

## Using Metabolomics To Estimate Unintended Effects in Transgenic Crop Plants: Problems, Promises, and Opportunities

Owen A. Hoekenga

Robert W. Holley Center for Agriculture and Health, USDA-ARS, Ithaca, NY

Transgenic crops are widespread in some countries and sectors of the agro-economy, but are also highly contentious. Proponents of transgenic crop improvement often cite the “substantial equivalence” of transgenic crops to their nontransgenic parents and sibling varieties. Opponents of transgenic crop improvement dismiss the substantial equivalence standard as being without statistical basis and emphasize the possible unintended effects to food quality and composition due to genetic transformation. Systems biology approaches should help consumers, regulators, and other stakeholders make better decisions regarding transgenic crop improvement by characterizing the composition of conventional and transgenically improved crop species and products. In particular, metabolomic profiling via mass spectrometry and nuclear magnetic resonance can make broad and deep assessments of food quality and content. The metabolome observed in a transgenic variety can then be assessed relative to the consumer and regulator accepted phenotypic range observed among conventional varieties. I briefly discuss both targeted (closed architecture) and nontargeted (open architecture) metabolomics with respect to the transgenic crop debate and highlight several challenges to the field. While most experimental examples come from tomato (*Solanum lycopersicum*), analytical methods from all of systems biology are discussed.

**KEY WORDS:** Metabolomics, transgenic crops/GMO, substantial equivalence, unintended effects, tomato (*Solanum lycopersicum*).

Not everything turns out quite the way people expect. Very few people shoot holes-in-one in golf, bowl perfect games or score 1600 on their SAT exams. We all try our best to accomplish our goals and largely do, no matter whether we are at work or at play. This is certainly true for plant breeding and improvement. Unintended effects to composition and quality occur at some frequency no matter what method is applied to make a bigger ear of corn or a tastier apple. Unintended effects represent statistically significant differences in phenotype, which, for example, could be the disappearance of a particular protein or alterations in polysaccharide composition.<sup>1</sup> These differences may or may not have an impact on quality or ultimate safety. Unintended effects can be further classified as predictable or unpredictable, where predictable changes can be explained based on an understanding of the underlying biology, from prior knowledge of the genetics of the parental varieties, or from the function of a transgene or the site of genomic integration. Unpredictable changes fall outside obvious explanation.<sup>1</sup> A National

Research Council (U.S.) taskforce estimated the probability of unintended effects due to a variety of methods for crop improvement.<sup>2</sup> They considered selection from homogenous populations to produce the smallest number of unintended effects (of any kind), with transformation of genes from closely related species producing similar outcomes to those observed by genetic crosses between existing germplasm pools. Genetic crosses between closely related species were estimated to produce a wider range of variance than the previous three methods. Transformation with genes from distantly related species was estimated to produce an even greater incidence of unintended effects, but fewer than those seen from mutational breeding methods, where deliberate mutagenesis using chemicals or ionizing radiation is used to induce novel genetic variation.

Most new plant varieties are generated by crossing highly related (homogenous) varieties together and selecting the small number that are more desirable than either parent. This style of plant improvement has led to steady gains in many traits, such as yield in maize, but is not likely to produce crops with radical changes in quality, composition, or adaptation to a particular stress. That this strategy will succeed presupposes that genetic diversity exists within the “elite” panel of varieties. In species with little genetic diversity, such as tomato, genetic

ADDRESS CORRESPONDENCE TO: Owen A. Hoekenga, Robert W. Holley Center for Agriculture and Health, USDA-ARS, Tower Road, Ithaca, NY 14853 (phone: +1 607 255 4502; fax: +1 607 255 1132; email: Owen.Hoekenga@ars.usda.gov).



hybridization between closely related species has introduced novel characteristics that are highly desirable. Cultivated tomato (*Solanum lycopersicum*) has been improved by replacing chromosomal segments with ones from wild relatives (e.g., *S. pennellii*), which contain pathogen resistance or fruit quality genes not found in *S. lycopersicum*.<sup>3,4</sup> Deliberate mutagenesis is a tool commonly used by geneticists to discover new genes and their functions. At least fourteen genes important for carbohydrate metabolism in maize seeds have been described by mutagenesis. Eight of these genes have been commercialized to some greater or lesser degree in the form of sweet corn (3 genes) or to provide specialty starches for various ethnic cuisines (5 genes).<sup>5</sup> Sweet corn varieties can carry the *sugary1* mutation, the *shrunken2* mutation, contain both *sugary1* and *sugary-enhancer1*, or carry all three mutations. Producers, consumers, regulators, and other stakeholders accept these products as safe. This implicitly means that the degree of phenotypic variation away from “normal” is acceptable as well.

Transgenic crop improvement is widespread for some commodities and in some parts of the world, but its use is hotly debated. In 2005, transgenic crops were planted on 87.2 million hectares around the world, including 47.4 million hectares in the United States.<sup>6</sup> In 2007, the majority of the major commodity crops in the United States were transgenic: soybean (91%), cotton (87%), and maize (73%), with wheat being the notable exception (crops with >5 million hectares planted).<sup>7</sup> Within a similar time frame, a poll conducted by the Mellman Group for the Pew Charitable Trusts determined that while 45% of Americans regard transgenic crops as “safe,” 29% regard them as “unsafe.”<sup>8</sup> A similar poll found that within the then 25-member state European Union, only 27% of respondents supported the use of transgenic crop improvement.<sup>9</sup>

There are many points of contention in the debate over transgenic crops, including arguments that rely upon cultural, economic, or scientific underpinnings.<sup>10–14</sup> Addressing cultural or economic concerns toward biotechnology is far beyond the scope of a minireview in a scientific journal. Thus, my comments will be confined to two of the scientific issues and how systems biology approaches, especially metabolomics, may help to constructively shape that debate.

### DEFINING THE DEBATE

Proponents of transgenic crop improvement often speak to the “substantial equivalence” (SE) of transgenic crops relative to their nontransgenic parents.<sup>15</sup> In this context, SE means that two varieties are so similar to one another that they can be taken to be same.<sup>16,17</sup> On one level, the notion of SE makes sense—a transgenic daughter con-

tains the construct of interest, which may express one or a small number of genes, in addition to whatever disturbance may have arisen from the site of genomic integration, relative to the nontransgenic mother variety. This degree of difference is much smaller in an absolute sense than the differences between two conventional varieties of the same market class. Maize has a high degree of nucleotide polymorphism, similar to that seen in potato, and far more than in tomato.<sup>18–20</sup> This is such a high degree that two maize varieties may be as different from each other as humans are from chimpanzees at the DNA level.<sup>21</sup> Any one would recognize that humans and chimpanzees are different, although it may take a highly trained botanist to tell the difference between, say cultivars B73 and Mo17 in maize.

In a plant-breeding context, near isogenic lines (NIL) are often considered to be SE to their progenitors. NILs are constructed by crossing back into one parental variety repeatedly over several (5–8) generations, so that the new stock is largely identical to that recurrent parent. If one assumes that a typical crop plant has 50,000 genes, after 5 generations of backcrossing, the parent and daughter will differ only 3.125% [i.e.,  $(1/2)^5$ ], which still represents differences at ~1600 genes. If 8 generations of backcrossing are used, parent and daughter will differ at only 0.4% of their genome, or 200 genes. This degree of difference between parent and daughter varieties is acceptable and does not keep NILs from being considered to SE to their progenitor mother varieties. However, the parent and daughter varieties are significantly different in at least one regard—the target trait that was improved in the daughter. Of course, the standard proposed by the Organization for Economic Cooperation and Development was of “substantial” equivalence rather than of “complete” or “total” equivalence.

Two principle objections are made to the SE concept, and by extension to the regulation of transgenic crops. First, opponents of transgenic crop improvement dismiss SE, as there is no specific, statistical basis for the standard.<sup>22</sup> In the original statement of principle, particular statistical tests were not explicitly defined to evaluate the SE standard.<sup>15</sup> “Substantial” is an adjective rather than an F-score or p-value, which begs the obvious question: How different is acceptably different versus unacceptably different? Opponents to transgenic crop improvement also may willingly miss the point that SE is a part of rather than the complete risk determination.<sup>23</sup> Second, opponents of transgenic crop improvement emphasize the possible consequences to food quality and composition due to unintended effects, either predictable or unpredictable ones. The original statement from the Organization for Economic Cooperation and Development emphasizes

testing for known toxins and quality biomarkers, examples of which can be found in the excellent review of Cellini et al. (2004).<sup>1</sup> This second objection begs the next obvious question: How can one investigate unintended effects using directed testing methods?

Systems biology can make a contribution to the transgenic crop debate by providing datasets that examine the composition of transgenic crops relative to their nontransgenic relatives. These hopefully comprehensive datasets can then be rigorously analyzed using the best statistical methods possible. To take full advantage of the literature available within a particular organism will require the existing data be easily compared from one study/laboratory to another and combined in meta-analyses. Transgenic varieties should not be the only subjects for analysis; a deeper understanding of naturally occurring variation, which represents the range of consumer-acceptable variation, should also be included and thus provide a frame of reference. Complete and unbiased information should facilitate decision-making by consumers, regulators, and other stakeholders and provide more substantial bases for those decisions. I would argue that the combination of analytical tools available to the community, in gene, protein, and metabolite expression analysis and identification, are mature and useful to examine transgenic crops and potential unintended effects. Rather, our ability to design and analyze these experiments will be the factors that limit their utility and contributions to the transgenic crop debate.

#### **METABOLOMIC DATA CAN BE GATHERED FROM MANY PLATFORMS**

To limit the scope of this review, I will confine my comments to studies within tomato, although excellent examples can be found in many other crops.<sup>1</sup> Tomato has several notable features in this context. It is an economically important crop, with market classes that have divergent properties (e.g., fresh eating versus processing tomatoes; heirloom varieties versus improved hybrids).<sup>24</sup> Tomato is a model organism for genomic studies, with genome sequencing underway, a large collection of molecular markers, cDNA and genomic libraries, and a worldwide community of researchers.<sup>25</sup> Tomato also represents an early false-start in transgenic crop improvement (e.g., *FlavrSavr*), serving as a cautionary tale for future transgenic crops.<sup>26</sup>

Perhaps the principal agronomic consideration for tomato is the quality of the fruit, which is due to a large number of factors and has been the subject of intense research for many years.<sup>27</sup> Research into fruit quality provides answers to basic scientific questions that have a significant relevance to applied research. Many different cat-

egories of chemical compounds contribute to fruit quality, including sugars, organic acids, amino acids, fatty acids, isoprenoids, and polyphenolic compounds. This wide range of quality biomarkers means that a wide range of separation chemistries have been used to investigate the tomato metabolome, using both targeted and nontargeted metabolomics. Targeted metabolomics are by far the most common studies, as most research programs are focused on understanding or improving a single target trait. Thus, a great deal of information exists that describes the range of stakeholder acceptable phenotypic variation, while this information may not exist in an easily accessible format.

Small molecules can have large effects. The variation in the ratio between sweetness and acidity can cause tomatoes to taste sharp, sweet, insipid, or lovely.<sup>28</sup> For breeding programs, simple, low-cost assays are required to accommodate the scale of research, where thousands of samples may need to be analyzed in a short period of time.<sup>29</sup> While rather limited in their depth, phenotypic surveys of diverse germplasm have a very broad scope and help define the range of acceptable phenotypic variation. These kinds of organic acid and sugar data can be leveraged using gene expression analysis to look for the underlying genetic causes of fruit quality, leveraging applied data sets into more basic research results.<sup>30</sup> The same trait information on carbohydrates and organic acids can also be developed using more sophisticated tools such as nuclear magnetic resonance (NMR) spectroscopy, which identify far more compounds per assay than enzymatic or colorimetric methods but at far lower throughput.<sup>31</sup> NMR spectroscopy also offers a high probability of an unambiguous structural determination for a novel metabolite of particular interest. Gas chromatography (GC) paired with mass spectrometry (MS) (GC-MS) is an alternative platform for broad-scope metabolomic profiling, with notable benefit of higher throughput relative to NMR.<sup>32</sup> On the other hand, GC-MS does require chemical derivatization, which may exclude classes of metabolites from the analysis, and also may not produce sufficient information for the unambiguous identification of a particular metabolite. However, the intersection of multiple data sets developed on these complementary analytical platforms offers a powerful strategy to analyze metabolomes.

Color and aroma are other targets for tomato improvement and study. Many of the pigments in tomato are isoprenoids, such as carotenoids, while others are polyphenolics, such as flavonoids.<sup>33</sup> Traditionally, liquid chromatography (LC) protocols with commercial standards have been sufficient for carotenoid profiling.<sup>34</sup> However, as investigators wish to build more complete estimates for various metabolomes, LC-MS is becoming more common for analysis of isoprenoids. The MS portion of the

experiment can occur either in-line with the LC or in an off-line mode.<sup>35,36</sup> In-line MS simplifies work flow, while off-line MS may provide greater sensitivity due to the greater reduction of sample complexity.<sup>36</sup> NMR spectroscopy is also an option for isoprenoid profiling, which is effective at distinguishing *E* from *Z* isomers while MS methods cannot.<sup>37</sup> This is significant, as different carotenoid isomers may have different biological activities and thus nutritive qualities.<sup>38</sup> Carotenoid composition can change through food preparation and processing, both in quality (i.e., isomerization) and identity (i.e., degradation by heat); analysis of both raw and cooked samples may be necessary to present a full description of the isoprenoid portion of a metabolome.<sup>39,40</sup> In addition to color, carotenoids also contribute to fruit aroma, as do fatty acid and amino acid derivatives.<sup>41</sup> As all three categories represent volatile compounds, GC and GC-MS are the platforms of choice for separation and identification.<sup>41,42</sup> A relatively small number of compounds contribute to aroma, which has facilitated the discovery of key biosynthetic and regulatory genes using traditional genetic and transgenic approaches.<sup>41,43–46</sup>

Flavonoids, like carotenoids, contribute to both the color and nutritional quality of tomato fruit.<sup>47</sup> Like carotenoids, LC-based approaches have been widely utilized to quantitate and identify flavonoids, using either commercially available standards or MS.<sup>48,49</sup> Flavonoids can degrade during food/sample preparation so that analysis of raw and cooked samples may need to be performed to insure the relevance of the results.<sup>40</sup> The regulation of flavonoid biosynthesis has been studied for decades and many genes are available for direct study.<sup>33</sup> As a result, many studies in tomato utilize either conventional mutants (e.g., *high-pigment1*) or transgenic strategies to manipulate the flavonoid pathway and products.<sup>33,50–55</sup> A third method to manipulate the flavonoid pathway (and others) is to observe changes through developmental time; the ripening process makes dramatic changes to fruit composition.<sup>56</sup> In addition, there is compartmentalization within fruit for many different compounds.<sup>35</sup> Tissue type explained far more variance in chemical species observed than did genetic differences in a small panel of fresh market/greenhouse adapted cultivars.<sup>56</sup> Even greater resolution of variance was obtained by using a time-course approach to dissect the tissue-specific metabolomes in ripening fruit. Nearly 70% of phenotypic variance observed was explained by a principal component analysis, with the obvious clustering according to tissue and developmental stage.<sup>56</sup> With further application of this sophisticated experimental design, more of the phenotypic variance could be explained; however, this dramatically increased

the number of samples analyzed, as five tissues and five developmental time points were examined.<sup>56</sup>

Nontargeted metabolomic approaches offer many advantages over targeted ones, if the goal is to characterize unintended effects.<sup>57,58</sup> If a primary concern is to understand the predictable and unpredictable unintended effects to composition and quality, then we need to look beyond the obvious quality and toxin biomarkers. However, this kind of survey does not require the unambiguous identification of every compound in the first draft of the analysis. Molecular fingerprinting is likely sufficient in the first draft of a metabolomic comparison between conventional and transgenic varieties.<sup>1,57,59</sup> For example, if there are 2000 negative ions produced by MS and only 10 are different between the conventional and transgenic varieties, then obviously those require the most attention in the second round of metabolomic analysis. Nontargeted metabolomics also offer advantages to the analysis of conventional varieties, as they have no preconceptions about what the so-called “interesting” molecules will be, beyond the limitations of whatever extraction or separation chemistry is utilized for sample preparation and analysis. Nontargeted surveys have been conducted using MS on conventional mutants, developmental time series, diverse germplasm, and transgenic varieties.<sup>35,56,60,61</sup> Similar studies have been conducted with NMR spectroscopy.<sup>55,62</sup>

One of the common difficulties with analyzing MS or NMR data is the need for a highly curated database to best understand the spectra produced during the course of the experiment. Fortunately, recent developments in tomato, in particular and within the larger metabolomics community, provide these resources. A metabolite survey of approximately 100 Dutch tomato cultivars was conducted using LC-MS and MS/MS as necessary.<sup>63</sup> A high-throughput methodology was described in a separate publication; software for comparison of LC-MS and GC-MS spectra (MetAlign) and the tomato metabolite database (MoToDB) are both freely available.<sup>60,63</sup> A more comprehensive (>20,000 compounds) public database has been organized to warehouse MS and NMR data.<sup>64</sup> The latter group has also made their software (Sesame) freely available to encourage community participation in the curation of the metabolite database.<sup>65</sup> As the Sesame software was originally written to manage proteomic data, comparative tools will hopefully be developed for the joint analysis of metabolite and protein data.<sup>66</sup>

#### EXPERIMENTAL DESIGN AND ANALYSIS ARE THE LIMITING FACTORS

Many experiments executed to date have had well-defined goals but have lacked thorough descriptions of the

experimental protocols and results. This is a general problem that is now being addressed by the establishment of community standards for reporting the details of experimental design and execution. This solution was first promoted in the gene expression community as the MIAME standard (the Minimum Information About a Microarray Experiment).<sup>67</sup> For tomato, several public databases are MIAME compliant, including the Tomato Expression Database, the Gene Expression Omnibus, and the Plant Expression Database.<sup>68–70</sup> One of the limitations for gene expression data mining is this distribution of community-generated results. In the cereals genomic community, computational biology tools (“middleware”) have been developed to facilitate the invisible exchange of information between model organism databases and clade-oriented databases.<sup>71</sup> Hopefully, the development of middleware tools will be expanded to promote the amalgamation of data between multiple web-accessible databases and thus encourage meta-analysis. The ability to consolidate all of the tomato gene expression data would be highly useful to help connect gene and metabolite expression analyses, especially as several microarray-based studies have been made on tomato fruit development.<sup>72–77</sup> Tools have been developed for the Arabidopsis community, such that both gene and metabolite expression analysis can be conducted within a single website.<sup>78</sup> At the SOL Genomics Network, the clade-oriented database that includes tomato, comparative genomics tools are apparently being developed, while some already exist at the Tomato Expression Database, a model organism database.<sup>25,69</sup>

The same issue of data exchange exists within the proteomic research community. As with MIAME, a similar community standard has been established for protein expression profiling (proteomics) and is called the MIAPE, for Minimum Information About a Proteomics Experiment.<sup>79</sup> For research groups that do apply both genomic and proteomic methods, tools have been developed to merge these data sets into single, coherent entities for consolidated analysis.<sup>80,81</sup> Studies have described the tomato fruit proteome as it changes through developmental time.<sup>82,83</sup> These methods have not been applied to tomato, although there is sufficient information in the literature to make that joint comparison of gene and protein expression.

There is at least one obstacle to effectively consolidating disparate data sources developed by different investigators. Data mining and middleware function are stymied by the tendency of researchers to use idiosyncratic or incomplete language to describe what they did and how they did it. The use of controlled vocabularies to describe genes, traits, and phenotypes can overcome some of these

difficulties.<sup>84,85</sup> Once controlled vocabularies are in place, gene and protein expression profiling experiments can be analyzed much more easily.<sup>86–88</sup> This is in large part due to the fact that more of the decision-making can be made automatically rather than subjectively by the particular researcher, which is a key consideration when there can be millions of data points. Controlled vocabularies allow data sets to be consolidated with confidence when assembled by multiple researchers since everyone has agreed upon the methods to be used for organization and for the classification of their results.

One of the driving forces for organization of metabolomic, genomic, and proteomic datasets is their scale. All of these approaches generate waves of information that can quickly overwhelm the investigator, which make the visualization and appreciation of the results difficult. In the elegant study of Fraser et al.,<sup>32</sup> normal and transgenic tomatoes overexpressing *phytoene synthase1* were examined using directed metabolomics, gene expression, protein activity, and physiological parameters. This multidisciplinary approach allowed them to construct the most complete understanding of changes in the metabolism of the transgenic tomato fruit relative to nontransgenic. Their data were summarized in a color-coded metabolic pathway diagram, which made for a highly effective transmission of the scope of the changes observed. However, these diagrams were drawn for only two of the ten possible comparisons that could have been made (two varieties, five developmental stages tested) and focused only on a small subset of overall metabolic processes.<sup>32</sup> For similar reasons, it seems highly likely that the sense of scale has interfered with reporting on the nondirected LC-MS studies on the panel of ~100 Dutch tomato varieties; all we have heard to date have been the descriptions of the methods necessary to process that many samples and the database that hosts some fraction of their results.<sup>60,63</sup> It must be a difficult process to analyze and visualize a dataset that contains hundreds of thousands of mass identifications derived from hundreds of LC runs, let alone parse this story into appreciable chapters. This work will provide a broad and deep description of the range of phenotypic variation observed and provide highly useful information to both basic and applied biologists working with tomato.

However, this is precisely the kind of information that is required to fully examine the predictable and unpredictable unintended effects to the composition of transgenic crops. Given the scope and depth of experiments on this scale, it is unlikely that single investigators will be able to conduct and analyze these experiments. I think that only through well-organized community efforts, where individuals laboratories divide responsibilities and conduct

experiments that can serve multiple functions, will the data necessary to judge the existence and importance of unintended effects be possible. These efforts need to be sufficiently large to describe the phenotypic range observed among a diverse selection of conventional varieties using metabolomic, genomic, and proteomic profiling methods. Once this is accomplished, the same metrics can be applied to a realistic panel of transgenic varieties with possible commercial application. While the characterization of transgenic varieties will likely serve only a biotechnology risk-assessment purpose, the natural diversity survey will provide broad and deep knowledge of food composition and quality. Such a systems biology approach should answer a large number of applied and basic biological questions, thus justifying the investment.

### FUTURE PROSPECTS

The MIAME and MIAPE standards are now requirements for publishing microarray and proteomic experiments in many journals. Hopefully, similar standards will be applied to metabolomic data as well. As more and more of these datasets are deposited in publicly accessible databases, meta-analyses that integrate multiple levels of information will allow us to ask many different systems biology questions. The adoption of controlled vocabularies for gene, trait, and phenotypic ontologies will further assist these meta-analyses. The benefit of this ability to leverage large collections of data should be obvious to the scientific community. Likewise, the identification of genetically informative populations has been very effective to address important biomedical and agronomic questions, such as the identification of cancer risk factors and genes important for carotenoid biofortification in staple crops.<sup>89,90</sup> If these genetically informative populations are studied using metabolomic, genomic, and proteomic methods, this should provide an immediately useful but also durable resource. From this base of knowledge, the range and identity of unintended effects to composition and quality of transgenic foods can be assessed in the most complete manner, and help inform consumers, regulators, and other stakeholders in their decision-making.

### ACKNOWLEDGMENT

This research program is supported by US Department of Agriculture, Agricultural Research Service base funds to O.A. Hoekenga, J.J. Giovannoni, and L.V. Kochian, "Determination and characterization of unintended effects in genetically modified crop plants" 1907-21000-028-00.

### REFERENCES

1. Cellini F, Chesson A, Colquhoun I, et al. Unintended effects and their detection in genetically modified crops. *Food Chem Toxicol* 2004;42(7):1089–1125.
2. Committee on Identifying and Assessing Unintended Effects of Genetically Engineered Foods on Human Health, N.R.C. *Safety of Genetically Engineered Foods: Approaches to Assessing Unintended Health Effects*. Washington DC: National Academies Press; 2004: 1–15.
3. Frary A, Nesbitt TC, Grandillo S, et al. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 2000;289(5476):85–88.
4. Martin GB, Williams JG, Tanksley SD. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proc Natl Acad Sci USA*, 1991;88(6):2336–2340.
5. Tracy W. Sweet corn. In Hallauer A (ed): *Specialty Corns*, Boca Raton, FL: CRC Press, 2001:155–197.
6. Brookes G, Barfoot P. GM Crops: The First Ten Years—Global Socio-Economic and Environmental Impacts. *ISAAA Policy Briefs*. Ithaca, NY, ISAAA; 2006:1–4. Available at: <http://www.isaaa.org/resources/publications/briefs/36/default.html>.
7. National Agricultural Statistics Service. USDA. Washington DC: US Government Printing Office; 2007:25. Available at: [http://www.nass.usda.gov/Publications/Ag\\_Statistics/2007/CHAP01.PDF](http://www.nass.usda.gov/Publications/Ag_Statistics/2007/CHAP01.PDF). Accessed February 13, 2008.
8. Mellman Group. Review of Public Opinion Research on Behalf of the Pew Initiative on Food and Biotechnology; 2006:4–5. Available at [http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Public\\_Opinion/Food\\_and\\_Biotechnology/2006summary.pdf](http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Public_Opinion/Food_and_Biotechnology/2006summary.pdf). Accessed 2/13/2008.
9. Gaskell G, Allansdottir A, Allum N, et al. Europeans and biotechnology in 2005: Patterns and trends. *Eurobarometer* 2006;64(3):19. Available at [http://ec.europa.eu/research/press/2006/pdf/pr1906\\_eb\\_64\\_3\\_final\\_report-may2006\\_en.pdf](http://ec.europa.eu/research/press/2006/pdf/pr1906_eb_64_3_final_report-may2006_en.pdf). Accessed February 13, 2008.
10. Sheldon I. Europe's regulation of agricultural biotechnology: precaution or trade distortion? *J Agr Food Ind Org* 2004;2(4) article 4. Available at: <http://www.bepress.com/jafio/vol2/iss2/art4/>. Accessed February 13, 2008.
11. Knight JG, Mather DW, Holdsworth DK, Ermen DF. Acceptance of GM food—an experiment in six countries. *Nat Biotechnol* 2007;25(5):507–508.
12. Tothova M, Oehmke J. Genetically modified food standards as trade barriers: harmonization, compromise and sub-global agreements. *J Agr Food Ind Org* 2004;2(5) article 5. Available at: <http://www.bepress.com/jafio/vol2/iss2/art5/>. Accessed February 13, 2008.
13. Kalaitzandonakes N, Alston JM, Bradford KJ. Compliance costs for regulatory approval of new biotech crops. *Nat Biotechnol* 2007;25(5):509–511.
14. Kalaitzandonakes N, Bijman J. Who is driving biotechnology acceptance? *Nat Biotechnol* 2003;21(4):366–369.
15. Organization for Economic Co-operation and Development (OECD). Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles. Paris. 1993:10–13. Available at: <http://www.oecd.org/dataoecd/57/3/1946129.pdf>. Accessed February 13, 2008.
16. Bradford KJ, Van Deynze A, Gutterson N, et al. Regulating transgenic crops sensibly: lessons from plant breeding, biotechnology and genomics. *Nat Biotechnol* 2005;23(4):439–444.
17. Ramessar K, Peremarti A, Gomez-Galera S, et al. Biosafety and risk assessment framework for selectable marker genes in transgenic crop plants: a case of the science not supporting the politics. *Transgenic Res* 2007;16(3):261–280.
18. Sharopova N, McMullen MD, Schultz L, et al. Development and mapping of SSR markers for maize. *Plant Mol Biol* 2002;48(5–6):463–481.
19. He C, Poysa V, Yu K. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. *Theor Appl Genet* 2003;106(2):363–373.
20. Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R. Isolation, characterisation and mapping of simple

- sequence repeat loci in potato. *Mol Gen Genet* 1998;259(3):233–245.
21. Buckler ES, Gaut BS, McMullen MD. Molecular and functional diversity of maize. *Curr Opin Plant Biol* 2006;9(2):172–176.
  22. Millstone E, Brunner E, Mayer S. Beyond 'substantial equivalence'. *Nature* 1999; 401(6753):525–526.
  23. Kok EJ, Kuiper HA. Comparative safety assessment for biotech crops. *Trends Biotechnol* 2003;21(10):439–444.
  24. Van Deynze AE, Stoffel K, Buell RC, et al. Diversity in conserved genes in tomato. *BMC Genomics* 2007;doi:10.1186/1471-2164-8-465.
  25. Mueller LA, Solow TH, Taylor N. The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol* 2005;138(3):1310–1317.
  26. Martineau, B. *First Fruit: The Creation of the Flavor Savr Tomato and the Birth of Biotech Food*. New York: McGraw-Hill Professional. 2001:1–269.
  27. Giovannoni, J. Molecular Biology of Fruit Maturation and Ripening. *Annu Rev Plant Physiol Plant Mol Biol* 2001;52:725–749.
  28. Baldwin EA, Scott JW, Shewmaker CK, Schuch W. Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *Hortsci* 2000;35:1013–1022.
  29. Velterop JS, Vos F. A rapid and inexpensive microplate assay for the enzymatic determination of glucose, fructose, sucrose, L-malate and citrate in tomato (*Lycopersicon esculentum*) extracts and in orange juice. *Phytochem Anal* 2001;12(5):299–304.
  30. Baxter CJ, Carrari F, Bauke A, et al. Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. *Plant Cell Physiol* 2005;46(3):425–437.
  31. Sobolev AP, Segre A, Lamanna R. Proton high-field NMR study of tomato juice. *Mag Reson Chem* 2003;41:237–245.
  32. Fraser PD, Enfissi EM, Halket JM, et al. Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism. *Plant Cell* 2007;19(10):3194–3211.
  33. Grotewold E. The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* 2006;57:761–780.
  34. Enfissi EM, Fraser PD, Lois LM, Boronat A, Schuch W, Bramley PM. Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. *Plant Biotechnol J* 2005;3(1):17–27.
  35. Capanoglu E, Beekwilder J, Boyacioglu D, Hall R, de Vos CH. Changes in antioxidant and metabolite profiles during production of tomato paste. *J Agric Food Chem* 2008;56(3):964–973.
  36. Fraser PD, Enfissi EM, Goodfellow M, Eguchi T, Bramley PM. Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant J* 2007;49(3):552–564.
  37. Tiziani S, Schwartz SJ, Vodovotz Y. Profiling of carotenoids in tomato juice by one- and two-dimensional NMR. *J Agric Food Chem* 2006;54(16):6094–6100.
  38. Yeum KJ, Russell RM. Carotenoid bioavailability and bioconversion. *Annu Rev Nutr* 2002;22:483–504.
  39. Dewanto V, Wu X, Adom KK, Liu RH. Thermal processing enhances the nutritional value of tomatoes by increasing total antioxidant activity. *J Agric Food Chem* 2002;50(10):3010–3014.
  40. Re R, Bramley PM, Rice-Evans C. Effects of food processing on flavonoids and lycopene status in a Mediterranean tomato variety. *Free Radic Res* 2002;36(7):803–810.
  41. Tieman DM, Zeigler M, Schmelz EA, et al. Identification of loci affecting flavour volatile emissions in tomato fruits. *J Exp Bot* 2006;57(4):887–896.
  42. Tikunov Y, Lommen A, de Vos CH, et al. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol* 2005;139(3):1125–1137.
  43. Tieman DM, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proc Natl Acad Sci USA* 2006;103(21):8287–8292.
  44. Tieman DM, Zeigler M, Schmelz EA, et al. Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry* 2007;68(21):2660–2669.
  45. Simkin AJ, Schwartz SH, Auldridge M, Taylor MG, Klee HJ. The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *Plant J* 2004;40(6):882–892.
  46. Basset GJ, Quinlivan EP, Ravel S, et al. Folate synthesis in plants: the p-aminobenzoate branch is initiated by a bifunctional PabA-PabB protein that is targeted to plastids. *Proc Natl Acad Sci USA* 2004;101(6):1496–1501.
  47. Ross JA, Kasum CM. Dietary flavonoids: bioavailability, metabolic effects, and safety. *Annu Rev Nutr* 2002;22:19–34.
  48. Mauri PL, Iemoli L, Gardana C, et al. Liquid chromatography/electrospray ionization mass spectrometric characterization of flavonol glycosides in tomato extracts and human plasma. *Rapid Commun Mass Spectrom* 1999;13(10):924–931.
  49. Stewart AJ, Bozonnet S, Mullen W, Jenkins GI, Lean ME, Crozier A. Occurrence of flavonols in tomatoes and tomato-based products. *J Agric Food Chem* 2000;48(7):2663–2669.
  50. Verhoeven ME, Bovy A, Collins G, et al. Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *J Exp Bot* 2002;53(377): 2099–2106.
  51. Davuluri GR, van Tuinen A, Fraser PD, et al. Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nat Biotechnol* 2005;23(7):890–895.
  52. Bovy A, de Vos CH, Kemper M, et al. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *Plant Cell* 2002;14(10):2509–2526.
  53. Nicoletti I, De Rossi A, Giovanazzo G, Corradini D. Identification and quantitation of stilbenes in fruits of transgenic tomato plants (*Lycopersicon esculentum* Mill.) by reversed phase HPLC with photodiode array and mass spectrometry detection. *J Agric Food Chem* 2007;55(9):3304–3311.
  54. Muir SR, Collins GJ, Robinson S, et al. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat Biotechnol* 2001;19(5):470–474.
  55. Le Gall G, DuPont MS, Mellon FA, et al. Characterization and content of flavonoid glycosides in genetically modified tomato (*Lycopersicon esculentum*) fruits. *J Agric Food Chem* 2003;51(9):2438–2446.
  56. Moco S, Capanoglu E, Tikunov Y, et al. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *J Exp Bot* 2007;58(15–16):4131–4146.
  57. Noteborn HP, Lommen A, van der Jagt RC, Weseman JM. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J Biotechnol* 2000;77(1):103–114.
  58. Kuiper HA, Kok EJ, Engel KH. Exploitation of molecular profiling techniques for GM food safety assessment. *Curr Opin Biotechnol* 2003;14(2):238–243.
  59. Catchpole GS, Beckmann M, Enot DP, et al. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci USA* 2005;102(40):14458–14462.
  60. de Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2007;2(4):778–791.

61. Bino RJ, de Vos CH, Lieberman M, et al. The light-hyperresponsive high pigment-2dg mutation of tomato: alterations in the fruit metabolome. *New Phytol* 2005;166(2):427–438.
62. Defernez M, Colquhoun IJ. Factors affecting the robustness of metabolite fingerprinting using 1H NMR spectra. *Phytochem* 2003. 62(6):1009–1017.
63. Moco S, Bino RJ, Vorst O, et al. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* 2006;141(4):1205–1218.
64. Cui Q, Lewis IA, Hegeman AD, et al. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* 2008;26(2):162–164.
65. Markley JL, Anderson ME, Cui Q, et al. New bioinformatics resources for metabolomics. *Pac Symp Biocomput* 2007:157–168.
66. Zolnai Z, Lee PT, Li J, et al. Project management system for structural and functional proteomics: Sesame. *J Struct Funct Genomics* 2003;4(1):11–23.
67. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29(4):365–371.
68. Barrett T, Suzek TO, Troup DB, et al. NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res* 2005;33(Database issue):D562–D566.
69. Fei Z, Tang X, Alba R, and Giovannoni JJ, Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Res* 2006. 34(Database issue): D766–D770.
70. Shen L, Gong J, Caldo RA, et al. BarleyBase—an expression profiling database for plant genomics. *Nucleic Acids Res* 2005;33(Database issue):D614–D618.
71. Casstevens TM, Buckler ES. GDPC: connecting researchers with multiple integrated data sources. *Bioinformatics* 2004;20(16):2839–2840.
72. Alba R, Payton P, Fei Z, et al. Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* 2005;17(11):2954–2965.
73. Kok EJ, Franssen-van Hal NL, Winnubst LN, et al. Assessment of representational difference analysis (RDA) to construct informative cDNA microarrays for gene expression analysis of species with limited transcriptome information, using red and green tomatoes as a model. *J Plant Physiol* 2007;164(3):337–349.
74. Kolotilin I, Koltai H, Tadmor Y, et al. Transcriptional profiling of high pigment-2dg tomato mutant links early fruit plastid biogenesis with its overproduction of phytonutrients. *Plant Physiol* 2007;145(2):389–401.
75. Lemaire-Chamley M, Petit J, Garcia V, et al. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiol* 2005;139(2):750–769.
76. Baxter CJ, Sabar M, Quick WP, Sweetlove LJ. Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. *J Exp Bot* 2005;56(416):1591–1604.
77. Eriksson EM, Bovy A, Manning K, et al. Effect of the colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. *Plant Physiol* 2004;136(4):4184–4197.
78. Kopka J, Schauer N, Krueger S, et al. GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* 2005;21(8):1635–1638.
79. Orchard S, Hermjakob H, Julian RK Jr, et al. Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics* 2004;4(2):490–491.
80. Waters KM, Pounds JG, Thrall BD. Data merging for integrated microarray and proteomic analysis. *Brief Funct Genom Proteom* 2006;5:261–272.
81. Fagan A, Culhane AC, Higgins DG. A multivariate analysis approach to the integration of proteomic and gene expression data. *Proteomics* 2007;7(13):2162–71.
82. Rose JK, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS. Tackling the plant proteome: practical approaches, hurdles and experimental tools. *Plant J* 2004;39(5):715–733.
83. Faurobert M, Mihr C, Bertin N, et al. Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiol* 2007;143(3):1327–1346.
84. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25–29.
85. Ware D, Jaiswal P, Ni J, et al. Gramene, a tool for grass genomics. *Plant Physiol* 2002. 130(4):1606–1613.
86. Whetzel PL, Parkinson H, Causton HC, et al. The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics* 2006;22(7):866–873.
87. Sun H, Fang H, Chen T, Perkins R, Tong W. GOFFA: Gene Ontology For Functional Analysis - A FDA Gene Ontology Tool for Analysis of Genomic and Proteomic Data. *BMC Bioinformatics* 2006;7(Suppl 2):S23.
88. Werner T. Bioinformatics applications for pathway analysis of microarray data. *Curr Opin Biotechnol* 2008;19(1):50–54.
89. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007;39(8):977–983.
90. Harjes, CE, Rocheford TR, Bai L, et al. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 2008;319(5861) 330–333.