucosylated benzoic acid metabolites as assessed by LC-MS were minimal. The transcriptome was also unaffected as assessed by microarray analyses. However, in fascinating contrast, *partial* expression of the dhurrin biosynthetic pathway through incorporation of only the two *S. bicolour* multifunctional cytochrome P450s with no concomitant glucosyltransferase incorporation led to alterations in the metabolome and transcriptome as well as stunted plant growth.

Fig. 1 Representation of the dhurrin biosynthetic pathway. Transgenic incorporation of the entire pathway or only the first catalytic step leads to minimal secondary effects. Incorporation of CYP79A1 and CYP71E1 but not UGT85B1 leads to dramatic changes in morphology and the metabolome as well as transcriptome alterations. These changes may be associated with increases in levels of *p*-hydroxymandelonitrile
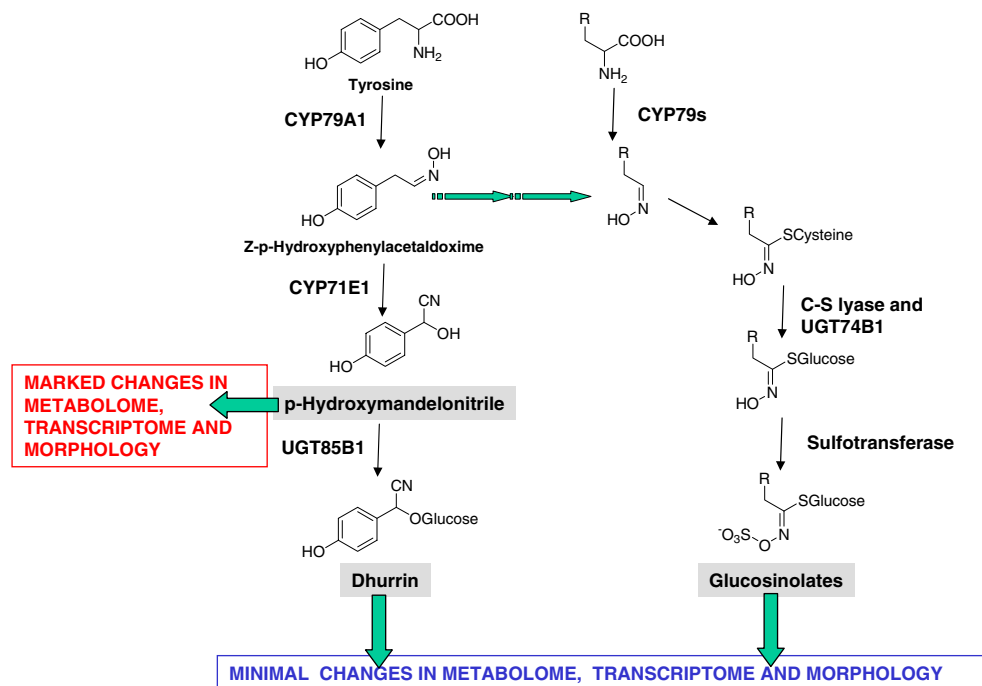
Metabolic profiling demonstrated that this observation coincided with accumulation of the dhurrin intermediate, *p*-hydroxymandelonitrile (Fig. 1). In *S. bicolour*, the two multifunctional cytochrome P450s and glucosyl transferase of the dhurrin biosynthetic pathway may form a metabolon allowing channelling of all intermediates. The complete incorporation of the entire metabolon appears to be required to minimize inadvertent secondary effects, whereas incomplete incorporation results in accumulation of an intermediate that induces profound phenotypic effects.

Significantly, Franks et al. (2006) introduced the *S. bicolour* derived dhurrin pathway into grapevine (*Vitis vinifera*) and noted no difference in patterns of gene expression between dhurrin-positive and dhurrin negative lines, consistent with the findings of Jørgensen et al. (2005), Kristensen et al. (2005) and Tattersall et al. (2001).

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Australia has performed some remarkable metabolic engineering in plants including the first example of opium-free poppies (Allen et al. 2004; Millgate et al. 2004; Larkin et al. 2007). This programme has utilized transgenic expression and RNAi silencing of enzymes involved in the benzylisoquinoline pathway to modulate expression of morphine and pharmaceutically important products such as thebaine which serves as a precursor for the chemical synthesis of analgesics such as buprenorphine and oxycodone

Metabolic profiling indicated that over-expression of codeinone reductase (COR) which catalyzes the penultimate step in morphine synthesis (see Fig. 2) resulted not only in the intended increase in morphine and codeine

levels, but was accompanied by an increase in levels of the upstream intermediate, thebaine. Details regarding regulation of the cellular and sub-cellular localization of the morphinan pathway still remain unclear, and no definite reasons for this secondary effect can be established at present. RNAi silencing of the same enzyme (Allen et al. 2004) was also characterized by unexpected changes in benzylisoquinoline metabolism. Whilst the intended decrease in morphine and codeine was observed, secondary accumulation of the biosynthetic precursor (*S*)-reticuline (as well as methylated derivatives), several enzymatic steps upstream of codeinone synthesis, was not predicted. This accumulation may indicate a feedback mechanism preventing intermediates from general benzylisoquinoline metabolism entering the morphine-specific branch. A testable hypothesis for the observed long-range effect is that COR comprises part of a metabolon that includes the enzyme that acts on (*S*)-reticuline, namely 1,2-dehydroreticulinium ion synthase; loss of COR would suppress the activity of this and other enzymes of the postulated metabolon.

Increasing levels of protein and free amino acid in plants represents a significant approach to enhancing the nutritional quality of foodstuffs. This area not only offers opportunities for emerging technical approaches such as RNAi silencing (e.g., Huang et al. 2005; Segal et al. 2003) but highlights the inappropriateness of selective associations of "unintended" effects with biotechnology-based approaches over genetic modification through conventional breeding. The opaque 2 mutation, discussed earlier, which was critical to the development of QPM is associated not
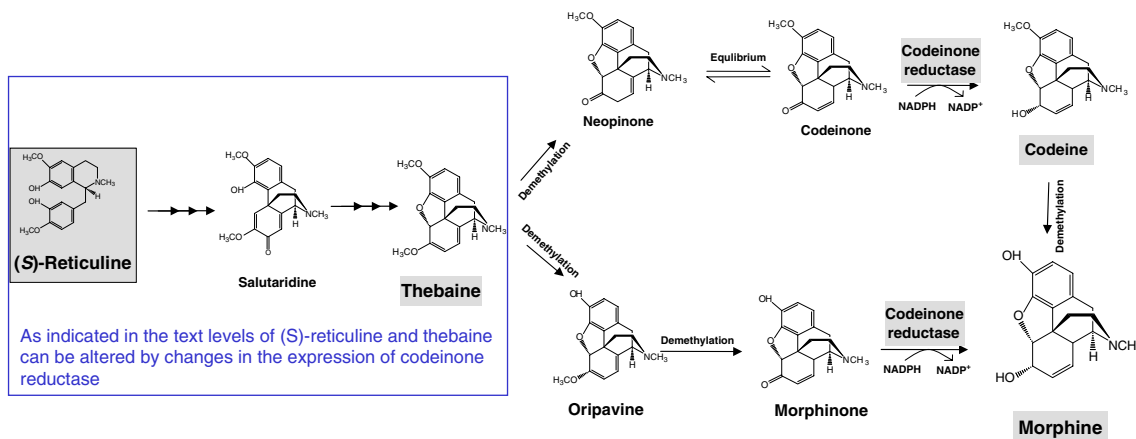
Fig. 2 A partial representation of the benzylisoquinoline/morphinan pathway. Over-expression of codeinone reductase leads to increases in the levels of codeine and morphine as expected but is also accompanied by an increase in thebaine. RNAi silencing of codeinone reductase leads to accumulation of (S) reticuline

only with the intended elevated levels of total lysine in grain due to down-regulation in the expression of lysine-poor zein storage proteins but also associated with changes in the levels of numerous free amino acids in the mature maize grain (Wang and Larkins 2001). The extent of this secondary effect appears to be germplasm dependent and several factors such as transport, increased amino acid synthesis, or reduced rates of amino acid incorporation into protein can, at least in principle, contribute (Wang and Larkins 2001). Regardless of mechanism, genetic modification of zein content through conventional breeding as a means to increase lysine levels is associated with pronounced secondary effects in free amino acid metabolism.

In contrast, transgenic incorporation of a mutant tryptophan-insensitive anthranilate synthase gene (*OSAID*) into a common rice cultivar is associated with markedly elevated levels of free tryptophan and only relatively minor changes in other free amino acids (Wakasa et al. 2006). Significantly, levels of phenylalanine and tyrosine, which as with tryptophan are derived from the shikimate pathway, were also relatively unaffected by *OSAID* incorporation. Analysis was conducted to determine whether changes in the levels of tryptophan impacted levels of precursors such as anthranilic acid and downstream intermediates including, tryptamine, serotonin, melanonin, indole and gramine, metabolites all known to be present in rice. Targeted analyses through LC-MS/MS showed no substantial differences (Dubouzet et al. 2007). Levels of tryptamine and serotonin were slightly increased, but fold changes were much less than that observed for tryptophan. The anthranilate levels were increased but remained at less than 0.1% of the levels of tryptophan indicating that it is effectively processed. Dubouzet et al. (2007) also pursued microarray analysis to investigate why so little tryptophan is chan-

nelled into downstream products. A monofactorial analysis of variance revealed only 70 genes that showed differential expression at $P < 0.01$, the statistical threshold used in this study. The analysis further revealed that there were no significant changes in the transcription of genes encoding enzymes of the phenylalanine and tyrosine metabolism or the shikimate pathway. A conclusion from this study was that the genes needed to convert tryptophan into other metabolites did not respond proportionally to the increased levels of substrate. The results may also imply that the sub-cellular and cellular distribution of intermediates, and the apparent lack of a metabolon of the type associated with dhurrin synthesis contribute to the minimal effects on the metabolome and transcriptome in rice modified by transgenic incorporation of *OSAID*.

3.5 Compositional comparisons

Compositional studies spanning over a decade of research on new biotechnology-derived crops have highlighted consistently the similarities of these crops with their conventional counterparts. These studies include assessments of soy (Padgette et al. 1996; Taylor et al. 1999), corn (Ridley et al. 2002; Sidhu et al. 2000; Herman et al. 2004), cotton (Bertrand et al. 2005; Nida et al. 1996), rice (Oberdoerfer et al. 2005), wheat (Obert et al. 2004), and alfalfa (McCann et al. 2006). As mentioned earlier, reports of omics studies on biotechnology-derived plants indicate that these are typically more closely related to the isogenic parental strain used in their development than the parental strain is to other members of the same species (see earlier references).

As described earlier, profiling analyses of metabolically engineered plants have highlighted that in some instances,

pronounced secondary effects on the metabolome and transcriptome can occur in crops modified with even exquisite genetic specificity, whereas in others, the same degree of specificity effects negligible secondary effects. It is most probable that the nature of any introduced trait is more relevant to the possibility of secondary effects than the means by which that trait is introduced. More specifically, the dynamic nature of a pathway, prospects for metabolon channelling, localization of major intermediates and co-factors and associated transport phenomena contributes more to the potential for secondary effects than the issue of whether genetic modification is pursued through modern biotechnology, conventional breeding or other approaches such as irradiation and mutagenesis. It is clear that targeted profiling can contribute to discussions on mechanism and biochemical theory and prove extremely valuable in determining which traits and pathways are most resistant or labile to secondary effects. Thus, targeted metabolomics should prove of immense value in contributing to evaluation of new traits at the early stages of discovery and selection of prospective candidates prior to regulatory submission. It must be acknowledged however, that other non-targeted metabolic profiling approaches to comparative compositional studies have proven quite informative. As an illustrative example, the hierarchical approach, described earlier, was recently applied to field-grown tubers from conventional potatoes and two biotechnology-derived varieties developed to yield high levels of inulin-type fructans (Catchpole et al. 2005). The first such variety was derived through incorporation of a gene encoding sucrose:sucrose 1-fructosyltransferase (SST), an enzyme that catalyzes the production of the trisaccharide 1-kestose, and oligofructans (Hellwege et al. 2000). The second experimental variety was derived through incorporation of a gene encoding fructan:fructan 1-fructosyltransferase (FFT), the product of which utilizes 1-kestose (and other oligofructans) to build inulin polymers (Hellwege et al. 2000). Preliminary flow-injection MS analyses of extracts from these varieties and the conventional potato cultivars highlighted extensive metabolic variation in all plants tested. A subsequent higher resolution gas chromatography-time-of-flight-MS approach allowed automated measurement of 252 analytes (90 positively identified, 89 assigned to a specific metabolite class, and 73 classified as unknowns) on over 2200 tubers (~160 per cultivar). This higher resolution data confirmed the large variation in the metabolic profiles and composition of the conventional cultivars. In summary, the metabolite composition of field-grown biotechnology-derived potatoes was consistent with the natural metabolite range of conventional cultivars and these lines differed from the progenitor line only in the intended elevation of levels of fructans and their derivatives.

4 Concluding remarks

Genetic diversity is required to elicit trait improvements in domesticated crops, yet concerns remain that crop domestication serves to deplete such diversity. Although plants demonstrate remarkable genetic plasticity including random DNA arrangements and transpositions, systematic approaches to enhancing agriculturally relevant diversity are required. Two major strategies were discussed in this review; the use of (i) non-domesticated sexually compatible relatives of major crops and (ii) biotechnology-based approaches including such emerging technologies as RNAi silencing. It was argued that metabolomics can contribute to both strategies and also play a major role in issues related to comparative compositional strategies. Notably, a previous document (ILSI 2004) reasons that application of metabolic profiling technologies to the comparative safety assessment process will require completion of efforts to standardize the reporting structure of such data, establishment of public repositories on baseline metabolomes of major crops including data generated from samples from diverse genetic backgrounds and environments. Thus, metabolic profiling can be used to identify biochemical traits in wild relatives of major crops and in introgression lines and correlation analyses further pursued to associate metabolic networks with morphological and other desired traits. The efficacy of transgene incorporation and the possibility of secondary effects can also be pursued through metabolomics, particularly targeted approaches.

Comparisons of results from biotechnology-based metabolic engineering with that of conventional breeding have established that the nature of any introduced trait or metabolic pathway is more relevant to the possibility of secondary effects than the breeding alternatives by which that trait is introduced. Furthermore, understanding pathway dynamics will prove critical not only to comparative compositional assessments but in enhancing the efficacy of transgenic trait enhancement. This was highlighted by the fact that transgenic expression in *A. thaliana* of the entire *S. bicolour* dhurrin pathway induced, not only its intended anti-herbivory effect, markedly less secondary effects than incorporation of only a subset of pathway members. Thus, the greatest contribution of metabolomics may lie in supporting the development of biochemical theory allowing predictive analyses to become a reality. It is worthwhile recognizing that the interaction of free amino acid metabolic networks with protein composition, an area one would expect to be thoroughly well researched in the pursuit of nutritionally enhanced crops remains only partially understood and remains an active area of research (e.g., Galili et al. 2006). Hopefully, the integration of metabolomics with biotechnology-based metabolic engineering and with conventional breeding will bring us closer

to fulfilling the promise of enhanced diversity and its implications for sustaining agricultural practices that benefit growers, consumers and the environment alike.

References

Allen, R. S., Millgate, A. G., Chitty, J. A., Thisleton, J., Miller, J. A. C., Fist, A. J., Gerlach, W. L., & Larkin, P. J. (2004). RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. *Nature Biotechnology, 22*, 1559–1566.

Auger, D. L., Gray, A. D., Ream, T. S., Kato, A., Coe, E. H. Jr., & Birchler, J. A. (2005). Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics, 169*, 389–397.

Baker, J. M., Hawkins, N. D., Ward, J. L., Lovegrove, A., Napier, J. A., Shewry, P. R., & Beale, M. H. (2006). A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnology Journal, 4*, 381–392.

Baudo, M. M., Lyons, R., Powers, S., Pastori, G. M., Edwards, K. J., Holdsworth, M. J., & Shewry, P. R. (2006). Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnology Journal, 4*, 369–380.

Benkeblia, N., Shinano, T., & Osaki, M. (2007). Metabolite profiling and assessment of metabolome compartmention of soybean leaves using non-aqueous fractionation and GC-MS analysis. *Metabolomics, 3*, in press, doi:10.1007/s11306-007-0078-y.

Bertrand, J. A., Sudduth, T. Q., Condon, A., Jenkins, T. C., & Calhoun, M. C. (2005). Nutrient content of whole cottonseed. *Journal of Dairy Science, 88*, 1470–1477.

Boggio, S. B., Palatnik, J. F., Heldt, H. W., & Valle, E. M. (2000). Changes in amino acid composition and nitrogen metabolizing enzymes in ripening fruits of *Lycopersicon esculentum* Mill. *Plant Science, 159*, 125–133.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Pertejo, M. A., Muir, S., Collins, G., Robinson, S., Verhoeyen, M., Hughes, S., Santos-Buelga, C., & van Tunen, A. (2002). High-flavonol tomatoes resulting from the heterolgous expression of the maize transcription factor genes LC and C1. *The Pllant Cell, 14*, 2509–2526.

Bradford, K. J., Gutterson, N., Parrott, W., Van Deynze, A., & Strauss, S. H. (2005a). Reply to "Regulatory regimes for transgenic crops". *Nature Biotechnology, 23*, 787–789.

Bradford, K. J., Van Deynze, A., Gutterson, N., Parrott, W., & Strauss, S. H. (2005b). Regulating transgenic crops sensibly: Lessons from plant breeding, biotechnology and genomics. *Nature Biotechnology, 23*, 439–444.

Broadhurst, D. I., & Kell, D. B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics, 2*, 171–196.

Browsher, C. G., & Tobin, A. K. (2001). Compartment of metabolism within mitochondria and plastids. *Journal of Experimental Botany, 52*, 513–527.

Catchpole, G. S., Beckmann, M., Enot, D. P., Mondhe, M., Zywicki, B., Taylor, J., Hardy, N., Smith, A., King, R. D., Kell, D. B., Fiehn, O., & Draper, J. (2005). Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 14458–14462.

Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H. V., Engel, K. H., Gatehouse, A. M., Karenlampi, S., Kok, E. J., Leguay, J. J., Lehesranta, S., Noteborn, H. P., Pedersen, J., & Smith, M. (2004). Unintended effects and their detection in genetically modified crops. *Food and Chemical Toxicology, 42*, 1089–1125.

Chen, Y., Long, L. K., Lin, X. Y., Guo, W. L., & Liu, B. (2006). Isolation and characterization of a set of disease resistance-rene analogs (RGAs) from wild rice, *Zizania latifolia* Griseb. I. introgression, copy number lability, sequence change, and DNA methylation alteration in several rice-*Zizania* introgression lines. *Genome, 49*, 150–158.

Cornish-Bowden, A. (1995). Metabolic control analysis in theory and practice. *Advances in Molecular Cell Biology, 11*, 21–64.

Corpillo, D., Gardini, G., Vaira, A. M., Basso, M., Aime, S., Accotto, G. P., & Fasano, M. (2004). Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: The case of a virus-resistant tomato. *Proteomics, 4*, 193–200.

Cubas, P., Vincent, C., & Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature, 401*, 157–161.

Davuluri, G. R., van Tuinen, A., Mustilli, A. C., Manfredonia, A., Newman, R., Burgess, D., Brummell, D. A., King, S. R., Palys, J., Uhlig, J., Bramley, P. M., Pennings, H. M. J., & Bowler C. (2004). Manipulation of DET1 expressin in tomato results in photomorphogenic phenotypes caused by post-transcriptional gene silencing. *The Plant Journal, 40*, 344–354.

Davuluri, G. R., van Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D. A., King, S. R., Palys, J., Uhlig, J., Bramley, P. M., Pennings, H. M. J., & Bowler C. (2005). Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnology, 23*, 890–895.

Defernez, M., Guning, Y. M., Parr, A. J., Shepher, L. V. T., Davies, H. V., & Coloquhoun, I. J. (2004). NMR and HPLC-UV profiling of potatoes with genetic modifications to metabolic pathways. *Journal of Agricultural and Food Chemistry, 52*, 6075–6085.

Dubouzet, J. G., Ishihara, A., Matsuda, F., Miyagawa, H., Iwata, H., & Wakasa, K. (2007). Integrated metabolomic and transcriptomic analyses of high-tryptophan rice expressing a mutant anthranilate synthase alpha subunit. *Journal of Experimental Botany* in press.

Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry, 24*, 285–294.

Enot, D., Beckmann, D., & Draper, J. (2007). Detecting a difference—Flow injection electrospray metabolome fingerprinting and machine learning to study genetically modified organisms. *Metabolomics, 3*, in press.

Enot, D., & Draper, J. (2007). Statistical strategies for validating plant genotype similarity assessments following multivariate analysis of metabolome fingerprint data. *Metabolomics, 3*, in press.

Facchini, P. J., Huber-Allanach, K. L., & Tari, L. W. (2000). Plant aromatic ʟ-amino acid decarboxylases: Evolution, biochemistry, regulation, and metabolic engineering applications. *Phytochemistry, 54*, 121–138.

Fiehn O (2002). Metabolomics—the link between genotype and phenotype. *Plant Molecular Biology, 48*, 155–171.

Fiehn, O., Wohlgemuth, G., & Scholz, M. (2005). Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. In Ludäscher, B. & Raschid, L. (Eds.), *Data integration in the life sciences: second interna-*

tional workshop (pp. 224–239, Vol. 3615). San Diego, CA, USA, July 20–22, 2005. Berlin, Heidelberg:. Springer-Verlag.

Franks, T. M., Powell, K. S., Choimes, S., Marsh, E., Iocco, P., Sinclair, B. J., Ford, C. M., & van Heeswijk, R. (2006). Consequences of transferring three sorghum genes for secondary metabolite (cyanogenic glucoside) biosynthesis to grapevine hairy roots. Transgenic Research, 15, 181–195.

Fraser, P. D., Truesdale, M., Bird, C. R., Schuch, W., & Bramley, P. M. (1994). Carotenoid biosynthesis during tomato fruit development. Plant Physiology, 105, 405–413.

Galili, G., Levanony, H., Slavikoava, S., Less, H., Angelovici, R., Ufaz, S., Tzin, V., & Koren, M. (2006). Elucidation of regulatory networks of plant metabolism and their exploitation to improve plant graoth and nutritional quality. http://www.weizmann.ac.il/Biology/open_day_2006/book/author/author_g.html

Gepts, P. (2002). A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Science, 42, 1780–1790.

Giovannoni, J. J. (2006). Breeding new life into plant metabolism. Nature Biotechnology, 26, 418–419.

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers - acquiring and understanding global metabolite data. Trends in Biotechnology, 22, 245–252.

Grotewold, E., Chamberlin, M., Snook, M., Siame, B., Butler, L., Swenson, J., Maddock, S., Clair, G., & Bowen, B. (1998). Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. The Plant Cell, 10, 721–740.

Grotewold, E., Sainz, M. B., Tagliani, L., Hernandez, J. M., Bowen, B., & Chandler, V. L. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proceedings of the National Academy of Sciences of the United States of America, 97, 13579–13584.

Grotewold, E. (2005). Plant metabolic diversity; a regulatory perspective. Trends in Plant Science, 10, 57–62.

Guo, M., Rupe, M. A., Yang, X., Crasta, O., Zinselmeier, C., Smith, O. S., & Bowen, B. (2006). Genome-wide transcript analysis of maize hybrids: Allelic additive gene expression and yield heterosis. Theoretical and Applied Genetics, 113, 831–845.

Hall, R. D. (2006). Plant metabolomics; from holistic hope, to hype, to hot topic. The New Phytologist, 169, 453–468.

Harrigan, G. G., & Goodacre, R. (2003). Metabolic profiling: Its role in biomarker discovery and gene function analysis. Boston: Kluwer Academic Publishers, p. 335.

Hellwege, E. M., Czapla, S., Jahnke, A., Willmitzer, L., & Heyer, A. G. (2000). Transgenic potato (Solanum tuberosum) tubers synthesize the full spectrum of inulin molecules naturally occurring in globe artichoke (Cynara scolymus) roots. Proceedings of the National Academy of Sciences of the United States of America, 97, 8699–8704.

Herman, R. A., Phillips, A. M., Collins, R. A., Tagliani, L. A., Claussen, F. A., Graham, C. D., Bickers, B. L., Harris, T. A., & Prochaska, L. M. (2004). Compositional equivalency of Cry1F corn event TC6275 and conventional corn (Zea mays L.). Journal of Agricultural and Food Chemistry, 52, 2726–2734.

Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J. M., Skovmand, B., Taba, S., & Warburton, M. (1999). Plant genetic resources: what can they contribute toward increased crop productivity? Proceedings of the National Academy of Sciences of the United States of America, 96, 5937–5943.

Huang, S., Kruger, D. E., Frizzi, A., D'ordine, R. L., Florida, C. A., Adams, W. R., Brown, W. E., & Luethy, M. H. (2005). High-lysine corn produced by the combination of enhanced lysine biosyntheis and reduced zein accumulation. Plant Biotechnology Journal, 3, 555–569.

Hughes, E. H., & Shanks, J. V. (2002). Metabolic engineering of plants for alkaloid production. Metabolic Engineering, 4, 414–418.

ILSI (International Life Sciences Institute). (2004). Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology. Comprehensive Reviews in Food Science and Food Safety 3, 35–104. Available from: http://www.members.ift.org/NR/rdonlyres/27BE106D-B616–4348-AE3A-091D0E536F40/0/crfsfsv3n2p00350104ms20040106.pdf

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R. J., Hall, R., Kopka, J., Lane, G. A., Lange, B. M., Liu, J. R., Mendes, P., Nikolau, B. J., Oliver, S. G., Paton, N. W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L. W., Wang, T., Walsh, S., Syrkin Wurtele, E., & Kell, D. B. (2004). A proposed framework for the description of plant metabolomics experiments and their results. Nature Biotechnology, 22, 1601–1606.

Jørgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, Bak, S., & Møller, B. L. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Current Opinion in Plant Biology, 8, 280–291.

Jorstad, T. S., Langaas, M., & Bones, A. M. (2007). Understanding sample size: what determines the required number of microarray for an experiment? Trends in Plant Science, 12, 46–50.

Kacser, H., & Burns, J. A. (1973). The control of flux. Symposia of the Society for Experimental Biology, 27, 65–104.

Kristensen, C., Morant, M., Olsen, C. E., Ekstrom, C. T., Galbraith, D. W., Moleer, B. L., & Bak, S. (2005). Metabolic engineering of dhurrin in transgenic plants with marginal inadvertent effects on the metabolome and transcriptome. Proceedings of the National Academy of Sciences of the United States of America, 102, 1779–1784.

Kutchan, T. M. (2005). Predictive metabolic engineering in plants: still full of surprises. Trends in Biotechnology, 23, 381–384.

Larkin P. J., Miller, J. A. C., Allen, R. S., Chitty, J. A., Gerlach, W. L., Frick, S., Kutchan, T. M., & Fist, A. J. (2007). Increasing morphinan alkaloid production by overexpressing codeinone reductase in transgenic Papaver somniferum. Plant Biotechnology Journal, 5, 26–37.

Larkin, P. J., & Harrigan, G. G. (2007). Opportunities and surprises of crops modified by transgenic technology—Metabolic engineering of benzylisoquinoline alkaloid, gossypol and lysine biosynthetic pathways. Metabolomics, 3, in press.

Last, R. L., Jones, A. D., & Yair Shachar-Hill, Y. (2007). Towards the plant metabolome and beyond. Nature Reviews. Molecular Cell Biology, 8, 167–174.

Lay, J. O. Jr., Borgmann, S., Liyanage, R., & Wilkins, C. (2006). The problem with the "omics". Trends in Analytical Chemistry, 25, 1046–1056.

Lee, M. (1998). Genome projects and gene pools: New germplasm for plant breeding? Proceedings of the National Academy of Sciences of the United States of America, 95, 2001–2004.

Lehesranta, S. J., Davies, H. V., Shepherd, L. V., Nunan, N., McNicol, J. W., Auriola, S., Koistinen, K. M., Suomalainen, S., Kokko, H. I., & Karenlampi, S. O. (2005). Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. Plant Physiology, 138, 1690–1699.

Lehesranta, S. J., Koistinen, K. M., Massat, N., Davies, H. V., Shepherd, L. V. T., McNicol, J. W., Cakmak, I., Cooper, J., Luck, L., Karenlampi, S. O., & Leifert, C. (2007). Effects of agricultural production systems and their components on protein profiles of potato tubers. Proteomics, 7, 597–604.

Liu, L., White, M. J., & MacRae, T. H. (1999). Transcription factors and their genes in higher plants. Functional domains, evolution

and regulation. *European Journal of Biochemistry, 262,* 247–257.

Liu, Z. L., Wang, Y. M., Shen, Y., Guo, W. L., Hao, S., & Liu, B. (2004). Extensive alterations in DNA methylation and transcription in rice caused by introgression from *Zizania latifolia*. *Plant Molecular Biology, 54,* 571–582.

Long, L., Lin, X., Zhai, J., Kou, H., Yang, W., & Liu, B. (2006). Heritable alteration in DNA methylation pattern occurred specifically at mobile elements in rice plants following hydrostatic pressurization. *Biochemical and Biophysical Research Communications, 340,* 369–376.

Lunn, J. E. (2007). Compartmentation in plant metabolism. *Journal of Experimental Botany, 58,* 3–547.

McCann, M. C., Rogan, G. J., Fitzpatrick, S., Trujillo, W. A., Sorbet, R., Hartnell, G. F., Riordan, S. G., & Nemeth, M. A. (2006). Glyphosate-tolerant alfalfa is compositionally equivalent to conventional alfalfa (Medicago sativa L.). *Journal of Agriculture and Food Chemistry, 54,* 7187–7192.

Memelink, J. (2004). Tailoring the plant metabolome without a loose stitch. *Trends in Plant Science, 7,* 305–307.

Millgate, A. G., Pogson, B. J., Wilson, I. W., Kutchan, T. M., Zenk, M. H., Gerlach, W. L., Fist, A. J., & Larkin, P. J. (2004). Analgesia: morphine-pathway block in top1 poppies. *Nature, 431,* 413–414.

Mounet, F., Lemaire-Charmley, M., Maucourt, M., Cabasson, C., Giraudel, J.-L. Deborde, C., Lessire, R., Gallusci, P., Bertrand, A., Gaudillere, M., Rothan, R., Roline, D., & Moing, A. (2007). Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics, 3,* in press, doi:10.1007/s11306-007-0059-1.

Mutch, D. M., Fauconnot, L., Grigorov, M., & Fay, L. B. (2006). Putting the 'Ome' in lipid metabolism. *Biotechnology Annual Review, 12,* 67–84.

Nida, D. L., Patzer, S., Harvey, P., Stipanovic, R., Wood, R., & Fuchs, R. L. (1996). Glyphosate-tolerant cotton: The composition of the cottonseed is equivalent to that of conventional cottonseed. *Journal of Agricultural and Food Chemistry, 44,* 1967–1974.

Oberdoerfer, R. B., Shillito, R. D., Beuckeleer, M., & Mitten, D. H. (2005). Rice (*Oryza sativa* L.) containing the bar gene is compositionally equivalent to the nontransgenic counterpart. *Journal of Agricultural and Food Chemistry, 53,* 1457–1465.

Obert, J. C., Ridley, W. P., Schneider, R. W., Riordan, S. G., Nemeth, M. A., Trujillo, W. A., Breeze, M. L., Sorbet, R., & Astwood, J. D. (2004). The composition of grain and forage from glyphosate tolerant wheat MON 71800 is equivalent to that of conventional wheat (*Triticum aestivum* L.). *Journal of Agricultural and Food Chemistry, 52,* 1375–1384.

Obiadalla-Ali, H., Fernie, A. R., Kossmann, J., & Lloyd, J. R. (2004). Developmental analysis of carbohydrate metabolism in tomato (*Lycopersicon esculentum* cv. Micro-Tom) fruits. *Physiologia Plantarum, 120,* 196–204.

Overy, S. A., Walker, H. J., Malone, S., Howard, T. P., Baxter, C. J., Sweetlove, L. J., Hill, S. A., & Quick, W. P. (2005). Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. *Journal of Experimental Botany, 56,* 286–297.

Padgette, S. R., Taylor, N. B., Nida, D. L., Bailey, M. R., MacDonald, J., Holden, L. R., & Fuchs, R. L. (1996). The composition of glyphosate-tolerant soybean seeds is equivalent to that of conventional soybeans. *Journal of Nutrition, 126,* 702–716.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature, 400,* 256–261.

Prasanna, B. M., Vasal, S. K., Kassahun, B., & Singh, N. N. (2001). Quality protein maize. *Current Science, 81,* 1308–1319.

Ratcliffe, R. G., & Schachar-Hill, Y. (2006). Measuring multiple fluxes through plant metabolic networks. *The Plant Journal, 45,* 490–511.

Raymond M. J., & Smirnoff, N. (2002). Proline metabolism and transport in maize seedlings at low water potential. *Annals of Botany (Lond), 89,* 813–823.

Rees, T. A., & Hill, S. A. (1994). Metabolic control analysis of plant-metabolism. *Plant, Cell & Environment, 17,* 587–599.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., Bohn, M., & Melchinger, A. E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics, 110,* 859–864.

Ridley, W. P., Sidhu, R. S., Pyla, P. D., Nemeth, M. A., Breeze, M. L., & Astwood, J. D. (2002). Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (*Zea mays* L.). *Journal of Agricultural and Food Chemistry, 50,* 7235–7243.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, B. K., Sherman, B. K., & Yu, G.-L. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science, 290,* 2105–2110.

Ruebelt, M. C., Leimgruber, N. K., Lipp, M., Reynolds, T. L., Nemeth, M. A., Astwood, J. D., Engel, K. H., & Jany, K. D. (2006a). Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 1. Assessing analytical validation. *Journal of Agricultural and Food Chemistry, 54,* 2154–2161.

Ruebelt, M. C., Lipp, M., Reynolds, T. L., Astwood, J. D., Engel, K. H., & Jany, K. D. (2006b). Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 2. Assessing natural variability. *Journal of Agricultural and Food Chemistry, 54,* 2162–2168.

Ruebelt, M. C., Lipp, M., Reynolds, T. L., Schmuke, J. J., Astwood, J. D., DellaPenna, D., Engel, K. H., & Jany, K. D. (2006c). Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 3. Assessing unintended effects. *Journal of Agricultural and Food Chemistry, 54,* 2169–2177.

Saneoka, H., Nagasaka, C., Hahn, D. T., Yang, W. J., Premachandra, G. S., Joly. R. J., & Rhodes, D. (1995). Salt tolerance of glycinebetaine-deficient and glycinebetaine-containing maize lines. *Plant Physiology, 107,* 631–638.

Sato F., Hashimoto, T., Hachiya, A., Tamura, K., Choi, K. B., Morishige, T., Fujimoto, H., & Yamada, Y. (2001). Metabolic engineering of plant alkaloid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 367–372.

Schauer, N., Zamier, D., & Fernie, A. R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: A survey of the *Solanum lycopersicum* complex. *Journal of Experimental Botany, 56,* 297–307.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Meleis, A., Bruedigam, C., Kopka, J., Willmitzer, Zamir, L., & Fernie, A. R. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology, 24,* 447–454.

Segal, G., Song, R., & Messing, J. (2003). A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics, 165,* 387–397.

Semel, Y., Schauer, N., Roessner, U., Zamier, D., & Fernie, A. R. (2007). Metabolite analysis for the compariosn of irrigated and

non-irrigated field grown tomato of varying genotype. *Metabolomics, 3,* in press, doi:10.1007/s11306-007-0055-5.

Shan, X. H., Liu, Z. L., Dong, Z. Y., Wang, Y. M., Chen, Y., Lin, X. Y., Long, L. K., Han, F. P., Dong, Y. S., & Liu, B. (2005). Mobilization of the active Mite transposons Mping and Pong in rice by introgression from wild rice (*Zizania latifolia* Griseb.). *Molecular Biology and Evolution, 22,* 976–990.

Sidhu, R. S., Hammond, B. G., Fuchs, R. L., Mutz, J.-N., Holden, L. R., George, B., & Olson, T. (2000). Glyphosate-tolerant corn: The composition and feeding value of grain from glyphosate-tolerant corn is equivalent to that of conventional corn (*Zea mays* L.). *Journal of Agricultural and Food Chemistry, 48,* 2305–2312.

Swanson-Wagner, R. A., Jia, Y., DeCook, R., Borsuk, L. A., Nettleton, D., & Schnable, P. S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 6805–6810.

Tarpley, L., Duran, A. L., Kebrom, T. H., & Sumner, L. W. (2005). Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. *BMC Plant Biology, 5,* in press, doi:10.1186/1471-2229-58.

Tattersall, D. B., Bak, S., Jones, P. R., Olsen, C. E., Nielsen, J. K., Hansen, M. L., Hoj, P. B., & Moller, B. L. (2001). Resistance to an hervivore through engineered cyanogenic glucoside synthesis. *Science, 293,* 1826–1828.

Taylor, N. B., Fuchs, R. L., MacDonald, J., Shariff, A. R., Padgette, S. R. (1999). Compositional analysis of glyphosate-tolerant soybeans treated with glyphosate. *Journal of Agricultural and Food Chemistry, 47,* 4469–4473.

Tohge, T., Nishiyama Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazahura, M., Inoue, E., Takahashi, H., Goodenow, D. B., Kitayama, M., Noji, M., Yamazaki, M., & Saitom K. (2005). A novel approach for non-targeted data anaysis for metabolomics: large-scale profiling of tomato fruit profiles. *The Plant Journal, 42,* 218–235.

Trethewey, R. N. (2004). Metabolite profiling as an aid to metabolic engineering in plants. *Current Opinion in Plant Biology, 7,* 196–201.

Vaidyanathan, S., Kell, D. B., & Goodacre, R. (2002). Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. *Journal of the American Society for Mass Spectrometry, 13,* 118–128.

Vaidyanathan, S., Harrigan, G. G., & Goodacre, R. (2005). *Metabolome analyses. strategies for systems biology.* New York: Springer, p. 381.

Verpoorte, R., & Memelink, J. (2002). Engineering secondary metabolite production in plants. *Current Opinion in Biotechnology, 13,* 181–187.

Voit, E. O., & Radivoyevitch, T. (2000). Biochemical systems analysis of genome-wide expression data. *Bioinformatics, 16,* 1023–1037.

Verslues P. E., & Sharp, R. E. (1999). Proline accumulation in maize (*Zea mays* L.) primary roots at low water potentials. II. Metabolic source of increased proline deposition in the elongation zone. *Plant Physiology, 119,* 1349–1360.

Vietmeyer, N. D. (2000). A drama in three long acts: the story behind the story of the development of quality-protein maize. *Diversity, 16,* 29–32.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medranao, D., Drake, R., Schuch, W., & Giovannoni, J. (2002). A MADS-Box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus. *Science, 296,* 343–346.

Wakasa, K., Hasegawa, H., Nemoto, H., Matsuda, F., Miyawaza, H., Tozawa, Y., Morino, K., Komatsu, A., Yamada, T., Terakawa, T., & Miyagawa, H. (2006). High-level tryptophan accumulation in seeds of transgenic rice and its limited effects on agronomic traits and seed metabolite profile. *Journal of Experimental Botany, 57,* 3069–3078.

Wang, X., & Larkins, B. A. (2001). Genetic analysis of amino acid accumulation in opaque-2 maize endosperm. *Plant Physiology, 125,* 1766–1777.

Wei, C. H., Sok, D.-E., Yang, Y.-H., Oh, S. H., Kim, H. C., Yoon, W. K., Kim, H. M., & Kim, M. R. (2006). Protein composition of domestic and glyphosate-tolerant soybean. *Han'guk Sikp'um Yongyang Kwahak Hoechi, 35,* 470–475.

Whitaker, B. D. (1988). Changes in the steryl lipid content and composition of tomato fruit during ripening. *Phytochemistry, 27,* 3411–3416.

Yang, W. J., Nadolska-Orczyk, A., Wood K. V., Hahn, D. T., Rich, P. J., Wood, A. J., Saneoka, H., Premachandra, G. S., Bonham, C. C., Rhodes, J. C., Joly, R. J., Samaras, Y., Goldsbrough, P. B., & Rhodes, D. (1995). Near-isogenic lines of maize differing for glycinebetaine. *Plant Physiology, 107,* 621–630.