

Gene Discovery and Product Development for Grain Quality Traits

Barbara Mazur, Enno Krebbers, Scott Tingey

The composition of oils, proteins, and carbohydrates in seeds of corn, soybean, and other crops has been modified to produce grains with enhanced value. Both plant breeding and molecular technologies have been used to produce plants carrying the desired traits. Genomics-based strategies for gene discovery, coupled with high-throughput transformation processes and miniaturized, automated analytical and functionality assays, have accelerated the identification of product candidates. Molecular marker-based breeding strategies have been used to accelerate the process of moving trait genes into high-yielding germplasm for commercialization. These products are being tested for applications in food, feed, and industrial markets.

Altered seed traits that result in grain with modified oil, protein, and carbohydrate content and composition can bring enhanced value to farmers, producers, and consumers. Such traits, referred to as seed quality or output traits, alter the nutritional or functional properties of the harvested plant for use in foods, animal feeds, or industrial products. Developing plants with improved grain quality traits involves overcoming a variety of technical challenges inherent in metabolic engineering programs. Continuing improvements in molecular and genomic technologies are contributing to the acceleration of product development.

Two complementary pathways have been used to identify and develop grain quality traits for corn and soybeans (1). The first is germplasm screening coupled with high-throughput analytical tests for trait identification, followed by the use of DNA-based molecular markers to accelerate the breeding process. An example of the use of these technologies comes from a program to identify soybeans with reduced concentrations of the antinutritional oligosaccharides, stachyose, raffinose, and galactose (2). Soybean lines were identified through screens of germplasm collections and mutagenized seed populations. These lines were combined to yield varieties that had still lower concentrations of these oligosaccharides; recombinant lines could be selected that also had higher concentrations of sucrose, a benefit in food applications. DNA markers were then used to direct a breeding process that moved the desired traits into commercial high-yielding germplasm. The use of the DNA markers accelerated the breeding program by several generations. The resultant soybeans are now being tested in feed, food, and beverage applications.

Similarly, high-oil varieties of corn were

developed at the University of Illinois through successive cycles of recurrent selection (3). These lines have improved energy content for animal feeding applications, but poor agronomic characteristics, including disease susceptibility and poor standability, precluded their commercial introduction on broad acreage. Quantitative trait analysis with molecular markers indicated that high oil content in the seed was controlled by more than 12 genes, making yield improvements through conventional breeding difficult (4). To circumvent this problem, the TopCross grain production system was developed and introduced (5). TopCross provides a means of capturing the genetic contribution from an inferior genotype through an elite hybrid corn variety. In this process, a male fertile high-oil corn variety is interplanted at low density with a male sterile elite hybrid variety of corn. The low planting density of the pollinator means that its contribution to grain yields is minimized. Pollen shed by the agronomically challenged male fertile line will fertilize the field of elite male sterile hybrid corn. In this case, the grain produced by the hybrid corn will have an oil content intermediate between that of both parents, with up to a 100% increase in energy content over that delivered by the elite variety alone. This system has been used successfully on hundreds of thousands of acres, under a variety of environmental conditions over a number of growing seasons, and has been vital to the broad introduction and commercial acceptance of the high-oil trait.

The other pathway taken has been the use of molecular biology to isolate, characterize, and modify individual genes, followed by plant transformation and trait analysis. The development of high-lysine corn and soybean for use in improved animal feeds illustrates the challenges that continually interlace metabolic engineering projects. Lysine is an essential amino acid that is limiting in corn- and soy-based animal feeds. Two of the key enzymes in the lysine biosynthetic pathway are aspartokinase (AK) and di-

hydrodipicolinic acid synthase (DHDPS), which are both feedback inhibited by lysine (6). Falco *et al.* (7) isolated bacterial genes encoding lysine-insensitive forms of AK and DHDPS from *Escherichia coli* and *Corynebacterium*, respectively. A deregulated form of the plant DHDPS was created by site-specific mutagenesis (8, 9). Despite assumed knowledge of the lysine biosynthetic pathway, expression of the deregulated enzymes in different tissues and species produced inconsistent results (7). Expression of these genes in tobacco leaves produced high concentrations of free lysine, as reported by others (10), but no accumulation was observed in tobacco seed with the use of either constitutive or seed-specific promoters. It was discovered that the failure to augment lysine concentrations in the seed was due to the presence of an active catabolic pathway for lysine in tobacco seed. In soybean and canola seeds, lysine accumulated sufficiently to more than double the total seed lysine content. However, the lysine catabolic products saccharopine or α amino adipic acid also accumulated in soybean and canola seeds, respectively, although in strikingly different quantities in the two species. In corn (9), expression with an endosperm-specific promoter did not lead to lysine accumulation, whereas expression with an embryo-specific promoter gave high concentrations of lysine, sufficient to raise overall lysine concentrations in seed 50 to 100% with only minor accumulation of catabolic products (see Table 1). Together these results offer several lessons. Microbial genetics provides a useful tool for isolating relevant genes, and microbes provide a rich resource for accessing diverse genes where enzymes with particular characteristics are needed. Bacterial or plant enzymes can be successfully expressed in specific plant tissues to alter regulation of a metabolic pathway. However, different tissues and different species may react to metabolic perturbations uniquely, and amounts of other metabolic products may unexpectedly change.

The development of soybeans with increased concentrations of oleic acid also provides an instructive example of seed trait modification. [See (11) for a review of fatty acid biosynthesis.] Unsaturated fatty acids are healthier than saturated fatty acids, and the monounsaturated form, oleic acid (18:1), is also more stable in frying and cooking applications than are the polyunsaturated forms, linoleic (18:2) and linolenic (18:3). Chemical hydrogenation is currently used to improve oxidative stability by increasing the concentrations of

DuPont Agricultural Products Experimental Station, Post Office Box 80402, Wilmington, DE 19880-0402, USA.

18:1 fatty acid, but hydrogenation also raises the concentration of trans fatty acids, which have been linked to higher health risks. Biochemical methods were used to purify and sequence the relevant soluble proteins from soybean, from which probes were created to isolate the genes that affect fatty acid composition. The membrane-bound enzymes in the pathway proved recalcitrant to biochemical purification strategies and were eventually cloned by T-DNA insertional mutagenesis in *Arabidopsis thaliana*. The cloning of the genes encoding each of the soybean FAD2 (12) desaturases enabled the cosuppression [silencing of an endogenous gene through the introduction of a homologous transgene; see (13)] of the seed-specific Fad2 desaturase, causing seed oleic acid concentrations to increase from 25% in wild-type lines to 85% in the transgenic lines. In contrast, a previously isolated line of canola, which carried a mutation in a non-seed-specific form of Fad2, was shown to have increased concentrations of oleic acid in roots as well as seeds (14). Subsequent testing revealed that this line had an unfavorable, cold-sensitive phenotype. The cosuppressed, seed-specific, soybean high oleic acid transgenic lines were shown to have excellent agronomic properties, demonstrating the importance of precise gene regulation. In related experiments, transgenic soybean lines with either increases or decreases in palmitic acid (15) were created by overexpressing or cosuppressing the soybean thioesterase gene in a seed-specific fashion. This work demonstrates that precise modifications of metabolic steps, when done correctly and when limited to storage tissues to avoid pleiotropic effects, can result in the desired alterations to seed composition.

The isolation of this set of fatty acid-modifying genes required a profound investment in time. In recognition of the limitations that conventional gene cloning imposed, an expressed sequence tag (EST) program was initiated to remove this barrier in the product development process. ESTs are developed by single-pass sequencing of the 5' end of complementary DNA clones made from selected biological samples. Since its inception, the DuPont EST program has generated a catalog of more than 1 million lanes of plant, microbe, and insect DNA sequences, thereby characterizing gene expression in a diverse set of biological states. This gene discovery process is extremely efficient, and the database now includes ESTs representing genes that code for most metabolic enzymes in the major crop species.

The program has been further extended through the adoption of an EST-based genomics paradigm in which genes are accessed and identified from a variety of plant and nonplant sources that have novel phenotypes. For example, vernolic acid, an epoxide fatty acid, and ricinoleic acid, a hydroxy fatty acid, are oleic acid derivatives used as hardeners in paints and

plastics (Fig. 1). Neither vernolic acid nor ricinoleic acid are made in soybean seeds. Complementary DNA libraries were made from Vernonia and castor bean seeds, plants known to produce the desired fatty acids, and were subjected to limited EST sampling. The genes that encoded the fatty acid-modifying enzymes were identified by DNA sequence homology to desaturase genes. After modification of the regulatory sequences, the genes were transferred to soybean, where the desired industrial oils were produced in seeds (16).

Presently, about half of the ESTs can be identified through DNA or protein sequence homology comparisons. Functional genomics programs have been established to aid in the identification of the remaining genes; in a number of cases, genes have been identified that have entered trait development projects. These programs include gene expression analyses with microarray and bead-based complementary DNA library comparisons and sequencing procedures, which allow the expression of a gene to be followed by tissue or organ type, by temporal or development stage, or under biotic, abiotic, or chemical stress conditions (17). Gene-tagging approaches (18), in which genetic elements interrupt or ectopically activate genes, are being used for forward and reverse genetic strategies, to identify genes associated with unusual phenotypes and conversely to identify phenotypes associated with specific genes. New protein analysis tools, such as the isolation of differentially expressed proteins by two-dimensional gel electrophoresis followed by identification with mass spectrometry (19), are also being used for gene identification. Finally, ge-

netically anchored physical maps of strategically important crops are being developed for use as scaffolds to localize ESTs to particular segments of a genome to facilitate map-based gene cloning and identification.

These transgenic trait development examples carry with them throughput constrictions associated with the plant transformation process. Transformation of many crop species, and particularly of elite lines of important crops, is difficult. High-throughput soybean transformation, whether achieved by biolistic or *Agrobacterium*-based methods (20), remains particularly challenging. Frequently coupled with the inefficiency of transformation is the inefficiency of achieving successful cosuppression events. Cosuppression typically occurs in only a limited number of transformants, for example, in soybean in only 10 to 20% of the transgenic events (21). Transgenes designed for overexpression can have their expression levels affected by the site at which they are inserted in the genome. Thus, in both cases, many transgenic lines must be identified initially so that lines that couple the desired phenotypic change in seed quality and acceptable agronomic performance can be isolated in subsequent field tests. Improvements to transformation frequencies for elite corn and soybean lines have been continual but incremental, so that it has been necessary to build high-throughput production laboratories to generate the numbers of transformation events required to meet product development needs.

The necessity of generating large numbers of transformation events can be partially reduced by improving the control of transgene

Table 1. Summary of results obtained from expressing a gene encoding a nonfeedback inhibited form of *dapA* in different plant species and tissues. The table shows whether a lysine increase was observed and the presence or absence of two major catabolites. The results demonstrate that the lysine catabolic pathway is regulated differentially in the different species and organs tested.

Tissue	Lysine increase	Saccharopine	α amino adipic acid
Tobacco leaf	Yes	No	No
Tobacco seed	No	No	Transient
Canola seed	Yes	No	High
Soybean seed	Yes	High	Low
Corn endosperm	No	No	No
Corn embryo	Yes	Low	Low

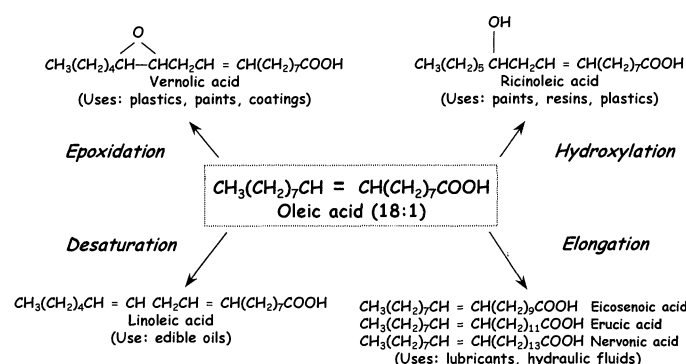


Fig. 1. Oleic acid can serve as the starting point for further fatty acid modifications with enzymes encoded by genes isolated from plant or nonplant sources. Some related fatty acids and their possible uses are shown.

expression. EST and microarray technologies are being used to identify promoters that meet specific quantitative, temporal, tissue-, or cell-specific gene expression requirements for a particular trait. Synthetic promoters composed of multiple cis-acting elements, combined with novel transcription factors, are being designed to increase gene expression levels. SARs and MARs, scaffold and matrix attachment regions, are being used to stabilize and enhance gene expression (22) and to reduce effects ascribed to the insertion site of a transgene (23). Specific transcription factors that regulate multiple genes in a particular metabolic pathway are being tested for pathway control and have the advantage of requiring the introduction of only a single transgene. The cre-lox site-specific recombination system has been used as a switch to control expression of a gene by excision of an intervening genetic tag (24). Finally, chemically regulated promoter systems have been used to regulate gene expression (25).

Another challenge for a robust product development process is that of functionality assessment. The earlier the properties of a transgene can be measured, the better. The soybean somatic embryo model system is an example of a preferred early screen (Fig. 2). Biochemical assays can be performed on transgenic somatic embryos generated during the soybean regeneration process, before plant development and seed maturation (26). Such assays have been used to identify gene constructs that effect desired changes in oil and protein content and composition. Early tests for biochemical parameters and functional properties on whole plants are also necessary and require high-

throughput screens. These assays serve as small-scale surrogate tests for more difficult biological measurements, such as oxidative stability for food applications or animal bioavailability for feed applications.

For many traits, functional attributes can only be assessed after the grain has been broken down into its component parts. Small-scale grain-processing facilities suitable for gram quantities of grain are necessary to isolate grain fractions similar to those produced in commercial-scale milling and processing plants. These small-scale facilities avoid the need for a time-consuming seed multiplication step that is necessary for large-scale assays. Tiered functional assays, in which high-throughput biochemical measurements are used to identify lines that subsequently enter low-throughput functional, nutritional, and sensory evaluations, also increase the efficiency of functionality assessments. An example of the utility of tiered compositional and functional assays comes from transgenic soybean lines with altered seed storage proteins. In soybean flour, two classes of seed storage proteins, 7S and 11S globulins, predominate, each with useful functional characteristics. Transgenic lines in which one or the other of these classes has been eliminated through cosuppression technology (27) were subjected to single seed-based protein polyacrylamide gel electrophoresis. Lines with altered 7S/11S ratios were then tested at increasing scale for emulsification, water and fat binding properties, gel-forming ability, and other parameters relevant to processing, cooking, and food manufacture. From these assessments, specific lines have been selected for seed in-

creases for testing in milk and meat replacement products.

The availability of multiple functionality tests that measure different end-use attributes increases the likelihood that value-added traits will be identified and used in multiple products. For example, when high oleic acid soybean oil is used as a food, the oil has heat and shelf life stability attributes that can be demonstrated in active oxygen-oxidative stability or frying tests (28). When these same soybeans are processed into flour, the flour has been shown to have improved sensory and shelf life properties due to the increased stability of the residual oil. Further, this same high oleic acid oil has been shown to have value as an industrial lubricant through metal wear tests with a small-scale hydraulic pump (29). Thus, a single soybean line was shown to have three distinct product applications because functionality tests for each application were available.

Soy or corn grain with a novel composition must be kept apart, or "identity preserved," from commodity grain throughout the production and distribution chains to retain its added value. Combining, or stacking, traits can add sufficient value to the final product to compensate for the cost of identity preservation, but the stacking process presents business and breeding challenges. Trait combinations must be chosen that address similar markets or, for ingredients, present similar acreage demands. Transgenes must be bred into the highest yielding germplasm, so that grain production does not incur yield penalties; the continuing improvement in overall production yield of elite lines makes this a challenging breeding target. The use of DNA markers has increased the efficiency of the breeding process. DNA-based molecular marker assays, including those for RFLP, RAPD, SSR, and SNP polymorphisms, have provided a method to accelerate germplasm analysis and trait mapping and a method to move traits rapidly into germplasm for different maturity or production zones (30). The use of DNA markers early in the trait introgression cycle has eliminated 1 to 2 years from breeding timelines. To accomplish this, laboratory facilities are needed where several million DNA-based diagnostic assays can be performed annually.

Another breeding challenge, specific to hybrid corn production, comes from the need to breed recessive or semidominant traits into both inbred parent lines for each production area. Dominant traits are preferred, in that they can be bred into only a single parent or into TopCross pollinators as described above. The TopCross system provides several advantages for transgene introductions in corn. Transgenic pollinator lines do not have to undergo multiple breeding cycles for yield parity, because the pollinator will be planted at a low seeding density and will have only a minimal effect on grain yields. Further, single pollinators can often be used in multiple F_1 hybrid combinations,

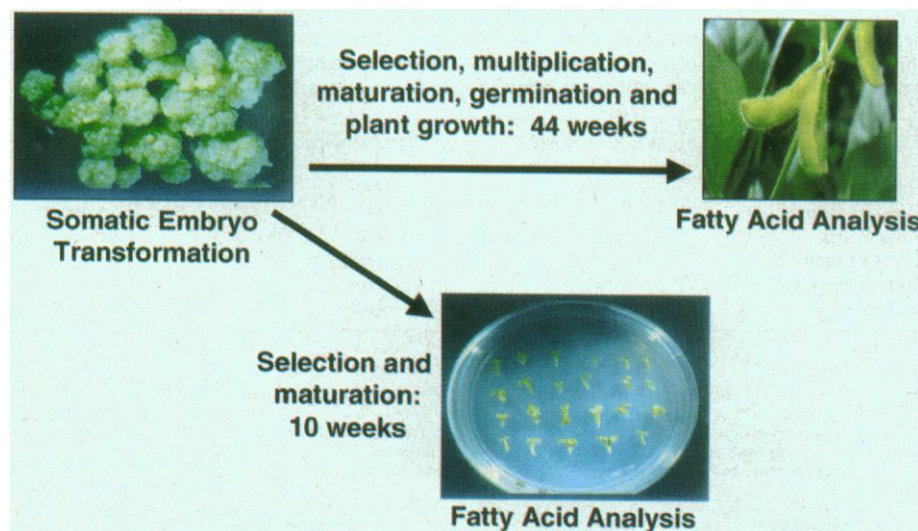


Fig. 2. The use of soybean somatic embryos as a more rapid tool for transgene testing. Several gene constructs can be tested with such embryos, and only those that produce a desirable phenotype are then used in the longer experiments necessary to produce whole plants and seeds. Soybean lines have been selected that produce large numbers of somatic embryos for use in the model system. Although the resulting somatic embryos cannot be regenerated to whole plants, they are ideal for testing multiple constructs. Lines used for regeneration of plants can also be tested at the somatic embryo stage of regeneration but produce smaller numbers of embryos and so are more suitable for production experiments with previously tested constructs.

thereby reducing the breeding requirements for transgene introductions. The TopCross system also provides a way to produce traits that have a negative effect on seed agronomic performance, such as on germination, but are desirable in the grain, as the trait can be introduced solely through the pollinator.

To date, a number of output traits have been developed at DuPont that are in various stages of functional testing or commercialization. These include corn and soybean lines for animal feed applications with increased concentrations of free lysine (7) or bound methionine (31). Soybean seeds with altered ratios of seed storage proteins (26) or reduced concentrations of trypsin inhibitors (32) have been produced, primarily for food applications. As described earlier, soybean lines with novel fatty acid compositions have been developed. High monounsaturate oils are useful in foods or food-processing applications where increased oxidative and thermal stability is desirable, and high saturate oils are useful as feedstocks for margarine manufacturing. All of these oils also have nutritional or health benefits. The soybean lines with reduced oligosaccharide concentrations described above and similar corn lines contain reduced concentrations of phytic acid and increased concentrations of free phosphorus. This combination provides nutritional value in animal feed and environmental value because of reduced bound phosphorus released in animal waste. Finally, a number of starches have been created for use in food and industrial applications, including polymers with altered

branching patterns, altered branch lengths, and novel monomer compositions (33).

Plants are capable of a stunning array of biochemical conversions. As an understanding of the regulation of these pathways is combined with the ability to engineer multiple genes simultaneously, the possibilities for producing novel products in plants will increase dramatically. Beyond the traditional agricultural products, opportunities exist for producing industrial feedstocks and polymers in crops and for producing pharmaceutical and nutraceutical products. Traits that have been initially produced in corn or soybeans may enter new markets through production in cereals such as wheat and rice or in plantation crops such as forest trees. The continuing integration of structural and functional genomics with trait development technologies will accelerate these advances.

References

1. B. J. Mazur, *Trends Biotechnol.* **3**, 319 (1995).
2. W. D. Hitz and S. A. Sebastian, World Patent Publication Number WO 98/45448 (1998).
3. D. E. Alexander, *Proc. 43rd Annu. Corn Sorghum Res. Conf.* **43**, 97 (1988).
4. S. V. Tingey, M. T. Jung, M. K. Hanafey, unpublished data.
5. R. R. Bergquist *et al.*, U.S. Patent 5706603 (1998); R. R. Bergquist, U.S. Patent 5704160 (1998).
6. G. Galili, *Plant Cell* **7**, 899 (1995).
7. S. C. Falco *et al.*, *Bio/Technology* **13**, 577 (1995).
8. J. M. Shaver *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1962 (1996).
9. S. C. Falco, personal communication.
10. K. Glassman *et al.*, World Patent Publication Number WO 89/11789 (1989); O. Shaul and G. Galili, *Plant J.* **2**, 203 (1992).
11. J. Ohlrogge and J. Browse, *Plant Cell* **7**, 957 (1995).
12. J. Okuley *et al.*, *ibid.* **6**, 147 (1994); W. D. Hitz *et al.*, *Plant Phys.* **105**, 635 (1994).
13. C. B. Taylor, *Plant Cell* **9**, 1245 (1997).
14. W. D. Hitz, unpublished observations.
15. ———, World Patent Publication Number WO 96/06936 (1996).
16. ———, U.S. Patent 5846784 (1998).
17. *Nature Genet.* **21**, 61 (1999) (and accompanying supplement); S. Brenner, U.S. Patent 5552278 (1996); U.S. Patent 5604097 (1997); U.S. Patent 5695934 (1997).
18. R. Walden *et al.*, *Plant Mol. Biol.* **26**, 1521 (1994); R. A. Martienssen, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2021 (1998).
19. A. R. Dongre *et al.*, *Trends Biotechnol.* **15**, 418 (1997); J. R. Yates, *Electrophoresis* **19**, 893 (1998).
20. T. Klein *et al.*, *Nature* **327**, 70 (1987); R. De Blaere *et al.*, *Methods Enzymol.* **153**, 277 (1987).
21. W. D. Hitz and T. Klein, unpublished observations.
22. S. Spiker and W. F. Thompson, *Plant Physiol.* **110**, 15 (1996); J. T. Odell and E. Krebbers, World Patent Publication Number WO 98/16650 (1998).
23. J. Jones *et al.*, *EMBO J.* **4**, 2411 (1985).
24. J. T. Odell *et al.*, *Plant Physiol.* **106**, 447 (1994).
25. H. Hershey *et al.*, U.S. Patent 5608143 (1997).
26. A. Kinney, *J. Food Lipids* **3**, 273 (1996).
27. G. Fader and A. Kinney, World Patent Publication Number WO 97/47731 (1997).
28. A. J. Kinney and S. Knowlton, in *Genetic Engineering for Food Industry: A Strategy for Food Quality Improvement*, S. Harlander and S. Roller, Eds. (Blackie Academic, London, 1998), pp. 193–213; S. Knowlton, World Patent Publication Number WO 97/40698 (1997).
29. J. L. Glancey *et al.*, *SAE Technical Paper 981999* (Society of Automotive Engineers, Warrendale, PA, 1998).
30. B. J. Mazur and S. V. Tingey, *Curr. Opin. Biotechnol.* **6**, 175 (1995).
31. C.-F. Chui *et al.*, World Patent Publication Number WO 92/14822 (1992).
32. G. Fader *et al.*, manuscript in preparation.
33. N. L. Hubbard *et al.*, World Patent Publication Number WO 97/22703 (1997); P. Caimi *et al.*, *Plant Physiol.* **110**, 355 (1996).

REVIEW

Nutritional Genomics: Manipulating Plant Micronutrients to Improve Human Health

Dean DellaPenna

The nutritional health and well-being of humans are entirely dependent on plant foods either directly or indirectly when plants are consumed by animals. Plant foods provide almost all essential vitamins and minerals and a number of other health-promoting phytochemicals. Because micronutrient concentrations are often low in staple crops, research is under way to understand and manipulate synthesis of micronutrients in order to improve crop nutritional quality. Genome sequencing projects are providing novel approaches for identifying plant biosynthetic genes of nutritional importance. The term "nutritional genomics" is used to describe work at the interface of plant biochemistry, genomics, and human nutrition.

Humans require a diverse, well-balanced diet containing a complex mixture of both macronutrients and micronutrients in order to maintain optimal health. Macronutrients—carbohydrates, lipids, and proteins (amino acids)—

make up the bulk of foodstuff and are used primarily as an energy supply. Micronutrients are organic or inorganic compounds present in small amounts and are not used for energy, but are nonetheless needed for good health. Essential micronutrients in the human diet include 17 minerals and 13 vitamins required at minimum levels to alleviate nutritional disorders (Table 1). Nonessential micronutrients

encompass a vast group of unique organic phytochemicals that are not strictly required in the diet, but when present at sufficient levels are linked to the promotion of good health.

Modifying the nutritional composition of plant foods is an urgent worldwide health issue as basic nutritional needs for much of the world's population are still unmet. Large numbers of people in developing countries exist on simple diets composed primarily of a few staple foods (cassava, wheat, rice, and corn) that are poor sources of some macronutrients and many essential micronutrients. Consequently, the diet of over 800 million people does not contain sufficient macronutrients, and micronutrient deficiencies are even more prevalent (1). As examples of the magnitude of micronutrient deficiencies, estimates place 250 million children at risk for

Department of Biochemistry/MS200, University of Nevada, Reno, NV 89557, USA. E-mail: della_d@med.unr.edu