The application of transcriptomics in the comparative safety assessment of (GMO-derived) plant products

E.J. Kok

Promotoren

Prof. dr. ir. I.M.C.M. Rietjens
Hoogleraar Toxicologie, Wageningen Universiteit

Prof. dr. M.R. Müller
Hoogleraar Voeding, Metabolisme en Genomics, Wageningen Universiteit

Co-promotoren

Dr. ir. H.A. Kuiper
Onderzoeker, RIKILT – Instituut voor Voedselveiligheid, Wageningen

Prof. dr. ir. J. Keijer
Groepshoofd, RIKILT – Instituut voor Voedselveiligheid, Wageningen

Promotiecommissie

Prof. dr. R.J. Bino (Wageningen Universiteit)
Prof. dr. H.V. Davies (Scottish Crop Research Institute, Dundee, UK)
Prof. dr. W.J. Stiekema (Wageningen Universiteit)
Prof. dr. W.M. de Vos (Wageningen Universiteit)

Dit onderzoek is uitgevoerd binnen de onderzoeksschool VLAG

The application of transcriptomics in the comparative safety assessment of (GMO-derived) plant products

Esther Jeannette (van Leeuwe-) Kok

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de Rector Magnificus
van Wageningen Universiteit,
Prof. dr. M.J. Kropff,
in het openbaar te verdedigen
op vrijdag 13 juni 2008
des morgens te elf uur in de Aula

Esther Kok

The application of transcriptomics in the comparative safety assessment of (GMO-derived) plant products

PhD thesis Wageningen University and Research Centre, the Netherlands
With summaries in English and Dutch.

ISBN 978-90-8504-929-6

Aan

mijn vader† en moeder,

Sarina† en Carolien

Abstract

National and international organizations have discussed current approaches to the safety assessment of complex (plant) food products in general and the safety assessment of GMO-derived food products in particular. One of the recommendations of different expert meetings was that the new analytical techniques, in particular the 'omics' approaches, need to be explored for their potential to improve the analysis and thereby the toxicological and nutritional assessment of complex (GMO-derived) plant products. This thesis aims to further explore this approach in general and, more specifically, has evaluated the potential added value of transcriptomics to assess unintended side effects in a newly bred (genetically modified) plant variety. As one of the first initiatives in this area a small food safety-related tomato-array was developed with pathway-selected cDNAs on the basis of two subtractive cDNA libraries. This tomato array was used to hybridise mRNA derived from tomatoes in five subsequent ripening stages from green, via breaker, turning, and light red, to red, to obtain a background library of gene expression profiles of different ripening stages for future comparisons. At the same time these initial series of experiments were aimed to assess the potential of the approach with respect to its sensitivity and specificity. In addition the tomato array was used to hybridise mRNA derived from GM tomato transformant lines and the traditionally bred parent line and the results were analysed for the presence of differential expression patterns in both transformant lines, that have incorporated the same genetic construct, compared to the parent line. A similar study was performed in Arabidopsis to assess the extent of unintended effects in GM lines that have incorporated different numbers of the introduced genetic construct. The resulting data show that the methodology of transcriptomics has the potential to detect large as well as small differences in gene expression. The first was primarily shown in the comparative study on the developmental stage, the latter in the comparison of transcriptomics profiles of the two transformant lines vs the parent variety. It was also shown that, for direct comparison, plants to be sampled need to be grown under very similar conditions and the sampling needs to be performed in a structured way taking into account the developmental stage of the selected plant organs and/or tissues. All experiments illustrated the necessity to establish the bandwidth of natural variation for comparative purposes in order to determine the biological as well as toxicological and/or nutritional significance of differences detected in GM lines or in lines resulting from other breeding procedures. Finally, the results of the international debate on the assessment of complex (GMO-derived) plant products, the knowledge on current breeding strategies, and the results of the first publications on experiments that aim to detect unintended effects in plant breeding strategies using 'omics' technologies, including the experiments described here, are combined to review current approaches. A new overall approach for the safety evaluation of complex plant products, including GMO-derived products, is proposed. This approach applies currently available tools, including the 'omics' technologies, to assess food safety aspects of newly developed plant varieties already during the plant breeding process. Undesired effects of the breeding procedures for the plant's physiology, can thus be traced at an early stage and this will help to further guarantee the safety of the final

plant-derived food products. Observed differences in the final plant product will form part of the comparative safety assessment. It can be envisioned on the basis of the data presented in this thesis as well as in other studies described in the scientific literature that in general few differences will be observed between GM lines and the WT counterparts that fall outside of the bandwidth of natural variation of commercial counterparts. The toxicological and nutritional evaluation of the final plant products will need to focus on the limited number of differences that are outside of this bandwidth and may affect the product's food safety characteristics. To avoid inequalities in food safety assessment procedures that do not have a sound scientific basis, it is argued that these developments should have an impact on all novel plant varieties, not just on GMO-derived food plant products.

Table of contents

Introduction

Genetically modified food products, i.e. products derived from genetically modified organisms (GMOs), were firstly introduced onto the world market in 1994 in the United States. The first commercial product was the FLAVR SAVR tomato with delayed ripening characteristics. Already before this initial introduction the international debate had started on the most appropriate procedure to assess the safety of GMO-derived food and feed products. In the following years international consensus was reached as to the basic approach for GMO safety assessment with the application of the comparative safety assessment (CSA). This CSA entails a two-tiered procedure: 1) comparison of the new GMO variety with its nearest conventional counterpart on the market and 2) assessment of detected differences in terms of toxicological and nutritional relevance for the safety and nutritional status of consumers of the GMO-derived products. The basis for the CSA is an elaborate compositional analysis of known nutrients and anti-nutritional factors. At the same time it was recognized that this approach based on known constituents has its limitations in terms of detecting unintended side effects of the genetic modification. It was therefore proposed that unbiased methods such as the 'omics' technologies should be further explored for their applicability to detect unintended and undesired alterations in GM plant varieties.

This thesis aims to evaluate the practical aspects of the application of one of the 'omics' technologies, transcriptomics, i.e. the generation of gene expression profiles using the microarray technology, to detect unintended effects in GMOs, as well as to assess the potential added value of transcriptomics in the comparative safety assessment of a newly bred (genetically modified) plant variety.

Chapter 1 and 2 are introductory chapters and provide an overview of the state-of-the-art of the safety assessment procedures at the onset of the experiments.

Chapter 1 introduces the concept of the Comparative Safety Assessment (CSA) as the basic principle in the safety assessment of GMOs and derived products, representing a tiered approach that would be an alternative for the similar and often cited Principle of Substantial Equivalence that is frequently misinterpreted as the endpoint of the safety assessment.

Chapter 2 focuses on the safety assessment of GMO-derived animal products with reference to the developments in the safety assessment of GMO-derived plants and on the basis of the traditional components of the risk analysis: hazard identification, hazard characterization, exposure assessment and risk characterization.

Chapter 3, 4, 5 and 6 are the chapters describing the experiments performed to assess the potential of transcriptomics as a tool in the comparative safety assessment of new (GMO-derived) plant varieties.

In Chapter 3 the development of food safety-related tomato cDNA libraries by application of Representational Difference Analysis (RDA) subtractive mRNA protocols is described that form the basis for a focused microarray. The results of initial hybridization experiments are analysed to assess the quality of the newly developed array.

In Chapter 4 the newly developed tomato array is used to obtain transcriptomics profiles from tomatoes in the green to red stage of ripening, as well of three intermediate ripening stages. These profiles can be used in the evaluation of detected differences in (GM) tomato lines of interest: when differences are observed they can be assessed taking known shifts in gene expression profiles that are due to the stage of ripening into account.

Chapter 5 describes hybridisation experiments with two GM tomato lines, both high beta-carotene expressors, that have the same genetic construct incorporated into their genome in different locations. In this way it is feasible to characterize observed effects as direct effects of the incorporated genetic fragment, or rather as effects that are not likely to be directly related to the introduced gene sequence, but to other, unintended, side effects.

Chapter 6 describes similar experiments analysing gene expression profiles for a series of Arabidopsis transformant lines, with single or multiple introduction of the genetic construct, compared to the parent variety. The nature of the differentially expressed genes is discussed, also in the light of targeted analyses of compounds that may have been altered by the genetic modification.

Chapter 7 and 8/9 evaluate the obtained experimental data in the preceding chapters as well as the safety-related transcriptomics data as published in the scientific literature in the broader perspective of comparative safety assessment procedures.

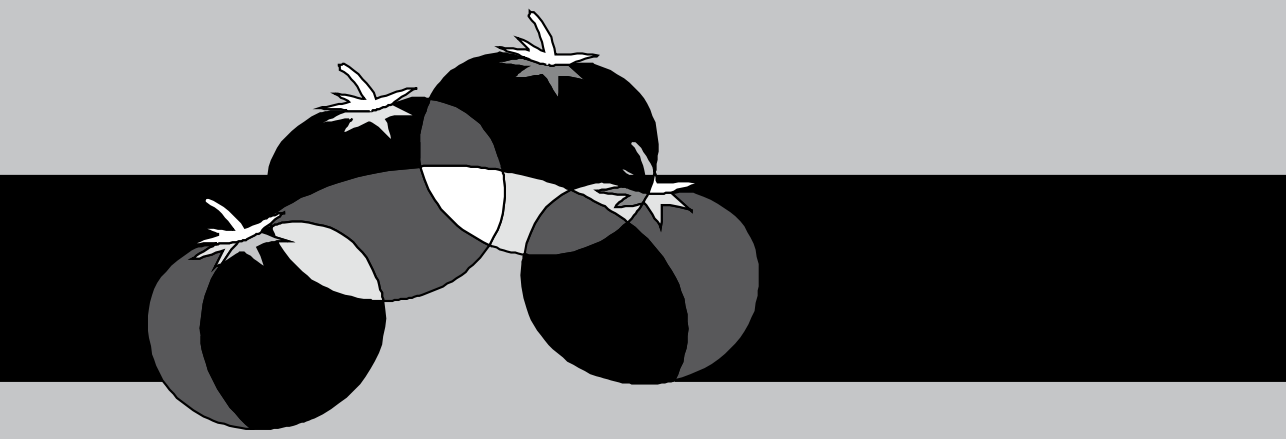
In Chapter 7 the assessment of plant-derived food products is discussed in the light of current developments in plant breeding, molecular biology, biochemistry and toxicology. It is argued that these developments (should) have an impact on all novel plant varieties, not just on GMO-derived food plant products.

In Chapter 8 (in English) and 9 (in Dutch) the main findings presented in the preceding chapters are summarized and discussed with respect to their implications for the applicability of transcriptomics in the food safety assessment of novel plant varieties, including GM plant varieties.

Chapter 1.

Abstract

Since the first discussions on strategies to assess the food safety of genetically modified (GM) crop plants, assessment of GM plants and derived tissues has been based on comparisons with their traditionally bred counterparts. This was termed the Principle of Substantial Equivalence. However, implementation of the principle led to controversy and hampered the precision of the actual safety assessment. Here, we propose the principle be rephrased into the Comparative Safety Assessment strategy. This describes the analytical nature of the first step of the entire (GM) food safety assessment in combination with consecutive toxicological and nutritional evaluations. Further development of advanced analytical methods will help to improve the efficacy of assessment strategies.



Comparative safety assessment for biotech crops

Esther J. Kok and Harry A. Kuiper

TRENDS in Biotechnology 2003, 21(10): 439-444.

1.1 Introduction

Genetically modified (GM) food crops were introduced commercially in 1994. The first commercial GM crop plant, which was introduced in the USA, was the FLAVR SAVR™ tomato that had delayed ripening characteristics. Since then, adoption of GM food crops has increased continuously, achieving a cultivated area of 58.7 million hectares worldwide [1] (and see Figure 1). Crops that are cultivated today carry foreign traits introduced by genetic modification that are predominantly of agronomic importance. The best-known examples include herbicide-resistant soybeans and insect-resistant maize, which have their own weed and insect control, respectively.

Currently, several cultivated GM crops have been modified with traits that affect the functional properties of the final product. For example, long-ripening tomatoes have favourable post-harvest texture characteristics for processing into tomato paste. Oilseed crops have a modified oil composition, including soybeans that are high in oleic acid (more stable during frying), and canola that is high in lauric acid (a desirable physical property). It is anticipated that, in the near future, more GM-crop-derived foods will have traits that are beneficial in food processing or that might positively influence the nutritional and health status of the crop for consumers and animals [2] (Table 1). Recently, GM crops have been designed – or are under development – to combat certain nutritional deficiencies. Well-cited examples include 'Golden Rice', in which provitamin A is introduced into the kernels [3], and iron-fortified GM rice [4]. The aim of these GM rice modifications is to alleviate vitamin A deficiency and/or anaemia in developing countries where rice is the staple crop. These modifications have been achieved through the insertion of genes encoding entire non-native metabolic pathways, or through targeted alterations in existing pathways.

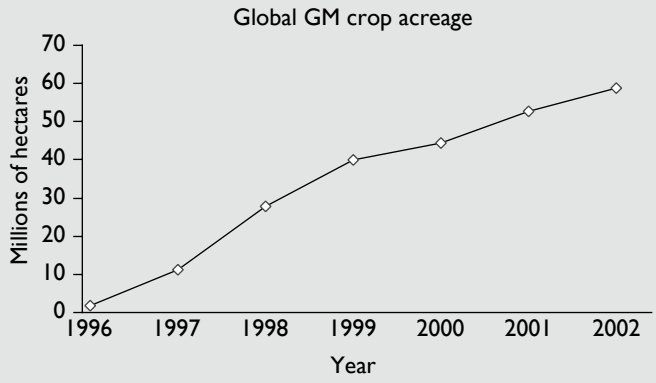


Figure 1. *Commercial genetically modified (GM) crop cultivation from 1996 until present [1].*

Table 1. *Experimental biotech food crops with potentials benefits for processors and consumers.*

Aim	Introduced trait	Crop	Refs
Nutritional			
Elevated levels of micronutrient	Synthesis of provitamin A from geranylgeranyldiphosphate (naturally present in kernels) by transgenic enzymes	Rice	3
	Iron-binding protein (ferritin) and two proteins for promotion of intestinal iron bioavailability (phytase, metallothionein)	Rice	4
Improved protein nutrition	Transgenic protein with favourable amino acid composition	Potato	35
Processing			
Bread baking	Transgenic glutenin protein associated with favourable dough characteristics	Tritordeum (cross between wheat and barley)	36
Less contamination of seed oil	Decreased synthesis of chlorophyll in seeds by antisense suppression	Canola	37
Improved starch degradation during malting	Transgenic amylase in kernels	Barley	38
Medical			
Edible vaccine	Polypeptides of heat-labile enterotoxin of *Escherichia coli* that raise immunity against diarrhoea-causing bacterial toxins	Maize	39

There is now a trend towards high-expression levels of foreign or endogenous proteins with an enhanced content of essential amino acids (e.g. high-lysine corn). Moreover, plants can be designed as 'protein factories' that serve as a medium for purification of a protein of interest, or to produce high levels of insecticidal proteins that decrease resistance development in insects. High protein expression levels have been achieved by plastid transformation; in one example, 45.3% of soluble leaf protein was transgenic [5].

1.2 Safety assessment strategies for GM-crop-derived foods

From the very first initiatives to establish globally agreed guidelines for the safety assessment of foods and food ingredients derived from GM organisms, comparison with the characteristics of relevant traditionally bred plant varieties was the leading principle [6]. The underlying assumption was – and still is – that traditional crop plant varieties currently on the market have not been elaborately tested in the laboratory before being marketed. However, because they have been consumed (after appropriate processing) for decades, they have gained a history of safe use. This history of safe use can be used as a baseline for the safety assessment of new GM plant varieties derived from established plant lines. The comparative concept for the safety evaluation of foods derived from GM crops has further been elaborated by the Organisation for Economic Cooperation and Development (OECD) and crystallised in the so-called Principle of Substantial Equivalence [7].

Food safety evaluation issues of foods derived from GM crops comprise:

- Molecular characterization of the introduced genetic fragment and resulting new proteins or metabolites (in addition, an increasing number of European member states routinely ask for characterization of the insertion point of the transgenic fragment);
- Analysis of the composition of the relevant plant parts with respect to key nutrients and anti-nutrients, including natural toxins and potential allergens;
- Potential for gene transfer of specific genes from the GM food to – particularly – microorganisms in the human and animal gastro-intestinal tract;
- Potential allergenicity of the new gene products, or alteration of the intrinsic allergenicity of the GM food organism;
- Estimated intake levels of the newly introduced proteins as well as of the final product, including any altered constituent;
- A toxicological and nutritional evaluation of the resulting data; and
- Additional toxicity testing (of the whole food) where necessary.

With regard to the last point, toxicity testing of the whole crop or derived plant products might be required. For example, cases where the composition of the whole crop has been changed significantly compared with the traditional counterpart, or where there is a need to further investigate potential unintended side effects of the genetic modification, warrant additional toxicity testing.

Specific guidance on these issues has been provided by: (l) the OECD [8], (2) the European Scientific Committee on Foodstuffs (SCF) [9], (3) the United Nations Food and Agriculture Organisation/World Health Organisation (FAO/WHO) [10-12], and (4) Codex [13]. A detailed overview of safety assessment practices relating to GM food crops has been published by Kuiper and colleagues [14]. A tiered approach for data generation and subsequent assessment is shown in Figure 2.

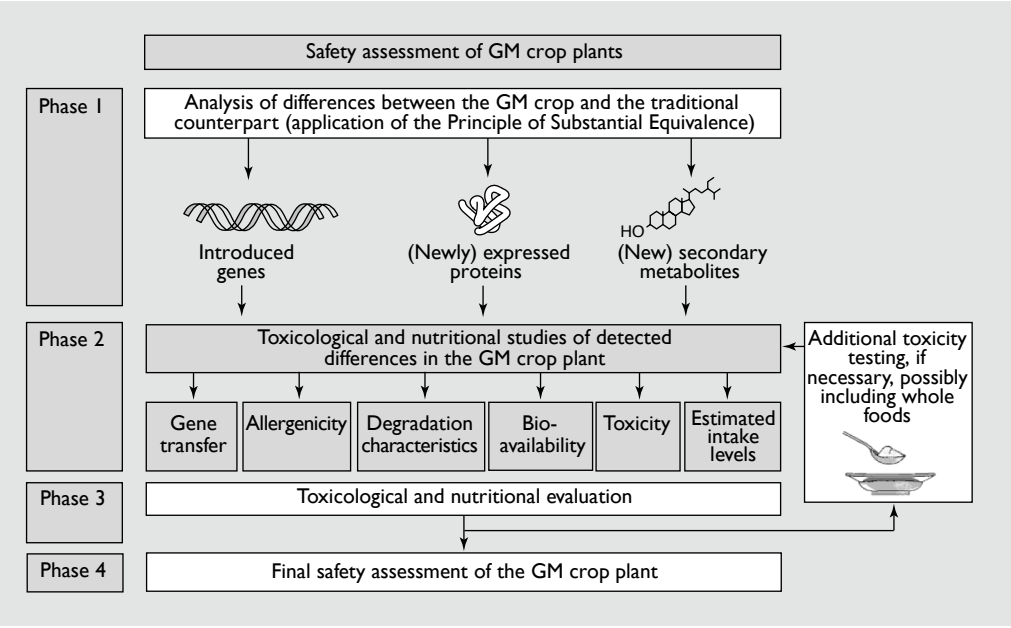


Figure 2. *Safety assessment strategies for genetically modified (GM)-crop-derived foods. Tiered approach for data generation and subsequent safety assessment of genetically modified (GM)-derived foods.*

1.3 Application of the Substantial Equivalence Principle needs to be improved

The approach of first comparing the GM line with the parent line and then with other traditionally bred varieties already on the market was predominantly formalized by the OECD [7]. The Principle of Substantial Equivalence was introduced with the aim of establishing a scientifically sound approach that would meet global acceptance. However, it soon became clear that the principle left much scope for individual (and national) interpretations. Further concerns established that the principle could only be applied on the basis of a thorough compositional analysis of the varieties under scrutiny (the GM line and its traditional counterpart). In addition, the compositional comparison is the starting-point of the food safety evaluation and not – as was misinterpreted in some publications [15,16] – an end-point in itself. Once differences in composition have been identified between the GM food plant and its appropriate comparator, targeted toxicological and nutritional studies should be carried out to assess the safety and nutritional impact on humans. Thus, toxicological and nutritional testing is an essential part of the safety assessment model for foods derived from GM crops. The Principle of Substantial Equivalence is merely a tool to identify potential differences and is part of a comprehensive comparative safety assessment approach. This issue was extensively discussed by the FAO/WHO Expert Consultation held in 2000 [11].

The OECD took up the challenge to formulate consensus documents on individual crop plants. This included an overview of the key macro- and micronutrients, as well as anti-nutritional factors, natural toxins and (where reported in the literature) their background values, for the different food crops [17]. This proved a difficult task because our knowledge, for example, of the levels and toxicity of anti-nutritional factors in crop plants is often fragmentary, especially in crops that are less economically important. Therefore, specific attention should be given to the quality and validation status of the analytical methods used to generate specific compositional data. In addition, the crop varieties and analytical methods used to generate the data might now be outdated, compared with present crops and methods. The ILSI Crop Composition Database, which has recently become available on the Internet, contains quality-controlled data and could be a valuable supplement to the OECD consensus documents.

Another complicating factor is the selection of plants to be analysed. The comparator of the GM line should preferably be the direct parent line. However, this line might no longer be available (e.g. it could be in possession of another breeding company). Furthermore, analysis of the plant line that will actually be marketed might reveal substantial changes to the parent line that are unrelated to the genetic modification. This is because, in general, a whole breeding programme separates the initial modification event from the breeding of the final genotype that will be marketed. Therefore, although comparison to several relevant lines is recommended, the data obtained might be less informative, and a proper analysis of these compositional data will be more complicated.

Environmental conditions also influence the physiology of the plant. It is therefore important that GM and non-GM plants to be analysed are grown under identical environmental conditions. In addition, it might be helpful to analyse plants grown under a range of environments and climates, which would influence the activity rate of individual metabolic pathways. However, it is unclear how much extra information could be obtained in this way; more unclear is how many environmental and climatological conditions should be assessed to improve significantly the food safety assessment of novel plant varieties by these extra analyses. All proposed conditions for the performance of field trials outlined in national and international guidelines thus far, are arbitrary and based on practical (breeding) experience with conventional crops, rather than on scientific evidence. Any extra information gained might therefore be limited.

1.4 Methods to detect and assess unintended effects of a genetic modification

Concerns that unintended and unexpected side effects might occur in GM organisms (GMOs) as result of the genetic modification process, thereby impacting on human and animal health, has attracted attention from both scientific and public groups. However, the potential occurrence of side effects in non-GM organisms must also be highlighted [14]. Compositional analyses of the GM plant and its traditional counterpart, in addition to the notion that relevant unintended side effects might remain undetected when analysing only specific compounds or intermediates

in important nutritional and anti-nutritional pathways, are complicated issues. It was therefore encouraged that more general, unbiased methods of analysis be developed to detect relevant changes in a much larger part of the physiology of the plant [8,11,18]. This could be of particular importance for GM plants that have multiple genes inserted, which possibly have a higher occurrence of unexpected and unintended effects (Table 1). As a result, specific projects were initiated to develop more informative, unbiased methods for different levels of integration of the physiology of the plant on mRNA, protein and metabolite levels.

The European Thematic Network, Entransfood, covers most of the current initiatives to develop new approaches for the food safety assessment of GM varieties (http://www.entransfood.com). The network serves as an umbrella project for five research groups and five working groups (Figure 3). Three of these are directly related to the food safety assessment of genetically modified organisms (GMOs). First, GMOCARE focuses on the development of new tools based on the unbiased analysis of the relevant plant tissues using fingerprinting techniques in the fields of genomics, transcriptomics, proteomics, metabolomics and glycomics. Second, SAFOTEST focuses on the development of new toxicological approaches to assess the safety of consumption of novel food products. Third, GMOBILITY investigates the possibility of gene

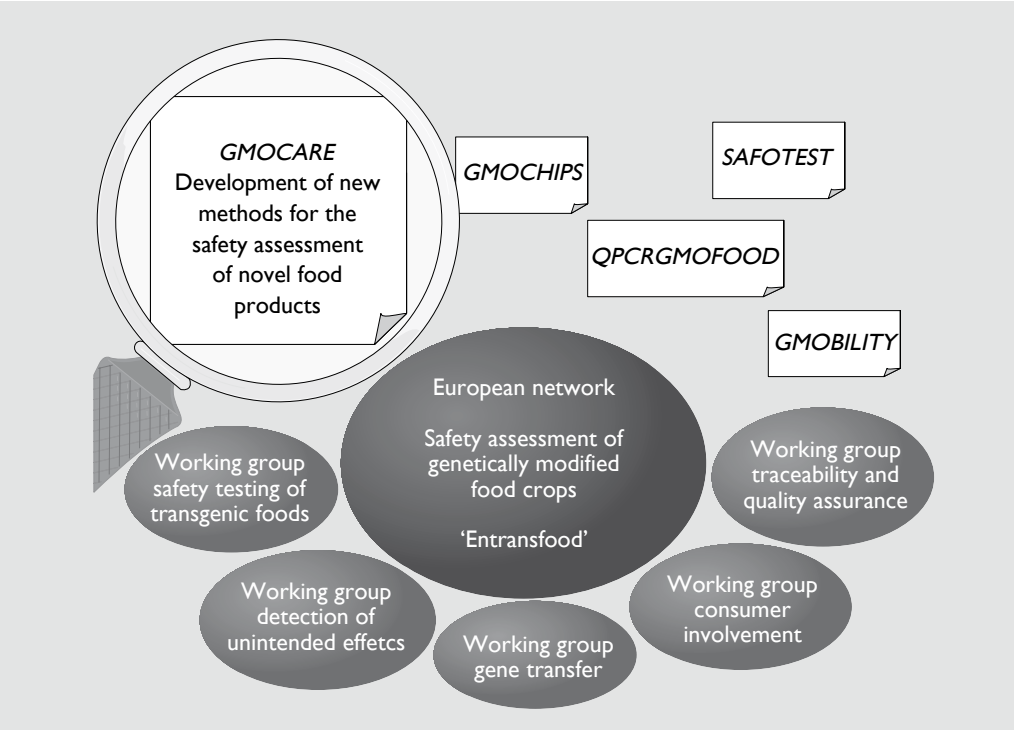


Figure 3. *Organization of Entransfood, the European Thematic Network on the Safety Assessment of Genetically Modified Food Crops.*

transfer in the human gastrointestinal tract using model systems. The remaining two projects relate to the detection, identification and quantification of GMOs in the food production chain. QPCRGMOFOOD focuses on the development of identification and quantification methods in the food production chain. GMOCHIPS aims to develop a chip-based approach for the screening of large numbers of GMO varieties in a single assay. Besides Entransfood, there is another important British initiative: the G02 Programme on the Safety of Novel Foods. This began in 2001 and includes projects that aim to investigate the potential of new methods for the safety assessment of novel food products (http://foodstandards.gov.uk/science/research/NoveIFoodsResearch/g02programme).

Unbiased fingerprinting approaches at the level of DNA, gene expression, proteins, metabolites and their secondary structures, could potentially provide a more thorough insight into any unpredicted changes in the physiology of the plant that might go undetected when focusing on single compounds [19]. For example, it is possible to introduce entirely new metabolic pathways, without any obvious phenotypic change [20]. Nevertheless, it is unlikely that new metabolic pathways do not affect existing pathways. Fingerprinting techniques might be a more efficient method of identifying such alterations. However, significant research will be required before these techniques produce confirmed and validated information. Fundamental setbacks that need to be addressed before these techniques can be included in a routine, integrated evaluation protocol are outlined in the following sections.

1.4.1 DNA level

Owing to the large sequencing projects of recent years, sequencing of large DNA stretches is now routine. Sequence analysis of the insertion point of the genetic fragment might be significant to evaluate whether it is possible to identify any potential side-effects, for example, based on the interruption of regulatory or gene sequences, or the presence of any such sequence in the vicinity. However, there is still limited knowledge of the genetic code of the organisms under investigation [21,22]. Additional knowledge, especially for regulatory elements, is crucial for the correct interpretation of DNA sequencing results.

1.4.2 Gene expression level

Microarrays enable altered gene expression to be screened in large numbers of genes simultaneously. However, correct interpretation of the resulting data is both difficult and dependent on many different factors. These include experimental set-up, available equipment, software, and knowledge of the organism under investigation [23,24].

1.4.3 Protein level

Given that altered gene expression levels might not correlate directly to shifts in protein levels [25], the most direct method of investigating unpredicted alterations is proteomic analysis

of the tissues of interest. Considerable expertise in 2D gel electrophoresis has enabled the simultaneous screening of large numbers of proteins, with subsequent characterization by mass spectrometry (MS) [26]. However, there are several important setbacks. Setting up an informative system for a single tissue is time-consuming. Furthermore, reliable quantification remains problematic, despite the availability of advanced software. The sensitivity of such an approach is affected by slight changes in isolation conditions, which, in turn, might profoundly affect the behavior of the proteins under investigation. Protein micro-arrays can theoretically expand more easily on the basis of increasing knowledge of the proteome. This could reduce the time-consuming setup of new protein analysis systems, and increase reproducibility and potential for quantification. Current issues relating to array production and assay performance still need to be addressed [27,28].

1.4.4 Metabolite level

Another direct approach is analysis of the secondary metabolites. Informative systems have been set-up for different organisms using gas and liquid chromatography *(GC/LC)* in combination with MS [29,30] or nuclear magnetic resonance (NMR) [31,32]. In theory, identification of large numbers of constituting compounds can be achieved using a combination of these techniques. However, in practice, there are several important drawbacks. These include a lack of reliable data on profile variation for relevant compounds in different matrices of the organism under study, and standardization of extraction procedures and measurement protocols.

Despite the technical hurdles, it is clear that these new developments have the potential to give increased insights into relevant changes in the physiology of plant products resulting from genetic modification or from the application of new and existing food processing techniques.

1.5 Concluding remarks

Although the Principle of Substantial Equivalence has received comments from all types of stakeholders (producers, regulators, consumers, evaluators, etc.) [15,33], the basic idea behind the principle remains untouched. When evaluating a new or GM crop variety, comparison with available data on the nearest comparator, as well as with similar varieties on the market, should form the initial part of the assessment procedure. The term 'substantial' has provoked interesting discussions, but has also led to misinterpretations. Therefore, the principle should be rephrased as the 'Comparative Safety Assessment (CSA)' approach. This phrase better outlines the comparative nature of the assessment, while avoiding the idea that it is a safety assessment in itself. Nutritional and toxicological assessment should be performed on the basis of the CSA, and might require additional safety tests.

Even where the idea of acceptable safety of conventional foods has gained worldwide approval, underlying assumptions of relative safety can still be questioned. For example, traditional plant breeding practices such as chemical mutagenesis might lead to a higher rate of

mutations compared with genetic changes induced by recombinant DNA technology [34]. Only in exceptional cases will a safety assessment of the resulting plant-lines be demanded. It is debatable whether the results of such generally accepted breeding practices should serve as the baseline for the safety assessment of new or GM plant-lines.

Perhaps it is time to rethink our philosophy on the safety of foods produced by different agricultural methods. This would result in a more-balanced universal risk analysis system and basic safety assessment protocol for all novel food crop varieties. In all cases, a CSA of available data on crop plant varieties with a history of safe use should serve as a starting-point for the consumer safety assessment. A system of reliable databases and informative profiles on individual compounds will provide significant progress towards a safe food supply, even as the concept of third-generation GMOs becomes reality.

Acknowledgements

We thank Gijs A. Kleter for his contribution. We also acknowledge the Dutch Ministry of Agriculture (LNV 390), the British Food Standards Agency, and the European Commission (Entransfood and associated RTD projects) for financial support to the projects related to the safety of novel food products highlighted in this publication.

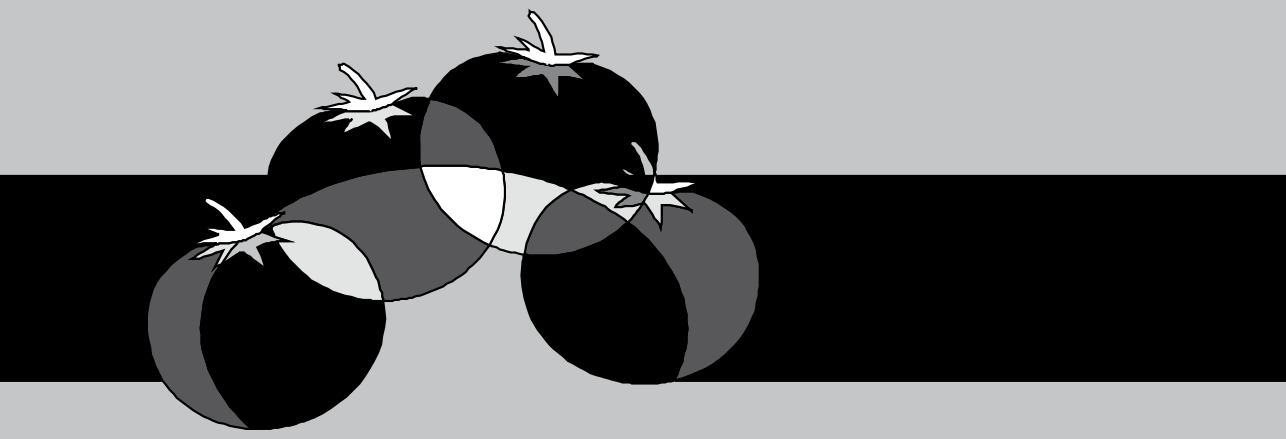
References

1. James, C. (2002) *Global Status of Commercialized Transgenic Crops: 2002,* ISAAA Briefs No. 27: Preview, International Service for the Acquisition of Agric-biotech Applications, Ithaca.
2. Kleter, G.A. *et al.* (2000) New developments in crop plant biotechnology and their possible implications for food product safety. *RIKILT Report* 2000.004, RIKILT, Wageningen (http://www.rikilt.dlo.nl/Publications/Publications/Tekstrapport2000%20004.htm).
3. Ye, X. *et al.* (2000) Engineering the provitamin A (ß-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287,303-305.
4. Lucca, P. *et al.* (2001) Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. *Theor. Appl. Genet.* 102, 392-397.
5. De Cosa, B. *et al.* (2001) Over expression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* 19,71-74.
6. IFBC International Food Biotechnology Council, (1990) Biotechnologies and food: assuring the safety of foods produced by genetic modification. *Regul. Toxicol. Pharmacol.* 12, Sl-S196.
7. OECD, (1993) *Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles,* Organisation for Economic Co-operation and Development, Paris, http://www.oecd.org/pdf/M00034000/M00034525.pdf.
8. OECD, (1996) *Food Safety Evaluation,* Organization for Economic Cooperation and Development, Paris.

9. EU *97/618/EC,* (1997) Commission Recommendations 29 July 1997 concerning the scientific aspects and the presentation of information necessary to support applications for the placing on the market of novel foods and novel food ingredients and the preparation of initial assessment reports under Regulation EC *259/97* of the European Parliament and of the Council. *Off. J. Eur. Commun.* L253, 1 – 36. http://europa.eu.int/eur-lex/en/lif/dat/1997/*en_397X0618.h*tml.
10. FAO/WHO, (1996) *Biotechnology and Food Safety. Report of a Joint FAO/WHO consultation, Rome, Italy,* 1996. *FAO Food and Nutrition Paper* 61, Food and Agriculture Organisation of the United Nations, Rome, ftp://ftp.fao.org/es/esn/food/biotechnology.pdf.
11. FAO/WHO, (2000) *Safety Aspects of Genetically Modified Foods of Plant Origin. Report of* a *Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland, 2000,* Food and Agriculture Organisation of the United Nations, Rome, ftp://ftp.fao.org/es/esn/food/gmreport.pdf.
12. FAO/WHO, (2001) *Allergenicity of Genetically Modified Foods. Report of* a *Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology. Rome, 2001,* Food and Agriculture Organisation of the United Nations, Rome, http://www.who.int/fsf/Documents/Biotech_Consult_Jan2001/report20.pdf.
13. FAO/WHO (2002) *Report of the Third Session of the Codex Ad Hoc Intergovernmental Task Force on Foods Derived from Biotechnology (ALINORM 01/34). Codex* Ad Hoc *Intergovernmental Task Force on Foods Derived from Biotechnology,* Food and Agriculture Organisation of the United Nations, Rome (ftp://ftp.fao.org/codex/alinorm03/Al03_34e.pdf).
14. Kuiper, H.A. *et al.* (2001) Assessment of the food safety issues related to genetically modified foods. *Plant J.* 27, 503-528.
15. Millstone, K *et al.* (1999) Beyond 'substantial equivalence'. *Nature* 401, 525-526.
16. UBA (2001) *Evaluating Substantial Equivalence. A Step Towards Improving the Risk/Safety Evaluation of GMOs.* Vienna, 19-20 October, 2001. Conference Papers Vol. 22, Austrian Federal Environment Agency (UBA), Vienna.
17. OECD, (2003) *Consensus Documents for the Work on the Safety of Novel Foods and Feed. Task Force for the Safety of Novel Foods and Feed,* Organisation for Economic Co-operation and Development, Paris, http://www.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-document-530-nodirectorate-no-27-24778-32,00.html.
18. EC Scientific Steering Committee, (2000) *Risk Assessment in* a *Rapidly Evolving Field: the Case of Genetically Modified Plants,* Scientific Steering Committee, European Commission, Brussels, http://europa.eu.int/comm/food/fs/sc/out148 _en.pdf.
19. Kuiper, H.A *et al.* (2003) Exploitation of molecular profiling techniques for GM food safety assessment. *Curro Opin. Biotechnol.* 14,238-243.
20. Tattersall, D.E. *et al.* (2001) Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science* 293, 1826-1828.
21. Bonifer, C. (2000) Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? *Trends Genet. 16,* 310-315.
22. Bell, A.C. *et al.* (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 291, 447-450.
23. Franssen van Hal, N.L.W. *et al.* (2001) Factors influencing cDNA micro array hybridization on silylated glass slides. *Anal. Biochem. 308,* 5-17.
24. Aharoni, A. and Vorst, O. (2001) DNA microarrays for functional plant genomics. *Plant Mol. Biol.* 48, 99-118.

25. Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol. 17,* 994-999.
26. Machuka, J. and Okeola, O.G. (2000) One- and two-dimensional gel electrophoresis identification of African yam bean seed proteins. *J. Agric. Food Chem.* 48,2296-2299.
27. MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289,1760-1763.
28. Templin, M.F. *et al.* (2002) Protein microarray technology. *Trends Biotechnol.* 20, 160-166.
29. Roessner, U. *et al.* (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J. 23,* 131-142.
30. Fiehn, O. *et al.* (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18,1157-1161.
31. Noteborn, H.P.J.M. *et al.* (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J. Biotechnol.* 77, 103-114.
32. Lommen, A *et al.* (1998) On the detection of environmental effects on complex matrices combining off-line liquid chromatography and H-NMR. *Biodegradation* 9, 513-525.
33. Levidow, L. and Murphy, J. (2002) *The decline of substantial equivalence: how civil society demoted* a *risky concept Proceedings of the Science and Citizenship in* a *Global Context: Challenges from New Technologies',* Institute of Development Studies, University of Sussex, Brighton, (http://technology.open.ac.uk/cts/bpg.htm).
34. Van Harten, AM. (1998) *Mutation Breeding: Theory and Practical Applications,* Cambridge University Press.
35. Chakraborty, S. *et al.* (2000) Increased nutritive value of transgenic potato by expressing a non allergenic seed albumin gene from *Amaranthus hypochondriacus. Proc. Natl. Acad. Sci. U. S. A. 97,* 3724-3729.
36. Barro, F. *et al.* (2003) Functional properties and agronomic performance of transgenic tritordeum expressing high molecular weight glutenin subunit genes 1AxI and 1Dx5. *J. Cereal Sci.* 37, 65-70.
37. Tsang, E.W.T. *et al.* (2002) Chlorophyll reduction in the seed of *Brassica napus* with a glutamate 1-semialdehyde aminotransferase antisense gene. *Plant Mol. Biol.* 51, 191-201.
38. Tull, D. *et al.* (2003) Enhanced amylolytic activity in germinating barley through synthesis of a bacterial α-amylase. *J. Cereal Sci.* 37, 71-80.
39. Chikwamba, R. *et al.* (2002) A functional antigen in a practical crop: LT-B producing maize protects mice against *Escherichia coli* heat-labile enterotoxin (LT) and cholera-toxin (CT). *Transgenic Res. 11,* 479-493.

Chapter 2.



The food safety risk assessment of GM animals

Esther J. Kok[1] and Wendelyn Jones[2]

[1] RIKILT Institute of Food Safety, Bornsesteeg 45, 6706 AL Wageningen, the Netherlands,
[2] Rockville, Maryland, USA

Working paper for the FAO/WHO Expert Consultation on GM animals, Rome, 17-21 November 2003. http://www.who.int/foodsafety/biotech/meetings/en/gmanimal_reportnov03_en.pdf; http://www.fao.org/ag/agn/food/risk_biotech_animal_en.stm; ftp://ftp.fao.org/es/esn/food/GMtopic4.pdf

2.1 Introduction

2.1.1 General

Over 98% of all genetically modified organisms (GMOs) that have been introduced into the environment worldwide are genetically modified plants. Micro-organisms and animals constitute only a very limited portion of all GMOs that have been 'field'-tested so far. This situation may, however, change as there are already a number of examples of genetically modified (GM) animals. Of these, GM fish varieties are perceived as closest to marketing. So far no GM animals have been introduced into either the US or EU market, but depending on regulatory developments and public opinion the number of different GM animals bred and marketed world wide may increase.

Two types of GM animals can be distinguished. The first group has been genetically modified to enhance overall performance and have, therefore, added value in an agronomic and/or economic sense. The whole animal will eventually be available for the food market. The second group of GM animals has been transformed to produce specific substances in milk, eggs or blood or serve as medical research model. The goal of this later GM technology is the production and isolation of the specific substance or tissue as a marketing material or to use the animal for medical research purposes or for toxicity testing; they are not intended for food production. However, animal disposition of these GM animals is still a concern. In addition, a distinction should be made between germ-line modified animals and somatic cell -modified animals. In the latter case only specific tissues will have incorporated the new trait(s), whereas the rest of the animal will remain genetically unaltered.

When discussing the food safety aspects of GM animals two scenarios should be considered: 1) the intentional introduction of GM animals into the food production chain and 2) the unintentional introduction of GM animals. Although in the latter case precautions will be taken to avoid GM animal materials entering the food production chain, such an unintentional event should nevertheless be part of the risk evaluation process.

2.1.2 Regulatory aspects

In recent years considerable expertise has been gained in the food safety assessment of GM plants. Although work is on-going to optimise current safety assessment approaches, it can be stated that solid strategies have already been developed that minimize the possibility of adverse health effects for the consumer of GMO-derived plant products. In fact, no adverse effects have been observed that can be related to an approved GM plant variety. These food safety assessment strategies for crop plant products are based on a series of expert reports initiated by the IFBC [1], that was taken over and carried further by OECD [2,3] and FAO/WHO [4,5,6] in order to come to globally agreed safety assessment approaches [7]. Current safety assessment strategies address a number of different issues related directly to the genetic

modification as well as to potential unintended side effects of the genetic modification on the food organism. In practice, information is asked on 1) the process of the genetic modification, 2) the safety of the newly introduced proteins, including information on potential allergenicity, 3) occurrence and potential implications of unintended side effects of the genetic modification, 4) possible effects of gene transfer and recombination, 5) the role of the new food in the diet and 6) the influence of food processing [8]. Based on this dossier additional questions of food safety may be asked.

The safety assessment of GM animal-derived food materials has also been subject to discussion in a number of expert meetings by OECD [2,3] and FAO/WHO [4,5,6] and the US National Research Council [9]. It was concluded that the minimal food safety assessment of GM animals should comprise 1) the molecular characterisation of the inserted foreign DNA, 2) the safety assessment of the introduced genes and their products, 3) any unintended effects of the insertion of foreign DNA in the organism, and 4) the effect of disease resistance brought about in transgenic food animals on consumer's exposure to disease-causing agents [10]. Depending on the method of gene transfer used, additional questions of food safety will have to be answered with relation to the infectivity of the vector, the assessment of potential effects of vector regulatory elements on the host cell and the possibility of recombination with endogenous viral sequences.

Experience in the safety assessment of GM animals is still very limited, but it is clear that evaluators of GM animal-derived food products will benefit from experiences with GM plants as basic approaches for GM plant materials also apply to GM food animals. This working paper will draw upon the experience in the food safety assessment of GM plants to discuss the food safety assessment of GM animal -derived food products.

2.1.3 Risk assessment approach and limitations

The traditional paradigm for conducting a risk assessment includes four steps: hazard identification, hazard characterization, exposure assessment and risk characterization [11]. With GM animals and derived foods all four steps of the risk assessment must be undertaken on a case-by-case basis within this still evolving area. Thus the authors propose a comparative safety assessment to enable the final risk characterization. Drawing from the experiences with GM plants, the safety assessment is often a two-tiered approach where initially information is gathered to identify parameters and the magnitude of these parameters that distinguish the GMO from its traditional counterpart. The next phase is then to gain further insight into the toxicological and nutritional relevance of the detected differences. These characterization steps, when matrixed with the exposure of the hazard, will allow for the final risk assessment.

2.2. GM animals

2.2.1 General

There are two types of GM animals: those whose alterations are stably incorporated throughout their genomes and those with non-heritable transgenic constructs. The former are often referred to as 'transgenic animals' while the latter techniques are often referred to as 'gene therapy'. Gene therapy modifications are not limited to modifications intended only to therapeutically treat animals. In fact, the distinction between heritable and non-heritable modifications is not dependent on the intent of the modification. Rather, it is a function of the technology chosen for the intended modification.

2.2.2 Methodologies used for gene transfer

2.2.2.1 Non-heritable modifications

Animals containing non-heritable changes are produced by the introduction of the gene of interest in a vector that targets the somatic cells of the animal. There are two types of vectors preferentially used: those based on viral sequences and those based on transposable elements.

Viral-based technologies take advantage of the integrative properties of retroviruses and adenoviruses. The integrative function is the ability of viruses to 'cut in' to the sequence of host DNA. Such interruptions may be benign or hazardous. Transposon-based technologies have also been developed. Transposons are often referred to as 'jumping genes' because of their ability to catalyze their own movement within the genome of the animal. Transposons were first discovered in the plant kingdom, but have recently been identified in animals, including humans.

Any gene therapy technique may give rise to insertional mutagenesis or unintended gene activation or silencing. The risk scenario for both viral and transposon-based vectors also includes concern for recombining with existing viruses in the intended target animals, in humans who are exposed to them, or in other animals that may be exposed to the target animals or their wastes. Recombination may give rise to viruses with increased host ranges (swine viruses becoming capable of infecting humans), increased virulence (innocuous viruses causing serious illness), or generation of entirely new, pathogenic viruses.

2.2.2.2 Heritable modifications

GM animals are produced as the result of the stable incorporation of genetic constructs in their nuclear chromosomes or mitochondrial genomes. In general, transgenic animals are produced by injecting early embryos with solutions of DNA that contain constructs that have all of the requisite information for directing the expression of the gene(s) of interest, but rely on the

cell's internal recombinatory enzymes for integration. Scientists have also used viral vectors or transposon-based vectors to produce transgenic animals with heritable traits.

Production of a transgenic line of animals is usually a two-step process. Mosaic transgenic animals are produced by the introduction of the transgenic construct into early stage embryos. The expectation is that most of the cells of that developing embryo will contain the gene of interest, including some germ cells. These animals are considered 'mosaics' as they are composed of two or more genetically distinct cells. Mosaics are then bred and the offspring tested to find animals with 100% transgenic cells (i.e., derived from a transgenic germ cell in the mosaic). A founder animal, in which all cells carry the transgene, is selected and bred to propagate the transgenic line.

2.2.3 Genetically modified animals and their products

2.2.3.1 Laboratory models

GM animals are now common tools used to investigate the mechanisms of both normal physiology and the pathophysiology of humans and animals. An example is the pig model for human retinitis pigmentosa, a progressive disease that begins with night blindness. This model is intended to help develop pharmaceutical strategies to slow the onset and progression of the disease.

2.2.3.2 Biopharm modification in food-animals

2.2.3.2a Human therapeutic agents

GM animals can be developed as bioreactors for the production of therapeutic proteins. In general, these protein products will be produced in the animal's milk (cows, sheep, and goats), eggs (chickens), semen (swine), or blood (large farm species). The advantages of producing these products in animals rather than cell or tissue cultures include high yields, mammalian glycosylation pattern and lower post-development costs [12]**.** Examples of therapeutic products from GM animals include alpha-1-antitrypsin (ATT) in goat milk. This human blood protein is intended to treat hereditary emphysema (ATT deficiency), cystic fibrosis, and chronic obstructive pulmonary disease [13]. Other examples include antibody production in GM animals for diagnostic and medicinal purposes from milk or blood [14].

2.2.3.2b Xenotransplantation

The field of xenotransplantation covers many procedures, ranging from implantation of single cells to treat Parkinson's disease to the transplantation of organs to treat organ failure. GM animal organ transplantation has yet to be successfully implemented in humans, although transplants of smaller tissues and individual cells are currently under active clinical investigation. Because of their physiological similarities to humans, pigs are attractive as a

potential organ donor species. Because tissue rejection appears to be the primary medical barrier, pigs have been modified to knock-out 1, 3-galactosyl transferase, a protein linked to acute human tissue rejection.

2.2.3.2c Industrial products

The use of GM animals in the production of industrial products provides a novel 'manufacturing' source, and a number of challenges to manufacturers, regulators, and the public. Perhaps the best known example is transgenic goats producing spider silk proteins in their milk. These proteins could be used in the manufacture of body armor. The larger part of this category of transgenic animals will be kept in containment and it is essential that they should not enter the food production chain. Nevertheless the unintended entry into the food supply chain should be part of the risk assessment procedure prior to the breeding of these GM animals.

2.2.3.3 Agronomic modification in food animals

2.2.3.3a Animal health and productivity

The most well-known products in this category are those incorporating growth hormone (somatotropin) genes into the genomes of the same or other species. Aquaculture provides several good examples. The main traits to be altered are growth rate, cold tolerance, disease resistance and sterility. Transgenic salmon, catfish, carp and tilapia have been developed to reach market weight sooner than their non-transgenic counterparts by using fish-derived somatotropin. However, earlier research involved somatotropins from other sources. The promoters used can be either tissue-specific or constitutive [15]. For cold tolerance antifreeze proteins, such as winter flounder-derived delta-9-desaturase, have been tested, but have not yet proved successful [15].

Disease resistance in animals can also be enhanced using GM techniques. Lysostaphin, a bacteriocidal enzyme, has been introduced into cows to decrease the incidence of mastitis caused by *Staphylococcus aureus*. Moth cecropin, a broad spectrum antimicrobial peptide, has been transgenically incorporated into catfish to decrease their susceptibility towards a broad range of bacterial diseases [16].

2.2.3.3b Enhanced animal nutrition

Enhanced animal nutrition and growth performance by modification is possible. For example, bovine lactalbumin and insulin-like growth factor-1 (IGF-1) have both been introduced into sow milk for the improvement of the growth characteristics of the piglets [17]. Attempts in fish are ongoing to alter the carbohydrate metabolism of especially salmonoids in order to be able to use vegetable products in the aquacultural systems [15]. The 'Enviropig' is another example of GM that affects the nutrition of the pig. In this specific case, phytase is introduced

into pigs to allow them to make better use of the phosphorus in their feed. This not only allows the farmer to decrease phosphate supplements, but also decreases the amount of phosphorus in pig manure [18].

2.2.3.3c Human foods

Foods derived from GM animals can be altered with respect to functionality and composition. For example, cows can be modified to make a more desirable milk: (1) producing milk more digestible for lactose intolerant individuals by lowering its lactose content, or (2) increasing the amount of a naturally occurring antimicrobial enzyme to increase the shelf life of milk. Althought the meat industry also has increasing interest in the improvement of the sensory and nutritional quality of their meat products [19], few GM experiments are currently performed in this area as yet.

Fish can also be modified to provide better, more nutritious food. One example is the transgenic modification of rainbow trout to increase the amount of the omega-3 fatty acid that they produce and store.

2.3. Comparative safety assessment

2.3.1 Principle of substantial equivalence, applied

The principle of substantial equivalence was originally described by the FAO/WHO [4], and subsequently named and detailed by the OECD [2]. The rationale behind the principle is that many food products we eat today are derived from organisms that we can not consider inherently safe. Nevertheless, we have been consuming these products for decades without any obvious deleterious effects. Because of this history of safe use, it is generally acknowledged that traditional food products should serve as a baseline for comparison and that novel GMO-derived food products should be at least as safe as the traditional products that they may replace in the diet. The principle has lead to much debate in recent years as interpretation of the principle differed between countries. Nevertheless, the basic idea of comparing new GMO-derived products with closely related traditional counterparts to assess the safety of the newly developed organisms is unchallenged. Substantial equivalence should represent a starting point of the assessment rather than the end point [20] and should not be confused with being an absolute safety standard.

Application of the principle is usually a tiered approach, a Comparative Safety Assessment (CSA) [21] where the initial step is comprised of a thorough comparison with the closely related traditional counterpart. This comparison includes both phenotypic characteristics as well as a compositional analysis. The phenotypic analysis should also include factors such as disease resistance to common diseases. Information should be supplied on:

- the transformation process of the genetic modification, including the sequence of the inserted material;
- the copy number and place(s) of insertion;
- stability of the integration;
- the safety of any newly introduced proteins, including allergenicity;
- occurrence and implications of unintended effects;
- potential effects of gene recombination;
- the role of the new GM animal food in the diet; and
- the influence of processing on the new GM food product.

Within Europe, sequence analysis of the place(s) of insertion is also part of initial phase of the CSA. More precise criteria for the molecular characterisation are currently being discussed in the OECD.

2.3.2 Hazard identification and characterization

The hazard identification step is typically the first step in any risk assessment. However, for complex GMO-derived foods, the hazard identification step will not be as readily completed as in the case of well-characterised single chemical compounds. Similarly, the hazard characterization is not as readily determined with complex GMO-derived foods. The variety and magnitude of unintended effects when testing complex food products, whether GMO-derived or not, may preclude straightforward hazard identification and characterization. The differences found as a result of the CSA serve as comparable to the hazard identification and hazard characterization steps in a traditional risk assessment paradigm.

2.3.3 Gene transfer

The DNA construct used to change the genetic make-up of the animal should be considered within an assessment especially if the gene or its promoter is derived from a viral source since horizontal transfer or recombination may occur. Additionally, bacterial host-derived materials may contain additional sequence fragments unrelated to the target gene. Because such fragments can be heterogenous in size and sequence, they are difficult to detect. This is particularly a problem with retroviral vectors. Host cells often contain large numbers of endogenous viruses and virus-like sequences [11,22,23]. Inadvertant introduction of such sequences into the germline of a GMO not only has the potential for creating unintended genetic damage but can also contribute by recombination to the generation of novel infectious viruses. A well known example is the generation of a replication-competent murine leukemia virus (MLV) during the growth of a vector containing a globin gene [24]. In a similar way prions may be introduced to the GM animal or derived products [25].

Furthermore, there is potential for horizontal transfer of the gene construct: food-ingested foreign DNA may not be completely degraded in the gastrointestinal tract of mice [26,27]. It

was shown that phage M13 mp18 DNA following oral ingestion by mice may reach peripheral leukocytes, the spleen and liver via the intestinal wall mucosa and was covalently linked to mouse DNA [26]. Similar results were obtained when a plasmid containing the gene for the green fluorescent protein was fed to mice [27]. However, these results have been criticized due to the complication of artifacts within the methodology [28]. For the food safety assessment it is prudent to assume that DNA fragments may survive the human gastrointestinal tract and be absorbed by either the gut microflora or somatic cells lining the intestinal tract.

Commonly used marker genes are genes that code for antibiotic resistance. Risk assessment of these selectable genes should focus on gene transfer to microorganisms residing in the gastrointestinal tract of humans or animals. There is general agreement that transfer of antibiotic resistance genes from plants to human gut micro flora is unlikely to occur and impact antibiotic efficacy [5,6,29]. Similarly, the likelihood of such transfer from GM animals to human gut microflora will also be low. However, as the potential of gene transfer can not be completely ruled out, the safety assessment should also consider information on the role of the antibiotic in human and veterinary medical uses. Furthermore, within the EU the use of antibiotic resistance marker genes in newly developed GMO-derived food products is not allowed.

2.3.4 Safety of the gene product

The safety of the gene product must be assessed on a case-by-case basis. Depending on the knowledge on the expressed product the assessment may range from a limited evaluation process of the available data on the protein, such as amino acid sequence and expression rates in different tissues, to, in the case of less well-documented proteins, extensive toxicity testing including animal studies. In theory, the advent of GM animals may lead to the introduction of many new proteins without a history of safe use into the human diet. The assessment of the novel proteins should be based on current knowledge of toxic substances, including a search for sequence homology with known toxins, and the function of the novel protein. In the case of unknown proteins a full classic toxicological safety assessment procedure will form part of the evaluation.

In this respect a distinction should be made between GM animal-derived food products that were developed, to improve agronomic characteristics and GM animal-derived food products developed for veterinary, pharmaceutical or industrial purposes. So far the number of different genes that is used for the production of GM food animals is still rather limited when compared with plants, but this situation may change with the progress of genome sequencing programmes that are likely to provide a wealth of data on important animal physiological pathways.

2.3.5 Allergenicity

Food whether developed by conventional means or through biotechnology is a potential source of allergens. All food allergies are mediated by antigen-specific IgE and are characteristic of

type-I reactions. In the case of new proteins being expressed in the GM animal, the allergenic potential of the protein will need to be established. In the case of production of specific well-characterised (medicinal) proteins by the GM animal, it needs to be established whether the post-translational modifications are comparable to the same substances being produced by more traditional sources in order to assess potential altered toxicological or allergenic properties of the newly synthesized proteins [30].

Efforts to characterize the mechanisms of allergies at both cellular and molecular levels, have produced only a limited understanding of the characteristics that allow a protein to induce sensitisation or a full allergenic reaction. Because of these complexities, it has long been recognized that there is no single parameter that can predict the allergenic potential of a substance. Recently, the strategy to address allergenicity of biotechnology products has been formulated [7, 31], which relies on the following parameters: source of the gene, sequence homology, serum testing of patients known to be allergenic to the source organism or to sources distantly related, pepsin resistance, the prevalence of the trait and animal models.

The source of the introduced protein should be part of the background material available to conduct an allergenicity assessment. Allergenic sources of genes would be defined as those organisms for which reasonable evidence of IgE mediated oral, respiratory, or contact allergy is available. Knowledge of the source of the introduced protein allows the identification of tools and relevant data to be considered in the allergenicity assessment.

Sequence homology is the initial step in the allergenicity assessment. When sequence homology to a known allergen is demonstrated, the product is considered allergenic and no further testing is typically undertaken. The FAO/WHO panel recommended using an amino acid window for the sequence homology that was scientifically justifiable. Research reports showed that an amino acid window size of less than eight amino acids may result in a high rate of false positives [32].

Specific serum screening is then undertaken irrespective of the prevalence of allergy to the source material in question when the source is a known food allergen. When no sequence homology has been found between the expressed protein and an allergen, targeted serum screening (direct source and related organisms) is undertaken. The use of larger numbers of sera is advocated whenever possible to increase the confidence associated with the results.

Additional assessment of the potential allergenicity of expressed proteins may be performed by pepsin degradation analysis or by using animal models. Pepsin digestion stability is believed to impart on the allergen an increased probability of reaching the intestinal mucosa intact where absorption of significant quantities may lead to sensitization. Protein stability in itself is, however, not sufficient to exclude potential allergenic properties as exceptions are known of less stable proteins that are allergenic. There are several animal models including the intraperitoneal (IP) murine model and the Brown Norway rat model. Failure to elicit IgE

antibody production after IP administration to the laboratory mice where immunogenicity is evident on the basis of IgG response may provide some reassurance that the protein lacks a significant potential to provoke allergic sensitization. In practice the predictive value of these systems for proteins that are new to the human diet may, however, be limited [33].

2.3.6 Unintended effects

Potential unintended effects represent a significant concern with GMOs including GM animals and these effects highlight the difficulty of establishing uniform considerations instead of case-by-case considerations. Unintended effects can be divided into insertional effects, related to the place of insertion of the transgenic fragment, and secondary effects, related to the nature of the expression products of the introduced genes. The major approach to detect any unintended side effects in the GM animal is a phenotypical and compositional analysis to compare the new food organism with the traditional counterpart. Whereas, there are databases on plant species describing the current knowledge (including a listing of (natural variation in) macro and micronutrients, natural toxins and other anti-nutritional factors) [34], a comparable database is not as readily available for food animals.

For GM animal-derived food products the same approach should apply. The edible tissues of the GM animal under investigation and comparable tissues from a genetically related non-GM animal should be phenotypically and compositionally analysed and screened for differences that may have toxicological or nutritional relevance. Similar to the GM plants, the key constituents of the tissue would have to be established. Because of the likeness between animals and humans, few animal tissue constituents can be considered anti-nutrients or natural toxins, but there are exceptions, such as thiaminase in different fish species and tetrodotoxin in puffer fish [10]. An important difference with GM plants is the average number of off-spring from one GM animal. The number of GM animals derived from a single GM founder animal will in general be much lower compared to GM plants. As the associated costs will be considerable, the selection process of the initial founders will be very limited compared to the plant breeding situation where thousands of GM calluses are screened for incorporation of the transgenic fragment and subsequently monitored for their phenotypic characteristics. This means that the information on the variation range between animals with the same genetic modification will be rather limited and that detected differences between individual animals will be difficult to interpret. In theory, the consequence of the smaller number in animal breeding may be that the selection process is less stringent with GM animals which may lead to higher chances for unintended effects. On the other hand, however, GM animals may be more vulnerable to smaller changes in their physiology and therefore selected transgenic organisms without obvious phenotypic aberrations may show relatively few physiological alterations when compared to GM plants. Further research may shed more light on these aspects with relation to the safety assessment.

As the number of key nutrients and/or anti-nutrients is limited in any species, a targeted compositional analysis will have its limitations in the information that can be provided.

For animal products where there is no tradition of composition analysis, unbiased profiling methodologies that are currently being developed may become a valuable addition to the present targeted approaches as part of the food safety assessment strategy, once they are validated [8]. The issue of sampling is crucial for both the targeted and profiling approach. For comparative compositional analyses, it is very important that the conditions for breeding of the animal and sampling of the edible animal parts are highly similar to avoid the detection of differences that are unrelated to the genetic modification. Animal tissues have to be analysed before any processing has taken place. At the same time, any potential effects of the subsequent processing steps should also be included in the overall risk evaluation.

There is likely to be expanded work in profiling food derived from GMOs including GM animals for safety evaluations as part of a CSA. The profiling approach can be roughly divided into holistic approaches on three different integration levels: genomics, proteomics, and metabolomics.

2.3.6.1 Genomics

Microarray technology is a powerful tool to study gene expression. The technology allows comparison of expression profiles of a large number of genes under different environmental or developmental conditions. cDNA or EST (expressed-sequence tag) libraries can be established of any organism under investigation [21]. If alterations in gene expression are found, the nature of the related gene will provide initial information on the toxicological or nutritional relevance of the alteration. Detected differences should be confirmed by additional targeted analysis preferably aiming also at the corresponding proteins or metabolites. The main advantage of the gene expression microarray approach over proteomic and/or metabolomic approaches is the scale of study. Where proteomics or metabolomics are likely to include at best 10% of the proteome or metabolome, respectively, the gene expression microarray makes it feasible to study whole transcriptomes of the organism.

2.3.6.2 Proteomics

Proteomics is the direct counterpart to transcriptomics. In general, correlation between mRNA expression and protein levels is rather poor as the rates of degradation of mRNA and proteins differ [35]. Therefore, understanding the biological complexities underlying gene function is facilitated by analysis of many proteins simultaneously. Methods used for analyzing differences in protein patterns include SDS-PAGE followed by peptide mass fingerprinting. There are, however, limits to what 2DGE can analyse as, in general, only highly expressed proteins will be detected [36]. Another approach that is currently being tested is the use of isotope-coded affinity tags to analyse fragmented proteins or multidimensional liquid chromatography coupled to mass spectrometry (http://www.foodstandards.gov.uk/science/research/NovelFoodsResearch/g02programme/g02projectlist/g02001). Also a protein microarray approach to accomplish the same end is under development based on the interactions of individual proteins with their

substrates or with other proteins. This development may lead in time to 'whole proteome' approaches that may reduce the necessity for initial gene expression profiling [37,38].

2.3.6.3 Metabolomics

Continuing down the cascade from genome and proteome is the metabolome. The metabolome consists of the metabolites that occur within a biological entity. A multi-compositional analysis of biologically active compounds (metabolites) may also indicate the presence of unintended effects. Metabolites can be determined by traditional chemical techniques including gas chromatography (GC), high pressure liquid chromatography (HPLC), coupled to mass spectrometry (MS) or nuclear magnetic resonance (NMR). These methods are capable of detecting, resolving, and quantifying a wide variety of compounds in a single sample. This type of chemical fingerprinting provides more details than can be obtained by single compound analysis. Once differences are observed these differences should be further analyzed by in vitro or in vivo testing. Before chemical fingerprinting can be readily exploited to determine substantial equivalence as related to GMO-derived foods, efforts to standardize sample collection and extraction are needed. Once again, background data on non-GMO comparator sources should be collected to acquire knowledge on the natural variability of the species.

2.3.7 Toxicology

In general, it will not be possible to test complex animal products by classical toxicological animal studies in the way they are routinely used to test single compounds. Classical studies measuring physiologic response relative to dose are complicated if the laboratory animal is receiving doses of the GM animal's edible tissue. If the genetic modification would result in the expression of novel proteins or if the compositional analysis revealed an alteration in an endogenous protein product or metabolite, the traditional toxicological approach would require the concentration of the product to be elevated in the laboratory animal's diet to the extent that the diet will often become unbalanced. This might result in toxicological observations that are unrelated to the product under investigation. To avoid this scenario, the concentration of the product can only take place within the limits of national and international recommendations on optimal laboratory animal diets, thereby limiting the sensitivity of the tests [39, 40]. On occasions where the genetic modification results in an increase in a specific (exogenous) protein, for instance directly derived from the gene construct, traditional testing would still be valid to assess that portion of the derived food. Alternatively, there may be instances wherein endogenous protein levels in the GM food are increased well above physiologic level in the given animal species and it might be prudent in specific cases to (also) test this elevated protein in animal studies.

2.3.8 Nutritional analysis

The nutritional analysis should focus on the potential replacement of nutritionally important food products by the novel GM animal-derived food products with possibly altered characteristics. The information for the nutritional analysis will largely be derived from the initial CSA, including the compositional analyses (especially macro-, micro- and anti-nutrients) and the estimated consumption rates. Detected alterations in the GM animal-derived food products compared to the traditional counterpart will be assessed by evaluating the significance of the compositional differences for the consumer in general and also, in specific cases, for specific consumer groups. Nutritional aspects of GMO-derived foods may become of increasing significance when the number of compositionally altered food products on the market increases. Therefore, the nutritional assessment of GM animal-derived food products is dependent on current consumption data of animal-derived food products in distinctive consumer groups and with respect to geographical and demographical differences. Special consumer groups perhaps worthy of special consideration include children, pregnant or lactating women and elderly persons.

Micronutrients are vitamins and minerals that are essential for normal physiology and biochemical functioning. Both deficiency and excess of a micronutrient can cause health problems which emphasizes the importance of this class of compounds. Macronutrients include dietary lipids, proteins and carbohydrates and these classes of compounds are present in the food and diet in substantial quantities. Assessment of the replacement factor of important animal-derived sources of micro- and macronutrients by GM animal products in the event of altered composition with relation to these nutrients is therefore of major importance. Bioavailability of the important micro- and macronutrients from GM animal-derived tissues is also of significant importance in this respect.

2.4 Exposure assessment

To assess the amount of food or food ingredient an individual or group is exposed to, represents the goal of an exposure assessment. No exact criteria have been formulated so far for the factors considered in an pre-market exposure assessment of a complex novel food product. Some exposure paradigms make assumptions based on per capita production while others use per capita distribution. An exposure assessment may also consider the cooking and food preparation process used. Some governments have instituted tracking of animal derived food and from this dataset, post-market consumption data may be determined. Exposure assessments will also include an estimate of the extent to which current food products will be replaced by the GM animal-derived novel food product. Thus, the accuracy of the exposure assessment for GMO-derived foods is dependent upon the available data on consumption patterns in consumer groups of interest and the validity of the underlying parameters [8].

The potential exposure of children of different age groups to growth factors in GM fish-derived tissues is an actual example as this GM animal-derived food product is requesting market entry in the US. The exposure assessment will be based on available consumption data and our knowledge on the bioavailability of the growth factors upon consumption. Mathematical models for integrating food consumption and distributions may be used in a so-called probabilistic approach to estimate future exposures more precisely. Alternatively, biomarker based methodologies for quantifying exposure to food chemicals are garnering interest but this approach is not yet validated for traditional food additives much less for GMO-derived food [41].

2.5 Risk characterization

Risk characterization typically refers to the probability that a hazard would produce a given adverse effect. Risk characterization is the stage of risk assessment that integrates information from exposure assessment and hazard characterization into advice suitable for use in decision making or risk management. It is prudent to highlight that risk characterization is typically viewed as an iterative and evolving process. With traditional food additives the risk characterization can take the form of establishment of an allowable daily intake level (ADI).

In the case of GMO-derived food the many facets of the CSA and the exposure assessment would need to be matrixed together. The baseline for the safety of novel food products derived from GMOs, including GM animals, in all cases will have to be the assessment that the novel GM animal-derived food products is at least as safe as its traditional counterpart. If any questions remain after the initial CSA with respect to the safety of the GM animal-derived food products additional tests may be required, including animal studies with the whole product or selected tissues/extracts. If, after a full safety assessment, the safety standard can not be satisfied the GM animal-derived product should not be approved for marketing. For food products derived from GM food animals this characterization should be established on a case-by-case basis.

2.5.1 Post-marketing surveillance

Closely related to the risk characterization is the issue of post-marketing surveillance. Post market surveillance could be useful in certain instances where a better estimate of dietary exposure and/or nutritional consequence of GMO-derived food are required. In general, potential safety issues should be addressed adequately through pre-market studies. However, given the complexities of food allergies it is conceivable that, for instance, allergenicity concerns could warrant post-market surveillance [42] as part of the risk management profile.

For GM animal-derived medicinal substances existing pharmacovigilance schemes will apply to monitor any unforeseen unintended side effects of the isolated medicinal substances. The same would apply in a veterinary sense with respect to the GM animal itself when modified with respect to the production of hormonal or disease-prevention substances: pharmacovigilance

schemes could help to detect unintended side effects of the introduced expression product to the GM food animal that were not detected in the pre-market phase. To this end the GM animals should then be included in such pharmacovigilance schemes on the basis of 'internal' administration of the specific veterinary substance.

Post-marketing surveillance systems for GM animal-derived food products need the establishment of adequate traceability systems of the GM animal products in the food production chain. Here, the food animal sector has clear advantages over the crop plant sector where basic traceability systems for individual farms, let alone plants, is still virtually lacking. In the animal production sector, such systems are already well-established for some animal food production chains in some countries and many other initiatives are ongoing in this field.

Traceability will in practice be most feasible for well-characterised GM animals dedicated for the production of specific substances or tissues that are kept in containment. Safety precautions should, however, be aimed at the prevention of any introduction of these GM animals into the food supply chain.The precautions should also include the development of analytical tests for detection and identification.

It is important to note that traceability and related control systems may be less straightforward in the case of chimeric organisms as different parts of the food animal will have different genetic constitutions and this may severely complicate analytical control of traceability systems.

Depending on the questions and risk management needs underlying the establishment of post-marketing surveillance systems the information conveyed to the consumer may, however, require adjustment. In order to enable consumers to relate potential adverse, for instance allergenic, effects to a GM animal-derived food product, it will be necessary to not only label the product as GMO-derived, but also provide information on the specific GM animal source, for instance by including in the label the unique identifier code specific for a single founder animal and its offspring.

2.6 Conclusions

The food safety evaluation of GM animals and derived products can largely be performed along the lines that have already been established for the evaluation of GM plants and derived products for the consumer. This means that the initial step of the food safety evaluation will be a CSA of the GM animal with its traditional counterpart, if available. This approach identifies potential differences between the GM animal-derived food products and its traditional counterpart as the first phase. The next phase is then to gain further insight into the toxicological and nutritional relevance of the detected differences. As every GM (founder) animal at this moment will have a different genetic constitution with respect to the integration of the genetic construct, the safety evaluation should be carried out on a case-by-case basis, even if the same genetic construct was used for the genetic modification. If homologous recombination will reduce the

possibility of insertional effects in the future, it may become more feasible to come to more harmonised approaches for the safety assessment of GM animals and products thereof.

Application of the concept of substantial equivalence allows for analysis of intended and unintended alterations in the GMO and is central to the CSA. The intended changes can be evaluated with knowledge of: the nature and source of the gene construct used in the modification, the process of the genetic modification, in situ characterization of the genetic modification in the animal, information on animal breeding and propagation of the GM animal, the amino acid sequence of expressed product from the gene construct, the expression rates in different tissues of the expressed product, and traditional toxicological testing of the expressed product.

In addition the food safety evaluation should focus on possible unintended side-effects of the genetic modification. Unintended effects can be divided into insertional effects, related to the place of insertion of the gene fragment, and secondary effects, related to the nature of the transcription products of the introduced genes. Allergenicity represents a possible hazard that most likely is an unintended effect of the modification of a food animal. To detect any unintended effects a comparative phenotypical and compositional analysis of the new food organism and the conventional counterpart should be carried out. This should currently be based on the known key micro- and macronutrients and anti-nutrients, if applicable, and may in the future also be based on unbiased profiling of the GMO-derived food and traditional counterpart. Techniques for the profiling approach are now under development and can be divided into three subsections: genomics, proteomics, and metabolomics to screen for differences in the GM animal with relation to the gene transcription products, proteins and metabolites, respectively. At this moment, however, none of these techniques is yet validated and ready for routine use in risk assessment. If applied, depending on the identity of differences detected further toxicological testing may be required to assess the safety and nutritional impact of the observed differences.

A few major differences can be seen when comparing the GM animal to the GM plant situation. Firstly, the numbers of GM animals derived from a single GM founder animal will in general be much lower compared to GM plant genetic modification events and numbers available in subsequent plant generations. This will result in less animals being available for the comparative safety assessment. This will have major influence on the reliability of the results of the comparative safety assessment. Knowledge on the natural background variation in animal tissue constituents will even be more important compared to the plant situation as it will be less feasible to obtain statistically significant results from analysis of the GM animals versus the conventional counterpart. An additional difference is the omnipresence of natural toxins in plant products and the very few cases of animal products that have proved to contain antinutritional substances for the consumer.

A third difference relates to the traceability systems that are (will be) available in the animal production sector and not yet in the plant sector. The presence of these traceability systems will make proper post-marketing surveillance systems much more feasible compared to the plant situation. Post-marketing surveillance studies may be advocated in the case of uncertainties relating to the nutritional or exposure assessment of the product or, in exceptional cases, to the potential allergenicity of the newly introduced protein(s). Other health- and nutrition-related aspects should be sufficiently dealt with during the pre-market assessment. Depending on the questions underlying a post-marketing study it may, however, be necessary in order to meet the goal, of the study to add information on the GM source animal to the label and inform the consumer of this additional label information.

Current food safety regulations for traditionally food (or food additives) are less stringent compared to those applied to GM foods. Pre-market safety assessment of GMO-derived foods must provide sufficient safety assurances, also in the case of GM animals or products derived thereof. The use of post-marketing surveillance as an instrument to gain information on the long-term effects of food either GMO derived or traditional should be further explored, but the requirement of routine application will entail large costs for limited amounts of information and should therefore be limited to exceptional cases.

The fact that the physiology of animals has major resemblance to our own physiology may in some aspects make the assessment of (GM) animal-derived food products 'easier'. On the other hand animal-derived food products form an important part of the human diet. Relatively small compositional changes may therefore have considerable effects on the nutritional status of the consumer. With increasing numbers of genetically altered plant- and animal derived food products the nutritional aspects, beside the safety aspects, will increasingly gain weight. The new developments in the area of GM animals further necessitate a harmonised approach to maintain our current standard for a safe and nutritious food supply in the light of growing numbers of different (GMO-derived) foods and food ingredients and increasingly complex food supply chains.

References

1. IFBC (1990) Biotechnologies and food: Assuring the safety of foods produced by genetic modification. Regulatory Toxicology and Pharmacology 12: S1-S196.
2. OECD (1993) *Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles*. Organisation for Economic Co-operation and Development, Paris, http://www.oecd.org/pdf/M00034000/M00034525.pdf.
3. OECD (1996) *Food Safety Evaluation*. Organisation for Economic Co-operation and Development, Paris.
4. FAO/WHO (1991) Strategies for Assessing the Safety of Foods Produced by Biotechnology. Report of a Joint FAO/WHO Consultation. World Health Organisation, Geneva, Switzerland.

5. FAO/WHO (1996) Biotechnology and Food Safety. Report of a Joint FAO/WHO Consultation. FAO Food and Nutrition Paper 61. Food and Agriculture Organisation of the United Nations, Rome, Italy. http://www.fao.org/es/esn/gm/biotec-e.htm.
6. FAO/WHO (2000) Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland. Food and Agriculture Organisation of the United Nations, Rome. http://www.fao.org/es/esn/gm/biotece.htm.
7. FAO/WHO (2001) Evaluation of Allergenicity of Genetically Modified Foods . Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland. Food and Agriculture Organisation of the United Nations, Rome. http://www.fao.org/es/esn/gm/biotece.htm.
8. Kuiper H.A., Kleter G.A., Noteborn H.P.J.M., and Kok E.J. (2001) Assessment of the food safety issues related to genetically modified foods. Plant Journal 27: 503-528.
9. National Research Council of the National Academies (U.S.), Committee on Defining Science based Concerns Associated with Products of Animal Biotechnology. (2002) Animal biotechnology: science-based concerns. The National Academies Press, Washington D.C., U.S.
10. Kleter G.A. and Kuiper H.A. (2002) Considerations for the assessment of the safety of genetically modified animals used for human food or animal feed. Livestock Production Science 74:275-285.
11. National Research Council of the National Academies (U.S.), Committee on Life Science (1983) Risk Assessment in the Federal Government: Managing the Process. The National Academies Press, Washington D.C., U.S.
12. Ziomek CA (1998) Commercialization of proteins produced in the mammary gland. Theriogenology. 9:139-44.
13. Colman A. (1999) Dolly, Polly and other 'ollys': likely impact of cloning technology on biomedical uses of livestock. Genet Anal. 15:167-73.
14. Houdebine L.-M. (2002) Antibody manufacture in transgenic animals and comparisons with other systems. Current Opinion in Biotechnology 13,6:625-629.
15. Maclean N. (2003) Genetically modified fish and their effects on food quality and human health and nutrition. Trends in Food Science & Technology 14, 242-252.
16. Zhang Q, Tiersch TR, Cooper RK. (1998) Inducible expression of green fluorescent protein within channel catfish cells by a cecropin gene promoter. Gene 216:207-13.
17. Wheeler MB, Bleck GT, Donovan SM. (2001) Transgenic alteration of sow milk to improve piglet growth and health. Reprod Suppl. 58:313-24.
18. Golovan SP, Meidinger RG, Ajakaiye A, Cottrill M, Wiederkehr MZ, Barney DJ, Plante C, Pollard JW, Fan MZ, Hayes MA, Laursen J, Hjorth JP, Hacker RR, Phillips JP, Forsberg CW. (2001) Pigs expressing salivary phytase produce low-phosphorus manure. Nat Biotechnol. Aug;19(8):741-5.
19. Garnier J.-P., Klont R., and Plastow G. (2003) The potential impact of current animal research on the meat industry and consumer attitudes towards meat. Meat Science 63,1:79-88.
20. Kuiper H.A., Kleter G.A., Noteborn H.P.J.M., and Kok E.J. (2002) Substantial equivalence – an appropriate paradigm for the safety assessment of genetically modified foods. Toxicology 181-182:427-431.
21. Kok E.J. and Kuiper H.A. (2003) Comparative safety assessment for biotech crops. Trends Biotechnol. 21: 439-44.
22. Chakraborty AK, Zink MA, Hodgson CP. (1994) Transmission of endogenous VL30 retrotransposons by helper cells used in gene therapy. Cancer Gene Ther. 1:113-8.

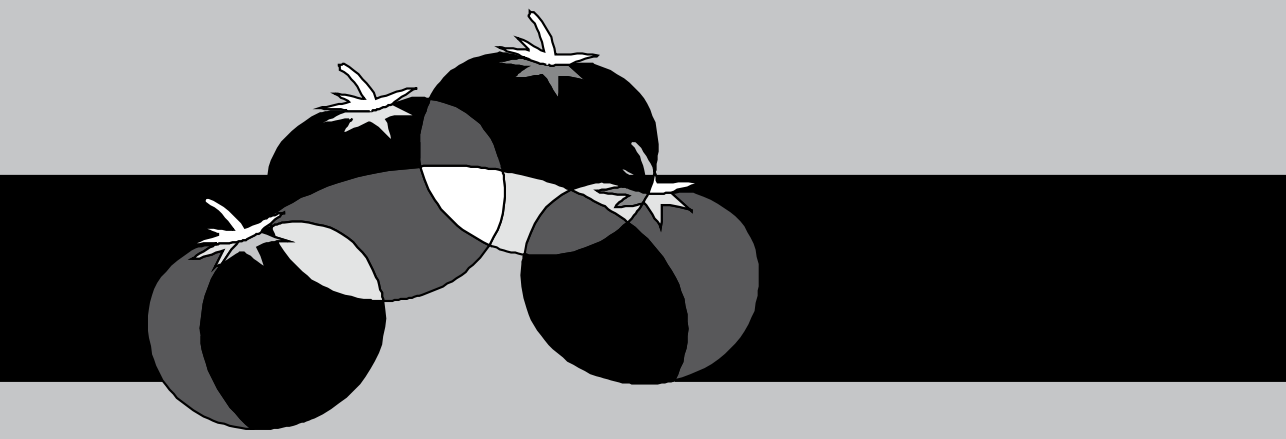
23. Scadden DT, Fuller B, Cunningham JM. (1990) Human cells infected with retrovirus vectors acquire an endogenous murine provirus. J Virol. 64:424-7.
24. Purcell DF, Broscius CM, Vanin EF, Buckler CE, Nienhuis AW, Martin MA. (1996) An array of murine leukemia virus-related elements is transmitted and expressed in a primate recipient of retroviral gene transfer. J. Virol. 70:887-97.
25. Faber D.C., Molina J.A., Ohlrichs C.L., Van der Zwaag D.F., and Ferré L.B. (2003) Commercialization of animal biotechnology. Theriogenology 59,1:125-138.
26. Schubbert R, Renz D, Schmitz B, Doerfler W. (1997) Foreign (M13) DNA ingested by mice reaches peripheral leukocytes, spleen, and liver via the intestinal wall mucosa and can be covalently linked to mouse DNA. Proc Natl Acad Sci U S A. 94:961-6.
27. Schubbert R, Hohlweg U, Renz D, Doerfler W. (1998) On the fate of orally ingested foreign DNA in mice: chromosomal association and placental transmission to the fetus. Mol Gen Genet. 259:569-76.
28. Beever D.E., and Kemp C.F. (2000) Safety issues associated with the DNA in animal feed derived from genetically modified crops. A review of scientific and regulatory procedures. Nutrition Abstracts 70: 175-182.
29. Van den Eede G., Aarts H., Buhk H.-J., Corthier G., Flint H.J., Hammes W., Jacobsen B., Midtvedt T., Van der Vossen J., Von Wright A., Wackernagel W., Wilcks A. The relevance of gene transfer to the safety of food and feed derived from gm-plant. Food and Chemical Toxicology, submitted.
30. Dyck M.K., Lacroix D., Pothier F., and Sirard M.-A. (2003) Making recombinant proteins in animals – different systems, different applications. Trends in Biotechnology 21,9:394-399.
31. Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL. (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit Rev Food Sci Nutr. 36 Suppl:S165-86.
32. Hileman RE, Silvanovich A, Goodman RE, Rice EA, Holleschak G, Astwood JD, Hefle SL. (2002) Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. Int Arch Allergy Immunol. 128:280-91.
33. König A., Cockburn A., Crevel R.W.R., Debruyne E., Grafstroem R., Hammerling U., Kimber I., KnudsenI., Kuiper H.A., Peijnenburg A.A.C.M., Penninks A.H., Poulsen M., Schauzu M., Wal J.M. (2004) Assessment of the Safety of Foods derived from Genetically Modified (GM) Crops. Food and Chemical Toxicology, 42, 1047-1088.
34. OECD (2003) Consensus Documents for the Work on the Safety of Novel Foods and Feed. Task Force for the Safety of Novel Foods and Feed, Organisation for Economic Co-operation and Development, Paris, http://www.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-document-530-nodirectorate-no-27-24778-32,00.html.
35. Gygi SP, Rochon Y, Franza BR, Aebersold R. (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 19:1720-30.
36. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A. 97:9390-5.
37. MacBeath G, Schreiber SL. (2000) Printing proteins as microarrays for high-throughput function determination. Science. 289:1760-3.
38. Templin MF, Stoll D, Schrenk M, Traub PC, Vohringer CF, Joos TO. (2002) Protein microarray technology. Drug Discov Today. 7:815-22.

39. Barlow S.M., Greig J.B., Bridges J.W., Carere A., Carpy A.J.M., Galli C.L., Kleiner J., Knudsen I., Koëter H.B.W.M., Levy L.S., Madsen C., Mayer S., Narbonne J.-F., Pfannkuch F., Prodanchuk M.G., Smith M.R., and Steinberg P. (2002) Hazard identification by methods of animal-based toxicology. Food and Chemical Toxicology 40,2-3:145-191.
40. Cockburn A. (2003) Assuring the safety of genetically modified (GM) foods: the importance of an holistic, integrative approach. Journal of Biotechnology 98,1:79-106.
41. Kroes, R., Muller, D., Lambe, J., Lowik, M.R.H., van Klaveren, J., Kleiner, J., Massey, R., Mayer, S., Urieta, I., Verger, P., Visconti, A., (2002) Assessment of intake from diet. Food and Chemical Toxciology 40: 327-385.
42. Hlywka, JJ., Reid, JE., Munro, IC., (2003) The use of consumption data to assess exposure to biotechnology-derived foods and the feasibility of identifying effects on human health through postmarket monitoring. Food and Chemical Toxicology 41: 1273-1282.

Chapter 3.

Abstract

Microarray technology makes it feasible to analyse the expression of thousands of different gene elements in a single experiment. Most informative are 'whole genome' arrays, where all gene expression products of a single species or variety are represented. Such arrays are now available for a limited number of model species. However, for other, less well-documented species other routes are still necessary to obtain informative arrays. This includes the use of cDNA libraries. To enhance the amount of information that can be obtained from cDNA libraries, redundancy needs to be minimised, and the number of cDNAs relevant for the conditions of interest needs to be increased. Here, we used representational difference analysis (RDA), a mRNA subtraction procedure, as a tool to enhance the efficiency of cDNA libraries to be used to generate microarrays. Tomato was chosen as a model system for a less well-documented species. cDNA libraries for two distinct physiological conditions of tomato fruits, red and green, were made. The libraries were characterized by sequencing and hybridisation analysis. The RDA procedure was shown to be effective in selecting for genes of relevance for the physiological conditions under investigation, and against constitutively expressed genes. At the same time, redundancy was reduced, but complete normalisation was not obtained, and subsequent sequence analysis will be required to obtain non-redundant arrays. Further, known and putative ripening-related cDNAs were identified in hybridisation experiments on the basis of RNA populations as isolated from the green and red stage of ripening.



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

Esther J. Kok, Nicole L.W. Franssen-van Hal, Lies N.W. Winnubst, Evelien H.M. Kramer, Wilko T.P. Dijksma, Harry A. Kuiper and Jaap Keijer

Journal of Plant Physiology 2007, 164: 337-349.

3.1 Introduction

A number of different methods have been developed to study gene expression in different physiological systems, and to identify differentially expressed genes. The first methods to assess populations of mRNA, such as differential display [1] or arbitrarily primed PCR [2,3], were based on the amplification of random mRNA subsets and subsequent analysis of the resulting fragment pool by gel electrophoresis. The methods are relatively laborious, and the resulting banding patterns are often difficult to analyse. Moreover, single differentiated bands can represent multiple cDNAs and do not give immediate clues as to the identity of the underlying genes. Serial analysis of gene expression (SAGE) was developed as an elegant means to analyse mRNA populations by large-scale sequence determination of short identifying stretches of individual messengers [4], but the required depth of sequencing for comparison of different conditions makes it also rather labour intensive and time consuming. More recent developments have made it possible to study altered gene expression in a more efficient and informative way by using DNA microarray technology [5,6,7,8]. Using DNA microarrays, the expression of a large number of genes can be analysed simultaneously and in a semi-quantitative manner. This allows for the analysis of different metabolic pathways in interaction and facilitates the identification of key responsive genes. For a limited number of species, microarrays have been constructed which represent all identified metabolic routes and genes active therein. These are the so-called whole genome arrays, oligo-arrays where all expressed gene sequences are represented by one or more short DNA sequences (usually up to 100 nucleotides [8]). For less well-documented species, such oligo-arrays will not be available in the near future. Therefore, other ways to obtain arrays with as many informative sequences as possible with relation to the questions underlying the investigation need to be identified. Depending on the amount of knowledge that is already available on the species and (physiological) condition being investigated, it may be appropriate to obtain new libraries enriched for the sequences of interest. For the detection of differential gene expression in less well-documented species, we chose the tomato as a model and developed a microarray-based strategy that combines the use of named, known genes and uncharacterised ESTs that are potentially related to the scope of the investigation. The named genes represent sequences that are already functionally described and can serve as controls and references. They allow for the description of the behaviour of selected processes under the physiological conditions that are studied. The uncharacterised genes are research question-related, pre-selected cDNAs obtained by the use of representational difference analysis (RDA, [9]). RDA is a hybridisation selection procedure in which mRNAs that are present in one condition (wanted), and not in the other (unwanted), are selectively amplified. Theoretically, the procedure has a preference for low abundant mRNAs, and the protocol results in a normalised cDNA library enriched for informative sequences, depending on the number of repetitions of the procedure and on the applied stringency. The RDA procedure was used to increase the informative nature of the array with respect to the questions to be answered. The resulting cDNAs can be further analysed to distinguish identified genes from unidentified ones and to further characterise metabolic pathways of interest. Possible applications of RDA-based cDNA libraries are manifold, with one being the generation of gene expression profiles of new plant varieties within the framework of

food safety assessment procedures. To test the potential of this latter approach, RDA was applied to obtain cDNA libraries of two different stages of ripening of tomato fruits. Tomato was used as a model because it is an important agricultural crop and a sufficient number of sequences have been annotated (NCBI/EMBL databases and TIGR tomato gene index database (http://www.tigr.org/tdb/tgi/plant.shtml). In addition, some expression information is available, which allows for assessment of the success of the approach. This may not have been the case if some other less well-documented target crop species had been used. By two independent RDA procedures, cDNA libraries were obtained that are specific for the green and red stage of ripening. In this paper, the approach is described, as well as experiments to assess the quality of the libraries and the derived arrays. The results strongly suggest that the selected approach is useful to obtain informative sequences specific for the selected conditions. At the same time, we show that RDA does not lead to the required normalisation of cDNA populations, and that additional sequence analysis will be required in most cases to (further) reduce the redundancy to the desired level.

3.2 Materials and methods

3.2.1 Total RNA isolation from tomatoes

Traditionally bred red and mature green tomatoes were used, obtained from a regular breeding programme. The stage of ripening was determined using a colour card as regularly applied by tomato breeders (The Greenery B.V., Breda, Kleurstadia Tomaten, Jan.2001.1050; green stage: stadium 1; red stage: stadia 9-12). For the isolation of total RNA from tomato tissue material, an extraction method based on TRizol Reagent (InVitrogen, Breda, the Netherlands) was used. Samples of peel and pulp with a volume of approximately 50 µl were freeze dried prior to crushing with Eppendorf micropestels in Eppendorf test tubes in liquid nitrogen and resuspension in TRizol Reagent at room temperature. The samples were vortexed prior to incubation at room temperature for 5 min. Subsequently, the samples were centrifuged for 15 min at 12000g at 4 °C to remove the cell debris. Then 0.2 ml chloroform/ isoamylalcohol (24:1) was added to the supernatants and mixed by inverting the tube for a period of 15 s. The samples were then incubated 2-3 min at room temperature prior to 15 min of centrifugation at 12000g at 4 °C. The supernatant was transferred to a new reaction tube. This purification step was repeated once with 500 µl chloroform/isoamylalcohol. The RNA was then precipitated in 0.5 ml ice-cold isopropanol, dried in a vacuum exsiccator and resuspended in 60 µl DEPC-treated H_2O. To resolve the pellet, the reaction tube was incubated for 10 min in a 55-60 °C water bath. The total RNA sample was subsequently DNase-treated, quantified using Ribogreen (Molecular Probes, Invitrogen, Breda, the Netherlands) and stored at -70 °C. The yield per sample was approximately 12-15 µg total RNA.

3.2.2 Construction of the subtractive cDNA libraries

cDNA synthesis from the total RNA isolates was performed using the Clontech SMART PCR cDNA Synthesis kit (Clontech, Mountain View, USA) according to the protocol supplied by

the manufacturer. For the cDNA syntheses, total RNA was used from a red tomato and from a green tomato. The Advantage PCR-Pure Kit was used for the purification of the resulting RsaI-digested PCR products. Adaptors were ligated to a part of the purified cDNAs resulting in the 'test' or 'wanted' population. The remaining cDNAs formed the 'driver' population. Subtractions were performed according to the protocol of the Clontech PCR-Select cDNA Subtraction kit (http://www.clontech.com/clontech/techinfo/manuals/PDF/ PT1117-1.pdf, Clontech, Mountain View, USA). The obtained PCR products were purified using either the Advantage PCR-Pure Kit (Clontech, Mountain View, USA) or the Qiaquick kit (Qiagen, Venlo, the Netherlands) according to the protocol of the supplier. Subsequently, a reaction was performed to obtain an A'overhang on both ends of the PCR products. This reaction was performed in a volume of 10 µl, containing 250 ng of the purified PCR-product, 5 U AmpliTaq (Perkin Elmer, Wellesley, USA), 1 x PCR-buffer (Perkin Elmer, Wellesley, USA), 2.5 mM $MgCl_2$ and 0.2 mM dATP. The reaction mixture was incubated for 20 min at 70 °C. Then 50 ng of the A-tailed PCR products was used for ligation into the pGEM-T Easy vector and subsequent transformation into XL-2 Blue ultra competent cells (Stratagene, La Jolla, USA). The cells were plated out onto LB-agar ampicilline plates and colonies were grown overnight at 37 °C. The plates were screened for white colonies that were grown in 100 µl LB-medium in 96-wells plates o/n. The next day, sterile glycerol was added to the grown cultures and the plates were stored in a -70 °C refrigerator prior to further examination.

3.2.3 Control clones

As positive controls for the reverse transcription reaction, four luciferase cDNA sequences were cloned. To obtain the luciferase control clones, a RT reaction was performed on luciferase mRNA (Promega, Leiden, the Netherlands) with an olig-T primer using the Superscript pre-amplification system for First Strand cDNA synthesis (InVitrogen, Breda, the Netherlands). Subsequently, a PCR was performed on 2.5 µl of the RT reaction, using the primer combinations as indicated below. RT-PCR products were cloned in the vector pGEM-T Easy (Promega, Leiden, the Netherlands) and transformed to the bacterial strain DH5α. The insert of the obtained clones was confirmed by sequence analysis.

Primers used for the RT-PCR of the luciferase clones:

LUC long	F: 5'-ATGGAAGACGCCAAAAAC-3'
	R: 5'-TTTACAATTTGGACTTTCC-3' (1714bp)
LUC 5'	F: 5'-ATGGAAGACGCCAAAAAC-3'
	R: 5'-GTCACGATCAAAGGACTC-3' (627 bp)
LUC middle	F: 5'-GAGTCCTTTGATCGTGAC-3'
	R: 5'-GGAACAACTTTACCGACC-3' (611 bp)
LUC 3'	F: 5'-CGGTCGGTAAAGTTGTTC-3'
	T: 5'-TTTACAATTTGGACTTTCC-3' (633 bp)

As a negative control, a Salmonella sequence (410 bp) was cloned into pUC18 and amplified using 1F and M13R primers that are present on pUC18. The Salmonella sequence was checked using BLAST-N (NCBI) in the Genbank/EMBL database for absence of similarity to sequences derived from the organisms under investigation.

3.2.4 Construction and hybridisation of the microarray

The clones that were randomly selected for the microarray were amplified by PCR using the primers F (5'-AGGCGATTAAGTTGGGTAAC-3' and 3R (5'-AGAATACTCAAGCTATGCATC-3') which are located on the pGEM-T Easy vector. Both primers contained a5'-C6-amine linker to allow binding to the aldehyde groups present on the silylated slides. PCR amplification was performed directly on the glycerol stocks of the clones in a volume of 100 µl, containing 40 pmol forward primer, 40 pmol reverse primer, 1.5 mM $MgCl_2$ (InVitrogen, Breda, the Netherlands), 1 x PCR buffer (InVitrogen, Breda, the Netherlands), 0.2 mM dNTPs, 5 U Taq polymerase (InVitrogen, Breda, the Netherlands) and 1 µl glycerol stock. The PCR started with a denaturation step of 2 min 94 °C, followed by 35 cycles: 40s 94 °C, 60s 55 °C and 2.5 min 72 °C. After a final elongation step of 10 min 72 °C the reactions were cooled down to 4 °C. For quality control 1 µl of the PCR products was checked on a 1% agarose gel.

The amplification products were purified either by Qiaquick purification kit (Westburg, Leusden, the Netherlands) or by ethanol precipitation. After ethanol precipitation, the pellets were dissolved in 10 µl 5 x SSC. After purification with the Qiaquick purification kit, eluates were evaporated overnight on a hot plate and dissolved in 10 µl 5 x SSC. A random selection of app. 1100 cDNAs of the resulting libraries was spotted in array format. As positive controls three different luciferase fragments corresponding to the 5', middle and 3' part of the cDNA, as well as the entire luciferase cDNA were spotted twice in each side of the array as well as a Salmonella gene fragment as a negative control. In addition the 3' end of the luciferase cDNA and the Salmonella negative control were spotted in each block of the array.

Microarrays were printed on silylated slides (CELAssociates, Pearland, USA) using a Cartesian Technologies, PixSys 7500 spotter (Biodot limited, Chichester, UK). Arrays were spotted by passive dispensing using Chipmaker 3 pins (Telechem, Sunnyvale, USA).

After printing, microarrays were dried at room temperature for at least 3 days. Free aldehyde groups were blocked with $NaBH_4$ according to the method of Schena *et al*. [10]. Slides were directly used for hybridisation or stored at room temperature in a slidebox until further use.

3.2.5 mRNA isolation

mRNA isolates for the hybridisation experiments were obtained from a second set of tomatoes at the mature green and red stage of ripening using the Dynabeads Direct kit (Dynal, InVitrogen, Breda, the Netherlands). The stage of ripening was determined using a colour card as regularly

applied by tomato breeders (The Greenery B.V., Breda, the Netherlands, Kleurstadia Tomaten, Jan.2001.1050; green stage: stadium 1; red stage: stadia 9-12). To this end, tomatoes were milled in liquid nitrogen and stored until use in a -70 °C freezer. Per isolation app. 1 ml of this tomato material was resuspended in lysis/binding buffer as provided by the supplier and mRNA was isolated according to the supplied protocol. A small sample of the mRNA isolate was used for quantification and quality control by gel electrophoresis.

3.2.6 Fluorescent labelling of the mRNA (cDNA) samples for hybridisation

Two microgram of sample mRNA was labelled by incorporation of either Cy3-dCTP or Cy5-dCTP during a RT reaction [10,11,12]. The labelled cDNA was dissolved in hybridisation buffer (5 x SSC, 0.2% SDS, 5 x Denhardt's, 50% (v/v) formamide, 0.2 mg/ml denatured herringsperm DNA). Prior to hybridisation, samples were heated for 3 min at 65 °C and spun to remove undissolved debris. Before hybridisation the microarrays were pre-hybridised in hybridisation buffer at 42 °C for at least 4 h. After the pre-hybridisation the arrays were washed twice in MilliQ water and once in isopropanol and dried by centrifugation (2 min at 470g). The hybridisation was performed under a cover slide (5 µl hybridisation buffer) or in a hybridisation frame (50 µl hybridisation buffer, Geneframe, Implen, Munich, Germany). Arrays were hybridized at 42 °C in a humid hybridization chamber overnight. After hybridisation the arrays were washed at room temperature, successively in 1 x SSC/0.1% SDS (5 min), 0.1 x SSC/0.1% SDS (5 min) and 0.1 x SSC (1 min) and subsequently dried by centrifugation (2 min, 470g).

3.2.7 Scanning and data analysis

Microarrays were scanned using a confocal laser scanner ScanArray 3000 (General Scanning Inc., Pittsfield, USA) containing a GreNe 543 nm laser for Cy3 measurement and a HeNe 633 nm laser for Cy5 measurement. Scans were made with a pixel resolution of 10 micron. The software package ArrayVision (Imaging Research, Waalwijk, the Netherlands) was used for image analysis of the TIFF-files as generated by the scanner. Fluorescent signals were collected for each individual spot and stored for further data processing using the Micro-soft Excel® and Genemaths XT® (Applied Maths, St-Martens-Latem, Belgium) software.

3.2.8 Sequence analysis of the cDNA libraries

Sequence analysis was performed by Greenomics, Plant Research International, Wageningen, the Netherlands. The resulting sequences were analysed on the basis of both the NCBI BLAST-N and BLAST-X programmes and the NCBI/Genbank database for homology with functionally identified cDNAs.

3.3 Results

Two tomato cDNA libraries were obtained by RDA subtraction of green tomato mRNA from red tomato mRNA and the reverse subtraction, resulting in libraries that are presumably specific for the red and green stage of ripening of tomato, respectively. The size of the fragments in the resulting libraries was analysed and is shown in Figure 1. The insert size ranged from 150 to 1500, and was comparable in the two libraries (average 500-700).

To investigate the quality of the libraries, sequencing and hybridisation experiments were performed. Altogether 2204 inserts (1151 from the green and 1053 from the red library) were sequenced and characterized by alignment and BLAST analysis. A larger number of different sequences were found in the green-specific library, i.e. the total of sequences obtained in the libraries, in basepairs. green-red subtractive procedure, also after correction for total number of sequences (Table 1). Further, 25% of the 'green' sequences were identified by BLAST

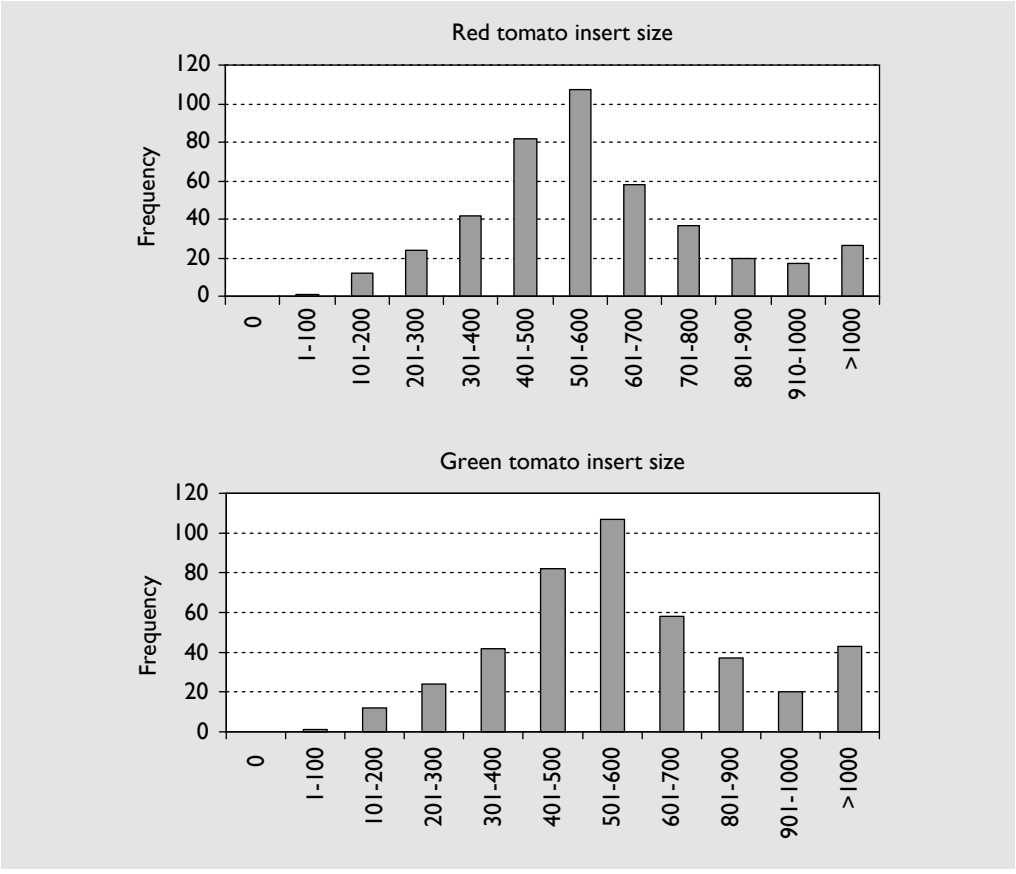


Figure 1. *Size distribution of the red and green tomato libraries, in basepairs.*

Table 1. *Numbers of total, different and unique sequences, present in the green and red tomato fruit RDA cDNA-libraries, based on sequence identity as determined by NCBI BLAST analysis*

GREEN	GREEN adjusted[1]	RED	
1151	1053	1053	total sequences
402	368	279	different sequences[2]
297	272	204	unique sequences[3]

[1] For numerical comparison.
[2] Based on principle N-BLAST hits.
[3] Represented once in the green or red library, based on principle N-BLAST hits.

analysis as being unique within the 'green' library, whereas 19% of the 'red' sequences are unique in the 'red' library, i.e. the total of sequences as obtained in the red-green subtractive procedure. Only a limited number of sequences (38) representing 424 clones, were observed in both libraries (Table 2). In the same table, it can also be seen that for a number of sequences in this category, the numbers can differ considerably between the two libraries. This is the case both for known tomato ripening-related genes, such as 'early light-inducible protein', a glucuronosyl transferase homologue and a beta-fructofuranosidase precursor, as well as for sequences that have not been previously described as related to tomato ripening, such as tomato 'glutathione S-transferase-like protein', '2-isopropylmalate synthase' and 'alcohol acyl transferase' from *Lycopersicon pennellii*, and sequences that show high homology to the tobacco sequence 'nictaba' (69%) and the 'steroleosin' sequence from *Sesamum indicum* (78%). In the green-specific library the sequences for 'tomato elongation factor 1-alpha', 'tobacco arginine decarboxylase' and the 'potato dehydrin homologue CI7' stand out, none of them known tomato ripening-related sequences.

Table 2. *Sequences that are represented in both the green and red tomato fruit RDA cDNA library, based on sequence identity as determined by NCBI BLAST analysis.*

Name	NCBI accession number	Green[1]	Red[2]
1-aminocyclopropane-1- carboxylate oxidase homolog	gi|119640	7	6
2-isopropylmalate synthase A [*Lycopersicon pennellii*]	gi|7387848	1	15
arginine decarboxylase [*Nicotiana tabacum*]	gi|7436502	12	1
hsr201 protein, hypersensitivity-related [*Nicotiana tabacum*]	gi|7489142	3	30
coatomer protein complex, putative [*Arabidopsis thaliana*]	gi|15218215	7	5

Table 2. *Continued.*

Name	NCBI accession number	Green[1]	Red[2]
cytochrome P450 family protein [*Arabidopsis thaliana*]	gi|15238202	1	1
dehydrin homolog CI7 protein [*Solanum tuberosum*]	gi|7489228	8	1
elongation factor 1-ALPHA [*Lycopersicon esculentum*]	gi|119150	9	1
endochitinase 4 precursor , potato (fragment)	gi|1705810	4	4
translation initiation factor 6, putative [*Arabidopsis thaliana*]	gi|15228161|	1	1
hypothetical protein; [*Arabidopsis thaliana*]	gi|15218692	1	1
putative protein [*Arabidopsis thaliana*]	gi|18413967	1	3
hypothetical protein; [*Arabidopsis thaliana*]	gi|15220409	1	1
expressed protein [*Arabidopsis thaliana*]	gi|22326566	3	4
early light inducible protein [*Lycopersicon esculentum*]	gi|14487954	20	75
formate dehydrogenase [*Solanum tuberosum*]	gi|11991527	1	1
putative alanine acetyl transferase [*Arabidopsis thaliana*]	gi|15225174	6	1
glucuronosyl transferase homolog, ripening-related - tomato	gi|629669	1	9
glucosyltransferase IS5a, salicylate-induced - common tobacco	gi|7433906	2	2
glutathione S-transferase-like protein [*Lycopersicon esculentum*]	gi|6653233	1	24
hypothetical 10.0K protein [*Zinnia elegans*]	gi|543565	3	1
beta-fructofuranosidase precursor [*Lycopersicon esculentum*]	gi|124701	1	15
Nt-gh3 deduced protein [*Nicotiana tabacum*]	gi|4887010	10	6
nictaba [*Nicotiana tabacum*]	gi|15088626	6	19
putative oxidoreductase [*Arabidopsis thaliana*]	gi|15219099	3	1
putative pectate lyase [*Arabidopsis thaliana*]	gi|18416424	1	2
peroxidase , cationic - adzuki bean [*Vigna angularis*]	gi|478409	2	3
unknown protein [*Arabidopsis thaliana*]	gi|15224494	1	1
glyceraldehyde-3-phosph dehydrogenase [*Zymomonas mobilis*]	gi|120716	9	5
putative pathogenesis related protein [*Oryza sativa*]	gi|5042456	1	3
hypothetical protein [*Arabidopsis thaliana*]	gi|25372796	2	2
similarity to nitrate-induced NOI protein [*Arabidopsis thaliana*]	gi|9294163	1	1
11-beta-hydroxysteroid dehydrogenase-like [*Arabidopsis thaliana*]	gi|15241205	1	5
steroleosin [*Sesamum indicum*]	gi|15824408	3	15
transcription factor JERF1 [*Lycopersicon esculentum*]	gi|22074046	1	1
putative glucosyl transferase [*Arabidopsis thaliana*]	gi|15228031	11	4
wound-induced protein Sn-1 [*Capsicum annuum*]	gi|2129926	4	2
expressed protein [*Arabidopsis thaliana*]	gi|18406725	1	1

[1]Number of times the sequences are present in the green tomato fruit RDA cDNA library.
[2]Number of times the sequences are present in the red tomato fruit RDA cDNA library.

When comparing the sequences that occur only in one of the libraries (Table 3) the most marked difference is seen for the 'polygalacturonase 2α precursor gene sequence', 89 times positively identified in the red library with nine additional identifications as 'putative polygalacturonases', and none in the green library. Other differences are the ripening-related '1-aminocyclopropane-1-carboxylate synthase – tomato' sequences, occurring 5 (ACC synthase 2), 6 (ACC synthase 4, gi100181) and 9 (ACC synthase 4, gi231357) times in the red library, as well as 'acid beta-fructofuranosidase precursor' (15 times in the red library), 'deoxyuridine 5'-triphosphate pyrophosphatase-like protein (*A. thaliana*)' (8 in red), 'endopolygalacturonase' (5 in red), 'gene C7 protein' (6 in red), and 'putative acyltransferases (*Cucumis melo*)' (8 in red), and 'putative lysosomal acid lipase' (*A. thaliana*, 8 in red). None of these sequences was found in the green library. For the green library, many more gene sequences were found that are solely observed in this library, but with lower frequencies, possibly reflecting less transcripts per gene in the green

Table 3. *Sequences that differ between the green and red tomato fruit RDA cDNA libraries. Sequences with a differential presence of three or more are included. RR= known ripening-related.*

Name		NCBI accession number	Green[1]	Red[2]
Predominantly red				
2-isopropylmalate synthase A [*Lycopersicon pennellii*]	--	gi|7387848	1	15
1-aminocyclopropane-1-carboxylate synthase 2 [*Lycopersicon esculentum*]	RR	gi|100182	-	5
1-aminocyclopropane-1-carboxylate synthase 4 - tomato	RR	gi|100181	-	6
branched-chain amino acid aminotransferase [*Capsicum annuum*]	--	gi|14280354	-	5
hsr201 protein, hypersensitivity-related [*Nicotiana tabacum*]	--	gi|7489142	3	30
carbonic anhydrase, chloroplast precursor, common tobacco	--	gi|8096277	-	4
dUTP pyrophosphatase-like protein [*Arabidopsis thaliana*]	--	gi|15232681	-	8
early light inducible protein [*Lycopersicon esculentum*]	RR	gi|46401440	20	75
endopolygalacturonase [*Lycopersicon esculentum*]	--	gi|225933	-	5
gene C-7 protein - common tobacco [*Nicotiana tabacum*]	--	gi|100318	-	6
glucuronosyl transferase homolog, ripening-related - tomato	RR	gi|629669	1	9
glutathione S-transferase-like protein [*Lycopersicon esculentum*]	--	gi|6653233	1	24
late-embryogenesis protein homolog - tomato heat-shock protein	--	gi|7489029	-	3
beta-fructofuranosidase precursor [*Lycopersicon esculentum*]	RR	gi|124701	1	15

Table 3. *Continued.*

Name		NCBI accession number	Green[1]	Red[2]
11-beta-hydroxysteroid dehydrogenase-like [*Arabidopsis thaliana*]	--	gi|15241205	1	5
LYCES ACC synthase 4	RR	gi|231357	-	9
nictaba [*Nicotiana tabacum*]	--	gi|15088626	6	19
putative nuclear protein of eukaryotic origin [*Caenorhabditis elegans*]	--	gi|17565906	-	3
phosphoenolpyruvate carboxylase kinase 2 [*Lycopersicon esculentum*]	RR	gi|15418703	-	4
polygalacturonase 2A precursor (Pectinase) [*Lycopersicon esculentum*]	RR	gi|129939	-	89
polygalacturonase, putative / pectinase, putative [*Arabidopsis thaliana*]	--	gi|15219756	-	9
probable anthranilate N-benzoyltransferase	--	gi|7484744	-	5
putative acyltransferase [*Cucumis melo*]	--	gi|18652312	-	8
putative lysosomal acid lipase [*Arabidopsis thaliana*]	--	gi|15226073	-	8
putative protein [*Arabidopsis thaliana*]	--	gi|15241334	-	8
unknown protein [*Arabidopsis thaliana*]	--	gi|15225276	-	3
putative protein; [*Arabidopsis thaliana*]	--	gi|18416995	-	8
unknown protein [*Arabidopsis thaliana*]	--	gi|14030723	-	5
steroleosin [*Sesamum indicum*]	--	gi|15824408	3	15
putative glucosyl transferase [*Arabidopsis thaliana*]	--	gi|15228033	-	3
Predominantly green				
3-dehydroquinate dehydratase / shikimate 5-dehydrogenase - tomato	--	gi|7436862	3	-
3-hydroxy-3-methylglutaryl-coenzyme A reductase 3 [*Solanum tuberosum*]	--	gi|11133016	3	-
arginine decarboxylase [*Nicotiana tabacum*]	--	gi|7436502	12	1
Nt-gh3 deduced protein [*Nicotiana tabacum*]	--	gi|4887010	10	6
auxin-responsive family protein [*Arabidopsis thaliana*]	--	gi|15236962	3	-
endochintinase [*Solanum tuberosum*]	--	gi|21465	6	-
CjMDR1 [*Coptis japonica*]	--	gi|14715462	4	-
coatomer protein complex, putative [*Arabidopsis thaliana*]	--	gi|15218215	7	5
cytochrome P450 [Madagascar periwinkle]	--	gi|7430703	5	-
cytochrome P450, putative [*Arabidopsis thaliana*]	--	gi|15220009	5	-
cytochrome P450 76A2 hydroxylase [*Solanum melongena*]	--	gi|584865	5	-

Table 3. *Continued.*

Name		NCBI accession number	Green[1]	Red[2]
dehydrin homolog CI7; cold-stress inducible protein [*Solanum tuberosum*]	--	gi|7489228	8	1
elongation factor 1-ALPHA (EF-1-ALPHA) [*Lycopersicon esculentum*]	--	gi|119150	9	1
embryo-abundant protein EMB [*Pisum sativum*]	--	gi|20339366	3	-
putative alanine acetyl transferase [*Arabidopsis thaliana*]	--	gi|15225174	6	1
glucan endo-1,3-beta-glucosidase B precursor - tomato	--	gi|461979	5	-
hsp70 (AA 6 - 651) [*Petunia x hybrida*]	--	gi|123650	6	-
glyceraldehyde-3-phosphate dehydrogenase [*Zymomonas mobilis*]	--	gi|120716	9	5
jasmonic acid 2 [*Lycopersicon esculentum*]	--	gi|6175246	3	-
malate dehydrogenase (oxaloacetate-decarboxylating) - tomato	--	gi|7431232	4	-
monosaccharide transport protein MST1 [*Nicotiana tabacum*]	--	gi|100347	3	-
P-glycoprotein, putative [*Arabidopsis thaliana*]	--	gi|15217785	5	-
nitrate reductase [NADH] (NR) - tomato [*Lycopersicon esculentum*]	--	gi|128195	5	-
oxalyl-CoA decarboxylase [*Arabidopsis thaliana*]	--	gi|12049590	3	-
peroxidase [*Manihot esculenta*]	--	gi|14029184	4	-
phenylalanine ammonia-lyase [*Lycopersicon esculentum*]	RR	gi|129587	3	-
phyA [*Lycopersicon esculentum*]	--	gi|3492795	3	-
probable glucosyltransferase twi1 [*Lycopersicon esculentum*]	--	gi|7433911	3	-
putative chloroplast thiazole biosynthetic protein [*Nicotiana tabacum*]	--	gi|2501187	4	-
putative glucosyl transferase [*Arabidopsis thaliana*]	--	gi|15228031	11	4
phenylpropanoid:glucosyltransferase 1 [*Nicotiana tabacum*]	--	gi|13492674	3	-
unknown [*Arabidopsis thaliana*]	--	gi|21593250	5	-
unknown [*Prunus armeniaca*]	--	gi|5031283	3	-
putative extracellular dermal glycoprotein precursor [*Arabidopsis thaliana*]	--	gi|15218740	4	-
putative zinc finger protein [*Arabidopsis thaliana*]	--	gi|7486672	4	-

[1]Number of times the sequences are present in the green tomato fruit RDA cDNA library.
[2]Number of times the sequences are present in the red tomato fruit RDA cDNA library.

stage of ripening. The most obvious differences are seen for different cytochrome sequences (15 in the green library with different principle NCBI hits, an 'endochitinase precursor' (six times in the green library), a 'glucan endo-1,3 -beta-glucosidase B precursor' (five in green), a 'hsp70 Petunia homologue' (six in green), 'putative P-glycoprotein (*A. thaliana*)' (five in green), 'nitrate reductase' (five in green), and a series of less well-defined sequences.

To analyse the correlation between the frequency of occurrence in the two individual libraries and the expression levels, hybridisation experiments were performed with mRNA isolates from a second set of mature green and red tomatoes. For this, the 2204 cDNAs were spotted in array format on glass slides. Furthermore, luciferase cDNA fragments from the 5', middle and 3' section as well as the entire luciferase cDNA were present on the array, both as positive controls and to obtain information on the effectiveness of the reverse transcription reaction as part of the labelling protocol (luciferase mRNA was spiked in the samples). A Salmonella-derived fragment was present as a negative control. The microarray was hybridised with fluorescently labelled mRNA that was extracted from green, unripe and red, ripe tomatoes and was spiked with the luciferase sequence. The hybridisation results were analysed in Microsoft Excel (spot selection, normalisation and categorisation of signal intensities) and Genemaths (analysis of differences in gene expression). Spots were selected on the basis of a signal-to-noise ratio larger than three. Positive spot numbers were calculated for the two types of hybridisation (back or self-hybridisation and cross-hybridisation) for the two sets of sequences, red-and green-specific, respectively (Figure 2). From the luciferase controls, we noted that the reverse transcription procedure was effective; the Salmonella negative control was negative. Average signals were higher after hybridisation with the red tomato, but the highest signals were found after the hybridisation with the 'green' material in the 'green' part of the array (results not shown). The obtained fluorescence signals and the frequency of occurrence were compared for a set of sequences that were found at least five times in the green or in the red EST library. The results are shown in Table 4. The signal intensities were categorized by classifying the normalised fluorescence signals into 4 quartiles relative to the highest fluorescent signals in the green and red hybridisations. For each sequence, the quartile was determined based on the average normalised fluorescence signal found in the different spots, representing the specific sequence in the two separate hybridisation experiments. Comparison of the hybridisation results for the sequences with the largest differences in gene expression profiles between the green and red stage of ripening with the profiles as published in the Tomato Expression Database (http://ted.bti.cornell.edu/ [13]) confirm the obtained expression data in the two developmental stages of tomato.

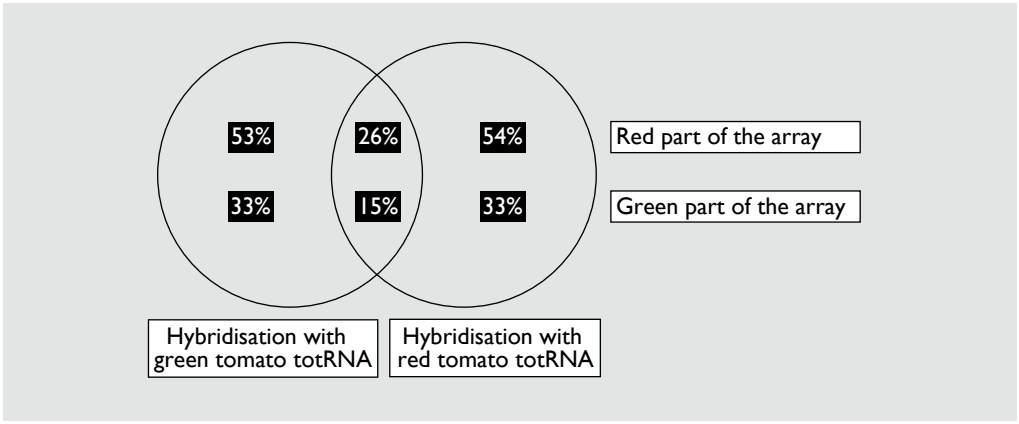


Figure 2: *Number of positive spots (3 times above background) in the green and red part of the array after 'self' or 'cross' hybridisations as well as percentage positive spots in both hybridisations. 'Self' hybridisation: green part of the array hybridised with totRNA isolated from the green tomato fruit or red part of the array hybridised with totRNA isolated from the red tomato fruit. 'Cross' hybridisation: green part of the array hybridised with totRNA isolated from the red tomato fruit or red part of the array hybridised with totRNA isolated from the green tomato fruit. The green and red specific parts of the array are defined as the sequences as obtained in respectively the green minus red and red minus green subtractive procedures.*

Table 4. *Comparison of abundance of sequences in the green and red tomato fruit[a] and the related RDA cDNA libraries.*

Nr of clones		BLASTOutput	Expression category	
GREEN	RED		GREEN	RED
11	4	gi|15228031| putative glucosyltransferase	C	B
9	1	gi|119150| LYCES elongation factor 1-alpha	B	A
5	0	gi|461979| LYCES glucan endo-1,3-beta-glucosidase B precursor	B	A
5	0	gi|7430703| cytochrome P450 - Madagascar periwinkle	B	A
5	0	gi|15217785| P-glycoprotein, putative	A	A
7	5	gi|15218215| coatomer complex subunit, putative	A	A
10	6	gi|4887010| Nt-gh3 deduced protein	A	A
5	0	gi|584865| SOLME cytochrome P450 76A2	A	A
12	1	gi|7436502| arginine decarboxylase	A	A

Table 4. *Continued.*

Nr of clones		BLASTOutput	Expression category	
GREEN	RED		GREEN	RED
0	89	gi|129939| LYCES polygalacturonase 2A precursor	A	C
0	5	gi|225933| endopolygalacturonase	A	C
1	24	gi|6653233| glutathione S-transferase/peroxidase [*Lycopersicon esc*]	A	C
0	5	gi|7484744| probable anthranilate N-benzoyltransferase	A	C
3	30	gi|7489142| hsr201 protein, common tobacco	A	C
0	8	gi|18652312| putative acyltransferase [*Cucumis melo*]	A	C
0	9	gi|231357| LYCES ACC synthase 4	A	B
1	9	gi|629669| glucuronosyl transferase homolog, ripening-related - tomato	A	B
0	5	gi|14030723| unknown protein [*Arabidopsis thaliana*]	A	B
1	15	gi|7387848| 2-isopropylmalate synthase A [*Lycopersicon pennellii*]	A	B
0	6	gi|100318| gene C-7 protein - common tobacco	A	B
3	15	gi|15824408| steroleosin [*Sesamum indicum*]	A	B
0	8	gi|18416995| putative protein; [*Arabidopsis thaliana*]	A	B
20	75	gi|14487954| early light inducible protein [*Medicago sativa*]	A	A
0	9	gi|15219756| polygalacturonase, putative [*Arabidopsis thaliana*]	A	A
0	6	gi|100181| 1-aminocyclopropane-1-carboxylate synthase	A	A
0	8	gi|15226073| putative lysosomal acid lipase [*Arabidopsis thaliana*]	A	A
0	8	gi|15241334| putative protein [*Arabidopsis thaliana*]	A	A
1	15	gi|124701| LYCES acid beta-fructonisidase precursor	A	A

Sequences are listed with a differential presence of five or more in the RDA cDNA libraries that were included in the hybridisation experiments.

[a]Signal intensities were categorized by classifying the normalised fluorescence signals in four quartiles relative to the highest fluorescent signals in the green respectively red hybridisations. For each sequence the quartile (A, B, C, D, A representing the lowest expression quartile, D representing the highest expression quartile) was determined based on the average normalised fluorescence signal found in the different spots representing the specific sequence in the two separate hybridisation experiments. As a result of taking the average values, it was found that none of the sequences could be placed in the highest signal (D) category in either the 'green' or the 'red' hybridisation experiment.

3.4 Discussion

DNA microarray technology analyses the expression of large numbers of genes in a single assay. For effectiveness, it is crucial to deposit cDNAs onto arrays that are as informative as possible. Ideally, this would mean a large, non-redundant array of cDNAs of known sequence that are functionally identified, and that relate to the physiological or developmental condition under investigation. At present, for many species, tissues and specific developmental or environmental conditions, oligonucleotide sets or cDNA libraries are not (yet) available. In cases where cDNA libraries are available, they are likely to contain many redundant sequences. By combining RDA and DNA microarray technology, potentially arrays can be created which contain cDNA sets that are non-redundant (normalised) and problem-targeted (enriched for the condition under investigation). RDA was applied here to generate informative cDNA libraries representative for red and green tomato, respectively. The obtained libraries have an average sequence length of 500-700 bp, with only a few short sequences being present (Figure 1). This is indicative of a successful subtraction procedure, as omitting the subtractive hybridisation selection step will result in a higher number of small inserts and a lower average insert length (500 bp, unpublished results).

Sequencing results show that there is a relatively large number of unique sequences, i.e. sequences that occur only once in the red-specific or green-specific library (19% and 25%, respectively). This likely indicates that normalisation occurred in both libraries. Moreover, there is little overlap between the two libraries. From this, it can be concluded that the procedure effectively selects for sequences that are uniquely present in green or red tomatoes, or that the two physiological conditions have very little in common. The latter may be true, since the green/breaker stages are reported to be the physiologically most active stages of ripening [14] with declining and markedly changed physiological activity towards the more mature, red stages [15,16,17,18]. Recent literature [19] suggests it is likely that the physiological activity in the developing green fruit is even higher, but this is outside the scope of the experiments discussed here. If the sequencing results represent either the distinct physiological states or the result of an effective subtraction, we would expect close to a 100% positive spots for self-' hybridisation, i.e. red RNA hybridised to red cDNAs and green RNA hybridised to green cDNAs, and a much lower number of positive signals when red RNA is hybridised to green cDNAs and green RNA is hybridised to red cDNAs. Low abundance of specific sequences may change these figures to some extent, but cannot explain the fact that we observe very similar percentages of positive spots for the 'self-' and 'cross-' hybridisation. Specifically, we observed 33% for both hybridisations in the green part of the array and 54% and 53% for the 'self-' and 'cross-' hybridisation, respectively, in the red part of the array (Figure 2). The discrepancy between the hybridisation results (>30% positive spots in the cross-hybridisation) and the sequencing 'data (<20% overlap between the libraries) indicates that, under the conditions employed, the number of hybridising spots is not an informative parameter to assess the efficiency of the procedure. The sequencing results not only imply a selection for ripening stage-specific genes, but also imply a strong selection against genes with comparable gene expression in the two conditions. Only 22 sequences were found

with similar representation (Table 2; the difference between the number of sequences was less than three) in the two libraries, which is only about 7% of the total number of red sequences and only approximately 5% of the green sequences. This is much lower than the reported number of constitutively expressed genes (http://www.plantcell.org/cgi/content/full/14/7/1441), assuming that the more conserved genes are indeed related to basic metabolic routes in the plant, and therefore implies selection during the subtractive procedure. Moreover, when looking at the number of spots that give positive signals in both hybridisation experiments and can therefore be regarded as more or less constitutively expressed genes, the number is also much higher (26% and 15% for the red and the green part of the array, respectively). This indicates further the successful redundancy reduction during the procedure. In almost all of these cases where both hybridisations give positive results, it can be seen that the highest fluorescence signal is found in the 'self-' hybridisation, as can be expected.

The higher number of positive 'red' spots is, however, likely to be reflected in the higher average signal intensity in the hybridisation experiments for 'red' as compared to 'green', which gives a lower average, but larger dynamic range, i.e. higher signal intensities for some sequences. This may indicate a smaller number of different mRNAs being present in red tomatoes, with higher abundance of individual mRNAs. This is in agreement with the higher physiological activity of green tomato as compared to red, ripe tomatoes. Indeed, we see the largest number of different sequences in the green library compared to the red library. From our (limited) dataset, it can be estimated that in the green tomato at least 20% more genes are actively expressed, compared to the red tomato, assuming that the subtraction protocol has been comparably efficient in both cases (Table 1). The average fluorescent signal values are at least 1.5 times higher for the self-hybridisation compared to the reverse hybridisation. For the red part of the array, the self hybridisation shows an average fluorescent signal value that is even 3.2 times higher than for the cross hybridisation. This considerably higher value for the average signal in the self-hybridization is likely explained by selection of RDA against redundancy, as the effect of a concentrated sample hybridised to a random set of sequences is the same for both types of hybridisation. This is in agreement with the data shown in Table 4, which indicates a relative discrepancy for the larger part of the sequences between the level of expression and observed redundancy in the library.

Pursuing this line of argument, the best indication for reduced redundancy comes from the sequences that show the highest hybridisation signals. For the green stage of ripening, the highest signal was found for a 'late embryogenesis protein homologue' from tomato (gi7489029, results not shown). The signal was decreased up to ten times from the green to the red stage of ripening. Remarkably, this sequence is present only three times in the red cDNA library, and not in the green cDNA library. Here, it seems that the redundancy reduction during the subtraction procedure has lead to the complete absence of the abundant sequence. With such high expression rates, it would be very unlikely that the sequence was eliminated from the random selection if no redundancy reduction had taken place. Similarly, the other way around most high 'red' hybridisation signals correspond to clearly ripening-related sequences

such as 'polygalacturonase 2A precursor' and putative ripening related sequences such as 'probable anthranilate N-benzoyltransferase'. However, here the sequence for tomato 'extensin' (gi|100209) is also found, and occurs only once in the red library. This latter finding also is likely to be the result of an efficient redundancy reduction of an abundant (red-specific) sequence during the subtraction procedure, and represents the desired situation at the onset of the subtraction procedure: maximal normalisation leading to optimal hybridisation signals. From these results, it can be concluded that redundancy reduction has indeed occurred during the subtraction phase.

The effectiveness of the procedure to select for differential expressed genes, against constitutively expressed genes and for reduction of redundancy, implies that the identified differentially expressed genes are likely to be ripening-related. These genes include known and novel ripening-related genes. For instance, for the '1-aminocyclopropane-1-carboxylate synthases 2' [20], the '1-aminocyclopropane-1-carboxylate synthases 4' [21], the 'glucuronosyl transferase homologue' [22], the 'beta-fructofuranosidase precursor' [23], the 'phosphoenolpyruvate carboxylase kinase 2' [24], 'early light-inducible protein' [25] and the 'polygalacturonase 2a precursor' [26], it has been established that they are upregulated in the red stage of ripening. Similarly, 'phenylalanine ammonia-lyase' may be involved in glycoalkaloid biosynthesis [27], http://www.tigr.org/tigr-scripts/tgi/mapTCs.pl?species=tomato&gi_dir=lgi&map=map00960) and is therefore more likely to be present in the green stage of ripening when more glycoalkaloids are formed. Other cDNAs that seem clearly ripening-related, such as '2-isopropylmalate synthase A', an important enzyme in leucine biosynthesis, and 'hsr201 protein, hypersensitivity response-related', have not previously been reported as such. Also 'glutathione S-transferase-like protein', which has so far been regarded as a stress-related enzyme [28], was shown to be differentially expressed in red tomatoes. In addition, two genes that show high homology with genes from *Nicotiana tabacum* (nictaba) and *Sesamum indicum* (steroleosin) seem to be related to the ripening process in tomato. The down-regulated genes are also, for most part, genes that show high homology to heterologous genes; for instance, to 'arginine decarboxylase' (*N. tabacum*, 93%), 'endochitinase' (*Solanum tuberosum*, 78%), cold stress inducible 'dehydrin homologue CI7' (*Solanum tuberosum*, 93%), and a 'putative glucosyl transferase' (*Arabidopsis thaliana*, 60%). The only tomato-specific cDNA that is clearly down-regulated from the green to the red stage of ripening is the 'elongation factor 1-alpha', with a central role in protein synthesis [29]. This decline in expression may reflect overall reduction in metabolic activity in the red stage of ripening.

Altogether, the RDA procedures were shown to be effective in selecting genes of relevance for the physiological conditions under investigation. At the same time, it resulted in reduced redundancy, but complete normalisation was not obtained, and subsequent sequence analysis will be required to obtain non-redundant arrays. RDA can thus contribute to obtain informative EST libraries that may be applied to assess differences in gene expression profiles. Ultimately, such arrays may be used within food safety assessment procedures, but to this end, the approach needs to be further assessed and validated. It is necessary to show its value in light of rapid developments in the area of especially protein profiling. At this point, RDA can be primarily

applied as an initial step to identify sequences of interest and to reduce the clones that must be sequenced to produce non-redundant cDNA sets for microarray experiments.

Acknowledgements

The authors would like to thank the European Committee (GMOCARE (QLK1-1999-00765) project) and the Dutch Ministry of Agriculture, Nature Conservation and Fisheries (scientific programme LNV 390) for their financial support of these studies

References

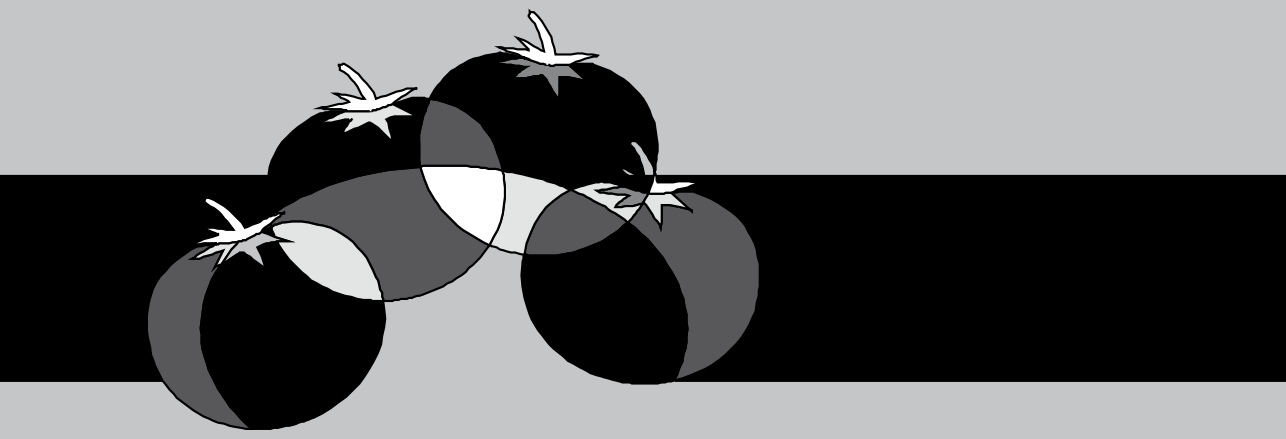
1. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 1992;257:967-71.
2. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res 1990;18: 7213-8.
3. Welsh J, Chada K, Dalal SS, Ralph D, Cheng L, MlClelland M. Arbitrarily primed PCR fingerprinting of RNA. Nucl Acids Res 1992;20:4965-70.
4. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:484-7.
5. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270: 467-70.
6. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000;405:827-36.
7. Panda S, Sato TK, Hampton GM, Hogenesch JB. An array of insights: application of DNA chip technology in the study of cell biology. Trends Cell Biol 2003;13(3): 151-6.
8. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. Genomics 2005;85(1):1-15.
9. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. Science 1993; 259(5097):946-51.
10. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray¬based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 1996;93:10614-9.
11. Boeuf S, Klingenspor M, Van Hal NLW, Schneider T, Keijer J, Klaus S. Differential gene expression in white and brown preadipocytes. Physiol Genomics 2001;7: 15-25.
12. Franssen-van Hal NLW, Vorst O, Kramer EHM, Hall RD, Keijer J. Factors influencing cDNA microarray hybridisation on silylated glass slides. Anal Biochem 2002;308(1):5-17.
13. Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J 2004;40:47-59.
14. Moore S, Vrebalov J, Payton P, Giovannoni J. Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. J Exp Bot 2002; 53(377):2023-30.
15. Fraser PD, Hedden P, Cooke DT, Bird CR, Schuch W, Bramley PM. The effect of reduced activity of phytoene synthase on isoprenoid levels in tomato pericarp during fruit development and ripening. Planta – Heidelberg 1995;196(2):321-6.
16. Kende H. Enzymes of ethylene biosynthesis. Plant Physiol 1989;91:1-4.

17. Grierson D, Fray R. Control of ripening in transgenic tomatoes. Euphytica 1994;79(3):251-63.
18. Giovannoni JJ, DellaPenna D, Bennett AB, Fischer RL. Expression of a chimeric polygalacturonse gene in transgenic rin (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. Plant Cell 1989;1(1):53-63.
19. Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. Plant Physiol 2005;139:750-69.
20. Rottmann WH, Peter GF, Oeller PW, Keller JA, Shen NF, Nagy BP. 1-Aminocyclopropane-1-carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence. J Mol Biol 1991;222(4):937-61.
21. Van der Straeten D, Van Wiemeersch L, Goodman HM, Van Montagu M. Cloning and sequence of two different cDNAs encoding 1-aminocyclopropane-1-carboxylate synthase in tomato. Proc Natl Acad Sci USA 1990; 87(12):4859-63.
22. Picton S, Gray J, Barton S, AbuBakar U, Lowe A, Grierson D. cDNA cloning and characterisation of novel ripen¬ing-related mRNAs with altered patterns of accumula¬tion in the ripening inhibitor (rin) tomato ripening mutant. Plant Mol Biol 1993;23(1):193-207.
23. Elliott KJ, Butler WO, Dickinson CD, Konno Y, Vedvick TS, Fitzmaurice L, *et al*. Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. Plant Mol Biol 1993;21(3): 515-24.
24. Bahrami AR, Chen ZH, Walker RP, Leegood RC, Gray JE. Ripening-related occurrence of phosphoenolpyruvate carboxykinase in tomato fruit. Plant Mol Biol 2001;47(4):499-506.
25. Bruno AK, Wetzel CM. The early light-inducible protein (ELIP) gene is expressed during the chloroplast-to-chromoplast transition in ripening tomato fruit. J Exp Bot 2004;55(408):2541-8.
26. Grierson D, Tucker GA, Keen J, Ray J, Bird CR, Schuch W. Sequencing and identification of a cDNA clone for tomato polygalacturonase. Nucl Acids Res 1986; 14(21):8595-603.
27. Lee SW, Robb J, Nazar RN. Truncated phenylalanine ammonia-lyase expression in tomato (*Lycopersicon esculentum*). J Biol Chem 1992;267(17):11824-30.
28. Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tsichlis PN, *et al*. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. J Biol Chem 2000;275(38):29207-16.
29. Pokalsky AR, Hiatt WR, Ridge N, Rasmussen R, Houck CM, Shewmaker CK. Structure and expression of elongation factor 1 alpha in tomato. Nucl Acids Res 1989; 17(12):4661-73.

Chapter 4.

Abstract

An important part of the comparative approach to assess the safety of new crop plant varieties is an extensive compositional analysis, including the measurement of all key nutrients and antinutrients in a specific crop. The study described here investigates the applicability of 'omics' technologies, transcriptomics and proteomics, as additional tools in this comparative safety assessment. The aim of the work was to assess the extent of the natural variation in ripening tomato fruits as a model crop and to determine whether it is possible to develop simple 'ripening stage' criteria for the sampling of fruits for 'omics' analyses. It is shown that the set-up of an 'omics' study is of crucial importance. Samples under scrutiny should be well-matched with relation to environmental conditions during growth and harvest, including the stage of ripening, as is stipulated in international guidance documents for the nutritional and toxicological assessment of genetically modified plants.



Changes in gene and protein expression during tomato ripening – consequences for the safety assessment of new crop plant varieties

Esther J. Kok[1], Satu J. Lehesranta[2], Jeroen P. van Dijk[1], J. Richard Helsdingen[1], Wilko T.P. Dijksma[1], Angeline M.A. Van Hoef[1], Kaisa M. Koistinen[2], Sirpa O. Kärenlampi[2], Harry A. Kuiper[1] and Jaap Keijer[1]

[1]RIKILT Institute of Food Safety, Bornsesteeg 45, P.O. Box 230, 6700 AE Wageningen, the Netherlands
[2]Institute of Applied Biotechnology, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland

Food Science and Technology International (in press).

4.1 Introduction

The arrival of genetically modified (GM) plant products on the market has led to fundamental questions on the safety of our present food supply. In traditional plant breeding, food safety issues are not consistently and systematically addressed prior to market introduction, unless there are clear indications that the composition is significantly altered. In only a limited number of food plant species specified antinutrient substances are routinely analysed in new varieties prior to marketing. This is quite the opposite in GM plants and products thereof. Here, extensive compositional analysis is required, in addition to information on the genetic modification and its direct effects, to ensure that no unintended side effects of the modification have occurred. In Europe the procedure for the food safety evaluation for foods and feed derived from GMOs (genetically modified organisms) is laid down in Regulation 1829/2003 [1] and associated guidelines for the producer have been formulated by the GMO panel of EFSA (European Food Safety Authority) [2].

The safety evaluation of GM crop plants is a comparative safety assessment (CSA) [3,4,5,6]. It includes an extensive compositional analysis, comprising the analysis of the most relevant crop-related nutrients and anti-nutrients. The outcome of these analyses are compared to the values of these constituents in their nearest comparator, e.g. the parent line, as well as to values from literature, taking into account the natural variation in composition as far as documented in the specific crop (variety). Ever since the food safety evaluation of the first GMO-derived plant products it emerged that our knowledge on the composition of most crop plant species is rather limited. The result of this is that the compositional analysis will focus on a large number of compounds in different metabolic pathways in order to cover the physiology of the plant as broadly as possible in search of possible unintended side effects of the genetic modification. Nevertheless, other metabolic routes, perhaps but not necessarily of less relevance, will remain unchallenged. Besides this 'biased' selection of key compounds for the compositional analysis, the availability of validated analytical methods for natural plant constituents is still rather poor.

In order to overcome this situation, projects were started in the late 1990's to develop new methodologies based on the 'omics'-approaches that, in theory, can supply more information, especially on unintended side effects of any form of plant breeding with a special focus on unintended effects of gentechnological changes [7,8, www.safefoods.nl). Transcriptomics, proteomics and metabolomics provide a wide overview of metabolism at, respectively, the mRNA, protein and metabolite level. In theory, the information they supply in selected samples is largely complementary, in practice it will be even more so as currently none of the approaches can guarantee full coverage of the transcriptome, proteome and metabolome, respectively. To assess the informative value of obtained profiles, the tomato was selected as one of the model species, because it is an important food crop and sufficient genomic information is available.

Comparative safety assessment is preferably performed on the red ripening stage of tomato, since this is the stage that is mostly consumed. Ideally for this experiment ripening profiles, i.e. gene and protein expression profiles at the subsequent stages of ripening, should be obtained for all genes and proteins that were monitored. In this way it would be possible to assess in future experiments whether detected changes are related to genetic modifications or merely to the stage of ripening of the sample under scrutiny. The aim of the work described here was to assess the extent of the natural variation in ripening tomato fruits and to investigate whether it is possible to develop simple 'ripening stage' criteria for the sampling of tomato fruits, which can be used in comparative compositional studies within food safety assessment strategies. In other words: is it possible to sample tomato fruits in a way that minimizes the natural variation that is due to the stage of ripening? To this end transcriptome and proteome profiles were obtained from five subsequent stages of ripening: green, breaker, turning, light red and red. To obtain comparable profiles for the transcriptomics and proteomics analyses, tomatoes were cut in four and opposite parts were used for either transcriptomics or proteomics. For transcriptomics a RDA (representational difference analysis) -based tomato array was used containing over 2000 EST-sequences that are specific for either the green or red stage of ripening. Two-dimensional electrophoresis (2-DE) was used to monitor over 600 protein spots, and differentially expressed proteins were identified. Using these approaches transcriptome and proteome profiles were obtained for the five subsequent ripening stages. The results were confirmed on the basis of literature reports and real-time PCR experiments. The implications of the findings are discussed in the light of current scientific literature on tomato ripening and with respect to the food safety evaluation of new (tomato) plant varieties.

4.2 Materials and methods

4.2.1 Tomatoes

For the isolation of the selected cDNA populations traditionally bred red, light red, turner, breaker and green tomatoes (*Solanum lycopersicum, var. Moneymaker*) were used that were obtained from a regular breeding programme. The stage of ripening was determined using a colour card as used by tomato breeders (The Greenery B.V., Breda, Kleurstadia Tomaten, Jan.2001.1050; green stage: stadium 1; breaker stage: stadium 2; turning stage: stadium 3; light red stage: stadia 5-7; red stage: stadia 9-12). Tomatoes were cut in 4 parts and immediately frozen in liquid nitrogen. Opposite parts were combined as a set. In this way two sets of tomato parts were obtained from each tomato. One set was used for transcriptomics analysis. The other set was used for proteomics analysis.

4.2.2 RNA isolation

For the isolation of total RNA from tomato tissue material an extraction method based on TRizol Reagent (InVitrogen, Breda, the Netherlands) was used. Frozen samples of peel and pulp were ground under liquid nitrogen and samples of approximately 2 ml were resuspended in 9

ml TRizol Reagent at room temperature. The samples were vortexed during 1 min or until all ground sample was mixed with the Trizol reagent prior to incubation at room temperature for 15 minutes. Subsequently the samples were centrifuged for 15 minutes at 11000g at 4 °C to remove the cell debris. 0.5 ml chloroform/isoamylalcohol (24:1) was added to the supernatants and mixed by inverting the tube during 15 seconds. The samples were then incubated 3 minutes at room temperature prior to 15 minutes of centrifugation at 11000g and 4 °C. The supernatant was transferred to a new reaction tube and 0.5 ml 8M LiCl per ml Trizol reagent as was used at the initial step of the isolation. The LiCl was mixed with the supernatant and incubated overnight at 4 °C. The following day, the tubes were centrifuged for 15 minutes at 11000g and the supernatant was removed, washed with 75% ice-cold ethanol (25% DEPC-treated water) and centrifuged again at 11000g during 10 min at 4 °C. The pellet was then dried and resuspended in 60 µl DEPC-treated water. Finally the RNA pellet was incubated in a 65 °C water bath for 10 minutes to completely dissolve. The total RNA sample was subsequently DNase-treated, quantified using Ribogreen (Molecular Probes, InVitrogen, Breda, the Netherlands) and stored at -80 °C. The yield per sample was approximately 12-15 µg total RNA.

4.2.3 Microarrays

Microarrays containing 2112 cDNAs of two RDA-based libraries were used: 960 clones of a green-tomato enriched cDNA library, and 1044 clones of a red-tomato enriched cDNA library [9]. In addition control sequences were spotted in each block of the array. As positive controls, four luciferase cDNA sequences were cloned, a full length sequence and partial sequences of the 5'-, 3'- and middle part of the luciferase sequence. As a negative control a Salmonella sequence (410 bp) was cloned. The Salmonella sequence was checked using BLAST-N (NCBI) in the Genbank/EMBL database for absence of similarity to sequences derived from tomato. Microarrays were printed on silylated slides (CELAssociates, Pearland, USA) using a Cartesian Technologies, PixSys 7500 spotter (Biodot limited, Chichester, UK) and Chipmaker 3 pins (Telechem, Sunnyvale, USA). After printing, microarrays were dried at room temperature for at least 3 days. Free aldehyde groups were blocked with $NaBH_4$ according to the method of Schena *et al*. [10].

4.2.4 Microarray hybridisation

25 µg of sample mRNA was labeled by incorporation Cy5-dCTP during a RT reaction [11]. The labeled cDNA was dissolved in 25 µl hybridisation buffer (5 x SSC, 0.2% SDS, 5 x Denhardt's, 50% (v/v) formamide, 0.2 mg/ml denatured herring sperm DNA). Prior to hybridisation, samples were heated for 3 min at 65 °C and spun to remove undissolved debris. As a reference samples were taken from all experimental RNA isolates, mixed and labelled by incorporation of Cy3-dCTP in a separate RT reaction according to the same labelling protocol. After the final dissolving step of the individually labelled cDNA fractions the Cy5- and Cy3-labelled fractions are mixed in a 1:1 ratio. Before hybridisation the microarrays were prehybridised in hybridisation buffer at 42 °C during at least 4 hours. After the prehybridisation the arrays were washed twice in MilliQ water and once in isopropanol and dried by centrifugation (2 min at 470 x g). The hybridisation

was performed in a hybridisation frame (50 µl hybridisation buffer, Geneframe, Implen, Munich, Germany). Arrays were hybridised at 42 °C in a humid hybridisation chamber overnight. After hybridisation the arrays were washed at room temperature, successively in 1 x SSC/0.1% SDS (5 min), 0.1 x SSC/0.1% SDS (5 min) and 0.1 x SSC (1 min) and subsequently dried by centrifugation (2 min, 470 x g).

4.2.5 Scanning and data analysis

Microarrays were scanned using a confocal laser scanner ScanArray 3000 (General Scanning Inc., Pittsfield, USA) containing a GreNe 543 nm laser for Cy3 measurement and a HeNe 633 nm laser for Cy5 measurement. Scans were made with a pixel resolution of 10 micron. The software package ArrayVision (Imaging Research, Diegem, Belgium) was used for image analysis of the TIFF-files as generated by the scanner. Fluorescent signals were collected for each individual spot and stored for further data processing (dot plot analysis) in Microsoft Excel (spot selection and normalisation) and Genemaths (principle component, discrimant analysis and ANOVA) software (Applied Maths, Sint-Martens-Latem, Belgium)

4.2.6 Total protein extraction and two-dimensional electrophoresis

Tomato material was ground in liquid nitrogen and total protein was extracted from 1 g of the powder as previously described by Koistinen *et al*. [12]. The protein pellet was dissolved in 2-DE sample buffer containing 9.5 M urea, 2% (w/v) CHAPS, 1% (w/v) DTT, 0.64% (v/v) Bio-Lyte 5/7 ampholyte (Bio-Rad, Hercules, CA, USA), 0.16% Bio-Lyte 3/10 ampholyte. Total protein was analysed using the Protein Assay Dye reagent (Bio-Rad). 2-DE was performed as previously described [13]. Gel image analysis was performed with PDQuest software (Bio-Rad). Protein spot intensities were normalised to the total intensity of valid spots to minimize possible errors due to differences in the amount of protein and staining intensity. The spot quantities were log(x + 1) transformed to normalise the data, and all subsequent data analyses were performed using the SPSS software package (SPSS Inc., Chicago, IL, USA).

4.2.7 Protein identification by HPLC-electrospray tandem mass spectrometry

In-gel digestion and identification of protein spots by HPLC-ESI-MS/MS was performed as previously described [12]. The peptides were identified using the NCBI nonredundant protein database and the TIGR tomato EST database.

4.2.8 Quantitative real-time PCR

1 µg total RNA (DNase treated) was reverse transcribed to cDNA using the Bio-Rad iScript cDNA Synthesis Kit in 20 µl volume. The cDNA reaction was replenished to 100 µl with 10 mM tris-

HCL (pH7.5). 1 µl of this cDNA solution was used for subsequent Q-PCR. EST sequences were selected from the microarray experiments that showed significant changes in gene expression in subsequent ripening stages. For the design of primers for quantitative RT-PCR, the EST sequences were blasted in Blast-nr and, if negative, in BLAST-est_others in the Genbank/EMBL database. Highly homologous gene sequences were aligned with the ESTs from the microarrays in AlignX (Vector NTI Inc., InVitrogen, Breda, the Netherlands). Homologous stretches between the ESTs and Genbank sequences were used to design primers in Beacon Designer software (Premier Biosoft International, Palo Alto, USA) for the application 'SYBR Green® design' taking into account secondary structures in the template and primers, self and cross dimer formation, hairpin structures, 3' C/G clamp, nucleotide runs and repeats, annealing temperature of 60 °C ± 1 °C, primer length of 15 to 28 bases, amplicon length of 50 to 200 bp. The primers were obtained from Biolegio (Nijmegen, the Netherlands), dissolved in 10 mM tris-HCl (pH 7.5) and the concentration was determined by measuring the O.D.260 (Nanodrop™, Thermo Fisher Scientific, Waltham, USA). In Table 1 the primer sequences are shown. Two to three primer sets were designed per EST. The primer set that performed best with relation to threshold position and amplification curve was selected for gene expression analyses. The optimal primer concentration was determined. For all primers the optimal concentration was 400 nM, except for the actin of which 800 nM was applied in the quantitative PCR.

The quantitative PCRs were performed in an iCycler iQ System (Bio-Rad) using the following conditions: 1 µl cDNA, 12.5 µl 2 x iQ™ SYBR® Green Supermix (Bio-Rad), 400 nM sense primer, 400 nM antisense primer (except for the actin primers) in a total volume of 25 µl. Amplification conditions were 3' 95 °C, 40 cycles of 15'' 95 °C and 45'' 60 °C followed by a melting curve analysis where the temperature increased from 60 °C to 95 °C in increments of 0.5 °C. The analysed samples were cDNAs from the five tomato ripening stadiums (green, breaker, turning, light red and red) and a dilution range down to 10,000 times of the ripening stadium that was expected to have the highest expression level based on the microarray results. A DNase treated total RNA sample and a negative control (H_2O) were also included. Q-PCRs were developed for actin as a reference gene as the actin sequence showed similar expression during all five subsequent stages of ripening. The actin Q-PCR was based on a consensus sequence for 4 different tomato actin gene sequences and situated around an intron sequence. The PCR efficiency as well as the quantities, the relative quantities gene X/actin and standard deviations per stadium, and the quantities relative to the green stadium and standard deviation were calculated in Microsoft Excel software.

4.3 Results

4.3.1 Transcriptomics

Tomato microarrays were constructed on the basis of green- and red-specific tomato EST libraries: 2112 sequences were spotted in array format, including control sequences [9] and hybridised with labelled total RNA fractions, isolated from five subsequent ripening stages

Table 1. *Primer sequences and amplicon size of the quantitative real-time PCR experiments.*

EST	Sense primer sequence	Antisense primer sequence	Amplicon length (bp)
Tomato fruit red ripe EST300761	GCTAGGTTTTGGGC TGATTTTATTGAC	TTGGCTGCCTCTTGATCTTCTCC	91
Calmodulin 7, *Arabidopsis thaliana*	CAGGACATGATAAACGAGGTGGATG	TCAGCAGCGGAGATGAAGCC	56
Glutathione-S-transfertase/ peroxidase	GTTGGAGGGAGCACTTGGAGAG	G-AAACCAGCAGTAGAACCCAATGAG	92
Unknown protein *Arabidopsis thaliana*	CAAACTTCGGTGGACCTCATAAGC	AAAGTCACTGATTCCTTCCCTCTCC	170
Ethylene-responsive transcriptional coactivator	TGTAAGGCAAGCAATACAGAAAGCG	CCATTCTCATACTCGGCAACAACC	108
Glucuronosyl transferase homolog	GTTAAGTGAAGAAGGCAAAGATGTGAAG	AGCGAGTCGTGATGGCAAGG	145
Oxalyl- CoA decarboxylase, *Arabidopsis thaliana*	AGAGGCAATTTCATCTGAGTTGTCTG	CTCCAATTTCTTCCTCCTCATCACTG	134
Invertase	GAGCAGCACGACTCTTTGTTTTC	TCTCCCTCTTCCCTTTCTTGATGG	195

(green, breaker, turning, light red, and red), in duplicate. The results of individual hybridisation experiments were quantified and spots with a signal-to-noise ratio larger than three were used for further analysis. Normalisation was performed on the basis of a reference sample, composed of a mixture (equal ratios) of all experimental samples that were labelled with an alternative cyanide label that was included in all hybridisation experiments. In addition, normalisation over the arrays was performed to correct for large differences in overall signals between the subsequent samples. Overall expression rates were calculated (Figure 1). It shows that overall expression rates vary considerably between the different stages of ripening with the highest level for the breaker stage of ripening. Subsequently, the resulting data were analysed by PCA (principle component analysis) and ANOVA (analysis of variance). The PCA and ANOVA analyses were performed for the data set with and without normalisation over arrays. Especially for the breaker stage of ripening the application of normalisation over the arrays resulted in a, slightly, different set of stage related genes (results not shown). PCA was performed on the basis of all hybridisation experiments, as well as on the basis of selected groups of experiments with two subsequent stages of ripening. In all cases the first PCA component already showed complete separation between the ripening stages (Table 2), indicating that the stage of ripening was the largest source of variation when comparing two subsequent stages of ripening. To identify genes that were primarily responsible for the difference between two subsequent ripening stages, PCA was performed on genes that were selected by an initial ANOVA analysis (selected genes showing differential gene expression, $p<0.05$). Based on the PCA in the components that clearly separated two subsequent stages of ripening, the genes with the largest differences in gene expression were selected, and listed in Table 3. It can be seen that significant differences occur between all subsequent stages of ripening, but the largest number of genes with an altered gene expression profile are found in the transition phase from breaker to turning.

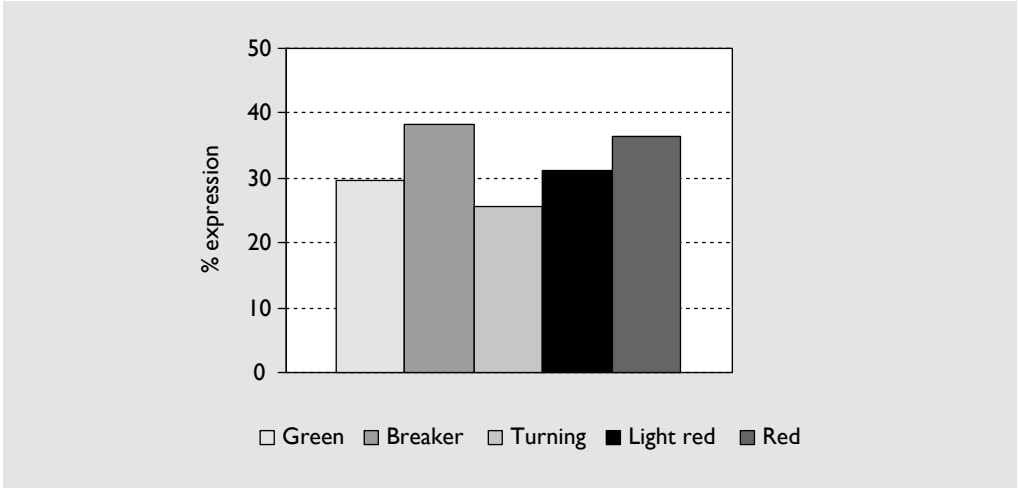


Figure 1. *Overall gene expression levels of tomato ESTs during the subsequent stages of ripening, i.e. the relative number of spots with a positive signal (signal-to-noise>3).*

Table 2. *Percentage of the variation that is explained by the 1st PCA component in the PCA between two subsequent ripening stages.*

Ripening stages	Percentage of the variation that is explained by the 1st PCA component
Green vs breaker stage of ripening	66.1%
Breaker vs turning stage of ripening	91.8%
Turning vs light red stage of ripening	95.4%
Light red vs red stage of ripening	96.3%

Table 3. *Differential gene expression in the subsequent ripening stages*[*].

Green versus breaker stage of ripening	
Down in breaker (38 ESTs)	
gi|100318| gene C-7 protein [*Nicotiana tabacum*]	
gi|11385579| In2-1 protein [*Glycine max*]	
gi|119150| elongation factor 1-alpha [*Lycopersicon esculentum*]	
gi|11994380| cucumisin-like serine protease; subtilisin-like protease [*Arabidopsis thaliana*]	
gi|123620| heat shock protein cognate 70 [*Lycopersicon esculentum*]	
gi|129939| polygalacturonase 2A precursor (PG-2A) (pectinase) [*Lycopersicon esculentum*]	RR
gi|13537555| mitogen PL-B [*Phytolacca americana*]	
gi|13549123| putative short-chain type alcohol dehydrogenase [*Solanum tuberosum*]	
gi|1370174| RAB1Y [*Lotus japonicus*]	
gi|14280354| branched-chain amino acid aminotransferase [*Capsicum annuum*]	
gi|15088626| nictaba [*Nicotiana tabacum*]	
gi|15217785| P-glycoprotein, putative [*Arabidopsis thaliana*]	
gi|15219756| putative polygalacturonase [*Arabidopsis thaliana*]	
gi|15220349| growth regulator protein [*Arabidopsis thaliana*]	
gi|15225230| hypothetical protein [*Arabidopsis thaliana*]	
gi|15225276| unknown protein [*Arabidopsis thaliana*]	
gi|15226073| putative lysosomal acid lipase [*Arabidopsis thaliana*]	
gi|15228161| eukaryotic translation initiation factor 6 (EIF-6) - like protein [*Arabidopsis thaliana*]	
gi|15231059| short-chain acyl CoA oxidase [*Arabidopsis thaliana*]	
gi|15237472| putative protein [*Arabidopsis thaliana*]	
gi|15239436| ABC transporter homolog PnATH - like [*Arabidopsis thaliana*]	
gi|15241334| putative protein [*Arabidopsis thaliana*]	
gi|287474| beta-fructosidase [*Lycopersicon esculentum*]	
gi|322743| phenylalanine ammonia-lyase - tomato	RR

Table 3. *Continued.*

Green versus breaker stage of ripening (continued)	
gi|399007| alcohol dehydrogenase [*Lycopersicon esculentum*]	RR
gi|416652| probable glutathione-S-transferase (auxin-induced protein PCNT107) [*Nicotiana tabacum*]	
gi|461812| cytochrome P450 72A1 (probable geranol-10-hydroxylase) [*Catharanthus roseus*]	
gi|529516| beta-fructosidase [*Solanum tuberosum*]	
gi|6634777| F26G16.16 [*Arabidopsis thaliana*]	
gi|6653233| glutathione S-transferase/peroxidase [*Lycopersicon esculentum*]	
gi|7433906| glucosyltransferase IS5a (EC 2.4.1.-), salicylate-induced [*Nicotiana tabacum*]	
gi|7484604| lectin-B - Virginian pokeweed	
gi|7489228| dehydrin homolog CI7 cold-stress inducible protein [*Solanum tuberosum*]	
gi|81850| histone H3 (AA 1-120) [*Medicago sativa*]	
gi|8778823| T6D22.2 [*Arabidopsis thaliana*]	
gi|25031822| hypothetical protein XP_207741 [*Mus musculus*]	
gi|4887010| Nt-gh3 deduced protein [*Nicotiana tabacum*]	
gi|671859| annexin VII [*Dictyostelium discoideum*]	
Up in breaker (25 ESTs)	
gi|100181| 1-aminocyclopropane-1-carboxylate synthase [*Lycopersicon esculentum*]	RR
gi|100230| phytoene synthase [*Lycopersicon esculentum*]	RR
gi|119290| early light-induced protein precursor [*Pisum sativum*]	
gi|119640| 1-aminocyclopropane-1-carboxylate oxidase homolog (protein E8) [*Lycopersicon esculentum*]	
gi|13174237| unknown protein [*Oryza sativa*]	
gi|14030723| [*Arabidopsis thaliana*]	
gi|14487954| early light inducible protein [*Medicago sativa*]	
gi|15218210| bZIP protein, putative [*Arabidopsis thaliana*]	
gi|15226534| putative glucosyltransferase [*Arabidopsis thaliana*]	
gi|15228031| putative glucosyl transferase [*Arabidopsis thaliana*]	
gi|15228033| putative glucosyl transferase [*Arabidopsis thaliana*]	
gi|15228034| putative glucosyl transferase [*Arabidopsis thaliana*]	
gi|15228037| putative glucosyl transferase [*Arabidopsis thaliana*]	
gi|15229253| putative O-linked GlcNAc transferase [*Arabidopsis thaliana*]	
gi|15235862| hypothetical protein [*Arabidopsis thaliana*]	
gi|15824408| steroleosin [*Sesamum indicum*]	
gi|231357| 1-aminocyclopropane-1-carboxylate synthase [*Lycopersicon esculentum*]	RR
gi|4887010| Nt-gh3 deduced protein [*Nicotiana tabacum*]	
gi|629669| glucuronosyl transferase homolog, ripening-related - tomato (fragment)	RR
gi|7484744| unknown [*Cucumis melo*]	

Table 3. *Continued.*

Green versus breaker stage of ripening (continued)	
gi|7489142| hypersensitivity-related gene [*Nicotiana tabacum*]	
gi|81896| early light-induced protein precursor - garden pea	
gi|1729887| 26 KD secreted antigen precursor [*Toxocara canis*]	
gi|18652312| putative acyltransferase [*Cucumis melo*]	
gi|7511819| cappuccino gene protein - fruit fly	
Breaker versus turning stage of ripening	
Down in turning (96 ESTs)	
gi|100181| 1-aminocyclopropane 1-carboxylate synthase [*Lycopersicon esculentum*]	RR
gi|100230| mutant phytoene synthase [*Lycopersicon esculentum*]	
gi|1076675| ubiquinol--cytochrome-c reductase (EC 1.10.2.2) Rieske iron-sulfur protein - potato	
gi|1076676| S-adenosylmethionine decarboxylase; SAMDC [*Solanum tuberosum*]	
gi|10798652| malate dehydrogenase [*Nicotiana tabacum*]	
gi|11252274| 2-oxoglutarate dehydrogenase, E1 subunit [*Arabidopsis thaliana*]	
gi|11358951| DNA-binding protein 4 [*Nicotiana tabacum*]	
gi|1173257| ribosomal protein S4 [*Solanum tuberosum*]	
gi|119150| elongation factor 1-alpha [*Lycopersicon esculentum*]	
gi|119290| early light-induced protein precursor [*Pisum sativum*]	
gi|119640| 1-aminocyclopropane-1-carboxylate oxidase homolog (protein E8) [*Lycopersicon esculentum*]	RR
gi|123650| heat shock cognate 70 KD protein [*Petunia x hybrida*]	
gi|12718824| MAPK [*Nicotiana tabacum*]	
gi|128195| nitrate reductase [*Lycopersicon esculentum*]	
gi|12831474| PAPS-reductase-like protein [*Catharanthus roseus*]	
gi|13174237| unknown protein [*Oryza sativa*]	
gi|1350878| chloroplast 30S ribosomal protein S16 [*Solanum tuberosum*]	
gi|13518432| PSII 43 KDa protein [*Lotus japonicus*]	
gi|14487954| early light inducible protein [*Medicago sativa*]	
gi|14715462| CjMDR1 [*Coptis japonica*]	
gi|15218210|r bZIP protein, putative [*Arabidopsis thaliana*]	
gi|15218726| unknown protein [*Arabidopsis thaliana*]	
gi|15218740| putative extracellular dermal glycoprotein EDGP precursor [*Arabidopsis thaliana*]	
gi|15218808| hypothetical protein [*Arabidopsis thaliana*]	
gi|15219097| putative oxidoreductase [*Arabidopsis thaliana*]	
gi|15219099| putative oxidoreductase [*Arabidopsis thaliana*]	
gi|15220409| hypothetical protein [*Arabidopsis thaliana*]	

Table 3. *Continued.*

Breaker versus turning stage of ripening (continued)
gi|15221079| unknown protein [*Arabidopsis thaliana*]
gi|15223386| hypothetical protein [*Arabidopsis thaliana*]
gi|15225276| unknown protein [*Arabidopsis thaliana*]
gi|15226073| putative lysosomal acid lipase [*Arabidopsis thaliana*]
gi|15226534| putative glucosyltransferase [*Arabidopsis thaliana*]
gi|15228031| putative glucosyl transferase [*Arabidopsis thaliana*]
gi|15228034| putative glucosyl transferase [*Arabidopsis thaliana*]
gi|15228381| lipase - like protein [*Arabidopsis thaliana*]
gi|15229761| putative cell division related protein [*Arabidopsis thaliana*]
gi|15231889| putative cytochrome P450 [*Arabidopsis thaliana*]
gi|15232212| putative protein [*Arabidopsis thaliana*]
gi|15232681| dUTP pyrophosphatase-like protein [*Arabidopsis thaliana*]
gi|15234245| putative phosphoribosylanthranilate transferase [*Arabidopsis thaliana*]
gi|15235862| hypothetical protein [*Arabidopsis thaliana*]
gi|15236442| hypothetical protein [*Arabidopsis thaliana*]
gi|15238961| putative protein [*Arabidopsis thaliana*]
gi|15241245| RAN2 small Ras-like GTP-binding nuclear protein (Ran-2) [*Arabidopsis thaliana*]
gi|15241334| putative protein [*Arabidopsis thaliana*]
gi|15242177| seed maturation -like protein [*Arabidopsis thaliana*]
gi|15451226| beta tubulin [*Arabidopsis thaliana*]
gi|15824408| steroleosin [*Sesamum indicum*]
gi|15983404| AT3g29360/MUO10_6 [*Arabidopsis thaliana*]
gi|1710840| adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) [*Nicotiana tabacum*]
gi|1915974| fructokinase [*Lycopersicon esculentum*]
gi|1934758| cytosolic glutamine synthetase [*Brassica napus*]
gi|20559| hsp70 (AA 6 - 651) [*Petunia x hybrida*]
gi|2078298| glyceraldehyde 3-phosphate dehydrogenase [*Lycopersicon esculentum*]
gi|2129926| wound-induced protein Sn-1, vacuolar membrane [*Capsicum annuum*]
gi|231687| catalase isozyme 1 [*Lycopersicon esculentum*]
gi|2369766| hypothetical protein [*Citrus x paradisi*]
gi|2492530| chloroplast aminopeptidase 2 precursor (leucine aminopeptidase) [*Lycopersicon esculentum*]
gi|2501187| thiazole biosynthetic enzyme, chloroplast precursor [*Alnus glutinosa*]
gi|279636| ubiquitin / ribosomal protein S27a [*Lycopersicon esculentum*]
gi|2982362| glutathione peroxidase [*Zantedeschia aethiopica*]
gi|2995990| dormancy-associated protein [*Arabidopsis thaliana*]
gi|3264767| AP2 domain containing protein [*Prunus armeniaca*]

Table 3. *Continued.*

Breaker versus turning stage of ripening (continued)	
gi|343516| ATPase alpha-subunit [*Nicotiana tabacum*]	
gi|3513758| phenylalanine ammonia-lyase [*Capsicum chinense*]	RR
gi|3777447|14-3-3 protein [*Solanum tuberosum*]	
gi|464981| ubiquitin-conjugating enzyme E2-17 KD (ubiquitin-protein ligase) [*Lycopersicon esculentum*]	
gi|4827251| plastidic aldolase [*Nicotiana paniculata*]	
gi|4836473| 17.6 kD class I small heat shock protein [*Lycopersicon esculentum*]	
gi|5031283 unknown [*Prunus armeniaca*]	
gi|5758306| S-adenosyl-l-homocysteine hydrolase [*Lycopersicon esculentum*]	
gi|629669| glucuronosyl transferase homolog, ripening-related - tomato (fragment)	
gi|629790| tubulin beta chain [*Oryza sativa*]	
gi|6469032| N-hydroxycinnamoyl/benzoyltransferase [*Ipomoea batatas*]	
gi|6714272| F6N18.8 [*Arabidopsis thaliana*]	
gi|6782421| DnaJ-like protein [*Lycopersicon esculentum*]	
gi|7430703| cytochrome P450 [*Catharanthus roseus*]	
gi|7436502| arginine decarboxylase (EC 4.1.1.19) ADC-1[*Nicotiana sylvestris*]	RR
gi|7446754| putative phosphate transporter [*Lycopersicon esculentum*]	
gi|7484604| lectin-B - Virginian pokeweed	
gi|7484744| probable anthranilate N-benzoyltransferase [*Cucumis melo*]	
gi|82088| histone H1-like protein - tomato (fragment)	
gi|8778823| T6D22.2 [*Arabidopsis thaliana*]	
gi|8895787| multiprotein bridging factor 1 [*Solanum tuberosum*]	
gi|9294562| contains similarity to CAF protein~gene_id:MQC12.21 [*Arabidopsis thaliana*]	
gi|9797761| contains similarity to PIR7A protein from *Oryza sativa* [*Arabidopsis thaliana*]	
gi|1174626| translationally controlled tumor protein homolog [*Solanum tuberosum*]	
gi|119150| elongation factor 1-alpha [*Lycopersicon esculentum*]	
gi|18652312| putative acyltransferase [*Cucumis melo*]	
gi|2129758| ubiquitin-protein ligase UBC7 [similarity] [*Arabidopsis thaliana*]	
gi|21362942| adenosylhomocysteinase [Lupinus luteus]	
gi|24745927| monooxygenase [*Solanum tuberosum*]	
gi|7433184| ripening protein E8 homolog [*Lycopersicon esculentum*]	RR
gi|7489029| late-embryogenesis protein homolog - tomato	
gi|7511819| cappuccino gene protein - fruit fly	
gi|9979196| translationally controlled tumor protein homolog [*Hevea brasiliensis*]	
Up in turning (25 ESTs)	
gi|100182| 1-aminocyclopropane 1-carboxylate synthase [*Lycopersicon esculentum*]	RR
gi|11994380| cucumisin-like serine protease; subtilisin-like protease [*Arabidopsis thaliana*]	

Table 3. *Continued.*

Breaker versus turning stage of ripening (continued)	
gi|129939| polygalacturonase 2A precursor (PG-2A) (pectinase) [*Lycopersicon esculentum*]	RR
gi|13430456| putative pectate lyase [*Arabidopsis thaliana*]	
gi|13492674| phenylpropanoid:glucosyltransferase 1 [*Nicotiana tabacum*]	
gi|13537555| mitogen PL-B [*Phytolacca americana*]	
gi|1370174| RAB1Y [*Lotus japonicus*]	
gi|14029184| peroxidase [*Manihot esculenta*]	
gi|14594815| putative phosphoenolpyruvate carboxylase kinase [*Beta vulgaris*]	
gi|15081688| AT5g14180/MUA22_18 [*Arabidopsis thaliana*]	
gi|15219756| putative polygalacturonase [*Arabidopsis thaliana*]	
gi|15225276| unknown protein [*Arabidopsis thaliana*]	
gi|15227152| putative CONSTANS-like B-box zinc finger protein [*Arabidopsis thaliana*]	
gi|15229339| putative protein [*Arabidopsis thaliana*]	
gi|15231046| putative protein [*Arabidopsis thaliana*]	
gi|15241334| putative protein [*Arabidopsis thaliana*]	
gi|15810407| putative endomembrane EMP70 precursor isolog protein [*Arabidopsis thaliana*]	
gi|1709692| peptide methionine sulfoxide reductase (fruit-ripening protein E4) [*Lycopersicon esculentum*]	RR
gi|416652| probable glutathione S-transferase (auxin-induced protein PCNT107) [*Nicotiana tabacum*]	
gi|5669634| ethylene-responsive transcriptional coactivator [*Lycopersicon esculentum*]	RR
gi|6653233| glutathione S-transferase/peroxidase [*Lycopersicon esculentum*]	
gi|7484604| lectin-B - Virginian pokeweed	
gi|7489142| hypersensitivity-related gene [*Nicotiana tabacum*]	
gi|225933| endopolygalacturonase	
gi|25031822| hypothetical protein XP_207741 [*Mus musculus*]	
Turning versus light red stage of ripening	
Down in light red (17 ESTs)	
gi|1076676| S-adenosylmethionine decarboxylase; SAMDC [*Solanum tuberosum*]	
gi|1110548| lectin-C, PL-C [*Phytolacca americana*]	
gi|119640| 1-aminocyclopropane-1-carboxylate oxidase homolog [*Lycopersicon esculentum*]	RR
gi|13549123| putative short-chain type alcohol dehydrogenase [*Solanum tuberosum*]	
gi|14487954| early light inducible protein [*Medicago sativa*]	
gi|15229784| calmodulin 7 [*Arabidopsis thaliana*]	
gi|15236437| hypothetical protein [*Arabidopsis thaliana*]	
gi|1709692| peptide methionine sulfoxide reductase (fruit-ripening protein E4) [*Lycopersicon esculentum*]	RR

Table 3. *Continued.*

Turning versus light red stage of ripening (continued)	
gi|2129926| wound-induced protein Sn-1, vacuolar membrane [*Capsicum annuum*]	
gi|4887010| Nt-gh3 deduced protein [*Nicotiana tabacum*]	
gi|5332353| phenylalanine ammonia lyase [*Lycopersicon esculentum*]	RR
gi|544011| basic 30 KD endochitinase precursor [*Lycopersicon esculentum*]	
gi|5669634| ethylene-responsive transcriptional coactivator [*Lycopersicon esculentum*]	RR
gi|586076| tubulin beta-1 chain [*Lupinus albus*]	
gi|6048743| chitinase [*Brassica juncea*]	
gi|629669| glucuronosyl transferase homolog, ripening-related - tomato (fragment)	RR
gi|6630683| EST C25991(C11435) [*Oryza sativa*]	
Up in light red (18 ESTs)	
gi|129939| polygalacturonase 2A precursor (pectinase) [*Lycopersicon esculentum*]	RR
gi|14030723| unknown protein [*Arabidopsis thaliana*]	
gi|14487954| early light inducible protein [*Medicago sativa*]	
gi|15219408| chlorophyll a oxygenase [*Arabidopsis thaliana*]	
gi|15220349| growth regulator protein [*Arabidopsis thaliana*]	
gi|15224843| hypothetical protein [*Arabidopsis thaliana*]	
gi|15235399| putative glycosyltransferase [*Arabidopsis thaliana*]	
gi|15824408| steroleosin [*Sesamum indicum*]	
gi|1708059| glutathione reductase, chloroplast precursor [*Nicotiana tabacum*]	
gi|629669| glucuronosyl transferase homolog, ripening-related - tomato (fragment)	RR
gi|7387848| 2-isopropylmalate synthase A [*Lycopersicon pennellii*]	
gi|7484744| probable anthranilate N-benzoyltransferase [*Cucumis melo*]	
gi|15224670| hypothetical protein [*Arabidopsis thaliana*]	
gi|15240947| putative protein [*Arabidopsis thaliana*]	
gi|18416995| putative protein [*Arabidopsis thaliana*]	
gi|24745927| monooxygenase [*Solanum tuberosum*]	
gi|384332| invertase	
gi|4887010| Nt-gh3 deduced protein [*Nicotiana tabacum*]	
Light red vs red stage of ripening	
Down in red (13 ESTs)	
gi|124701| beta-fructofuranosidase precursor - vacuolar invertase precursor [*Lycopersicon esculentum*]	
gi|14029184| peroxidase [*Manihot esculenta*]	
gi|14030723| unknown protein [*Arabidopsis thaliana*]	
gi|14487954| early light inducible protein [*Medicago sativa*]	

Table 3. *Continued.*

Light red vs red stage of ripening (continued)	
gi|15220349| growth regulator protein [*Arabidopsis thaliana*]	
gi|4887010| Nt-gh3 deduced protein [*Nicotiana tabacum*]	
gi|6688560| putative ferredoxin [*Lycopersicon esculentum*]	
gi|15221783| protein translation factor Sui1 homolog, putative; protein [*Arabidopsis thaliana*]	
gi|15224670| hypothetical protein [*Arabidopsis thaliana*]	
gi|15240947| putative protein; protein [*Arabidopsis thaliana*]	
gi|18416995|| putative protein [*Arabidopsis thaliana*]	
gi|20069537| papillary renal cell carcinoma protein [*Xenopus laevis*]	
gi|384332| invertase	
Up in red (24 ESTs)	
gi|100182| 1-aminocyclopropane-1-carboxylate synthase [*Lycopersicon esculentum*]	RR
gi|115473| carbonic anhydrase, chloroplast precursor [*Nicotiana tabacum*]	
gi|12641619| histone H3 [*Beta vulgaris*]	
gi|129939| polygalacturonase 2A precursor (pectinase) [*Lycopersicon esculentum*]	RR
gi|14280354| branched-chain amino acid aminotransferase [*Capsicum annuum*]	
gi|15218726| unknown protein [*Arabidopsis thaliana*]	
gi|15219097| putative oxidoreductase [*Arabidopsis thaliana*]	
gi|15219465| unknown protein [*Arabidopsis thaliana*]	
gi|15229784| calmodulin 7 [*Arabidopsis thaliana*]	
gi|15236437| hypothetical protein [*Arabidopsis thaliana*]	
gi|15238323| putative protein [*Arabidopsis thaliana*]	
gi|2129926| wound-induced protein Sn-1, vacuolar membrane [*Capsicum annuum*]	
gi|4902525| carbonic anhydrase [*Glycine max*]	
gi|544011| basic 30 KD endochitinase precursor [*Lycopersicon esculentum*]	
gi|585746| phytoene synthase 1 precursor (fruit ripening specific) [*Lycopersicon esculentum*]	RR
gi|6634777| F26G16.16 [*Arabidopsis thaliana*]	
gi|7433906| glucosyltransferase IS5a (EC 2.4.1.-), salicylate-induced - [*Nicotiana tabacum*]	
gi|7489228| dehydrin homolog CI7 [*Solanum tuberosum*]	
gi|119640| 1-aminocyclopropane-1-carboxylate oxidase homolog [*Lycopersicon esculentum*]	RR
gi|120559| fruit-specific protein [*Lycopersicon esculentum*]	
gi|21592804| ribosomal protein, putative [*Arabidopsis thaliana*]	
gi|225933| endopolygalacturonase	
gi|3378652| CaM-1 [*Nicotiana plumbaginifolia*]	
gi|7433184| ripening protein E8 homolog [*Lycopersicon esculentum*]	RR

* RR = known from literature to be ripening-related.

4.3.2 Proteomics

Total proteins from tomatoes of the five subsequent stages of ripening, with three replicates of each stage, were analyzed by 2-DE. As most protein spots appeared in the pI range of 4-7 (data not shown), this gradient was chosen. Proteins outside of this pI range were therefore not detected in this study. 740-1500 protein spots were resolved and detected in individual gels, and 655 of these could be quantified reproducibly across the gels. In PCA, the ripening stages were separated in the first component, which accounted for 26.7% of variation in the 2-DE data set, indicating that the stages of ripening were the largest source of variation variation also in this dataset. 53 individual spots were found differentially expressed during ripening. Changes in the expression of a spot were considered significant, when the difference in spot intensity between two or more stages was statistically significant ($p<0.05$ in the Kruskal-Wallis test). The expression profiles of individual protein spots varied, but an overall increase during ripening was detected in 26 spots, a decrease was seen in 27 spots, and two spots reached their maximum at the breaker or light red stage. The rest of the spots remained fairly constant or their expression varied in individual tomatoes from the same ripening stage. Differentially expressed spots were further characterised by HPLC-ESI-MS, and several of them were tentatively identified (Figure 2, Table 4). In some cases, the identification was based on similar proteins in other plant species or tomato EST sequences. For example, the EST sequences to which spot 1117 was matched, TPRAA46TV and TPRAA46TH, were found to be similar to 'thioredoxin peroxidases from various species'.

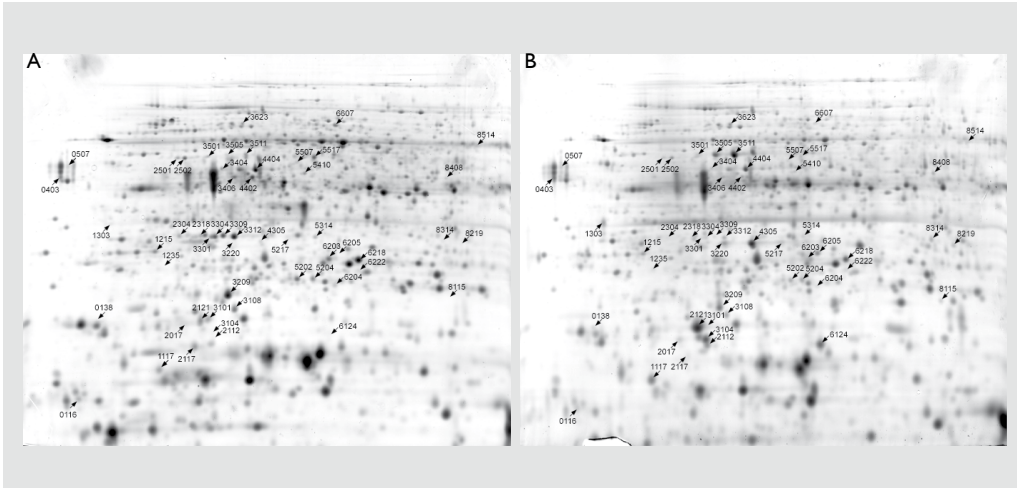


Figure 2. *Two-dimensional gel electrophoresis images of tomato fruit protein from two ripening stages: A, green; B, red. Differentially expressed proteins are marked with arrows and numbers that correlate with protein identifications in Table 3.*

Table 4. *Tomato proteins identified with HPLC-ESI-MS (corresponding proteins are shown in Figure 2).*

Spot nr	Protein	Accession nr	Apparent MW (kDa)[1]	MW (kDa)[2]	pI[3]	Peptides identified	Expression profile[4]
0138	EST	TRZDB68TH	20.0			5 (AGEESGSTTELFASAK APAAAPVAEEAK LDETQGIGQYVEK VAEAAADVLDAAQK VLADAAQSQFNK)	Increase
0403	Suberization-associated anionic peroxidase 1 precursor	P15003	47.0	38.7	4.9	2	Increase
0507	Suberization-associated anionic peroxidase 1 precursor	P15003	48.0	38.7	4.9	1	Increase
1117	EST	TOFAE86THB	18.0			3 (VLNLEEGGAFNVSSA HLPGFVEK YAMLVDDGVVK)	Increase
1215	EST	TFBBK29TH	30.0			6 (FGSQAAPAGAAPAK LASSFPGK LASSFPGKAVGVR LVPVGYGIK SVNDHLSGK TYISGDQLTK)	Decrease
1303	Suberization-associated anionic peroxidase 1 precursor	CAA33852	38.0	38.7	4.9	4	Increase
2112	Acid beta-fructofuranosidase precursor	P29000	19.5	70.1	5.5	3	Increase
2121	Acid beta-fructofuranosidase precursor	P29000	20.0	70.1	5.5	1	Increase
2501	26S protease regulatory subunit 6 homolog	P54776	52.0	47.5	4.9	2	Decrease
2502	26S protease regulatory subunit 6 homolog	P54776	52.0	47.5	4.9	6	Decrease
3104	Acid beta-fructofuranosidase precursor	P29000	19.5	70.1	5.5	6	Increase
3108	Putative mitochondrial ATP synthase (*Solanum demissum*)	AAT40531	21.0	19.8	5.3	4	Decrease
3209	Oxygen-evolving enhancer protein 2	P29795	23.0	27.7	8.3	9	Decrease
3301	Oxygen-evolving enhancer protein 1	P23322	33.0	34.9	5.9	5	Decrease

3304	Oxygen-evolving enhancer protein 1	P23322	33.0	34.9	5.91	7	Decrease
3309	Oxygen-evolving enhancer protein 1	P23322	33.0	34.9	5.91	4	Decrease
3312	Oxygen-evolving enhancer protein 1	P23322	33.0	34.9	5.91	10	Decrease
3404	Actin 51/41	Q96482, Q96483	45.9	37.1	5.46	2+1	Decrease
3406	Ribulose bisphosphate carboxylase/ oxygenase activase (*L. pennellii, A. thaliana*)	O49074, G96756	43.5	50.7	8.61	2+1	Decrease
3505	Acid invertase	AAB30874	52.0	70.1	5.6	2	Increase
3511	Acid invertase	AAB30874	52.0	70.1	5.6	5	Increase
3623	Beta-galactosidase	T04340	65.0	80.5	9.0	2	Increase
4305	Acid invertase	AAB30874	35.0	70.1	5.6	5	Increase
4402	Ribulose bisphosphate carboxylase/ oxygenase activase (*L. pennellii, A. thaliana*)	O49074, G96756	43.5	50.7	8.61	1+1	Decrease
4404	Actin 41	Q96483, AAL66196	45.0	37.1	5.3	2+6	Decrease
5517	S-adenosyl-L-methionine synthetase	CAA80865	50.0	43.3	5.5	2	Decrease
6124	Polygalacturonase-2a	CAA01256	19.5	50.1	6.4	3	Increase
6203	Ascorbate peroxidase	AAX84654	30.0	27.3	5.9	3	Decrease
6205	EST	TPRAN02TH	31.0			2 (LKPLLEAAGHK VTALDLAASIGDLR)	Increase
6218	Ascorbate peroxidase	CAB58361, BAA12918	30.0	27.4	5.7	7+2	Decrease
6607	D-3-phosphoglycerate dehydrogenase, chloroplast precursor (*A. thaliana*)	O04130	65.0	66.5	5.8	2	Decrease
8219	EST	TFCAY75TH	32.0			4 (ALGMQVVK GGSSVIAFVK QPGSIPGLNTK VVNLAIQELGGK)	Increase

[1] Estimated from the 2-DE gel

[2,3] Calculated from the theoretical protein sequence; not applicable to EST sequences

[4] The overall expression pattern of the protein spot during ripening

4.3.3 Quantitative PCR

For confirmation of the microarray results primers for 8 quantitative real time RT-PCRs (Q-PCRs) were developed (Table 1). The selection of the sequences for confirmational experiments was based on the criteria that the sequence is annotated and that the gene expression profile had to have distinctive features over the subsequent ripening stages based on the microarray experiments. Some of the PCRs were developed to confirm altered gene expression profiles in the transition between different subsequent ripening stages. In this way 11 observed changes in microarray-based gene expression profiles were assessed on the basis of confirmatory Q-PCR experiments. The results of the confirmative Q-PCR experiments are shown in Figure 3. Actin was selected out of the stable expressors as a reference. For the actin Q-PCR, a concensus sequence over all known tomato actin genes was used. In eight of eleven Q-PCR analyses the microarray results were confirmed. In two cases the decrease from the green to the breaker stage of ripening could not be confirmed. In one case, the decrease from the turning to the light red stage of ripening could not be confirmed.

4.4 Discussion

4.4.1 Transcriptomics and proteomics

Microarray-based gene expression analysis (transcriptomics) and two-dimensional electrophoresis (2-DE)-based proteome analysis (proteomics) have the potential to screen many metabolic routes simultaneously for alterations in gene expression and protein levels. In this way altered profiles can be detected that may be relevant for the toxicological and/or nutritional assessment of newly developed plant varieties. To test this approach, tomato was used as a model species. For transcriptomics, RDA-based (representational difference analysis) [14] microarrays (app. 2000 sequences) were obtained, containing sequences specific for the red and green stage of ripening, respectively [9]. The red-specific EST-library is assumed to contain ESTs that are related to the nutritional and perhaps even health-protecting properties of the tomato, the green-specific EST library is assumed to consist in part of sequences that are related to antinutritional metabolic routes. For proteomics over 600 protein spots were monitored.

As a first step in the assessment of the validity of the approach, it was assessed whether it is possible to develop simple criteria for the sampling of the ripening tomatoes for the purpose of a comparative safety assessment. At the same time a first insight into the relevance of observed changes in gene expression and protein profiles could be obtained. Total RNA was isolated from the green and the red, ripe stage of ripening, as well as from the three intermediary ripening stages. In 'breaker' the first yellowing of the fruit is observed. This stage occurs at approximately 10 days after the mature green stage. 'Turning' is characterized by the first colouring with orange tints and in 'light red' the whole tomato fruit is just full orange coloured. The final stages of ripening, from the breaker stage to the red-ripe stage of ripening, also take place in approximately 10 days, with increasingly short intervals between the subsequent

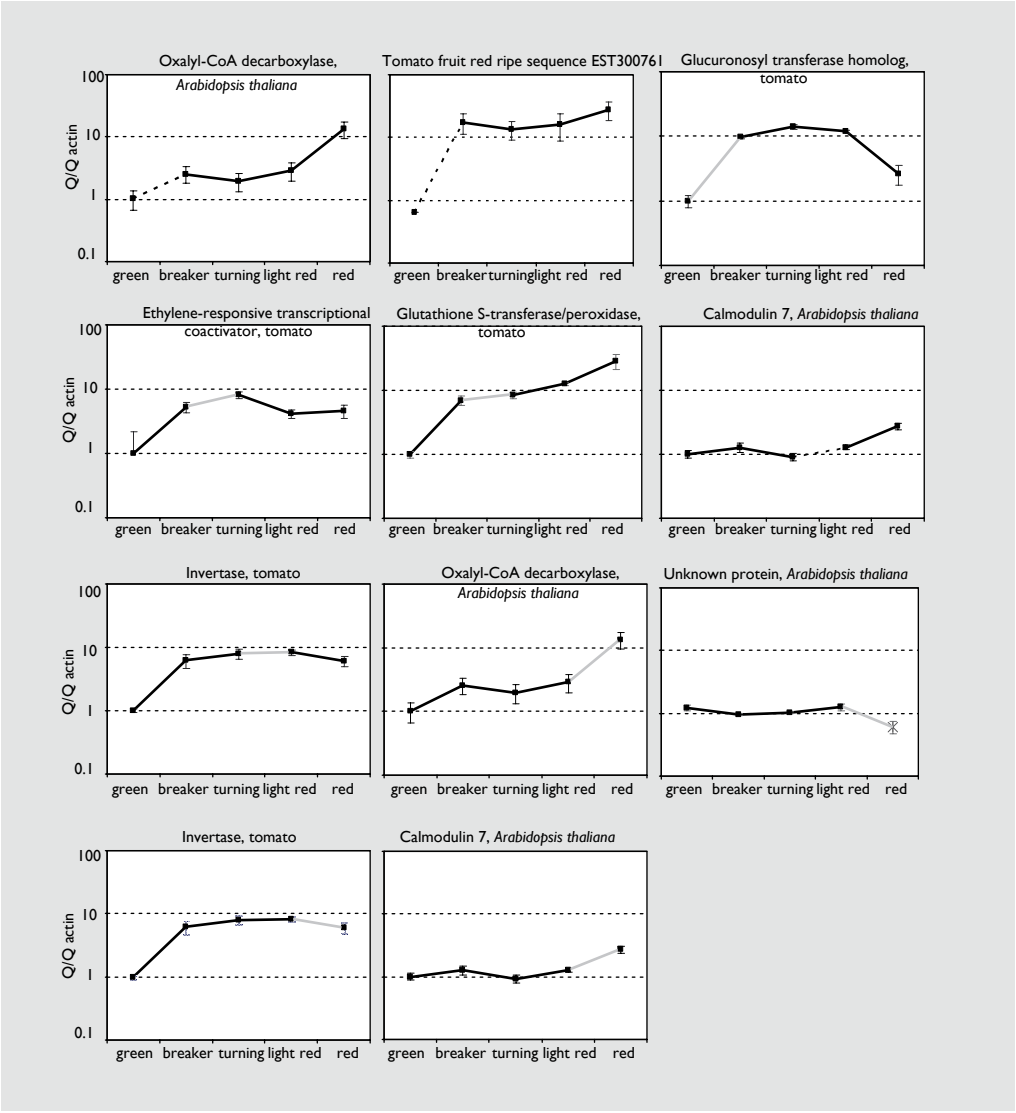


Figure 3. *Quantitative real-time PCR results of differentially expressed genes in the subsequent stages of ripening, based on the microarray hybridization results: differential expression identified by micro array analysis is indicated by a grey line when confirmed and by a dotted line when not confirmed.*

stages [15]. Total RNAs were used for transcriptomics. To obtain similar proteomic information protein isolates were obtained from the other half of the tomatoes and comparison of the resulting profiles was performed.

In all cases the stage of ripening was the largest source of variation between samples, based on the observation that a clear separation was obtained between the ripening stages in the first component of the PCA (Table 2).

The principle PCA results to identify the most relevant genes for altered expression profiles from the comparisons of two subsequent ripening stages are shown in Table 3. In general, it can be concluded that the most pronounced differences in gene expression are observed between the green and the breaker stage and breaker and turning. Differences between the later stages of ripening are clear but less pronounced. This is also reflected in the overall expression levels for the spotted EST fragments: 38% overall expression for the breaker stage, compared to around 30% in other ripening stages (Figure 1). From the light red to the red stage of ripening also an increase in overall expression level can be observed, which is interpreted as an increase in metabolic activity. This is especially the case, as can be expected, for EST sequences derived from the the red-tomato enriched EST-library. It should, however, be taken into account that from earlier experiments [9] as well as from literature [16,17], it is known that in the green stage the number of moderately expressed genes may be larger resulting in limited numbers of mRNA molecules available for hybridisation per spot, while in the red stage of ripening less genes are likely to produce more RNA sequences abundantly, resulting in higher average signal values per spot.

The same tomatoes in the subsequent ripening stages were analysed for changes in proteome composition by 2-DE. Here out of the 655 protein spots that were analysed further, 53 spots were found to be differentially expressed during ripening. An overall increase during ripening was detected in 26 spots, a decrease was seen in 27 spots, and two spots reached their maximum at the breaker or light red stage. Due to technical limitations of 2-DE it is likely that only the most abundant proteins in 2-DE from the selected pI and MW range were visible. In the current experiment the number of replicates was low and this may influence the number of differentially expressed proteins, as smaller changes may not end up being significant because of the variation between replicates.

4.4.2 Significance of observed differences

To confirm observed differences in the transcriptome Q-PCR experiments were performed. 8 Q-PCRs were developed to allow analysis of 11 observed differences in subsequent stages of ripening, of which 8 were confirmed (Figure 3). In two cases, the decrease from the green to the breaker stage of ripening could not be confirmed. In these two cases it is, however, likely that the over the array data correction may have lead to a misinterpretation of the data. The data correction assumes that overall expression levels are constant. However, the breaker stage is metabolically very active and the increased overall signal may be a reflection of increased overall expression rather than a difference that is due to technical aspects that needs to be corrected for. The Q-PCR experiments in these two cases are in line with the array results, when the second, over the array correction is not applied. In one case a difference observed in the

turning to the light red stage of ripening could not be confirmed, but from the quantitative PCR experiments it seems clear that the difference between the two stages is very limited. It should be noted that for the Q-PCR experiments total RNA from different tomatoes was used than those used for the hybridisation experiments. In view of this it is reassuring that most gene expression differences could be confirmed showing the relative robustness of phenotypic scoring of the ripening stage.

When comparing the proteomics results with the transcriptomics data there is only one identified agreement: acid beta-fructofuranosidase is analysed in both 'omics' approaches and found to be upregulated in gene expression in the breaker stage, downregulated in the subsequent turning and light red stages and then once again upregulated in the red stage of ripening. In the proteome analysis it is found to be overall upregulated during the five subsequent stages of ripening (Figure 4). While polygalacturonase 2a mRNA increased between the green and breaker stages, the quantity of the protein spot remained low until it increased sharply from turning to

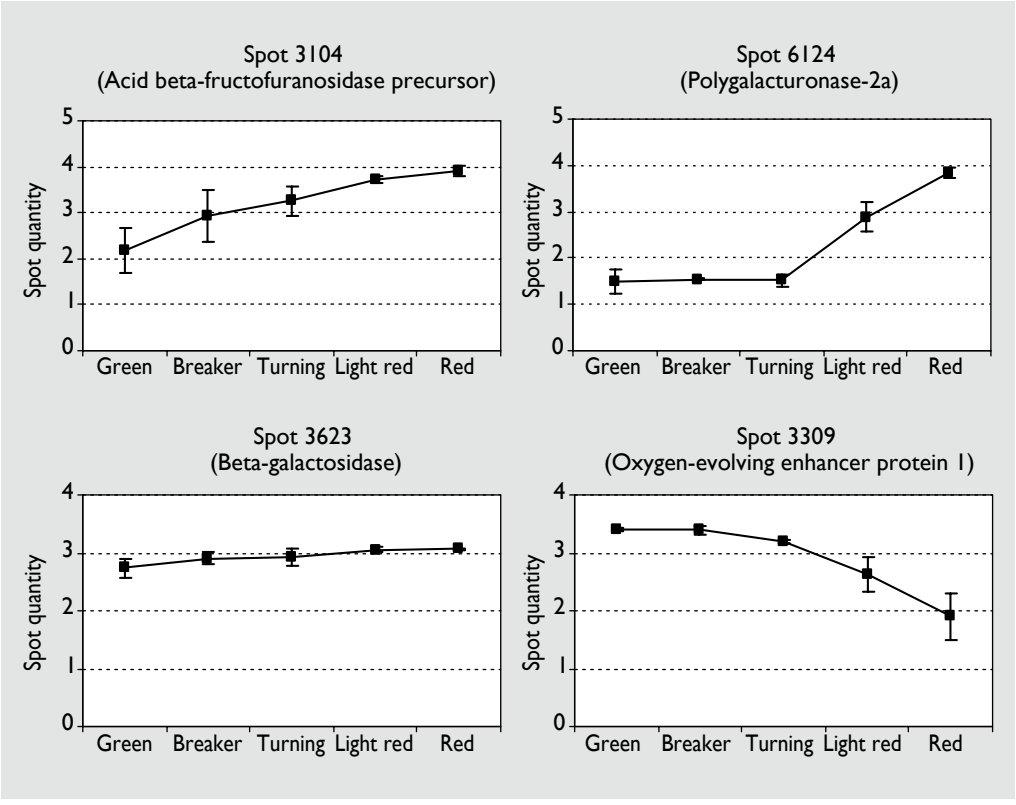


Figure 4 *Protein quantity patterns of selected protein spots during the subsequent tomato ripening stages. The quantity of the spots is shown as the mean of log(x+1) transformed spot quantity ± SD. The spot numbers correlate with protein identifications in Figure 2 and Table 3.*

light red and red stages. No other clear agreements are seen here. This is not surprising, as for proteome analysis not the same selection was made with respect to metabolic pathways as for transcriptomics. When looking at the large numbers of transcripts expressed per cell or tissue, estimates varying from 30.000-120.000 and the, probably, even larger number of resulting proteins, it is still a chance event when the data match, considering the limited numbers of different transcripts and proteins included in this study. Furthermore, there will be a delayed-phase effect between the transcriptome and proteome in developmental stages, as the gene expression profiles will change before a changed protein profile in the same cell system is seen [18,19]. For this type of analysis, it can be concluded that, for the time being, the data from transcriptomics and proteomics are likely to be complementary rather than overlapping.

4.4.3 Ripening stages

When comparing the results from these experiments with information as can be found in the scientific literature, it is found to be very much in agreement. This information is limited and is mostly linked to economically important metabolic routes, such as cell wall degradation during the final stages of ripening and, more recently, metabolic routes that are related to the production of nutritional compounds such as carotenoids and flavonoids. One example is phytoene synthase, an important enzyme in the synthesis of lycopene, a nutritionally relevant carotenoid that is formed in the tomato fruit [20]. Phytoene synthase is also upregulated in the red stage of ripening in our microarray analysis. However, we found the largest increase in the breaker stage of ripening. This may be explained by the fact that phytoene synthase can occur in a soluble, inactive form as well as in a membrane-bound, active form, which may explain that the highest increase in mRNA production is already seen in the breaker stage of ripening, whereas the highest level of lycopene formation is likely to be in the final, red stages of ripening [21]. Polygalacturonase, ACC (1-aminocyclopropane-1-carboxylate) synthase and ACC oxidase, that are ripening-related [21], are also upregulated during the breaker/turning and red stages of ripening. Giovannoni [21] also described the regulation of a ripening protein E8 homolog in the red stage of ripening, which is confirmed by the hybridization results. Our data also agree with Moore *et al*.,[16] who described a large increase in the breaker/turning versus green for phytoene synthase and polygalacturonase-2α. In agreement with published data [16, 22] the cell wall degrading enzymes 'putative pectate lyase (*Arabidopsis thaliana*)' (transcriptomics) and beta-galactosidase (proteomics, Figure 4) were increased during fruit maturation in the proteomics analysis. The arginine decarboxylase gene is also described to peak at the breaker stage [23], which is in line with the microarray results. Phenylalanine ammonia-lyase is seen to decrease in the breaker, turning and light red stage of ripening, which is fully in line with its described activity in glycoalkaloid biosynthesis which is known to decrease in the ripening tomato [24,25]. In the same way the protein profile of oxygen-evolving enhancer protein 1 shows a steady decrease from the breaker to the red stage of ripening (Figure 4). This protein is active in the oxygen evolution complex of photosystem II and thus it can be explained that it decreases in ripening tomatoes that show a decrease in photosynthesis.

4.4.4 Omics-studies for food safety assessment

This study has shown that the set-up of an 'omics' study is of crucial importance. It is essential that the samples under scrutiny are well-matched with relation to environmental conditions during growth, harvest and storage as is stipulated in, for instance, the FAO/WHO [4] and European (EFSA, [2]) guidance documents for the risk assessment of GM plants, and this includes matching the samples with relation to the stage of development of the tissues to be analysed. Failing to do so is likely to result in detected differences that are unrelated to the differences in genotype that are investigated. At the same time it was shown that it will be difficult to determine the stage of ripening accurately as there is a clear increase in metabolic activity from, for instance, the light red to the red stage of ripening. In practice this may hamper the applicability of the approach as it may be necessary to perform larger numbers of analyses of control plants in order to better catch the natural variation in the transcriptome and/or proteome in the harvested tissues. Other studies have confirmed these observations [26]. This will, however, also hold true for targeted studies and/or studies of the metabolome in developing fruits. Alternatively, the 'omics' strategies may be applied primarily for screening of large differences in gene and/or protein expression in newly developed plant varieties. This study has shown that ripening-related changes that are relevant for the nutritional value as well as the toxicological charasteristics of the tomato fruit can be monitored using 'omics' approaches. Examples of this are the upregulation of genes that are involved in carotenoid and flavonoid synthesis as well as the downregulation of a gene that is involved in natural toxin production, respectively. The combination of 'omical' screening of new plant varieties (in an early stage of the breeding procedures) and targeted analyses on key compounds with relation to the toxicologically and nutritionally relevant metabolic pathways, conform EFSA guidelines, will make it highly unlikely that undesired or detrimental effects of any breeding process will remain unnoticed. This study also shows that profiles in the less mature stages of ripening can supply additional information as another set of ESTs and/or proteins will be assessed. This may furthermore prove helpful for the interpretation of the data in the final stages of ripening. When applied in this way the holistic 'omics' approaches can be used to make an overall estimate of differences in gene expression and protein formation, and have the potential to pick up differences that may not have been picked up by current targeted analysis. The direct analysis of the proteome is likely to become more relevant in that respect, but for the time being transcriptomics has shown its value by making it feasible already at this stage to include many different metabolic pathways in a single analysis. For further use in quality and safety assessment strategies of novel plant varieties the experimental protocols need to be further refined and validated.

Acknowledgements

This research was financially supported by the European Commission (GMOCARE project QLRT-1999-00765) and the Dutch Ministry of Agriculture, Nature and Food Quality. The authors thank Seppo Auriola (Department for Pharmaceutical Chemistry, University of Kuopio) for his work on mass spectrometry analysis.

References

1. European Union (2003). Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. Official Journal of the European Union L268/1: p1-23.
2. European Food Safety Authority (2004). Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants and derived food and feed, EFSA Journal 99: 1-94, http://www.efsa.eu.int/science/gmo/gmo_guidance/660/guidance_docfinal1.pdf
3. Organisation for Economic Cooperation and Development (1996). Food Safety Evaluation. Report Workshop held in Oxford UK, September, 12-15, 1994, ISBN 92-64-14867-1.
4. Food and Agriculture Organisation / World Health Organisation (2000). Safety aspects of genetically modified foods of plant origin. Report of a Joint FAO/WHO Expert Consultation on Foods derived from Biotechnology, World Health Organization, Geneva, Switzerland.
5. Kuiper H.A., Kleter G.A., Noteborn H.P.J.M., and Kok E.J. (2001). Assessment of the food safety issues related to genetically modified foods. Plant Journal 27(6): 503-528.
6. Kok E.J., and Kuiper H.A. (2003). Comparative Safety Assessment for Biotech Crops. Trends in Biotechnology 21(10): 439-444.
7. GMOCARE (2004). New methodologies for assessing the potential of unintended effects in genetically modified food crops, http://www.entransfood.nl.
8. Food Standards Agency (2005). Safety assessment of genetically modified foods research programme, http://www.food.gov.uk/science/research/researchinfo/foodcomponentsresearch/novelfoodsresearch/g02programme.
9. Kok E.J., Franssen-van Hal N.L.W., Winnubst L.N.W., Kramer E.H.M., Dijksma W.T.P., Kuiper H.A. and Keijer J. (2007). 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. Journal of Plant Physiology 164: 337-349.
10. Schena M., Shalon D., Heller R., Chai A., Brown P.O., and Davis R.W. (1996). Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proceedings of the National Academy of Sciences USA 93:10614-19.
11. Franssen-van Hal N.L.W., Vorst O., Kramer E.H.M., Hall R.D., and Keijer J. (2002). Factors influencing cDNA microarray hybridisation on silylated glass slides. Analytical Biochemistry 308(1):5-17.
12. Koistinen K.M., Hassinen V.H., Gynther P.A.M., Lehesranta S.J., Keinänen S.I., Kokko H.I., Oksanen E.J., Tervahauta A.I., Auriola S., Kärenlampi S.O., and Birch P.R. (2002). 10c is induced by factors causing oxidative stress but appears not to confer tolerance to these agents. New Phytologist 155 (3), 381-391.

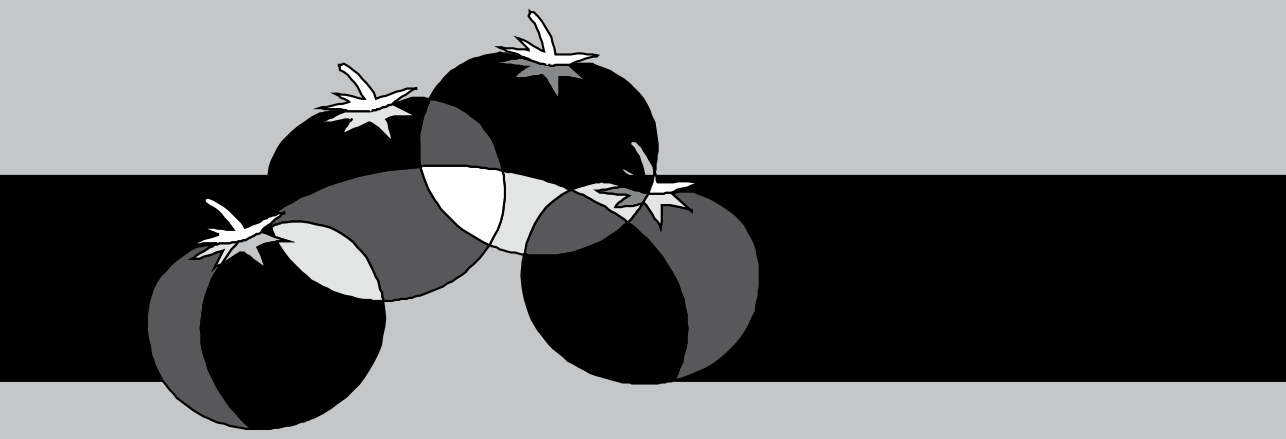
13. Lehesranta S.J., Davies H.V., Shepherd L.V.T., Nunan N., McNicol J.W., Auriola S., Koistinen K.M., Suomalainen S., Kokko H.I., and Kärenlampi S.O. (2005). Comparison of tuber proteomes of potato (Solanum sp.) varieties, landraces and genetically modified lines. Plant Physiology 138: 1690-1699.
14. Lisitsyn N., Lisitsyn N., and Wigler M. (1993). Cloning the differences between two complex genomes. Science 259(5097) :946-51.
15. Sargent S. (2000). Ripening tomatoes with ethylene. Department of Horticultural Sciences VC-29 Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, http://edis.ifas.ufl.edu/CV206#beginning.
16. Moore S., Vrebalov J., Payton P., and Giovannoni J. (2002). Use of genomic tools to isolate key ripening genes and analyse fruit maturation in tomato. Journal of Experimental Botany, 53(377) Fruit Development and Ripening Special Issue: 2023-2030.
17. Alba R., Payton P., Fei Z., McQuinn R., Debbie P., Martin G.B., Tanksley S.D., and Giovannoni J.J. (2005). Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. Plant Cell 17: 2954-2965.
18. Gallie D.R. (2002). Protein-protein interactions required during translation. Plant Molecular Biology 50:949-970.
19. Kaufman R.J. (2004). Regulation of mRNA translation by protein folding in the endoplasmic reticulum. Trends in Biochemical Sciences 29(3): 152-158.
20. Bramley P.M. (2002). Regulation of carotenoid formation during tomato fruit ripening and development. Journal of Experimental Botany 53 (377): 2107-2113.
21. Giovannoni J. (2001). Molecular biology of fruit maturation and ripening. Annual Review of Plant Physiology and Plant Molecular Biology 52: 725-49.
22. Marin-Rodriguez M.C., Orchard J., and Seymour G.B. (2002). Pectate lyases, cell wall degradation and fruit softening. Journal of Experimental Botany 53(377) Fruit Development and Ripening Special Issue: 2115-2119.
23. Rastogi R., Dulson J., and Rothstein S.J. (1993). Cloning of tomato (*Lycopersicon esculentum* Mill.) arginine decarboxylase gene and its expression during fruit ripening. Plant Physiology 103: 829-834.
24. Lee S.W., Robb J., and Nazar R.N. (1992). Truncated phenylalanine ammonia-lyase expression in tomato (*Lycopersicon esculentum*). Journal of Biological Chemistry 267(17):11824-30.
25. Ramamoorthy V., Raguchander T., and Samiyappan R. (2002). Induction of defense-related proteins in tomato roots treated with Pseudomonas fluorescens Pf1 and Fusarium oxysporum f. sp. Lycopersici. Plant Soil 239: 55-68.
26. Metzdorff S.B., Kok E.J., Knuthsen P., and Pedersen J. (2006). Evaluation of a non-targeted 'omic' approach in the safety assessment of genetically modified plants. Plant Biology 8:662-672.

Chapter 5.

Abstract

In the present study the use of transcriptomics as a tool in the comparative safety assessment (CSA) of genetically modified plant varieties was investigated. To this end dedicated arrays with ripening related genes were used, thereby focussing indirectly on nutritional and antinutritional ingredients. Transcriptome profiles were obtained from two different genetically modified tomato transformant lines, GM7 and GM8, transformed with the same genetic construct. The gene expression changes in the transformant lines were small. Compared to wild-type (WT), the majority of the differentially expressed genes was analogously changed in the two transformant lines; of the detected differences in GM7, 82% was also differentially expressed in GM8. Most changes were in ripening related genes, in agreement with the functional role of the inserted sequence.

Of all known ripening related genes present on the array, 37% and 48 % were differently expressed for GM7 and GM8, respectively, suggesting the limited extent of the modification also in the light of the intended alterations. These changes are in nutritional and health-protective metabolic routes, rather than in routes linked to anti-nutritional compounds. All together, the data show that differential gene expression in two GM transformant lines is limited, especially when the known natural variation in tomato lines is taken into account.



Food safety analysis of genetically modified tomato lines using an 'omics' approach

Esther J. Kok[1], Jeroen P. van Dijk[1], A. M. Angeline Van Hoef[1], Paul D. Fraser[2], Peter M. Bramley[2], Arnaud G. Bovy[3], Michael R. Muller[4], Ivonne M.C.M. Rietjens[5], Harry A. Kuiper[1] and Jaap Keijer[1]

[1]RIKILT Institute of Food Safety, Bornsesteeg 45, P.O. Box 230, 6700 AE Wageningen, the Netherlands
[2]School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, United Kingdom
[3] Plant Research International, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA Wageningen, the Netherlands
[4]Wageningen University, Department of Human Nutrition, Bomenweg 2, P.O. Box 8129, 6700 EV Wageningen, the Netherlands
[5] Wageningen University, Department of Toxicology, Tuinlaan 5, P.O. Box 8000, 6700 EA Wageningen, the Netherlands

Plant Biotechnology Journal, submitted.

5.1 Introduction

In recent years food and feed safety assessment strategies have been developed for the specific group of new plant varieties that have been developed using gentechnological approaches. The resulting plant varieties are generally addressed as genetically modified organisms (GMOs) or genetically modified plants (GMPs). The strategies for their safety assessment are based on a comparative safety assessment (CSA) [1,2,3,4] that uses the plant varieties already on the market, with a history of safe use, as the basis for the safety assessment. An important cornerstone of this assessment is an elaborate compositional analysis, comprising the analysis of the crop-related key nutrients and anti-nutritional factors. This approach has, however, its limitations, especially in the area of micronutrients and antinutrients, including natural toxins, in terms of available validated methodology and of knowledge of background levels. Only already well-identified end-points of metabolic routes (e.g. vitamins, flavonoids, known natural toxins or groups of toxins) will be assessed in order to detect potential alterations in the toxicological and/or nutritional characteristics of the new plant variety. Furthermore, in the approval process of new plant varieties from other types of breeding practices, no such elaborate compositional studies are generally required.

Already a number of years ago the importance of the development of new tools for a more 'holistic' characterisation of the new varieties was recognized [1,3]. Since that time different European projects have focused on the development of the so-called 'omics'-approaches as a tool to gain increased insight into potential unintended alterations in the plant's physiology due to the breeding process, including the genetic modification, that are outside of the natural variation in the crop plant species. One of the largest projects in this area was the European project 'New methodologies for assessing the potential of unintended effects in genetically modified food crops' (GMOCARE) that started in 2000, under the umbrella of the European Entransfood network project on the safety assessment of genetically modified food crops [5]. The tomato was selected as one of the model species. One of the 'omics'-methodologies that has been applied within the project, the results of which are discussed in this paper, is transcriptomics. To test the potential of transcriptomics as a tool for obtaining relevant data in the safety analysis of newly developed (GM) plant varieties, gene expression profiles were obtained from GM tomato lines as well as from the WT parent variety. To this end RDA-based (representational difference analysis) [6] EST-libraries were used, specific for the red and the green stage of ripening, respectively, to construct a food safety-related microarray [7]. The red-specific EST-library is assumed to contain ESTs that are related to the nutritional and health-protecting properties of the tomato, whereas the green-specific EST library is assumed to consist in part of sequences that are related to antinutritional metabolic routes and potentially toxic metabolites that are known to be high in the unripe tomato and decline upon ripening [8,9]. The resulting array, although comprising gene sequences from a broad network of metabolic routes [7] can therefore be considered to be a focused array with gene sequences related to tomato ripening and to relevant nutritional and toxicological pathways.

In the present study transcriptomics profiles were obtained from two genetically modified tomato lines, transformant lines GM7 and GM8, that have been transformed with the same genetic construct. GM7 and GM8 were modified for increased β-carotene production by incorporation of the bacterial phytoene desaturase gene *crt*I under the control of a constitutive CaMV 35S promoter sequence and the pea-derived ribulose biphosphate carboxylase small subunit (SSU) transit sequence for correct processing within the cell. A kanamycin resistance gene was used as a selective marker gene in the genetic construct. The bacterial phytoene desaturase (*Erwinia uredovora*) converts phytoene directly into lycopene that is subsequently transformed by cyclisation into α- and β- carotene, whereas in the plant two different desaturases, phytoene desaturase and zeta-carotene desaturase, are required for this transition. The result is a tomato with elevated relative levels of provitamin A, β-carotene, amounting to 40-55% of the tomato carotenoids, but an approximately two-fold reduction in total carotenoids [10,11]. Lycopene, phytofluene and phytoene were decreased 3-, 7- and 12-fold, respectively, whereas lutein levels were relatively increased. Additionally, production of alpha-carotene, bisdehydrolycopene, zeaxanthin, neoxanthin, and antheraxanthin were observed in the transformant lines and not in the WT lines. Vitamin E levels were increased in the transformant lines. Phenotypically the only observed difference was the fruit colour that varied from red to orange, as a result of increased β-carotene levels [11].

Here, differences in the gene expression profiles of the two GM lines compared to the WT variety have been analysed with relation to the (intended) genetic modification and to similarities and differences in gene expression in the two transformant lines. Similarities in differential gene expression may point at direct effects of the altered metabolic routes, whereas differences in gene expression profiles between the two GM lines may indicate somaclonal or insertional effects or different levels of secondary effects (dosage effects) as a result of the introduced gene sequence. The obtained results were also evaluated based on established gene expression profiles of the subsequent stages of ripening in tomato in order to exclude the differences to be related to slight differences in ripening stage of the samples rather than to the genetic modification as such. The potential of transcriptomics to contribute to the safety evaluation of genetically altered tomato varieties is discussed.

5.2 Results

Dedicated tomato microarrays focusing on transcripts derived from food safety related metabolic networks were used to generate the transcriptomics data for 26 hybridisations with Ailsa Craig WT tomato control samples, 17 hybridisations with tomato samples from 'high β-carotene expressor' line GM7 and 17 hybridisations with tomato samples from 'high β-carotene expessor' GM8. Using PCA, the components that clearly separated either of the GM lines from the wild type control were selected, and the contributing genes with the largest differences in gene expression were identified. A further selection was made on the basis of T-testing of the results for either of the two GM lines vs those for the WT. The resulting differentially expressed genes are listed in Table 1. The observed differences in expression are generally small. The largest

Table 1. *Overview of differentially expressed genes in GM7 and GM8 (ANOVA p< 0.05).*

Sequence
Down regulation in GM line 7/8 vs WT
gi|15232681| dUTP pyrophosphatase-like protein [*Arabidopsis thaliana*]
gi|2129926| wound-induced protein Sn-1, vacuolar membrane [*Capsicum annuum*]
gi|1135895| DNA-binding protein 4 [*Nicotiana tabacum*]
gi|124701| beta-fructofuranosidase precursor - vacuolar invertase precursor [*Lycopersicon esculentum*]
gi|15236437| hypothetical protein [*Arabidopsis thaliana*]
gi|7436502| arginine decarboxylase [*Nicotiana sylvestris*]
gi|13430456| putative pectate lyase [*Arabidopsis thaliana*]
gi|15236442| hypothetical protein [*Arabidopsis thaliana*]
gi|7387848| 2-isopropylmalate synthase [*Lycopersicon pennellii*]
gi|461812| cytochrome P450 72A [*Catharanthus roseus*]
gi|7258518| exterior membrane protein gp120 [Human immunodeficiency virus 1]
gi|23483659| hypothetical protein [Plasmodium yoelii yoelii]
gi|15234173| putative hypoersensitive response protein [*Arabidopsis thaliana*]
gi|13492674| phenylpropanoid:glucosyltransferase 1 [*Nicotiana tabacum*]
gi|15228381| lipase - like protein [*Arabidopsis thaliana*]
gi|15242177| seed maturation -like protein [*Arabidopsis thaliana*]
gi|100318| gene C-7 protein [*Nicotiana tabacum*]
gi|3913001| MADS box protein homolog, transcription factor [*Solanum tuberosum*]
gi|416652| auxin-induced protein [*Nicotiana tabacum*]
gi|15228037| putative glucosyl transferase [*Arabidopsis thaliana*]
gi|6688560| putative ferredoxin [*Lycopersicon esculentum*]
gi|585746| phytoene synthetase [*Lycopersicon esculentum*]
gi|7433906| salicylate-induced glucosyltransferase [*Nicotiana tabacum*]
gi|231357| 1-aminocyclopropane-1-carboxylate synthase [*Lycopersicon esculentum*]
gi|7489142| hypersensitivity-related gene [*Nicotiana tabacum*]
gi|3061271| NPCA1 [*Nicotiana paniculata*]
gi| 39933020| Solanum tuberosum L374 gene
gi| 47105748| Lycopersicon esculentum clone 133606F
gi| 124222132| Solanum lycopersicum cDNA, clone FC10DE01
gi|12049590| Oxalyl-CoA decarboxylase [*Arabidopsis thaliana*]
gi|119640|1-aminocyclopropane-1-carboxylate oxidase homolog, E8 protein [*Lycopersicon esculentum*]
gi|7406669| putative ripening-related protein [*Vitis vinifera*]
gi|24582637| CG7093 gene product [*Drosophila melanogaster*]
gi|18413291| lipase-like protein [*Arabidopsis thaliana*]
gi|6625560| beta-glucan binding protein [*Phaseolus vulgaris*]
gi|15219465| unknown protein [*Arabidopsis thaliana*]
gi|7263070| putative water channel protein [*Lycopersicon esculentum*]

Ripening related (RR)[1] Green tomato-specific (G)	GM7 and GM8	GM 7 -fold change	GM 7 -stat significance	GM 8 -fold change	GM 8 -stat significance
RR	X	-1,75	4,96E-04	-1,51	3,36E-03
RR	X	-1,65	8,55E-07	-1,49	6,70E-01
RR	X	-1,61	3,74E-05	-2,10	3,54E-07
RR	X	-1,60	8,67E-07	-1,39	1,62E-02
RR		-1,58	1,94E-03		
RR	X	-1,58	1,56E-02	-3,01	9,27E-07
RR	X	-1,56	1,39E-08	-1,22	1,83E-03
RR	X	-1,56	9,02E-03	-3,89	4,81E-13
RR	X	-1,55	1,45E-04	-1,30	3,84E-03
	X	-1,54	9,97E-05	-1,99	9,71E-10
	X	-1,53	4,75E-02	-1,33	3,58E-02
	X	-1,50	8,36E-05	-1,37	7,16E-03
	X	-1,40	4,00E-02	-2,50	1,95E-09
RR	X	-1,39	8,94E-04	-1,36	1,94E-03
RR	X	-1,37	3,64E-02	-1,39	1,46E-02
RR	X	-1,37	4,26E-03	-1,58	2,51E-03
	X	-1,36	1,43E-02	-1,52	1,56E-03
	X	-1,35	4,12E-02	-1,65	2,11E-03
RR		-1,34	2,26E-03		
	X	-1,29	4,47E-02	-1,52	2,51E-03
RR	X	-1,28	8,75E-03	-1,39	3,63E-04
RR		-1,27	1,75E-03		
RR	X	-1,27	1,11E-02	-1,40	3,31E-04
RR[2]	X	-1,26	4,63E-02	-1,66	3,99E-03
RR	X	-1,25	2,83E-02	-1,32	1,95E-03
		-1,24	1,86E-02		
				-3,11	*1,99E-08*
				-2,92	*5,87E-07*
				-2,85	*2,66E-11*
G				-2,66	*6,34E-07*
RR				-2,51	1,79E-03
RR				-2,50	9,90E-08
G				-2,36	1,09E-06
				-2,17	2,27E-06
G				-2,06	4,48E-05
RR				-2,04	2,76E-05
				-1,93	5,72E-05

Table 1. *Continued.*

Sequence
gi|120559| fruit specific protein [*Lycopersicon esculentum*]
gi|100302| auxin-induced protein - common tobacco
gi|629669| glucuronosyl transferase homolog, ripening-related - tomato (fragment)
gi|23019847| hypothetical protein [*Thermobifida fusca*]
gi|7484744| probable anthranilate N-benzoyltransferase [*Cucumis melo*]
gi|15231785| unknown protein [*Arabidopsis thaliana*]
gi|18423187| 11-beta-hydroxysteroid dehydrogenase-like [*Arabidopsis thaliana*]
gi|15238914| indole-3-glycerol phosphate synthase [*Arabidopsis thaliana*]
gi|6630683| catalyzing the hydroxylation of phenazine-1-carboxylic acid [*Oryza sativa*]
gi|15241205| 11-beta-hydroxysteroid dehydrogenase-like [*Arabidopsis thaliana*]
gi|7331143| chaperonin 21 precursor [*Lycopersicon esculentum*]
gi|100182| 1-aminocyclopropane-1-carboxylate synthase [*Lycopersicon esculentum*]
gi|15238202| cytochrome P450 - like protein [*Arabidopsis thaliana*]
gi|15224153| C2H2 zinc finger protein FZF [*Arabidopsis thaliana*]
Up regulation in GM line 7/8 vs WT
gi|82088| histone H1-like protein - tomato (fragment)
gi|13958032| polygalacturonase [*Pisum sativum*]
gi|15241334| putative protein [*Arabidopsis thaliana*]
gi|1110548| lectin-C [*Phytolacca americana*]
gi|15233218| unknown protein [*Arabidopsis thaliana*]
gi|23603058| cadherin, EGF LAG seven-pass G-type receptor 2 [*Mus musculus*]
gi|279636| ubiquitin / ribosomal protein S27a [*Lycopersicon esculentum*]
gi|15239735| thiazole biosynthetic enzyme precursor [*Arabidopsis thaliana*]
gi|24745927| monooxygenase [*Solanum tuberosum*]
gi|15220224| unknown protein [*Arabidopsis thaliana*]
gi|225933| endopolygalacturonase
gi|129587| phenylalanine ammonia-lyase [*Lycopersicon esculentum*]
gi|15233976| extensin-like protein [*Arabidopsis thaliana*]
gi|15220590| hypothetical protein [*Arabidopsis thaliana*]
gi|11385579| In2-1 protein [*Glycine max*]
gi|12229671| Potential phospholipid-transporting ATPase 6 [*Arabidopsis thaliana*]
gi|25404278| probable polygalacturonase [*Arabidopsis thaliana*]

[1]Based on Kok *et al*., 2008.
[2]Based on Olsen *et al*., 1991, Xie *et al*., 2006.
[3]Based on Lee *et al*., 1992; Ramamoorthy *et al*., 2002.

Ripening related (RR)[1] Green tomato-specific (G)	GM7 and GM8	GM 7 -fold change	GM 7 -stat significance	GM 8 -fold change	GM 8 -stat significance
RR				-1,91	1,95E-04
				-1,82	5.73E-04
RR				-1,64	2,47E-05
				-1,63	2,01E-03
RR				-1,60	1,82E-07
				-1,59	1,82E-02
				-1,52	9,78E-07
G				-1,48	1,33E-02
RR				-1,46	4,56E-04
				-1,40	7,63E-03
				-1,40	2,99E-02
RR				-1,40	1,76E-02
				-1,38	2,24E-02
G				-1,37	1,33E-02
RR	X	2,40	1,09E-03	1,52	7,25E-03
RR	X	1,74	3,06E-02	1,20	3,29E-02
	X	1,38	2,52E-02	1,60	5,99E-03
RR	X	1,28	4,29E-02	3,79	3,10E-04
	X	1,24	2,03E-02	1,26	4,09E-02
		1,23	3,62E-02		
RR		1,20	2,72E-02		
				4,94	9,04E-03
RR				1,81	3,81E-03
				1,78	1,56E-02
RR, G				1,66	1,50E-03
RR[3], G				1,49	5,50E-03
				1,42	1,66E-03
				1,38	9,17E-03
				1,34	1,01E-02
				1,31	6,90E-03
				1,30	1,22E-02

differences appeared in the comparison between the WT and GM8, both in terms of fold-changes as well as statistical significance. The differentially expressed genes were compared in both transformant lines (Table 2). The larger part of the differentially expressed genes, compared to WT, was analogous in the two transformant lines, in GM7 these amounted to 27 out of 33 differentially expressed genes (82%) and in GM8 to 27 out of 61 differentially expressed genes (44%). To assess the toxicological relevance of any observed difference, differentially expressed genes were also compared to the selection of genes on this array that were previously identified as ripening-related, i.e. to be subject to alteration in the development from the turning stage of ripening (the stage in the normal developing tomato the first orange colouring is observed) to the red, ripe stage of ripening [12] (Table 2). In addition, literature data were included to annotate ripening-related sequences [13,14,15,16]. On the basis of these data it was observed that 18 of the 27 common differentially expressed genes were ripening-related. When looking at the differentially expressed gene sequences in the individual GM lines 67% and 48% were ripening-related in GM7 and GM8, respectively. Only a limited number of non-ripening-related genes were differentially expressed between the transformant lines and WT: 2 for GM7 and 24 for GM8, respectively. Here also the fold changes are small. To further investigate whether this differential expression was indeed biologically significant, gene expression of a selection of these differentially expressed genes in GM7 or GM8 was analysed on the basis of available data from other comparative transcriptomics arrays. Unfortunately, no data were available from experiments with different wild-type varieties that had used both the same array and reference mRNA population. Therefore data were selected derived from experiments that included WT tomato varieties and had used a reference that was unlikely to have influenced the resulting data significantly. The results are shown in Figure 1. Although no definite conclusions can be drawn, the data seem to be comparable, as no experimental group shows aberrant results for the

Table 2. *Overview of the characteristics of the differentially expressed genes in GM7 and GM8.*

	Total number	In both GM lines		Ripening-related genes		GM-line specific, not ripening-related	
		Nr	%	Nr	%	Nr	%
GM line 7 vs WT, upregulation	7	5	71%	4	57%	1	14%
GM line 7 vs WT, downregulation	26	22	85%	18	69%	1	4%
GM line 8 vs WT, upregulation	15	5	33%	6	40%	7	46%
GM line 8 vs WT, downregulation	46	22	48%	23	50%	17	37%

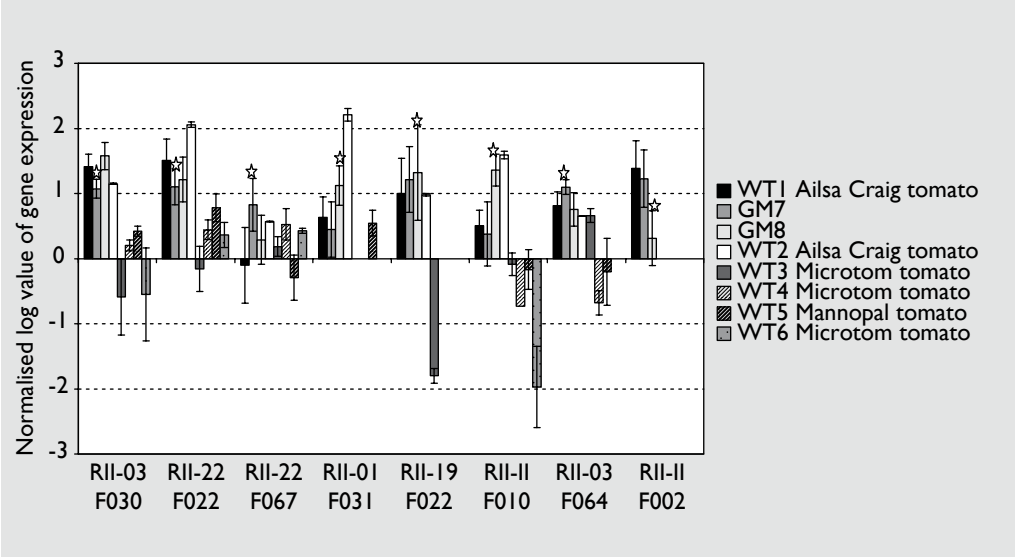


Figure 1. *Gene expression values in GM transformant and wild-type tomatoes. Gene expression values for differentially expressed genes in one or both of the GM transformant lines (GM7 and GM8) vs WT1 (☆ indicates statistical significance), compared to gene expression profiles of the same sequences in different WT varieties from other comparable transcriptomics experiments, in a first attempt to estimate the biological significance of detected differences.*

selected genes as such. Furthermore, it is shown that the variation in these genes in different WTs is considerable, the variation between different varieties by far exceeding the variation between different samples within the WT that in some cases represent technical controls and biological controls in others. These results furthermore indicate that the observed differential gene expression in the GM7 and GM8 lines as compared to the WT tomato variety is likely to fall within the bandwidth of natural variation in most if not all cases. These comparisons underline the necessity to come to common gene expression platforms and associated reference cRNA populations in order to be able to interpret data from transcriptomics analysis of GM varieties more efficiently.

For further confirmation of the results quantitative PCR of selected genes, that either showed relatively large fold changes in the microarray data and/or were of interest on the basis of their assigned function, were performed (Table 3). For the tomato-derived gene sequences of alcohol acyl transferase, early light inducible protein, and vacuolar invertase (β-fructofuranosidase precursor), downregulation on the basis of the microarray results was confirmed. For a cDNA sequence that was identified as a putative protein in *Arabidopsis thaliana* (gi|15242071) the downregulation in GM8 was also confirmed. In the case of the gene sequence that shows homology with a tomato 11-β-hydroxysteroid dehydrogenase-like protein and that was downregulated in both GM7 and GM8, a tendency was shown towards downregulation, but

Table 3. *Primer sequences for the confirmative PCR experiments.*

	Genbank access nr.	Description	(anti-)sense primer	primer sequence	Amplicon length (bp)
1	BT013646	*Lycopersicon esculentum*, mRNA sequence.	BT013646-1465s BT013646-1634as	CAAACTTCGGTGGACCTCATAAGC AAAGTCACTGATTCCTTCCCTCTCC	170
2	M81081	TOMACIN, Tomato (vacuolar) acid invertase	TOMACIN-1815s TOMACIN-2010as	GAGCAGCACGACTCTTTGTTTTC TCTCCCTCTTCCCTTTCTTGATGG	195
3	A24194	*L.esculentum* polygalacturonase	A24194-556s A24194-803as	GAGGAGGAGGAACTATCAATGGCAATG GACTCCATCAGTATTTGGGCTCTTTG	247
4	AY552528	*Lycopersicon esculentum* 18S ribosomal RNA gene	18S-28s 18S-137as	ACCGCCCGTCGCTCCTAC AATGATAAGGTTCAATGGACTTCTCGC	109
5	S36691	*Lycopersicon esculentum* phytoene desaturase mRNA	CrtI 1689s CrtI 1899as	ATGGTTTGACAGAAAACTGAAGAACAC CGTTGCTAGTTCCTTCATCGTTGC	211
6	AY534531	*Lycopersicon esculentum* alcohol acyl transferase mRNA	TomAAT-812s TomAAT-943as	CCATTGCTCTTAATTTGCACTCTGATG CAACAACTGGAGTAATAAACGCATTCC	132
7	AW040030	EST28252I tomato mixed elicitor, 11-beta-hydroxysteroid dehydrogenase-like protein	AW040030-436s AW040030-492as	ACCAGATATGGACTTGCCGTAGG CATCAGAACACGAACAAGCCAGAC	57
8	BT014305	*Lycopersicon esculentum* early light inducible protein (ELIP) gene	ELIP-230s ELIP-321as	GTGTTAGTGTTAAGTGTATGGCTGAGG AGGTGTTGGCTTGGGAACTGTAG	92
9	Actin	Consensus sequence derived from: U60478, U60480, U60481, U40482	Act-291a Act-425as	ACCGAGAGAAGATGACCCAGATTATG TCACACCATCACCAGAGTCCAAC	135

the expression was too low in both the GM and control lines to obtain technically and hence statistically valid results. The same was the case for a gene sequence that shows homology with a IH109 gene *Iris hollandica*. For the abundant polygalacturonase gene sequences multiple spots were present on the array despite the subtractive PCR procedure to obtain normalised libraries. Most of the spots showed a tendency towards downregulation in both GM7 and GM8, but overall the results were ambivalent. The PCR for polygalacturonase matched the sequence of a selection of the polygalacturonase spots present on the array. When looking only at those spots, statistically significant downregulation was observed in a limited number of cases, 4 out of 7 spots for GM8 and 1 out of 7 spots for GM7. The quantitative PCR results confirm these ambivalent results as downregulation is observed in both GM7 and GM8 as compared to WT, but for both lines the downregulation is not statistically significant (results not shown).

Finally, a single control experiment was performed to quantify the expression of the plant phytoene desaturase in the GM lines vs the control plants. No sequence for tomato phytoene desaturase was present on the array and therefore no other gene expression data were available for this gene. The quantitative PCR experiment showed downregulation of the tomato phytoene desaturase in the GM lines compared to the WT, average ratios for GM7 and GM8 versus WT being -1.41 and -1.56, respectively. This downregulation may be the result of a negative feedback mechanism as a result of the expression of the bacterial phytoene desaturase gene in the GM lines, but further experiments would be required to shed further light on this. In Römer *et al*. [11] and Fraser *et al*. [10] it was already shown that the phytoene synthase -1 was downregulated in the crtI transformants (1.69 +/- 0.046, compared to WT). On the basis of the microarray experiments a downregulation of -1.28 +/- 0.317 for the plant phytoene synthase was found in GM7. In GM8 no statistically significant downregulation is observed.

5.3 Discussion

To assess the usefulness of transcriptomics as a methodology that may be helpful in identifying relevant changes in the plant's physiology as a result of genetic modification, we obtained in the present study gene expression profiles for two transformant lines, i.e. two genetically modified tomato lines that have incorporated the same genetic construct, as well as for the parent WT variety. The advantage of this double transformant line-approach is that it allows to better discriminate between the direct and/or secondary effects of the expression products of the introduced sequence and the unintended effects that are for the larger part not directly related to the nature of the introduced genetic construct and its expression products. Direct and construct-related secondary effects will occur similarly in both transformant lines, as both have incorporated the same genetic construct. On the other hand, the unintended side effects that are the result of other effects will probably not match between the two lines. These other effects can be, for instance, somaclonal variation, i.e. alterations in the genetic make-up of plants that have gone through one or several tissue culture steps, or insertional effects, i.e. mutations that are related to the insertional event: genes may be interrupted or otherwise affected by the insertion and this may also have consequences for the physiology

of the plant. A third option is that there are different levels of secondary effects in the two transformant lines as a result of different transcription or translation levels of the introduced gene sequence.

In the present study WT Ailsa Craig tomatoes and two homozygous genetically modified 'high β-carotene expressor' lines, GM7 and GM8, were analysed for differential gene expression. GM8 shows more statistically significant differentially expressed genes compared to Ailsa Craig WT than does GM7. When comparing the detected differences in both transformant lines the resemblance is remarkable: of the detected differences in GM7, 82% was also differentially expressed in GM8. This constituted 44 % of the differentially expressed genes in GM8, as more differential gene expression is observed in GM8. When focusing on the downregulated genes (being the majority) the similarity between differentially expressed genes in GM7 and GM8 is even higher, 22 out of 26 downregulated genes for GM7 (85%) and 22 out 46 downregulated genes for GM8 (48%). Looking at the total numbers of genes represented on the array and the limited numbers showing statistically significant differences in gene expression in both lines, it may be assumed that the coinciding effects for GM7 and GM8 versus WT are directly related to the intended physiological changes introduced by the genetic construct. This is even more likely as the majority, over two-third of the sequences being differentially expressed in both GM lines, is known to be ripening-related and the genetically altered metabolic pathway is a ripening-related pathway. It may be argued that the type of array used here may lead to an array-related bias, but in earlier experiments only approximately 3% of the selected sequences were shown to vary largely in the final stages of ripening and this 3% selection of all sequences represented on the array was used as the basis for comparison in this study. When considering that 3% of all array sequences were identified as ripening-related, the observation that 67% and 48% of the differentially expressed genes were ripening-related in GM7 and GM8, respectively, provides very strong indications that this differential gene expression is indeed directly related to the introduced genetic alteration. The observed ripening-related differential gene expression patterns are in line with the data from Römer *et al*. [11] showing that in both GM lines substantial alterations have occurred in the pathways leading to the formation of carotenoids and other isoprenoids, i.e. ripening-related metabolic networks. From their data it can be seen that other physiologically linked metabolic routes may have been changed, perhaps to a lesser extent, as well.

In earlier transcriptomics experiments it was shown that clear differences in gene expression profiles between samples that differed in stage of ripening could in most cases easily be confirmed by quantitative real-time PCR [12]. Similarly quantitative PCR experiments were performed in this study to confirm the most distinctly observed differences between the GM lines and the wild type control samples. Differences were, however, much smaller in this study compared to the previous study on ripening. Nevertheless, in those cases where it proved feasible to develop good quality quantitative PCRs the results confirmed the microarray data.

In theory, an explanation for the observed differences between the GM and WT lines, that are ripening related, could also be that the sampling was not performed in exactly the same state of maturation, which is indeed more difficult to assess in the transformant lines because their colouring is different from WT control tomatoes. However, for these experiments all tomatoes were harvested at the same post anthesis stage. And as by far most of the ripening-related differential gene expression patterns occur in both transformant lines, it seems more plausible that in this case the differential gene expression is a direct result of the inserted gene sequence and its expression products, rather than the consequence of differences in sampling procedures.

For GM7 only 1 of the identified upregulated and 1 of the identified downregulated gene sequences, that is not also differentially expressed in GM8, is not known to be ripening-related. For GM line 8 these numbers are 7 and 17, respectively. These are very low numbers, when taking the average natural variation in gene expression due to internal and external conditions into account. It is likely that many of these observed differences will fall within the band width of natural variation if more WT tomatoes would be analysed, as the fold-changes are quite limited. The tentative comparison with transcriptomics data from other WT varieties seems to confirm this as well. On the other hand, these numbers may represent the second category of effects that may result from the insertional event or from somaclonal effects, as they do not occur in both transformant lines. If so, it seems from these results that these effects have been very small in these two transformation events. At the same time it can not be excluded that this differential gene expression that is observed in one of the GM lines, but not in the other, is also construct-related, but that there are different levels of secondary effects of the gene products in both transformant lines, leading to effects in GM8 that are not statistically significant in GM7. As earlier experiments have shown that the methodology can pick up major changes in the plant's physiology, it seems plausible, on the basis of these results, to assume that no major effects, including the intended effects, have occurred in both GM lines. When focusing on the largest observed differences that are only observed in GM line 8 and are not known to be ripening-related, four cDNA sequences stand out that combine an approximately three-fold change with larger statistical significance (Table 2, in italics). Only one of these sequences is functionally identified, i.e. a gene coding for oxalyl-CoA decarboxylase in *Arabidopsis thaliana*, a basic enzyme in the glyoxylate and dicarboxylate metabolism. Finally, it was analysed which of the differentially expressed genes was derived from the green tomato-specific cDNA library (Table 2): the green-specific cDNAs may comprise anti-nutrient related genes, as tomato antinutrients, for instance tomatin, are specific for the green stage of ripening. From this analysis it can be seen that 7 gene sequences, that are only differentially expressed in GM8, are derived from the green-specific cDNA library. Only one of these, phenylalanine ammonia-lyase, is involved in the synthesis of anti-nutritional factors [13,15]. Others relate to basic metabolism (oxalyl-CoA decarboxylase), defense mechanisms (beta-glucan binding protein, [17], auxin biosynthesis (indole-3-glycerol phosphate synthase, [18] or have unknown functions. Here also, the fold changes are small and in most cases with low statistical significance. Under other conditions then the strictly environmentally controlled conditions in the experiments

described here, it will probably be difficult to distinguish any of the observed differences in the background of natural variation. Support for this view can be derived from the earlier study on natural variation in the final stages of ripening [12] as well from other transcriptomics studies. In a transcriptomics study to compare conventionally bred and GM wheat lines Baudo *et al*. [19] showed that more differences are observed in conventionally bred new varieties compared to similar GM varieties. There are also studies on the application of other 'omics' technologies on (GM) novel plant varieties. A study linked to the study by Baudo *et al*. analysed the metabolome of similar wheat lines using NMR and GC-MS [20]. This work showed that the differences observed between the GM lines and their parent counterparts fall within the range of natural variation in the control lines as harvested in subsequent years. Other metabolomics studies have shown similar results [21, 22, 23]. Lehesranta *et al*. [24] analysed protein profiles using two-dimensional gel electrophoresis proteomics, for GM as well as a range of different non-GM potato varieties. They also concluded that there was less variation between GM lines and their non-GM counterparts compared to the natural variation they observed in the different WT potatoes.

The safety assessment of new (GM) plant varieties includes the identification and characterisation of hazards associated with intended and unintended alterations in the plant's physiology. Transcriptomics results can never form the sole basis for the food safety assessment of these new varieties, but the results in this and earlier papers show that the methodology can be helpful to broadly assess the extent and nature of both intended and unintended physiological changes. This manuscript shows that the use of different transformant lines can be a further help to distinguish changes that are directly related to the expression products of the introduced gene sequence and changes that are more likely related to other mutational or secondary effects, which may or may not be linked to the genetic modification per se. For the further evaluation of the toxicological and nutritional relevance of observed differences the differential gene expression patterns will subsequently need to be assessed in the light of available knowledge on the natural variation in similar plant varieties that are already on the market. All together the data obtained show that differential gene expression in two GM transformant lines is relatively limited, especially when taking into account the known natural variation in tomato lines. Furthermore changes detected in the GM transformant lines were primarily reflecting changes also observed in the ripening-related processes linked to nutritional and health-protective metabolic routes rather than in routes linked to identified antinutritional compounds.

5.4 Experimental procedures

5.4.1 Tomatoes

Genetically modified tomato lines as well as control lines were supplied by the Royal Holloway University of London, United Kingdom. All tomatoes were harvested at the same post anthesis stage and selected for similar size at the time of harvesting. For unmodified tomatoes this

stage corresponded to the red stage of ripening. Tomatoes from all lines were frozen in liquid nitrogen and shipped on dry ice to the Netherlands for transcriptomics studies. The GM lines used were two high carotenoid expressors, 35 *crt*I E312 C5S2-14 (GM7) and 35 *crt*I E313 c6S2-14 (GM8) of which the latter shows lower expression of the bacterial phytoene desaturase [11]. The comparator is the Ailsa Craig WT parent line (line 3).

5.4.2 RNA isolation (for hybridisation)

For the isolation of total RNA from tomato tissue material an extraction method based on TRizol Reagent (InVitrogen, Breda, the Netherlands) was used. Frozen samples of peel and pulp (combined) were ground under liquid nitrogen and samples of approximately 2 ml volume were resuspended in 9 ml TRizol Reagent at room temperature. The samples were vortexed during 1 min or until all ground sample was mixed with the Trizol reagent prior to incubation at room temperature for 15 minutes. Subsequently the samples were centrifuged for 15 minutes at 11000g at 4 °C to remove the cell debris. Then 0.5 ml chloroform/isoamylalcohol (24:1) was added to the supernatants and mixed by inverting the tube during 15 seconds. The samples were then incubated 3 minutes at room temperature prior to 15 minutes of centrifugation at 11000g and 4 °C. The supernatant was transferred to a new reaction tube and 0.5 ml 8M LiCl per ml Trizol reagent as was added for the initial step of the isolation. The LiCl was mixed with the supernatant and the mixture was incubated overnight at 4 °C. The next day the tubes were centrifuged for 15 minutes at 11000g and the supernatant removed, the pellet was washed with 75% ice-cold ethanol (25% DEPC-treated water) and centrifuged again at 11000g during 10 min at 4 °C. The pellet was then dried and resuspended in 60 µl DEPC-treated water. Finally the RNA pellet was incubated in a 65 °C water bath for 10 minutes to completely dissolve. The total RNA sample was quantified using the Nanodrop™ ND1000 spectrophotometer (NanoDrop Technologies) Subsequently the total RNA sample was DNase-treated according to the manufacturers protocol (RNase-Free DNase, Promega) and purified by phenol/chloroform/isoamylalcohol extraction and ethanol precipitation. The total RNA was again quantified, the quality was checked by agarose gel electrophoresis (1% agarose (Invitrogen, Breda, the Netherlands), 0.5 x TBE, 5% formamide). And then the RNA was stored at -80 °C. The yield per sample was approximately 12-15 µg total RNA.

5.4.3 Microarrays

Microarrays used were dedicated arrays and contained 2112 cDNAs of two RDA-based libraries [7], i.e. 960 clones of a green-tomato enriched cDNA library, and 1044 clones of a red-tomato enriched cDNA library. In addition control sequences were spotted in each block of the array. As positive controls, four luciferase cDNA sequences were cloned, a full length sequence and partial sequences of the 5'-, 3'- and middle part of the luciferase sequence. As a negative control a Salmonella sequence (410 bp) was cloned. The Salmonella sequence was checked using BLAST-N (NCBI) in the Genbank/EMBL database for absence of similarity to sequences derived from tomato. Microarrays were printed on silylated slides (CELAssociates, Pearland,

USA) using a Cartesian Technologies, PixSys 7500 spotter (Biodot limited, Chichester, UK) and Chipmaker 3 pins (Telechem, Sunnyvale, USA). After printing, microarrays were dried at room temperature for at least 3 days. Free aldehyde groups were blocked with $NaBH_4$ according to the method of Schena *et al*. [25].

5.4.4 Microarray hybridisation

25 µg of sample mRNA was labeled by incorporating Cy5-dCTP in a RT reaction [26]. The labeled cDNA was dissolved in 25 µl hybridisation buffer (5 x SSC, 0.2% SDS, 5 x Denhardt's, 50% (v/v) formamide, 0.2 mg/ml denatured herring sperm DNA). Prior to hybridisation, samples were heated for 3 min at 65 ºC and spun to remove undissolved debris. As a reference samples were taken from all experimental RNA isolates, mixed and labelled by incorporation of Cy3-dCTP in a separate RT reaction according to the same labelling protocol. After the final dissolving step of the individually labelled cDNA fractions the Cy5- and Cy3-labelled fractions are mixed in a 1:1 ratio. Before hybridisation the microarrays were prehybridised in hybridisation buffer at 42 ºC during at least 4 hours. After the prehybridisation the arrays were washed twice in MilliQ water and once in isopropanol and dried by centrifugation (2 min at 470 x g). The hybridisation was performed in a hybridisation frame (50 µl hybridisation buffer, Geneframe, Implen, Munich, Germany). Arrays were hybridised at 42 ºC in a humid hybridisation chamber overnight. After hybridisation the arrays were washed at room temperature, successively in 1 x SSC/0.1% SDS (5 min), 0.1 x SSC/0.1% SDS (5 min) and 0.1 x SSC (1 min) and subsequently dried by centrifugation (2 min, 470 x g).

A series of 60 hybridisations was performed with 60 different tomato totRNA (total RNA) samples: 26 hybridisations with WT Ailsa Craig tomato samples (3 different plants, 9 different tomatoes, 2-4 repeats per tomato), 34 hybridisations with tomato samples from two genetically modified 'high carotenoid expressor' lines: 17 GM7 (transformant line 35*crt*I E312 C5S2-14, 3 different plants, 9 different tomatoes, 1-2 repeats per tomato), and 17 with tomato samples from 'high carotenoid expessor' GM8 (transformant line 35*crt*I E313 C6S2-14, 3 different plants, 9 different tomatoes, 1-2 repeats per tomato).

5.4.5 Scanning and data analysis

Microarrays were scanned using a confocal laser scanner ScanArray 3000 (General Scanning Inc., Pittsfield, USA) containing a GreNe 543 nm laser for Cy3 measurement and a HeNe 633 nm laser for Cy5 measurement. Scans were made with a pixel resolution of 10 micron. The software package ArrayVision (Imaging Research, Waalwijk, the Netherlands) was used for image analysis of the TIFF-files as generated by the scanner. Fluorescent signals were collected for each individual spot and stored for further data processing (dot plot analysis) in Microsoft Excel (spot selection, normalisation and statistical T-test) and Genemaths (principle component) software. The results of individual hybridisation experiments were quantified and spots with a signal-to-noise ratio larger than three were used for further analysis. Normalisation

was performed on the basis of the reference sample, composed of a mixture (equal ratios) of all experimental samples that was included in all hybridisation experiments. In addition normalisation over the arrays was performed to correct for large differences in overall signals between the subsequent samples. Normalised data were analysed by PCA (principle component analysis). The PCA analyses were performed on the basis of direct comparisons between the WT Ailsa Craig transcriptomics data and the transcriptomics data as derived from either of the GM7 or GM8 experimental lines. To identify genes that were potentially altered between the parent and the GM line, the PCA was used to make a broad selection of potentially interesting genes. Subsequent further selection was performed by T-testing for statistical significance between WT and the individual GM lines ($p<0.05$).

5.4.6 Quantitative real-time PCR

1 µg of RNA (DNase treated) was reverse transcribed to cDNA using the Bio-Rad iScript cDNA Synthesis Kit (BioRad, Veenendaal, the Netherlands) in 20 µl volume. The cDNA reaction mixture was filled up to 100 µl with 10 mM Tris-HCl (pH 8.0). 1 µl of this cDNA solution was used for subsequent Q-PCR. For design of the primers, the selected EST sequences were re-blasted in Blast-nr and, if negative, in BLAST-est_others in the Genbank/EMBL database. Largely homologous gene sequences were aligned with the ESTs as spotted in AlignX (Vector NTI Inc., InVitrogen, Breda, the Netherlands). Homologous stretches between the selected ESTs and Genbank sequences were used to design primers in Beacon Designer® (Premier Biosoft International, Palo Alto, California, USA). The primers were obtained from Biolegio (Nijmegen, the Netherlands), dissolved in H_2O and the concentration was determined by measuring the OD 260 (Nanodrop™, Isogen, Nieuwegein, the Netherlands). In Table 3 the target genes and associated primer sequences are shown. Two to three primer sets were designed per EST in order to select the primer combination that performed best. Two primer concentrations (400 and 800 nM) were tested. Most primers performed best at 800 nM, except for 18S that performed best at a concentration of 400 nM. Tomato ribosomal 18S RNA was chosen as a reference gene as it was most constantly expressed at high levels with the lowest variation (average of duplo results of all samples: Ct = 14.35, SD 1.08) compared to BT13646 (unidentified tomato mRNA sequence showing unvariant gene expression in the subsequent ripening stages in tomato, average Ct = 29.79, SD 1.56) and β-actin (average Ct = 22.96, SD 3.31). The actin primers were derived from the consensus sequence of the NCBI genbank hits, by alignment in AlignX (Vector NTI™ suite (Invitrogen), with the aim to amplify different actin gene products simultaneously, assuming that the total expression may be more constant than the expression of individual actin genes.

A dilution series (1-10,000 times dilution) from a 'high expression' sample was included in the analysis, as well as a negative control (H_2O). The quantitative PCRs were performed in an iCycler iQ System (BioRad, Veenendaal, the Netherlands) using the following conditions: 1 µl cDNA, 12.5 µl 2 x iQ™ SYBR® Green Supermix (Bio-Rad), 400 or 800 nM sense primer, 400 or 800 nM antisense primer in a total volume of 25 µl. Amplification conditions were 3' 95 °C, 40

cycles of 15" 95 °C and 45" 60 °C followed by a melting curve analysis in which the temperature increased from 60 °C to 95 °C in increments of 0.5 °C. Relative expression levels were calculated in Microsoft Excel per sequence as well as relative expression levels for the WT and GM lines.

Acknowledgements

This research was financially supported by the European Commission (GMOCARE project QLRT-1999-00765) and the Dutch Ministry of Agriculture, Nature and Food Quality.

References

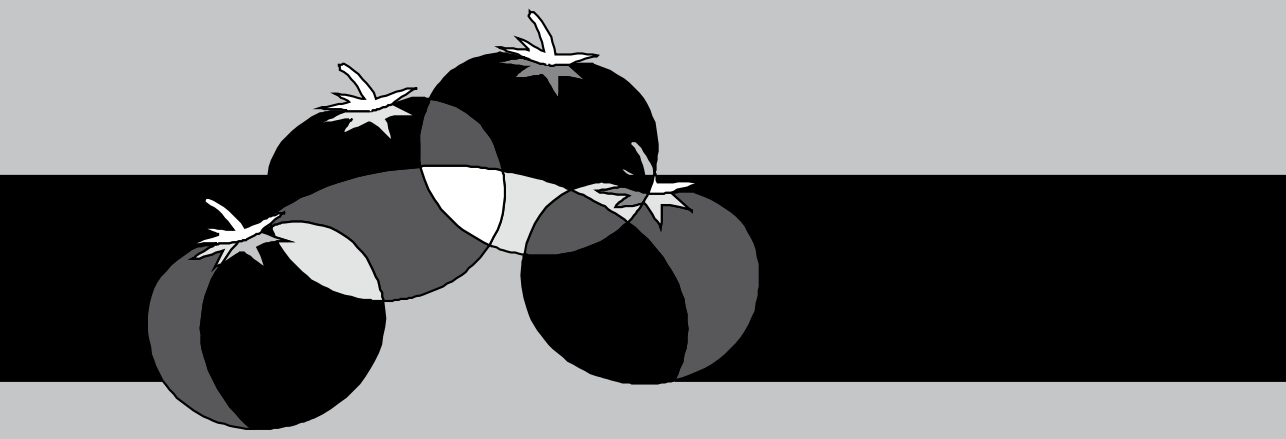
1. FAO/WHO (2000) Safety aspects of genetically modified foods of plant origin. *Report of a Joint FAO/WHO Expert Consultation on Foods derived from Biotechnology*, World Health Organization, Geneva, Switzerland.
2. Kok, E.J., and Kuiper, H.A. (2003) Comparative Safety Assessment for Biotech Crops. *TIBTECH* 21, 10, 439-444.
3. Kuiper, H.A., Kleter, G.A., Noteborn, H.P.J.M. and Kok, E.J. (2001) Assessment of the food safety issues related to genetically modified foods. *Plant J.*27, 6, 503-528.
4. OECD (1996) Food Safety Evaluation. *Report Workshop held in Oxford UK, September, 12-15, 1994*, OECD, ISBN 92-64-14867-1.
5. Breslin, L., and Kuiper, H.A. (2004) Safety Assessment, Detection and Traceability, and Societal Aspects of Genetically Modified Foods. European Network on Safety Assessment of Genetically Modified Food Crops (ENTRANSFOOD). Preface and introduction. *Food Chem. Tox*. 42, 7, 1043-1045.
6. Lisitsyn, N., Lisitsyn, N., and Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259, 5097, 946-51
7. Kok, E.J., Franssen-van Hal, N.L.W., Winnubst, L.N.W., Kramer, E.H.M., Dijksma, W.T.P., Kuiper, H.A. and Keijer, J. (2007) 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*. 164, 337-349.
8. Friedman, M. (2002) Tomato glycoalkaloids: role in the plant and in the diet. *J. Agric. Food Chem*. 50, 5751-5780.
9. Van Gelder, W.M.J. and De Ponti, O.M.B. (1987) α-Tomatine and other steroidal glycoalkaloids in fruits of tomato lines resistant to the glasshouse whitefly (*Trialeurodes* vaporariorum Westw.). *Euphytica* 36, 555-561.
10. Fraser, P.D., Romer, S., Kiano, J.W., Shipton, C.A., Mills, P.B., Drake, R., Schuch, W., Bramley, P.M. (2001) Elevation of carotenoids in tomato by genetic manipulation. *J. Sci. Food Agric*.81, 9, 822-827.
11. Römer, S., Fraser, P.D., Kiano, J.W., Shipton, C.A., Misawa, N., Schuch, W., Bramley, P.M. (2000) Elevation of the provitamin A content of transgenic tomato plants. *Nature Biotech* 18, 6, 666-669.
12. Kok, E.J., Lehesranta, S.J., Van Dijk, J.P., Helsdingen J.R., Dijksma, W.T.P., Van Hoef, A.M.A., Koistinen, K.M., Kärenlampi, S.O., Kuiper, H.A., and Keijer, J. Changes in gene and protein expression during tomato ripening – consequences for the safety assessment of new crop plant varieties. *Food Sci.Technol. Int*., in press.

13. Lee, S.W., Robb J., and Nazar, R.N. (1992). Truncated phenylalanine ammonia-lyase expression in tomato (*Lycopersicon esculentum*). *J. Biol.Chem*. 267, 17, 11824-30.
14. Olsen, D.C., White, J.A., Edelman, L., Harkins, R.N., and Kende, H. (1991) Differential expression of two genes for 1-aminocyclopropane-1-carboxylate synthase in tomato fruits, *Proc. Natl. Acad. Sci.* 88, 5340-5344.
15. Ramamoorthy, V., Raguchander, T., and Samiyappan, R. (2002) Induction of defense-related proteins in tomato roots treated with Pseudomonas fluorescens Pf1 and Fusarium oxysporum f. sp. Lycopersici. *Plant Soil* 239, 55-68.
16. Xie, D.-Y., Sharma, S.B., Wright, E., Wang, Z.-Y. and Dixon, R.A. (2006). Engineering plants for introduction of health-beneficial proanthocyanidins. *Plant J.* 45, 895-907.
17. Zipfel, C., and Felix, G. (2005) Plants and animals: a different taste for microbes? *Curr. Opin.Plant Biol.* 8, 4, 353-360.
18. Trainotti, L., Tadiello, A., and Casadoro, G. The involvement of auxin in the ripening of climacteric fruits comes of age: the hormone plays a role of its own and has an intense interplay with ethylene in ripening peaches. *Journal of Experimental Botany*, doi:10.1093/jxb/erm178.
19. Baudo, M.M., Lyons, R., Powers, S., Pastori, G.M., Edwards, K.J., Holdsworht, M.J., and Shewry, P.R. (2006) Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnol. J.* 4, 369-380.
20. Baker, J.M., Hawkins, N.D., Ward, J.L., Lovegrove A., Napier, J.A., Shewry, P.R., and Beale, M.H. (2006). A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnol. J.* 4, 381-392
21. 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., and Draper, J. (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc. Natl. Acad. Sci. USA* 102, 40, 14458-14462.
22. Colquhoun, I.J., Le Gall, G., Elliott, K.A., Mellono, F.A., and Michael, A.J. (2006) Shall I compare thee to a GM potato? *TIG* 22, 10, 525-528.
23. Le Gall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J., and Verhoeyen, M.E. (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using 1H NMR Spectroscopy as a tool to detect potential unintended effect following a genetic modification. *J. Agric. Food Chem*. 51, 2447-2456.
24. Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., Nunan, N., McNicol, J.W., Auriola, S., Koistinen, K.M., Suomalainen, S., Kokko, H.I., Kärenlampi, S.O. (2005) Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. *Plant Physiol* 138, 1690-1699.
25. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad.Sci. USA* 93, 10614-19.
26. Franssen-van Hal, N.L.W., Vorst, O., Kramer, E.H.M., Hall, R.D., and Keijer, J. (2002). Factors influencing cDNA microarray hybridisation on silylated glass slides. *Analytical Biochemistry* 308, 1, 5-17

Chapter 6.

Abstract

Genetically modified plants must be approved before release in the European Union, and the approval is generally based upon a comparison of various characteristics between the transgenic plant and a conventional counterpart. As a case study, focusing on safety assessment of genetically modified plants, we here report the development and characterisation of six independently transformed *Arabidopsis thaliana* lines modified in the flavonoid biosynthesis. Analyses of integration events and comparative analysis for characterisation of the intended effects were performed by PCR, quantitative Real-time PCR and High Performance Liquid Chromatography. Analysis by cDNA microarray was used as a non-targeted approach for the identification of potential unintended effects caused by the transformation. The results revealed that although the transgenic lines possessed different types of integration events, no unintended effects were identified. However, we found that the majority of genes showing differential expression were identified as stress-related genes and that environmental conditions had a large impact on the expression of several genes, proteins and metabolites. We suggest that the microarray approach has the potential to become a useful tool for screening of unintended effects, but state that it is crucial to have substantial information on the natural variation in traditional crops in order to be able to interpret 'omics' data correctly within the framework of food safety assessment strategies of novel plant varieties, including genetically modified plant varieties.



Evaluation of a non-targeted 'omic' approach in the safety assessment of genetically modified plants

Stine B. Metzdorff [1], Esther J. Kok[2], Pia Knuthsen[1], Jan Pedersen [1]

[1] Danish Institute for Food and Veterinary Research, 19 Mørkhøj Bygade, DK 2860 Søborg, Denmark
[2] RIKILT Institute of Food Safety, 45 Bornsesteeg, 6708 PD Wageningen, the Netherlands

Plant Biology 8 (2006): 662-672.

6.1 Introduction

Prior to release and marketing of genetically modified (GM) plants in the European Union (EU), an approval, based on a safety assessment of the impact on the environment, is mandatory (Directive 2001/18/EC, [1]). In addition, a specific authorisation needs also to be given before food or feed derived from GM -organisms can be marketed in the EU (GMO Food and Feed Regulation (EC) No 1829/2003). A few transgenic crops, all altered by the sense/antisense technique, such as a tomato (applicant Zeneca) with prolonged maturation, soybean (applicant DuPont) modified for high oleic acid and a potato (applicant Amylogene HB) with modified starch content, have been notified for approval under the EU GMO regulation [2,3]. The approval, when given, will be based on a scientific assessment, and it will be carried out according to the EU guidance document for risk assessment of genetically modified plants and derived food and feed derived from this organism (EFSA, [4]). The concept of substantial equivalence (also mentioned as the comparative safety approach) has been recommended as a guiding tool for the safety assessment. The concept is based on the idea that an existing organism, used as food or feed with a history of safe use, can serve as a comparator when assessing the safety of the genetically modified food/feed [4,5,6,7]. Moreover, it has been recognized that the application of the concept of substantial equivalence contributes to a robust safety assessment framework and it has been adopted in the EU guidelines as well as in the Codex guidelines [8].

A safety assessment of GM crops includes comparison with a counterpart for molecular, biological and compositional characteristics to reveal both intended and unintended effects. Intended effects are those that are expected to occur from the introduction of the gene(s) in question. Unintended effects are considered to be differences between the GM plant and its appropriate control line, which go beyond the primary expected effect(s) of introducing the target gene(s) (Regulation (EC) No 1829/2003). A detailed characteristic of the transgenic plant includes determination of DNA integration events, such as segregation ratio, copy number, and identification of putative vector backbone sequences with the presence of marker genes. Furthermore, precise determination of the integration site and detection of putative rearrangements should be carried out.

The current work formed part of an EU-funded project, GMOCARE, where genetically modified *Arabidopsis*, tomato, potato, and tobacco plants were used as subjects to different non-targeted techniques, the so-called 'omics' approaches, for evaluation of these techniques in relation to their applicability in safety assessment of GM plants [9]. As a case study, focussing on safety assessment of genetically modified plants, we generated transgenic *Arabidopsis* plants by *Agrobacterium*-mediated transformation to characterise and compare different transgenic lines and their appropriate control. The plants were transformed with an antisense chalcone synthase (*CHS*) gene, which encodes the rate-limiting enzyme in the flavonoid biosynthesis. The purpose was to obtain transgenic lines with altered flavonoid synthesis due to silencing of the endogenous *CHS* gene through the phenomenon termed homology-dependent gene silencing [10]. Flavonoids are secondary plant metabolites and are generally believed to have

possible health benefits in edible food plants [11,12,13]. Future transgenic crops produced for food, feed or industrial uses may very well include plants that are altered in their production of secondary plant metabolites [14]. Consequently, in relation to future risk assessment of GM plants with improved health characteristics, plants with modification in the synthesis of flavonoids, using the sense/antisense technique, were chosen in this study as a model for such upcoming GM crops. For the comparative analyses, PCR, quantitative Real-time PCR (QPCR) and high performance liquid chromatography (HPLC) analysis were used for characterisation of the intended effects, while cDNA microarray analysis was performed as a non-target approach for identification of unintended effects caused by the transformation.

6.2 Material and methods

6.2.1 Plasmid construction

The full-length chalcone synthase (*CHS*) gene was amplified by PCR from *A. thaliana*, ecotype Wassilewskija genomic DNA, using primers *CHS*-f: 5´-AAGCTCTCACTCTCCGGT-3' and *CHS*-r: 5´-TCGTGTGAGTCCCTTGCT-3'. The PCR-fragment was blunted and cloned into a *Sma*I site of pUC18 (Pharmacia). The cloned full-length *CHS* gene contains an *Xba*I restriction site positioned 1052 bases downstream the *CHS*-f specific site and 320 bases upstream the stop codon. A scheme showing the primer localisation on the genomic *CHS* sequence (Accession M20308, NCBI) is shown in Figure 1. The pBI121 plasmid (Clontech) contains a β-glucuronidase (*GUS*) gene regulated by CaMV 35S promoter. By excising the *GUS* gene in pBI121 with *Xba*I/*Sac*I and replacing it with the 1052 bp *Xba*I/*Sac*I fragment in the 5´-end of the *CHS* gene, pBI121/*CHS* was formed. The pBI121/*CHS* contains a neomycin phosphotransferase II (*nptII*) gene under control of the nopaline synthase promoter and the 5´-end *CHS* fragment in antisense direction driven by the CaMV 35S promoter. An illustration of the transformation vector is shown in Figure 2.

6.2.2 Plant material

Seeds of *Arabidopsis thaliana* (L.) Heynh. ecotype Wassilewskija were obtained from the *Arabidopsis* Stock Centre Nottingham centre (NASC) (Ws-0 stock line N1602). *Agrobacterium*

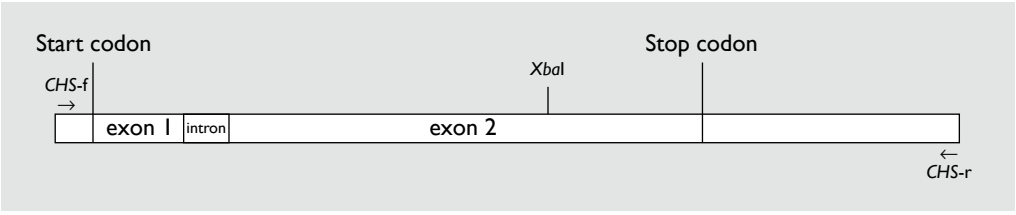


Figure 1. *A schematic illustration of the genomic* CHS *gene with marked locations of start and stop codons, intron and exons. The positions of primers used for amplifying the gene, and the* Xba*I restriction site used for producing the transformation vector are indicated as well.*

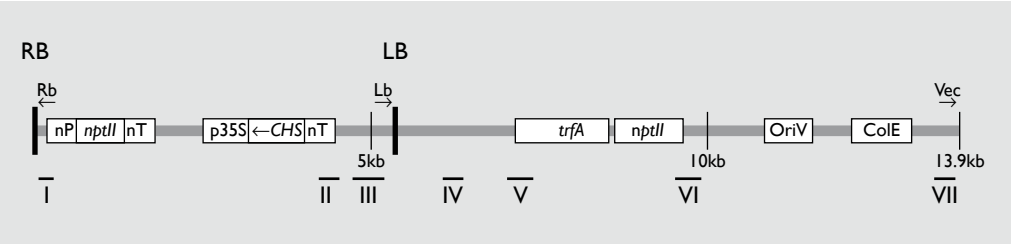


Figure 2. *A schematic illustration of the colinear vector pBI121/*CHS *used for transformation. RB: Right border. LB: Left border. Pn: nos promoter. nptII: neomycin phosphotransferase II gene. Tn: nos terminator. P35S: CaMV 35S promoter. I-VII: Amplified fragments for integration size analysis. Rb, Lb, Vec: Primers used for tandem repeat analysis and border determination.*

tumefaciens pGV3101 was transformed with pBI121/*CHS* by electroporation and subsequently used for transformation of *Arabidopsis thaliana,* ecotype Wassilewskija, (Ws-0) using the floral dipping method [15]. Seeds were liquid sterilized [15] and germinated on Murashige and Skoog (MS) medium containing kanamycin (50 mg/l), 3% (w/v) sucrose, 0.1 % (v/v) Gamborg vitamin (Sigma), and 0.8% (w/v) agar. Transformants were selected after 10 days and transferred to pots. Seeds from self-fertilised primary transformant lines and the subsequent generations were surface-sterilised and plated on MS medium with kanamycin (50 mg/l) for determination of segregation ratio for the selectable marker *nptII* gene, in order to estimate the number of integration loci and for registration of *nptII*-silencing. For lines with 100% kanamycin sensitive seedlings or non-Mendalian ratios, the segregation ratios were determined from PCR analysis with primers specific for the *nptII* gene using 40 siblings from each line germinated in pots.

6.2.3 Analysis of genomic DNA

Genomic DNA used for analysing different integration events was isolated from rosette leaves of homozygous transgenic lines and of the wild type (control) using DNeasy Plant Mini Kit (Qiagen). Part of T-DNA and vector backbone integrated in the plant genome were analysed by PCR, using 7 primer pairs specific for fragment I-VII (Figure 2, primer sequences in Table 1) covering the entire vector. Integration site was identified from flanking sequences obtained by a PCR-walking method, [16] using an asymmetrical adaptor and primers specific for the adaptor combined with primers for the right and left border of the T-DNA and the vector backbone. Fragments obtained from nested PCR were cloned into pGEM®-T Easy vector (Promega) and sequenced. Determination of T-DNA and vector backbone repeats was performed by PCR using specific primers orientated in opposite directions in the construct, which only amplified PCR products when insertions contained direct or inverted repeats. Rb, Lb and Vec primers (Figure 2 and Table 1) were used in different combinations and alone for the analysis.

Table 1. *List of primers categorised by name of primer or PCR product.*

Name	Forward primer, 5´-3´	Reverse primer, 5´-3´
I	GAAGGCGGGAAACGACAATCTG	CGGTTCTGTCAGTTCCAAAC
II	AAGACCGGCAACAGGATTCAAT	AGCAGTGCCAATAGCCAA
III	TGCCGGTCTTGCGATGATTAT	GGAGCCCCCGATTTAGAGC
IV	CACGCCAGGCCGAGAACATTG	GGCGGTCCTGGGGGCTATTTG
V	ACGGACCCGAACATCTCTGGA	ACCCGCTCAAGCTGGAAACCT
VI	CGACGGAGCCGATTTTGAA	GCGCGATTTAGCCCCGACATAG
VII	CCGGCTGCTGAACCCCCAACC	GCCGCCGCCCGAAACGAT
Rb	CGTCATCGGCGGGGGTCAT	
Lb	TTCGGAACCACCATCAAACAGG	
Vec	AGGGCGGCACGGATCACTGTATTC	
CHS	CGTGGTGGTCGAAGTCCCTAAGC	TACCGCCGGCGAAGCAACC
CHS2	AGGTCCAGCGATCCTAGAC	TCTAGTATGAAGAGAACGCAC
VSPI	CATAAACTAAACAATAAACCATACAAA	GGGGTAGTTGATGGACAGTCC
18S	CCTTGGGATGGGTCGGC	AGCCGGCCGTGAAGGG
CHS probe	FAM-ATGTCGTCTTCTGCACTACCTCCGGCGT-TAMRA	
CHS2 probe	FAM-TTCCATACTCGCTCAACACGTGA-TAMRA	
VSPI probe	FAM-CTCTTGCTCTTGGCCGCTACGGTC-TAMRA	
18S probe	FAM-CAATGATCCTTCCGCAGG-TAMRA	

6.2.4 Copy number by quantitative real-time PCR (QPCR)

QPCR was performed in order to estimate the relative copy number in the transgenic lines using standard TaqMan® technology with fluorescence monitoring (LightCycler, Roche) [17]. By comparing the amplification of the total number of *CHS* sequence targets (*CHS* from the inserted gene, exogenous *CHS* (exo*CHS*), plus the endogenous *CHS* (endo*CHS*)) with the amplification of only endo*CHS* sequence targets, a relative determination of copy number was made. Two sets of oligos were used for the estimation. CHS forward and reverse and the hybridisation CHS probe are specific for both the exo- and endo*CHS* gene (Table 1) and corresponds to the 5'end of the *CHS* gene upstream the *Xba*I site. CHS2 forward and reverse and the hybridisation CHS2 probe are specific for the endo*CHS* gene and correspond to the 3'end of the *CHS* gene downstream of the *Xba*I site. The endo*CHS* was used as an internal standard for one copy. A serial dilution of a linear plasmid containing the full-length *CHS* gene was used as template to generate standard curves for both targets. Each PCR master mix, with a total volume of 10 µL, contained 20 ng of DNA sample, 4 mM $MgCl_2$, 0.2 µM probe, 2.0 µM primers (except for *CHS*-I-for where 0.5 µM was used), and 1 µL LightCycler FastStart DNA Master Hybridisation Probes reaction mix (Roche).

PCR conditions for amplifying both targets were 95 °C for 3 min, 35 cycles of 95 °C with holding time set to 0 sec, and 60 °C for 30 sec. Two samples of DNA, independently isolated, for each transgene were amplified. The samples were tested in duplicate within the reaction. All PCR amplification was with one colour detection. Second Derivatives Maximum Method from the Lightcycler Software (Roche) was used for the calculation of concentrations in the unknown samples. The signals of the unknown samples were quantified from standard curves for the two targets, endo and exo*CHS*, and the ratio between the concentration of the exo*CHS* and the endo*CHS* for each sample was calculated. The exo/endo ratio for *CHS* in the wild type equals a single *CHS* target, and by dividing the ratio of the exo/endo for *CHS* in the different samples by the exo/endo for *CHS* in the wild-type, the relative copy number integrated in the GM lines was estimated: relative copy no.= (exo/endo$CHS_{transgene}$)/(exo/endoCHS_{WT}).

6.2.5 HPLC analysis

Seeds were sterilised and plated on Petri dishes containing Murashige and Skoog medium, stored at 4 °C for 48 h for stratification and then placed in the growth chamber. 12 days old seedlings were harvested, and, as seedlings from each line were plated on 4 dishes and 2 samples were harvested from each dish, a total of 8 samples from each line were analysed separately for flavonoids, thus compensating for the biological variation between plants. An earlier reported procedure was optimised for small sample amounts. Samples of 100 mg were homogenised and extracted with 600 µL methanol and water (80:20%, v/v) by Ultra Turax. The resulting extract was made up to 50% methanol in 1.2 M hydrochloric acid, hydrolysed for 2 h, and analysed by HPLC, as described earlier [18]. Contents of flavonoid aglycones were found by UV detection at optimal wavelengths and comparison to external standards.

6.2.6 Gene expression analysis by QPCR

The expression of *CHS* and *VSP1* relative to the housekeeping gene *18SrRNA* was analysed by standard TaqMan® technology using fluorescence monitoring (LightCycler, Roche) [17]. Rosette leaves were harvested from 7-week-old plants (vegetative phase), from the second and third inner whorl. Flowers were harvested when the plants had produced a bushy secondary inflorescence. All plants were grown randomly, side by side, and, to diminish plant-to-plant variation, tissue from 8 individual plants was combined. Seedlings were produced and harvested using a similar process to seedlings used for HPLC analysis. All tissue was stored in RNA*Later* (Ambion), total RNA was isolated using RNeasy Plant Mini Kit (Qiagen), DNase-treated (DNA-*free*, Ambion) and cDNAs were produced (Omniscript Qiagen) using oligo T25 and an oligo specific for 18SrRNA. The following three primer and probe sets were used for identifying transcripts of *VSP1*, *CHS* and *18SrRNA*: VSP1 primers and probe, CHS2 primers and probe (targeting only the endogenous *CHS* gene), and 18S primers and probe (Table 1). Each PCR master mix, with a total volume of 10 µL, contained 1 ng cDNA, 4 mM $MgCl_2$, 0.2 µM probe, 0.5 µM primers and 1 µL LightCycler FastStart DNA Master Hybridisation Probes reaction mix (Roche). PCR conditions for all targets were 95 °C for 10 min, 45 cycles of 95 °C with holding time set to 0 sec, and 60 °C for 30 sec. Three

experimental samples, independently isolated and cDNA synthesised, for each transgene were analysed. A minimum of three samples were amplified for detection of *VSP1*, *CHS* or *18SrRNA*, quantified from standard curves for the three targets, and expression of *VSP1* and *CHS* relative to 18SrRNA was determined.

6.2.7 Gene expression analysis by Microarray

Subtractive hybridisation (Clontech) was used in order to obtain cDNAs representing genes that are differentially expressed in different tissues. Both forward and reverse subtraction of cDNAs from leaf and flower, and forward subtraction of cDNAs from leaf and root and from leaf and stressed seedlings (treated with 200 mM NaCl, 5 mM DTT, or 5% EtOH, respectively, in aqueous solutions prior to RNA isolation) were performed. cDNAs were cloned in pGEM-T Easy (Promega) and a total of 1536 clones, including controls, were amplified by PCR with 5'hexamine-modified primers. Purified PCR products, using Sephadex G-50 in Milli Multiscreen-HV plates, were precipitated with NaAc and isopropanol, washed, dissolved and spotted on silylated glass slides with the PixSys 7500 arrayer (Cartesian Technologies).

35 individual plants from each of the six transgenic lines and wild type were grown, placed randomly side by side. Rosette leaves from 7 week-old plants (vegetative phase) were harvested, from the second and third inner whorl. 1 leaf from each of 5 individual plants was combined into a sample. Flowers were harvested over a period of 5 weeks due to variation in flowering time between plants. All tissue was stored in RNA*Later* (Ambion). Total RNA was isolated (Qiagen), DNase-treated (DNA-*free*, Ambion) and ultimately, the total RNA sample was pooled for each line and tissue in order to minimise the biological variation between plants.

Two hybridisations were performed. In experiment 1, arrays were hybridised with flower mRNA. Leaf mRNA was used in experiment 2. Test RNA (20 µg total RNA) was labelled with FluoroLink™ Cy5-dCTP (Amersham Pharmacia Biotech) during cDNA synthesis. Also a large pool of RNA from all tested lines was labelled with FluoroLink™ Cy3-dCTP (Amersham Pharmacia Biotech), and this pool of labelled cDNA was used as reference. The reference mRNA sample was used for all hybridisations in order to relate individual experiments within a series to each other. The two labelled cDNA pools were then mixed 1:1 and placed on the array. Two slides for each test RNA were made and, after competitive hybridization overnight, slides were washed and scanned at the two wavelengths for Cy3 and Cy5. Spots and backgrounds on the array images were quantified in ArrayVision, and then analysed in Microsoft Excel for data normalisation. Data were normalised by correcting for the reference and for median of the signals. An X-Y scatter plot for each duplex hybridisation was analysed and only hybridisations with acceptable variation (under two-fold differences) within each duplex were used for further analysis. Clones showing more than two-fold differential expression compared to the wild-type were sequenced and blasted in the TAIR database for identification.

6.2.8 Sequencing and sequence data analysis

Sequence analysis was performed using CEQ 2000 Dye Terminator Cycle Sequencing with Quick Start Kit (Beckman Coulter) and analysed on CEQ™ 2000 DNA Analysis System. Analysis of sequences was performed with DNASTAR (Version 4.0, Lasergene), BLAST and BLAST 2 sequences of NCBI (National Center for Biotechnology Information, http:// www.ncbi.nlm.nih.gov/) and BLAST and SeqViewer of TAIR (The Arabidopsis Information Resource, http:// www.arabidopsis.org/).

6.3 Results

6.3.1 GM lines

From limited analyses of several independent primary transformation lines, six lines, a, b, c, d, f, and o were selected for a complete characterisation of homozygous T_4 progenies. The six lines were selected on the basis of their different types of integration events, such as integration size, copy number, and the presence of vector backbone and repeats. The integration site could be determined for four lines. In line c, the T-DNA has been inserted in chromosome 3 in the 5′ UTR- region in a gene with unknown function. Line o was interrupted in an inter-genic region, and the inserts of line f and d had both been integrated in coding regions of genes with unknown function. The latter three lines had their insertion integrated in chromosome 1. All integration events are summarised in Table 2, and these events show that the six lines thereby

Table 2. *Summary of integration events in the six GM lines estimated by various PCR techniques. bb: vector backbone. nd: not detected. Chr: chromosome.*

Line	Transgenic loci	Copy no.	Integration size	Direct repeat from	Integration site	*nptII* silencing	Transgene *CHS* transcript[1]
a	I	3-4	T-DNA	T-DNA	Unknown	+	÷
b	I	15+	T-DNA + bb.	T-DNA + bb.	Unknown	+	÷
c	I	I	T-DNA	nd.	Chr. III[2]	÷	+
d	I	5	T-DNA + bb.	T-DNA + bb.	Chr. I	+	÷
f	I	I-2	T-DNA	nd.	Chr. I	÷	+
o	I	I	T-DNA trunc.	nd.	Chr. I	÷	÷

[1]Analysed in leaf, flower and seedlings.
[2]Verified by PCR.

represent both 'optimal' GM lines as well as 'worst cases' with regard to potential unintended side effects of the genetic modification.

In heterozygous generations, segregation ratios revealed that c, f and o had a ratio of 3:1 (resistance:sensitive) in relation to the inserted kanamycin resistance gene (nptII), demonstrating insertion at a single locus for the nptII gene. Line a, b and d displayed kanamycin-sensitivity in a non-Mendelian ratio, but when analysed by PCR for the nptII gene, the lines showed the expected segregation ratio of 3:1, showing single locus integration as well. The aforementioned lines continued being kanamycin sensitive (described as nptII silenced) in homozygous progenies (Table 2). When homozygous plants were grown in soil without selection, only one GM line (line c), showed diminished red pigmentation in the rosette leaves and stem, which is a direct consequence of down regulation of the expression of the CHS gene induced by the antisense construct.

6.3.2 Tandem repeat in multi-copy lines

Copy number estimations by Real-Time PCR analysis revealed a low copy number with one or two CHS copies integrated in line c, f and o. Line a has three to four copies and line d has five copies, while line b had more than fifteen CHS copies integrated (Table 2). PCR amplifications, using primer pairs specific for tandem repeat detection, revealed repeats from the T-DNA at different lengths in lines a, b and d. Furthermore, the PCR analysis also indicated that lines b and d contain direct repeats of the whole vector.

6.3.3 No transgenic CHS transcripts in multi-copy lines

In order to examine if the transgene *CHS* sequence, called exogenous *CHS* (exo*CHS*), was transcribed from the T-DNA, leaf, flower, and seedling cDNA samples (same as for QPCR, see 'Materials and Methods') were used as template in an end-point PCR analysis, using *18SrRNA* as a positive control and *CHS-Tnos* (fragment II in Figure 2) as the exo*CHS* specific fragment. Gel electrophoresis of the PCR products revealed that exo*CHS* was transcribed in all samples from line *c* and *f*, and although it is not possible to perform a quantification of the expression levels from an end-point PCR analysis, the results imply that the exo*CHS* transcript is expressed at similar levels in leaf, flower and seedling. No exo*CHS* transcription was detected in any of the samples from line *a*, *b* and *d* (Figure 3 and summarized in Table 2).

6.3.4 Targeted analyses of flavonoid biosynthesis

The transgenic lines were examined for intended alteration in *CHS* mRNA levels by QPCR, and at the metabolite level by HPLC. The expression of the endogenous *CHS* mRNA relative to *18SrRNA* (Figure 4) revealed that the *CHS* expression in leaves was significantly decreased in all the GM lines compared to the wild type, except for line *a* where no alteration was observed. The *CHS* mRNA level was decreased in flowers in all six GM lines, while in seedlings the level was

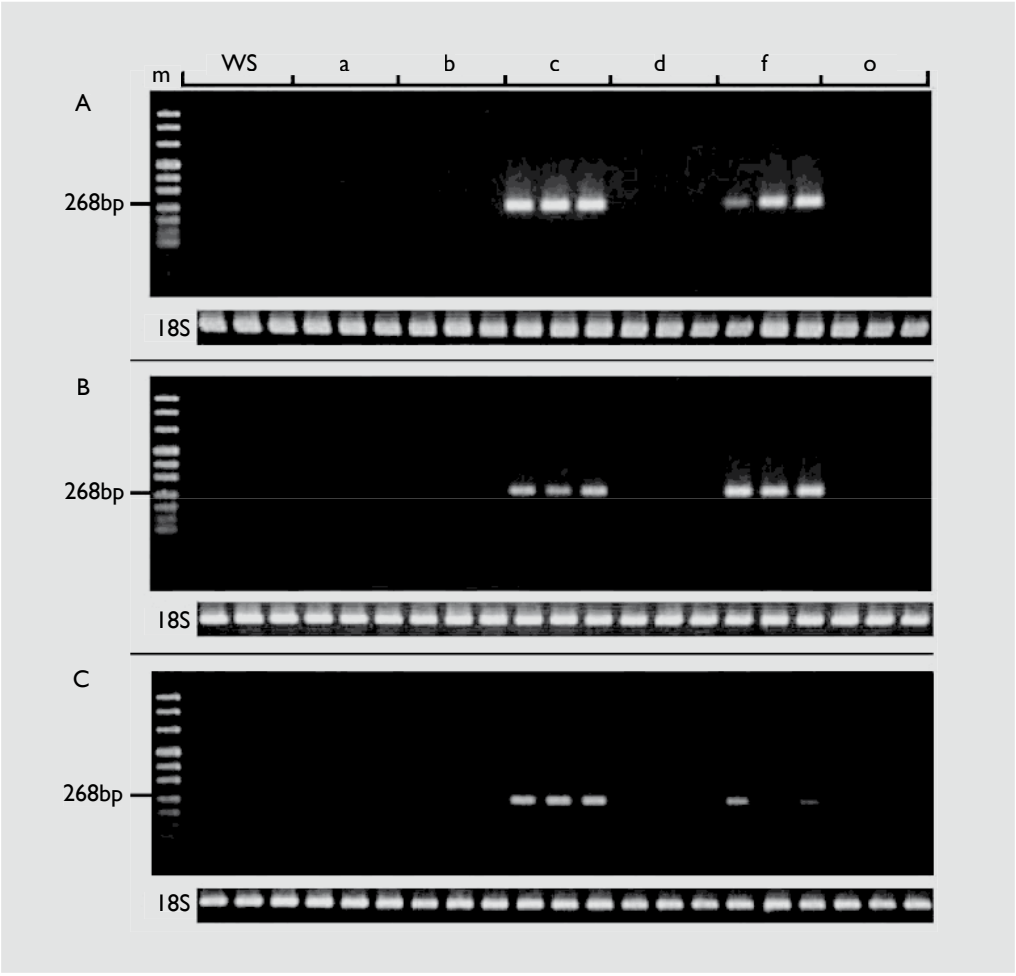


Figure 3. *Gel electrophoresis of an end-point PCR analysis of the exogenous* CHS *fragment (268 bp) using* 18SrRNA *as a positive control in cDNA samples from wild type and* a, b, c, d, f *and* o *(n=3). A: leaf, B: flower and C: seedling. m: pUC Mix Marker, 8 (Fermentas).*

decreased in *c, f* and *o,* and increased in *a, b,* and *d* compared to the wild type. Generally, the analysis implies that the expression of *CHS* was affected at the transcript level in all six GM lines and that *c* had a notably lower expression of *CHS* in flowers and seedlings, compared to the wild type and to the other transgenic lines.

HPLC analyses revealed a significant decrease ($P<0.05$) in the amount of the three flavonoid aglycones: quercetin, kaempferol and isorhamnetin in seedlings of all the GM lines except line *d* (Figure 5).

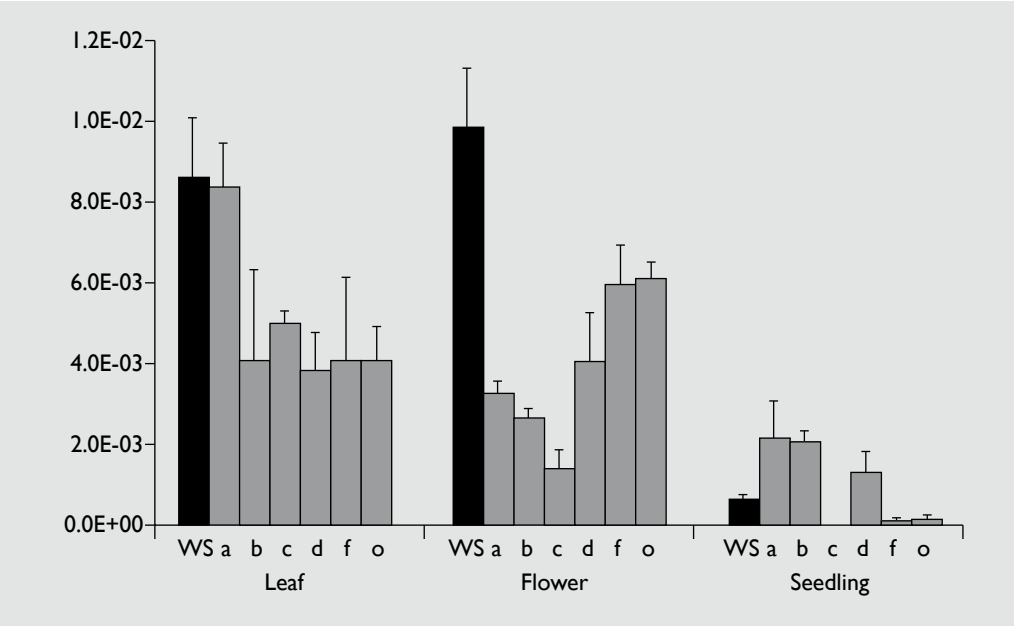


Figure 4. *Expression of the endogenous* CHS *gene relative to the housekeeping gene* 18SrRNA *(arbitrary units) in rosette leaf, flower and seedling from wild type and* a, b, c, d, f *and* o *based on Real-Time PCR. Mean value (n=3) ± S.D.*

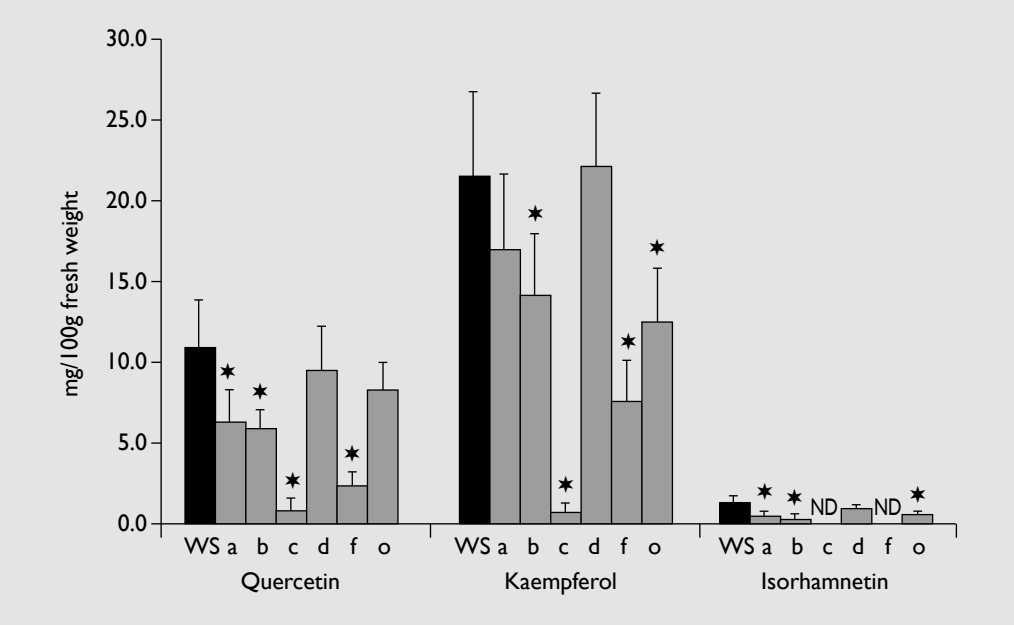


Figure 5. *HPLC analysis of three flavonoid aglycones: quercetin, kaempferol and isorhamnetin in 12-day-old seedlings from wild type and* a, b, c, d, f *and* o. *Mean value (n=8) ± S.D.*

6.3.5 Non-targeted analysis by microarray

Unintended effects caused by the genetically modification were analysed by the microarray technique, using a cDNA array consisting of approximately 1500 randomly selected *Arabidopsis* cDNAs from leaf, flower, root and stressed seedlings tissues, obtained by subtractive hybridisation. Competitive hybridisation between a reference and leaf or flower mRNA from each of the GM lines and the wild type was performed where each experimental RNA sample was pooled from 35 individual plants per line in order to minimize the effect of the biological variation between plants. The analysis revealed that 25 different clones had a differential expression level of more than two fold compared to the wild type. Annotation and locus name for the clones showing statistically significant (P<0.05) expression alteration in experiment 1 (flower), and 2 (leaf) are shown in Tables 3 and 4, respectively. In the flower, variations

Table 3. *Experiment 1; hybridisation with flower RNA. Gene expression differing from the wild-type in magnitude. *: Involved in the photosynthetic apparatus. **: Involved in defence response. ***: Response to oxidative stress.*

Line	Annotation	Locus name	Fold
c	Vegetative storage protein, *VSP1***	At5g24780	2.2
	Putative jasmonate inducible protein**	At1g54040	2.7
	Pseudogene, polygalacturonase inhibitor	At3g12145	2.2
d	Light-harvest. chlorophyll a/b binding prot.*	At2g05070	-2.1
	Putative protein, proline-rich protein APG	At5g33370	2.9
f	Photosystem I subunit 0, *Psa0**	At1g08380	2.5
	Light-harvest. chlorophyll a/b binding prot. CP26 in PSII*	At4g10340	2.2
	Photosystem II polypeptid, putative*	At1g79040	2.4
	Light-harvest. chlorophyll a/b binding prot.*	At2g05070	2.5
	Light-harvest. chlorophyll a/b binding prot.*	At1g29930	2.3
	Light-harvest. chlorophyll a/b binding prot.*	At1g61520	2.4
	Chlorophyll binding protein*	At1g15820	2.7
	Putative chlorophyll a/b-binding protein*	At3g08940	2.5
	Alanine-glyoxylate aminotransferase, *AGT**	At2g13360	2.5
	Putative jasmonate-inducible protein**	At1g54040	2.4
	Peroxidase, putative***	At4g21960	3.9
	Adenylosuccinate lyase-like protein	At4g18440	2.2
	Non-specific lipid-transfer protein, *LP1*	At2g38540	-2.2
	Putative protein, proline-rich protein APG	At5g33370	-2.2
	Putative thioredoxin-m	At1g03675	2.2

Table 4. *Experiment 2; hybridisation with leaf RNA. Gene expression differing from the wild-type in magnitude. *: Involved in the photosynthetic apparatus. **: Involved in defence response.*

Line	Annotation	Locus name	Fold
a	Vegetative storage protein, *VSP1***	At5g24780	-3.3
d	Photosystem I subunit V precursor, putative*	At1g55670	2.5
	Photosystem I subunit X precursor*	At1g30380	2.1
	Photosystem I subunit 0, *Psa0**	At1g08380	2.8
	Photosystem II type I chlorophyll a/b binding protein*	At2g34420	2.4
	Photosystem II polypeptid, putative*	At1g79040	2.3
	Light-harvest. chlorophyll a/b binding prot. CP26 in PSII*	At4g10340	2.7
	Light-harvest. chlorophyll a/b bind. prot. (LHCI typeIII CAB-4)*	At3g47470	2.6
	Light-harvest. chlorophyll a/b binding protein, putative*	At5g54270	2.3
	Light-harvest. chlorophyll a/b binding prot.*	At2g05070	2.2
	Light-harvest. chlorophyll a/b binding prot.*	At3g54890	2.2
	Light-harvest. chlorophyll a/b binding prot.*	At1g29930	2.2
	Bet v I allergen family, sim. to pollen allergen Bet v I	At1g24020	2.2
	Glutathione transferase, putative	At4g02520	2.7

in expression were recorded in line *c*, *d* and *f* in 3, 2 and 15 genes, respectively. In the leaf, variations in expression were only found in line *a* and *d* in, respectively, 1 and 13 genes. One important finding was that the majority (~72%) of the genes showing differential expression were identified as genes that are influenced by different types of stress. The majority of the stress-related genes (~83%) was identified to be involved in the photosynthetic apparatus, which has been reported to be sensitive to abiotic stress, such as excess light, salt, and temperature, all of which can cause oxidative stress [19,20]. Furthermore, stress-responsive genes, such as genes induced by jasmonic acid, considered to be an important stress-signalling molecule in plants [21], and the wound-inducible vegetative storage protein gene (*VSP1*) [22], were also found to be differential expressed.

VSP1 was selected as one of the differentially expressed genes identified by the microarray technique, in order to reveal if the expression pattern, when analysed by QPCR, was consistent in a newly grown set of plants. QPCR results from these new plants revealed that the expression in flowers is decreased in *a,* with no observed significant differences in the remaining GM lines (Figure 6). In the leaf, *VSP1* was down-regulated 3.3 fold in line *a* compared to the wild type in the microarray experiment (Table 4), while QPCR analysis revealed that the *VSP1* expression in the leaf was up-regulated in lines *a* and *b,* and *o* compared to the wild type, and *VSP1* was down-regulated in line *d* and *f* in the new plant set.

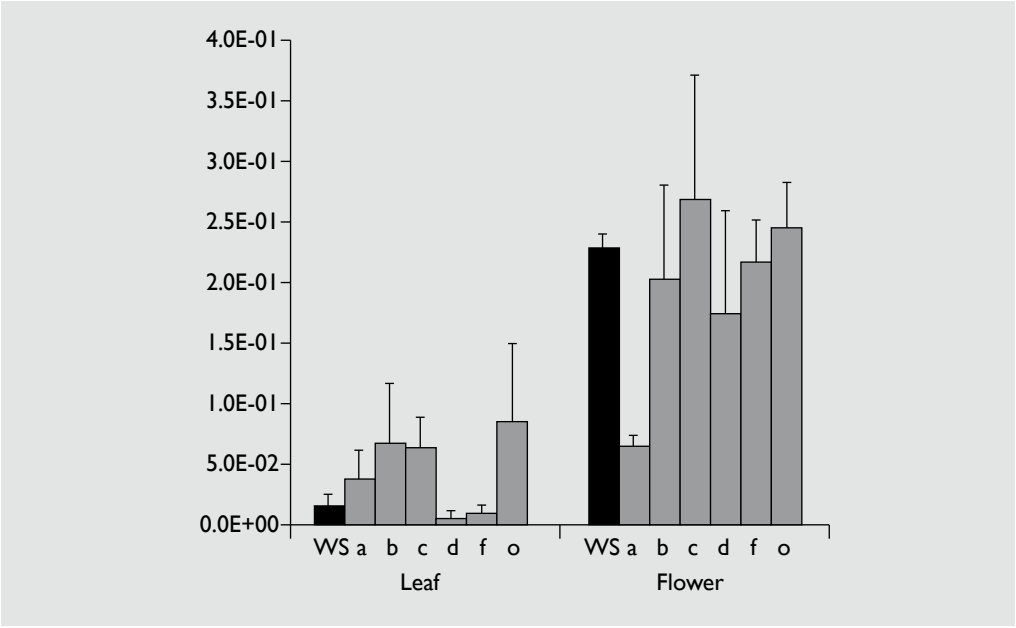


Figure 6. *Expression of* VSP1 *relative to the housekeeping gene* 18SrRNA *(arbitrary units) in flower and rosette leaf from a new-grown set in wildtype and* a, b, c, d, f *and* o *based on Real-Time PCR. Mean value (n=3) ± S.D.*

6.4 Discussion

Several genetically modified, agronomically and horticulturally important plants, such as strawberry [23], apple [24], potato [25,26], and tomato [27] have been generated by *Agrobacterium*-mediated transformation using constructs with sense or antisense sequences of target genes. Consequently, focusing on safety assessment of future GM crops, we characterised transgenic *Arabidopsis* plants as a case study with the aim of evaluating current profiling methods for the detection of unintended effects and for evaluation of their potential applicability in safety assessment of GM plants. The six selected transgenic lines showed different types of integration events, where three lines had a low copy number integrated with no repeats, while three had a high number integrated in repeats. These findings are similar to previous findings of multiple T-DNA integrations forming direct repeats at single loci [28,29]. The transgenic *CHS* transcript was not detectable in the three lines with a high copy number integrated in repeats. Recent studies indicate that the mechanisms behind the silencing and variations in expression of inserted genes in transgenic plants are related to the number of inserted genes rather than positional effects and that identical marker sequences above a certain number, ~4-5 in the genome, triggers RNA silencing of the transgenic transcript [30]. This strongly suggest that transcript-level-mediated silencing, via post-transcriptional gene silencing (PTGS), may account for the observed silencing of transgenic *CHS* transcripts in lines *a*, *b* and *d*. However, line *a*, *b* and *d* also showed *nptII* silencing in all progeny of the three lines in all homozygous

generations tested (2 to 5). In contrast to PTGS, TGS is mitotically and meiotically heritable and TGS in GM plants, initiated by repetitive transgene insertion arrays, have been reported frequently [31,32]. Thus, it cannot be excluded that the *nptII* gene in line *a, b* and *d* has been immediately subjected to TGS in the primary transformants by *de novo* methylation. This may also be the case for the region of the exogenous *CHS*, resulting in TGS of this transgene. However, until further research has been performed, it remains unclear whether PTGS, TGS, or both types of silencing are involved in the observed *nptII* and transgenic *CHS* silencing in lines *a, b* and *d*.

Publications of large sets of T-DNA insertion sites reveal that GM lines with a high copy number integrated in a complex structure complicates both the identification and verification of the integration site [33,34,35]. This observation is consistent with our difficulties in the determination of integration sites for some of the GM lines studied here. Results from the molecular analysis of the integration sites as part of the current safety assessment may, therefore, not always give sufficient information to reveal unintended effects from insertion in the GM lines. However, it can be argued that similar mutational events are likely to occur rather frequently in conventional breeding, especially where strategies are applied that intend to enhance the mutation rate, such as chemical mutagenesis. So far, in both traditional and GM plant breeding practices, hardly any observed adverse effects have been seen on the safety and nutritional value of the resulting plant product. It should also be mentioned that even in plants that are well characterised in relation to the locus of insertion, rearrangement outside this locus may very well be present due to naturally occurring recombination. With respect to food safety, we do believe that mutations caused by insertion events should in many cases rank rather low in the hierarchy of potential hazards as, again, this is a regular phenomenon in traditional plant breeding with so far no documented adverse effects that are not eliminated in the subsequent plant breeding phase. Nevertheless the most straightforward approach to assess the effects of insertion events is to start with the analysis of the place of transgene integration. We found that lines *f, d* and probably *c* had their insertion integrated in such a way that encoding gene regions might have been interrupted. All three genes encode products with unknown functions, which therefore give no indications of the effect the insertion might have. In the case of integration of elements of the vector backbone, it is clear that the entire inserted sequence, as well as its expression products, such as selection markers, will be subject to the risk assessment. For markers that have been used in approved GM plant lines there is no evidence that they pose a health risk to humans or domestic animals [7]. Nevertheless, the EU, to meet public concerns, aims to fade out the use of any antibiotic resistance markers in new GM lines.

Targeted analyses of the flavonoid biosynthesis by QPCR and HPLC indicated that the flavonoid synthesis was affected not only at the transcriptional, but also at the metabolic level. Finally, two hybridisation series were performed on the basis of an *Arabidopsis* cDNA array with 1536 sequences that were derived from four different subtractive hybridisation procedures to obtain cDNAs that are specific for *Arabidopsis* leaf, flower and root tissues, as well as cDNAs

that may be involved in metabolic stress responses in the plant. By using the subtractive hybridisation procedure, the aim was to include as many relevant metabolic pathways as possible, as subtractive hybridisation has been shown to enrich the subtracted cDNA pool for differential and lowly expressed genes from the selected tissues, while at the same time reducing the redundancy of abundantly expressed genes (unpublished results). This approach, unlike the theoretically more optimal 'whole transcriptome' array, is currently feasible for all crop species. The advantage of using *Arabidopsis* as a model is the fact that selected sequences are often annotated in public databases, making it more feasible to assess the toxicological and nutritional relevance of any detected differences in gene expression.

Despite the fact that the lines showed different types of integration events, the non-targeted approach, using a cDNA microarray, did not reveal any clear unintended effects. This could be due to the fact that the array did not contain the affected genes or that the sensitivity was too low for detecting all differences, or simply that no unintended effects were present in the plants. We did, however, find that several stress-related genes were deregulated. A range of known plant genes involved in cellular functions such as detoxification, transport, and metabolism, and also in transcriptional regulation are highly affected by stressful conditions (reviewed by Bray [36] and Reymond [37]), and it is therefore not unexpected to find stress-related genes are differently expressed in different plants unless the growing conditions for all plants are exactly the same, which is almost never the case. Also, given the limited extent of our experiments, it is very unlikely that the entire scale of the biological variation in gene expression of stress-responsive genes under 'normal' conditions was reached. As a representative for the differentially expressed genes found by the non-targeted approach, the expression of *VSP1* was analysed in a second set of plants. Here, it was revealed that the pattern was not consistent and it did not confirm the microarray results obtained from the first set of plants. This can be explained by a large biological variation for this gene, and it also reveals that expression of stress-induced genes can vary greatly. The fact that both up- and downregulation of the *VSP1* gene was seen in the GM lines compared to the control lines makes it even more feasible that the natural variation in traditional lines was not caught in the control samples.

Le Gall *et al*. [38] analysed the flavonoid content in leaf material from exactly the same plants as used for our microarray experiment. In this batch (Set 1), Le Gall's group found a severe to mild depletion of four kaempferols in all six lines, including line *d*, which did not have any flavonoid depletion in seedlings in this study. As phenols, including flavonoids, function as ultraviolet (UV) screening pigments that reduce the penetration of UV-B into mesophyll tissue [39,40], it could be hypothesized that the relative high number of genes involved in the photosynthetic apparatus, identified to be differentially expressed in our GM lines, is an indirect consequence of the altered flavonoid synthesis. If this is the case, it may be designated as an unintended, although expected and likely effect, but further investigations are necessary in order to verify this hypothesis.

Our results indicate that the microarray technique is able to detect differences, but since the expression for stress-related genes can vary greatly, such genes might be less useful to include in the array for revealing unintended effects, unless more is known about the variation under different conditions. Traditionally a two-fold induction or reduction has been used to select differentially expressed genes in microarray experiments, and therefore a two-fold criterion was also used in this experiment. This approach can be argued to be somewhat undifferentiated, but necessary as long as the knowledge of the variation of each gene is not clear. Therefore, if the microarray technique should be useful for identifying unintended effects, the natural variation in gene expression of individual genes must be known or needs to be included in the experimental set-up, leading to high numbers of hybridisations, that were beyond the scope of this project.

As part of the *Arabidopsis* work in the EU project GMOCARE, proteomic, glycomic and metabolomic analyses were also performed on leaves from the six GM lines. Results from protein quantification analysis by two-dimensional electrophoresis (2-DE) (S.O. Kärenlampi, pers. communication), metabolic profiling using 31P-, 1H and/or 13C-NMR (Le Gall *et al*., 2005 and J. Leguay, pers. communication), and protein glycosylation by mass spectrometry (MALDI-TOF MS) (F. Altmann, pers. communication) did not reveal significant modification in the GM lines compared to the corresponding wild-type (data not shown), but the different 'omic' approaches revealed that environmental conditions and genetic differences had a large impact on the expression of several proteins and metabolites. These findings are consistent with our observations, and this further underlines the necessity to have sufficient insight into the natural variation to be able to use 'omics' data in food safety assessment procedures. Publicly available databases for data mining on specific genes or proteins, focusing on different species and metabolic pathways and based on well-standardised arrays and protocols can in the near future play an important part in evaluating this information. Initiatives towards harmonised approaches for microarray databases have already been launched with MIAME (minimal information about a microarray experiment) [41, 42].

In relation to food safety aspects, the microarray technology seems to have the potential to reveal differences in gene expression, and the advantage of the microarray approach is that it can contain large numbers of expressed sequences, up to full coverage of the transcriptome. In that sense, it can be a powerful tool to trace potential unintended alterations in a GM plant. However, as seen from our results, the endogenous *CHS* increase in seedlings from line *a, b* and *c* is not correlated with the flavonol amount (not modified or lower). This indicates the problem mentioned by others [43], that differences in RNA levels might not be directly correlated with effects on other levels e.g. the protein level. These differences can therefore not be used directly for the evaluation of food safety aspects unless such a correlation has been established. In this sense, gene expression profiling should be regarded as an informative screening method, and detected differences in gene expression profiles on the basis of the cDNA microarray analysis should be considered as indicative. Any differences detected by the microarray technology will therefore in general need to be confirmed by targeted analyses and,

when confirmed, need to be further assessed for their toxicological and nutritional implications in the light of our current knowledge on the natural variation in the respective metabolic pathways. In conclusion, it can be stated that as long as no method is available for analysis of the full proteome and/or metabolome, the microarray approach for analysis of gene expression profiles has the potential to be a useful tool for screening for unintended effects in a wide variety of metabolomic routes within a growing number of food plant species.

Acknowledgements

We are grateful to Heidi B. Letting and Gunnar R. Rasmussen for technical assistance. This work was financially supported by grants from the European Commission 5th Framework Programme (Project QLK1-1999-00765, GMOCARE).

References

1. Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organism and repealing Council Directive 90/220/EEC. Official Journal of the European Communities L106: 1-39. http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/l_106/l_10620010417en00010038.pdf.
2. EU (2005) Applications under Regulation (EC) N° 258/97 of the European Parliament and of the Council, as of April 2005. http://europa.eu.int/comm/food/food/biotechnology/novelfood/app_list_en.pdf
3. Joint Research Centre (2006). Deliberate releases and placing on the EU market of Genetically Modified Organisms (GMOs). http://gmoinfo.jrc.it/gmc_browse.asp
4. EFSA (2004) http://www.efsa.eu.int/science/gmo/gmo_guidance/catindex_en.html
5. OECD (1993) Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles. Organization for Economic Cooperation and Development, OECD, Paris.
6. WHO (1996) Biotechnology and food safety. Report of a Joint FAO/WHO Expert Consultation on Biotechnology and Food Safety. Rome.
7. WHO (2000) Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology. WHO, Geneva.
8. FAO (2003) ftp://ftp.fao.org/es/esn/food/guide_plants_en.pdf.
9. Cellini, F., Chesson, A., Colquhoun, I.J., Constable, A., Davies, H.V., Engel, K.H., Gatehouse, A.M.R., Kärenlampi, S., Kok, E.J., Leguay, J.J., Lehesranta, S., Noteborn, H.P.J.M., Pedersen, J., Smith, M. (2004) Unintended effects and their detection in genetically modified crops. Food and Chemical Toxicology 42, 1089-1125.
10. Matzke, M.A. and Matzke, A.J.M. (1995) How and Why Do Plants Inactivate Homologous (Trans)genes? Plant Physiology 107(3), 679-685.
11. Wang, H., Provan, G.J., Helliwell, K. (2000) Tea flavonoids: their function, utilisation and analysis. Trends in Food Science & Technology. 11, 152-160.
12. van den Berg, H., Faulks, R., Granado H.F., Hirschberg, J., Olmedilla, B., Sandmann, G., Southon, S., Stahl, W. (2000) The potential for the improvement of carotenoid levels in foods and the likely systemic effects. Journal of the Science of Food and Agriculture 80, 880-912.

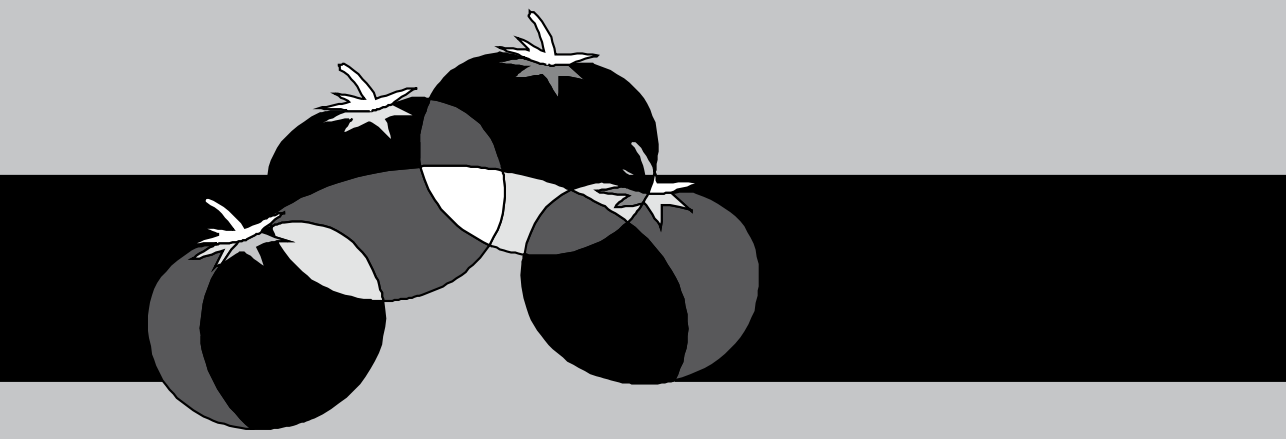
13. Cos, P., De Bruyne, T., Hermans, N., Apers, S., Berghe, D.V., Vlietinck, A.J. (2004) Proanthocyanidins in health care: current and new trends. Current Medicinal Chemistry 11(10), 1345-1359.
14. Ducreux, L.J.M., Morris, W.L., Hedley, P.E., Shepherd, T., Davies, H.V., Millam, S., Taylor, M.A. (2005). Metabolic engineering of high carotenoid potato tubers containing enhanced levels of b-carotene and lutein. Journal of Experimental Botany 65(409), 81-89.
15. Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. The Plant Journal 16(6), 735-743.
16. Balzergue, S., Dubreucq, B., Chauvin, S., Le-Clainche, I., Le Boulaire, F., de Rose, R., Samson, F., Biaudet, V., Lecharny, A., Cruaud, C., Weissenbach, J., Caboche, M. and Lepiniec, L. (2001) Improved PCR-Walking for Large-Scale Isolation of Plant T-DNA Borders. Biotechniques 30, 496-504.
17. Roche Molecular Biochemicals (2000) LightCycler Operator's Manual. Version 3.5. Oktober 2000.
18. Justesen U, Knuthsen P, Leth T (1998). Quantitative analysis of flavonols, flavones, and flavanones in fruits, vegetables and beverages by high-performance liquid chromatography with photo-diode array and mass spectrometric detection. Journal of Chromatography A, 799, 101-110.
19. Rossel, J.B., Wilson, I.W., Pogson, B.J. (2002) Global changes in gene expression in response to high light in Arabidopsis. Plant Physiology 130, 1109-1120.
20. Mahalingam, R., Gomez-Buitrago, A., Eckardt, N., Shah, N., Guevara-Garcia, A., Day, P., Raina, R., Fedoroff, N.V. (2003) Characterizing the stress/defense transcriptome of Arabidopsis. Genome Biology. 4(3), R20.
21. Staswick, PE, Su, W, Howell, SH. (1992) Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an *Arabidopsis thaliana* mutant. Proc Natl Acad Sci U S A. 89(15), 6837-6840.
22. Berger, S., Mitchell-Olds, T., Stotz, H.U. (2002) Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis thaliana*. Physiologia Plantarum 114, 85-91.
23. Jiménez-Bermúdez, S., Redondo-Nevado, J., Muñoz-Blanco, J., Caballero, J.L., López-Aranda, J.M., Valpuesta, V., Pliego-Alfaro, F., Quesada, M.A., Mercado, J.A. (2002) Manipulation of Strawberry Fruit Softening by Antisense Expression of a Pectate Lyase Gene. Plant Physiology 128(2), 751-759.
24. Dandekar, A. M., Teo, G., Defilippi, B. G., Uratsu, S. L., Passey, A. J., Kader, A. A., Stow, J. R., Colgan, R. J., James, D. J. (2004) Effect of down-regulation of ethylene biosynthesis on fruit flavour complex in apple fruit. Transgenic Research 13, 373-384.
25. Rung, J.H., Draborg, H.H., Jørgensen, K., Nielsen, T.H. (2004) Carbon partitioning in leaves and tubers of transgenic potato plants with reduced activity of fructose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase. Physiological Plantarum 121(2), 204-214.
26. Fulton, D.C., Edwards, A., Pilling, E., Robinson, H.L., Fahy, B., Seale, R., Kato, L., Donald, A.M., Geigenberger, P., Martin, C., Smith, A.M. (2002) Role of granule-bound starch synthase in determination of amylopectin structure and starch granule morphology in potato. Journal of Biological Chemistry 277(13), 10834-10841.
27. Lashbrook, C. C., Giovannoni, J. J., Hall, B. D., Fischer, R. L., and Bennett, A. B. (1998) Transgenic analysis of tomato endo-β-1,4-glucanase gene function. Role of cel1 in floral abscission. The Plant Journal 13, 303-310.
28. De Neve, M., De Buck, S., Jacobs, A., Van Montagu, M., Depicker, A. (1997) T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs. The Plant Journal 11(1), 15-29.

29. De Buck, S., Jacobs, A., Van Montagu, M., Depicker, A. (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. The Plant Journal 20(3), 295-304.
30. Schubert, D., Lechtenberg, B., Forsbach, A., Gils, A., Bahadur, S., Schmidt, R. (2004) Silencing in Arabidopsis T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. The Plant Cell. 16, 2561-2572
31. Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A. and Matzke, A.J.M. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. The EMBO Journal 19(19), 5194-5201.
32. Qin, H., Dong, Y., von Arnim, A.G. (2003) Epigenetic interactions between Arabidopsis transgenes: characterisation in light of transgene integration sites. Plant Molecular Biology 52(1), 217-231.
33. Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., Goff, S.A. (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell 14(12), 2985-2994.
34. Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M., Lecharney, A. (2002) T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites. EMBO Reports 3(12), 1152-1157.
35. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301(5633), 653-657.
36. Bray, E.A. (2002) Classification of genes differentially expressed during water-deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data. Annals of Botany 89, 803-811.
37. Reymond, P. (2001) DNA microarrays and plant defence. Plant Physiology and Metabolism 39, 313-321
38. Le Gall, G., Metzdorff, S.B., Pedersen, J., Bennet, R.N., Colquhoun, I. (2005) Primary and secondary metabolites profiling in *Arabidopsis thaliana* (L.) ecotype Wassilewskija with an antisense chalcone synthase transgene. Metabolomics 1(2),181-198.
39. Landry L.G., Chapple C.C., Last R.L. (1995) Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiology 109(4), 1159-1166.
40. Booij-James, I.S., Dube, S.K., Jansen, M.A., Edelman, M., Mattoo, A.K. (2000) Ultraviolet-B radiation impacts light-mediated turnover of the photosystem II reaction center heterodimer in Arabidopsis mutants altered in phenolic metabolism. Plant Physiology 124(3), 1275-1284.
41. Brazma A. (2001). On the importance of standardisation in life sciences. Bioinformatics 17, 113-4.
42. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature Genetics. 29, 365-71.
43. Nordic Council (2003) Use of cDNA microarray technology in the safety assessment of GM food plants. TemaNord 2003:558. ISBN 92-893-0955-5

Chapter 7.

Abstract

The second generation of genetically modified (GM) plants that are moving towards the market are characterized by modifications that may be more complex and traits that more often are to the benefit of the consumer. These developments will have implications for the safety assessment of the resulting plant products. In part of the cases the same crop plant can, however, also be obtained by 'conventional' breeding strategies. The breeder will decide on a case-by-case basis what will be the best strategy to reach the set target and whether genetic modification will form part of this strategy. This article discusses important aspects of the safety assessment of complex products derived from newly bred plant varieties obtained by different breeding strategies. On the basis of this overview, we conclude that the current process of the safety evaluation of GM versus conventionally bred plants is not well balanced. GM varieties are elaboratedly assessed, yet at the same time other crop plants resulting from conventional breeding strategies may warrant further food safety assessment for the benefit of the consumer. We propose to develop a general screening frame for all newly developed plant varieties to select varieties that cannot, on the basis of scientific criteria, be considered as safe as plant varieties that are already on the market.



Comparative safety assessment of plant-derived foods

E.J. Kok, J. Keijer, G.A. Kleter and H.A. Kuiper

Regulatory Toxicology and Pharmacology 50 (2008) 98-113.

7.1 Introduction

Novel foods are defined as foods or food ingredients that have 'hitherto not been used for human consumption to a significant degree within the Community' (EU Regulation 258/97). A number of categories are included: among others foods or ingredients with a new or modified primary molecular structure, novel plant products that do not have a history of safe use and products that have been obtained by a new production process that changes the compositional characteristics of the plant product significantly.
A genetically modified organism (GMO) is 'an organism in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination' (EU Directive 2001/18). Techniques that fall under this definition of 'unnatural' include recombinant DNA techniques, methods for direct introduction of DNA and cell or protoplast fusion techniques. In vitro fertilisation, natural transformation and polyploidy induction are excluded from the definition.

One of the central questions regarding the safety assessment of GMO-derived products is whether they require a different type of food safety assessment than other novel food products. This stems from the argument that the technology to produce GM-plants makes it feasible to cross genetic barriers that can not (easily) be crossed by conventional breeding strategies. The added specification of the technologies involved in the legal GMO definition, i.e. the use of recombinant DNA techniques, methods for direct introduction of DNA and cell or protoplast fusion techniques, already shows that the statement on the potential to cross species barriers is not sufficient to identify the group of techniques. In fact, characteristics of the different breeding strategies indeed overlap and are often not well defined.

In recent years, developments in plant breeding have led to an increased interest in breeding varieties with a direct advantage for the consumer. This development holds pace with other developments in the food industry to meet current consumer trends for functional, healthy and tasty foods. As a result of this, the complexity of the food supply increases, leading to new questions regarding the safety and nutritional characteristics of these different types of novel foods, which have never been asked before. These developments do not leave the plant breeding industry untouched.

In this review we discuss the different characteristics of food products derived from novel plant varieties that have been obtained by different breeding strategies as well as the consequences this may have for the food and feed safety assessment of these products.

7.2 Current plant breeding strategies

7.2.1 Classical plant breeding

Classical plant breeding is based on genetic diversity. If the desired traits to improve agricultural crop varieties are already present within the species, improved plant varieties can be obtained through crossing of a well-performing line with a line with the desired trait. If the desired characteristic is not within the species, but in distantly related varieties, a number of cell tissue culture techniques may help the plant breeder to obtain fertile generations from normally sterile crossings. If the desired trait is also not present in distantly related lines then, in some cases, mutant lines with the desired characteristics can be obtained by chemical mutagenesis using substances such as the alkylating agents ethane-methyl-sulphonate (EMS) [1] or N-ethyl-N-nitrosourea (ENU) [2], or by mutagenesis based on irradiation [3] or the use of transposons, i.e. naturally occurring transposable genetic elements [4]. The different agents may have different characteristics: for instance the application of EMS in Arabidopsis lines produces primarily C to T changes in the DNA, resulting in C/G to T/A transition mutations in the DNA molecules [5]. The International Atomic Energy Agency [6] mentions the release of at least 620 cultivars of 44 crop species that were produced with the application of nuclear technologies to obtain the desired mutations in the last few decades in China alone. The total number of (irradiation) mutant varieties and mutant-derived varieties that was registered in the FAO/IAEA Mutant Varieties Database at the end of 2001 was 2276. In their report the IAEA suggests, that this may be only the tip of the iceberg, as further information is still missing on the actual numbers of mutant varieties that were released in different crop species and also on mutagen agents that were applied, selection processes and the use of mutants in cross breeding. Numbers of chemical or transposon mutant varieties in crop species are not available. Large volumes have been published on classical and modern plant breeding (e.g. Agrawal [7] and Jain and Kharkwal [8]. At the same time Wilson *et al*. [9] find that little exact information is available on current plant breeding schemes to obtain novel plant varieties.

The increasing knowledge on the genetic constitution of production organisms, including crop varieties, has led to the identification of many quantitative trait loci (QTLs), i.e. genomic regions that are associated with phenotypic differences in values for specific traits. Usually a continuous distribution of values for these traits is seen in the segregating progeny, which is considered to be due to relatively small changes in several related genes. This knowledge is nowadays often used to apply marker-assisted-selection (MAS) in traditional plant breeding strategies to select for beneficial QTLs and thus obtain improved elite crop varieties [10].

7.2.2 Genetic modification

Genetic modification, similar to classical breeding, aims to alter metabolic routes in order to render the resulting plant varieties more favourable characteristics in terms of, for instance, agronomic, nutritional and/or processing quality. Alteration of metabolic pathways is usually

still limited to the insertion of single genes that either express a new protein or enhance or reduce the expression of genes already present in the pathway. The introduction of more than one trait is often achieved by crossing individual single-gene GM lines, resulting in so-called stacked gene varieties. However, the first new plant varieties with simultaneous insertion of multiple genes are now in development, the most well-known being the 'golden rice' lines [11]. In the introduction of multigenic sequence also new strategies are employed. Quesada-Vargas *et al*. [12] show for example that chloroplasts can process multigenic sequences that are effectively expressed via the chloroplast genome without significant intervention of the chloroplast regulatory systems. The expression or manipulation of multiple genes is, however, still difficult and forms one of the major technical challenges for further extension of the potential of gene technology in plant breeding strategies [13].

Homologous recombination, a common phenomenon in the plant that serves as a DNA repair mechanism, may in the future be applied also as a means for routinely directing genetic alterations [14]. It has been shown already that there are substantial differences between numbers of homologous recombination events depending on the genomic positions of the inserted transgene [15,16]. This may hamper the future applicability of homologous recombination in breeding strategies that aim to alter metabolic routes in other plant parts or tissues. Further research will need to provide insight into the potential of this approach for directed genetic modification as part of plant breeding schemes.

Another development within recombinant DNA technology strategies that is gaining importance is the development of cisgenic plants, i.e. plants that have been genetically modified with a natural gene from a sexually compatible plant. The largest advantage of the application of cisgenesis compared to conventional breeding in the case of sexually compatible plants, is that the gene of interest can be transferred in a 'clean' way and the so-called 'linkage drag' of deleterious genes associated with the desired trait in e.g. a wild relative will not hamper or impede the breeding process [17].

Thomson [18] identifies four specific areas in plant breeding where gene technology may contribute: abiotic stress resistance, biotic stress resistance, herbicide resistance and food crops with improved nutritional characteristics. In 2003, the International Council for Science (ICSU) [19] perceives four categories of potential health benefits from novel plant products: enhanced food security, enhanced nutrient security, targeted health benefits and reduction of diet-related diseases. Different reviews [20,21] provide an overview of developments in gene technological applications in plant breeding. An interesting development is the production of low-allergen (or even allergen-free) crop varieties. Examples are low-allergen wheat varieties [22] and hypo-allergen apples [23,24]. At the same time, an increasing number of different applications of novel plant varieties for industrial products can be envisaged [25]. This includes altered composition of plants with respect to oils, starch, fibre, protein, but also includes plants that may produce specific chemicals, natural polymers, pharmaceuticals, decontamination agents, or fuels. The website of the International Service for the Acquisition of Agri-biotech

Applications (ISAAA, www.isaaa.org) provides an up-to-date overview of commercialised GM varieties worldwide.

The two current developments in gene technology applications in the food sector, i.e. the more complex alteration of metabolic routes using multiples genes and the aim to focus on consumer-related characteristics, is often referred to as the second generation of genetically modified food products [20]. This may benefit the Western world, but also consumers in developing countries [26]. Gene technology has the clear potential to improve the health and nutritional status of consumers in developing countries, by increasing the available volume of staple foods, but perhaps more importantly, by reducing the prevalent micronutrient under-nutrition in these countries. This can be achieved by either increasing the levels or the bioavailability of micronutrients in new plant lines [21,27]. Examples of such developments are the often cited rice with increased levels of vitamin A precursors [11,28,29], (cooking) plant oils with improved fatty acid composition [22,30,31], grains enriched with vitamin E [32,33], vegetables with enhanced folate levels, and cassava with improved nutritional characteristics [22,34].

7.3 Safety assesment of plant-derived food products

7.3.1 Safety assessment of single compounds

The safety assessment of *defined food constituents* or mixtures of food chemicals is usually performed according to the traditional toxicological concept of risk characterisation as part of the overall risk assessment, including the concepts of hazard identification, hazard characterisation and exposure assessment [35]. The two possible outcomes, in case of toxicity, are that there is either a threshold effect and, for instance, an acceptable daily intake (ADI) can be established, or there is a non-threshold effect, in which case, for instance, a virtual safe dose (VSD), i.e. the intake corresponding to an estimated risk of one in a million, can be established [35,36]. In 2002 the results of the FOSIE (Food Safety in Europe) project of the International Life Sciences Institute (ILSI) were published [37]. This project examined the available qualitative and quantitative methodologies to assess risks from food chemicals and provides a valuable overview of tools, established and in development, for risk assessment of food chemicals, including animal-based toxicology [38], in-vitro toxicology [39], hazard characterisation [40,41], and exposure assessment [42].

With respect to the safety assessment of nutrients in fortified foods and nutritional supplements, it is reasoned that the classical approach can not be fully applied to nutrients, as adverse effects will result both from intakes that are too low (deficiency) or too high and both effects need to be taken into account [40]. This generally leads to two reference levels, the recommended daily allowance (RDA) or reference dose, which is associated with a low probability of deficiency and the tolerable upper intake level (UL), which sets the limit for intakes with a low probability of toxic effects. The range between the two levels may be small and varying for different subpopulations [40]. Also in the case of macronutrients or (other

complex) food ingredients this shows that both toxicological as well nutritional aspects need to be taken into account.

7.3.2 Safety assessment of plant-derived food products

In practice, the evaluation of plant-derived food products derived from novel plant varieties for human consumption has limited traditional background yet. In exceptional cases specific defined components were routinely assessed in new plant varieties for food safety purposes, usually after earlier food incidents with newly developed crop varieties. Examples of this are the routine analysis of glycoalkaloid levels in new potato varieties after food incidents with new potato varieties that proved to have high levels of glycoalkaloids [43], the analysis of (new) leavy vegetables for nitrate levels, as this is a health concern in especially the northern countries in parts of the year [44], and the analysis for psoralen levels in new varieties of the celery family [45]. In other cases new plant products could usually be brought onto the market without a specific food safety assessment. The general assumption was that newly bred varieties that are based on crop varieties with a history of safe use would not cause any reason for a food safety concern. Nowadays regulations such as the European Novel Foods Regulation (258/97) do require specific data to assess novel plant products that are not GMO-derived. This includes all plant-derived products that were not on the EU market in substantial amounts prior to the enforcement of the Novel Foods Regulation in 1997. If it is decided that a novel plant product is to be assessed for its food safety, its assessment will, to a certain extent, be comparable to the food safety assessment of GM-plants. So far there have been very few cases of safety assessment procedures under the Novel Foods Regulation for 'conventionally bred' crop varieties with altered compositions, such as the (extract of) high-lycopene tomatoes (http://www.food.gov.uk/multimedia/pdfs/lycosummary.pdf). Moreover, in practice only plant products with *intended* compositional alterations are assessed under this regulation. Compositional analysis of novel plant varieties for *unintended* physiological changes as a result of breeding procedures is not routinely performed under this regulation, but in principle it would follow the same approach as described below.

With the advent of irradiated food products [46,47], genetically modified organisms and their derived food products, and functional foods, it became necessary to determine whether the classical toxicological approach can also be applied to this type of food products. The safety assessment strategies for new (GM) plant varieties was discussed in the scientific platforms of IFBC, FAO, WHO, OECD, EU and ILSI (reviewed by Kuiper *et al*. [48] and König *et al*. [49]). More or less in parallel, similar developments were seen in the area of non GM-derived functional foods, such as botanical mixtures [50,51]. The result of this is a worldwide consensus on the basic approach for the safety evaluation of plant-derived (GM) foods, the so-called comparative safety assessment [52]. In 2004, the results of the European project Entransfood, a scientific consortium with over 60 members from 14 countries, on the safety assessment of genetically modified food crops were published [53]. An extensive overview was provided on the different aspects of the safety assessment of GMO-derived foods, including a general overview of food

safety assessment strategies [49], an overview of methodologies to detect unintended side effects in GM crops [54], and potential gene transfer [55].

An important part of this globally harmonized approach is the hazard identification phase. During this phase, the novel crop is extensively compared with its traditional counterpart, preferably the direct parent line, which has a so-called 'history of safe use'. In practice, this latter aspect will, however, be difficult to determine for the parent varieties involved, as the necessary data to determine this 'history of safe use' will often be poorly documented. The process of genetic alteration as a result of the breeding process is routinely scrutinized to identify hazards associated with it. In the case of new GM varieties, the genetic sequences that are incorporated into the GM-plant, and their expression products, will be analysed. This includes extensive in-silico analysis, as well as an analysis with relation to the mode of introduction of the genetic sequences into the recipient host genome and the place of insertion into the host genome. The comparison also includes the general phenotypic characteristics of the new plant line and the parent line(s), as well as an elaborate compositional analysis of key macro- and micronutrients, anti-nutrients and natural toxins.

The OECD has developed so-called consensus documents for important crops that summarize the existing knowledge on the crop and its key components for use during the regulatory assessment of a particular food or feed product [56]. These key components (nutrients and anti-nutrients) are used to screen the plant product for unintended effects of the breeding process. Based upon this basic information further hazard characterisation is performed, including the assessment of potential toxicity and/or allergenicity of the newly expressed products, the effects of potential gene transfer and of other potential unintended effects in the whole crop that may be related to the genetic modification. Further extension of these OECD consensus documents with additional key components related to (crop-specific) metabolic networks would be valuable for the assessment of the so-called predictable unintended effects when breeding aims to modify these specific networks.

Hazard characterisation of plant-derived food products, however, has its limitations. For example, the compositional analysis of key components may detect differences that are unrelated to the genetic alterations, but that may be due to the set-up of the analytical comparison, e.g when the growing conditions differ slightly between the sets of samples. On the other hand, differences between the lines under investigation may remain unnoticed, because the differential components may not be regarded as key components, and are thus not measured, or the natural variation between individual plants is already large for a specific component, reducing the sensitivity of the analytical approach for small changes in the limited number of plants to be analysed. But the differences may also not show up under the selected growing conditions. To overcome this latter aspect, different international advisory reports [57,58,59] advocate the compositional analysis to be performed on samples from plants that

were grown under a range of different environmental (different locations) and climatological (different years of growth) conditions.

Depending on the outcomes in the hazard identification phase and the outcome of the compositional analysis, additional toxicity testing may be required in the hazard characterisation phase. This may include a range of specific toxicological tests, including in-silico testing, in-vitro digestibility testing and different types of in-vivo animal toxicity studies, ranging from subacute to chronic toxicity testing. Which studies are most appropriate for adequate hazard characterisation will need to be determined on a case-by-case basis. Here it should be taken into account that the extended dose ranges that are commonly used for testing single compounds in animal studies can in most cases not be applied to the testing of food plant products due to possible nutritional imbalances in the diet and diet formulation technical requirements [49,60]. As a result of this, in the case of whole foods plant products, the sensitivity of animal toxicity studies may be compromised to the extent that the aimed results are no longer feasible, especially when the differences are small. In this light it may need to be considered whether in specific cases current analytical developments, once validated, may not be better apt to detect and characterise (unintended) effects in new plant products (see also 7.5.2). Nevertheless, the toxicological characterisation of the identified differences will, in all cases, remain the crucial final step. An overview of the history of animal-based toxicology as well as an informative listing of available tests with their strengths and weaknesses, as well as current developments in this area is provided by Barlow *et al*. [38]. The EFSA Working Group on Animal Feeding Trials [61] provides more recent information on the prospects and limitations of animal feeding trials in the hazard identification and/or assessment of food products. This working group concluded that the testing of whole plant-derived foods in animal feeding trials is only indicated when the composition of the (GM) plant is substantially modified or when there are indications for or remaining uncertainties with respect to the occurrence of unintended effects. Barlow *et al*. [38] state that more emphasis will need to be placed on food safety assessment strategies for the situation where nutritional and toxicological doses come in close ranges. They advocate more emphasis in these cases on *human* pre-marketing studies as well as post-marketing surveillance.

In addition to the hazard characterisation, data on the potential intake of the new plant products in comparison to the counterpart will need to be obtained. This should include information on general exposure assessment within the population, but also information on potential intake for specific consumer groups, where relevant. In this context, information from food consumption surveys carried out in different (sub)populations in different parts of the world is of great importance. The food intake assessment data will need to be integrated with the results of the hazard characterisation in the risk characterisation. In practice, this may lead to additional questions with regard to, for instance, the composition of the novel plant product and may therefore prove to be an iterative process until sufficient insight has been gained to complete the risk characterisation phase.

To equate the first phase of the safety assessment of all novel plant products, a comparative compositional analysis of newly bred plant varieties with near comparators should become a standard requirement. This is much in line with developments in other areas of food production, where producers need to know the detailed composition of the products that they want to put on the market. This initial comparison could provide insight in possible changes in the physiology of the plant as a basis for further safety assessment, if needed.

7.4 Regulatory aspects of plant-derived food products

7.4.1 Common basis

The general strategy to assess GMO-derived plant products as a separate food product group has nowadays been incorporated in regulations in many countries. Below, we describe the current regulations for novel plant varieties, including GM-plants, in the EU, the US and Japan. Despite the common ground, regulatory differences are observed and may form the basis for prolonged trading disputes between countries and international organisations [62]. These disputes, however, are often on minor aspects and tend to conceal the fact that in practice there is, to a large extent, consensus between scientists worldwide on how to evaluate (GMO-derived) novel plant products.

Europe

In Europe different regulations have been formulated for GM foods and feeds (Regulation 1829/2003) and other types of novel foods, including newly developed plant products or plant products that were not yet on the European market in significant amounts (Regulation 258/97). Regulation 258/97 describes the risk assessment procedures of all types of novel food products, including novel plant products, but, since 2004, has excluded GM-derived plant products. The basis for the risk assessment under regulation 258/97 is a comparative safety assessment, where relevant, and a full toxicological and nutritional assessment of novel plant products, if no traditional counterpart is available for the comparative approach. In practice only novel plant varieties with a large intended compositional effect nowadays go through the 258/97 authorisation procedure. The regulation is technology-based, products derived from new traditionally bred plant varieties are not assessed for their safety on a routine basis.

Regulation 1829/2003 provides the legal basis for the market approval process for GM foods and feed, including food and feed ingredients, additives, or enzymes [63]. The European Commission will seek advise from the European Food Safety Authority (EFSA, scientific panel on genetically modified organisms), as well as solicit comments from member states. The dossier information requirements may vary from case to case and guidance in this respect is provided by the 'guidance document for the risk assessment of genetically modified plants and derived food and feed' [59]. The basis for the approval procedure is also a comparative safety assessment of the newly

developed GM line and derived products. The comparator is preferably the direct parent line or a genetically close relative.

US

In the US regulatory system, companies need to demonstrate that a novel (GM) crop has equivalent composition and nutritional status compared to its traditional counterpart, except for the introduced trait. For crops that express pesticidal proteins the US EPA (Environmental Protection Agency) demands that it is shown that the proteins are safe, also in terms of allergenic properties under the Federal Insecticide, Fungicide, and Rodenticide Act, (other) crops and their expression products are evaluated by (voluntary) consultations with the FDA (Food and Drug Administration) under the Food Drug and Cosmetic Act (http://usinfo.state.gov/journals/ites/1005/ijee/regulation.htm). At the same time, companies are liable for the health risks of the new crop varieties [22]. In 1992, the FDA published their position that food and feed derived from GM crops pose no unique safety concerns and should be regulated no differently from other plant products [64]. Products from so-called traditional breeding schemes would need to be evaluated, on a voluntary basis, when their composition would be substantially altered compared to the parent lines [65]. Products derived from new traditionally bred plant varieties are, however, not assessed for unintended effects or their safety on a routine basis. In this respect there is no difference between the US and the EU.

Japan

The Japanese Food Sanitation Law stipulates the requirements for GM foods and additives under the Food Sanitation Law (http://www.jetro.go.jp/en/market/regulations/pdf/food-e.pdf). Since 2001 the safety assessment of GM foods has been mandatory under the Food Sanitation Law. In 2003 the Food Safety Commission was established to conduct the safety assessments of foods including GMO-derived foods and food additives according to the Standards for the Safety Assessment of Genetically Modified Foods (Seed Plants, http://www.fsc.go.jp/english/standardsforriskassessment/gm_kijun_english.pdf). This document states, among others, that animal feeding trials will only be required if the analytical results do not sufficiently confirm the safety of the GMO-derived product. Products derived from traditionally bred plant varieties are not assessed for unintended side effects of the breeding process on a routine basis.

7.5 New developments with relation to the comparative safety assessment of plant-derived food products

In recent years the developments in the plant breeding sector and the subsequent discussions among scientists on adequate safety assessment strategies for (GMO-derived) food plant products, have promoted new developments that aim to improve the tools to assess the safety of food products in a broad sense, not just limited to GM-plant-derived food products. In this

section a number of developments and insights are described that may provide better tools for the evaluation of novel plant products, now or in the near future.

7.5.1 Molecular characterisation

Molecular characterisation of a novel plant variety is, in the case of GM varieties, first performed to identify the exact genetic construct that was inserted into the host plant genome and to assess the possibility of insertional mutagenic effects. Sequence information on the place of insertion is used to assess whether plant genes or regulatory elements have been disrupted or affected in another way. Alternative types of (genome) analysis may provide further information, such as chromatin immunoprecipitation, to detect sequences that bind to transcription factors, methylome analysis, to investigate epigenetic regulatory processes, and polymorphism analysis and genome resequencing [66]. These developments are, however, current research tools and can therefore only be used within food safety evaluation settings in very specific cases to provide further information on specified questions.

In practice, actual sequencing of the construct and the place of insertion is not yet routinely performed as part of the food safety assessment in different parts of the world. The EU requires a characterisation of the introduced sequence and a detailed analysis of the place of insertion. Other countries usually ask more global molecular biological data.

Kohli *et al*. [67] discuss T-DNA insertion and particle bombardment as means to transfer the gene(s) of interest. They conclude that when T-DNA insertion is used, DNA disruptions may occur, that can be minor or major, and extra DNA may be inserted with the intended fragment. When particle bombardment is the method of choice they estimate that in most cases genomic disruption will take place and additional DNA fragments may be incorporated into the host plant genome. In addition, the plant tissue culture step, a necessary part of the transformation procedure, but also an important step in many traditional breeding strategies, alone will result in many genome-wide mutations, i.e. the so-called somaclonal variation [68]. A distinction should be made here between the tissue culture of undifferentiated and dedifferentiated plant cells, as many more mutations can be observed in the latter group [69].

Unintended mutations may also occur in plant varieties obtained via mutation breeding. In this respect it is worth mentioning that in the case of vegetatively propagated crop species all mutations occurring in the same cell, i.e. the desired mutations and the additional mutations, should, in the absence of further recombination steps, be considered as genetically linked [70].

Another phenomenon, the detection of RNA variants as transcribed from the transgenic construct in herbicide resistant soybeans, was described by Rang *et al*. [71]. They showed that the 250 bp fragment of the epsps-gene, coding for herbicide resistance, that was erroneously introduced downstream of the nos-terminator sequences, was also (partly) transcribed in the transgenic soybeans. They showed that the nos-terminator sequence was ignored during transcription.

This may lead to the formation of fusion genes and in some cases to fusion proteins. The nos-terminator sequence is often used in transgenic constructs and it can be envisaged that this phenomenon may therefore occur more often. At the same time it seems plausible that this may also occur in the case of other terminator sequences, and in transgenic as well as non-transgenic settings. Little information is available in literature on these basic questions with relation to molecular biological aspects of plant breeding. Homologous recombination may reduce the chance of this type of unintended side effects of a genetic modification as a result of the insertion of the new genetic sequence into well-characterised parts of the plant genome, but the approach is not yet routinely used in plant breeding.

It seems realistic to state that in any type of breeding procedure unintended genetic effects may occur, the frequency of unintended genetic effects will, however, vary depending on the procedures followed during the subsequent genetic alteration and selection steps as part of the breeding programme. Directed molecular characterisation can only be performed in the case of GM varieties, in other cases unintended genetic effects will only be observed if they lead to distinct phenotypic, including compositional, alterations.

7.5.2 Compositional analysis

A comparative compositional analysis of novel plant-derived products, GM or traditional, with their appropriate counterparts already on the market forms the basis for the detection of any potential unintended effects in the novel plant variety. In order to obtain meaningful compositional data, it is necessary that the data meet a number of quality criteria with relation to the selection of the counterpart for comparison, the setup of the field experiments, the phenotypic characteristics and key compositional parameters to be analysed, the validation status of the applied methods of analysis and the data analysis and reporting.

In different (international) guidelines for the compositional analysis of new plant varieties, including GM lines, it is advocated that plants are analysed that have been grown in different locations that are representative for the area of growth of the commercial varieties, and in subsequent growth seasons [57,58,59]. In practice this recommendation leads to the accumulation of large amounts of data on the composition of the new variety and its counterpart. In many cases statistically significant differences are observed for individual compounds between the new (GM) line and its counterpart, but usually this difference is not consistent over all locations. Originally the idea behind the demand for multiple locations was that potential differences between the new (GM) line may show under some circumstances, but not under other. However, in all cases the exact bandwidth of the natural variation under the selected circumstances will not be known and therefore it will not be possible to establish, on the basis of the limited numbers of plants analysed, that the observed differences may be related to the genetic modification as a result of the followed breeding strategy. It may, however, be questioned whether the current approach to detect unintended effects is sufficiently effective. It may be worthwhile to investigate whether the current numbers of

field experiments underpinning regulatory decisions should not be replaced by a limited set of experiments under controlled conditions that mimic a selected set of informative (extreme) growth conditions that stretch the physiological equilibrium maximally in order to more easily show physiological imbalances as a result of the chosen breeding strategy.

With respect to the key components to be analysed, it is clear that both the important nutrients, anti-nutrients and natural toxins should be included, as well as those compounds that are known to be part of intentionally altered metabolic networks. For comparative purposes an interesting development has been the development of the ILSI Crop Composition Database [72]. The database is online searchable (www.cropcomposition.org) with compositional data that were gathered by six biotechnology companies. It compiles information derived from controlled field trials that were performed on worldwide locations in a 6-year period. Important, and new, was that all analyses were performed with validated analytical methods and mostly under GLP. This makes it, in principle, a database that can be used when assessing the toxicological and/or nutritional relevance of detected differences between plant varieties already on the market and derived new ones, GM or not, under scrutiny. This development complements the international effort by OECD countries to formulate OECD consensus documents that compile general information on individual crop species as well as information on key nutrients, anti-nutrients and natural toxins [56].

The development of analytical profiling or 'omics' technologies is changing the 'analytical world'. These technologies include transcriptomics, i.e. the generation of profiles of mRNA transcripts, proteomics, i.e. the generation of protein profiles, and metabolomics, i.e. the generation of profiles of secondary metabolites. Also glycomics approaches have been developed, that profile the sugar moiety attachments of proteins or sets of proteins. All approaches can be applied to single cells or cell systems, tissues or entire organisms and profiles can include hundreds to several thousands of transcripts, proteins or metabolites, depending on the selected approach. Different projects have (had) the aim to develop these 'omics' technologies for the aim of detecting and identifying differences in plant lines and products thereof compared to the parent line(s) as part of a comparative safety assessment approach. Examples of such projects have been the European GMOCARE project (http://www.entransfood.nl/RTDprojects/GMOCARE/GMOCARE.html), the UK FSA project (http://www.food.gov.uk/science/research/researchinfo/foodcomponentsresearch/novelfoodsresearch/g02programme/g02projectlist/g02001/), and the current European Safefoods project (www.safefoods.nl). In these projects the microarray technology has been assessed to obtain informative gene expression profiles that can be used in the safety assessment of novel (GM) plant varieties [73,74]. Although microarray protocols become more and more standardised, it will always be questionable how relevant the obtained information on transcripts is for the actual physiological status of the plant, which may be more related to shifts in active proteins and/or metabolites. Therefore in the same projects, methods for proteome and metabolome analysis were assessed [54,75,76,77]. For proteomics, 2-dimensional gel electrophoresis in combination with mass spectrometry is still the method

that is usually applied. Chen and Harmon [78] provide an overview of the developments in this area, with emphasis on sample preparation and spectrometric analysis.

Metabolomics may be the optimal approach if comparisons are made for the purpose of the food safety evaluation, as this approach generates profiles of primarily the low-molecular weight molecules, which may be most directly related to the plant phenotype and nutritional and toxicological characteristics. Metabolites will in many cases not be just the end-product of a metabolic route, but are active regulators themselves, making it perhaps even more interesting and relevant for food safety issues to monitor alterations in these networks. For metabolomics usually chromatographic methods in combination with mass spectrometry is used. This approach generates large amounts of information but also complex data to handle. Dixon *et al* [79] describe applications of metabolomics in agriculture with possible consequences for the food and environmental safety assessment. They conclude that despite large progress there is still much work to do before the complex information generated by metabolomics is sufficiently understood to be applied in agriculture on a regular basis. Hollywood *et al*. [80] discuss the specific advantages and drawbacks of metabolomics in comparison to transcriptomics and proteomics. They also consider metabolomics as the approach that is mostly related to the (plant's) phenotype, but bottlenecks are the sensitivity of the metabolomics analyses and specific characteristics of metabolites, their lability, diversity and the general wide dynamic range of their presence in specific tissues. These authors also stress the shift from thinking in terms of metabolic pathways to metabolic networks, or neighbourhoods, but advocate the identification of key metabolite markers for alterations in specific networks as an approach for near future application in a broad range of disciplines.

When applying microarray gene expression analysis to GM lines and conventionally bred wheat lines Baudo *et al*. [81] found many more differences in new varieties that were obtained by conventional breeding strategies compared to similar GM varieties. Although the authors present their results as remarkable, underlining the substantial equivalent nature of the GM lines, these findings are easily explained by the extent of genetic alteration that occurs in conventional breeding compared to genetic modification approaches. Of more scientific interest in that respect is a related study that analysed the metabolome of similar wheat lines under field conditions using NMR and GC-MS [82]. Applying a two-tiered analysis, it was concluded that the differences observed between the GM lines and their control counterparts fall within the range of natural variation in the control lines grown on different sites and in different years. Moreover, differences between the GM lines and the control counterparts were not observed in all field locations, showing the environmental effects on the metabolome that may affect the potential to observe phenotypic differences that are related to genetic alterations. Similar studies are performed by others using targeted metabolite analysis [83] and metabolomics [77,84,85]. These studies show that work still needs to be done in terms of standardisation and validation. At the same time, it also seems clear that these more or less unbiased approaches can become meaningful in risk assessment approaches of plant products that also aim to include unintended effects in novel crop varieties.

While the technological possibilities for application of the 'omics' profiling techniques in the safety assessment of (plant) products constantly improve, this is also the case for the associated bioinformatics tools. Where initially effects were identified based on fold-changes, analysis has now shifted to statistics. Especially methods that allow analysis of coordinate changes allow for identification of several small effects that occur in the same process or pathway or interaction cluster. Indeed, for one of these methods, Gene Set Enrichment Analysis [86], it was elegantly shown that it allowed for identification of functional effects that could not be identified by the examination of individual genes [86,87,88]. Pathway analysis tools also assess coordinate changes, but focus on biological interpretation. Different (web-based) pathway tools exist, for instance PathMAPA [89], Pathway Processor [90], GiGA [91] and KaPPa-view [92]. These programs not only help to identify changes, but may also help to integrate different types of 'omics' data. At present, these are mainly research tools that are being further developed and in time will help to better interpret observed differences in metabolic routes between novel plant varieties and varieties with a 'history of safe use'.

7.5.3 Toxicological assessment

As stated before, current toxicological assessment is mainly focused on well-characterised single compounds and, until recently, not so much on complex products. This means that in specific cases the classical approach can well be applied on single substances that as a result of the breeding process are newly present in the novel plant varieties, provided it is possible to obtain sufficient quantities of the specific product. If the new plant variety has altered macronutrient characteristics, the traditional toxicological approach may not be easily applicable to the altered component(s) individually. The amounts that can be administered in animal studies may be limited because of bulk, palatability or imbalance of the animal's diet [38]. A major challenge in such studies is the formulation of corresponding diets from the novel plant variety and the control variety with the control, standard diet that are identical with respect to all relevant measured components and caloric value [61,93]. Large differences can be seen in the set up of the studies with feeds used containing up to 60% of the crop plant ingredient under investigation. It has been proposed to adjust the test animal diets to be able to increase the incorporation of the foods to be assessed in the diet to levels over 60% [94].

Taking into account these practical aspects as well as ethical issues that are related to the performance of animal toxicity studies, it is clear that this type of study should only be performed if there are clear-cut questions on the basis of the study, i.e. if the study is hypothesis-driven with relation to hazard characterisation, or as a method of reassurance, but only in those cases where the study can be performed with sufficient sensitivity. In all cases where other available (in-vitro or analytical) approaches are likely to result in at least as much information on potential unintended effects in the plant-derived food products, it is obvious that animal toxicity studies will not be the method of choice.

The pro's and cons of toxicogenomics as a future tool for hazard identification and characterisation are summarized by the OECD [95]. In this report, toxicogenomics is defined as 'a study of the response of a genome to hazardous substances, using genomic-scale mRNA expression (transcriptomics), cell and tissue wide protein expression (proteomics) and cell and tissue wide metabolite profiling (metabolomics) in combination with bioinformatics methods and conventional toxicology'. The report states that toxicogenomics can help 'to improve the understanding of mechanisms of toxicity, identification of biomarkers of toxicity and exposure, and possibly alternative methods for chemical screening, hazard identification and characterisation'. Toxicogenomics can already provide additional information on changed profiles in a growing number of specific cell systems as a result to exposure to toxic substances [96]. However, much work still needs to be done before toxicogenomics can be applied on a routine basis, e.g. standardisation of protocols is required and the results should unequivocally be interpretable in terms of risk. Also here it can be foreseen that toxicogenomics profiling will be used in the research phase, while specific validated and interpretable markers derived from toxicogenomic studies are the ones that will ultimately be used in the safety analysis. In the future, in vitro exposures in combination with toxicogenomics profiling may, in specific cases, replace animal testing procedures as part of food safety assessment strategies. However, validation of this approach, which is novel and complex, is a challenge in itself and will likely result in the combined application of the in-vitro and in-vivo approaches for many years to come.

7.5.4 Nutritional assessment

With the advent of novel plant products with altered nutritional characteristics, GMO-derived and non-GMO, insight into suboptimal, nutritional and toxic levels for a range of food constituents becomes necessary. While this is already very difficult for most macro and micronutrients, it becomes a clear challenge when also secondary metabolites, such as polyphenols, are taken into account. At the same time it can be expected that in the western world most attention will be focussed on exactly these bioactive food components, in view of their potential beneficial health properties [97]. Not only do a general RDI and UL need to be defined, it is clear that these limits will differ between different consumer groups: elderly, children, pregnant and breast-feeding women, diseased people, vegetarians, groups varying in genetic constitution, etc. [42,98,99]. As a consequence of the nutritional assessment being an integral part of the newly developed safety assessment strategies for novel food products, it became necessary to have reliable validated food-nutrient databases as well as food consumption data for different parts of Europe, or in global perspective, of the world. In response to this, the European project EuroFIR is setting up a validated 'food table' for macronutrients, micronutrients and bioactive food components. At the same time, a global initiative is setting up the FAOSTAT database by the UN Food and Agriculture Organization, which provides consumption data for 200 primary commodities in 200 countries and regions, but not yet for specific types of consumer groups (http://faostat.fao.org/site/345/default.aspx).

A major achievement in the area of exposure assessment has been the development of probabilistic approaches that enable safety evaluators to assess the entire range, from minimum to maximum, of potential exposure, as well as their likelihood of occurrence. The underlying data can be based on empirical as well as theoretical distributions. The approach provides data on potential exposure for the consumer population at large or specific consumer groups, but also gives insight into the associated uncertainties of the exposure assessment [98,100]. The necessary further perfection of the probabilistic approach obviously depends on the availability of appropriate food consumption data of good quality [101].

7.5.5 Allergenicity

Current approaches to test for potential allergenicity are based on the assessment of a series of characteristics of the novel gene product, such as source, structural similarities, digestibility, degradability, and the results of tests such as antisera binding tests, animal test models and clinical tests [102]. In 1993, the often-cited brazil nut experiment took place where a gene from brazil nut was transferred to a soy variety to improve its nutritional value. Experiments in the developmental phase showed, however, that the resulting soy variety tested positive in the serum assay for allergenicity by cross-reactivity with Brazil nut [103]. The development of the soy variety was then stopped. Lehrer and Bannon [104] state that the current risk assessment strategy with relation to potential allergenicity has resulted in the absence of transgenic proteins in foods that have been shown so far to cause allergic reactions. This indicates that the current approach is robust. At the same time they state that new information and technologies will further help the assessment process of novel proteins for potential allergenicity. Similarly Orruño and Morgan [105] also conclude that further efforts are needed to develop new systems for the prediction of allergenic characteristics of (novel) food proteins, as well as for the derived peptide sequences, especially in the light of the development of future crop varieties with more profound metabolic alterations.

An inventory on current knowledge on predictive assays for the potential allergenicity of (novel) proteins concludes that no assay is completely predictive [106]. To detect allergenic proteins an additional two-track approach is proposed. The first step is a literature search for data on IgE-epitopes of known allergenic proteins that have features in common with the protein under investigation, nowadays already a standard procedure within the EU GMO food safety evaluation. The second step should be an analysis on the basis of an antigenicity prediction programme that reveals the most probable site of antigenicity, based on hydrophilicity characteristics of the novel protein as well as of the allergen with similar sequence stretches. This software tool is, however, not yet validated. Using it, it was shown that also novel proteins expressed in conventional crops may have allergenic potential based on the amino acid sequence comparison of the proteins with those of known allergenic proteins [107]. As a consequence, it was advocated to include a general allergen identification check in the safety screening of all new plant varieties.

7.5.6 Gene transfer

The potential risks of the consumption of food or feed derived from GM crop with relation to horizontal gene transfer has been reviewed by van den Eede *et al*. [55]. Horizontal gene transfer was defined as the transfer of genetic material directly to a living cell or an organism followed by its expression, a common phenomenon among prokaryotes. The risk assessment with relation to possible gene transfer from GM varieties to the microflora in the GI tract should be based on the fact that, although chances are very small, this may occur [59]. This assumption should be used as a starting-point for the evaluation of the possible harmful consequences of such events.

A number of studies have by now been performed to assess the occurrence of gene transfer in practice. Nielsen *et al*. [108] describe a study on the uptake of dietary DNA in the blood and organs of Atlantic salmon. The uptake of DNA fragments was indeed observed with the highest concentrations in the liver and kidney. At the same time the authors report the difficulties relating to this type of experiments, especially the risk of contamination from the feed to the tissues under investigation. Another study on potential transfer of transgenic DNA from feed to animal tissues was published by Mazza *et al* [109]. Piglets were fed either GM or a conventional maize in their diets during 35 days, and blood, muscle and different organs were subsequently assessed for the present of maize and GMO-derived sequences. Maize sequences were found in all tissues except for muscles, fragmented transgenic sequences (app. 500 bp) were detected in blood and organs of the piglets. The authors conclude that uptake of foreign DNA molecules, and not just the nucleotides, may be a natural process and that the risk of gene transfer from GM food is not different from the risk of gene transfer in conventional feeds.

Netherwood *et al*. [110] have studied the fate of the the epsps transgene from GM soy in consumers. They report that the gene survived the passage through the small intestine, as analysed in human ileostomists, but not the passage throught the intact gastrointestinal tract. They propose to take the fact that transgenic gene sequences may survive the small intestine passage into account in safety assessment procedures for GMOs.

7.5.7 Post-marketing monitoring

It is generally acknowledged that in specific cases post-marketing monitoring may be useful to confirm assumptions, e.g. on intake level in specific consumer groups, that were used in the premarket safety assessment. Also, post-marketing monitoring may be useful to monitor specific nutritional effects in the population at large, or in specific consumer groups. But in general it will be very difficult for consumers to relate any adverse effects to a specific ingredient in a food product, or even to the food product as a whole [48].

To this end, informative traceability systems will need to be developed, or put in place. It can be envisaged that a broad range of certification processes for traceability will emerge as a result

of the different requirements with respect to identity preservation for value-added products [69]. For post-marketing monitoring only strict regimes would enable the linkage of potential perceived dietary effects with specific (GM) traits. This type of regime is, however, likely to be restricted to, for instance, crops that are altered to produce pharmaceutical substances, that are not supposed to enter the food production chain at any point. Other control systems will provide less options to link potential observed dietary effects to specific novel plant varieties and -derived products.

7.6. Comparative safety assessment for (all) novel plant-derived food products, discussion and conclusions

7.6.1 GM versus traditional breeding strategies

The safety assessment of plant products derived from new plant varieties is now divided into two mainstream approaches. For GMO-derived plant products stringent regulations are formulated that need to be applied, on a case-by-case basis, to all plant varieties within this category, irrespective of the nature of the actual changes in comparison with the direct parent lines. Alternatively, the other large category of new plant varieties that have not been obtained by gene technology, but may have been obtained by breeding strategies that are known to cause genetic alterations in the genome of the plant, such as classical hybridisation strategies and mutation breeding approaches, and the inclusion of (multiple) cell culture steps, are subject to the requirements as described under the Novel Food Regulation. This Regulation does not require an assessment of potential unintended side effects of the breeding process on a routine basis.

The area of plant breeding is not a static field, developments are continuous. New methodologies are being developed in what is sofar called 'conventional' breeding, as well as in the area of gene technology. Mutation breeding, both on the basis of chemical agents as well with the use of nuclear or transposon technologies, is an important school within plant breeding that may have gained in importance with the formulation of stringent food safety assessment procedures for GM plant varieties, but the relative contribution of mutation breeding to the total number of novel plant varieties is not documented. It is generally acknowledged that many mutations may occur in the mutational process in addition to the desired mutations, which will only to a limited extent be selected out in the subsequent steps of the breeding protocol. Also in this way new proteins and other types of expression products or metabolites can be introduced into new plant varieties that have not been part of the human diet before and would need to be assessed for their safety to the human consumer. This fact may warrant some additional precautionary safety assessment measures for this group of 'conventionally bred' crop varieties. On the other hand, the increasing knowledge on plant genomes and the function of plant genes has in recent years increased the possibilities for marker assisted breeding, i.e. the application of knowledge on gene function in traditional plant breeding strategies.

Breeding strategies can be speeded up by this approach, as after each step the presence of desired genes and/or the absence of undesired or deleterious genes can be monitored. In this way new plant varieties can be developed along traditional lines that may already be better characterised, also in aspects of food safety.

The application of gene technology has widened the pool of genes that can be used in plant breeding as species barriers can be crossed more easily compared to traditional breeding strategies. In many cases it is the only option to (easily) introduce specific genes from unrelated species into new plant varieties. The newly introduced genes and their expression products as well as any potential effects of the transformation process will need to be assessed for their food safety. Because of the two aspects of second generation GM products, i.e. the use of multiple genes or entire metabolic routes and often the nutritional nature of the genetic modification, it is likely that also the safety assessment of this type of products will become more complex: not only will the safety of multiple genes and potential interactions need to be assessed, but also the nutritional and toxicological assessment will become more challenging when metabolic pathways are modified that directly influence the nutritional characteristics of the resulting food products.

7.6.2 Cisgenesis

A development in this respect that may change the future of the application of gene technology in plant breeding is the appearance of cisgenic varieties that make use of genes that are already within the available natural genetic pool. Schouten *et al*. [17] state that cisgenic plants should be exempt from current GMO regulatory requirements. It is not clear, however, how unambiguous the definition of cisgenesis is in terms of food safety, as it may not exclude wild relatives that are not part of the human diet so far that can only be crossed under laboratory conditions. A number of food safety aspects that are key to novel transgenic varieties, such as the safety assessment of the expressed proteins, may indeed be much less relevant for these cisgenic plant varieties, as the genes involved were already within the available gene pool when using traditional breeding strategies. If the (distant) relative is also being used as a food source, the safety assessment of the newly introduced protein may obviously benefit from the knowledge that it is already part of the human diet. The food safety assessment should take this into account and be conducted accordingly. On the other hand, the wild relative may not form part of the human diet yet, and in that case it would be prudent to assess the safety of the newly introduced sequences and protein(s). But indeed this can only be justified if the same approach would be followed when a novel plant variety that would be the result of a traditional cross of the same parent lines, with any (remaining) linkage drag, would also be assessed within the regulatory frameworks of the food safety assessment of new plant varieties. Furthermore, other aspects such as unintended effects in the new plant variety as a result of insertional mutagenesis will not be different for cisgenic varieties compared to new transgenic varieties.

7.6.3 Safety assessment of new plant varieties

In practice, selected breeding strategies will often not be limited to one of the approaches described here. New plant varieties are likely to be the result of combined strategies; the plant breeder will decide the strategy of preference on a case-by-case basis, taking direct breeding-related aspects as well as the subsequent category-dependent regulatory requirements into account.

Because of the recent developments in breeding and processing technologies in food production, the scientific community has put large efforts in (further) developing tailored risk assessment strategies for plant-derived food products. At this moment it can be stated that international consensus has been reached on the basic approach. At the same time important developments are ongoing that may benefit risk assessment strategies in future times. They are now in different stages of development and validation. Worth mentioning here are the development of molecular screening tools to look at genetic alterations at the DNA level, the development of 'omics' technologies for improved detection and interpretation of compositional alterations as a result of the breeding process, the efforts to adapt current animal-based toxicology protocols for complex products, the further development of informative in-vitro toxicological tools, including the development of toxicogenomics and bioinformatics tools and the use of early markers for toxicological effects, improved food intake assessment strategies and improved (in silico) tools to assess potential allergenicity of newly introduced proteins. These are all important developments that may in time lead to increased efficacy in the food safety assessment of food (plant) products.

In the last two decades national and international regulatory bodies have decided that products derived from gene technology should be assessed separately. At the same time approved GM varieties can currently be crossed with any traditional variety, including wild relatives without a history of safe use, and the resulting cross will be evaluated as a conventional variety. In practice this means that in both Europe and the US in general no further safety evaluation of this novel plant variety will be required. On the other hand in Europe crosses of two GM varieties do need to go through the regulatory process: a dossier needs to be compiled with the relevant information on the new GMxGM variety, based on the knowledge of the assessment of the GM parent varieties. At the same time other crosses, perhaps including poorly characterised lines derived from not well-documented (mutagenic) breeding strategies, remain more or less unassessed because they are not considered under the current definition of gene technology, especially in those cases where there are no intended substantial physiological alterations.

7.6.4 Optimised framework for the safety assessment of all new plant varieties

The application of a first screening step, i.e. a hazard identification phase, in all cases of new plant-derived food products, including a compositional analysis, may serve to decide the most

appropriate subsequent steps in the food safety assessment on a case-by-case basis. Other elements, such as environmental and societal aspects, may form part of this first screening as well.

For GM crosses this may form the basis to decide whether it is necessary to supply the details of e.g. all (approved) parent GM varieties in detail. If there are no indications for unintended metabolic alterations in a novel GM cross variety compared to the direct parent lines, there may be no need for further safety assessment, as a logical consequence of the confidence in the premarket safety assessment of the GM parent lines. A similar view on the subject has been expressed by EFSA [111] and De Schrijver *et al*. [112] in their publication on the subject of the risk assessment of these GM stacked events. For conventionally bred lines, where different breeding technologies may have been applied, it should also form the basis to decide, case-by-case, whether a further more in-depth safety assessment may be required on specific aspects of the new plant product, or on the product as a whole.

It is important that the safety assessment of food products derived from new plant varieties closely follows the developments in plant breeding in order to assure that the safety assessment is optimally tailored to the novel food products that are moving towards the market. It may be that the current distinction between GMO-derived and so-called conventionally bred new plant varieties does not in all cases provide the best framework for an adequate safety assessment of new plant varieties as the basis for a safe food supply also in the years to come. It seems advisable to screen *all new plant varieties* for their new characteristics by applying the comparative safety assessment, which may have different end-points. These endpoints may range from an informative fact sheet with compositional data in the case of crosses of known crop plant varieties that confirms the absence of new hazards up to the full toxicological and nutritional characterisation for novel plant varieties where clear hazards have been identified or where no conventional counterpart is available for the comparative approach. The requirement of compositional data in all cases is less revolutionary than it may look at first instance: it is common practice that producers of new food products characterise their products well before the marketing phase.

At the end of the day the aim of food safety assessment strategies, as part of general food product approval procedures, should be, as Garza and Stover [113] formulate it, 'to improve the human condition and to minimise inequalities'. This can only be achieved if we try to formulate optimal and objective criteria for the food safety assessment for all food products, including plant-derived food products, and develop a screening-frame to select against new plant varieties that we can not, on the basis of well-thought criteria, consider to be as safe as plant varieties that are already on the market or that on similarly formulated criteria should be considered as otherwise unacceptable.

Acknowledgements

The authors would like to acknowledge the Dutch Ministry of Agriculture, Nature and Food Quality for their financial support of this study as part of the thematic programme on 'biological agents', as well as the European Commission for their support of the study as part of the European integrated research project Safefoods.

References

1. Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J., and Last, R.L., 2003. Ethylmethanesulphonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant Physiol. 131, 139-146.
2. Gichner, T., 2003. Differential gentoxicity of ethyl methanesulphonate, N-ethyl-N-nitrosourea and maleic hydrazide in tobacco seedlings based on data of the Comet assay and two recombination assays. Mutation Res. 538(2003)171-179.
3. Ahloowalia, B.S., Maluszynski M., 2001 Induced mutations – a new paradigm in plant breeding. Euphytica 118, 167-173.
4. Fitzmaurice, W.P., Nguyen, L.V., Wernsman, E.A., Thompson, W.F., Conkling, M.A., 1999. Transposon tagging of the sulphur gene of tobacco using engineered maize Ac/Ds elements. Genetics 153, 1919-1928.
5. McCallum, C.M., Comai, L., Greene, E.A., Henikoff, S., 2000. Targeted screening for induced mutations. Nature Biotechnology, 18, 455-457.
6. IAEA (International Atomic Energy Agency), 2004. Liu, L., Van Zanten, L., Shu, Q.Y., and Maluszynski, M. Officially released mutant varieties in China. International Atomic Energy Agency, Mutation Breeding Review 14, ISSN 1011-2618.
7. Agrawal, R.L., 1998. Fundamentals of plant breeding and hybrid seed production. Science publishers, Enfield, New Hampshire, USA.
8. Jain, H.K., and Kharkwal, M.C., 2004. Plant breeding: Mendelian to molecular approaches. Agritech publications, Shrub Oak, New York, USA.
9. Wilson, A.W., Latham, J., Steinbrecher, R., 2004. Genome scrambling – myth or reality. EcoNexus Technical Report – October 2004.
10. Maloof, J.N., 2003. QTL for plant growth and morphology. Curr. Opin.Plant Biol. 6, 85-90.
11. Hoa, T.T.C., Al-Babili, S., Schaub, P., Potrykus, I., and Beyer, P., 2003. Golden Indica and Japonica rice lines amenable to deregulation. Plant Physiol. 133, 161-169.
12. Quesada-Vargas, T., Ruiz, O.N., and Daniell, H., 2005. Characterization of heterologous multigene operons in transgenic chloroplasts. Transcription, processing and translation. Plant Physiol 138, 1746-1762.
13. Halpin, C., 2005. Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology. Plant Biotechn. J. 3, 141-155.
14. Terada, R., Johzuka-Hisatomi, Y., Saitoh, M., Asao, H., and Iida, S., 2007. Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. Plant Physiol 144, 846-856.
15. Swoboda, P., Gal, S., Hohn, B., Puchta, H., 1994. Intrachromosomal homologous recombination in whole plants. EMBO J 13, 2, 484-489.

16. Filkowski, J., Yeoman, A., Kovalchuk, O., and Kovalchuk I., 2004. Systemic plant signal triggers genome instability. Plant J. 38, 1,1-11.
17. Schouten, H.J., Krens, F.A., and Jacobsen, E., 2006. Cisgenic plants are similar to traditionally bred plants. EMBO reports 7, 8, 750-753.
18. Thomson, J., 2003. Genetically modified food crops for improving agricultural practice and their effects on human health. TIFS 14, 210-228.
19. International Council for Science (ICSU), 2003. Genetically modified foods for human health and nutrition: the scientific basis for benefit/risk assessment. TIFS 14, 173-181.
20. ILSI (International Life Sciences Institute) Europe, 2001. Genetic modification technology and food. Consumer health and safety. ILSI Europe concise monograph series, Clare Robinson, ed. ISBN 1-57881-125-2.
21. Chassy, B., Hlywka, J.J., Kleter, G.A., Kok, E.J., Kuiper, H.A., McGloughlin M., Munro, I.C., Phipps, R.H., Reid, J.E., 2004. Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology – Prepared by a task force of the ILSI International Food Biotechnology Committee. CRFSFS 3, 2, 35-104, http://www.ift.org/pdfs/crfsfs/crfsfsv3n2p0035-0104ms20040106.pdf
22. Schmidt, C.W., 2005. Genetically modified foods: breeding uncertainty. Environmental Health Perspectives 113 (8).
23. Hoffmann-Sommergruber, K., SAFE consortium., 2005. The SAFE project 'plant food allergies: field to table strategies for reducing their incidence in Europe', an EC-funded study. Allergy 60, 4, 436-42.
24. Gao, Z.S., Van de Weg, W.E., Schaart, J.G., Van der Meer, I.M., Kodde, L., Laimer, M., Breiteneder, H., Hoffmann-Sommergruber, K., Gilissen, L.J.W.J., 2005. Linkage map positions and allelic diversity of two Mald3 (non-specific lipid transfer protein) genes in the cultivated apple (Malus domestica). Theor. App. Genet. 110, 3, 479-91.
25. McKeon,T.A., 2003. Genetically modified crops for industrial products and processes and their effects on human health. TIFS 14, 229-241.
26. ILSI (Internation Life Sciences Institute), 2002. Biotechnology-derived nutritious foods for developing countries : needs, opportunities, and barriers. Bouis, H.E., Lineback, D., and Schneeman, B., eds. Food and Nutrition Bulletin 23, 4.
27. Bouis, H.E., Chassy, B.M., and Ochanda, J.O., 2003. Genetically modified food crops and their contribution to human nutrition and food quality. TIFS 14, 191-209.
28. Khush, G.S., 2002. The promise of biotechnology in addressing current nutritional problems in developing countries. Food and Nutrition Bulletin 23, 4, 354-357.
29. Paine, J.A., Shipton, C.A., Chaggar, S., Howells, R.M., Kennedy M.J., 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nature biotechnol. 23, 4, 482.
30. Napier, J.A., Sayanova. O., 2005. The production of very-long-chain PUFA biosynthesis in transgenic plants: towards a sustainable source of fishy oils. Proc Nutr Soc 64, 387-393.
31. Wildung, M.R., and Croteau, R.B., 2005. Genetic engineering of peppermint for improved essential oil composition and yield. Transgenic Research 14, 365-372.
32. Cahoon, E.B., Hall, S.E., Ripp, K.G., Ganzke, T.S.,. Hitz, W.D., and Coughlan, S.J., 2003. Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. Nature Biotechnol. 21, 9, 1082-1087.
33. Dörmann, P., 2003. Corn with enhanced antioxidant potential. Nature Biotechnol. 21, 9, 1015-1016.

34. Fauquet, C.M., and Taylor, N.,2002. The potential for biotechnology to improve the nutritional value of cassava. Food and Nutrition Bulletin 23, 4, 364-366.
35. Renwick, A.G., 2004. Risk characterization of chemicals in food. Toxicology Letters 149, 163-176.
36. Gaylor, D.W., and Gold, L.S., 1995. Quick estimate of the regulatory virtually safe dose based on the maximum tolerated dose for rodent bioassays. Regul. Toxicol. Pharmacol. 22, 1, 57-63.
37. Smith, M., 2002. Food Safety in Europe (FOSIE): risk assessment of chemicals in food and diet: overall introduction. Food Chem Toxicol 40, 2/3,141-144.
38. Barlow, S.M., Greig, J.B., Bridges, J.W., Carere, A., Carpy, A.J.M., Galli, C.L., Kleiner, J., Knudsen, I., Koëter, H.B.W.M., Levy, L.S., Madsen, C., Mayer, S., Narbonne, J.F., Pfannkuch, F., Prodanchuk, M.G., Smith, M.R., and Steinberg, P., 2002. Hazard identification by methods of animal-based toxicology. Food Chem. Toxicol. 40, 2/3,145-191.
39. Eisenbrand, G., Pool-Zobel, B., Baker, V., Balls, M., Blaauboer, B.J., Boobis, A., Carere, A., Kevekordes, S., Lhuguenot, J-C., Pieters, R., and Kleiner, J., 2002. Methods of in-vitro toxicology. Food Chem. Toxicol. 40, 2/3, 193-236.
40. Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A.G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., and Younes, M., 2002. Hazard characterisation of chemicals in food and diet: dose response, mechanisms and extrapolation issues. Food and Chemical. Toxicol. 40, 2/3, 237-282.
41. Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., and Würtzen, G., 2002. Mathematical modelling and quantitative methods. Food Chem. Toxicol. 40, 2/3, 283-326.
42. Kroes, R., Müller, D., Lambe, J., Löwik, M.R.H., Van Klaveren, J., Kleiner, J., Massey, R., Mayer, S., Urieta, I., Verger, P., and Visconti, A., 2002. Assessment of intake from the diet. Food Chem. Toxicol. 40, 2/3, 327-385.
43. Van Gelder, W.M.J., Vinke, J.H., Scheffer, J.J.C., 1988. Steroidal glycoalkaloids in tubers and leaves of Solanum species used in potato breeding. Euphytica Suppl:147-158
44. EC (European Commission), 2005. Commission Regulation (EC) 1822/2005 of 8 November 2005 amending Regulation (EC) No. 466/2001 as regards nitrate in certain vegetables. Official Journal of the European Union dd. 9.11.2005, L293/11-13.
45. Seligman, P.J., Mathias, C.G., O'Malley, M.A., Beier, R.C., Fehrs, L.J., Serrill, W.S., Halperin, W.E., 1987. Phytophotodermatitis from celery among grocery store workers. Arch Dermatol 123, 11, 1478-82.
46. Hammond, B., Rogers, S.G., Fuchs, R.L., 1996. Limitations of whole food feeding studies in food safety assessment. In: OECD (Ed.), Food Safety Evaluation. Organisation for Economic Cooperation and Development, Paris, 85-97.
47. Diehl, J.F., 2002. Food irradiation, past, present and future. Radiation Phys. Chem. 63, 211-215.
48. Kuiper, H.A., Kleter, G.A., Noteborn, H.P.J.M., and Kok, E.J., 2001. Assessment of the food safety issues related to genetically modified foods. Plant J. 27, 6, 503-528.
49. König, A., Cockburn, A., Crevel, R.W.R., Debruyne, E., Grafstroem, R., Hammerling, U., Kimber, I., Knudsen, I., Kuiper, H.A., Peijnenburg, A.A.C.M., Penninks, A.H., Poulsen, M., Schauzu, M., Wal, J.M., 2004. Assessment of the safety of foods derived from genetically modified (GM) crops. Food Chem. Toxicol. 42, 1047-1088.
50. Schilter, B., Andersson, C., Anton, R., Constable, A., Kleiner ,J., O'Brien, J., Renwick, A.G., Korver, O., Smit, F., and Walker, R., 2003. Guidance for the safety assessment of botanicals and botanical preparations for use in food and food supplements. Food Chem Toxicol 41, 1625-1649.

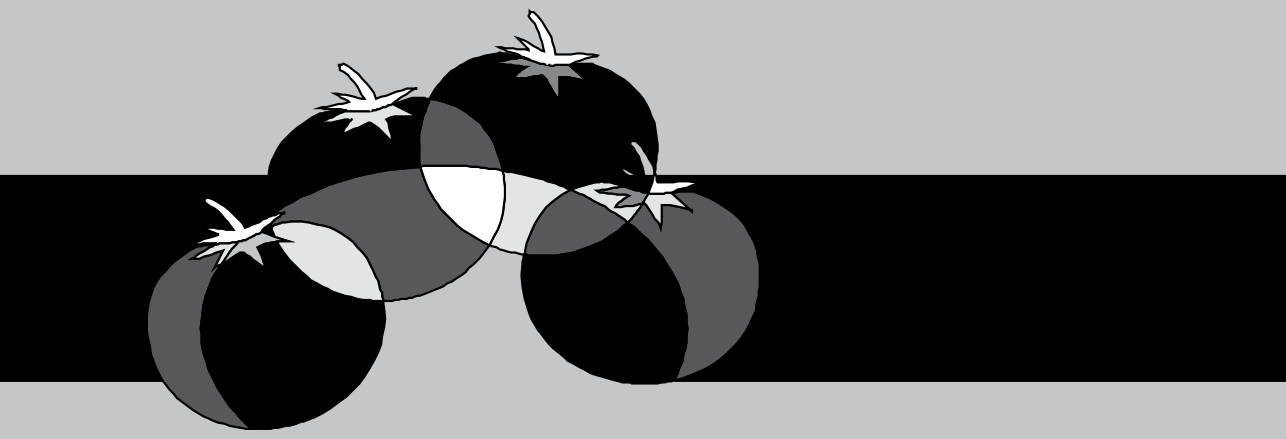
51. Kroes, R., and Walker, R., 2004. Safety issues of botanicals and botanical preparations in functional foods. Toxicology 198, 213-220.
52. Kok, E.J., and Kuiper, H.A., 2003. Comparative Safety Assessment for Biotech Crops. TIB 21, 10, 439-444.
53. Kuiper, H.A., 2004. Introduction to Entransfood. Food Chem. Toxicol. 42, 1044-1045.
54. Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H.V., Engel, K.H., Gatehouse, A.M.R., Kärenlampi, S., Kok. E,J., Leguay, J.-J., Lehesranta, S., Noteborn, H.P.J.M., Pedersen, J., and Smith, M., 2004. Unintended effects and their detection in genetically modified crops. Food Chem Toxicol 42, 1089-1125.
55. Van den Eede, G., Aarts, H., Buhk, H.-J., Corthier, G., Flint, H.J., Hammes, W., Jacobsen, B., Midtvedt, T., Van der Vossen, J., Von Wright, A., Wackernagel, W., and Wilcks, A., 2004. The relevance of gene transfer to the safety of food and feed derived from genetically modified (GM) plants. Food Chem. Toxicol. 42, 1127-1156.
56. OECD (Organisation for Economic Cooperation and Development), 2006. An introduction to the food/feed safety consensus documents of the Task Force. Series on the Safety of Novel Foods and Feeds, no 14, ENV/JM/MONO(2006)10. http://appli1.oecd.org/olis/2006doc.nsf/43bb6130e5e86e5fc12569fa005d004c/f6714efc9beea840c1257156005638c2/$FILE/JT03207766.PDF
57. FAO/WHO (Food and Agriculture Organization/World Health Organization), 1996. Biotechnology and food safety. Report of a Joint FAO/WHO consultation, Rome, Italy. FAO Food and Nutrition Paper 61, Food and Agriculture Organisation of the United Nations, Rome, ftp://ftp.fao.org/es/esn/food/biotechnology.pdf
58. OECD, 1996. Food Safety Evaluation. Organisation for Economic Cooperation and Development, Paris.
59. EFSA (European Food Safety Authority), 2006. Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. The EFSA Journal 99, 1-100, http://www.efsa.europa.eu/en/science/gmo/gmo_guidance/660.html
60. Kleter, G.A., and Kuiper, H.A., 2006. Regulation and risk assessment of biotech food crops. Plant Genetic Engineering vol 8: Metabolic Engineering and Molecular Farming-II. Jaiwal, P.K., and Singh, R.P. (eds). Studium Press, LLC, Huston, Texas 77272, USA, pp 311-337.
61. EFSA (European Food Safety Authority), 2007b. Safety and nutritional assessment of GM Plant derived foods/feed. The role of animal feeding trials, http://www.efsa.europa.eu/etc/medialib/efsa/science/gmo/gmo_consultations/gmo_animalfeedingtrials.Par.0002.File.dat/gmo_AnimalFeedingTrials_consultation.pdf
62. World Trade Organisation (2006) European Communities – Measures affecting the approval and marketing of biotech products. Reports of the Panel. WTO reports WT/DS291/R, WT/DS292/R, WT/DS293/R, 29 September 2006, World Trade Organisation, Geneva, Switzerland.
63. EC (European Commission), 2003. Regulation (EG) Nr. 1829/2003 of the European Parliament and the Council of 22 September 2003 with regard to genetically modified foods and feed. Official Journal of the European Union dd. 18.10.2003, L268/1-23.
64. FDA (US Food and Drug Administration), 1992. Statement of policy: foods derived from new plant varieties. Federal Register, May 29.
65. Freese, W., and Schubert, D., 2004. Safety testing and regulation of genetically engineered foods. Biotechnol. Genet. Engineering Rev. 21, 299-324.
66. Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R., 2005. Applications of DNA tiling arrays for whole-genome analysis. Genomics 85, 1-15.

67. Kohli, A., Twyman, R.M., Abranches, R., Wegel, E., Stoger, E., and Christou, P., 2003. Transgene integration, organization and interaction in plants. Plant Mol. Biol. 52, 247-258.
68. Filipecki, M., and Malepsy, S., 2006. Unintended consequences of plant transformation: a molecular insight. J Appl Genet 47, 4, 277-286.
69. Wilson, W.W., Janzen, E.L., and Dahl, B.L., 2003. Issues in development and adoption of genetically modified (GM) wheats. AgBioForum 6, 3, 1-12.
70. Van Harten, A.M., 1998. Mutation Breeding: Theory and Practical Applications. Cambridge University Press, London.
71. Rang, A., Linke, B., and Jansen, B., 2005. Detection of RNA variants transcribed from the transgene in Roundup Ready soybean. Eur Food Res Technol 220, 438-443.
72. Ridley, W.P., Shillito, R.D., Coats, I., Steiner, H-Y., Shawgo, M., Phillips. A,, Dussold. P,, Kurtyka. L., 2004. Development of the International Life Sciences Institute Crop Composition Database. J Food Comp Anal 17, 423-438.
73. Nordic Council, 2003. Use of the cDNA microarray technology in the safety assessment of GM food plants. TemaNord report 2003:558, Nordic Council of Ministers, Copenhagen, Denmark.
74. Kok, E.J., Franssen-van Hal, N.L.W., Winnubst, E.N.W., Kramer, E.H.M., Dijksma, W.T.P., Kuiper, H.A., and Keijer, J., 2007. 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, 164, 337-349.
75. Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., 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 Physiol. 138,3, 1690-1699.
76. Lehesranta, S.J., Koistinen, K.M., Massat, N., Davies, H.V., Shepherd, L.V.T., McNicol, J.W., Cakmak, I., Cooper, J., Lück, L., Kärenlampi, S.O., and Leifert, C., 2007. Effects of agricultural production systems and their components on protein profiles of potato tubers. Proteomics 7, 597-604.
77. Colquhoun, I.J., Le Gall, G., Elliott, K.A., Mellon, F.A., and Michael, A.J., 2006. Shall I compare thee to a GM potato? TIG 22, 10, 525-528.
78. Chen, S., and Harmon, A.C., 2006. Advances in plant proteomics. Proteomics 6, 5504-5516.
79. Dixon, R.A., Gang, D.R., Charlton, A.J., Fiehn, O., Kuiper, H.A., Reynolds, T.L., Tjeerdema, R.S., Jeffery, E.H., German, J.B., Ridley, W.P., and Seiber, J.N., 2006. Applications of metabolomics in agriculture. J Agric. Food Chem. 54, 8984-8994.
80. Hollywood, K., Brison, D.R., and Goodacre, R., 2006. Metabolomics: current technologies and future trends. Proteomics 6, 4716-4723.
81. Baudo, M.M., Lyons, R., Powers, S., Pastori, G.M., Edwards, K.J., Holdsworht, M.J., and Shewry, P.R., 2006 Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. Plant Biotechnol. J. 4, 369-380.
82. Baker, J.M., Hawkins, N.D., Ward, J.L., Lovegrove A., Napier, J.A., Shewry, P.R., and Beale, M.H., 2006. A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotech J. 4, 381-392.
83. Shepherd LVT., McNichol JW., Razzo R., Taylor MA., Davies HV. Assessing the potential for unintended effects in genetically modified potatoes perturbed in metabolic and developmental processes. Targeted analyses of key nutrients and anti-nutrients. (2006) Transgenic Research 15, 409-425.

84. LeGall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J., Verhoeven, M.E., 2004. Metabolite profiling of tomato (*Lycopersicon esculentum*) using 1H-NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. J. Agricult. Food Chem. 51, 9, 2447-2456.
85. 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. and Draper. J., 2005. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. PNAS 102, 40, 14458-14462.
86. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihaq, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamavo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D., Groop, L.C., 2003. Nature Gen. 34, 3, 267-273.
87. Kim, S-Y., and Volsky, D.J., 2005. PAGE: Parametric analysis of gene set enrichment. BMC Bioinformatics 6, 144.
88. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. PNAS, 102, 43, 15545-15550.
89. Pan, D., Sun, N., Cheung, K.-H., Guan, Z., Ma, L., Holford, M., Deng, X., and Zhao, H., 2003. PathMAPA: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for Arabidopsis. BMC Bioinformatics 4, 56.
90. Grosu, P., Townsend, J.P., Hartl, D.L., and Cavalieri D., 2002. Pathway Processor: A tool for integrating whole-genome expression results into metabolic networks. Genome Res. 12, 1121-1126.
91. Breitling, R., Amtmann, A., and Herzyk P., 2004. Graph-based iterative group analysis enhances microarray interpretation. BMC Bioinformatics 5, 100.
92. Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., Umezawa, T., Misawa, N., Saito, K., and Shibata, D., 2005. KaPPa-view. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol 138, 1289-1300.
93. Pryme, I., and Lembcke, R., 2003. In vivo studies on possible health consequences of genetically modified food and feed with particular regard to ingredients consisting of genetically modified plant materials. Nutrition and Health (Bicester) 17, 1, 1-8.
94. Hugget, A.C., Marchesini, M., Perrin, I., Schilter, B., Tschantz, J.C., Donnet, A., Morgenthaler, P., Sunhaler, G., Wurzner, H.P., 1996. The application of human-type diets in rodent feeding studies for the safety evaluation of novel foods. In: OECD (Ed.), Food Safety Evaluation. Organisation for Economic Cooperation and Development, Paris, 135-150.
95. OECD, (Organisation for Economic Cooperation and Development), 2005. Report of the OECD/IPCS Workshop on Toxicogenomics, Kyoto, 13-15 October 2004. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 50, OECD, Paris.
96. Moggs, J.G., 2005. Molecular responses to xenoestrogens: mechanistic insights from toxicogenomics. Toxicology 213(2005)177-193.
97. Hollman, P.C.H., and Arts, I.C.W., 2005. Polyphenols and disease risk in epidemiological studies. Am J Clin Nutr 81 (suppl), 317-325.
98. Lambe, J., 2002. The use of food consumption data in assessments of exposure to food chemicals including the application of probabilistic modelling. Proc Nutr Soc 61, 11-18.

99. Hesketh, J., Wybranska, I., Dommels, Y., King, M., Elliott, R., Pico, C., Keijer, J., 2006. Nutrient-gene interactions in benefit-risk analysis (workshop report). Br. J. Nutr. 95, 1232-1236.
100. Gilsenan, M.B., Lambe, J., and Gibney, M.J., 2003. Assessment of food intake input distributions for use in probabilistic exposure assessments of food additives. Food Additives and Contaminants 20, 11, 1023-1033.
101. Leclercq, C., Arcella, D., Armentiaz, A., Boon, P.E., Kruizinga, A.G., Gilsenan, M.B., and Thompson, R.L., 2003. Development of databases for use in validation studies of probabilistic models of dietary exposure to food chemicals and nutrients. Food Additives and Contaminants, 20, 1, 27-35.
102. FAO/WHO (Food and Agriculture Organization/World Health Organization), 2001. Evaluation of allergenicity of genetically modified foods. Report of a joint FAO/WHO expert consultation of allergenicity of foods derived from biotechnology. Food and Agriculture Organization, Rome, 22-25 January.
103. Nordlee, J.A., Taylor, S.L., Townsend, J.A., Thomas, L.A., and Bush, R.K., 1996. Identification of a brazil-nut allergen in transgenic soybeans. New England Journal of Medicine 334, 688-92.
104. Lehrer, S.B., and Bannon, G.A., 2005. Risks of allergic reactions to biotech proteins in foods: perception and reality. Allergy 60, 559-564.
105. Orruño, E., and Morgan, M.R.A., 2006. IgE binding to proteins from sesame and assessment of allergenicity: implications for biotechnology? Biotechnology Letters 28, 1877-88.
106. Kleter, G.A., and Peijnenburg, A.A.C.M., 2004. Prediction of the potential allergenicity of novel proteins. Food Sci. Technol. 18, 3, 28-31.
107. Kleter, G.A., Peijnenburg, A.A.C.M., 2003. Presence of potential allergy-related linear epitopes in novel proteins from conventional crops and the implication for the safety assessment of these crops with respect to the current testing of genetically modified crops. Plant Biotechnol. J. 1, 371-380.
108. Nielsen, C.R., Berdal, K.G., Bakke-McKellep, A.M., Holst-Jensen, A., 2005. Dietary DNA in blood and organs of Atlantic salmon (Salmo salar L.). Eur. Food Res.Technol., 221, 1-8.
109. Mazza, R., Soave, M., Morlacchini, M., Piva, G., and Marocco, A., 2005. Assessing the transfer of genetically modified DNA from feed to animal tissues. Transgenic Res. 14, 775-784.
110. Netherwood, T., Martin-Orue, S.M., O'Donnell, A.G., Gockling, S., Graham, J., Mathers, J.C., and Gilbert, H.J., 2004. Assessing the survival of transgenic plant DNA in the human gastrointestinal tract. Nat Biotech 22,2, 204-209.
111. EFSA (European Food Safety Authority), 2007a. Guidance Document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants containing stacked transformation events. The EFSA Journal (2007) 512, 1-5, http://www.efsa.europa.eu/etc/medialib/efsa/science/gmo/gmo_guidance/gmo_guidance_ej512.Par.0001.File.tmp/gmo_guidance_ej512_GM_stacked_events_en.pdf
112. De Schrijver, A., Devos, Y., Van den Bulcke, M., Cadot, P., De Loose, M., Reheul, D., and Sneyers, M., 2007. Risk assessment of GM stacked events obtained from crosses between GM events. TIFS 18, 101-109.
113. Garza, C., and Stover, P., 2003. General introduction – the role of science in identifying common ground in the debate on genetic modification of foods. TIFS 14, 5, 182-190.

Chapter 8.



Summary and conclusions

The major omics technologies include transcriptomics, i.e. the generation of profiles of mRNA or gene transcripts, proteomics, i.e. the generation of protein profiles, and metabolomics, i.e. the generation of profiles of secondary metabolites. Metabolomics uses a range of different technologies, most often ^{1}H-NMR spectroscopy or the more sensitive, so-called hyphenated approaches such as liquid or gas chromatography coupled to mass spectrometry (LC-MS or GC-MS, respectively), but also high performance liquid chromatography (HPLC) or electrochemical arrays (ECs) [1,2,3]. The advantage of (most of) these approaches is that the techniques are high throughput and relatively cheap on a per sample basis, but metabolomics faces the difficulty that the metabolite profiles will include metabolites across a concentration range of 10^9 and a log polarity range of approximately -6 to 14 [1]. The number of metabolites included in a profile will generally not exceed thousand, which may even be less than 1% of the metabolites present in a single organism [4], but this percentage may be higher in individual cell or organ systems. Proteomics is usually based on two-dimensional (2D) electrophoresis coupled to mass spectrometry or on multidimensional protein identification techniques [5,6]. Identified proteins will usually be high-copy and not include hydrophobic proteins, such as membrane proteins, and the dynamic range may be up to 10^6 [7]. Developments are ongoing to increase the scope and power of proteomics analyses. At this moment up to a few hundred proteins are generally analysed in a single experiment, which is likely to be a small part (1% or less) of the plant proteome, considering the fact that the number of genes in a plant genome (estimated at 25.000-40.000) can all result in multiple proteins as a result of alternative gene transcription and RNA splicing. Again, for individual cells or tissues the percentage may be higher, possibly up to 10%, but actual numbers of proteins are difficult to assess Transcriptomics, finally, uses either (PCR-amplified) cDNA or oligonucleotide microarrays. Microarrays in general form a robust platform for gene expression analysis and provide reproducible data, also because all mRNAs have similar biochemical characteristics. To date, for plants so-called whole transcriptome arrays are only available for the model species Arabidopsis and rice, for other species arrays with up to 30.000 expressed sequences are available, which may represent up to >50% of the total transcriptome [8,9]. The most relevant setback of transcriptome analysis is that observed differences may not be equally reflected in differences in the proteome and/or metabolome and as such in the plant's physiology. It will therefore always be necessary to confirm observed physiological alterations by targeted analyses of the potentially affected metabolic networks.

The present thesis evaluates current developments in the safety assessment of complex food products derived from genetically modified organisms (GMOs) as well as produced by conventional breeding methods. It especially evaluates the potential use of 'omics' technologies, and transcriptomics in particular, as part of these food safety assessment strategies.

In Chapter 1 the concept of the Comparative Safety Assessment (CSA) is introduced as the guiding principle in the safety assessment of complex agricultural food products, replacing the similar but controversial Principle of Substantial Equivalence that was frequently interpreted as being an endpoint of the assessment. The CSA focuses on the analysis of the newly bred

variety and a near(est) comparator as the initial part of the assessment procedure. Detected differences should be evaluated in the light of intended effects of the breeding strategy, taking natural variations in the varieties analysed, as well as in other (related) commercial varieties, into account. The results of this comparison should subsequently guide the consecutive toxicological and nutritional evalulation. It is argued that this tiered approach is not only applicable to GMO-derived plant products, but has wider applicability as a universal safety assessment strategy for new products produced by a broad range of breeding strategies currently in use in the different areas of agriculture.

In Chapter 2 food safety assessment strategies are discussed and evaluated for food products derived from genetically modified (GM) animals, in the light of the preceding and coinciding developments in the plant breeding sector. One of the important differences between GM animal versus GM plant products is the number of organisms involved: where hundreds GM plants can be obtained in a single transformation experiment, the number of GM animals resulting from a single transformation series will usually be very small, depending on the species, often below ten. This may in theory result in less stringent selection procedures and will result in less animals being available for the safety assessment procedures. As this will be the case for both the GM animal as for the comparators, it may also prove to be more difficult to assess detected differences for their toxicological and/or nutritional relevance. Another relevant difference is the fact that antinutritional substances are very rare in animal products compared to the large numbers of identified natural toxins in plants. Assessment of the GM animal-derived food products for altered levels of antinutrients is therefore an aspect that can rarely be applied, leaving the primary focus of the CSA to altered nutrient profiles. A last difference relates to the traceability systems that are more generally available in the animal production sector and not yet in the plant sector, making different post-marketing surveillance systems much more feasible compared to the plant situation. Post-marketing surveillance studies may be advocated in the case of uncertainties relating to the nutritional or exposure assessment of the product or, in exceptional cases, to the potential allergenicity of the newly introduced protein(s). It is concluded that the CSA principle is also feasible and applicable in the case of GM animal-derived food products, but some adaptations with relation to the aforementioned aspects may be required in the practical elaboration of the principle.

One of the recommendations of international expert meetings on the issue of the safety assessment of GMO-derived food products is that the new analytical techniques, such as the 'omics' approaches, need to be explored for their potential to identify unintended effects as part of the CSA and thereby of the toxicological and nutritional assessment of complex (GMO-derived) products. Chapter 3 describes one of the first experimental approaches using transcriptomics in this area: a small tomato-array was developed with informative cDNAs on the basis of two subtractive cDNA libraries. One library was obtained by subtracting the green tomato mRNA population from the red mRNA population, thereby aiming to enrich the resulting cDNA population for genes that are related to the nutritional and health-promoting characteristics of the red tomato. The other library was obtained by the reverse subtraction,

resulting in an enriched green tomato cDNA library that may be more related to metabolic routes that are involved in the formation of anti-nutrients and natural toxins, that are primarily known to occur in the green, unripe tomato. The Representational Difference Analysis (RDA) -approach to obtain informative plant cDNA libraries and construct derived arrays is evaluated: although normalization of the mRNA population, i.e. single representation of each unique mRNA sequence, is not obtained, the approach seems valuable as an initial step to reach this goal.

Chapter 4 describes how the tomato array developed in chapter 3 was used to analyse mRNA derived from tomatoes in five subsequent ripening stages from green, via breaker, turning, and light red, to red, to obtain a background library of gene expression profiles of different ripening stages for future comparisons as part of a CSA. At the same time these initial series of experiments aimed at determining whether it is possible to develop simple developmental or 'ripening stage' criteria for the sampling of fruits for 'omics' analyses in a CSA. The experiments show that the setup of the transcriptomics experiments is of crucial importance: small differences in ripening phase of the selected tomatoes for a CSA may cause substantially altered gene expression profiles that are unrelated to the genetic alterations. It is shown that samples under scrutiny should be well-matched with relation to the stage of development of the tissues. This also pertains to the application of proteomics, as clear differences were also observed in protein profiles during tomato ripening. It is furthermore advocated that application of the 'omics' technologies should preferably already take place in early stages of the breeding procedures, to include the factor 'food safety' into the breeding process. It is shown that the 'omics' technologies have the potential to pick up differences that may not have been picked up by current targeted analysis as part of a CSA.

In chapter 5 the same tomato array has been used to analyse two GM transformant lines and the parent WT variety. Because the same genetic construct has been introduced in both transformant lines, but presumably in different locations in the random transformation process, it is possible to assess the differences in terms of effects that are directly related to the expression products of the introduced genetic construct, and other, mostly unintended effects. The results of the comparative transcriptomics experiments are analysed in view of what is known from the targeted compositional analyses of the transformant lines and in view of our knowledge on ripening-related gene expression profiles. It is shown that the differences between the two GM lines and the WT variety are limited and to a large extent similar in both lines, making it very plausible that the observed effects are indeed construct-related and can not as such be considered to be basically unintended effects.

In chapter 6 similar experiments are described to assess the extent of unintended effects in Arabidopsis transformant lines. A number of different transformant lines were included in this study with different copy numbers and orientations of the introduced genetic construct, but all of the lines had a single place of insertion. Differential gene expression was observed primarily in genes that are known to be stress-related. But here also the limited knowledge on the extent of the natural variation in gene expression of the genes under consideration

complicates the toxicological and nutritional evaluation of the differences observed, especially when slight differences in the environmental conditions of growth can not be excluded. These experiments show that it is crucial to have substantial information on the natural variation in gene expression in the crops and tissues under scrutiny in order to be able to interpret 'omics' data correctly within the framework of a CSA.

Chapter 7 provides an overview of the international debate on the safety assessment of complex (GMO-derived) plant products, based on current breeding strategies, and the results of the first publications on experiments that aim to detect unintended effects due to application of different plant breeding strategies using 'omics' technologies, including the experiments described in this thesis. The results are combined to review current approaches and propose a new overall approach for the safety evaluation of complex plant products, including GMO-derived products. It is advised to screen all new plant varieties, not only GM varieties, for their new characteristics by applying the subsequent steps of the CSA principle on a case-by-case basis. As a result, the end-points of the CSA may range from an informative fact sheet with compositional data in the case of crosses of known crop plant varieties that confirms the absence of new hazards, up to the full toxicological and nutritional characterisation for novel plant varieties where clear hazards have been identified or where no conventional counterpart is available for the comparative approach. 'Omics' technologies may form part of the leading initial analytical comparative CSA-step to identify differences with known comparable plant varieties on the market, if available.

A limited number of other studies have already used omics technologies to assess differences between GM and conventionally bred (parent) crop varieties. Transcriptomics has been applied by Baudo *et al*. [10] in different GM and conventional wheat varieties. They primarily investigated the band width of natural variation and found the natural variation in gene expression patterns in conventionally bred wheat varieties to be much larger than the variation between different GM lines, which is in agreement with the assumption that with the application of gentechnological techniques only a limited genetic alteration is realized, compared to traditional breeding strategies where recombination events are numerous, resulting in altered phenotypes, including altered gene expression profiles.

Metabolomics studies have been applied in comparative studies between GM and non-GM varieties. Examples are the NMR and GC-MS studies of field-grown wheat varieties [11]. GM and control lines were analysed and it was found that differences can be observed, but not necessarily in all test locations that were used in the experiment. This may indicate that under different (environmental) conditions differences that are related to genetic alterations may or may not show up. On the other hand it may also indicate that the differences are indeed small and will not in all cases be statistically significant. The latter may be further underlined by the fact that all differences observed between GM and control lines were within the band width of natural variation of the control lines in the different locations. Similar studies were performed by others: Le Gall *et al*. [12] studied GM tomatoes and control lines on the basis of NMR spectra.

They observed differences in a number of metabolites that could not be linked to the genetic modification, but all observed differences were limited and within the band width of natural variation. It was their conclusion that NMR can trace even small differences in metabolite profiles that may be related to unintended effects in genetically modified crops. Catchpole *et al*. [13] describe a study on field grown potatoes, GM lines and conventional controls, that were analysed on the basis of a two-step MS approach. They found a large variation in metabolite profiles between the tested conventional cultivars. Observed differences between the GM line and the parent variety fell within this band width of natural variation. Colquhoun *et al*. [14] further analysed the work by Catchpole *et al*. [13] and concluded that subtle metabolic changes can indeed be revealed by metabolomics.

Proteomics studies comparing GM and control varieties were performed by a few groups. Corpillo *et al*. [15] used two-dimensional electrophoresis (2-DE) to study GM virus-resistant tomato plants and their parent line counterpart. They found no significant differences between the two lines, neither qualitatively nor quantitatively. Lehesranta *et al*. [16] applied 2-DE proteomics to obtain protein profiles for GM as well as a range of different non-GM potato varieties. They concluded that there was less variation between GM lines and their non-GM counterparts compared to the natural variation they observed in the different WT potatoes.

The data in this thesis show that, although work still needs to be done in terms of standardisation and validation, the methodology of transcriptomics has the potential to detect large as well as small differences in gene expression. The detection of large differences in gene expression profiles was primarily shown in the comparative studies of gene expression profiles of subsequent ripening stages in the tomato. Considerable alterations in gene expression profiles were seen, a selected number of observed differences was put to the test and the observations could in most cases be confirmed by real-time PCR analysis. The possibility to detect small alterations in gene expression was convincingly illustrated by the analogously altered profiles as observed in the two GM tomato transformant lines. These coincidental transcriptional changes were small, but formed by far the majority of the differentially expressed genes in these two lines. Similar effects were seen in the Arabidopsis transformant lines.

At the same time the experiments showed that the interpretation of observed differences, i.e. the significance of observed differences as well as the toxicological and nutritional relevance thereof, is highly depended on sufficient insight into the natural variation in expression of the selected genes. New data analysis tools, such as Gene Set Enrichment Analysis [17] or other pathway analysis tools, such as KaPPa-view [18] (http://kpv.kazusa.or.jp/kappa-view/), Visant [19,20] (http://visant.bu.edu/), and GScope 3 [21](http://omicspace.riken.jp/osml/), are, however, now becoming available to further aid in the assessment of the biological and physiological significance of observed differences using pathway- and network-based analysis of altered gene expression profiles.

The CSA is a tiered approach. Therefore it needs to established for each observed differential gene expression whether it is biologically relevant. This holds true for transcriptomics as well as for other 'omics' technologies [14,15]. Further standardization of array platforms and data mining tools is furthermore crucial for the routine determination of biological significance in the light of natural variation. A major development in this respect has been the introduction of (Plant) MIAME (Minimum Information About a Microarray Experiment), that requires a structured representation of the data and associated metadata of the experiment to enable future use of microarray data [22,23] (http://www.mged.org/Workgroups/MIAME/MIAME-plant_Dec2005.pdf).

If on the basis of this first step in the tiered CSA approach biologically significant differential gene expression can not be excluded, the toxicological and nutritional significance of the observed differences needs to be evaluated in the second step. Also in this phase the detected differences are considered in the light of the established natural variation. It is decided on a case-by-case basis whether the combination of knowledge of the expression products of the differentially expressed gene, the bandwidth of natural variation, the data on the basis of this bandwidth, and the observed differences in gene expression for this gene as well as for related ones within a metabolic network, should warrant additional confirmative analyses. And subsequently, if the findings are confirmed, it needs to be determined whether additional toxicological or nutritional investigations are required to establish the food safety of the new plant variety. This may seem a lengthy exercise but it is anticipated that the additional information obtained will increasingly facilitate data interpretation.

The data in the thesis furthermore show that differential gene expression in two GM transformant lines is relatively (very) limited, especially when taking into account the natural variation that was observed in similar tomato lines during two subsequent stages of ripening. This is in line with findings reported in other scientific publications on omics technologies showing that differences in GM lines as compared to wt lines are generally small and well within natural variation limits, especially when current methods of genetic modification are applied. Furthermore, it was proven plausible that the majority of the observed effects were indeed effects that were directly related to the introduced genetic construct and its expression products.

So far, the safety assessment of new plant varieties is still limited to the assessment of intended alterations, unintended effects are only assessed in the case of GMO-derived plant products. This thesis advocates incorporating safety criteria as additional criteria in plant breeding selection schemes for all new plant varieties. Omics methodologies, including transcriptomics, can be valuable tools to screen for unintended alteration in the physiology of the plant. Transcriptomics is for the moment the omics methodology that has best coverage over all plant metabolic networks, but in time proteomics and/or metabolomics may provide more direct insight into potential unintended physiological alterations. To this end it will be necessary to develop improved protocols that take into account procedures for sampling, experimental procedures and data analysis. These protocols can also build on already

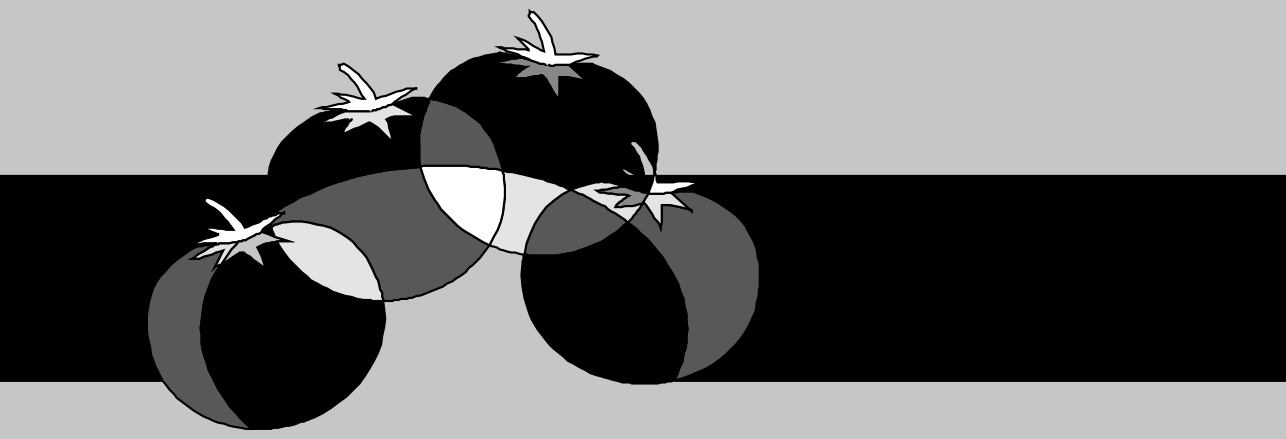
available (standardized) information on the natural variation for the genes included in the (transcript)omics experiments. When applying the omics technologies during the initial phases of the plant breeding process, irrespective of the applied breeding strategy, compilation of the food safety-related data can indeed be performed as part of the entire plant breeding process and at the same time the food safety of the final plant variety may be better assured. This combination of the application of 'omics' technologies to detect potentially unintended effects during the subsequent steps of the breeding process and the targeted analyses on key compounds with relation to the toxicologically and nutritionally relevant metabolic pathways will make it highly unlikely that undesired or detrimental effects of any breeding process will remain unnoticed.

References

1. Griffin J.L., 2006. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? Phil. Trans. R. Soc. B., 361, 147-161.
2. Hollywood K., Brison D.R., and Goodacre R., 2006. Metabolomics: current technologies and future trends. Proteomics 6, 4716-4723.
3. Last R.L., Jones A.D., and Shachar-Hill Y.,, 2007. Towards the plant metabolome and beyond. Nature Reviews Molecular Cell Biology 8, 167-174.
4. De Luca V., and St Pierre B., 2000. The cell and developmental biology of alkaloid biosynthesis. Trends in Plant Science 5, 168-173.
5. Rampitsch C., and Srinivasan M., 2006. The application of proteomics to plant biology: a review. Canadian Journal of Botany 84,883-892
6. Wu W.W., Wang G., Baek S.J., and Shen R.-F., 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. Journal of Proteome Research 5, 651-658.
7. Chen S., and Harmon A.C. 2006. Advances in plant proteomics. Proteomics 6, 5504-5516.
8. Rensink W.A., and Buell C.R., 2005. Microarray expression profiling resources for plant genomics. Trends in Plant Science 10(12), 603-609.
9. Childs K.L., Hamilton J.P., Zhu W., Ly E., Cheung F., Wu H., Rabinowicz P.D., Town C.D., Buell R., and Chan A.P., 2007. The TIGR plant transcript assemblies database. Nucleic Acids Research, 35, 846-851.
10. Baudo, M.M., Lyons, R., Powers, S., Pastori, G.M., Edwards, K.J., Holdsworht, M.J., and Shewry, P.R., 2006 Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. Plant Biotechnol. J. 4, 369-380.
11. Baker, J.M., Hawkins, N.D., Ward, J.L., Lovegrove A., Napier, J.A., Shewry, P.R., and Beale, M.H., 2006. A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotech J. 4, 381-392.
12. LeGall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J., Verhoeven, M.E., 2004. Metabolite profiling of tomato (*Lycopersicon esculentum*) using 1H-NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. J. Agricult. Food Chem. 51, 9, 2447-2456.

13. 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. and Draper. J., 2005. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. PNAS 102, 40, 14458-14462.
14. Colquhoun, I.J., Le Gall, G., Elliott, K.A., Mellon, F.A., and Michael, A.J., 2006. Shall I compare thee to a GM potato? TIG 22, 10, 525-528.
15. Corpillo D., Gardini G., Vaira A.M., Basso M., Aime S., Accotto G.P., and 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(1), 193-200.
16. Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., 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 Physiol. 138,3, 1690-1699.
17. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihaq, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamavo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D., Groop, L.C., 2003. Nature Gen. 34, 3, 267-273.
18. Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., Umezawa, T., Misawa, N., Saito, K., and Shibata, D., 2005. KaPPa-view. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol 138, 1289-1300.
19. Hu, Z., Mellor, J., Wu, J. and DeLisi, C. 2004. VisANT: an online visualization and analysis tool for biological interaction data. BMC Bioinformatics, 5, 17.
20. Hu, Z., Mellor J., Wu J., Kanehisa M., Stuart JM., and DeLisi C., 2007. Towards zoomable multidimensional maps of the cell. Nat Biotechnol, 25(5): p. 547-54.
21. Hasegawa Y., Seki M., Mochizuki Y., Heida N., Hirosawa K., Okamoto N., Sakurai T., Satou M., Akiyama K., Iida K., Lee K., Kanaya S., Demura T., Shinozaki K., Konagaya A., and Toyada T., 2006. A flexible representation of omic knowledge for thorough analysis of microarray data. Plant Methods 2, 5.
22. Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J., Vingron M., 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature Genetics 29, 365-371.
23. Zimmermann Ph., Schildknecht B., Craigon D., Garcia-Hernandez M., Gruissem W., May S., Mukherjee G., Parkinson H., Rhee S., Wagner U., and Hennig L., 2006. MIAME/Plant – adding value to plant microarray experiments. Plant Methods 2, 1.

Chapter 9.



Samenvatting en conclusies

De belangrijkste omics-technologieën zijn 1) transcriptomics: het maken van profielen van mRNA populaties of gentranscripten, 2) proteomics: het maken van eiwitprofielen, en 3) metabolomics: het maken van profielen van secundaire metabolieten. Metabolomics maakt gebruik van een serie verschillende technologieën, meestal ^{1}H-NMR-spectroscopie of de meer gevoelige combinaties van methoden zoals LC-MS of GC-MS ((liquid of gas) chromatography-mass spectrometry), HPLC (high performance liquid chromatography) of EC (electrochemical arrays) [1,2,3]. Het voordeel van (de meeste van) deze methoden is dat ze 'high throughput' zijn en dat ze per monster relatief goedkoop zijn, maar metabolomics heeft tegelijkertijd als doel metabolieten in kaart te brengen die aanwezig kunnen zijn met een 10^{9} variatie in concentratie en met een log polariteit die varieert van ongeveer -6 tot 14 [1]. Het aantal metabolieten in een metabolietprofiel is in het algemeen maximaal 1000, en daarmee waarschijnlijk minder dan 1% van de metabolieten die in een organisme voorkomen [4], maar dit percentage kan hoger zijn in individuele celsystemen of organen. Proteomics wordt meestal uitgevoerd als twee-dimensionele (2D) electroforese gekoppeld aan massa-spectrometrie of op basis van multidimensionele eiwit-identificatietechnieken [5,6]. Geïdentificeerde eiwitten in proteomics-studies zijn meestal hydrofiele eiwitten die in hoge concentratie aanwezig zijn, de dynamische range kan tot 10^{6} groot zijn [7]. Er worden in het algemeen geen hydrofobe eiwitten, zoals membraaneiwitten, gedetecteerd. Er zijn ontwikkelingen gaande om het bereik van proteomics-methodologieën verder te vergroten en de analyse van proteomics-profielen te verbeteren. Op dit moment kunnen er meestal tot enkele honderden eiwitten tegelijk geanalyseerd worden in een enkel experiment, wat waarschijnlijk maar een beperkt deel van het plantenproteoom is (1% of minder), wanneer uitgegaan wordt van het totaal aantal genen in een plantengenoom (geschat op 25.000-40.000) en het gegeven dat al deze genen kunnen resulteren in verschillende eiwitten door alternatieve manieren van gentranscriptie en RNA-splitsing. Ook hier geldt dat het percentage eiwitten in een eiwitprofiel groter kan zijn in specifieke celsystemen of weefsels, mogelijk zelfs tot zo'n 10%, maar het blijft lastig om dit in proteomics-experimenten enigszins accuraat te schatten. Transcriptomics, ten slotte, maakt gebruik van (door middel van Polymerase Chain Reaction (PCR)-verveelvoudigde) cDNA of oligonucleotide microarrays. Microarrays zijn in het algemeen robuuste systemen voor genexpressie-analyse en ze resulteren in reproduceerbare gegevens, ook al omdat alle te analyseren mRNA-moleculen vergelijkbare biochemische eigenschappen hebben. Tot dusver zijn voor planten in beperkte mate zogenaamde 'whole transcriptome arrays' beschikbaar, arrays waarop alle mRNAs die in een plant voor kunnen komen vertegenwoordigd zijn. Dit is met name het geval voor de modelgewassen Arabidopsis en rijst, voor andere gewassen zijn er soortspecifieke arrays beschikbaar met tot zo'n 30.000 transcriptsequenties, die tesamen meer dan 50% van het totale transcriptoom kunnen vertegenwoordigen [8,9]. Het belangrijkste nadeel dat aan transcriptomics kleeft is het feit dat gevonden verschillen in genexpressie niet in alle gevallen overeen hoeven te komen met verschillen in het proteoom en/of metaboloom en daarmee met relevante verschillen in de fysiologie van de plant. Het zal om die reden altijd nodig zijn om gevonden verschillen in genexpressie-patronen in de plant te bevestigen door middel van specifieke analyses van plantbestanddelen (eiwitten en/of metabolieten) die bij de mogelijk veranderde metabole netwerken betrokken zijn.

Dit proefschrift evalueert de ontwikkelingen in de veiligheidsanalyse van complexe voedselproducten die afkomstig zijn van zowel genetisch gemodificeerde organismen (GGO's) als van conventioneel veredelde gewassen. Specifieke aandacht wordt hierbij gegeven aan de mogelijkheid om omics-technologieën, en met name transcriptomics, toe te passen als onderdeel van deze voedselveiligheidsanalyse.

In hoofdstuk 1 wordt het concept 'Comparative Safety Assessment' (CSA), letterlijk de 'vergelijkende veiligheidsevaluatie', geïntroduceerd als het leidend principe in de veiligheidsanalyse van complexe producten afkomstig van voedselgewassen. Dit CSA-principe vervangt het vergelijkbare, maar meer controversiële 'Principle of Substantial Equivalence', letterlijk het principe van aanzienlijke gelijkwaardigheid, dat vaak geïnterpreteerd werd als zouden de resultaten het eindpunt van de analyse zijn. De CSA legt de nadruk op de analyse van de nieuw ontwikkelde en een vergelijkbare, verwante plantenvariëteit. Deze vergelijking vormt de eerste stap van de veiligheidsanalyse. Gevonden verschillen worden vervolgens geëvalueerd met inachtneming van de beoogde verschillen ten gevolge van het veredelingsproces en van de natuurlijke variatie in de onderzochte variëteiten en in andere (gerelateerde) variëteiten die al op de markt zijn. De resultaten van deze vergelijking vormen vervolgens de basis voor de toxicologische en nutritionele evaluatie. Er wordt benadrukt dat deze benadering niet alleen van toepassing is op plantproducten die afkomstig zijn van GGO's, maar de benadering kan breder toegepast worden als een universele veiligheidsanalyse voor nieuwe producten die geproduceerd zijn met het scala aan verschillende veredelingsstrategieën die tegenwoordig in gebruik zijn binnen de landbouw.

In hoofdstuk 2 worden de verschillende strategieën voor voedselveiligheidsbeoordeling van producten afkomstig van genetisch gemodificeerde (GG) dieren besproken en geëvalueerd in het licht van de voorafgaande en gelijktijdige ontwikkelingen in de plantenveredeling. Een van de belangrijkste aspecten bij de vergelijking van de veiligheidsevaluatie van producten afkomstig van GG dieren en GG planten is het aantal GG organismen dat beschikbaar is voor de evaluatie: bij planten kan een enkel transformatie-experiment resulteren in honderden GG planten, terwijl het aantal GG dieren afkomstig van één experimentele transformatiereeks vaak zeer klein is, vaak minder dan 10. In theorie kan dit enerzijds leiden tot minder stringente selectiecriteria in het geval van GG dieren, maar ook tot een lager aantal dieren dat beschikbaar is voor de veiligheidsevaluatie. Daar dit geldt voor zowel de GG dieren als voor de verwante dieren in de vergelijkende veiligheidsanalyse, kan dit ertoe leiden dat het moeilijker wordt om de toxicologische en nutritionele betekenis van gevonden verschillen goed in te schatten. Een ander verschil bij de veiligheidsevaluatie van GG dieren is het feit dat antinutritionele stoffen nauwelijks bekend zijn in dierlijke producten, in vergelijking tot de grote aantal geïdentificeerde natuurlijke toxinen in verschillende plantensoorten. Het gevolg is dat het in het algemeen niet mogelijk is om een vergelijking te maken van mogelijk veranderde niveaus van deze antinutriënten, zoals bij planten gebruikelijk is. De nadruk van de CSA komt hiermee voor GG dieren te liggen bij veranderde nutriëntenprofielen. Een laatste verschil betreft de beschikbaarheid van traceerbaarheidssystemen: deze zijn algemeen in de dierlijke

productiesector, maar niet in de plantaardige, waardoor de mogelijkheid van 'post-marketing surveillance' in het geval van GG dieren meer reëel is dan bij GG planten. Toepassing van post-marketing surveillance kan worden aanbevolen wanneer er na de CSA nog onzekerheden zijn ten aanzien van geschatte inname van het product door de consument of, in uitzonderlijke gevallen, wanneer er onzekerheid is ten aanzien van mogelijke allergeniteit van (een) nieuw(e) eiwit(ten) in het GG product. De conclusie wordt getrokken dat het CSA principe ook van toepassing is op voedselproducten afkomstig van GG dieren, maar dat er ten aanzien van deze hierboven genoemde aspecten specifieke aanpassingen in de praktische uitwerking van het principe nodig kunnen zijn.

Eén van de aanbevelingen van internationale expertgroepen op gebied van de veiligheidsevaluatie van voedselproducten afkomstig van GGO's, is dat de nieuwe analytische technieken, de zogenaamde omics-technologieën, geëvalueerd zouden moeten worden ten aanzien van de mogelijkheid om, als onderdeel van de CSA, onbedoelde effecten op te kunnen sporen in complexe producten al dan niet afkomstig van GGO's. De omics-technologieën zouden daarmee tevens een bijdrage kunnen leveren aan de toxicologische en nutritionele evaluatie van de gevonden verschillen. Hoofdstuk 3 beschrijft één van de eerste experimentele toepassingen van transcriptomics op dit terrein: een beperkte tomatenarray werd ontwikkeld met informatieve cDNAs die afkomstig zijn van twee specifieke cDNA-banken. Eén bank is het resultaat van subtractie van de mRNA populatie van de groene tomaat van de mRNA populatie van de rode tomaat. De resulterende cDNAs zouden als gevolg van de subtractieprocedure meer sequenties bevatten die betrokken zijn bij de metabole routes die verband houden met de nutritionele en gezondheidsbevorderende eigenschappen van de rode tomaat. De andere cDNA-bank is op de omgekeerde wijze verkregen, resulterend in een verrijkte CDNA-bank, die met name sequenties bevat die betrokken zijn bij vorming van anti-nutriënten en natuurlijke toxinen, die met name voorkomen in de groene, onrijpe tomaat. Deze 'Representational Difference Analysis' (RDA) -benadering voor het maken van informatieve plantenarrays wordt geëvalueerd: hoewel normalisatie van de mRNA populatie, waarbij iedere unieke mRNA sequentie slechts één maal is vertegenwoordigd, niet wordt bereikt, wordt er wel aangetoond dat de procedure efficiënt is als eerste stap om dit doel te bereiken.

Hoofdstuk 4 beschrijft hoe de in hoofdstuk 3 ontwikkelde tomatenarray is gebruikt om mRNA te hybridiseren van tomaten in vijf opeenvolgende rijpingsstadia: groen, 'breaker' (het stadium waarin de tomaat geel begint te worden), 'turning' (het stadium waarin de tomaat oranje begint te worden), lichtrood en rood. Hierdoor wordt een genexpressiedatabestand verkregen van de tomaat in verschillende rijpingsfasen, die gebruikt kan worden bij toekomstige vergelijkingen als onderdeel van een CSA. Tegelijkertijd wordt deze eerste serie experimenten gebruikt om te bepalen of het mogelijk is om eenvoudige criteria vast te stellen ten aanzien van het ontwikkelings- of rijpingsstadium bij het bemonsteren van vruchten voor omics-analyses in het kader van een CSA. De experimenten tonen aan dat de opzet van omics-experimenten, i.e. van transcriptomics, van groot belang is voor de interpretatie van resultaten: kleine verschillen in de rijpingsfase van de voor een CSA geselecteerde tomaten kunnen leiden tot grote verschillen in

het genexpressieprofiel, die niet gerelateerd zijn aan de genetische veranderingen. Aangetoond werd dat de te onderzoeken monsters goed op elkaar afgestemd moeten zijn voor wat betreft de ontwikkelings- of rijpingsfase van de weefsels. Dit geldt ook voor proteomics want ook eiwitprofielen laten duidelijke verschillen zien in verschillende rijpingsstadia. Aanbevolen wordt om de omics-technologieën al in een vroeg stadium van de veredelingsprocedures toe te passen, zodat de factor 'voedselveiligheid' meegenomen wordt in het veredelingsproces. De conclusie is dat de omics-technologieën verschillen aan kunnen tonen die mogelijk niet opgemerkt zouden zijn bij de gerichte analyse van de samenstelling van de plant, zoals die nu als onderdeel van een CSA wordt uitgevoerd.

In hoofdstuk 5 wordt dezelfde tomatenarray gebruikt om twee GG transformantlijnen en de bijbehorende ouderlijn te analyseren. Omdat in beide transformantlijnen hetzelfde genetische construct is ingebouwd, maar in verschillende locaties van het genoom tijdens het random transformatieproces, is het mogelijk om de gevonden verschillen tussen transformantlijn en ouderlijn te verdelen in enerzijds effecten die direct gerelateerd zijn aan het geïntroduceerde genetische construct, en anderzijds andere, met name onbedoelde, effecten. De resultaten van de vergelijkende transcriptomics-experimenten worden geanalyseerd in het licht van wat bekend is van gerichte analyses van de transformantlijnen en op basis van de genexpressieprofielen in de rijpingsstadia van de tomaat. Er wordt aangetoond dat verschillen tussen de twee GG lijnen en de WT ouderlijn beperkt zijn en dat deze verschillen voor de beide GG lijnen in belangrijke mate overeenkomen. Hierdoor is het aannemelijk dat de gevonden verschillen inderdaad gerelateerd zijn aan het ingebouwde genetisch construct en dus niet als onbedoelde effecten kunnen worden beschouwd.

In hoofdstuk 6 worden vergelijkbare experimenten beschreven om de onbedoelde neveneffecten te bepalen in Arabidopsis-transformantlijnen. In deze studie worden een aantal verschillende transformantlijnen geanalyseerd met verschillende aantallen ingebouwde constructen en verschillende orientaties van het ingebrachte genetische construct. Wel hadden alle lijnen slechts een enkele insertieplaats. Verschillen in genexpressie werden vooral gevonden in genen waarvan bekend is dat ze stress-gerelateerd zijn. Maar hier bemoeilijkt de beperkte kennis van de natuurlijke variatie in genexpressie van de onderzochte genen de toxicologische en nutritionele evaluatie van de gevonden verschillen, vooral ook omdat kleine verschillen in de omgevingscondities bij de kweek van de planten niet kunnen worden uitgesloten. Deze experimenten tonen aan dat het cruciaal is om voldoende informatie te hebben met betrekking tot de natuurlijke variatie in genexpressie in de te onderzoeken voedselgewassen en afgeleide weefsels om omics-gegevens op een juiste manier te kunnen interpreteren wanneer de omics-technologie wordt toegepast in het kader van een CSA.

Hoofdstuk 7 geeft een overzicht van het internationale debat met betrekking tot de veiligheidsevaluatie van complexe plantproducten, al dan niet afkomstig van GGO's, op basis van moderne veredelingsprocedures, en van de publicaties die omics-experimenten beschrijven die zich richten op het vinden van onbedoelde effecten van verschillende veredelingstechnieken,

inclusief de experimenten die in dit proefschrift worden beschreven. Op basis van de resultaten wordt de huidige veiligheidsanalyse van (GG) complexe plantproducten geëvalueerd en een nieuwe overall benadering voorgesteld. Geadviseerd wordt om alle nieuwe plantvariëteiten te screenen, niet alleen de GG variëteiten, met betrekking tot hun nieuwe eigenschappen door de opeenvolgende stappen van het CSA-principe van geval tot geval toe te passen. Op die manier kunnen de eindpunten van een CSA variëren van een informatieve 'fact sheet' met samenstellingsgegevens wanneer het een kruising van bekende plantvariëteiten betreft, die de afwezigheid van nieuwe veiligheidsrisico's bevestigt, tot de volledige toxicologische en nutritionele karakterisering van nieuwe plantvariëteiten waar duidelijke risico's zijn geïdentificeerd of waar geen traditionele tegenhanger beschikbaar is voor de vergelijkende analyse. Omics-technologieën kunnen onderdeel zijn van de richtinggevende eerste analytisch vergelijkende stap van de CSA om verschillen te identificeren met bekende, vergelijkbare plantvariëteiten die al op de markt zijn, voor zover beschikbaar.

Een beperkt aantal studies heeft omics-technologieën al toegepast om verschillen tussen GG gewasvariëteiten en traditioneel veredelde tegenhangers op te sporen. Baudo *et al*. [10] analyseerden verschillende GG en conventionele tarwevariëteiten met behulp van transcriptomics. Zij onderzochten daarbij met name de bandbreedte van de natuurlijke variatie en vonden dat de natuurlijke variatie in genexpressiepatronen in conventionele tarwevariëteiten veel groter is dan de variatie tussen GG lijnen. Dit is overigens overeenkomstig de aanname dat de toepassing van gentechnieken in het algemeen maar een beperkte genetische verandering teweegbrengt in vergelijking tot traditionele veredelingsstrategieën waar recombinatie veelvuldig voorkomt en resulteert in veranderde fenotypes, inclusief veranderde genexpressieprofielen.

Metabolomics studies zijn toegepast in vergelijkende studies tussen GG en non-GG variëteiten. Voorbeelden hiervan zijn de NMR en GC-MS-studies van in het veld geteelde tarwevariëteiten [11]. Zowel GG als controlelijnen werden geanalyseerd en er werden verschillen gevonden, maar niet in alle testlocaties in het experiment. Dit kan erop wijzen dat sommige effecten alleen onder bepaalde omgevingscondities tot uiting kunnen komen, maar het kan ook zijn oorzaak vinden in het feit dat de verschillen tussen beide lijnen klein zijn en daarmee niet in alle gevallen statistisch significant. De laatste mogelijkheid wordt verder ondersteund door het gegeven dat alle gevonden verschillen tussen GG en controlelijnen binnen de bandbreedte van de natuurlijke variatie vielen in de verschillende locaties. Dergelijke studies zijn ook uitgevoerd door anderen: Le Gall *et al*. [12] onderzochten verschillen tussen extracten van GG tomaten en controlelijnen op basis van NMR spectra. Zij vonden verschillen in een aantal metabolieten die geen direct verband hielden met de genetische modificatie, maar de verschillen waren beperkt en binnen de bandbreedte van natuurlijke variatie. Zij concludeerden dat NMR kleine verschillen in metabolietprofielen op kan sporen, die mogelijk gerelateerd zijn aan onbedoelde effecten in genetisch gemodificeerde gewassen. Catchpole *et al*. [13] beschrijven een studie met in het veld geteelde aardappelen, een GG lijn en conventionele controle-aardappelen, die beide geanalyseerd werden met een tweestaps MS benadering. Zij vonden een grote variatie in metabolietprofielen tussen de geteste conventionele cultivars. De gevonden verschillen tussen

de GG lijn en de ouderlijn viel binnen deze vastgestelde natuurlijke variatie. Colquhoun *et al*. [14] hebben deze studie van Catchpole *et al*. [13] verder geanalyseerd en concludeerden dat subtiele metabole veranderingen aangetoond kunnen worden met metabolomics.

Er zijn enkele proteomics studies uitgevoerd, waarbij GG lijnen en controlelijnen zijn vergeleken. Corpillo *et al*. [15] gebruikten twee-dimensionele electroforese (2-DE) om GG virusresistente tomatenplanten te vergelijken met de ouderlijn. Er werden geen duidelijke kwantitatieve noch kwalitatieve verschillen gevonden tussen beide lijnen. Lehesranta *et al*. [16] pasten 2-DE proteomics toe voor de vergelijking van eiwitprofielen in zowel GG lijnen als een serie verschillende non-GG aardappelvariëteiten. Zij concludeerden dat er minder variatie is tussen GG lijnen en de vergelijkbare conventionele tegenhangers dan tussen verschillende WT aardappelvariëteiten.

Al deze studies laten alle zien dat er nog verdere standaardisering en validatie nodig is van de technologieën. Tegelijkertijd lijkt het wel duidelijk dat deze min of meer open analytische benaderingen een zinvolle bijdrage kunnen leveren aan de veiligheidsevaluatie van plantproducten, waarbij ook de onbedoelde effecten in nieuwe plantvariëteiten meegenomen kunnen worden in de beoordeling.

De gegevens in dit proefschrift laten zien dat de transcriptomics methodologie de mogelijkheid heeft om zowel grote als kleine verschillen in genexpressie te detecteren. De detectie van grote verschillen werd met name aangetoond in de vergelijkende studies van genexpressieprofielen van opeenvolgende rijpingsstadia in de tomaat. Er werden aanzienlijke verschillen gevonden tussen de profielen. Een aantal van deze verschillen werden onderworpen aan bevestigingsexperimenten op basis van PCR (polymerase chain reaction), waarbij in de meeste gevallen de resultaten van de transcriptomics-studie inderdaad bevestigd konden worden. De mogelijkheid om ook kleine verschillen in genexpressie te kunnen waarnemen, werd geïllustreerd aan de hand van de identiek veranderde profielen in de twee GG tomatentransformantlijnen. De overeenkomende veranderingen in genexpressie ten opzichte van de WT lijn waren klein, maar vormden de meerderheid van de differentieel tot expressie komende genen in de beide lijnen. Vergelijkbare effecten werden waargenomen in de verschillende Arabidopsislijnen.

Tegelijkertijd laten de experimenten zien dat de interpretatie van de waargenomen verschillen, i.e. de biologische relevantie, maar ook de toxicologische en nutritionele relevantie, zeer afhankelijk is van voldoende inzicht in de natuurlijke variatie in de expressie van de geselecteerde genen. Nieuwe data-analyse-programma's als Gene Set Enrichment Analysis [17], of andere analyse-programma's ten aanzien van metabole routes, zoals KaPPa-view [18] (http://kpv.kazusa.or.jp/kappa-view/), Visant [19,20] (http://visant.bu.edu/), of GScope 3 [21] (http://omicspace.riken.jp/osml/) komen nu beschikbaar en kunnen een belangrijke bijdrage leveren aan het bepalen van de biologische en fysiologische betekenis van gevonden genexpressieverschillen op basis van metabole route en netwerkanalyse.

De CSA is een getrapte benadering. De eerste stap is de detectie van verschillen en het vaststellen voor elk waargenomen verschil of het een biologisch relevant verschil is. Om die reden is de monstername van cruciaal belang: er wordt in dit proefschrift aangetoond dat, voor een directe vergelijking, de plantmonsters zoveel mogelijk gekweekt moeten worden onder identieke condities en dat de monstername rekening moet houden met het ontwikkelingsstadium van de geselecteerde plantenorganen en/of -weefsels. Dit geldt voor zowel transcriptomics als voor de andere omics-technologieën [14,15]. Verdere standaardisering van arrays en data-opslagprogramma's is van groot belang voor de routinematige bepaling van de biologische significantie in het licht van de natuurlijke variatie. Een belangrijke ontwikkeling op dit terrein is de introductie van (Plant) MIAME (Minimum Information About a Microarray Experiment), dat zowel een gestructureerde weergave van de data vereist als de bijbehorende metadata van het experiment om de microarray data ook in de toekomst te kunnen gebruiken [22,23] (http://www.mged.org/Workgroups/MIAME/MIAME-plant_Dec2005.pdf).

Wanneer op basis van deze eerste stap biologische significantie van de gevonden differentiële genexpressie niet kan worden uitgesloten, dient de toxicologische en nutritionele significantie van de waargenomen verschillen in de tweede stap te worden geëvalueerd. Ook in deze fase moeten de gevonden verschillen geëvalueerd worden in het licht van de vastgestelde natuurlijke variatie. Van geval tot geval dient te worden bezien of de combinatie van kennis van het expressieproduct van het differentieel tot expressie komende gen, de bandbreedte van natuurlijke variatie, de data die aan deze bandbreedte ten grondslag liggen en de waargenomen verschillen in genexpressie voor dit gen en voor gerelateerde genen binnen een metabool netwerk leidt tot de noodzaak voor additioneel, bevestigend onderzoek. Wanneer het verschil bevestigd wordt, moet vervolgens verder bekeken worden of aanvullend toxicologisch of nutritioneel onderzoek nodig is. Dit lijkt een tijdrovende exercitie, maar verondersteld wordt dat de aanvullende informatie in toenemende mate de data-interpretatie zal vereenvoudigen.

De gegevens in dit proefschrift laten zien dat de differentiële genexpressie relatief (zeer) beperkt is in twee GG transformantlijnen, met name wanneer rekening gehouden wordt met de natuurlijke variatie zoals die waargenomen is in vergelijkbare tomatenlijnen tijdens twee opeenvolgende rijpingsstadia. Dit komt overeen met de bevindingen in verschillende wetenschappelijke publicaties waarin ook wordt gemeld dat de gevonden verschillen in het algemeen klein zijn en binnen de limieten van de natuurlijke variatie vallen, met name wanneer de huidige methoden van genetische modificatie worden toegepast. Het is verder aannemelijk gemaakt dat de meerderheid van de waargenomen effecten effecten waren die direct gerelateerd zijn aan het geïntroduceerde genetische construct en de afgeleide expressieproducten daarvan.

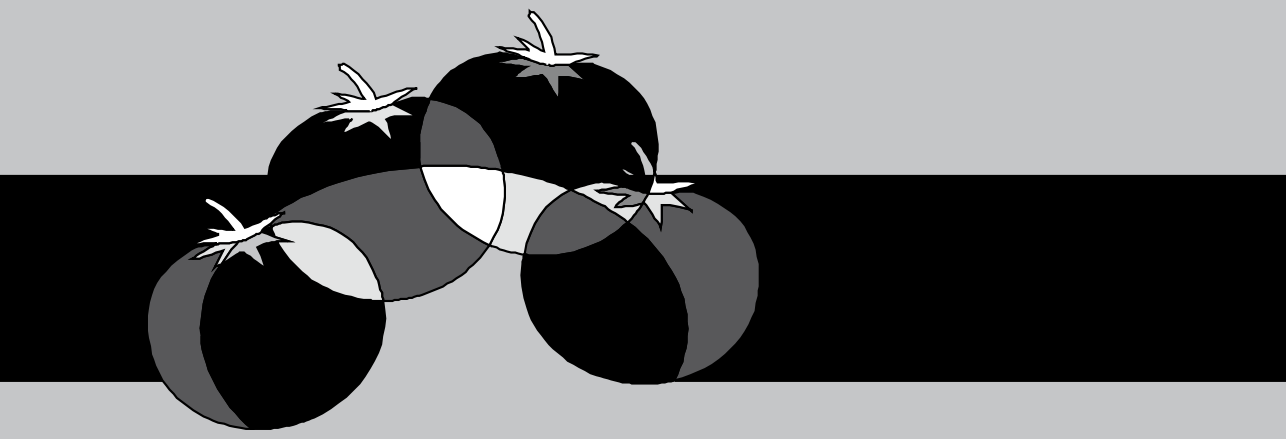
Tot dusver wordt de veiligheidsevaluatie van nieuwe plantenvariëteiten beperkt tot de evaluatie van bedoelde veranderingen, onbedoelde veranderingen worden alleen onderzocht in het geval van plantproducten afkomstig van GGO's. Dit proefschrift stelt voor om veiligheidscriteria als additionele criteria mee te nemen in schema's voor plantenveredeling voor alle nieuwe

plantvariëteiten. Omics-methodologieën, waaronder transcriptomics, kunnen een waardevolle bijdrage leveren aan de screening van onbedoelde veranderingen in de fysiologie van de plant. Transcriptomics is op dit moment de omics-technologie die de metabole netwerken in de plant het breedst in kaart kan brengen, in de toekomst zullen proteomics en metabolomics mogelijk een meer direct inzicht kunnen geven in mogelijk onbedoelde fysiologische veranderingen. Om dit op een verantwoorde manier te kunnen doen, is het van belang dat er gestandaardiseerde procedures komen voor monstername, opzet en uitvoering van omics-experimenten en procedures voor data-analyse waarbij verder gebouwd kan worden op reeds aanwezige (gestandaardiseerde) gegevens ten aanzien van de natuurlijke variatie voor de genen en (afgeleide) expressieproducten die geanalyseerd worden in de omics-experimenten.

Wanneer de omics-technologieën op deze wijze toegepast worden in de initiële fasen van een plantenveredelingsproces, dan kunnen de verzamelde, voedselveiligheidsgerelateerde data een onderdeel vormen van het gehele veredelingsproces en tegelijkertijd kan de voedselveiligheid van de uiteindelijke nieuwe plantvariëteit beter verzekerd zijn. Dit geld niet alleen voor producten afkomstig van GG plantvariëteiten, maar voor alle nieuwe plantproducten, zonder onderscheid naar veredelingsprocedure. De combinatie van de toepassing van omics-technologieën om mogelijke onbedoelde effecten te detecteren tijdens de opeenvolgende stappen van het veredelingsproces en de gerichte analyses van belangrijke componenten met betrekking tot de toxicologische en nutritioneel relevante metabole routes maakt de kans erg klein dat mogelijke ongewenste of negatieve effecten van het veredelingsproces onopgemerkt blijven.

Referenties

Zie Chapter 8.



About the author

Publication list

Peer-reviewed papers

Comparative safety assessment of plant-derived foods. **EJ Kok**, J. Keijer, GA Kleter and HA Kuiper. Regulatory Toxicology and Pharmacology. 2008, 50: 98-113

Changes in gene and protein expression during tomato ripening – consequences for the safety assessment of new crop plant varieties. **Esther J Kok**, Satu J Lehesranta, Jeroen P van Dijk, J Richard Helsdingen, Wilko TP Dijksma, Angeline MA Van Hoef, Kaisa M Koistinen, Sirpa O Kärenlampi, Harry A Kuiper and Jaap Keijer. Food Science and Technology International (in press).

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. **Esther J. Kok**, Nicole L.W. Franssen-van Hal, Lies N.W. Winnubst, Evelien H. M. Kramer, Wilko T.P. Dijksma, Harry A. Kuiper and Jaap Keijer. Journal of Plant Physiology 2007, 164: 337-349

Evaluation of a non-targeted 'omic' approach in the safety assessment of genetically modified plants. Stine B. Metzdorff, **Esther J. Kok**, Pia Knuthsen, Jan Pedersen. Plant Biology 2006, 8 (5) : 662-672

Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology. B. Chassy, J.J. Hlywka, G.A. Kleter, **E.J. Kok**, H.A. Kuiper, M. McGloughlin, I.C. Munro, R.H. Phipps, J.E. Reid, J. Stein, J. Zabik. Food and Nutrition Bulletin 2005, 26:436-442

Unintended effects and their detection in genetically modified crops. F. Cellini, A. Chesson, I. Colquhoun, A. Constable, H.V. Davies, K.H. Engel, A.M.R. Gatehouseg, S. Kärenlampi, **E.J. Kok**, J.-J. Leguay, S. Lehesranta, H.P.J.M. Noteborn, J. Pedersen, M. Smith. Food and Chemical Toxicology 2004, 42(7, special issue ENTRANSFOOD): 1089-1125

Detection and traceability of genetically modified organisms in the food production chain. M. Miraglia, K.G. Berdal, C. Brera, P. Corbisier, A. Holst-Jensen, **E.J. Kok**, H.J.P. Marvin, H. Schimmel, J. Rentsch, J.P.P.F. van Rie, J. Zagon. Food and Chemical Toxicology 2004, 42(7, special issue ENTRANSFOOD): 1157-1180

Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology – Prepared by a task force of the ILSI International Food Biotechnology Committee. B. Chassy, J.J. Hlywka, G.A. Kleter, **E.J. Kok**, H.A. Kuiper, M. McGloughlin, I.C. Munro, R.H. Phipps, J.E. Reid. Comprehensive Reviews in Food Science and Food Safety 2004, 3(2), 35-104. http://www.ift.org/pdfs/crfsfs/crfsfsv3n2p0035-0104ms20040106.pdf

Comparative Safety Assessment for Biotech Crops. **Esther J. Kok** and Harry A. Kuiper. TRENDS in Biotechnology 2003, 21(10): 439-444

Exploitation of molecular profiling techniques for GM food safety assessment. Kuiper H.A., **Kok E.J.,** and Engel K.-H. Current Opinion in Biotechnology 2003, 14:238-243

Voeding en gezondheid – genetisch gemodificeerd voedsel. H.A. Kuiper, G.A. Kleter en E.J. Kok. Ned. Tijdschr. Geneeskd. 2003, 11 januari 147(2):56-60

Substantial equivalence – an appropriate paradigm for the safety assessment of genetically modified foods? Harry A. Kuiper, Gijs A. Kleter, Hub P.J.M. Noteborn and **Esther J. Kok**. Toxicology 2002, 181-182: 427-431

Safety aspects of novel foods. Harry A. Kuiper, Hub P.J.M. Noteborn, **Esther J. Kok**, Gijs A. Kleter. Food Research International 2002, 35:267-271

Traceability of genetically modified organisms. Henk J.M. Aarts, Jean-Paul P.F. van Rie and **Esther J. Kok**. Expert Rev. Mol. Diagn. 2(2002)1:89-96

DNA methods: Critical Review of Innovative Approaches. **E.J. Kok**, HJM Aarts, AMA Van Hoef, and HA Kuiper. Journal of AOAC International.2002, 85 (3): 797-800

Assessment of the food safety issues related to genetically modified foods. Harry A. Kuiper, Gijs A. Kleter, Hub P.J.M. Noteborn and **Esther J. Kok**. The Plant Journal 2001, 27(6):503-528

The application of cDNA microarrays for analysis of gene expression. Nicole L.W. van Hal, **Esther J. Kok**, Oscar F.J. Vorst, Lies N.W. van der Wal-Winnubst, Ad Peijnenburg, Harry A. Kuiper and Jaap Keijer. Journal of Biotechnology 2000, 78: 271-280

Development and application of a selective detection method for genetically modified soy and soy-derived products. A.M.A. Van Hoef, **E.J. Kok**, E. Bouw, H.A. Kuiper and J. Keijer. Food Additives and Contaminants 1998, 15(7):767-774

Food safety assessment of marker genes in transgenic crops. **Esther J. Kok**, Hub P.J.M. Noteborn, Harry A. Kuiper. Trends in Food Science & Technology 1994, 5:294-298.

Novel food products from genetically modified plants: do they need additional food safety regulations? **E.J. Kok**, A. Reynaerts, H.A. Kuiper. Trends in Food Science & Technology 1993, 4: 42-48

The sequential appearance of components of the synaptonemal complex during meiosis of the female rat. A.J. Dietrich, **E.J. Kok**, H.H. Offenberg, C. Heijting, P. de Boer, A.C.G. Vink. Genome 1992, 35: 492-497

Non peer-reviewed scientific publications (selected)

Use of the cDNA microarray technology in the safety assessment of GM food plants. **E.J. Kok**, G.A. Kleter, J.P. van Dijk. TemaNord report 2003:558. Nordic Council of Ministers, Copenhagen, Denmark, 2003

The food safety risk assessment of GM animals. **Esther J. Kok** and Wendelyn Jones. Working paper for the FAO/WHO Expert Consultation on GM animals, Rome, 17-21 November 2003

Regulation and exploitation of genetically modified crops. Functional foods, nutraceuticals, and plant-derived medicines may improve public relations for transgenic crop technology, but could leave manufacturers with tricky business decisions to make. Gijs A. Kleter, Wim. M. van der Krieken, **Esther J. Kok**, Dirk Bosch, W. Jordi, Luud J.W.J. Gilissen. Nature Biotechnology 2001, 19: 1105-1110

Differential display of mRNA. **E.J. Kok**, E.N.W. van der Wal-Winnubst, A.M.A. Van Hoef and J. Keijer. Molecular Microbial Ecology Manual 8.1.1.:1-10, 2001

Method for detecting differences in gene expression between wild type plants and newly developed plants. Patent number WO0127321. Keijer Jaap (NL), **Esther Kok** (NL), Wilhelmus Maria van der Krieken (NL).

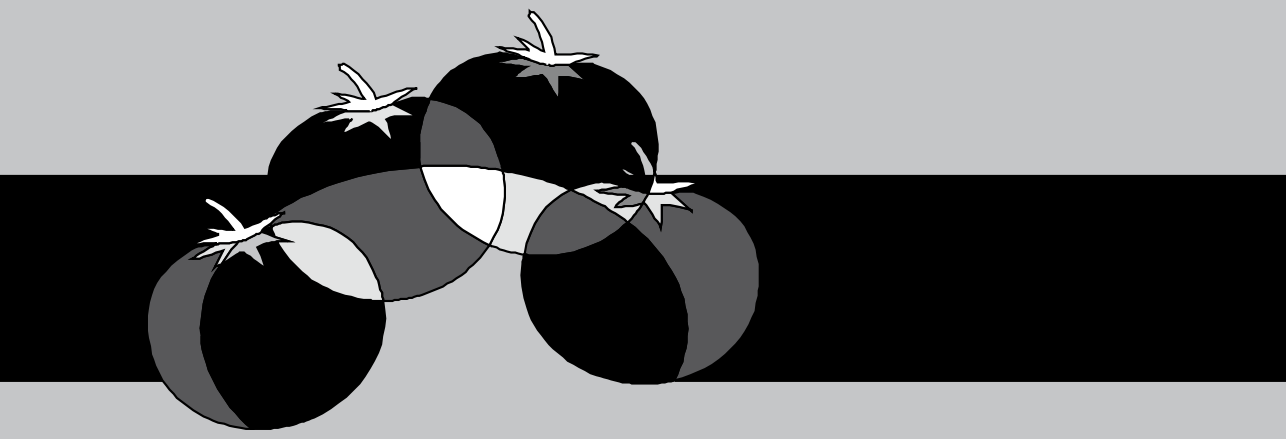
Profiling Techniques to Identify Differences between Foods Derived from Biotechnology and their Counterparts. Harry A. Kuiper, **Esther J. Kok** and Hub J.P.M. Noteborn. FAO/WHO Expert Consultation on foods derived from Biotechnology, 29 May – 2 June, 2000, Geneva, Switzerland

Evaluation of Strategies for Food Safety Assessment of Genetically Modified Agricultural Products – Information Needs. **E.J. Kok** and H.A. Kuiper. Food Safety Evaluation (1996): 80-84. Proceedings OECD Workshop on Food Safety Evaluation, Oxford, September 12-15, 1994

Food safety assessment of the use of marker genes in rDNA products: ongoing projects. H.P.J.M. Noteborn, **E.J. Kok** and H.A. Kuiper. WHO workshop on Health aspects of the use of marker genes in plants and possibilities for their use in identification and control of genetically modified plants. Rungsted, Denmark, september 1993. Working paper WHO/FNU/FOS/93-6, 1993

Curriculum Vitae

Esther Jeannette (van Leeuwe-) Kok was born on 23 July, 1964 in Alkmaar, the Netherlands, third daughter of Willem Cornelis Kok and Riekje Kok-Lycklama à Nijeholt. She completed secondary school (Gymnasium, CSG Jan Arentsz, Alkmaar) in 1982. From 1983 till 1989 she studied Biology at Wageningen Agricultural University. As part of her study she stayed six months in Australia and obtained the Australian Vacation Scholarship for her study at the Australian National University. After her graduation she was appointed in 1998 at the Institute for Animal Husbandry in Zeist as scientific researcher in the field of animal biotechnology. She performed a feasibility study for the application of recombinant-DNA techniques in pig breeding. She also organized a study tour for a small group of scientists to 22 scientific and policy institutions in the USA with relation to application of biotechnology in animal breeding strategies. In 1991 she started working at the State Institute of Quality Control of Agricultural Products (RIKILT), Department of Toxicology, that was headed by Harry Kuiper, in the group that was recently started on the subject of the safety of genetically modified organisms (GMOs). In the current RIKILT Institute of Food Safety she is now senior scientist and coordinator of the Novel Foods group within the Cluster Microbiology and Novel Foods. Since her start at RIKILT she has been participant in and project leader of a number of national and international projects and performs national and international advisory tasks in 1) the area of the safety assessment of novel foods, including GMOs, and 2) the area of detection and traceability of GMOs, and other specialty crops, in food and feed production chains. With relation to the first subject she has been member of the Dutch Health Council from 2000 till 2005 and is currently member of the Dutch committee on the Safety Assessment of Novel Foods and of the Working Group Food and Feed Safety of the European Food Safety Authority. She has furthermore participated in the major projects in this area Entransfood, GMOCARE and NOFORISK, as well as in the UK programme on the Safety Assessment of GMOs of the British Food Safety Authority, and currently participates in the European integrated project Safefoods. With relation to the subject of detection and traceability (of GMOs), she participates in the European Network of GMO Laboratories, and in the current European integrated projects Co-Extra and TRACE, as well as in a number of national advisory committees. She has performed a number of other side activities, such as participation in the Wageningen UR Ethical Committee from 2000 till 2005, referee for different scientific journals, member of the editorial board of the Dutch food magazine Voeding Nu, guest lecturer at a.o. Wageningen University, Amsterdam Higher Professional Education, International Centre for Genetic Engineering and Biotechnology, during a number of different courses, and (co-)organizer of different training courses in the field of the safety of novel foods, including GMOs, as well as the detection and traceability of GMOs in the food and feed supply chains. The research for this thesis has taken place since 2000 within the framework of different national and EU projects.



Dankwoord

Dit proefschrift heeft een minder gebruikelijke voorgeschiedenis. Het is niet het eindproduct van een vast omlijnd promotieproject, maar het resultaat van onderzoek over een langere periode, waarbij het hier beschreven onderzoek tegelijkertijd maar een deel van mijn aktiviteiten op het RIKILT betrof. Juist in de tijd dat de eerste artikelen van dit proefschrift gepubliceerd werden, was ook de periode dat het RIKILT-onderzoek naar detectie en traceerbaarheid van GGO's breder gestalte kreeg, en het was soms jongleren met veel ringen in de lucht om te voorkomen dat zaken spaak liepen. Gelukkig is dat nu met de komst van een aantal heel goede onderzoekers en onderzoeksassistenten in een rustiger vaarwater terechtgekomen. Gevolg van deze langere 'aanloopperiode' tot het proefschrift is wel dat de lijst van mensen die op enig moment mee hebben gewerkt aan het onderzoek dat hier beschreven is, lang is. Ik wil graag allen hier hartelijk bedanken, het proefschrift is de weerslag van de samenwerking met een brede groep mensen, waarvan ik er hieronder slechts een aantal met naam kan noemen, maar mijn dank gaat ook zeker uit naar andere (RIKILT-)collega's, die mij soms alleen al verder hebben geholpen door een leuk en/of nuttig gesprek bij de koffie, het kopieerapparaat of in de zo belangrijke wandelgangen. Maar toch een aantal met name:

De promotoren

Ivonne en Michael, we hebben maar weinig bijeenkomsten (nodig) gehad, maar de bijeenkomsten die we bij Michael hebben gehad, heb ik als zeer stimulerend ervaren. Ik ben ook erg blij met het laatste gezamenlijke manuscript dat daar uit voortgevloeid is. Ivonne, speciaal dank voor je doortastend optreden om zaken vlot te regelen en je duidelijke opmerkingen bij manuscripten.

De co-promotoren

Harry en Jaap, jullie ben ik het meest dank verschuldigd, jullie zijn van het begin af aan de drijvende krachten geweest achter het idee van een proefschrift en jullie hebben het idee, en mij, niet laten vallen in periodes dat het allemaal niet zo hard opschoot. Verder hebben jullie altijd met veel interesse commentaar geleverd op manuscripten, terwijl ik weet hoe vol jullie agenda's meestal zijn.

De grondleggers van de ideeën

Harry, jij bent degene die de visie had dat de komst van eerst de bestraalde producten en daarna de GGO's, i.e. complexe producten, een nieuwe dimensie zouden geven aan de toxicologie, jouw vakgebied. Hierdoor heeft het RIKILT baanbrekend onderzoek kunnen doen op dit terrein en is de RIKILT-naam voorgoed verbonden aan de toxicologie van complexe producten in het algemeen en GGO's in het bijzonder. Dat jij nu al jaren het Europese EFSA-panel op het gebied van de veiligheidsbeoordeling GGO's voorzit onderstreept je vooruitziende blik van destijds.

Jaap en Henk, jullie hebben de visie gehad dat de micro-arraytechnologie revolutionaire veranderingen teweeg zou brengen in het onderzoek van levende organismen. Ik kan me nog

herinneren dat wij buiten bij de RIKILT-kantine zaten en jullie voorstelden om van de 'differential display'-methode over te stappen naar de micro-arraytechnologie om genexpressieprofielen te maken. Het bleek een gouden greep, voor dit proefschrift, maar ook voor het RIKILT in veel bredere zin. Jaap, ik vind het heel mooi dat jij je carrière nu als hoogleraar zult vervolgen, de vakgroep zal er wel bij varen.

De onderzoeksmanagers

Robert, jij hebt mij in de gelegenheid gesteld om het tomatenonderzoek op een goede manier analytisch af te ronden, zodat het gepubliceerd kon worden in twee extra publicaties. Hierdoor was het mogelijk om een samenhangend proefschrift te schrijven. Belangrijk is ook geweest dat je me op een gegeven moment, in drukke tijden, er op hebt gewezen dat ik ook op dit terrein verplichtingen was aangegaan. Voor mij heeft dat het mogelijk gemaakt om op zeker moment mijn prioriteiten zo te stellen dat het proefschrift ook daadwerkelijk is afgerond. Henk, jij hebt mij ook als clusterleider altijd de ruimte gegeven om, waar nodig, keuzes te maken om dit proefschrift tot een goed einde te kunnen brengen.

De paranimfen

Angeline, jij was al een vriendin voordat je indertijd bij het RIKILT kwam werken en mee ging doen aan het onderzoek met de 'differential display'-methode, die je al in korte tijd wist op te zetten. Ik vind het heel bijzonder dat we nu al zoveel jaren samenwerken en dat je nu mijn paranimf wilt zijn. Lies, jij bent ook van het begin af aan betrokken geweest bij het onderzoek dat tot dit proefschrift heeft geleid. Inmiddels heb je een prachtige woonwinkel in Zutphen, maar ik ben erg blij dat je nu als vriendin mijn paranimf wilt zijn.

De vraagbaken

Maryvon en Gijs, al vele jaren werken we nu samen en altijd als ik vragen heb op het terrein van m.n. regelgeving en toxicologie (Maryvon) en GGO's in heel brede zin, en heel veel andere zaken (Gijs), dan kan ik bij jullie terecht voor de juiste informatie. Daarnaast is het altijd heel gezellig om met je op de kamer te zitten, Maryvon, en erg goed voor de nodige relativering hier en daar. Voorlopig ga ik ervanuit dat we straks ook samen in de nieuwbouw zullen zitten.

Nog meer vraagbaken

Irene, dank je wel voor je hulp bij het klaarmaken van het proefschriftmanuscript, en bij al die andere zaken, waar je altijd uitkomst biedt. Ruud en Jan, dank voor jullie assistentie bij alle computerzaken, zonder de vpn-verbinding was het niet gelukt. Jeannette, dank je wel voor je adviezen om een en ander esthetisch verantwoord te presenteren. En tot slot ook dank aan de mensen die alles ten aanzien van de financiën altijd soepel in goede banen leiden: Mart, Arjen, Nico, Willem, Marion en Jelte.

De collega's

Angeline, Bonnie, Henriek, Ingrid, Jeroen, Lieke, Marleen, Peter en Theo: sommigen van jullie zijn co-auteur van dit proefschrift, maar ik wil jullie graag allen danken voor de prettige sfeer en de samenwerking binnen de groep bij het onderzoek naar GGO's en andere novel foods. Ook andere collega's, als Angela, El, Peter (H.), Ad, Hans, Richard, Hakan, Jenneke wil ik graag noemen, zij hebben een uiteenlopende, maar waardevolle inbreng in het hier beschreven onderzoek gehad.

I also wish to thank the people from abroad that have participated in our research and have shared their views and insights with us: Edna (Israel), Aniol (Spain), Ahmed (Egypt), Barbara (Italy), Stine (Denmark), Ming (China), Katja (Slovenia) and Xiaotong (China), and others who worked on other subjects within our group but have helped us with their questions and discussions.

De uitgevers

Jessica, Inge en Mike, hartelijk dank voor de prettige samenwerking, met als resultaat dit mooie proefschrift volgens Jessica's ontwerp!

De achterban

Dit proefschrift is opgedragen aan mijn ouders en mijn twee zussen, het fijne gezin waar de basis voor mijn mogelijkheden en interesses is gelegd. Ik had graag gewild dat jullie er alle vier bij hadden kunnen zijn, en in gedachten zal dat ook zo zijn. Mijn dank gaat ook uit naar mijn schoonouders die eerst samen, later ma alleen, kwamen oppassen, dat vinden de kinderen, en wij, altijd erg gezellig. En ook vanuit de vriendenkring heb ik veel belangstelling ondervonden, met toch wel een speciale vermelding voor Piet en Brigitta ☺.

En dan de coachploeg, onmisbaar: Flip, spin docter en voor de day-to-day-logistiek, je weet overal een mouw aan te passen; Gerard, voor de ICT, voor het minste computerprobleempje ben je altijd bereid om er een (huiswerk)middag mee aan de slag te gaan; Sieta, lay-out, voor de vele, creatieve adviezen ten aanzien van de mooie vormgeving (ik heb ze niet allemáál kunnen opvolgen), Aiso, duidelijk vooral gericht op de randvoorwaarden voor goede prestaties, op tijd ontspanning en zo, en ten slotte Jente voor de PR en communicatie naar derden ('als je nu wat tegen haar wilt zeggen, moet je vlak bij gaan staan en het dan nog eens zeggen'). Met zulke coaches heb je geen tegenstander nodig ☺!

Dank jullie wel!